NGA-West2 Equations for Predicting PGA, PGV, and 5% Damped PSA for Shallow Crustal Earthquakes

David M. Boore,[a] Jonathan P. Stewart,[b] M.EERI, Emel Seyhan,[c] M.EERI, and Gail M. Atkinson,[d] M.EERI

We provide ground motion prediction equations for computing medians and standard deviations of average horizontal component intensity measures (IMs) for shallow crustal earthquakes in active tectonic regions. The equations were derived from a global database with **M** 3.0–7.9 events. We derived equations for the primary **M**- and distance-dependence of the IMs after fixing the V_{S30}-based nonlinear site term from a parallel NGA-West2 study. We then evaluated additional effects using mixed effects residuals analysis, which revealed no trends with source depth over the **M** range of interest, indistinct Class 1 and 2 event IMs, and basin depth effects that increase and decrease long-period IMs for depths larger and smaller, respectively, than means from regional V_{S30}-depth relations. Our aleatory variability model captures decreasing between-event variability with **M**, as well as within-event variability that increases or decreases with **M** depending on period, increases with distance, and decreases for soft sites. [DOI: 10.1193/070113EQS184M]

INTRODUCTION

Ground motion prediction equations (GMPEs) are used in seismic hazard applications to specify the expected levels of shaking as a function of predictor variables such as earthquake magnitude, site-to-source distance, and site parameters. GMPEs for active crustal regions are typically developed from an empirical regression of observed amplitudes against an available set of predictor variables (Douglas 2003, 2011).

In this paper, we present GMPEs developed as part of the NGA-West2 project (Bozorgnia et al. 2014). As with the other NGA-West2 GMPEs, we use the database described by Ancheta et al. (2014) in which ground motions are taken as the average horizontal component (as defined by Boore 2010) and the intensity measures (IMs) consist of peak ground acceleration and velocity (PGA, PGV) as well as 5% damped pseudo-spectral acceleration (PSA) for periods ranging from 0.01 s to 10 s.

[a] U.S. Geological Survey, MS 977, 345 Middlefield Rd., Menlo Park, CA 94025
[b] University of California, Los Angeles, CA, USA (corresponding author)
[c] Risk Management Solutions, Newark, CA (formerly UCLA Civil & Environmental Engineering Department)
[d] Western University, London, Ontario, Canada

Earthquake Spectra, Volume 30, No. 3, pages 1057–1085, August 2014;

We used a three-phase model building approach that strikes a balance between prediction accuracy and simplicity of form and application. Our philosophy was as follows: The primary variables that control ground motion at a site are earthquake moment magnitude **M** (the primary source variable), distance to the fault (the primary path variable), and time-weighted average shear-wave velocity over the upper 30 m of the profile V_{S30} (the primary site variable). After constraining site response and some additional effects based on an initial analysis of the data, in what we call Phase 1 of our study, we performed classical two-stage regressions to develop base-case GMPEs based on a simple functional form using just **M**, distance, and fault type (Phase 2). We then performed mixed effects analysis of residuals (defined by the difference, in natural log units, between the observed and predicted amplitude of motion) and examined trends of between- or within-event residuals against secondary predictor variables. These secondary parameters include the region in which the event occurs, whether the source is Class 1 or 2 (hereafter CL1 and CL2) per Wooddell and Abrahamson (2014; roughly main shocks or aftershocks, respectively), various source depths, and basin depth. We assessed the extent to which these additional variables improve the accuracy of the GMPEs in a way that is both statistically significant and practically meaningful. We implemented the inclusion of secondary variables, where warranted, as optional adjustment factors that may be applied to the base-case GMPEs. In this way, we aimed to ensure that our GMPEs are centered for the general case of future events in regions for which site-specific path and site parameters may be unknown.

Our GMPEs are generally applicable for earthquakes of **M** 3.0 to **M** 8.5 (except for lack of constraint for $\mathbf{M} > 7$ normal slip events), at distances from 0 to 400 km, at sites having V_{S30} in the range from 150 m/s to 1,500 m/s, and for spectral periods (T) of 0.01–10 s. We considered regional variability in source, path and site effects, but did not address directivity effects.

We first review prior work that helps establish the motivation for this research. We then discuss the data and predictor variables used in the analysis, which is followed by a section giving the recommended equations. Subsequent sections describe the derivation of the coefficients in the GMPEs, compare our results to prior studies, and give guidelines for application. Because of space limitations, some details of our results are deferred to Boore et al. (2013; hereafter BSSA13). Tables of coefficients can be found in Electronic Supplement materials. To facilitate application of the GMPEs, Fortran programs and routines for other software platforms such as Matlab and Excel can be found at http://www.daveboore.com/software_online.html and at the PEER website (http://peer.berkeley.edu).

PRIOR RELATED WORK

The research described in this paper and BSSA13 builds on the GMPEs of Boore and Atkinson (2008; BA08), which were part of the NGA-West1 project (Power et al. 2008 and references therein). The NGA-West2 project was formed to take advantage of new data available since NGA-West1, address some weaknesses in the NGA-West1 database, and allow the developers to reconsider their functional forms.

One improvement needed to the NGA-West1 equations involved adding data at small-to-moderate magnitudes. The need to enrich the database at low magnitudes to ensure robust magnitude scaling was highlighted by several studies (Atkinson and Morrison 2009,

Chiou et al. 2010, Atkinson and Boore 2011), and two of the NGA-West1 developers provided amendments to improve their equation performance at low magnitudes (Chiou et al. 2010, Atkinson and Boore 2011); the revised Boore and Atkinson GMPEs that account for this adjustment are referred to as BA08′. Studies by Scasserra et al. (2009), Atkinson and Morrison (2009), and Chiou et al. (2010) also pointed to the need to consider regional variability of path effects, as the attenuation of motions with distance is faster in some active regions than in others.

Finally, the richer database available for NGA-West2 allows us to improve on prior work by considering additional variables that could not previously be adequately resolved. However, we maintain the same basic functions for the equations as used in BA08.

DATABASE

DATA SOURCES

We use a NGA-West2 flatfile that contains site and source information, along with distance parameters and computed ground motion intensity measures (IMs; Ancheta et al. 2014). Various versions of the flatfile were used during GMPE development as the file evolved (details in BSSA13).

We used variable subsets of the data for different analysis phases. Consistent criteria (i.e., applied in all analysis phases) were applied with respect to the following:

- Availability of metadata: We required the presence of magnitude, distance, and site metadata in order to include a record in the analysis.
- Co-located stations: We do not use more than one record when multiple records from the same earthquake were recorded at the same site (e.g., in a differential array or different sensors at the same site).
- Single-component motions: We only use records having two horizontal-component recordings.
- Inappropriate crustal conditions: We exclude recordings from earthquakes originating in oceanic crust or in stable continental regions.
- Soil-structure interaction (SSI): We exclude records thought to not reasonably reflect free-field conditions as a result of SSI that potentially significantly affects the ground motions at the instrument. Our primary guide to stations not to be used was the Geomatrix 1st letter code, available in the flatfile, as indicated in Table 2.1 of BSSA13.
- Proprietary data: Data not publicly available are not used.
- Problems with record: Based on visual inspection, we exclude records with S-triggers, second trigger (i.e., two time series from the same event due to consecutive triggers), noisy records, or records with time step problems.
- Usable frequency range: We only use PSAs for periods less than the inverse of the lowest usable frequency, as specified in the flatfile; we did not exclude any records based on the high frequency-filter used in the record processing (Douglas and Boore 2011).
- Data were screened using **M**, distance, and recording-type criteria, as shown in Figure 1. These criteria are intended to minimize potential sampling bias, which

can occur at large fault distances where ground motions have low amplitudes and instruments may only be triggered by unusually strong shaking. Including such records would bias the predicted distance decay of the ground motion towards slower distance attenuation than is present in the real motions.

- An earthquake is only considered if it has at least four recordings within 80 km after applying the other selection criteria.

GROUND MOTION INTENSITY MEASURES

The ground motion IMs comprising the dependent variables of the GMPEs include horizontal component PGA, PGV, and 5% damped PSA. These IMs were computed using the RotD50 parameter (Boore 2010), which is the median single-component horizontal ground motion across all non-redundant azimuths. This is a departure from the GMRotI50 parameter used in BA08. Shahi and Baker (2014) describe how the maximum component can be computed from RotD50; Boore (2010) describe how RotD50 compares with GMRotI50.

We do not include equations for peak ground displacement (PGD), which we believe to be too sensitive to the low-cut filters used in the data processing to be a stable measure of ground shaking (details in Appendix C of Boore and Atkinson 2007).

PREDICTOR VARIABLES

The main predictor variables used in our regression analyses are moment magnitude **M**, R_{JB} distance (closest distance to the surface projection of the fault plane), site parameter V_{S30}, and fault type. Fault type represents the classification of events as strike slip (SS), normal slip (NS), or reverse slip (RS), based on the plunge of the *P*- and *T*-axes (see Table 2.2 in

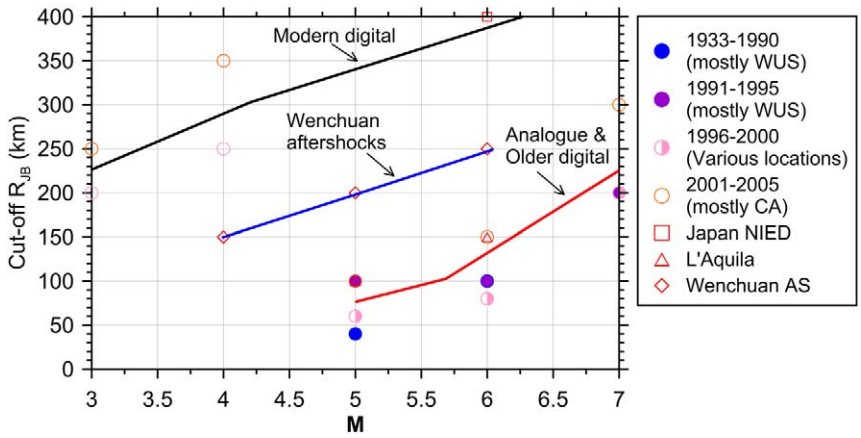

Figure 1. Magnitude- and distance-dependent cutoff criteria for using records. Data for a given earthquake with magnitude **M** were only considered for use if R_{JB} was less than the cutoff distance shown in the figure for that magnitude and type of recording. The symbols in the figure represent judgment-based cutoffs of data reliability. Symbols in figure developed from NGA-West2 GMPE developer interactions; lines represent our implementation.

BSSA13). Almost the same fault type assignments would be obtained using rake angle, with SS events being defined as events with rake angles within 30 degrees of horizontal, and RS and NS being defined for positive and negative rake angles not within 30 degrees of horizontal, respectively. Secondary parameters considered in residuals analysis include:

- Depth to top of rupture Z_{tor} and hypocentral depth Z_{hypo}.
- Basin depth z_1 (depth from the ground surface to the 1.0 km/s shear-wave horizon).
- Event type, being either Class 1 (CL1: main shocks) or Class 2 (CL2: aftershocks), using the minimum centroid R_{JB} separation of 10 km from Wooddell and Abrahamson (2014) based on subjective interpretation of results from exploratory analysis.

We did not consider hanging wall effects, as our use of the R_{JB} distance measure implicitly accounts for larger motions over the hanging wall (Donahue and Abrahamson 2014). Each of the predictor variables was taken from the NGA-West2 database flatfile.

DISTRIBUTION OF DATA BY M, R_{JB}, FAULT TYPE, AND V_{S30}

The **M** and R_{JB} distribution of data used to develop our GMPEs is shown in Figure 2, differentiated by fault type. There are many more small magnitude data than used in BA08, as well as data from a few new large events such as the 2008 **M**7.9 Wenchuan, China, earthquake. The magnitude range is widest for SS earthquakes and narrowest for NS earthquakes, suggesting that magnitude scaling will be better determined for SS than for NS—a problem we circumvented by using common magnitude scaling for all fault types.

Figure 3 shows the numbers of recordings and earthquakes used in equation development, differentiated by fault type. There is a decrease in available data for periods longer

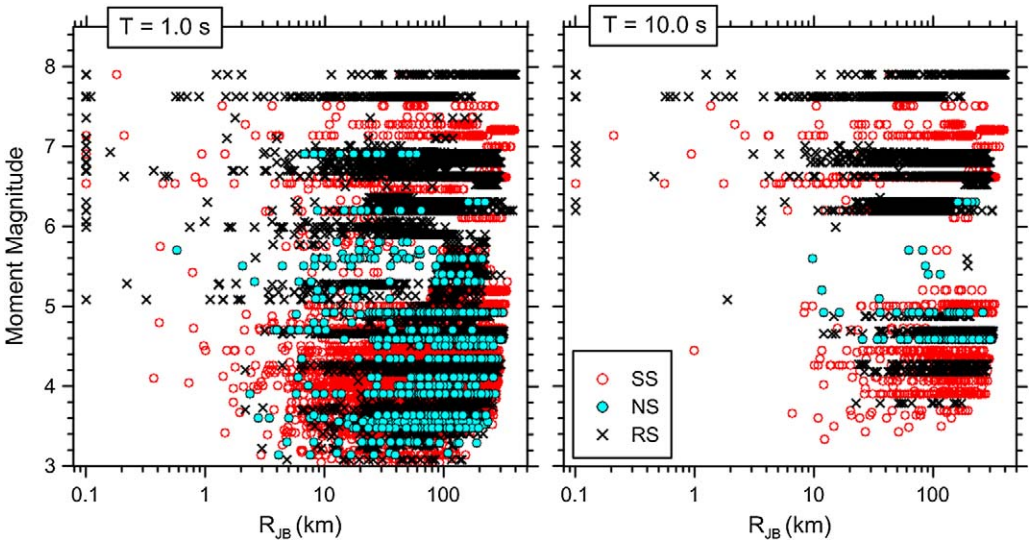

Figure 2. Distribution of data, according to fault type, used to develop the GMPEs. SS = strike-slip; NS = normal-slip; RS = reverse-slip.

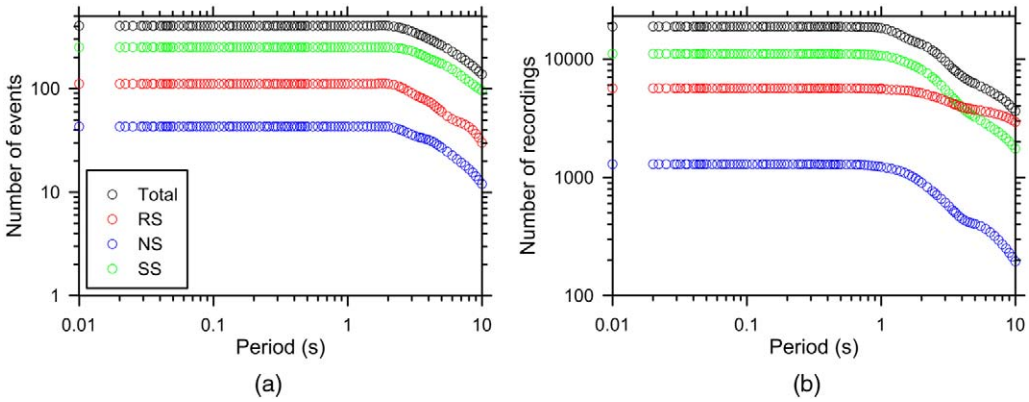

Figure 3. (a) Number of events and (b) recordings used to develop the GMPEs. The numbers are differentiated by fault type. SS = strike-slip; NS = normal-slip; RS = reverse-slip.

than several seconds, but there are many more data available at the longest periods than in BA08. Figure 4 shows the data distribution by V_{S30}; most of the data are for soil and soft rock sites (NEHRP categories C and D) but there are markedly more data for rock sites (mostly B) than in BA08. The V_{S30} data include measured and inferred velocities (Seyhan et al. 2014).

The data distributions over the predictor variable space necessarily influence the GMPEs. Note in particular the lack of data at close distances for small earthquakes. This

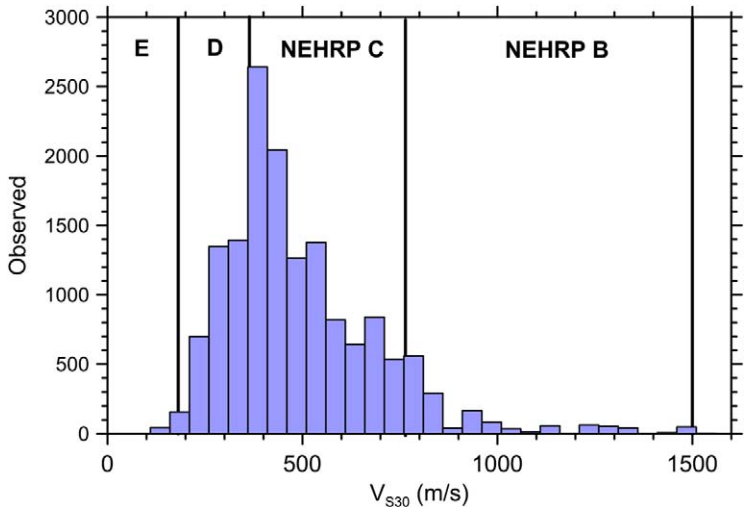

Figure 4. Histogram of V_{S30} for records used in deriving the GMPEs, with NEHRP site classes indicated by the vertical lines. Two records had V_{S30} values beyond the range of the plot corresponding to NEHRP class A (1,526 m/s and 2,016 m/s).

means that the near-source ground motions for small events will not be constrained by observations. In addition, there are many fewer small **M** data for long periods than for short periods, meaning that the small-earthquake **M** scaling will be less well determined for long periods.

THE GROUND MOTION PREDICTION EQUATIONS

As with Boore et al. (1997) and BA08, we sought simple functions for our GMPEs, with the smallest number of predictor variables required to provide a reasonable fit to the data. We call these the "base-case GMPEs." We subsequently derived adjustment factors for these base-case GMPEs to account for additional predictor variables. The selection of functions was heavily guided by subjective inspection and study of nonparametric plots of data such as in Figure 5, which shows the magnitude and distance dependence of PSA at four periods for strike slip events. The data have been adjusted to $V_{S30} = 760$ m/s, using the site amplification function of Seyhan and Stewart (2014; hereafter SS14). Inspection of these and similar plots revealed several features that the functions used for the GMPEs must accommodate: **M**-dependent geometric spreading; anelastic attenuation effects evident from curvature in the decay of log ground motions versus log distance for distances beyond about 80 km; and strongly nonlinear (and period dependent) magnitude dependence of amplitude scaling at a fixed distance.

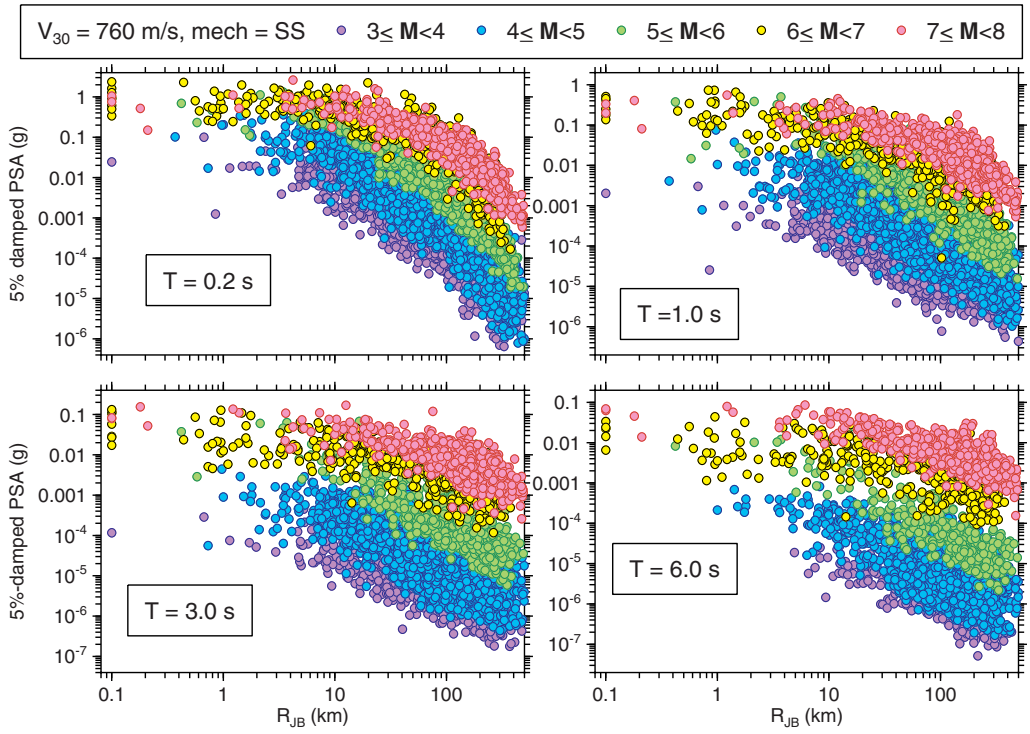

Figure 5. PSA at four periods for strike-slip earthquakes. All amplitudes adjusted to $V_{S30} = 760$ m/s using the soil amplification factors of this study.

Our predictions of ground motion are given by the following equation:

$$\ln Y = F_E(\mathbf{M}, mech) + F_P(R_{JB}, \mathbf{M}, region) + F_S(V_{S30}, R_{JB}, \mathbf{M}, region, z_1) + \varepsilon_n \sigma(\mathbf{M}, R_{JB}, V_{S30}) \quad (1)$$

where ln Y represents the natural logarithm of a ground motion IM (PGA, PGV, or PSA); F_E, F_P, and F_S represent functions for source ("E" for "event"), path ("P"), and site ("S") effects, respectively; ε_n is the fractional number of standard deviations of a single predicted value of ln Y away from the mean (e.g., $\varepsilon_n = -1.5$ is 1.5 standard deviations smaller than the mean); and σ is the total standard deviation of the model. The predictor variables are $\mathbf{M}$, *mech*, R_{JB} (in km), *region*, V_{S30} (in m/s), and z_1 (in km). Parameter *mech* = 0, 1, 2, and 3 for unspecified, SS, NS, and RS, respectively. Parameter *region* is 0 if no regional correction is to be made (default value); 1 for California, New Zealand, and Taiwan (this also provides no correction); 2 for China and Turkey; and 3 for Italy and Japan. The units of PGA and PSA are g; the units of PGV are cm/s.

Equation 1 is a combination of a base-case function and adjustments derived from analysis of residuals. These equations are given separately in BSSA13, but are combined here into a single equation.

ELEMENTS OF MEDIAN MODEL (SOURCE, PATH, AND SITE FUNCTIONS)

The source (event) function is given by:

$$F_E(\mathbf{M}, mech) = \begin{cases} e_0 U + e_1 SS + e_2 NS + e_3 RS + e_4(\mathbf{M} - \mathbf{M}_h) + e_5(\mathbf{M} - \mathbf{M}_h)^2 & \mathbf{M} \le \mathbf{M}_h \\ e_0 U + e_1 SS + e_2 NS + e_3 RS + e_6(\mathbf{M} - \mathbf{M}_h) & \mathbf{M} > \mathbf{M}_h \end{cases} \quad (2)$$

where U, SS, NS, and RS are dummy variables, with a value of 1 to specify unspecified, strike-slip, normal-slip, and reverse-slip fault types, respectively, and 0 if the fault type is unspecified; the hinge magnitude $\mathbf{M}_h$ is period-dependent, and e_0 to e_6 are model coefficients.

The path function is given by:

$$F_P(R_{JB}, \mathbf{M}, region) = [c_1 + c_2(\mathbf{M} - \mathbf{M}_{ref})] \ln(R/R_{ref}) + (c_3 + \Delta c_3)(R - R_{ref}) \quad (3)$$

where

$$R = \sqrt{{R_{JB}}^2 + h^2} \quad (4)$$

and c_1, c_2, c_3, Δc_3, $\mathbf{M}_{ref}$, R_{ref} and h are model coefficients. Parameter Δc_3 depends on the geographic region, as discussed later.

The site function is given by:

$$F_S(V_{S30}, R_{JB}, \mathbf{M}, region, z_1) = \ln(F_{lin}) + \ln(F_{nl}) + F_{\delta z_1}(\delta z_1) \quad (5)$$

where F_{lin} represents the linear component of site amplification, F_{nl} represents the nonlinear component of site amplification, and $F_{\delta z_1}$ represents the effects of basin depth.

Terms F_{nl} and $F_{\delta z_1}$ are region-dependent. Justification for the functional form of terms F_{lin} and F_{nl} is given in SS14.

The linear component of the site amplification model (F_{lin}) describes the scaling of ground motion with V_{S30} for linear soil response conditions (i.e., small strains) as follows:

$$\ln(F_{lin}) = \begin{cases} c\ln\left(\frac{V_{S30}}{V_{ref}}\right) & V_{S30} \le V_c \\ c\ln\left(\frac{V_c}{V_{ref}}\right) & V_{S30} > V_c \end{cases} \tag{6}$$

where c describes the V_{S30}-scaling, V_c is the limiting velocity beyond which ground motions no longer scale with V_{S30}, and V_{ref} is the site condition for which the amplification is unity (taken as 760 m/s). Parameters c and V_c are period-dependent, but not region-dependent (details in SS14). The function for the F_{nl} term is as follows:

$$\ln(F_{nl}) = f_1 + f_2\ln\left(\frac{PGA_r + f_3}{f_3}\right) \tag{7}$$

where f_1, f_2, and f_3 are model coefficients and PGA_r is the median peak horizontal acceleration for reference rock (for a given R_{JB}, **M**, and *region*, PGA_r is obtained by evaluating Equation 1 with $V_{S30} = 760$ m/s). Parameter f_2 represents the degree of nonlinearity as a function of V_{S30} and is formulated as:

$$f_2 = f_4[\exp\{f_5(\min(V_{S30}, 760) - 360)\} - \exp\{f_5(760 - 360)\}] \tag{8}$$

where f_4 and f_5 are model coefficients.

The term $F_{\delta z_1}$ is an adjustment to the base model to consider the effects of basin depth on ground motion amplitude. This adjustment is as follows:

$$F_{\delta z_1}(\delta z_1) = \begin{cases} 0 & T < 0.65 \\ f_6\delta z_1 & T \ge 0.65 \;\&\; \delta z_1 \le f_7/f_6 \\ f_7 & T \ge 0.65 \;\&\; \delta z_1 > f_7/f_6 \end{cases} \tag{9}$$

where f_6 and f_7 are model coefficients, f_7/f_6 has units of km, and δz_1 (also in km) is computed as:

$$\delta z_1 = z_1 - \mu_{z1}(V_{S30}) \tag{10}$$

where $\mu_{z1}(V_{S30})$ is the prediction of an empirical model relating z_1 to V_{S30}. For convenience, we give below relations for $\mu_{z1}(V_{S30})$ derived from data in California and Japan (modified from B. Chiou, *pers. communication*, 2013):

$$\text{California:}\quad \ln(\mu_{z1}) = \frac{-7.15}{4}\ln\left(\frac{V_{S30}^4 + 570.94^4}{1360^4 + 570.94^4}\right) - \ln(1000) \tag{11}$$

$$\text{Japan:}\quad \ln(\mu_{z1}) = \frac{-5.23}{2}\ln\left(\frac{V_{S30}^2 + 412.39^2}{1360^2 + 412.39^2}\right) - \ln(1000) \tag{12}$$

where μ_{z1} and V_{S30} have units of km and m/s, respectively. These relationships can be used to estimate a representative depth z_1 for any given V_{S30}. We realize that in many applications z_1 may be unknown; in such cases we recommend using the default value of $\delta z_1 = 0.0$, which turns off this adjustment factor (i.e., $F_{\delta z_1} = 0$). This is a reasonable default condition because the remaining elements of the model are "centered" on a condition of no $F_{\delta z_1}$ adjustment as described further below.

ALEATORY-UNCERTAINTY FUNCTION

The total standard deviation σ is partitioned into components that represent between-event variability (τ) and within-event variability (ϕ) as follows:

$$\sigma(\mathbf{M}, R_{JB}, V_{S30}) = \sqrt{\phi^2(\mathbf{M}, R_{JB}, V_{S30}) + \tau^2(\mathbf{M})} \tag{13}$$

The **M**-dependent between-event standard deviation τ is given by

$$\tau(\mathbf{M}) = \begin{cases} \tau_1 & \mathbf{M} \le 4.5 \\ \tau_1 + (\tau_2 - \tau_1)(\mathbf{M} - 4.5) & 4.5 < \mathbf{M} < 5.5 \\ \tau_2 & \mathbf{M} \ge 5.5 \end{cases} \tag{14}$$

and the **M**-, R_{JB}-, and V_{S30}-dependent within-event standard deviation ϕ is given by

$$\phi(\mathbf{M}, R_{JB}, V_{S30}) = \begin{cases} \phi(\mathbf{M}, R_{JB}) & V_{S30} \ge V_2 \\ \phi(\mathbf{M}, R_{JB}) - \Delta\phi_V \left(\frac{\ln(V_2/V_{S30})}{\ln(V_2/V_1)} \right) & V_1 \le V_{S30} \le V_2 \\ \phi(\mathbf{M}, R_{JB}) - \Delta\phi_V & V_{S30} \le V_1 \end{cases} \tag{15}$$

where

$$\phi(\mathbf{M}, R_{JB}) = \begin{cases} \phi(\mathbf{M}) & R_{JB} \le R_1 \\ \phi(\mathbf{M}) + \Delta\phi_R \left(\frac{\ln(R_{JB}/R_1)}{\ln(R_2/R_1)} \right) & R_1 < R_{JB} \le R_2 \\ \phi(\mathbf{M}) + \Delta\phi_R & R_{JB} > R_2 \end{cases} \tag{16}$$

and where the **M**-dependent ϕ is given by:

$$\phi(\mathbf{M}) = \begin{cases} \phi_1 & \mathbf{M} \le 4.5 \\ \phi_1 + (\phi_2 - \phi_1)(\mathbf{M} - 4.5) & 4.5 < \mathbf{M} < 5.5 \\ \phi_2 & \mathbf{M} \ge 5.5 \end{cases} \tag{17}$$

We recognize that this is a relatively complex form for aleatory uncertainty, being dependent on **M**, R_{JB}, and V_{S30}. We comment further on this subsequently in the paper.

DEVELOPMENT AND INTERPRETATION OF REGRESSION RESULTS

We developed our GMPEs in three phases. In Phase 1, we analyzed subsets of data and simulation results to evaluate elements of the model that would not be well-constrained if left as free parameters in the regression. Model elements evaluated in this way are c_3 (for apparent anelastic attenuation) and F_S (for site response), which are then fixed in subsequent

analysis phases. Phase 2 comprised the main regression for the base-case model. As in BA08, this was a two-stage regression, the first solving for path function coefficients, and the second stage solving for source function coefficients. Phase 3 consisted of mixed-effects regression analysis to check model performance and to develop adjustment factors for various secondary parameters beyond the base-case predictor variables of R_{JB}, **M**, *mech*, and V_{S30}. The standard deviation model was also developed from Phase 3 analysis.

As described by SS14, the V_{S30}-dependent site amplification model (terms F_{lin} and F_{nl}) was developed in an iterative manner with our GMPEs; that is, Phase 2 results were used as the basis for the analysis establishing the site coefficients (Phase 1), which were then used to redo the Phase 2 analysis, and so on.

PHASE 1: SETTING OF FIXED PARAMETERS

There were three parameters and/or functions determined in Phase 1 analyses that were held fixed in Phases 2 and 3: the apparent anelastic attenuation coefficient c_3 and the site amplification function. Development of the site amplification (not including the basin adjustment) is discussed by SS14 and is not repeated here. As such, in this section, we focus on the determination of the apparent anelastic attenuation coefficient c_3.

Due to trade-offs between apparent geometric spreading and apparent anelastic attenuation, regression cannot simultaneously determine both robustly; this arises because we cannot distinguish between the slope and the curvature of the distance decay from data with significant scatter. Accordingly, we undertook regressions to constrain the apparent anelastic attenuation term, c_3, using the large inventory of data from small events ($\mathbf{M} \leq 5.0$) in California, which are now included in the flatfile. Low-magnitude earthquakes were chosen to minimize complexities in the data associated with possible finite fault effects and nonlinear site effects. We applied the site factors from SS14 to adjust each observation to a reference V_{S30} of 760 m/s (denoted $\ln Y_{ij}$). The data were then grouped into bins 0.5 magnitude units in width and regressed using the following expression, which does not include an **M**-dependent geometric spreading term:

$$\ln Y_{ij} = \eta_i' + c_1' \ln(R/R_{ref}) + c_3(R - R_{ref}) \tag{18}$$

where η_i' is the event term for event i, j indicates a particular observation (and is implicitly contained in the distance R), $R_{ref} = 1.0$ km, and c_1' and c_3 are parameters set by the regression. The c_1' term represents the apparent geometric spreading for the **M** bin; it would be expected to change with **M**. The prime ($'$) is used on the event term and apparent geometric spreading term to indicate these are associated with the present analyses of binned data and are distinct from the Phase 2 and 3 regressions.

As suggested by the low **M** bins of Figure 5 (greater details are provided in Figure 4.1 of BSSA13), high-frequency IMs exhibit substantial curvature (indicating negative c_3), whereas the low-frequency data exhibit negligible curvature (nearly zero c_3). The regressions using Equation 18 found c_3 terms to be relatively independent of **M**, which is expected if they represent anelastic attenuation effects properly (**M**-dependence is contained in the c_1' term). These values of c_3 were used in subsequent phases of work.

PHASE 2: BASE-CASE REGRESSIONS

The objective of the two-stage Phase 2 analyses is to derive the coefficients for the F_P and F_E terms in Equation 1, with the exception of c_3 (from Phase 1) and Δc_3 (from Phase 3). The

analyses were performed using the two-stage regression discussed by Joyner and Boore (1993, 1994). In addition to the data selection criteria described previously, Stage 2 analyses also exclude CL2 events (aftershocks) and data with $R_{JB} > 80$ km (as described further below). Prior to Phase 2 regressions, we adjust all selected observations to the reference velocity of 760 m/s, using the SS14 site amplification model (excluding the basin term).

Stage 1 Analysis for Path Term

In Stage 1, path coefficients are evaluated by regressing observations ln Y_{ij} against the following base-case path relationship $F_{P,B}$ (the base-case excludes Δc_3):

$$F_{P,B}(R_{JB}, \mathbf{M}) = (\overline{\ln Y})_i + [c_1 + c_2(\mathbf{M} - \mathbf{M}_{ref})] \ln(R/R_{ref}) + c_3(R - R_{ref}) \qquad (19)$$

where $(\overline{\ln Y})_i$ represents average observations for event i adjusted to $R = R_{ref}$. As explained further in BSSA13, we set $R_{ref} = 1$ km and $\mathbf{M}_{ref} = 4.5$. With c_3 constrained, these regressions establish c_1, c_2, and h, as well as $(\overline{\ln Y})_i$ for each earthquake. Parameters c_1 and c_2 describe geometric spreading, with c_2 capturing its **M**-dependence. This process is complicated by the appearance of **M** in both the path and source function; the resulting coupling led to unrealistic results in initial Stage 2 analyses whereby the **M**-scaling was small, particularly for small **M**. Following the suggestion of K. Campbell (*pers. communication*, 2012), we recalculated Stage 1 regressions using data for $R_{JB} \leq 80$ km (still including fixed values of c_3), which stabilized the results.

Stage 2 Analysis for Source Term

In Stage 2, the $(\overline{\ln Y})_i$ terms from Stage 1 (subsequently referred to as $\overline{\ln Y}$) were used in weighted regressions to evaluate source terms e_0 to e_6, which control **M**-scaling and source type effects. The function for the source term (Equation 2), which was arrived at after many trials, consists of two polynomials hinged at $\mathbf{M}_h$; a quadratic for $\mathbf{M} < \mathbf{M}_h$; and a linear function for $\mathbf{M} > \mathbf{M}_h$. The hinge magnitude $\mathbf{M}_h$ was selected from visual inspection of many plots similar to those in Figure 6. Unlike BA08, which used a fixed $\mathbf{M}_h = 6.75$, we transition $\mathbf{M}_h$ from 5.5 for $T < \sim 0.1$ s to 6.2 for $T > \sim 0.4$ s. The major considerations in the selections of these values were ensuring robust scaling on both sides of the hinge, while allowing scaling behavior to be data-driven. Unlike BA08, we did not constrain the slope of the linear function (e_6) to be positive, resulting in negative slopes for high-frequency IMs, as shown in Figure 6. The negative slopes are suggestive of oversaturation, which has been implied by the data since NGA-West1 and supported by some simulations (Schmedes and Archuletta 2008). The negative slopes, however, do not result in IM decreases with increasing **M** at close distances (i.e., there is no true oversaturation), which can be understood as follows:

- Oversaturation in the source function occurs at $R = R_{ref} = 1$ km.
- Distance R (Equation 4) can never be less than the pseudo-depth h, but h is always greater than 4.0 km.
- Therefore, regardless of site location relative to the fault, the magnitude-dependent apparent geometrical spreading (Equation 3) affects predicted median motions. Those effects more than compensate for the apparent oversaturation from the source term (Equation 2).

The fault type coefficients in the Stage 2 regression were computed simultaneously with the coefficients describing **M**-dependence. The coefficients for unspecified fault type (e_0) were then computed as a weighted average of the SS, NS, and RS coefficients (e_1, e_2, and e_3; the weights are given in BSSA13).

Smoothing of Coefficients

Coefficients were obtained separately for each period using the above two-stage regression analysis. Resulting response spectra are generally smooth, with the exception of jaggedness for large and small magnitudes at periods exceeding about 2 s. This jaggedness is probably due to the decrease in the number of recordings available at longer periods (Figure 3). Accordingly, we undertook a smoothing process in which we first smoothed

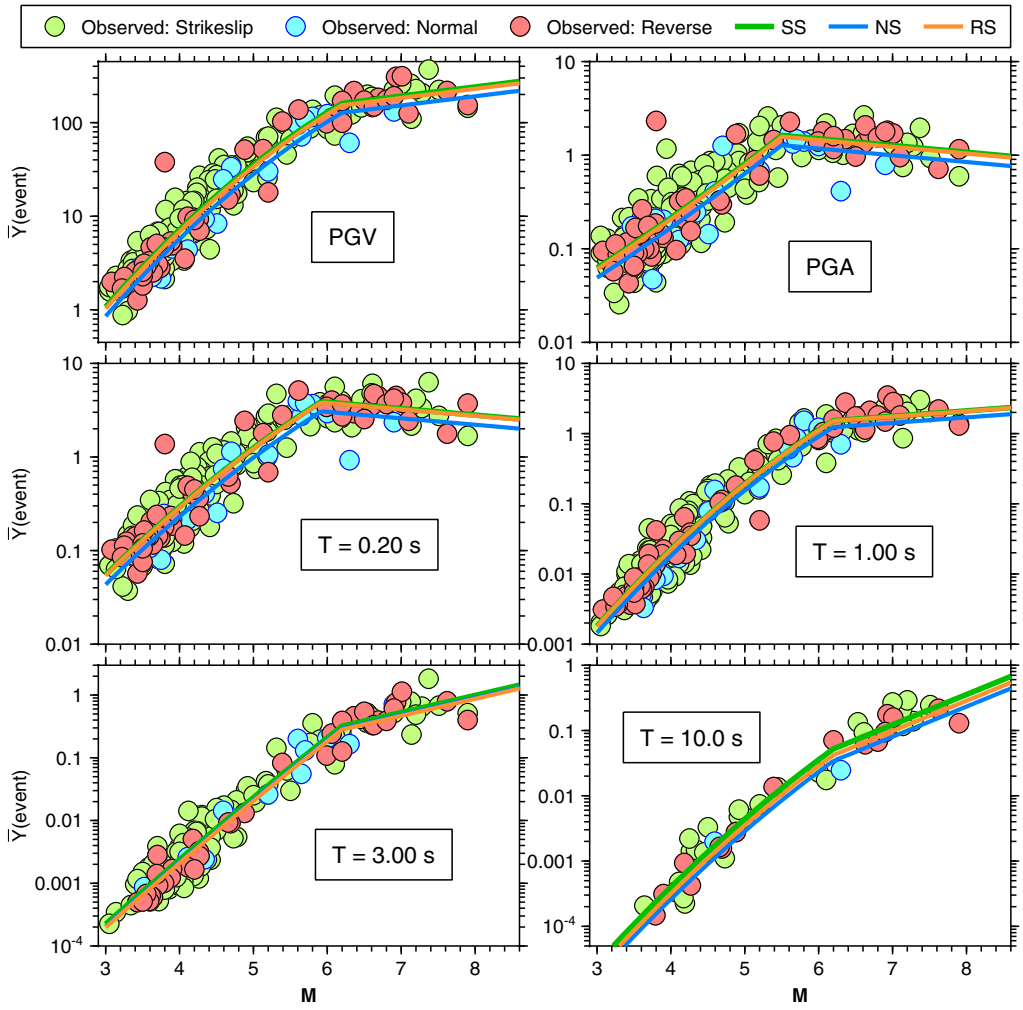

Figure 6. $\bar{Y}$ data points for each event and fit of **M**-scaling function.

the h parameter, re-regressed the model using those values, then computed 11-pt running means of the resulting coefficients.

Regression Results and Comparison to BA08′

Some well-known features of ground motion variations with the considered predictor variables are shown in plots of the base-case GMPEs in Figures 7 and 8. From the PGA plot in Figure 7, we see pronounced **M**-scaling of the motions for $\mathbf{M} < 6$, as shown by

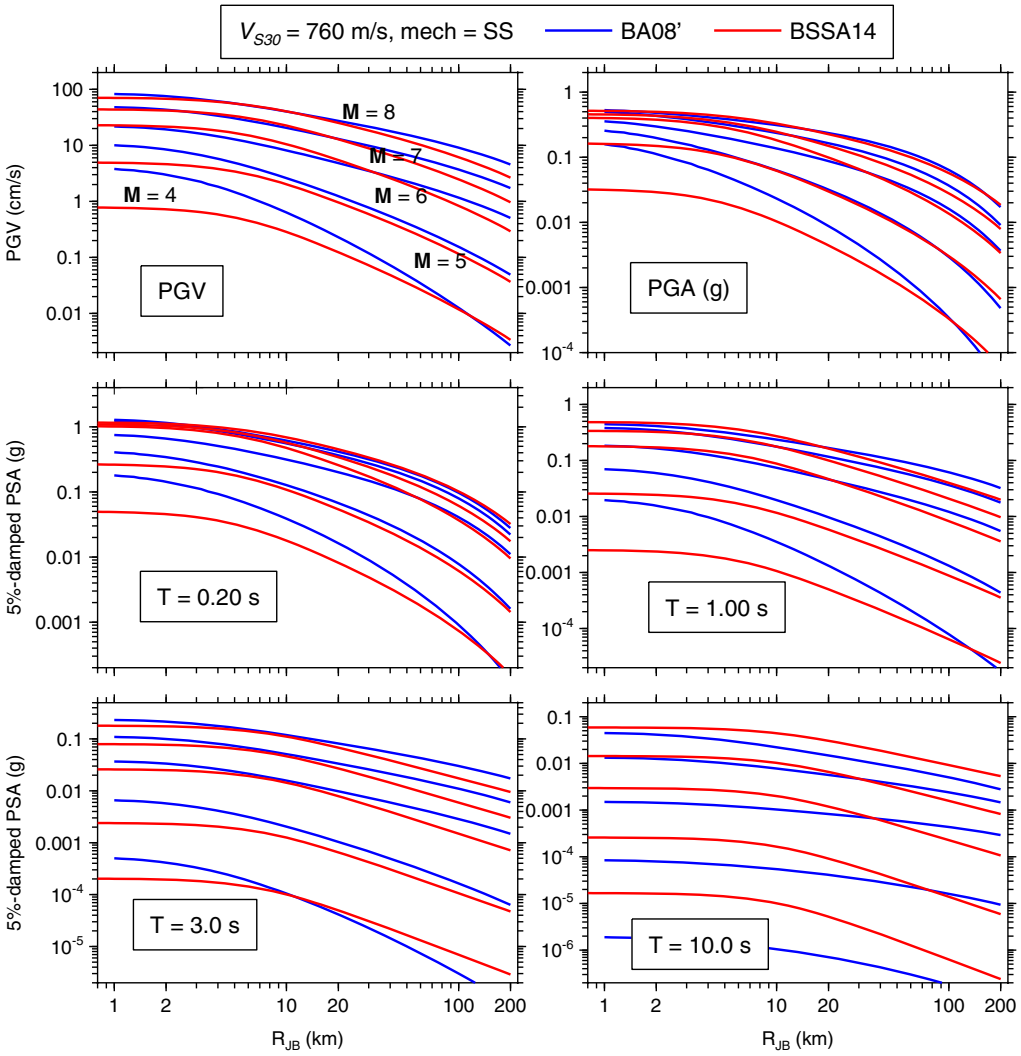

Figure 7. Median trends of proposed GMPEs as compared to BA08′, as a function of distance for strike-slip earthquakes and $V_{S30} = 760$ m/s. The BA08′ values have been adjusted to RotD50 using the ratios RotD50/GMRotI50 in Boore (2010; maximum adjustment of 1.06 for $T = 10$ s).

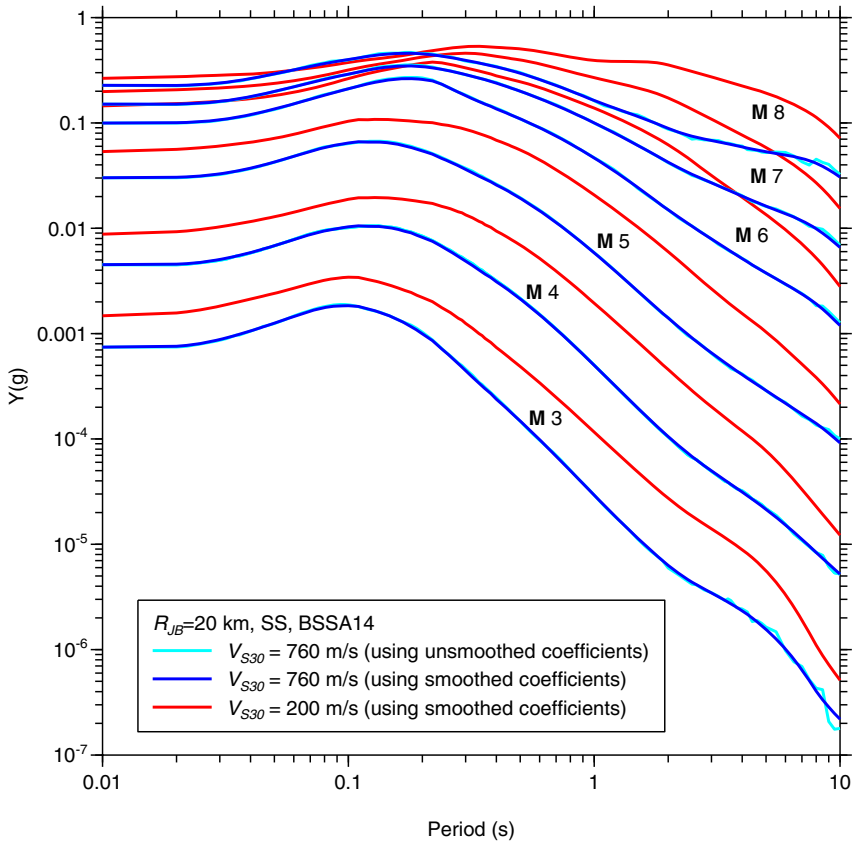

Figure 8. Variation of median predicted PSA versus period (T) for **M** 3, 4, 5, 6, 7, and 8 strike-slip earthquakes for $R_{JB} = 20$ km and $V_{S30} = 200$ m/s and 760 m/s.

the spread of the medians, but very little sensitivity at larger **M**. Conversely, for long-period PSA, we see substantial **M**-scaling over the full range of magnitudes. The distance attenuation trends in Figure 7 indicate amplitude-saturation, with amplitudes being nearly constant (flat) for distances $R_{JB} < \sim 3 - 5$ km. For $R_{JB} > \sim 10$ km, the distance attenuation transitions to a nearly linear slope for $T > \sim 1$ s, while for shorter periods, apparent anelastic attenuation produces downward curvature for $R_{JB} > \sim 80$ km. For short periods and $R_{JB} > \sim 10$ km, the curves steepen as **M** decreases due to **M**-dependent geometrical spreading that may be associated with finite-fault effects (i.e., as **M** increases, geometrical spreading from the small portion of a fault closest to the site is partially offset by contributions from other parts of the finite fault; Andrews 2001) and duration effects. For the same amount of energy radiated from a source, the longer the duration, the smaller the peak amplitudes of the ground motion. Duration is made up of source and path contributions, which combine through summation. Path duration generally increases with distance, and is relatively more important than the source duration for smaller events. Therefore, the ground motions for smaller events will attenuate more rapidly with distance than for large events (Boore 2003).

Also shown for comparative purposes in Figure 7 are median predictions of BA08′. The present medians are substantially smaller than those from BA08′ for small **M** at $R_{JB} < 10$ to 20 km and $T < \sim 3.0$ s. For $T = 10$ s the medians are now substantially larger. The two situations noted (small magnitudes and long periods) correspond to the conditions having the greatest data increase relative to BA08′.

In Figure 8, many of these features are illustrated by the relative positions of response spectra. This figure is useful in that it shows the **M** dependence of the predominant period (T_p), defined as the period of the peak in the spectrum. For rock-like sites, T_p ranges from approximately 0.1 s for **M** 3 to 0.2 s for **M** > 6. Note that for softer site conditions, corresponding to deeper sites, T_p increases to $\sim 0.4-0.5$ s due to site response effects. The effect of the aforementioned smoothing is also shown in Figure 8 by comparing median PSA from smoothed and unsmoothed coefficients.

The 2008 **M** 7.9 Wenchuan earthquake has been debated regarding its suitability for use in GMPE development. As shown in Figure 9, if the Phase 2 regressions are repeated without that event, short-period IMs are practically unchanged, but long-period IMs increase at large **M**. As we see no justification for excluding data from the Wenchuan earthquake, our GMPEs are those developed with the data including Wenchuan. We show the comparison in Figure 9 only to answer the inevitable question of what influence the Wenchuan data have on our GMPEs.

PHASE 3: RESIDUALS ANALYSIS

Phase 3 is comprised of mixed-effects residuals analyses having two purposes: (1) to check that the base-case GMPEs developed through the Phase 1 and 2 analyses are not biased with respect to **M**, R_{JB}, or site parameters; and (2) to examine trends of residuals against parameters not considered in the Phase 1 and 2 analyses, including regional effects.

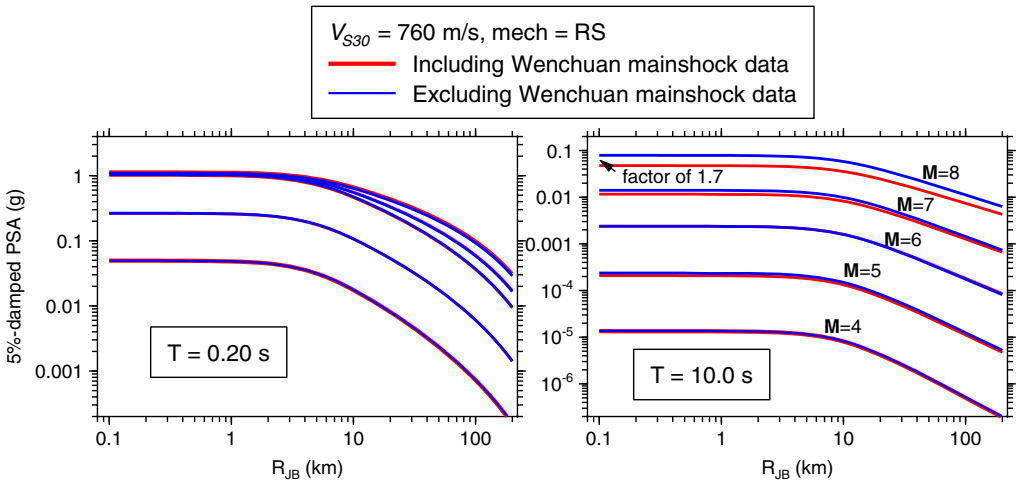

Figure 9. PSA versus R_{JB} for two periods, $V_{S30} = 760$ m/s, and reverse-slip (RS) events, from equations derived with and without the 2008 **M** 7.9 Wenchuan earthquake.

We used the data selection criteria given in the "Data Sources" section above, which differ from Phase 2 by including CL2 events and data with $R_{JB} > 80$ km (but still subject to the distance cutoffs given in Figure 1). This process resulted in some changes to the base-case equations, specifically, the introduction of the Δc_3 coefficient in the F_P term (for regional anelastic attenuation effects) and the $F_{\delta z_1}$ term for basin effects. The effects of CL2 source type and source depth were investigated but did not require model adjustments.

Methodology and Model Performance

The methodology for the analysis of residuals employed here is similar to that described in Scasserra et al. (2009). We begin by evaluating residuals between the data and the base-case GMPEs. Residuals are calculated as:

$$R_{ij} = \ln Y_{ij} - \mu_{ij}(\mathbf{M}, R_{JB}, V_{S30}) \tag{20}$$

Index i refers to the earthquake event and index j refers to the recording within event i. Term $\mu_{ij}(\mathbf{M}, R_{JB}, V_{S30})$ represents the base-case GMPE median in natural log units. Hence, R_{ij} is the residual of data from recording j in event i as calculated using the base-case GMPEs.

The analysis of residuals with respect to **M**, distance, and site parameters requires between-event variations to be separated from within-event variations. This is accomplished by performing a mixed effects regression (Abrahamson and Youngs 1992) of residuals according to the following function:

$$R_{ij} = c_k + \eta_i + \varepsilon_{ij} \tag{21}$$

where c_k represents a mean offset (or bias) of the data relative to GMPE k (we consider only the present set of GMPEs, so k is singular), η_i represents the event term for event i, and ε_{ij} represents the within-event residual for recording j in event i. Event terms are used to evaluate GMPE performance relative to source predictor variables, such as **M**. Event terms have zero mean and standard deviation $= \tau$ (natural log units). Within-event error ε has zero mean and standard deviation $= \phi$. Mixed-effects analyses with Equation 21 were performed using the NLME routine in program R (Pinheiro et al. 2013).

Checks of the base-case GMPEs using these residuals were undertaken by plotting η_i against **M** (to check the **M**-scaling in function F_E), η_i against rake angle (to check the focal mechanism terms in F_E), ε_{ij} against R_{JB} (to check the path function F_P), and ε_{ij} against V_{S30} (to check the site amplification terms). Plots for these effects are given in BSSA13 and support the following principal findings:

- There are some complexities in the plots of η_i with **M** when both CL1 and CL2 events are considered, which are largely related to CL2 aftershocks from China (Figure 4.28 of BSSA13). When these data are excluded, no trends are evident (Figure 4.29 of BSSA13), indicating adequate performance of the **M**-scaling function over the **M** range of 3 to 8.
- Plots of η_i against rake angle show zero bias, with the following exceptions—positive residuals for NS events with $\mathbf{M} < 5$ (the bias would have been zero without the focal mechanism adjustment in the model) and positive bias for $T > 1.0$ s PSA

for RS events with **M** > 5. We do not consider these trends to be sufficiently compelling to warrant adjustments to the model terms.

- There is no trend of ε_{ij} with distance up to 400 km when the full data set is used; this finding demonstrates that the California-based c_3 provides a good match to the global NGA-West2 data set (Figure 4.19 of BSSA13).
- There is no trend of ε_{ij} with V_{S30}, indicating satisfactory performance of the SS14 site amplification function. The lack of trend is present for the global data set (Figure 4.24 of BSSA13) and regional subsets.

Since a major region of application for this work will be California, we investigated the relative influence of non-California earthquakes on the GMPEs by examining CL1 (main shock) event terms (η_i) by region and fault type, as given in Figure 10. There is no overlap

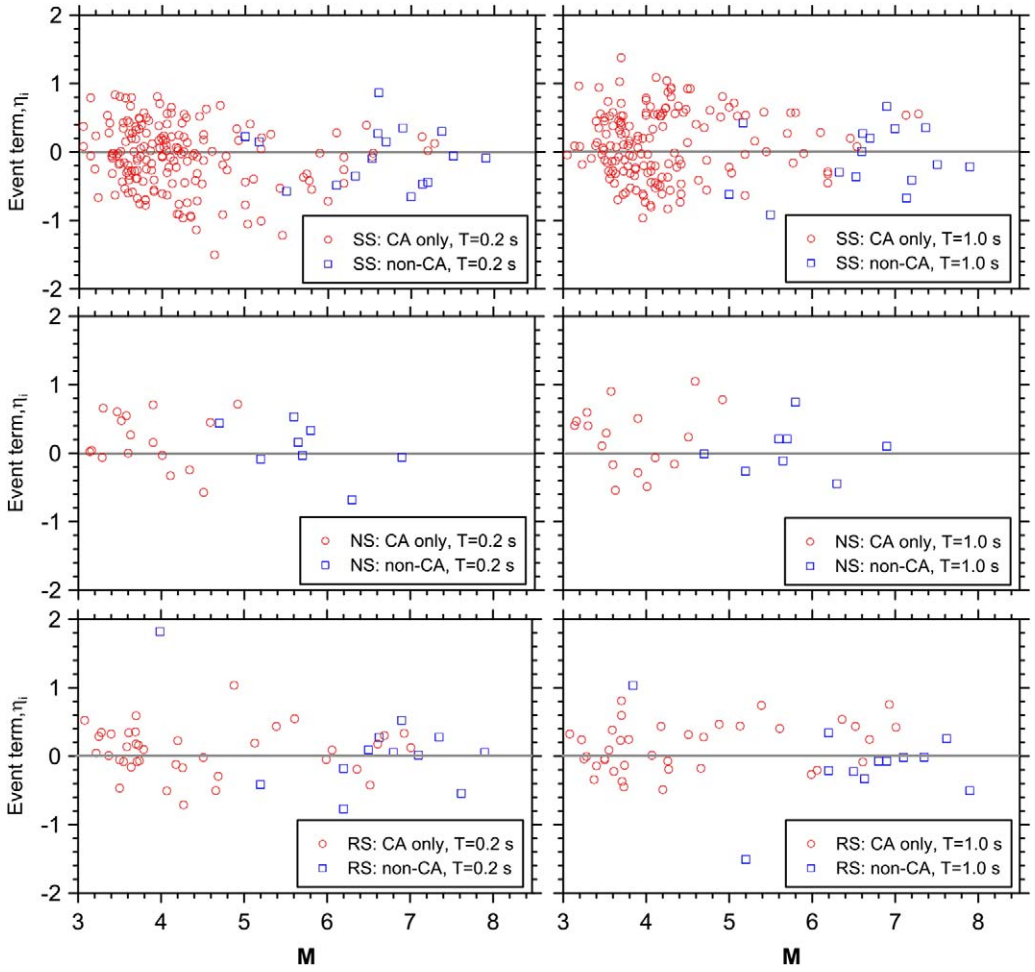

Figure 10. Between-event residuals for CL1 (main shock) events from Phase 3 analysis discriminated by fault type and CA versus non-CA regions. Note that there is almost no overlap in **M** for the CA and non-CA NS events, so it is difficult to judge possible bias for that condition.

in the magnitudes of the California and non-California NS events, so we cannot comment on regional differences in that case. For SS, the California and non-California event terms appear similar. For RS events at $T = 1.0$ s the California residuals are higher than the non-California events, at least for the larger magnitudes for which there is **M** overlap. This indicates that on average, the motions for California RS events may be under-predicted by our GMPEs for $T = \sim 1.0$ s (not shown here, a similar difference is suggested for longer periods). As our GMPEs are intended for global use, however, we have chosen not to provide a set only applicable to California.

Regional Adjustments to the Apparent Anelastic Attenuation Coefficient c_3

The lack of distance trends in the within-event residuals (ε_{ij}; bullet three above) demonstrates that the base-case path-scaling terms for apparent geometric spreading and apparent anelastic attenuation reasonably represent the global data. Recalling that the anelastic term c_3 was constrained from small **M** data in California, we plotted ε_{ij} against distance for various combinations of regions to investigate possible regional variations in crustal damping. As shown in Figure 11, we found California, New Zealand, and Taiwan to have flat trends relative to the global model (indicated as "Average Q"); Japan and Italy to have downward trends indicating faster distance attenuation (indicated as "Low Q"); and China and Turkey to have an upward trend (indicated as "High Q"). (We caution that country names are used as a convenient shorthand to describe the regions, realizing that results for the region may well be applicable beyond the political boundaries of the country and that regional differences of attenuation may occur within the countries. At this time, we do not have sufficient data to establish the geographic limits of our results nor to parse the data more finely.) Similarly flat distance attenuation trends for New Zealand were observed by Bradley (2013). Similarly fast distance attenuation trends have been observed in Japan and Italy (e.g., Stewart et al. 2013, Scasserra et al. 2009). The slower attenuation for China is not surprising given the location of the Wenchuan event near the western boundary of a stable continental region (Johnston et al. 1994), with the recordings being located in both stable continental and active crustal regions (Kottke 2011). This result for Turkey was not expected, but has been observed by others using larger Turkish data sets (Z. Gülerce, *pers. communication*, 2013).

For the Low and High Q cases, we fit a linear expression through the data in Figure 11 according to:

$$\varepsilon = \Delta c_3 (R - R_{ref}) + \bar{\varepsilon}_{lR} \tag{22}$$

where Δc_3 is the additive regional adjustment to the c_3 term from Equation 3 and $\bar{\varepsilon}_{lR}$ is the mean value of the residuals at close distance in a given region. In order to prevent the relatively sparse data at the closest distances from affecting the slope Δc_3, we limited the data range used in the regression to $R_{JB} > 25$ km; this includes a portion of the "flat" region in the residuals before anelastic effects become significant ($<\sim 80$ km), as well as larger distances to capture applicable data trends. Adjustments according to Equation 22 are plotted in Figure 11. Values of Δc_3 are given in the figure and are compiled as model coefficients. These regional adjustments are used in the computation of residuals for subsequent Phase 3 analyses.

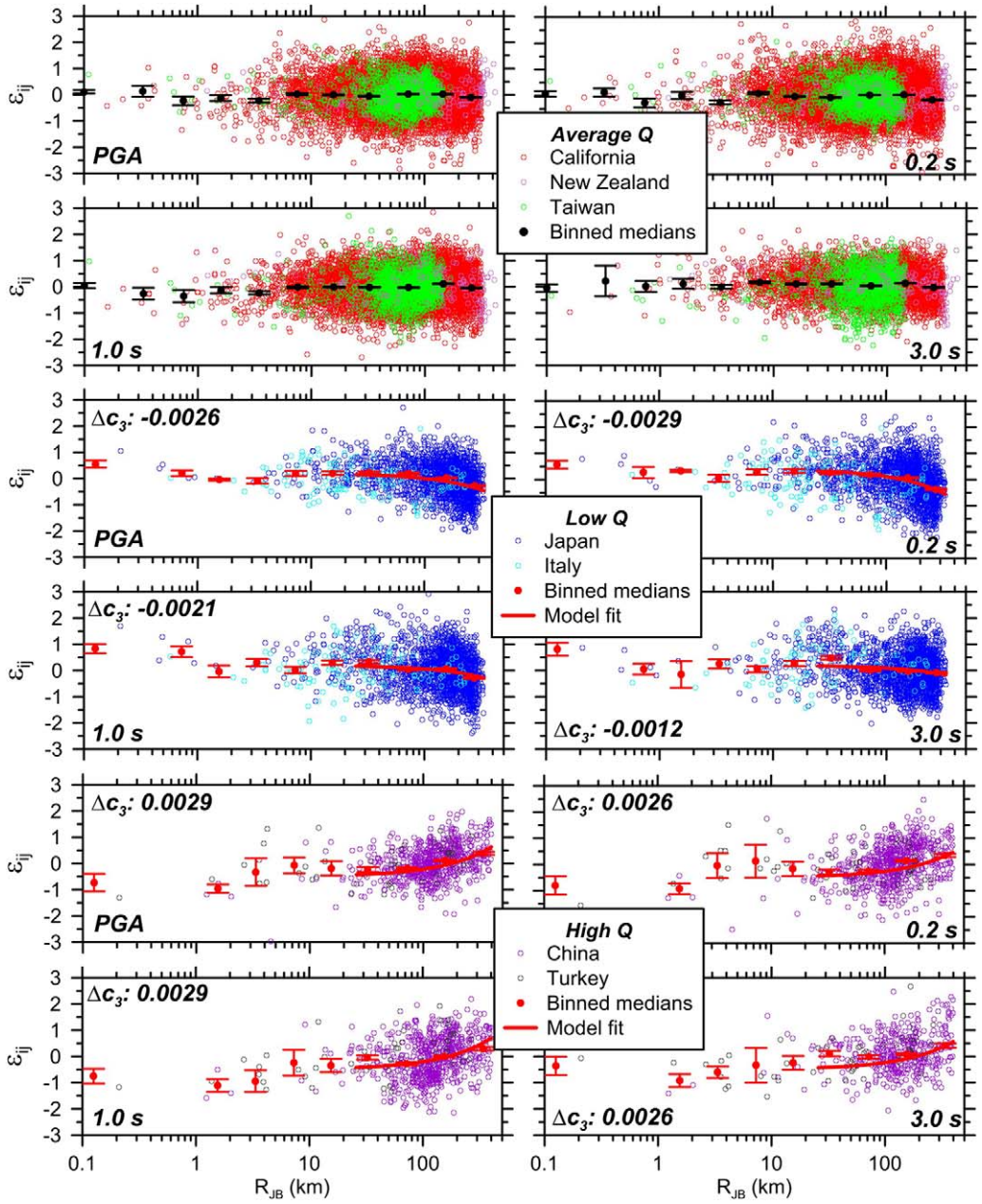

Figure 11. Within-event residuals for regions identified as "Average Q" (California, New Zealand, and Taiwan), "Low Q" (Japan and Italy), and "High Q" (China and Turkey). Also shown is the fit line per Equation 22 for $R_{JB} > 25$ km.

Adjustment of the Base-Case GMPEs for Sediment Depth

The site parameter V_{S30} strictly describes only the characteristics of sediments in the upper 30 m, even though it has been shown to be correlated to deeper structure (e.g., Boore et al. 2011). Several previous studies have found site amplification effects related to the depth of the deposit that are not captured by the use of V_{S30} alone as a site descriptor, including three GMPEs from the NGA-West1 project (Abrahamson and Silva 2008, Campbell and Bozorgnia 2008, and Chiou and Youngs 2008). Since residual trends in prior work are typically strongest at long spectral periods ($T > \sim 1.0$ s), the depth parameter is descriptive of low-frequency components of the site response, which may be related to resonances of sedimentary basin structures. The BA08 model did not include a basin-depth term; here we investigate whether the data support the use of such a term in the present equations.

As indicated in the "Predictor Variables" section, we consider basin-depth parameter z_1, which is the shallowest of the depth metrics considered in prior work. This choice is motivated by its greater practicality and lack of evidence (from Day et al. (2008) and this study) that deeper metrics are more descriptive of site amplification. We found stronger trends of residuals (ε_{ij}) against depth differential δz_1 (Equation 10) than basin depth itself (z_1), so δz_1 was adopted. In Figure 12, we plot residuals ε_{ij} against δz_1 along with the model fit from

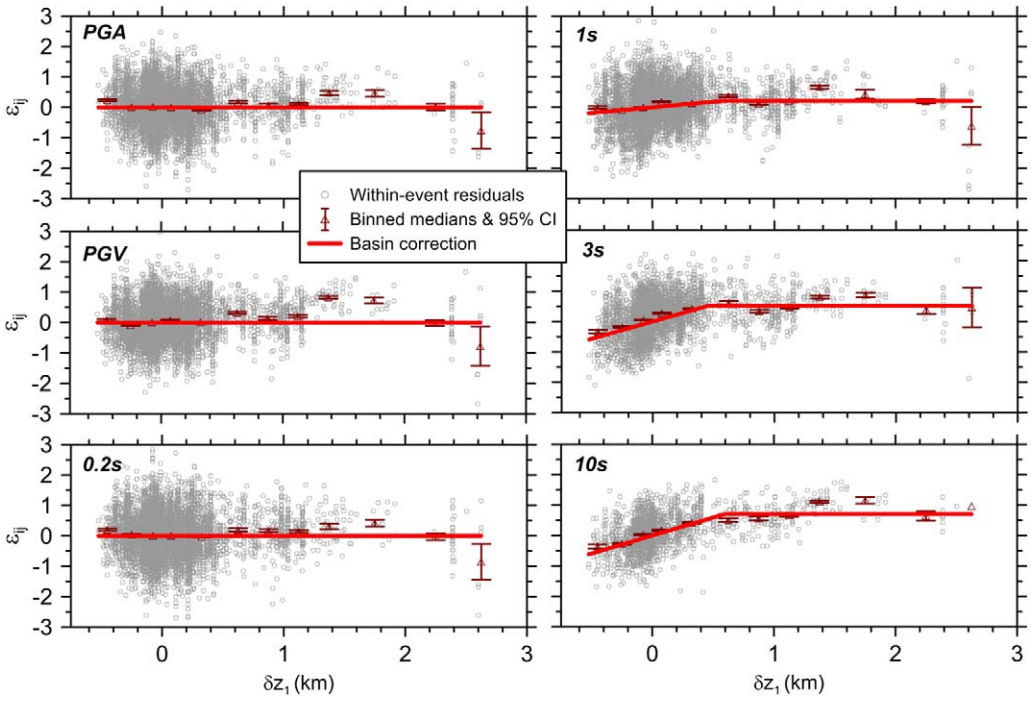

Figure 12. Within-event residuals against sediment depth differential δz_1 along with proposed basin model. Results plotted are for Southern California, the San Francisco Bay Area, and Japan.

Equation 9. At short periods there are no clear trends. For $T \geq 1$ s we find negative residuals for negative δz_1, a positive gradient for $\delta z_1 < 0.5$ km, and relatively flat trends thereafter. Model coefficients in Equation 9 represent the slope f_6 of the $\varepsilon_{ij} - \delta z_1$ relation and the limiting value f_7 of ε_{ij} for $\delta z_1 > 0.5$ km in the equation. The coefficients are regressed using all available data (from southern California, San Francisco Bay Area, and Japan); variations between regions were investigated and found to be modest and are not included in the recommended model.

Evaluation of Source Effects Using Between-Event Residuals

Event terms η_i derived from Equation 21 are used to investigate differences between main shock and aftershock motions (using the CL1 and CL2 designation) and effects of source depth (using depth to top rupture, Z_{tor}).

We found a modest correlation between η_i values from "parent" CL1 events and their "children" CL2 events (details in BSSA13). Accordingly, we examine differences between main shock and aftershock ground motions in the form of differences between CL1 event terms η_{CL1} and the mean of their "children" CL2 event terms $\bar{\eta}_{\mathrm{CL2}}$:

$$\Delta\eta = \bar{\eta}_{\mathrm{CL2}} - \eta_{\mathrm{CL1}} \tag{23}$$

There are 13 such pairs in our data set, and the resulting $\Delta\eta$ values are plotted in Figure 13 against the magnitude of the CL1 event. The results show no systematic departure of $\Delta\eta$ from zero, indicating that on average CL2 events do not have any more bias relative to the GMPEs than do their parent CL1 events. Accordingly, we consider our GMPEs equally applicable to both event types.

Because our distance metric R_{JB} is the closest horizontal distance of the site to the surface projection of the fault, rupture depth is not considered. This could conceivably lead to overprediction of motions from deep events (because the ground motions for such events have a longer travel path to reach recording sites), although such effects could possibly be offset by an increase in the value of the stress parameter with earthquake source depth (e.g., Fletcher et al. 1984). Figure 14 shows trends of η_i with Z_{tor} for CL1 and CL2 events with $\mathbf{M} \geq 5$,

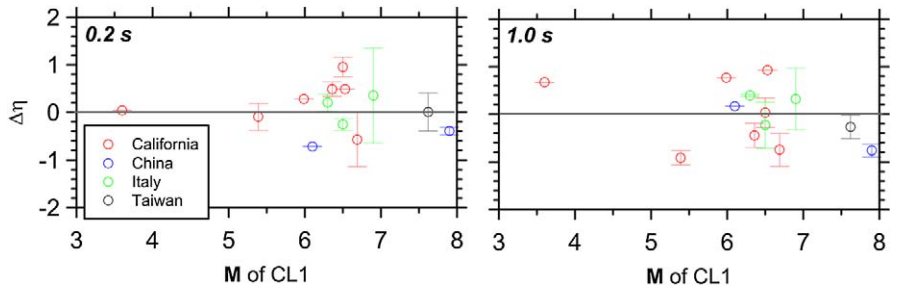

Figure 13. CL2 event term differential $\Delta\eta$ as function of CL1 magnitude (including binned means and their 95% confidence intervals). Results show no systematic offsets from zero, implying no increased level of misfit of CL2 "children" events as compared to their "parent" CL1 event.

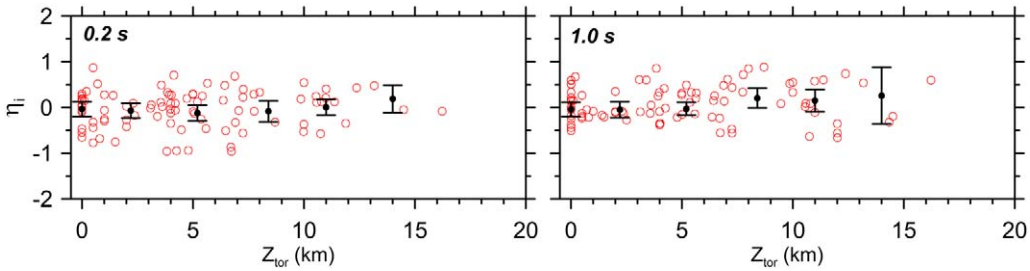

Figure 14. Event term variation with depth to top of rupture (Z_{tor}) for $\mathbf{M} \geq 5$ CL1 and CL2 events (including binned means and their 95% confidence intervals). Results show no significant offset from zero or trend, indicating lack of dependence of ground motion residuals on rupture depth.

which are essentially flat. Similar trends were observed for hypocentral depth. BSSA13 examined these trends by region, finding no effects, and for small magnitudes, for which trends were observed. Since the hazard for most engineering applications is governed by $\mathbf{M} \geq 5$ events, we have not included a source depth parameter in our GMPEs.

Aleatory Uncertainty Model

Our model for aleatory uncertainty (Equations 13–17) is derived on the basis of Phase 3 analyses, due to the relatively large database (as compared to Phase 2), which allows the standard deviation models to cover a broader range of **M**, R_{JB}, and V_{S30} than was used in Phase 2. Our approach was to bin event terms η_i by **M** to evaluate between-event standard deviation τ and bin residuals ε_{ij} by **M**, distance, and V_{S30} to enable evaluations of within-event standard deviation ϕ. The residuals analyses were performed using the median model given in Equations 1–9, including the anelastic attenuation and basin adjustments. Figure 15 shows example plots of binned values of τ and ϕ against the respective predictive variables for 0.2 and 1.0 s PSA (results for many additional periods were used to guide model development). Described further in BSSA13, the principal findings of this process are as follows:

- As shown in Figure 15a, τ decreases with **M**, but is nearly constant for $\mathbf{M} > 5.5$, which is the range of principal engineering interest. We account for this effect through Equation 14. Conversely, ϕ decreases with **M** at short periods, but increases with **M** for $T > 0.6$ s (Equation 17 and Figure 4.40 of BSSA13). Terms τ_1 and τ_2 and ϕ_1 and ϕ_2 are computed using all residuals (η or ε) within the respective **M** ranges (e.g., $\mathbf{M} < 4.5$ for τ_1 and $\mathbf{M} > 5.5$ for τ_2), not as the average of the binned values shown in Figure 15a.
- Figure 15b shows the R_{JB}-dependence of ϕ for $\mathbf{M} > 5.5$ residuals (indicated as ϕ_2). We see that ϕ increases with R_{JB}, but only beyond about 80 km to 130 km. Although only partially shown in the figure, at closer distances ϕ is approximately constant with respect to distance. We account for this effect through Equation 16. The increases in ϕ for $R_{JB} > 80 - 130$ km may reflect regional variability in anelastic attenuation. Thus, we expect that this increase is influenced by epistemic uncertainty in regional attenuation rates.

- Figure 15c shows the V_{S30}-dependence of ϕ for **M** > 5.5 and R_{JB} < 80 km. We see that ϕ decreases with V_{S30} at short periods (T < ~1.0 s), but only below about 300 m/s, which is captured by Equation 15. Results for many IMs in the format of Figure 15c show no trend of ϕ for V_{S30} > 300 m/s. Parameter $\Delta\phi_V$ is the difference in ϕ values computed from bins of ε for $V_{S30} > V_2$ and $V_{S30} < V_1$. We attribute the lower ϕ values for soft sites to nonlinear site response, which amplifies weak motions and de-amplifies strong motions, thus reducing ϕ relative to underlying reference site conditions (Choi and Stewart 2005).

Seismic hazard analysis for active crustal regions will most often be controlled by **M** > 5.5 events at R_{JB} < 80 km and V_{S30} > 300 m/s. These conditions correspond to $\tau = \tau_2$ and $\phi = \phi_2$, which are plotted in Figure 16, along with the standard deviation terms provided in BA08. While the ϕ terms are similar, the τ terms have increased notably relative to BA08 for T < 0.2 s and decreased for T > 1.0 s. Reductions from the tabulated ϕ values can be made if site response (and potentially path effects) are evaluated on a site-specific basis (Atkinson 2006, Al Atik et al. 2010). Moreover, the actual variability may be smaller than that obtained from these regression statistics if applications are for a more controlled set of region and site conditions. For example, Atkinson (2013) finds that aleatory variability for ground motions recorded on rock sites (V_{S30} > 1,000 m/s) in a region of eastern North America is significantly lower than the values obtained here—presumably because of the restriction in both region and site condition of the included data.

We recognize that the proposed standard deviation model is more complex than in BA08, which was necessitated by the significant expansion of the data set in

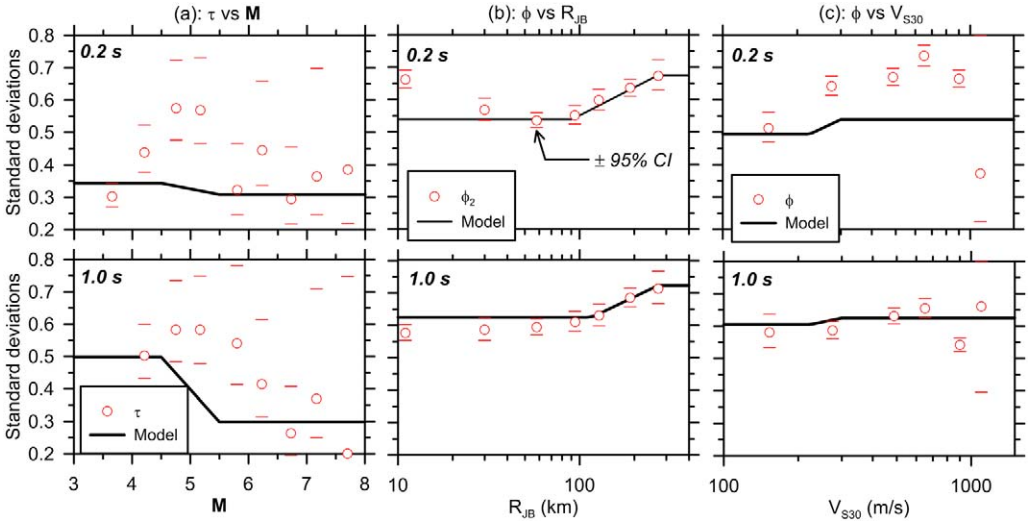

Figure 15. Binned standard deviation terms and their 95% confidence intervals from Phase 3 analyses showing: (a) Between-event standard deviation τ against **M**; (b) within-event standard deviation ϕ against R_{JB} for **M** > 5.5, and (c) ϕ against V_{S30} for **M** > 5.5 and R_{JB} < 80 km. Lines in the figures represent model fits per Equations 14 to 17.

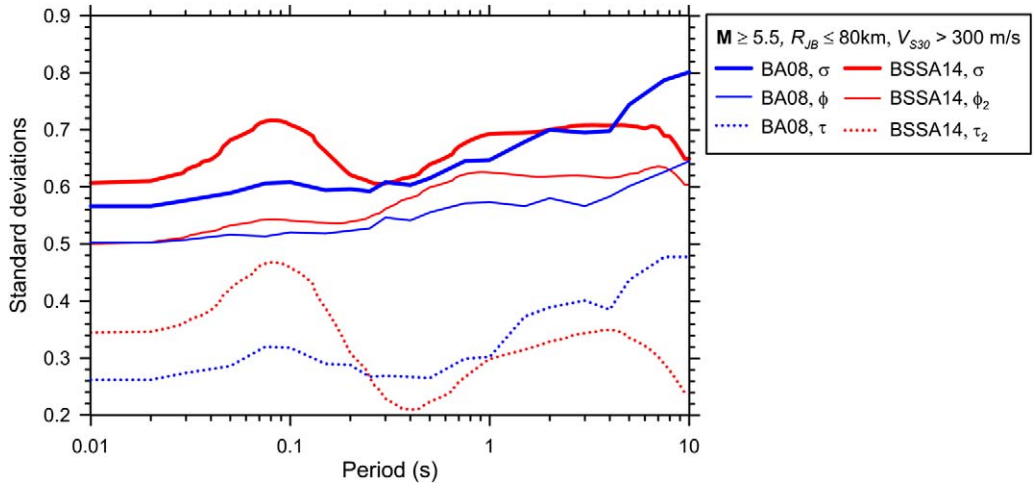

Figure 16. Comparison of the standard deviation terms in the BSSA14 GMPEs for **M** ≥ 5.5, R_{JB} < 80 km and V_{S30} > 300 m/s with the **M**- and R_{JB}-independent standard deviations from BA08.

NGA-West2, including a wider range of regions, magnitudes, and distances. Had we kept **M**-, R_{JB}-, and site-independent ϕ and τ, our standard deviations would be high relative to BA08 and the other NGA-West2 GMPEs. This increase is largely due to the substantial amount of small **M** data introduced to the data set. The additional complexity in the standard deviation model is worthwhile given its strong effect on PSHA results (Bommer and Abrahamson 2006).

As shown in Figure 16, there is a bump (i.e., increase) in the standard deviations centered near 0.08 s, which is controlled by the τ component. This bump is stable when the data is parsed in various ways, including CL1-only events, events in various **M** ranges (<4.5, ≥5.5), and events in various regions, with one exception. The exception is California CL1 events with **M** > 5.5, for which the bump is absent and the τ_2 terms are lower (i.e., nearly 0.24 for T < 0.5 s) than those given by our model. Due to the relatively modest number of large-**M** California events, and persistence of the bump for all other conditions, we have retained the bump in our model.

We believe that there are physical justifications for the bump: PGA and very short-period PSA are controlled by ground motion periods in the approximate range of 0.2–0.5 s, so dispersions for the very short periods might be expected to be similar to those in the 0.2–0.5 s range, as observed. Any processes that produce large variability in short-period energy (T < ~0.5 s) would therefore be expected to produce effects on PSA variability that are concentrated between about 0.03 s and 0.2 s because outside of that range, the spectral ordinates are dominated by relatively low-frequency ground motions. Based on point source simulations conducted by the first author and R. Youngs (*pers. communication*, 2013), we suggest two such processes could contribute to the bump in τ at T ~ 0.08 s: (1) variations in the source stress parameter for small **M** earthquakes (effect not present for larger earthquakes) and (2) variations in κ_0 for larger **M** earthquakes (smaller **M** events were not investigated).

Event-to-event variability in κ_0 could be postulated to result from regional crustal or geological variations that are not presently well understood. Further study of the short-period peak in τ is needed to verify and extend these postulations.

SUMMARY AND DISCUSSION

We have presented a set of ground motion prediction equations that we believe are the simplest formulation demanded by the NGA-West2 database used for the regressions. The new relations are a significant improvement over BA08′ and provide a demonstrably reliable description of recorded ground motion amplitudes for shallow crustal earthquakes in active tectonic regions over a wide range of magnitudes, distances, and site conditions.

Our GMPEs are intended for application in tectonically active crustal regions. The equations should not be used for other tectonic regimes such as subduction zones or stable continental regions unless their applicability can be verified. The data controlling the equations are derived principally from California, Taiwan, Japan, China, the Mediterranean region (Italy, Greece, Turkey), and Alaska. We have demonstrated some regional variations in ground motions, so application of the GMPEs to other areas considered to be active crustal regions carries an additional degree of epistemic uncertainty. This includes regions in the western United States outside of the portions of California included in our data set.

We recommend the following limits for the predictor variables used in our GMPEs:

- Strike-slip and reverse-slip earthquakes, **M** = 3 to 8.5.
- Normal-slip earthquakes, **M** = 3 to 7.
- Distance, R_{JB} = 0 km to 400 km.
- Time-averaged shear wave velocities of V_{S30} = 150 m/s to 1,500 m/s.
- Basin depth, z_1 = 0 km to 3.0 km.
- CL1 and CL2 event types (main shocks and aftershocks).

These limits are subjective estimates based on the distributions of the recordings used to develop the equations. We note that our recommended limit of **M** 8.5 for strike-slip and reverse-slip earthquakes is actually beyond the data limits. Events as large as **M** 8.5 are important for hazard applications. Therefore, we checked that extrapolation of our GMPEs to **M** 8.5 appears to be reasonable, based on agreement of the magnitude scaling in the GMPEs with predictions from simple SMSIM simulations (Boore 2005).

ACKNOWLEDGMENTS

This study was sponsored by the Pacific Earthquake Engineering Research Center (PEER) and funded by the California Earthquake Authority, the California Department of Transportation, and the Pacific Gas & Electric Company. Participation of the fourth author was funded by the Natural Sciences and Engineering Research Council of Canada. This study has been peer-reviewed and approved for publication consistent with U.S. Geological Survey Fundamental Science Practices policies, but does not necessarily reflect the opinions, findings, or conclusions of the other sponsoring agencies. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The work described in this report benefitted from the constructive discussions among

the NGA developers N.A. Abrahamson, Y. Bozorgnia, K.W. Campbell, B. S.-J. Chiou, I.M. Idriss, W.J. Silva, and R.E. Youngs. Input from supporting researchers was also invaluable, including T.D. Ancheta, J.W. Baker, A.M. Baltay, J. Donahue, C. Goulet, T. Hanks, R. Kamai, A. Kottke, T. Kishida, K. Wooddell, P. Spudich, and J. Watson-Lamprey. It was a pleasure to collaborate with this distinguished group of researchers. We thank S. Akkar, G. Ameri, F. Cotton, C. Mueller, and the two anonymous reviewers for their useful comments.

ELECTRONIC SUPPLEMENT

Please refer to the online edition of this paper to access the coefficient tables.

REFERENCES

Abrahamson, N. A., and Silva, W. J., 2008. Summary of the Abrahamson & Silva NGA ground motion relations, *Earthquake Spectra* **24**, 67–97.

Abrahamson, N. A., and Youngs, R. R., 1992. A stable algorithm for regression analyses using the random effects model, *Bull. Seismol. Soc. Am.* **82**, 505–510.

Al Atik, L., Abrahamson, N. A., Bommer, J. J., Scherbaum, F., Cotton, F., and Kuehn, N., 2010. The variability of ground-motion prediction models and its components, *Seismol. Res. Lett.* **81**, 794–801.

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., and Donahue, J. L., 2014. NGA-West2 database, *Earthquake Spectra* **30**, 989–1005.

Atkinson, G. M., 2006. Single-station sigma, *Bull. Seismol. Soc. Am.* **96**, 446–455.

Atkinson, G. M., 2013. Empirical evaluation of aleatory and epistemic uncertainty in eastern ground motions, *Seismol. Res. Letters* **84**, 130–138.

Atkinson, G. M., and Boore, D. M., 2011. Modifications to existing ground-motion prediction equations in light of new data, *Bull. Seismol. Soc. Am.* **101**, 1121–1135.

Atkinson, G. M., and Morrison, M., 2009. Regional variability in ground motion amplitudes along the west coast of North America, *Bull. Seismol. Soc. Am.* **99**, 2393–2409.

Andrews, D. J., 2001. A suggestion for fitting ground motion attenuation near an extended earthquake source, *Seismol. Res. Letters* **72**, 454–461.

Bommer, J. J., and Abrahamson, N. A., 2006. Why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates?, *Bull. Seismol. Soc. Am.* **96**, 1967–1977.

Boore, D. M., 2003. Prediction of ground motion using the stochastic method, *Pure and Applied Geophysics* **160**, 635–676.

Boore, D. M., 2005. SMSIM—Fortran programs for simulating ground motions from earthquakes: Version 2.3, U.S. Geological Survey, *Open-File Report 00-509*, revised 15 August 2005, 55 pgs.

Boore, D. M., 2010. Orientation-independent, non geometric-mean measures of seismic intensity from two horizontal components of motion, *Bull. Seismol. Soc. Am.* **100**, 1830–1835.

Boore, D. M., and Atkinson, G. M., 2007. *Boore-Atkinson NGA Ground Motion Relations for the Geometric Mean Horizontal Component of Peak and Spectral Ground Motion Parameters*, PEER Report No. 2007/01, Pacific Earthquake Engineering Research Center, Berkeley, CA, 234 pp.

Boore, D. M., and Atkinson, G. M., 2008. Ground motion prediction equations for the average horizontal component of PGA, PGV, and 5% damped PSA at spectral periods between 0.01 s and 10.0 s, *Earthquake Spectra* **24**, 99–138.

Boore, D. M., Joyner, W. B., and Fumal, T. E., 1997. Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work (with 2005 Erratum), *Seismol. Res. Lett.* **68**, 128–153.

Boore, D. M., Thompson, E. M., and Cadet, H., 2011. Regional correlations of V_{S30} and velocities averaged over depths less than and greater than 30 m, *Bull. Seismol. Soc. Am.* **101**, 3046–3059.

Boore, D. M., Stewart, J. P., Seyhan, E., and Atkinson, G. A., 2013. *NGA-West2 Equations for Predicting Response Spectral Accelerations for Shallow Crustal Earthquakes*, PEER Report No. 2013/05, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, 134 pp.

Bozorgnia, Y., Abrahamson, N. A., Al Atik, L., Ancheta, T. D., Atkinson, G. M., Baker, J. W., Baltay, A., Boore, D. M., Campbell, K. W., Chiou, B. S.-J., Darragh, R., Day, S., Donahue, J., Graves, R. W., Gregor, N., Hanks, T., Idriss, I. M., Kamai, R., Kishida, T., Kottke, A., Mahin, S. A., Rezaeian, S., Rowshandel, B., Seyhan, E., Shahi, S., Shantz, T., Silva, W., Spudich, P., Stewart, J. P., Watson-Lamprey, J., Wooddell, K., and Youngs, R., 2014. NGA-West2 research project, *Earthquake Spectra* **30**, 973–987.

Bradley, B. A., 2013. A New Zealand-specific pseudo spectral acceleration ground motion prediction equation for active shallow crustal earthquakes based on foreign models, *Bull. Seismo. Soc. Am.* **103**, 1801–1822.

Campbell, K. W., and Bozorgnia, Y., 2008. NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear-elastic response spectra for periods ranging from 0.01 and 10.0 s, *Earthquake Spectra* **24**, 139–171.

Chiou, B. S.-J., and Youngs, R. R., 2008. An NGA model for the average horizontal component of peak ground motion and response spectra, *Earthquake Spectra* **24**, 173–216.

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

Choi, Y., and Stewart, J. P., 2005. Nonlinear site amplification as function of 30 m shear wave velocity, *Earthquake Spectra* **21**, 1–30.

Day, S. M., Graves, R. W., Bielak, J., Dreger, D., Larsen, S., Olsen, K. B., Pitarka, A., and Ramirez-Guzman, L., 2008. Model for basin effects on long-period response spectra in southern California, *Earthquake Spectra* **24**, 257–277.

Donahue, J. L., and Abrahamson, N. A., 2014. Simulation-based hanging wall effects, *Earthquake Spectra* **30**, 1269–1284.

Douglas, J., 2003. Earthquake ground motion estimation using strong-motion records: A review of equations for the estimation of peak ground acceleration and response spectral ordinates, *Earth-Science Reviews* **61**, 43–104.

Douglas, J., 2011. *Ground Motion Prediction Equations 1964–2010*, PEER Report No. 2011/102, Pacific Earthquake Engineering Research Center, Berkeley, CA.

Douglas, J., and Boore, D. M., 2011. High-frequency filtering of strong-motion records, *Bull. Eqk. Eng.* **9**, 395–409.

Fletcher, J., Boatwright, J., Haar, L., Hanks, T., and McGarr, A., 1984. Source parameters for aftershocks of the Oroville, California, earthquake, *Bull. Seismo. Soc. Am.* **74**, 1101–1123.

Johnston, A. C., Coppersmith, K. J., Kanter, L. R., and Cornell, C. A., 1994. *The Earthquakes of Stable Continental Regions*, Technical Report TR-102261-VI, Electric Power Research Institute (EPRI), Palo Alto, CA.

Joyner, W. B., and Boore, D. M., 1993. Methods for regression analysis of strong-motion data, *Bull. Seismol. Soc. Am.* **83**, 469–487.

Joyner, W. B., and Boore, D. M., 1994. Errata: Methods for regression analysis of strong-motion data, *Bull. Seismol. Soc. Am.* **84**, 955–956.

Kottke, A., 2011. *Selection of Wenchuan stations based on geology*, Unpublished internal PEER white paper, 11 November 2011.

Pinheiro, H., Bates, D., DebRoy, S., Sarkar, D., and the R Development Core Team, 2013. *NLME: Linear and Nonlinear Mixed Effects Models*, R package version 3.1-108.

Power, M., Chiou, B.S.-J., Abrahamson, N., Bozorgnia, Y., Shantz, T., and Roblee, C., 2008. An overview of the NGA project, *Earthquake Spectra* **24**, 3–21.

Scasserra, G., Stewart, J. P., Bazzurro, P., Lanzo, G., and Mollaioli, F., 2009. A comparison of NGA ground motion prediction equations to Italian data, *Bull. Seismol. Soc. Am.* **99**, 2961–2978.

Schmedes, J., and Archuleta, R. J., 2008. Near-source ground motion along strike-slip faults: Insights into magnitude saturation of PGV and PGA, *Bull. Seismo. Soc. Am.* **98**, 2278–2290.

Seyhan, E., and Stewart, J. P., 2014. Semi-empirical nonlinear site amplification from NGA-West2 data and simulations, *Earthquake Spectra* **30**, 1241–1256.

Seyhan, E., Stewart, J. P., Ancheta, T. D., Darragh, R. B., and Graves, R.W., 2014. NGA-West2 site database, *Earthquake Spectra* **30**, 1007–1024.

Shahi, S. K., and Baker, J. W., 2014. NGA-West2 models for ground motion directionality, *Earthquake Spectra* **30**, 1285–1300.

Stewart, J. P., Midorikawa, S., Graves, R. W., Khodaverdi, K., Kishida, T., Miura, H., Bozorgnia, Y., and Campbell, K. W., 2013. Implications of M_w 9.0 Tohoku-oki Japan earthquake for ground motion scaling with source, path, and site parameters, *Earthquake Spectra* **29**, S1–S21.

Wooddell, K. E., and Abrahamson, N. A., 2014. Classification of main shocks and aftershocks in the NGA-West2 database, *Earthquake Spectra* **30**, 1257–1267.

(Received 1 July 2013; accepted 8 October 2013)