NGA-Subduction Global Ground Motion Models with Regional Adjustment Factors

Grace A. Parker, M.EERI, Jonathan P. Stewart, M.EERI, David M. Boore, Gail M. Atkinson, M.EERI, and Behzad Hassani

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 sec, for the median orientation-independent horizontal component of subduction earthquake ground motion. 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 are applicable to interface and intraslab subduction earthquakes in Japan, Taiwan, Mexico, Central America, South America, Alaska, the Aleutian Islands, and Cascadia. 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, $V_{S30}$-scaling, and sediment depth terms. We develop models for the aleatory between-event variability ($\tau$), within-event variability ($\phi$), single-station within-event variability ($\phi_{SS}$), and site-to-site variability ($\phi_{S2S}$). Ergodic analyses should use the median GMM and aleatory variability computed using the between-event and within-event variability.

1 Courtesy-Review DISCLAIMER: This draft manuscript is distributed solely for the purpose of courtesy review. Any comments received will be addressed and treated as appropriate to ensure there is no conflict of interest. Its content is merely being considered for publication, so it is not to be disclosed or released by reviewers. Until the manuscript has been approved for publication by the U.S. Geological Survey (USGS), it does not represent any official USGS finding or policy.

2 U.S. Geological Survey, Earthquake Science Center, P.O. Box 158, Moffett Field, CA, 94035; gparker@usgs.gov
3 University of California Los Angeles, 5731 Boelter Hall, Los Angeles, CA, 90095
4 Western University, London, 1151 Richmond St., London, ON, CAN
5 BC Hydro, 1800 Gilmore Ave., Burnaby, BC, CAN
models. An analysis incorporating nonergodic site response should use the median GMM at the reference shear-wave velocity condition, a site-specific site response factor, 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 deviation on the regional model constants, which facilitates scaled-backbone representations of model uncertainty in hazard analyses.

**INTRODUCTION**

Subduction zones produce interface earthquakes at the boundary between subducting and overriding tectonic plates, typically reverse in mechanism, and intraslab earthquakes within the subducting plate, typically normal in mechanism (Stern, 2002). This type of tectonic environment occurs in many highly populated regions globally, such as the Pacific Northwest (PNW) region of the U.S. and Canada, making the resulting seismic hazards 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 (e.g., Atkinson, 1997; Youngs et al., 1988; 1997; and Crouse et al., 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 national seismic hazard model.

As the size and reliability of ground-motion databases increased, this ergodic assumption—that ground motion should behave similarly across global regions—was disproven (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 two 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 or the 2011 M9.1 Tohoku, Japan interface earthquakes. The analyses of Abrahamson et al. (2016) found that the same magnitude-scaling slope can be used for interface and intraslab events, but different distance-scaling slopes are 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 $V_{s30}$-scaling terms.

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

The Next Generation Attenuation-Subduction (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 encompasses 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 (Mazzoni et al. 2020b, 202x; Kishida et al. 2020) 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 0.01-10 sec, for the orientation-independent horizontal component (RotD50; Boore, 2010). The reference site condition of the model is $V_{s30} = 760$ 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, 202x). Coded versions of the model in Excel, MatLab, R, and Python are available from Mazzoni et al. (2020a).

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, \( V_{30} \)-scaling, 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 manuscript is adapted from Parker et al. (2020), which presents the model and its derivation in greater detail.

**MODEL FUNCTIONAL FORM**

This section presents the suite of model functions, including the median interface, intraslab, regional, aleatory variability, and epistemic uncertainty models. Subsequent sections describe data selection and model development. The site response functions are included here for completeness, but the development and features of these models are presented in Parker and Stewart (20xx). All model input parameters are defined in Table 1, all model coefficients are defined in Tables 2-3, with period-independent coefficients given there, and period-dependent coefficient values are given in Tables E1-E4 in the electronic supplement to this article.

**MEDIAN MODEL**

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 (Eq. 1-13) with the exception of the near-source saturation term (Eq. 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 g and PGV in units of cm/s \( (\mu_{lnY}) \):

\[
\mu_{lnY} = c_0 + F_p + F_m + F_d + F_s
\]  

(1)
The path model (Eq. 2) incorporates magnitude-dependent geometrical spreading and anelastic attenuation. Near-source saturation is incorporated via parameter $h$ (Eq. 4), which is combined with site-to-source rupture distance ($R_{rup}$; Eq. 3). The near-source saturation term $h$ depends on event type.

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

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

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

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

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

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

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

Source-depth scaling is described with a tri-linear function conditioned on hypocentral depth, with 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; i.e., $F_D = 0$. 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 Source-Depth Scaling section).

The ergodic site response model, $F_S$, has three components (Parker and Stewart, 202x): (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:

\[ F_s = F_{\text{lin}} + F_{\text{nl}} + F_b \]  

The linear term is tri-linear in \( V_{S30} \) space, with natural log site amplification scaling linearly with the natural log of \( V_{S30} \) between corner velocities \( V_1 \) and \( V_2 \) and going through zero at \( V_{\text{ref}} = 760 \) m/sec (Table 2; Eq. 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, 2018a; Parker et al., 2019).

\[
F_{\text{lin}} = \begin{cases} 
  s_1 \ln \left( \frac{V_{S30}}{V_1} \right) + s_2 \ln \left( \frac{V_1}{V_{\text{ref}}} \right) & \text{if } V_{S30} \leq V_1 \\
  s_2 \ln \left( \frac{V_{S30}}{V_{\text{ref}}} \right) & \text{if } V_1 < V_{S30} \leq V_2 \\
  s_2 \ln \left( \frac{V_2}{V_{\text{ref}}} \right) & \text{if } V_{S30} > V_2 
\end{cases} 
\]  

(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:

\[
F_{\text{nl}} = f_1 + f_2 \ln \left( \frac{PGA_r + f_3}{f_3} \right) 
\]  

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; \( f_3 = 0.05g \) for all periods; and \( f_2 \) is given as (Chiou and Youngs 2008):

\[
f_2 = f_4 \left[ \exp \{ f_5(\min(V_{S30}, 760) - 200)) - \exp \{ f_5(760 - 200)\} \right] 
\]  

(10)

The basin depth term \( (F_b) \) depends on region, differential depth \( \delta Z_{2.5} \), and IM. Region options are Japan, and Cascadia, with further regionalization available for the Seattle basin, and Cascadia outside of basin the boundaries defined in Parker and Stewart (202x), but with a \( Z_{2.5} \) estimate. For the four specified regions, the basin depth function is:

\[
F_b = \begin{cases} 
  e_1 & \delta Z_{2.5} \leq \frac{e_1}{e_3} \\
  e_3 \delta Z_{2.5} & \frac{e_1}{e_3} < \delta Z_{2.5} < \frac{e_2}{e_3} \\
  e_2 & \delta Z_{2.5} \geq \frac{e_2}{e_3} 
\end{cases} 
\]  

(11)
where $e_1, e_2,$ and $e_3$ are dimensionless region- and period-specific model coefficients. For sub-regionalized Cascadia models, $e_2$ and $e_3$ can be modified using $\Delta e$. Differential depth is defined as:

$$\delta Z_{2.5} = \ln(Z_{2.5}) - \ln(\mu_{Z2.5}(V_{S30}))$$

where $Z_{2.5}$ is in units of m and the centering depth, $\mu_{Z2.5}$ (also in m), is conditioned on $V_{S30}$ using the functional form of Nweke et al. (2020), with separate coefficients for Cascadia and Japan:

$$\ln(\mu_{Z2.5}) = \ln(10) \times \theta_1 \left[ 1 + \text{erf} \left( \frac{\ln_{10}(V_{S30}) - \ln_{10}(\nu_{s})}{\sigma \sqrt{2}} \right) \right] + \ln(10) \times \theta_0$$

### ALEATORY VARIABILITY MODEL

The aleatory variability model represents the ground motion dispersion relative to the median model. The total standard deviation ($\sigma$) is related to between-event ($\tau$) and within-event variabilities ($\phi$) as,

$$\sigma = \sqrt{\tau^2 + \phi^2}$$

Our $\tau$ model is period-dependent but independent of source, path, and site parameters, and thus is simply given as period-dependent coefficients (Table E3).

Our $\phi$ model is dependent on $R_{rup}$ and $V_{S30}$ (Eqs. 15-17). Model coefficients for the aleatory variability are given in Table 2 and Table E3 in the electronic supplement of this article.
Eq. 16 describes a flat-ramp-flat trilinear relationship for variance as a function of distance that has period-independent corner distances \(R_1\) and \(R_2\) (Table 2). Eq. 16 reduces \(\phi^2\) 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/sec, is null for stiff sites \((V_{S30} \geq V_2 = 500\) m/sec), and has a linear transition between \(V_1\) and \(V_2\) controlled by \(\phi_V\).

For applications involving the use of non-ergodic site response, we partitioned the within-event variability \(\phi\) into site-to-site variability \((\phi_{S2S})\) and within-event single-station variability \((\phi_{SS})\) as (Al Atik et al., 2010):

\[
\phi = \sqrt{\phi_{S2S}^2 + \phi_{SS}^2} \quad (18)
\]

Models for \(\phi_{S2S}\) and \(\phi_{SS}\) are provided in Eqs. (19-20) and (21-23), respectively. The model components are formulated such that their sum is similar to the total within-event variance \(\phi^2\).

\[
\phi_{S2S}(V_{S30}, R_{rup}) = \phi_{S2S,0} + \Delta Var_{S2S}(V_{S30}, R_{rup}) \quad (19)
\]

\[
\Delta Var_{S2S}(V_{S30}, R_{rup}) = \begin{cases} 
    a_1 \ln \left( \frac{V_3}{V_M} \right) \left( \frac{\ln \left( \frac{R_4}{\max \left( R_3, \min \left( R_4, R_{rup} \right) \right)} \right)}{\ln \left( \frac{R_4}{R_5} \right)} \right) & V_{S30} \leq V_3 \\
    a_1 \ln \left( \frac{V_{S30}}{V_M} \right) \left( \frac{\ln \left( \frac{R_4}{\max \left( R_3, \min \left( R_4, R_{rup} \right) \right)} \right)}{\ln \left( \frac{R_4}{R_5} \right)} \right) & V_3 < V_{S30} < V_M \\
    a_1 \ln \left( \frac{V_{S30}}{V_M} \right) & V_M \leq V_{S30} < V_4 \\
    a_1 \ln \left( \frac{V_4}{V_M} \right) & V_{S30} \geq V_4
\end{cases}
\]

\[
\phi_{SS}(R_{rup}, V_{S30}) = \phi_{SS,0} + \Delta Var_{SS}(V_{S30}) \quad (21)
\]

\[
\phi_{SS}(R_{rup}) = \begin{cases} 
    \phi_{SS,1} & R_{rup} \leq R_5 \\
    \phi_{SS,2} + \phi_{SS,1} \left( \frac{\ln \left( \frac{R_{rup}}{R_5} \right)}{\ln \left( \frac{R_6}{R_5} \right)} \right) & R_5 < R_{rup} < R_6 \\
    \phi_{SS,2} & R_{rup} \geq R_6
\end{cases}
\]


\[ \Delta \text{Var}_{SS}(V_{30}) = \begin{cases} 
    a_2 \ln \left( \frac{V_3}{V_M} \right) \frac{\ln(R_4)}{\ln(R_3)} & \text{if } V_{30} \leq V_3 \\
    a_2 \ln \left( \frac{V_{30}}{V_M} \right) \frac{\ln(R_4)}{\ln(R_3)} & \text{if } V_3 < V_{30} < V_M \\
    a_2 \ln \left( \frac{V_{30}}{V_M} \right) & \text{if } V_M \leq V_{30} < V_4 \\
    a_2 \ln \left( \frac{V_4}{V_M} \right) & \text{if } V_{30} \geq V_4 
\end{cases} \]

In Eqs. (19-20), \( \phi_{SS}, a_1, \) and \( V_M \) are period-dependent model coefficients, and \( V_3, V_4, R_3, \) and \( R_4 \) are period-independent (Table 2). In Eqs. (21-23), \( \phi_{SS,1}, \phi_{SS,2}, a_2, \) and \( V_M \) are period-dependent model coefficients, and \( R_5 \) and \( R_6 \) are period-independent (Table 2).

**EPISTEMIC UNCERTAINTY MODEL**

We use a scaled backbone approach to represent epistemic uncertainty in the median GMM (Atkinson et al., 2014). In our approach, the median model is scaled up and down in ground-motion space uniformly with respect to all independent variables. This is achieved by varying the median model constant terms \( (c_0; \text{Eq. 1}) \) with an epistemic standard deviation \( \sigma_{e} \). We take \( \sigma_{e} \) as a function of period \( (T; \text{Eq. 24}) \), where model coefficients depend on event type and region. Model coefficients for Eq. 24 are given in Table E4 in the electronic supplement to this article.

\[ \sigma_{e}(T) = \begin{cases} 
    \sigma_{e1} & \text{if } T < T_1 \\
    \sigma_{e1} - (\sigma_{e1} - \sigma_{e2}) \frac{\ln(T/T_1)}{\ln(T_2/T_1)} & \text{if } T_1 < T < T_2 \\
    \sigma_{e2} & \text{if } T > T_2
\end{cases} \]

**Table 1.** Definitions of model input parameters and their units.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_{rup} )</td>
<td>Rupture distance; closest 3-dimensional distance to the rupture plane</td>
<td>km</td>
</tr>
<tr>
<td>( M )</td>
<td>Earthquake moment magnitude</td>
<td>unitless</td>
</tr>
<tr>
<td>( Z_{hyp} )</td>
<td>Hypocentral depth</td>
<td>km</td>
</tr>
<tr>
<td>( \bar{Z}_{hyp} )</td>
<td>Mean hypocentral depth given the depth to top of rupture ( (Z_{tor}) ). See Eqs. 25-26.</td>
<td>km</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
<td>Unit</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>$V_{S30}$</td>
<td>Time-averaged shear wave velocity in the upper 30m</td>
<td>m/sec</td>
</tr>
<tr>
<td>$PGA_r$</td>
<td>Estimate of peak ground acceleration at 760 m/sec reference condition</td>
<td>g</td>
</tr>
<tr>
<td>$Z_{2.5}$</td>
<td>Depth to the 2.5 km/sec shear wave velocity isosurface</td>
<td>m</td>
</tr>
<tr>
<td>$T$</td>
<td>PSA oscillator period</td>
<td>sec</td>
</tr>
</tbody>
</table>
### Table 2. Coefficient definitions, units, and values.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Model Component</th>
<th>Units</th>
<th>Global or Regional</th>
<th>Value</th>
<th>Coefficient</th>
<th>Model Component</th>
<th>Units</th>
<th>Global or Regional</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>Magnitude-independent geometrical spreading</td>
<td>Global</td>
<td>Tables E1 and E2</td>
<td>$\phi_1^2$</td>
<td>$\phi$ distance-scaling</td>
<td>Global</td>
<td>Table E3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b_4$</td>
<td>Magnitude-dependent geometrical spreading</td>
<td>Global</td>
<td>Tables E1 and E2</td>
<td>$\phi_4^2$</td>
<td>$\phi$ distance-scaling</td>
<td>Global</td>
<td>Table E3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_6$</td>
<td>Anelastic attenuation</td>
<td>Regional</td>
<td>Tables E1 and E2</td>
<td>$R_4$</td>
<td>$\phi$ distance-scaling corner</td>
<td>km</td>
<td>Global</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>$m_4$</td>
<td>Magnitude breakpoint</td>
<td>Regional</td>
<td>Table 3</td>
<td>$R_5$</td>
<td>$\phi$ distance-scaling corner</td>
<td>km</td>
<td>Global</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>$c_4$</td>
<td>Small-magnitude linear scaling</td>
<td>Global</td>
<td>Tables E1 and E2</td>
<td>$\phi_5^2$</td>
<td>$\phi$ VS30-scaling</td>
<td>Global</td>
<td>Table E3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_5$</td>
<td>Small-magnitude parabolic scaling</td>
<td>Global</td>
<td>Tables E1 and E2</td>
<td>$V_1$</td>
<td>$\phi$ VS30-scaling corner</td>
<td>m/s</td>
<td>Global</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>$c_6$</td>
<td>Large magnitude scaling</td>
<td>Global</td>
<td>Tables E1 and E2</td>
<td>$V_2$</td>
<td>$\phi$ VS30-scaling corner</td>
<td>m/s</td>
<td>Global</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>Source-depth scaling slope</td>
<td>Global</td>
<td>Tables E1 and E2</td>
<td>$\alpha_1$</td>
<td>$\phi$ VS30-scaling slope</td>
<td>Global</td>
<td>Table E3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D$</td>
<td>Source-depth scaling cap value</td>
<td>Global</td>
<td>Tables E1 and E2</td>
<td>$\alpha_2$</td>
<td>$\phi$ VS30-scaling slope</td>
<td>Global</td>
<td>Table E3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{11}$</td>
<td>Source-depth scaling corner</td>
<td>km</td>
<td>Global</td>
<td>20</td>
<td>$V_M$</td>
<td>$\phi_{225}$ and $\phi_{35}$ model reference $V_{s30}$</td>
<td>m/s</td>
<td>Global</td>
<td>Table E3</td>
</tr>
<tr>
<td>$d_{12}$</td>
<td>Source-depth scaling corner</td>
<td>km</td>
<td>Global</td>
<td>67</td>
<td>$V_3$</td>
<td>$\phi_{225}$ and $\phi_{32}$ $V_{s30}$-scaling corner</td>
<td>m/s</td>
<td>Global</td>
<td>200</td>
</tr>
<tr>
<td>$s_1$</td>
<td>Low-velocity $V_{s30}$-scaling slope</td>
<td>Regional</td>
<td>Tables E1 and E2</td>
<td>$V_4$</td>
<td>$\phi_{225}$ and $\phi_{32}$ $V_{s30}$-scaling corner</td>
<td>m/s</td>
<td>Global</td>
<td>800</td>
<td></td>
</tr>
<tr>
<td>$s_2$</td>
<td>Mid-velocity $V_{s30}$-scaling slope</td>
<td>Regional</td>
<td>Tables E1 and E2</td>
<td>$R_3$</td>
<td>$\phi_{225}$ and $\phi_{32}$ distance-scaling corner</td>
<td>km</td>
<td>Global</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>$V_1$</td>
<td>$V_{s30}$-scaling corner</td>
<td>m/s</td>
<td>Global</td>
<td>$R_4$</td>
<td>$\phi_{225}$ and $\phi_{32}$ distance-scaling corner</td>
<td>km</td>
<td>Global</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>$V_2$</td>
<td>$V_{s30}$-scaling corner</td>
<td>m/s</td>
<td>Global</td>
<td>$R_5$</td>
<td>$\phi_{32}$ distance-scaling corner</td>
<td>km</td>
<td>Global</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>$V_{rd}$</td>
<td>$V_{s30}$-scaling reference value</td>
<td>m/s</td>
<td>Global</td>
<td>760</td>
<td>$R_6$</td>
<td>$\phi_{33}$ distance-scaling corner</td>
<td>km</td>
<td>Global</td>
<td>800</td>
</tr>
<tr>
<td>$f_1$</td>
<td>Value of nonlinear site response for PGA, $&lt;&lt; 0.05g$</td>
<td>Global</td>
<td>0</td>
<td>$\sigma_{e1}$</td>
<td>Epistemic uncertainty</td>
<td>Regional</td>
<td>Table E4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_3$</td>
<td>Transition shaking intensity for nonlinear site effects begin</td>
<td>g</td>
<td>Global</td>
<td>0.05</td>
<td>$\sigma_{e2}$</td>
<td>Epistemic uncertainty</td>
<td>Regional</td>
<td>Table E4</td>
<td></td>
</tr>
<tr>
<td>$f_4$</td>
<td>Relationship between nonlinear site response model slope and $V_{s30}$</td>
<td>Global</td>
<td>Tables E1 and E2</td>
<td>$T_1$</td>
<td>Epistemic period corner</td>
<td>s</td>
<td>Regional</td>
<td>Table E4</td>
<td></td>
</tr>
<tr>
<td>$f_5$</td>
<td>Relationship between nonlinear site response model slope and $V_{s30}$</td>
<td>Global</td>
<td>Tables E1 and E2</td>
<td>$T_2$</td>
<td>Epistemic period corner</td>
<td>s</td>
<td>Regional</td>
<td>Table E4</td>
<td></td>
</tr>
<tr>
<td>$e_1$</td>
<td>Basin sediment-depth scaling</td>
<td>Regional</td>
<td>Tables E1 and E2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_2$</td>
<td>Basin sediment-depth scaling</td>
<td>Regional</td>
<td>Tables E1 and E2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_3$</td>
<td>Basin sediment-depth scaling</td>
<td>Regional</td>
<td>Tables E1 and E2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta_0$</td>
<td>Basin depth- $V_{s30}$ centering</td>
<td>Regional</td>
<td>Japan: 3.05 Cascadia: 3.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>Basin depth- $V_{s30}$ centering</td>
<td>Regional</td>
<td>Japan: -0.8 Cascadia: -0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_\mu$</td>
<td>Basin depth- $V_{s30}$ centering</td>
<td>m/sec</td>
<td>Regional</td>
<td>Japan: 500 Cascadia: 200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_\sigma$</td>
<td>Basin depth- $V_{s30}$ centering</td>
<td>Regional</td>
<td>Japan: 0.33 Cascadia: 0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EMPIRICAL 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., 2020b, 202x). The relational database can be used to extract a single summary flatfile with one line per recording for use in the development of GMMs. The subset of records used for model development were selected as follows, keeping records that met these criteria:

1. Metadata necessary for model development are available in the NGA-Sub database (Contreras et al., 2020, 202x; Ahdi et al. 2020, 202x) such as $M$, 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 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., 202x).
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. Pseudo-spectral acceleration 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., 202x);
9. Earthquakes without multiple event flags; these are events for which the recordings appear to indicate that more than one seismic source affected the ground motions;
10. Earthquakes with source review flags = 0, 1, 2 or 4, which indicate earthquakes that underwent quality control checks and meet 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 $M_{5-9.1}$, recorded at rupture distances of 15-1000 km, and the intraslab records are from events with $M_{4.5-8.3}$, 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 due to a lack of mainshock-aftershock classifications (criterion 3). Subduction events that have occurred since 2016 are not included in the database.

Figure 1. Magnitude-distance distribution of recordings from interface and intraslab events, color-coded by region.
DEVELOPMENT OF MEDIAN MODEL

Model development took place in a series of steps during which we progressively constrained model components in order 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 our reference site condition of $V_{S30} = 760 \text{ m/s}$ 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. Lastly, we iteratively adjusted the SS14 site response model, including a basin depth term where appropriate (Japan and Cascadia; Parker and Stewart 202x), and our global and regional constants. 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 Eq. 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 two potential geometric effects (Yenier and Atkinson, 2014; Rogers and Perkins, 1996): (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 typical offshore or deep locations of 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. Additionally, 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. We adopt 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 the present work. The simulations are performed with EXSIM, which
is an open-source stochastic finite-source simulation algorithm (Motazedian and Atkinson, 2005;
Boore, 2009; and Assatourians and Atkinson, 2012).

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 Eq. 4a (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 using matching
between empirical and simulated response spectra. As part of this work, they considered $h$ models
from YA14 and Atkinson and Silva (2000; shown in Figure 2), along with event-specific empirical
estimates of $h$ from YA14, Boore et al. (2014b), and earthquakes in the Christchurch, New
Zealand, sequence. They proposed a parameterization using Eq. 4a (Figure 2). Lastly, 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 ($M_{1.5–3.6}$). They applied a similar
method of fitting event-specific distance-scaling as was applied in YA14. As shown in Figure 2,
their results are consistent with the near-source saturation model of YA15 for global earthquakes.

Figure 2. Comparison of recommended near-source saturation models for interface and intraslab
earthquakes (Eq. 4, shown in red) with models constrained using active tectonic region data (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).

To constrain $h$ at large magnitudes, we performed EXSIM simulations for interface events with $M = 3.75–9.5$ in 0.25-magnitude unit intervals, with five runs per magnitude. Details on the simulations are provided in Parker et al. (2020). Ground motions were generated at various azimuths and for distances between 10 and 1000 km. Simple magnitude-independent path models were fit for each magnitude bin (e.g., Eqs. 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 peak ground acceleration (PGA) and velocity (PGV) increase with magnitude as shown in Figure 2 (similar trends were observed for other intensity measures). The resulting $h$ values are similar to the YA14, YA15, and AS00 models at moderate magnitudes, but are lower than the models for events with $M > 6.5$ (Figure 2).

Our near-source saturation model for interface events (Eq. 4a) applies to all intensity measures. As shown in Figure 2, we constrained the model to be similar to YA15 for $M \leq 5.5$, but to follow trends from EXSIM simulations for $M > 5.5$. For those larger magnitudes, the model is more similar to Atkinson and Silva (2000).

Residuals analyses using the near source saturation model of Eq. (4a) reveal that the intraslab GMM overpredicts median ground motions by a factor of ~1.3 for large intraslab events ($M_{6.5+}$) at short distances ($\leq 75$ km), indicating that the $h$ model given by Eq. (4a) saturates at rupture distances that are too small. Removing this bias requires increasing $h$ in the $M_{6.5-7}$ range for intraslab events. Once such adjustments were made, however, very large values of $h$ are given by the exponential function Eq. (4a) for larger values of $M$, and we consider these 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 $M_{7+}$ region (Figure 1), we examine existing simulation results for intraslab events from Ji and Archuleta (2018). We fit Eq. (4b) to the Ji and Archuleta (2018) response spectral values at 0.2-sec and 1.0-sec PSA for a $M_8$ intraslab earthquakes in Japan using alternate
fixed values of $h = 29$ km (value from YA15 for M8), 35 km, 40 km, 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 enforce this upper limit at the regional corner magnitude ($m_c$) used in the magnitude-scaling model. We re-fit Eq. (4b) over the magnitude range of $M_4 - m_c$ such that the resulting model predicts similar values of $h$ as YA15 at $M_4$ and $h = 35$ km at $m_c$ (Figure 2).

The regional corner magnitude $m_c$ (as used in Eq. 4b) in the function for $h$ was chosen in part to produce a smooth model in ground motion–magnitude space, so that the magnitude corner in $h$ and the magnitude-scaling corner in Eq. 5 are the same. Values of intraslab $m_c$ are based on the seismogenic thickness of subducting slabs (Ji and Archuleta 2018). As seismogenic thickness and $m_c$ increases, 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 Eq. 4b when $m_c$ is increased because $h$ is reduced for $M < m_c$. 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 $R^{-1}$, which implies a function of $\ln R$ in Eq. 2, given that the ground motion is modeled in natural log space. However, heterogeneities in the earth produce scattering, reflections, refractions, and wave-type conversions. As a result, the empirical exponent (i.e., $c_1$ in Eq. 2) is not -1.0. The transition from Fourier amplitude spectra (FAS) to response spectra (RS) introduces a magnitude-dependence in this term, since the response spectrum at a single period is sensitive to a range of Fourier periods (Yenier and Atkinson 2014; Hassani and Atkinson 2018a), which can be represented as $(b_3 + b_4 M) \ln R$. The $b_3$ term is not used in Eq. (2) as it is combined with $c_1$. The anelastic attenuation term represents the per-cycle energy dissipation; it is a property of the material through which the seismic wave is traveling. This term controls curvature in the decay of natural log ground motion with the natural 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 Eqn. (2) (similar to Joyner and Boore 1981, 1993, 1994). The geometrical spreading coefficients derived from this process ($c_1, b_4$) are the same across all magnitudes, but a preliminary event term ($\eta_E'$) was determined for each individual event and intensity measure (IM). These event terms, which have a trend with $M$, were considered in the subsequent derivation of source terms (see the Magnitude-Scaling section). Originally, we adopted values of $b_4$ from the simulation-based GMM of Hassani and Atkinson (2018a; HA18). However, we found that $b_4$ values from 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 the observations.

Despite the large size of the NGA-Sub database, it is not possible to constrain both the geometric 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 (e.g., Atkinson and Boore, 2003;., and Abrahamson et al., 2016).

With the geometrical spreading coefficients fixed, we fit the anelastic attenuation coefficient, $a_0$, as a random effect to produce both a global value and regional values. We smoothed $a_0$ with respect to period and constrained it to go to zero at 10 sec, 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 $a_0$. 
Figure 3. Comparison of the distance dependence in the global interface and intraslab PGV data corrected to $V_{S30} = 760$ m/sec using Eq. 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: $M = 6.64$).

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.

SOURCE MODEL

Magnitude-Scaling

Once the path model was set (Eqs. 2–4), we used preliminary event terms $\eta E'$ (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 $\eta E'$ is expected. To model this dependence, Eq. (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 of the magnitude-scaling for $M > m_c$. Lastly, 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; Archuleta and Ji, 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.

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).

<table>
<thead>
<tr>
<th>Region</th>
<th>Interface $m_c$ (Campbell 2020)</th>
<th>Intraslab $m_c$ (Ji and Archuleta 2018)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>7.90</td>
<td>7.60</td>
</tr>
<tr>
<td>Alaska</td>
<td>8.60</td>
<td>7.20</td>
</tr>
<tr>
<td>Aleutian Islands</td>
<td>8.00</td>
<td>7.98</td>
</tr>
<tr>
<td>Cascadia</td>
<td>7.70</td>
<td>7.20</td>
</tr>
<tr>
<td>Northern Central America &amp; Mexico</td>
<td>7.40(^6)</td>
<td>7.40</td>
</tr>
<tr>
<td>Southern Central America &amp; Mexico</td>
<td>7.40(^5)</td>
<td>7.60</td>
</tr>
<tr>
<td>Japan – Kuril-Kamchatka Trench (Pacific Plate)</td>
<td>8.50</td>
<td>7.65</td>
</tr>
<tr>
<td>Japan – Nankai-Ryukyu Trench (Philippine Sea Plate)</td>
<td>7.70</td>
<td>7.55</td>
</tr>
<tr>
<td>Northern South America</td>
<td>8.50</td>
<td>7.30</td>
</tr>
<tr>
<td>Southern South America</td>
<td>8.60</td>
<td>7.25</td>
</tr>
<tr>
<td>Taiwan</td>
<td>7.10</td>
<td>7.70</td>
</tr>
</tbody>
</table>

\(^6\) For 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.
Figure 5. Global interface and intraslab magnitude-scaling models ($F_M$; Eq. 5) and path model event terms ($\eta_{\text{in}}$, in ln units, residuals from Eq. 2) as a function of $M$ for 0.2 and 2.0-sec PSA. For plotting purposes, the recommended global $m_c$ values were used for the intraslab and interface model (Table 3).

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., $\mu_{\text{inY}} = c_0 + F_P + F_M + F_S$). Those event terms were examined for trends with hypocentral depth ($Z_{\text{hyp}}$). We select $Z_{\text{hyp}}$ 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 (e.g., Bilek and Lay, 1998; 1999; Baltay et al., 2019). See additional details in Parker et al. (2020).

We do not observe statistically significant trends in interface event terms with $Z_{\text{hyp}}$ (Figure 6). For intraslab events, we initially fit Eq. (6) to the event terms using a nonlinear 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 $m$ goes to 0 at the lower end of the depth range populated with data, 20 km. The depth adjustment increases over the range 20-67 km, which we interpret as an increase of stress drop with depth as also suggested by previous GMMs for active tectonic regions (Yenier and Atkinson, 2015a; Hassani and Atkinson, 2018a), stable continental regions (Yenier and Atkinson, 2015b), and induced earthquakes (Novakovic et al., 2018). The depth-dependence goes to zero for $T > 2.0$ sec.

Figure 6. Variation of path model and magnitude-scaling event terms (ln units) as a function of hypocentral depth at PGA, 0.2- and 1.0-sec PSA for interface and intraslab events. Best-fit depth scaling model (Eq. 6) shown for intraslab events. Straight line at 0 is a reference indicating no model bias.
We recognize that $Z_{hyp}$ is cumbersome in hazard applications because it requires randomization of the hypocenter location on the fault, which involves an additional loop in the hazard integral. Where possible, we recommend considering hypocenter location in an additional hazard analysis integral 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 plane, $\bar{Z}_{hyp}$ (Eqs. 25-26). The value of $\bar{Z}_{hyp}$ is fully determined once a fault rupture plane is defined and does not require an additional loop in the hazard integral. We define $Z_{dd}$ as the down-dip depth of hypocenters given other source parameters, $\theta_W$ as a normalized version of this depth, and $W$ as the fault rupture down-dip width (Figure 7):

\[
Z_{dd} = \frac{Z_{hyp} - Z_{TOR}}{\sin(dip)} \quad (25a)
\]

\[
\theta_W = \frac{Z_{dd}}{W} \quad (25b)
\]

\[
\bar{Z}_{hyp} = Z_{TOR} + \theta_W W \sin(dip) \quad (26)
\]

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.
Mai et al. (2005) used a database of 80 finite-fault models (Mai, 2004) to examine hypocenter positions within fault planes. They computed Z_{dd} for each earthquake, compared the normalized down-dip depths (θ_W) for strike–slip and dip–slip events, and looked for any trends with magnitude. We performed similar analyses using finite-fault models for 25 intraslab 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 planes. This compares favorably to the mean value from Mai et al. (2005) of 0.43 for dip–slip events (not only subduction).

When hypocentral depth Z_{hyp} is represented in Eq. (6) by its mean value (Z_{hyp}), no further modification to either the model coefficients or the between-event variability is required. However, in order to capture the effects of epistemic uncertainties in Z_{hyp}, alternate realizations of θ_W = \tilde{\theta}_{W} should be considered. We recommend a central logic tree branch for the mean of \tilde{\theta}_{W}=0.48 and additional branches for the mean ± a representation of θ_W variability (its 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 (Eq. 1), and final event terms, η_E, through a mixed-effects residuals analysis. Total residuals \( R_{ijk} = \ln(Y_{ijk}) - \mu_{nY} \) for earthquake i, station j and region k were computed using the GMM mean without a constant (i.e., \( \mu_{nY} = F_P + F_M + F_P + F_S \)). Before our subduction-specific site response model was developed (Parker and Stewart, 202x), \( F_S \) was taken from the Seyhan and Stewart (2014) model to develop preliminary estimates of regional model constants. This process was iterated as the site response model evolved, and the final sets of global and regional constants are developed from, and are compatible with, the recommended subduction-specific site response model.

The total residuals were partitioned into constants, c_{0,k}, for each region k, and event terms η_{E,i}, for each earthquake i using linear mixed effects in the R software environment (R Core Team, 2019; Bates et al., 2015):

\[
R_{ijk} = c_{0,k} + \eta_{E,i} + \delta W_{ij}
\]  

(27)
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 Archuleta and Ji (2018). In other regions (Central America and Mexico, Japan, and Taiwan), variations of the constant 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 $M < m_c$. The Japan-Philippine Sea Plate constant is set such that the Japan-Pacific Plate median ground motion is matched for $M < m_c$. The South America–northern region constant is set such that the South America–southern region median ground motion is matched for $M < m_c$. 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 $M > m_c$. The match at larger magnitudes was applied because such events are more hazard-critical than events with $M < m_c$.

We set the global constant to be compatible with the weighted average of regional median reference-rock ($V_{30} = 760$ m/sec) ground motions at the center of the data in the distance range $\leq 100$ km to avoid effects of regional anelastic attenuation differences. For interface events, this was set at $M7.0$ and $R_{rup} = 65$ km, and for intraslab events this was set at $M6.0$, $R_{rup} = 75$ 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) this 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 Model Behavior and Regional Comparison section.

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

DEVELOPMENT OF VARIABILITY AND UNCERTAINTY MODELS

**BETWEEN- AND WITHIN-EVENT ALEATORY VARIABILITY**

Aleatory variability represents the random variability of data relative to a model, and for ground motions is usually expressed using log-normal standard deviations. As indicated in Eq. (14), our GMMs include models for between-event variability ($\tau$) and within-event variability ($\phi$).

Prior studies for active tectonic regions (e.g., Gregor et al., 2014) show decreases in $\tau$ with magnitude. We investigated this for subduction zones by computing event terms and their standard deviations ($\tau$) from mixed-effects analyses for 0.5 $M$ units bins between $M_{4.5}$ and 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, our $\tau$ model is independent of $M$ and event type.
The peak near 0.1-sec occurs across the considered subduction zone regions and has also been observed for small magnitude shallow earthquakes in active tectonic regions (e.g., Boore et al. 2014a; Campbell and Bozorgnia 2014).

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

Using the same binning and inspection approach as for the between-event variability, we look for trends in total within-event variance ($\phi^2$) with magnitude, rupture distance, and $V_{S30}$. Variance rather than standard deviation is used to combine model terms (Eqs. 15-17). At short periods (e.g., PGA), no trend in $\phi^2$ with magnitude is apparent. At 1.0-sec PSA, $\phi^2$ for $M < 7$ and at $M_9$ are approximately equivalent, with an increase for intermediate magnitudes (approximately $M_7 - 8.75$), but with large uncertainty compared to values at lower magnitudes and at $M_9$ (see Figure 6.3 of Parker et al., 2020). As a result, we do not include magnitude terms in our $\phi^2$ model, which departs from previous findings for active tectonic regions (e.g., Boore et al., 2014a).

Parker et al. (2020) present a series of plots of binned $\phi^2$ with respect to distance and $V_{S30}$. Dispersion increases for distances beyond 200 km for PGA and other short-period parameters. We anticipate this is caused by complexities related to ground-motion attenuation that are not fully captured by regional terms in the path model, perhaps due 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/sec. This is likely related to site-response nonlinearity, which reduces the dispersion of site response. These effects are not observed at long periods (>1.0-sec PSA). Similar features have been observed previously for active tectonic regions.
We model $\phi^2$ using a piecewise function for within-event variance conditioned on $R_{rup}$ and $V_{S30}$ (Eqs. 15–17). 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/sec, the conditions where we expect soil nonlinearity to be the most prevalent. This minimum value is period-independent. Distance-dependence (Eq. 16) was evaluated by setting corner distances $R_1$ and $R_2$ based on visual inspection and then computing $\phi_1^2$ and $\phi_2^2$ 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 $\phi^2$. The data considered in the distance analyses had $V_{S30} \leq 500$ m/sec. In order to incorporate $V_{S30}$-dependence (Eq. 17), we selected corner velocities $V_1 = 200$ m/sec and $V_2 = 500$ m/sec by visual inspection, computed weighted variances in $V_{S30} \leq V_1$ and $V_{S30} > V_2$ bins, and then took $\phi_V^2$ 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 Eq. (17) 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-sec PSA.

**Figure 10.** Total within-event aleatory variability (in ln units) from direct model (black lines) and from combination of partitioned models for $\phi_{S2S}^2$ and $\phi_S^2$ using Eq. 15 (red lines), showing dependence on $R_{rup}$ and $V_{S30}$ for 0.2- and 1.0 sec PSA. Note that while we develop the model using variances, this figure shows the resulting models for standard deviations.
The models for between- and direct within-event variability models presented in Eqs. (14-17) are meant to be used in ergodic ground motion analyses. The next section discusses a partitioned within-event variability model that is required for partially nonergodic (i.e. site-specific) seismic hazard analyses (Stewart et al., 2017).

PARTITIONED WITHIN-EVENT ALEATORY VARIABILITY

In addition to the total within-event variance ($\phi^2$), we provide models for the partitioned components, $\phi_{s2s}^2$ and $\phi_{s3s}^2$. These were computed using mixed-effects analyses on binned datasets to estimate site terms, from which site term variances were computed ($\phi_{s2s}^2$) along with variances of remaining within-event residuals ($\phi_{s3s}^2$). Parker et al. (2020) present plots of both variances binned with respect to various predictor variables.

We start by visually inspecting trends of binned $\phi_{s2s}^2$ with respect to magnitude, rupture distance, and $V_{S30}$. We do not find a trend in $\phi_{s2s}^2$ with magnitude. Although we observe a slight trend in $\phi_{s2s}^2$ at large $R_{rup}$ for PGA and other short-period parameters, we do 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 ($\phi_{s2s}^2$). Parameter $\phi_{s2s,0}^2$ (Eq. 19) represents the distance-independent dispersion when $V_{S30}$ is not considered, and is computed from the weighted average of binned values between 50–200 km. Next, we examine the dependence of $\phi_{s2s}^2$ on $V_{S30}$ by subtracting $\phi_{s2s,0}^2$ from $V_{S30}$-binned values of $\phi_{s2s}^2$, and plot these differential variances at the median $V_{S30}$ for each bin (Figure 11). The four-segment model in Eq. (20) 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 $\Delta Var_{s2s}$ goes to zero at long periods where nonlinear effects in site response diminish. Because $\Delta Var_{s2s}$ 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 $R_{rup} > R_4$. 
Figure 11. Values of \( \phi_{SS}^2 \) (in ln units) with 95% confidence intervals from \( V_{S30} \)-binned site terms for PGA and 1.0-sec PSA. Solid line shows model for \( \Delta Var_{SS} \) (Eq. 20).

Within-event single-station variance \( \phi_{SS}^2 \) is 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 (\( \phi_{SS}^2 \)) with earthquake magnitude, rupture distance, and \( V_{S30} \). We model \( \phi_{SS}^2 \) as magnitude-independent. As shown in Figure 12, using data for all velocities, we observe no appreciable distance-dependence in binned values of \( \phi_{SS}^2 \) up to 500 km, and then a sharp 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 makes 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 \( \phi_{SS,2}^2 \) (Eq. 22, Figure 11). The physical basis for distance-dependent \( \phi_{SS}^2 \) is path-to-path variability as with \( \phi \). 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 \( R_{rup} < R_3 \) (\( \Delta Var_{SS} \)). A trilinear model 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 \( \phi_{SS,2}^2 \) (Eq. 23). We observed an increase in \( \phi_{SS}^2 \) for fast \( V_{S30} \) and a decrease for slow \( V_{S30} \) (Figure 12). The tri-linear function has corner velocities of 200 and 800 m/sec. As in the \( \phi_{SS}^2 \) model, because \( \Delta Var_{SS} \) is associated with soil nonlinearity, we applied its full effect for \( R_{rup} < R_3 \) and scaled this effect to zero for \( R_{rup} > R_4 \).
Figure 12. Values of \( \phi_{SS}^2 \) (in ln units) for PGA with 95% confidence intervals binned by distance and \( V_{S30} \). Solid lines show models for \( \phi_{SS}^2( R_{rup} ) \) (Eq. 22) and \( \Delta \text{Var}_{SS} \) (Eq. 23).

As given in Eq. 15, the sum of the within-event single-station (\( \phi_{SS}^2 \)) and site-to-site (\( \phi_{S2S}^2 \)) variances is equivalent to the total within-event variance (\( \phi^2 \)). In Figure 9, we compare the direct model for \( \phi \) (Eqs. 16-18) to the estimate from the component models using Eq. 15. For 0.2-sec PSA, the two models provide similar estimates of \( \phi \) for close distances (\( R_{rup} < 200 \) km) and soft soils (\( V_{S30} < 400 \) m/sec). 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 \( \phi \) model. At longer periods (e.g., 1.0-sec PSA) the models are similar for intermediate to large distances and (\( R_{rup} > 300 \) km) and stiff soil to rock site conditions (\( V_{S30} > 300 \) m/sec). At short distances and soft site conditions (\( V_{S30} = 200 \) m/sec, \( R_{rup} = 100 \) km), the partitioned model has larger estimates of variability by about 0.1 natural log unit.

For applications, we recommend the \( \phi_{SS}^2 \) model for partially non-ergodic seismic hazard analyses in which the \( F_{lin} \) and \( F_b \) terms (Eq. 7) are replaced with a site-specific model, and a site response epistemic uncertainty model is used (which should be less than \( \phi_{S2S}^2 \), 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 \( \phi \) model.
Uncertainty in Model Constants

We provide an epistemic uncertainty model for the median GMM that uses the scaled backbone approach (Atkinson et al., 2014), in which the median model can be scaled independently of predictor variables in ground-motion space by adjusting constant terms for each region. Thus the epistemic uncertainty is a constant shift in amplitudes that depends on event type, intensity measure and region, but is independent of magnitude, distance, and site condition. These uncertainties are represented by epistemic standard deviation $\sigma_e$, given by Eq. (24).

Different $\sigma_e$ values are estimated for interface and intraslab events (Figure 13). The global value of $\sigma_e$ is taken as the standard deviation of the regional median ground motions for the same conditions used to define the global constant: $M7.0$ and $R_{rup} = 65$ km for interface, and $M6.0$ and $R_{rup} = 75$ km for intraslab events. For all regions other than Cascadia, $\sigma_e$ 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 $\sigma_e$ 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:

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

(28)

where $c_4$ and $c_5$ are magnitude scaling coefficients (Eq. 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 Eq. (28), and take $\sigma_e$ 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, $\sigma_e$ for interface and intraslab PGA is 0.43 and 0.35, respectively. This translates to a 84th / 16th percentile range of
about 2.4 and 2.0 in ground motions. For all regions $\sigma_\varepsilon$ is larger at short periods and decreases by about 0.1 ln units at long periods (Figure 13). The 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 (Eq. 24) are given in Table E4 of the electronic supplement to this article. We refer to Gregor et al. (202x) for comprehensive recommendations on logic tree approaches when using the suite of NGA-Subduction GMMs.

**Figure 13.** Epistemic uncertainty ($\sigma_\varepsilon$; ln units) in regional and global model constants (open circles) and model fit recommended for application (solid lines) for interface and intraslab events.
**Other Considerations**

We recommend use of the scaled backbone approach, implemented through variations on constant terms via $\sigma_e$, as the principle means by which to capture epistemic uncertainties in the application of the GMMs presented in this paper. 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 $M > m_c$, 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 Eq. (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 Eq. (28) to maintain the same level of ground motion for $M < m_c$. 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 Campbell (2020) for interface events and Ji and Archuleta (2018) for slab events.

As described in the **Source Depth Scaling** section, if the mean hypocentral depth ($\bar{Z}_{\text{hyp}}$; Eqs. 25-26) is used in place of hypocentral depth as a directly applied independent variable, there is uncertainty in the mean depths that affects mean ground motions. This can be accounted for by varying the dimensionless down-dip location of the hypocenter $\theta_w$ relative to its mean value. Details are provided in the **Source Depth Scaling** section.

**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 ($\delta W_{ij}$), the site-to-site component of within-event residuals (also known as site terms, $\eta_{S,i}$) and between-event residuals (also known as event terms, $\eta_{E,i}$). Here we focus on $\delta W_{ij}$ and $\eta_{E,i}$ to evaluate overall model performance with respect to source and path parameters (Eq. 29).

$$R_{ij} = c + \eta_{E,i} + \delta W_{ij}$$  \hspace{1cm} (29)

A more complete presentation of residuals analyses results is presented in Parker et al. (2020). Site terms are examined further in Parker and Stewart (202x).

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

Event terms are shown as a function of $M$ for PGA, 0.2, 1.0, and 5.0 sec 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 $M$ range ($> \sim 2$ $M$ units) to judge model effectiveness, for example, interface Taiwan, Central America and Mexico, 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 sec for the interface and intraslab models (see Figures 4.13-4.14 of Parker et al., 2020).

Within-event residuals are shown as a function of distance for PGA, 0.2, 1.0, and 5.0 sec 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 reasonably flat. Site terms as a function of $V_{S30}$ are presented in the companion paper Parker and Stewart (202x).
Figure 14. Event terms as a function of moment magnitude for PGA and 0.2-, 1.0-, and 5.0-sec 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 filled symbol.

Figure 15. Within-event residuals as a function of rupture distance for PGA and 0.2-, 1.0-, and 5.0-sec PSA for (a) interface earthquakes and (b) intraslab earthquakes. Binned means are plotted as filled circles along with standard errors.
In this section we demonstrate the global model behavior and compare the various regional models. Comparisons with other NGA-Subduction models (Kuehn et al., 2020; Chiou et al., 2020; Abrahamson and Gulerce, 2020; and Si et al., 2020) and existing models for subduction zones (Atkinson and Boore, 2003; Zhao et al., 2006; Zhao et al, 2016a and b; Atkinson and Macias, 2009; and Abrahamson et al., 2016 and 2018) are provided by Gregor et al. (202x). 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., Atkinson and Boore, 2003; Abrahamson et al. 2016) the magnitude scaling was taken as the same 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-sec PSA for an M9 interface event with $V_{S30} = 760$ m/sec. Regional variations are smaller at short distances (< 200 km) and increase at larger distances (> 600 km). Even at short distances 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 separately. Figure 17b presents similar ground motion comparisons for an M8 intraslab event. It may be noted that regional variations appear less pronounced for large intraslab events relative to those for large interface events. This may be because an M8 intraslab scenario falls within the range of observations for many regions, whereas the M9 interface scenario is an extrapolation for all regions except Japan. One notable feature is that the Cascadia distance scaling shows stronger empirical anelastic attenuation than the other regions (also shown in Figure 4).
Figure 16. Pseudo-spectral acceleration (PSA) predictions for interface and intraslab earthquakes for $M_5$-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.

Figure 17. Regional comparisons of distance scaling of 1.0-sec PSA for $V_{S30} = 760$ m/sec for (a) $M_9$ interface events, and (b) $M_8$ intraslab event (CAM = Central America and Mexico; SA = South America). A hypocentral depth of 55 km was used for the intraslab analyses.
SUMMARY AND DISCUSSION

We describe the development of a horizontal-component ground motion model (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 attenuation coefficients ($a_0$). Site response is also regionalized (Parker and Stewart 202x), with regional $V_{30}$-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 $V_{30}$ 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. (2020a). Model coefficients for all model components are given in Tables 2, 3, and the electronic supplement to this article.

The GMM predicts PGA, PGV, and PSA at oscillator periods between 0.01–10.0 sec for interface and intraslab subduction-zone events. The reference condition of the GMM is $V_{30} = 760$ m/sec, but it can be applied to a range of site conditions using the site response model given herein and discussed in Parker and Stewart (202x). The interface model is valid for $M_{4.5–9.5}$, $R_{rup} = 20–1000$ km, $Z_{hyp} \leq 40$ km, and $V_{30} = 150–2000$ m/sec. The intraslab model is valid over $M_{4.5–8.5}$, $R_{rup} = 35–1000$ km, $Z_{hyp} = 20–200$ km, and $V_{30} = 150–2000$ m/sec. 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 backarcs (as defined in Contreras et al., 202x) 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 E5 of the electronic supplement. 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 U.S. and Canada, but we lack sufficient data to support model development for interface events (Figure 1). The situation is somewhat more well-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 (e.g., the 2001 M6.8 Nisqually earthquake, shown as a filled symbol in Figure 14b).

Aleatory variability models are developed that encompass both event types, with different coefficients for each intensity measure. Models are provided for four components of ground-motion variability: (1) between-event variability, $\tau$; (2) within-event variability, $\phi$; (3) single-station within-event variability, $\phi_{ss}$; and (4) site-to-site variability, $\phi_{s2s}$. 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 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 ($\sigma_c$). 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 to be accounted for.
ACKNOWLEDGMENTS

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 Pacific Earthquake Engineering Research center (PEER) or the Regents of the University of California. We thank Yousef Bozorgnia and PEER for organizing the NGA-Sub project. We thank the many researchers of the broader NGA-Sub team, and, in particular, Silvia Mazzoni, Sean Ahdi, Victor Contreras, and Tadahiro Kishida 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, and Annemarie Baltay for their reviews of this manuscript.

REFERENCES


Ji C., Archuleta R. (2018). Scaling of PGA and PGV deduced from numerical simulations of intraslab earthquakes, Department of Earth Science, University of California, Santa Barbara, CA.


Kuehn N., Bozorgnia Y., Campbell K.W., Gregor N. (2020); Partially non-ergodic ground-motion model for subduction regions using NGA-Subduction database, *PEER Report No. 2020/03*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA (headquarters).


PEER (2020). Data resources for NGA-Subduction project, *PEER Report 2020/02*, Y. Bozorgnia (PI), J.P. Stewart (ed.), Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA (headquarters).


