PACIFIC EARTHQUAKE ENGINEERING
RESEARCH CENTER

Boore-Atkinson NGA Ground Motion Relations for the Geometric Mean Horizontal Component of Peak and Spectral Ground Motion Parameters

David M. Boore
U.S. Geological Survey, Menlo Park, California

and

Gail M. Atkinson
University of Western Ontario, Canada

PEER 2007/01
MAY 2007

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
PEER 2007/01		

4. Title and Subtitle	5. Report Date
Boore-Atkinson NGA Ground Motion Relations for the Geometric Mean Horizontal Component of Peak and Spectral Ground Motion Parameters.	May 2007
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
David M. Boore and Gail M. Atkinson	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Pacific Earthquake Engineering Research Center 325 Davis Hall MC 1792 University of California Berkeley, CA 94720	
	11. Contract or Grant No. 65A0058

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
California Department of Transportation Engineering Service Center 1801 30th St., West Building MS-9 Sacramento, CA 95807	Technical report through June 2006
	14. Sponsoring Agency Code 06681

15. Supplementary Notes

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.

16. Abstract

This report contains ground motion prediction equations (GMPEs) for a particular measure of horizontal-component ground motions as a function of earthquake mechanism, 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. The equations were derived by empirical regression of the PEER NGA strong-motion database. For periods of less than 1 s, the analysis used 1574 records from 58 mainshocks in the distance range from 0 km to 400 km (the number of available data decreased as period increased).

17. Key Words	18. Distribution Statement
Ground motion prediction equations, site amplification, record processing, response spectra, NEHRP site class	Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	242	

Boore-Atkinson NGA Ground Motion Relations for the Geometric Mean Horizontal Component of Peak and Spectral Ground Motion Parameters

David M. Boore
U.S. Geological Survey, Menlo Park, California

and

Gail M. Atkinson
Department of Earth Sciences
University of Western Ontario, Canada

PEER Report 2007/01
Pacific Earthquake Engineering Research Center
College of Engineering
University of California, Berkeley

May 2007

ABSTRACT

This report contains ground motion prediction equations (GMPEs) for a particular measure of horizontal-component ground motions as a function of earthquake mechanism, 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. The equations were derived by empirical regression of the PEER NGA strong-motion database. For periods of less than 1 s, the analysis used 1574 records from 58 mainshocks in the distance range from 0 km to 400 km (the number of available data decreased as period increased).

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, with an absolute lack of ego and no power politics. In addition, we thank these people, in alphabetical order: Sinan Akkar, John Douglas, Art Frankel, Vladimir Graizer, Steve Harmsen, Robert Herrmann, Tom Holzer, Charles Mueller, Maury Power, and Rakesh Saigal.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 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. (Some terminology surrounding these equations is discussed in Appendix A.) 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. 2006), using an extensive database of thousands of records compiled from active shallow crustal tectonic environments worldwide. These equations represent a substantive 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 increase 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.

2 Data

2.1 DATA SOURCES

The source of the strong ground motion data for the development of the GMPEs of this study is the database compiled in the PEER–NGA project (Power et al. 2006); the aim of this 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. We used what we call v.7.27 of the flatfile. This is the version given in the file *NGA Flatfile V7.2 (07-11-05).xls*, with corrections provided by the file *727Brian.xls* sent by Brian Chiou, in an email dated February 17, 2006. The values of the response variable are different in the two versions. In v.7.2, the rotation angle used to combine the two horizontal components (see next section) for each recording was based on all periods for which PSA was computed, regardless of the low-cut filter frequency applied to each record. In v.7.27, the rotation angle was determined using periods only up to the maximum useable period (T_{MAX}). As apparently a number of the NGA developers have used v7.2, we include a comparison of the response variables for the two versions in Appendix B. Although the bulk of the data are not changed between the two versions, there are some differences. The comparisons in Appendix B suggest that it should make little or no difference to the GMPEs whether version 7.2 or 7.27 of the flatfile is used; the best test of this, of course, is to develop GMPEs using both versions. We used the corrected version of the flatfile; testing the sensitivity of the GMPEs to the version of the flatfile is beyond the scope of our study.

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 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 U.S. 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).

2.2 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 geometric mean of the two horizontal component (as was used in BJF97), but rather the values 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 50^{th} percentile values of the geometric means computed for all nonredundant rotation angles and all periods of 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 component direction; in this case the geometric mean would be 0.

This report 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 dataset from records for which the low-cut filtering was not performed as part of the NGA project. Appendix C 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 aftershock

recordings 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 dataset almost in half because about half 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, taken from the accompanying spreadsheet *flag_definitions.xls.* Another spreadsheet, *recnum_flag.xls***,** contains a flag entry for each record in the NGA flatfile; only data with flag = 0 were used in our analysis. Response variables were excluded for oscillator periods greater than T_{MAX} (the inverse of the lowest useable frequency entry in the NGA flatfile).

Table 2.1 Exclusion criteria.

Flag	Meaning
1	one h component
2	Jensen Admin Bldg?
3	V_{S30} missing
4	spikes, baseline problems (see, e.g., Appendix H)
5	dam abutments
6	dam toe
7	base of column
8	base of pier
9	basement, 12.7 m below ground, 1.8 m above bedrock
10	basement, 6.4 m below ground
11	greater than or equal to 3 stories
12	S triggers
13	older events not included in Boore et al. (1997), probably because distances are too uncertain
14	proprietary records with restrictions on use
15	earthquake in oceanic crust
16	stable continental region (SCR) events
17	basement recordings
18	Geomatrix C, D, E, F, G, H, J (but not including Lexington Dam for LP89 and LA Dam for NR94, and Martis Creek Dam for 2001 Mohawk)
19	duplicate record?
20	aftershocks
21	Chi-Chi_quality D (Lee et al. 2001)
22	chi_chi_colocated (remove record from older instrument), Many such records were removed earlier because they are quality class D
23	second trigger
24	dam crest
25	only SMART1 data for this quake, should be considered a one observation earthquake (recall Boore et al. (1997) criteria).

A potential bias in regression results can result from not including low-amplitude data for distance ranges for which larger amplitude data for the same earthquake are included in the dataset. Low-amplitude data can be below trigger thresholds of instruments, can cause the recording to begin sometime during the S-wave arrival, can be too small to digitize, or 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 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, nontriggered 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 can also exist in nontriggered digital recordings because of the presence of long-period noise. It would be possible to devise a strategy to avoid potential bias by using information available in the NGA flatfile. For example, one could exclude all data beyond the closest station to a given event for which the maximum useable period of the processed recording is less than the oscillator period of interest. We did not attempt to define such a criterion, however, and as a result, we think that our distance dependence for small earthquakes and long periods is biased toward a decay that is less rapid than the true decay (discussed later).

2.3 PREDICTOR VARIABLES

The 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 continuous V_{S30} (time-averaged shear-wave velocity over the top 30 m) for site characterization. We also considered the effect of fault type (i.e., normal, strike-slip, and reverse). Each of these predictor variables was taken from the NGA database. The values of V_{S30} were taken from column IE of the NGA flatfile; the values in this column are based on measured values when available and

estimates otherwise (only about 30% of the stations that provided data in the NGA flatfile had a value of V_{S30} based on a local measurement). Column IE includes W. Silva's interpretation of NCREE measurements for Taiwan as well as Brian Chiou's correlation method for Taiwan, for sites in Taiwan for which measurements are not available. The values in column IE were updated from the file *Update 1 (02-17-06) to NGA Flatfile V7.2 (07-11-05).xls*, which uses some of Rob Kayen's V_{S30} estimates based on SASW measurements. The fault type was specified by the plunge of the *P*- and *T*-axes, as described in Appendix D. The R_{JB} distances estimated by Youngs (2005) were used for earthquakes with unknown fault geometry.

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

The distribution of data used to develop our GMPEs are shown in Figures 2.1–2.2 by **M** and R_{JB}, with the symbols representing different fault types. The distribution by V_{S30} (as given by binning the data by NEHRP site class) is given in Figures 2.3(a)–(b). The total number of recordings for the analysis (after all exclusions) is shown as a function of oscillator period in Figure 2.4 (which also breaks down the total number of recordings into digital and analog recordings). Figure 2.4 shows that the full dataset is available 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; note the dearth of analog recordings for periods of 9 and 10 s.

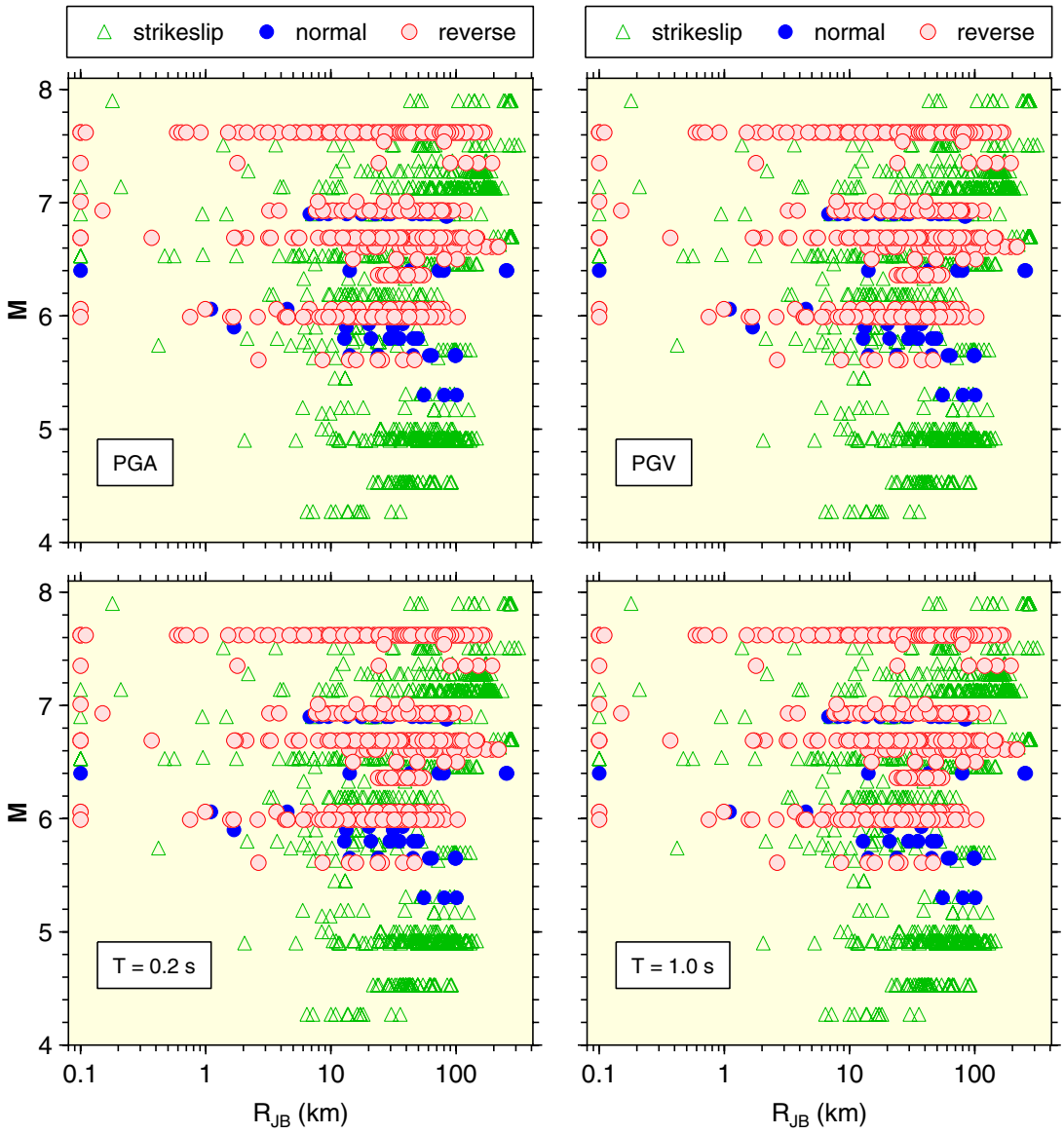


Fig. 2.1 Distribution of data used in BA07 regression equations for PGA, PGA, and PSA at 0.2 s and 1.0 s, differentiated by fault type (points with R_{JB} less than 0.1 km plotted at 0.1 km).

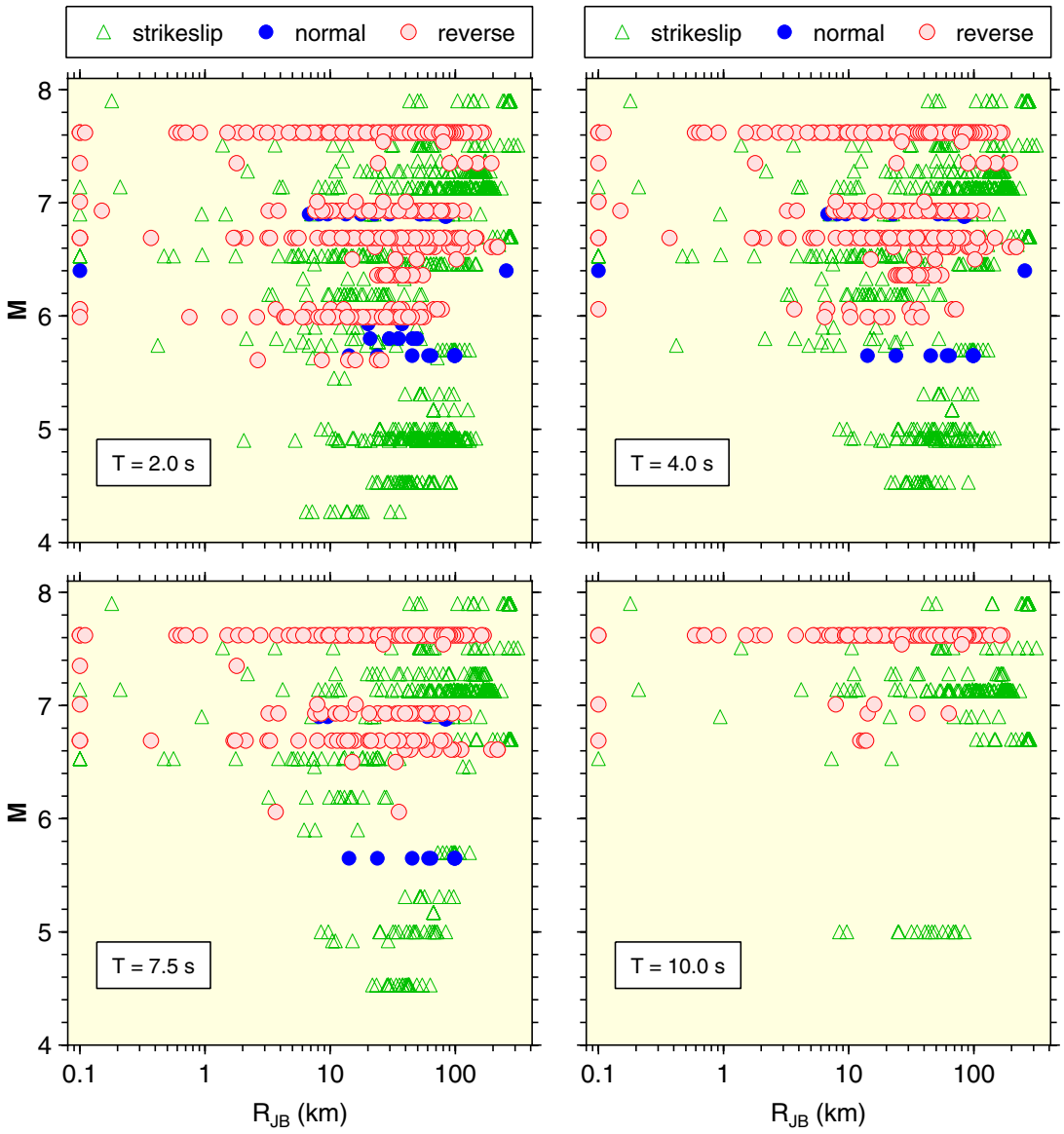
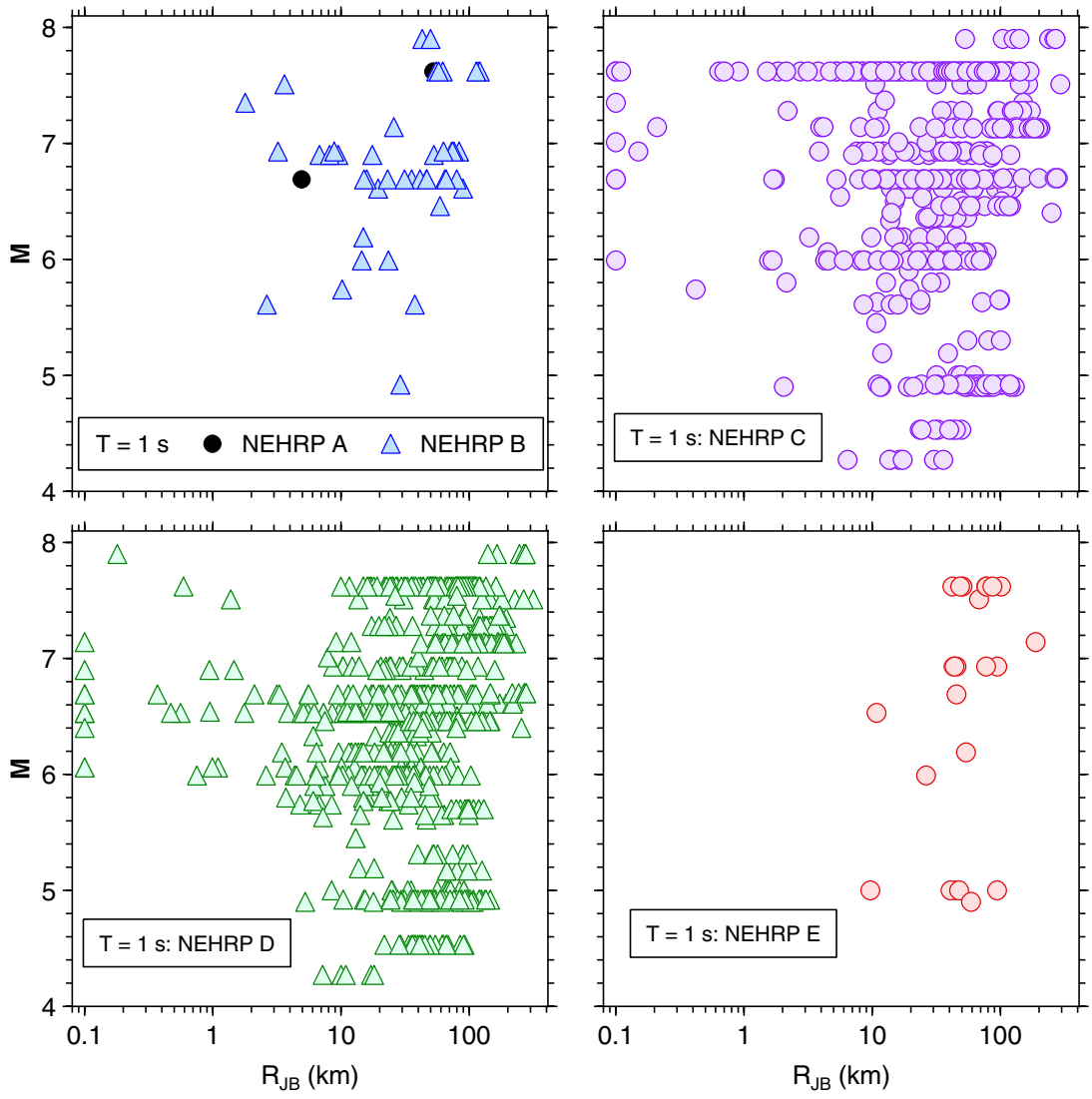
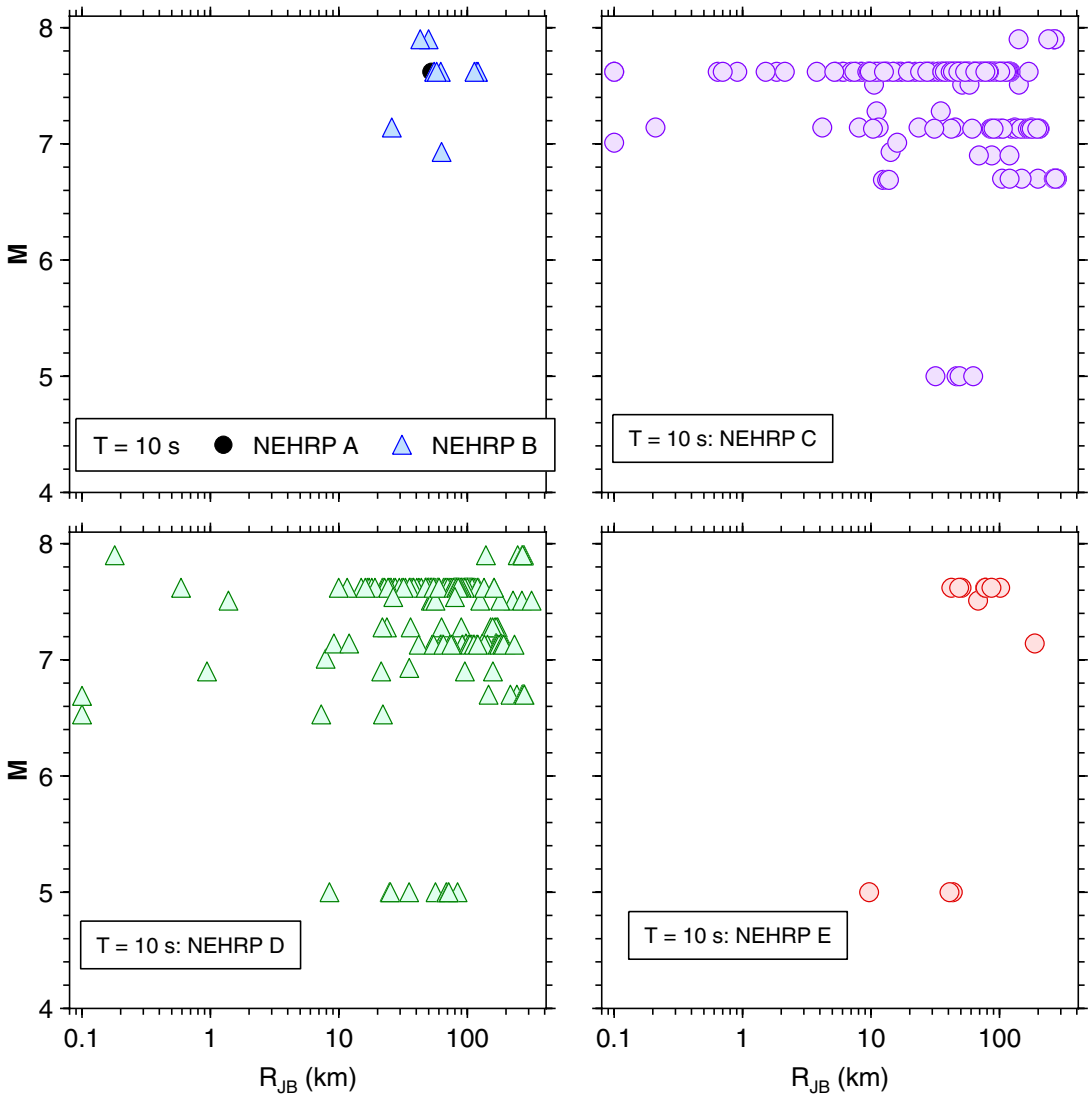


Fig. 2.2 Distribution of data used in BA07 regression equations for PSA at periods of 2.0, 4.0, 7.5, and 10.0 s, differentiated by fault type (points with R_{JB} less than 0.1 km plotted at 0.1 km).



(a)

Fig. 2.3 Distribution of data used in BA07 regression equations for PSA at 1.0 s for various NEHRP site classes (points with R_{JB} less than 0.1 km plotted at 0.1 km).



(b)

Fig. 2.3—*Continued*

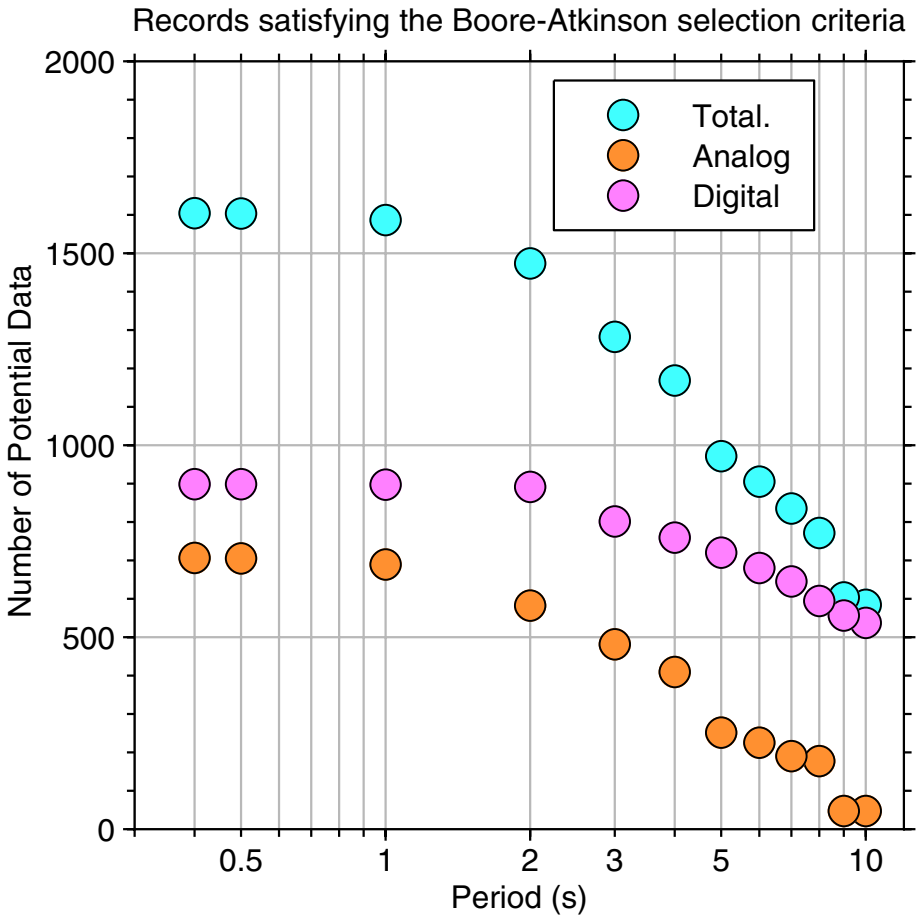


Fig. 2.4 Number of data in NGA flatfile satisfying our selection criteria, including limitation based on highest useable period, as function of oscillator period. Count includes a few recordings from singly recorded earthquakes that did not contribute to our GMPEs.

The distribution of the data by fault type, rake angle, and dip angle is shown in Figure 2.5. This figure also shows that the classification scheme used by BJF97 (shown by horizontal gray lines) produces essentially the distributions of fault type as the method used here, based on the plunge of the *P*- and *T*-axes.

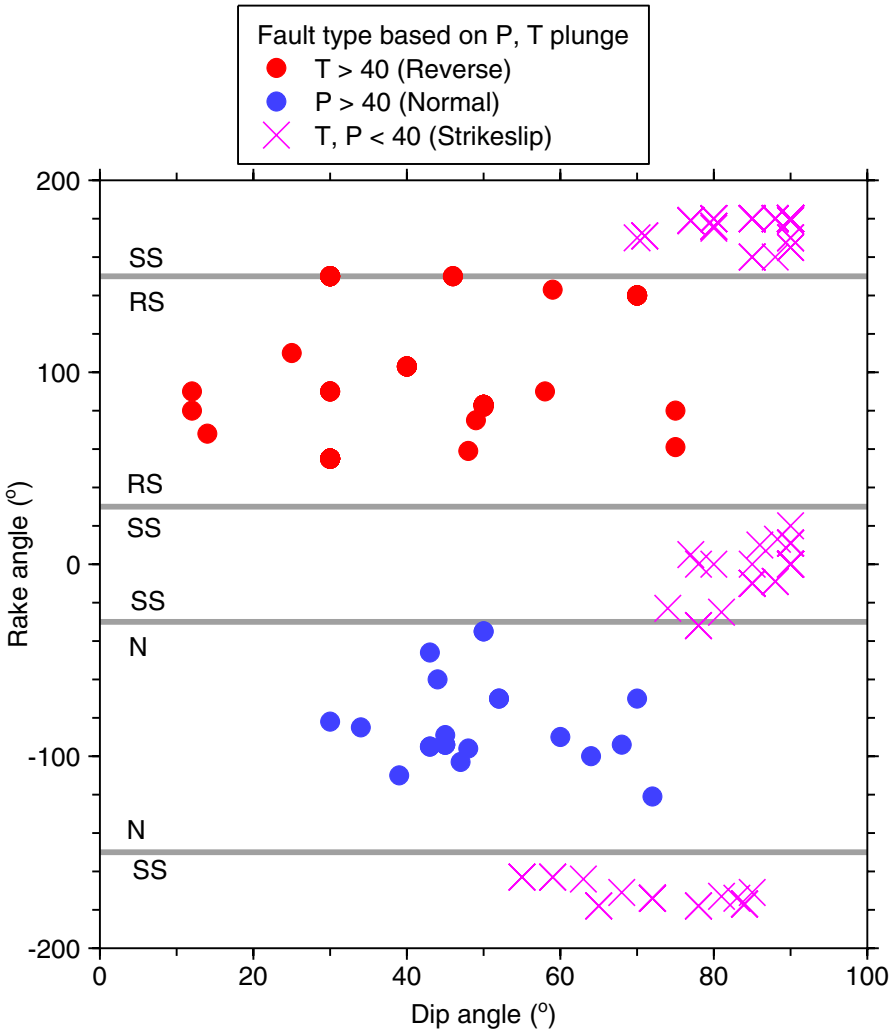


Fig. 2.5 Distribution of data we used in rake-angle and dip-angle space. Horizontal gray lines indicate boundaries between fault types used by BJF97, and symbols and colors indicate our classification based on plunges of P- and T-axes (our classification scheme indicated in legend).

The distributions of the data over the predictor variable space, as shown in Figures 2.1–2.5, 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 Fig. 2.2 for $T = 10$ 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, while the narrowest range is for normal-slip earthquakes. 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 distribution by site class shows that very few data were from class A sites (hard rock). The bulk of the data are from class C and D sites, which range from soft rock to firm soil. More detail can be found in Appendix E, which includes two possible sets of V_{S30} values to use in evaluating our equations for a particular NEHRP site class.

3 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 amplitudes with distance by a single magnitude-independent effective geometric spreading factor—a straight line amplitude decay of log amplitude versus log distance. This form appeared sufficient for the distance range of <80 km that BJF97 used. 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:

$$\ln Y = F_{\boldsymbol{M}}(\mathbf{M}) + F_D(R_{JB}, \mathbf{M}) + F_S(V_{S30}, R_{JB}, \mathbf{M}) + \varepsilon\sigma_T \,, \tag{3.1}$$

In this equation, $F_{\boldsymbol{M}}$, 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 **M**<6), and V_{S30} is the time-averaged shear-wave velocity over the top 30 m of the site. 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 (3.5) 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:

$$\sigma_T = \sqrt{\sigma^2 + \tau^2} \,, \tag{3.2}$$

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).

3.1 DISTANCE AND MAGNITUDE FUNCTIONS

The distance function is given by:

$$F_D(R_{JB}, \mathbf{M}) = [c_1 + c_2(\mathbf{M} - \mathbf{M}_{ref})]\ln(R / R_{ref}) + c_3(R - R_{ref}) \,, \tag{3.3}$$

where

$$R = \sqrt{{R_{JB}}^2 + h^2} \tag{3.4}$$

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) $\mathbf{M} \le \mathbf{M}_h$

$$F_M(\mathbf{M}) = e_1 U + e_2 SS + e_3 NS + e_4 RS + e_5(\mathbf{M} - \mathbf{M}_h) + e_6(\mathbf{M} - \mathbf{M}_h)^2 \,, \tag{3.5a}$$

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

$$F_M(\mathbf{M}) = e_1 U + e_2 SS + e_3 NS + e_4 RS + e_7(\mathbf{M} - \mathbf{M}_h), \quad (3.5b)$$

where *U*, *SS*, *NS*, and *RS* are dummy variables used to specify unspecified, strike-slip, normal-slip, and reverse-slip fault type, respectively, as given by the values in Table 3.1, and $\mathbf{M}_h$, the "hinge magnitude" for the shape of the magnitude scaling, is a coefficient to be set during the analysis.

The determination of the coefficients in the distance and magnitude functions is discussed in the section of the report following the next section on site amplification.

Table 3.1 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

3.2 SITE AMPLIFICATION

We did not solve for the site amplification terms in our analysis, but rather used a slight modification of the site amplification given by Choi and Stewart (2005). The equation and the coefficients of the equation are provided in this section. The site amplification equation is given by:

$$F_S = F_{LIN} + F_{NL}, \quad (3.6)$$

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

The linear term is given by:

$$F_{LIN} = b_{lin} \ln(V_{S30} / V_{ref}), \quad (3.7)$$

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); they are empirically based, but were not determined by the regression analysis in our study.

The nonlinear term is given by:

(a) $pga4nl \le a_1$:

$$F_{NL} = b_{nl} \ln(pga\_low / 0.1) \tag{3.8a}$$

(b) $a_1 < pga4nl \le a_2$:

$$F_{NL} = b_{nl} \ln(pga\_low / 0.1) + c[\ln(pga4nl / a_1)]^2 + d[\ln(pga4nl / a_1)]^3 \tag{3.8b}$$

(c) $a_2 < pga4nl$:

$$F_{NL} = b_{nl} \ln(pga4nl / 0.1) \tag{3.8c}$$

where a_1 (=0.03g) and a_2 (=0.09g) are assigned to 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 an initial estimate of the predicted PGA in g for $V_{ref} = 760$ m/s, as given by Equation (3.1) with F_S = 0 and ε = 0. The period-dependent and V_{S30}-dependent coefficients b_{nl}, *c,* and *d* are prescribed based on a slight modification of the empirical analysis results presented by Choi and Stewart (2005), where the modification was designed to smooth the predicted soil amplifications more effectively over amplitude and V_{S30}. As discussed below, the three equations for the nonlinear portion of the soil response (Eq. 3.8) are required for two reasons: (1) to prevent the nonlinear amplification from increasing indefinitely as *pga4nl* decreases and (2) to smooth the transition from amplification to no amplification.

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

(a) $V_{S30} \le V_1$:

$$b_{nl} = b_1 . \tag{3.9a}$$

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

$$b_{nl} = (b_1 - b_2) \ln(V_{S30} / V_2) / \ln(V_1 / V_2) + b_2 . \tag{3.9b}$$

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

$$b_{nl} = b_2 \ln(V_{S30} / V_{ref}) / \ln(V_2 / V_{ref}) . \quad (3.9c)$$

(d) $V_{ref} \le V_{S30}$:

$$b_{nl} = 0.0 . \quad (3.9d)$$

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}).

The coefficients *c* and *d* in Equation (3.8b) are given by:

$$c = (3\Delta y - b_{nl}\Delta x) / \Delta x^2 \quad (3.10)$$

and

$$d = -(2\Delta y - b_{nl}\Delta x) / \Delta x^3 , \quad (3.11)$$

where

$$\Delta x = \ln(a_2 / a_1) \quad (3.12)$$

and

$$\Delta y = b_{nl} \ln(a_2 / pga\_low) . \quad (3.13)$$

The coefficients needed to evaluate the site-response equations are listed in Tables 3.2 and 3.3. 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. It is important to emphasize that the site-response equations were prescribed, based on the work of Choi and Stewart (CS05), rather than determined by our regression. The reason for this is that we were concerned that the NGA database would be insufficient to simultaneously determine 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 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. However, note that our prescribed soil response terms 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.

Table 3.2 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

Table 3.3 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

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 3.1, they are similar to the linear soil coefficients derived by BJF97. For periods not provided by CS05, we

interpolated the b_{lin} values as shown in Figure 3.1. As periods get very large (>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 the log period shown in Figure 3.1 as the basis for choosing the values for the longer periods.

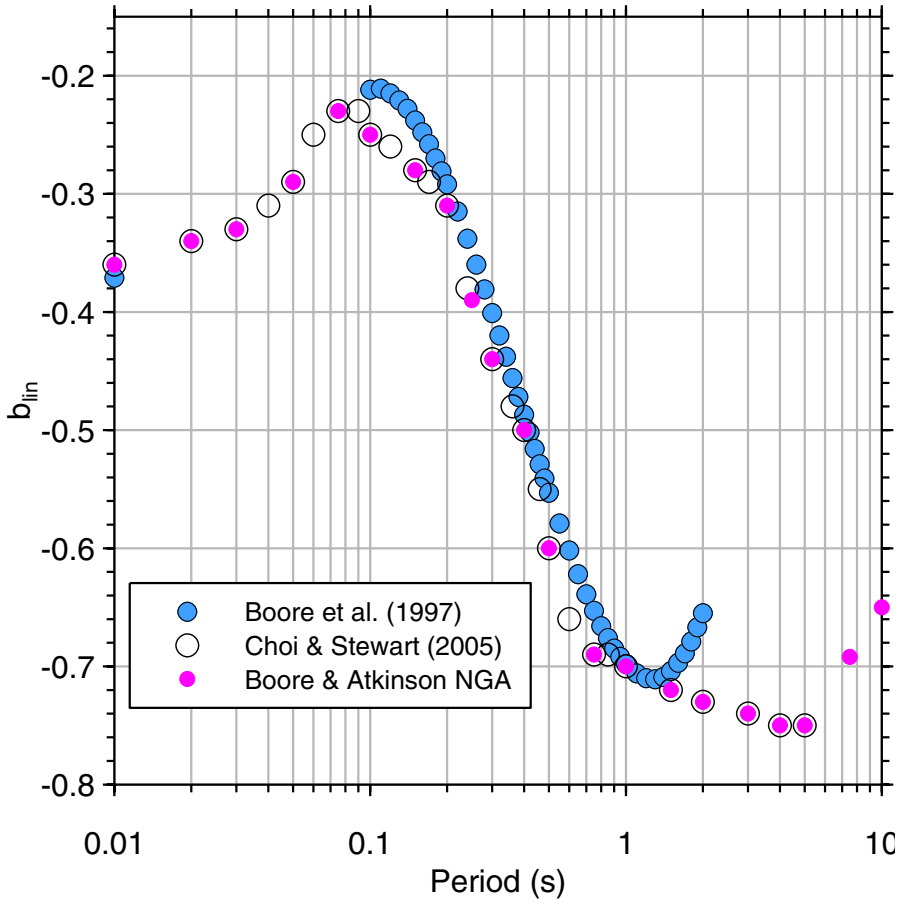


Fig. 3.1 Coefficient controlling linear amplification, as function of period. Values used in equations in this report indicated by magenta circles.

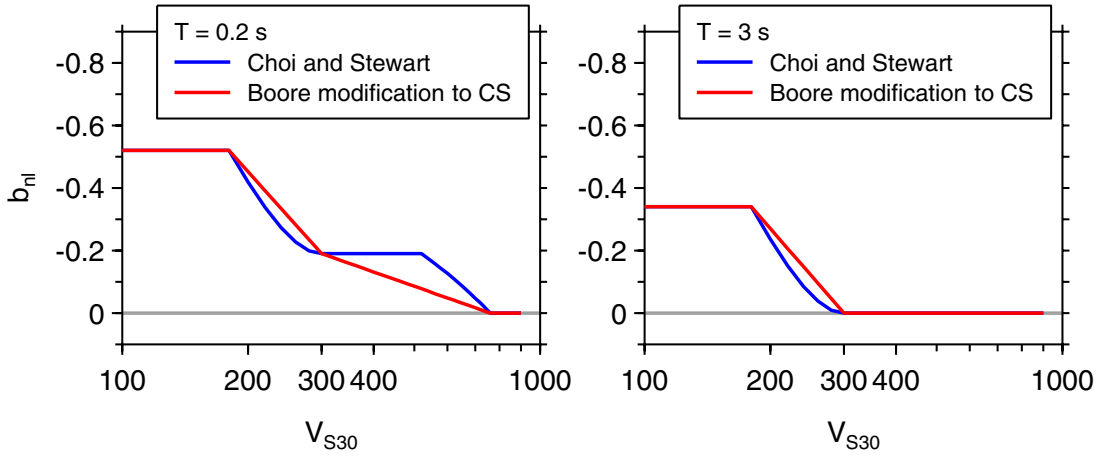


Fig. 3.2 Comparison of slope that controls nonlinear amplification function.

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 3.2. 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, as shown in Figures 3.3 and 3.4. To extend the value of b_1 to periods longer than 5 s, we fit two quadratic curves to their values: one for all of the values and another for values corresponding to periods greater than 0.2 s. As shown in Figure 3.3, the results were similar. 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. 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.

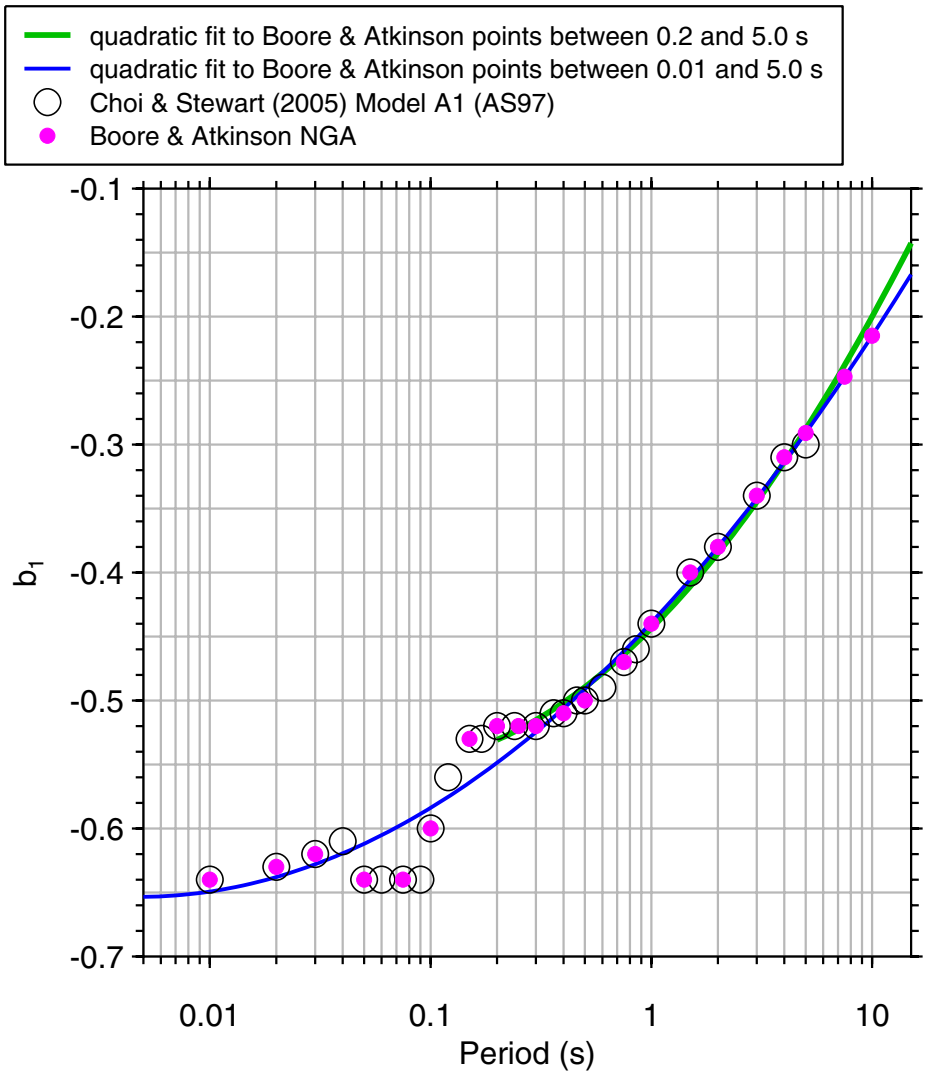


Fig. 3.3 Basis for choice of nonlinear amplification coefficient b_1. We used Choi and Stewart (2005) (CS05) when available, except for periods of 5, 7.5, and 10 s, for which we used values from quadratic fit to all of CS05's values.

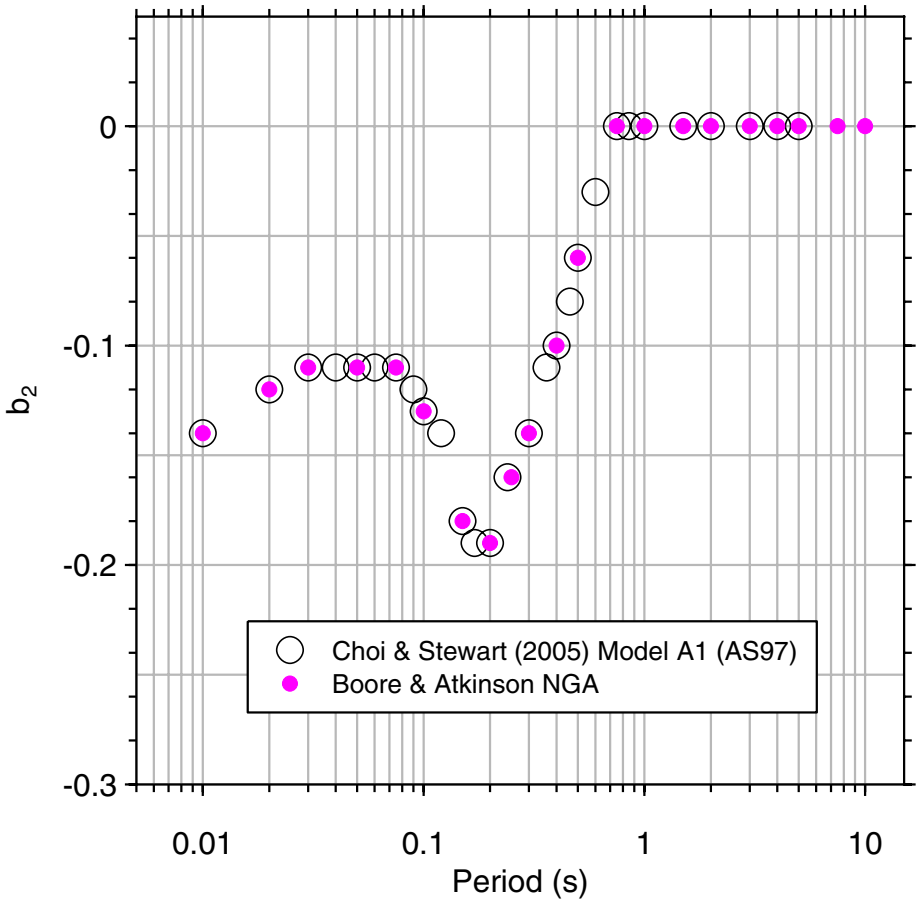


Fig. 3.4 Basis for choice of nonlinear amplification coefficient b_2. We used Choi and Stewart (2005) (CS05) when available, except for periods of 7.5 and 10 s, for which we used CS05's value for period of 5 s.

We point out a potential confusion in terminology: according to Equation (3.8c), $F_{NL} = 0.0$ when $pga4nl = 0.1$ g. Does this mean that there is no nonlinear amplification for this level of rock motion? No. The amplification for this value of *pga4nl* is given entirely by the F_{LIN} term because CS05 derived the "linear" amplifications (F_{LIN}) for motions with a mean PGA on rock 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 can be nonlinear for values of *pga4nl near* 0.1 g.

CS05 use only Equation (3.8c) 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 their comment (Choi and Stewart 2005, p. 24) that they consider Equation (3.8c) 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*. Simply terminating the nonlinear amplification at *pga*4*nl* = *pga_low* results in kinks in plots of ground motion vs. distance. For that reason we included a transition curve, as given in Equation (3.8b). A plot of the nonlinear amplification that shows the effect of this transition region is given in Figure 3.5, where the "smoothed" curve represents our formulation through Equations (3.8a, b, c).

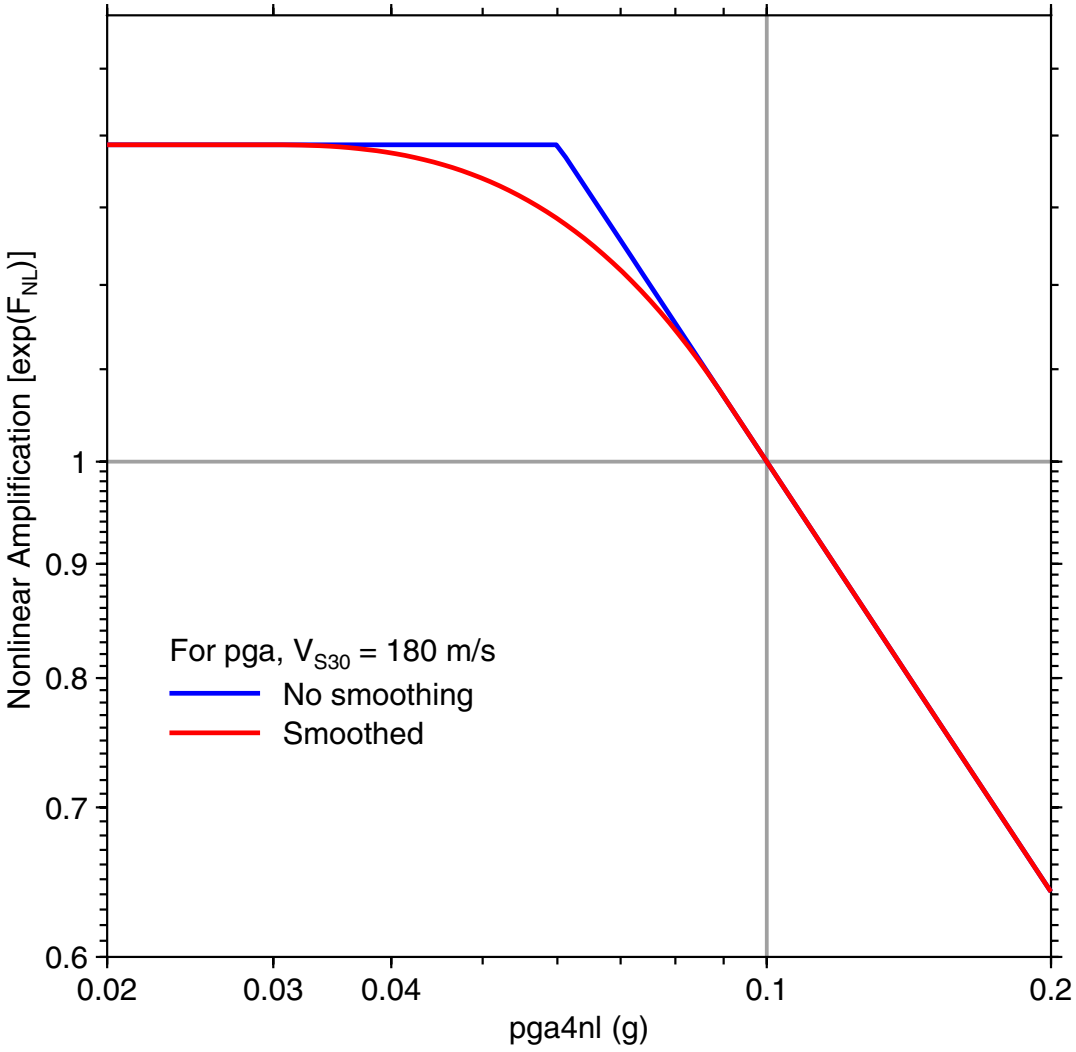


Fig. 3.5 Nonlinear contribution to site amplification showing how cubic polynomial gives smoothed version of amplification. Amplification is for $V_{S30} = 180$ m/s.

The total amplification for a short (0.2 s) and a long (3.0 s) period oscillator is shown in Figure 3.6 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 can still affect ground motions, but the large "linear" amplification is not outweighed by the nonlinear effect for large values of *pga4nl*. For periods longer than 0.75 s (see Table 3.2) there is no nonlinear contribution to the amplification for $V_{S30} > 300$ m/s.

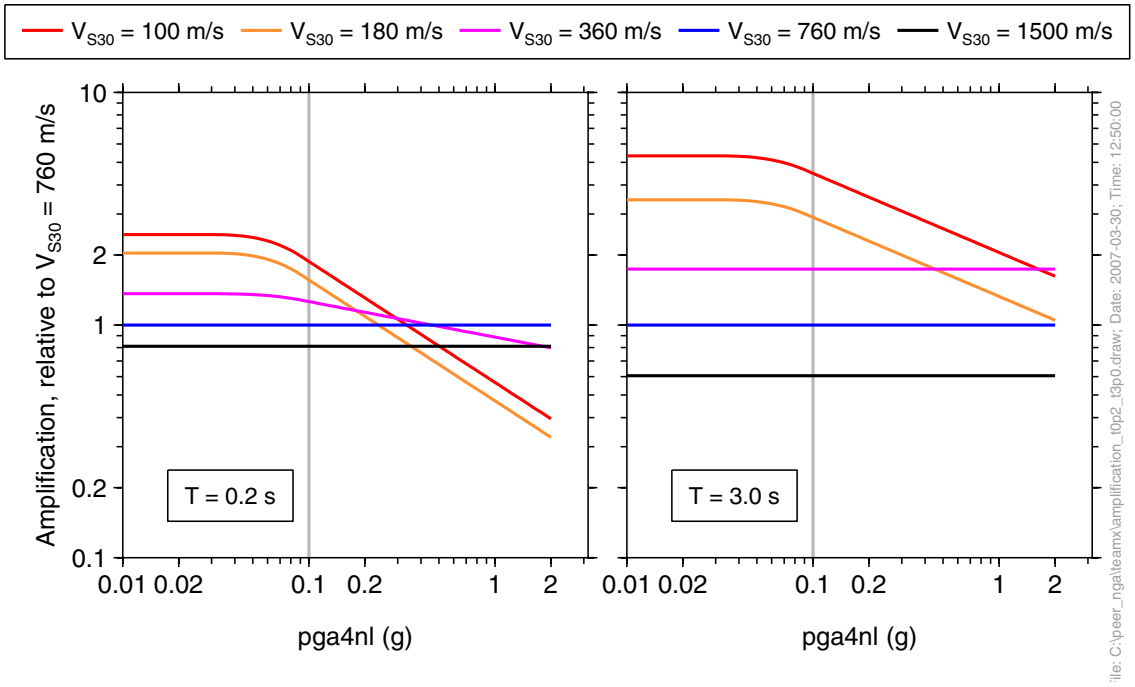


Fig. 3.6 Combined amplification for $T = 0.2s$ and $T = 3.0s$ as function of *pga4nl*, for suite of V_{S30}. Note at short periods (left graph), purely linear amplification does not occur on soft soils until $pga4nl < 0.03$ g.

It should be noted that the empirical studies on which the soil amplification functions were based contained very few data for hard sites, with V_{S30} >1000 m/s. The amplification functions are probably reasonable for values of V_{S30} up to about 1300 m/s, but should not be applied for very hard rock sites ($V_{S30} \geq 1500$ m/s).

4 Determination of Coefficients

4.1 METHODOLOGY

The selected response variables in the NGA database were regressed to Equation (3.1) to determine F_D and F_M, after first correcting all observations to the reference velocity of 760 m/s, using Equations (3.6)–(3.13). This allows us to perform the regression analyses for the case where $F_S = 0$. 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 inter-event aleatory variability, σ), and the second stage determines the magnitude dependence (and the intra-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).

4.1.1 Stage 1: Distance Dependence

The distance dependence 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 (3.6)–(3.13) from $\ln Y_{observed}$. The corrected response variables for our selected subset of the NGA dataset (using the exclusion criteria of Table 2.1, with distances out to 400 km) are regressed against distance using Equation (4.1), which is the same as Equation (3.2) but with dummy variables ($c_0(event)$) added to represent the event term for each earthquake (the event term is thus the average motion of the observations for each earthquake corrected to $R = R_{ref}$).

$$F_D(R_{JB}, \mathbf{M}) = c_0(event) + [c_1 + c_2(\mathbf{M} - \mathbf{M}_{ref})]\ln(R / R_{ref}) + c_3(R - R_{ref}) \quad (4.1)$$

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

$$(c_0)_1\delta_1 + (c_0)_2\delta_2 + \cdots + (c_0)_{NE}\delta_{NE}, \tag{4.2}$$

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), 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 80–100 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 dataset in order to determine the remaining coefficients.

Determination of c3 (anelastic term): The data used to determine c_3 include 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 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 4.1 (see also Appendices M and N).

Table 4.1 Comparisons of numbers of stations in NGA flatfile and in extended dataset 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

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 had no site class information. 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 are fixing a single straight-line slope (c_1) and then determining the curvature, c_3, required to match the more rapid decay of the data at greater distances (c_3 must be less than 0) and the near-source effective depth coefficient, h, 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 also determined for each of the four earthquakes (these are the coefficients of the dummy variables for each event). Figure 4.1 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 also found the best values of c_3 and h to fit the distance functions determined in southern California by Raoof et al. (1999). The equivalent values from the Raoof et al. (1999) analysis were similar to those from our analysis of the four-event California subset described above. 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 dataset 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). To assign values of c_3 for all periods for which the NGA equations were to be determined, we fit a quadratic to the c_3 values from the analysis of our four-event data subset. This is shown in Figure 4.2. 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>5 s).

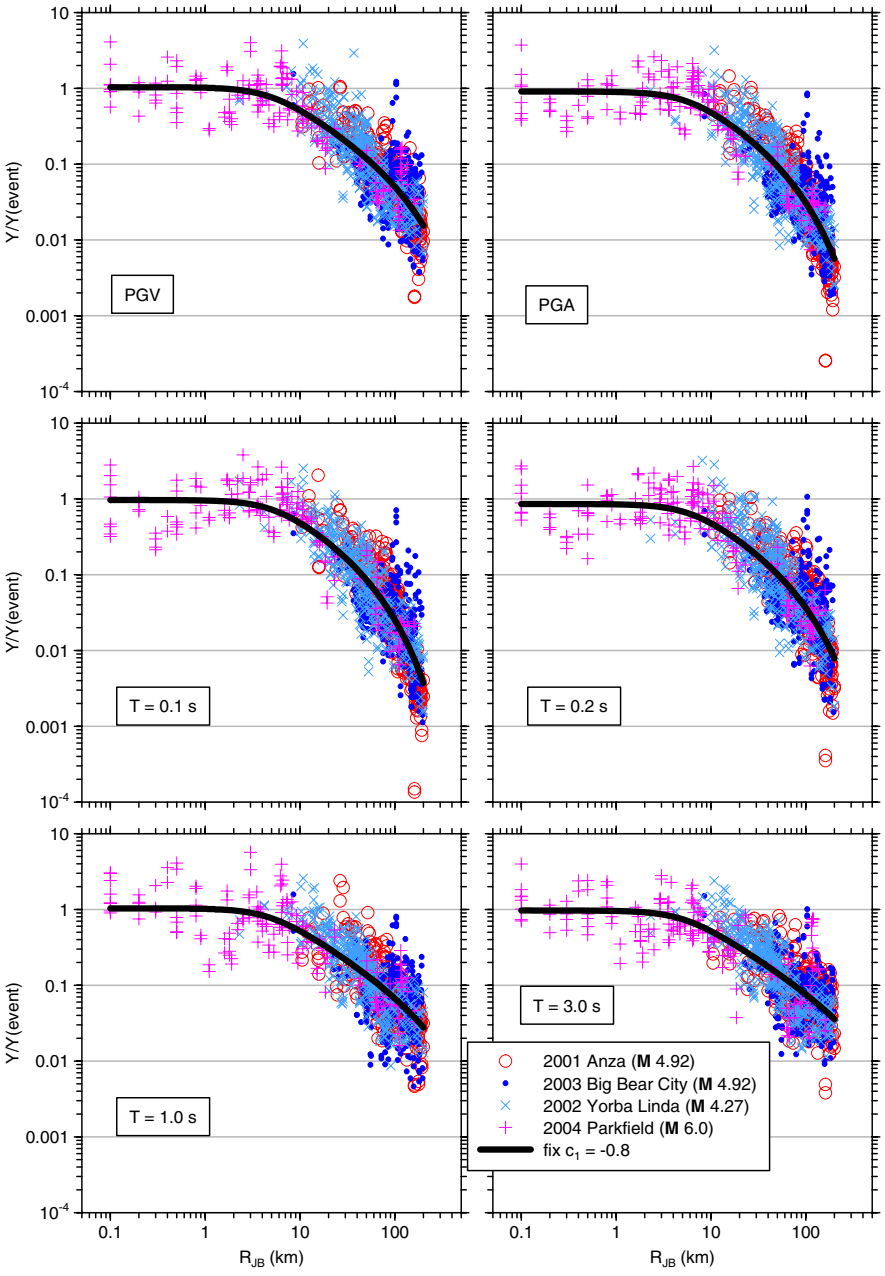


Fig. 4.1 Normalized ground motions for four events, using extended dataset (more data than in NGA flatfile). Black curve is regression fit obtained with constraints $c_1 = -0.8$ and $c_2 = 0.0$.

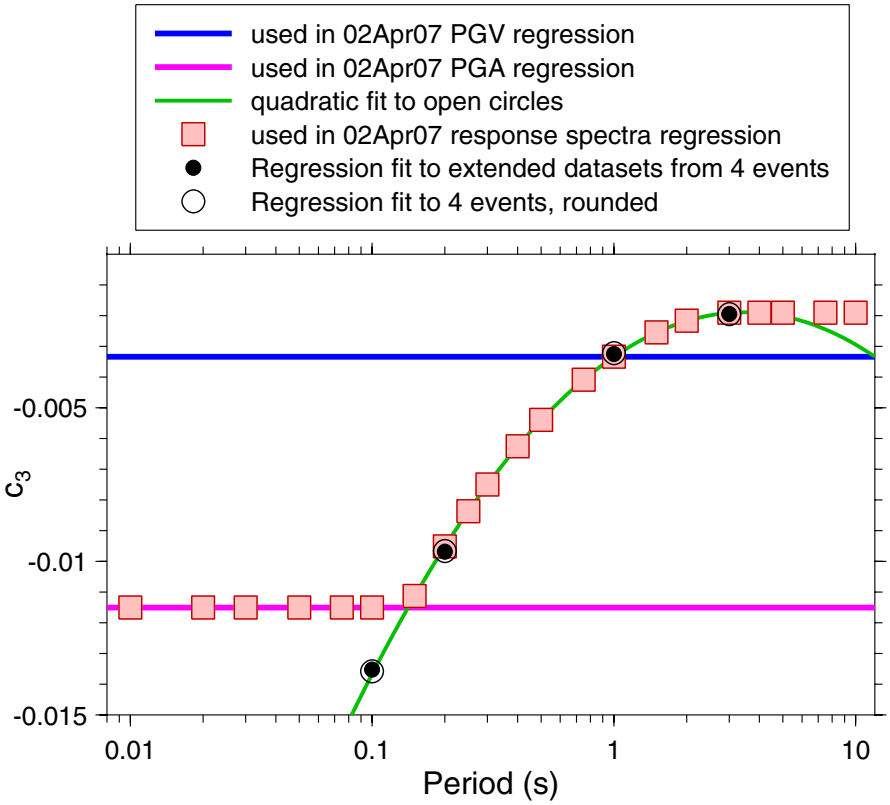


Fig. 4.2 Basis for choice of c_3 at periods other than analyzed in regression in extended dataset from four earthquakes (see text).

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 magnitude scaling that we observed between PGV and PSA at 3 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 obtained 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 the values in use at the time. Our procedure is shown in Figure 4.3, along with the final values of h (squares). The black dots were determined when h was a free parameter. We fit the values with a quadratic (dashed green line), but observed that the h value at 0.05 s 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 (solid green line) 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.

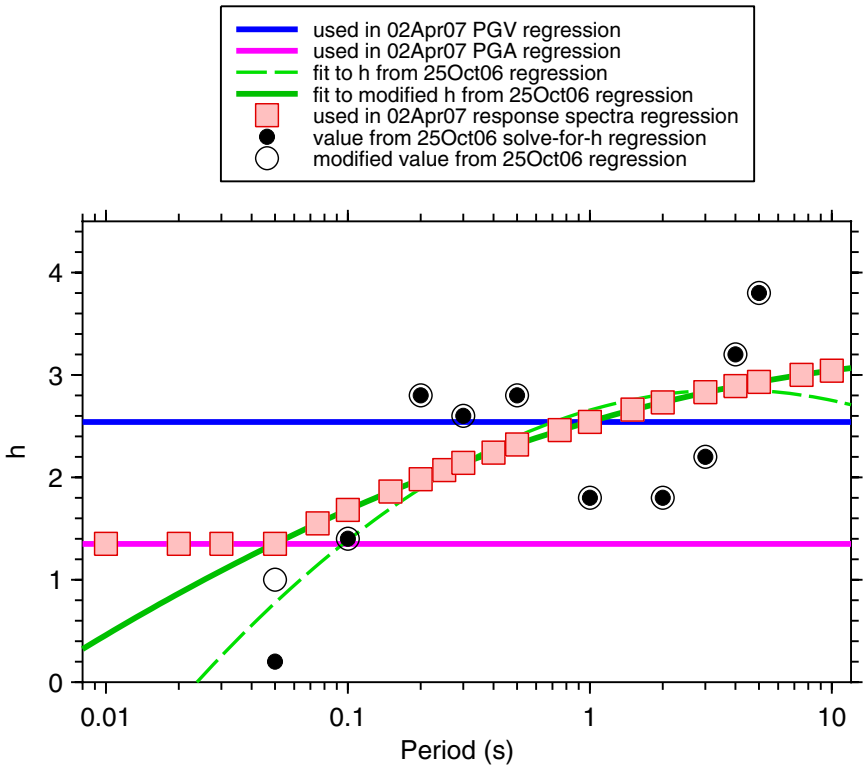


Fig. 4.3 Basis for choice of h at all periods (see text).

These pre-analyses establish 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 (Eq. 3.3), along with the event terms (c_0) for each earthquake, using all data (subject to the exclusions of Table 2.1) for distances less than 400 km (we originally included a magnitude dependence to the anelastic term but found that the

resulting ground motion predictions tended to increase with increasing distance). The c_1 coefficient is the effective geometric spreading rate (slope) for an event of **M**= $\mathbf{M}_{\text{ref}}$, 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 $\mathbf{M}_{\text{ref}}$). 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, $\boldsymbol{M}_{ref}$, 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 $\boldsymbol{M}_{ref}$, 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 overlap of curves at distances less than the pivot distance. This was avoided this by choosing $R_{ref} = 1\,\text{km}$, 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 $R_{JB} = 0\,\text{km}$).

4.1.2 Stage 2: Magnitude Dependence

The event terms (coefficients $(c_0)_j$ in Eq. (4.1)) 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. Only events with more than one observation were used in the regression. The basic form we selected for the magnitude scaling is a quadratic, similar to the form used by BJF93. However, we imposed a constraint that the quadratic not reach its maximum at **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 **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 $\mathbf{M} \le \mathbf{M}_h$ 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 of less than or equal to 1.0 s required the constraint of zero slope; this is saying that for short periods the data actually indicated 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 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. On this basis, we initially set $\mathbf{M}_h = 7.0$, but there are a few periods for which $\mathbf{M}_{\max}$ was less than 7.0. Consequently the use of $\mathbf{M}_{\max}$ =7 would lead to a slight decrease of magnitude scaling between $\mathbf{M}_{\max}$ and $\mathbf{M}_h$, which is contrary to our requirement that our equations give no oversaturation of ground motion for sites with purely linear amplification. The adoption of $\mathbf{M}_h$ slightly lower than the minimum value of $\mathbf{M}_{\max}$ over all periods prevented this problem.

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. 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 (3.5), 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 subscript "U" and "M" distinguish between unspecified and specified fault type, respectively, in the table of aleatory uncertainties.

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.

4.2 RESULTS

4.2.1 Coefficients of Equations

The coefficients for the GMPEs are given in Tables 3.2–3.3 and 4.2–4.5. 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 coefficients of the equation for *pga*4*nl* are given in the first row in Tables 4.2 and 4.4 (with $\mathbf{M}_{ref}$ and R_{ref} as given in Table 4.3); there is no site amplification for *pga*4*nl* (it applies to sites with the reference velocity of 760 m/s). The coefficients for *pga*4*nl* were developed as an initial estimate early in the project, using only data for which $R_{JB} \le 80$ km and $V_{S30} > 360$ m/s, and therefore the predictions of PGA from this equation will differ somewhat from the predictions of PGA for $V_{S30} = 760$ m/s using the coefficients in the third row of the tables. The equation for *pga*4*nl* need only give approximately-correct values for the peak acceleration on rock-like sites. The equation provided for *pga*4*nl* maintains internal consistency; the site amplifications were used to reduce the observations to a reference velocity before doing the regressions, and thus the same site amplifications should be used when predicting ground motions using the results of the regressions.

Table 4.2 Period-dependent distance-scaling coefficients.

period	c_1	c_2	c_3	h
pga4nl	-0.55000	0.00000	-0.01151	3.00
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

Table 4.3 Period-independent distance scaling coefficients.

Coefficient	Value
$\mathbf{M}_{ref}$	4.5
R_{ref}	1.0

Table 4.4 Magnitude-scaling coefficients.

Period	e_1	e_2	e_3	e_4	e_5	e_6	e_7	$\mathbf{M}_h$
pga4nl	-0.03279	-0.03279	-0.03279	-0.03279	0.29795	-0.20341	0.00000	7.00
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	0.00000	-2.14635	0.40387	-0.48492	0.00000	8.50

Table 4.5 Aleatory uncertainties.

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

4.2.2 Discussion of "Geometrical Spreading" Coefficients

Figure 4.4 plots the "geometrical spreading terms" c_1 and c_2; Figure 4.5 shows the effective "geometrical spreading" factor obtained by the combination $c_1 + c_2(\mathbf{M} - \mathbf{M}_{ref})$. The coefficients have more variation with period than seems reasonable, but to some extent this may be a result of forcing very smooth period dependence for the other distance-related coefficients. Note that the signs of the c_1 and c_2 coefficients differ for periods of less than 4 s. As a result, the effective geometrical spreading factor decreases with magnitude for periods of less than about 4 s, as shown in Figure 4.5. This is expected from simulation studies. On the other hand, the magnitude coefficient c_2 becomes negative for periods greater than 4 s, leading to a decrease of the geometrical spreading factor for smaller earthquakes. This is not expected. The reason for this

apparently anomalous behavior might be the relative sparseness of data for small magnitudes at long periods, in combination with the bias discussed earlier due to data censoring. For these reasons, the predictions of PSA at period above about 4 or 5 s and magnitudes less than about 5.6 should be treated with some skepticism. Judging from the plot in Figure 4.4, the coefficient controlling the magnitude dependence of the geometrical spreading (c_2) would be 0.0 at about T= 4.6 s, thus explaining the convergence of the curves at that period.

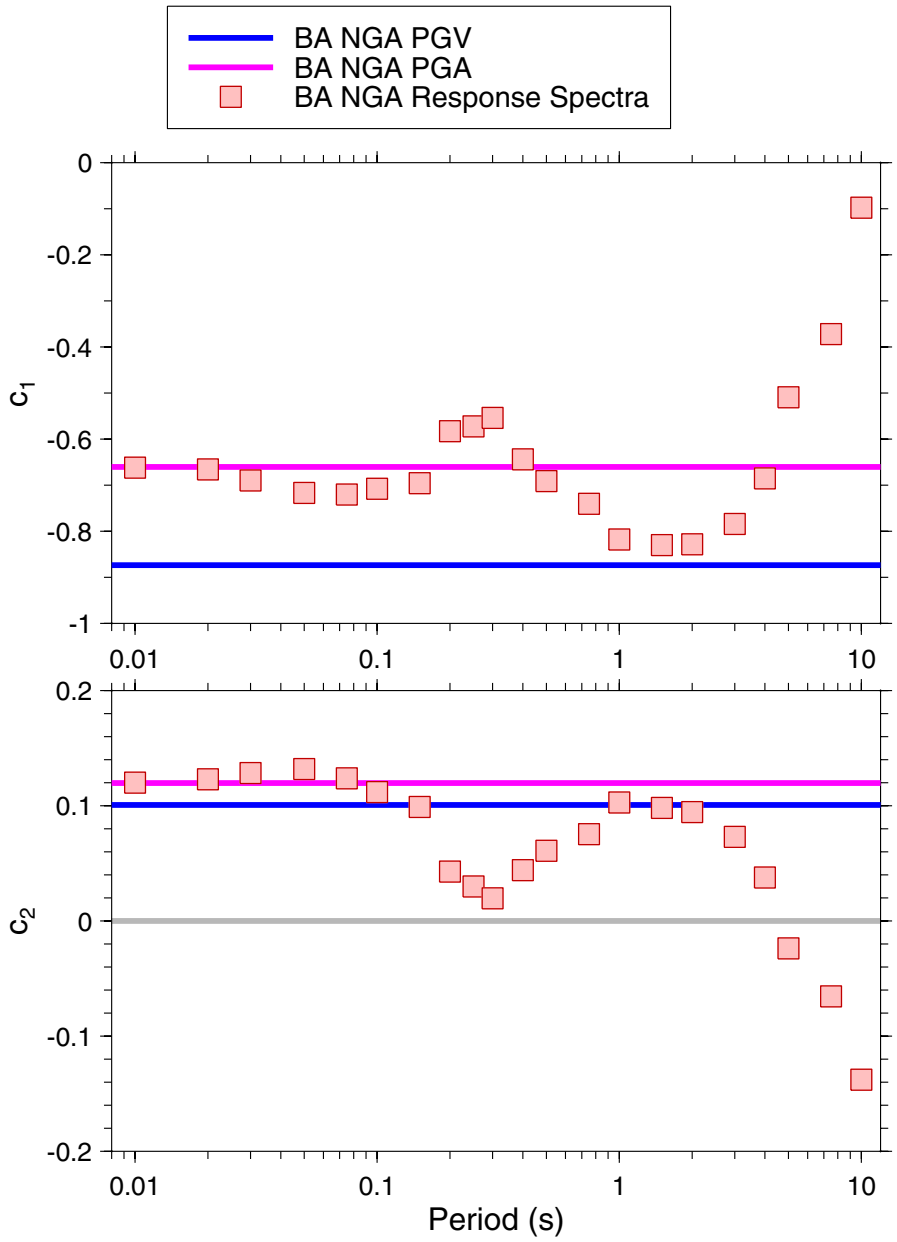


Fig. 4.4 Coefficients c_1 and c_2 from regression analysis. Horizontal lines show values for PGA and PGV (see legend).

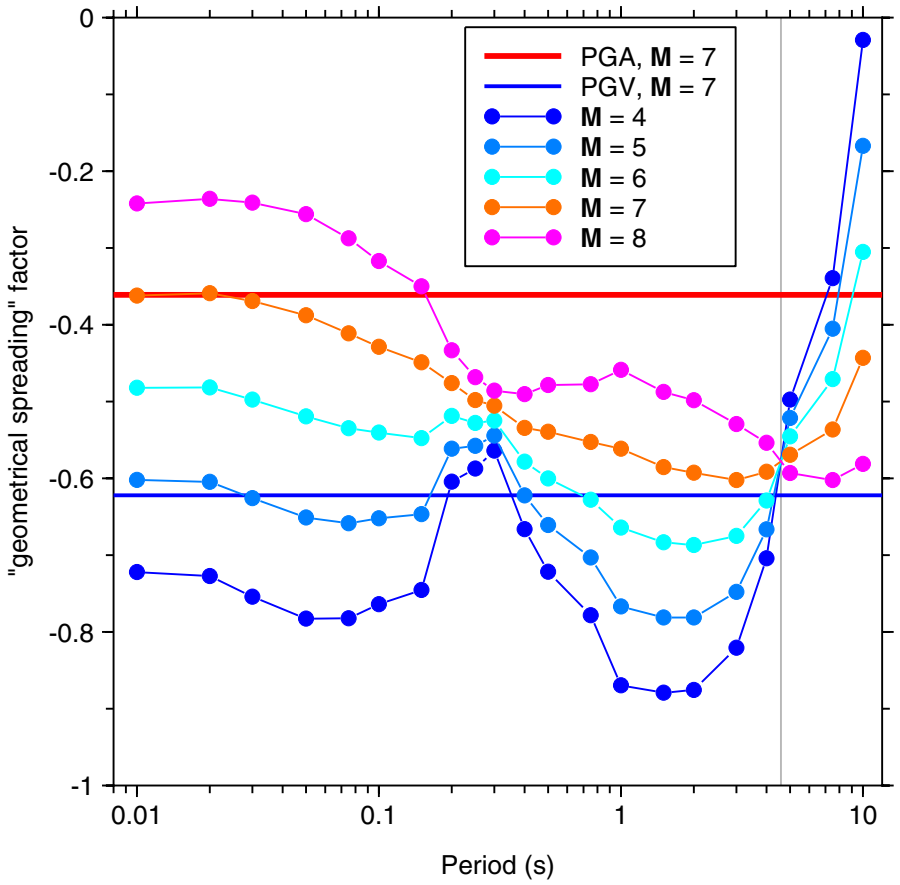
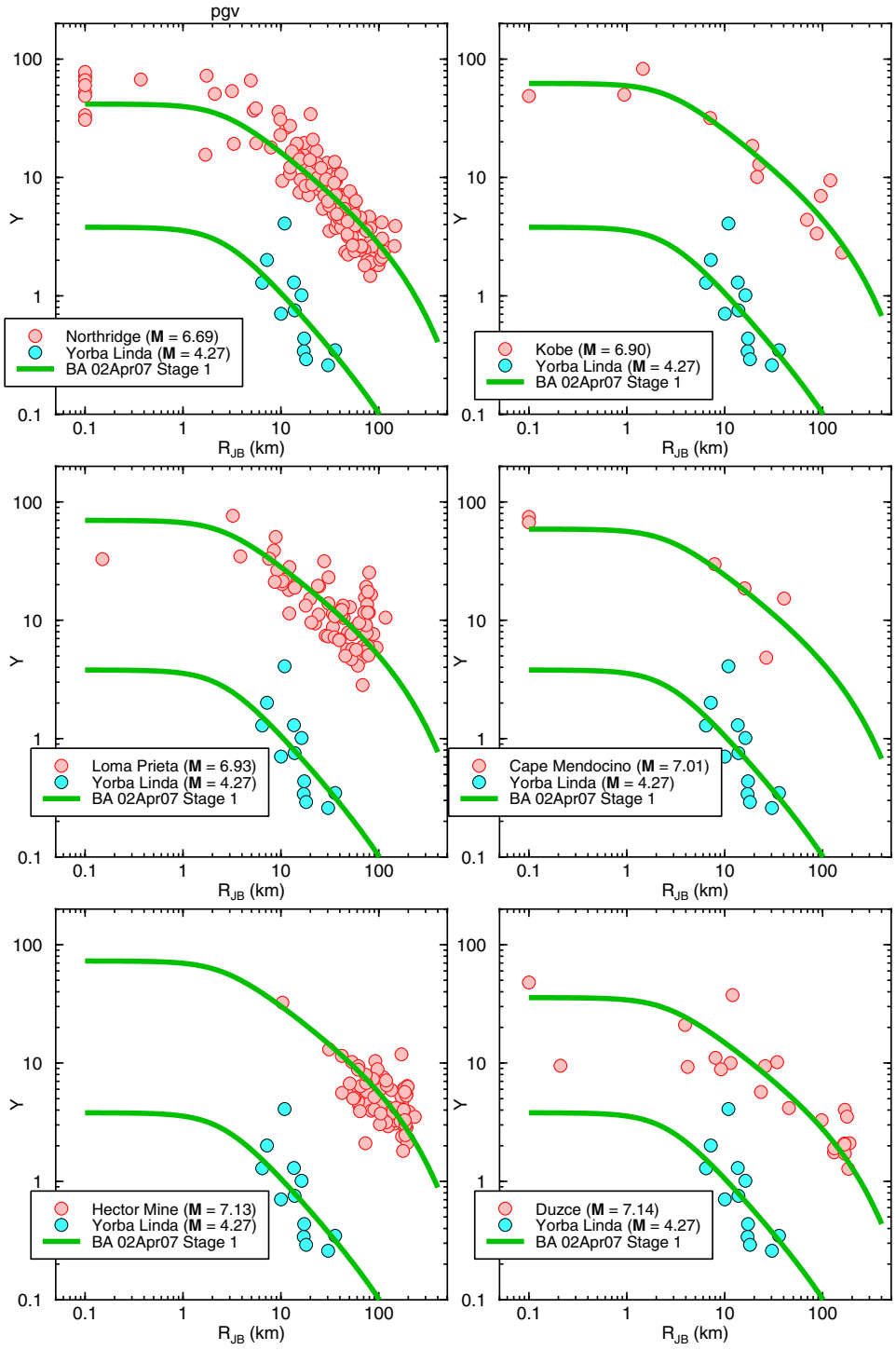


Fig. 4.5 Effective geometrical spreading coefficient, given by $c_1 + c_2(\mathbf{M} - \mathbf{M}_{ref})$.

4.2.3 Fit of Stage 1 Regressions

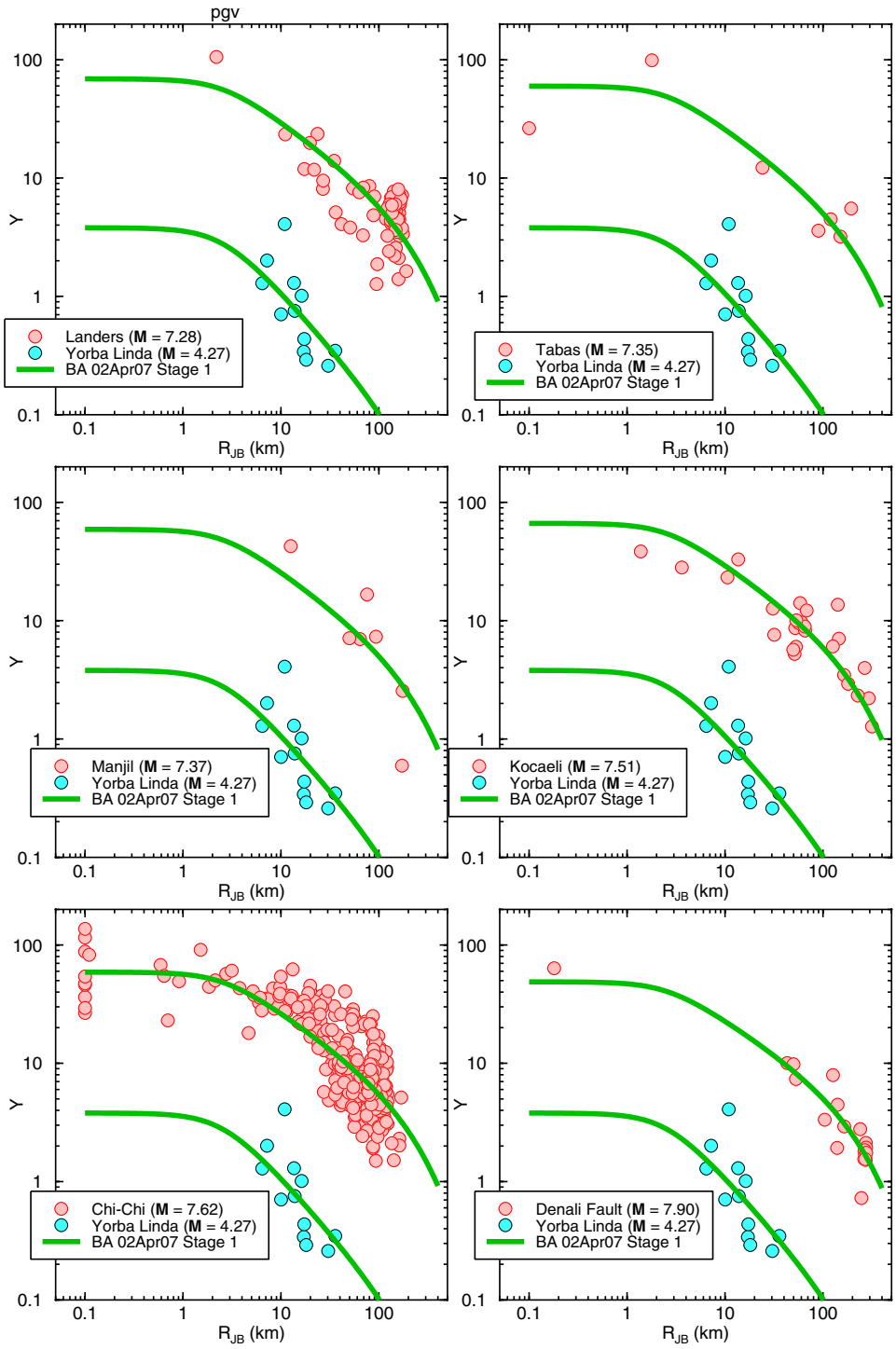
Figures 4.6–4.11 are a series of graphs showing the observations in comparison to the Stage 1 regression predictions. These figures provide a visual test of the ability of our functional form to represent the distance dependence of the response variables. The Yorba Linda data points are from the smallest earthquake in the NGA flatfile and are included on each figure to provide a basis for judging the magnitude scaling (note that data for that earthquake is missing for $T = 3$ s because the maximum useable period is 2.6 s). In all plots, records with R_{JB} less than 0.1 km have been plotted at 0.1 km. The curves are from the regression fits and include the event terms found for the specific regression—they are intended to help assess the Stage 1 regression and do not include event-to-event variability (see the plots of event terms for the Stage 2 regressions, shown later, for this variability). The figures contain data from all but one (St. Elias) of the **M** 7

earthquakes, in addition to the 1994 Northridge, 1995 Kobe, and 1989 Loma Prieta earthquakes. Overall, the agreement between our simple functional form and the observations appears to be quite good.



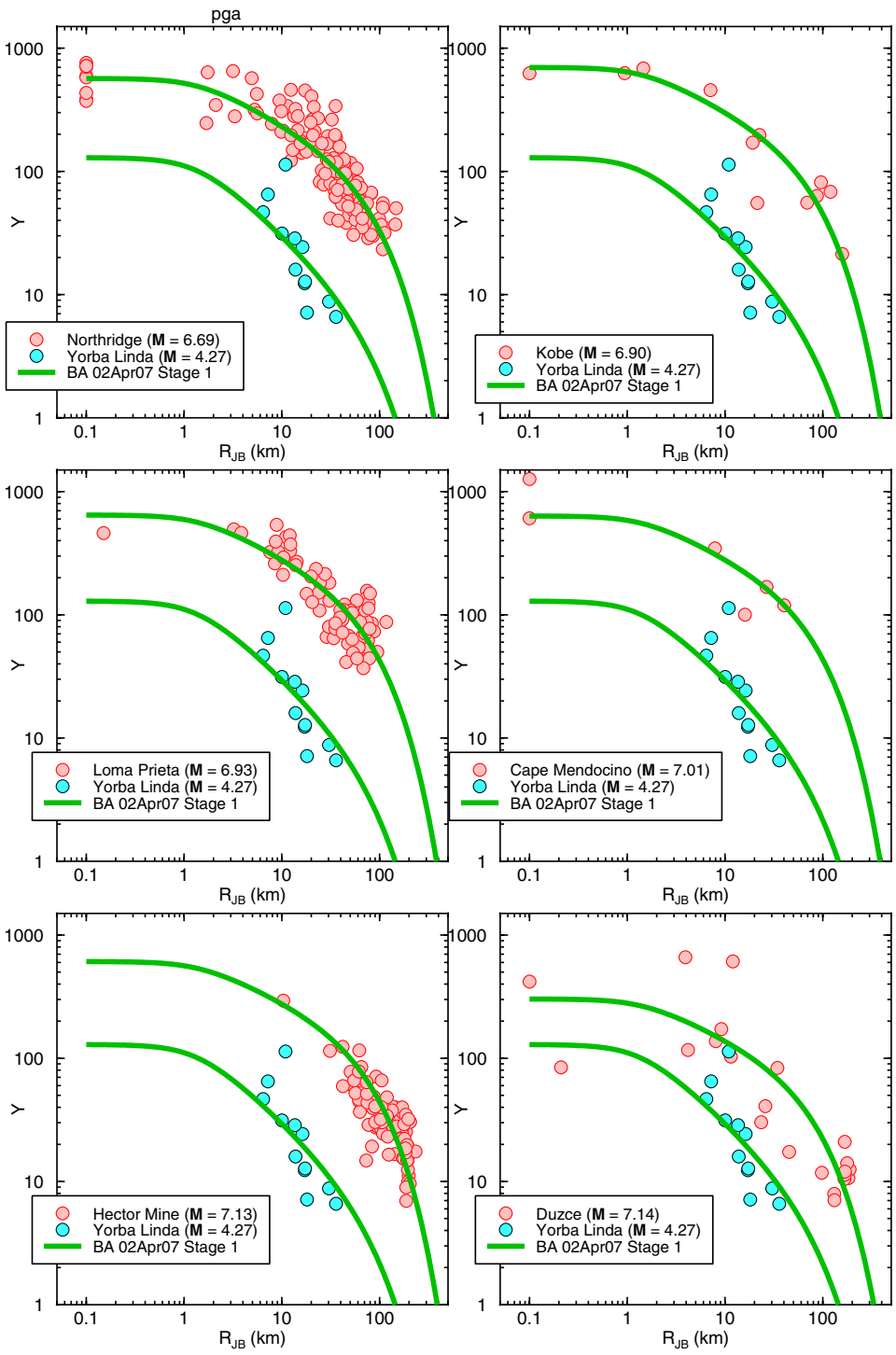
(a)

Fig. 4.6 Symbols: PGV observations, corrected to $V_{S30} = 760$ m/s, as function of distance for indicated events; Curves: Stage 1 regression fits.



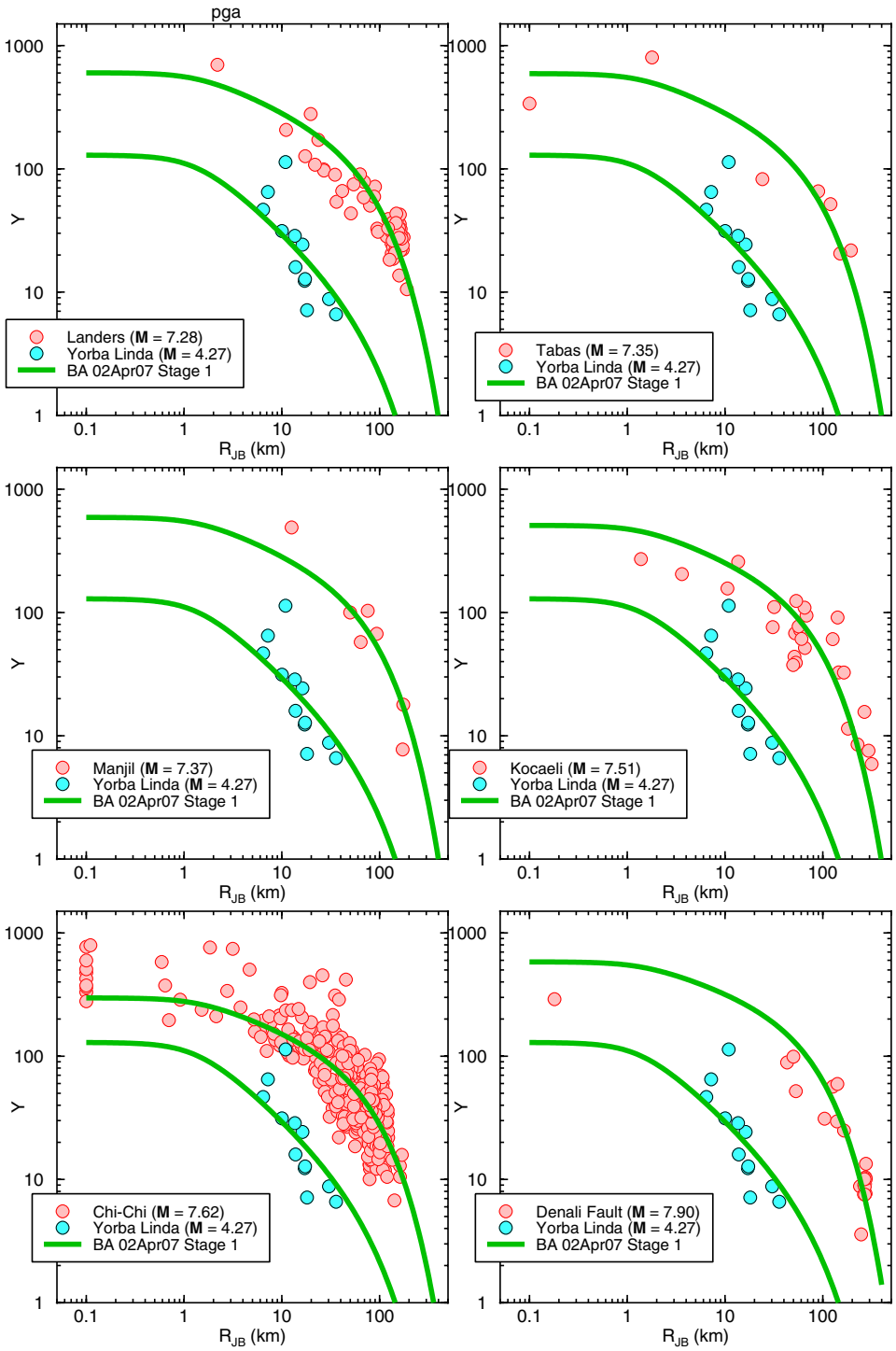
(b)

Fig. 4.6 —*Continued*



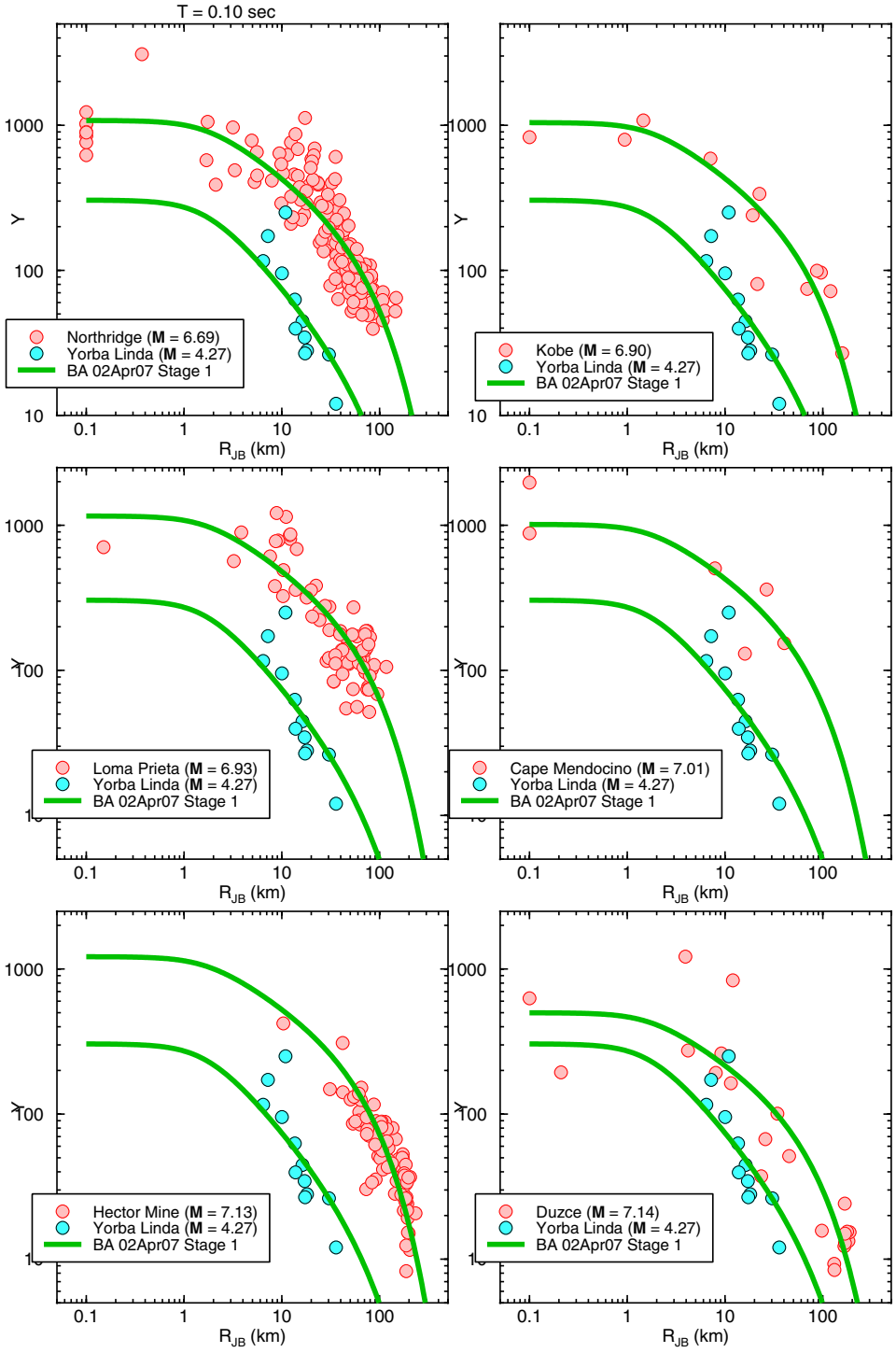
(a)

Fig. 4.7 Symbols: PGA observations, corrected to $V_{S30} = 760$ m/s, as function of distance for indicated events; Curves: Stage 1 regression fits.



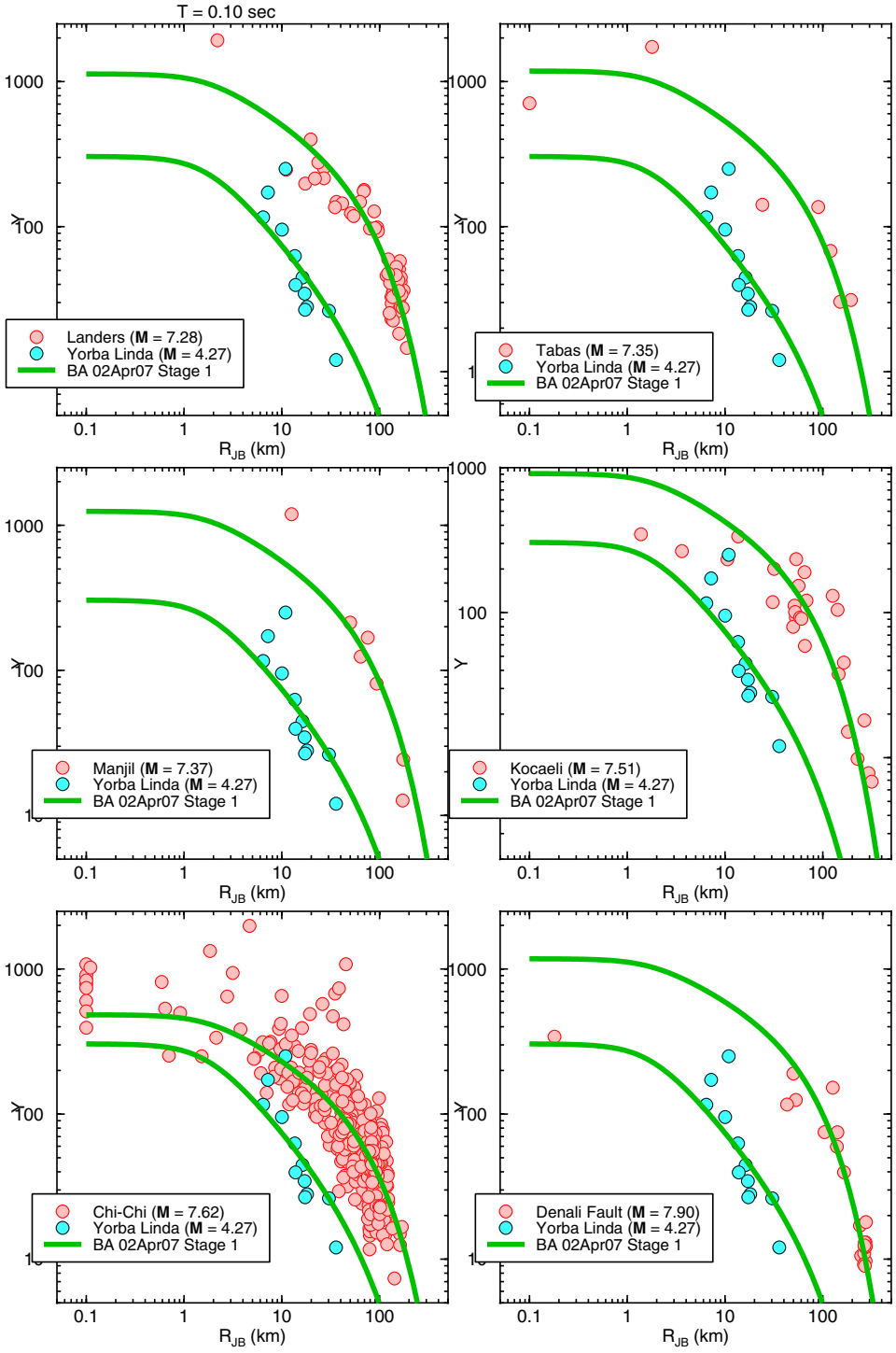
(b)

Fig. 4.7—*Continued*



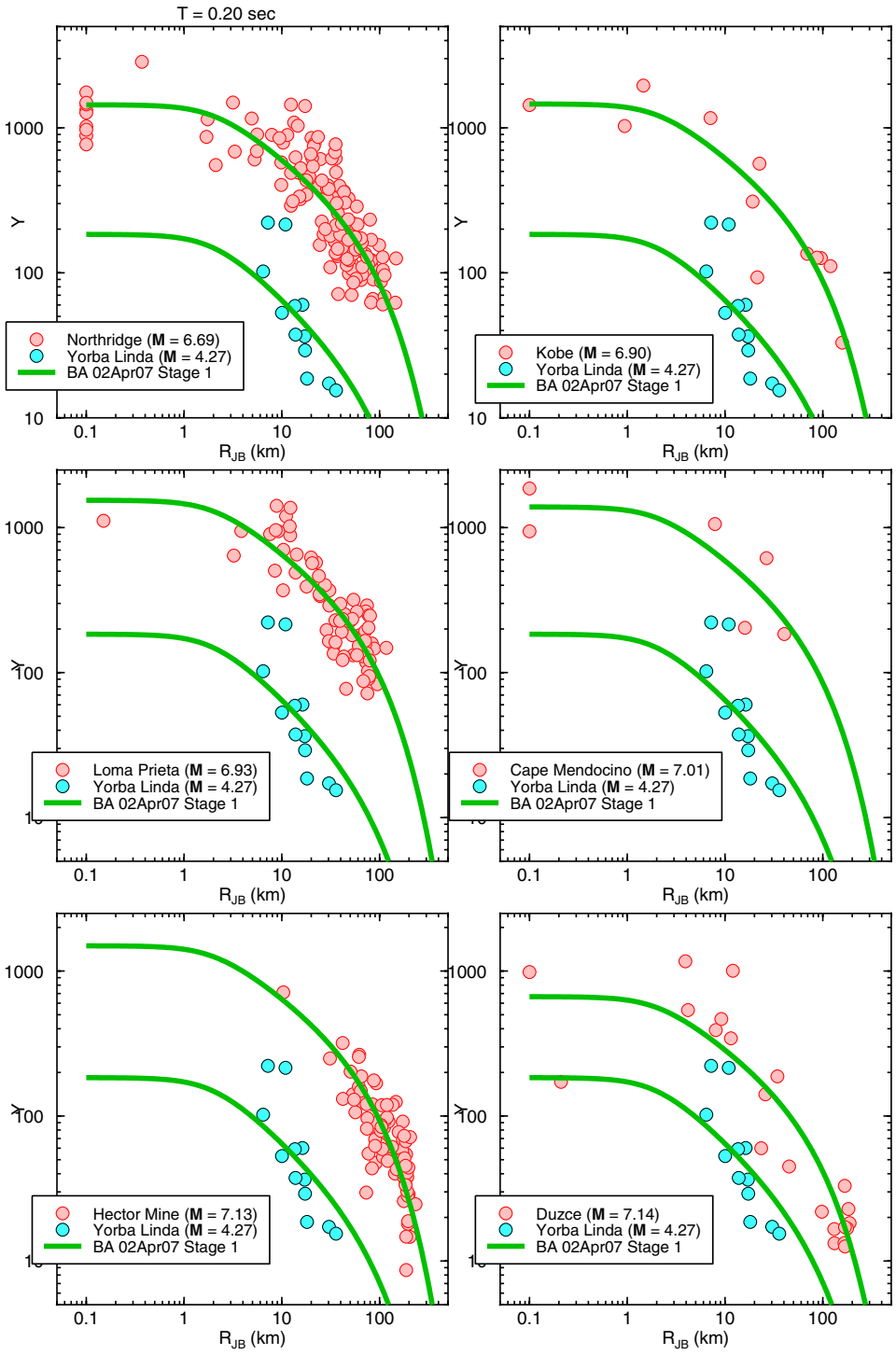
(a)

Fig. 4.8 Symbols: 0.1 s PSA observations, corrected to $V_{S30} = 760$ m/s, as function of distance for indicated events; Curves: Stage 1 regression fits.



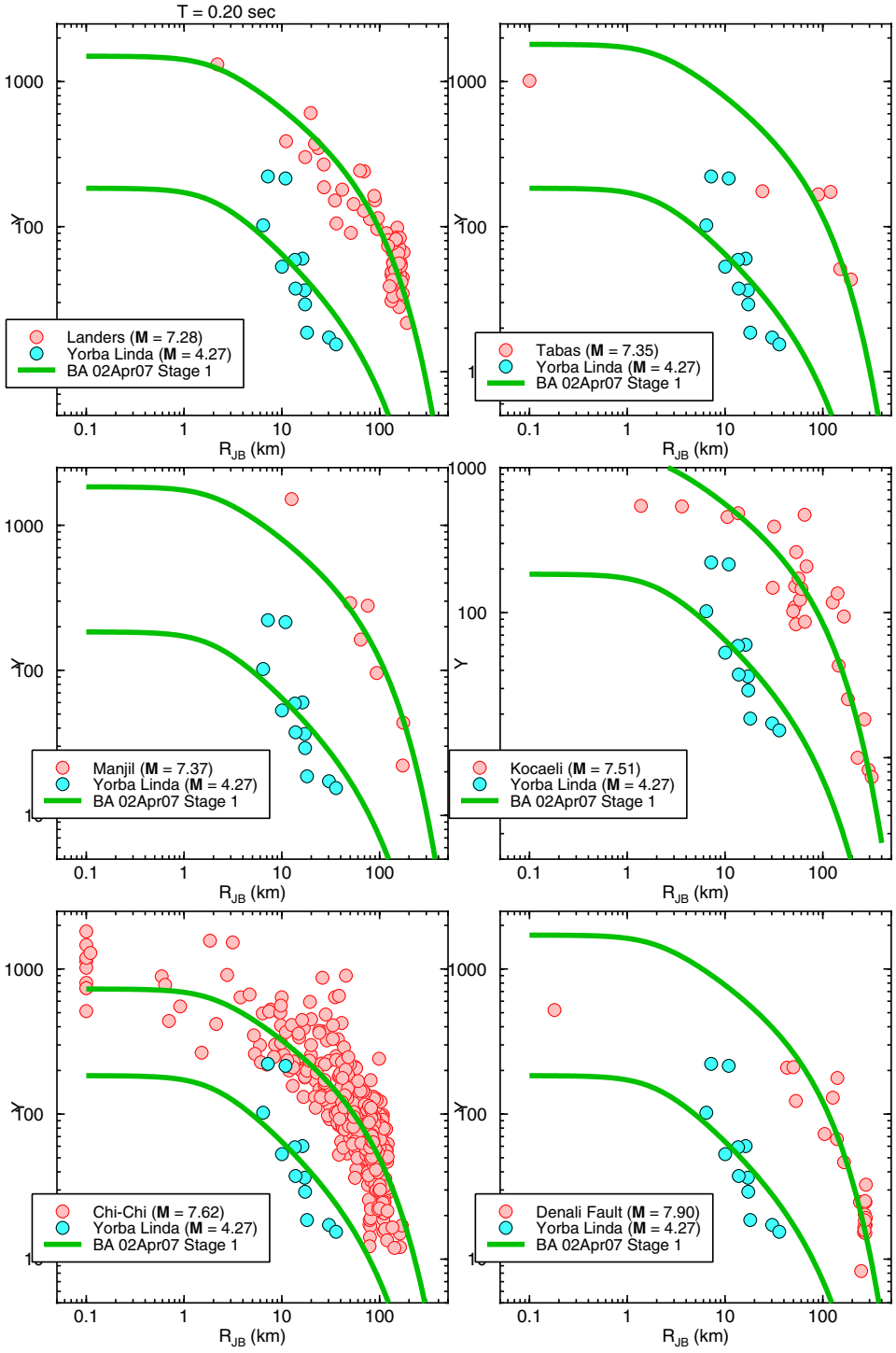
(b)

Fig. 4.8—*Continued*



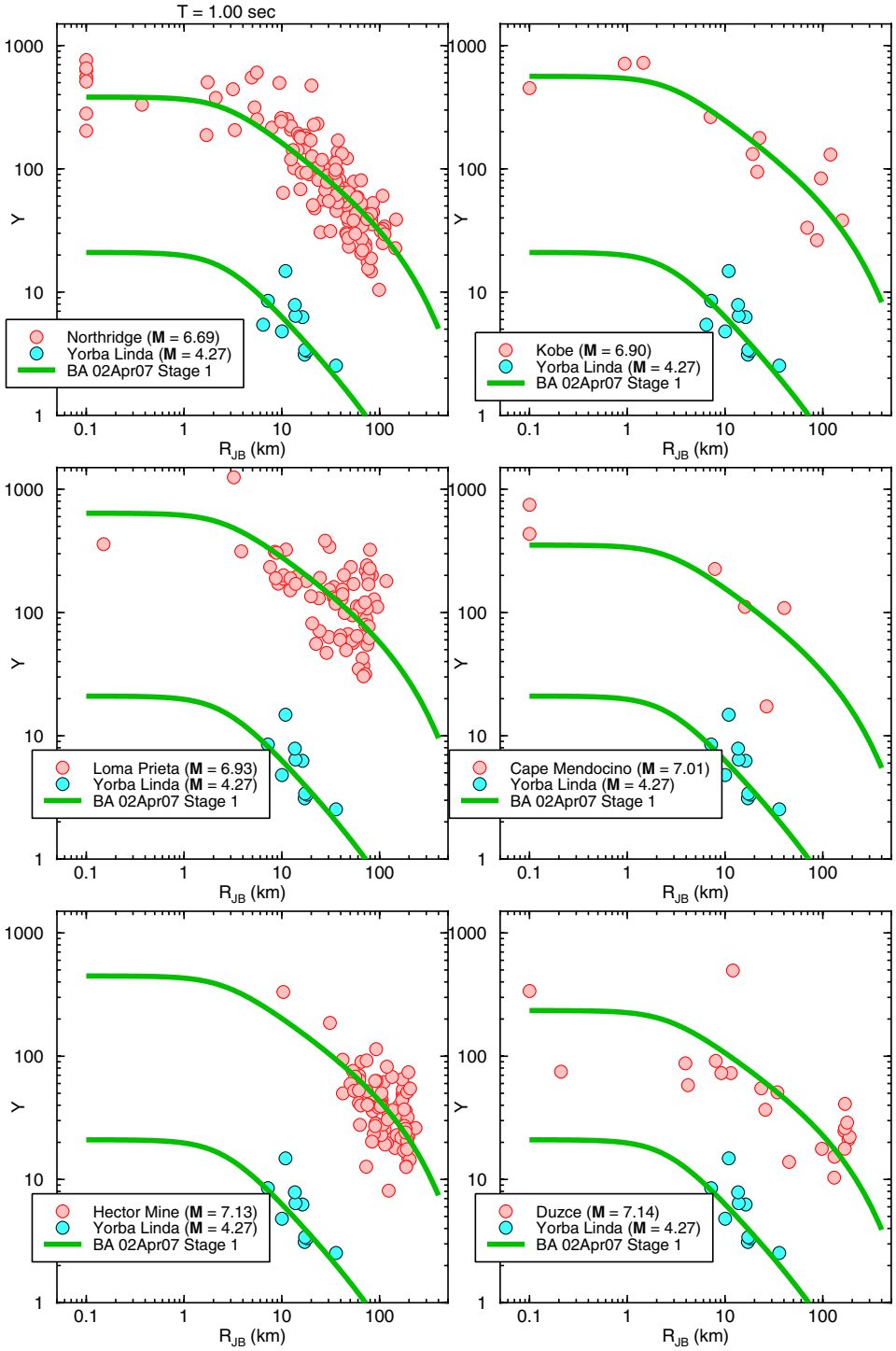
(a)

Fig. 4.9 Symbols: 0.2 s PSA observations, corrected to $V_{S30} = 760$ m/s, as function of distance for indicated events; Curves: Stage 1 regression fits.



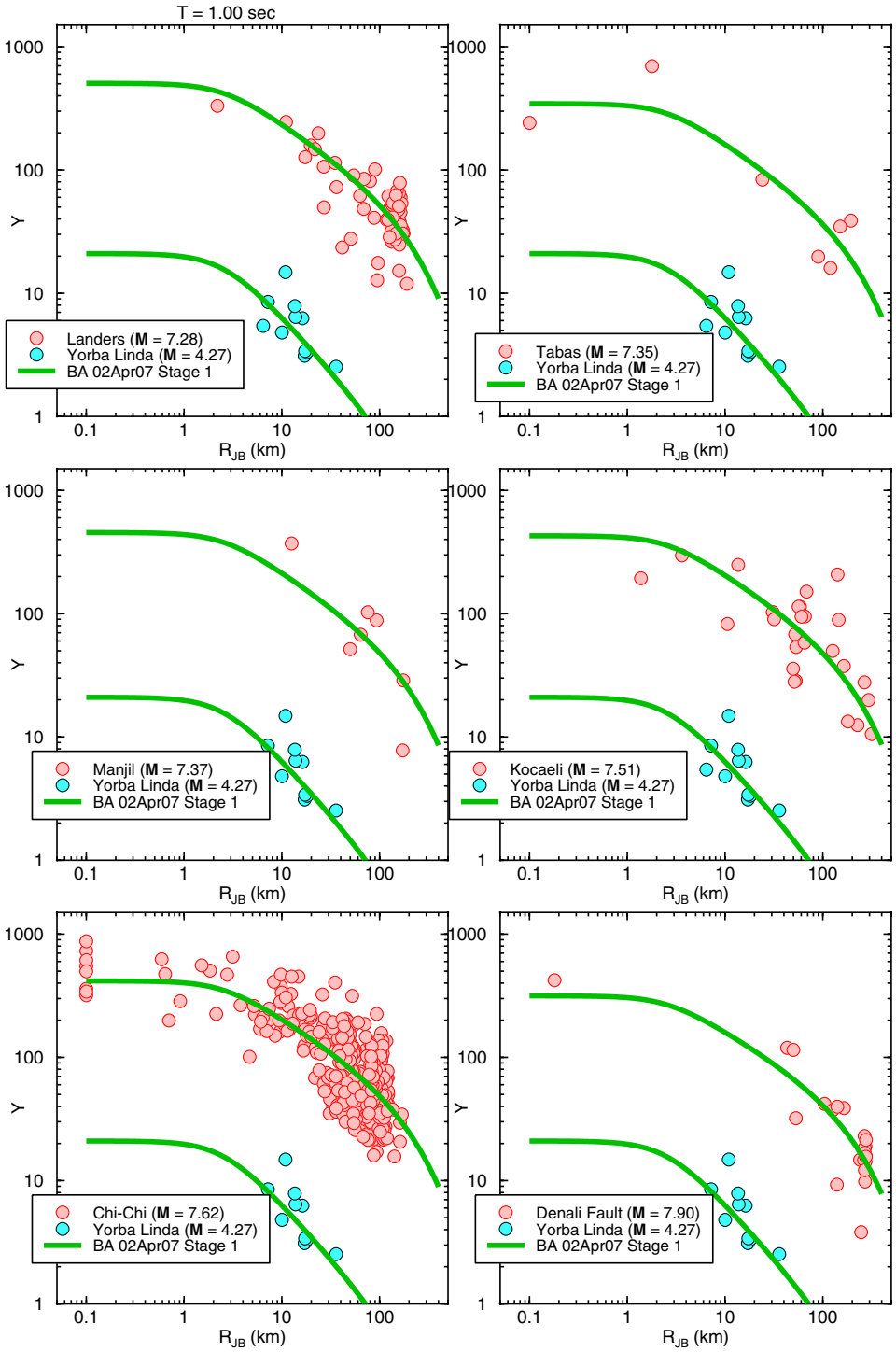
(b)

Fig. 4.9—*Continued*



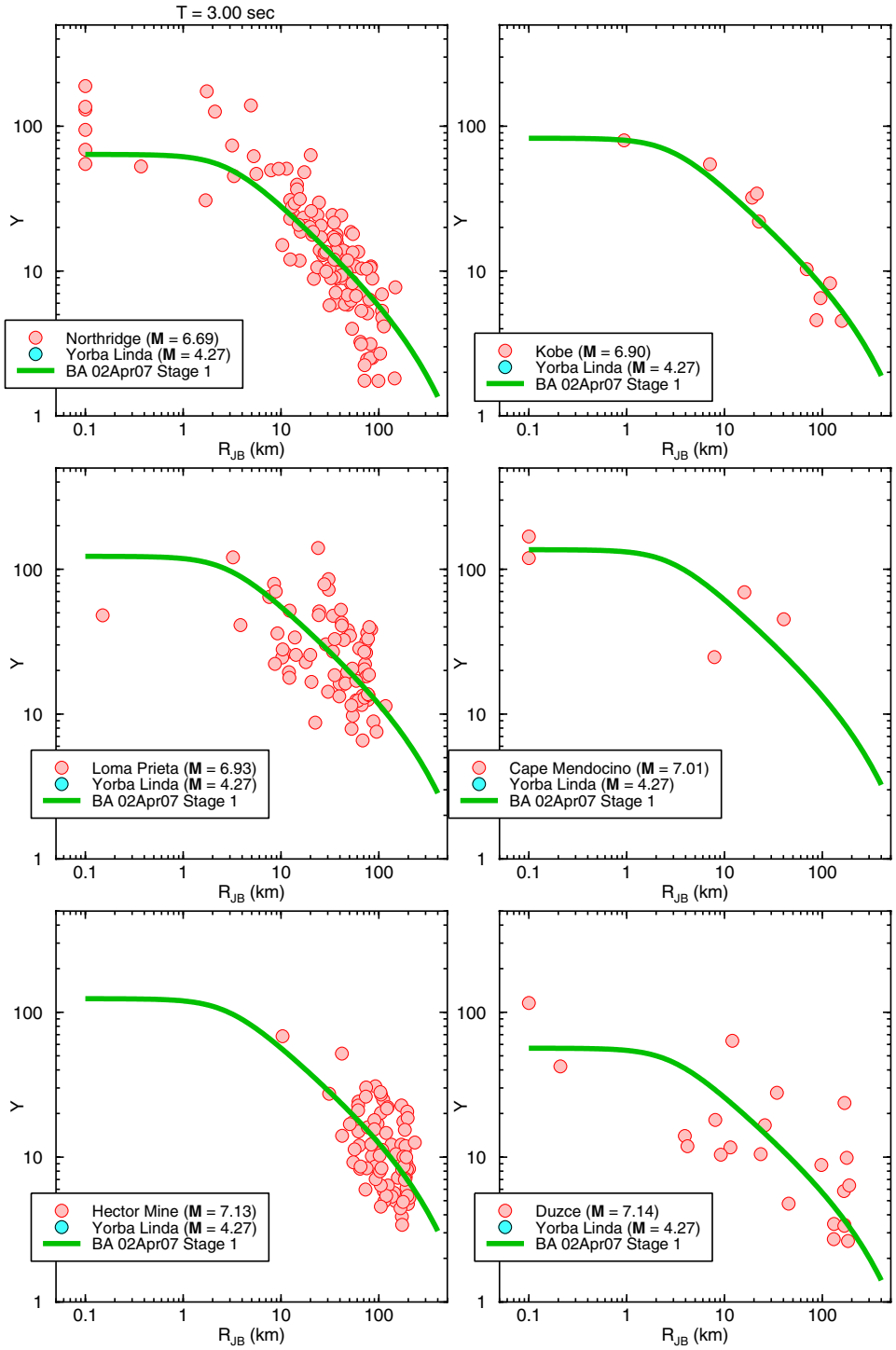
(a)

Fig. 4.10 Symbols: 1.0 s PSA observations, corrected to $V_{S30} = 760$ m/s, as function of distance for indicated events; Curves: Stage 1 regression fits.



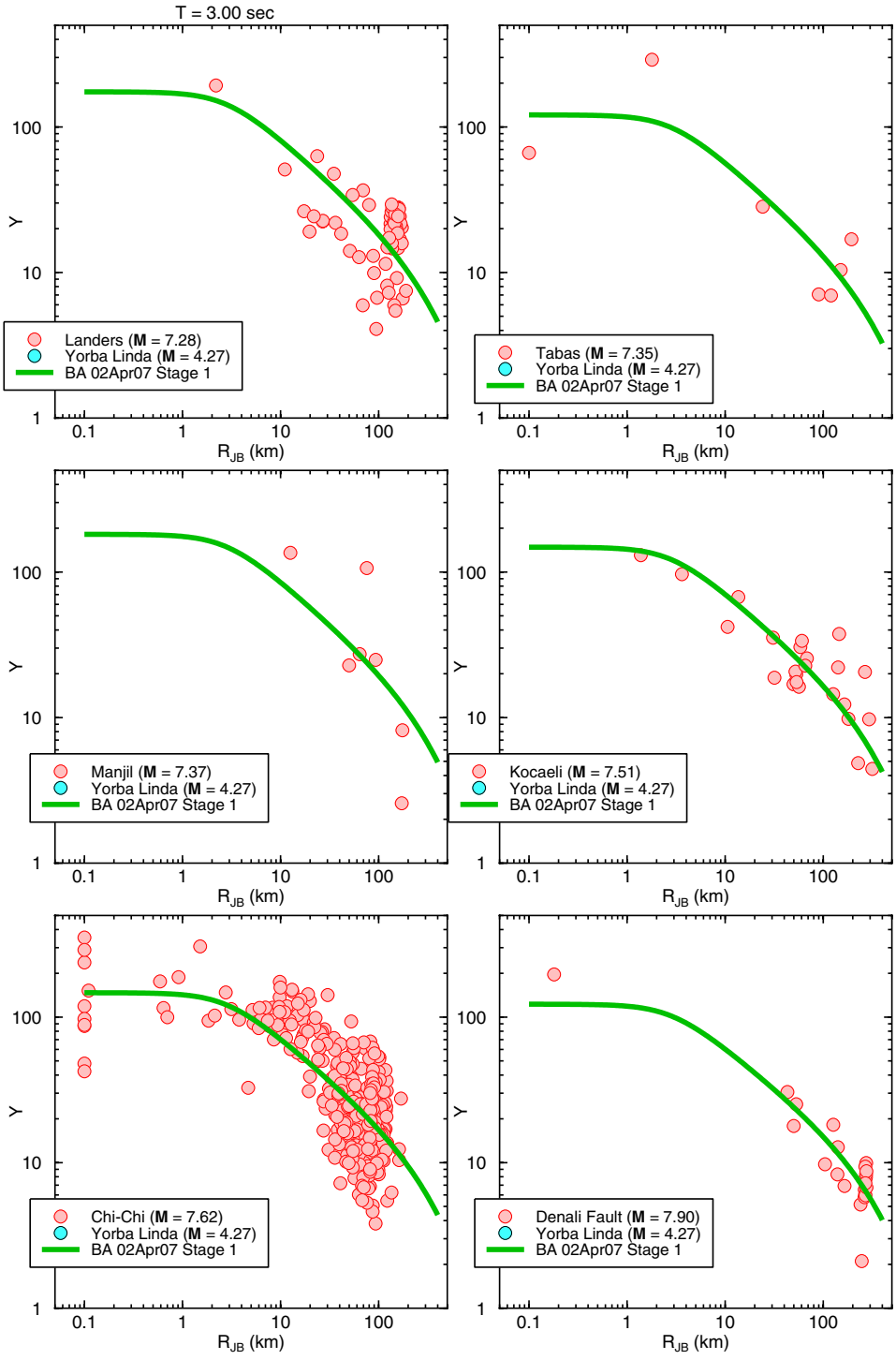
(b)

Fig. 4.10—*Continued*



(a)

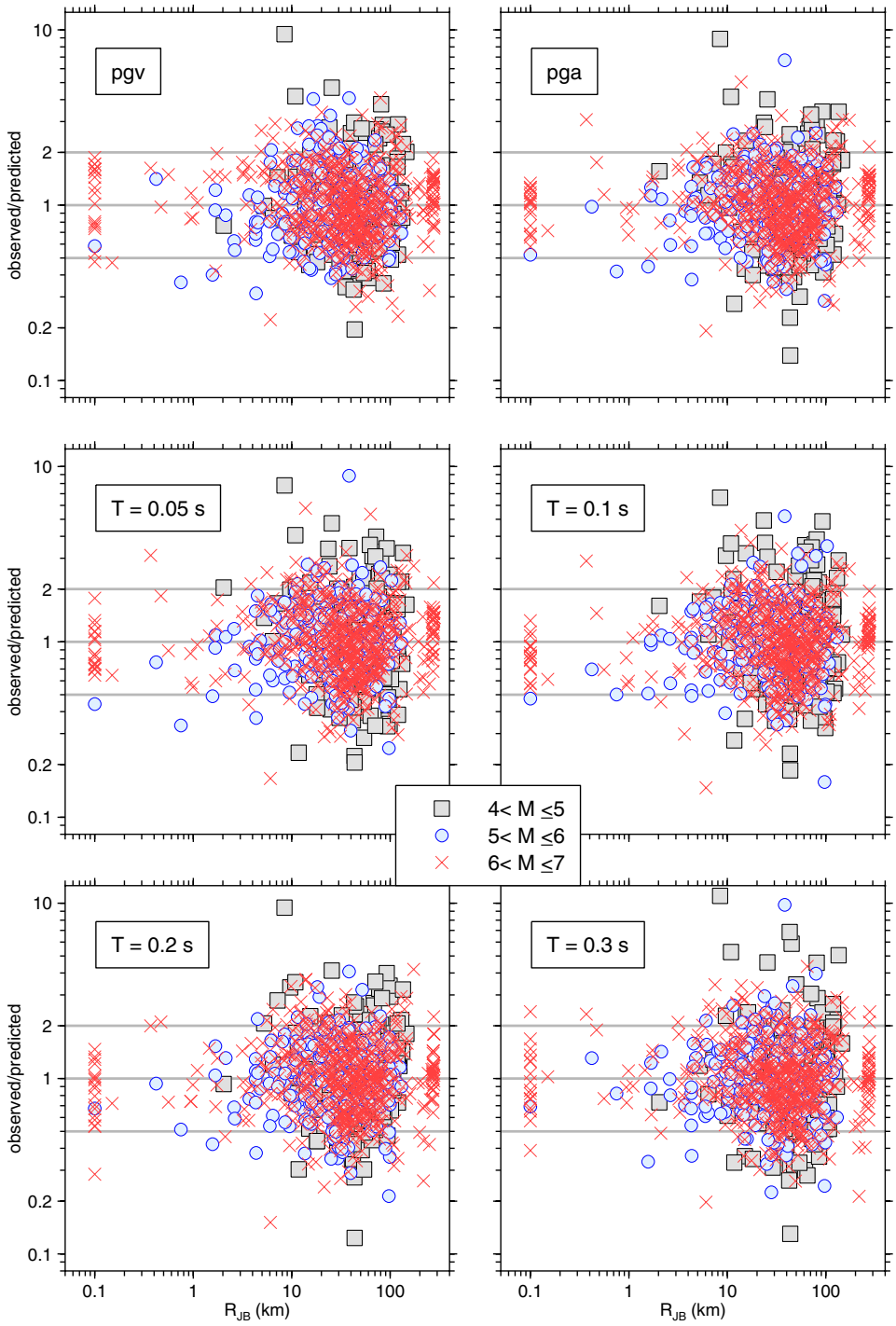
Fig. 4.11 Symbols: 3.0 s PSA observations, corrected to $V_{S30} = 760$ m/s, as function of distance for indicated events; Curves: Stage 1 regression fits.



(b)

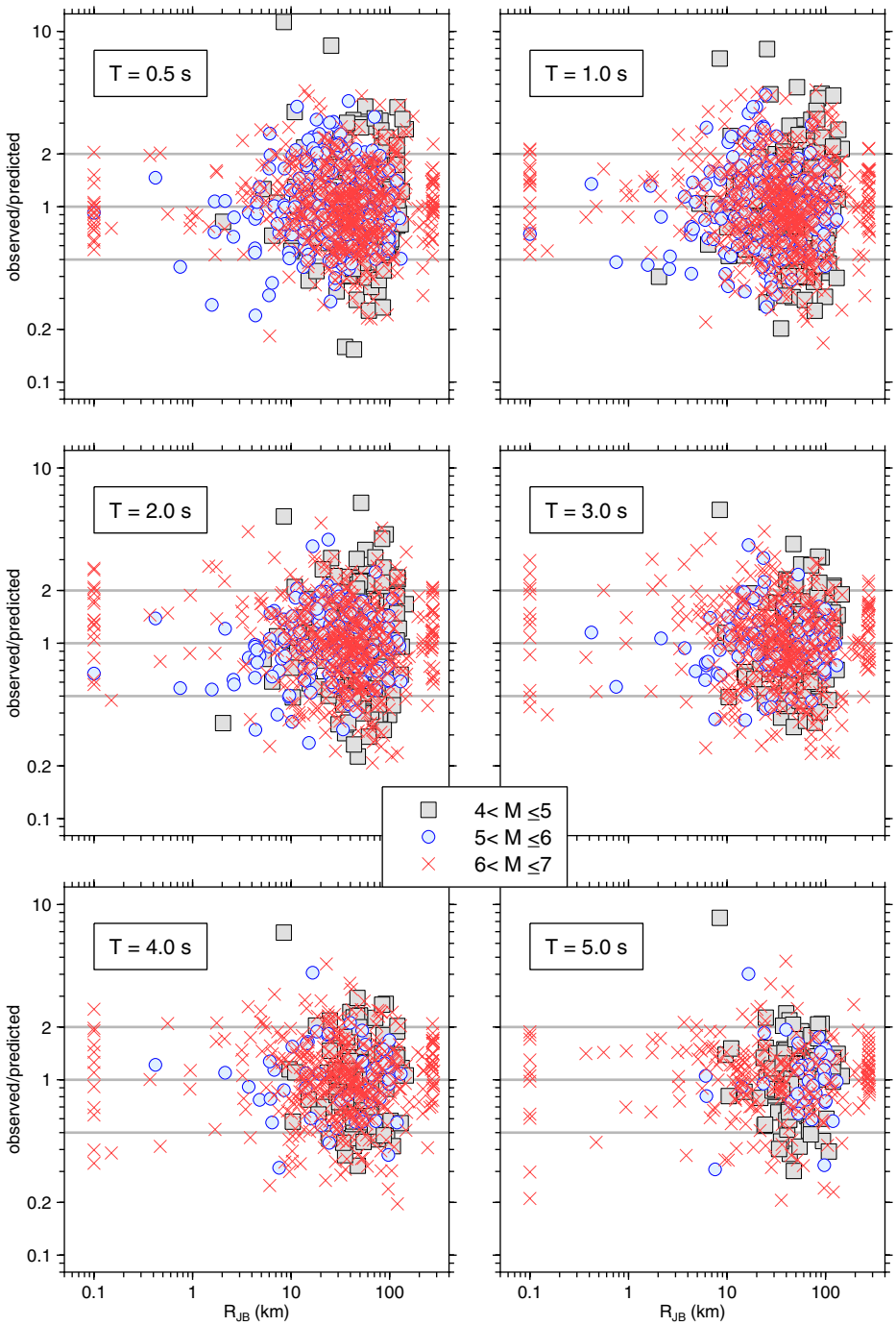
Fig. 4.11—*Continued*

A more precise way of looking for systematic mismatches between predictions and observations is to plot the residuals, defined as the ratio of observed to predicted ground motions. Figures 4.12(a)–(c) show residuals as a function of distance for earthquakes less than or equal to magnitude 7; Figures 4.13(a)–(c) show residuals from larger earthquakes, excluding the 1999 Chi-Chi earthquake, and Figures 4.14(a)–(c) show residuals just for the Chi-Chi earthquake. While there are some systematic departures for various periods and distances, the fit between observations and our predictions was judged to be adequate.



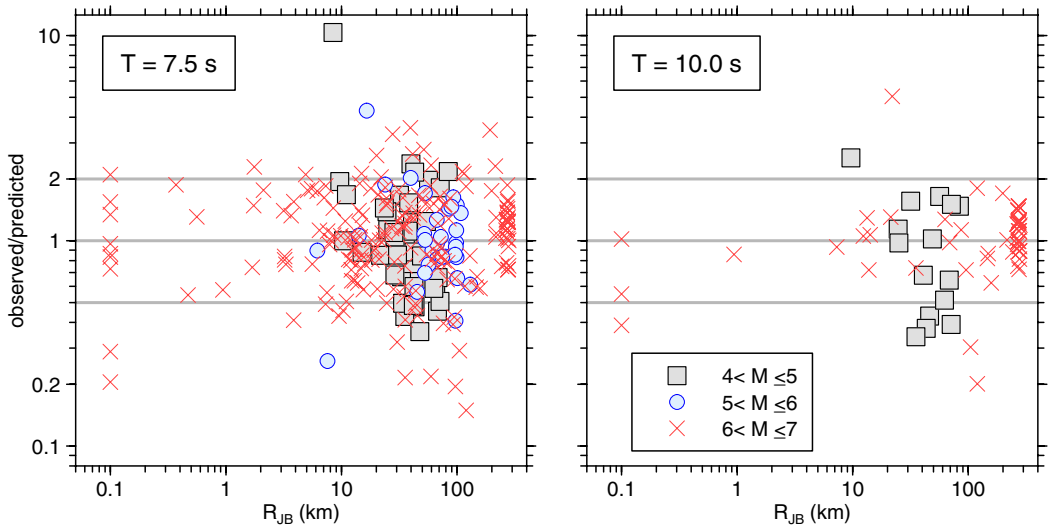
(a)

Fig. 4.12 Stage 1 residuals for earthquakes less than or equal to 7.0.



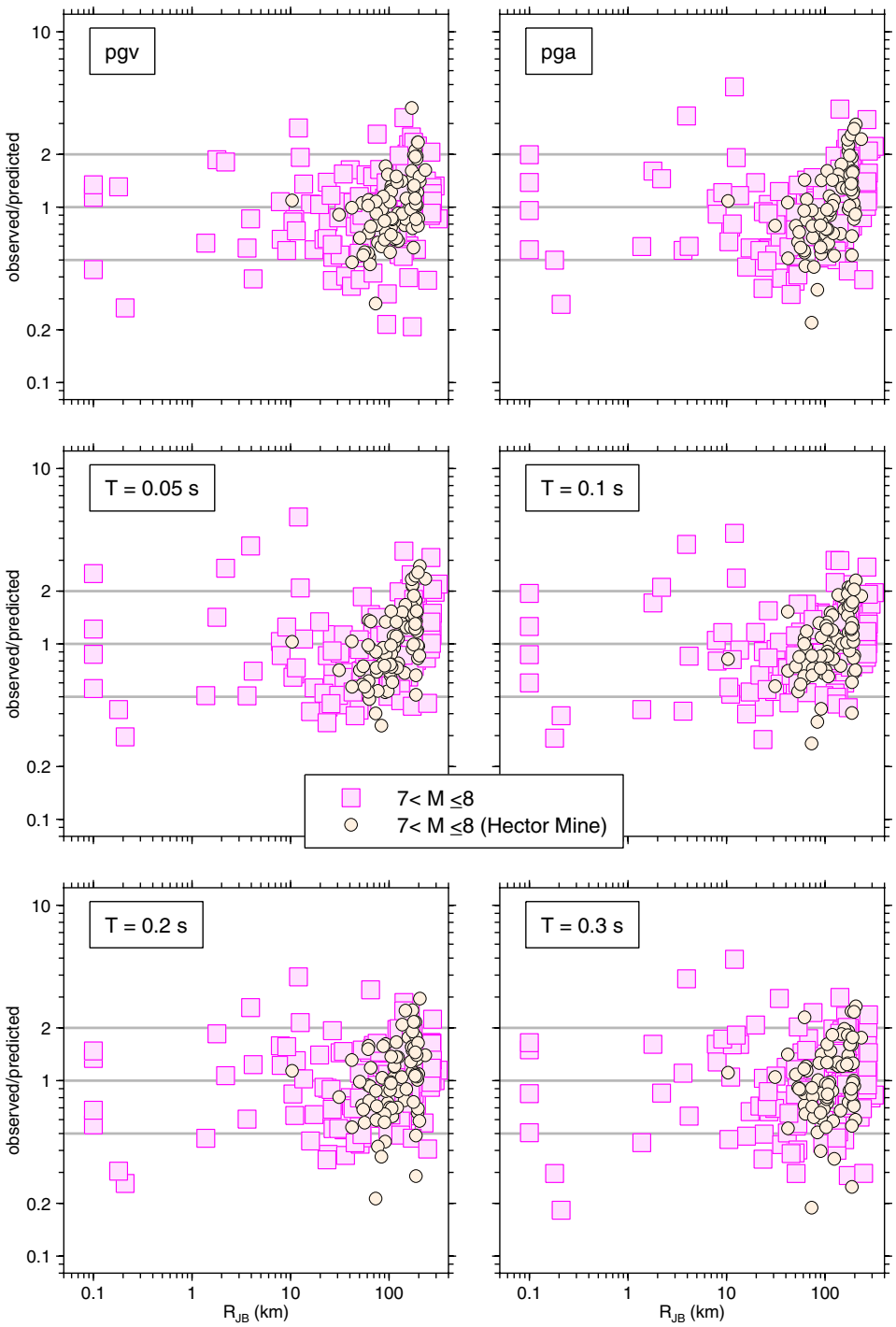
(b)

Fig. 4.12—*Continued*



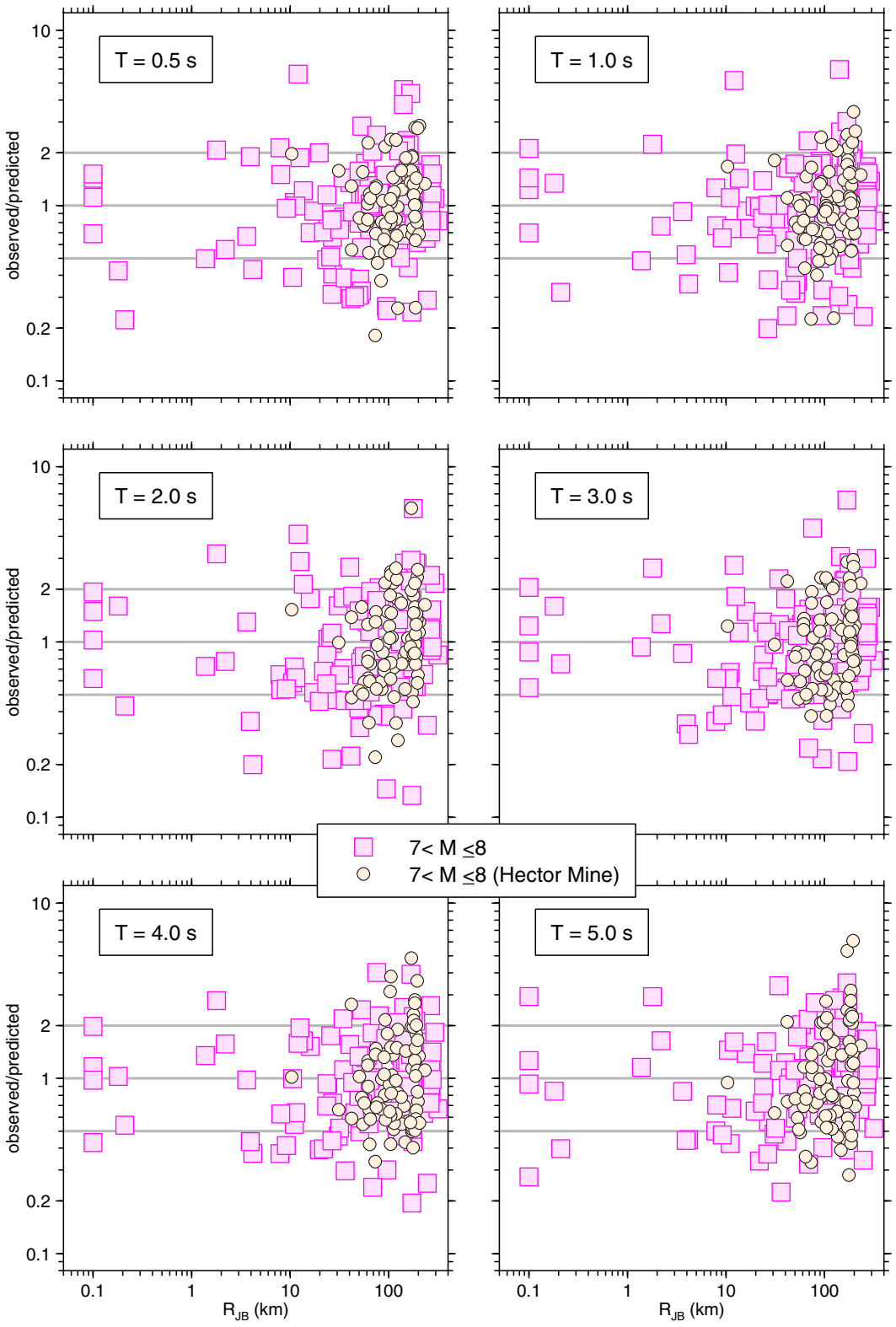
(c)

Fig. 4.12—*Continued*



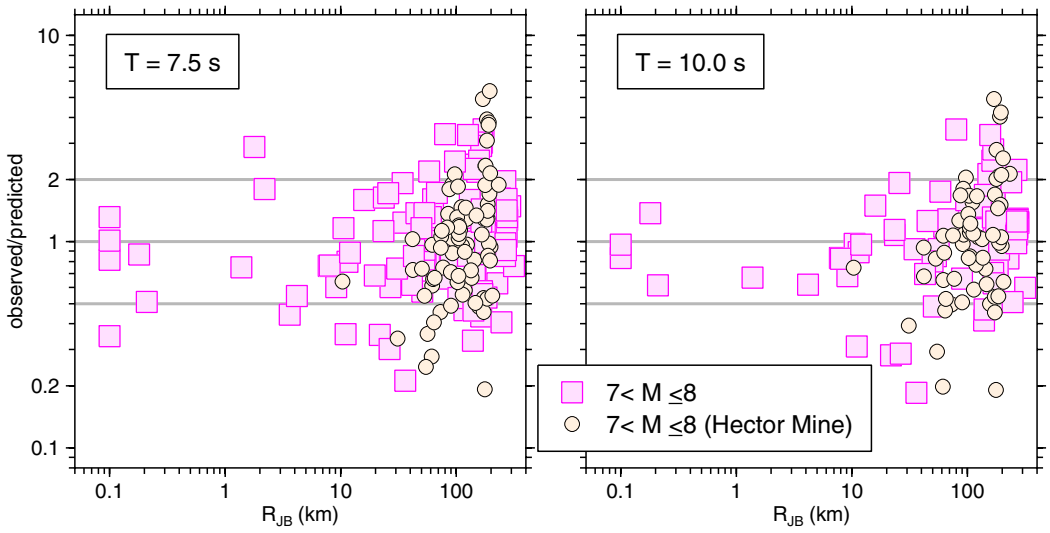
(a)

Fig. 4.13 Stage 1 residuals for earthquakes greater than 7.0, excluding 1999 Chi-Chi.



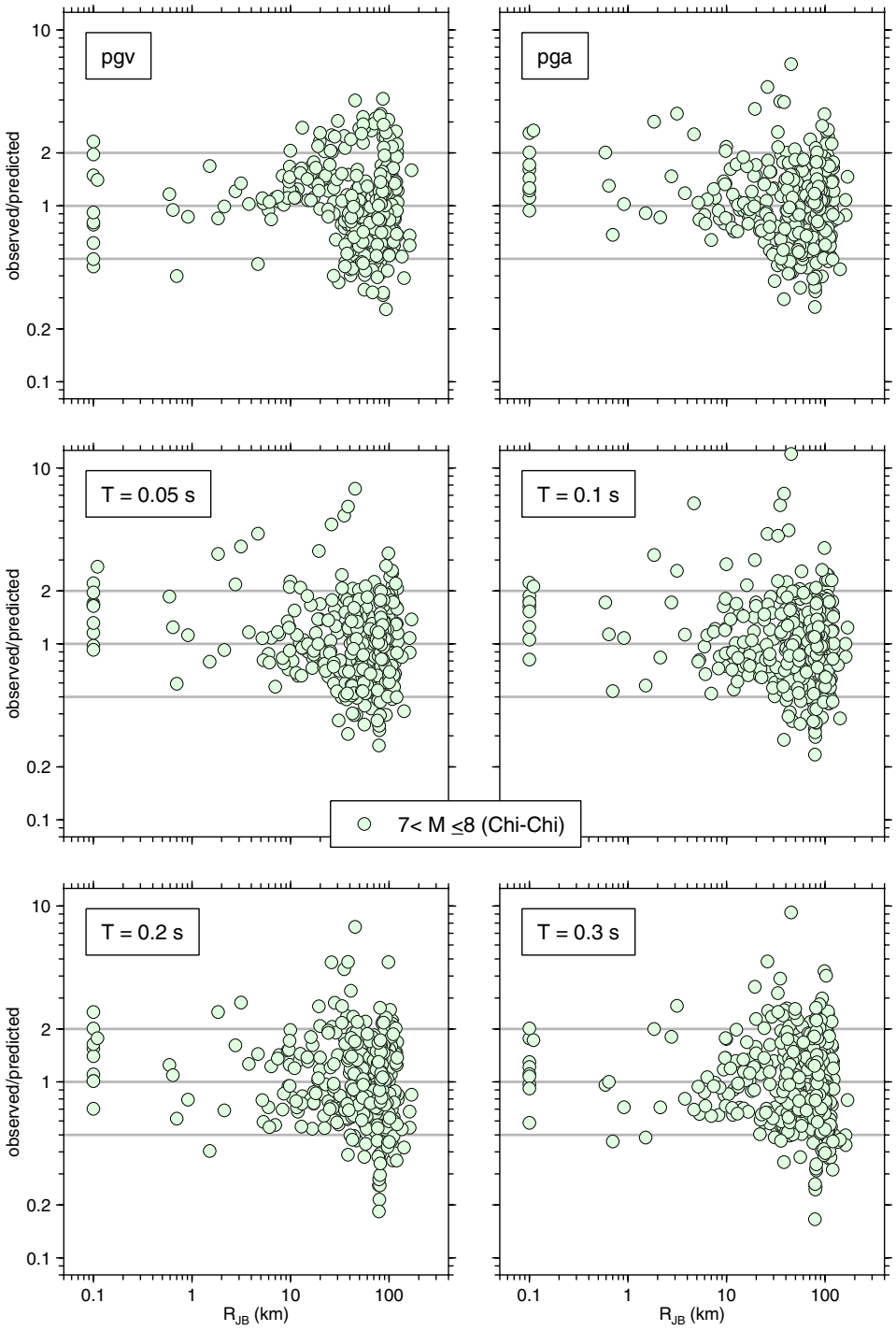
(b)

Fig. 4.13—*Continued*



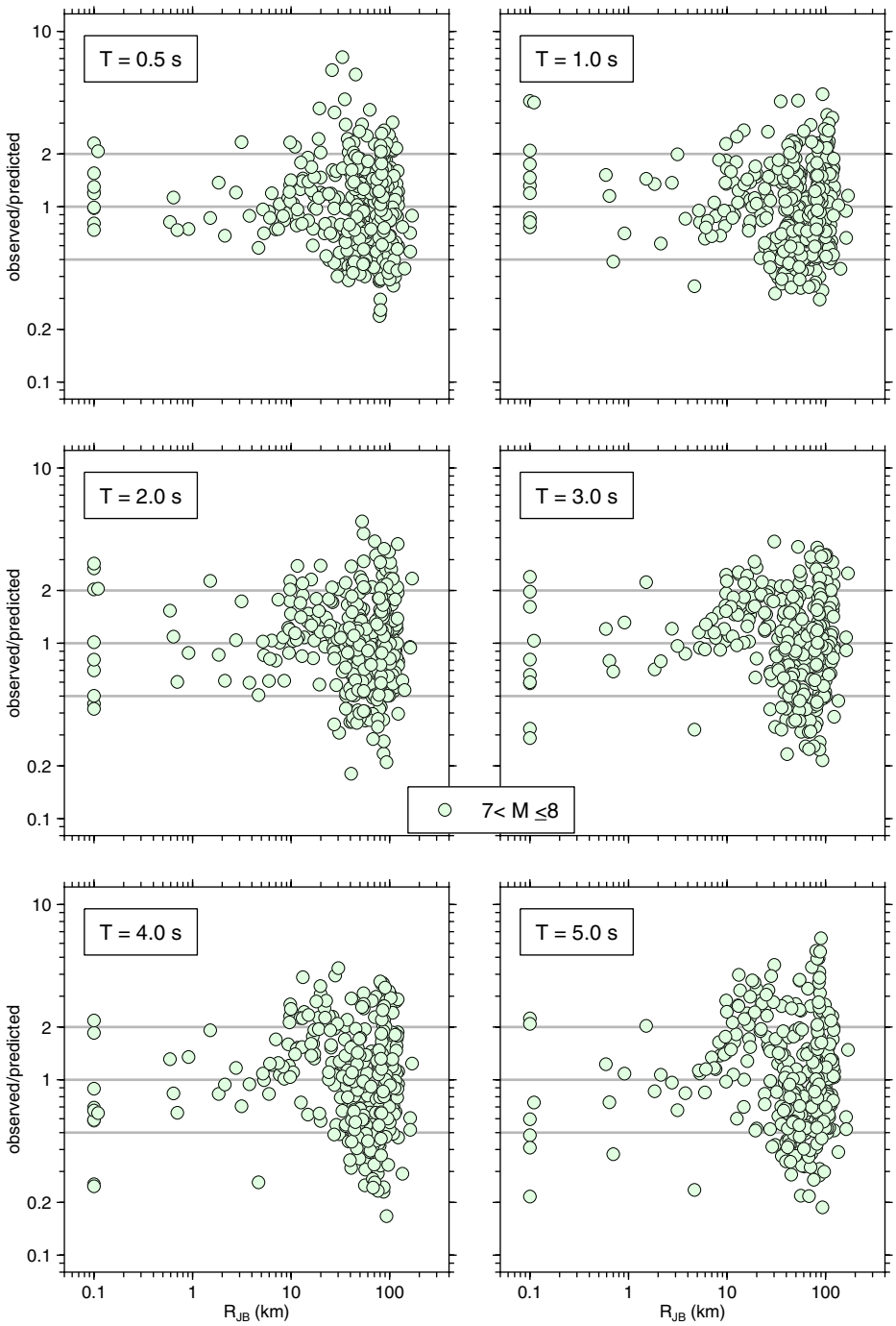
(c)

Fig. 4.13—*Continued*



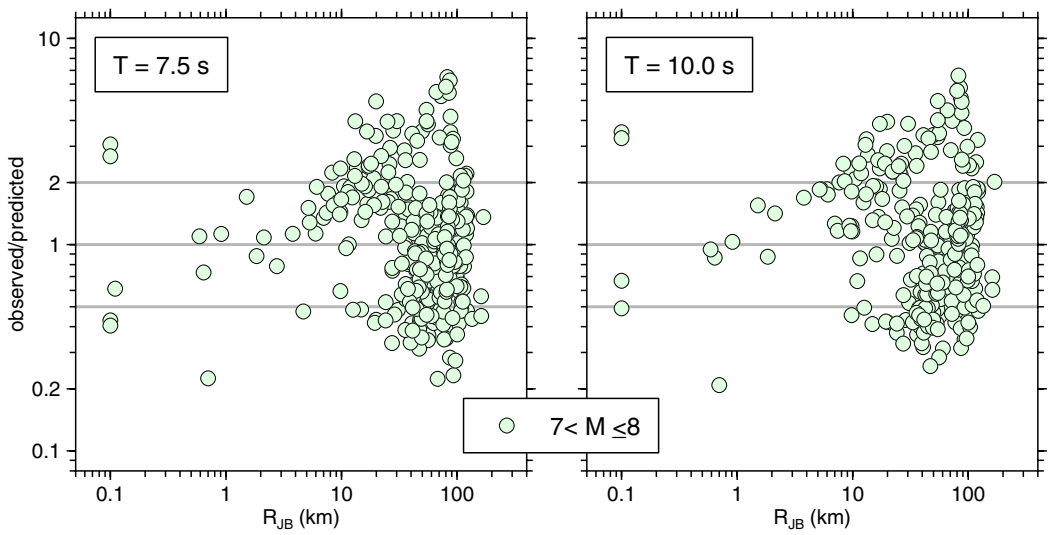
(a)

Fig. 4.14 Stage 1 residuals for 1999 Chi-Chi mainshock.



(b)

Fig. 4.14—*Continued*

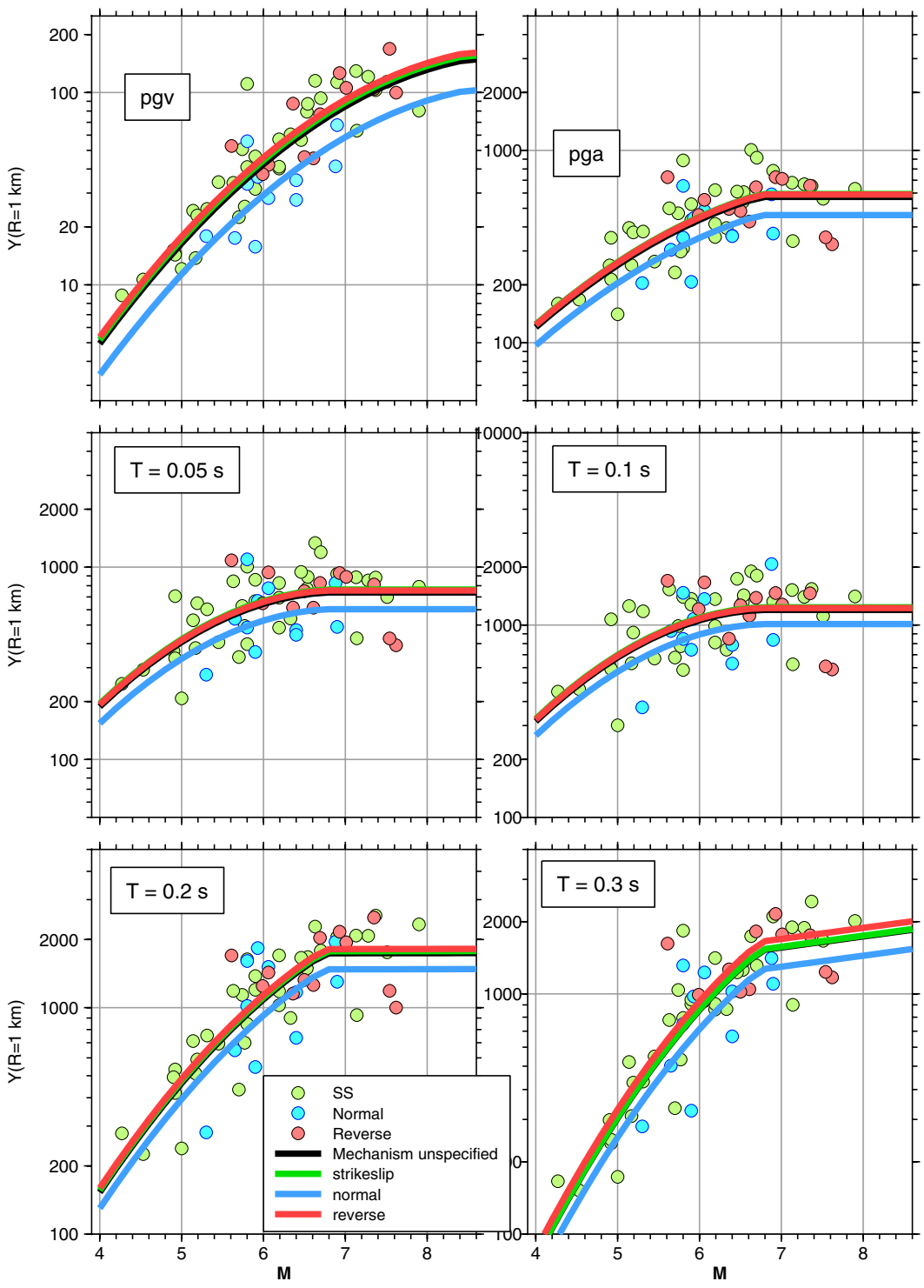


(c)

Fig. 4.14—*Continued*

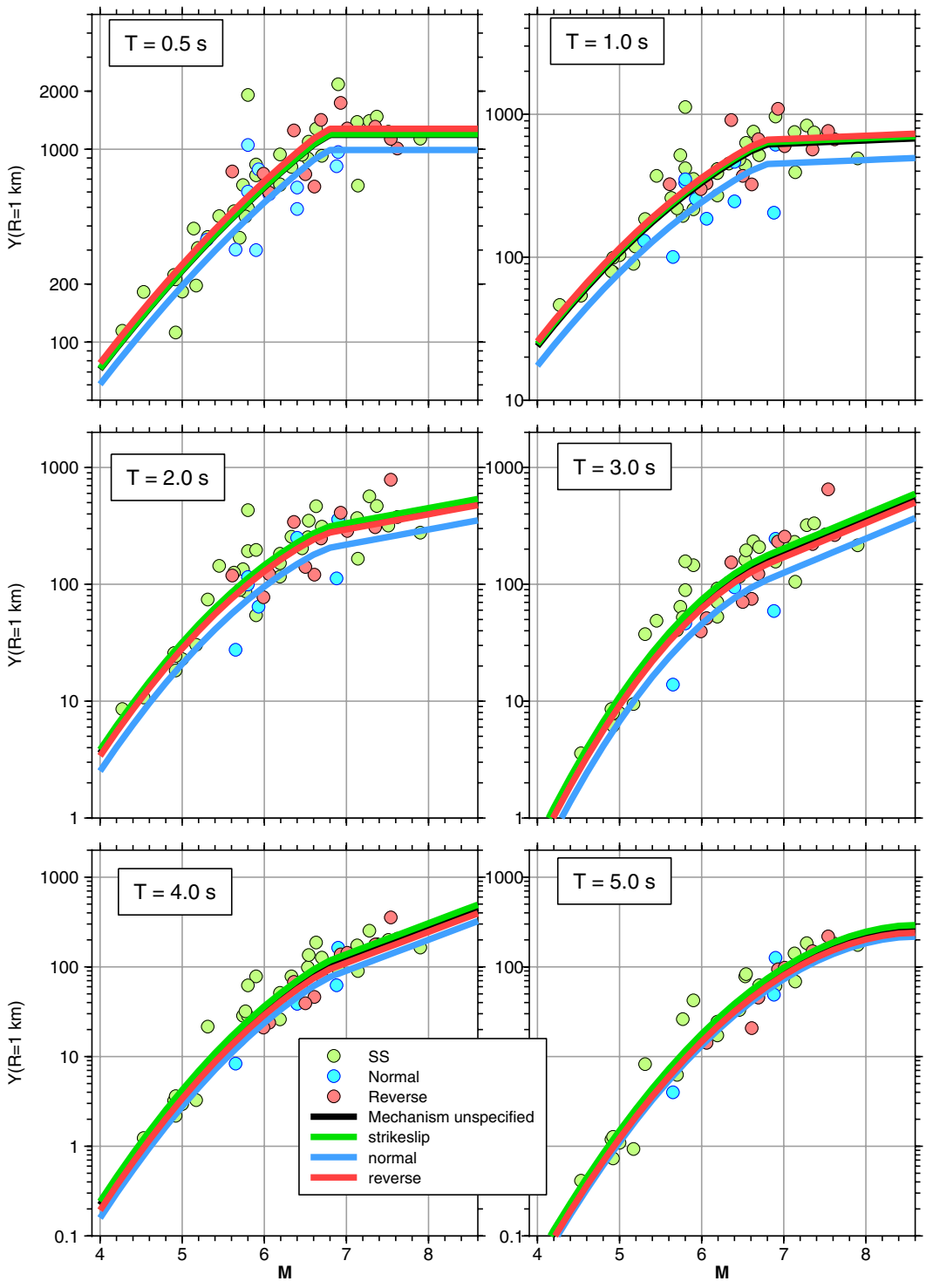
4.2.4 Fit of Stage 2 Regressions

Figures 4.15(a)–(c) are plots 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 amplitude data. Note that the magnitude scaling for T=10 s at **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 **M**<6.5.



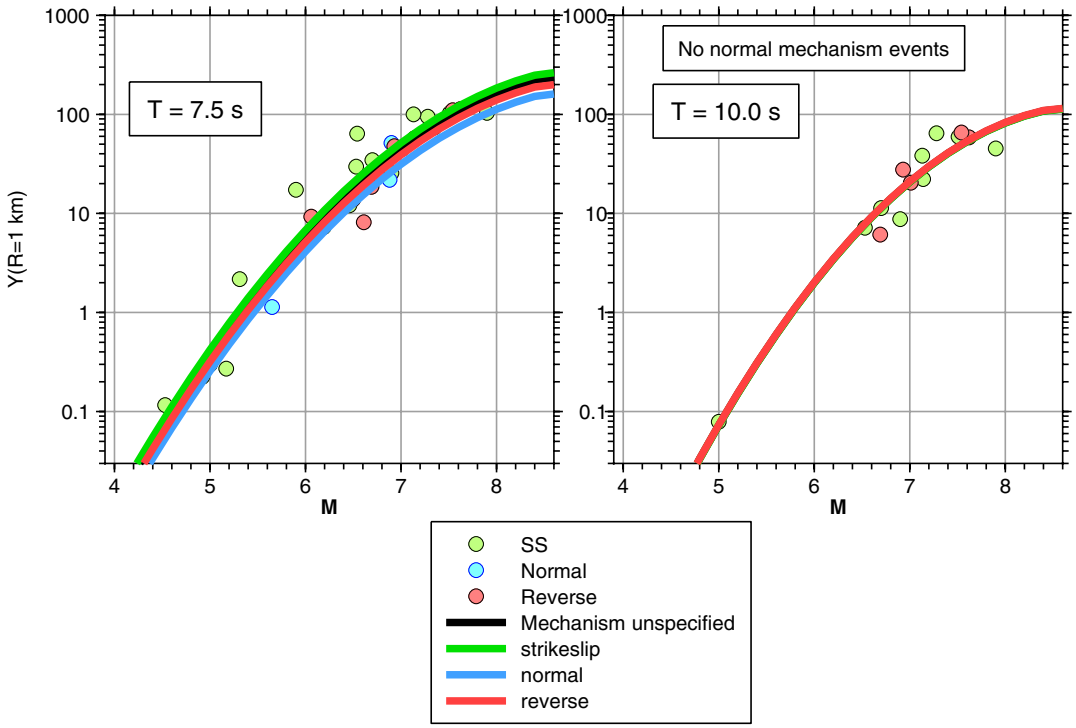
(a)

Fig. 4.15 Event terms and Stage 2 regression fits.



(b)

Fig. 4.15—*Continued*



(c)

Fig. 4.15—*Continued*

4.2.5 Predictions of PSA from Combined Stage 1 and Stage 2 Regressions

Graphs of PSA and PGA predicted from our equations for three values of R_{JB} and four magnitudes are shown in Figure 4.16. The curves for the larger earthquakes tend to pinch 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 (Fig. 4.5). But otherwise the PSA are quite smooth, especially considering that many of the coefficients were determined independently for each period.

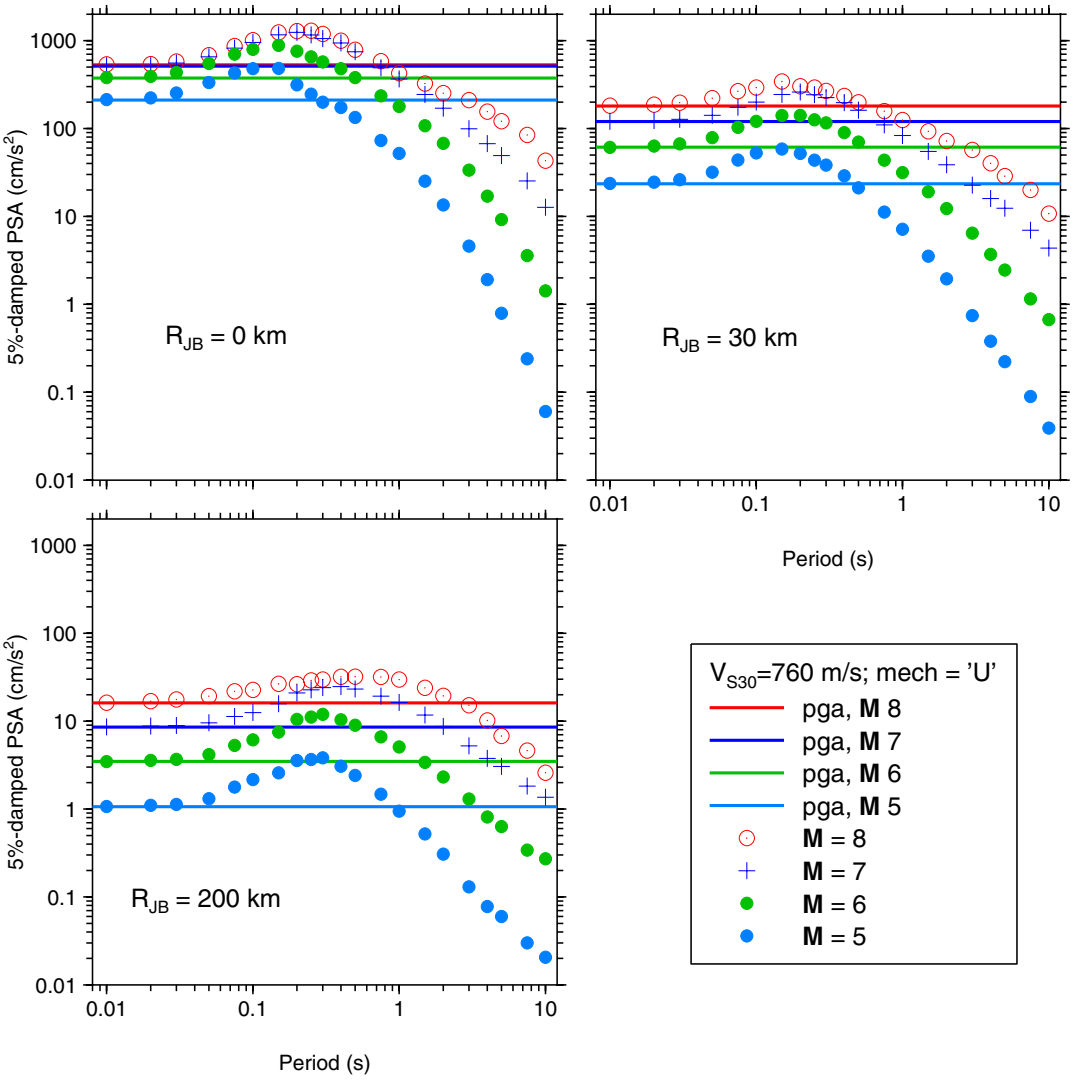
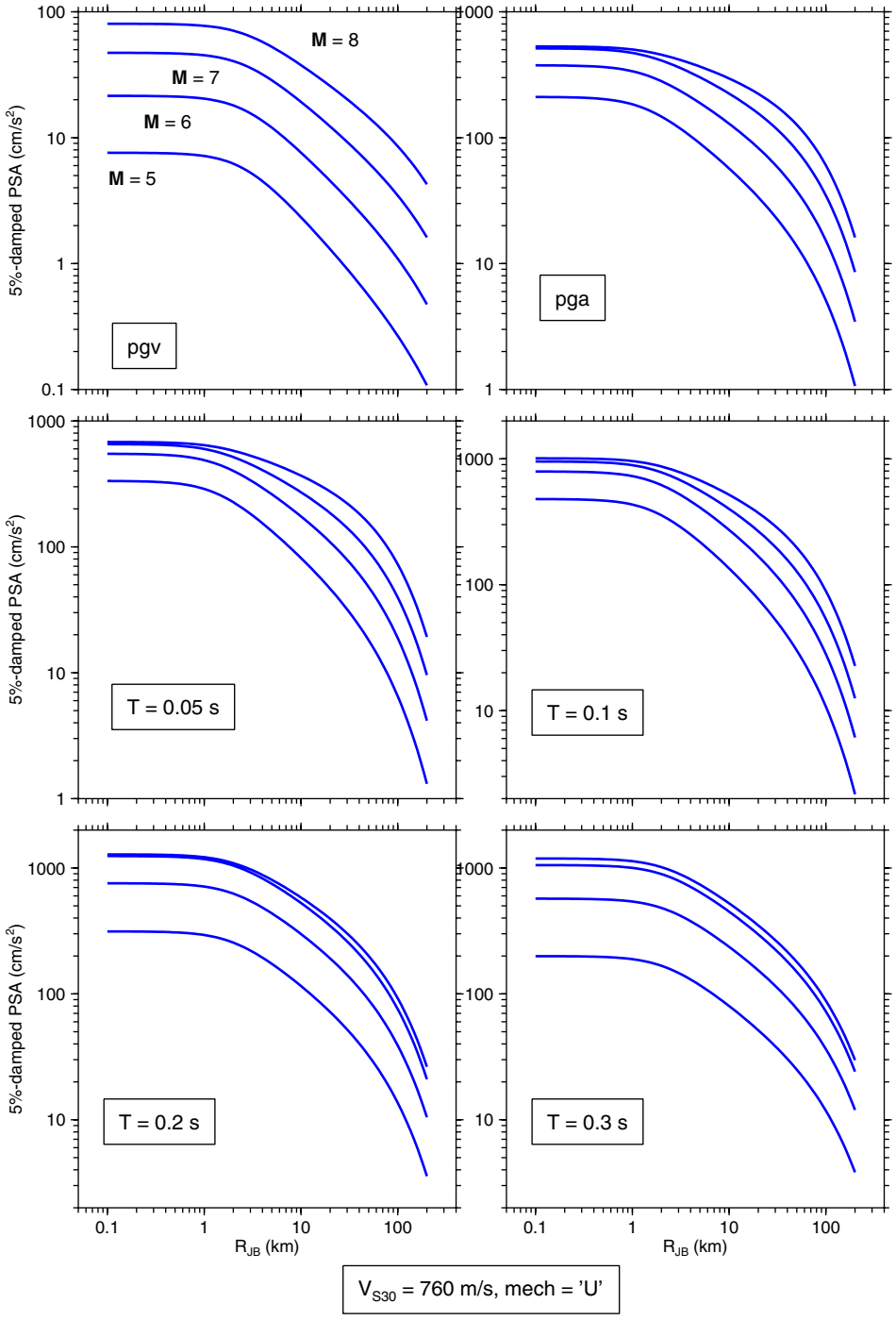


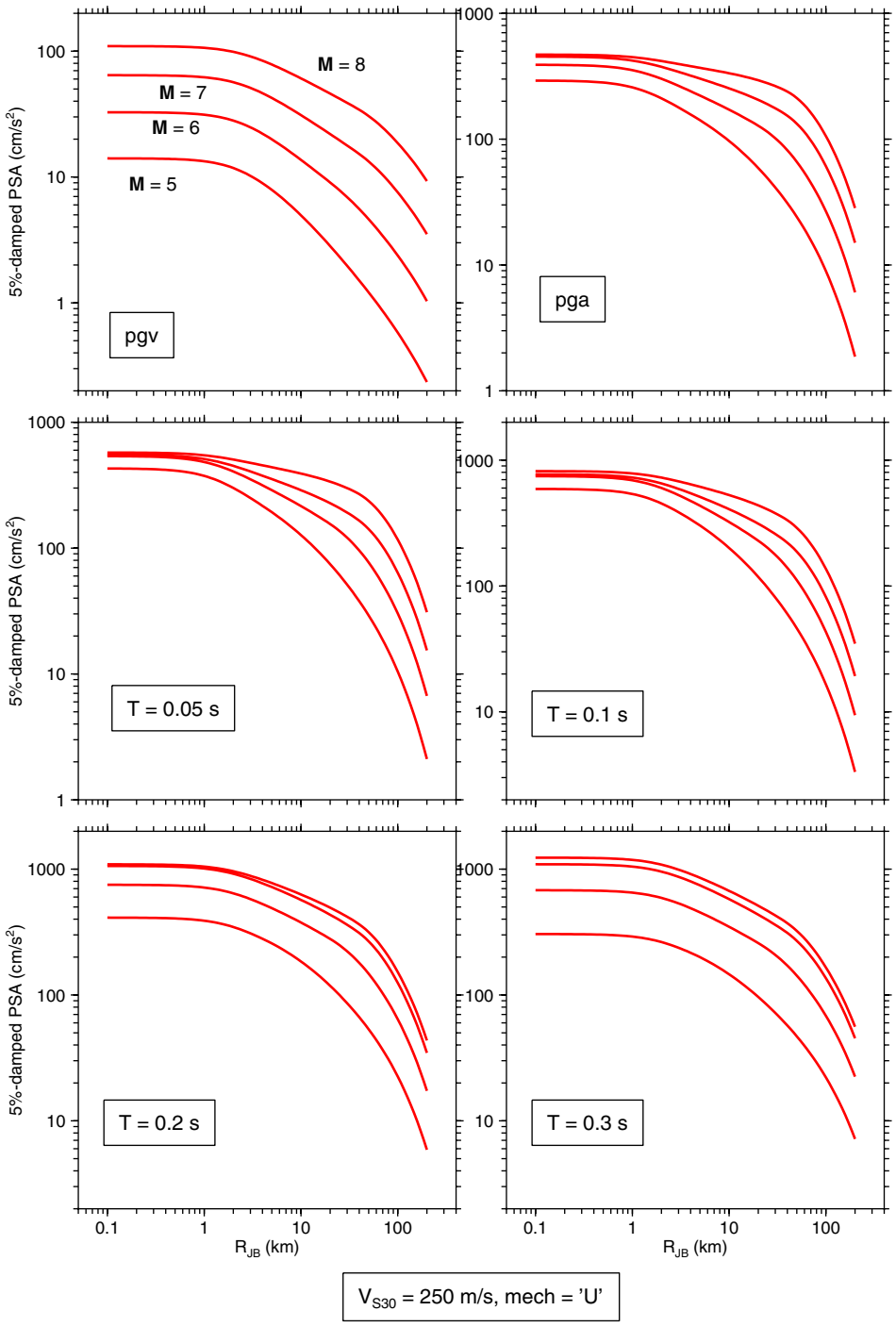
Fig. 4.16 PSA from our equations, as function of period; see legend for details.

Plots of PSA as a function of distance are shown in Figures 4.17(a)–(f) for the whole range of periods. The figures are in pairs, one for $V_{S30} = 760$ m/s (NEHRP B/C boundary), followed by one for $V_{S30} = 250$ m/s (NEHRP D; see Appendix E).



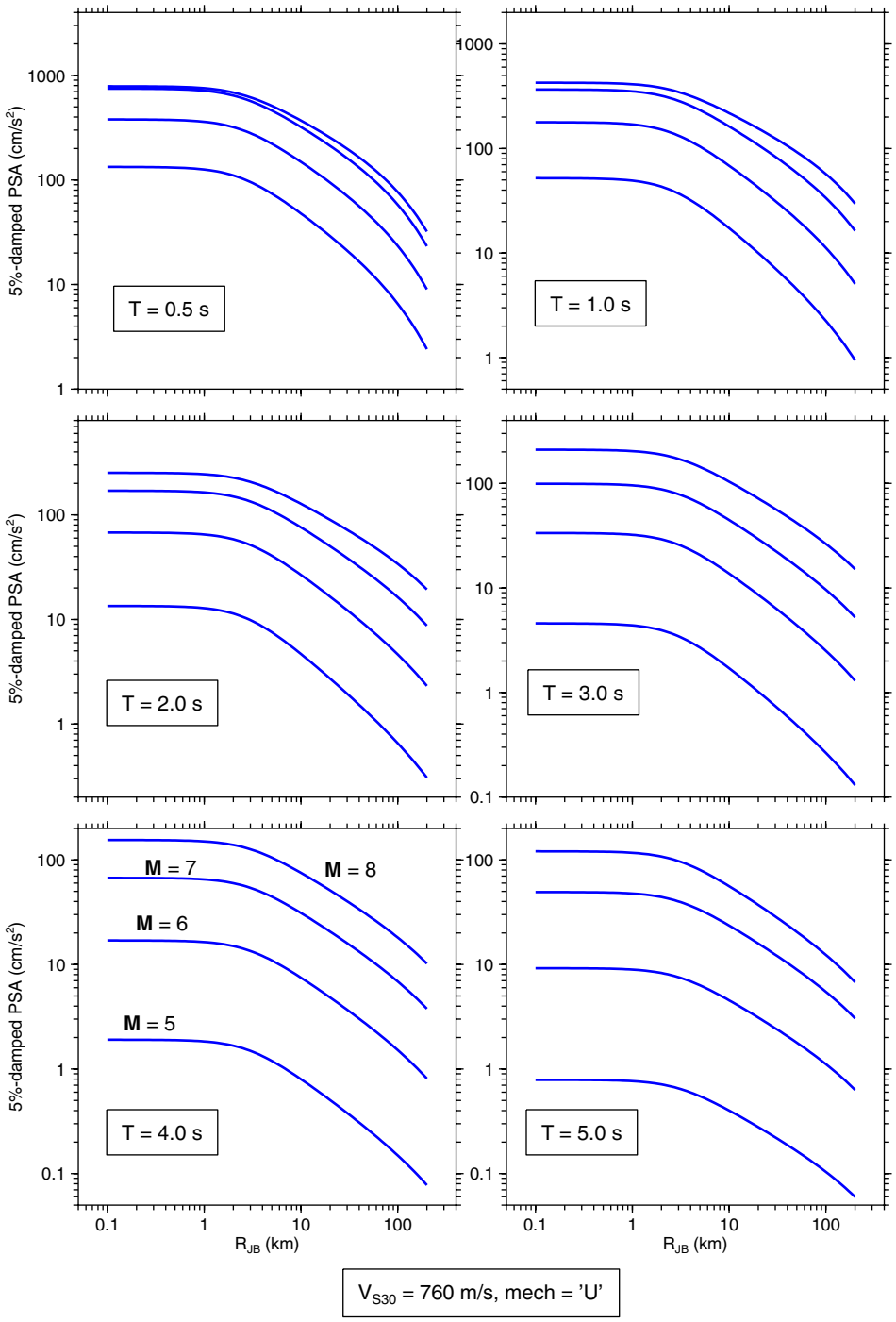
(a)

Fig. 4.17 PSA from our equations, as function of distance; see legend for details.



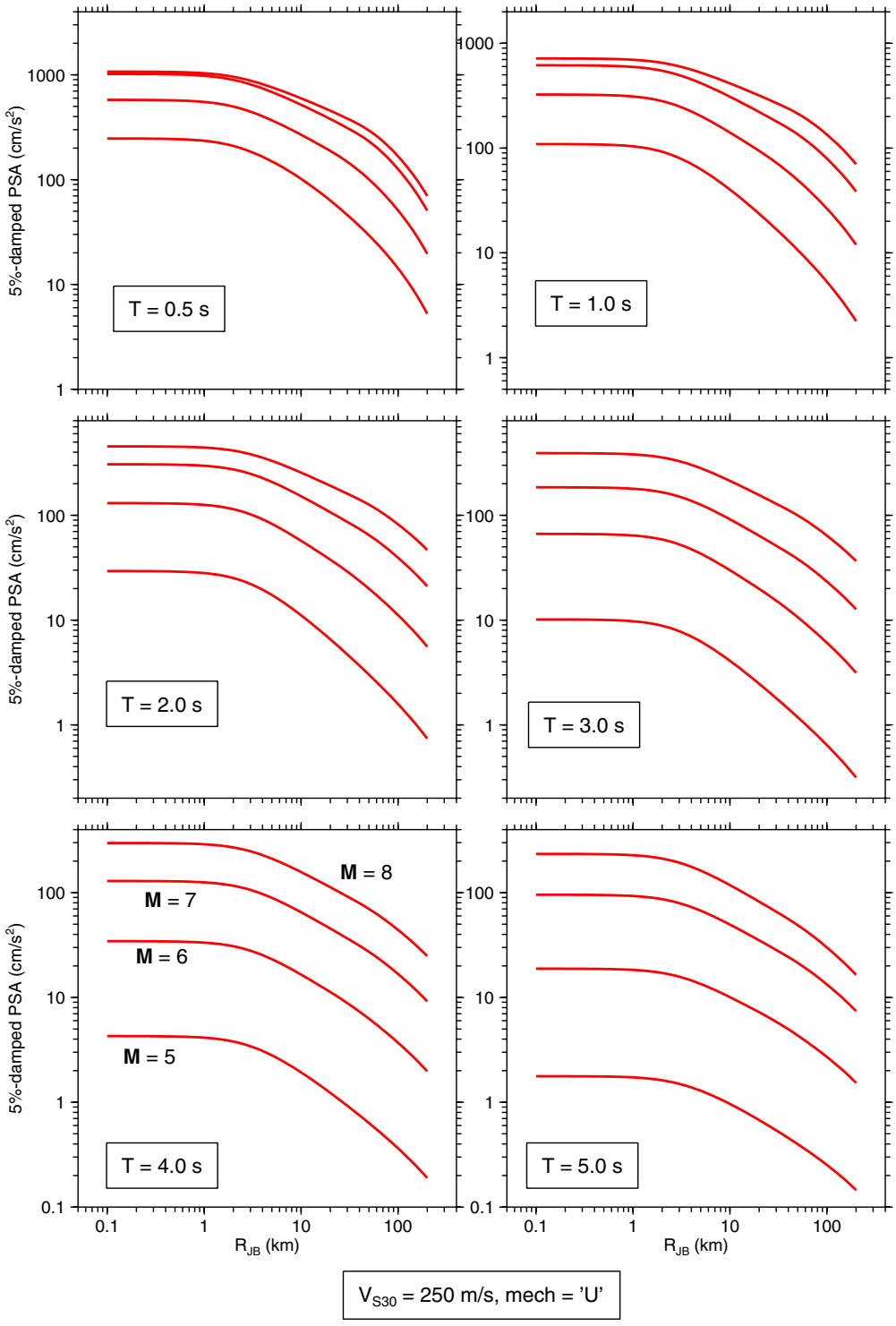
(b)

Fig. 4.17—*Continued*



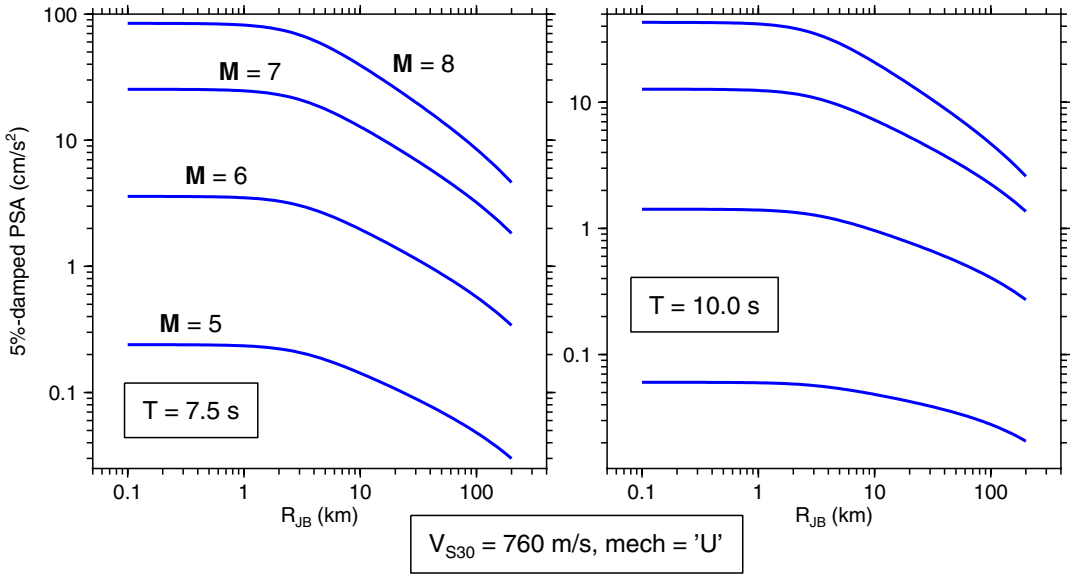
(c)

Fig. 4.17—*Continued*



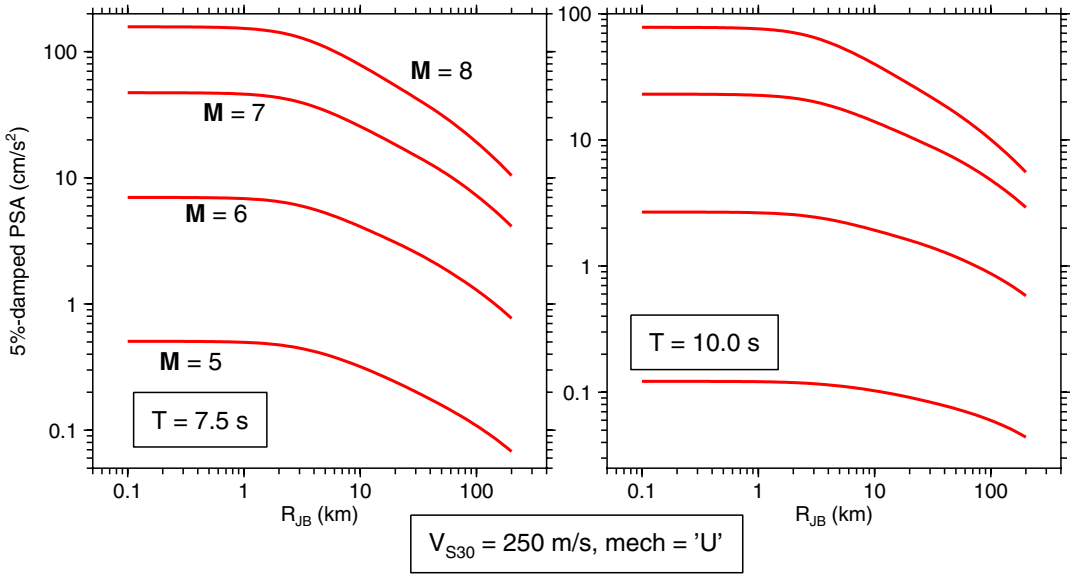
(d)

Fig. 4.17—*Continued*



(e)

Fig. 4.17—*Continued*



(f)

Fig. 4.17—*Continued*

The effect of V_{S30} on predicted ground motion amplitude is shown more directly in Figure 4.18. Nonlinear soil amplification causes the curves to cross, such that at close distances lower 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.

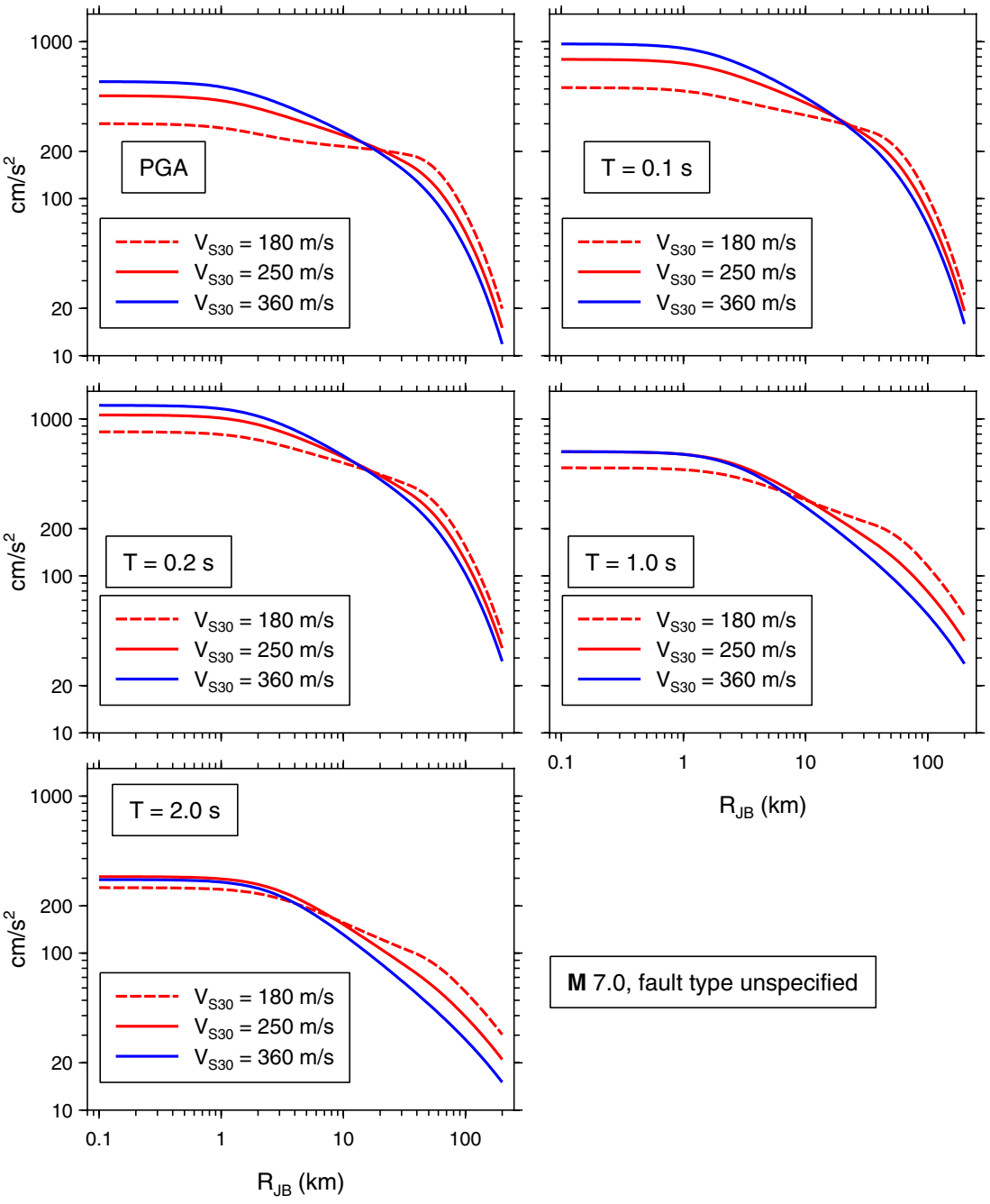
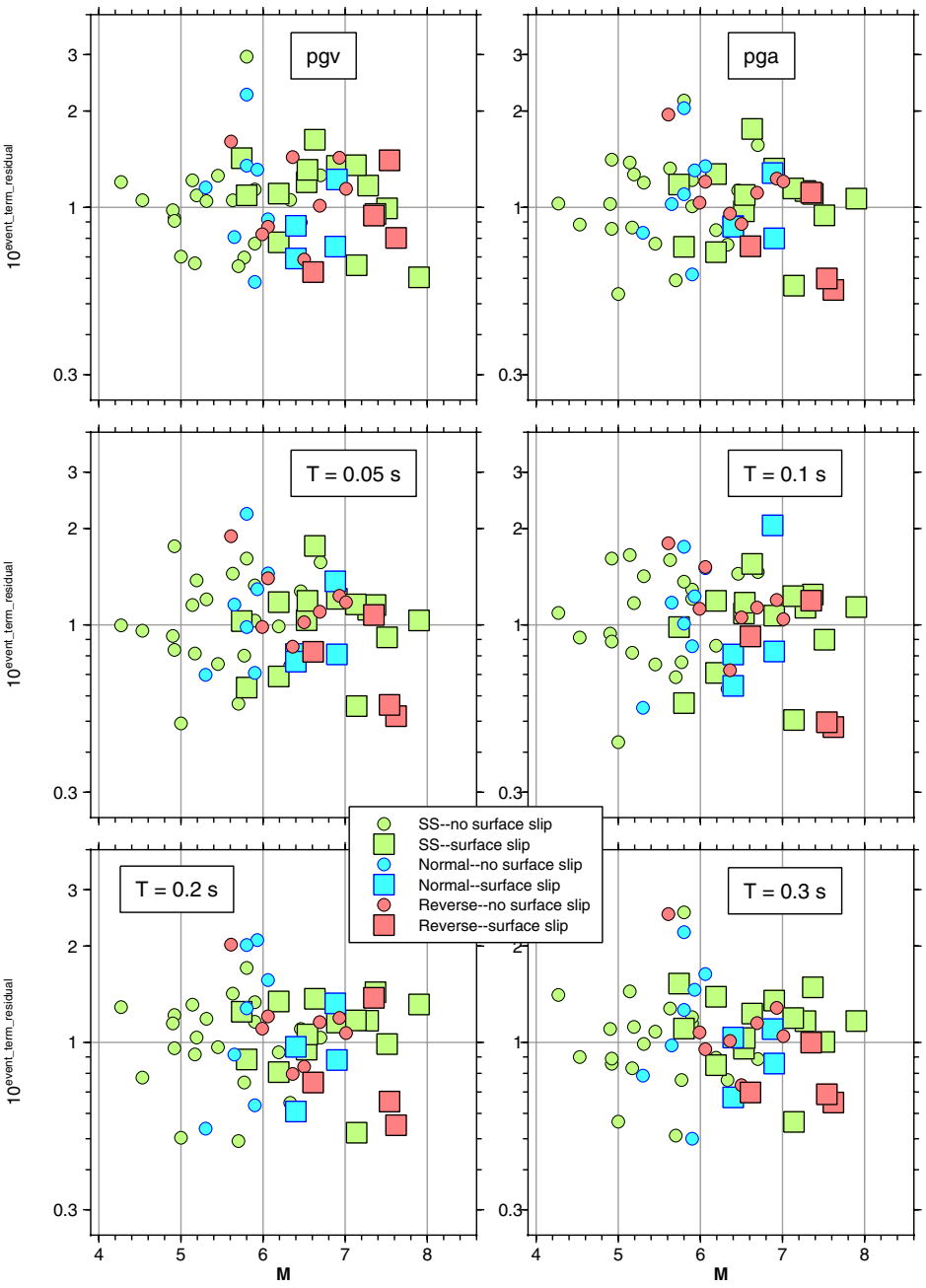


Fig. 4.18 PSA from our equations, as a function of distance; see legend for details.

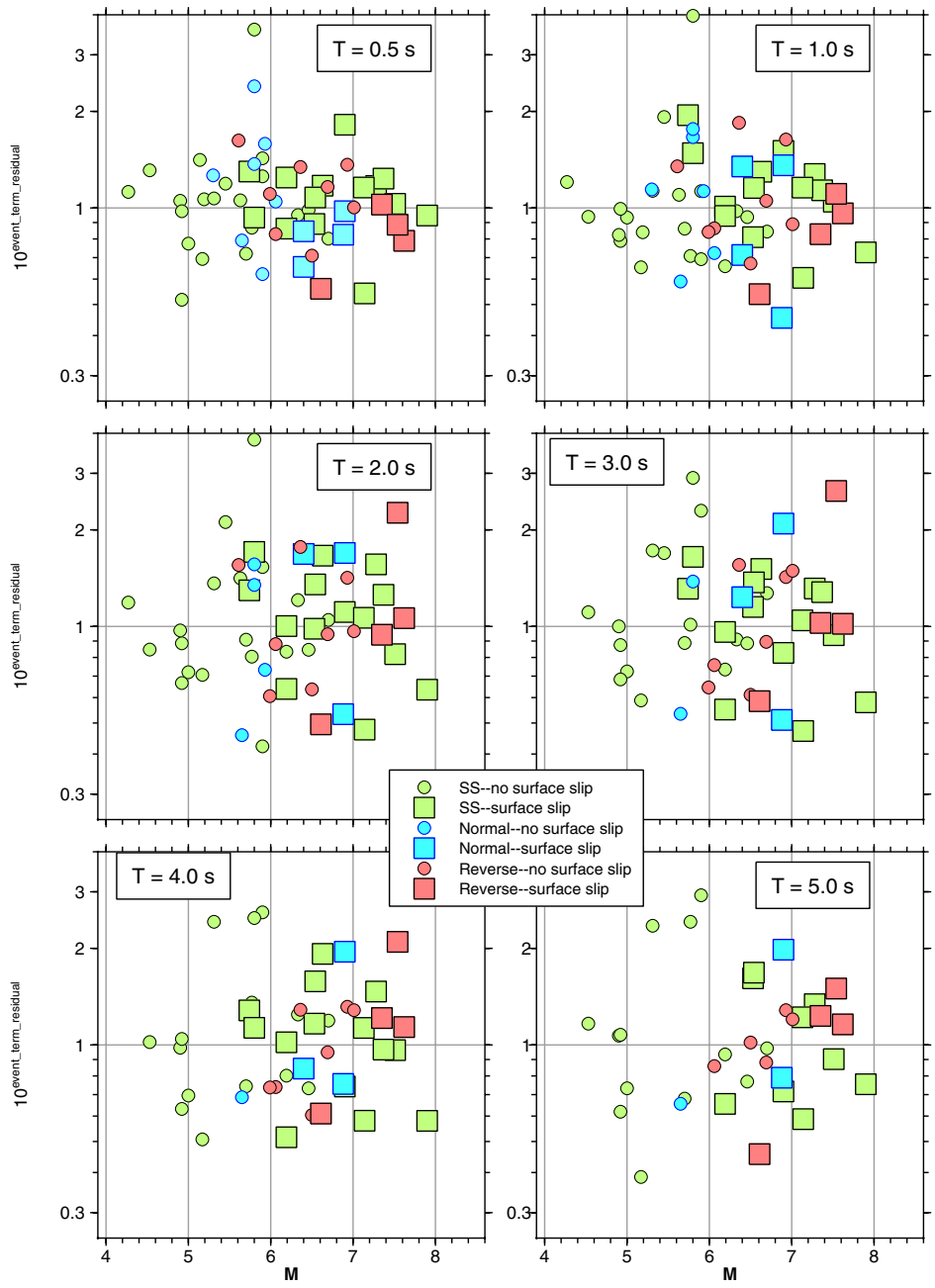
4.2.6 Surface-Slip vs. No-Surface-Slip Earthquakes

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 Figures 4.19(a)–(c), which show the event terms 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 of magnitudes. As seen in Figures 4.19(a)–(c), it is only for strike-slip earthquakes that there are 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 terms 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, it might be that in the future we will add a buried/surface faulting term to the equations, even though the data do not demand it. 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 (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).



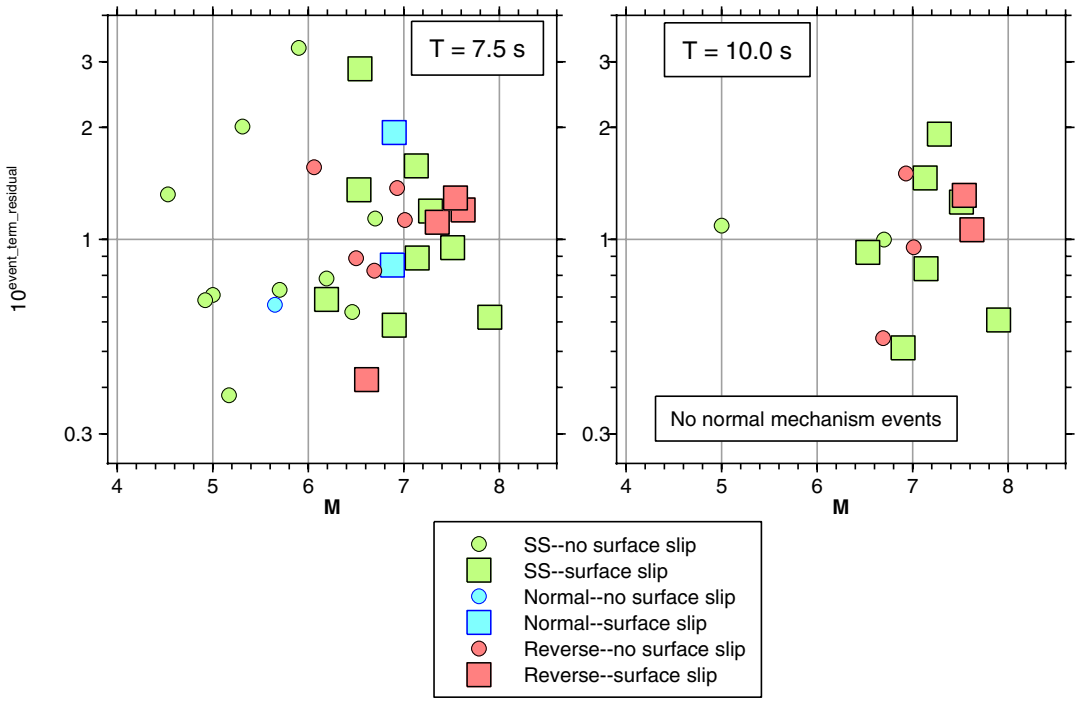
(a)

Fig. 4.19 Antilogarithms of event terms, plotted against magnitude and differentiated by events of different fault types, for which faults did or did not break to surface.



(b)

Fig. 4.19—*Continued*

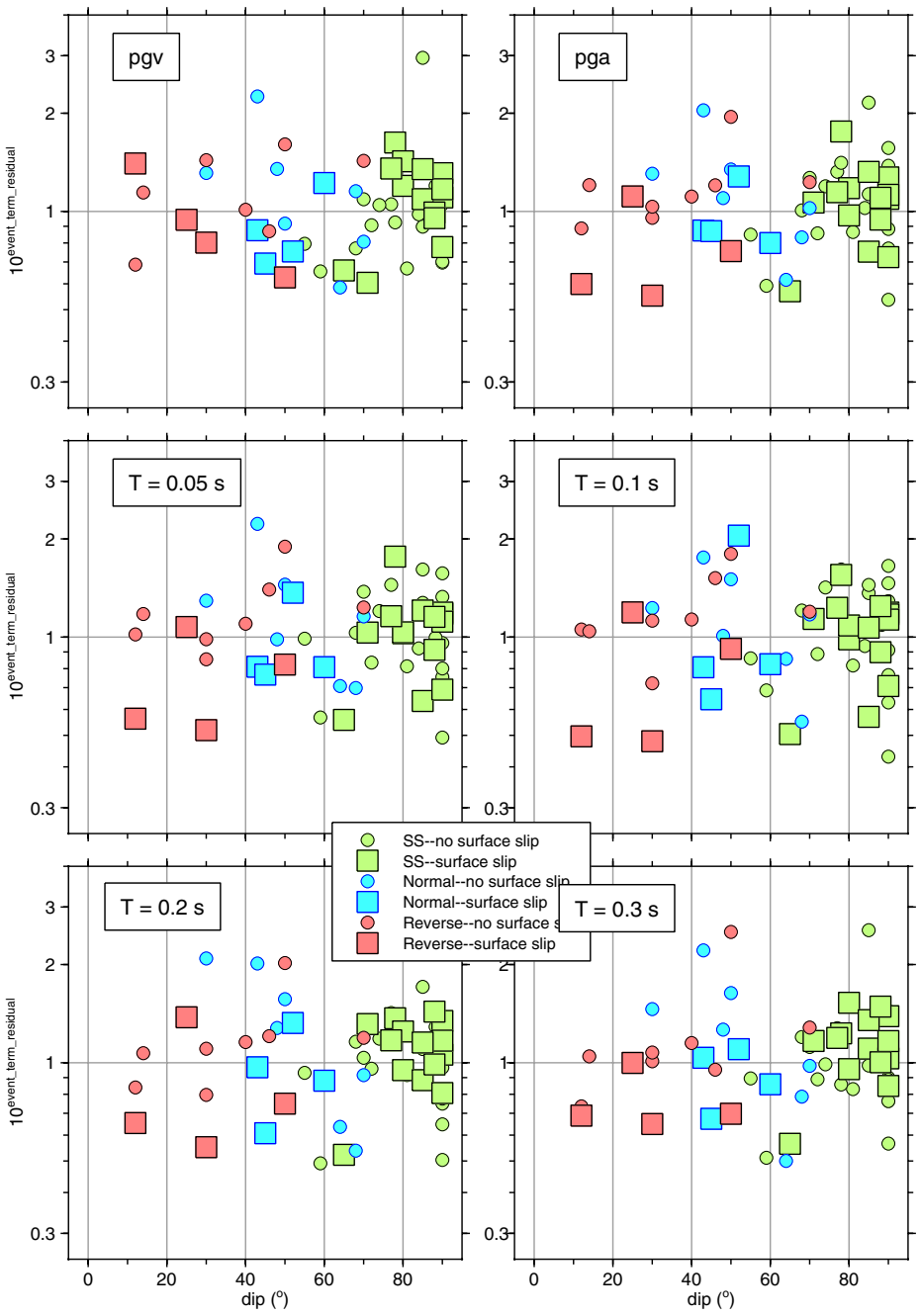


(c)

Fig. 4.19—*Continued*

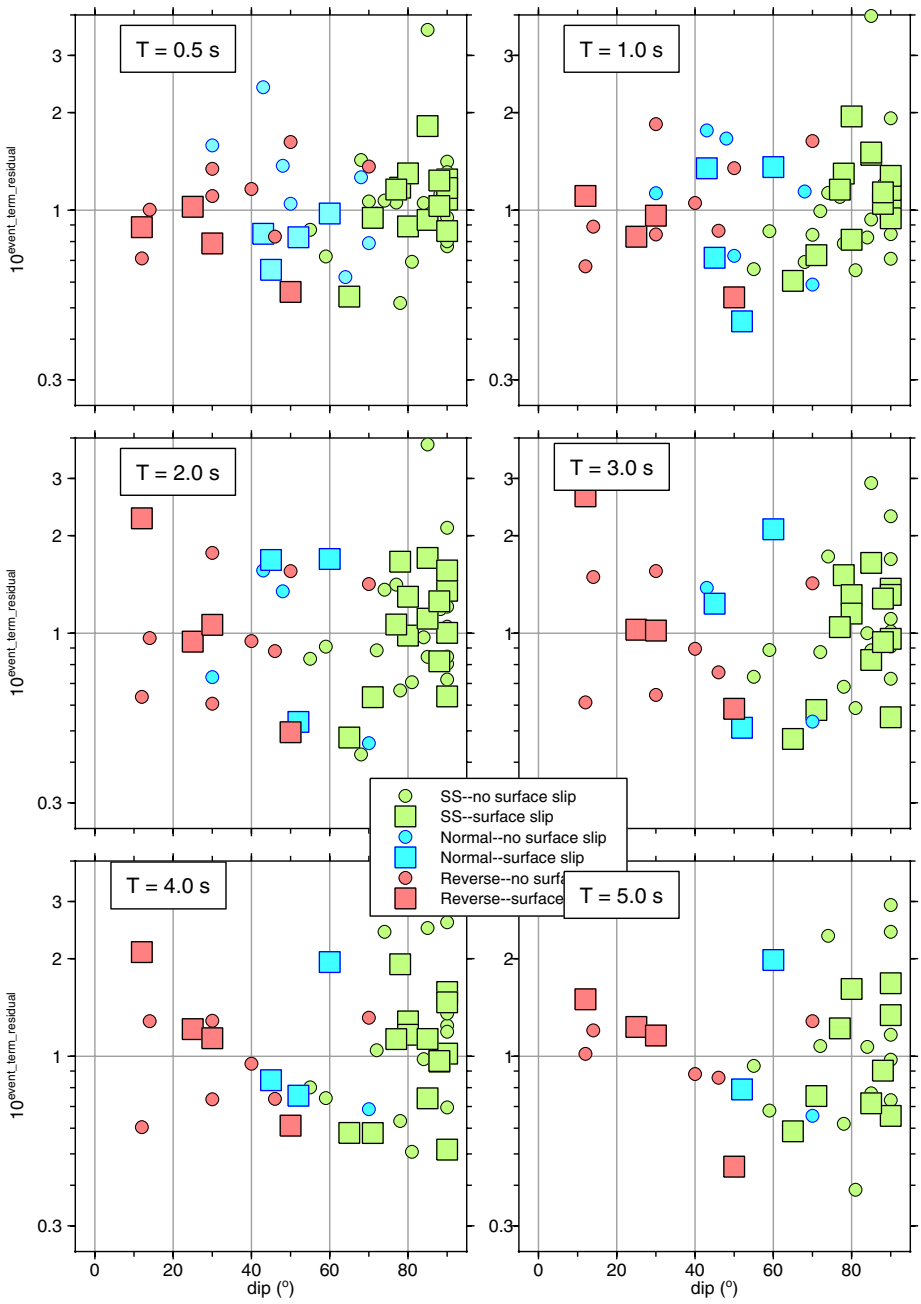
4.2.7 Dependence of Event Terms on Dip Angle

Figures 4.20(a)–(c) plot the event terms against dip angle. There are no obvious systematic effects of dip angle on the ground motion amplitudes.



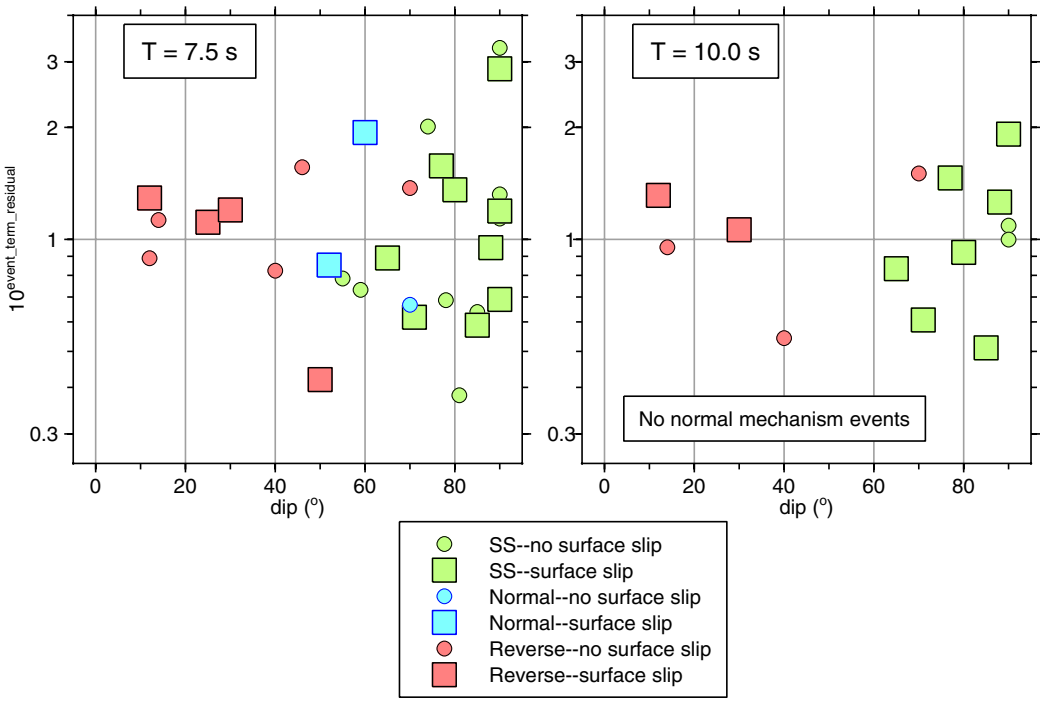
(a)

Fig. 4.20 Antilogarithms of event terms, plotted against dip and differentiated by events of different fault types, for which faults did or did not break to surface.



(b)

Fig. 4.20—*Continued*



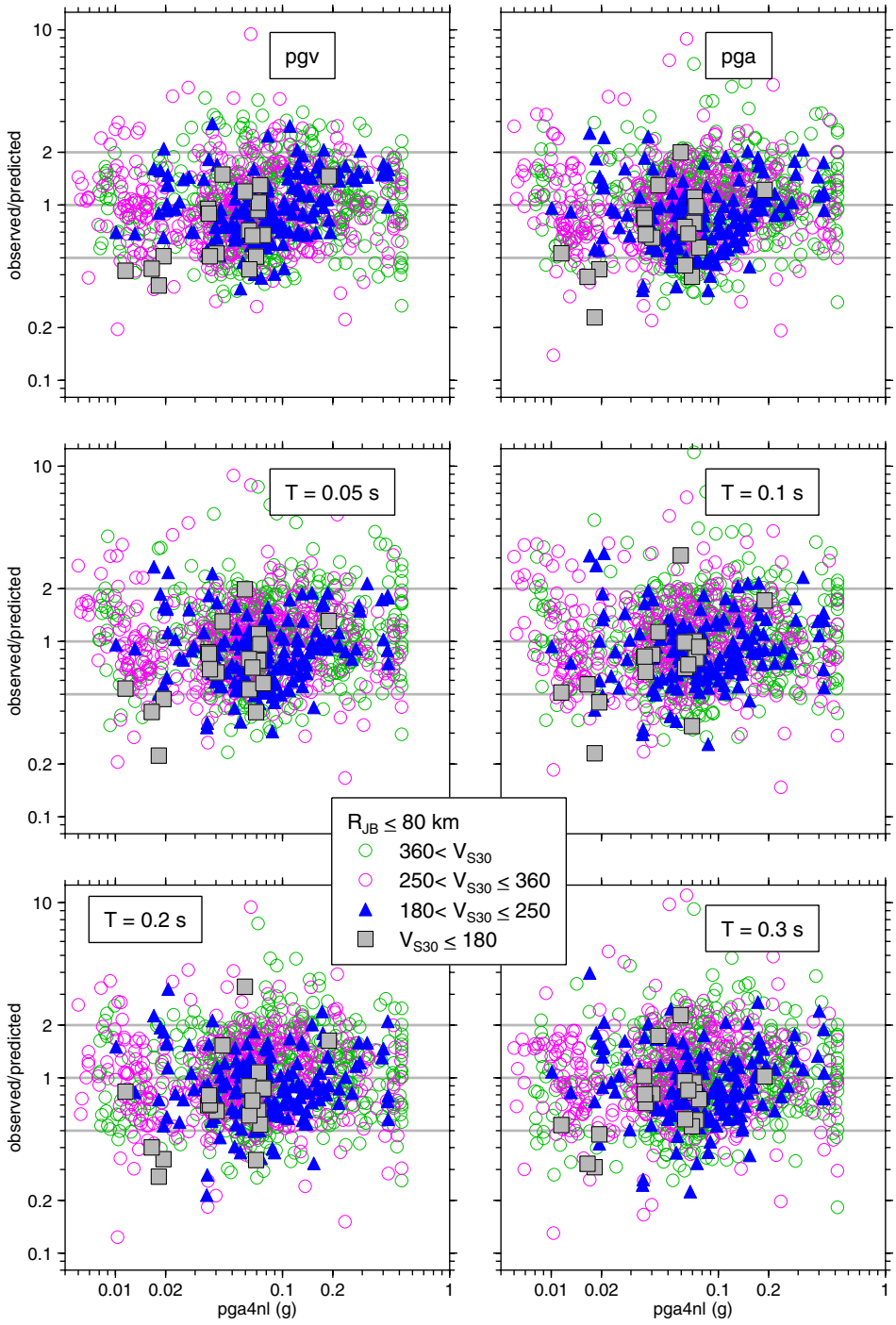
(c)

Fig. 4.20—*Continued*

4.2.8 Dependence of Stage 1 Residuals on *pga4nl*

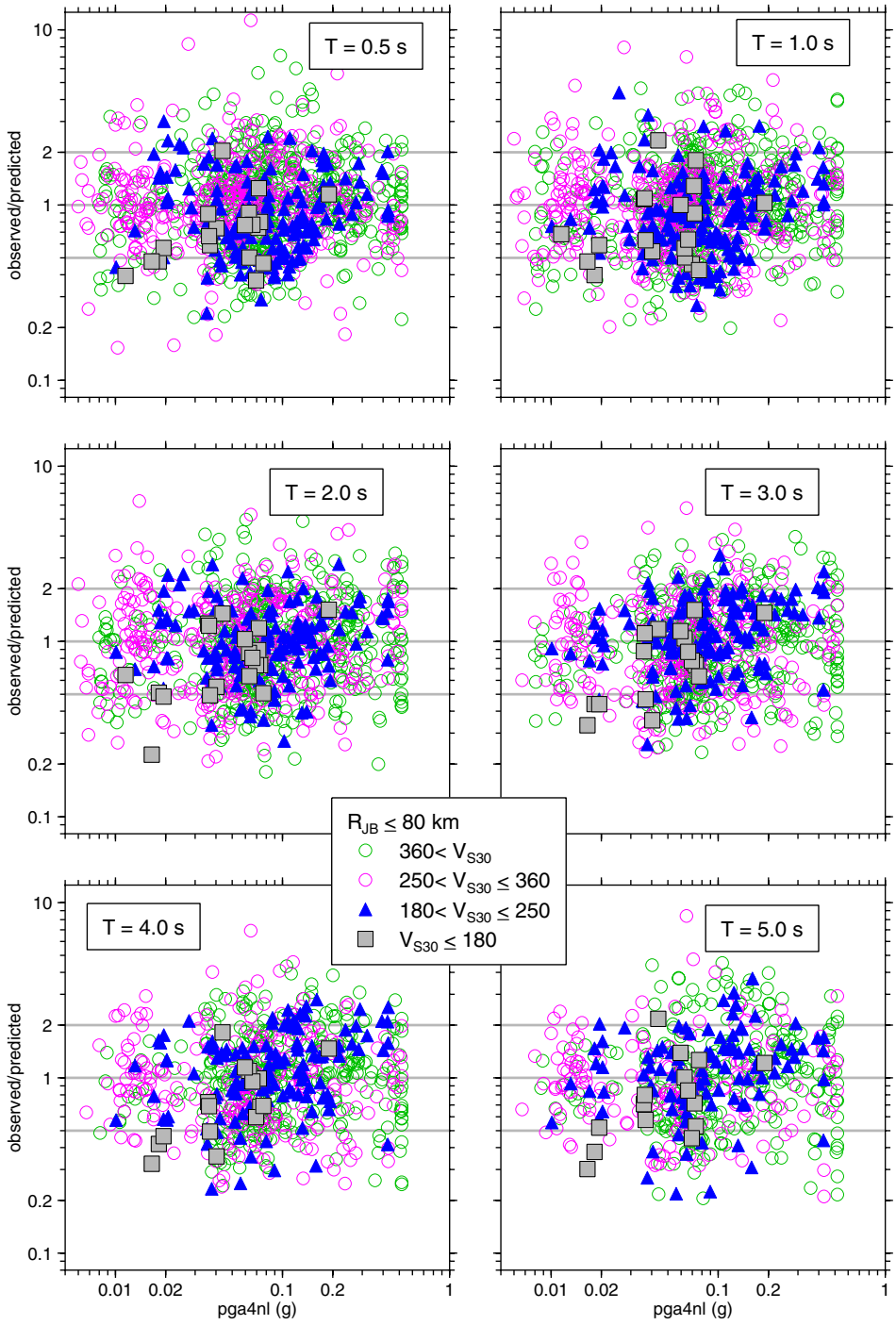
Our amplification factors were taken from an empirical analysis by Choi and Stewart (2005). There is a large overlap in the dataset they used and in the NGA flatfile, but there are also many small differences, including added data in the NGA flatfile and redeterminations of both ground motion intensity measures and V_{S30}. For these reasons it is important to search for systematic trends in the Stage 1 residuals for different ranges of V_{S30} and for different *pga4nl* (it is best to examine Stage 1 residuals for this analysis to exclude event-to-event variation that may obscure actual trends). This serves as a test of the effectiveness of the applied site-amplification factors in removing site effects. Figures 4.21(a)–(c) show the Stage 1 residuals plotted against *pga4nl*, with the residuals grouped into different bins of V_{S30} (NEHRP site classes). We split the NEHRP class D into two parts in order to see better any systematic trends in this class (about half the data come from this NEHRP class). In addition, we consider only data for $R_{JB} \leq 80$ km to avoid

mapping inadequacies in the distance function into systematic trends that might manifest themselves in the site-response functions. Figures 4.21(a)–(c) address the adequacy of the site-response functions by site class (which speaks mainly to the linear component of the correction) and by the strength of shaking (addressing the nonlinear component). Overall, the residuals do not depend on site class or the strength of ground shaking, with some significant apparent exceptions. In particular, the class E ($V_{S30} \leq 180$ m/s) residuals appear to be low (average residual <1) for most periods, indicating an overcorrection of the data to the reference velocity of 760 m/s. In addition, the data with 180 m/s $< V_{S30} \leq 250$ m/s appear to be systematically high for *pga4nl* > 0.1 g, but low for smaller values of *pga4nl*. This suggests that the nonlinear component of the site response may have been overestimated for soft sites (<250 m/s). Future work may be able to improve upon the soil-amplification factors.



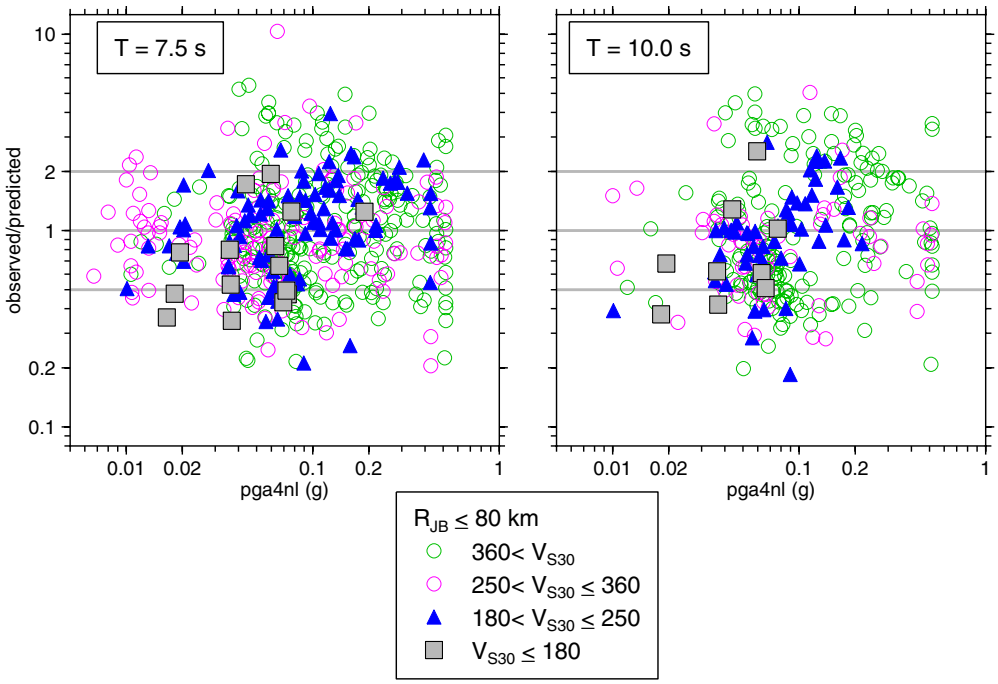
(a)

Fig. 4.21 Stage 1 residuals plotted against pga4nl, differentiated by V_{S30}, for $R_{JB} \leq 80$ km.



(b)

Fig. 4.21—*Continued*



(c)

Fig. 4.21—*Continued*

4.2.9 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), and from simulations (Day et al. 2005; Day et al. 2006). One of the reasons that we did not include a basin-depth term in our equations is indicated in Figure 4.22, which shows the distribution of V_{S30} and two measures of basin depth. The plot shows all data in the NGA flatfile for which both V_{S30} has been measured and basin depth has been estimated. It is clear that the softer sites are in basins, and hence basin depth and V_{S30} are strongly correlated. Therefore any basin depth effect will tend to 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. For example, one could use simulations to estimate the basin-depth effects and incorporate them into the GMPE model. 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.

Instead, our equations implicitly capture the basin-depth effect (attributing it to site amplification), provided that applications of our equations are in similar situations for which the data were recorded.

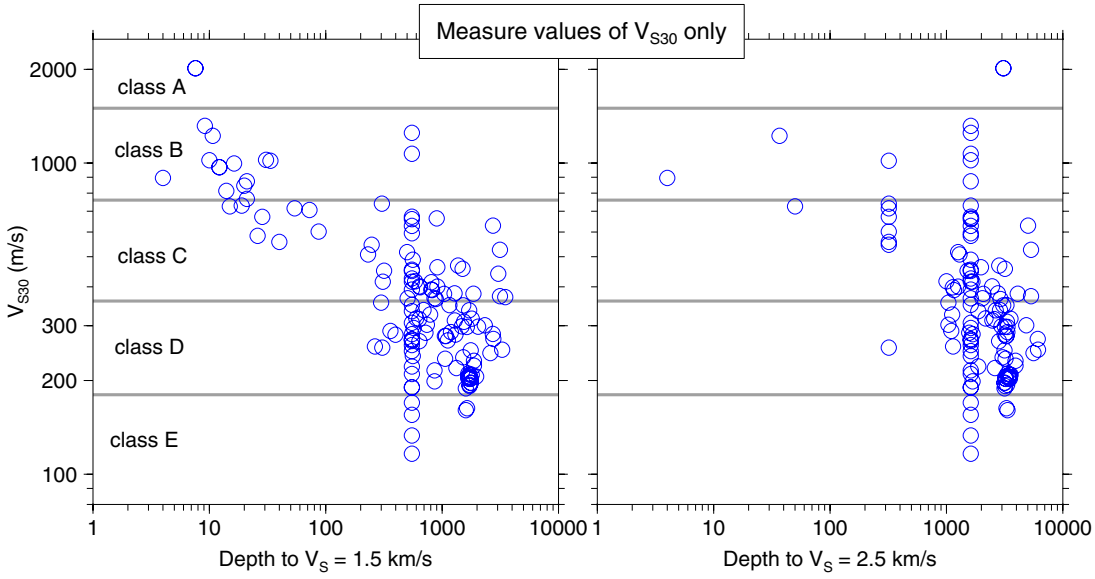


Fig. 4.22 V_{S30} plotted against two measures of basin depth. All values in NGA flatfile with basin depths and measured values of V_{S30} shown.

To show that there is no significant basin-depth effect that is not already captured implicitly via the site-amplification function, Figures 4.23–4.28 plot the Stage 1 residuals against the depth to $V_{S30} = 1.5$ km/s (plots are shown in pairs, for all distances and for $R_{JB} \leq 80$ km). There is no obvious dependence of the residuals on basin depth. 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 if the depth to 2.5 m/s was used as the measure of basin depth.)

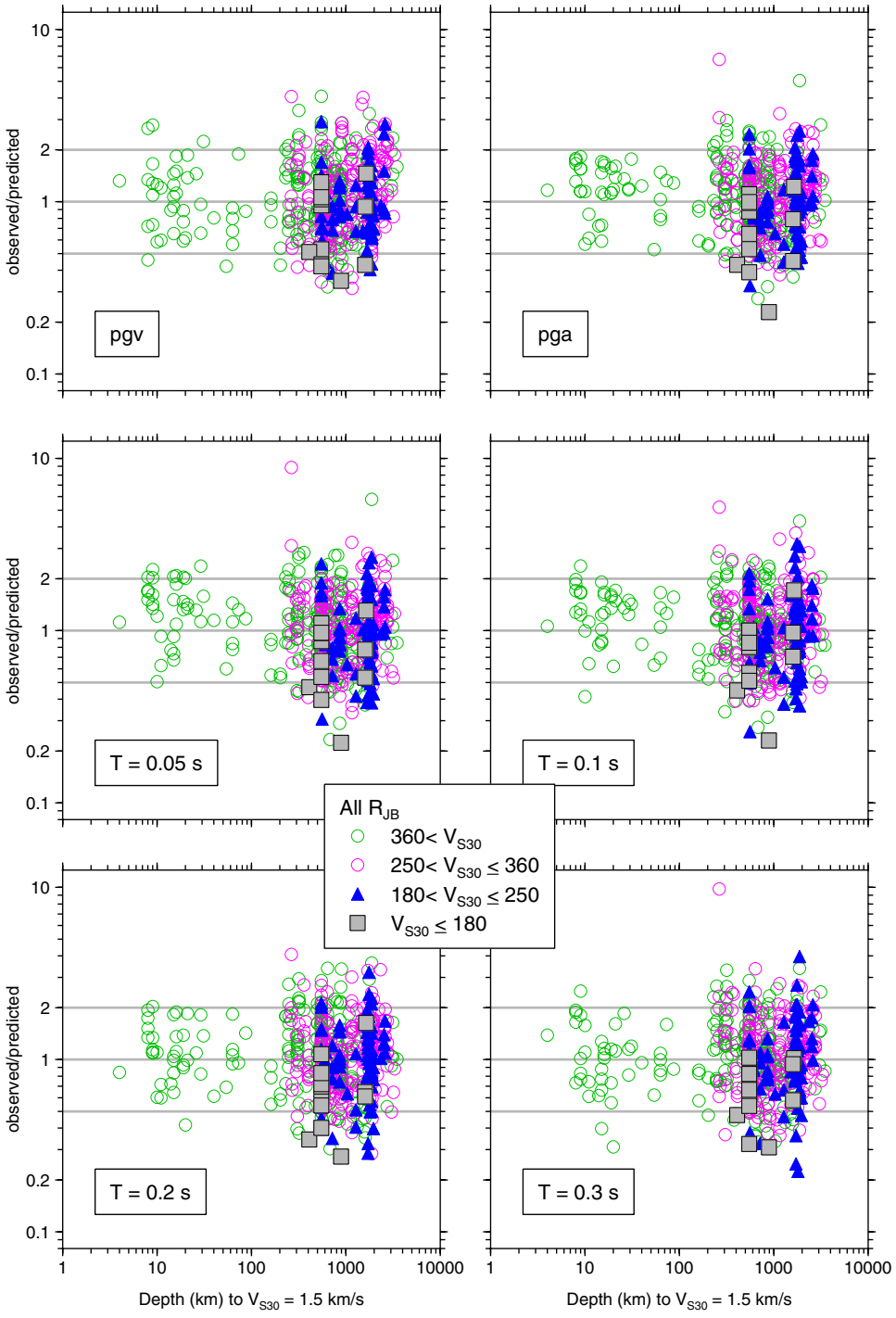


Fig. 4.23 Stage 1 residuals plotted against depth to $V_S = 1.5$ km/s, differentiated by V_{S30}, for all distances.

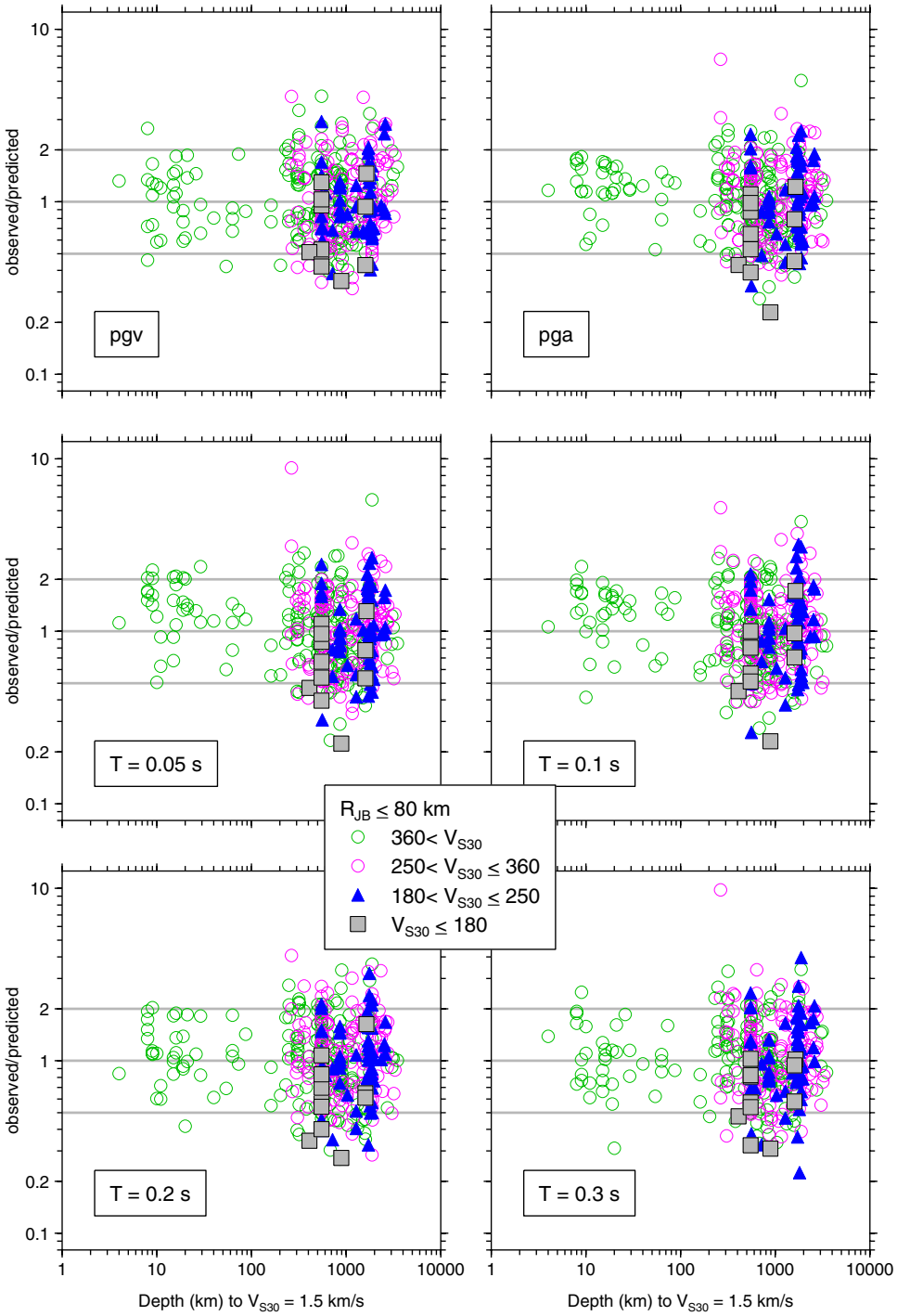


Fig. 4.24 Stage 1 residuals plotted against depth to $V_S = 1.5$ km/s, differentiated by V_{S30}, for $R_{JB} \leq 80$ km.

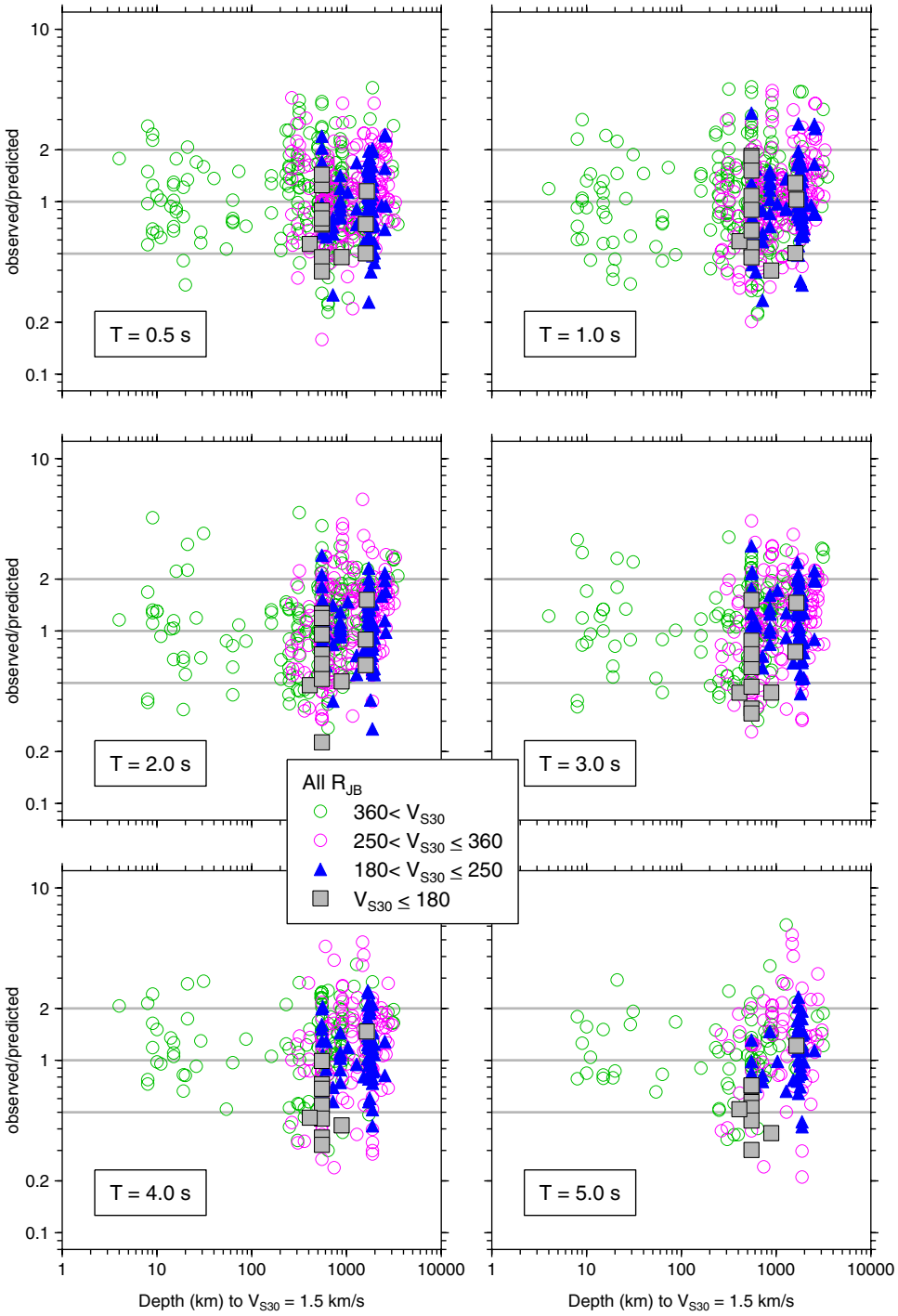


Fig. 4.25 Stage 1 residuals plotted against depth to $V_S = 1.5$ km/s, differentiated by V_{S30}, for all distances.

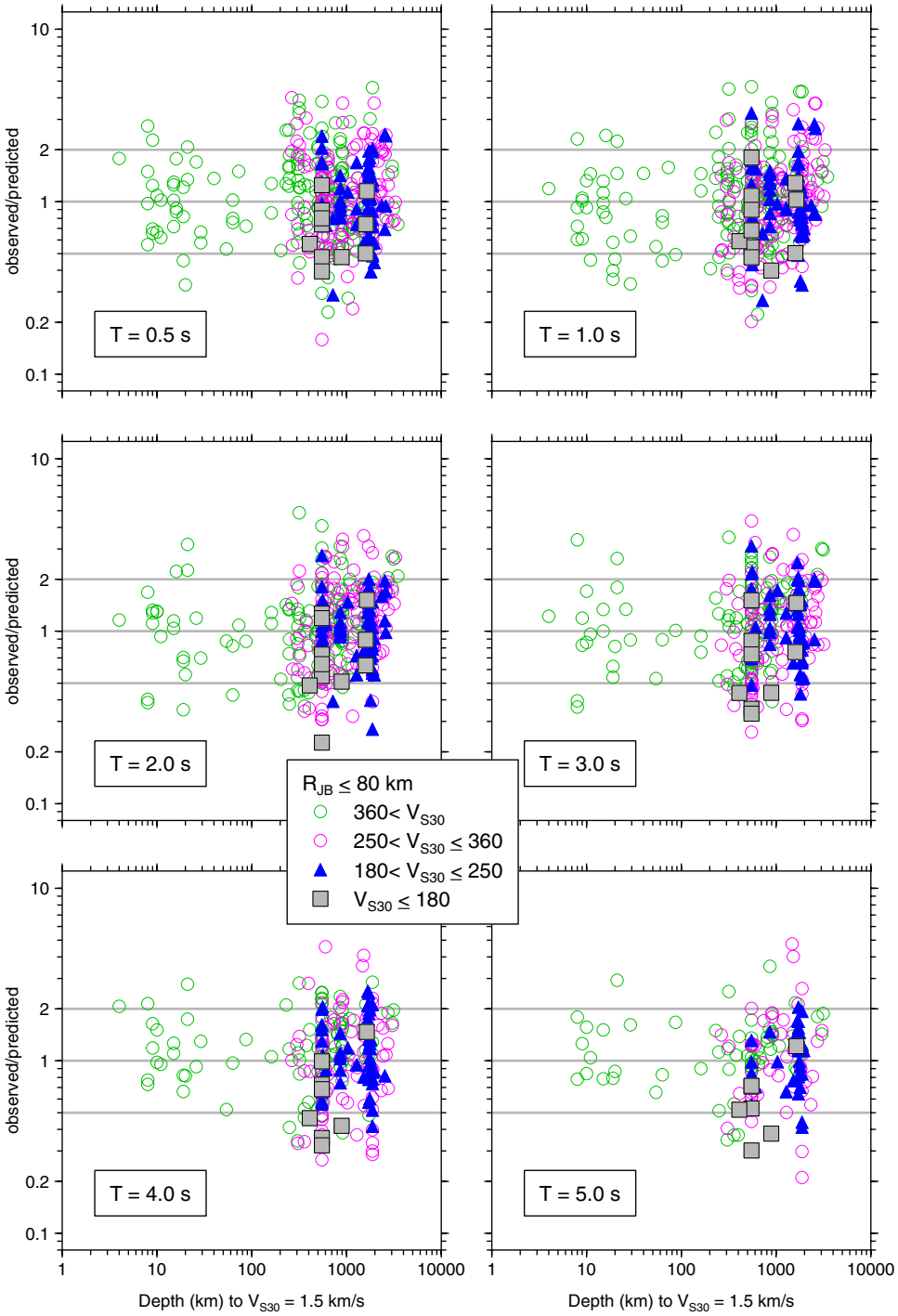


Fig. 4.26 Stage 1 residuals plotted against depth to $V_S = 1.5$ km/s, differentiated by V_{S30}, for $R_{JB} \leq 80$ km.

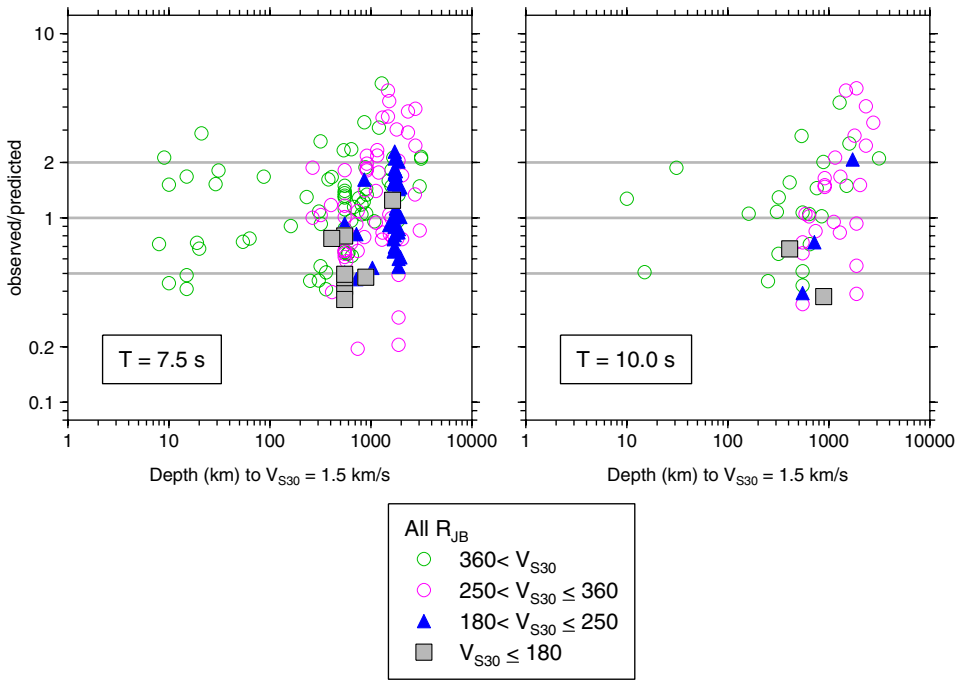


Fig. 4.27 Stage 1 residuals plotted against depth to $V_S = 1.5$ km/s, differentiated by V_{S30}, for all distances.

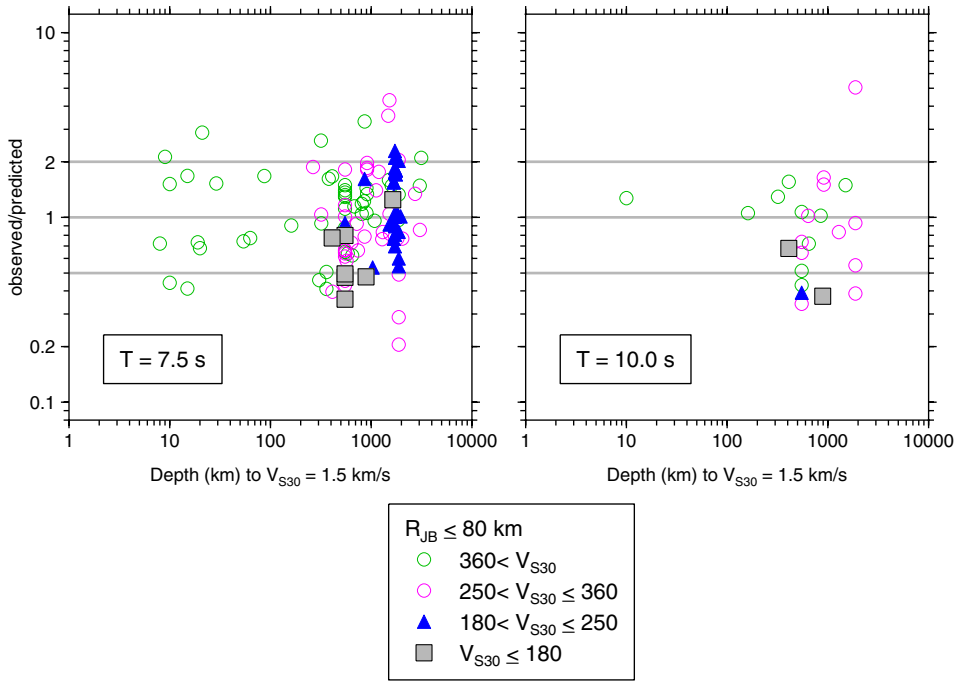


Fig. 4.28 Stage 1 residuals plotted against depth to $V_S = 1.5$ km/s, differentiated by V_{S30}, for $R_{JB} \leq 80$ km.

4.2.10 Comparison of GMPEs from One- and Two-Stage Regressions

We used the two-stage regression method discussed in Joyner and Boore (1993, 1994) as it was most computationally convenient; it has been previously shown that one-stage and two-stage regression methods (when properly implemented) are equivalent (Joyner and Boore 1993, 1994). Dr. John Douglas has implemented the Joyner and Boore one-stage regression scheme in Matlab. Prof. Sinan Akkar revised the program to accommodate our functional form and used it to compute regression coefficients. The ground motions predicted from our equations at a sample period (T= 5 s), based on the two-stage method, are compared with those from the equations developed from the one-stage regression (as implemented by Douglas and Akkar) in Figure 4.29. The comparison is excellent, which serves as a useful check on the regression results.

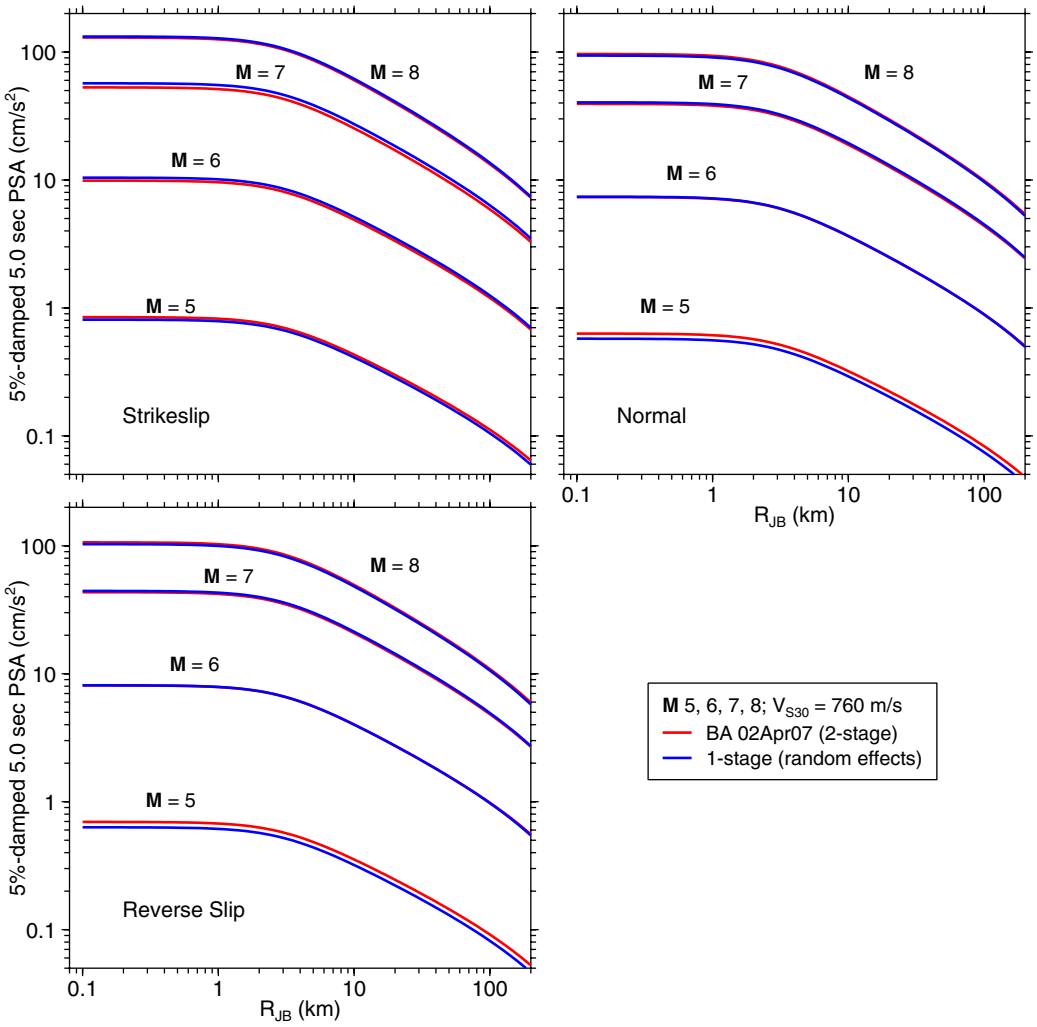


Fig. 4.29 Comparison of PSA from equations developed using one-stage weighted regression and two-stage regression (using algorithms of Joyner and Boore 1993, 1994).

4.2.11 Comparison of GMPEs Developed with and without 1999 Chi-Chi Earthquake

Because the Chi-Chi earthquake forms a significant fraction of the dataset 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 analysis without the Chi-Chi data. Figure 4.30 compares selected ground motion intensity measures given by the two sets of equations. The figures also show 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 dataset contributed by the Chi-Chi earthquake increases with period, reaching 64% of the dataset 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 4.30 are the same, to facilitate comparisons of the relations between the two predictions between periods). At intermediate to short periods, the differences are not dramatic, but are significant even at small magnitudes (despite the fact that we include only the Chi-Chi mainshock, not its aftershocks). We think this is because 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.

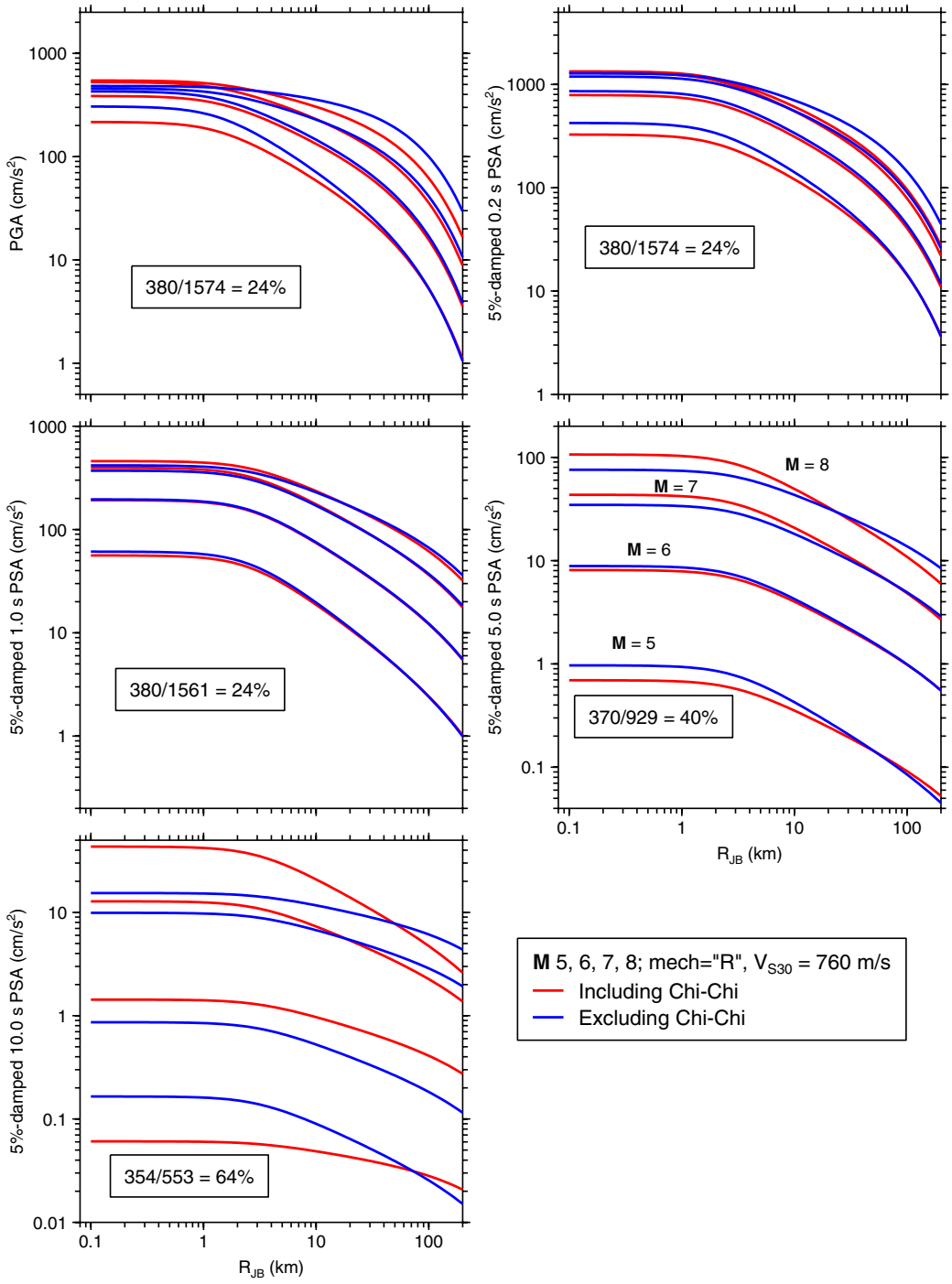


Fig. 4.30 Comparisons of PSA from equations developed with and without 1999 Chi-Chi mainshock. Seismic intensity measure is given by ordinate title. Ratios are number of Chi-Chi recordings used to develop final equations divided by total number of recordings.

4.2.12 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 4.31 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 dataset. For this reason, the new equations provide a more robust prediction of ground motion amplitudes over a wide range of magnitudes and distances.

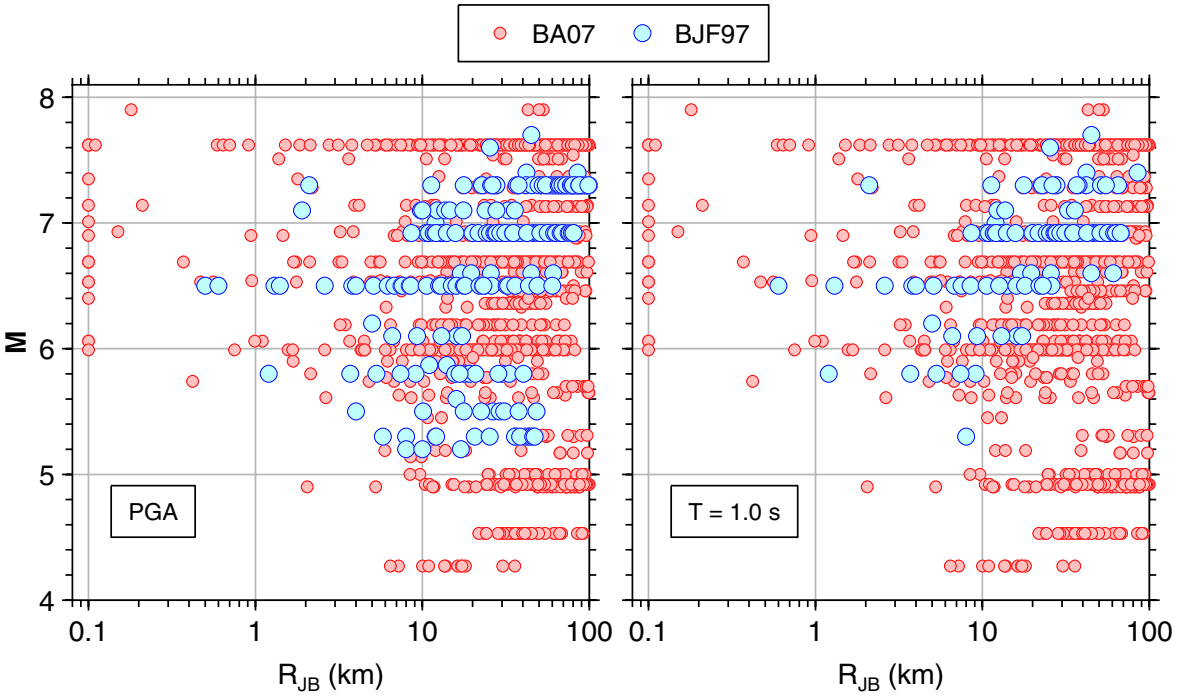


Fig. 4.31 Comparison of magnitude-distance distribution of data used by BJF97 and by us in developing GMPEs, for PGA and 1.0 s PSA.

We compare predicted ground motions from the BJF97 equations and from our current equations in Figure 4.32, 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 use the same scale for the ordinates in all 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.

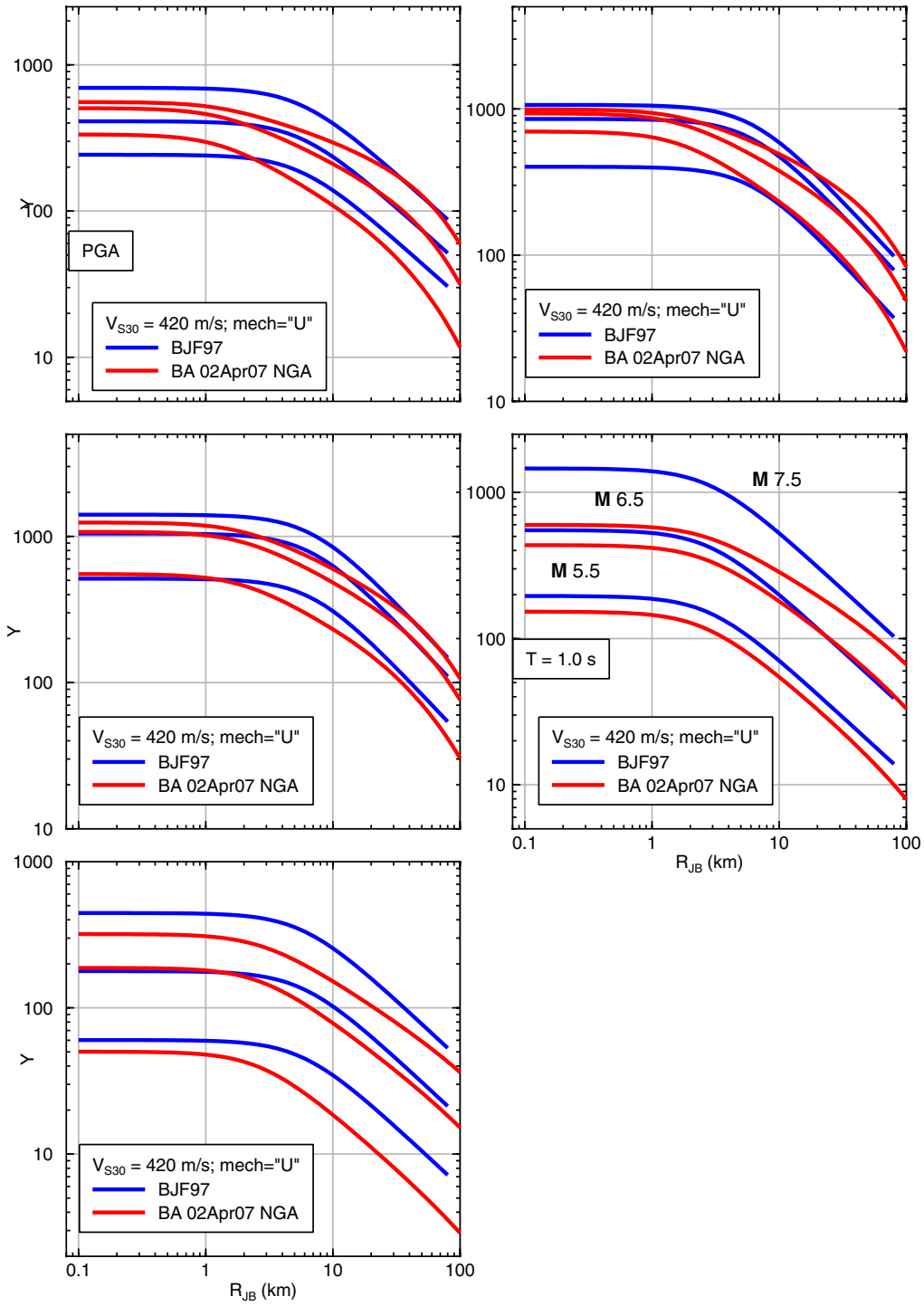


Fig. 4.32 Comparisons of PSA from BJF97 equations and our new equations.

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 $T = 1\,\text{s}$ and $\mathbf{M} = 7.5$. Almost no data were available in BJF97 for **M**~7.5 and $R_{JB} < 10\,\text{km}$ (see Fig. 4.31), so discrepancies 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 $R_{JB} < 10\,\text{km}$. Observed differences at $R_{JB} \approx 30\,\text{km}$ 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 $R_{JB} \approx 30\,\text{km}$ data). The current equations allow for the **M**-dependent distance scaling.

A comparison of aleatory uncertainties for the new equations and for the BJF97 equations is given in Table 4.6. Note that the total aleatory uncertainties, as well as the intra- and inter-event uncertainties are significantly larger for the new equations. The larger uncertainties will offset to some extent the smaller ground motions for large magnitudes in the construction of seismic hazard maps.

Table 4.6 Comparison of intra-event (σ), inter-event (τ), and total (σ_T) standard errors for BA 02Apr07 (mechanism specified) and Boore et al. (1997) equations. For latter, intra-event error is S1, which does not include component-to-component variation. This is appropriate in view of measure of ground motion intensity being used in this report.

per	σ: ba_02apr07	σ: bjf97	τ: ba_02apr07	τ: bjf97	σ_T: ba_02apr07	σ_T: bjf97
pga	0.502	0.431	0.260	0.184	0.564	0.469
0.10	0.520	0.440	0.318	0.000	0.608	0.440
0.20	0.523	0.435	0.288	0.009	0.596	0.435
0.30	0.546	0.440	0.269	0.048	0.608	0.443
0.50	0.555	0.454	0.265	0.115	0.615	0.468
1.00	0.573	0.474	0.302	0.214	0.647	0.520
2.00	0.580	0.495	0.389	0.276	0.700	0.567

4.2.13 Some Comments on Pseudo-Depth Variable *h*

There are correlations between many of the variables in the prediction equations; these correlations can affect the reliability of the GMPEs, especially when applied in magnitude-distance ranges with few data. A significant correlation issue involves the pseudo-depth variable *h*. Figure 4.33 shows *h* determined from the extended four-event dataset that we used to explore the anelastic attenuation coefficient, as determined under different assumptions about the geometrical spreading and anelastic distance-scaling coefficients. The largest values of *h* are for regressions assuming no anelastic attenuation, in which case the geometrical spreading term c_1 is large and negative (approximately -1.5 for the periods shown). On the other hand, when c_1 is constrained to be -1.0, -0.8, and -0.5, the values of *h* decrease systematically. Thus the value of *h* is linked to the geometric spreading term.

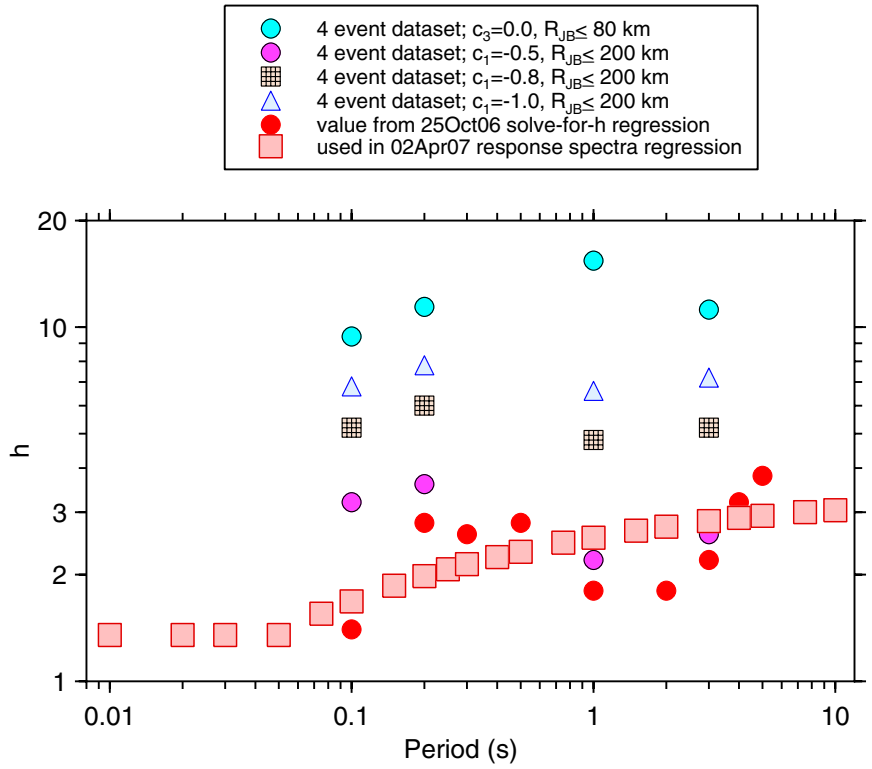


Fig. 4.33 Pseudo-depth variable from regressions on extended four-event dataset, compared with values from NGA dataset

In addition to the correlations between h, c_1, and c_3 just demonstrated, we notice an inconsistency between values of h that we determined from the four-event regressions and those subsequently determined from the regression on our subset of the entire NGA dataset. Recall that in order to constrain h to be a smooth and well-behaved function, we performed a regression of the NGA dataset in which h was a free parameter, but in which we constrained c_3; the constrained c_3 values were associated with $c_1 = -0.8$ in the four-event database that we used to study attenuation. We then defined a smooth function for h based on these results. From Figure 4.5, the effective geometrical spreading factor for our equations is somewhere between -0.5 and -0.8 for the magnitudes in the extended four-event dataset. The values of h for the four-event dataset, however, are larger than those obtained from regression of the NGA dataset (Fig. 4.8). At one NGA Developer's workshop, N. Abrahamson suggested that the relatively small value of h in our equations might be a consequence of not including a hanging wall/footwall term in our equations. It is possible that high values of ground motion on the hanging wall close to the fault outcrop might force small values of h. In future work we will investigate this possibility by determining h using only strike-slip earthquakes.

4.2.14 Comparison of BA07 and Data from Four Earthquakes Wholly or Largely Missing from NGA Flatfile

In this section the ground motion predictions from our equations (BA07) are checked with data wholly or largely missing from the NGA flatfile. These earthquakes, discussed before, are the 2001 Anza, 2002 Yorba Linda, 2003 Big Bear City, and 2004 Parkfield earthquakes. Figures 4.34–4.35 show the comparisons for the first three earthquakes and with the Parkfield earthquake, respectively. The comparisons are relatively good, but recall that the data in the figures were used in deriving the c_3 coefficient. It is worth remembering that the NGA dataset flatfile contains many fewer points for the three events than shown in Figure 4.34 (see Table 4.1) and no data from the Parkfield earthquake. For this reason, Figures 4.34–4.35 constitute a check of the predictions against data not used in the derivation of the equations (other than the coefficient c_3).

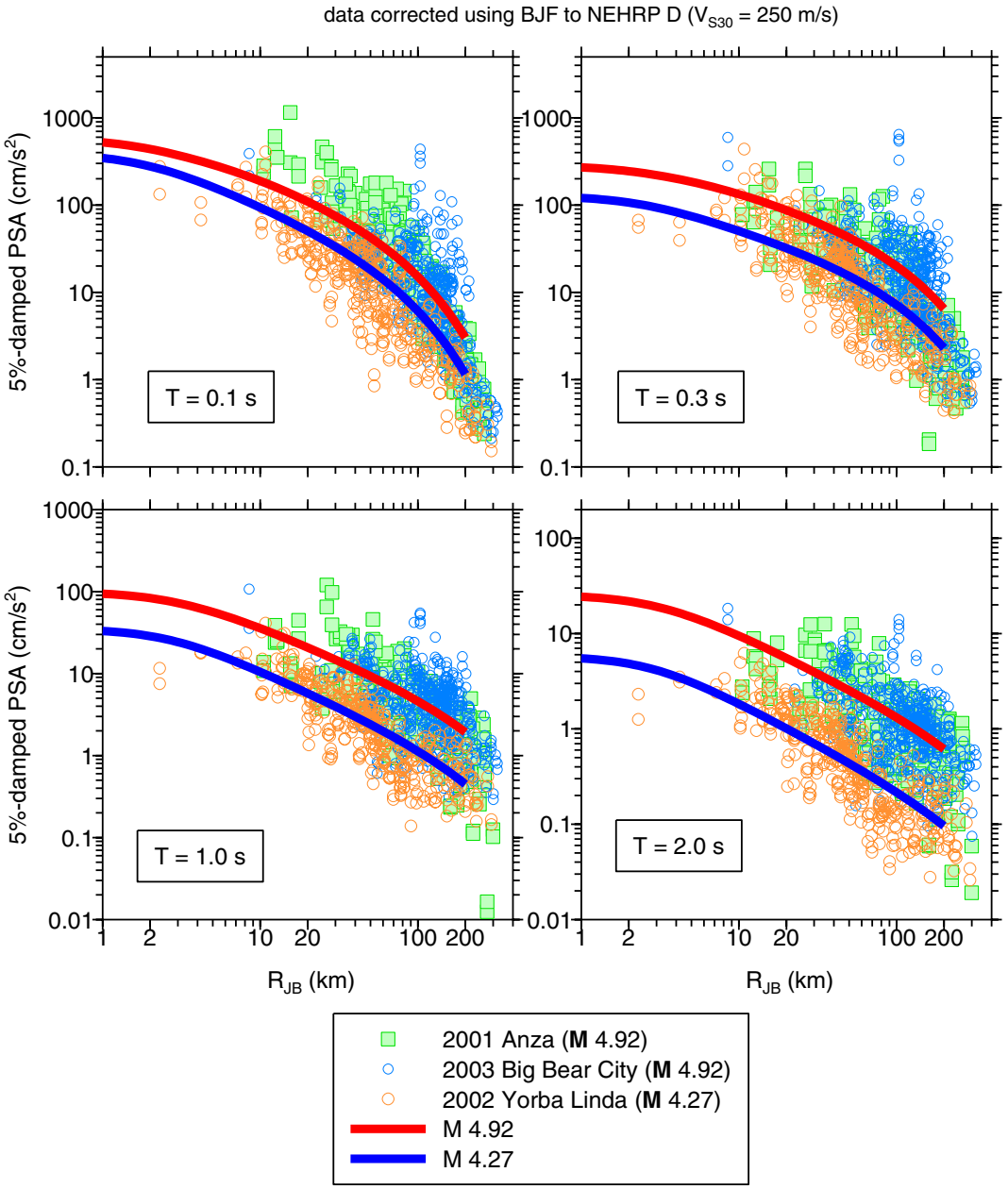


Fig. 4.34 Comparison of data from three small earthquakes with predictions from BA07 GMPEs.

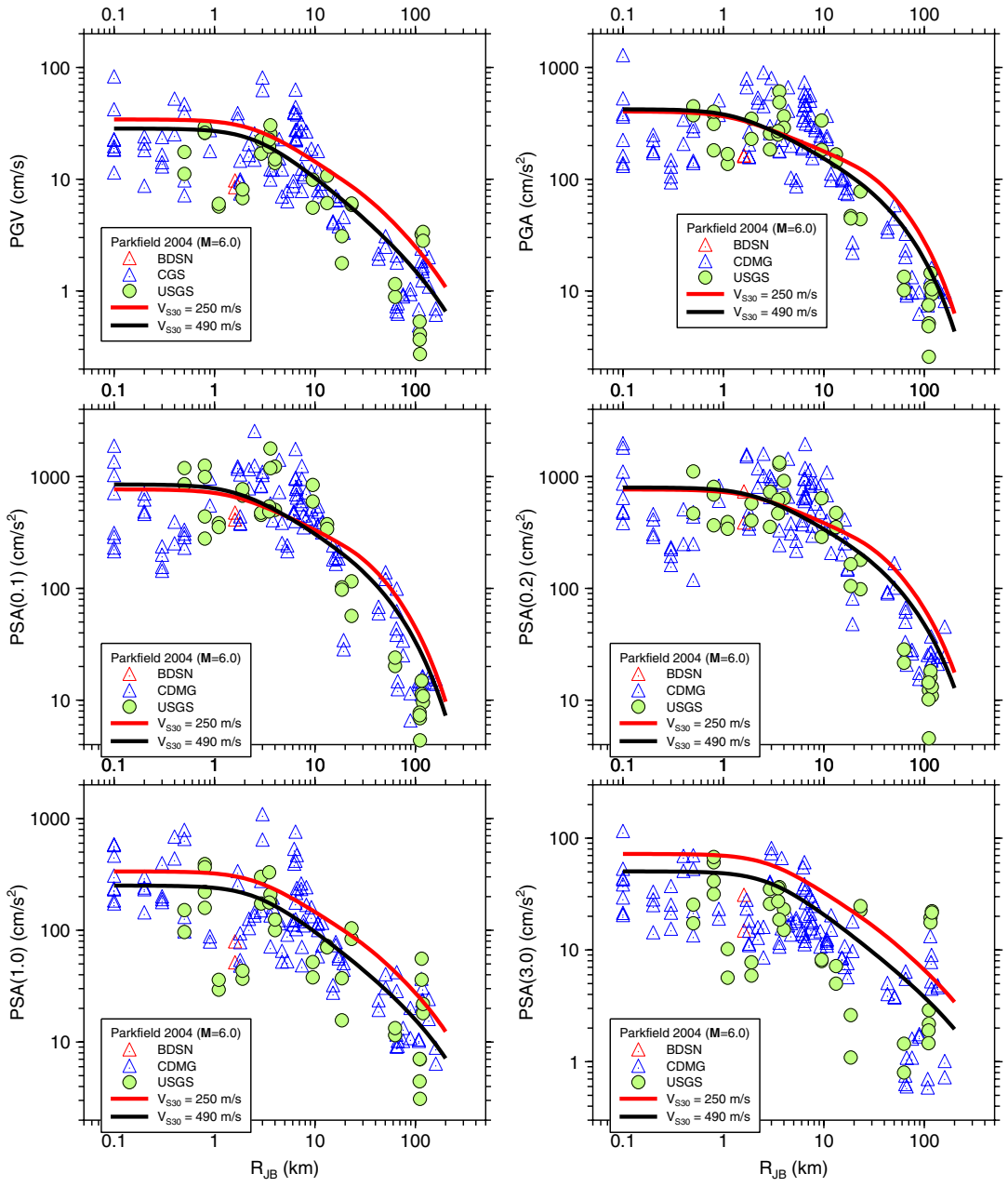


Fig. 4.35 Comparison of data from 2004 Parkfield earthquake with predictions from BA07 GMPEs.

4.2.15 Comparisons of PSA M-Scaling with Simulations

The most striking difference between previous GMPEs and the NGA GMPEs (not just those of BJF97 and BA07, but of all NGA developers) is the prediction of near- or complete saturation of PSA at short periods in the NGA equations (e.g., Figs. 4.16 and 4.17(a)). As a first step in understanding the reason for this saturation, Figure 4.36 shows the magnitude scaling at $R = 30$ km predicted by two point-source scaling models that were defined to mimic finite-fault effects: the Atkinson and Silva (2000) and the Joyner (1984) models. The simulations were made using the SMSIM program *tmr_rv_drvr* (Boore, 2000). For both source models we used the Raoof et al. (1999) distance attenuation parameters. Given that the motions predicted from SMSIM are simple point-source predictions of absolute motion from the source to the site (as opposed to motion relative to some particular distance, magnitude, period, or site condition), with no effort made to adjust parameters to match the motions predicted by the BA07 equations, the agreement is quite good. Adjustments of the geometrical spreading and Q functions, as well as the site amplifications (which used the generic rock amplifications of Boore and Joyner, 1997), could be made to provide better agreement between the BA07 and the simulated ground motions. Finite-fault stochastic-simulation models (Motazedian and Atkinson 2005), or more detailed broadband-simulation methods (Hartzell et al. 1999) could also be applied to better understand the observed scaling. It is interesting, though, that the scaling of motions for large magnitudes seems to fall between that predicted by two simple point-source models.

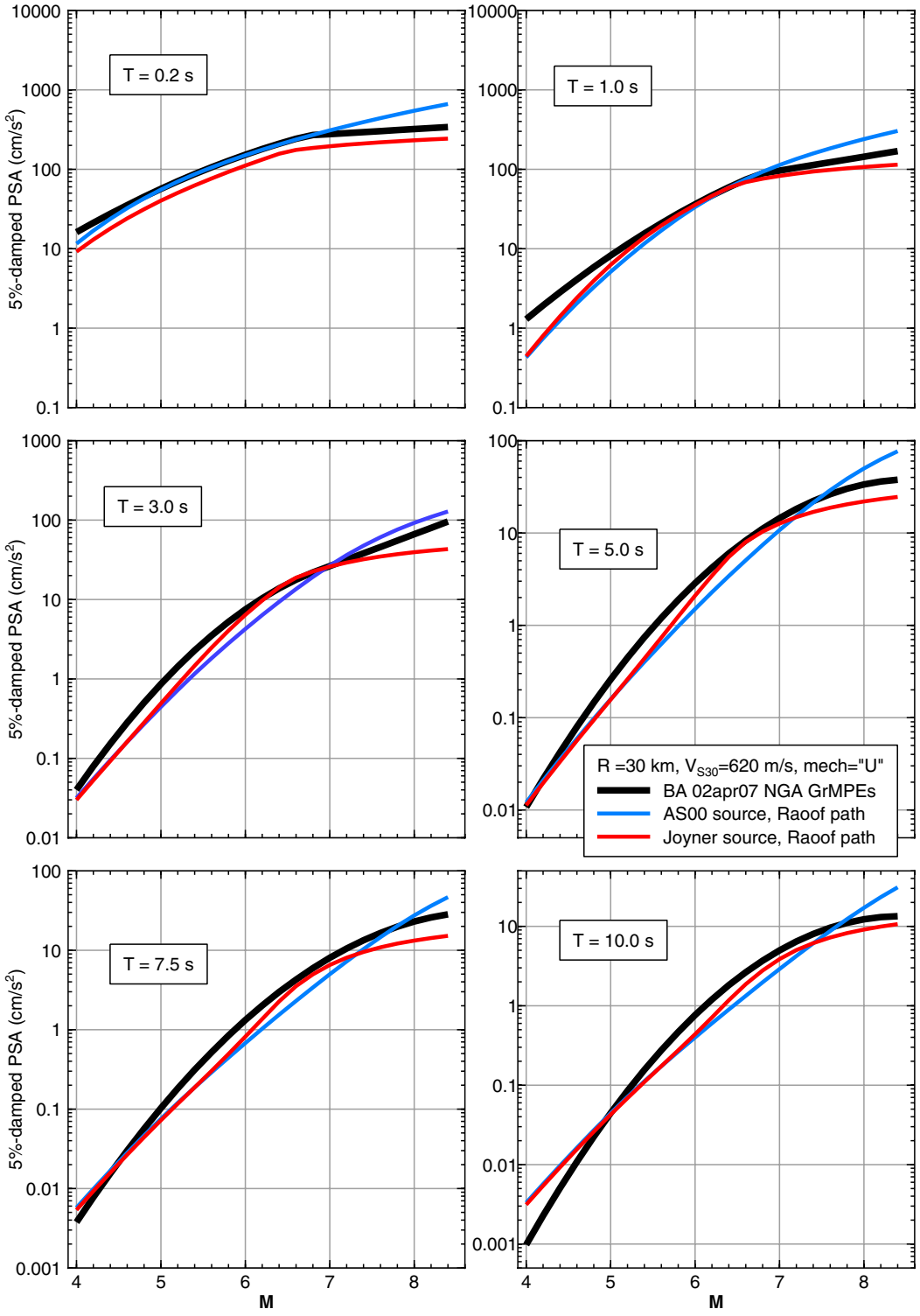


Fig. 4.36 Observed and predicted scaling with magnitude at $R = 30$ km.

5 Guidelines for Usage

5.1 LIMITS ON PREDICTOR VARIABLES

Although we know perfectly well that people will ignore the following limits for the predictor variables, for the record we state that our equations should be used only for predictor variables in these ranges:

- **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.

5.2 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 (this volume).

6 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 if these can be unambiguously supported by data. Many ground motion observations that have not been included in the NGA flatfile, or should be reprocessed in an attempt to provide more data at long periods; additional 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 potential 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 dataset. 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.

REFERENCES

Ambraseys, N. N., J. Douglas, S. K. Sarma, and P. M. Smit (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 Engineering* **3:** 1–53.

Atkinson, G. M. (1993). Earthquake source spectra in eastern North America. *Bull. Seism. Soc. Am.* **83:** 1778–98.

Bakun, W. H., and W. B. Joyner (1984). The M_L scale in central California. *Bull. Seism. Soc. Am.* **74:** 1827–43.

Atkinson, G. M., and D. M. Boore (2003). Empirical ground-motion relations for subduction zone earthquakes and their application to Cascadia and other regions. *Bull. Seism. Soc. Am.* **93:** 1703–29.

Bakun, W. H., and W. B. Joyner (1984). The M_L scale in central California. *Bull. Seism. Soc. Am.* **74:** 1827–43.

Benz, H. M., A. Frankel, and D. M. Boore (1997). Regional *Lg* attenuation for the continental United States. *Bull. Seism. Soc. Am.* **87:** 606–19.

Beyer, K., and J. J. Bommer (2006). Relationships between median values and between aleatory variabilities for different definitions of the horizontal component of motion. *Bull. Seism. Soc. Am.* **96**: 1512–22.

Boatwright, J., H., Bundock, J. Luetgert, L. Seekins, L. Gee, and P. Lombard (2003). The dependence of PGA and PGV on distance and magnitude inferred from northern California ShakeMap data. *Bull. Seism. Soc. Am.* **93:** 2043–55.

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

Bommer, J. J., J. Douglas and F. O. Strasser (2003). Style-of-faulting in ground-motion prediction equations. *Bull. Seism. Soc. Am.* **93:** 171–203.

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

Boore, D. M. (2000). SMSIM–Fortran programs for simulating ground motions from earthquakes: version 2.0 — A revision of OFR 96-80-A. *U. S. Geological Survey Open-File Report* **00-509**, 55 pp.

Boore, D. M. (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. (2005). On pads and filters: Processing strong-motion data. *Bull. Seism. Soc. Am.* **95:** 745–50.

Boore, D. M., and S. Akkar (2003). Effect of causal and acausal filters on elastic and inelastic response spectra. *Earthquake Engineering and Structural Dynamics* **32:** 1729–48.

Boore, D. M., and G. M. Atkinson (1989). Spectral scaling of the 1985 to 1988 Nahanni, Northwest Territories, earthquakes. *Bull. Seism. Soc. Am.* **79:** 1736–61.

Boore, D. M., and W. B. Joyner (1997). Site amplifications for generic rock sites. *Bull. Seism. Soc. Am.* **87**: 327–41.

Boore, D. M., W. B. Joyner, and T. E. Fumal (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.

Boore, D. M., W. B. Joyner, and T. E. Fumal (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.

Boore, D. M., W. B. Joyner, and T. E. Fumal (1997). Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work. *Seism. Research Letters* **68:** 128–53.

Boore, D. M., J. Watson-Lamprey, and N. A. Abrahamson (2006). GMRotD and GMRotI: Orientation-independent measures of ground motion. *Bull. Seism. Soc. Am.* **96:** 1502–11.

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

Day, S. M., J. Bielak, D. Dreger, R. Graves, S. Larsen, K. Olsen, A. Pitarka (2005). *3D ground motion simulation in basins.* Final report prepared for the Pacific Earthquake Engineering Research Center, Project 1A03, by San Diego State University, Carnegie-Mellon University, University of California at Berkeley, Lawrence Livermore National Laboratory, and URS Corporation.

Day, S. M., J. Bielak, D. Dreger, R. Graves, S. Larsen, K. Olsen, A. Pitarka, and L. Ramirez-Guzman (2006). Numerical simulation of basin effects on long-period ground motion. In *Proceedings, Eighth National Conference on Earthquake Engineering,* Paper No. 1857.

Ellsworth, W. L., M. Celebi, M., J. R. Evans, E. G. Jensen, R. Kayen, M. C. Metz, D. J. Nyman, J. W. Roddick, P. Spudich, and C. D. Stephens (2004). Near-field ground motions of the 2002, Denali Fault, Alaska, earthquake recorded at Pump Station 10. *Earthquake Spectra* **20:** 597–615.

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. Seism. Soc. Am.* **90:** S209–21.

Frohlich, C., and K. D. Apperson (1992). Earthquake focal mechanisms, moment tensors, and the consistency of seismic activity near plate boundaries. *Tectonics* **11:** 279–96.

Hartzell, S., S. Harmsen, A. Frankel, and S. Larsen (1999). Calculation of broadband time histories of ground motion: Comparison of methods and validation using strong-ground motion from the 1994 Northridge earthquake. *Bull. Seism. Soc. Am.* **89:** 1484–1504.

Hutton, L. K., and D. M. Boore (1987). The M_L scale in southern California. *Bull. Seism. Soc. Am.* **77:** 2074–94.

Joyner, W. B. (1984). A scaling law for the spectra of large earthquakes. *Bull. Seism. Soc. Am.* **74:** 1167–88.

Joyner, W. B. and D. M. Boore (1993). Methods for regression analysis of strong-motion data, *Bull. Seism. Soc. Am.* **83:** 469–87.

Joyner, W. B., and D. M. Boore (1994). Errata, *Bull. Seism. Soc. Am.* **84:** 955–56.

Komatitsch, D., Q. Liu, J. Tromp, P. Süss, C. Stidham, and J. H. Shaw (2004). Simulations of ground motion in the Los Angeles Basin based upon the Spectral-Element Method. *Bull. Seism. Soc. Am.* **94**: 187–206.

Lee, W. H. K., T. C. Shin, K. W. Kuo, K. C. Chen, and C. F. Wu (2001). CWB free-field strong-motion data from the 21 September Chi-Chi, Taiwan, earthquake. *Bull. Seism. Soc. Am.* 91: 1370–76.

Liu, Q., J. Polet, D. Komatitsch, and J. Tromp (2004). Spectral-element moment tensor inversions for earthquakes. *Bull. Seism. Soc. Am.* **94:** 1748–61.

Mori, J., and D. Helmberger (1996). Large-amplitude Moho reflections (SmS) from Landers aftershocks, southern California. *Bull. Seism. Soc. Am.* **86**: 1845–52.

Motazedian, D., and G. Atkinson (2005). Stochastic finite fault modeling based on a dynamic corner frequency. *Bull. Seism. Soc. Am.* **95:** 995–1010.

Oglesby, D. D., and S. M. Day (2001). The effect of fault geometry on the 1999 Chi-Chi (Taiwan) earthquake. *Geophysical Research Letters* **28:** 1831–34.

Power, M., B. Chiou, N. Abrahamson, and C. Roblee (2006). The "Next Generation of Ground Motion Attenuation Models" (NGA) project: An overview. In *Proceedings, Eighth National Conference on Earthquake Engineering,* Paper No. 2022.

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

Rathje, E. M., K. H. Stokoe II, and B. Rosenblad (2003). Strong motion station characterization and site effects during the 1999 earthquakes in Turkey. *Earthquake Spectra* **19:** 653–75 (for more details, see http://www.ce.utexas.edu/prof/rathje/research/turkey_velocity.html).

Rathje, E. M., K. H. Stokoe II, and B. Rosenblad (2004). Erratum: ''Strong motion station characterization and site effects during the 1999 earthquakes in Turkey''[Earthquake Spectra 19: 653–75 (2003)], *Earthquake Spectra* **20:** 283.

Schmedes, J., and R. J. Archuleta (2007). Oversaturation of peak ground velocity near strike slip faults (abs). *Seismological Research Letters* **78:** 272.

Somerville, P., and A. Pitarka (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.

Trifunac, M. D., M. I. Todorovska, and V. W. Lee (1998). The Rinaldi strong motion accelerogram of the Northridge, California earthquake of 17 January 1994. *Earthquake Spectra* **14:** 225–39.

Watson-Lamprey, J. A., and D. M. Boore (2007). Beyond *SaGMRotI*: Conversion to Sa_{Arb}, Sa_{SN}, and Sa_{MaxRot}. *Bull. Seism. Soc. Am.* **97:** (in press).

Youngs, R. (2005). Estimation of distance and geometry measures for earthquakes without finite rupture models, unpublished notes, dated December 8, 2005, distributed to PEER NGA developers.

Zoback, M. L. (1992). First- and second-order patterns of stress in the lithosphere: The World Stress Map Project. *J. Geophysical Research* **97:** 11,703–28.