NGA-West2 Database

Timothy D. Ancheta,[a)] M.EERI, Robert B. Darragh,[b)] M.EERI, Jonathan P. Stewart,[c)] M.EERI, Emel Seyhan,[c)] M.EERI, Walter J. Silva,[b)] M.EERI, Brian S.-J. Chiou,[d)] M.EERI, Katie E. Wooddell,[e)] M.EERI, Robert W. Graves,[f)] M.EERI, Albert R. Kottke,[g)] M.EERI, David M. Boore,[f)] Tadahiro Kishida,[h)] M.EERI, and Jennifer L. Donahue,[i)] M.EERI

The NGA-West2 project database expands on its predecessor to include worldwide ground motion data recorded from shallow crustal earthquakes in active tectonic regimes post-2000 and a set of small-to-moderate-magnitude earthquakes in California between 1998 and 2011. The database includes 21,336 (mostly) three-component records from 599 events. The parameter space covered by the database is **M** 3.0 to **M** 7.9, closest distance of 0.05 to 1,533 km, and site time-averaged shear-wave velocity in the top 30 m of V_{S30} = 94 m/s to 2,100 m/s (although data becomes sparse for distances >400 km and V_{S30} > 1,200 m/s or <150 m/s). The database includes uniformly processed time series and response spectral ordinates for 111 periods ranging from 0.01 s to 20 s at 11 damping ratios. Ground motions and metadata for source, path, and site conditions were subject to quality checks by ground motion prediction equation developers and topical working groups. [DOI: 10.1193/070913EQS197M]

INTRODUCTION

In 2003, the Pacific Earthquake Engineering Research Center (PEER) initiated a large research program to develop next generation ground motion prediction equations (GMPEs) for shallow crustal earthquakes in active tectonic regions. The project (now called NGA-West1) concluded in 2008 and provided several important products, including a strong motion database of recorded ground motions (Chiou et al. 2008) and a set of peer-reviewed GMPEs (Power et al. 2008). Many researchers, practitioners, and organizations throughout the world are now using the NGA-West1 models and database for research and engineering applications. As successful as the original NGA-West1 program was, some problems were identified over time with the resulting GMPEs (e.g., poor predictions for small **M** events) that

[a)] Risk Management Solutions, Newark, CA
[b)] Pacific Engineering and Analysis, El Cerrito, CA
[c)] University of California, Los Angeles, CA
[d)] California Department of Transportation, Sacramento, CA
[e)] Pacific Gas & Electric, San Francisco, CA
[f)] U.S. Geological Survey, Pasadena CA (RWG) and Menlo Park CA (DMB)
[g)] Bechtel Corporation, San Francisco, CA
[h)] Pacific Earthquake Engineering Research Center, Berkeley, CA
[i)] Geosyntec Consultants, San Francisco, CA

Earthquake Spectra, Volume 30, No. 3, pages 989–1005, August 2014;

could be addressed with further work benefitting from additional data and simulation results. Therefore, a major thrust of the NGA-West2 project was to expand the database and improve its quality and consistency. Specific issues targeted by this expansion include:

1. Ground motion scaling for small-to-moderate **M** events (**M** < 5).
2. Possible regionalization of path and site effects within and outside California.
3. Possible constraint of single-station (i.e., intra-site) standard deviation (sigma) in the NGA-West2 GMPEs (Al Atik et al. 2010).
4. Improved uniformity in the metadata and increased transparency in its development.
5. Addition of post-2003 (the latest event date in the NGA-West1 database) shallow crustal events.

In this paper, we summarize the principal aspects of the NGA-West2 database (records and metadata), which are presented in the NGA-West2 flatfile. A more complete documentation of the database is given by Ancheta et al. (2013, hereafter AEA13) and of the site database in Seyhan et al. (2014, hereafter SEA14) in this volume. The flatfile contains the information used by NGA-West2 GMPE developers and is provided on the PEER NGA-West2 website: http://peer.berkeley.edu/ngawest2/databases/. An explanation of the flatfile columns is provided as an Electronic Supplement to this paper. The flatfile was created from four data sources: (1) earthquake source table, (2) site database, (3) propagation path table, and (4) record catalog.

The paper begins with an overall summary of the NGA-West2 database, followed by a summary of the key information from the four data sources. We then summarize the flatfile created from the four data sources, highlighting the key aspects that are useful to the research community.

OVERVIEW OF NGA-WEST2 DATABASE

The NGA-West2 database development used a data collection and record processing methodology similar to that used in the NGA-West1 database (Chiou et al. 2008). New events were selected based on similar criteria used in the NGA-West1 project to ensure consistency within the data set. Events are considered shallow crustal if they occur within continental crust. Independent of region, the majority of the collected events hypocenters are within a depth of 20 km. The region used to collect events is considered "tectonically active" if it is not a stable continental region (SCR), within a subducting slab or on the interface between the slab and the continental lithosphere; typically these events are near a plate boundary. Events were not excluded if they occurred in close proximity (time and space) with a previous event.

The NGA-West2 database is a combination of two data sets gathered in separate efforts and then combined. The first data set is global and concentrates on relatively large **M** > 5 events. This portion of the data set includes 173 NGA-West1 events and 160 added events. The second data set is 266 small-to-moderate (SMM) **M** events from California (3.0 < **M** < 5.45). Figure 1 shows the distribution of the epicenter locations and Figure 2 shows the magnitude-closest distance distribution of the selected recordings. The number of stations in the 2008 and current versions of the database for the five most populated

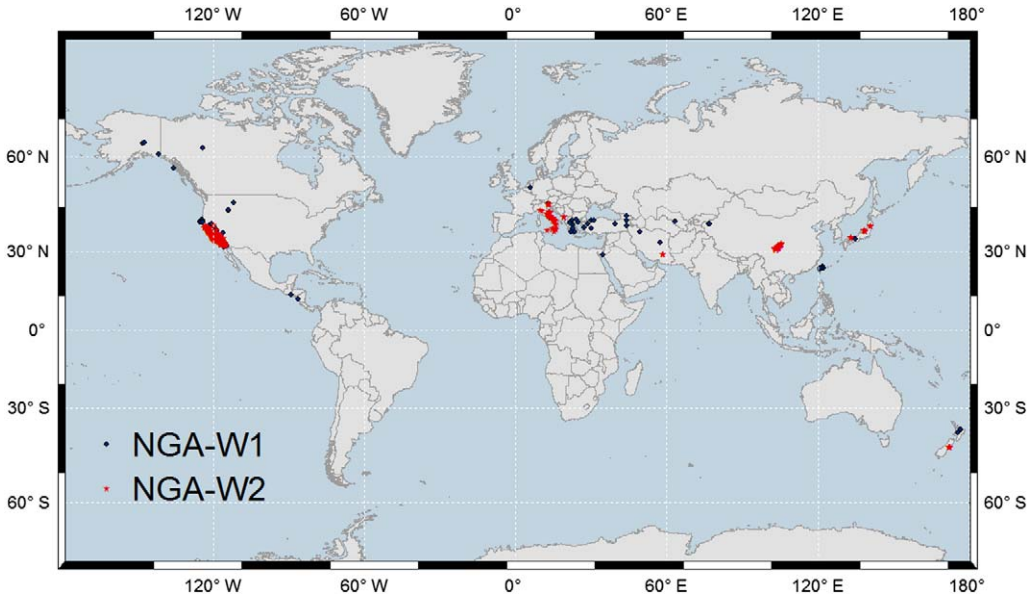

Figure 1. Epicenter distribution of 599 events included in the NGA-West2 database.

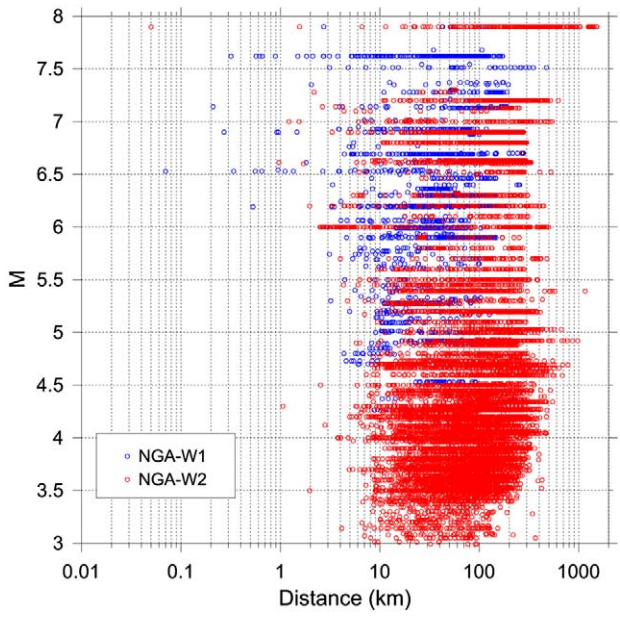

Figure 2. Magnitude-distance distribution of records in the NGA-West2 database (**M** 3.0 to 7.9).

regions are shown in Figure 3. The current database has a large increase in the number of records from Japan, China, and California over the 2008 version.

Significant well-recorded shallow crustal events since 2000 that have been added to the data set include the 2003 **M** 6.6 Bam (Iran), 2004 **M** 6.0 Parkfield (CA), 2008 **M** 7.9 Wenchuan (China), 2009 **M** 6.3 L'Aquila (Italy), 2010 **M** 7 El Mayor-Cucupah (CA and Mexico), 2010 **M** 7 Darfield (NZ), 2011 **M** 6.3 Christchurch (NZ), and four well-recorded **M** 6.6–6.9 shallow crustal earthquakes in Japan between 2000 and 2008. The present data set has more than double the number of **M** > 5.5 recordings than NGA-West1.

Records were collected from a number of agencies managing strong-motion networks worldwide. Record selection followed a similar methodology as in NGA-West1 as follows:

1. Recording station is a ground-level instrument (not within a structure) or is located at the lowest structural level of a structure. Records excluded per these criteria include downhole recordings from vertical arrays and recordings from structural arrays (other than the lowest structural level). The decision of which records to include in regressions (representing approximate free-field conditions) is left to individual GMPE developer teams.
2. Recording has adequate signal-to-noise ratio from visual inspection sometimes both in the time and frequency domains.
3. Location of recording station is known (latitude and longitude).
4. When multiple instruments are co-located on the same foundation pad, a single recording is selected.

The collection of the SMM component of the data set was motivated by findings of misfits of the NGA-West1 GMPEs against small **M** data (Atkinson and Morrison 2009;

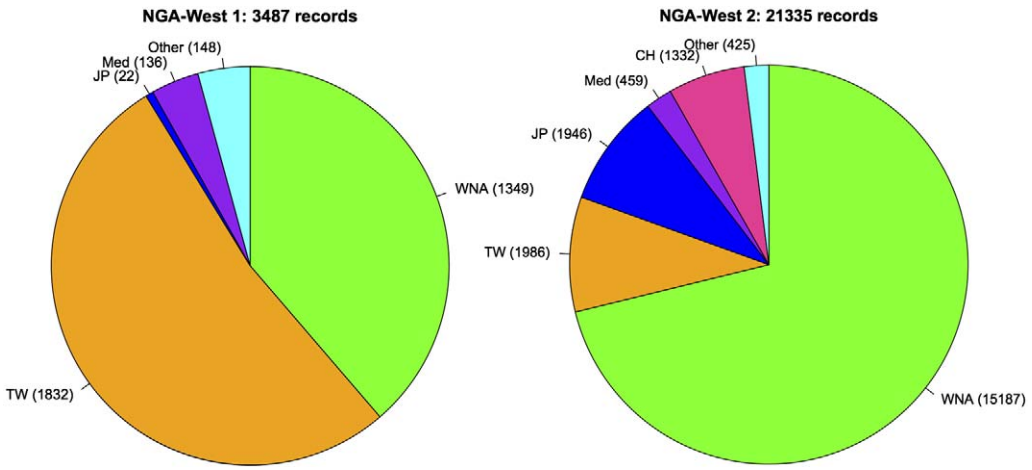

Figure 3. Pie chart of record numbers in the NGA-West1 and NGA-West2 for different regions. CH: China, JP: Japan, Med: Mediterranean, TW: Taiwan, WNA: Western North America (i.e., mostly California).

Chiou et al. 2010; Atkinson and Boore 2011). The selected events are all from California and include those considered in Chiou et al. (2010), events since the 2009 cut-off in the Chiou et al. (2010) data set, and several events from aftershock sequences and earthquake swarms. Records were obtained from the four major combined networks operating in California: (1) Berkeley Digital Seismic Network (BDSN)/Northern California Seismic Network (NCSN); (2) California Institute of Technology (CIT)/Southern California Seismic Network (SCSN)/Southern California Earthquake Center (SCEC); (3) California Geological Survey (CGS)/California Strong Motion Program (CSMIP); and (4) the United States Geological Survey (USGS). In total, we received in excess of 200,000 digital recordings.

The same record selection criteria described above were used for the SMM recordings. Additional considerations specific to this data set included availability of instrument parameters (e.g., natural frequency, damping, sample rate, and anti-alias filter corner) and complexity in the baseline error (e.g., baseline errors with multiple jumps or time dependent trends were excluded). Therefore, record selection was generally stricter than for the worldwide data. After the criteria was applied 12,818 three-component records were utilized from 266 events. The data set spans a magnitude and closest distance range of **M** 3.0–5.45 and 2–1,100 km, respectively. Figure 4 shows the distribution of the epicenter locations.

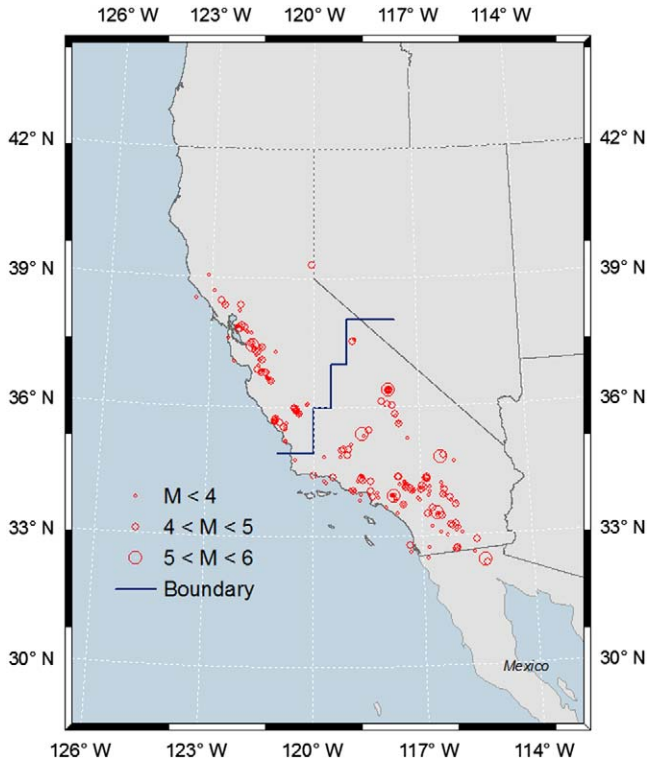

Figure 4. Map of the epicentral distribution of the 266 SMM events added to the NGA-West2 database. For events north of boundary, we preferred metadata from NCSN\BDSN, to the south we use preferred metadata from SCSN\CIT. Boundary from Allen et al. (1965).

Metadata in the flatfile accompanies each recording containing information on source, path, and site parameters. Metadata was originally prepared in tables assembled by internal NGA-West2 working groups with experts in the source modeling, site response, and strong motion processing. A significant degree of vetting and review occurred within the working groups and with the GMPE developers. The metadata tables include: earthquake source table, site database, propagation path table, and record catalog. This basic structure of each table was retained from the NGA-West1 project, but was expanded to accommodate the new NGA-West2 data. A summary flatfile created from the metadata tables contains information used for record selection and as model predictor variables during GMPE development.

EARTHQUAKE SOURCE TABLE

Information in the earthquake source table includes moment magnitude (**M**), hypocenter location, fault rupture dimensions, and focal mechanism prepared by the NGA-West2 Source Working Group (Robert B. Darragh, Robert W. Graves, Brian S.J. Chiou, and Jack Boatwright). Each earthquake is given a unique ID number (EQID). Figure 5 shows the distribution of the 599 NGA-West2 events by rake and dip angles. Relative to NGA-West1 there are approximately two and three times the number of reverse and normal events. The relative percentage of strike, normal, and reverse slip events is 54%, 19%, and 25%, with 2% unknown.

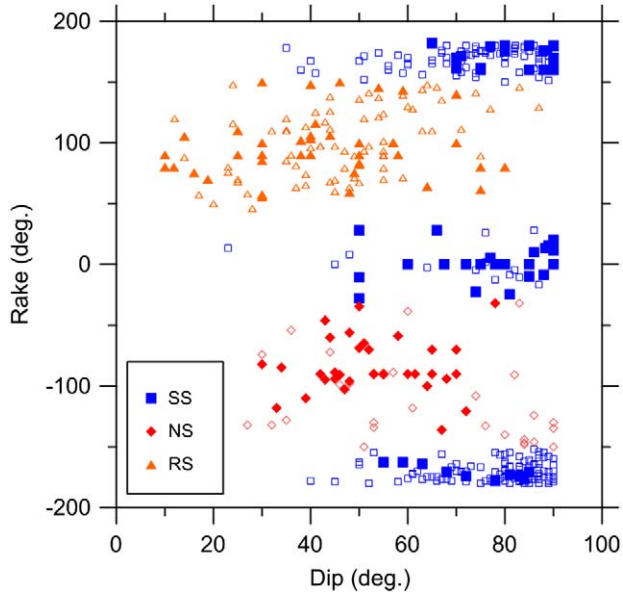

Figure 5. Rake versus dip scatter plot for events in the NGA-West2 database. Open symbols represent events that were added, solid symbols represent events in the NGA-West1 database. The three predominant mechanisms (strike slip, or SS; normal slip, or NS; and reverse slip, or RS) are illustrated as indicated by legend, based on rake angle assignments in the flatfile (AEA13). These boundaries are for illustrative purposes and are not universally applied.

Contents of the source table fed into the flatfile include event time, magnitude, location, mechanism, and finite fault geometry (strike, dip, length, width, top of rupture, etc.). A website link to the electronic version of the source table and explanation file can be found in AEA13. The following sub-sections explain specific issues related to finite fault models, location and magnitude of the SMM events, parameterization of multi-segment ruptures, and event classification with respect to main shock or aftershock.

FINITE FAULT MODELS

Finite fault models describe the earthquake source geometry in terms of a plane or series of planes in the earth's crust, and are used to derive multiple source and path parameters used as predictor variables in GMPE development. The finite fault geometry is defined by the end points of the top edge of rupture, the depth to the bottom edge of rupture, the fault dip angle, and the strike direction. The finite fault geometry was typically obtained, in the order of preference, from field observations of primary surface rupture, coseismic slip distribution from inversions of waveform and geodetic data, and aftershock distributions. When available, slip inversion models were also used to extract information about rise time, rupture velocity, and other data related to the spatial characteristics of (coseismic) fault slip (such as existence of a shallow asperity producing significant (>20% of the total) moment release in the top 5 km of crust).

Finite fault models from NGA-West1 were retained. An additional 14 models were added for NGA-West2 (Table 1). Information about each finite fault model was obtained directly from the researchers or extracted from publications. The models were converted to a uniform format using a latitude/longitude coordinate system. For some older models presented in figures, coordinates defining the fault rupture limits were manually digitized.

Table 1. Finite fault models for earthquakes added in NGA-West2 database

EQID	Earthquake name	Primary choice	Others
262	Montenegro, Yugoslavia	Benetatos and Kiratzi (2006)	
146	Joshua Tree, CA	Hough and Dreger (1995)	
176	Tottori, Japan	Piatanesi et al. (2007)	
177	San Simeon, CA	Ji et al. (2004)	Hardebeck et al. (2004), Dreger et al. (2005)
178	Bam, Iran	Jackson et al. (2006)	
180	Niigata, Japan	Asano and Iwata (2009)	Hikima and Koketsu (2005)
179	Parkfield, CA	Custudio et al. (2005)	
278	Chuetsu-oki, Japan	Miyake et al. (2010)	Cirella et al. (2008)
279	Iwate, Japan	Suzuki et al. (2010)	
274	L'Aquila, Italy	Scognamiglio et al. (2010)	
277	Wenchuan, China	Koketsu et al. (2008)	Wang et al. (2008)
280	El Mayor-Cucupah, MX	Wei et al. (2011)	
281	Darfield, NZ	Beavan et al. (2010)	
346	Christchurch, NZ	Holden (2011)	Beavan et al. (2011)

When there is more than one rupture model for an earthquake, the models were carefully reviewed to select a preferred model. Table 1 lists the preferred finite fault model for each added event along with some of the alternate models considered. The areal extent of the published finite fault models was generally trimmed to remove low-slip areas on the periphery (details are in Chiou et al. 2008; an example for L'Aquila is in Stewart et al. 2012).

SIMULATED FINITE FAULT GEOMETRY

When a published finite fault model is not available, we develop an approximate simulated finite fault based on hypocenter location, **M**, and focal mechanism (or style of faulting). The goal of the simulation routine is to obtain an approximate fault rupture geometry that can be used to compute distance metrics and other path parameters (such as the source-site angle θ_{SITE}, the fault rupture width W, and the depth to the top of the rupture, Z_{TOR}). This process began in NGA-West1 and was significantly expanded for NGA-West2.

The methodology for simulating the fault plane consists of random sampling of probabilistic distributions of fault rupture area, aspect ratio of ruptured area, and hypocenter position on the fault plane. The simulation routine generates 101 random fault ruptures that are rotated and translated in space but fixed on the given hypocenter. The strike and dip are fixed if reported. For a grid of pseudo stations (~500 spaced around the epicenter at epicentral distances from 0 km to 300 km), the median R_{RUP} from the 101 simulated ruptures is calculated. The selected fault plane is the simulated rupture that best fits the set of median R_{RUP}. Use of a grid of pseudo stations instead of the observed recordings in the Chiou and Youngs (2006) approach ensures that adding or subtracting observed stations from the list would not affect the simulated rupture model.

EARTHQUAKE SOURCE PARAMETERS FOR THE SMM CALIFORNIA DATA

Finite fault inversions are typically not available for events with $\mathbf{M} < \sim 5.0$. Therefore, event source parameters for SMM events were chosen from high-quality relocation and moment tensor catalogs, as available. Parameters that are typically available in the catalogs included the event time, magnitude, strike, dip, rake, hypocenter location, and hypocenter depth. The seismic network in California is split between the Northern California Seismic Network (NCSN) and the Southern California Seismic Network (SCSN), as marked by the boundary in Figure 4. For Northern California earthquakes, hypocenter location is taken in order of preference from Waldhauser and Schaff (2008) and then from the Northern California Earthquake Data Center (NCEDC) catalog. Seismic moment, conjugate fault planes (strike and dip), and rake angle are taken from the NCEDC moment tensor catalog. For Southern California earthquakes, hypocenter location is taken in order of preference from Hauksson et al. (2012) and then from the SCSN\SCEC catalog. Seismic moment, conjugate fault planes (strike and dip), and rake angle are taken from Yang et al. (2012) and from the SCSN\SCEC moment tensor catalog. Details on the methods for selecting the preferred location/moment tensor solution, resolving fault plane ambiguity from conjugate fault planes, and selection of unknown parameters are given in AEA13.

PARAMETERS FROM MULTISEGMENT EVENTS

Several earthquakes with finite source models have multiple segments, for example, Kern County, San Fernando, Kobe, Landers, Denali, and Hector Mine. A segment within a multi-segment rupture is included if it had significant slip, generally $> \sim 50$ cm from a slip inversion. A multi-segment rupture consists of two or more contiguous planar quadrilaterals joined along their down-dip edges, sharing a single hypocenter. A multi-fault rupture occurs on two or more non-contiguous surfaces (which might each be multi-segment) and each surface having its own hypocenter. A list of the events with multi-segment ruptures along with the event parameters for each rectangular fault segments are included in Appendix A of AEA13.

The multi-segment/multi-fault nature of these events may have a significant impact on the resulting ground motions and fault/path parameters. Affected parameters include some directivity parameters, source-to-site distances, and the reported event parameters. Distance parameters were measured relative the closest rupture segment. For applications requiring fault dimensions and orientations (e.g., for directivity), the fault length and area were generally summed over the rectangular segments. Exceptions to this include events with significantly overlapping faults either along-strike or down-dip (e.g., 1971 San Fernando earthquake shown in Figure 6). The reported fault strike and dip was evaluated as the weighted average of these parameters across segments, with the weights proportional to the segment length and

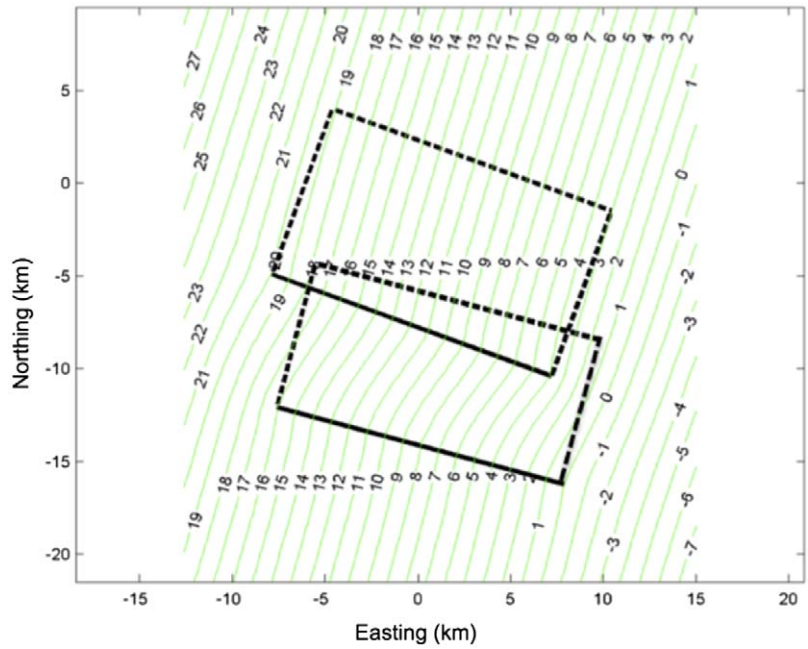

Figure 6. Contour lines (thin green) of U in km for the 1971 San Fernando earthquake calculated from the Spudich and Chiou GC. The solid and dashed lines represent the top of rupture and the surface projection of the finite fault planes respectively. The rupture length here is estimated as 20 km.

width, respectively. Most multi-segment events in the data set have consistent faulting mechanism (rake) across segments. Three earthquakes (Denali, El-Mayor Cucapah, and Wenchuan) have variable faulting mechanisms (rake) along strike. The reported fault rake was evaluated as the weighted average of these parameters across segments, with the weights proportional to the segment area.

EVENT CLASSIFICATION

A topic of interest in the NGA-West2 project was to evaluate possible differences between median ground motions from aftershocks and main shocks. Past experience has reached conflicting findings, in some cases finding different magnitude scaling for aftershocks relative to main shocks (Boore and Atkinson 1989, Atkinson 1993), while in other cases finding similar motions for similar magnitudes (Douglas and Halldórsson 2010). In the NGA-West1 project, GMPE developers either removed aftershocks from the data set (Boore and Atkinson 2008, Campbell and Bozorgnia 2008), included them with a term to account for a constant scale factor difference between main shock and aftershock motions (Abrahamson and Silva 2008, Chiou and Youngs 2008), or included them without a term accounting for a difference (Idriss 2008).

Past practice on event classification (as aftershock and main shock) was rather ad hoc and variable between GMPE developer teams. In NGA-West2, earthquakes are grouped into two classes based on their distance to the rupture plane of the main event and their time with respect to the main event. Following Wooddell and Abrahamson (2014), Class 1 (CL1) earthquakes are main shocks, triggered events, or foreshocks that occur off the surface projection of the main shock rupture plane, and Class 2 (CL2) earthquakes occur within or near the surface projection of the main shock rupture plane and within a time window for aftershocks. The CL1 and CL2 distinction for a given event following a CL1 main shock is based on the Gardner and Knopoff (1974) time window and a distance window. The distance window uses the shortest horizontal distance between the centroid of the event rupture surface and the closest point on the edge of the CL1 main shock rupture surface, which is denoted the centroid Joyner-Boore distance "CR_{JB}." The flatfile contains CL1 and CL2 event designations based on alternate CR_{JB} criteria. The cutoff CR_{JB} values used are 0 km, 2 km, 5 km, 10 km, 20 km, and 40 km. Further information on the CL1 and CL2 criteria are given in Wooddell and Abrahamson (2014).

SITE DATABASE

This section briefly describes key information and changes in the site database, a collection of metadata for each recording station, from the NGA-West1 project. A more detailed description of the site database, data collection, processing, and new protocols is included in SEA14 in this volume and Chapter 3 of AEA13.

The site database contains station information such as station coordinate, the time-averaged shear wave velocity in the upper 30 m (V_{S30}), various proxies used for V_{S30} estimation, and basin depth parameters $z_{1.0}$, $z_{1.5}$, and $z_{2.5}$ which define the depth in meters to shear-wave velocities of 1.0, 1.5, and 2.5 km/s. Each station has a unique station sequence number (SSN). Key contents of the site database merged into the flatfile include station name/ID, location, housing/siting conditions, V_{S30}, V_{S30} uncertainty, and the aforementioned depth

parameters. Relative to the NGA-West1 site database, the principal changes for NGA-West2 described in SEA14 are:

1. A substantial increase in the number of sites with V_{S30} based on measurements (24% of sites in NGA-West1; 49% of sites in NGA-West2). The increase in sites with measurements results principally from a major site investigation program in California (Yong et al. 2012) and the use of many sites in Japan, most of which have V_S profiles.
2. Basin depth parameters are re-evaluated from updated 3-D velocity models and shear wave velocity profiles. The re-evaluation was a result of updated basin models developed since NGA-West1.
3. Proxy values used for V_{S30} estimation have been substantially expanded using geology, geotechnical, geomorphology, and slope based methods. This was an extensive effort, the details of which are given in SEA14.
4. An updated method to infer V_{S30} from proxies, especially when multiple proxies are available. In the updated approach, proxies that fit data for a given region better with respect to lack of bias and low dispersion are given high weight (SEA14).
5. An updated evaluation of the epistemic variability of mean $\ln(V_{S30})$ using sites having multiple credible measured V_S profiles. As described in SEA14, the epistemic variability is computed as the log standard deviations of V_{S30} for sites with multiple measurements.

PROPAGATION PATH TABLE

The propagation path table is a set of computed site-source path parameters used by directivity modelers (Spudich et al. 2013) and GMPE developers. Unlike the source table and site database, the propagation path metadata parameters are not collected but are calculated based on the geometry of the finite fault rupture plane and the location of the station. Data in the propagation path table include six source-to-site distance measures (epicentral distance, R_{EPI}; hypocentral distance, R_{HYP}; shortest distance to rupture plane, R_{RUP}; Joyner-Boore distance, R_{JB}; distance measured perpendicular to the fault strike from the surface projection of the up-dip edge of the fault plane, R_X; and distance measured parallel to the fault strike from the midpoint of the surface projection of the fault plane, R_Y), a hanging wall indicator, radiation pattern coefficients, source-to-site azimuth, and directivity parameters. Path parameters introduced in the NGA-West2 project include the local strike-parallel direction, the distance measure R_Y, and directivity parameters given in Spudich et al. (2014).

The computation of distance and directivity parameters is non-trivial for events with multiple-segments or faults. Distances R_X and R_Y were computed from U and T using the generalized coordinate system (GC) which provides a smooth transition in the distances around bends in the fault segments (i.e., changes in strike) developed by Spudich and Chiou (2008). Distance R_X is equal to T while $R_Y = \mathrm{U} - \mathrm{L}/2$ (where L is the rupture length). For a single fault plane U and T are defined as the fault parallel and transverse distance from the upper left corner of the fault to the site. An example contour plot of U for the 1971 San Fernando finite fault planes is given in Figure 6. All other distances were based on the closest segment or minimum distance when estimated for each segment. Information on how directivity models incorporate multi-segment or multiple faults ruptures is given in Spudich et al. (2013, 2014).

The strike-parallel direction at a recording site is ambiguous when the finite fault model has multiple segments. The consensus definition reached by the NGA developers is that the local strike-parallel direction is the fault strike direction averaged over a 20 km (or less) stretch of fault segment beginning at the closest point on the fault and extending towards the epicenter.

RECORD CATALOG

The record catalog contains computed ground motion intensity measures (IMs) and filter corner information for each ground motion recording included in the database. The IMs compiled are pseudo-spectral accelerations and peak ground motion measures (PGA, PGV, and PGD), Arias intensities, and durations calculated from the Husid (1969). Each record is given a unique record sequence number (RSN). IMs are provided in uniformly formatted text files containing the three-component as-recorded pseudo-acceleration response spectra and peak parameters, orientation-independent horizontal spectra and peak parameters (RotDnn, Boore 2010), and Arias intensity and related duration parameters. The available pseudo-spectral accelerations are provided at 111 periods ranging from 0.01 s to 20 s and for 11 different damping ratios ranging from 0.5% to 30%.

ROTD50

GMPE development in the NGA-West2 project for the horizontal components of ground motion use the RotDnn spectra (Boore 2010). RotDnn spectra are computed from rotated time series where "nn" represents the fractile of the spectra sorted by amplitude and "D" indicates that rotation angle depends on the period of the oscillator. The horizontal component GMPEs are computed using RotD50, which is the 50th percentile (median) amplitude among all possible azimuths. The motivation for its use comes from the simplicity of its definition and calculation. Shahi and Baker (2014) describe how 50th percentile ground motions can be adjusted for the maximum component, RotD100.

RECORD PROCESSING

A brief description of the processing decisions is discussed here but the full methodology is described in AEA13. Staff at PEER and Pacific Engineering and Analysis (PE&A) performed record processing of the NGA-West2 database. If possible, the processing began with "raw" (unprocessed) accelerograms collected from various agencies around the world or velocity time series (from broadband high gain instruments in California networks). Strong-motion record processing has two major objectives: (1) reduction of low- and high-frequency noise in the recorded time series and (2) correction for the instrument response. The processing methodology, applied consistently to all raw data, remains generally unchanged from the NGA-West1 project except for the filter type. The NGA-West2 processed time series used an acausal Butterworth filter (discussed in Douglas and Boore 2011) where the NGA-West1 used causal Butterworth filters (Chiou et al. 2008). The selection of filters is guided by an examination of the Fourier amplitude spectrum on a component-by-component basis at both low and high frequencies. The useable frequency bandwidth depends on the selected high- and low-pass filter corners and the number of poles in the selected Butterworth filter. Due to the filter response shape, a factor of 1.25 (typical value), along with the corner frequencies, are used to determine

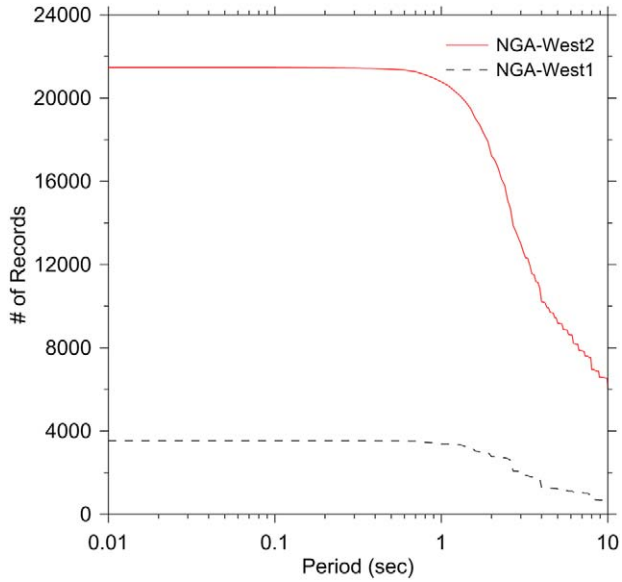

Figure 7. Comparison of number of records with a highest useable period versus period.

the useable bandwidth. A more complete description on the filter selection is included in AEA13.

All previously processed records in the NGA-West1 database were carefully reviewed and reprocessed in an effort to expand the useable bandwidth especially at long periods. As shown in Figure 7, the reprocessing and addition of new records has increased the number records with a useable bandwidth up to a period of 10 s by a factor of 10 from the NGA-West1 data set. Selected filter corners are provided for each component. We also visually inspected the raw and processed time series to evaluate effects of instrument triggering and pre-event memory (i.e., missing P-wave onset) from recordings that were from triggered (e.g., analog or digital with inadequate pre-event memory) instrumentation, rather than continuously recorded instrumentation. For example, if the recording does not include the P-wave onset because the instrument was triggered after the P-wave onset, we flagged that recording as late-P trigger. Visual inspections are similarly performed for S-wave onset.

NGA-WEST2 FLATFILE

Subsets of the pertinent information in all the tables are compiled in the database flatfile, which is used for GMPE development. The three ID numbers (RSN, SSN, and EQID) facilitate a linkage between the tables and aid in the creation of the flatfile. An Electronic Supplement to this paper contains the flatfile in Excel format and an explanation of column headings.

SUMMARY AND CONCLUSIONS

This paper describes the data included in the NGA-West2 ground motion database. The database includes updates to the NGA-West1 database (processing; improved metadata),

additional recordings from large magnitude global events ($\mathbf{M} > 6$) since 2003, and recordings from small-to-moderate-magnitude events ($3.0 < \mathbf{M} < 5.45$) from California between 1998 and 2011. The number of three-component recordings in the NGA-West2 database is 21,336 (some recordings have only horizontal components) from 599 shallow crustal events. The included time series were uniformly processed using the PEER methodology using an acausal filter. Spectral accelerations at eleven different damping levels (0.5%, 1%, 2%, 3%, 5%, 7.5%, 10%, 15%, 20%, 25%, and 30%) were computed for periods ranging from 0.01 s to 20 s. The horizontal-component intensity measure reported is RotD50.

Metadata were collected to define important event, path, and station information used during GMPE development and for other engineering applications. Throughout the project, numerous experts and the NGA-West2 GMPE developers systematically checked and reviewed the data. New metadata added to the flatfile include an event classification for main shock versus aftershock, additional directivity parameters, late-trigger flags, and a new distance parameter, R_Y. Site parameters have been substantially updated and expanded, with many more V_{S30} values now coming from measurements, as well as more complete and transparent application of proxies to estimate V_{S30} and its uncertainty in the absence of velocity measurements.

As in the NGA-West1 project, summary metadata tables, flatfiles, and time series data files have been created for dissemination to the research and engineering communities.

ACKNOWLEDGMENTS

This project was sponsored by the Pacific Earthquake Engineering Research Center (PEER) and funded by the California Earthquake Authority (CEA), California Department of Transportation, and the Pacific Gas & Electric Company. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the above-mentioned agencies. Many people and organizations contributed data to this effort. Space precludes a complete listing of them here, but we refer the interested reader to the Acknowledgements section of AEA13.

REFERENCES

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

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

Allen, C. R., Amand, P., Richter, C. F., and Nordquist, J. M., 1965. Relationship between seismicity and geologic structure in the Southern California region, *Bull. Seismo. Soc. Am.* **55**, 753–797, available at http://www.bssaonline.org/content/55/4/753.full.pdf.

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., and Donahue, J. L., 2013. *PEER NGA-West2 Database*, PEER Report No. 2013/03, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, 134 pp.

Asano, K., and Iwata, T., 2009. Surface rupture process of the 2004 Chuetsu, Mid-Nagata Prefecture, Japan, earthquake inferred from waveform inversion with dense strong-motion data, *Bull. Seismo. Soc. Am.* **99**, 123–140.

Atkinson, G. M., 1993. Source spectra for earthquakes in eastern North America, *Bull. Seismo. Soc. Am.* **83**, 1778–1798, available at http://www.bssaonline.org/content/83/6/1778.abstract.

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

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

Beavan, J., Samsonov, S., Motagh, M., Wallace, L., Ellis, S., and Palmer, N., 2010. The Darfield (Canterbury) earthquake: Geodetic observation and preliminary source model, *Bull. New Zealand Soc. Earthq. Eng.* **43**, 228–235, available at http://insar.ca/pubs/beavan_nzseeb2010.pdf.

Beavan, J., Fielding, E., Motagh, M., Samsonov, S., and Donnelly, N., 2011. Fault location and slip distribution of the 22 February 2011 Mw 6.2 Christchurch, New Zealand, earthquake from geodetic data, *Seismo. Res. Lett.* **82**, 789–799.

Benetatos, C., and Kiratzi, A., 2006. Finite-fault slip models for the 15 April 1979 (M_w 7.1) Montenego earthquake and its strongest aftershock of 24 May 1979 (M_w 6.2), *Tectonophysics* **421**, 129–143.

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

Boore, D. M., and Atkinson, G. M., 1989. Spectral scaling of the 1985–1988 Nahanni, Northwest Territories, earthquakes, *Bull. Seismo. Soc. Am.* **79**, 1736–1761, available at http://www.bssaonline.org/content/79/6/1736.short.

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

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

Chiou, B. S.-J., and Youngs, R. R., 2006. *Chiou and Youngs PEER-NGA Empirical Ground Motion Model for the Average Horizontal Component of Peak Acceleration, Peak Velocity, and Pseudo-Spectral Acceleration for Spectral Periods of 0.01–10 sec*, Interim Report submitted to PEER, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

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

Chiou, B. S.-J., Darragh, R. B., Gregor, N., and Silva, W. J., 2008. NGA project strong-motion database, *Earthquake Spectra* **24**, 23–44.

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

Cirella, A., Piatanesi, A., Tinti, E., and Cocco, M., 2008. Rupture process of the 2007 Niigata-ken Chuetsu-oki earthquake by non-linear joint inversion of strong motion and GPS data, *Geophys. Res. Lett.* **35**, L16306.

Custódio, S., Liu, P. P., and Archuleta, R. J., 2005. The 2004 M_w 6.0 Parkfield, California, earthquake: Inversion of near-source ground motions using multiple data sets, *Geophys. Res. Lett.* **32**, L23312.

Douglas, J., and Halldórsson, B., 2010. On the use of aftershocks when deriving ground motion prediction equations, in *Proc, 9th U.S. Nat. & 10th Canadian Conf. Earthq. Eng.*, Paper 220, 10 pgs., Toronto, ON, Canada.

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

Dreger, D., Gee, S. L., Lombard, P., Murray, M. H., and Romanowicz, B., 2005. Rapid finite-source analysis and near-fault strong-ground motions: Application to the 2003 M_w 6.5 San Simeon and 2004 M_w 6.0 Parkfield earthquakes, *Seismo. Review Lett.* **76**, 40–48.

Gardner, J. K., and Knopoff, L., 1974. Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?, *Bull. Seismo. Soc. Am.* **64**, 1363–1367.

Hardebeck, J. L., Boatwright, J., Dreger, D., Goel, R., Graizer, V., Hudnut, K., Ji, C., Jones, L., Langbein, J., Lin, J., Roeloffs, E., Simpson, R., Stark, K., Stein, R., and Tinsley, J. C., 2004. Preliminary report on the 22 December 2003 M_w 6.5 San Simeon, California earthquake, *Seismo. Review Lett.* **75**, 155–172.

Hauksson, E. W., Yang, W., and Shearer, P. M., 2012. Waveform relocated earthquake catalog for Southern California (1981 to 2011), *Bull. Seismo. Soc. Am.* **102**, 2239–2244.

Hikima, K., and Kodetsu, K., 2005. Rupture process of the 2004 Chuetsu (mid-Niigata prefecture) earthquake, Japan: A series of events in a complex system, *Geophys. Res. Lett.* **32**, L18303, available at http://onlinelibrary.wiley.com/doi/10.1029/2005GL023588/full.

Holden, C., 2011. Kinematic source model of the 22 February 2011 M_w 6.2 Christchurch earthquake using strong motion data, *Seismo. Res. Lett.* **82**, 783–788.

Hough, S. E., and Dreger, D. S., 1995. Source parameters of the 23 April 1992 M 6.1 Joshua Tree, California, earthquake and its aftershocks: Empirical Green's function analysis of GEOS and TERRAscope data, *Bull. Seismo. Soc. Am.* **85**, 1576–1590, available at http://www.bssaonline.org/content/85/6/1576.short.

Husid, L. R., 1969. Características de terremotos, Análisis general, Revista del IDIEM 8, Santiago del Chile, 21–42.

Idriss, I. M., 2008. An NGA empirical model for estimating the horiztonal spectral values generated by shallow crustal earthquakes, *Earthquake Spectra* **24**, 217–242.

Jackson, J., Bouchon, M., Fielding, E., Funning, G., Ghorashi, M., Hatzfeld, D., Nazari, H., Parsons, B., Priestley, K., Talebian, M., Tatar, M., Walker, R., and Wright, T., 2006. Seismotectonic, rupture process, and earthquake-hazard aspects of the 2003 December 26 Bam, Iran, earthquake, *Geophys. J. Int.* **166**, 1270–1292.

Ji, C., Larson, K. M., Tan, Y., Hudnut, K. W., and Choi, K., 2004. Slip history of the 2003 San Simeon earthquake constrained by combining 1-Hz GPS, strong motion and teleseismic data, *Geophys. Res. Lett.* **31**, L17608.

Koketsu, K., Yokota, Y., Ghasemi, H., Hikima, K., Miyake, H., and Wang, Z., 2008. *Source Process and Ground Motions of the 2008 Wenchuan Earthquake*, Investigation Report of the 2008 Wenchuan Earthquake, China, Grant-in-Aid for Special Purposes of 2008, MEXT No. 2090002.

Miyake, H., Koketsu, K., Hikima, K., Shinohara, M., and Kanazawa, T., 2010. Source fault of the 2007 Chuetsu-oki, Japan, earthquake, times, *Bull. Seismo. Soc. Am.* **100**, 384–391.

Piatanesi, A., Cirella, A., Spudich, P., and Cocco, M., 2007. A global search inversion for earthquake kinematic rupture history: Application to the 2000 western Tottori, Japan earthquake, *J. Geophys. Res.* **112**, B07314, available at http://onlinelibrary.wiley.com/doi/10.1029/2006JB004821/full.

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

Scognamiglio, L., Tinti, E., Michelini, A., Dreger, D. S., Cirella, A., Cocco, M., Mazza, S., and Piatanesi, A., 2010. Fast determination of moment tensors and rupture history: What has been learned from the 6 April 2009 L' Aquila earthquake sequence, *Seismo. Res. Lett.* **81**, 892–906.

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

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

Spudich, P., Bayless, J. R., Baker, J. W., Chiou, B. S.-J., Rowshandel, B., Shahi, S. K., and Somerville, P., 2013. *Final Report of the NGA-West2 Directivity Working Group*, PEER Report No. 2013/09, Pacific Earthquake Engineering Research Center, University of California, Berkeley, 129 pp.

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

Spudich, P., Rowshandel, B., Shahi, S. K., Baker, J. W., and Chiou, B. S.-J., 2014. Comparison of NGA-West2 directivity models, *Earthquake Spectra* **30**, 1199–1221.

Stewart, J. P., Lanzo, G., Pagliaroli, A., Scasserra, G., DiCapua, G., Peppolini, S., Darragh, R., and Gregor, N., 2012. Ground motion recordings from the M_w 6.3 2009 L'Aquila earthquake in Italy and their engineering implications, *Earthquake Spectra* **28**, 317–345.

Suzuki, W., Aoki, S., and Sekiguchi, H., 2010. Rupture process of the 2008 Iwate-Miyagi Nairku, Japan, earthquake derived from near-source strong-motion records, *Bull. Seismo. Soc. Am.* **100**, 256–266.

Waldhauser, F., and Schaff, D., 2008. Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.* **113**, B08311, available at http://onlinelibrary.wiley.com/doi/10.1029/2007JB005479/full.

Wang, W., Zhao, L., Li, J., and Yao, Z.-X., 2008. Rupture process of the MS 8.0Wenchuan earthquake of Sichuan, China, *Chinese J. Geophys.* **51**, 1403–1410.

Wei, S., Fielding, E., Leprince, S., Sladen, A., Avouac, J-P., Helmberger, D., Hauksson, E., Chu, R., Simons, M., Hudnut, K., Herrring, T., and Briggs, R., 2011. Superficial simplicity of the 2010 M_w 7.2 El Mayor-Cucapah earthquake of Baja, California, *Nature Geoscience* **4**, 615–618.

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

Yang, W., Hauksson, E., and Shearer, P., 2012. Computing a large refined catalog of focal mechanisms for Southern California (1981–2010): Temporal Stability of the Style of Faulting, *Bull. Seismo. Soc. Am.* **102**, 1179–1194.

Yong, A., Hough, S. E., Iwahashi, J., and Braverman, A., 2012. Terrain-based site conditions map of California with implications for the contiguous United States, *Bull. Seismol. Soc. Am.* **102**, 114–128.

(Received 9 July 2013; accepted 26 January 2014)