

NGA-subduction global ground motion models with regional adjustment factors

Earthquake Spectra

1–38

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/87552930211034889

journals.sagepub.com/home/eqs

Grace A Parker, M.EERI¹ , Jonathan P Stewart, M.EERI² , David M Boore¹, Gail M Atkinson, M.EERI³, and Behzad Hassani⁴

Abstract

We develop semi-empirical ground motion models (GMMs) for peak ground acceleration, peak ground velocity, and 5%-damped pseudo-spectral accelerations for periods from 0.01 to 10 s, for the median orientation-independent horizontal component of subduction earthquake ground motion. The GMMs are applicable to interface and intraslab subduction earthquakes in Japan, Taiwan, Mexico, Central America, South America, Alaska, the Aleutian Islands, and Cascadia. The GMMs are developed using a combination of data inspection, data regression with respect to physics-informed functions, ground-motion simulations, and geometrical constraints for certain model components. The GMMs capture observed differences in source and path effects for interface and intraslab events, conditioned on moment magnitude, rupture distance, and hypocentral depth. Site effect and aleatory variability models are shared between event types. Regionalized GMM components include the model constant (that controls ground motion amplitude), anelastic attenuation, magnitude-scaling break point, linear site response, and sediment depth terms. We develop models for the aleatory between-event variability (τ), within-event variability (ϕ), single-station within-event variability (ϕ_{SS}), and site-to-site variability (ϕ_{S2S}). Ergodic analyses should use the median GMM and aleatory variability computed using the between-event and within-event variability models. An analysis incorporating non-ergodic site response should use the median GMM at the reference shear-wave velocity condition, a site-specific site response model, and aleatory variability computed using the between-event and single-station within-event variability models. Epistemic uncertainty in the median model is represented by standard deviations on the regional model constants, which facilitates scaled-backbone representations of model uncertainty in hazard analyses.

¹Earthquake Science Center, U.S. Geological Survey, Moffett Field, CA, USA

²University of California Los Angeles, Los Angeles, CA, USA

³Western University, London, ON, Canada

⁴BC Hydro, Burnaby, BC, Canada

Corresponding author:

Grace A Parker, Earthquake Science Center, U.S. Geological Survey, P.O. Box 158, Moffett Field, CA 94035, USA.

Email: gparker@usgs.gov

Keywords

Subduction zone, ground motion model, earthquake hazard, seismic hazard, earthquake ground motion characterization

Date received: 16 November 2020; accepted: 21 June 2021

Introduction

Subduction zones produce interface earthquakes at the boundary between subducting and overriding tectonic plates, which are typically reverse in mechanism. They also produce intraslab earthquakes within the subducting plate, which are variable in mechanism, with normal faulting being predominant at intermediate depths (Astiz et al., 1988; Stern, 2002). Shaking from subduction zone earthquakes can affect many highly populated regions globally, including the Pacific Northwest (PNW) region of the United States and Canada, and as such accurate characterization of the associated seismic hazards is critical.

Early studies of empirical ground motions from subduction zones did not investigate regional differences in ground motions nor did they distinguish between event types, or did so only through adjustment of a constant term (Atkinson, 1997; Crouse et al., 1988; Youngs et al., 1997, 1988). For example, Youngs et al. (1997) presented an ergodic ground motion model (GMM) developed using a mixed-effects regression of 350 recordings from Alaska, the Cascadia Subduction Zone of the Pacific Northwest, Japan, Mexico, Peru, and the Solomon Islands. This model was used for earthquakes in the Cascadia region in the Frankel et al. (1996) U.S. Geological Survey (USGS) national seismic hazard model (NSHM).

As the size and reliability of ground-motion databases increased, this assumption—that ground motion should behave similarly across global regions—was disproven (e.g., Anderson and Brune, 1999). Atkinson and Boore (2003) used 1200 recordings from global events to develop a subduction GMM with significant regional differences. For example, ground-motion amplitudes in Cascadia were found to be reduced at short oscillator periods by up to a factor of 2 relative to those in Japan for the same event type, magnitude, source-to-site distance, and site class. They also found that intraslab events produce larger ground motions than interface events within 100 km of the fault, but decay faster with distance, leading to smaller motions at larger distances.

The recent BC Hydro model (Abrahamson et al., 2016) was developed using a dataset consisting of 9946 horizontal time series pairs from 292 earthquakes, although the model did not directly consider the 2010 M8.8 Maule, Chile nor the 2011 M9.1 Tohoku, Japan interface earthquakes. The analyses of Abrahamson et al. (2016) found that the same magnitude-scaling slope could be used for interface and intraslab events, but different distance-scaling slopes were needed in the forearc region, between the subduction trench and the volcanic arc, and the backarc region on the far side of the volcanic arc opposite the trench. This GMM was formulated as a global model, with a range of epistemic uncertainty in the constant term that can be used to represent regional variations in ground-motion amplitudes, but without regionalized anelastic attenuation or linear site terms. Abrahamson et al. (2018) produced an interim update of the BC Hydro GMM using an early version of the Next Generation Attenuation-Subduction (NGA-Sub) database for consideration in the 2018 USGS NSHM (Powers et al., 2021). The magnitude-dependent geometrical spreading, the large magnitude scaling, the magnitude-dependent finite-fault effect, and the quadratic magnitude terms were not modified from BC Hydro, but

regionalization was introduced in the anelastic attenuation, linear site, and model constant terms to produce a model that could be applied for Cascadia.

Due to the observed global differences in ground motions, regional GMMs for data-rich subduction regions have been developed, such as in Japan (Kanno et al., 2006; Si and Midorikawa, 1999; Zhao et al., 2006, 2016a, 2016b) and Taiwan (Chao et al., 2020; Lin and Lee, 2008; Phung et al., 2020). In regions with limited available data, such as Cascadia, simulations have been used to investigate subduction interface ground motions (Atkinson and Macias, 2009; Frankel et al., 2018; Gregor et al., 2002; Wirth et al., 2018a).

The NGA-Sub project began in 2014 with the goal of producing a uniformly processed ground-motion database and a suite of improved subduction zone GMMs to represent epistemic uncertainties in predicted median ground motions. This project encompassed global subduction zones, including those in Japan, Taiwan, Alaska (including the Aleutian Islands), New Zealand, Mexico, Central America, and South America. The database includes over 71,000 three-component ground motion records from interface and intraslab subduction events (Kishida et al., 2020; Mazzoni et al., this issue) and is the largest ever developed for an NGA project.

Here, we develop semi-empirical global GMMs with regional adjustment factors for interface and intraslab subduction events using the NGA-Sub ground-motion database. The models provide median ground motion, aleatory variability, and epistemic uncertainty for peak ground acceleration (PGA), peak ground velocity (PGV), and 5%-damped pseudo-spectral acceleration (PSA) for 26 periods between 0.01 and 10 s, for the orientation-independent horizontal component (RotD50; Boore, 2010). The reference site condition of the model is a time-averaged shear wave velocity in the upper 30 m (V_{S30}) of 760 m/s; information on applying a site response model for other velocity conditions is presented herein, but the details of site response model development are given in a companion paper (Parker and Stewart, 2021). Coded versions of the model in Excel, MatLab, R, and Python are available from Mazzoni et al. (2020).

The NGA-Sub project and the GMM presented here improve upon prior subduction ground motion modeling efforts by utilizing a much larger dataset; considering regionalization in the constants that control ground-motion amplitude, anelastic attenuation, magnitude-scaling, linear site response, and sediment depth terms; treating the amplitude and distance-, magnitude-, and depth-scaling terms differently between interface and intraslab event types; including the dependencies of within-event aleatory variability on rupture distance and site condition; and distinguishing the single-station and site-to-site variances from the total within-event variance. This article is adapted from Parker et al. (2020), which presents the model and its derivation in greater detail.

Median model functional form

This section presents the functions for the global and regional median interface and intraslab models. Subsequent sections describe data selection, model development, and the aleatory variability and epistemic uncertainty models. The site response functions are included here for completeness, but the development and features of these models are presented by Parker and Stewart (2021). All model input parameters are defined in Table 1 and all model coefficients are defined in Tables 2 and 3. Period-independent coefficients are given in Tables 2 and 3 while period-dependent coefficients are given in Tables S1 to S4 in the Electronic Supplement to this article.

Table 1. Definitions of model input parameters and their units

Parameter	Definition	Units
R_{rup}	Rupture distance; closest three-dimensional distance to the rupture plane	km
M	Earthquake moment magnitude	unitless
Z_{hyp}	Hypocentral depth	km
\bar{Z}_{hyp}	Mean hypocentral depth given the depth to top of rupture (Z_{tor}). See Equations 25–26.	km
V_{s30}	Time-averaged shear wave velocity in the upper 30 m	m/s
PGA_r	Estimate of peak ground acceleration computed using the GMM at the $V_{s30} = 760$ m/s reference condition	g
$Z_{2.5}$	Depth to the 2.5 km/s shear wave velocity isosurface	m
T	PSA oscillator period	s

Due to differences in path and source-scaling attributes, separate GMMs are provided for interface and intraslab earthquakes. Both models share a common functional form (Equations 1–13) with the exception of the near-source saturation term (Equation 4), with some coefficients being the same for both event types, and others varying. Each median model has five terms: a constant (c_0) that controls the overall amplitude of the predicted ground motion; a path model (F_P) that describes the decay of ground motion with distance; a magnitude-scaling model (F_M); a source depth-scaling model (F_D); and a site-amplification model (F_S). These models are additive in natural log space to produce the natural log of PGA and PSA in units of gram and PGV in units of centimeters per second (μ_{lnY}):

$$\mu_{lnY} = c_0 + F_P + F_M + F_D + F_S \quad (1)$$

The path model (Equation 2) incorporates magnitude-dependent geometrical spreading and anelastic attenuation. Near-source saturation is incorporated via parameter h (Equation 4), which is combined with site-to-source rupture distance (R_{rup} ; Equation 3). The near-source saturation term h depends on event type.

$$F_P = c_1 \ln R + b_4 \mathbf{M} \ln(R/R_{ref}) + a_0 R \quad (2)$$

$$R = \sqrt{R_{rup}^2 + h^2} \quad (3a)$$

$$R_{ref} = \sqrt{1 + h^2} \quad (3b)$$

$$h = 10^{-0.82 + 0.252\mathbf{M}} \text{ (interface events)} \quad (4a)$$

$$h = \begin{cases} 10^{\left(\frac{1.050}{m_c - 4}\right)(\mathbf{M} - m_c) + 1.544} & \mathbf{M} \leq m_c \text{ (intraslab events)} \\ 35 \text{ km} & \mathbf{M} > m_c \end{cases} \quad (4b)$$

Magnitude-scaling is described using a piecewise function with parabolic and linear segments, transitioning at a corner magnitude m_c (Table 3):

$$F_M = \begin{cases} c_4(\mathbf{M} - m_c) + c_5(\mathbf{M} - m_c)^2 & \mathbf{M} \leq m_c \\ c_6(\mathbf{M} - m_c) & \mathbf{M} > m_c \end{cases} \quad (5)$$

Table 2. Coefficient definitions, units, and values

Median models				Aleatory and epistemic models					
Coefficient	Model component	Units	Global or regional	Value	Coefficient	Model component	Units	Global or regional	Value
c_1	Magnitude-independent geometrical spreading		Global	Tables S1 and S2	ϕ_1^2	ϕ distance-scaling		Global	Table S3
b_4	Magnitude-dependent geometrical spreading		Global	Tables S1 and S2	ϕ_2^2	ϕ distance-scaling		Global	Table S3
d_0	Anelastic attenuation		Regional	Tables S1 and S2	R_1	ϕ distance-scaling corner	km	Global	200
m_c	Magnitude breakpoint		Regional	Table 3	R_2	ϕ distance-scaling corner	km	Global	500
c_4	Small-magnitude linear scaling		Global	Tables S1 and S2	ϕ_3^2	ϕ V_{530} -scaling		Global	Table S3
c_5	Small-magnitude parabolic scaling		Global	Tables S1 and S2	V_1	ϕ V_{530} -scaling corner	m/s	Global	200
c_6	Large magnitude scaling		Global	Tables S1 and S2	V_2	ϕ V_{530} -scaling corner	m/s	Global	500
M	Source-depth scaling slope		Global	Tables S1 and S2	a_1	ϕ_{525} -scaling slope		Global	Table S3
D	Source-depth scaling cap value		Global	Tables S1 and S2	a_2	ϕ_{55} -scaling slope		Global	Table S3
d_{b1}	Source-depth scaling corner	km	Global	20	V_M	ϕ_{525} and ϕ_{55} model reference V_{530}	m/s	Global	Table S3
d_{b2}	Source-depth scaling corner	km	Global	67	V_3	ϕ_{525} and ϕ_{55} V_{530} -scaling corner	m/s	Global	200
s_1	Low-velocity V_{530} -scaling slope		Regional	Tables S1 and S2	V_4	ϕ_{525} and ϕ_{55} V_{530} -scaling corner	m/s	Global	800
s_2	Mid-velocity V_{530} -scaling slope		Regional	Tables S1 and S2	R_3	ϕ_{525} and ϕ_{55} distance-scaling corner	km	Global	200
V_1	V_{530} -scaling corner	m/s	Global	Tables S1 and S2	R_4	ϕ_{525} and ϕ_{55} distance-scaling corner	km	Global	500
V_2	V_{530} -scaling corner	m/s	Global	Tables S1 and S2	R_5	ϕ_{55} distance-scaling corner	km	Global	500
V_{ref}	V_{530} -scaling reference value	m/s	Global	760	R_6	ϕ_{55} distance-scaling corner	km	Global	800
f_1	Value of nonlinear site response for $PGA_i < 0.05$ g		Global	0	σ_{ϵ_i}	Epistemic uncertainty		Regional	Table S4
f_3	Transition shaking intensity for nonlinear site effects begin	g	Global	0.05	σ_{ϵ_2}	Epistemic uncertainty		Regional	Table S4

(continued)

Table 3. Regional saturation magnitudes for interface events computed using seismogenic fault width (Campbell, 2020) and for intraslab events computed using slab thickness (Ji and Archuleta, 2018)

Region	Interface m_c (Campbell 2020)	Intraslab m_c (Ji and Archuleta 2018)
Global	7.90	7.60
Alaska	8.60	7.20
Aleutian Islands	8.00	7.98
Cascadia	7.70	7.20
Northern Central America and Mexico	7.40 ^a	7.40
Southern Central America and Mexico	7.40 ^a	7.60
Japan—Kuril-Kamchatka Trench (Pacific Plate)	8.50	7.65
Japan—Nankai-Ryukyu Trench (Philippine Sea Plate)	7.70	7.55
Northern South America	8.50	7.30
Southern South America	8.60	7.25
Taiwan	7.10	7.70

^aFor central America and Mexico, the interface m_c is not varied for northern and southern regions, but instead is taken as the average for the whole margin.

Source-depth scaling is described with a piecewise-linear function of hypocentral depth, with three segments and two corner depths d_{b1} and d_{b2} :

$$F_D = \begin{cases} m (d_{b1} - d_{b2}) + d & Z_{hyp} < d_{b1} \\ m (Z_{hyp} - d_{b2}) + d & d_{b1} < Z_{hyp} \leq d_{b2} \\ d & Z_{hyp} > d_{b2} \end{cases} \quad (6)$$

where $d_{b1} = 20$ km and $d_{b2} = 67$ km for intraslab events. There is no source-depth scaling for interface events; that is, $F_D = 0$. The model is conditioned on the hypocentral depth (Z_{hyp}), but that can be replaced with a mean hypocentral depth (\bar{Z}_{hyp}) that depends on the depth to top of rupture (Z_{tor}), fault width (W), and fault dip angle (see the section “Source-Depth Scaling”).

The ergodic site response model, F_S , has three components (Parker and Stewart, 2021): (1) a linear term, F_{lin} , that represents the site amplification at small strains; (2) a nonlinear term, F_{nl} , that accounts for attenuation of high-frequency components of ground motion due to soil damping that occurs under strong shaking conditions at soil sites; and (3) a basin-depth term, F_b , that represents site response effects related to sediment depth. The three terms are summed in natural logarithmic space:

$$F_S = F_{lin} + F_{nl} + F_b \quad (7)$$

The linear term is piecewise-linear with three segments in V_{S30} space. The natural log of site amplification scales linearly with the natural log of V_{S30} between corner velocities V_1 and V_2 , and goes through zero at $V_{ref} = 760$ m/s (Table 2; Equation 8). Data from Taiwan and Japan show a break in slope ($s_1 \neq s_2$) at V_1 , which is similar to prior observations in Japan (Campbell and Bozorgnia, 2014) and central and eastern North America (Hassani and Atkinson, 2018; Parker et al., 2019).

$$F_{lin} = \begin{cases} s_1 \ln\left(\frac{V_{S30}}{V_1}\right) + s_2 \ln\left(\frac{V_1}{V_{ref}}\right) & V_{S30} \leq V_1 \\ s_2 \ln\left(\frac{V_{S30}}{V_{ref}}\right) & V_1 < V_{S30} \leq V_2 \\ s_2 \ln\left(\frac{V_2}{V_{ref}}\right) & V_{S30} > V_2 \end{cases} \quad (8)$$

The nonlinear term has the same functional form as the NGA-West2 Seyhan and Stewart (2014) model for shallow earthquake in active tectonic regions:

$$F_{nl} = f_1 + f_2 \ln\left(\frac{PGA_r + f_3}{f_3}\right) \quad (9)$$

where $f_1 = 0$, which means that the effect of nonlinearity disappears as the PGA at the reference velocity condition (PGA_r) goes to zero; $f_3 = 0.05$ g for all periods; and f_2 is given as (Chiou and Youngs, 2008):

$$f_2 = f_4 [\exp\{f_5(\min(V_{S30}, 760) - 200)\} - \exp\{f_5(760 - 200)\}] \quad (10)$$

The basin depth term (F_b) for a particular intensity measure (IM) depends on region, and differential depth $\delta Z_{2.5}$. Region options are Japan and Cascadia. Application of these models requires a measurement or estimate of the depth to the 2.5 km/s shear wave velocity horizon ($Z_{2.5}$). For the four specified regions, the basin depth function is:

$$F_b = \begin{cases} e_1 & \delta Z_{2.5} \leq \frac{e_1}{e_3} \\ e_3 \delta Z_{2.5} & \frac{e_1}{e_3} < \delta Z_{2.5} < \frac{e_2}{e_3} \\ e_2 & \delta Z_{2.5} \geq \frac{e_2}{e_3} \end{cases} \quad (11)$$

where e_1 , e_2 , and e_3 are dimensionless region- and period-specific model coefficients. Differential depth is defined as:

$$\delta Z_{2.5} = \ln(Z_{2.5}) - \ln(\mu_{Z_{2.5}}(V_{S30})) \quad (12)$$

where $Z_{2.5}$ is in units of meters and the centering depth, $\mu_{Z_{2.5}}$ (also in meters), is conditioned on V_{S30} using the functional form from Nweke et al. (2020), with separate coefficients for Cascadia and Japan:

$$\ln(\mu_{Z_{2.5}}) = \ln(10) \times \theta_1 \left[1 + \operatorname{erf}\left(\frac{\log_{10}(V_{S30}) - \log_{10}(\nu_\mu)}{\nu_\sigma \sqrt{2}}\right) \right] + \ln(10) \times \theta_0 \quad (13)$$

Coefficients c_0 , a_0 , m_c , s_1 , s_2 , and basin depth scaling coefficients (Equations 11–13) are regionalized (Table 2). The c_0 value changes the amplitude of ground motion, which likely reflects regionally variable source properties such as stress drop; a_0 controls the anelastic attenuation, or the per-cycle damping of seismic waves, related to the regional crustal quality factor (Q); m_c is related to the slab width for intraslab events (Ji and Archuleta, 2018) and to the seismogenic width for interface events (Campbell, 2020); the V_{S30} -scaling slopes s_1 and s_2 are controlled by the correlation between V_{S30} and deeper seismic

velocities, which can be regionally dependent. Finally, the basin depth scaling models are unique to the regional geological structures in Cascadia and Japan.

Data selection

The NGA-Sub database contains more than 71,000 three-component time series from 1883 earthquakes acquired from subduction zone regions around the world. The overall relational database combines an earthquake source database, an earthquake recording database, and a recording station database (Mazzoni et al., this issue). The relational database can be used to produce a single summary flatfile with one line per recording for use in the development of GMMs. The records used for model development were selected according to the following criteria:

1. Metadata necessary for model development are available in the NGA-Sub database (Ahdj et al., this issue; Contreras et al., this issue) such as \mathbf{M} , rupture distance (R_{rup}), hypocentral depth (Z_{hyp}), and V_{S30} ;
2. Earthquake classified with high confidence as being interface, intraslab, or in the lower double seismic zone of Japan (Suzuki et al., 1983);
3. Earthquake is a mainshock (Class 1; C1) rather than an aftershock (Class 2; C2) according to the Wooddell (2018) method 2 using an 80 km cutoff distance;
4. $R_{rup} \leq \min(R_{max}, 1000 \text{ km})$, where R_{max} is a maximum distance limit set based on seismic network properties (Contreras et al., this issue).
5. Sensor depth $\leq 2 \text{ m}$;
6. Interface events with hypocentral depths (Z_{hyp}) $\leq 40 \text{ km}$ and intraslab events with $Z_{hyp} \leq 200 \text{ km}$;
7. PSA at oscillator periods $T \leq T_{LU}$, where T_{LU} refers to the longest useable period based on the corner frequencies used to process the record;
8. Earthquake epicenter and the station are both located in the forearc region (boundaries defined in Contreras et al., this issue);
9. Earthquakes without multiple event flags; these are events for which there is no indication that more than one seismic event is contained in the ground motion record;
10. Earthquakes with source review flags = 0, 1, 2, or 4, which indicate earthquakes that underwent quality control checks and met metadata quality standards;
11. Records that capture the start of the P-wave (i.e. those without a late P-wave trigger flag);
12. After applying criteria 1–11, we only used records from events having at least three recordings.

The number of events and recordings used for model development varies as a function of period due to criterion 7, with a range of 3215–6374 records and 90–122 events for combined data from both event types. The interface records in the screened database are from events with $\mathbf{M}5$ -9.1, recorded at rupture distances of 15–1000 km, and the intraslab records are from events with $\mathbf{M}4$.5–8.3, recorded at rupture distances of 18–1000 km (Figure 1). The majority of interface records are from Japan and South America, with contributions from Alaska, Central America, Mexico, and Taiwan. The majority of intraslab records are from Japan and Taiwan, with contributions from Alaska, Cascadia, South America, Central America, and Mexico (Figure 1). We did not consider records from New Zealand as the source table for New Zealand was adopted directly from Van

Houtte et al. (2017) and did not contain C1-C2 classifications (criterion 3). Subduction events that have occurred since 2016 are not included in the database.

Development of the median model

Model development occurred in a series of steps during which we progressively constrained model components to capture meaningful trends in the data due to source, path, and site processes and to avoid trade-offs in model coefficients. We started by adjusting the ground motion metrics in our screened database to a reference site condition of $V_{S30} = 760$ m/s using the Seyhan and Stewart (2014; hereafter SS14) site response model. We then developed distance-scaling models, including near-source saturation, using a two-step approach similar to Joyner and Boore (1981, 1993, 1994). With a path model established, we fit the magnitude and source-depth scaling. Finally, we returned to iteratively adjust our global and regional constants and the SS14 site response model, including a basin depth term where appropriate (Japan and Cascadia; Parker and Stewart, 2021). The following subsections describe the path and source model development.

Path model

Near-source saturation. It is typical for ground-motion path models to express the decay of ground-motion intensity, both geometrical spreading and anelastic attenuation, with a distance metric R that combines the rupture distance (closest distance from a point to the source, R_{rup}) with a near-source saturation term, h , as in Equation 3a. The use of h in the expression for R causes ground motions to saturate at close rupture distances, where $R_{rup} < h$. This term—also sometimes called the finite-fault term or fictitious depth—is necessary due to the following two potential geometric effects (Rogers and Perkins, 1996; Yenier and Atkinson, 2014): (1) whereas path models are based on the distance to the closest part of the finite fault, other portions of the fault at greater distance also contribute to the observed ground-motion amplitudes; and (2) to the extent that ground motions are controlled by slip on the closest part of the fault, the slip at that location would, in isolation from other parts of the fault, correspond to a smaller seismic event than the full earthquake rupture; essentially, a nearby site can only see part of the fault.

Initially we considered using the subduction data to constrain the near-source saturation; however, due to the offshore or deep locations of typical subduction earthquakes and the lack of recording stations at close source-to-site distances (Figure 1), there are not enough data close to the source to constrain this feature. In addition, because models for h are magnitude-dependent, attempts to regress them from data are difficult because of trade-offs with other magnitude-dependencies in the data, such as magnitude scaling and magnitude-dependent geometrical spreading. Accordingly, we instead adopted an approach in which h for subduction zones is constrained jointly from empirical estimates from active tectonic regions at small-to-large magnitudes (Atkinson et al., 2016; Yenier and Atkinson, 2014) and simulations of moderate-to-large subduction interface events performed as part of this work. The simulations were performed with EXSIM, which is an open-source stochastic finite-source simulation algorithm (Assatourians and Atkinson, 2012; Boore, 2009; Motazedian and Atkinson, 2005).

Yenier and Atkinson (2014; YA14) used records from 11 shallow earthquakes in global active tectonic regions with $M \geq 6$ to fit distance-scaling functions to each event to estimate h , and produced a magnitude-dependent model for h in the form of Equation 4a

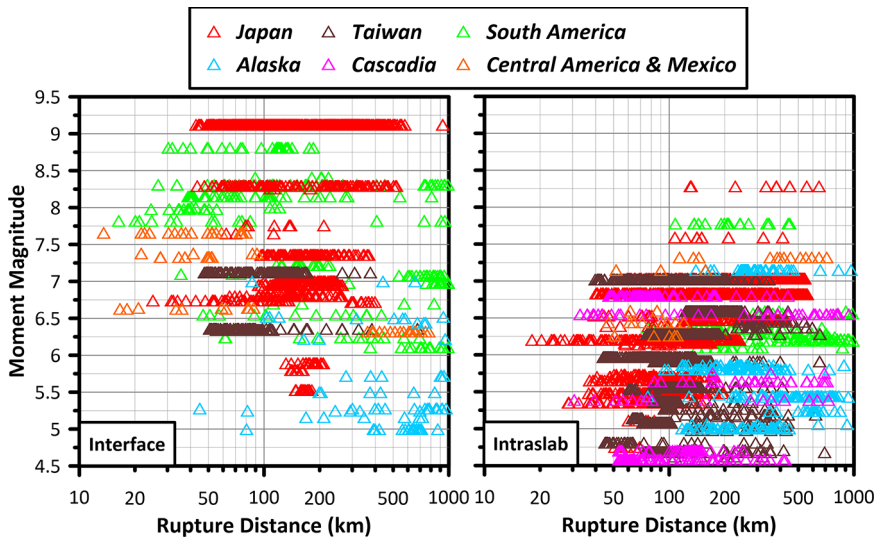


Figure 1. Magnitude–distance distribution of recordings from interface and intraslab events, color-coded by region.

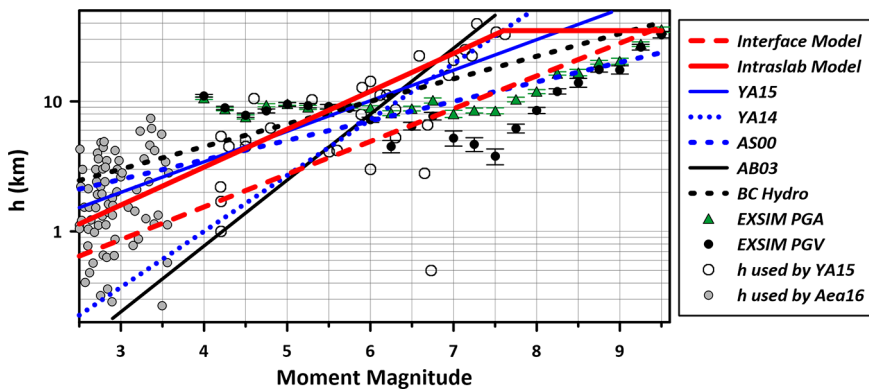


Figure 2. Comparison of recommended near-source saturation models for interface and intraslab earthquakes (Equation 4, shown in red) with models constrained using active tectonic region data (YA14, YA15, and Atkinson and Silva, 2000, labeled AS00), and models from published subduction zone GMMs (Abrahamson et al., 2016, labeled BC Hydro; and Atkinson and Boore, 2003, labeled AB03). Also shown are the empirical estimates of h used to constrain the YA15 model, empirical estimates of h from small induced events in the Geyser region of California (Atkinson et al., 2016, labeled Aea16), and estimates of h and their standard errors for PGA and PGV developed using EXSIM for interface earthquakes (this study).

(data and model shown in Figure 2). Yenier and Atkinson (2015a; YA15) examined the NGA-W2 database (Ancheta et al., 2014) to estimate the best-fit source parameters for each California earthquake by matching empirical and simulated response spectra. As part of this work, they considered h models by YA14 and the study by Atkinson and Silva (2000; shown in Figure 2), along with event-specific empirical estimates of h by YA14,

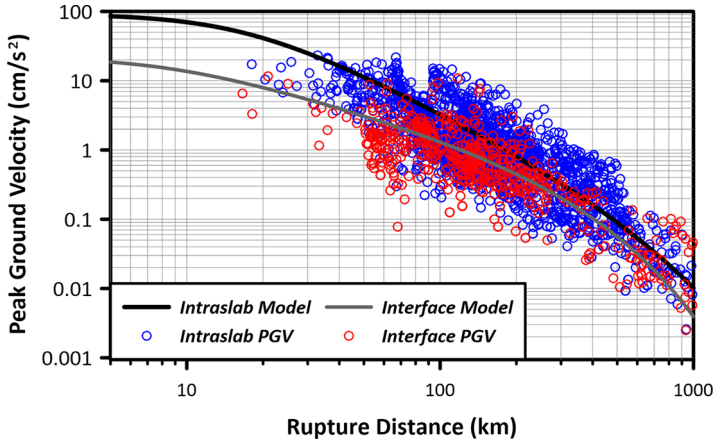


Figure 3. Comparison of the distance dependence in the global interface and intraslab PGV data corrected to $V_{S30} = 760$ m/s using Equation 7 from earthquakes with $M 6.0$ – 7.0 . Global model predictions shown for the average parameters of the binned data (intraslab: $M = 6.5$, $Z_{hyp} = 71$ km; interface: $M = 6.64$).

Boore et al. (2014a), and earthquakes in the Christchurch, New Zealand, sequence. They proposed a parameterization using Equation 4a (Figure 2). Finally, Atkinson et al. (2016) examined a number of small, induced events that are well recorded at short source-to-site distances from the Geysers region of California (Sharma et al., 2013) to better constrain near-source saturation effects for small magnitude earthquakes ($1.5 \leq M \leq 3.6$). They applied a similar method of fitting event-specific distance-scaling as was applied in the study by YA14. As shown in Figure 2, their results are consistent with the near-source saturation model of YA15 for global earthquakes.

For interface events, to constrain h at large magnitudes, we performed EXSIM simulations with $M = 3.75$ – 9.5 in 0.25-magnitude unit intervals, with five runs per magnitude. For each run, fault length and width were generated using the scaling relationships of Strasser et al. (2010); a fault dip between 15° and 28° was randomly assigned, consistent with estimates of dip for interface events in the database. Hypocenter locations were randomly sampled from a uniform distribution over the fault plane. Stress drop was taken as 150 bars. Ground motions were generated at various azimuths and for distances between 10 and 1000 km. See Parker et al. (2020) for additional details.

Simple magnitude-independent path models were fit to the simulated data in each magnitude bin (e.g. Equations 2–4a without the b_4 term). First, coefficients representing geometrical spreading and anelastic attenuation effects were fit to the simulated ground motions at $R_{rup} \geq 40$ km to avoid the influence of near-source saturation effects at closer distances. With the attenuation coefficients fixed, h was then fit using the simulated ground motions over the entire distance domain. The resulting best-fit h values for PGA and PGV vary with magnitude as shown in Figure 2 (similar trends were observed for other IMs). The resulting h values are similar to the YA14, YA15, and AS00 models for $M \approx 6$, but are larger and smaller than the models for events with $M < 5.5$ and $M > 6.5$, respectively (Figure 2).

Our near-source saturation model for interface events (Equation 4a) applies to all IMs. As shown in Figure 2 as a dashed red line, we constrained the model to have a similar slope to that of YA15 for $\mathbf{M} \leq 5.5$, but to follow trends from EXSIM simulations for $\mathbf{M} > 5.5$.

For intraslab events, we found that residuals computed using the interface near source saturation model of Equation 4a with the intraslab GMM indicate over-prediction of median ground motions by a factor of ~ 1.3 for large intraslab events ($\mathbf{M} > 6.5$) at short distances (≤ 75 km). This indicates that the interface h model saturates at rupture distances that are too small. Removing this bias required increasing h in the $6.5 \leq \mathbf{M} \leq 7$ range for intraslab events. Once such adjustments were made, however, very large values of h are given by the exponential function Equation 4a for larger values of \mathbf{M} , and we considered these to be non-physical if h is related to fault-plane dimensions. Having no data to constrain a maximum value of h for intraslab events in the $\mathbf{M} > 7$ region (Figure 1), we examined simulation results for intraslab events from the study by Ji and Archuleta (2018). We fit Equation 4b to the Ji and Archuleta (2018) response spectral values at 0.2 and 1.0 s PSA for an $\mathbf{M} = 8$ intraslab earthquakes in Japan using alternate fixed values of $h = 29$ km (value from YA15 for $\mathbf{M} = 8$), 35, 40, and 50 km. We adopt an upper limit of $h = 35$ km because it has the best agreement with simulation results at the closest distances. We enforced this upper limit at the regional corner magnitude (m_c) used in the magnitude-scaling model. We re-fit Equation 4b over the magnitude range of $4.0 \leq \mathbf{M} \leq m_c$ such that the resulting model predicts similar values of h as YA15 at $\mathbf{M} = 4$ and $h = 35$ km at m_c (Figure 2).

The regional corner magnitude m_c (as used in Equation 4b) in the function for h was chosen, in part, to produce only one corner in ground motion–magnitude space. Values of intraslab m_c are based on the seismogenic thickness of subducting slabs (Ji and Archuleta, 2018). As seismogenic thickness and m_c increase, the likelihood of a large fraction of the rupture occurring directly beneath a site, rather than most of the rupture being located some distance down-strike, also increases. In turn, this produces increased ground motions within the distance range controlled by saturation. This effect is provided by Equation 4b when m_c is increased because h is reduced for $\mathbf{M} < m_c$. Therefore, we argue that connecting h to m_c has physical justification.

Distance-scaling. The path model has two components that attenuate ground motion with increasing distance: a geometrical spreading term and an anelastic attenuation term. The geometrical spreading term represents the decay of energy as it moves from a point source along a spherical wave front. In an idealized homogeneous elastic half-space, the energy at any point on the radius of the sphere will decay as R^{-1} . Given that the ground motion is modeled in natural log units, we capture this behavior using $c_1 \ln R$ in Equation 2, where c_1 would be equal to -1.0 . However, heterogeneities in the earth produce scattering, reflections, refractions, and wave-type conversions. As a result, the empirical exponent (c_1 in Equation 2) is not -1.0 . Moreover, the transition from Fourier amplitude spectra to PSA introduces magnitude-dependence in this term, since the PSA at a single period is sensitive to a range of Fourier periods near the oscillator period (Hassani and Atkinson, 2018; Stafford et al., 2017; Yenier and Atkinson, 2014). The ratio between the characteristic dimension of the rupture and the distance from the rupture will also have an influence that is magnitude dependent. This can be represented as $(b_3 + b_4 \mathbf{M}) \ln R$; however, in Equation 2, we combine the b_3 term with c_1 . The anelastic attenuation term represents the per-cycle energy dissipation; it is a property of the material through which the seismic waves travel. This term controls curvature in the decay of natural log ground motion with the natural

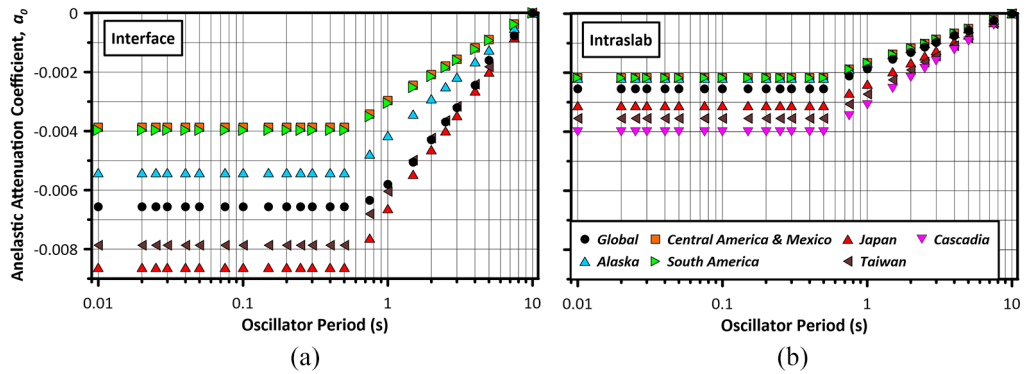


Figure 4. Anelastic attenuation coefficient, a_0 , as a function of oscillator period for (a) interface events and (b) intraslab events. Due to lack of Cascadia interface events, the global interface value is recommended. Note that in (a), South America overlays Central America and Mexico; and in (b), both South America and Central America and Mexico overlay Alaska.

log of rupture distance, which strongly influences the rate of attenuation at large distances (Figure 3).

To fit the path model independently from the source terms, we use a two-step regression for F_P in Equation 2 (similar to Joyner and Boore, 1981, 1993, 1994). The geometrical spreading coefficient derived from this process (c_1) was the same across all magnitudes, and a preliminary event term (η_E') was evaluated for each individual event and IM (η_E' acts as a constant added to Equation 2). These event terms, which have a trend with M , were considered in the subsequent derivation of source terms (see the section “Magnitude Scaling”). Originally, we adopted values of b_4 from the simulation-based GMM of Hassani and Atkinson (2018; HA18). However, we found that b_4 values of HA18 were too small to adequately capture the magnitude-dependent component of geometrical spreading observed in the data, especially for intraslab events. We set $b_4 = 0.1$ for both event types, which improved the model fit to data.

Despite the large size of the NGA-Sub database, it was not possible to constrain both the geometrical spreading and anelastic components of the path model simultaneously due to substantial trade-offs between these two model components. We address this by fitting c_1 to the subset of data with $R_{rup} \leq 100$ km to avoid the portion of the data with the most curvature due to anelastic attenuation at large distances. Data from intraslab events show steeper geometrical spreading in comparison to data from interface events, and as a result our GMMs have different values of c_1 for the two event types (Figure 3). This is consistent with the results of some previous studies (Atkinson and Boore, 2003; Abrahamson et al., 2016).

With the geometrical spreading coefficients fixed, we fit the anelastic attenuation coefficient, a_0 , as a random effect to produce global and regional values. We smoothed a_0 with respect to period and constrained it to go to zero at 10 s, as the per-cycle damping at long oscillator periods is negligible. The anelastic attenuation rate is slower for intraslab events than for interface events (Figure 4), indicating that although the intraslab data show more overall distance attenuation, there is less curvature in the data at large distances (Figure 3). For both event types relative to the global model, the anelastic attenuation is slower in South America and Alaska, and faster in Japan and Taiwan. For Cascadia, the attenuation

is faster than the global model for intraslab events, and for interface events we recommend adopting the global value of a_0 . Because of the coupling between anelastic attenuation and geometrical spreading terms (Equation 2), they should only be used together; mixing of values between regions or between GMMs is not recommended.

Source model

Magnitude scaling. Once the path model was set (Equations 2–4), we used preliminary event terms η_E' (event-specific mean residuals from the path model) to visualize trends in the data with respect to magnitude, which informed the formulation of the magnitude-scaling model. Event terms represent the average bias over all ground-motion recordings for one event relative to a particular model; following the first regression stage, magnitude dependence of η_E' is expected. To model this dependence, Equation 5 was fit with the parabolic term $c_5 = 0$, allowing the two linear slopes (c_4 , c_6) to be set by regression. The linear magnitude-scaling coefficients were treated as fixed effects, and the constant c_0 was treated as a random effect conditioned on region and NGA-Sub unique earthquake identifier. The coefficients were constrained to enforce $c_6 \leq c_4$, which ensures slowing (or no change) of the magnitude scaling for $M > m_c$. Finally, the parameter that controls the parabolic behavior of the model below the break point, c_5 , was fit to the event terms with all other coefficients fixed to their values from the first regression iteration (Figure 5). Values of m_c were constrained based on geometrical considerations specific to each subduction zone region (Campbell, 2020; Ji and Archuleta, 2018; Table 3). Global values of m_c were taken as weighted averages over the values for regions considered, where each region was weighted equally; in other words, Cascadia and Japan are given equal weight, so the Cascadia m_c value was given double the weight of each of the two m_c values from Japan.

Source-depth scaling. The source depth-scaling model was derived based on event terms computed using site-adjusted data and the source and path models described in previous sections (i.e. $\mu_{lnY} = c_0 + F_P + F_M + F_S$). Those event terms were examined for trends with hypocentral depth (Z_{hyp}). We select Z_{hyp} in lieu of depth to top of rupture due to greater certainty in estimates and a stronger correlation of hypocenter location with earthquake stress drop in past work (Baltay et al., 2019; Bilek and Lay, 1998, 1999). See additional details in Parker et al. (2020).

We did not observe statistically significant trends in interface event terms with Z_{hyp} (Figure 6). For intraslab events, we initially fit Equation 6 to the event terms using a non-linear least-squares regression with all parameters unconstrained. Based on these results, a single corner depth, $d_b = 67$ km was chosen for all periods. The regression was repeated with the corner depth constrained, and the slope m and coefficient d were fit and smoothed. The model slope goes to 0 at the lower end of the depth range populated with data, 20 km. The depth adjustment increases ground motions over the depth range 20–67 km, which we interpret as an increase of stress drop with depth as also applied in previous GMMs for active tectonic regions (Hassani and Atkinson, 2018; Yenier and Atkinson, 2015a), stable continental regions (Yenier and Atkinson, 2015b), and induced earthquakes (Novakovic et al., 2018). The depth-dependence goes to zero for $T > 2.0$ s.

We recognize that Z_{hyp} is cumbersome in probabilistic seismic hazard applications because it requires randomization of the hypocenter location on the fault, which involves an additional integral in the hazard calculation. Where possible, we recommend

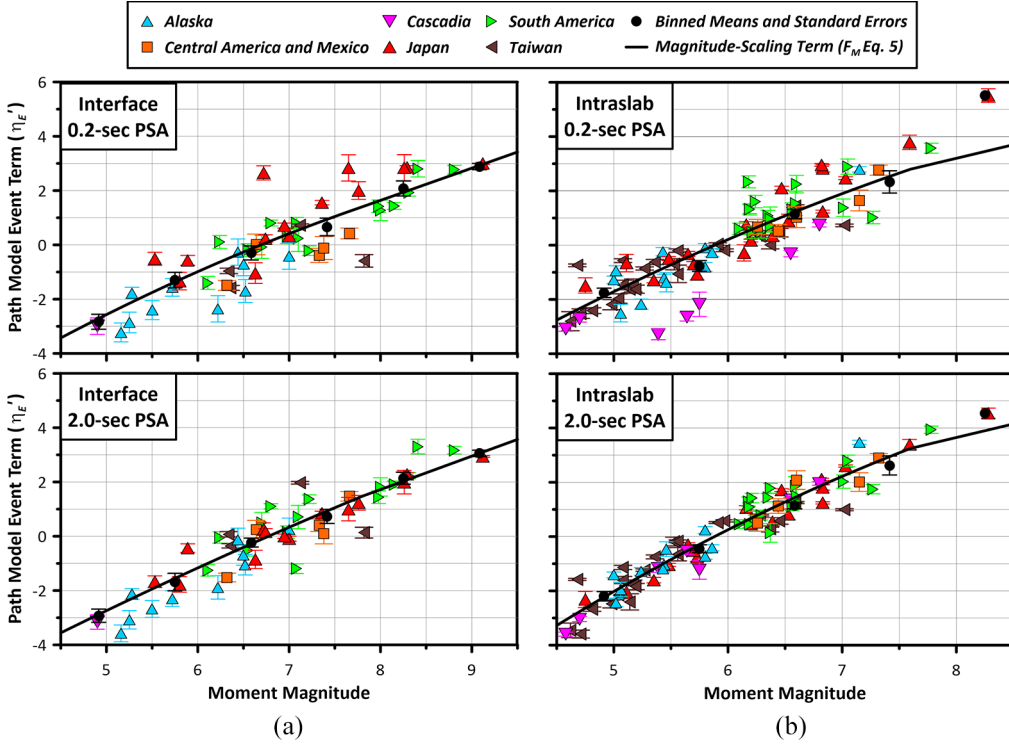


Figure 5. Global (a) interface and (b) intraslab magnitude-scaling models (F_M ; Equation 5) and path model event terms (η_E^i , in ln units, with 95% confidence intervals) as a function of M for 0.2 and 2.0 s PSA. For plotting purposes, the recommended global m_c values were used for the intraslab and interface model (Table 3).

considering hypocenter location in hazard analyses as it represents realistic variability in earthquake source processes. However, it is possible to replace the event-specific hypocenter depth with the mean depth expected for a given fault rupture surface, \bar{Z}_{hyp} (Equations 14–15). The value of \bar{Z}_{hyp} is fully determined once the location of a fault rupture rectangle is defined (length, width, strike, and dip angles, and Z_{TOR}). Accordingly, when \bar{Z}_{hyp} is used, an additional loop in the hazard integral is not required. We define Z_{dd} as the down-dip depth of the hypocenter given other source parameters, θ_W as a normalized version of this depth that is set empirically, and W as the fault rupture down-dip width (Figure 7). \bar{Z}_{hyp} can then be computed as:

$$Z_{dd} = \frac{Z_{hyp} - Z_{TOR}}{\sin(dip)} \quad (14a)$$

$$\theta_W = \frac{Z_{dd}}{W} \quad (14b)$$

$$\bar{Z}_{hyp} = Z_{TOR} + \theta_W W \sin(dip) \quad (15)$$

Data presented in the studies by Mai et al. (2005) and Melgar and Hayes (2019) can be used to estimate normalized down-dip depths (θ_W). Those studies used databases of 80

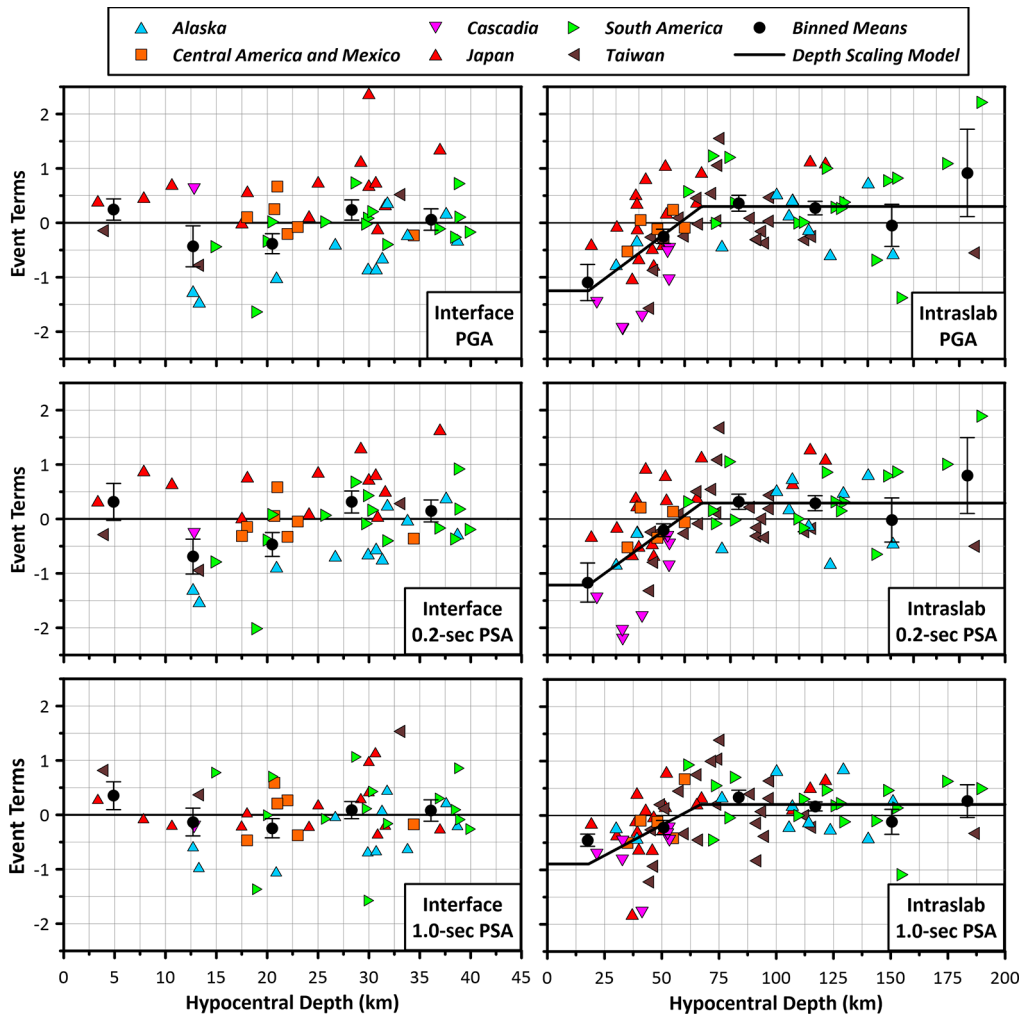


Figure 6. Variation of path model and magnitude-scaling event terms (In units) as a function of hypocentral depth for PGA, 0.2 and 1.0 s PSA for interface and intralab events. Best-fit depth scaling model (Equation 6) shown for intralab events. Straight line at 0 is a reference indicating no model bias.

and 192 finite-fault models, respectively, to examine hypocenter positions within fault planes. Mai et al. (2005) computed Z_{dd} for each earthquake, compared θ_W for strike-slip and dip-slip events, and looked for trends with magnitude. Melgar and Hayes (2019) investigated fault slip distributions modeled as spatially random fields, and examined correlation lengths. We performed similar analyses to those of Mai et al. (2005) using finite-fault models for 25 intralab events in the NGA-Sub database. As shown in Figure 7, we found that θ_W does not have a trend with magnitude, and therefore, we adopt the average value of 0.48. This indicates that on average the earthquakes are nucleating near the center of their rupture surfaces. The mean value from the study by Mai et al. (2005) is 0.43 for dip-slip events (not only subduction). The distribution of down-dip hypocenter positions from the study by Melgar and Hayes (2019) was found to be normally distributed with a

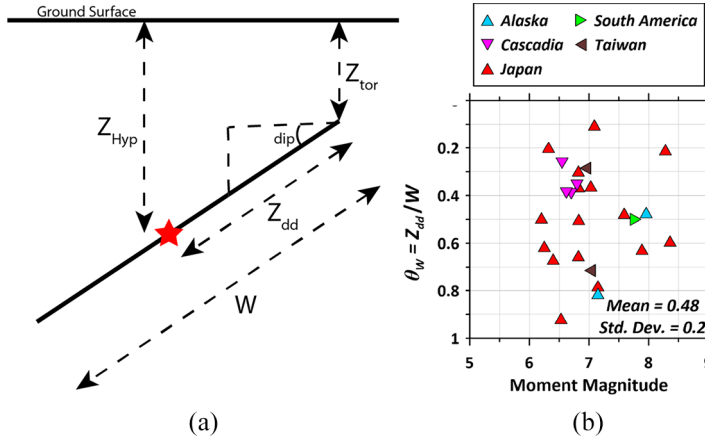


Figure 7. (a) Schematic showing fault plane geometry used to derive relationship between Z_{tor} and Z_{hyp} for forward use in hazard analyses. (b) Normalized down-dip hypocentral depth ($\theta_W = Z_{dd} / W$) for intraslab events in the NGA-Subduction database, color-coded by region.

mean equivalent to $\theta_W = 0.55$. Hence, the present result using the NGA-Sub data is intermediate between the two prior estimates.

When hypocentral depth Z_{hyp} is represented in Equation 6 by its mean value (\bar{Z}_{hyp}), no further modification to either the model coefficients or the between-event variability is required. However, to capture the effects of epistemic uncertainties in \bar{Z}_{hyp} , alternate realizations of $\theta_W =$ should be considered. We recommend a central logic tree branch for the mean of $\bar{\theta}_W = 0.48$ and additional branches for the mean \pm a representation of θ_W variability (the standard deviation is 0.2; Figure 7b).

Regional and global constants

The last step in model development for the reference-rock GMM was to determine global and regional model constants, c_0 (Equation 1), and final event terms, η_E , through a mixed-effects residuals analysis. Total residuals ($R_{ijk} = \ln(Y_{ijk}) - \mu_{\ln Y}$) for earthquake i , station j , and region k were computed using the GMM mean without a constant (i.e. $\mu_{\ln Y} = F_P + F_M + F_D + F_S$). Before our subduction-specific site response model was developed (Parker and Stewart, 2021), F_S was taken from the Seyhan and Stewart (2014) model to develop preliminary estimates of regional model constants. This constant fitting was iterated as the site response model evolved, and the final sets of global and regional constants were developed from, and are compatible with, the recommended subduction zone site response model.

The total residuals were partitioned into constants, $c_{0,k}$, for each region k , event terms $\eta_{E,i}$, for each earthquake i , and the remaining within-event residual for each record recorded at station j , δW_{ij} , using linear mixed effects in the R software environment (Bates et al., 2015; R Core Team, 2019):

$$R_{ijk} = c_{0,k} + \eta_{E,i} + \delta W_{ij} \quad (16)$$

Where adequate data in regions exist, $c_{0,k}$ was set from data. When data are sparse, constraints were applied in setting $c_{0,k}$.

In the case of intraslab events, $c_{0,k}$ was generally set from data. Through residuals analyses, we found that some regions (Alaska and South America) had large sub-regional variations in event terms. These variations corresponded to geography; the earthquakes in the Aleutian Islands on average have larger ground motions than earthquakes in mainland Alaska, and the earthquakes in the southern part of South America (i.e. Chile) have larger short-period ground motions than earthquakes in the northern section of the subduction zone (e.g. Ecuador and Colombia). In South America, this could be due to the subduction of different tectonic plates (e.g. the Nazca versus the Caribbean, or microplates within the South American plate, including the Altiplano and North Andes plates; Bird, 2003). Because of these sub-regional variations, we allow our intraslab model to have different sub-regional constants for these two regions, with spatial definitions consistent with the study by Ji and Archuleta (2018). In other regions (Central America and Mexico, Japan, and Taiwan), variations of the constant (c_0) between subregions was checked and found not to be required.

In the case of interface events, $c_{0,k}$ was set from data where possible (Aleutian Islands, Central America and Mexico, Japan—Pacific Plate, South America—southern region, and Taiwan—northwest region; region boundaries from Campbell, 2020). For the remaining regions with sparse data, constants were set with constraints. In particular, the Alaska constant is set such that the Aleutian median ground motion is matched for $\mathbf{M} < m_c$. The Japan—Philippine Sea Plate constant is set such that the Japan—Pacific Plate median ground motion is matched for $\mathbf{M} < m_c$. The South America—northern region constant is set such that the South America—southern region median ground motion is matched for $\mathbf{M} < m_c$. There are no data for the southwest subregion of Taiwan, so the constant set empirically for the northeast subregion is applied for the full region. For Cascadia, due to lack of data, the constant was set such that the global median ground motion is matched for $\mathbf{M} > m_c$. The match at larger magnitudes was applied because such events were considered to be more hazard-critical than events with $\mathbf{M} < m_c$.

We set the global constant to be compatible with the weighted average of regional median reference-rock ($V_{S30} = 760$ m/s) ground motions at the center of the data in the distance range ≤ 100 km to avoid effects of regional anelastic attenuation differences. For interface events, this was set at $\mathbf{M} = 7.0$ and $R_{rup} = 65$ km, and for intraslab events this was set at $\mathbf{M} = 6.0$, $R_{rup} = 75$ km. The weights were taken as proportional to the inverse of the constant parameter variances. The global constant was not set by mixed effects analysis for two reasons: (1) the constant is strongly correlated to m_c and should be set for compatibility with the global m_c ; and (2) such a constant would be an unweighted average of the regional constants, which would give too much weight to regions with small data populations.

There is more region-to-region variability in the constants for interface events than for intraslab events (Figure 8). However, given the correlation between regional constants and m_c , the increased spread of interface constants can be understood to largely reflect the larger range of m_c (about 1.5 magnitude units for interface, 0.8 for intraslab). Regional model comparisons in ground motion space are given in the section “Model behavior and regional comparison.”

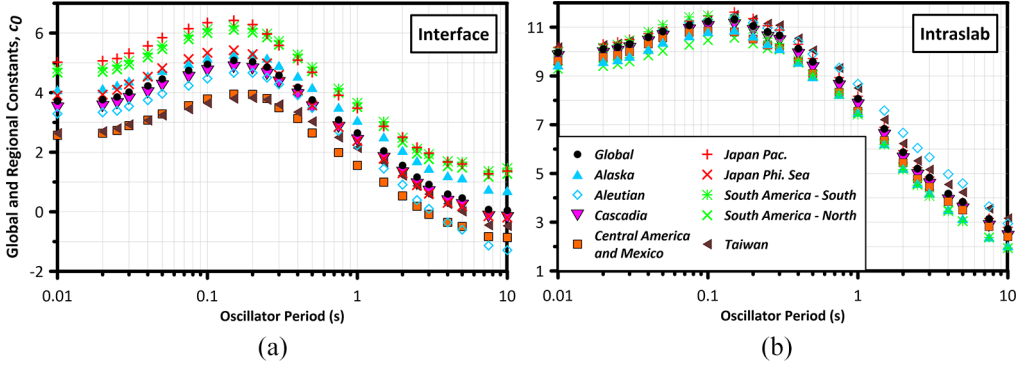


Figure 8. Comparison of global and regional constants (c_0 ; Equation 1) for **(a)** the interface model, and **(b)** the intraslab model (Japan Pac = Japan—Pacific Plate; Japan Phi. Sea = Japan—Philippine Sea Plate).

Variability and uncertainty models

Aleatory variability model

Functional forms. Aleatory variability represents the random variability of data relative to a model, and for ground motions is usually expressed using log-normal standard deviations (σ). The total standard deviation has between-event (τ) and within-event components (ϕ) that are assumed to be independent, and hence are related as,

$$\sigma = \sqrt{\tau^2 + \phi^2} \quad (17)$$

The τ model is period-dependent but independent of source, path, and site parameters, and thus is simply given as period-dependent coefficients (Table S3 in the Electronic Supplement).

The ϕ model is dependent on R_{rup} and V_{S30} (Equations 18–20). Model coefficients for the aleatory variability are given in Table 2 and Table S3 in the Electronic Supplement of this article.

$$\phi = \sqrt{\phi^2(R_{rup}) + \Delta Var(V_{S30})} \quad (18)$$

$$\phi^2(R_{rup}) = \begin{cases} \phi_1^2 & R_{rup} \leq R_1 \\ \frac{\phi_2^2 - \phi_1^2}{\ln(R_2/R_1)} \ln\left(\frac{R_{rup}}{R_1}\right) + \phi_1^2 & R_1 < R_{rup} < R_2 \\ \phi_2^2 & R_{rup} \geq R_2 \end{cases} \quad (19)$$

$$\Delta Var(V_{S30}) = \begin{cases} \phi_V^2 \left(\frac{\ln\left(\frac{R_2}{\max(R_1, \min(R_2, R_{rup}))}\right)}{\ln\left(\frac{R_2}{R_1}\right)} \right) & V_{S30} \leq V_1 \\ \phi_V^2 \left(\frac{\ln\left(\frac{V_2}{V_{S30}}\right)}{\ln\left(\frac{V_2}{V_1}\right)} \right) \left(\frac{\ln\left(\frac{R_2}{\max(R_1, \min(R_2, R_{rup}))}\right)}{\ln\left(\frac{R_2}{R_1}\right)} \right) & V_1 < V_{S30} < V_2 \\ 0 & V_{S30} \geq V_2 \end{cases} \quad (20)$$

Equation 19 describes a flat-ramp-flat, linear, piecewise relationship for variance as a function of \ln distance that has period-independent corner distances R_1 and R_2 (Table 2).

Equation 20 modifies the aleatory variance ϕ^2 in the initial flat and sloped portions ($R_{rup} < R_2$), but not at larger distances. This reflects the conditions under which site response nonlinearity is most prevalent. The reduction is maximized for sites with $V_{S30} \leq V_1 = 200$ m/s, is null for stiff sites ($V_{S30} \geq V_2 = 500$ m/s), and has a linear transition between V_1 and V_2 controlled by ϕ_V^2 .

For partially non-ergodic applications involving location-specific site response, we partitioned the within-event variability ϕ into site-to-site variability (ϕ_{S2S}) and within-event single-station variability (ϕ_{SS}) as (Al Atik et al., 2010):

$$\phi = \sqrt{\phi_{SS}^2 + \phi_{S2S}^2} \quad (21)$$

Models for ϕ_{S2S} and ϕ_{SS} are provided in Equations 22–23 and 24–26, respectively. The model components are formulated such that their sum is similar to the total within-event variance ϕ^2 .

$$\phi_{S2S}^2(V_{S30}, R_{rup}) = \phi_{S2S,0}^2 + \Delta Var_{S2S}(V_{S30}, R_{rup}) \quad (22)$$

$$\Delta Var_{S2S}(V_{S30}, R_{rup}) = \begin{cases} a_1 \ln\left(\frac{V_3}{V_M}\right) \left(\frac{\ln\left(\frac{R_4}{\max(R_3, \min(R_4, R_{rup}))}\right)}{\ln\left(\frac{R_4}{R_3}\right)}\right) & V_{S30} \leq V_3 \\ a_1 \ln\left(\frac{V_{S30}}{V_M}\right) \left(\frac{\ln\left(\frac{R_4}{\max(R_3, \min(R_4, R_{rup}))}\right)}{\ln\left(\frac{R_4}{R_3}\right)}\right) & V_3 < V_{S30} < V_M \\ a_1 \ln\left(\frac{V_{S30}}{V_M}\right) & V_M \leq V_{S30} < V_4 \\ a_1 \ln\left(\frac{V_4}{V_M}\right) & V_{S30} \geq V_4 \end{cases} \quad (23)$$

$$\phi_{SS}^2(R_{rup}, V_{S30}) = \phi_{SS}^2(R_{rup}) + \Delta Var_{SS}(V_{S30}) \quad (24)$$

$$\phi_{SS}^2(R_{rup}) = \begin{cases} \phi_{SS,1}^2 & R_{rup} \leq R_5 \\ \frac{\phi_{SS,2}^2 - \phi_{SS,1}^2}{\ln\left(\frac{R_6}{R_5}\right)} \left(\ln\frac{R_{rup}}{R_5}\right) + \phi_{SS,1}^2 & R_5 < R_{rup} < R_6 \\ \phi_{SS,2}^2 & R_{rup} \geq R_6 \end{cases} \quad (25)$$

$$\Delta Var_{SS}(V_{S30}) = \begin{cases} a_2 \ln\left(\frac{V_3}{V_M}\right) \left(\frac{\ln\left(\frac{R_4}{\max(R_3, \min(R_4, R_{rup}))}\right)}{\ln\left(\frac{R_4}{R_3}\right)}\right) & V_{S30} \leq V_3 \\ a_2 \ln\left(\frac{V_{S30}}{V_M}\right) \left(\frac{\ln\left(\frac{R_4}{\max(R_3, \min(R_4, R_{rup}))}\right)}{\ln\left(\frac{R_4}{R_3}\right)}\right) & V_3 < V_{S30} < V_M \\ a_2 \ln\left(\frac{V_{S30}}{V_M}\right) & V_M \leq V_{S30} < V_4 \\ a_2 \ln\left(\frac{V_4}{V_M}\right) & V_{S30} \geq V_4 \end{cases} \quad (26)$$

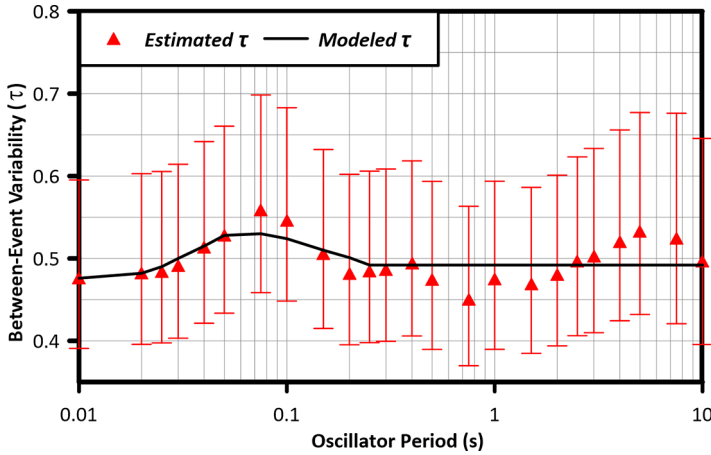


Figure 9. Period-dependence of between-event variability (in ln units) with 95% confidence intervals (triangles and bars), and the smoothed modeled τ for forward applications (line).

In Equations 22–23, $\phi_{S2S,0}^2$, a_1 , and V_M are period-dependent model coefficients, and V_3 , V_4 , R_3 , and R_4 are period-independent (Table 2). In Equations 24–26, $\phi_{SS,1}^2$, $\phi_{SS,2}^2$, a_2 , and V_M are period-dependent model coefficients, and R_5 and R_6 are period-independent (Table 2).

Between- and within-event aleatory variability. As indicated in Equation 17, our GMMs include models for between-event variability (τ) and within-event variability (ϕ). Prior studies for active tectonic regions (Gregor et al., 2014) showed decreases in τ with magnitude. We investigated this for subduction zones by computing event terms and their standard deviations (τ) from mixed-effects analyses for 0.5 M units bins between $M = 4.5$ and $M = 9.5$. The results did not reveal appreciable trends with magnitude, nor are there appreciable differences between event types (see Parker et al., 2020). As a result, the τ model is independent of M and event type (Figure 9). The peak near 0.1 s occurs across the considered subduction zone regions and has also been observed for small magnitude shallow earthquakes in active tectonic regions (Boore et al., 2014b; Campbell and Bozorgnia, 2014).

Using the same binning and inspection approach as for the between-event variability, we looked for trends in total within-event variance (ϕ^2) with magnitude, rupture distance, and V_{S30} . Variance rather than standard deviation is used to combine model terms (Equations 18–20). At short periods (e.g. PGA), no trend in ϕ^2 with magnitude was apparent. At 1.0 s PSA, ϕ^2 for $M < 7$ and at $M9$ are approximately equivalent, with an increase for intermediate magnitudes (approximately $M7$ – 8.75), but with large uncertainty compared to values at lower magnitudes and at $M9$ (see Figure 6.3 in Parker et al., 2020). As a result, we did not include magnitude terms in our ϕ^2 model, which departs from previous findings for active tectonic regions (Boore et al., 2014b).

Parker et al. (2020) presented a series of plots of binned ϕ^2 with respect to distance and V_{S30} . Dispersion increases for distances beyond 200 km for PGA and other short-period parameters. We anticipate that this is caused by complexities related to ground-motion attenuation that are not fully captured by regional terms in the path model, perhaps due

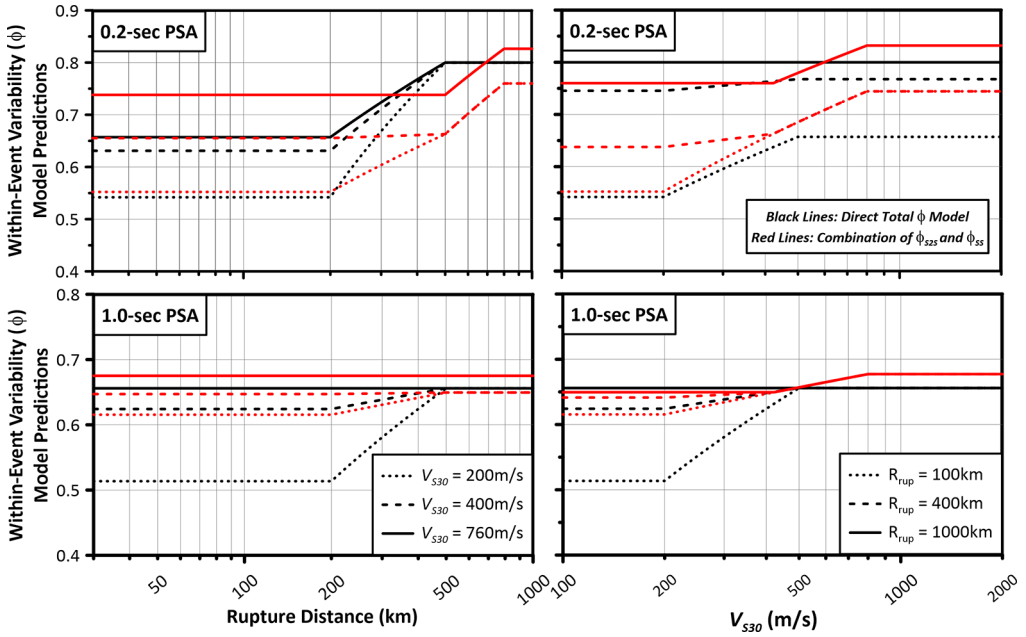


Figure 10. Total within-event aleatory variability (in ln units) from direct model (black lines) and from combination of partitioned models for ϕ_{S25}^2 and ϕ_{S5}^2 using Equation 21 (red lines), showing dependence on R_{rup} and V_{S30} for 0.2 and 1.0 s PSA. Note that while we develop the model using variances, this figure shows the resulting models for standard deviations.

to scattering and wave-type conversions at large distances. Within-event dispersion for PGA and other short-period parameters decreases for sites with V_{S30} below 500 m/s. This is likely related to site-response nonlinearity, which reduces the dispersion of site response (Al Atik and Abrahamson, 2010). These effects are not observed at long periods (> 1.0 s PSA). Similar features have been observed previously for active tectonic regions.

We modeled ϕ^2 using a piecewise function for within-event variance conditioned on R_{rup} and V_{S30} (Equations 18–20). This model was developed by first setting a minimum value of $\phi^2 = 0.30$ based on records with $R_{rup} \leq 200$ km and $V_{S30} \leq 200$ m/s, the conditions where we expect soil nonlinearity to be the most prevalent. This minimum value is period-independent. Distance-dependence (Equation 19) was evaluated by setting corner distances R_1 and R_2 based on visual inspection and then computing ϕ_1^2 and ϕ_2^2 from weighted least-squares regression using variances in $R_{rup} \leq R_1$ and $R_{rup} > R_2$ bins, respectively. The weights used in the regressions were taken as the inverse of the standard error of binned ϕ^2 . The data considered in the distance analyses had $V_{S30} \geq 500$ m/s. To incorporate V_{S30} -dependence (Equation 20), we selected corner velocities $V_1 = 200$ m/s and $V_2 = 500$ m/s by visual inspection, computed weighted variances in $V_{S30} \leq V_1$ and $V_{S30} > V_2$ bins, and then took ϕ_V^2 as the differences in these variances. Because these dispersion reductions are related to nonlinearity, the data considered had $R_{rup} \leq 200$ km. The distance terms in Equation 19 transition the site dependence to zero for $R_{rup} > R_2$. Figure 10 shows the principle features of the total within-event aleatory variability model at 0.2 and 1.0 s PSA.

The models for between- and direct within-event variability models presented in Equations 17–20 are meant to be used in ergodic ground motion analyses. The next

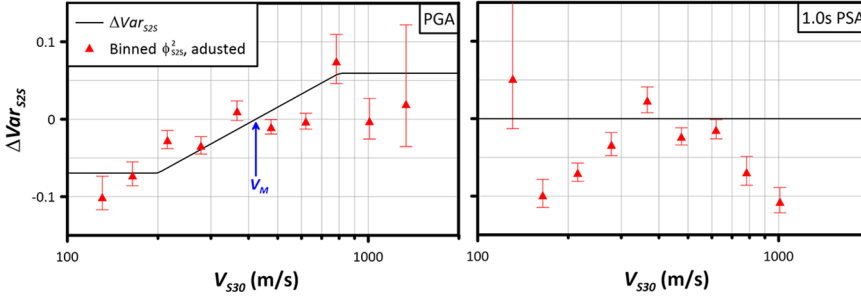


Figure 11. Values of ϕ_{S2S}^2 (in ln units) with 95% confidence intervals from V_{S30} -binned site terms for PGA and 1.0 s PSA. Solid line shows model for ΔVar_{S2S} (Equation 23).

section discusses a partitioned within-event variability model that is required for partially non-ergodic (i.e. site-specific) seismic hazard analyses (Stewart et al., 2017).

Partitioned within-event aleatory variability. In addition to the total within-event variance (ϕ^2), we provide models for the partitioned components, ϕ_{S2S}^2 and ϕ_{SS}^2 . These were computed using mixed-effects analyses on binned datasets to estimate site terms, from which site term variances were computed (ϕ_{S2S}^2) along with variances of remaining within-event residuals (ϕ_{SS}^2). Parker et al. (2020) present plots of both variances binned with respect to various predictor variables.

We started by visually inspecting trends of binned ϕ_{S2S}^2 with respect to magnitude, rupture distance, and V_{S30} . We did not find a trend in ϕ_{S2S}^2 with magnitude. Although we observe a slight trend in ϕ_{S2S}^2 at large R_{rup} for PGA and other short-period parameters, we did not model this dependence because there is not a physical basis for distance-dependent site-to-site variability. The distance trend may be an artifact of path-to-path variability that is mapped into site terms and hence into their variability (ϕ_{S2S}^2). Parameter $\phi_{S2S,0}^2$ (Equation 22) represents the distance-independent dispersion when V_{S30} is not considered, and is computed using data in the distance range 50 to 200 km. Next, we examine the dependence of ϕ_{S2S}^2 on V_{S30} by subtracting $\phi_{S2S,0}^2$ from V_{S30} -binned values of ϕ_{S2S}^2 , and plot these differential variances at the median V_{S30} for each bin (Figure 11). The four-segment model in Equation 23 passes through zero at the median V_{S30} for the population, (i.e. V_M). Site-to-site variability increases for stiff sites relative to $\phi_{S2S,0}^2$ and decreases for soft sites, which is consistent with the expected effects of site response nonlinearity. The value of ΔVar_{S2S} goes to zero at long periods where nonlinear effects in site response diminish. Because ΔVar_{S2S} is associated with soil nonlinearity—and thus shaking intensity—we applied its full effect for $R_{rup} < R_3$, and scaled it to zero for $R_{rup} > R_4$.

Within-event single-station variance ϕ_{SS}^2 was computed from the residuals remaining after fixed source and site effects are removed. Accordingly, it reflects the impact of path-to-path variability and event-to-event variability in site response for a given site. Thus, we examined trends in binned single-station variance (ϕ_{SS}^2) with earthquake magnitude, rupture distance, V_{S30} , and region. We modeled ϕ_{SS}^2 as magnitude-independent. Differences in ϕ_{SS}^2 by region were investigated and no consistent patterns were found; accordingly, the ϕ_{SS}^2 model is derived from global data. As shown in Figure 12, using data for all velocities, we observe no appreciable distance-dependence in binned values of ϕ_{SS}^2 up to 500 km, and then a sharp

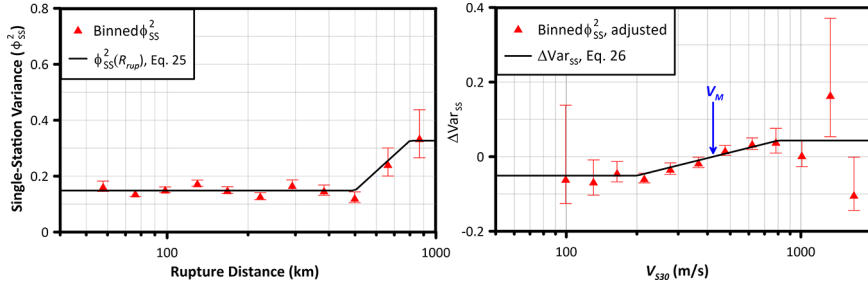


Figure 12. Values of ϕ_{SS}^2 (in ln units) for PGA with 95% confidence intervals binned by distance and V_{S30} . Solid lines show models for $\phi_{SS}^2(R_{rup})$ (Equation 25) and ΔVar_{SS} (Equation 26).

increase occurs. This could be caused by increased wave scattering, wave-type conversions, or disparate phase arrivals at long distances. We investigated whether this increase is a result of significant changes of regional contributions to data; in particular, Japanese data make up 43% of recordings with $R_{rup} < 500$ km and 20% with $R_{rup} > 500$ km. While we cannot exclude the possibility that the increase is at least in part regional, we nonetheless retained this feature in the model. Accordingly, we fit a piecewise linear function to the binned values, with corner distances at $R_5 = 500$ km, below which the variance is equal to $\phi_{SS,1}^2$, and $R_6 = 800$ km, above which the variance is equal to $\phi_{SS,2}^2$ (Equation 25, Figure 11). The physical basis for distance-dependent ϕ_{SS}^2 is path-to-path variability as with ϕ . We evaluated the V_{S30} -dependence of single-station variance by taking differences between V_{S30} -binned variances and the global average variance for sites with $R_{rup} < R_5$ (ΔVar_{SS}). A piecewise-linear model with three segments was fit to the results, which is forced to go through zero at the median V_{S30} value (V_M) used in determining $\phi_{SS,1}^2$ and $\phi_{SS,2}^2$ (Equation 26). We observed an increase in ϕ_{SS}^2 for fast V_{S30} and a decrease for slow V_{S30} (Figure 12). The piecewise linear function has three segments, with corner velocities of 200 and 800 m/s. As in the ϕ_{SS}^2 model, because ΔVar_{SS} is associated with soil nonlinearity, we applied its full effect for $R_{rup} < R_3$ and scaled this effect to zero for $R_{rup} > R_4$.

As given in Equation 21, the sum of the within-event single-station (ϕ_{SS}^2) and site-to-site (ϕ_{S2S}^2) variances is equivalent to the total within-event variance (ϕ^2). In Figure 10, we compare the direct model for ϕ (Equations 18–20) to the estimate from the component models using Equation 21. For 0.2 s PSA, the two models provide similar estimates of ϕ for close distances ($R_{rup} < 200$ km) and soft soils ($V_{S30} < 400$ m/s). For other conditions, the partitioned model shows more V_{S30} -dependence (including at large distances and stiff sites) and less distance-dependence than in the direct ϕ model. At longer periods (e.g. 1.0 s PSA), the models are similar for intermediate to large distances ($R_{rup} > 300$ km) and stiff-soil to rock-site conditions ($V_{S30} > 300$ m/s). At short distances and soft-site conditions ($V_{S30} = 200$ m/s, $R_{rup} = 100$ km), the partitioned model has larger estimates of variability by about 0.1 natural log unit.

For applications, we recommend the ϕ_{SS}^2 model for partially non-ergodic seismic hazard analyses in which the F_s term (Equation 7) is replaced with a site-specific model, and a site response epistemic uncertainty model is used (which should provide values less than ϕ_{S2S}^2 , informed either by empirical site response at the site of interest, or from uncertainty in one-dimensional ground response simulations, e.g. Stewart and Afshari, 2021). For ergodic analyses, we recommend using the total ϕ model.

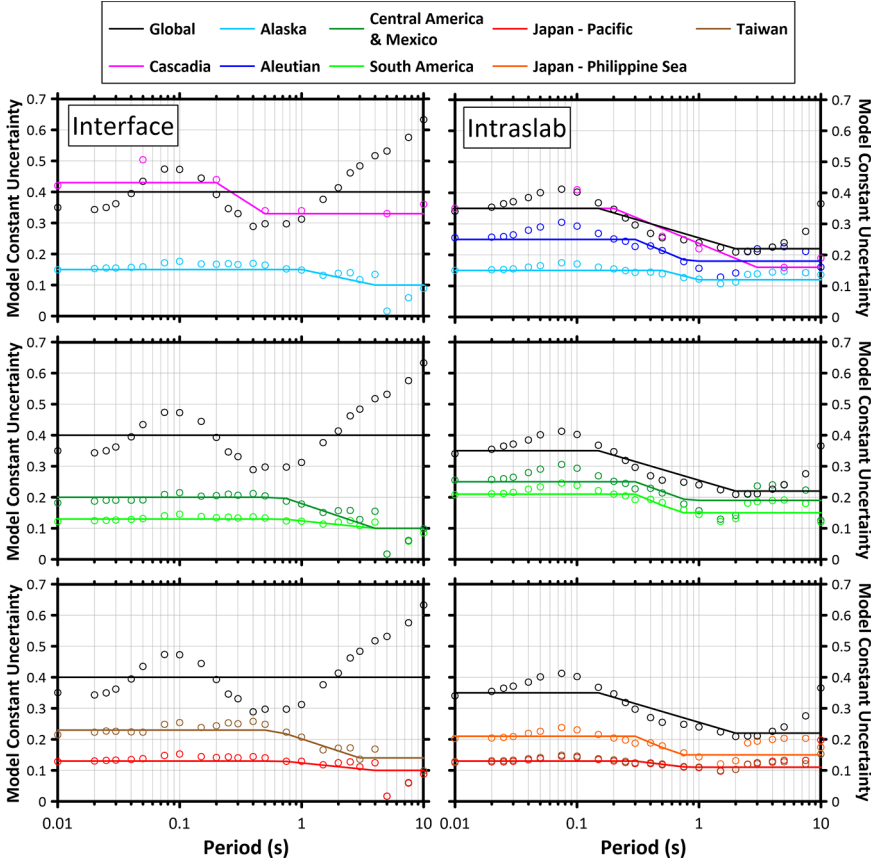


Figure 13. Epistemic uncertainty (σ_ε ; In units) in regional and global model constants (open circles) and model fit recommended for application (solid lines) for interface and intraslab events.

Epistemic uncertainty model

Functional form. We use a scaled backbone approach to represent epistemic uncertainty in the median GMM (Atkinson et al., 2014). In this approach, the median model is adjusted up and down in ground-motion space uniformly with respect to all independent variables. This is achieved by adjusting the median model constant terms (c_0 ; Equation 1) by $\varepsilon \times \sigma_\varepsilon$, where ε is the standard normal variate (mean of zero, standard deviation of 1.0) and σ_ε is an epistemic standard deviation. The constant c_0 can be summed with $\varepsilon \times \sigma_\varepsilon$ in natural log units, or the ground motion can be scaled by multiplying by $\exp(\varepsilon \times \sigma_\varepsilon)$. We take σ_ε as a function of period (T ; Equation 27), where model coefficients depend on event type and region. Model coefficients for Equation 27 are given in Table S4 in the Electronic Supplement to this article.

$$\sigma_\varepsilon(T) = \begin{cases} \sigma_{\varepsilon 1} & T < T_1 \\ \sigma_{\varepsilon 1} - (\sigma_{\varepsilon 1} - \sigma_{\varepsilon 2}) \frac{\ln(T/T_1)}{\ln(T_2/T_1)} & T_1 < T < T_2 \\ \sigma_{\varepsilon 2} & T > T_2. \end{cases} \quad (27)$$

Derivation of uncertainty model parameters. Different σ_ε values are estimated for interface and intraslab events (Figure 13). The global value of σ_ε is taken as the standard deviation of the regional median ground motions for the same conditions used to define the global constant: $\mathbf{M} = 7$ and $R_{rup} = 65$ km for interface, and $\mathbf{M} = 6$ and $R_{rup} = 75$ km for intraslab events. For all regions other than Cascadia, σ_ε was taken as the standard error of regional constants, which are small in data-rich regions (e.g. Japan; Figure 13, bottom) and relatively large in data-sparse regions (e.g. Central America and Mexico; Figure 13, middle).

Special consideration was given to Cascadia to represent the large epistemic uncertainty in median ground motions. Unlike most other regions, we lack empirical estimates of the constant or its standard error. Instead, we develop values of σ_ε for Cascadia based on the spread of constants across all regions considered. However, because these constants are correlated with regional corner magnitudes (m_c ; Table 3), adjustments are needed so that all of the constants are applicable to the Cascadia value of m_c (7.7 for interface and 7.2 for intraslab). The constant for region k ($c_{0,k}$) can be adjusted to an equivalent Cascadia constant ($c_{0,k}^{adj}$), where the superscript *adj* indicates an adjusted value, as:

$$c_{0,k}^{adj} = c_{0,k} + c_4 \Delta m_{c,k} + c_5 \Delta m_{c,k}^2 \quad (28)$$

where c_4 and c_5 are magnitude scaling coefficients (Equation 5) and $\Delta m_{c,k} = m_{c,cas} - m_{c,k}$.

We adjusted regional constants to be compatible with the Cascadia value of m_c using Equation 28, and take σ_ε as the weighted standard deviation of the adjusted interface and intraslab constants. The weights are inversely proportional to the standard errors of $c_{0,k}^{adj}$. For example, σ_ε for interface and intraslab PGA are 0.43 and 0.35, respectively. This translates to an 84th/16th percentile range of about 2.4 and 2.0 in ground motions. For all regions, σ_ε is larger at short periods and decreases by about 0.1 ln units at long periods (Figure 13). This epistemic uncertainty for Cascadia is larger than other regional values due to the relative sparsity of data.

Model coefficients for the epistemic uncertainty about the constants (Equation 27) are given in Table S4 of the Electronic Supplement to this article. We refer to Gregor et al. (2020) for recommendations on logic tree approaches when using the suite of NGA-Subduction GMMs.

Discussion

We recommend use of the scaled backbone approach, implemented through variations on constant terms via $\varepsilon \times \sigma_\varepsilon$, as the principle means by which to capture epistemic uncertainties in the application of the GMMs presented in this article. However, there are other sources of epistemic uncertainty that can be relatively easily accounted for within the framework of the GMM functional form.

One of the main limitations of our implementation of the scaled backbone approach is that the epistemic uncertainty does not change with independent variables. In general, larger uncertainty would be expected near the limits of the data (e.g. large magnitudes, short distances) than near the center of the data (e.g. Al Atik and Youngs, 2014), which is not accounted for via uncertainty on the constant c_0 . For the case of ground motion estimation at large magnitudes, specifically $\mathbf{M} > m_c$, uncertainties in the magnitude-scaling breakpoint (m_c) can be applied to shift ground motion amplitudes (as recommended by

Abrahamson et al., 2016). As shown in Equation 5, changes in m_c will alter the predicted ground motions over the entire magnitude range considered, whereas the present intent would be to adjust only large-magnitude ground motions. To correct for this, alternate values of m_c should be coupled with adjustments to the constant using Equation 28 to maintain the same level of ground motion for $\mathbf{M} < m_c$. In this case, $c_{0,k}$ should be taken as the constant for the region of interest, and $\Delta m_{c,k}$ represents the change in m_c that is applied to account for its epistemic uncertainty. Recommended mean values of m_c are given in Table 3; variations on this parameter are given in the studies by Campbell (2020) for interface events and Ji and Archuleta (2018) for slab events.

As described in the section “Source-depth scaling,” if the mean hypocentral depth (\bar{Z}_{hyp} ; Equations 14–15) is used in place of hypocentral depth as a directly applied independent variable, there is uncertainty in the mean depths that affects median ground motions. This can be accounted for by varying the dimensionless down-dip location of the hypocenter θ_W relative to its mean value. Additional discussion of uncertainties related to intraslab earthquake sources is provided by Weatherill et al. (2017).

Model residuals

Residuals analyses were performed to check model performance with respect to predictor variables. Three types of model residuals were considered (Al Atik et al., 2010): within-event residuals (δW_{ij}), the site-to-site component of within-event residuals (also known as site terms, $\eta_{S,j}$) and between-event residuals (also known as event terms, $\eta_{E,i}$). Here, we focus on δW_{ij} and $\eta_{E,i}$ to evaluate overall model performance with respect to source and path parameters (Equation 29).

$$R_{ij} = c + \eta_{E,i} + \delta W_{ij} \quad (29)$$

A more complete presentation of residuals analyses results is presented in Parker et al. (2020).

The overall model bias when region-specific constants and other terms are used, c , is relatively low, generally ± 0.1 ln units. We expect a small but nonzero bias because of the manual adjustments to constant terms to improve model performance; see the section “Regional and global constants.”

Event terms are shown as a function of \mathbf{M} for PGA, 0.2, 1.0, and 5.0 s for the interface model in Figure 14a and the intraslab model in Figure 14b. The event terms were computed using regional terms where applicable and are color-coded by region in each plot. Not all regional data have a sufficient number of events over a wide enough \mathbf{M} range ($> \sim 2$ \mathbf{M} units) to judge model effectiveness, for example, interface Taiwan, and intraslab Alaska, Central America and Mexico. For the other regions, the event terms do not appear to trend with magnitude. Similarly, the event terms do not trend appreciably with hypocentral depth for PGA, 0.2, 1.0, and 5.0 s for the interface and intraslab models (see Figures 4.13 and 4.14 in Parker et al., 2020).

Within-event residuals are shown as a function of distance for PGA, 0.2, 1.0, and 5.0 s for the interface model in Figure 15a, and for the intraslab model in Figure 15b. Residuals were computed using regional terms where applicable and are color-coded by region. For both the overall dataset and regional datasets, the trend of residuals with distance are

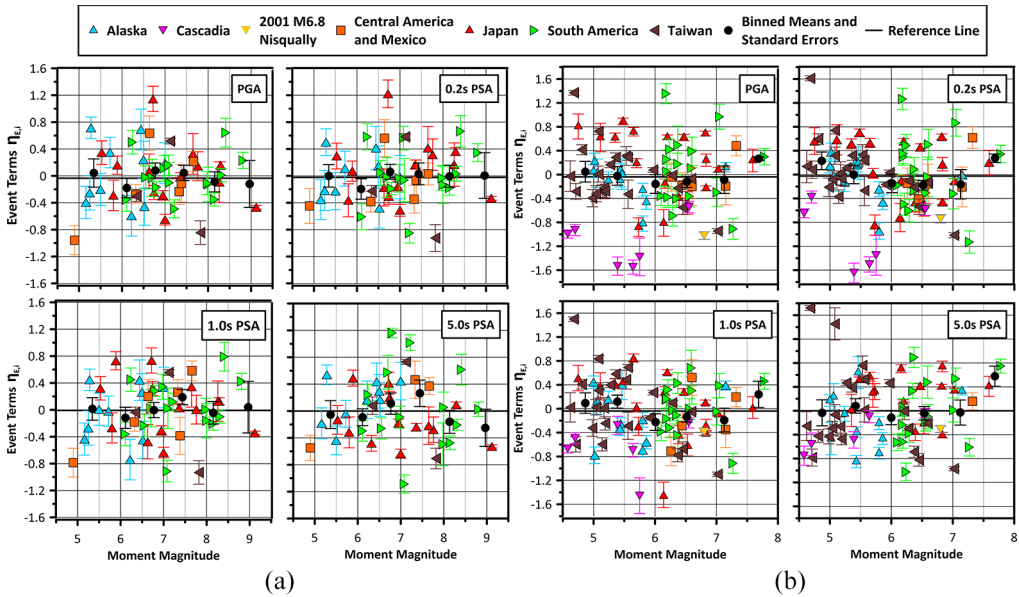


Figure 14. Event terms as a function of moment magnitude for PGA and 0.2, 1.0, and 5.0 s PSA for (a) interface earthquakes and (b) intraslab earthquakes. Binned means are plotted as filled circles along with standard errors. In (b), the 2001 $M6.8$ Nisqually intraslab earthquake is shown as a gold triangle.

reasonably flat. Site terms as a function of V_{S30} are presented in the companion article by Parker and Stewart (2021).

Model behavior and regional comparisons

In this section, we demonstrate the global model behavior and compare the various regional models. Comparisons with other NGA-Subduction models (Abrahamson and Gulerce, 2020; Kuehn et al., 2020; Si et al., 2020) and existing models for subduction zones (Abrahamson et al., 2016, 2018; Atkinson and Boore, 2003; Atkinson and Macias, 2009; Zhao et al., 2006, 2016a, 2016b) are provided by Gregor et al. (this issue). Additional comparisons to existing models for subduction zones are provided in Chapter 7 of Parker et al. (2020).

In general, the intraslab model predicts larger ground motions than the interface model for a common magnitude and distance, in particular, for short periods and close distances (Figure 16). The magnitude scaling for intraslab events is steeper than for interface, and saturates less at large magnitudes (larger spacing between spectra for different magnitudes in Figure 16). In previous GMMs for subduction zones (e.g. Abrahamson et al., 2016; Atkinson and Boore, 2003), the magnitude scaling was taken as the same as between-event types. Interface events in general exhibit stronger anelastic attenuation than intraslab events (Figure 4), but those differences are minimal at the short distances shown in Figure 16.

Figure 17a presents regional comparisons of distance-scaling of 1.0 s PSA for an $M = 9$ interface event with $V_{S30} = 760$ m/s. Regional variations are smaller at short distances (< 200 km) and increase at larger distances (> 600 km). Even at short distances,

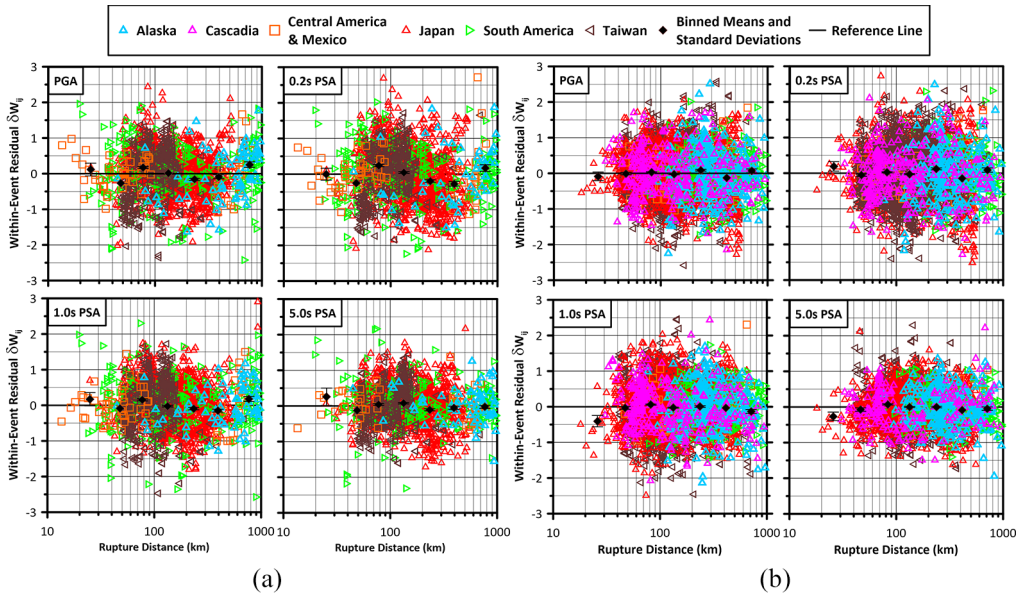


Figure 15. Within-event residuals as a function of rupture distance for PGA and 0.2-, 1.0-, and 5.0-s PSA for (a) interface earthquakes and (b) intraslab earthquakes. Binned means are plotted as filled circles along with standard errors.

the spread in the GMMs is a factor of 2.5 for these large magnitude scenarios. At large distances, Central America and Mexico, South America, Alaska, and the Aleutian Islands have the strongest motions due to slower decay with distance than Japan and Taiwan. The global model largely tracks that for Japan (Pacific Plate), but amplitudes are slightly higher at large distances due to slower anelastic attenuation. The Cascadia interface model predictions are equal to global predictions and are not shown. Figure 17b presents similar ground motion comparisons for an $M = 8$ intraslab event. It may be noted that regional variations appear less pronounced for large intraslab events relative to those for large interface events. As noted previously, the relatively limited range of m_c for intraslab as compared to interface events largely explains the different levels of regional variability. One notable feature is that the Cascadia distance scaling shows stronger empirical anelastic attenuation than the other regions (also shown in Figure 4).

Summary and discussion

We describe the development of a horizontal-component GMM applicable to global subduction zone regions. The models for both interface and intraslab earthquakes account for near-source saturation, geometrical spreading, anelastic attenuation, magnitude-scaling, and site response. The model for intraslab earthquakes also accounts for source-depth scaling. The GMM is formulated with adjustment factors that can be used to customize the models for regional conditions in Alaska, the Aleutian Islands, Cascadia, Central America and Mexico, Japan–Pacific Plate, Japan–Philippine Sea Plate, southern and northern portions of South America, and Taiwan. The regional modifications apply to the model amplitude (constant, c_0), magnitude scaling breakpoint (m_c), and anelastic

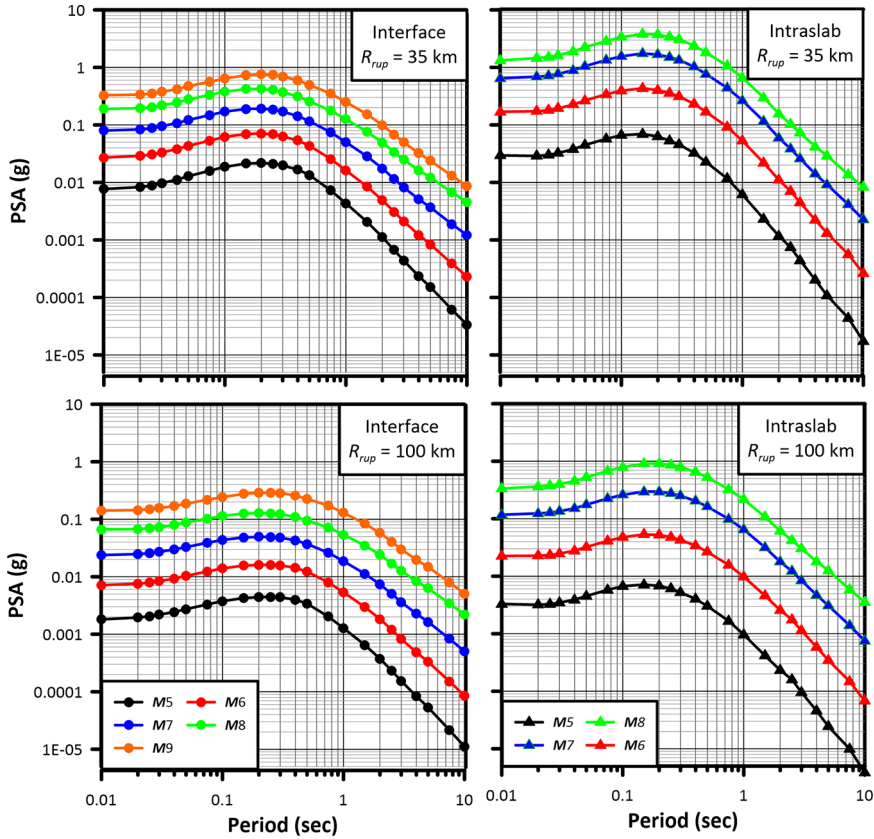


Figure 16. Pseudo-spectral acceleration (PSA) predictions for interface and intraslab earthquakes for **M5–9** and **5–8**, respectively. All plots are for the reference shear wave velocity condition $V_{S30} = 760$ m/s. Top panels show predictions for $R_{rup} = 35$ km, the bottom panels for $R_{rup} = 100$ km. Slab predictions are made for $Z_{hyp} = 60$ km.

attenuation coefficients (a_0). Site response is also regionalized (Parker and Stewart, 2021), with regional V_{S30} -scaling and basin sediment depth terms, and a global soil nonlinearity term.

For locations where regional factors are not defined, a global version of the model can be applied with larger epistemic uncertainty. The aleatory variability model provides estimates of the between-event, within-event, site-to-site, and within-event single-station standard deviations as a function of V_{S30} and R_{rup} . We provide a recommended epistemic uncertainty model on the GMM constant (c_0) that depends on region, event type, and PSA oscillator period (T). Coded versions of the median, aleatory variability, and epistemic uncertainty models are provided by Mazzoni et al. (2020). Model coefficients for all model components are given in Tables 2 and 3 and the Electronic Supplement to this article (Tables S1–S4 in the Electronic Supplement).

The GMM predicts PGA, PGV, and 5%-damped PSA at oscillator periods between 0.01 and 10.0 s for interface and intraslab subduction-zone events. The reference condition of the GMM is $V_{S30} = 760$ m/s, but it can be applied to a range of site conditions using the site response model given herein and discussed in Parker and Stewart (2021). The

interface model is valid for $4.5 \leq M \leq 9.5$, $R_{rup} = 20\text{--}1000$ km, $Z_{hyp} \leq 40$ km, and $V_{S30} = 150\text{--}2000$ m/s. The intraslab model is valid over $4.5 \leq M \leq 8.5$, $R_{rup} = 35\text{--}1000$ km, $Z_{hyp} = 20\text{--}200$ km, and $V_{S30} = 150\text{--}2000$ m/s. Both models are applicable only to sites in the forearc region of subduction zones. Based on preliminary residuals analyses, in general, we recommend that the model can be applied as-is to sites in subduction zone backarc (as defined in Contreras et al., this issue) in the Aleutian Islands, Mexico, and in Cascadia up to 400 km east of the volcanic arc, but should not be applied without further adjustment to South America, Central America, Japan, or Alaska. Additional details on the regional applicability of the model to backarc regions are given in Table 4. Future work is planned to create additional anelastic attenuation terms in these regions.

Cascadia is a region of practical importance for hazard applications in the United States and Canada, but we lack sufficient data to support model development for interface events (Figure 1). The situation is somewhat better resolved for intraslab events, although they are mostly limited to small magnitudes. We have developed a model that we consider technically defensible for hazard applications, but (1) epistemic uncertainties are larger than in other relatively data-rich regions (Figure 13) and (2) the model development was aimed at hazard-critical large-magnitude scenarios, which may lead to the overprediction of ground motion from small-to-moderate magnitude intraslab events, especially at short periods (e.g. the 2001 $M6.8$ Nisqually earthquake, shown as a gold triangle in Figure 14b).

Aleatory variability models are developed that encompass both event types, with different coefficients for each IM. Models are provided for four components of ground-motion variability: (1) between-event variability, τ ; (2) within-event variability, ϕ ; (3) single-station within-event variability, ϕ_{SS} ; and (4) site-to-site variability, ϕ_{S2S} . The aleatory variability models are magnitude-independent, but within-event variabilities depend on R_{rup} and V_{S30} . Ergodic analyses should use the median GMM and aleatory variability computed using the between-event and within-event variability models. An analysis incorporating non-ergodic site response (i.e. partially non-ergodic) should use the median GMM for reference-rock, a site-specific site response model with appropriate epistemic uncertainty, and

Table 4. Can the path model developed for subduction zone forearc regions (Equations 2–4) be used to estimate ground motions at locations in backarcs?

Region	Interface	Intraslab
Alaska	No	No
Aleutian Islands	Yes	Yes
Cascadia	Not enough data to inform recommendation	Yes, up to 400 km east of the volcanic arc as defined by Contreras et al. (this issue)
Central America	Not enough data to inform recommendation	Not enough data to inform recommendation
Japan	Yes, if earthquake is in Region 3 and station is in Region 1 or 3 (as defined by Contreras et al., this issue). No, otherwise	Yes, if earthquake is in Region 3 and station is in Region 1 or 3 (as defined by Contreras et al., this issue). No, otherwise
Mexico	Yes	Yes
South America	No	No
Taiwan	Not enough data to inform recommendation	Not enough data to inform recommendation

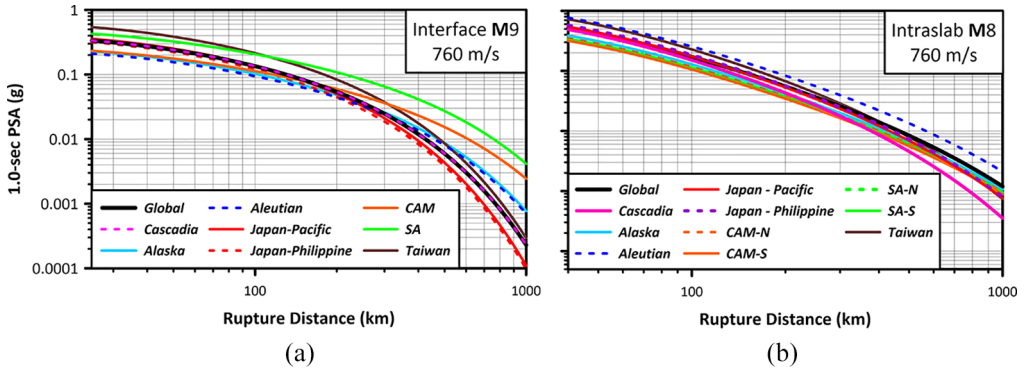


Figure 17. Regional comparisons of distance scaling of 1.0 s PSA for $V_{530} = 760$ m/s for (a) $M = 9$ interface events and (b) $M = 8$ intralab event (CAM = Central America and Mexico; SA = South America). A hypocentral depth of 55 km was used for the intralab analyses.

aleatory variability computed using the between-event and single-station within-event variability models.

Epistemic uncertainty in the median model is represented by standard deviation terms on region-dependent model constant terms (σ_e). This epistemic standard deviation facilitates scaled-backbone representations of model uncertainty in hazard analyses. The GMM functional form allows for additional sources of epistemic uncertainty related to large-magnitude ground motions and source depth effects.

Future work is needed to develop recommendations on epistemic uncertainty in other model components such as magnitude and distance-scaling. Future work could also consider additional subduction zone regions such as New Zealand, Greece, Italy, and Indonesia; partially non-ergodic path terms that consider in-slab wave propagation; calibration for class 2 aftershock events; more nuanced basin models that consider basin shape or relative distances to the basin edge; or additional basin models for structures such as the Santiago, Taipei, and Mexico City basins.

Acknowledgments

The Principal Investigator for the NGA-Sub project was Yousef Bozorgnia, and the project was organized through the Pacific Earthquake Engineering Research (PEER) center. We thank the many researchers of the broader NGA-Sub team, and, in particular, Silvia Mazzoni, Sean Ahdi, Victor Contreras, Tadahiro Kishida, and Robert Darragh for their efforts in producing the NGA-Sub database. We thank the other NGA-Sub model developers for fruitful discussions that substantially improved this model. We are grateful to Nico Kuehn, Sanaz Rezaeian, Annemarie Baltay, Fleur Strasser, Suzanne Hecker, Michael Diggles, and two anonymous reviewers for their reviews of this manuscript.

Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work of the first two authors within the NGA-Sub project was

sponsored by the U.S. Geological Survey contract G16AP00181, as well as the UCLA Graduate Division and Civil & Environmental Engineering Department. The support of these organizations is gratefully appreciated. The opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of PEER center or the Regents of the University of California.

ORCID iDs

Grace A Parker  <https://orcid.org/0000-0002-9445-2571>

Jonathan P Stewart  <https://orcid.org/0000-0003-3602-3629>

Supplemental material

Supplemental material for this article is available online.

References

- Abrahamson NA and Gulerce Z (2020) *Regionalized ground-motion models for subduction earthquakes based on the NGA-sub database*. Report 2020/05. Berkeley, CA: Pacific Earthquake Engineering Research Center, University of California, Berkeley.
- Abrahamson NA, Gregor N and Addo K (2016) BC Hydro ground motion prediction equations for subduction earthquakes. *Earthquake Spectra* 32: 23–44.
- Abrahamson NA, Kuehn N, Gulerce Z, Gregor N, Bozorgnia Y, Parker GA, Stewart JP, Chiou B, Idriss IM, Campbell K and Youngs R (2018) *Update of the BC hydro subduction ground-motion model using the NGA-subduction dataset*. Report No. 2018/02, Berkeley, CA: Pacific Earthquake Engineering Research Center, University of California, Berkeley.
- Ahdi SK, Kwak DY, Ancheta TD, Contreras V, Kishida T, Kwok AO, Mazzoni S, Ruz F and Stewart JP (this issue) Site parameters applied in NGA-Sub database. *Earthquake Spectra*.
- Al Atik L, Abrahamson NA, Bommer JJ, Scherbaum F, Cotton F and Kuehn N (2010) The variability of ground-motion prediction models and its components. *Seismological Research Letters* 81: 794–801.
- Al Atik L and Abrahamson N (2010) Nonlinear site response effects on the standard deviations of predicted ground motions. *Bulletin of the Seismological Society of America* 100: 1288–1292.
- Al Atik L and Youngs RR (2014) Epistemic uncertainty for NGA-West2 models. *Earthquake Spectra* 30: 1301–1318.
- Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BS-J, Wooddell KE, Graves RW, Kottke AR, Boore DM, Kishida T and Donahue JL (2014) NGA-West 2 database. *Earthquake Spectra* 30: 989–1005.
- Anderson JG and Brune JN (1999) Probabilistic seismic hazard analysis without the ergodic assumption. *Seismological Research Letters* 70: 19–28.
- Assatourians K and Atkinson GM (2012) EXSIM12: A stochastic finite-fault computer program in FORTRAN. Available at: <http://www.seisimtoolbox.ca> (accessed November 2017)
- Astiz L, Lay T and Kanamori H (1988) Large intermediate-depth earthquakes and the subduction process. *Physics of the Earth and Planetary Interiors* 53: 80–166.
- Atkinson GM (1997) Empirical ground motion relations for earthquakes in the Cascadia region. *Canadian Journal of Civil Engineering* 24: 64–77.
- Atkinson GM and Boore DM (2003) Empirical ground-motion relations for subduction-zone earthquakes and their application to Cascadia and other regions. *Bulletin of the Seismological Society of America* 93: 1703–1729.
- Atkinson GM and Macias M (2009) Predicted ground motions for great interface earthquakes in the Cascadia subduction zone. *Bulletin of the Seismological Society of America* 99: 1552–1578.
- Atkinson GM and Silva WJ (2000) Stochastic modeling of California ground motions. *Bulletin of the Seismological Society of America* 90: 255–274.

- Atkinson GM, Bommer JJ and Abrahamson NA (2014) Alternative approaches to modeling epistemic uncertainty in ground motions in probabilistic seismic-hazard analysis. *Seismological Research Letters* 85: 1141–1144.
- Atkinson GM, Yenier E, Sharma N and Convertito V (2016) Constraints on the near-distance saturation of ground-motion amplitudes for small-to-moderate induced earthquakes. *Bulletin of the Seismological Society of America* 106: 2104–2111.
- Baltay AS, Hanks TC and Abrahamson NA (2019) Earthquake stress drop and Arias intensity. *Journal of Geophysical Research-Solid Earth* 124: 3838–3852.
- Bates D, Maechler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48.
- Bilek SL and Lay T (1998) Variation of interplate fault zone properties with depth in the Japan subduction zone. *Science* 281: 1175–1178.
- Bilek SL and Lay T (1999) Rigidity variations with depth along interplate megathrust faults in subduction zones. *Nature* 400: 443–446.
- Bird P (2003) An updated digital model of plate boundaries. *Geochem Geophys Geosyst* 4(3): 1–52.
- Boore DM (2009) Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM. *Bulletin of the Seismological Society of America* 99: 3202–3216.
- Boore DM (2010) Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. *Bulletin of the Seismological Society of America* 100: 1830–1835.
- Boore DM, Di Alessandro C and Abrahamson NA (2014a) A generalization of the double-corner-frequency source spectral model and its use in the SCEC BBP validation exercise. *Bulletin of the Seismological Society of America* 104: 2387–2398.
- Boore DM, Stewart JP, Seyhan E and Atkinson GM (2014b) NGA-West 2 equations for predicting PGA, PGV, and 5%-damped PSA for shallow crustal earthquakes. *Earthquake Spectra* 30: 1057–1085.
- Campbell KW (2020) Proposed methodology for estimating the magnitude at which subduction megathrust ground motions and source dimensions exhibit a break in magnitude scaling: Example for 79 global subduction zones. *Earthq. Spectra* 36: 1271–1297.
- 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: 1087–1115.
- Chao SH, Chiou BS-J, Hsu CC and Lin PS (2020) A horizontal ground-motion model for crustal and subduction earthquakes in Taiwan. *Earthquake Spectra* 36: 463–506.
- Chiou BS-J and Youngs RR (2008) An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra* 24: 173–215.
- Contreras V, Stewart JP, Kishida T, Darragh RB, Chiou BSJ, Mazzoni S, Youngs RR, Kuehn NM, Ahdi SK, Wooddell K, Boroschek R, Rojas F and Ordenes J (this issue) NGA-Sub source and path database. *Earthquake Spectra*.
- Crouse CB, Vyas YK and Schell BA (1988) Ground motions from subduction-zone earthquakes. *Bulletin of the Seismological Society of America* 78: 1–25.
- Frankel A, Wirth E, Marafi N, Vidale J and Stephenson W (2018) Broadband synthetic seismograms for magnitude 9 earthquakes on the Cascadia megathrust based on 3D simulations and stochastic synthetics, part 1: Methodology and overall results. *Bulletin of the Seismological Society of America* 108: 2347–2369.
- Frankel AD, Mueller CS, Barnhard T, Perkins D, Leyendecker E, Dickman N, Hanson S and Hopper M (1996) *National seismic hazard maps: Documentation*. U.S. Geological Survey Open-file Report 96-532. Washington, DC: U.S. Geological Survey.
- Gregor N, Abrahamson NA, Addo K, Al Atik L, Atkinson GM, Boore DM, Bozorgnia Y, Campbell KW, Chiou BSJ, Hassani B, Kishida T, Kuehn N, Midorikawa S, Mazzoni S, Parker

- GA, Si H, Stewart JP and Youngs RR (2020) *Comparison of NGA-Sub ground motion models*. Report 2020/07. Berkeley, CA: Pacific Earthquake Engineering Research Center, University of California, Berkeley.
- Gregor N, Abrahamson NA, Atkinson GM, Boore DM, Bozorgnia Y, Campbell KW, Chiou BS-J, Idriss IM, Kamai R, Seyhan E, Silva WJ, Stewart JP and Youngs RR (2014) Comparison of NGA-West2 GMPEs. *Earthquake Spectra* 30: 1179–1197.
- Gregor N, Silva WJ, Wong I and Youngs RR (2002) Ground-motion attenuation relationships for Cascadia subduction zone megathrust earthquakes based on a stochastic finite-fault modeling. *Bulletin of the Seismological Society of America* 92: 1923–1932.
- Hassani B and Atkinson GM (2018) Adjustable generic ground-motion prediction equation based on equivalent point-source simulations: Accounting for Kappa effects. *Bulletin of the Seismological Society of America* 108: 913–928.
- Ji C and Archuleta R (2018) *Scaling of PGA and PGV Deduced from Numerical Simulations of Intraslab Earthquakes*. Santa Barbara, CA: Department of Earth Science, University of California, Santa Barbara.
- Joyner WB and Boore DM (1981) Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. *Bulletin of the Seismological Society of America* 71: 2011–2038.
- Joyner WB and Boore DM (1993) Methods for regression analysis of strong-motion data. *Bulletin of the Seismological Society of America* 83: 469–487.
- Joyner WB and Boore DM (1994) Errata: Methods for regression analysis of strong-motion data. *Bulletin of the Seismological Society of America* 84: 955–956.
- Kanno T, Narita A, Morikawa N, Fujiwara H and Fukushima Y (2006) A new attenuation relation for strong ground motion in Japan based on recorded data. *Bulletin of the Seismological Society of America* 96: 879–897.
- Kishida T, Darragh RB, Chiou BSJ, Bozorgnia Y, Mazzoni S, Contreras V, Boroschek R, Rojas F and Stewart JP (2020) Chapter 3 – Ground motions and intensity measures. in Report No. 2020/02 (ed Stewart JP). Berkeley, CA: Pacific Earthquake Engineering Research Center, University of California, Berkeley.
- Kuehn N, Bozorgnia Y, Campbell KW and Gregor N (2020) *Partially non-ergodic ground-motion model for subduction regions using NGA-subduction database*. Report No. 2020/04. Berkeley, CA: Pacific Earthquake Engineering Research Center, University of California, Berkeley.
- Lin PS and Lee CT (2008) Ground-motion attenuation relationships for subduction-zone earthquakes in northeastern Taiwan. *Bulletin of the Seismological Society of America* 98: 220–240.
- Mai PM, Spudich P and Boatwright J (2005) Hypocenter locations in finite-source rupture models. *Bulletin of the Seismological Society of America* 95: 965–980.
- Mazzoni S, et al. (2020) NGA-subduction ground motion characterization tools. *Natural Hazards Risk and Resiliency Research Center, UCLA*. <https://www.risksciences.ucla.edu/nhr3/gmtools/#NGASubGMTool>
- Mazzoni S, Kishida T, Stewart JP, Contreras V, Darragh RB, Ancheta TD, Chiou BSJ, Silva WJ and Bozorgnia Y (this issue) Relational database used for ground-motion model development in the NGA-Sub project. *Earthquake Spectra*.
- Melgar D and Hayes GP (2019) The correlation lengths and hypocentral positions of great earthquakes. *Bulletin of the Seismological Society of America* 109: 2582–2593.
- Motazedian D and Atkinson GM (2005) Stochastic finite-fault modeling based on a dynamic corner frequency. *Bulletin of the Seismological Society of America* 95: 995–1010.
- Novakovic M, Atkinson GM and Assatourians K (2018) Empirically calibrated ground-motion prediction equation for Oklahoma. *Bulletin of the Seismological Society of America* 108: 2444–2461.
- Nweke CC, Stewart JP and Brandenberg SJ (2020) *Site response of southern California sedimentary basins and other geomorphic provinces*. Report GIRS-2020-12202012. Los Angeles, CA: B. John Garrick Risk Institute, Natural Hazards Risk and Resiliency Research Center, UCLA.

- Parker GA and Stewart JP (this issue) Ergodic site response model for subduction zone regions. *Earthquake Spectra*.
- Parker GA, Stewart JP, Boore DM, Atkinson GM and Hassani B (2020) *NGA-Subduction global ground-motion models with regional adjustment factors*. Report 2020/03. Berkeley, CA: Pacific Earthquake Engineering Research Center, University of California, Berkeley.
- Parker GA, Stewart JP, Hashash YMA, Rathje EM, Campbell KW and Silva WJ (2019) Empirical linear seismic site amplification in central and eastern North America. *Earthquake Spectra* 35: 849–881.
- Phung VB, Loh CH, Chao SH and Abrahamson NA (2020) Ground motion prediction equation for Taiwan subduction zone earthquakes. *Earthquake Spectra* 36: 1331–1358.
- Powers PM, Rezaeian S, Shumway AM, Petersen MD, Luco N, Boyd OS, Moschetti MP, Frankel AD and Thompson EM (2021) The 2018 update of the US national seismic hazard model: Ground motion models in the western US. *Earthquake Spectra*. Epub ahead of print 14 May. DOI: 10.1177/87552930211011200.
- R Core Team (2019) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rogers AM and Perkins DM (1996) Monte Carlo simulation of peak-acceleration attenuation using a finite-fault uniform-patch model including isochrone and extremal characteristics. *Bulletin of the Seismological Society of America* 86: 79–92.
- Seyhan E and Stewart JP (2014) Semi-empirical nonlinear site amplification from NGA-West2 data and simulations. *Earthquake Spectra* 30: 1241–1256.
- Sharma N, Convertito V, Maercklin N and Zollo A (2013) Ground-motion prediction equations for the Geysers geothermal area based on induced seismicity records. *Bulletin of the Seismological Society of America* 103: 117–130.
- Si H and Midorikawa S (1999) New attenuation relationships for peak ground acceleration and velocity considering effect of fault type and site condition. *Journal of Structural and Construction Engineering* 523: 63–70 (in Japanese).
- Si H, Midorikawa S and Kishida T (2020) *Development of NGA-sub ground motion model of 5%-damped pseudo-spectral acceleration based on database for subduction earthquakes in Japan*. Report 2020/06. Berkeley, CA: Pacific Earthquake Engineering Research Center, University of California, Berkeley.
- Stafford PJ, Rodriguez-Marek A, Edwards B, Kruiver PP and Bommer JJ (2017) Scenario dependence of linear site-effect factors for short-period response spectral ordinates. *Bulletin of the Seismological Society of America* 107: 2859–2872.
- Stern RJ (2002) Subduction zones. *Reviews of Geophysics* 40: 3–1.
- Stewart JP, Afshari K and Goulet CA (2017) Non-ergodic site response in seismic hazard analysis. *Earthquake Spectra* 33: 1385–1414.
- Stewart JP and Afshari K (2021) Epistemic uncertainty in site response as derived from one-dimensional ground response analyses. *Journal of Geotechnical and Geoenvironmental* 147: 04020146.
- Strasser FO, Arango MC and Bommer JJ (2010) Scaling of the source dimensions of interface and intraslab subduction-zone earthquakes with moment magnitude. *Seismological Research Letters* 81: 941–950.
- Suzuki S, Sasatani T and Motoya Y (1983) Double seismic zone beneath the middle of Hokkaido, Japan, in the southwestern side of the Kurile arc. *Tectonophysics* 96: 59–76.
- Van Houtte C, Bannister S, Holden C, Bourguignon S and McVerry G (2017) The New Zealand strong motion database. *Bulletin of the New Zealand Society for Earthquake Engineering* 50: 1–20.
- Weatherill G, Pagani M and Garcia J (2017) Modelling in-slab subduction earthquakes in PSHA: Current practice and challenges for the future. In: *Proceedings of the 16th World Conference on Earthquake Engineering*, Santiago, Chile, 9–13 January.
- Wirth EA, Frankel AD, Marafi N, Vidale JE and Stephenson WJ (2018a) Broadband synthetic seismograms for magnitude 9 earthquakes on the Cascadia megathrust based on 3D simulations and stochastic synthetics, part 2: Rupture parameters and variability. *Bulletin of the Seismological Society of America* 108: 2370–2388.

- Wooddell KE (2018) New earthquake classification for the NGA-subduction project. In: *Proceedings of the 11th National Conference on Earthquake Engineering*, Los Angeles, CA, 25–29 June. Oakland, CA: Earthquake Engineering Research Institute.
- Yenier E and Atkinson GM (2014) Equivalent point-source modeling of moderate-to-large magnitude earthquakes and associated ground-motion saturation effects. *Bulletin of the Seismological Society of America* 104: 1458–1478.
- Yenier E and Atkinson GM (2015a) An equivalent point-source model for stochastic simulation of earthquake ground motions in California. *Bulletin of the Seismological Society of America* 105: 1435–1455.
- Yenier E and Atkinson GM (2015b) Regionally adjustable generic ground-motion prediction equation based on equivalent point source simulations: Application to central and eastern North America. *Bulletin of the Seismological Society of America* 105: 1989–2009.
- Youngs RR, Chiou SJ, Silva WJ and Humphrey JR (1997) Strong ground motion attenuation relationships for subduction zone earthquakes. *Seismological Research Letters* 68: 58–73.
- Youngs RR, Day SM and Stevens JL (1988) Near field ground motions on rock for large subduction earthquakes. In: JL Von Thun (editor). *Earthquake Engineering and Soil Dynamics II: Recent Advances in Ground-Motion Evaluation*. Reston, VA: American Society of Civil Engineers, pp. 445–462.
- Zhao JX, Jiang F, Shi P, Xing H, Huang H, Hou R, Zhang Y, Yu P, Lan X, Rhoades DA and Somerville PG (2016a) Ground-motion prediction equations for subduction slab earthquakes in Japan using site class and simple geometric attenuation functions. *Bulletin of the Seismological Society of America* 106: 1535–1551.
- Zhao JX, Liang X, Jiang F, Xing H, Zhu M, Hou R, Zhang Y, Lan X, Rhoades DA, Irikura K and Fukushima Y (2016b) Ground-motion prediction equations for subduction interface earthquakes in Japan using site class and simple geometric attenuation functions. *Bulletin of the Seismological Society of America* 106: 1518–1534.
- Zhao JX, Zhang J, Asano A, Ohno Y, Oouchi T, Takahashi T, Ogawa H, Irikura K, Thio HK, Somerville PG and Fukushima Y (2006) Attenuation relations of strong ground motion in Japan using site classification based on predominant period. *Bulletin of the Seismological Society of America* 96: 898–913.