NGA Project Strong-Motion Database

Brian Chiou,[a] **Robert Darragh,**[b] M.EERI, **Nick Gregor,**[b] M.EERI, **and Walter Silva,**[b] M.EERI

A key component of the NGA research project was the development of a strong-motion database with improved quality and content that could be used for ground-motion research as well as for engineering practice. Development of the NGA database was executed through the Lifelines program of the PEER Center with contributions from several research organizations and many individuals in the engineering and seismological communities. Currently, the data set consists of 3551 publicly available multi-component records from 173 shallow crustal earthquakes, ranging in magnitude from 4.2 to 7.9. Each acceleration time series has been corrected and filtered, and pseudo absolute spectral acceleration at multiple damping levels has been computed for each of the 3 components of the acceleration time series. The lowest limit of usable spectral frequency was determined based on the type of filter and the filter corner frequency. For NGA model development, the two horizontal acceleration components were further rotated to form the orientation-independent measure of horizontal ground motion (GMRotI50). In addition to the ground-motion parameters, a large and comprehensive list of metadata characterizing the recording conditions of each record was also developed. NGA data have been systematically checked and reviewed by experts and NGA developers. [DOI: 10.1193/1.2894831]

INTRODUCTION

A high-quality ground-motion database was the key to the success of NGA model development. In the early stage of the NGA project developers embraced the idea of a common strong-motion database, from which individual records could be selected or excluded at the discretion of each developer. There were two motivations for developing a common database. First, a common database could potentially reduce unwarranted model-to-model variation. It has been recognized that some of the noted differences between previous attenuation models can be attributed to the unintentional differences in data selection. Such differences also made it difficult to conduct a fair and systematic comparison of attenuation models. Secondly, developing a high-quality database that meets the NGA project scope was a challenging task which demanded close collaborations between seismologists and engineers. A common database fostered such collaboration and therefore increased the chance of achieving the goals of the project.

The scope of the database development was drafted by the database development

[a] California Department of Transportation, Office of Infrastructure Research, 5900 Folsom Blvd., MS-5, Sacramento, CA 95819
[b] Pacific Engineering and Analysis, 856 Seaview Drive, El Cerrito, CA 94530

Earthquake Spectra, Volume 24, No. 1, pages 23–44, February 2008; © 2008, Earthquake Engineering Research Institute

team and reviewed by a large group of participants at the NGA kickoff meeting on October 24, 2002 and two subsequent NGA workshops in 2003. The final scope covered a wide range of information, including many predictor variables judged by NGA developers and other researchers to be worthy of consideration for the model development and other future research.

This paper is to serve as a general reference for the NGA database. The first part of this paper gives an overview of the NGA database development. The flat file document (Appendix A, on line) provides the basic documentation of the NGA data. The second part of this paper documents in more detail the strong-motion record processing, finite fault models, and site conditions, all of which are essential to the NGA model development.

NGA DATABASE

The NGA database was built on the existing PEER Strong-Motion Database (http://peer.berkeley.edu/smcat/), which was developed by Pacific Engineering & Analysis (Pacific Engineering) during the 1990s. The goals of the NGA database development were straightforward, (1) fill the data gap in existing PEER database with recent earthquakes, and (2) greatly expand the supporting information (metadata) to meet the needs of the ambitious scope of model development set by the NGA project (Power et al., 2008, this volume). Another significant effort of the database development is the systematic reviews and checking of collected data. When conflicts were identified, resolution typically involved data review and as appropriate soliciting comments from NGA developers and other researchers. The efforts of updating and expanding the PEER database began in 2003. By the end of 2004 the main database was completed. The subsequent efforts focused on reviews, errata reports, and limited updates as information, such as V_{S30} (average shear-wave velocity to 30-meter depth) of strong-motion stations in Taiwan and California, became available. Many individuals and agencies contributed data and expertise during the course of database development. In several cases the data were made available to NGA project prior to publication.

The NGA data collection included acceleration time series of the multi-component recordings and supporting information about the earthquakes and recordings. Time series were stored on a hard disk as text files. Each acceleration time series has been reviewed by Pacific Engineering in a consistent manner, as will be discussed later. Compatible velocity and displacement time series as well as response spectra at multiple spectral damping levels were calculated and also stored as text files, all in a common format. For the development of NGA models, the two horizontal acceleration components were further rotated to form the orientation-independent measure of horizontal ground motion (GMRotI50, Boore et al., 2006). A record was also rotated to the strike-normal and strike-parallel directions, if the strong-motion sensor orientation and strike-parallel direction were known.

Metadata were stored in 4 tables: record catalog, earthquake source, strong-motion station, and propagation path. To facilitate the collection, maintenance, and query of metadata, we implemented a simple data linkage model to relate the tables. The record

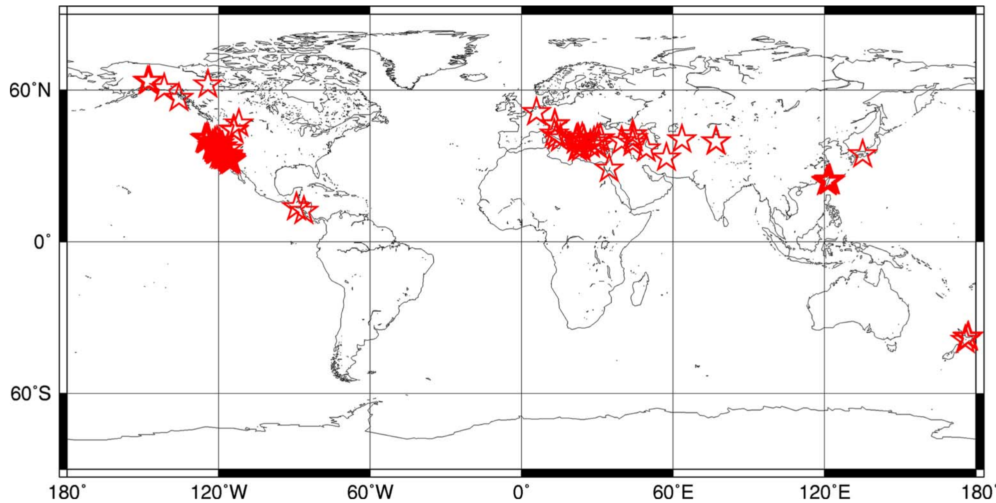

Figure 1. Map showing the epicenter distribution of the 173 earthquakes in the NGA database.

catalog is the master table as it holds the list of available strong-motion records and, for each record, the directory and file names of time series and response spectra. It also holds the unique identifiers of each record, earthquake, and strong-motion station from which a record was collected. These identifiers served as the primary keys linking each record to supporting information stored in the other three tables. Earthquake identification was created according to the order by which earthquakes were entered into the catalog, not by the chronological order of earthquake occurrence. The same ordering criterion was applied to record identification and station identification. If a strong-motion station was moved more than 100 meters from its original location we followed California Strong Motion Instrumentation program (CSMIP) practice and assigned a new station identification to it. Examples of such stations are Tarzana—Cedar Hills and Lake Hughes #12.

Earthquake identification number ranges from 0001 to 0175, but two earthquakes (ID 0146, the 1992 Joshua Tree, California, earthquake; and ID 0158, the 2000 Loma Linda, California, earthquake) were not populated with data due to time and resource constraints, making the total number of earthquakes in the NGA database 173. A map showing the geographic distribution of earthquakes is given in Figure 1. The type of earthquake targeted by NGA was shallow crustal earthquake from active tectonic regions world wide, but with a focus on earthquakes in California. The new earthquakes (relative to those used by, say, Abrahamson and Silva, 1997) include the 1999 Hector Mine, California, earthquake; 1999 Kocaeli and Duzce, Turkey, earthquakes; 1999 Chi-Chi, Taiwan, earthquake and five major aftershocks; several well-recorded moderate earthquakes in California; 2003 Nenana Mountain and Denali, Alaska, earthquakes, and several earthquakes from extensional tectonic regimes.

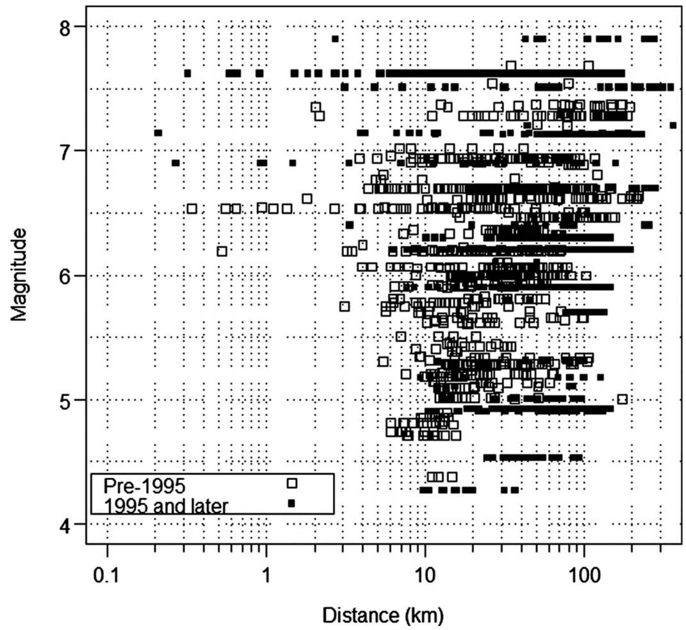

Figure 2. Magnitude and distance distribution of strong-motion records in the NGA database.

RECORD CATALOG

The NGA record catalog currently contains 3551 strong-motion records obtained from about 35 agencies. Most of them are from the strong-motion networks operated by the California Geological Survey Strong Motion Instrumentation Program (CSMIP), the U.S. Geological Survey (USGS), and the Central Weather Bureau of Taiwan (CWB). About half of the records came from the 1999 Chi-Chi, Taiwan, earthquake and its 5 aftershocks. Figure 2 shows the magnitude-distance distribution of the new strong-motion records, superimposed on the pre-1995 data. Earthquakes in regions outside California are the primary source for data at large ($\mathbf{M} > 7$) magnitudes.

EARTHQUAKE SOURCE TABLE

The earthquake source table contains basic information about the seismic source, including origin date and time, moment magnitude (**M**), hypocenter location, faulting mechanism, occurrence of primary surface rupture, and tectonic environment, among others. In addition, finite fault models for 63 earthquakes were collected and systematically evaluated. The finite fault model provided additional information such as the dimension of fault rupture and depth to the top of rupture. Discussions of the collection and evaluation of finite fault models are provided later in this paper in the section 'Finite Fault Models'.

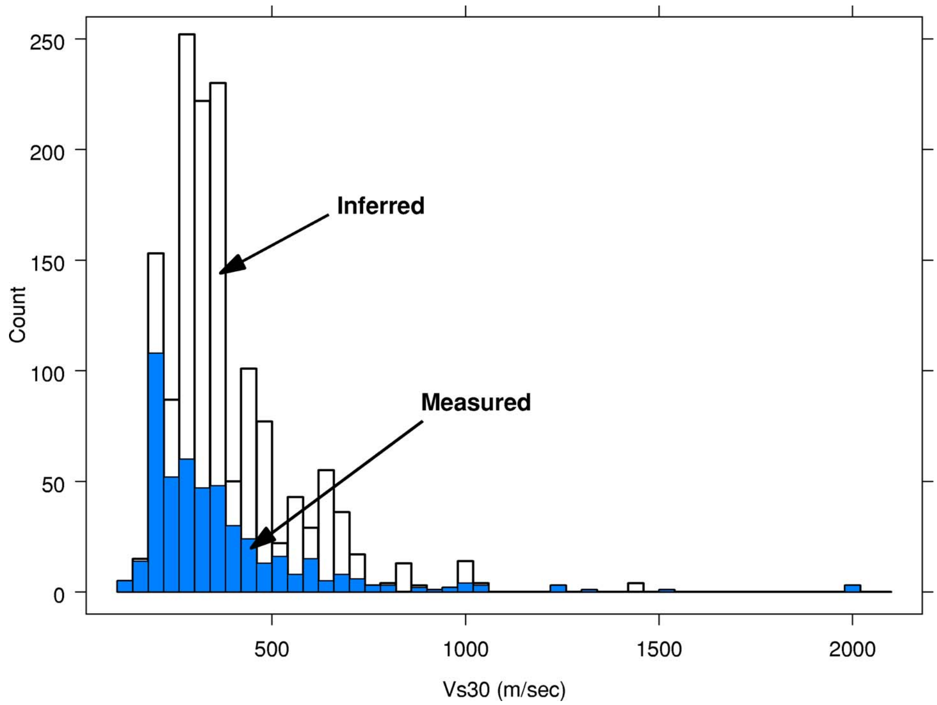

Figure 3. Histograms of measured and inferred V_{S30} at the recording station sites.

STRONG-MOTION STATION TABLE

The strong-motion station table contains information about surface geology and shallow subsurface condition at about 1600 strong-motion stations, of which 1456 stations recorded one or more NGA records. Compared to other strong-motion databases, the NGA database is particularly rich in site characterizations. In addition to V_{S30} (average shear-wave velocity to 30-meter depth), site classifications based on classification schemes developed by NEHRP (BSSC, 1994), Geomatrix (Wells, 2005, personal communication), Campbell and Bozorgnia (2003), Bray and Rodriguez-Mark (1997), Rodriguez-Marek et al. (2001), Wills et al., (NEHRP-extended, 2000), Lee et al. (2001) (for CWB Taiwan stations), Spudich et al. (1999), and other groups were included. Definitions for some of the aforementioned site classification schemes are provided in Appendix B. Figure 3 shows the histograms of V_{S30} at the NGA strong-motion stations. The majority of stations are in NEHRP categories C and D. A total of 485 stations (about 33% of the NGA stations) have measured V_{S30}, and 103 of them were based on borehole measurements obtained by NCREE (National Center for Research on Earthquake Engineering in Taiwan). Details on the collection and evaluation of V_{S30} and site classifications are provided in a later section.

In addition, depths to 1.0, 1.5 and 2.5 km/sec shear-wave velocity horizons for stations in southern California were contributed by Robert Graves and Jonathan Stewart

(2004, personal communication) based on the SCEC-3D model (Magistrale et al., 1996), by Boatwright et al. (2004) for stations in the San Francisco Bay area, and by Graves (1994) for stations in the Eel River basin in northern California. Depth data at a station were superseded by values obtained from borehole data, provided a borehole exists and penetrates the velocity horizon. Additional information about sedimentary basins and depth to basement rock were contributed by Somerville et al. (2002) and Campbell (2003, personal communication), respectively.

The station table also contains information on instrument location. Coordinates of the strong-motion stations (latitude and longitude) were compiled from information published by the operating agencies. Information on housing structure type and location of instrument inside the structure were used by NGA developers to determine if a record is appropriate to be used as a free-field record. The key instrument location information is Geomatrix's classification of structure type and instrument location (Geomatrix's 1st letter; see Appendix B for definition). Donald Wells systematically reviewed and updated Geomatrix's classification. Campbell and Bozorgnia (2003) contributed additional information about instrument housing and location inside a structure.

PROPAGATION PATH TABLE

Key metadata in the propagation path table include various distance measures, hanging wall indicator, directivity parameters, and radiation pattern coefficients. They were derived from information in the earthquake source table, the finite fault models, and the station coordinates. Nancy Collins and her colleagues at the URS Corporation reviewed and checked the path metadata of an early version of the NGA database.

Calculation of the path metadata was somewhat complicated for a non-vertical finite fault with multiple fault segments. An algorithm developed by Robert Youngs (personal communication) was used to project the down-dip extension of a multi-segment fault. A brief description of this algorithm is given in Appendix A of Spudich and Chiou (2008, this issue), along with the definitions of several directivity parameters for the case of a multi-segment fault.

The strike-parallel direction at a recording site is ambiguous when the finite fault model has a variable strike direction. The consensus definition reached by the NGA developers is that the strike-parallel direction is the average fault strike direction over no more than a 20-km stretch of fault length beginning at the closest point on the fault and moving towards the epicenter.

NGA FLAT FILE AND DATA DOCUMENTATION

To deliver data to NGA developers in a single table for use in regression analysis, a flat file was created by merging ground-motion parameters and key metadata. The NGA flat file[1] included PGA, PGV, PGD, and (5%-damped) pseudo absolute spectral acceleration at 105 periods, all being GMRotI50. It also included 126 columns of metadata, even though many of these were not used by the NGA developers. The flat file also fa-

[1] Available via http://peer.berkeley.edu/products/nga_project.html

cilitated the timely and efficient dissemination of the core NGA data to the research and engineering communities. Flat files for other components of ground-motion parameters, such as pseudo spectral accelerations for strike-normal, strike-parallel, and vertical components, were also created.

A document was prepared in 2005 with the objective to provide basic definitions and some explanations of each data column in the flat file. A revised copy is included here as Appendix A (on line). The flat file document has been and will continue to be the basic documentation of NGA data. The rest of this paper gives supplemental and enhanced documentation, specifically about strong-motion record processing, finite fault models, and site conditions.

STRONG-MOTION RECORD PROCESSING

Strong-motion record processing has two major objectives to make the data useful for engineering analysis: (1) correction for the response of the strong-motion instrument itself; and (2) reduction of random noise in the recorded time series. A large portion of the NGA recordings were processed by Pacific Engineering using the PEER processing procedure (Figure 4); the remaining recordings were entered into the database without additional processing (pass-through records).

PEER RECORD PROCESSING PROCEDURE

The PEER processing concentrated on extending both the high- and low-frequency ranges of the useable signal in the recordings (spectral accelerations) on an individual component basis. This processing scheme (Figure 4) consists generally of low- and high-pass causal or acausal Butterworth filters applied in the frequency domain. Corner frequencies were selected by visual examination of the Fourier amplitude spectra and integrated displacements. If necessary, a simple baseline correction was applied for cases where filtering did not remove non-physical trends in the displacement time series. The baseline correction consisted of fitting a polynomial (degree greater than two) to the displacement time series and subtraction of the corresponding acceleration from the filtered acceleration time series. Examples of PEER processing results at both high- and low-frequencies can be found in Darragh et al. (2004).

PASS-THROUGH RECORDS

There are two main reasons why many records were entered into the NGA database without additional processing (Darragh et al., 2004). First, more recent digitally-recorded data generally do not benefit from additional processing and were entered into the database after review of the Fourier amplitude spectra and time series (for example, for glitches). Second, some acceleration data (for example, CSMIP data starting with the 1992 Cape Mendocino earthquake) were available only in Volume II format (filtered and base-line corrected). These data were simply entered into the database in standard format after a similar review. A pass-through record was identified as '#' in the column 'PEA Processing Flag' of NGA flat file. Filter type and filter corner frequency used in the processing of a pass-through record were also entered into the database, if they were available.

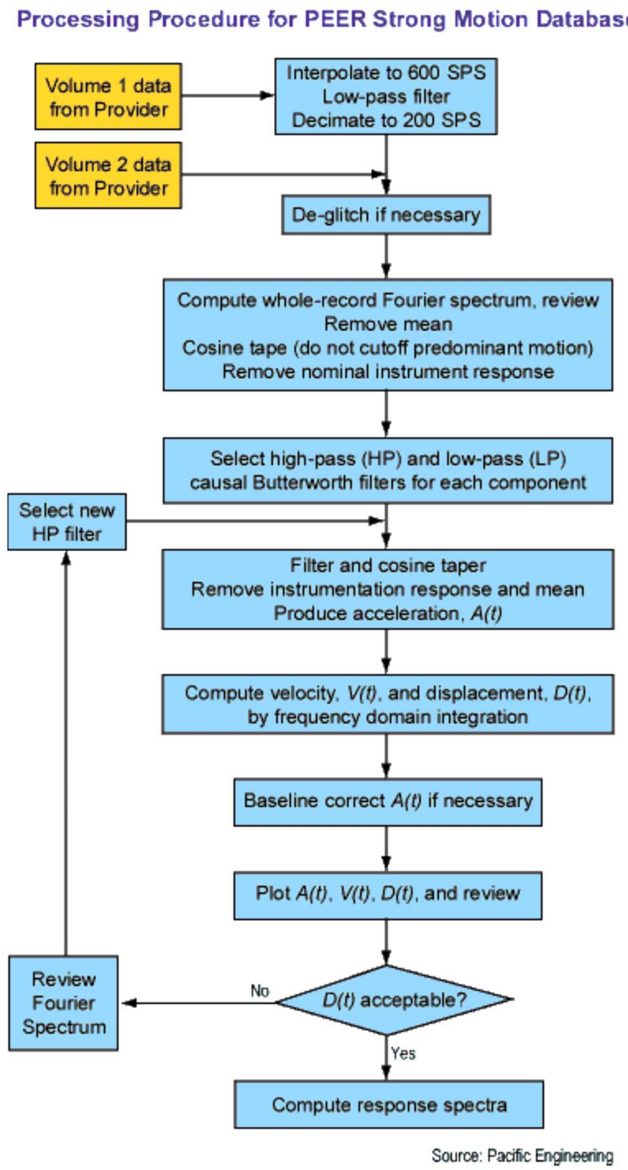

Figure 4. PEER record processing procedure.

REVIEWS OF TIME SERIES AND RESPONSE SPECTRA

The processing by Pacific Engineering was in general different than the processing done by the agency that collected the data. This was necessarily the case as record processing largely relies on judgment as to where (in frequency) noise has significantly contaminated a recording at both high- and low-frequency ranges. More importantly, record processing (filtering) must, by definition, distort a record (side effects) and different processing procedures result in different side effects or distortions. Record processors are faced with the dilemma as to which set of side effects are the most or least desirable. The use of causal versus acausal filters discussed later is an example of such a dilemma.

The NGA strong-motion time series and (5%-damped) pseudo spectral accelerations were extensively reviewed at the NGA-COSMOS Joint Working Group Meeting on Data Processing on March 17th, 2004, and a summary was presented at the International Workshop on Strong-Motion Record Processing sponsored by COSMOS. At the NGA-COSMOS joint meeting the results of a large number of time domain and spectral domain (5%-damped pseudo spectral acceleration) comparisons were presented and discussed. Two hundred and seventy-one time domain and spectral domain comparisons were made between the NGA and the California Geological Survey Strong Motion Instrumentation Program (CSMIP) Volume II and III data sets. The comparison included all records common to both data sets. The processed records are from 34 California earthquakes ranging from the 1979 Imperial Valley earthquake to the 1989 Loma Prieta earthquake. Similar comparisons were made on strong motion data from a smaller set of 16 earthquakes processed by the USGS. These data started with the 1974 Hollister earthquake and ended with the 1989 Loma Prieta earthquake. Forty-eight time domain and 108 spectral domain comparisons were presented.

The differences in response spectra between NGA records and those processed by either CGS or USGS were mainly associated with: 1) selection of the high-pass and low-pass filter corner frequencies that define the effective passband; and 2) the type of filter used (causal or acausal). The presented comparisons showed that the different processing procedures produced zero difference, on average, in elastic response spectra across the useable (common) frequency band. Outside of the common useable bandwidth, large spectra differences may be observed due to differences in the filter corner frequencies. Incidentally, on a number of records the PEER processing resulted in an expanded bandwidth due to the selection of filters independently on each component rather than an entire record (3 components) or on a policy basis.

Acausal filtering results in fewer phase distortions as discussed in Boore and Akkar (2003), Boore and Bommer (2005), and Boore (2005). The greater distortion present with causal filtering may affect spectral values, especially inelastic spectra, at frequencies much higher than the high-pass filter corner frequency. This occurs because the response spectrum measures a peak value in the time domain, and this measurement is affected by the phasing of the ground motion over a wide frequency range. The Boore and Akkar (2003) work presented the analysis from two recordings from the Hector Mine earthquake and a limited number of other recordings. They state “the question on whether to use causal or acausal filters depends on the intended use of the data, desirability for compatible processed acceleration, velocity and displacement time series and considerations of computer storage space.” An advantage of causally filtered time histories is compatibility; that is, velocity, displacement and response spectra computed from the acceleration time series will match the data provided. In contrast, acausal filters require that the padded portions of the processed time series also be distributed to maintain compatibility between time histories and spectra (see Boore and Bommer, 2005, for more discussion). Additionally, causal filters have a significantly steeper spectral falloff compared to acausal filters that results in a wider useable bandwidth. The phase distortion, however, can distort the displacement time series, particularly at periods near the filter corner period. On the other hand, acausal filters generally result in an artificial and significant ramp in displacement, preceding the arrival of long period energy from the source. For analyses of spatial arrays, where relative timing is important, causal filtering is preferred, as is done in seismological observatories for earthquake locations.

The sensitivity to filtering method presented by Pacific Engineering at the NGA-COSMOS joint working group meeting showed that for most of the nearly 1000 components studied the elastic response spectra differences associated with the different filtering methods are small and they do not appear to result in systematic high or low bias of spectra within the common pass band. Bazzurro et al. (2004), in another large spectral domain study, support the above observation. They found that elastic and inelastic spectra from causal and acausal filtered records are statistically indistinguishable from each other provided the same filter order and corner frequencies have been used. The causally filtered records however result in a slightly larger variability in both elastic and inelastic response. It was concluded that the effects of filter causality on NGA regression results are considered to be insignificant.

An exception to the above conclusion is for a small group of near-source records having large static displacements (i.e. tectonic displacements). Standard PEER processing of the time histories does not allow for the displacements to have a static offset or residual displacement (i.e., frequency = 0 Hz). To preserve the static displacement, a static baseline correction method such as those described in Iwan et al. (1985), Grazier (1979), and Darragh et al. (2004) could be used in lieu of a high-pass filtering. The peak ground displacement and, to a lesser extent, peak ground velocity values for PEER processed records are typically lower than for static baseline corrected cases. Interestingly, a comparison of the peak-to-peak displacement shows that the value from the standard PEER processed time history is approximately the same as the value from the static baseline corrected time history (Darragh et al., 2004). This suggests that the standard processing, which does not preserve static fields, may result in similar dynamic loads to structures. As noted by Boore (2001), the difference in the acceleration response spectra between time histories which have been processed using a standard approach and those using a static baseline correction approach are relatively small for periods less than about 20 seconds, which is greater than the maximum 10 second period used in the PEER-NGA data set.

ASSESSMENT OF USABLE FREQUENCY RANGE

Response spectral values were provided to a highest frequency of 100 Hz in the NGA database. For sites in western North America, peak ground acceleration (PGA) is equivalent to the 100 Hz spectral value (5% damped PSA) even at hard rock locations.

In contrast, at low frequencies, the minimum useable frequency is a critical issue. For example, the causal 5-pole Butterworth filters commonly used in PEER processing has a significant reduction (0.707, or −3 db) in response at the filter corner frequency. Hence it was recommended that the usable bandwidth of these records for the purpose of engineering analysis extend from 100 Hz to the high-pass corner frequency multiplied by a factor of 1.25 (Abrahamson and Silva, 1997). With the 1.25 factor (1.5 factor for acausal 5-pole Butterworth filter), the lowest usable spectral frequency is the Fourier frequency at which the filter response is about $-1/2$ db down from the maximum response. Using the same $-1/2$ db criterion, the recommended low-frequency limit for each pass-through record was selected according to the Butterworth filter order and the number of filter passes used in the record processing, as recommended by Boore (2004, personal communication). For records filtered with an Ormbsy filter, a factor of 1 was used because of the slow decay of filter response beyond the corner frequency.

FINITE FAULT MODELS

An earthquake's finite fault model is a critical piece of information from which numerous other source and path metadata were derived. In the NGA database the finite fault geometry was defined by the end points on the top edge of rupture, depth to the bottom edge of rupture, and fault dip angle. The finite fault geometry was typically obtained, in the order of preference, from field observation of primary surface rupture, slip model obtained by inversions of waveform and geodetic data, and observation of aftershock distribution. When a slip model was available, that model was also used to extract information about the rise time, rupture velocity, and other metadata related to the spatial distribution of (coseismic) fault slip.

The NGA finite fault models were built on three model collections previously used in ground-motion attenuation studies: PEER-NEAR (Silva et al., 1999a), USGS-YM (Spudich et al., 1996), and Chiou and others (2000). PEER-NEAR is a set of finite fault models for shallow crustal earthquakes collected by Pacific Engineering. USGS-YM, developed by USGS for the Yucca Mountain Project, is for earthquakes in extensional regimes. These two model collections supplement each other, with only eight earthquakes overlap between them. The third collection (Chiou et al., 2000) overlaps considerably with PEER-NEAR but contains several obscure models that are not in PEER-NEAR. We also expanded the collection by adding models for other earthquakes, especially more recent events. In total, finite fault models for 63 earthquakes were collected. Information about each finite fault model was obtained directly from the researchers or was extracted from their publications. Some of the older models were presented in figures, therefore coordinates defining the fault rupture limits were manually digitized from those figures. We converted every model to a uniform format and to a latitude/longitude coordinate system.

The areal extent of the rupture was a main issue in evaluating the finite fault model. When a model included regions of zero or low level of slip near the edges, the model area was reduced or trimmed back. The consensus reached among the NGA developers and other attending seismologists at the June 2004 NGA developers meeting was that regions with more than 50 cm of coseismic slip should not be trimmed off. Also, it was agreed that the final model should maintain the primary surface rupture observed in the field.

If there were more than one rupture model for an earthquake, a careful evaluation of

Table 1. Spudich et al. (1999) site classification with Geomatrix correlation matrix from all profiles in the Pacific Engineering profile database

	Geomatrix A	Geomatrix B	Geomatrix C	Geomatrix D	Geomatrix E	Total
Spudich 0	4	0	1	0	0	5
Spudich 1	10	1	0	0	0	11
Spudich 2	3	3	2	1	0	9
Spudich 5	0	1	5	3	0	9
Spudich 6	0	3	15	39	0	57
Spudich 7	0	4	2	0	0	6
Total	17	12	25	43	0	97

the available models was conducted to develop a preferred model. In general, the dimension of the preferred model is close to the average of the available models. The preferred finite fault models went through several iterations of reviews by NGA Working Group #4 (during a meeting in September 2003) and the NGA developers (during two developer meetings in May and June of 2004). Furthermore, Paul Somerville, Nancy Collins, and their colleagues at URS Corporation systematically reviewed the finite fault models and provided useful feedback and recommendations that were incorporated into the final models.

SITE CONDITIONS

The PEER-NGA project attempted to collect all publicly available site condition information at strong-motion stations in the NGA strong-motion database. Appendix B provides definitions for several of the site classifications collected during the project. These are the Geomatrix 3-letter site classification (Wells, 2005, personal communication), NEHRP site classification (BSSC, 1994), Spudich et al., (1999) site classification for extensional regimes, extended NEHRP (Wills et al., 2000), and Campbell and Bozorgnia GEOCODE (2003).

The project supported various investigations to systematically fill in the missing site information with an emphasis on V_{S30}. For example, Geomatrix's site classification at the Chi-Chi recording stations initially was missing but later assigned by Donald Wells using geological maps of Taiwan. Wills and Clahan (2004, 2006) provided estimates of V_{S30} for many California stations using surface geology (Table 2). Kayen et al. (2005) acquired SASW (Spectral Analysis of Surface Wave) data and developed shear-wave velocity profiles at 60 California strong-motion stations identified by the NGA project and some of the resulting V_{S30} values were incorporated into the NGA database.

In addition, two approaches were adopted to assign site categories for regions of sparse data. These were classifications in special studies available to the project and site classification correlation matrices (e.g. Table 1). Table 1 shows the correlation matrix for Spudich et al. (1999) and Geomatrix site classifications. This table, along with other available information, was used to assign a Geomatrix classification at strong-motion stations in extensional regimes that had a Spudich et al. (1999) classification.

ESTIMATION OF V_{S30}

The following hierarchy was used to estimate a V_{S30} value for each strong-motion station in the NGA database. First, measured V_{S30} values were obtained from the Pacific Engineering profile data set of over 1500 interpreted V_S profiles. The USGS, ROSRINE, CUREE, NCREE, Agbabian and Associates, Shannon and Wilson, Caltrans, and other organizations measured the shear-wave velocity for these profiles. Only profiles with V_S measurements at depth greater than or equal to 20 m were usually considered, with the 20 m shear-wave velocity extrapolated to 30 m to estimate V_{S30}. To be consistent with the USGS in assignment of V_{S30} values to a location, measurements within 300 m of a site were considered to be located at the recording location (Borcherdt, 2002).

Second, V_{S30} values were inferred from site geology for California stations that recorded the Northridge earthquake from the analysis by Borcherdt (2003, personal communication; Borcherdt and Fumal, 2002) and for other California stations from the surface geology assignments by Wills and Clahan (2004, 2006) (Table 2).

Third, V_{S30} values were inferred for non-California sites from regional V_{S30} profiles developed by Pacific Engineering (e.g. Kobe, Japan, Silva et al., 1999b). Also, V_{S30} values were inferred from the Geomatrix site classification or Spudich et al. (1999) site classification (Table 3).

An alternative inference of V_{S30} was developed for Central Weather Bureau, Taiwan sites using a Taiwan-specific relation of V_{S30} as a function of Geomatrix's classification and station elevation (Chiou and Wen, 2006, personal communication). This relation, described in Appendix C of Chiou and Youngs (2006), is reproduced in Table 4.

Fourth, V_{S30} values were obtained from measured VIC (Vibration Instruments Company Ltd., Tokyo, Japan) data (high-frequency Rayleigh wave measurement) or inferred from maps of V_{S30} for Anchorage, Alaska (Martirosyan et al., 2002).

V_{S30} UNCERTAINTY

The preferred V_{S30} value for a site is considered to reflect the median estimate of V_{S30}, or the mean estimate of $ln(V_{S30})$. The assignment of uncertainty (epistemic variability) to V_{S30} is dependent on the estimation method used and judgment by Pacific Engineering (Figure 5). The estimate of the V_{S30} uncertainty for sites with measured shear-wave velocities was the outcome of an analysis of variance on closely spaced V_{S30} measurements in the Pacific Engineering profile database. The uncertainty at these sites is nonzero because measurements may have been made up to 300 m (Borcherdt, 2002) from the recording station and hence includes considerable epistemic variability (uncertainty) due to varying shear-wave velocity near the site. For example, at Gilroy #2 (NEHRP D site) there are 16 nearby measurements of V_{S30} estimated from velocities obtained from a variety of surface and borehole methods. These measurements yield a standard deviation of 0.08 in natural logarithmic units. A value of 0.10 is used in Figure 5 based on these results and similar cluster analyses at other sites.

Table 2. Statistics of V_{S30} for site geology from profiles in the Pacific Engineering profile data base (from Wills and Clahan, 2004)

Surface Geology	Number of profiles	Mean (m/sec)	Standard Deviation (m/sec)	Median (m/sec)
Qi	20	160	39	155.43
af/qi	44	217	94	202.45
Qal, deep, Imperial Valley	53	209	31	207.47
Qal, fine grained	13	236	55	229.79
Qal deep, LA basin	64	281	85	270.19
Qal, deep (including LA)	161	280	74	271.44
Qal, thin, west LA	41	297	45	294.25
Qs	15	302	46	297.92
Qal, thin, including west LA	115	331	79	322.54
Qal, thin	65	349	89	338.54
Qal, coarse	18	354	82	345.42
Qoa	132	387	142	370.79
Tsh	55	390	112	376.07
QT	18	455	150	438.34
Tss	24	515	215	477.65
Kss	7	540	195	513.70
pCg	3	609	289	554.00
Tv	3	609	155	597.12
J metamorphic	4	641	159	622.86
Serpentine	6	653	137	641.56
Crystalline Rock	28	748	430	660.52
Kgr	21	788	480	684.94
KJf	32	782	359	712.82

The assigned uncertainty is also a function of profile stiffness which has been represented by the NEHRP classification in Figure 5. For example, a 1 km separation between a recording station and the nearest borehole with velocity measurements implies a smaller uncertainty for NEHRP D sites in a large basin compared to a NEHRP B site in the nearby mountains. Also, Schneider and Silva (1994) found that the shear-wave velocity profiles at rock sites are more variable than profiles at soil sites.

The observed V_{S30} aleatory variability (within category randomness) within surface geological units (Table 2) formed the basis for the assigned epistemic variability in the "Inferred" from site geology class in Figure 5. Similarly, Geomatrix or Spudich within category randomness (Table 3), NEHRP category variability (Table 5), along with judgment formed the basis of assigned epistemic variability for the other cases shown in Figure 5.

Table 3. Statistics of V_{S30} for Geomatrix, Campbell and Bozorgnia GEOCODE, and Spudich et al. site classifications from all profiles in the Pacific Engineering profile database

Geomatrix 3rd Letter	Median V_{S30} (m/s)	Standard Deviation of $ln(V_{S30})$	Mean V_{S30} (m/s)	Standard Deviation of V_{S30} (m/s)
A	659.6	0.416	720.2	324.2
B	424.8	0.431	464.3	211.0
C	338.6	0.203	345.4	70.4
D	274.5	0.335	291.4	110.5
E	191.3	0.290	199.4	61.4
Campbell Geocode	**Median V_{S30} (m/s)**	**Standard Deviation of $ln(V_{S30})$**	**Mean V_{S30} (m/s)**	**Standard Deviation of V_{S30} (m/s)**
A	259.4	0.268	269.8	84.3
B	375.7	0.386	387.7	101.7
C	463.4	0.242	476.2	107.6
D	824.6	0.346	869.6	274.4
E	749.5	0.387	801.4	282.9
F	195.5	0.393	215.5	134.1
Spudich	**Median V_{S30} (m/s)**	**Standard Deviation of $ln(V_{S30})$**	**Mean V_{S30} (m/s)**	**Standard Deviation of V_{S30} (m/s)**
2	362.4	⋯	362.4	⋯
6	215.2	0.174	218.7	44.1

Table 4. Taiwan-specific Vs30 relation based on profiles from Taiwan. $ln(Vs_{30}) = ln(\phi_1) + ln(\phi_2) - ln(\phi_1) / 1 + e^{(ln(\phi_3) - ln(Elv))/\phi_4}$.

Geomatrix 3rd Letter	ϕ_1	ϕ_2	ϕ_3	ϕ_4	σ	Number of Data Points
A	552	680[a]	244	0.1154	0.3174	15
B	418	579	107.1	0.3850	0.2294	35
C	—	—	—	—	—	4[b]
D	228	509	39.4	0.373	0.2953	91
E	201	405	38.2	0.087	0.1810	18
						Total = 163[c]

Elv is the elevation of recording station in meters.
[a] This parameter value is fixed by judgment.
[b] There are insufficient data to derive a relationship. To estimate Vs30, one could use the relationship for category D.
[c] Two data points in Geomatrix-B category were removed.

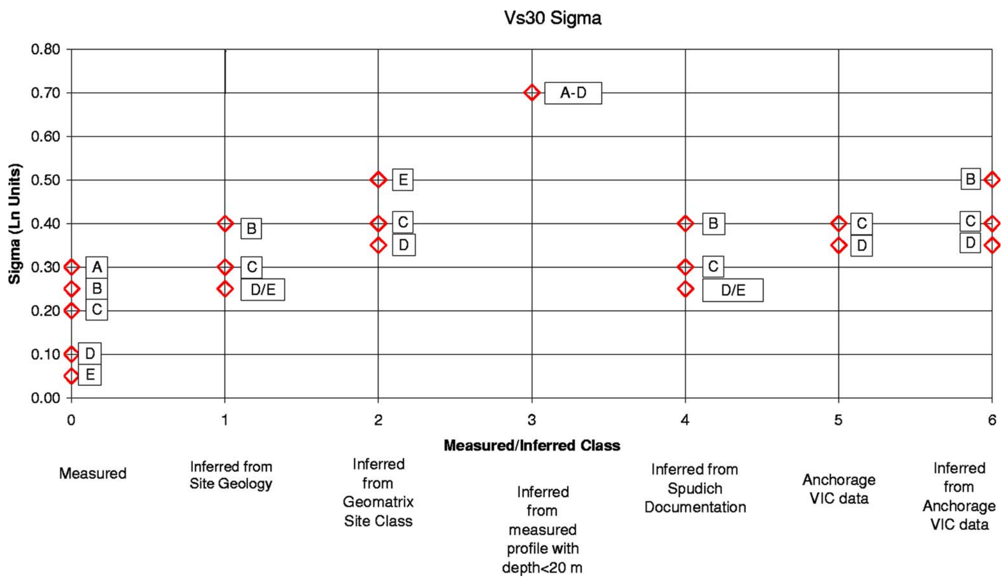

Figure 5. Epistemic variability (uncertainty) of mean $ln(V_{S30})$ estimate for a recording site.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the sharing of strong-motion catalog\database and other documentation by the NGA model developers. This project would not have succeeded without their unselfish contributions of time, data, informal reviews, and guidance. In particular, the authors want to thank Dave Boore for his thoughtful suggestions and careful reviews which resulted in many improvements to the database.

In addition, many individuals and organizations also contributed data and expertise to the database development. These individuals and organizations are Agbabian and Associates, ALYESKA, Jack Boatwright, Roger Borcherdt, Jon Bray, Central Weather Bu-

Table 5. NEHRP Category mean (median) V_{S30} estimates and aleatory variability computed from all profiles in the Pacific Engineering profile database.

NEHRP Category	Median V_{S30} (m/sec)	Standard Deviation of $ln(V_{S30})$	Mean V_{S30} (m/sec)	Standard Deviation of V_{S30} (m/sec)	Number of Profiles
A	1745.3	0.122	1756.0	212.1	6
B	967.6	0.169	981.6	173.5	56
C	450.5	0.189	459.1	94.6	479
D	258.2	0.208	263.8	54.2	886
E	152.5	0.172	154.5	22.3	129

reau of Taiwan, California Department of Transportation, CEORKA, CGS/SMIP Data Center, Kandilli Observatory and Earthquake Engineering Research Institute of Boğaçizi University, Kevin Clahan, COSMOS, CUREE, Doug Dreger, Bill Ellsworth, ESD, Vladimir Graizer, Rob Graves, Moh Huang, Istanbul Technical University, Japan Meteorological Association, Japan Railroad, Rob Kayen, LADWP, Kandilli Observatory and Earthquake Engineering Research Institute of Boğaçizi University, William H.K. Lee Martin Mai, NCREE, Bob Nigbor, Mark Petersen, Maury Power, Ellen Rathje, Cliff Roblee, Kyle Rollins, ROSRINE, SCEC, Linda Seekins, Seismic Networks (CIT, ISC, UCB, USGS, SCSN), Shannon And Wilson, Tony Shakal, Jaime Steidl, Chris Stevens, Paul Somerville, Paul Spudich, Jon Stewart, Ken Stokoe, USC, USGS/NSMP Data Center, URS Corporation, Dave Wald, Jennie Watson-Lamprey, Donald Wells, Kuo-Liang Wen, Chris Wills, and Yuehua Zeng. The staff and agencies listed above that record, process and disseminate strong motion data are acknowledged for their long-term efforts in this field.

This study was sponsored by PEER 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 & 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 PEER Center. Any opinions, findings, and conclusion or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation.

Constructive suggestions from Maury Power allowed the authors to improve this manuscript. Finally, thanks to the editors of the special issue for their patience and assistance in publishing this paper.

APPENDIX B: SITE CLASSIFICATION DEFINITIONS

GEOMATRIX 3-LETTER SITE CLASSIFICATION

FIRST LETTER: Instrument Structure Type (Donald Wells, personal communication, 2005)

I = Free-field instrument or instrument shelter. Instrument is located at or within several feet of the ground surface, and not adjacent to any structure.

A = One-story structure of lightweight construction. Instrument is located at the lowest level and within several feet of the ground surface.

B = Two- to four-story structure of lightweight construction, or tall one-story warehouse-type building. Instrument is located at the lowest level and within several feet of the ground surface.

C = One- to four-story structure of lightweight construction. Instrument is located at the lowest level in a basement and below the ground surface.

D = Five or more story structure or heavy construction. Instrument is located at the lowest level and within several feet of the ground surface.

E=Five or more story structure or heavy construction. Instrument is located at the lowest level in a basement and below the ground surface.

F=Structure housing instrument is buried below the ground surface, e.g. tunnel or seismic vault.

G=Structure of light or heavyweight construction, instrument not at lowest level of structure.

H=Earth dam (station at toe of embankment or on abutment).

J=Concrete Dam (none in PEER-NGA database).

K=Near a one-story structure of lightweight construction. Instrument is located outside on the ground surface, within approximately 3 m of the structure.

L=Near a two- to four-story structure of lightweight construction. Instrument is located outside on the ground surface, within approximately 6 m of the structure.

M=Near a two- to four-story structure of lightweight construction with basement. Instrument is located outside on the ground surface, within approximately 6 m of the structure.

N=Near a five- to eight-story structure or heavy construction. Instrument is located outside on the ground surface, within approximately 10 m of the structure.

O=Near a five- to eight-story structure or heavy construction with basement. Instrument is located outside on the ground surface, within approximately 10 m of the structure.

SECOND LETTER: Mapped local geology sedimentary or metasedimentary rocks:

H=Holocene (Recent) Quaternary (<11,000 ybp).

Q=Pleistocene Quaternary (<1.8 my bp).

P=Pliocene Tertiary (<5 my bp).

M=Miocene Tertiary (<24 my bp).

O=Oligocene Tertiary (<34 my bp).

E=Eocene Tertiary (<55 my bp).

L=Paleocene Tertiary (<65 my bp).

K=Cretaceous (<144 my bp).

F=Franciscan Formation (Cretaceous/Late Jurassic).

J=Jurassic (<206 my bp).

T=Triassic (<248 my bp).

Z=Permian or older (<248 my bp). Igneous or meta-igneous:

V=Volcanic (extrusive).

N=Intrusive.

G=Granitic.

THIRD LETTER: Geotechnical subsurface characteristics

A=Rock. Instrument on rock (Vs > 600 mps) or <5 m of soil over rock.

B=Shallow (stiff) soil. Instrument on/in soil profile up to 20 m thick overlying rock.

C=Deep narrow soil. Instrument on/in soil profile at least 20 m thick overlying rock, In a narrow canyon or valley no more than several km wide.

D=Deep broad soil. Instrument on/in soil profile at least 20 m thick overlying rock, in a broad valley.

E=Soft deep soil. Instrument on/in deep soil profile with average Vs < 150 mps.

NEHRP SITE CLASSIFICATION (BSSC, 1994)

Average shear-wave velocity to a depth of 30 m is:

A > 1500 m/s

B=760 m/s–1500 m/s

C=360 m/s–760 m/s

D=180 m/s–360 m/s

E < 180 m/s

SITE CLASSIFICATION FOR EXTENSIONAL TECTONIC REGIMES

Spudich et al., (1999)

0=Unknown rock site

1=Hard rock site, soil <5 m over hard rock

2=Soft rock site, soil <5 m over soft rock

5=Soil, unknown depth

6=Deep soil, >20 m thick

7=Shallow soil, 5 m⇐thickness⇐20 m

NEHRP—UBC (extended) Site Classification

Wills et al. (2000), personal communication (2003)

B: Plutonic and metamorphic rocks, most volcanic rocks, coarse sedimentary rocks of Cretaceous age and older.

BC: Franciscan Complex rocks of the Transverse Ranges which tend to be more sheared, Cretaceous siltstones, or mudstone.

C: Franciscan mélange and serpentine, sedimentary rocks of Oligocene to Cretaceous age or coarse-grained sedimentary rocks of younger age.

CD: Sedimentary rocks of Miocene and younger age, unless formation is notably coarse grained, Plio-Pleistocene alluvial units, older (Pleistocene) alluvium, some areas of coarse younger alluvium.

D: Younger (Holocene) alluvium

DE: Fill over bay mud in the San Francisco Bay Area, fine-grained alluvial and estuarine deposits elsewhere along the coast.

E: Bay mud and similar intertidal mud.

GEOCODE: Campbell-Bozorgnia site class

Campbell and Bozorgnia (2003), personal communications (2002, 2003)

Suggested V_{S30} from (Wills and Silva, 1998)

A=Firm Soil: Holocene; recent alluvium, alluvial fans, undifferentiated Quaternary deposits. $V_{S30}=298\pm92$ m/sec; NEHRP D

B=Very Firm Soil: Pleistocene; older alluvium or terrace deposits. $V_{S30}=368\pm80$ m/sec; NEHRP CD

C=Soft Rock: Sedimentary rock, soft volcanic deposits of Tertiary age, "softer" Franciscan, low grade metamorphic rocks such as mélange, serpentine, schist. $V_{S30}=421\pm109$ m/sec; NEHRP CD

D=Firm Rock: Older sedimentary rock and hard volcanic deposits, high grade metamorphic rock, crystalline rock, "harder" Franciscan $V_{S30}=830\pm339$ m/sec; NEHRP BC

E=Shallow Soils (≤10 m deep)

F=Extremely soft or loose Holocene age soils such as beach sand or recent floodplain, lake, swamp estuarine, and delta deposits.

REFERENCES

Abrahamson, N. A., and Silva, W. J., 1997. Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seismol. Res. Lett.* **68**(1), 94–127.

Bazzurro, P., Sjoberg, B., Luco, N., Silva, W., and Darragh, R., 2004. Effects of strong motion processing procedures on time histories, elastic and inelastic spectra, *Proceedings of COSMOS Workshop on Strong-Motion Record Processing*, Richmond, California, May 26–27, 2004, 1–39.

Boatwright, J., Blair, L., Catchings, R., Goldman, M., Perosi, F., and Steedman, C., 2004. *Using twelve years of USGS refraction lines to calibrate the Brocher and others (1997) 3D velocity model of the Bay Area*, U.S. Geological Survey Open File Report *2004-1282*.

Boore, D. M., 2001. Effect of baseline corrections on displacements and response spectra from several recordings of the 1999 Chi-Chi, Taiwan, earthquake, *Bull. Seismol. Soc. Am.* **91**(5), 1199–1211.

——, 2005. On pads and filters: Processing strong-motion data, *Bull. Seismol. Soc. Am.* **95**(2), 745–750.

Boore, D. M., and Akkar, S., 2003. Effects of causal and acausal filters on elastic and inelastic response spectra, *Earthquake Eng. Struct. Dyn.* **32**, 1729–1748.

Boore, D. M., and Bommer, J. J., 2005. Processing of strong-motion accelerograms: Needs, options and consequences, *Soil Dyn. Earthquake Eng.* **25**, 93–115.

Boore, D. M., Watson-Lamprey, J., and Abrahamson, N. A., 2006. GMRotD and GMRotI: Orientation-independent measures of ground motion, *Bull. Seismol. Soc. Am.* **96**(4a), 1502–1511.

Borcherdt, R. D., 2002. Empirical evidence for acceleration-dependent amplification factors, *Bull. Seismol. Soc. Am.* **92**(2), 761–782.

Borcherdt, R. D., and Fumal, T. E., 2002. *Shear-wave velocity compilation for Northridge strong-motion recording sites*, U.S. Geological Survey Open File Report *2002-107*.

Bray, J. D., and Rodriguez-Marek, A., 1997. Geotechnical site categories proceedings, First PEER-PG&E Workshop on Seismic Reliability of Utility Lifelines, San Francisco, CA, August.

BSSC, 1994. *NEHRP recommended provisions for seismic regulations for new buildings, Part 1—Provisions*, *FEMA 222A*, Federal Emergency Management Agency.

Campbell, K. W., and Bozorgnia, Y., 2003. Updated near-source ground-motion (attenuation) relations for the horizontal and vertical components of peak ground acceleration and acceleration response spectra, *Bull. Seismol. Soc. Am.* **93**(1), 314–331.

Chiou, S.-J., Makdisi, F. I., and Youngs, R. R., 2000. *Style-of-faulting and footwall/hanging wall effects on strong ground motion*, FY 1995 NEHRP Award Number *1434-95-G-2614*, final report, 21 p.

Chiou, B. S.-J., and Youngs, R. R., 2006. *Chiou and Youngs PEER-NGA empirical ground motion model for the average horizontal component of peak acceleration and pseudo-spectral acceleration for spectral periods of* 0.01 to 10 Seconds, Report submitted to PEER, June, 2006.

Darragh, B., Silva, W., and Gregor, N., 2004. Strong motion record processing procedures for the PEER center, *Proceedings of COSMOS Workshop on Strong-Motion Record Processing*, Richmond, California, May 26–27, 2004, 1–12.

Graves, R. W., 1994. *Simulating the 3D basin response in the Portland and Puget Sound regions from large subduction zone earthquakes*, USGS Award: *1434-93-G-2327*, Annual Technical Report.

Grazier, V. M., 1979. Determination of the true ground displacement by using strong motion records, Izvestiya Academy of Sciences, *Izv., Acad. Sci., USSR, Phys. Solid Earth* **15**(12), 875–885.

Iwan, W. D., Moser, M. A., and Peng, C.-Y., 1985. Some observations on strong-motion earthquake measurement using a digital accelerograph, *Bull. Seismol. Soc. Am.* **75**(5), 1225–1246.

Kayen, R., Thompson, E., Minasian, D., and Carkin, B., 2005. *Shear-wave velocity of the ground near sixty California strong motion recording sites by the Spectral Analysis of Surface Waves (SASW) method and harmonic-wave sources*, U.S. Geological Survey Open-File Report *2005-1366*.

Lee, C.-T., Cheng, C.-T., Liao, C.-W., and Tsai, Y.-B., 2001. Site classification of Taiwan free-field strong-motion stations, *Bull. Seismol. Soc. Am.* **91**(5), 1283–1297.

Magistrale, H., McLaughlin, K., and Day, D., 1996. A geology-based 3D velocity model of the Los Angeles basin sediments, *Bull. Seismol. Soc. Am.* **86**(4), 1161–1166.

Martirosyan, A., Dutta, U., Biswas, U., Papageorgiou, N., and Combellick, R., 2002. Determination of site response in Anchorage, Alaska, on the basis of spectral ratio methods, *Earthquake Spectra* **18**(1), 85–104.

Power, M., Chiou, B., Abrahamson, N., Bozorgnia, Y., Shantz, T., and Roblee, C., 2008. An overview of the NGA project, *Earthquake Spectra* **24**(1), 3–21.

Schneider, J. F., and Silva, W. J., 1994. What is rock? Implications for site response estimation, *Seismol. Res. Lett.* **65**(1), 44.

Silva, W. J., Gregor, N., and Darragh, B., 1999a. *Near fault ground motions*, A PEARL report to PG&E/CEC/Caltrans, Award No. *SA2193-59652*.

Silva, W. J., Costantino, C., and Iwasaki, Y., 1999b. *Assessment of liquefaction potential for the 1995 Kobe, Japan earthquake including finite-source effects*, Prepared for U.S. Army Waterways Experiment Station, Corps of Engineers Contract *#DACW39-97-k-0015*.

Somerville, P. G., Collins, N., Graves, R., and Pitarka, A., 2002. Development of an engineering model of basin generated surface waves, *Proceedings of SMIP03 Seminar on Utilization of Strong-Motion Data*, p. 1–21.

Spudich, P., Fletcher, J. B., Hellweg, M., Boatwright, J., Sullivan, C., Joyner, W. B., Hanks, T. C., Boore, D. M., McGarr, A., Baker, L. M., and Lindh, A. G., 1996. *Earthquake ground motions in extensional tectonic regimes*, U.S. Geological Survey Open File Report *96-292*.

Spudich, P., Joyner, W. B., Lindh, A. G., Boore, D. M., Margaris, B. M., and Fletcher, J. B., 1999. SEA99: A revised ground motion prediction relation for use in extensional tectonic regimes, *Bull. Seismol. Soc. Am.* **89**(5), 1156-1170.

Spudich, P., and Chiou, B. S. J., 2008. Directivity in NGA earthquake ground motions: analysis using isochrone theory, *Earthquake Spectra* **24**(1), 279–298.

Rodriguez-Marek, A., Bray, J. D., and Abrahamson, N. A., 2001. An empirically based geotechnical seismic site response procedure, *Earthquake Spectra* **17**(1), 65–87.

Wills, C. J., Petersen, M., Bryant, W. A., Reichle, M., Saucedo, G. J., Tan, S., Taylor, G., and Treiman, J., 2000. A site-conditions map for California based on geology and shear-wave velocity, *Bull. Seismol. Soc. Am.* **90**(6B), S187-S208.

Wills, C. J., and Silva, W., 1998. Shear wave velocity characteristics of geologic units in California, *Earthquake Spectra* **14**(3), 533–556.

Wills, C. J., and Clahan, K. B., 2004. *NGA: Site condition metadata from geology*, final project *1L05* to PEER.

——, 2006. Developing a map of geologically defined site-condition categories for California, *Bull. Seismol. Soc. Am.* **96**(4A), 1483–1501.

(Received 7 December 2007; accepted 30 January 2008)