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

David M. Boore[a] and Gail M. Atkinson,[b] M.EERI

This paper contains ground-motion prediction equations (GMPEs) for average horizontal-component ground motions as a function of earthquake magnitude, distance from source to site, local average shear-wave velocity, and fault type. Our equations are for peak ground acceleration (PGA), peak ground velocity (PGV), and 5%-damped pseudo-absolute-acceleration spectra (PSA) at periods between 0.01 s and 10 s. They were derived by empirical regression of an extensive strong-motion database compiled by the "PEER NGA" (Pacific Earthquake Engineering Research Center's Next Generation Attenuation) project. For periods less than 1 s, the analysis used 1,574 records from 58 mainshocks in the distance range from 0 km to 400 km (the number of available data decreased as period increased). The primary predictor variables are moment magnitude (**M**), closest horizontal distance to the surface projection of the fault plane (R_{JB}), and the time-averaged shear-wave velocity from the surface to 30 m (V_{S30}). The equations are applicable for **M**=5–8, $R_{JB}<200$ km, and V_{S30}=180–1300 m/s. [DOI: 10.1193/1.2830434]

INTRODUCTION

Ground-motion prediction equations (GMPEs), giving ground-motion intensity measures such as peak ground motions or response spectra as a function of earthquake magnitude and distance, are important tools in the analysis of seismic hazard. These equations are typically developed empirically by a regression of recorded strong-motion amplitude data versus magnitude, distance, and possibly other predictive variables. The equations in this report were derived as part of the Pacific Earthquake Engineering Research Center's Next Generation Attenuation project (PEER NGA; Power et al. 2008), using an extensive database of thousands of records, compiled from shallow crustal earthquakes in active tectonic environments worldwide. These equations represent a substantial update to GMPEs that were published by Boore and his colleagues in 1997 (Boore et al. 1997, hereafter "BJF97"; note that BJF97 summarized work previously published by Boore et al. in 1993 and 1994). The 1997 GMPEs of Boore et al. were based on a fairly limited set of data in comparison to the results of this study. The in-

[a] U.S. Geological Survey, MS 977, 345 Middlefield Rd., Menlo Park, CA 94025
[b] Department of Earth Sciences, University of Western Ontario, London, Ont. Canada N6A 5B7

Earthquake Spectra, Volume 24, No. 1, pages 99–138, February 2008; © 2008, Earthquake Engineering Research Institute

crease in data quantity, by a factor of approximately 14, is particularly important for PSA; in addition, PGV equations are provided in this study (but were not given in BJF97). The amount of data used in regression analysis is an important issue as it bears heavily on the reliability of the results, especially in magnitude and distance ranges that are important for seismic hazard analysis.

This paper is a condensation of our final project report published by PEER (Boore and Atkinson 2007); the reader may refer to that document for more details and a number of relevant appendices. We will refer to that report as "BA07".

DATA

DATA SOURCES

The source of the strong ground-motion data for the development of the GMPEs in this study is the database compiled in the PEER NGA project (Chiou et al. 2008); the aim of that project was to develop empirical GMPEs using several investigative teams to allow a range of interpretations (this paper is the report of one team). The use of this database, referred to as the "NGA Flatfile," was one of the "ground rules" of the GMPE development exercise. However, investigators were free to decide whether to use the entire NGA Flatfile database, or to restrict their analyses to selected subsets.

In addition to the data in the NGA Flatfile, we also used data compiled by J. Boatwright and L. Seekins for three small events, and data from the 2004 Parkfield, California, mainshock from the Berkeley Digital Seismic Network station near Parkfield, as well as data from the Strong-Motion Instrumentation Program of the California Geological Survey and the National Strong-Motion Program of the United States Geological Survey. These additional data were used in a study of the distance attenuation function that constrained certain regression coefficients, as discussed later, but were not included as part of the final regression (to be consistent with the NGA "ground rules" regarding the database for regression).

RESPONSE VARIABLES

The ground-motion parameters that are the dependent variables of the GMPEs (also called response variables or ground-motion intensity measures) include peak ground acceleration (PGA), peak ground velocity (PGV), and response spectra (PSA, the 5%-damped pseudo-acceleration), all for the horizontal component. In this study, the response variables are not the simple geometric mean of the two horizontal component (as was used in BJF97), but rather are measures of geometric mean not dependent on the particular orientation of the instruments used to record the horizontal motion. The measure used was introduced by Boore et al. (2006). In that paper a number of orientation-independent measures of ground motion were defined. In this report we use GMRotI50 (which we abbreviate "GMRotI"); this is the geometric mean determined from the 50th-percentile values of the geometric means computed for all non-redundant rotation angles and all periods less than the maximum useable period. The advantage of using an orientation-independent measure of the horizontal component amplitude can be appreciated by considering the case in which the motion is perfectly polarized along one com-

ponent direction; in this case the geometric mean would be 0. In most cases, however, the differences between the geometric mean and GMRotI are not large, so that our response variable can be thought of in simple terms as an average horizontal component.

This paper includes GMPEs for PGA, PGV, and 5%-damped PSA for periods between 0.01 s and 10 s. Equations for peak ground displacement (PGD) are not included. In our view, PGD is too sensitive to the low-cut filters used in the data processing to be a stable measure of ground shaking. In addition there is some bias in the PGD values obtained in the NGA data set from records for which the low-cut filtering was not performed as part of the NGA project. Appendix C in BA07 contains a short discussion of these points. We recommend using response spectra at long periods instead of PGD.

Data were excluded from our analysis based on a number of criteria, the most important of which (in terms of number of records excluded from the analysis) is that no recordings from obvious aftershocks were used. Aftershock records were not used because of some concern that the spectral scaling of aftershocks differs from mainshocks (see Boore and Atkinson 1989 and Atkinson 1993). This restriction cut the data set almost in half because a substantial number of the records in the NGA Flatfile are aftershocks of the 1999 Chi-Chi earthquake. The other exclusion criteria that were applied are listed in Table 2.1 of BA07. Response variables were excluded for oscillator periods greater than T_{MAX} (the inverse of the lowest useable frequency entry in the NGA Flatfile).

We did not use singly recorded earthquakes. Table 1 lists all earthquakes used in our data analysis, along with the number of stations used per earthquake (for an oscillator period of 0.2 s).

A potential bias in regression results can result from not including low-amplitude data from distance ranges for which larger amplitude data for the same earthquake are included in the data set. There are several reasons that low-amplitude data might not be included: it can be below trigger thresholds of instruments, which will cause the recording to begin at some point during the S-wave arrival, it can be too small to digitize, or it can be below the noise threshold used in determining low-cut filter frequencies. Any collection of data in a small distance range will have a range of amplitudes because of the natural variability in the ground motion (due to such things as source, path, and site variability). At distances far enough from the source (depending on magnitude), some of the values in the collection will be below the amplitude cutoff and would therefore be excluded. If only the larger motions (above the amplitude cutoff) were included, this would lead to a bias in the predicted distance decay of the ground motion—there would be a tendency for the predicted ground motions to decay less rapidly with distance than the real data. BJF97 attempted to avoid this bias by excluding data for each earthquake beyond the closest distance to an operational, non-triggered station (most of the data used by BJF97 were obtained on triggered analog stations). Unfortunately, information is not available in the NGA Flatfile that would allow us to apply a similar distance cutoff, at least for the case of triggered analog recordings. Furthermore, a similar bias might also exist in digital recordings because of the presence of long-period noise that is indepen-

Table 1. Events used in analysis, for a period of 0.2 s, giving type of earthquake (S=strike-slip, N=normal, R=reverse), number of observations (NOBS), range of R_{JB} in km, and NGA Flatfile event identification number

NAME	YEAR	MODY	M	DIP	DEPTH (km)	TYPE	NOBS	RJB RANGE	EQID
Parkfield	1966	0628	6.19	90	10	S	4	10–18	25
Borrego Mtn	1968	0409	6.63	78	8	S	2	129–222	28
San Fernando	1971	0209	6.61	50	13	R	31	14–218	30
Hollister-03	1974	1128	5.14	90	6	S	2	9–10	34
Friuli, Italy-01	1976	0506	6.50	12	5	R	5	15–102	40
Tabas, Iran	1978	0916	7.35	25	6	R	7	0–194	46
St Elias, Alaska	1979	0228	7.54	12	16	R	2	26–80	142
Coyote Lake	1979	0806	5.74	80	10	S	7	0–34	48
Norcia, Italy	1979	0919	5.90	64	6	N	3	2–31	49
Imperial Valley-06	1979	1015	6.53	80	10	S	33	0–49	50
Livermore-01	1980	0124	5.80	85	12	S	5	15–53	53
Anza (Horse Canyon)-01	1980	0225	5.19	70	14	S	5	6–39	55
Mammoth Lakes-01	1980	0525	6.06	50	9	N	2	1–5	56
Victoria, Mexico	1980	0609	6.33	90	11	S	4	6–39	64
Irpinia, Italy-01	1980	1123	6.90	60	10	N	12	7–60	68
Westmorland	1981	0426	5.90	90	2	S	6	6–19	73
Coalinga-01	1983	0502	6.36	30	5	R	44	24–55	76
Borah Peak, ID-01	1983	1028	6.88	52	16	N	2	83–85	87
Morgan Hill	1984	0424	6.19	90	9	S	24	3–71	90
Lazio-Abruzzo, Italy	1984	0507	5.80	48	14	N	5	13–49	91
Hollister-04	1986	0126	5.45	90	9	S	3	11–13	98
N Palm Springs	1986	0708	6.06	46	11	R	30	0–78	101
Chalfant Valley-01	1986	0720	5.77	90	7	S	5	6–24	102
Chalfant Valley-02	1986	0721	6.19	55	10	S	10	6–51	103
San Salvador	1986	1010	5.80	85	11	S	2	2–4	108
Whittier Narrows-01	1987	1001	5.99	30	15	R	106	0–82	113
Superstition Hills-02	1987	1124	6.54	90	9	S	11	1–27	116
Loma Prieta	1989	1018	6.93	70	18	R	73	0–117	118
Upland	1990	0228	5.63	77	5	S	3	7–72	143
Manjil, Iran	1990	0620	7.37	88	19	S	7	13–175	144
Sierra Madre	1991	0628	5.61	50	12	R	8	3–46	145
Roermond, Netherlands	1992	0413	5.30	68	15	N	3	55–101	122
Cape Mendocino	1992	0425	7.01	14	10	R	6	0–40	123
Landers	1992	0628	7.28	90	7	S	68	2–190	125
Big Bear-01	1992	0628	6.46	85	13	S	39	7–147	126
Little Skull Mtn, NV	1992	0629	5.65	70	12	N	8	14–99	152
Northridge-01	1994	0117	6.69	40	18	R	154	0–148	127
Kobe, Japan	1995	0116	6.90	85	18	S	12	0–158	129
Kozani, Greece-01	1995	0513	6.40	43	13	N	3	14–79	130
Dinar, Turkey	1995	1001	6.40	45	5	N	4	0–255	134

Table 1. (cont.)

NAME	YEAR	MODY	M	DIP	DEPTH (km)	TYPE	NOBS	RJB RANGE	EQID
Northwest China-01	1997	0405	5.90	68	23	S	2	12–49	153
Northwest China-02	1997	0406	5.93	30	31	N	2	20–37	154
Northwest China-04	1997	0415	5.80	43	22	N	2	21–35	156
Kocaeli, Turkey	1999	0817	7.51	88	15	S	26	1–316	136
Chi-Chi, Taiwan	1999	0920	7.62	30	7	R	380	0–169	137
Hector Mine	1999	1016	7.13	77	5	S	82	10–233	158
Düzce, Turkey	1999	1112	7.14	65	10	S	22	0–188	138
Yountville	2000	0903	5.00	90	10	S	24	8–94	160
Big Bear-02	2001	0210	4.53	90	9	S	41	22–92	161
Mohawk Val, Portola	2001	0810	5.17	81	4	S	6	67–126	162
Anza-02	2001	1031	4.92	78	15	S	72	10–133	163
Gulf of California	2001	1208	5.70	59	10	S	11	72–130	164
CA/Baja Border Area	2002	0222	5.31	74	7	S	9	40–97	165
Gilroy	2002	0514	4.90	84	10	S	34	2–130	166
Yorba Linda	2002	0903	4.27	88	7	S	12	6–36	167
Nenana Mountain, Alaska	2002	1023	6.70	90	4	S	33	105–280	168
Denali, Alaska	2002	1103	7.90	71	5	S	23	0–276	169
Big Bear City	2003	0222	4.92	72	6	S	33	24–146	170

dent of the distance from the source to the station. Consequently, the obtained distance dependence for small earthquakes and long periods may be biased towards a decay that is less rapid than the true decay.

PREDICTOR VARIABLES

The primary predictor variables (independent variables in the regression analysis) are moment magnitude **M**, R_{JB} distance (closest distance to the surface projection of the fault plane), and V_{S30} for site characterization. (V_{S30} is the time-averaged shear-wave velocity over the top 30 m, calculated as the inverse of the average shear-wave slowness from the surface to a depth of 30 m [although slowness is simply the inverse of velocity, it has a number of useful properties, as discussed in Boore and Thompson 2007].) The R_{JB} distances estimated by R. Youngs, as described in Appendix B of Chiou and Youngs (2006), were used for earthquakes with unknown fault geometry.

We also considered the effect of fault type (i.e., normal, strike-slip, and reverse). The fault type was specified by the plunge of the *P*- and *T*-axes, as shown in the legend to Figure 1 (Appendix D of BA07 contains a more complete description). While there are some advantages to using *P*- and *T*-axes, Figure 1 shows that the simple classification used by BJF97 (rakes angles within 30° of horizontal are strike-slip, angles from 30° to

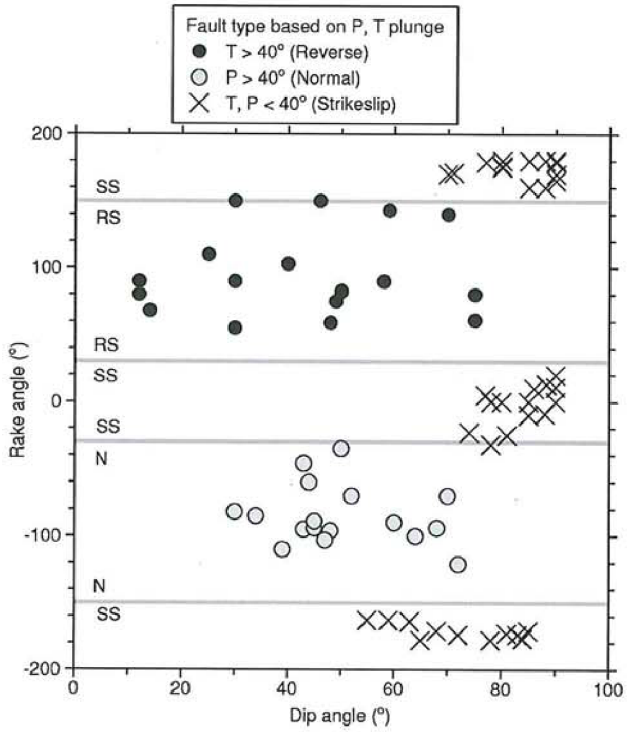

Figure 1. Distribution of the data we used in rake-angle and dip-angle space. The horizontal gray lines indicate boundaries between fault types used by BJF97, and the symbols and colors indicate our classification based on the plunges of the P- and T-axes (our classification scheme is indicated in the legend; see Appendix D in BA07).

150° are reverse, and angles from −30° to −150° are normal) gives essentially the same classifications as obtained using *P*- and *T*-axes, at least for the data set we used.

All of the predictor variables were taken from the NGA database.

DISTRIBUTION OF DATA BY M, R_{JB}, FAULT TYPE, AND SITE CLASS

Representative **M** and R_{JB} distributions of the data used in developing our GMPEs are shown in Figure 2, with the symbols representing different fault types. The total number of recordings for the analysis (after all exclusions) is 1,574 for periods out to 1 s, with a slight decrease at 2 s, and a rapid fall off in the number of available data at periods longer than 2 s. The distributions of the data over the predictor variable space necessarily influence the GMPEs. Note in particular the lack of data at close distances for small earthquakes. This means that the near-source ground motions for small events will not be constrained by observations. For long oscillator periods, there are very few data for small earthquakes at any distance (the points in Figure 2 for **M**=5 and *T*

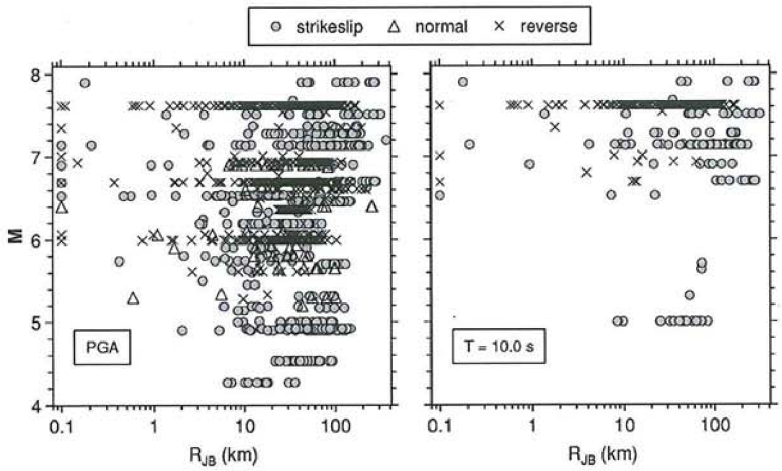

Figure 2. Distribution of data used to derive our regression equations for PGA and for PSA at a period 10.0 s, differentiated by fault type (points with R_{JB} less than 0.1 km plotted at 0.1 km). The overall distributions for periods less than about 4 s are similar to those for PGA, although there are fewer recordings (the number of available recordings decreases noticeably for periods longer than 2 s).

=10 s are all from a single event—the 2000 Yountville, California earthquake), so the magnitude scaling at long periods will be poorly determined for small magnitudes.

The widest range of magnitudes is for strike-slip earthquakes (4.3–7.9), while the narrowest range is for normal-slip earthquakes (5.3–6.9). This suggests that the magnitude scaling is better determined for strike-slip than for normal-slip earthquakes—a problem that we circumvented by using a common magnitude scaling for all types of events, as discussed later.

The bulk of the data are from class C and D sites, which range from soft rock to firm soil; very few data were from class A sites (hard rock). More detail can be found in Appendix A, which includes two possible sets of V_{S30} values to use in evaluating our equations for a particular NEHRP site class.

THE EQUATIONS

Following the philosophy of Boore et al. (1993, 1994, 1997), we seek simple functional forms for our GMPEs, with the minimum required number of predictor variables. We started with the simplest reasonable form for the equations (that used in BJF97), and then added complexity as demanded by comparisons of the predictions of ground motions from the simplest equations with the observed ground motions. The selection of functional form was heavily guided by subjective inspection of nonparametric plots of data; many such plots were produced and studied before commencing the regression analysis. For example, the BJF97 equations modeled the far-source attenuation of am-

plitudes with distance by a simple function that had no magnitude dependence and no curvature at greater distances. This form appeared sufficient for the maximum distance range of 80 km specified for the BJF97 GMPEs. The data, however, clearly show that curvature of the line is required to accommodate the effects of anelastic attenuation when modeling data beyond 80 km; furthermore, the data show that the effective geometric spreading factor is dependent on magnitude. To accommodate these trends, we (1) added an "anelastic" coefficient to the form of the equations, in which $\ln Y$ is proportional to R (where Y is the response variable), and (2) introduced a magnitude-dependent "geometrical spreading" term, in which $\ln Y$ is proportional to $\ln R$ and the proportionality factor is a function of **M**. These features allow the equations to predict amplitudes to 400 km; the larger size of the NGA database at greater distances and for larger magnitudes, in comparison to that available to BJF97, enabled robust determination of the additional coefficients. Our functional form does not include such factors as depth-to-top of rupture, hanging wall/footwall terms, or basin depth, because residual analysis does not clearly show that the introduction of such factors would improve their predictive capabilities on average. The equations are data-driven and make little use of simulations. They include only those terms that are truly required to adequately fit the observational database, according to our analysis. Our equations may provide a useful alternative to the more complicated equations provided by other NGA models, as they will be easier to implement in many applications.

Our equation for predicting ground motions is:

$$\ln Y = F_M(\mathbf{M}) + F_D(R_{JB}, \mathbf{M}) + F_S(V_{S30}, R_{JB}, \mathbf{M}) + \varepsilon\sigma_T, \tag{1}$$

In this equation, F_M, F_D, and F_S represent the magnitude scaling, distance function, and site amplification, respectively. **M** is moment magnitude, R_{JB} is the Joyner-Boore distance (defined as the closest distance to the surface projection of the fault, which is approximately equal to the epicentral distance for events of $\mathbf{M} < 6$), and the velocity V_{S30} is the inverse of the average shear-wave slowness from the surface to a depth of 30 m. The predictive variables are **M**, R_{JB}, and V_{S30}; the fault type is an optional predictive variable that enters into the magnitude scaling term as shown in Equation 5a and 5b below. ε is the fractional number of standard deviations of a single predicted value of $\ln Y$ away from the mean value of $\ln Y$ (*e.g.*, $\varepsilon = -1.5$ would be 1.5 standard deviations smaller than the mean value). All terms, including the coefficient σ_T, are period dependent. σ_T is computed using the equation

$$\sigma_T = \sqrt{\sigma^2 + \tau^2}, \tag{2}$$

where σ is the intra-event aleatory uncertainty and τ is the inter-event aleatory uncertainty (this uncertainty is slightly different for cases where fault type is specified and where it is not specified; we distinguish these cases by including a subscript on τ).

Table 2. Values of dummy variables for different fault types

Fault Type	U	SS	NS	RS
Unspecified	1	0	0	0
Strike-slip	0	1	0	0
Normal	0	0	1	0
Thrust/reverse	0	0	0	1

THE DISTANCE AND MAGNITUDE FUNCTIONS

The distance function is given by:

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

where

$$R = \sqrt{R_{JB}^2 + h^2} \tag{4}$$

and c_1, c_2, c_3, $\mathbf{M}_{ref}$, R_{ref}, and h are the coefficients to be determined in the analysis.

The magnitude scaling is given by:

a) $\mathbf{M} \leq \mathbf{M}_h$

$$F_M(\mathbf{M}) = e_1 U + e_2 SS + e_3 NS + e_4 RS + e_5(\mathbf{M} - \mathbf{M}_h) + e_6(\mathbf{M} - \mathbf{M}_h)^2, \tag{5a}$$

b) $\mathbf{M} > \mathbf{M}_h$

$$F_M(\mathbf{M}) = e_1 U + e_2 SS + e_3 NS + e_4 RS + e_7(\mathbf{M} - \mathbf{M}_h), \tag{5b}$$

where U, SS, NS, and RS are dummy variables used to denote unspecified, strike-slip, normal-slip, and reverse-slip fault type, respectively, as given by the values in Table 2, and $\mathbf{M}_h$, the "hinge magnitude" for the shape of the magnitude scaling, is a coefficient to be set during the analysis.

SITE AMPLIFICATION FUNCTION

The site amplification equation is given by:

$$F_S = F_{LIN} + F_{NL}, \tag{6}$$

where F_{LIN} and F_{NL} are the linear and nonlinear terms, respectively.

The linear term is given by:

$$F_{LIN} = b_{lin} \ln(V_{S30}/V_{ref}), \tag{7}$$

where b_{lin} is a period-dependent coefficient, and V_{ref} is the specified reference velocity (=760 m/s), corresponding to NEHRP B/C boundary site conditions; these coefficients

were prescribed based on the work of Choi and Stewart (2005; hereafter "CS05"); they are empirically based but were not determined by the regression analysis in our study.

The nonlinear term is given by:

a) $pga4nl \leq a_1$:

$$F_{NL} = b_{nl} \ln(pga_low/0.1) \tag{8a}$$

b) $a_1 < pga4nl \leq a_2$:

$$F_{NL} = b_{nl} \ln(pga_low/0.1) + c[\ln(pga4nl/a_1)]^2 + d[\ln(pga4nl/a_1)]^3 \tag{8b}$$

c) $a_2 < pga4nl$:

$$F_{NL} = b_{nl} \ln(pga4nl/0.1) \tag{8c}$$

where a_1 (=0.03 g) and a_2 (=0.09 g) are assigned threshold levels for linear and nonlinear amplification, respectively, pga_low (=0.06 g) is a variable assigned to transition between linear and nonlinear behaviors, and $pga4nl$ is the predicted PGA in g for V_{ref} =760 m/s, as given by Equation 1 with F_S=0 and ε=0. The three equations for the nonlinear portion of the soil response (Equation 8a–8c) are required for two reasons: 1) to prevent the nonlinear amplification from increasing indefinitely as $pga4nl$ decreases and 2) to smooth the transition from linear to non-linear behavior. The coefficients c and d in Equation 8b are given by

$$c = (3\Delta y - b_{nl}\Delta x)/\Delta x^2 \tag{9}$$

and

$$d = -(2\Delta y - b_{nl}\Delta x)/\Delta x^3, \tag{10}$$

where

$$\Delta x = \ln(a_2/a_1) \tag{11}$$

and

$$\Delta y = b_{nl} \ln(a_2/pga_low). \tag{12}$$

The nonlinear slope b_{nl} is a function of both period and V_{S30} as given by:

a) $V_{S30} \leq V_1$:

$$b_{nl} = b_1. \tag{13a}$$

b) $V_1 < V_{S30} \leq V_2$:

$$b_{nl} = (b_1 - b_2)\ln(V_{S30}/V_2)/\ln(V_1/V_2) + b_2. \tag{13b}$$

c) $V_2 < V_{S30} < V_{ref}$:

$$b_{nl} = b_2 \ln(V_{S30}/V_{ref})/\ln(V_2/V_{ref}). \tag{13c}$$

d) $V_{ref} \leq V_{S30}$:

$$b_{nl} = 0.0. \tag{13d}$$

where $V_1 = 180$ m/s, $V_2 = 300$ m/s, and b_1 and b_2 are period-dependent coefficients (and consequently, b_{nl} is a function of period as well as V_{S30}). These equations are a simplified version of those used by CS05.

DETERMINATION OF COEFFICIENTS

METHODOLOGY

The selected response variables in the NGA database were first corrected to obtain the equivalent observations for the reference velocity of 760 m/s, using Equations 6, 7, 8a–8c, 9–12, and 13a–13d and an equation for *pga4nl* developed early in the project, using only data for which $R_{JB} \leq 80$ km and $V_{S30} > 360$ m/s (see BA07 for details). We then regressed the site-corrected observations to Equation 1 to determine F_D and F_M. Because the observations had all been corrected to the reference condition, we set $F_S = 0$, simplifying the regression. The analyses were performed using the two-stage regression discussed by Joyner and Boore (1993, 1994); the first stage determines the distance dependence (as well as event terms used in the second stage and the intra-event aleatory variability, σ), and the second stage determines the magnitude dependence (and the inter-event variability, τ). All regressions were done period-by-period; there was no smoothing of the coefficients that were determined by the regression analyses (although some of the constrained coefficients were smoothed). (Our "event term" is the average of the ln Y values for a given event, adjusted to the reference velocity and a reference distance (760 m/s and 1 km, respectively, in our study). This differs from the "event term" that is derived in a random effects model. This latter term is more precisely called a "random effect for a given event" (e.g., Abrahamson and Youngs 1992). The residuals from our Stage 2 regression are equivalent to this alternate meaning of "event term.")

Site Amplification

Because corrections for site amplification were made before doing the first-stage and second-stage regressions, we discuss the determination of the site amplification coefficients first. The coefficients in the site-response equations were based on the work of CS05, rather than determined by our regression analysis. The coefficients needed to evaluate the site-response equations are listed in Tables 3 and 4. Note that for the reference velocity of 760 m/s, $F_{LIN} = F_{NL} = F_S = 0$. Thus the soil amplifications are specified relative to motions that would be recorded on a B/C boundary site condition.

The rationale for pre-specifying the site amplifications is that the NGA database may be insufficient to determine simultaneously all coefficients for the nonlinear soil equations and the magnitude-distance scaling, due to trade-offs that occur between parameters, particularly when soil nonlinearity is introduced. It was therefore deemed preferable to "hard-wire" the soil response based on the best-available empirical analysis in

Table 3. Period-dependent site-amplification coefficients

Period	b_{lin}	b_1	b_2
PGV	−0.600	−0.500	−0.06
PGA	−0.360	−0.640	−0.14
0.010	−0.360	−0.640	−0.14
0.020	−0.340	−0.630	−0.12
0.030	−0.330	−0.620	−0.11
0.050	−0.290	−0.640	−0.11
0.075	−0.230	−0.640	−0.11
0.100	−0.250	−0.600	−0.13
0.150	−0.280	−0.530	−0.18
0.200	−0.310	−0.520	−0.19
0.250	−0.390	−0.520	−0.16
0.300	−0.440	−0.520	−0.14
0.400	−0.500	−0.510	−0.10
0.500	−0.600	−0.500	−0.06
0.750	−0.690	−0.470	0.00
1.000	−0.700	−0.440	0.00
1.500	−0.720	−0.400	0.00
2.000	−0.730	−0.380	0.00
3.000	−0.740	−0.340	0.00
4.000	−0.750	−0.310	0.00
5.000	−0.750	−0.291	0.00
7.500	−0.692	−0.247	0.00
10.000	−0.650	−0.215	0.00

the literature, and allow the regression to determine the remaining magnitude and distance scaling factors. It is recognized that there are implicit trade-offs involved, and that a change in the prescribed soil response equations would lead to a change in the derived magnitude and distance scaling. Note, however, that our prescribed soil response terms

Table 4. Period-independent site-amplification coefficients

Coefficient	Value
a_1	0.03 g
pga_low	0.06 g
a_2	0.09 g
V_1	180 m/s
V_2	300 m/s
V_{ref}	760 m/s

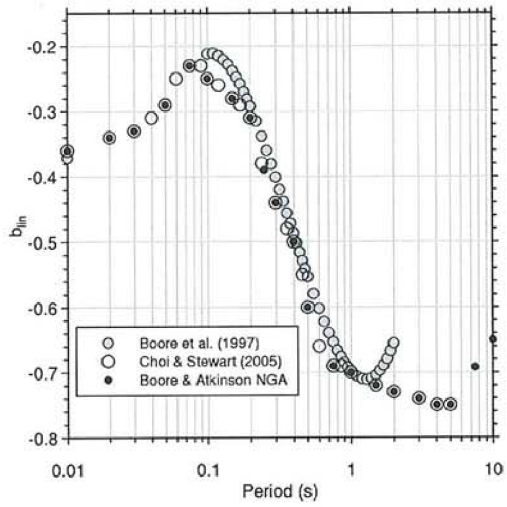

Figure 3. Coefficient controlling linear amplification, as function of period. Values used in equations in this report indicated by the black dots.

are similar to those adopted by other NGA developers who used different approaches; thus there appears to be consensus as to the appropriate level for the soil response factors.

The details of setting the coefficients for the soil response equations are as follows. The linear amplification coefficients b_{lin} were adopted from CS05. As shown in Figure 3, they are similar to the linear soil coefficients derived by BJF97. CS05 do not provide coefficients for periods beyond 5 s. To determine coefficients for longer periods, we extrapolated the b_{lin} values as shown on Figure 3. As periods get very long (>5 s), we would expect the relative linear site amplification to decrease (and a trend in this direction has been found by some of the other NGA developers). For this reason, we subjectively decided on the linear trend in terms of log period shown in Figure 3 as the basis for choosing the values for the longer periods.

The nonlinear slope factor b_{nl} depends on V_{S30} through the equations given above. Our equations define a somewhat simpler relation than that used by CS05. We compare the two definitions of the coefficient b_{nl} for periods of 0.2 and 3.0 s in Figure 4. The values of b_{nl} at the hinge points $V_{S30}=V_1$ and $V_{S30}=V_2$ are given by the coefficients b_1 and b_2, respectively, and these are functions of period. We use CS05's values for most periods. To extend the value of b_1 to periods longer than 5 s we fit two quadratic curves to the CS05 values: one for all of the values and another for values corresponding to periods greater than 0.2 s; the results were similar (see BA07 for a graph). We based our value of b_1 at periods of 7.5 s and 10 s on the quadratic fit to all of the CS05 values.

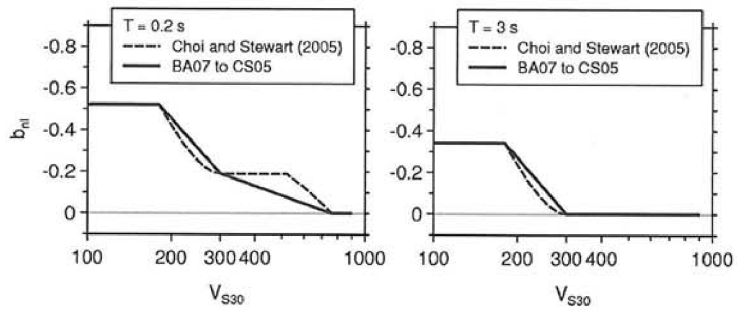

Figure 4. Comparison of slope that controls nonlinear amplification function.

This curve was also used for the value at 5 s, but the results of using the CS05 value at 5 s versus our value makes almost no difference in the predicted ground motions for 5 s periods.

We point out a potential confusion in terminology: according to Equation 8c, $F_{NL}=0.0$ when $pga4nl=0.1$ g. Does this mean that there is no nonlinear amplification for this level of rock motion? Not necessarily. The amplification for this value of $pga4nl$ is given entirely by the F_{LIN} term because the motions used by CS05 to derive the "linear" amplifications (F_{LIN}) had an approximate mean log PGA for most site categories close to 0.1 g. F_{NL} is not necessarily zero, however, for values of $pga4nl$ less than and greater than 0.1 g. So although the amplification at $pga4nl=0.1$ g is completely determined by F_{LIN}, the amplification could implicitly include the nonlinear component that applies for values of $pga4nl$ near 0.1 g. CS05 use only Equation 8c to describe the nonlinear amplification, and they do not limit the nonlinear response to $pga4nl>0.1$ g. It is clear from Figure 3 of CS05 and the comment on p. 24 of their paper that they consider Equation 8c to be valid for $pga4nl$ from 0.02 to 0.8 g. This means that the total amplification (F_S) can be greater than the "linear" amplification (F_{LIN}) for small values of $pga4nl$; their nonlinear amplification continues to increase without bound as $pga4nl$ decreases. We made an important modification to the CS05 procedure to prevent nonlinear amplification from extending to small values of $pga4nl$, by capping the amplifications at a low value of $pga4nl$ (0.03 g). Simply terminating the nonlinear amplification at a fixed value of $pga4nl$ results in a kink in plots of ground motion vs. distance. For that reason we included a transition curve, as given in Equation 8b.

The total amplification for a short (0.2 s) and a long (3.0 s) period oscillator is shown in Figure 5 as a function of $pga4nl$ for a range of V_{S30}. At short periods the nonlinear term can result in a significant reduction of motions on sites underlain by relatively low velocities. At long periods soil nonlinearity is still important, but the net soil response effect is an amplification, even for large values of $pga4nl$. For periods longer than 0.75 s (see Table 3) there is no nonlinear contribution to the amplification for $V_{S30}>300$ m/s

It should be noted that the empirical studies on which the soil amplification functions

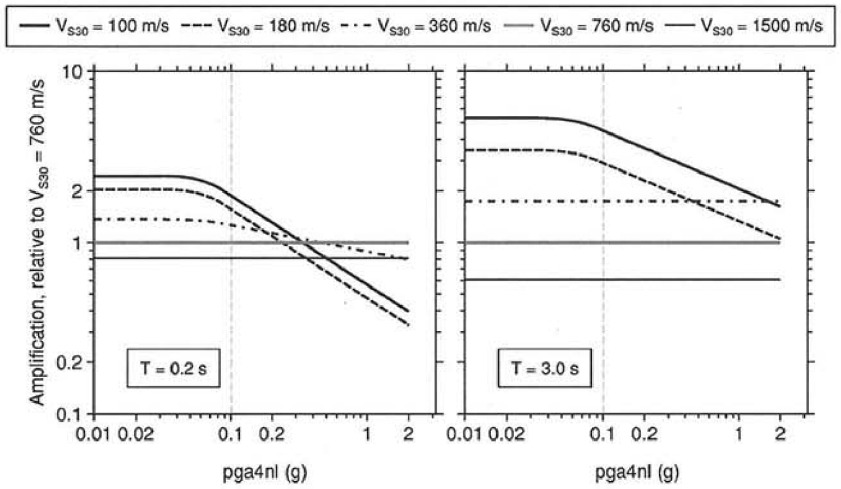

Figure 5. Combined amplification for T=0.2 s and T=3.0 s as function of *pga4nl*, for suite of V_{S30}. Note at short periods (left graph), purely linear amplification does not occur on soft soils until $pga4nl < 0.03$ g.

were based contained very few data for hard sites, with $V_{S30} > 1{,}000$ m/s. The amplification functions are probably reasonable for values of V_{S30} up to about 1,300 m/s, but should not be applied for very hard rock sites ($V_{S30} \geqslant 1500$ m/s).

Distance Dependence (Stage 1)

The distance dependence of ground motion is determined in the first-stage regression, where the dependent response variable is PGA, PGV, or PSA at a selected period, in each case corrected to the reference velocity of 760 m/s by subtracting F_S as defined in Equations 6, 7, 8a–8c, 9–12, and 13a–13d from $\ln Y_{observed}$. The corrected response variables for our selected subset of the NGA data set (using the exclusion criteria discussed earlier), with distances out to 400 km, are regressed against distance using Equation 14, which is the same as Equation 2 but with dummy variables ($c_0(event)$) added to represent the event term for each earthquake.

$$F_D(R_{JB}, \mathbf{M}) = c_0(event) + [c_1 + c_2(\mathbf{M} - \mathbf{M}_{ref})]\ln(R/R_{ref}) + c_3(R - R_{ref}) \qquad (14)$$

In this equation, "$c_0(event)$" is shorthand for the sum

$$(c_0)_1\delta_1 + (c_0)_2\delta_2 + \cdots + (c_0)_{NE}\delta_{NE}, \qquad (15)$$

where $(c_0)_j$ is the event term for event j, δ_j equals 1 for event j and zero otherwise, and NE is the number of earthquakes.

There are several significant issues in performing this regression. One is that regional differences in attenuation are known to exist (e.g., Boore 1989, Benz et al. 1997),

Table 5. Comparisons of numbers of stations in NGA Flatfile and in extended data set used to determine anelastic coefficient

Earthquake	# of Stations in NGA	# of Stations Used by BA
2001 Anza (**M** 4.92)	73	197
2002 Yorba Linda (**M** 4.27)	12	207
2003 Big Bear City (**M** 4.92)	37	262
2004 Parkfield (**M** 6.0)	0	74

even within relatively small regions such as California (e.g., Bakun and Joyner 1984, Boatwright et al. 2003, Hutton and Boore 1987, Mori and Helmberger 1996). We ignore this potential pitfall and assume that the distance part of the GMPEs apply for crustal earthquakes in all active tectonic regimes represented by the NGA database. This is a reasonable initial approach, as the significance of regional effects can be tested later by examining residual trends (model errors) for subsets of data organized by region. The second difficulty is more problematic: the data in the NGA Flatfile become increasingly sparse for distances beyond about 80 to 100 km, especially for moderate events. This makes it difficult, if not impossible, to obtain a robust simultaneous determination of c_1 and c_3 (slope and curvature). To overcome this database limitation, we have used additional ground-motion data from California that are not in the NGA Flatfile, to first define the "anelastic" term, c_3, as a function of period. We then used these fixed values of c_3 in the regression of the NGA data set in order to determine the remaining coefficients.

Determination of c3 (anelastic term): The data used to determine c_3 includes the data compiled in the NGA database for three small California events, plus many more data for these same events recorded by accelerometers at "broadband" stations in California; these additional data, compiled by J. Boatwright and L. Seekins, were not available from the traditional strong-motion data agencies used in compiling the NGA Flatfile. We also used response variables that we computed from 74 two-component recordings of the 2004 Parkfield mainshock (**M** 6.0) in the determination of c_3; these data were recorded after the compilation of the NGA database had concluded. The numbers of stations providing data for our analysis and the corresponding numbers of stations in the NGA Flatfile are given in Table 5 (see also Appendices M and N in BA07).

For the additional data for the three small California earthquakes, we used site classes assigned by Boatwright and Seekins to correct the response spectra to V_{S30} =760 m/s. For the Parkfield recordings we did not correct to a common value of V_{S30}, as we were interested only in determining the distance function, and also because measured values of V_{S30} were available at only a few sites. For all of the data from the four events we used spectra from the two horizontal components as if they were separate recordings (we did not combine the horizontal components). We did the regressions on this data subset with c_1 fixed at −0.5, −0.8, and −1.0. We set c_2 to zero and solved for c_3 and h. In other words, we fixed a single straight-line slope (c_1), and then determined the curvature, c_3, required to match the more rapid decay of the data at greater distances (c_3 must be less than 0). We also solved for the near-source effective depth coefficient, h,

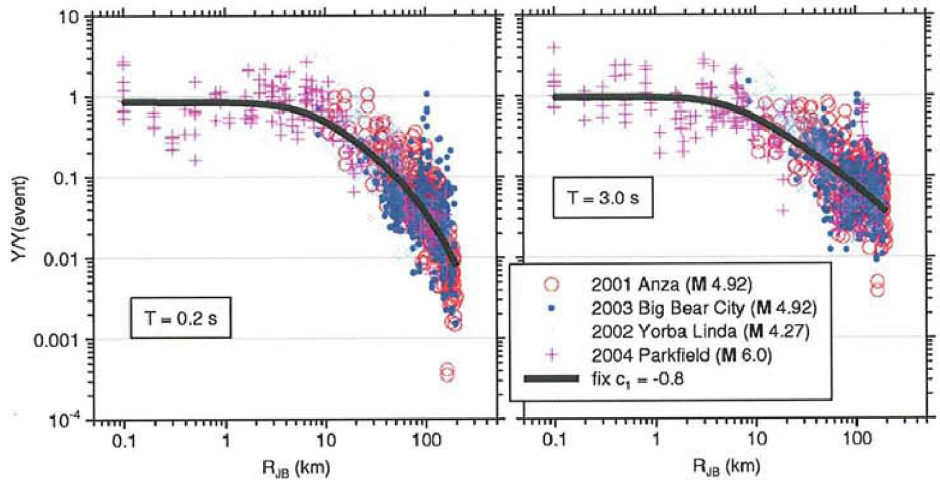

Figure 6. Normalized ground motions for four events, using extended data set (more data than in NGA Flatfile). Black curve is regression fit obtained with constraints $c_1=-0.8$ and $c_2=0.0$.

required to match the less rapid increase of the data as distance decreases at close distances. An event term that gives the relative amplitude level, (c_0), is determined for each of the four earthquakes (these are the coefficients of the dummy variables for each event). Figure 6 compares the regression fits to the observations, where the observations have been normalized to a common amplitude level by subtracting the event terms (c_0). We chose the c_3 values determined for the case $c_1=-0.8$ as the fixed c_3 values to apply in the regression of the NGA data set because $c_1=-0.8$ is a typical value determined in empirical regressions for the effective geometric slope parameter at intermediate periods (BJF97; this study).

As a broader check on the results from our four-event attenuation data set, we determined the best values of c_3 and h to fit the distance functions determined in southern California from a much larger database, by Raoof et al. (1999). The equivalent values of c_3 and h implied by the Raoof et al. (1999) attenuation results are similar to those that we determined from our four-event analysis.

To assign values of c_3 over the full period range required in the NGA project, we fit a quadratic to the c_3 values from the analysis of our four-event data subset. We did not allow the value of c_3 at short periods to be less than that for PGA, thus placing an upper limit on $|c_3|$ at $|c_3|=0.01151$. Similarly, we fixed the values for long periods to be that determined for $T=3$ s, thus placing a lower limit on $|c_3|$ of $|c_3|=0.00191$ (we did not think it physically plausible for the anelastic attenuation to increase with period at T >3 s).

We also constrained the c_3 values for the PGV regressions to be that for the $T=1.0$ s regression. This choice is a compromise between the similarity in larger-

magnitude scaling that we observed between PGV and PSA at 2 s and the recommendation of Bommer and Alarcón (2006) that PGV is related to PSA at 0.5 s.

Determination of h: It is desirable to constrain the pseudo-depth h in the regression in order to avoid overlap in the curves for large earthquakes at very close distances. We did this by performing initial regressions with h as a free parameter, then modifying the resultant values of h as required to avoid overlap in the spectra at close distances (for the reference site condition of 760 m/s). In this regression, c_1 was a free variable and c_3 was constrained to a set of initially-determined values. We fit the values determined with h as a free parameter with a quadratic, but we observed that the h value at 0.05 s from the quadratic fit was very small, much below that determined for PGA. We increased the h value at 0.05 s to match the value for a regression of PGA with h unconstrained, and refit the quadratic with this change in the data points. We used the modified quadratic as the basis for assigning h for all periods. The value of h at short periods was guided by the unequivocal statement that PSA is equal to PGA at periods much less than 0.1 s. For PGA, we adopted the value implied by the modified quadratic for the $T=0.05$ s oscillator. We then assigned values of h for periods between 0.01 s and 0.05 s to be the same as that for 0.05 s. Consistent with the convention adopted for the c_3 coefficient, we used the value of h at 1 s for PGV.

These analyses established smooth, constrained values for c_3 and h that facilitated robust and well behaved determinations of the remaining parameters by regression of the NGA database.

Determination of c1, c2, and σ: With h and c_3 constrained, we regressed the response variables of the NGA database to solve for c_1 and c_2 (Equation 3), along with the event terms (c_0) for each earthquake, using all data (subject to the exclusions discussed earlier) for distances less than 400 km. The c_1 coefficient is the effective geometric spreading rate (slope) for an event of $\mathbf{M}=\mathbf{M}_{\mathrm{ref}}$, while the c_2 coefficient provides a means to describe magnitude-dependent distance decay (it changes the slope for events that are greater or smaller than $\mathbf{M}_{\mathrm{ref}}$). The intra-event aleatory uncertainty σ is given by the standard deviation of the residuals from the Stage 1 regression.

The regression used assigned values for the reference distance, R_{ref}, at which near-source predictions are pegged, and for the reference magnitude, M_{ref}, to which the magnitude dependence of the geometric spreading is referenced. The assigned values for these reference values are arbitrary and are largely a matter of convenience. For M_{ref}, we chose a value of 4.5, since this is the approximate magnitude of much of the data used to determine the fixed c_3 coefficients; this choice means that the magnitude dependence of the slope will be referenced to that observed for small events. For R_{ref}, we use the value of 1 km. This is convenient because the curves describing the distance dependence pivot around $R=R_{ref}$. The curves for larger magnitudes are flatter than for smaller magnitudes, which can lead to those curves being below the curves for smaller magnitudes at distances less than the pivot distance. This was avoided by choosing $R_{ref}=1$ km, although any value such that $R_{ref}<\min(h)$, where the minimum is taken over all periods,

would prevent undesirable overlapping of prediction curves near the source (i.e. we want to ensure that R will always be greater than the pivot distance of R_{ref}, even when R_{JB} $=0$ km).

Magnitude Dependence (Stage 2)

The event terms (coefficients $(c_0)_j$ in Equation 14 from the Stage 1 regression were used in a weighted Stage 2 regression to determine the magnitude scaling of the response variables. As discussed in Joyner and Boore (1993), the Stage 2 weighted regression was iterative in order to solve for the inter-event variability τ. The basic form we selected for the magnitude scaling is a quadratic, similar to the form used by BJF97. However, we imposed a constraint that the quadratic not reach its maximum at $\mathbf{M}$ <8.5, in order to prevent "oversaturation" (the prediction of decreasing amplitudes with increasing magnitude). The following algorithm was used to implement the constrained quadratic magnitude dependence:

1. Fit the event terms $(c_0)_j$ for a given period to a second-order polynomial. If the $\mathbf{M}$ for which the quadratic starts to decrease ($\mathbf{M}_{\max}$) is greater than 8.5, we adopt this regression for the magnitude dependence for this period.
2. If $\mathbf{M}_{\max}$ for a given period is less than 8.5, we perform a two-segment regression, hinged at $\mathbf{M}_h$ (described below), with a quadratic for $\mathbf{M} \leq \mathbf{M}_h$ and a linear function for $\mathbf{M}_h < \mathbf{M}$. If the slope of the linear function is positive, we adopt this two-segment regression for the magnitude dependence for this period.
3. If the slope of the linear segment is negative, we redo the two-segment regression for that period, constraining the slope of the line above $\mathbf{M}_h$ to be 0.0. Note that the equations for almost all periods less than or equal to 1.0 s required the constraint of zero slope; this is telling us that for short periods the data actually indicate oversaturation. We felt that because of limited data and knowledge, oversaturation was too extreme at this stage of equation development, and we chose to impose saturation rather than allow the data to dictate an oversaturated form. More observations from ground motions near large earthquakes, as well as theoretical simulations using dynamic rupture models (e.g., Schmedes and Archuleta, 2007) may give us confidence in allowing oversaturation in future versions of GMPEs.

Choice of $\mathbf{M}_h$: The parameter $\mathbf{M}_h$ is the hinge magnitude at which the constrained magnitude scaling in the two-segment regression changes from the quadratic form to the linear form. Subjective inspection of nonparametric plots of data clearly indicated that near-source ground motions at short periods do not get significantly larger with increasing magnitude, beyond a magnitude in the range of 6.5 to 7, and therefore we set $\mathbf{M}_h$ within this range.

Fault-Type Dependence: Plots of event terms against magnitude (presented later) showed that normal-fault earthquakes have amplitudes that are consistently below those for strike-slip and reverse earthquakes for most periods (others have found similar results, including Spudich et al. 1999, Bommer et al. 2003, and Ambraseys et al. 2005). We used this observation to guide our determination of the dependence on fault type. We first grouped the data from all fault types together and solved for the coefficients e_1, e_5,

e_6, e_7, and e_8 in Equation 5a and 5b), setting e_2, e_3, and e_4 to 0.0. The regression was then repeated, fixing the coefficients e_5, e_6, e_7, and e_8 to the values obtained when lumping all fault types together, and solving for the coefficients e_2, e_3, and e_4 of the fault type dummy variables SS, NS, and RS. Thus we have constrained the relative scaling of amplitudes with magnitude to be the same for all event types, but we allow an offset in the average predicted amplitude level according to the fault mechanism. The inter-event aleatory uncertainty (τ) was slightly different for these two cases, so subscripts "U" and "M" were used to distinguish between unspecified and specified fault type, respectively, in the table of aleatory uncertainties. Note that the term "unspecified" is strictly applicable to a random selection of an earthquake from the distribution of fault types used in our analysis; it is an accurate description of a truly random selection from all earthquakes only to the extent that the distribution of all fault types is equal to the distribution used in our analysis.

All analyses were done using Fortran programs developed by the first author, in some cases incorporating legacy code from programs and subroutines written by W. B. Joyner.

RESULTS

COEFFICIENTS OF THE EQUATIONS

The coefficients for the GMPEs are given in Tables 3, 4, and 6–8. The coefficients are for ln Y, where Y has units of g for PSA and PGA and cm/s for PGV. The units of distance and velocity are km and m/s, respectively. The equation for *pga4nl* is the same as for PGA, with $V_{S30} > 760$ m/s (for which $F_S = 0$) (Boore and Atkinson, 2008).

There are no normal-fault data are in our data set for an oscillator period of 10 s, and thus formally we could not obtain the coefficient e_3 for that period; the value in Table 7 was obtained using the assumption that the ratio of motions for normal and unspecified faults is the same for periods of 7.5 s and 10 s. With this assumption, $e_3(10s) = e_1(10s) + (e_3(7.5s) - e_1(7.5s))$.

Fit of the Stage 1 Regressions

BA07 contains a series of graphs showing the observations in comparison to the Stage 1 regression predictions. These figures provide a visual test of the ability of our functional form to represent the distance dependence of the response variables. A more precise way of looking for systematic mismatches between predictions and observations is to plot the residuals from the Stage 1 analysis, defined as the ratio of observed to predicted ground motions. Figure 7 shows residuals as a function of distance for PGA and for PSA at 10 s; these span the range of seismic intensity measures included in our equations (the graphs of residuals for PSA at most oscillator periods are similar to that for PGA in Figure 7—see BA07 for a complete set of graphs). For the sake of clarity, we have separated the residuals into different magnitude ranges and for two specific earthquakes in Figure 7. Log residuals averaged over distance bins 0.1 log unit in width and magnitude bins 1 unit in width are shown in Figure 8 for two representative periods; the graphs are grouped by values of V_{S30}. While there are some small departures from a null residual (values of 1 and 0 in Figures 7 and 8, respectively), there are no significant

Table 6. Distance-scaling coefficients ($\mathbf{M}_{ref}$=4.5 and R_{ref}=1.0 km for all periods, except R_{ref}=5.0 km for *pga4nl*)

Period	c_1	c_2	c_3	h
PGV	−0.87370	0.10060	−0.00334	2.54
PGA	−0.66050	0.11970	−0.01151	1.35
0.010	−0.66220	0.12000	−0.01151	1.35
0.020	−0.66600	0.12280	−0.01151	1.35
0.030	−0.69010	0.12830	−0.01151	1.35
0.050	−0.71700	0.13170	−0.01151	1.35
0.075	−0.72050	0.12370	−0.01151	1.55
0.100	−0.70810	0.11170	−0.01151	1.68
0.150	−0.69610	0.09884	−0.01113	1.86
0.200	−0.58300	0.04273	−0.00952	1.98
0.250	−0.57260	0.02977	−0.00837	2.07
0.300	−0.55430	0.01955	−0.00750	2.14
0.400	−0.64430	0.04394	−0.00626	2.24
0.500	−0.69140	0.06080	−0.00540	2.32
0.750	−0.74080	0.07518	−0.00409	2.46
1.000	−0.81830	0.10270	−0.00334	2.54
1.500	−0.83030	0.09793	−0.00255	2.66
2.000	−0.82850	0.09432	−0.00217	2.73
3.000	−0.78440	0.07282	−0.00191	2.83
4.000	−0.68540	0.03758	−0.00191	2.89
5.000	−0.50960	−0.02391	−0.00191	2.93
7.500	−0.37240	−0.06568	−0.00191	3.00
10.000	−0.09824	−0.13800	−0.00191	3.04

trends in magnitude, distance, or shear-wave velocity. We therefore judge the fit between observations and our predictions to be reasonable. In particular, we note that the imposed soil response coefficients appear to be adequate, as evidenced by the apparent fit over the three distinct ranges of shear-wave velocity used in Figure 8; the fit is good at both short and large distances over all magnitude ranges, which implicitly supports the degree of nonlinearity that was specified.

Fit of the Stage 2 Regressions

Figure 9 is a plot of the antilog of the event terms $(c_0)_j$ from the Stage 1 regression as a function of magnitude, with the Stage 2 regression fit to these terms superimposed. The fault type for each earthquake is indicated, as are curves for fault type unspecified and for strike-slip, normal, and thrust/reverse faults (the fault type is indicated by the color of the symbols). The functional form provides a reasonable fit to the near-source

Table 7. Magnitude-scaling coefficients

Period	e_1	e_2	e_3	e_4	e_5	e_6	e_7	$\mathbf{M}_h$
PGV	5.00121	5.04727	4.63188	5.08210	0.18322	−0.12736	0.00000	8.50
PGA	−0.53804	−0.50350	−0.75472	−0.50970	0.28805	−0.10164	0.00000	6.75
0.010	−0.52883	−0.49429	−0.74551	−0.49966	0.28897	−0.10019	0.00000	6.75
0.020	−0.52192	−0.48508	−0.73906	−0.48895	0.25144	−0.11006	0.00000	6.75
0.030	−0.45285	−0.41831	−0.66722	−0.42229	0.17976	−0.12858	0.00000	6.75
0.050	−0.28476	−0.25022	−0.48462	−0.26092	0.06369	−0.15752	0.00000	6.75
0.075	0.00767	0.04912	−0.20578	0.02706	0.01170	−0.17051	0.00000	6.75
0.100	0.20109	0.23102	0.03058	0.22193	0.04697	−0.15948	0.00000	6.75
0.150	0.46128	0.48661	0.30185	0.49328	0.17990	−0.14539	0.00000	6.75
0.200	0.57180	0.59253	0.40860	0.61472	0.52729	−0.12964	0.00102	6.75
0.250	0.51884	0.53496	0.33880	0.57747	0.60880	−0.13843	0.08607	6.75
0.300	0.43825	0.44516	0.25356	0.51990	0.64472	−0.15694	0.10601	6.75
0.400	0.39220	0.40602	0.21398	0.46080	0.78610	−0.07843	0.02262	6.75
0.500	0.18957	0.19878	0.00967	0.26337	0.76837	−0.09054	0.00000	6.75
0.750	−0.21338	−0.19496	−0.49176	−0.10813	0.75179	−0.14053	0.10302	6.75
1.000	−0.46896	−0.43443	−0.78465	−0.39330	0.67880	−0.18257	0.05393	6.75
1.500	−0.86271	−0.79593	−1.20902	−0.88085	0.70689	−0.25950	0.19082	6.75
2.000	−1.22652	−1.15514	−1.57697	−1.27669	0.77989	−0.29657	0.29888	6.75
3.000	−1.82979	−1.74690	−2.22584	−1.91814	0.77966	−0.45384	0.67466	6.75
4.000	−2.24656	−2.15906	−2.58228	−2.38168	1.24961	−0.35874	0.79508	6.75
5.000	−1.28408	−1.21270	−1.50904	−1.41093	0.14271	−0.39006	0.00000	8.50
7.500	−1.43145	−1.31632	−1.81022	−1.59217	0.52407	−0.37578	0.00000	8.50
10.000	−2.15446	−2.16137	−2.53323	−2.14635	0.40387	−0.48492	0.00000	8.50

amplitude data. Note that the magnitude scaling for T=10 s at $\mathbf{M}<6.5$ is strongly controlled by the data from only one small earthquake (2000 Yountville, **M** 5.0), and may therefore be unreliable for $\mathbf{M}<6.5$.

Predictions of PSA from Combined Stage 1 and Stage 2 Regressions

Graphs of PSA predicted from our equations for three values of R_{JB} and four magnitudes are shown in Figure 10. The curves for the larger earthquakes tend to squeeze together for periods near 0.2–0.3 s, probably a reflection of the pinching together of the effective geometric spreading factor for these periods. But otherwise the PSA are quite smooth, especially considering that many of the coefficients were determined independently for each period.

Plots of PSA as a function of distance are shown in Figure 11 for two representative periods (see BA07 for plots at other periods). The figure is for V_{S30}=760 m/s (NEHRP B/C boundary).

The effect of V_{S30} on predicted ground-motion amplitude is shown in Figure 12. Nonlinear soil amplification causes the curves to cross, such that at close distances lower

Table 8. Aleatory uncertainties (σ: intra-event uncertainty; τ: inter-event uncertainty; σ_T: combined uncertainty ($\sqrt{\sigma^2+\tau^2}$); subscripts U, M for fault type unspecified and specified, respectively)

Period	σ	τ_U	σ_{TU}	τ_M	σ_{TM}
PGV	0.500	0.286	0.576	0.256	0.560
PGA	0.502	0.265	0.566	0.260	0.564
0.010	0.502	0.267	0.569	0.262	0.566
0.020	0.502	0.267	0.569	0.262	0.566
0.030	0.507	0.276	0.578	0.274	0.576
0.050	0.516	0.286	0.589	0.286	0.589
0.075	0.513	0.322	0.606	0.320	0.606
0.100	0.520	0.313	0.608	0.318	0.608
0.150	0.518	0.288	0.592	0.290	0.594
0.200	0.523	0.283	0.596	0.288	0.596
0.250	0.527	0.267	0.592	0.267	0.592
0.300	0.546	0.272	0.608	0.269	0.608
0.400	0.541	0.267	0.603	0.267	0.603
0.500	0.555	0.265	0.615	0.265	0.615
0.750	0.571	0.311	0.649	0.299	0.645
1.000	0.573	0.318	0.654	0.302	0.647
1.500	0.566	0.382	0.684	0.373	0.679
2.000	0.580	0.398	0.702	0.389	0.700
3.000	0.566	0.410	0.700	0.401	0.695
4.000	0.583	0.394	0.702	0.385	0.698
5.000	0.601	0.414	0.730	0.437	0.744
7.500	0.626	0.465	0.781	0.477	0.787
10.000	0.645	0.355	0.735	0.477	0.801

values of V_{S30} (softer sites) will have lower predicted amplitudes than stiffer sites, due to nonlinear deamplification. The effect is more pronounced at short periods than at long periods.

Surface Slip vs. No-Surface Slip Earthquakes

Several authors (e.g., Somerville and Pitarka 2006) have proposed that the high-frequency ground motions from earthquakes with faults that break to the surface are smaller than from those with faults that remain buried. We search for evidence of this effect in Figure 13, which shows the event-term residuals from the Stage 1 regression plotted against **M** for the two classes of earthquakes. The first thing to notice is that most surface-slip earthquakes correspond to larger magnitudes, with almost no buried ruptures for magnitude greater than **M**=7. For this reason any reduction in motions for surface-slip earthquakes will be mapped into reduced magnitude scaling in the Stage 2 magnitude regression. In order to differentiate magnitude scaling from the effects of surface versus buried rupture, data from both class of rupture are needed for the same range

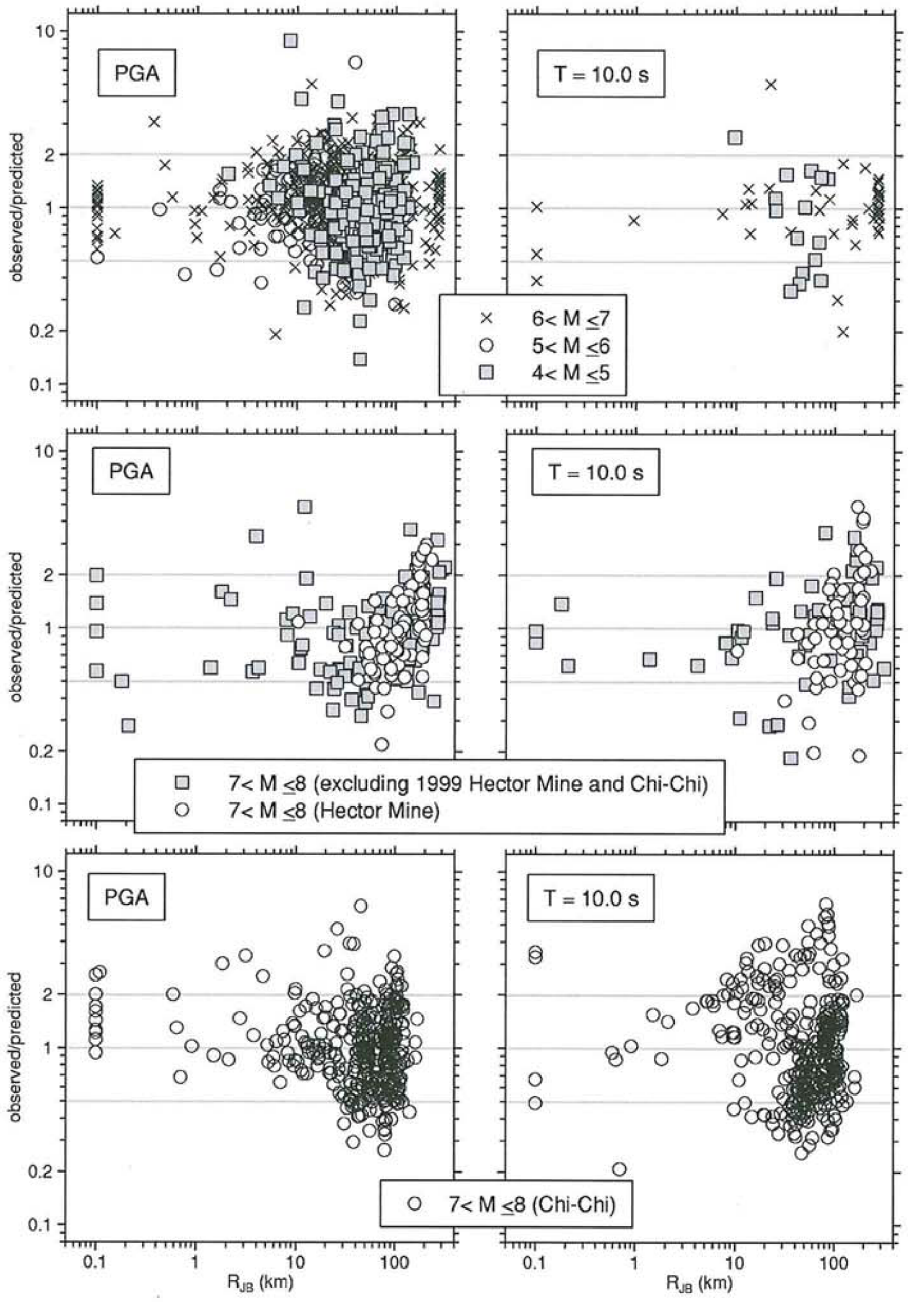

Figure 7. Stage 1 residuals, separated by magnitude, for PGA and T=10 s PSA (these span the range of possibilities; see BA07 for more plots). The residuals for the 1999 Chi-Chi earthquake are shown separately, in the bottom two graphs.

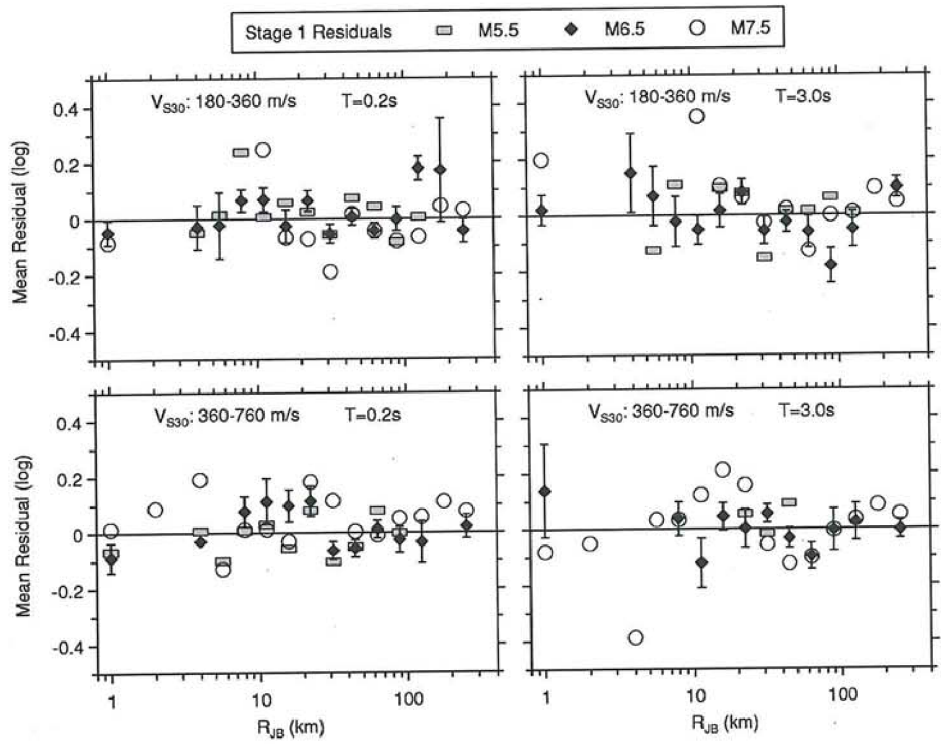

Figure 8. Stage 1 regression residuals (log10 units) for 0.2-s and 3-s response spectra, averaged over distance bins 0.1 log unit in width and magnitude bins 1 unit in width. Only bins with at least three observations are plotted. The standard error of the mean is shown for the middle magnitude bin (6 to 7) only. Two ranges of shear-wave velocity are shown: 180–360 m/s (top), 360–760 m/s (bottom). No residuals are shown for $V_{S30} > 760$ m/s because of the small number of observations in that velocity range.

of magnitudes. As seen in Figure 13, it is only for strike-slip earthquakes that there is more than one of each class of earthquake in a common magnitude range (there are several strike-slip events of 5.7–6.7 in both classes). There is no indication for these earthquakes that the event-term residuals are systematically different for the two classes of data. Therefore, there was no need to include dummy variables for surface-slip/buried earthquakes in our functional forms. As confidence in simulations from dynamic models of rupture propagation increases, or if additional data change our understanding, it might be that in the future we will add a buried/surface faulting term to the equations. By doing so, the apparent saturation of the magnitude scaling would not be as dramatic (i.e., the larger earthquakes are entirely surface slip events, and if these produce smaller ground motions than buried events, as has been suggested by Somerville and colleagues

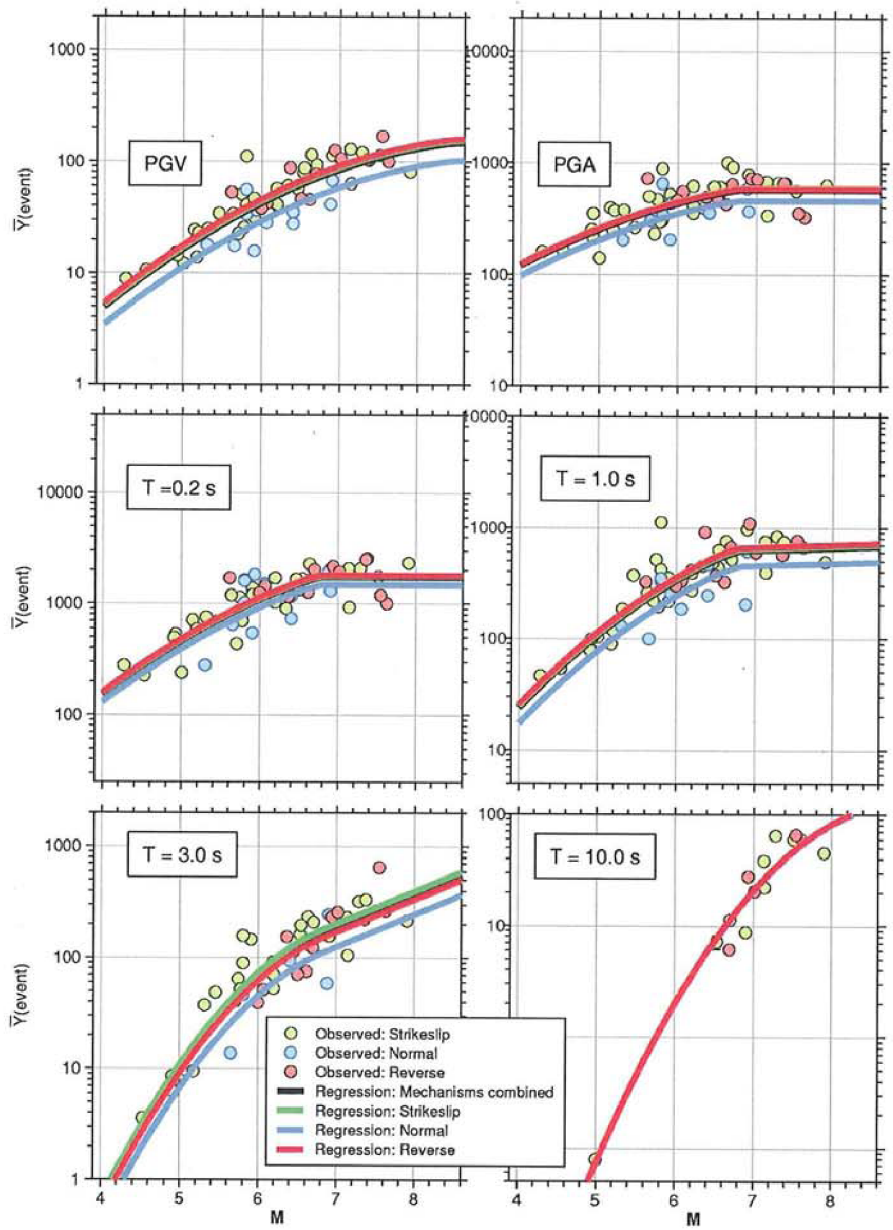

Figure 9. $\bar{Y}(event)$ (the antilog of the average of the log of the motions for each event adjusted to the reference distance of 1 km and the reference velocity of 760 m/s; in the terminology of Equations 14 and 15, $\bar{Y}(event) = \exp(c_0(event)))$ and Stage 2 regression fits. Note that the same vertical scale was used for all graphs in order to compare the magnitude scaling from period to period.

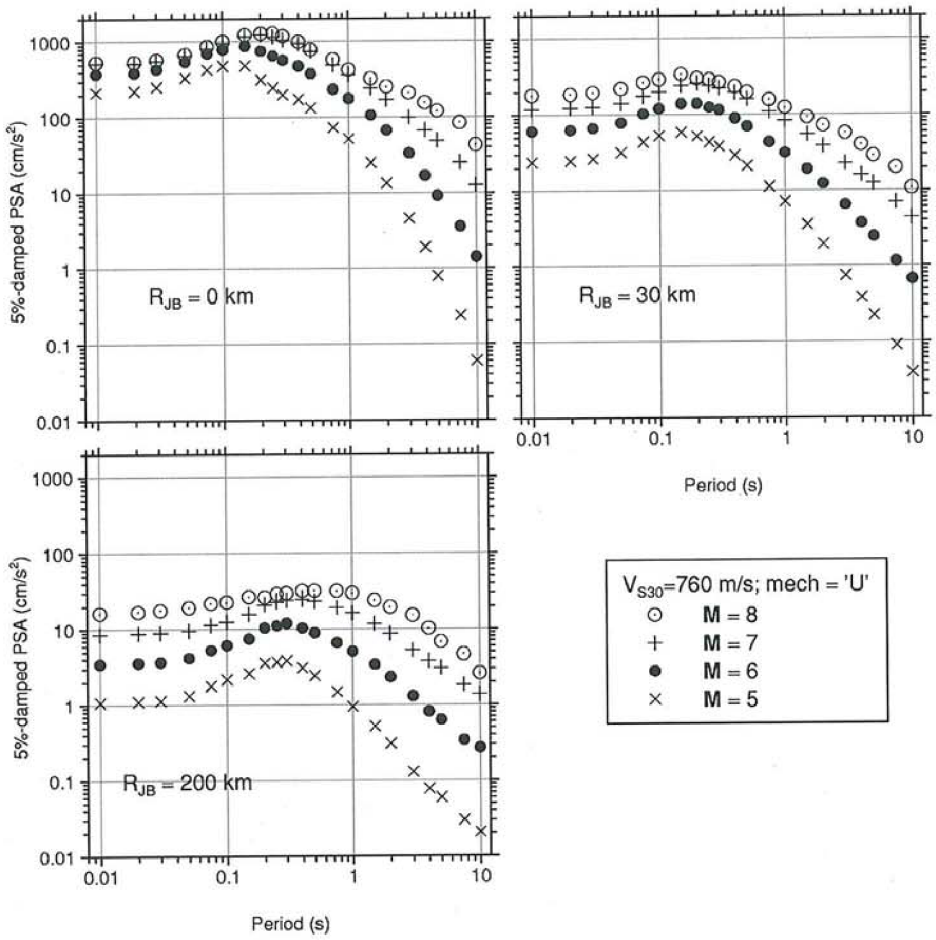

Figure 10. PSA from our equations, as function of period. The spectra are shown for three distances and four magnitudes, for fault type unspecified and V_{S30}=760 m/s.

(e.g., Somerville and Pitarka 2006), then there will be an apparent tendency for saturation if the events are not separated into two classes according to whether they break to the surface or not).

Dependence of Stage 1 Residuals on Basin Depth

Another ground-motion effect that we searched for in the residuals of the Stage 1 regression was that of basin depth. Basin-depth effects on ground-motion amplitudes have been reported in empirical studies (Field, 2000; Choi et al. 2005), and from simulations (Day et al. 2008). One of the reasons that we did not include a basin-depth term in our equations is indicated in Figure 14, which shows the distribution of V_{S30} and a

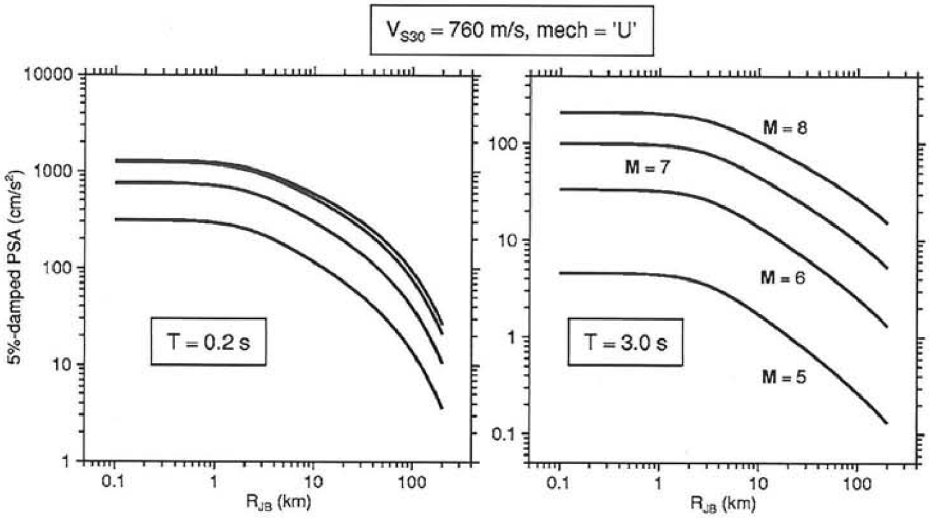

Figure 11. PSA from our equations, as a function of distance. The spectra are shown for magnitudes 5, 6, 7, and 8, for fault type unspecified and V_{S30}=760 m/s. Note that the same vertical scale was used for both graphs in order to compare the magnitude scaling for the two periods.

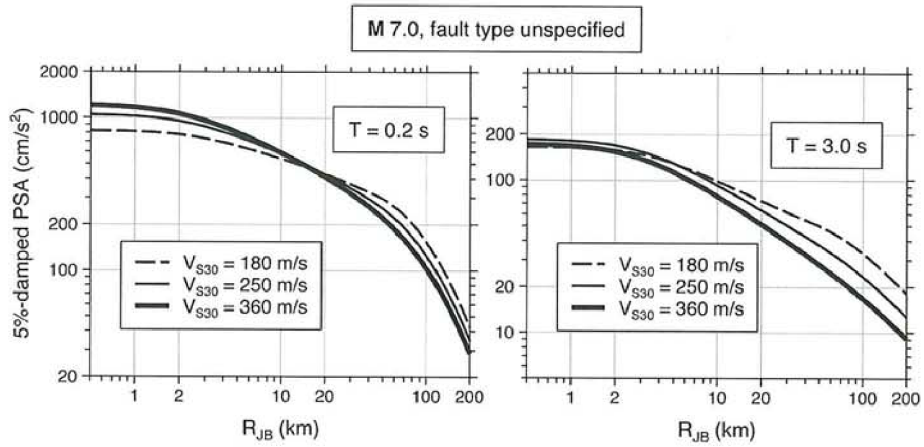

Figure 12. PSA from our equations, as a function of distance for magnitude 7, fault type unspecified, and three values of V_{S30}. Note that the same vertical scale was used for both graphs in order to compare the magnitude scaling for the two periods.

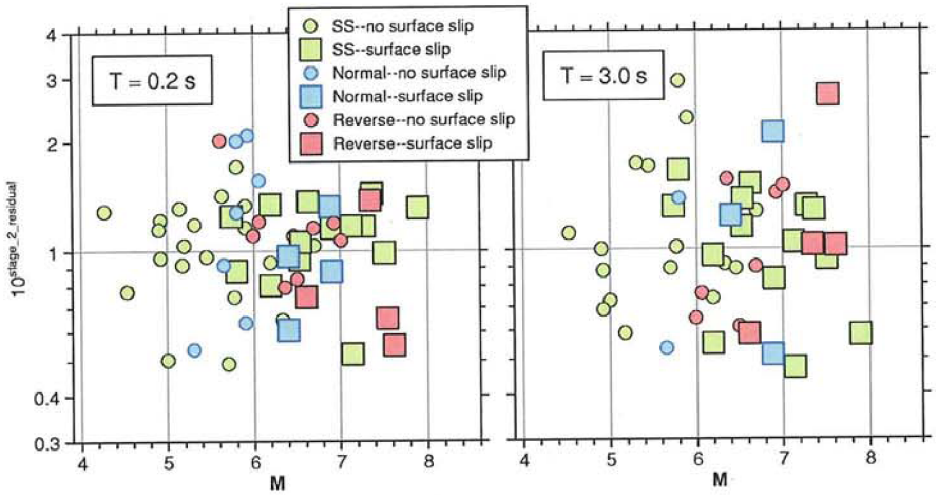

Figure 13. Antilogarithms of Stage 2 residuals, plotted against magnitude and differentiated by events of different fault types, for which faults did or did not break to surface.

measure of basin depth. The plot shows all data in the NGA Flatfile for which both V_{S30} have been measured and basin depth have been estimated. It is clear that the softer sites are in basins, and hence basin depth and V_{S30} are strongly correlated (this was found previously by Choi et al. 2005, Figure 7). Therefore any basin depth effect will tend to

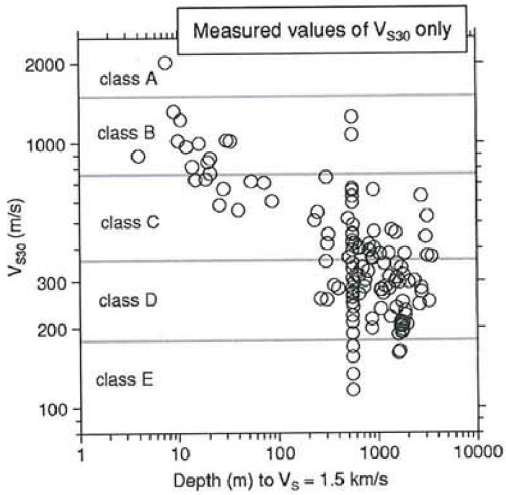

Figure 14. V_{S30} plotted against one measures of basin depth: the depth to a shear-wave velocity of 1.5 km/s. All values in the NGA Flatfile with basin depths and measured (rather than estimated) values of V_{S30} are shown.

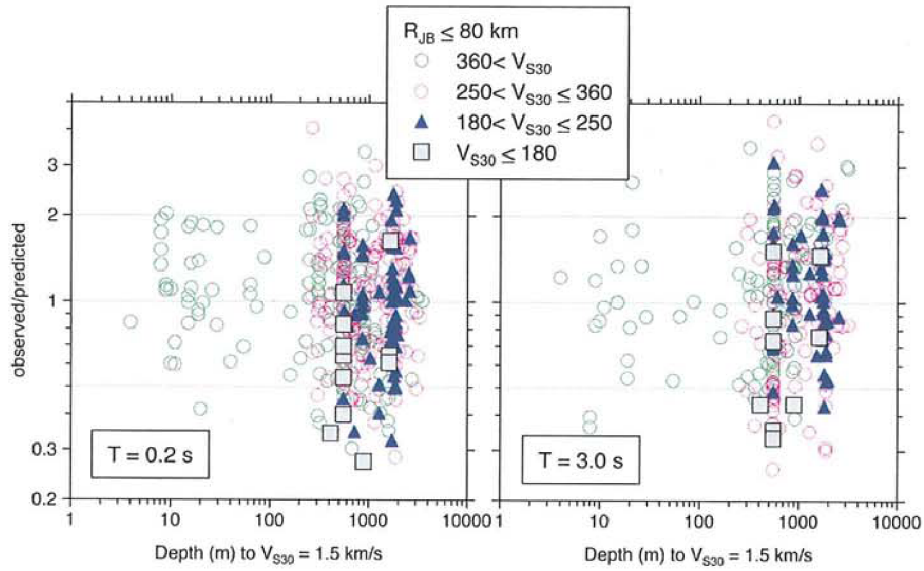

Figure 15. Stage 1 residuals plotted against depth to V_S=1.5 km/s, differentiated by V_{S30}, for $R_{JB} \leq 80$ km.

have been captured by the empirically-determined site amplification. To try to separate the amplification and the basin-depth effects in the data would require use of additional information or assumptions. Since we are opting for the simplest equations required by the data, no attempt was made to break down the site-response function into basin depth and the amplification terms.

We searched for any uncaptured basin depth effect by examining the residuals of the Stage 1 regressions. We find that the residuals have no dependence on basin depth, except for a trend to positive residuals (underprediction) for long periods at distances beyond 80 km (by about a factor of 1.6), for sites having depth-to-1.5 km/s > 700 m. (The trend for greater distances is shown in a figure not included here.) These residuals could also be due to regional variations in the distance function, along with correlations between distance and the basin depth. Figure 15 contains plots of the Stage 1 residuals against the depth-to-V_{S30}=1.5 km/s; only residuals for $R_{JB} \leq 80$ km are shown, in order not to map mismatches in the more distant attenuation into the residuals. There is no obvious dependence of the residuals on basin depth. But assuming that the positive residuals at distances greater than 80 km are due solely to a basin depth effect, the trends indicate that our equations may underpredict long-period motions at large distances from sites in deep basins. For shallower basins and at shorter distances, we find no basin depth effect. This is not surprising in light of the observations made above regarding the correlation between basin parameters and V_{S30}. (Note: similar results were obtained when the depth to 2.5 km/s was used as the measure of basin depth.) Another reason for

little apparent basin effect has to do with the way in which basins affect incoming waves; Choi et al. (2005) found a basin effect for sources inside the basin in which the motions were recorded, but little effect for sources outside the basin; they attribute the difference to the manner in which incoming waves are converted and refracted upon entering the basin. As many of our data come from earthquakes that occurred outside the basins in which they were recorded, a similar explanation might apply to our finding.

Comparison of GMPEs Developed With and Without the 1999 Chi-Chi Earthquake

Because the Chi-Chi earthquake forms a significant fraction of the data set we used in developing our equations, it is important to see how the equations would change if the data from the Chi-Chi earthquake were eliminated from both the Stage 1 and the Stage 2 regressions. We therefore repeated the complete analysis without the Chi-Chi data. Figure 16 compares selected ground-motion intensity measures given by the two sets of equations. The figure also shows the percent of data used in the regression analysis from the Chi-Chi earthquake (the number of Chi-Chi recordings is the numerator of the ratio). It is clear that the fraction of the data set contributed by the Chi-Chi earthquake increases with period, reaching 64% of the data set for a period of 10 s. For this reason it is not surprising that the predictions of 10 s PSA are quite different for the equations developed with and without the Chi-Chi data (the ordinate scales of all graphs in Figure 16 are the same, to facilitate comparisons of the relations between the two predictions between periods); at intermediate to short periods, however, the differences are not dramatic. Interestingly, the differences can occur even at small magnitudes (despite the fact that we include only the Chi-Chi mainshock, not its aftershocks). We think the explanation of this apparent paradox is that the Chi-Chi earthquake is very well recorded and thus dominates the Stage 1 regression, for which each recording of an earthquake has equal weight in determining the distance terms in the equations. These distance terms then affect the event terms, and this in turn controls the magnitude scaling. We conclude that although the Chi-Chi earthquake affects the GMPEs, it is only a major controlling factor in the predictions of PSA at periods of greater than 5 s.

Comparison of BA07 and BJF97 GMPEs

It is interesting to compare our new predicted ground motions with those from the Boore et al. (1997) (BJF97) equations. Figure 17 (top row) compares the magnitude-distance distribution of the data used in each study. It is apparent that many more data are used in the new equations; the NGA data fill gaps at close distances for all magnitudes, add more data at small magnitudes at all distances, add data for large magnitudes, and fill out the distribution so that no longer is there a strong correlation between distance and magnitude in the data set. For this reason, the new equations provide a more robust prediction of ground-motion amplitudes over a wide range of magnitudes and distances.

We compare predicted ground motions from the BJF97 equations and from our current equations in Figure 17 (bottom row), for V_{S30}=420 m/s, which is near the weighted geometric mean of the velocities for the sites used in the BJF97 regression analysis. We

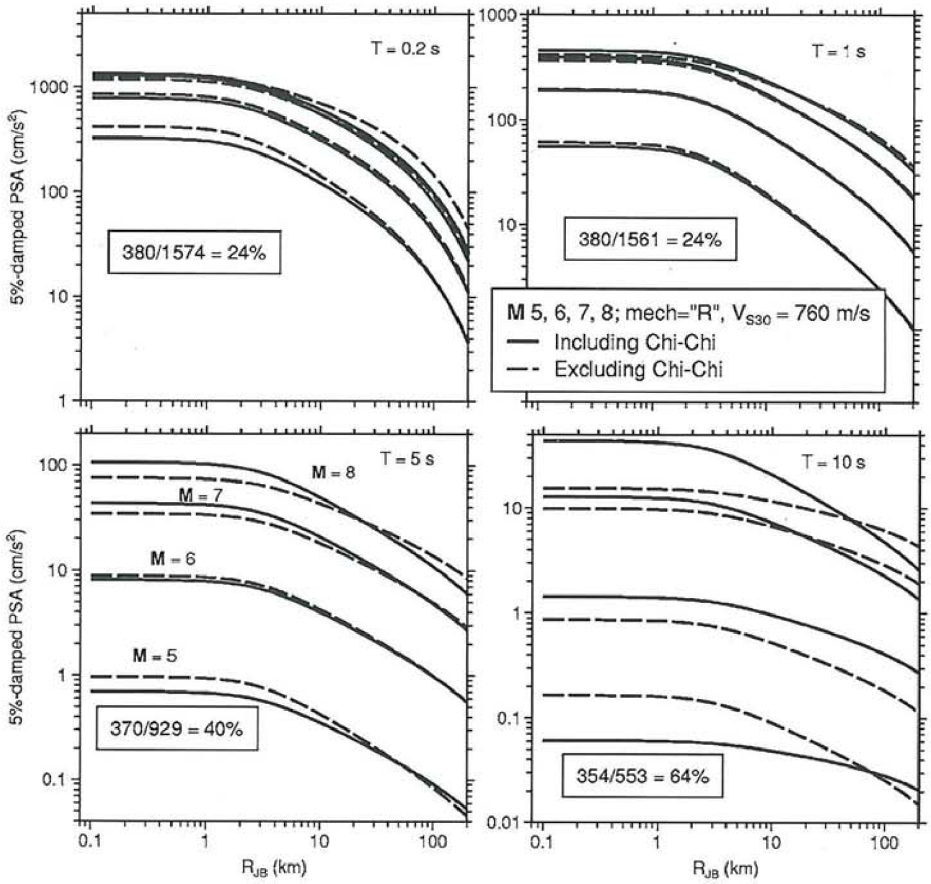

Figure 16. Comparisons of PSA for four periods from equations developed with and without 1999 Chi-Chi mainshock. Ratios are number of Chi-Chi recordings used to develop final equations divided by total number of recordings.

use the same scale for the ordinates in both graphs. The new and old equations predict similar amplitudes for **M** and R_{JB} ranges for which data were available for the BJF97 equation development. Large differences occur in regions of the magnitude-distance space for which data were not available in BJF97; the differences in the predicted values of seismic ground-motion intensity are largely attributable to the overly-simplified distance-independent magnitude scaling used in the BJF97 equations.

At all periods, the new equations predict significantly smaller motions than do the BJF97 equations for large magnitudes. This is probably the most important change in the new equations compared to the old equations. The difference in the predicted motions is particularly large for $T=1$ s and $\mathbf{M}=7.5$ (a factor of 2.4 at $R_{JB}=1$ km). Almost no data were available in BJF97 for $\mathbf{M}\sim 7.5$ and $R_{JB}<10$ km (see Figure 17), so discrepancies

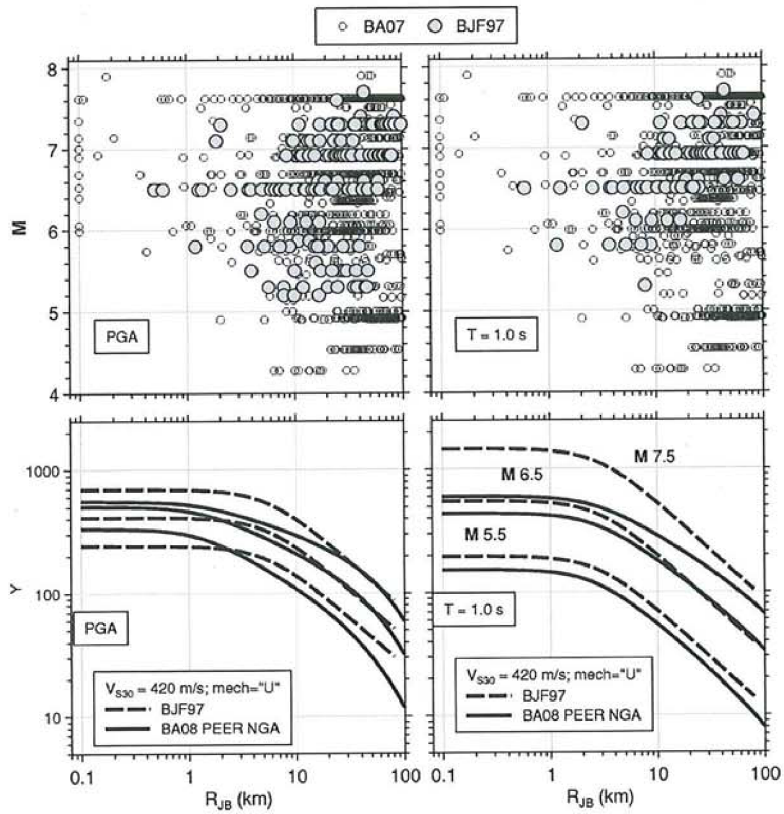

Figure 17. Top two graphs: Comparison of magnitude-distance distribution of data used by BJF97 and by BA08; bottom two graphs: PGA and T=1.0 s PSA predictions from BJF97 and from BA08 GMPEs.

are not surprising. The BJF97 data were for R_{JB} centered about 30 km. The discrepancy between the predictions from the BJF97 and the new equations is not nearly as strong for R_{JB} near 30 km as it is for $R_{JB} < 10$ km. Observed differences at $R_{JB} \approx 30$ km are likely due to including more data for large earthquakes in our current equations. The values of the BJF97 motions at close distances are strongly controlled by the assumption of distance-independent **M** scaling (and therefore the scaling at close distances is driven by the $R_{JB} \approx 30$ km data). The current equations allow for **M**-dependent distance scaling. Another effect that can reduce motions predicted from our equations at close distances from large earthquakes is nonlinear site response, which is not included in BJF97.

The total aleatory uncertainties, as well as the intra- and inter-event uncertainties, are significantly larger for the new equations than for the BJF97 equations (e.g., for a period of 0.2 s the total aleatory uncertainty is 0.60 for our equations and 0.44 for BJF97; more comparisons can be found in Table 4.6 of BA07). We are not sure of the reasons for the

differences, but we suspect the differences are due to a combination of more data providing a better sample of the true uncertainty, as well as an increase in the uncertainties produced by mixing data from regions for which the attenuation of the motions might be different (see Douglas 2007, for a study of regional differences in ground-motion prediction). The larger sigma values will offset to some extent the smaller ground motions for large magnitudes in the construction of seismic hazard maps. However, it is a point for further investigation how the aleatory uncertainties should be implemented in hazard analyses for a particular site, given that part of the aleatory uncertainty arises from mixing data from numerous regions.

GUIDELINES FOR USAGE

LIMITS ON PREDICTOR VARIABLES

We wish to emphasize that our equations should be used only for predictor variables in these ranges:

- $\mathbf{M}=5-8$
- $R_{JB}<200$ km
- $V_{S30}=180-1300$ m/s

These limits are subjective estimates based on the distributions of the recordings used to develop the equations.

PREDICTIONS FOR OTHER MEASURES OF SEISMIC INTENSITY

The NGA GMPEs are for the GMRotI measure of seismic intensity. Simple conversion factors between GMRotI and other measures of seismic intensity are given by Beyer and Bommer (2006) and Watson-Lamprey and Boore (2007), as well as by Campbell and Bozorgnia (2008).

DISCUSSION AND SUMMARY

We have presented a set of ground-motion prediction equations that we believe are the simplest formulation demanded by the NGA database used for the regressions. Future versions of the equations might include additional terms, such as basin depth, if these can be unambiguously supported by data. Expansion of the NGA database by way of additional or reprocessed data could potentially support the inclusion of more predictive variables. In spite of this, we note that the aleatory uncertainties in our equations are similar to those of other NGA developers who included more predictive variables. Therefore we do not think that our simplified analysis limits the usefulness of our equations, at least for those situations for which predictor variables not included in our equations are not crucial in site-specific hazard analysis.

One modification we would like to address in future versions of our equations is to account for regional variations in distance attenuation, particularly at distances beyond about 80 km. The near-source data could be used to constrain magnitude scaling for all regions, which could be patched onto regionally-dependent distance functions. The approach taken in this study, in which the anelastic coefficient was constrained using data

from a few earthquakes in central and southern California, is not optimal. Furthermore, there are inconsistencies in the pseudo-depths that might be attributed to forcing the values of the anelastic coefficient into the regression of the worldwide data set. Notwithstanding these limitations, the new relations developed here provide a demonstrably reliable description of recorded ground-motion amplitudes for shallow crustal earthquakes in active tectonic regions over a wide range of magnitudes and distances.

ACKNOWLEDGMENTS

This study was sponsored by the Pacific Earthquake Engineering Research Center's Program of Applied Earthquake Engineering Research of Lifelines Systems supported by the California Department of Transportation, the California Energy Commission, and the Pacific Gas and Electric Company.

This work made use of the Earthquake Engineering Research Centers Shared Facilities supported by the National Science Foundation, under award number EEC-9701568 through the Pacific Earthquake Engineering Research (PEER) Center. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation.

We have benefited from discussions and comments from many people. First and foremost, we want to thank the whole PEER NGA project team for the opportunity to participate in the project; all interactions with the members of the team were extraordinarily open and supportive. In addition, we thank these people, in alphabetical order: Sinan Akkar, Jack Boatwright, John Douglas, Art Frankel, Vladimir Graizer, Steve Harmsen, Robert Herrmann, Tom Holzer, Charles Mueller, Maury Power, Rakesh Saigal, Linda Seekins, and Chris Stephens. John Douglas and Jon Stewart provided excellent reviews that substantially improved the paper. Finally, we acknowledge the contributions of Bill Joyner to the development of the regression procedures and codes used in this work.

APPENDIX A. CHOICE OF V_{S30} FOR A NEHRP CLASS

The need sometimes arises to evaluate GMPEs for a particular NEHRP site class. Because the PEER NGA GMPEs use the continuous variable V_{S30} as the predictor variable for site amplification, the question naturally arises as to what value of V_{S30} to use for a specific NEHRP class. To explore that question, we used the distribution of V_{S30} values from the borehole compilation given in Boore (2003) and from the NGA Flatfile, and computed the geometric means of the average of the V_{S30} values in each NEHRP class.

We used the geometric mean of V_{S30} in each NEHRP class, as these will give the same value of ln Y as the average of the ln Y's obtained using the actual V_{S30} values in the data set. Here is the analysis:

Because

$$\ln Y \approx b \ln V_{30}$$

the average of ln Y for a number of V_{S30}'s in a site class is:

$$\overline{\ln Y} \approx b \frac{1}{N} \sum_{i=1}^{N} \ln(V_{30})_i$$

and the same value of ln Y is obtained using the value of V_{S30} given by:

$$\ln \overline{V_{30}} = \frac{1}{N} \sum_{i=1}^{N} \ln(V_{30})_i$$

But does that mean that the values of V_{S30} in the NGA database should be used to determine the average value of V_{S30} that will be substituted into the GMPEs for a given NEHRP site class? Yes, under the assumption that the distribution of V_{S30} in the NGA database is similar to the one that would be obtained if a random site were selected. We discuss this in more detail at the end of this appendix.

To determine the geometric means of V_{S30} from the NGA Flatfile, we used the Excel function *vlookup* to select only one entry per station. Figure A1 shows the histograms. For the Boore (2003) data set, we used values of V_{S30} for which the borehole velocities had to be extrapolated less than 2.5 m to reach 30 m. The top graph shows histograms for the Boore (2003) velocities; the middle graph shows histograms for NGA velocities for which the values of V_{S30} are based on measurements (source=0 and 5); and the bottom graph is for NGA values from measurements and estimations (source=0, 1, 2, and 5). In choosing the most representative value of V_{S30} for each NEHRP class, we gave most weight to the middle graph in Figure A1. Those histograms used more data than in Boore (2003), but they are not subject to the possible bias in using an estimated value of V_{S30}, in which the value might be based on the assignment of a NEHRP class to a site, with someone else's correlation between NEHRP class and V_{S30} (correlations that may or may not have used the geometric mean of V_{S30}). We are trying to find the appropriate value independently.

The gray vertical lines in Figure A1 are the geometric means in each NEHRP class for the data used for each graph; the black vertical lines in Figure A1 are the V_{S30} values we recommend be used for each NEHRP class; they are controlled largely by the analysis of the source=0 and 5 NGA data. Table A1 contains the values of V_{S30} determined for the different histograms. Based on these values, the second-to-last column in the table contains the observation-based representative values that could substituted into the NGA GMPEs for specific NEHRP classes. The last column contains another possible set of values for evaluating the GMPEs for a specific NEHRP class; these values are the geometric means of the velocities defining each NEHRP class, rounded to the nearest 5 m/s (e.g., for NEHRP class D the value from the class definition is $\sqrt{180 \times 360}$ =255 m/s).

As mentioned before, the values in the second-to-last column of Table A1 are valid representations of the different NEHRP classes if the distribution of velocities in the

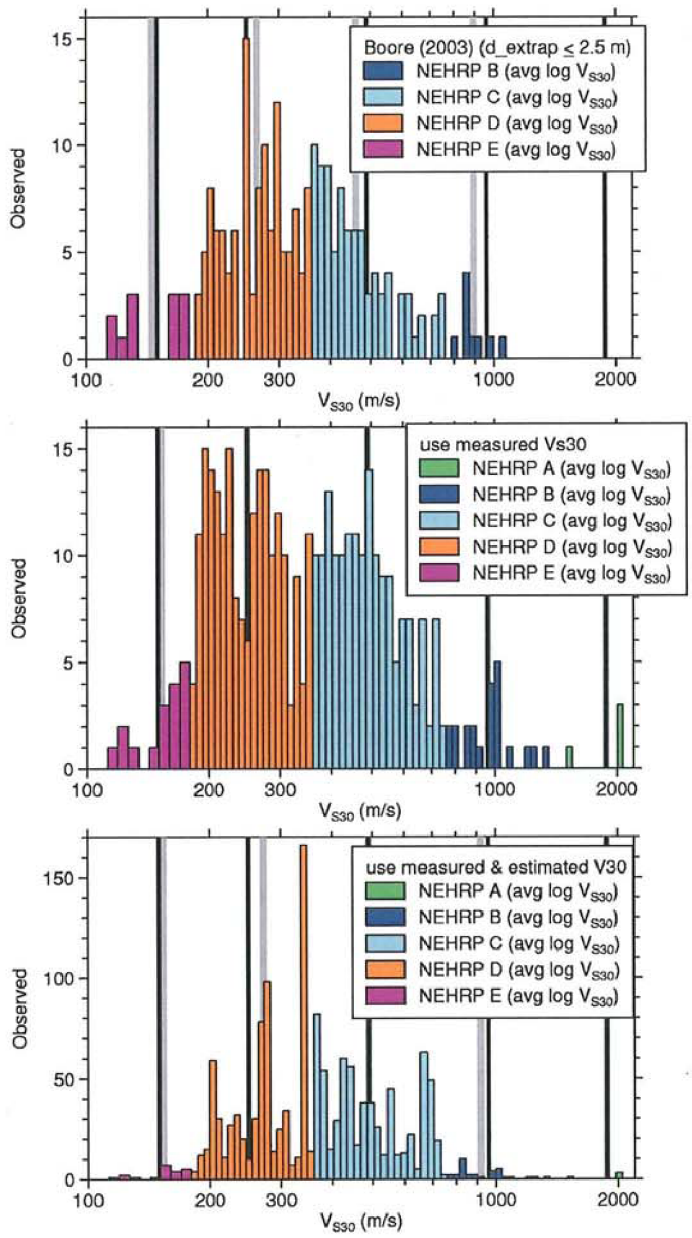

Figure A1. Histograms used to determine value of V_{S30} to use in evaluating NGA GMPEs for a particular NEHRP class (gray vertical lines are the geometrical means of the V_{S30} in each NEHRP site class, and black vertical lines are recommended values for each NEHRP class, as given in the second-to-last column in Table A1).

Table A1. Correspondence between NEHRP class and geometric mean V_{S30} (see text)

NEHRP Site Class	Geometric Mean of Measured V_{S30} in NGA Flatfile	Geometric Mean of Measured & Inferred V_{S30} in NGA Flatfile	Geometric Mean of V_{S30} in Boore (2003)	Suggested V_{S30} Based on Measured Velocities in NGA Flatfile	Geometric Mean of NEHRP Class Boundaries (rounded)
A	1880.5	1880.5		1880	
B	962.3	919.6	891.2	960	1070
C	489.8	489.9	461.4	490	525
D	249.8	271.5	263.7	250	255
E	153.3	153.7	145.0	150	

geographic region of interest is the same as that for the data used in the analysis above. Most of the measured values in the NGA database, however, come from the Los Angeles and San Francisco areas of California, so there is the potential for a bias if the V_{S30} values for those regions are not representative of a generic site. An alternative set of representative V_{S30} values for each NEHRP site class is given by the geometric mean of the velocities defining the site-class boundaries. These are given in the last column of Table A1. The values in the last two columns of Table A1 are similar, but to assess the impact of the two sets of representative values, we evaluated the ratios of ground motions for the two values for each NEHRP class, for a wide range of periods and distances. The differences in ground motions using the two possible sets of V_{S30} values are less than 8%, 5%, and 3% for NEHRP classes B, C, and D, respectively. The differences are largest at long periods for classes B and C and for short periods for class D. The differences in ground motions for each site class obtained using the alternative sets of representative V_{S30} values are so small that either set of could be used.

REFERENCES

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

Ambraseys, N. N., Douglas, J., Sarma, S. K., and Smit, P. M., 2005. Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: Horizontal peak ground acceleration and spectral acceleration, *Bull. Earthquake Eng.* **3**, 1–53.

Atkinson, G. M., 1993. Earthquake source spectra in eastern North America, *Bull. Seismol. Soc. Am.* **83**, 1778–1798.

Bakun, W. H., and Joyner, W. B., 1984. The M_L scale in central California, *Bull. Seismol. Soc. Am.* **74**, 1827–1843.

Benz, H. M., Frankel, A., and Boore, D. M., 1997. Regional *Lg* attenuation for the continental United States, *Bull. Seismol. Soc. Am.* **87**, 606–619.

Beyer, K., and Bommer, J. J., 2006. Relationships between median values and between aleatory

variabilities for different definitions of the horizontal component of motion, *Bull. Seismol. Soc. Am.* **96**, 1512–1522.

Boatwright, J., Bundock, H., Luetgert, J., Seekins, L., Gee, L., and Lombard, P., 2003. The dependence of PGA and PGV on distance and magnitude inferred from northern California ShakeMap data, *Bull. Seismol. Soc. Am.* **93**, 2043–2055.

Bommer, J. J., and Alarcón, J. E., 2006. The prediction and use of peak ground velocity, *J. Earthquake Eng.* **10**, 1–31.

Bommer, J. J., Douglas, J., and Strasser, F. O., 2003. Style-of-faulting in ground-motion prediction equations, *Bull. Earthquake Eng.* **1**, 171–203.

Boore, D. M., 1989. The Richter scale: Its development and use for determining earthquake source parameters, *Tectonophysics* **166**, 1–14.

——, 2003. *A compendium of P- and S-wave velocities from surface-to-borehole logging: Summary and reanalysis of previously published data and analysis of unpublished data, U. S. Geological Survey Open-File Report 03-191*, 13 pp.

Boore, D. M., and Atkinson, G. M., 1989. Spectral scaling of the 1985 to 1988 Nahanni, Northwest Territories, earthquakes, *Bull. Seismol. Soc. Am.* **79**, 1736–1761.

——, 2007. *Boore-Atkinson NGA ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters*, PEER *Report 2007/01*, Pacific Earthquake Engineering Center, Berkeley, California.

——, 2008. Notes on the equation to use for *pga4nl, unpublished notes*, available from http://quake.wr.usgs.gov/~boore/pubs_online.php.

Boore, D. M., Joyner, W. B., and Fumal, T. E., 1993. *Estimation of response spectra and peak accelerations from western North American earthquakes: An interim report, U. S. Geological Survey Open-File Report 93-509*, 72 pp.

——, 1994. *Estimation of response spectra and peak accelerations from western North American earthquakes: An interim report, part 2, U. S. Geological Survey Open-File Report 94-127*, 40 pp.

——, 1997. Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work, *Seismol. Res. Lett.* **68**, 128–153.

Boore, D. M., and Thompson, E. M., 2007. On using surface-source downhole-receiver logging to determine seismic slownesses, *Soil Dyn. Earthquake Eng.* **27**, 971–985.

Boore, D. M., Watson-Lamprey, J., and Abrahamson, N. A., 2006. GMRotD and GMRotI: Orientation-independent measures of ground motion, *Bull. Seismol. Soc. Am.* **96**, 1502–1511.

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 to 10 s, *Earthquake Spectra* **24**, 139–171.

Chiou, B. S.-J., Darragh, R., and Silva, W., 2008. An overview of the NGA database, *Earthquake Spectra* **24**, 23–44.

Chiou, B. S.-J., and Youngs, R. R., 2006. Chiou and Youngs PEER-NGA empirical ground motion model for the average horizontal component of peak acceleration and pseudo-spectral

acceleration for spectral periods of 0.01 to 10 seconds, available at: http://peer.berkeley.edu/products/Chiou_Youngs_NGA_2006.html

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

Choi, Y., Stewart, J. P., and Graves, R. W., 2005. Empirical model for basin effects that accounts for basin depth and source location, *Bull. Seismol. Soc. Am.* **95**, 1412–1427.

Day, S. M., Graves, R., 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.

Douglas, J., 2007. On the regional dependence of earthquake response spectra, *ISET J. Earthquake Technol.* **44**, 71–99.

Field, E. H., 2000. A modified ground-motion attenuation relationship for southern California that accounts for detailed site classification and a basin-depth effect, *Bull. Seismol. Soc. Am.* **90**, S209–S221.

Hutton, L. K., and Boore, D. M., 1987. The M_L scale in southern California, *Bull. Seismol. Soc. Am.* **77**, 2074–2094.

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

——, 1994. Errata, *Bull. Seismol. Soc. Am.* **84**, 955–956.

Mori, J., and Helmberger, D., 1996. Large-amplitude Moho reflections (SmS) from Landers aftershocks, southern California, *Bull. Seismol. Soc. Am.* **86**, 1845–1852.

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

Raoof, M., Herrmann, R., and Malagnini, L., 1999. Attenuation and excitation of three-component ground motion in sourthern California, *Bull. Seismol. Soc. Am.* **89**, 888–902.

Schmedes, J., and Archuleta, R. J., 2007. Oversaturation of peak ground velocity near strike slip faults [abstract], *Seismol. Res. Lett.* **78**, 272.

Somerville, P., and Pitarka, A., 2006. Differences in earthquake source and ground motion characteristics between surface and buried earthquakes, in *Proceedings, Eighth National Conference on Earthquake Engineering*, Paper No. 977.

Spudich, P., Joyner, W. B., Lindh, A. G., Boore, D. M., Margaris, B. M., and Fletcher, J. B., 1999. SEA99: A revised ground motion prediction relation for use in extensional tectonic regimes, *Bull. Seismol. Soc. Am.* **89**, 1156–1170.

Watson-Lamprey, J. A., and Boore, D. M., 2007. Beyond *SaGMRotI*: Conversion to Sa_{Arb}, Sa_{SN}, and Sa_{MaxRot}, *Bull. Seismol. Soc. Am.* **97**, 1511–1524.

(Received 21 June 2007; accepted 15 November 2007)