Bulletin of the Seismological Society of America, Vol. 93, No. 4, pp. 1703–1729, August 2003

Ⓔ

Empirical Ground-Motion Relations for Subduction-Zone Earthquakes and Their Application to Cascadia and Other Regions

by Gail M. Atkinson and David M. Boore

Abstract Ground-motion relations for earthquakes that occur in subduction zones are an important input to seismic-hazard analyses in many parts of the world. In the Cascadia region (Washington, Oregon, northern California, and British Columbia), for example, there is a significant hazard from megathrust earthquakes along the subduction interface and from large events within the subducting slab. These hazards are in addition to the hazard from shallow earthquakes in the overlying crust. We have compiled a response spectra database from thousands of strong-motion recordings from events of moment magnitude (**M**) 5–8.3 occurring in subduction zones around the world, including both interface and in-slab events. The 2001 **M** 6.8 Nisqually and 1999 **M** 5.9 Satsop earthquakes are included in the database, as are many records from subduction zones in Japan (Kyoshin-Net data), Mexico (Guerrero data), and Central America. The size of the database is four times larger than that available for previous empirical regressions to determine ground-motion relations for subduction-zone earthquakes. The large dataset enables improved determination of attenuation parameters and magnitude scaling, for both interface and in-slab events. Soil response parameters are also better determined by the data.

We use the database to develop global ground-motion relations for interface and in-slab earthquakes, using a maximum likelihood regression method. We analyze regional variability of ground-motion amplitudes across the global database and find that there are significant regional differences. In particular, amplitudes in Cascadia differ by more than a factor of 2 from those in Japan for the same magnitude, distance, event type, and National Earthquake Hazards Reduction Program (NEHRP) soil class. This is believed to be due to regional differences in the depth of the soil profile, which are not captured by the NEHRP site classification scheme. Regional correction factors to account for these differences are proposed for Cascadia and Japan.

The results of this study differ significantly from previous analyses based on more limited data and have important implications for seismic-hazard analysis. The ground-motion relations predict that a great megathrust earthquake (**M** $\geq$8) at a fault distance of about 100 km would produce pseudoacceleration (PSA), 5% damped, horizontal component on soil sites of about 110 cm/sec^2 at 0.5 Hz, 660 cm/sec^2 at 2.5 Hz, and 410 cm/sec^2 at 5 Hz, with a peak ground acceleration of about 180 cm/sec^2. These damaging levels of motion would be experienced over a very large area, corresponding to a rectangular area about 300 km wide by 500 km long. Large in-slab events (**M** 7.5) would produce even higher PSA values within 100 km of the fault, but the in-slab motions attenuate much more rapidly with distance. Thus the hazard posed by moderate to large in-slab events such as the 2001 Nisqually earthquake is modest compared to that of a Cascadia megathrust earthquake of **M** $\geq$8, in terms of the area that would experience damaging levels of ground motion.

Online material: table of data used in the regression analysis.

Introduction

Ground-motion relations for earthquakes that occur in subduction zones are an important input to seismic-hazard analyses for the Cascadia region (Washington, Oregon, northern California, and British Columbia). There is a significant hazard from megathrust earthquakes along the subduction interface and from large events within the subducting slab. We refer to these types of earthquakes as interface and in-slab, respectively. Despite recent moderate in-slab earthquakes in Washington, such as the moment magnitude (**M**) 6.8 2001 Nisqually and 1999 **M** 5.9 Satsop events, there remains a paucity of ground-motion data in the Cascadia region from which to develop regression relations. Therefore it is common practice in regional seismic-hazard analyses to employ empirical ground-motion relations based on a global subduction database. For example, the national seismic-hazard maps developed by both the U.S. Geological Survey (Frankel *et al.,* 1996) and the Geological Survey of Canada (Adams *et al.,* 1999), and incorporated into current building code provisions, model the ground motions from in-slab and interface earthquakes using the global subduction relationships developed by Youngs *et al.* (1997). Preliminary studies of limited empirical databases have suggested that there are no detectable differences between ground motions among different subduction regions, for a given magnitude and distance (Crouse *et al.,* 1988; Youngs *et al.,* 1988; Atkinson, 1997). Thus it would appear that the collation of data from different subduction regimes around the world into a single database for regression analyses is both necessary and reasonable. This assumption is one of many that will be tested in this study.

We have compiled a response spectra database from thousands of ground-motion recordings from events of **M** 5–8.3 occurring in subduction zones around the world, including both interface and in-slab events. The 2001 Nisqually and 1999 Satsop earthquakes are included in the database, as are many records from subduction zones in Japan (Kyoshin-Net [KNET] data), Mexico (Guerrero data), and Central America. The size of the database is much larger than that used in previous regressions for subduction-zone earthquakes. For example, the Youngs *et al.* (1997) database contains 350 horizontal-component response spectra, compiled for earthquakes occurring through 1989. The new database, compiled for earthquakes occurring through 2001, contains 1200 horizontal-component spectra in the magnitude–distance range of interest for regression (e.g., roughly **M** >5 within 100 km, or **M** >7 within 300 km), plus many thousands more records at greater distances that can be used to explore various aspects of the ground-motion scaling with magnitude and distance. All spectra are horizontal component, for 5% of critical damping. The large dataset enables better determination of attenuation parameters and magnitude scaling, for both types of events, than has previously been possible. Soil response parameters are also better determined by the data. The new data suggest that significant revisions to current estimates of seismic hazard may be required.

Database for Regression

The database for regression builds on past work and adds many new ground-motion recordings that have become available in the last decade. The Youngs *et al.* (1997) global subduction database was compiled in about 1989 by adding to an early version of the subduction database compiled by Crouse (1991). We refer the reader to Youngs *et al.* for a description of these data. We noticed that the Crouse (1991) catalog contains events not in the Youngs et al. (1997) catalog, and vice versa; this occured because compilation of the Crouse catalog continued after an early version was provided to Youngs *et al.* We therefore began by merging the global subduction databases compiled by Crouse (1991) and Youngs *et al.* (1997) to create a single historical subduction database containing records from subduction zones around the world through 1989. We then added readily available data from more recent in-slab and interface events. These included events in Cascadia (strong-motion and broadband seismographic records), Japan (KNET strong-motion data), Mexico (Guerrero strong-motion data) and Central America (El Salvador strong-motion data). The compiled database forms an electronic supplement to this article. (Ⓔ Available online at the SSA Web site.)

This is a heterogeneous catalog, containing events from many regions and including many tectonic and soil-type environments. Some of the conditions represented in the database, such as data from soft soil sites in Mexico City, may be quite unique. There is always a danger in an analysis of global databases such as this one that such conditions may bias the final results. Thus the applicability of the global relations to individual areas should be evaluated on a region-by-region basis as more data become available. For any particular region, global ground-motion relations are a good beginning assumption only.

For each record, the database lists the moment magnitude, as obtained from the historical database or from a search of the Harvard Centroid Moment Tensor or Japanese Fundamental Research on Earthquakes and Earth's Interior Anomalies (FREESIA) websites; focal mechanism information and depths were also obtained from these sources. The tabulated distance measure in the database is the closest distance to the earthquake fault plane. For events in the Youngs *et al.* (1997) global database, we adopted these values directly from their database. For the Crouse (1991) database, we estimated the closest fault distance based on the epicentral distance and the event magnitude, assuming that (1) the size of the fault plane is given by the empirical relationships of Wells and Coppersmith (1994) that predict fault length and area as a function of moment magnitude; and (2) the epicenter lies above the geometric center of a dipping fault plane. The empirical relationships of Wells and Coppersmith (1994) were derived for crustal earthquakes and are not nec-

essarily a good description of subduction-zone earthquakes; however they suffice for our purposes, which is to make a modest correction for fault size in deriving the closest fault distances for events of unknown geometry. This correction was only needed for one event of **M** >7 that appeared in the Crouse (1991) dataset but not in the Youngs *et al.* (1997) dataset (a 1978 Japan event of **M** 7.8). In all other cases, these corrections were applied only to events of **M** <7, for which the fault dimensions are small, and so they have little influence on the computed distance. For recent events of moderate magnitude (**M** <7), we also used the epicentral distance and event magnitude to estimate closest distance to fault (in the same way as described earlier). We checked that this approach provides reasonable estimates of the closest fault distance on average, by plotting these estimated closest fault distances against actual values for several of the large events given in Youngs *et al.* (1997) for which the fault geometry is known.

The event classification by type was based on both focal depth and mechanism. Event location and depth were first used to establish that the event was a subduction event (i.e., either interface or in-slab). Among the subduction events, normal faulting mechanisms are always in-slab events. Thrust mechanisms imply interface events for earthquakes that occur at depths of less than 50 km on shallow dipping planes; in this depth range the subducting oceanic plate is in contact with the overriding continental crust. Thrust mechanisms are assumed to represent in-slab events if the events occur at depths greater than 50 km (i.e., below the crustal contact zone) or if they occur on steeply dipping planes. Events of unknown type are not included in the regression, nor are events of focal depth greater than 100 km. Events that occur within the crust, above the subduction zone, are not included in the database.

Each record is assigned a NEHRP site class (see Dobry *et al.,* 2000). For the Japanese KNET data, the classification was based on the shear-wave velocity profiles of the sites, as determined by borehole measurements. The procedure used to assign NEHRP site classes based on the KNET borehole information is described in the Appendix. For the Guerrero data, the sites are all classified as rock, which we have assumed equivalent to NEHRP B (J. Anderson, personal comm., 1999; Chen and Atkinson, 2002). Broadband stations in Washington and British Columbia are also sited on rock. We assume that these sites are best classed as NEHRP B; the assumed average shear-wave velocity is about 1100 m/sec (see Atkinson and Cassidy, 2000). Strong-motion sites in Washington are largely soil and have been classified using a map of site classes prepared by R. Haugerud (personal comm., 2001). The map was based on surficial geology, using correlations between the geology and the average shear-wave velocity to a depth of 30 m from data in Washington. Blind comparisons of the site classes predicted from the Haugerud map with those from eight sites with actual shear-wave velocity measurements (Williams *et al.,* 1999; R. Williams, personal comm., 2001) were in excellent agreement. For the historical data in the Youngs *et al.* (1997) catalog, the assigned Geomatrix classification scheme was converted to the NEHRP equivalent by assuming Geomatrix A = NEHRP B, Geomatrix B = NEHRP C, Geomatrix C/D = NEHRP D, and Geomatrix E = NEHRP E. This equivalence is based on shear-wave velocities and descriptions quoted for the Geomatrix classification scheme (e.g., Abrahamson and Silva, 1997). For the data in the Crouse (1991) catalog, the Youngs *et al.* (1997) station list was used to classify each of the stations to be consistent with the Youngs *et al.* (1997) catalog.

The distribution of the database used in the regressions is shown in Figure 1 for both in-slab and interface events. The database is available in digital form from the electronic supplement to this article (www.seismosoc.org). Observe in Figure 1 that the data are relatively plentiful in the most important magnitude–distance ranges. Specifically, we are most interested in two "design earthquake" scenarios that contribute most to seismic hazard in the Cascadia region, according to deaggregations of typical hazard results (e.g., Frankel *et al.,* 1999; Adams and Halchuk, 2000). These are (1) in-slab earthquakes of **M** 6.5–7.5 at fault distances of 40–100 km (since events are within the slab, at depths that are typically 50 km or more) and (2) interface earthquakes of **M** ≥7.5 at fault distances of 20–200 km (typical depths of about 20 km). It is the in-slab events that have multiplied within the database most markedly in recent years, particularly with the occurrence of the 2001 Nisqually, Washington, and 2001 Geiyo, Japan, events of **M** 6.8. In total, there are 1148 horizontal-component records (where each earthquake–station pair contributes two horizontal components) from 77 earthquakes plotted on Figure 1, which represents a fourfold increase in the number of records as compared to the Youngs *et al.* (1997) database.

Records from moderate events at large distances recorded on the KNET network in Japan are not reliable at high frequencies due to limitations of the instrumentation (Heenan, 2002). To ensure that all analyses are well within the reliable range, the KNET data are excluded from the regressions for events of **M** <6 at distances beyond 100 km; for larger events the KNET data are not used for distances beyond 200 km. KNET data outside of the reliable range are not shown on Figure 1 and have not been used in the regressions. Some of the data that we eliminated may be useful at lower frequencies, but we have excluded them to avoid any potential problems at higher frequencies.

Regression Analysis

Functional Form

Regression of the dataset has been performed using the maximum likelihood method (Joyner and Boore, 1993, 1994). Both horizontal components are included in the dataset for regression. Separate regressions are performed for the interface and in-slab events, since our analyses indicated that

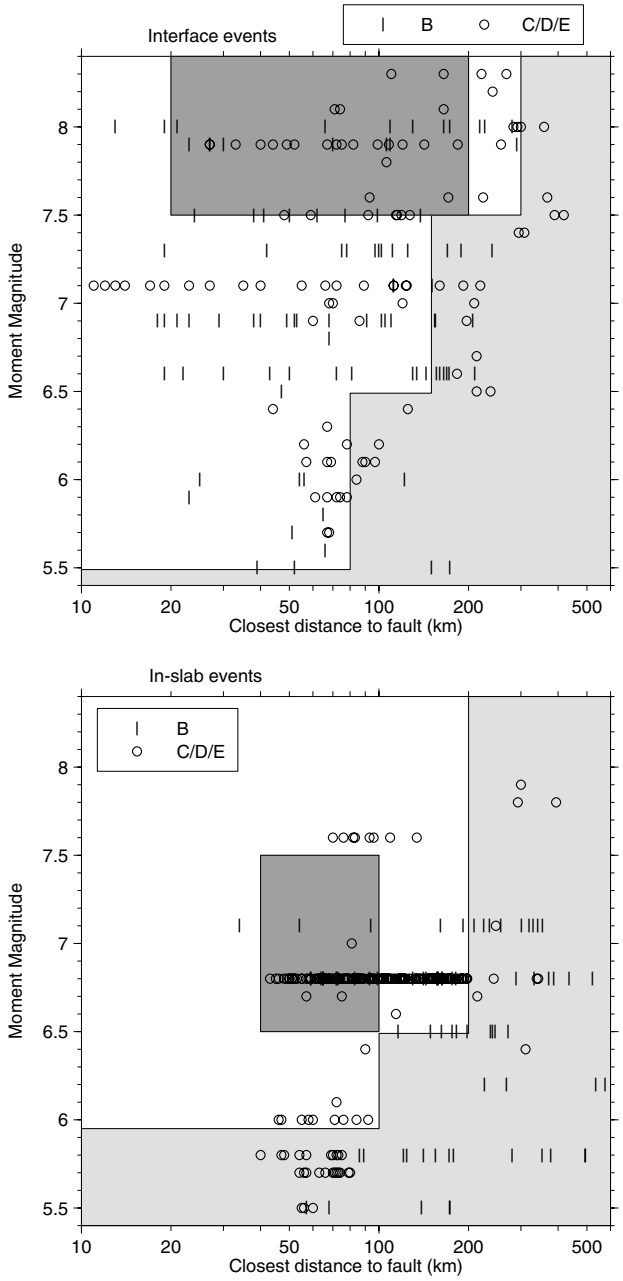

Figure 1. Database for subduction-zone earthquakes. The top frame shows data available for interface events, by NEHRP site class; the lower frame shows data available for in-slab events, by NEHRP site class. Data of **M** <5.4 are not shown. The magnitude–distance range of most engineering interest is shaded dark gray. Magnitude–distance cutoffs imposed on final regressions are shaded light gray. KNET data that are believed to be unreliable at higher frequencies (moderate magnitudes at large distances; see text) are not included.

there are extensive differences in the amplitudes, scaling, and attenuation of these different event types. After detailed experimentation with a variety of function forms, the adopted functional form (all logs base 10) is

$$\log Y = fn(\mathbf{M}) + c_3 h + c_4 R - g\log R + c_5\, sl\, S_C + c_6\, sl\, S_D + c_7\, sl\, S_E, \quad (1)$$

where:

Y = peak ground acceleration or 5% damped pseudoacceleration (PSA) in cm/sec random horizontal component

M = moment magnitude (use **M** 8.5 for interface events of **M** >8.5, **M** 8.0 for in-slab events of **M** ≥8);

$fn(\mathbf{M}) = c_1 + c_2\,\mathbf{M}$ (this is the selected final form; an initial form of $fn(\mathbf{M}) = [c_0' + c_1'\,(\mathbf{M} - 6) + c_2'\,(\mathbf{M} - 6)^2]$ was also explored, as described later)

h = focal depth in kilometers

$R = \sqrt{(D_{\text{fault}}{}^2 + \Delta^2)}$ with D_{fault} being the closest distance to fault surface, in kilometers (use h = 100 km for events with depth >100 km) and Δ, a near-source saturation term, given by

$$\Delta = 0.00724 \times 10^{0.507\mathbf{M}}$$

and

S_C = 1 for NEHRP C soils (360 < β ≤ 760 m/sec), = 0 otherwise

S_D = 1 for NEHRP D soils (180 ≤ β ≤ 360 m/sec), = 0 otherwise

S_E = 1 for NEHRP E soils (β < 180 m/sec), = 0 otherwise

$g = 10^{(1.2-0.18\mathbf{M})}$ for interface events,

$g = 10^{(0.301-0.01\mathbf{M})}$ for in-slab events

sl = 1. for $PGA_{rx} \le 100$ cm/sec^2 or frequencies ≤ 1 Hz

$sl = 1. - (f - 1)\,(PGA_{rx} - 100.)/400.$ for 100 < PGA_{rx} < 500 cm/sec^2 (1 Hz < f < 2 Hz)

sl = 1. − (f − 1) for $PGA_{rx} \ge 500$ cm/sec^2 (1 Hz < f < 2 Hz)

$sl = 1. - (PGA_{rx} - 100.)/400.$ for 100 < PGA_{rx} < 500 cm/sec^2 ($f \ge 2$ Hz and PGA)

sl = 0. for $PGA_{rx} \ge 500$ cm/sec^2 ($f \ge 2$ Hz and PGA);

PGA_{rx} is predicted PGA on rock (NEHRP B), in cm/sec and σ is the standard deviation of residuals, equal to $\sqrt{(\sigma_1^2 + \sigma_2^2)}$ where 1, 2 denote estimated intra- and inter-event variability, respectively. Note that β is the shear-wave velocity averaged over the top 30 m of the soil profile.

The selected functional form incorporates the results of analyses into specific features of the data, such as the magnitude dependence of the geometrical spreading coefficient g, the functional form for the scaling of amplitudes with magnitude, and amplitude-dependent soil nonlinearity, as described later. There are attributes of this functional form

that may be considered unusual. The distance variable for regression, R, is approximately equal to the average distance to the fault surface. The Δ term that combines with D_{fault} to form R is defined from basic fault-to-site geometry: for a fault with length and width given by the empirical relations of Wells and Coppersmith (1994) for all fault types, the average distance to the fault for a specified D_{fault} is calculated (arithmetically averaged from a number of points distributed around the fault), then used to determine Δ, which is a function of fault size. Thus the distance measure depends on the closest distance to the fault and the earthquake magnitude; the magnitude dependence of R arises because large events have a large spatial extent, so that even near-fault observation points are far away from most of the fault. It is important to note that the coefficients that appear in the definition of Δ were defined analytically, so as to represent average fault distance. They were not determined by the regression. However, trial regressions were performed to verify that alternative definitions of the distance measure (using different coefficients to define Δ, over a wide range of possible values) would not result in improved accuracy of model predictions, relative to the average-fault-distance measure that we selected.

The magnitude dependence of the geometric spreading coefficient g was determined by preliminary regressions of the data for both interface and in-slab events. These preliminary regressions looked at slices of data in 1-unit magnitude increments (e.g., $5 \leq \mathbf{M} < 6$, $5.2 \leq \mathbf{M} < 6.2$, etc.) to determine the slope of the attenuation as a function of magnitude. For this exercise, we looked only at the low-frequency (0.5 and 1 Hz) PSA data, within the distance range from 50 to 300 km. (Note: we use data at closer distances for the final regressions.) The selection of data with $D_{\text{fault}} \geq$ 50 km was made to avoid near-source distance saturation effects, which are handled by the Δ term. For the data within each magnitude bin, the regression was made to a simple functional form, given by

$$\log Y' = a_1 + a_2 \mathbf{M} - g \log R + a_3 S, \qquad (2)$$

where Y' is the PSA for 0.5 or 1 Hz, corrected for curvature of the attenuation line due to anelasticity assuming $Y' = Y \exp(0.001R)$, R is the average distance to the fault as defined in equation (1), and $S = 0$ for rock (B) or 1 for soil (C, D, E). The coefficient g is the far-field slope determined for each magnitude bin. As illustrated in Figure 2, the obtained values of the attenuation slope were plotted as a function of magnitude to obtain the form assigned to g in the final regression. Note that there is a marked difference in the slope of the attenuation for interface and in-slab events, with the interface events showing a much more pronounced magnitude dependence to the attenuation. This contrast in attenuation behavior between interface and in-slab events is readily apparent in plots of data amplitudes, even without doing any regression analyses. This is illustrated in Figure 3. The very flat attenuation of large interface events relative to large in-slab events has important implications for the relative impact of these two types of events on seismic hazard. It is evident that large interface events will cause damaging ground motions over a much greater area than will large in-slab events.

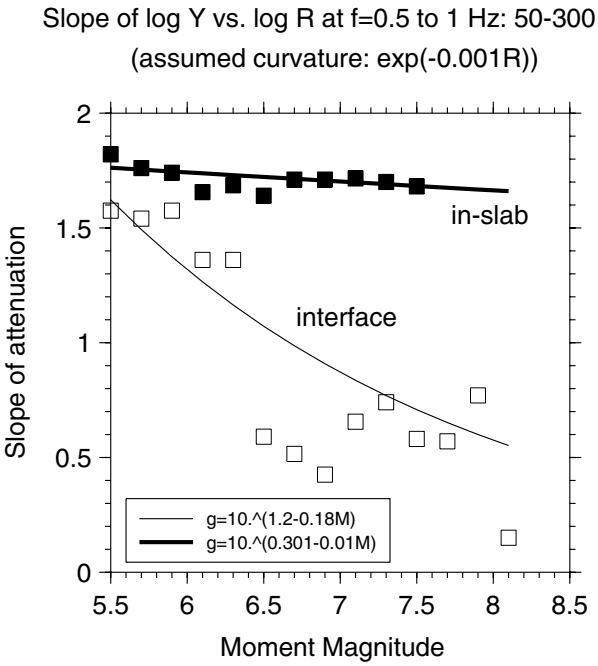

Figure 2. Slope of log Y' versus log R for records within 50–300 km of the fault plane, where $Y' =$ PSA(0.5 or 1 Hz) exp(0.001 R) and R is average distance to the fault as defined by equation (1). Plotted slopes are obtained as the average values for 0.5 and 1 Hz, for data within a 1-unit magnitude bin. For example, the values plotted at **M** 6.5 are the slopes for interface (open squares) and in-slab (filled squares) data of $6.0 \leq \mathbf{M} < 7.0$. The lines show the adopted slopes for the final regressions.

The soil response terms in the final regression include both linear and nonlinear effects. Nonlinear soil effects are not strongly apparent in the database or upon examination of the residuals from preliminary regression results, as most records have PGA < 200 cm/sec^2, but may be important in using the relations for the largest magnitudes and closest distances. To determine the linear soil effects, we performed separate preliminary regressions of each dataset (interface and in-slab) to determine the coefficients c_5, c_6, and c_7 (assuming linear soil response). We used these preliminary results, weighted by the number of observations in each of the two datasets, then lightly smoothed by inspection, to assign fixed values to c_5, c_6, and c_7 (independent of earthquake type) for subsequent regressions. The reason that we fix these terms for the final regressions is to ensure that the soil response coefficients are equal for the in-slab and interface datasets. The soil linearity term, *sl,* was assigned by looking for evidence of nonlinearity in residual plots from these preliminary regressions and from consideration of the NEHRP guidelines for nonlinearity (Dobry *et al.,* 2000). We con-

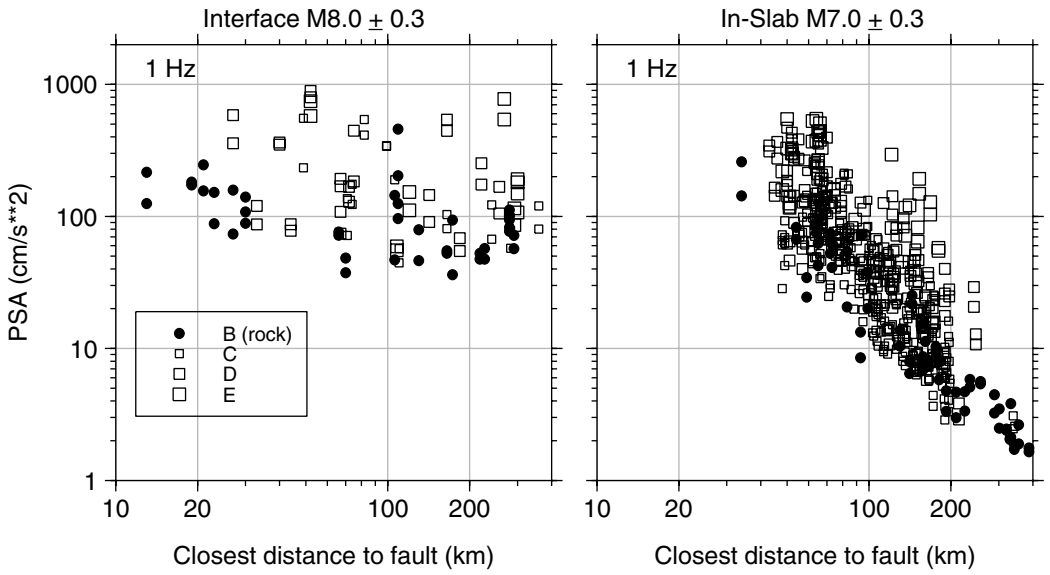

Figure 3. Illustration of the difference in attenuation behavior of interface events versus in-slab events. Plot shows 1-Hz PSA data for interface events of **M** 8 ± 0.3 (left) in comparison to 1-Hz PSA data for in-slab events of **M** 7 ± 0.3 (right).

cluded that there is weak evidence of nonlinearity in the dataset for the limited number of records with $PGA_{rx} > 100$ cm/sec^2 for E soils at frequencies above 1 Hz, and that the data are consistent with the concept that soil and rock amplitudes converge at high amplitudes. Based on these observations and the factors suggested by NEHRP, we assume that the soil coefficients may be multiplied by a soil linearity factor that decreases from a value of 1.0 (fully linear) at $PGA_{rx} = 100$ cm/sec^2 to a value of 0.0 (fully nonlinear) at $PGA_{rx} = 500$ cm/sec^2. This scheme was used to fix the term *sl* for the final regressions, using the form given in equation (1). Figure 4 shows the soil amplification for NEHRP site classes C, D, and E, respectively, and its dependence on PGA_{rx}.

The final regression is performed by taking all terms with fixed coefficients to the left side of the equation; thus we add $g \log R - c_5\, sl\, S_C - c_6\, sl\, S_D - c_7\, sl\, S_E$ to $\log Y$ before the regression to determine the remaining coefficients. This process needs to be iterated a few times because the resulting PGA_{rx} is used in the definition of the dependent variable for the regression. After a few iterations, the coefficients for PGA_{rx} remain unchanged by further iteration and the regression is complete.

In order to optimize the fit for the magnitude–distance range of engineering interest, we limited the final regression to data that fell within the following criteria (where D_{fault} is the closest distance to the fault):

- Interface events:

$$5.5 \le \mathbf{M} < 6.5 \quad D_{fault} \le 80 \text{ km}$$
$$6.5 \le \mathbf{M} < 7.5 \quad D_{fault} \le 150 \text{ km}$$
$$\mathbf{M} \ge 7.5 \quad D_{fault} \le 300 \text{ km}$$

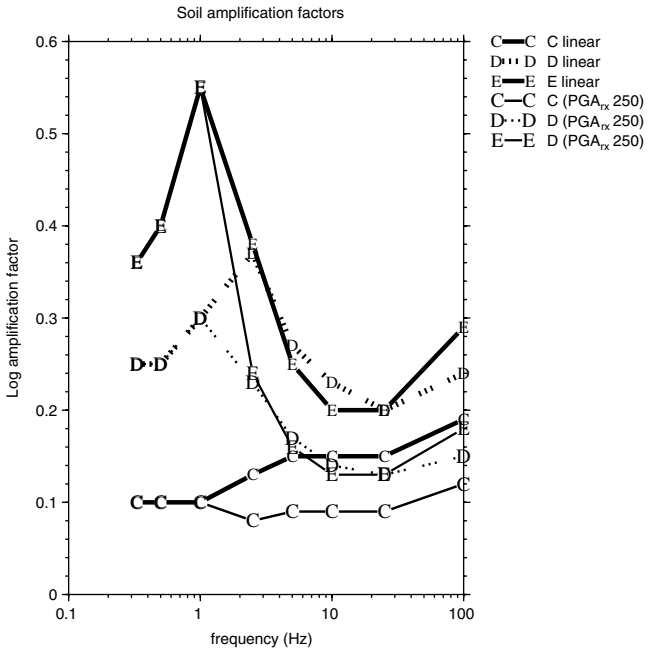

Figure 4. Soil amplification factors for NEHRP site classes C, D, and E relative to rock (NEHRP B), as given by regression coefficients c_5, c_6, and c_7 (linear), respectively. The linear factors apply for $PGA_{rx} \le 100$ cm/sec^2. Amplifications for stronger ground motions (250 cm/sec^2) are also shown.

- In-slab events:

$$6.0 \le \mathbf{M} < 6.5 \quad D_{fault} \le 100 \text{ km}$$
$$\mathbf{M} \ge 6.5 \quad D_{fault} \le 200 \text{ km}$$

These criteria were refined by experimenting with the regression results until an optimal fit of the regression equation to the data was achieved for the events that are important to seismic-hazard analysis: namely interface events of **M** $\geq$7.5 and in-slab events of **M** $\geq$6.5, at distances of up to 200 km, as well as smaller earthquakes at close distances. The need to restrict the magnitude–distance range for regression arises from the fact that there are many more data recorded for moderate events and intermediate distances than for large events and close distances. Thus the database is dominated by the former, while the hazard is dominated by the latter. This needs to be considered in selection of the data for regression, in order to avoid a result that is dominated by data availability. Specifically, if all of the data are included, out to the largest available distances, then the regression will be biased toward fitting the plentiful data from moderate events at regional distances. The method that we have chosen for counteracting this biasing effect is to choose magnitude–distance windows for regression that are relevant to the data we wish to fit well, then checking the fit and adjusting the windows as needed to ensure we have achieved this goal. An alternative strategy would be to use a Monte Carlo technique to draw a limited number of records from the dataset, in such a way that the data were more evenly distributed over magnitude and distance; even in this case, however, one might want to draw more records from the most important magnitude–distance ranges. We note that although the final regression uses a subset of the data, the larger database is still useful in exploring several aspects of the functional form, such as those discussed earlier, and in quantifying the fit of the relations in the magnitude–distance ranges not specifically included in the final regressions.

Finally, after performing the regressions to the selected functional form, the resulting coefficients were lightly smoothed (using a weighted three-point smoothing) over frequency. This ensures a smooth spectral shape against frequency and allows for reliable interpolation of coefficients for frequency values not explicitly used in the regression.

In our initial regressions, a quadratic term in magnitude was used to model the scaling of amplitudes with magnitude: our initial form had a term in $\mathbf{M}^2$ as well as in **M.** The quadratic term led to a better fit than the linear magnitude scaling overall, but the sign of the quadratic term is positive, rather than negative as would be expected. To ensure the best fit in the magnitude range that is both important to hazard and constrained by data, while providing realistic scaling outside of this range, the quadratic source terms in the equation were refit to a linear form. The linear model was constrained to provide the same results as given by the quadratic model, in the range from **M** 7.0 to 8.0 for interface events and from **M** 6.5 to 7.5 for in-slab events. Thus the initial source description of $[c_0' + c_1'(\mathbf{M} - 6) + c_2'(\mathbf{M} - 6)^2]$ was refit to $[c_1 + c_2\mathbf{M}]$. The scaling of amplitudes with magnitude for near and far distances, for both the initial quadratic and the adopted linear form, is plotted for rock sites in Figures 5–8, for interface and in-slab events. The curvature displayed for the linear magnitude scaling results from the interplay of the magnitude-dependent slope of the attenuation with the magnitude scaling. From these plots, it is apparent that the ground-motion relations require the imposition of a maximum magnitude of **M** 8.5 for prediction of interface amplitudes or **M** 8.0 for prediction of in-slab amplitudes. At greater magnitudes (beyond the range of the data), the predicted amplitudes will grow smaller rather than larger. Thus estimates of ground motion for interface events of **M** $>$ 8.5 should be made using **M** 8.5, while estimates of ground motion for in-slab events of **M** $>$ 8.0 should be made using **M** 8.0. This saturation effect is mainly an issue for interface events, since in-slab events of **M** $>$8 are not generally expected. The use of a maximum magnitude of **M** = 8.5 for prediction of the ground motions from interface events is not meant to imply that a maximum magnitude of 8.5 should be assigned in the hazard analysis. Rather, the ground motions for larger events should be calculated using the value of **M** 8.5 in the ground-motion equations. This saturation of the predicted amplitudes at very large magnitudes is not unreasonable in view of the data plotted in Figure 5, especially when the effect of nonlinear soil response is considered.

Restrictions should also be placed on the depth term, to prevent the prediction of unrealistically large amplitudes for earthquakes deeper than the 100-km-depth cutoff applied in the regressions. If the equations are extrapolated for depths greater than 100 km, a value of h = 100 km should be assumed. The validity of the equations for deeper events has not been evaluated. There may be some underestimation if the equations are applied to deeper events, as there is evidence that ground motions (for a fixed fault distance) continue to increase with increasing earthquake depth (Crouse, 1991; Youngs *et al.,* 1997; Molas and Yamazaki, 1995). Similarly, the validity of the equations for earthquakes greater than **M** 8.5 cannot be evaluated due to the lack of data. In general we would expect events of **M** $>$8.5 to produce larger motions at very long periods, but similar motions to those of a **M** 8.5 event at high frequencies.

The regression strategy outlined above is potentially confusing, due to the use of several steps in arriving at the final form and the use of different data to determine different aspects of the regression. A summary of the steps described earlier and data used in each step is as follows:

1. Regression of far-field data to determine the geometric spreading term, g, that describes the decay of log amplitudes with distance at low frequencies (equation 2) (at distances beyond those for which saturation effects occur). This regression was done separately for the interface and in-slab events and was used to determine the dependence of the attenuation slope g on magnitude. Data used were the PSA data for frequencies of 0.5 and 1 Hz, within the fault distance range from 50 to 300 km (this is the far-field subset of the data plotted in Fig. 1).

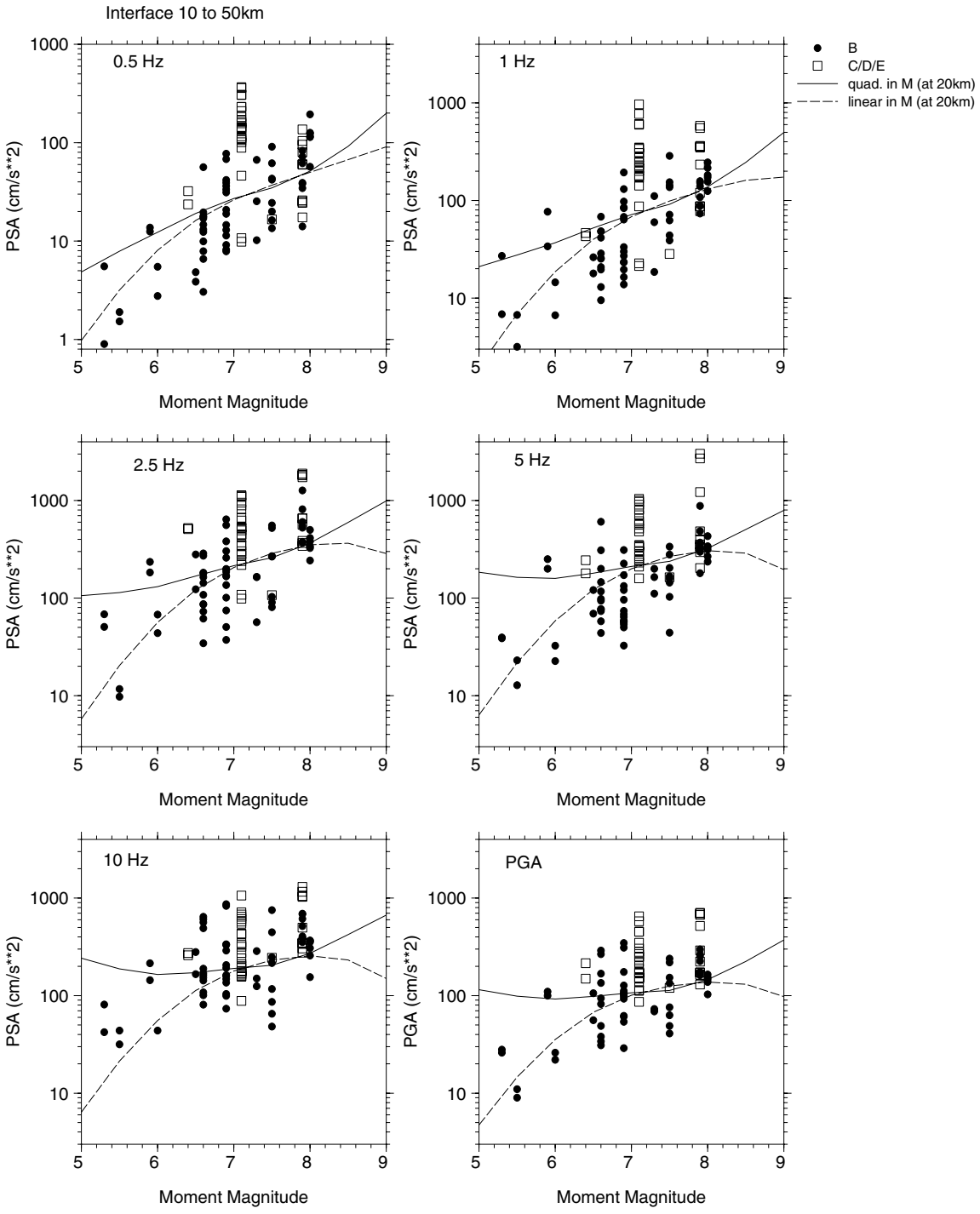

Figure 5. Scaling of ground-motion amplitudes with moment magnitude for interface events in the distance range (closest distance to fault) from 10 to 50 km. Assumed event depth is 20 km. Filled symbols show data for rock (NEHRP B), while open symbols show data for soil (NEHRP C, D, E). Solid lines show scaling behavior using the quadratic term in magnitude, while dashed lines show behavior using the linear term in magnitude, both for rock sites at a distance of 20 km. Note that prediction lines for soil sites (not shown) would be higher than those for rock (by about a factor of 1.5 at high frequencies to 2.5 at low frequencies, depending on soil type).

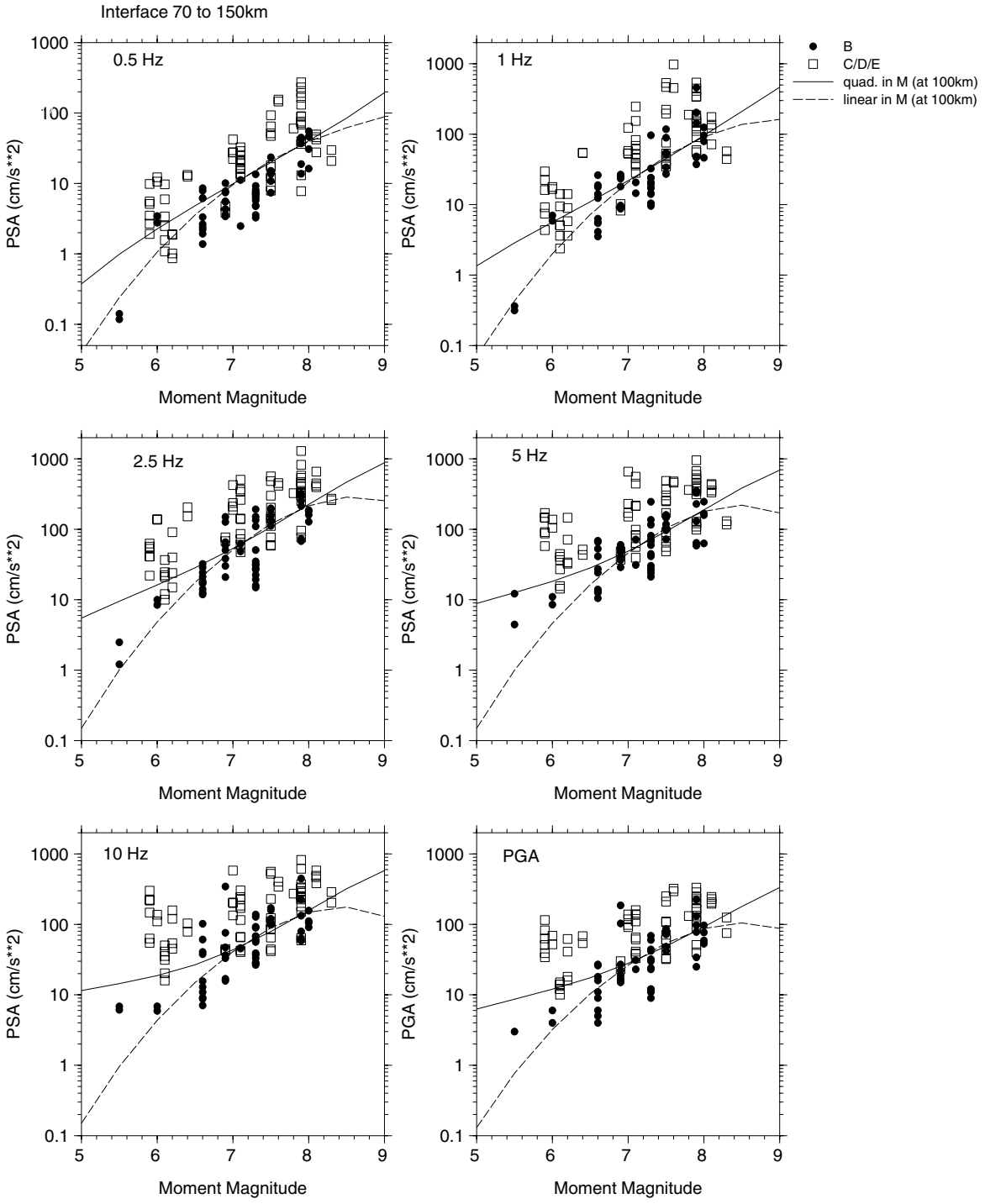

Figure 6. Scaling of ground-motion amplitudes with moment magnitude for interface events in the distance range (closest distance to fault) from 70 to 150 km. Assumed event depth is 20 km. Filled symbols show data for rock (NEHRP B), while open symbols show data for soil (NEHRP C, D, E). Solid lines show scaling behavior using the quadratic term in magnitude, while dashed lines show behavior using the linear term in magnitude, both for rock sites at a distance of 100 km. Note that prediction lines for soil sites (not shown) would be higher than those for rock (by about a factor of 1.5 at high frequencies to 2.5 at low frequencies, depending on soil type).

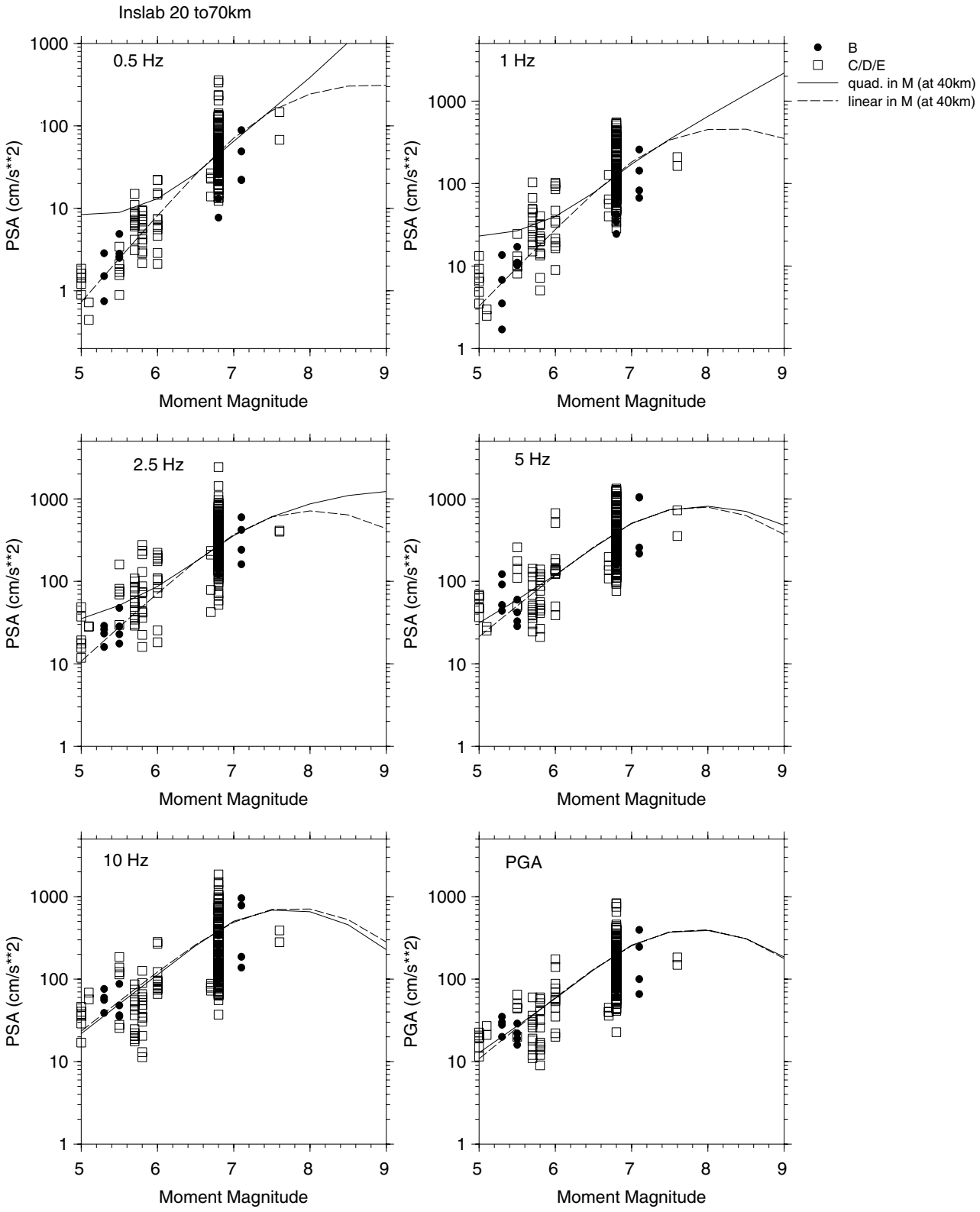

Figure 7. Scaling of ground-motion amplitudes with moment magnitude for in-slab events in the distance range (closest distance to fault) from 20 to 70 km. Assumed event depth is 50 km. Filled symbols show data for rock (NEHRP B), while open symbols show data for soil (NEHRP C, D, E). Solid lines show scaling behavior using the quadratic term in magnitude, while dashed lines show behavior using the linear term in magnitude, both for rock sites at a distance of 40 km.

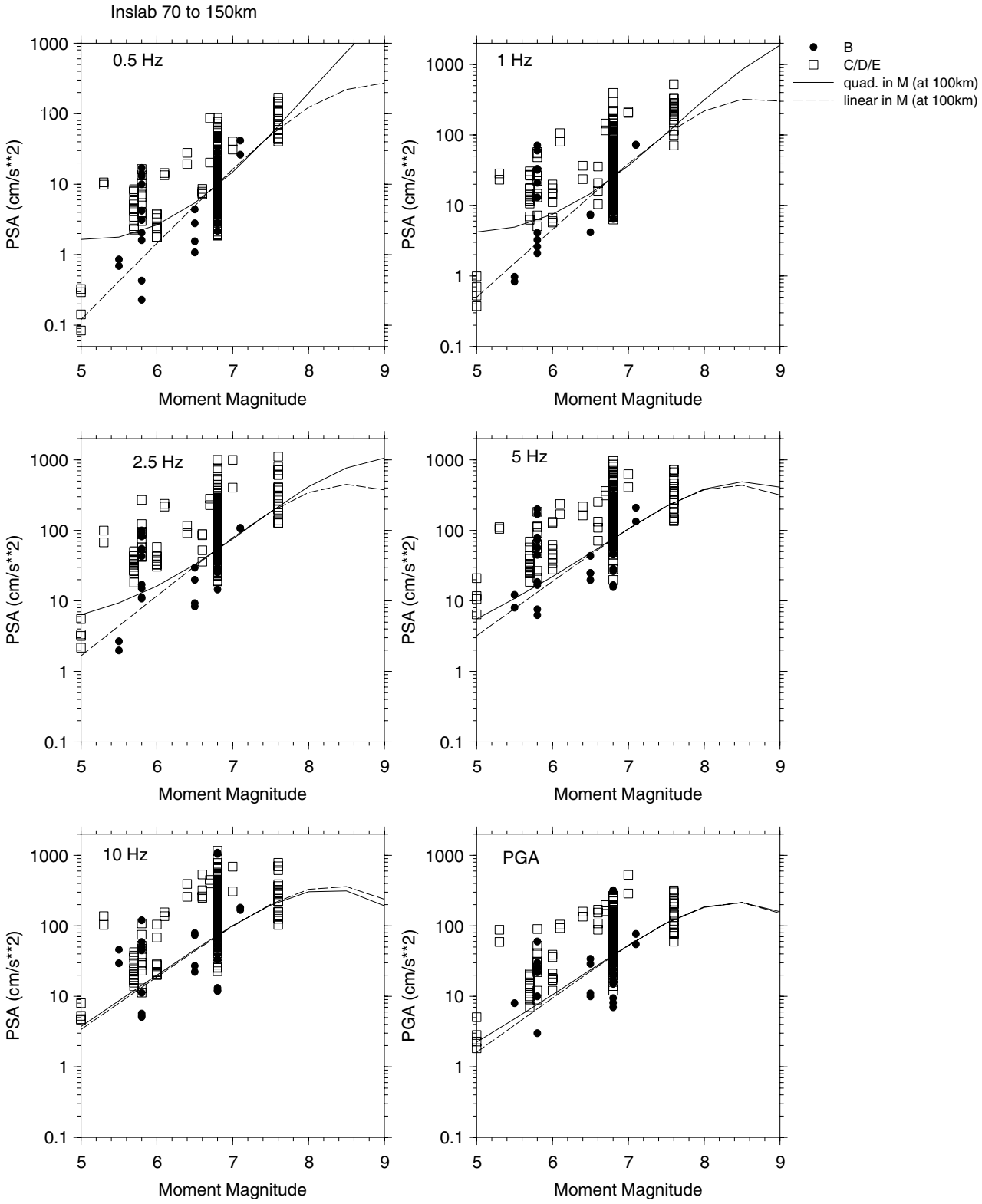

Figure 8. Scaling of ground-motion amplitudes with moment magnitude for in-slab events in the distance range (closest distance to fault) from 70 to 150 km. Assumed event depth is 50 km. Filled symbols show data for rock (NEHRP B), while open symbols show data for soil (NEHRP C, D, E). Solid lines show scaling behavior using the quadratic term in magnitude, while dashed lines show behavior using the linear term in magnitude, both for rock sites at a distance of 100 km.

2. Preliminary regressions to the selected functional form (equation 1, with the set values of g as determined in step 1 and the initial quadratic form for the source description), but setting $sl = 1$ (i.e., assuming soil linearity) in order to determine the linear soil response coefficients (c_5, c_6, and c_7). These preliminary regressions were also used to explore different possible definitions for the near-source saturation term, Δ. The data used for these preliminary regressions were the same as those in the final regressions (same data as given for step 3). The separate results from the interface and in-slab datasets were weighted as described earlier and used to fix the values of the linear soil coefficients (c_5, c_6, and c_7) for the final regressions, to ensure that soil amplification would be the same for interface and in-slab events.
3. Final regressions to selected functional form (equation 1, using the quadratic form for the source description), with the set values of g as determined in step 1, the set value of Δ as tested in step 2, the set values of c_5, c_6, and c_7 as determined in step 2, and the assigned values of sl. The data used in the final regression are the following subset of the data plotted on Figure 1.

   Interface events:

$$5.5 \le \mathbf{M} < 6.5 \; D_{\text{fault}} \le 80 \text{ km}$$
$$6.5 \le \mathbf{M} < 7.5 \; D_{\text{fault}} \le 150 \text{ km}$$
$$\mathbf{M} \ge 7.5 \; D_{\text{fault}} \le 300 \text{ km}$$

   In-slab events:

$$6.0 \le \mathbf{M} < 6.5 \; D_{\text{fault}} \le 100 \text{ km}$$
$$\mathbf{M} \ge 6.5 \; D_{\text{fault}} \le 200 \text{ km}$$

4. Refitting of the quadratic source description, $c_0' + c_1'(\mathbf{M} - 6) + c_2'(\mathbf{M} - 6)^2$ to the final form of the source term, $c_1 + c_2\mathbf{M}$.

The coefficients of the regression are given in Table 1. Values for other frequencies can be obtained by linear interpolation of each coefficient value in log frequency space. Inspection of the coefficients c_5–c_7 reveals that soil motions exceed rock motions by as much as a factor of 4 for soft soils at low frequencies. There is a weak depth effect given by the coefficient c_3, with deeper events causing larger motions.

The recommended standard deviation of the residuals is also listed in Table 1. The values of σ listed in Table 1 are calculated based on records within 100 km of the fault, considering all soil types. For interface events, the standard deviation is calculated considering all events of $\mathbf{M} \ge 7.2$, while for in-slab events it is calculated considering all events of $\mathbf{M} \ge 6.5$. These magnitude ranges were selected in order to obtain the variability that is applicable to the magnitude ranges most relevant to hazard calculations. In a separate study, Atkinson and Casey (2003) have found that the KNET data from Japan appear to have a greater high-frequency site response than data of the same soil class from other regions, due to the prevalence of shallow soil sites in Japan. Thus their inclusion could artificially inflate the expected scatter of high-frequency amplitudes in any one region. For this reason the KNET data are not used in computing the standard deviation of residuals listed in Table 1. The regional variability of ground-motion residuals due to gross regional differences in site characteristics is evaluated in more detail later.

A rough estimate of how much of the total variability (σ) is attributable to intraevent variability (σ_1) is made by calculating this value for several of the larger-magnitude events for which the data are most plentiful and determining an average value. The interevent variability (σ_2) is then calculated assuming that $\sigma_2 = \sqrt{(\sigma^2 - \sigma_1^2)}$.

Figure 9 compares PGA on rock and soil sites for both interface and in-slab events. Nonlinearity of soil response is only important for records with $\text{PGA}_{\text{rx}} > 100$ cm/sec^2 and so exerts only a minor influence on the curves (for interface events of $\mathbf{M} \ge 7.5$ or in-slab events of $\mathbf{M} > 6.5$).

In Figure 10 we plot in-slab versus interface ground motions on NEHRP D sites for frequencies of 0.5 and 5 Hz for **M** 6.5. We have also shown the previous regression results of Youngs *et al.* (1997) for comparison, as these are the most widely used relations at present. Interface ground motions for **M** 8.5 are also shown. Our results are similar to those of Youngs *et al.* (1997) for the largest interface events, for which our databases are similar, but differ greatly for events of **M** 6.5, for which the size of our database exceeds theirs by more than a factor of 10. Note the very slow attenuation for great interface events in comparison to moderate in-slab events.

Great interface events pose a hazard to large areas because of the lack of significant attenuation within 300 km of the fault plane. In-slab events, by contrast, produce large near-source amplitudes; near-source amplitudes of in-slab events of **M** 7.5 are greater than those from interface amplitudes for events of **M** 8.5, by about a factor of 2. However, the in-slab amplitudes attenuate rapidly with distance from the fault, resulting in hazard to a much smaller area. The finding of significantly different attenuation rates for in-slab versus interface events is a novel result of this analysis, with significant implications for seismic-hazard analysis. Previous studies (Youngs *et al.,* 1988, 1997) recognized differences in amplitudes between event types but did not describe these differences in terms of their dependence on distance.

Evaluation of Results

The most relevant measure of any empirical regression result is how accurately it models the database that it purports to represent. With the large database available, we are able to present a more comprehensive analysis of the residuals than has been given in any previous regression study

Table 1
Regression Coefficients

Freq	c_1	c_2	c_3	c_4	c_5	c_6	c_7	σ	σ_1	σ_2
Coefficients for Interface Events										
0.33	2.301	0.02237	0.00012	0.000	0.10	0.25	0.36	0.36	0.31	0.18
0.5	2.1907	0.07148	0.00224	0.000	0.10	0.25	0.40	0.34	0.29	0.18
1	2.1442	0.1345	0.00521	−0.00110	0.10	0.30	0.55	0.34	0.28	0.19
2.5	2.5249	0.1477	0.00728	−0.00235	0.13	0.37	0.38	0.29	0.25	0.15
5	2.6638	0.12386	0.00884	−0.00280	0.15	0.27	0.25	0.28	0.25	0.13
10	2.7789	0.09841	0.00974	−0.00287	0.15	0.23	0.20	0.27	0.25	0.10
25	2.8753	0.07052	0.01004	−0.00278	0.15	0.20	0.20	0.26	0.22	0.14
PGA	2.991	0.03525	0.00759	−0.00206	0.19	0.24	0.29	0.23	0.20	0.11
Coefficients for In-Slab Events										
0.33	−3.70012	1.1169	0.00615	−0.00045	0.10	0.25	0.36	0.30	0.29	0.08
0.5	−2.39234	0.9964	0.00364	−0.00118	0.10	0.25	0.40	0.30	0.28	0.11
1	−1.02133	0.8789	0.00130	−0.00173	0.10	0.30	0.55	0.29	0.27	0.11
2.5	0.005445	0.7727	0.00173	−0.00178	0.13	0.37	0.38	0.28	0.26	0.10
5	0.51589	0.69186	0.00572	−0.00192	0.15	0.27	0.25	0.28	0.26	0.10
10	0.43928	0.66675	0.01080	−0.00219	0.15	0.23	0.20	0.28	0.27	0.07
25	0.50697	0.63273	0.01275	−0.00234	0.15	0.20	0.20	0.25	0.24	0.07
PGA	−0.04713	0.6909	0.01130	−0.00202	0.19	0.24	0.29	0.27	0.23	0.14

The regression equation is $\log Y = c_1 + c_2\,\mathbf{M} + c_3 h + c_4 R - g \log R + c_5\, sl\, S_C + c_6\, sl\, S_D + c_7\, sl\, S_E$. See text near equation (1) for definitions of variables. The c_1 coefficient may be refined to better model the Japan or Cascadia region, as described in the text. The revised c_1 values specific to Japan and Cascadia are listed in Table 3.

for subduction-zone earthquakes. We have analyzed the residuals for both interface and in-slab events in a number of ways. The residual is measured in log (base 10) units and is defined as the difference between the log of the observed value and the log of the predicted value; thus a residual of +0.1 represents an underprediction by a factor of 1.26, for example. Figure 11 shows the interface residuals as a function of distance, in three magnitude ranges, while Figure 12 shows the corresponding plot for in-slab events. (The reason for distinguishing the Cape Mendocino residuals in Fig. 11 is discussed in the next section.) In making these plots, all records of $\mathbf{M} \geq 5.5$ in the database within 200 km of the fault were used, while for the large events ($\mathbf{M} \geq 6.5$) all records within 300 km were used (including data from all regions). Note that this is a larger distance range than that used in the regressions, to enable a broader sense of how well the equations represent the larger database. Thus we do not necessarily expect zero average residuals over all magnitude and distance ranges. The residual plots indicate a large amount of random variability in the data. Overall, the average residuals are near zero for both types of events in the important distance range within 100 km of the fault (with some exceptions).

In Figures 13 and 14, we plot the same set of residuals but distinguish between them based on soil condition instead of magnitude. These figures are used to provide an indication of how effective the assigned soil coefficients are in providing an unbiased fit for all NEHRP site classes. We note in Figure 14 that there is an apparent bias toward positive residuals for rock sites (NEHRP B) in the distance range from 60 to 70 km for in-slab events. Closer examination reveals that these data are records of the Nisqually earthquake made in the Seattle area. An evaluation of site effects present in the Nisqually data has shown that motions in Seattle from this event were particularly high for all site classes (Atkinson and Casey, 2003) and may reflect unmodeled basin effects in this region (Frankel *et al.,* 1999). We do not have detailed information on the sites in the Seattle area that have been classified as rock (NEHRP B), but in view of the regional geology (see Frankel *et al.,* 1999) many of these sites may actually be hard till deposits within the basin, while others would be on Tertiary rock outcrops. Thus we should not expect the regression results to match this particular subset of the data. It should be noted, however, that motions in Seattle and other basin regions may be significantly higher than predicted by the regression equations, because of these unusual site effects (i.e., the regression equations predict motions for average site conditions only).

The standard deviation of the residuals (σ) for records of $\mathbf{M} \geq 6$ within 100 km of the fault is shown on Figure 15 as a function of frequency (using data from all regions). The random variability is an important parameter for seismic-hazard analysis. Overall, the variability is similar for interface and in-slab events. There is a significant tendency to lower variability for higher magnitudes ($\mathbf{M} \geq 7.2$) and for higher frequencies, in agreement with the findings of Youngs *et al.* (1997). The standard deviation of the residuals is typically in the range 0.20–0.32 log units for the larger events and 0.25–0.40 for the smaller events. Youngs *et al.* (1997), by comparison, reported slightly larger standard deviations, in the range of 0.28–0.33 for the largest magnitudes or 0.37–0.41 for small magnitudes.

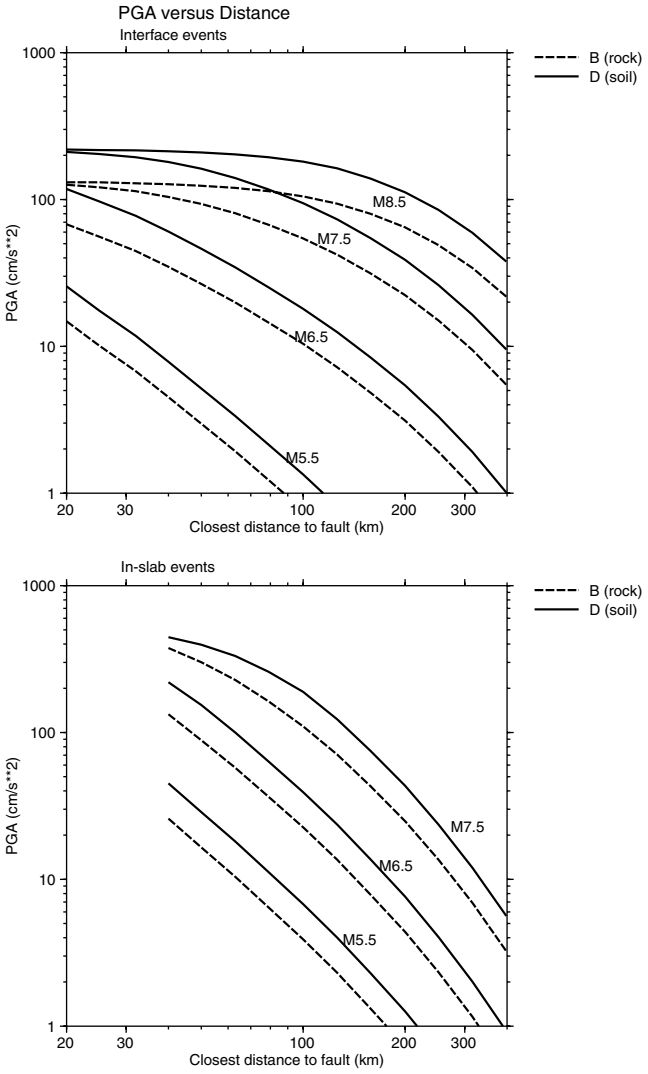

Figure 9. Peak ground acceleration for rock (NEHRP B) and soil (NEHRP D) for interface events (depth = 20 km) of **M** 5.5, 6.5, 7.5, and 8.5 (top frame) and in-slab events (depth = 50 km) of **M** 5.5, 6.5, and 7.5. Nonlinear soil response is assumed for records with $PGA_{rx} > 100$ cm/sec^2.

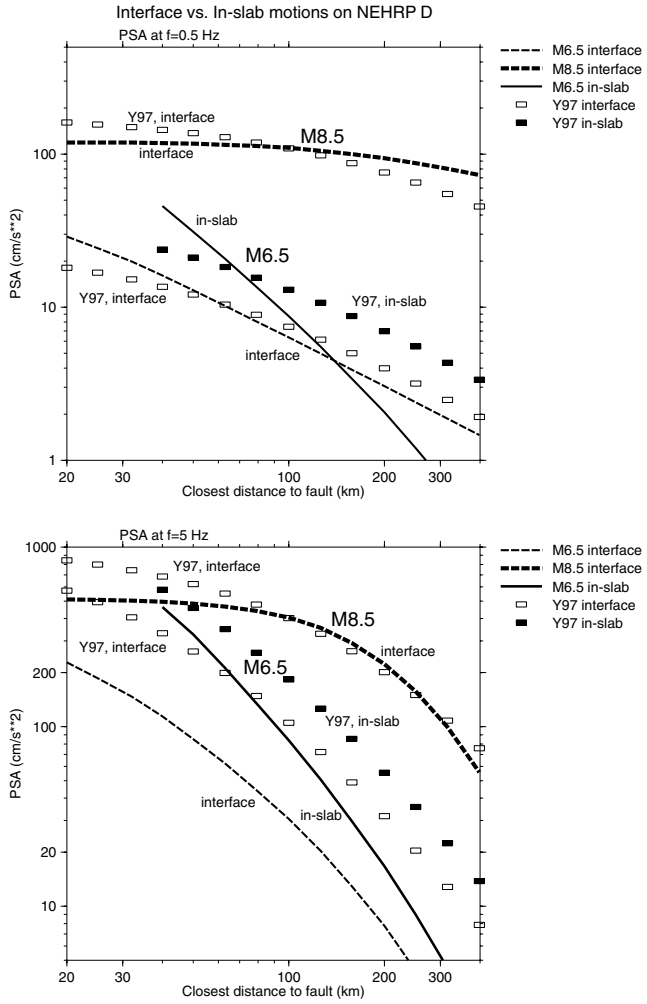

Figure 10. Comparison of response spectra amplitudes at frequencies 0.5 Hz (top frame) and 5 Hz (lower frame), for interface (depth = 20 km) and in-slab (depth = 50 km) events of **M** 6.5 on NEHRP D soil sites. The corresponding predictions of Youngs *et al.* (1997) for interface (open symbols) and in-slab (filled symbols) events are also shown. Interface ground motions for **M** 8.5 (NEHRP D) are also shown in comparison to the Youngs *et al.* (1997) predictions.

Figures 16 and 17 compare the predictions of the regression to the observed ground-motion amplitudes, at a range of frequencies, for the largest events that are well represented in the database. These are interface earthquakes of **M** 8 ± 0.3 (there are no larger events in the database) and in-slab earthquakes of **M** 6.8 ± 0.3. In these figures, the predictions of our regression equations are plotted for the central magnitude value, for NEHRP B (rock) and D (soil). The corresponding predictions of the Youngs *et al.* (1997) relations are also shown. For large interface events (Fig. 16), our predictions are generally similar to those of Youngs *et al.* (1997), with differences between the predictions being within a factor of 2 over the distance range from 20 to 200 km. Both sets of relations fit the data to about the same degree. We also show in Figure 16 the predictions of Gregor *et al.* (2002) for interface subduction earthquakes of **M** 8. Gregor *et al.* (2002) developed these relations, for rock and soil sites, using a stochastic finite-fault model. Thus the Gregor *et al.* relations are not empirical fits to the data, but rather expectations based on seismological modeling. The Gregor *et al.* (2002) predictions generally exceed our empirical predictions (or those of Youngs *et al.* [1997]) for rock sites, especially at close distances, but match our empirical relations well for soil sites at frequencies of 2.5 Hz and greater. At lower frequencies (0.5 and 1 Hz), the Gregor *et al.* stochastic-model predictions are significantly higher than the empirical relations for both rock and soil sites (both ours and those of Youngs *et al.* [1997] and appear to significantly overestimate the data at distances less than 100 km from the

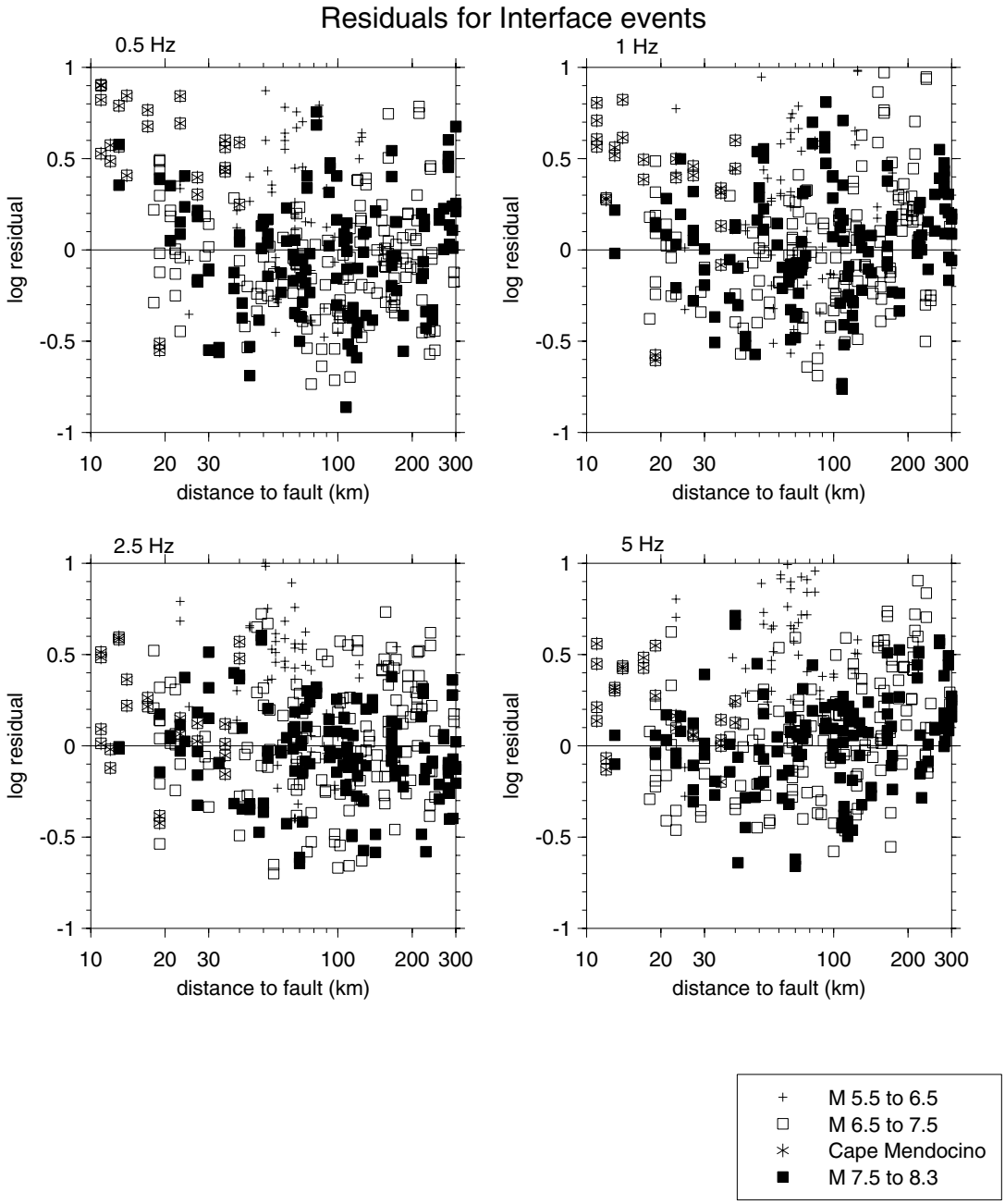

Figure 11. Log residuals (= log observed value − log predicted value) based on regression equation, for interface events, distinguished by magnitude. The residuals from the Cape Mendocino event (discussed in text) are denoted with an asterisk.

fault. This reflects the very low attenuation rates that are observed in the data for large subduction events at low frequencies. In our regression form, the flat shape of the near-source attenuation for large events is controlled by the Δ term. In our preliminary regressions we explored a wide range of potential definitions of this term and found that the data are best fitted by using large Δ values for large magnitudes, such as we have adopted (e.g., our form gives $\Delta =$ 82 km for **M** 8). It is interesting that the data plotted in Figure 16 appear to support the flat near-source attenuation at low frequencies that is given in both our model and that of Youngs *et al.* (1997), but it is difficult to reconcile these observations with the modeled attenuation predictions of Gregor *et al.* (2002). These aspects of the attenuation are intriguing and warrant further investigation. The comparisons of Figure 16 also suggest that the low-frequency source amplitudes for large subduction events are smaller than predicted by the Gregor *et al.* (2002) model. The source and attenuation features of these large subduction events should be investigated through further modeling, in an attempt to

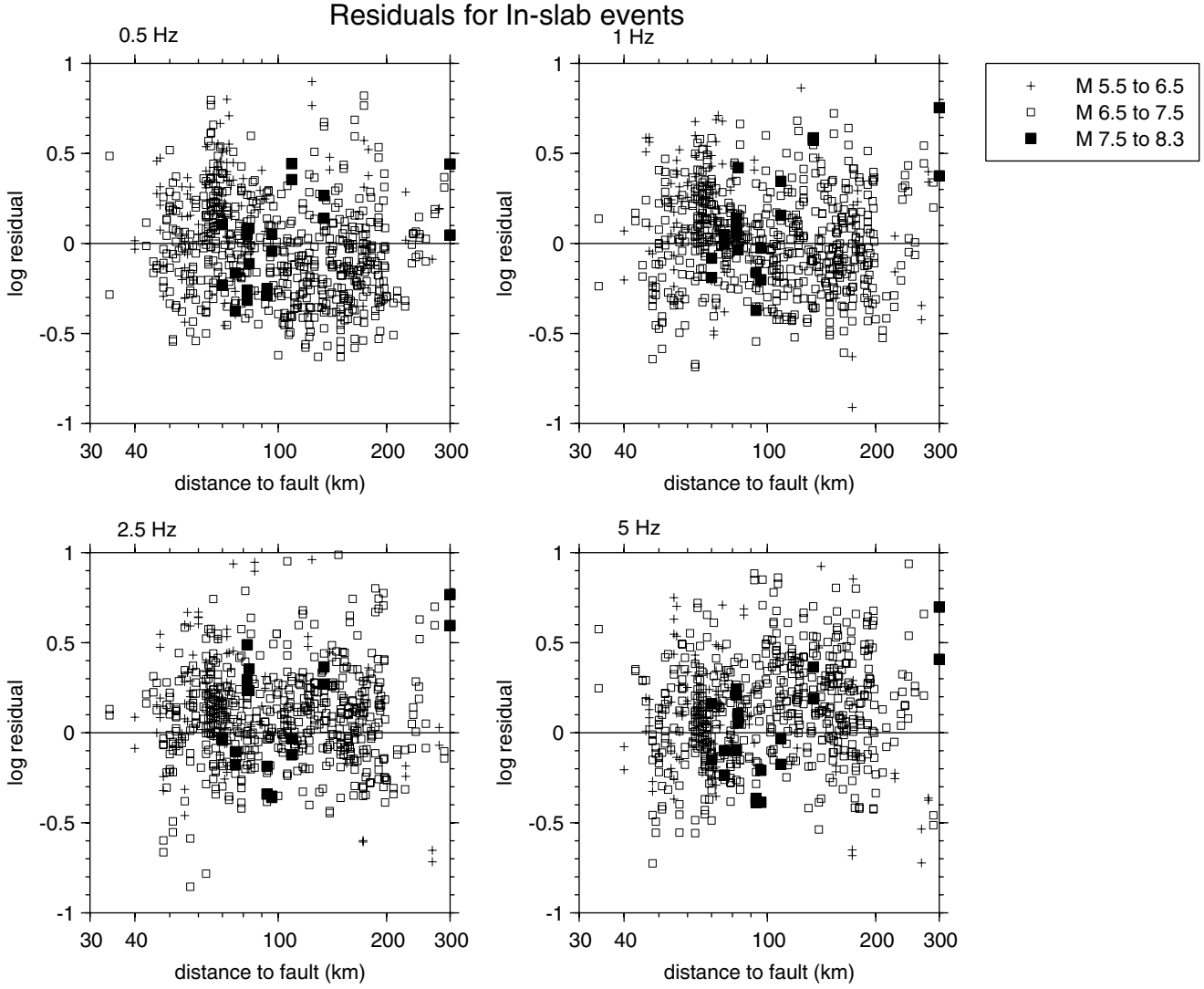

Figure 12. Log residuals (= log observed value − log predicted value) based on regression equation, for in-slab events, distinguished by magnitude.

explain these apparent features of the data.

For in-slab events (Fig. 17) we predict a significantly faster decay of amplitudes with distance than do the Youngs *et al.* (1997) relations, and this is clearly supported by the data. The much slower attenuation of ground-motion amplitudes for megathrust earthquakes as compared to moderate to large in-slab earthquakes is apparent in comparing Figure 16 to Figure 17.

The relations developed by Youngs *et al.* (1988, 1997) were the first ground-motion relations to recognize the difference between in-slab and interface amplitudes. The Youngs *et al.* relations have been widely used in engineering practice, including in the development of current seismic-hazard maps for building code applications. The Youngs *et al.* (1997) study followed a similar regression methodology to that used in this article, based on maximum likelihood. There are two reasons for the differences between our results and theirs. First, they used a different functional form, and thus the relations may behave quite differently in the magnitude–distance ranges that are not well constrained by data. The second and more important reason for the differences lies in the increased database that we were able to employ: we had 12 more years of records, representing an order-of-magnitude increase in database size, particularly for the in-slab events. This enables us to better distinguish differences in the amplitudes and distance dependence of in-slab and interface events, as well as to improve the modeling of other effects such as the magnitude dependence of attenuation and soil response.

In Figures 16 and 17, it appears that the rock observations are underpredicted at close distances for some frequencies (2.5 and 10 Hz). This might suggest that nonlinearity is stronger than that assumed in the regressions. Yet when the regression residuals are plotted against the predicted rock amplitude, there are no apparent trends (analysis similar to that presented in Figs. 11–14; plots not shown).

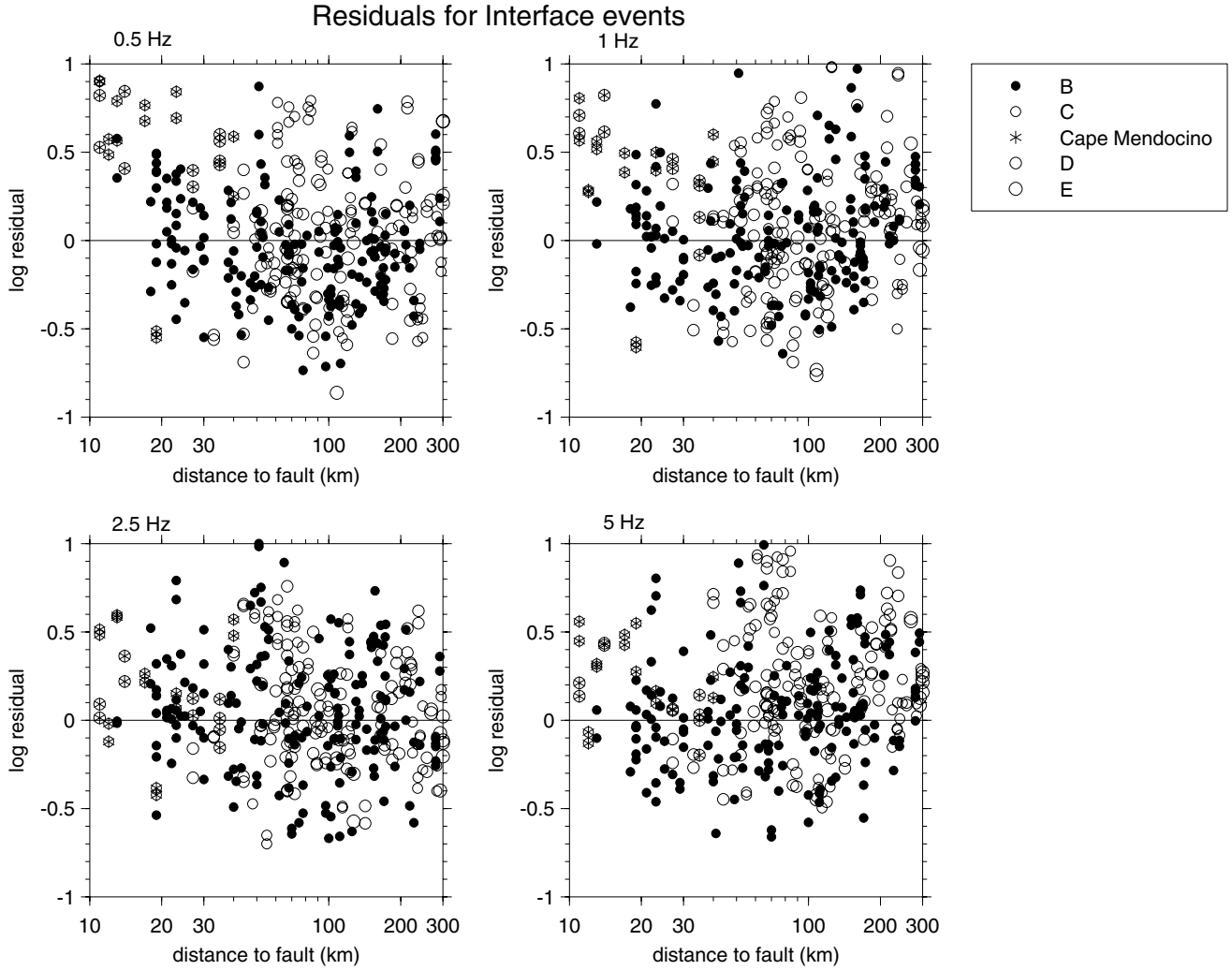

Figure 13. Log residuals (= log observed value − log predicted value) based on regression equation, for interface events, distinguished by site class. The residuals from the Cape Mendocino event (discussed in text) are denoted with an asterisk.

Furthermore, examination of Figures 13 and 14 suggests that underprediction of amplitudes for B sites in some selected magnitude ranges is balanced by overprediction in others, to produce no net bias for rock sites. Finally, as discussed earlier, we believe there is a bias in the Nisqually B data at close distances due to unusual site effects in the Seattle area.

Figure 18 plots the predicted spectral amplitudes for moderate to large interface and in-slab events, at distances of 50 and 100 km. For reference, the corresponding predictions for shallow crustal earthquakes in California (from Atkinson and Silva, 2000) are also shown. The ground motions predicted by the Atkinson and Silva relations are very similar to those predicted by the empirical relations of both Abrahamson and Silva (1997) and Sadigh *et al.* (1997); thus the choice of California relations for the comparison is not critical. All relations are plotted for NEHRP C sites. Observe that for large-magnitude ($\mathbf{M} > 7$) events within 100 km of the fault, the in-slab motions are higher than both the corresponding California crustal motions and the interface motions by more than a factor of 2 at most frequencies, in spite of the more rapid decay of motions with distance for in-slab events. The interface motions for large events are broadly similar to those for California crustal events of a similar size and distance.

Analysis of Ground-Motion Residuals by Magnitude, Soil Type, and Region

To examine whether the large data variability includes systematic over- or underprediction of records from certain regions or soil types, we examined figures and statistics showing subsets of the data plotted in Figures 11 and 12, broken down by magnitude, soil type, and region. Table 2 summarizes the mean residuals, for records of $\mathbf{M} \geq 6$ within 100 km of the fault, for both interface and in-slab events. Examination of this table shows that there are significant positive residuals (denoting underprediction) for events of

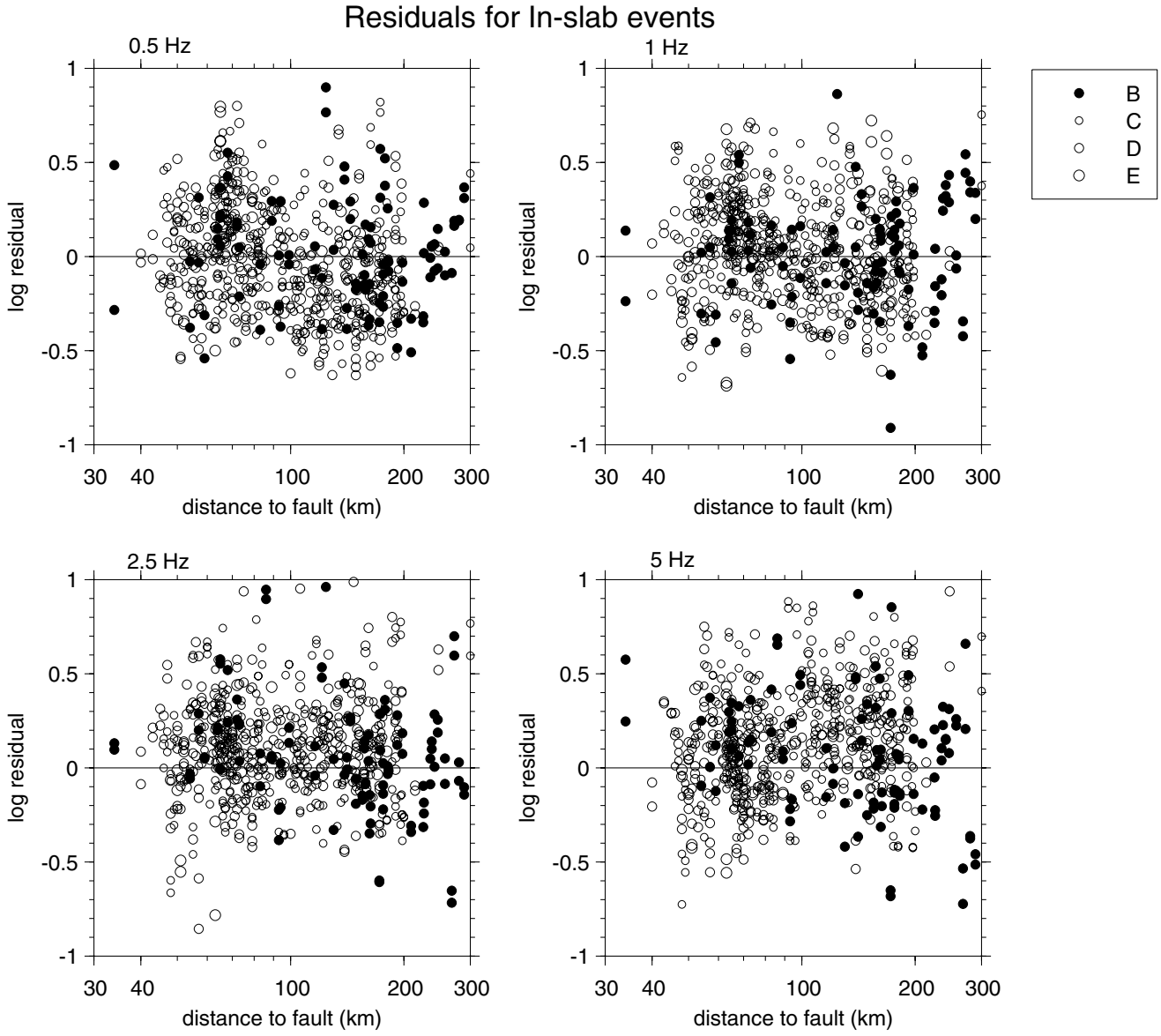

Figure 14. Log residuals (= log observed value − log predicted value) based on regression equation, for in-slab events, distinguished by site class.

M <6.6. This is attributable to our decision to force a linear scaling of amplitudes with magnitude, despite apparently high amplitudes at low magnitudes that would suggest a positive quadratic scaling (Fig. 5–8). The positive residuals are considered acceptable because this magnitude range is below that which contributes most strongly to hazard (at least for regions such as Cascadia). For larger magnitudes, the fit to the data is excellent.

Looking at the analysis of residuals by site class, there are large positive residuals for class C sites for interface events; most of these records are from Japan. However the class C residuals for in-slab events, which are from both Japan and Cascadia, do not show this trend. For other site classes, there are no overwhelming trends, except that positive residuals for E sites at high frequencies from interface events appear to be balanced by negative residuals for in-slab events. The reason for the positive residuals for interface events on NEHRP C sites will soon become apparent, as we examine the regional variability of residuals.

The analysis of residuals by region shows that the Cascadia in-slab data tend to have positive residuals at lower frequencies, with negative residuals at higher frequencies. Note that, because we are looking at the residuals for events of **M** ≥6 at distances less than 100 km (the important range for hazard analysis), the Cascadia in-slab data are almost entirely from the 2001 **M** 6.8 Nisqually event (plus a few records from the 1949 and 1965 Seattle events). The positive trend in the Cascadia residuals at low frequencies is balanced out by the reverse trend in the Japan in-slab data. A detailed analysis of Nisqually earthquake versus the 2001 **M** 6.8

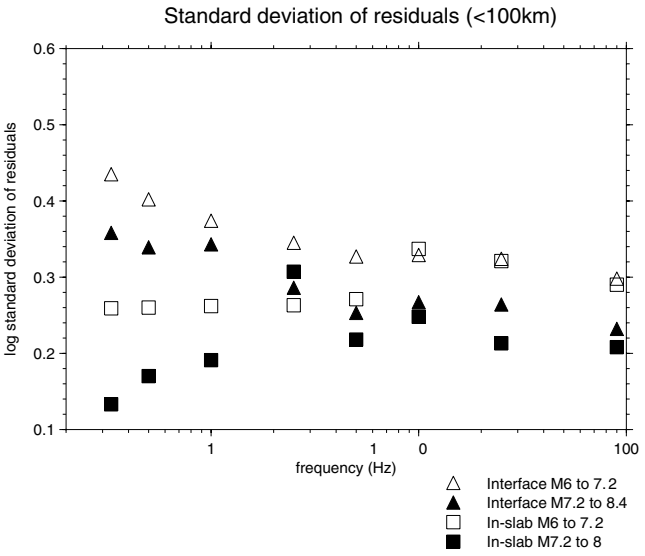

Figure 15. Variability (standard deviation of log residuals) of ground-motion estimates. Triangles show variability for interface events, while squares show variability for in-slab events. Open symbols used for events of **M** 6–7.2, with filled symbols for larger events.

Geiyo, Japan, records suggests that this is likely due to differences in typical soil profiles, resulting in systematic regional differences in site amplification within the same NEHRP class (Atkinson and Casey, 2003). Sites in Japan, of all NEHRP classes, are typically shallow soil over rock, which tends to amplify high frequencies. In the Cascadia region, the NEHRP B class represents rock or glacial till, while the NEHRP C/D/E classes represent relatively deep layers over rock or till. Computation of theoretical amplification functions for typical generic soil profiles for each region, using the method of Boore and Joyner (1997), indicates that sites in Japan are expected to have larger amplifications at high frequencies, while sites in Cascadia will have larger amplifications at low frequencies (Atkinson and Casey, 2003).

The high positive residuals for interface events in some of the groupings are influenced by inclusion of the 1992 **M** 7.1 Cape Mendocino, California, earthquake in our subduction-zone earthquake database (the residuals from this event are identified in Figs. 11 and 13). We initially classified this as an interface event and thus included it in our analyses. However, there are arguments that could be made for treating it as a shallow crustal event. The argument in favor of considering it as an interface event is that the projection of the fault plane crops out near the seaward edge of the subduction zone (Oppenheimer *et al.,* 1993). On the other hand, this may be a crustal event because it was significantly shallower than the pre-earthquake seismicity in the area, and it occurred in a volume having seismic velocities that match crustal velocities rather than subducted Gorda plate velocities (Oppenheimer *et al.,* 1993). Thus it may be that this event is not really representative of interface events, but might be more nearly considered a California crustal earthquake (this is the only California earthquake in the database). Excluding the Cape Mendocino event from the database has only a minor effect (a few percent) on the regression, reducing the predictions somewhat for interface events of $\mathbf{M} < 7.5$.

In Figure 19 the regional variability of the ground-motion residuals is examined more closely, combining residuals from all records of $\mathbf{M} \geq 6$ within 100 km of the fault for both interface and in-slab events, for all soil types. The average residual over all regions is near zero, but does not equal zero at all frequencies since this data range does not exactly match that used in the regressions (in our analysis of residuals we focus on the variability for larger magnitudes at distances less than 100 km from the fault). Relative to this average, the Cascadia region residuals are noticeably positive at low frequencies and noticeably negative at high frequencies. By contrast, the Japan residuals show the opposite trend. For other regions, the averages are not much different from zero (some of these regions have a very limited number of records). As discussed above, the opposing trends for Japan versus Cascadia can be attributed to regional differences in the depth of typical soil profiles. This points to a limitation of the NEHRP site classes in characterizing the soil types; the NEHRP classes capture the amplification effects due to the average shear-wave velocity but not its frequency dependence due to the depth of the soil profile. Consequently, there are limitations on the applicability of global relations such as this one to any specific region. It is preferable to develop relations using data just from the region of interest. When that is not possible due to limitations of the database, an alternative would be to first adjust all recordings to a common site condition using regional soil amplification factors, prior to regression. If these factors are not known in advance, they can be determined from analysis of the residuals, as has been done here. In this case, the applicability of the relations to a specific region may be improved by using these average residuals as regional correction factors. Table 3 lists such regional correction factors for Cascadia and Japan. They were derived by subtracting the average residual over all regions from the corresponding residual for the specified region. To implement these correction factors, they should be added to the c_1 coefficients of Table 1. The revised c_1 coefficients that result are listed in Table 3.

As a final check on these adjusted equations, we replot the residuals for in-slab events by region, after applying the regional correction factors to Cascadia and Japan, as shown on Figure 20. The residuals for both Cascadia and Japan are acceptable, although there are still some low-amplitude residuals for Cascadia events of $\mathbf{M} < 6.5$ at high frequencies (mostly from the Satsop event). Table 4 provides statistics on the region-adjusted residuals (data from all regions are included) for the magnitude–distance range of most interest to hazard analysis: $\mathbf{M} \geq 7.5$ at $D_{\text{fault}} < 300$ km for interface events and $\mathbf{M} \geq 6.5$ at $D_{\text{fault}} < 100$ km for in-slab events.

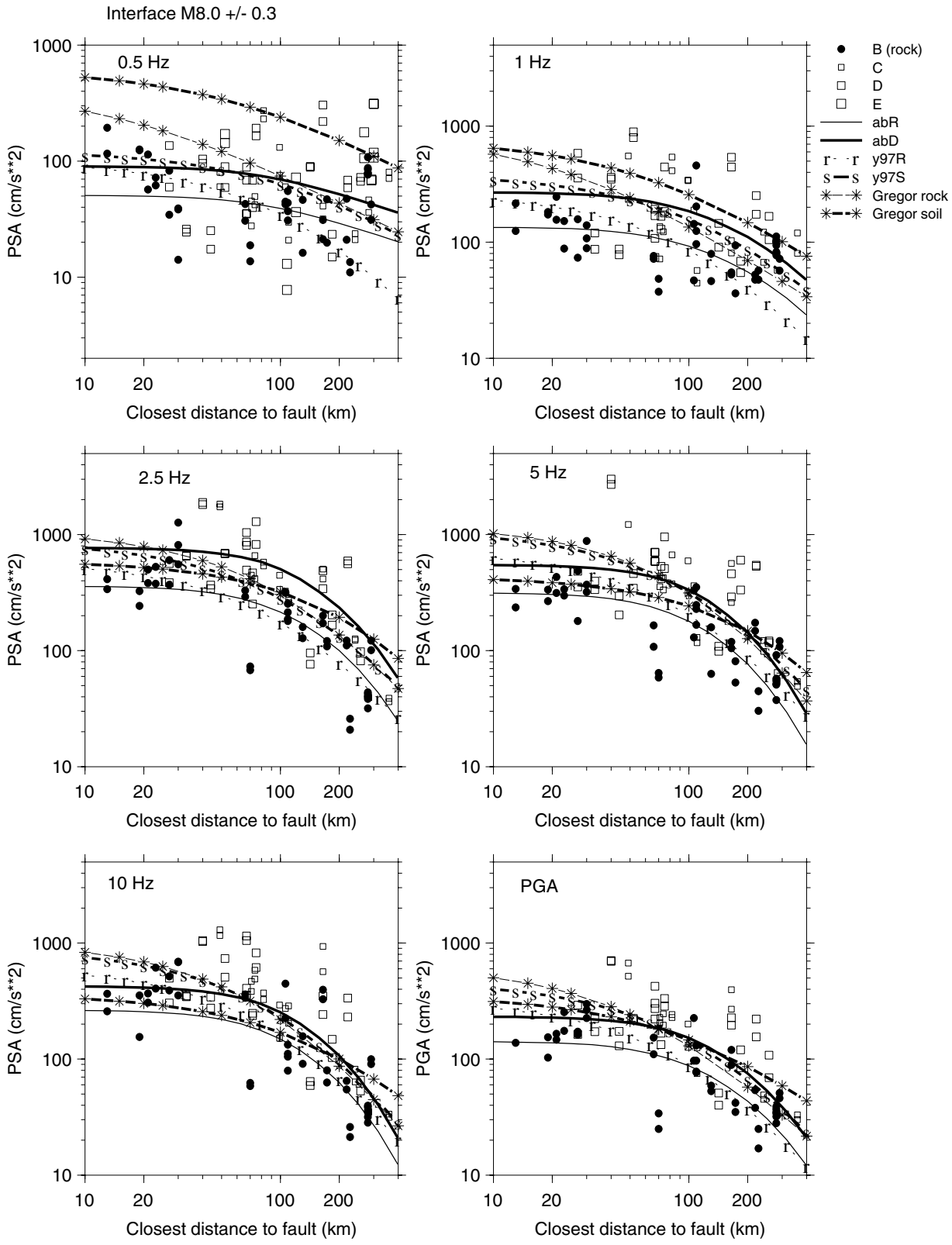

Figure 16. Comparison of amplitudes predicted by our regression equations for rock (abR) and NEHRP D soil (abD) to observed amplitudes, for B (rock), C, D, and E data, for interface events ($h = 20$ km) of **M** 8 $\pm$ 0.3. Corresponding predictions of Youngs *et al.* for rock (y97R) and soil (y97S) are also shown, along with stochastic model predictions of Gregor *et al.* (2002).

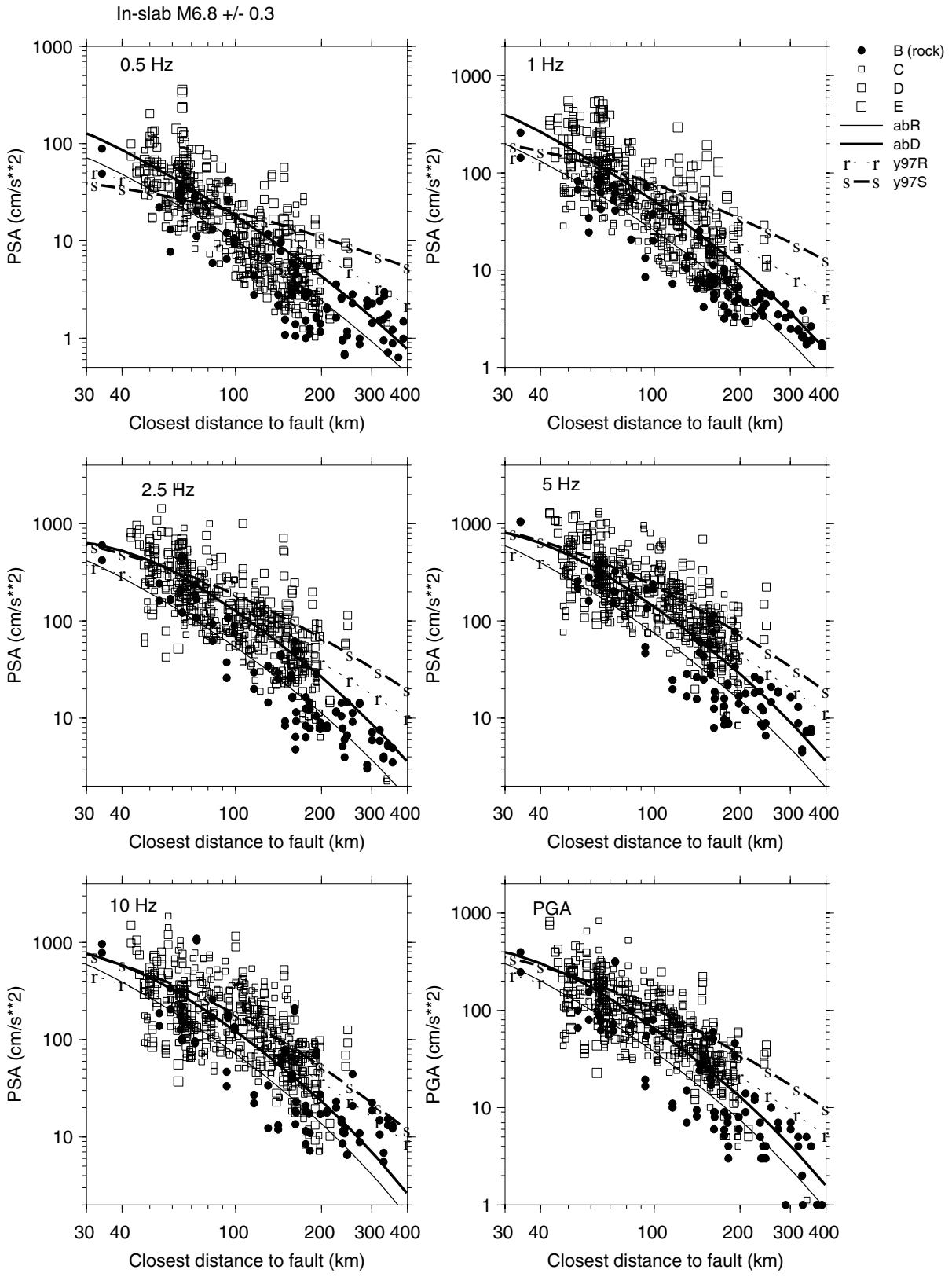

Figure 17. Comparison of amplitudes predicted by our regression equations for rock (abR) and soil (abD) to observed amplitudes, for B (rock), C, D, and E data, for in-slab events ($h = 50$ km) of **M** 6.8 $\pm$ 0.3. Corresponding predictions of Youngs *et al.* for rock (y97R) and soil (y97S) are also shown.

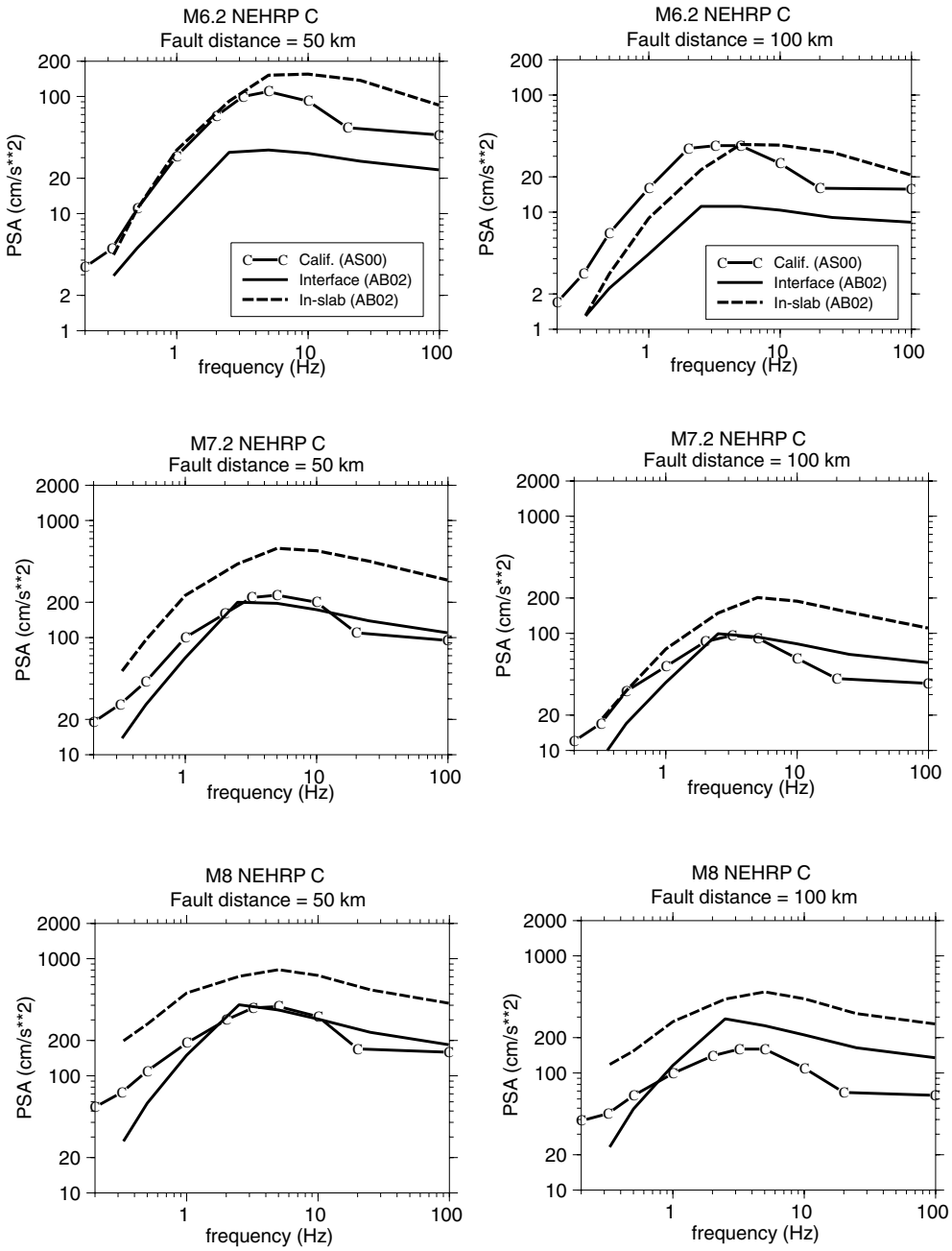

Figure 18. Predicted spectra for earthquakes of **M** 6.2 (top frames), 7.2 (middle frames), and 8.0 (lower frames), at distances of 50 km (left) and 100 km (right) from the fault, for NEHRP C site conditions, for interface (solid lines) and in-slab (dashed lines) events. Corresponding predictions for shallow California earthquakes (from Atkinson and Silva, 2000) are denoted by lines with "C."

Note that in some cases the standard deviation has been reduced from that quoted in Table 1, due to the slightly different magnitude–distance criteria applied for selection of records and due to the application of the regional factors for Cascadia and Japan. In a few instances the standard deviation is slightly higher due to the broader distance range considered for the interface events. The user may select either set of standard deviations (Table 1 or Table 4) depending on which better matches the magnitude–distance range of interest for their application. The intraevent component of variability (σ_1) would be the same in either case.

Conclusions

We have presented ground-motion relations for interface and in-slab earthquakes based on regression of a large

Table 2
Mean Residuals (log units) for Records of **M** ≥ 6.0 within 100 km of the Fault

Case	No. Records	0.5 Hz	1.0 Hz	2.5 Hz	5.0 Hz	10.0 Hz	Std. Error*
Interface Events							
Overall	194	0.04	0.05	0.07	0.07	0.17	0.03
By magnitude							
6.0 ≤ **M** < 6.6	30	−0.04	−0.02	0.24	0.34	0.43	0.05
6.6 ≤ **M** < 7.2	87	0.15	0.11	0.09	0.06	0.19	0.03
7.2 ≤ **M** ≤ 7.7	30	−0.11	0.01	−0.13	−0.07	−0.01	0.05
7.7 ≤ **M** ≤ 8.3	47	−0.01	0.01	0.03	0.02	0.10	0.04
By site class							
A/B	93	−0.02	−0.02	0.04	−0.04	0.17	0.03
C	35	0.28	0.25	0.21	0.24	0.17	0.05
D	62	−0.01	0.04	0.02	0.13	0.18	0.04
E	4	0.17	0.05	0.07	0.25	0.14	0.15
By region							
Cascadia†	14	0.39	0.23	0.07	0.18	0.07	0.08
Japan	39	0.04	0.08	0.09	0.27	0.25	0.05
Other Pacific‡	125	−0.08	−0.05	0.02	−0.03	0.14	0.03
In-slab Events							
Overall	302	0.03	0.05	0.08	0.06	−0.02	0.02
By magnitude							
6.0 ≤ **M** < 6.6	24	0.03	0.12	0.15	0.14	0.00	0.06
6.6 ≤ **M** < 7.2	264	0.03	0.05	0.07	0.06	−0.02	0.02
7.2 ≤ **M** ≤ 7.7	14	−0.12	−0.01	0.03	−0.09	−0.06	0.08
By Site Class							
A/B	24	−0.02	−0.06	0.09	0.17	0.12	0.06
C	158	0.05	0.09	0.10	0.06	−0.02	0.02
D	88	−0.02	0.02	0.05	0.05	−0.02	0.03
E	32	0.08	0.01	0.04	−0.02	−0.16	0.05
By Region							
Cascadia†	162	0.19	0.11	0.07	−0.03	−0.19	0.02
Japan	136	−0.07	0.05	0.12	0.21	0.20	0.03
Other Pacific‡	20	−0.10	−0.03	0.02	−0.01	0.03	0.06

*Typical standard error of the mean residuals (all frequencies).
†Cascadia interface is entirely the Cape Mendocino event (crustal?).
‡Includes Alaska, Mexico, Central and South America.

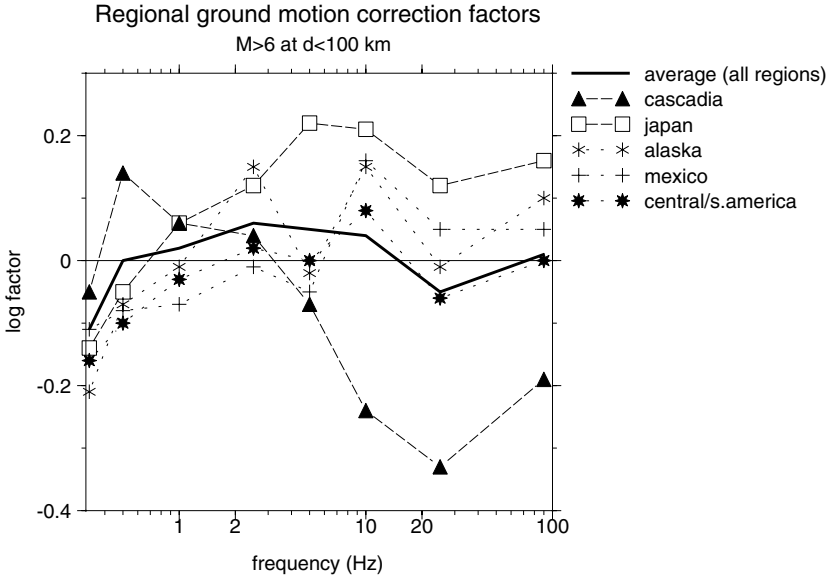

Figure 19. Correction factors to account for regional differences in ground-motion amplitudes. Factors are the average residual (in log units) for all records of **M** ≥6 within 100 km of the fault, within a given region (symbols). Heavy lines show average residual for these records over all regions.

Table 3
Recommended Regional Ground-Motion Correction Factors (log units)

Frequency (Hz)	Factor		C_1 Interface		C_1 In-Slab	
	Cascadia	Japan	Cascadia	Japan	Cascadia	Japan
0.33	0.06	−0.03	2.36	2.27	−3.64	−3.73
0.5	0.14	−0.05	2.33	2.14	−2.25	−2.44
1.0	0.04	0.04	2.18	2.18	−0.98	−0.98
2.5	−0.02	0.06	2.50	2.58	−0.01	0.07
5.0	−0.12	0.18	2.54	2.84	0.40	0.70
10.0	−0.28	0.17	2.50	2.95	0.16	0.61
25.0	−0.28	0.17	2.60	3.05	0.23	0.68
PGA	−0.20	0.15	2.79	3.14	−0.25	0.10

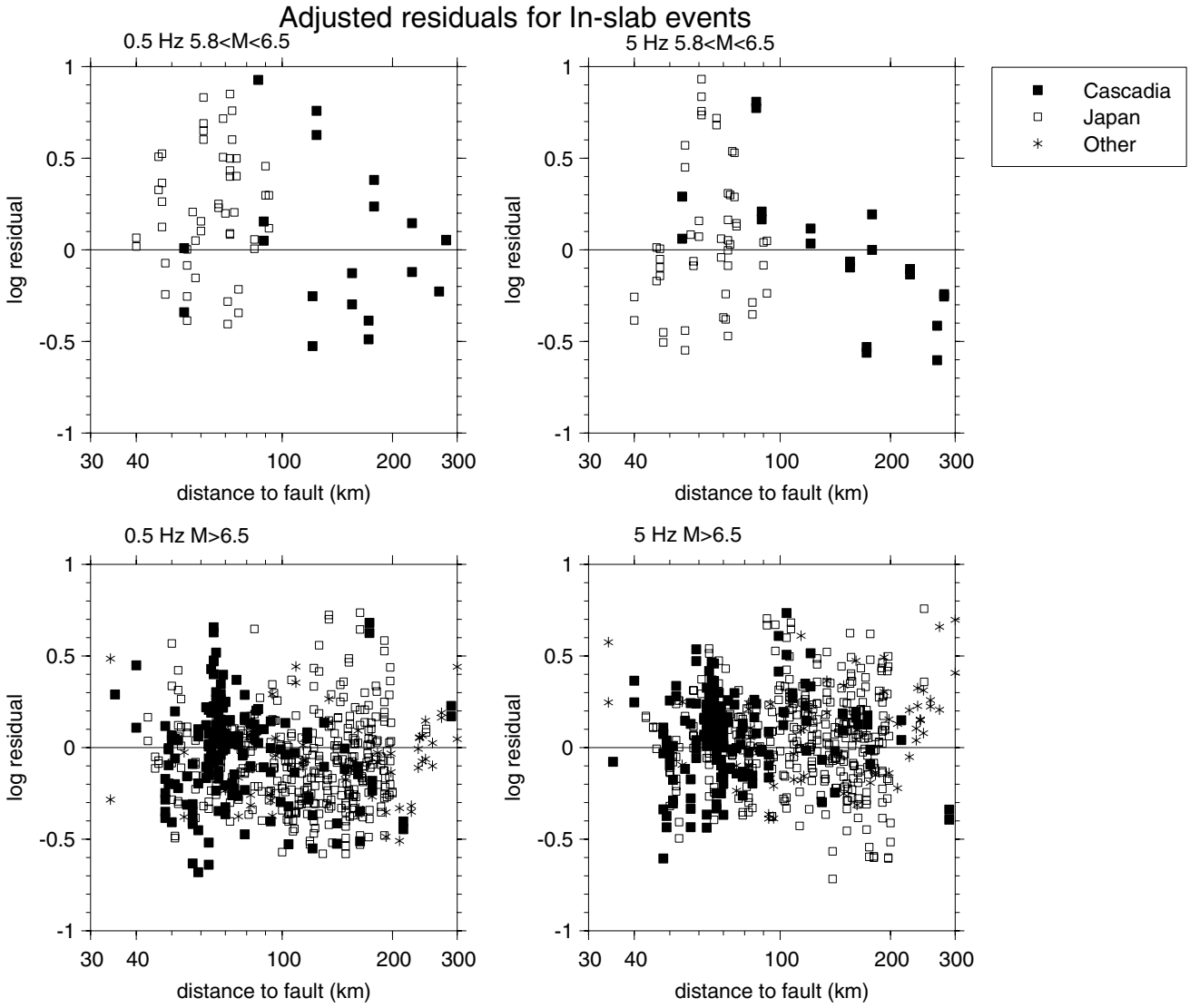

Figure 20. Replotted residuals for in-slab events at frequencies of 0.5 and 5 Hz, after correcting the Cascadia and Japan equations using the factors of Table 3.

Table 4
Residual Statistics after Application of Regional Correction Factors

	Interface $\mathbf{M} \geq 7.5$ at $D_{fault} < 300$ km		In-slab $\mathbf{M} \geq 6.5$ at $D_{fault} < 100$ km	
Frequency (Hz)	Mean Residual	Standard Deviation	Mean Residual	Standard Deviation
0.33	−0.09	0.36	−0.15	0.23
0.5	−0.02	0.33	−0.03	0.23
1.0	0.03	0.31	0.01	0.25
2.5	−0.05	0.26	0.06	0.26
5.0	−0.01	0.28	0.04	0.23
10.0	0.01	0.30	0.05	0.25
25.0	−0.06	0.26	−0.04	0.23
PGA	0.00	0.24	0.01	0.23

global database, containing thousands of ground-motion records for events of **M** ≥5 at distances up to several hundred kilometers. The results of this study differ significantly from previous analyses based on more limited data and contain more detailed analyses of the variability of ground motion with region, event type, and soil class. The new results have important implications for seismic-hazard analysis. The ground-motion relations for global events predict that a great megathrust earthquake (**M** ≥8) at a fault distance of about 100 km would produce PSA on soil sites of about 110 cm/sec^2 at 0.5 Hz, 660 cm/sec^2 at 2.5 Hz, and 410 cm/sec^2 at 5 Hz, with a PGA of about 180 cm/sec^2. These damaging levels of motion would be experienced over a very large area, corresponding to a rectangle about 300 km wide by 500 km long. Large in-slab events (**M** 7.5) would produce even higher PSA values within 100 km of the fault, but the in-slab motions attenuate much more rapidly with distance. Thus the damage potential of events like the 2001 **M** 6.8 Nisqually earthquake is modest relative to that of a large megathrust event. This is true not only for low-frequency ground motion, but even for 5 Hz. This suggests that the contemporary view of seismic hazard for southwestern British Columbia, namely that in-slab events dominate the hazard for 5 Hz for the 2% in 50-year probability level (Adams and Halchuk, 2000), may need revision. The hazard posed by moderate to large in-slab events such as the 2001 Nisqually earthquake is modest compared to that of a Cascadia megathrust earthquake of **M** ≥8, in terms of the area that would experience damaging levels of ground motion.

Analysis of regional variability of ground-motion amplitudes suggests that the use of regression equations based on a global subduction database is not well justified. In particular, amplitudes in Cascadia differ by more than a factor of 2 from those in Japan for the same magnitude, distance, event type, and NEHRP soil class. This is believed to be due to regional differences in the depth of the soil profile, which are not captured by the NEHRP site classification scheme. Regional correction factors to account for these differences are proposed for Cascadia and Japan.

Acknowledgments

We gratefully acknowledge the generous contributions of C. B. Crouse and Bob Youngs in making their subduction ground-motion databases available to us, Ralph Haugerud and Rob Williams for providing information about site classes for the region around Seattle, Washington, and Pat McCrory for discussions and information about whether the 1992 Cape Mendocino earthquake should be classified as an interface or a crustal earthquake. We appreciate the constructive reviews by Art Frankel, Chuck Mueller, and Ivan Wong, and two anonymous referees for the Bulletin. This work was supported by the USGS National Earthquake Hazards Reduction Program under Grant 99HQGR0021. All figures were prepared using the graphics software package CoPlot (www.cohort.com).

References

Abrahamson, N., and W. Silva (1997). Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seism. Res. Lett.* **68,** 94–127.

Adams, J., and S. Halchuk (2000). Knowledge of in-slab earthquakes needed to improve seismic hazard estimates for southwestern British Columbia, in *Proc. of the Workshop on In-Slab Earthquakes in the Cascadia Subduction System,* www.seismo.nrcan.gc.ca/hazards/inslab2000/inslab2000.html (last accessed 2002).

Adams, J., D. Weichert, and S. Halchuk (1999). Seismic hazard maps of Canada—1999: 2%/50 Year values for selected Canadian cities, Geological Survey of Canada Open File, 95 pp., www.seismo.nrcan.gc.ca (last accessed 2002).

Atkinson, G. (1997). Empirical ground motion relations for earthquakes in the Cascadia region, *Can. J. Civil Eng.* **24,** 64–77.

Atkinson, G., and R. Casey (2003). A comparative study of the 2001 Nisqually, Washington and Geiyo, Japan in-slab earthquakes *Bull. Seism. Soc. Am.* (in press).

Atkinson, G., and J. Cassidy (2000). Integrated use of seismograph and strong-motion data to determine soil amplification in the Fraser Delta: results from the Duvall and Georgia Strait earthquakes, *Bull. Seism. Soc. Am.* **90,** 1028–1040.

Atkinson, G., and W. Silva (2000). Stochastic modeling of California ground motions, *Bull. Seism. Soc. Am.* **90,** 255–274.

Boore, D., and W. Joyner (1997). Site amplifications for generic rock sites, *Bull. Seism. Soc. Am.* **87,** 327–341.

Centroid Moment Tensor (CMT) Catalog, www.seismology.harvard.edu/CMTsearch.html (last accessed 2002).

Chen, S., and G. Atkinson (2002). Global comparisons of earthquake source spectra, *Bull. Seism. Soc. Am.* **92,** 885–895.

Crouse, C. (1991). Ground-motion attenuation equations for Cascadia subduction zone earthquakes, *Earthquake Spectra* **7,** 201–236.

Crouse, C., Y. Vyas, and B. Schell (1988). Ground motions from subduction zone earthquakes, *Bull. Seism. Soc. Am.* **78,** 1–25.

Dobry, R., R. Borcherdt, C. Crouse, I. Idriss, W. Joyner, G. Martin, M. Power, E. Rinne, and R. Seed (2000). New site coefficients and site classification system used in recent building seismic code provisions, *Earthquake Spectra* **16,** 41–67.

Frankel, A., D. Carver, E. Cranswick, M. Meremonte, T. Bice, and D. Overturf (1999). Site response for Seattle and source parameters of earthquakes in the Puget Sound region, *Bull. Seism. Soc. Am.* **89,** 468–483.

Frankel, A., C. Mueller, T. Barnhard, D. Perkins, E. Leyendecker, N. Dickman, S. Hanson, and M. Hopper (1996). National seismic hazard maps: documentation June 1996. *U.S. Geol. Surv. Open-File Rept. 96-532,* 69 pp.

Fundamental Research on Earthquakes and Earth's Interior Anomalies (FREESIA), www.fnet.bosai.go.jp (last accessed 2002).

Gregor, N., W. Silva, I. Wong, and R. Youngs (2002). Ground motion attenuation relationships for Cascadia subduction zone megathrust

earthquakes based on a stochastic finite-fault model, *Bull. Seism. Soc. Am.* **92,** 1923–1932.

Heenan, S. (2002). Empirical study of the ground motion of earthquakes originating in the subduction zones of Japan and Cascadia. *Master's Thesis,* Carleton University, Ottawa.

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

Joyner, W. B., and D. M. Boore (1994). Erratum, *Bull. Seism. Soc. Am.* **84,** 955–956.

Kyoshin-Net (KNET). Japan strong-motion data center, www.k-net.bosai.go.jp (last accessed 2002).

Molas, G., and F. Yamazaki (1995). Attenuation of earthquake ground motions in Japan including deep focus events, *Bull. Seism. Soc. Am.* **85,** 1343–1358.

Oppenheimer, D., G. Beroza, G. Carver, L. Dengler, J. Eaton, L. Gee, F. Gonzalez, A. Jayko, W. H. Li, M. Lisowski, M. Magee, G. Marshall, M. Murray, R. McPherson, B. Romanowicz, K. Satake, R. Simpson, P. Somerville, R. Stein, and D. Valentine (1993). The Cape Mendocino, California, earthquakes of April 1992: subduction at the triple junction, *Science* **261,** 433–438.

Sadigh, K., C. Chang, J. Egan, F. Makdisi, and R. Youngs (1997). Attenuation relationships for shallow crustal earthquakes based on California strong-motion data, *Seism. Res. Lett.* **68,** 180–189.

Wells, D., and K. Coppersmith (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seism. Soc. Am.* **84,** 974–1002.

Williams, R. A., W. J. Stephenson, A. D. Frankel, and J. K. Odum (1999). Surface seismic measurements of near-surface *P*- and *S*-wave seismic velocities at earthquake recording stations, Seattle, Washington, *Earthquake Spectra* **15,** 565–583.

Youngs, R., S. Chiou, W. Silva, and J. Humphrey (1997). Strong ground motion attenuation relationships for subduction zone earthquakes, *Seism. Res. Lett.* **68,** 58–73.

Youngs, R., S. Day, and J. Stevens (1988). Near field ground motions on rock for large subduction zone earthquakes, in *Earthquake Engineering and Soil Dynamics II,* American Society of Civil Engineers Geotech. Special Publication no. 20, 445–462.

Appendix A

Assigning NEHRP Site Classes for KNET Stations

Shear-wave velocities have been determined for all but one KNET station from measurements in boreholes (the exception is CHB023). Unfortunately, with one exception, the depths of all holes are between 10 and 20 m (the one exception is station AKT019, with a depth to the bottom of 5.04 m). For this reason, the average velocity to 30 m, and therefore the NEHRP site classes, cannot be precisely determined for the KNET stations. Rather than abandon the KNET data, we assigned site classes based on a statistical approach, using data gathered by the U.S. Geological Survey from 69 boreholes in California.

A provisional NEHRP site class was assigned for each KNET station by assuming that the velocity at the bottom of the associated borehole (V_{bot}) extends to 30 m. Because velocities generally increase with depth, however, there is a finite chance that the velocities between the bottom of the borehole and 30 m are high enough that the NEHRP site class should be increased to a stiffer (faster) class. To account for this in a statistical way, we computed the ratio of V_{bot} to the effective constant velocity (V_{eff}) from the bottom of the borehole to 30 m needed to raise the site class to the next stiffer class and then used the probability of the required V_{eff}/V_{bot}, given the depth of the bottom of the KNET borehole, to decide if the provisional site class should be changed. We derived a probability distribution for V_{eff}/V_{bot} using data from 69 boreholes in California. V_{eff}/V_{bot} from the 69 boreholes were computed for depths between 10 and 20 km, and then for each depth, the values of V_{eff}/V_{bot} were tabulated in increments of 0.1 unit, starting from 0.4, and a cumulative distribution was computed. A variety of functions were fit to the empirical distribution. A power law gave the best fit for $V_{eff}/V_{bot} > 1.0$ (see Fig. A1 for the empirical and power-law fits for depths of 10 and 20 m). The power-law distribution was then used to decide if the site class should be changed, given the KNET borehole depth and the value of V_{eff}/V_{bot} for that borehole.

A concrete example will help explain our procedure. Assume that the borehole at a particular KNET station was 14 m deep and that $V_{eff}/V_{bot} = 1.31$ in order for the site class to change. Substituting 1.31 into the power-law fit for

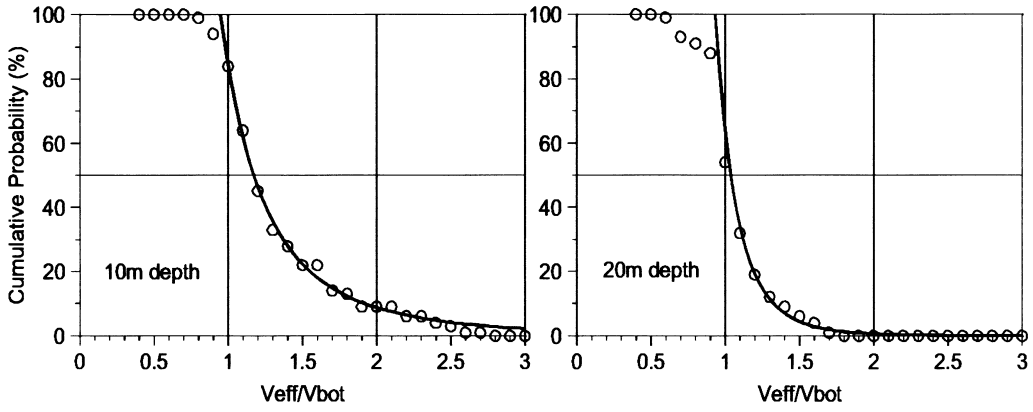

Figure A1. Cumulative probability of the ratio V_{eff}/V_{bot} from 69 boreholes in California, for bottom depths of 10 and 20 m. Also shown are the power-law fits to the observations.

14 m depth gives a cumulative percentage of 21% (21% of the California boreholes have a velocity that increases enough below 14 m to give a V_{03} large enough to change the site class). We then decided whether or not to change the site class for each recording at that particular KNET station by generating a random number between 0 and 1 and then multiplying that number by 100. If the resulting number was less that 21, we changed the site class. Using this procedure, we changed the site class for 89 out of 1001 stations. We are not claiming that the site classes assigned to the KNET stations are correct for each station; this procedure only makes sense in a statistical way, when all the data are being thrown into a big pot without identifying each station.

We only consider cases when the velocity might have bumped the class into the next firmer class, not the other way, although as Figure A1 shows, there is a finite probability that it could go the other way. The power-law fit is clearly invalid for cases where $V_{\text{eff}}/V_{\text{bot}} < 1$ (as would be required to change to a softer site class), so in order to use the procedure to reassign site to softer class, we would have to derive a different functional form. We decided that physically it is more likely that the site class would move to a stiffer rather than a softer class, and therefore we decided to retain the simplicity of the power-law fit by only considering increases in site class.

The power-law distributions were only derived for borehole depths between 10 and 20 m, and therefore they could not be used to assess the site class at the KNET station AKT019, for which the borehole was only 5 m deep. For this station, $V_{\text{eff}}/V_{\text{bot}}$ is so close to unity (1.23) that we changed the site class for this station (from D to C).

As mentioned before, the site classes at 89 out of 1001 stations were changed; this resulted in changes in site class for 441 of the 6307 KNET recordings considered in the analysis in this article. This seems reasonable.

Dept. of Earth Sciences
Carleton University
Ottawa, Ontario
Canada K1S 5B6
gma@ccs.carleton.ca
(G.M.A.)

U.S. Geological Survey
345 Middlefield Rd.
Menlo Park, California 94025
boore@usgs.gov
(D.M.B.)

Manuscript received 15 July 2002.