17th World Conference on Earthquake Engineering, 17WCEE
Sendai, Japan - September 13th to 18th 2020

DATA RESOURCES FOR NGA-SUBDUCTION PROJECT

V. Contreras (1), S. Mazzoni (2), T. Kishida (3), S.K. Ahdi (4), R.B. Darragh (5), R.R. Youngs (6), B.S.J. Chiou (7), N. Kuehn (8), K. Wooddell (9), Y. Bozorgnia (10), and J.P. Stewart (11)

(1) *Ph.D. Student, University of California-Los Angeles, vcontreras@ucla.edu*
(2) *Research Scientist and Lecturer, University of California-Los Angeles, smazzoni@ucla.edu*
(3) *Assistant Professor, Khalifa Univ. of Science and Technology-UAE, tadahiro.kishida@ku.ac.ae*
(4) *Associate, Exponent Failure Analysis, Los Angeles, CA, sahdi@exponent.com*
(5) *Senior Seismologist, Pacific Engineering and Analysis, El Cerrito, CA, bbal@cwnet.com*
(6) *Principal Engineer, Wood E&IS, Oakland, CA, bob.youngs@woodplc.com*
(7) *California Department of Transportation, Sacramento, CA, brian.chiou@dot.ca.gov*
(8) *Research Scientist, University of California-Los Angeles, kuehn@g.ucla.edu*
(9) *Pacific Gas and Electric Co, San Francisco, CA, katie.wooddell@gmail.com*
(10) *Professor, University of California-Los Angeles, yousefbozorgnia@g.ucla.edu*
(11) *Professor, University of California-Los Angeles, jstewart@seas.ucla.edu*

Abstract

A relational database was developed over a five-year period to support ground motion model (GMM) development for the Next Generation Attenuation-Subduction (NGA-Sub) project. The relational database has components that interact according to a database schema, including a source and path component used to describe attributes of seismic sources in global subduction regions and to compute source-to-site distances, a site component that describes attributes of sites where recordings have been made, and a ground motion component.

The source component of the database has information for 1880 earthquakes, mainly from the following regions: the Pacific Northwest region of North America, Alaska and the Aleutian Islands, Japan, Taiwan, New Zealand, South America, Central America, and Mexico. Of the 1880 earthquakes, 88 have finite fault models (FFMs) from the literature that were systematically reviewed, distilled to one more rectangular shapes, and trimmed according to procedures based on percentage of total slip. For earthquakes without FFMs, a simulation routine is used to represent finite fault effects required for distance calculations. This simulation routine was adjusted and made more uniform in its application than in prior NGA projects. All earthquakes are classified as interface, intraslab, shallow crustal, or outer rise, using uniform protocols developed for this project. All earthquakes are also assigned class designations adapted from a prior NGA project for active regions, that allows foreshock, mainshock, and aftershock events to be distinguished.

The site component of the database is described in a companion paper (Ahdi et al. 2020 [1]).

The ground motion component of the database consists of median – and maximum – horizontal component peak parameters (peak ground acceleration, PGA and peak ground velocity, PGV) and pseudo-spectral accelerations (PSa) at 111 oscillator periods and 11 damping ratios. Response spectra were also computed for the vertical component. Fourier amplitude spectra (FAS) and duration metrics were also computed. The ground motion recordings were obtained from collaborating organizations world-wide as uncorrected (Vol 1) digital recordings, that were corrected (component-specific low – and high – pass filters and baseline correction, as needed) following Pacific Earthquake Engineering Research Center (PEER)/NGA protocols.

The relational database operates on each of these (and other) database components to dynamically draw relevant parameters into a single file, known as a flatfile, that is used by researchers engaged in GMM development. The flatfiles used in model development are being published with the NGA-Sub GMMs as products of the NGA-Sub project.

Keywords: Ground motion model; relational database; subduction zones

1. Introduction

NGA-Sub is one in a series of NGA projects directed towards database and GMM development for applications in seismic demand characterization. Whereas prior projects had targeted shallow crustal earthquakes and active tectonic regions (NGA-West1, Power et al. 2008 [2]; NGA-West2, Bozorgnia et al. 2014 [3]) and stable continental regions (NGA-East, Goulet et al. 2018 [4]), NGA-Sub is the first to specifically address subduction zones, which are a dominant source of seismic hazard in many regions globally. The objectives of the NGA-Sub project are (1) to develop a state-of-the-art database, (2) to develop a series of GMMs that operate over the parameter range (magnitudes, event types, distances, and site condition) required for typical hazard applications; and (3) to provide guidance on application of the models, including epistemic uncertainties. This paper describes the development of data resources for the NGA-Sub project.

Regionalization of certain ground motion attributes is an important feature of NGA-Sub. Fig. 1 shows a global map with locations of the strong motion recording stations in the NGA-Sub project. As presented in this figure, the database has been organized in seven major regions: Alaska (ALK), Cascadia (CAS), Central America and Mexico (CAM), Japan (JPN), New Zealand (NZL), South America (SAM), and Taiwan (TWN). The last set of truly global GMMs is that from NGA-West1. Regionalization in anelastic attenuation and site effects was introduced in NGA-West2 and was integral to NGA-East as well (mostly in relation to path effects). Some of the NGA-Sub GMMs are region-specific (e.g., Si et al. 2020 [5]; Youngs et al. 2020 [6] for Japan), while the remainder are global but include regional adjustments for various source, path, and site effects.

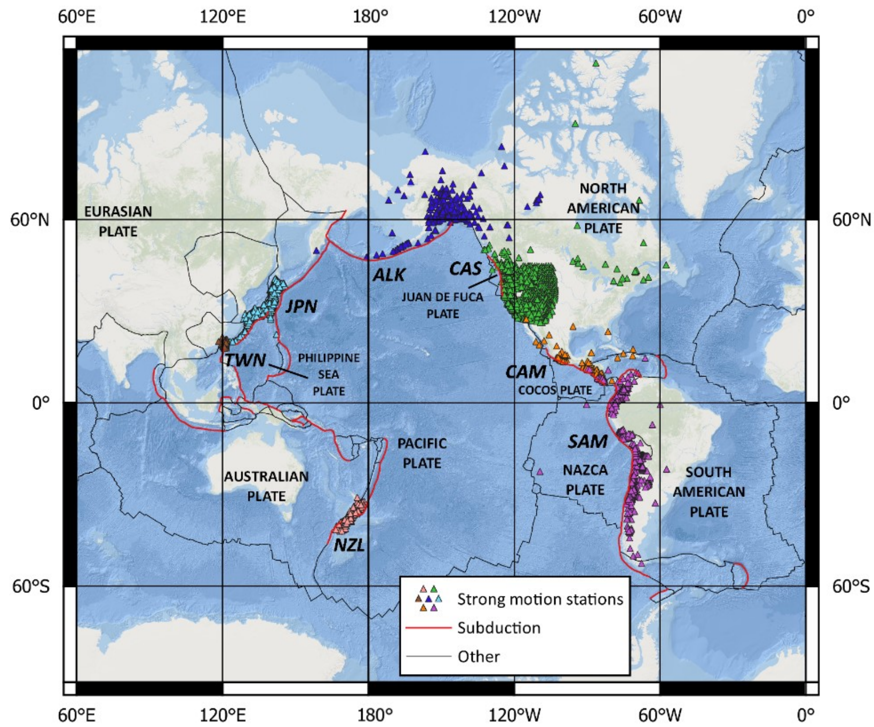

Fig. 1 – Locations of strong motion recording stations with recordings in the NGA-Sub database

Many other regions are known to have subduction zone earthquakes, but are not represented in this database (e.g., Indonesia, Greece, Calabria/Italy). These omissions were not accidental (we were aware of the significance of subduction earthquake hazards). However, data from these additional regions was not incorporated into the database either because we anticipated not having ready access to sufficient data to benefit the project or we anticipated that the necessary data simply were not available.

2. Relational Database

An extensive effort was undertaken to coordinate the NGA-Sub project with local agencies in areas affected by subduction earthquakes, with an emphasis on the seven regions presented in Fig. 1: Alaska, Cascadia, Central America and Mexico, Japan, New Zealand, South America, and Taiwan. This included communications to identify relevant sources of ground motion data and either (*i*) identifying public repositories of data that could be accessed by the project or (*ii*) forming collaborative agreements to enable data sharing between the NGA-Sub project and individual network operators. We obtained digital but unprocessed versions of records (Vol 1) from either accelerometers (accelerograms) or seismometers (velocity time series). The agencies from each of the above regions from which we obtained uncorrected records are summarized in Section 4.

As data was obtained, it was initially organized into a series of spreadsheets, as had been the practice in prior NGA projects. However, due to the enormous size of the collected data, we ultimately organized it into a relational database consisting of several tables containing various data, metadata, and outputs of various codes that perform calculations to compute desired quantities (e.g., intensity measures, distances). The database developed for NGA-Sub is a relational database, meaning that it has a well-defined data structure and can be queried using structured query language (SQL). Previous NGA projects have applied the term "database" to collections of spreadsheet files that were linked in an ad hoc manner using Excel macros (Chiou et al. 2008 [7]; Ancheta et al. 2014 [8]). In contrast, a relational database is organized into a schema that describes the tables, fields, and relationships among tables. Tables are collections of information organized into fields (or columns). The contents of tables (i.e., each row) are identified using keys. Every entry in the database (i.e., a given row within a field) is assigned a primary key that uniquely identifies it. In some cases, a field from one table might appear in another table to relate the two tables. In such cases, the primary key from the host table appears as a foreign key in the other table to map the relationship. While in some applications a key, primary or foreign, may consist of a unique combination of more than one field, for NGA-Sub keys are associated with a single field.

As shown in Fig. 2, the NGA-Sub database broadly contains information on source, site, path, ground motions (contains instrument/time-series metadata), and intensity measures. This information is contained in 23 tables. The contents of fields in these tables are mapped via database keys. There are three main keys in the database: (1) the Record Sequence Number (NGAsubRSN) applies to a given ground motion, (2) NGAsubEQID applies to an earthquake event, and (3) NGAsubSSN applies to a recording site.

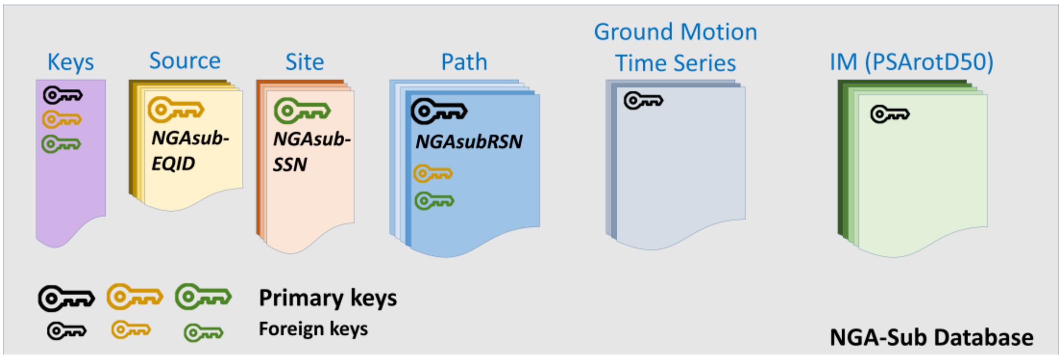

Fig. 2 – Components of NGA-Sub relational database

As with prior NGA projects, model developers and others mostly interact with the data using a flatfile, which is a single file extracted from the database containing all fields of interest (e.g. pseudo-spectral accelerations at a certain oscillator damping level). The flatfile can be readily generated from the relational database using an SQL command (MS Access was used in this project, using Visual Basic to write SQL commands). Such flatfiles are a time-stamped output of the database.

The fields within the NGA-Sub database are of three types: instrument recordings, various metadata, and computed data. The ground motion time series exist as a series of files that are used in database development, but which are not part of the database tables – the filepath, as well as additional characteristics of these files, are stored in the database. Digitized and processed instrument recordings are stored in individual ASCII files of acceleration, velocity and displacement histories; the files have headers structured to convey information on the causative event, the station that produced the recording, and the time step. Record processing is briefly described in Section 4.

Metadata defines attributes of a record, such as earthquake location and magnitude, location and site conditions at the recording site, characteristics of the recording instrument, and attributes of the path between the earthquake source and the recording station. There are two types of metadata: (1) independent metadata, which are either measured or taken directly from literature or catalogs, and (2) dependent metadata, which are computed from the independent metadata fields. Source-to-site distance metrics and fault geometry are an example of dependent metadata. Damped elastic response spectra, Fourier amplitude spectra, and duration metrics are an example of the third type of data within the NGA-Sub database – computed data, often referred to as Intensity Measures, IMs. These IMs are computed from the processed instrument recordings and are stored in individual tables within the database. Table 1 presents a summary of how the metadata are organized in NGA-Sub; further details can be found in Mazzoni et al. 2020 [9].

Table 1 – NGA-Sub metadata structure

Metadata	**Description**
Key	Single table that contains primary and foreign keys. The primary purpose of this table is to map the source (NGAsubEQID) and site (NGAsubSSN) keys to the primary record key (NGAsubRSN).
Source	Primary key is the Earthquake Identification Number (NGAsubEQID), a unique value which was assigned to each event. Main fields are: region, hypocentral geodetic coordinates (latitude, longitude), hypocentral depth, date, time, seismic moment (M_0), moment magnitude (**M**), preferred moment tensor parameters (rake, strike, slip), focal mechanism, event type, FFM flag, number of rectangles, coordinates of upper left corner of fault as viewed from hanging wall, along-strike length (L), down-dip width (W), event class flag, volcanic arc flag.
Site and Station	Primary key is the Station Sequence Number (NGAsubSSN), a unique value which is assigned to each site. Main fields are: site name and station ID; station location (including geodetic coordinates, elevation, depth, housing); recommended V_{S30} (m/s); proxies used for V_{S30} prediction, as available, including surface geology, ground slope, geomorphic terrain class; basin depth information, including depth to a particular V_S horizon (values of 1.0 and 2.5 km/s are used).
Path	Primary key is the Record Sequence Number (NGAsubRSN). Main fields are: various distance parameters (R_{rup}, R_{JB}, R_X, R_Y, R_{Y0}, R_{epi}, R_{hyp}, R_{rms}), location on fault surface from which the closest distance is measured (geodetic coordinates and depth), directivity parameters, maximum recommended distance R_{max}.
Ground Motion Times Series and IMs	Primary key for a given ground motion recording is the Record Sequence Number (NGAsubRSN). Main fields are: as-recorded horizontal azimuths, processing details (low-cut and high-cut corner frequencies), computer path to access time series, time step, pseudo-spectral acceleration (components H1, H2, V, RotD0, RotD50, RotD100), and Arias Intensity, FAS, and CAV (components H1, H2, V).

3. Source and Path Metadata

For the earthquakes considered in the NGA-Sub project, a series of descriptive source parameters is needed to support GMMs development. For a given ground motion recording site, source parameters also allow path parameters to be defined, so the issues of source and path are strongly linked in the NGA-Sub database development. This section describes the manner by which those parameters are compiled and assembled in a source and path database file.

3.1 Summary of NGA-Sub events

Fig. 3 shows a global map with locations of the epicenters in the NGA-Sub database, along with the main tectonic plates and plate boundaries as defined in a digital model assembled by Bird 2003 [10]. The boundaries shown in red are mostly classical oceanic-beneath-continental subduction boundaries whereas other plate boundaries are shown in black. The NGA-Sub source and path database contains source information on event date, origin time, seismic moment, moment magnitude, hypocenter location, nodal planes, and finite-fault geometric parameters, among other parameters. A key aspect is classification of each earthquake into one of four types: interface, intraslab, shallow crustal, or outer-rise. While the NGA-Sub project focuses on subduction-zone events (i.e., interface and intraslab), there is an important number of shallow crustal events and a small number of outer-rise events. The presence of these events in the database is a byproduct of the manner in which the database was developed; ground motion recordings were collected in the seven study regions in Fig. 3 without establishing a priori earthquake type. Once the data had been collected and processed, instead of discarding data from non-subduction sources, it was retained and flagged based on the event-type.

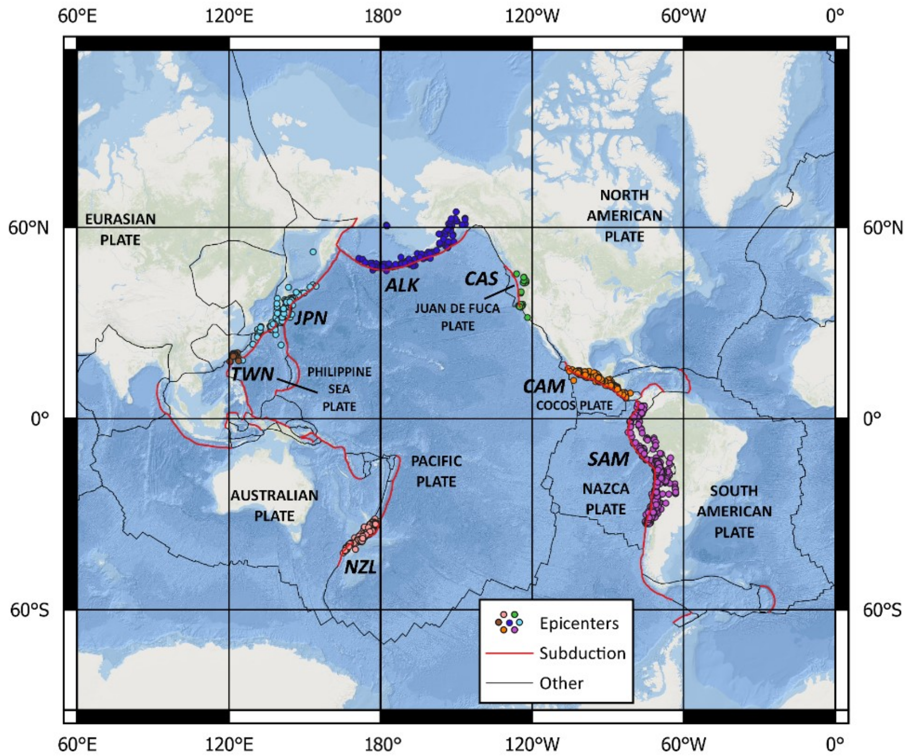

Fig. 3 – Locations of epicenters in the NGA-Sub database. Regions are indicated by color of the epicenters and labeled as ALK (Alaska), CAS (Cascadia), CAM (Central America and Mexico), JPN (Japan), NZL (New Zealand), SAM (South America), and TWN (Taiwan).

Fig. 4 presents the locations of earthquakes included in the NGA-Sub database for the example of the Japan region, with differentiation by magnitude and type of earthquake. In the north, the northwest-dipping Pacific Plate subducts beneath the Okhotsk Plate (an extension of the North American Plate) at the Japan Trench. To the west, in the Sea of Japan, a convergent plate boundary occurs between the Okhotsk Plate to the east and the Amur plate to the west. Near the middle of the main island (Honshu), the Pacific Plate's western boundary bends south and east, and the Philippine Sea Plate subducts beneath Japan at the Nankai Trough. The northern sector has many interface and intraslab events in the vicinity (and inboard of) the Japan trench, including the 2011 **M**9.12 Tohoku earthquake (an interface event). The largest intraslab earthquake in this region and proximate to Japanese islands is the 1994 **M**8.28 Hokkaido Tohu-oki earthquake. Interface earthquakes also occur west of the island at the convergent boundary in the Sea of Japan. The southern sector has fewer events, with the region east of Osaka having primarily intraslab events, while the southern terminus of Honshu (near Kagoshima) has a series of interface and intraslab events.

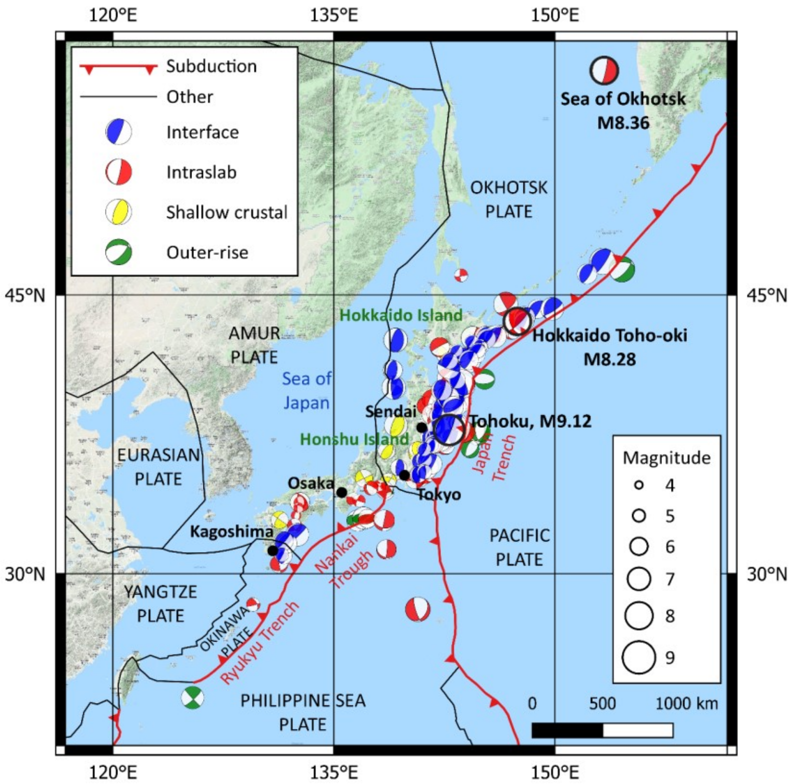

Fig. 4 – Epicentral locations of earthquakes with recordings in Japan.

The total number of events in the source database with an assigned earthquake identification number (NGAsubEQID) is 1,880. QA procedures eliminate some events, mostly because of missing magnitudes or hypocenter locations, decreasing the number of potentially usable earthquakes to 1,782. Further screening to remove events without an assigned event-type reduces the number of events to 991. Fig. 5(a) shows the distribution of these 991 events by region. South America is the region which contributes the largest number of subduction earthquakes, followed by Japan and New Zealand. The events tallied in Fig. 5 are those in the NGA-Sub database with event-type classifications. For most regions, subduction-type events (interface or intraslab) are dominant; New Zealand is an exception, with 135 shallow crustal and 139 subduction earthquakes. Fig. 5(b) shows the event-type distribution. The dataset is dominated by interface and intraslab earthquakes which are nearly evenly distributed. A significant number (221) of shallow crustal earthquakes are present in the database, mostly from New Zealand. While not directly useful for NGA-Sub modeling, these data were retained in the database. The contribution of outer-rise earthquakes is small. Fig.5(c) shows the distribution of recordings by event-type, which generally mirrors the event distribution.

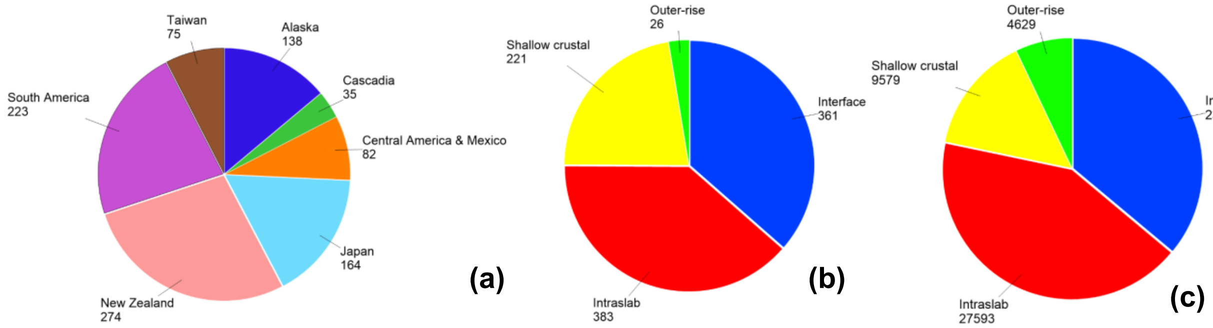

Fig. 5 – (a) Regional distribution of earthquakes with event-type classifications, (b) Distribution of the events by type of earthquake, (c) Distribution of the recordings by type of earthquake.

3.2 Source parameters for events with FFMs

We identified FFMs mainly by reviewing compilations of past studies in the literature presented at the following websites: a) SRCMOD website, available at http://equake-rc.info/SRCMOD/, b) Source Models of Large Earthquakes, Caltech Tectonic Observatory, available at http://www.tectonics.caltech.edu/slip_history/index.html, c) Rupture processes of global large earthquakes (M > 7), UC Santa Barbara, available at http://www.geol.ucsb.edu/faculty/ji/big_earthquakes/home.html (all websites last accessed Nov 2019). We also performed independent literature searches for the largest magnitude events (2010 **M**8.81 Maule, Chile and 2011 **M**9.12 Tohoku, Japan) and other recent, large events, some of which occurred contemporaneously with the data compilation for NGA-Sub (e.g., the 2001 **M**8.41 Arequipa earthquake in Southern Peru and the 2007 **M**7.75 Tocopilla, 2014 **M**8.15 Iquique, and 2015 **M**8.31 Illapel earthquakes in Chile).

For some earthquakes, more than one FFM is available in the literature. The three major considerations in evaluation of FFMs are (1) model is generated using default (automated) procedures vs. an inversion process managed and interpreted by experts; (2) the data sources considered in the inversion; and (3) peer-review, or lack thereof, of the model and the process by which is was derived. The data sources used in FFM development can include: Permanent crustal displacement caused by the earthquake, typically measured from GPS sensors, InSAR, or measurements of on-land elevation change (typically in coastal areas); teleseismic waveforms from global network; broadband ground motion sensors in reasonably close proximity to the source; tsunami-related data (run up heights, wave heights as measured by ocean buoys; spatial distribution of aftershocks, typically within 24 to 60 hrs. of the mainshock event. We prefer FFMs that have been reviewed/developed by experts (not preliminary or automatic solutions), have been developed using multiple data sources (inclusive of ground motion data, preferably from proximate stations), and have appeared in peer-reviewed documents.

Once a published FFM model is selected, it is typically necessary to apply some trimming of the rupture dimensions. This is important because faults are often set as large geometric objects at the outset of the inversion so as to avoid "missing" areas of potential rupture. As a result, the inverted fault may contain broad regions with relatively little slip, in addition to concentrated areas of high slip. This need for trimming is not unique to NGA-Sub, and was addressed earlier in the NGA-West1 project (Power et al., 2008). At that time, on average, a threshold of 50 cm of slip was generally applied, meaning that portions of the fault having slip below this value were trimmed (excluded) in the development of representative fault geometries used for distance calculations. Similar procedures were subsequently used in NGA-West2. Because the amounts of slip on subduction sources can be very large relative to the crustal sources considered in NGA-West1 and NGA-West2, we were concerned that the 50 cm threshold used in previous NGA projects may not provide a reliable basis for fault trimming in all cases. Accordingly, we re-examined this issue, starting with a fresh look at the source models used to develop the 50 cm threshold. The models were of the seven shallow crustal events in California. These events had maximum slips in the approximate range of 45 to 790 cm, so that on average the 50 cm threshold corresponded to approximately 15% of the maximum. For NGA-Sub, we consider this percentage of the maximum slip, in lieu of the 50 cm threshold directly. This was considered to be appropriate given the large rupture dimensions and slip values involved in subduction-zone earthquakes when compared to the **M**6-7 shallow crustal events upon which the original criteria had been based. When the 15% criteria was applied to large subduction events with FFMs, the results were judged to be reasonable by the source working group. Accordingly, we trim the FFMs by applying a threshold of 15% of the maximum slip and then drawing one or more rectangles around the high slip areas

3.3 Source parameters for events without FFMs

Most earthquakes with event-types in the NGA-Sub database do not have FFMs in literature (903 out of 991). Because there is a need for a finite fault representation of each earthquake source, we apply a simulation procedure for events without published models. We require models for the rectangular dimensions of finite faults (along-strike length *L* and width *W*), the orientation of the rectangles, and hypocenter location within the rectangle. Models for fault area (*A*) and aspect ratio (*L/W*) are provided for interface subduction events by Murotani et al. (2013) and Skarlatoudis et al. (2016).

Figure 6 shows fault rupture areas for the 47 interface events with FFMs in the NGA-Sub database. Also shown are source dimension data from Skarlatoudis et al. (2016) and regression fits. For rupture area data, linear regression fits are shown from Murotani et al. (2013), Skarlatoudis et al. (2016), and the present study using the following linear expression:

$$lnA = a_1 + a_2\mathbf{M} + \varepsilon_{n1}\sigma_A \quad (1)$$

where A is rupture area in km^2, a_1 and a_2 are regression coefficients, ε_{n1} is the standard normal variate (zero mean, standard deviation of 1) and σ_A is the standard deviation. Regressions were performed at two stages in the NGA-Sub project. The first regression was performed using data from 29 earthquakes available at that time (approximately Nov 2017) along with data from Skarlatoudis et al. (2016); the 29 events used at that time are depicted in Fig. 6(a) and the resulting coefficients are indicated in the figure (marked as "Applied"). The next regression was performed near the end of the project (Dec. 2019) using all 47 events and Skarlatoudis et al. (2016) data for non-redundant events with the results shown in Fig. 6(b) (marked as "Complete"). There is no appreciable difference between the "Applied" and "Complete" versions of the mean model, although dispersion increases in the update. The fits derived in the present study are similar to those derived by Murotani et al. (2013).

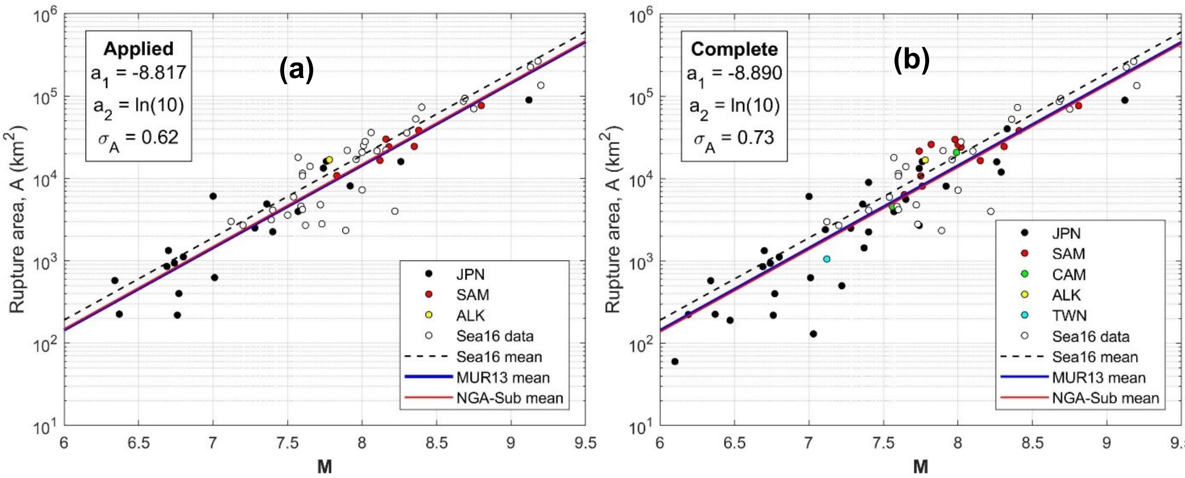

Fig. 6 – (a) Applied and (b) Complete geometric relations for rupture area for subduction interface earthquakes. JPN = Japan, SAM = South America, CAM = Central America and Mexico, ALK = Alaska, TWN = Taiwan, Sea16 = Skarlatoudis et al. (2016), MUR13 = Murotani (2013).

4. Ground Motions and Intensity Measures

4.1 Data sources and time series processing

The NGA-Sub database contains 70,107 three-component records from 1,880 earthquakes from seven global subduction zone regions: Alaska, Cascadia, Central America and Mexico, Japan, New Zealand, South America, and Taiwan. Table 2 shows the agencies from each of the above regions from which we obtained uncorrected records. All these recordings were processed for NGA-Sub from all sources.

Table 2 –Ground motion catalogs contributing data to NGA-Sub database

Region	Catalog	Processed motions
Alaska	CESMD	36
	IRIS	2,812
	GSC	178
Cascadia	CESMD	29
	COSMOS	100
	IRIS	1,432
	NSMP	112
	NCEDC	219
	GSC	217
Central America and Mexico	NORSAR	292
	COSMOS	349
	NOAA	727
	IRIS	908
	MARN	227
	Universidad de Costa Rica	145
Japan	NIED, K-NET	20,869
	NIED, KIK-NET	18,836
	JMA	444
	PARI	303
	NOAA	44
	HI-NET	72
	Electricity Companies (TEPCO, EPCO)	149
South America	CESMD	213
	NOAA	40
	IRIS	1,689
	GFZ	1,193
	RENADIC (U Chile, CEE Dept.)	1,274
	CSN (U Chile, seismology)	1,076
	CISMID	213
	RNAC (Colombia)	409
	RENAC (Ecuador)	89
Taiwan	CWB	11,176
	IES	1,196
	K-NET	62
	JMA	69

Data processing starts with digital time series (accelerograms or seismograms) with a sample rate (time step), and in the case of signal from modern digital instruments, with a time-stamp (a known reference time at the start of the record). The major steps in data processing are: (1) screening of time series to select the ground motions to process, (2) application of window functions that reduce the signal to zero outside of a time interval, (3) computation of Fourier amplitude spectra, (4) filtering of the record to remove noise-dominated features over selected frequency intervals, and (5) baseline correction. The procedures applied here are similar to those used in previous NGA projects, including NGA-West1 (Chiou et al. 2008[7]), NGA-West2 (Ancheta et al. 2013[12]), and NGA-East (Goulet et al. 2014[13]). Corrected time series are saves as .AT2 files for use in the computation of ground motion parameters. This procedure strictly applies for accelerograms. For seismograms, before step (2), instrument corrections are applied to the time series and the signals are time-differentiated once to acceleration.

4.2 Computation of ground motion parameters

Acceleration time series processed using the procedures described previously were used to compute two types of pseudo spectral accelerations (PSa). One is "as-recorded spectral acceleration" which computes the PSa for three components independently. The other PSa is "RotDnn" which is an orientation-independent combination of the two horizontal components (Boore 2010 [14]). PSa is computed from spectral displacement (Sd), which is the maximum relative displacement of the single-degree-of-freedom (SDOF) elastic oscillator with a specific period and damping. A damping ratio of 5% was used. SD is converted to PSa as shown in Eq. (1).

$$PSa = Sd \cdot \left(\frac{2\pi}{T}\right)^2 \qquad (2)$$

where T is the structural period. As described in Boore 2010 [14], the RotDnn spectra represent the range of oscillator responses to a given pair of horizontal input motions. The responses are computed across all non-redundant rotation angles, and 'nn' represents the fractile of the spectra sorted by amplitude. The 'D' indicates that rotation angle will be specific to the period of the oscillator. RotDnn spectra for a given azimuth can be computed from the horizontal ground motion for that same azimuth. The ground motion for a particular azimuth (rotated an amount θ from the azimuths of the original recordings) can be computed from the orthogonal horizontal-component time series, $a_1(t)$ and $a_2(t)$, as follows:

$$a_{ROT}(t,\theta) = a_1(t)\cos\theta + a_2(t)\sin\theta \qquad (3)$$

where a_{ROT} is the rotated time series, and θ is the rotation angle from the a_1 axis. Using the RCTC code from Wang et al. 2017 [15] (or an equivalent Fortran code from Dave Boore, 2020[16]), response spectra for the rotated time series are calculated for non-redundant rotation angles between zero and 180°. Three fractiles are saved: the minimum (RotD00), median (RotD50), and the maximum (RotD100).

As pointed out in Section 4.1, there are 70,107 three-component records from 1,880 events. Following screening to remove events without magnitudes, hypocenter locations, assigned event types, or distances, the database is reduced to 65,276 recordings from 976 events. Fig. 7a shows the data distribution for these 976 events in magnitude-distance space for peak acceleration. Figs. 7(b),(c) show PGA distributions for 360 events classified as interface (23,552 recordings) and 383 intraslab events (27,547 recordings), respectively. Fig. 7(d) shows the number of records as a function of oscillator period. There is fall-off in the amount of data as period increases as some records become outside of their usable range based on the selected low-cut frequency. Fig.7(d) shows this fall-off begins at 1 sec and 2/3 of the data is lost at 20 sec period.

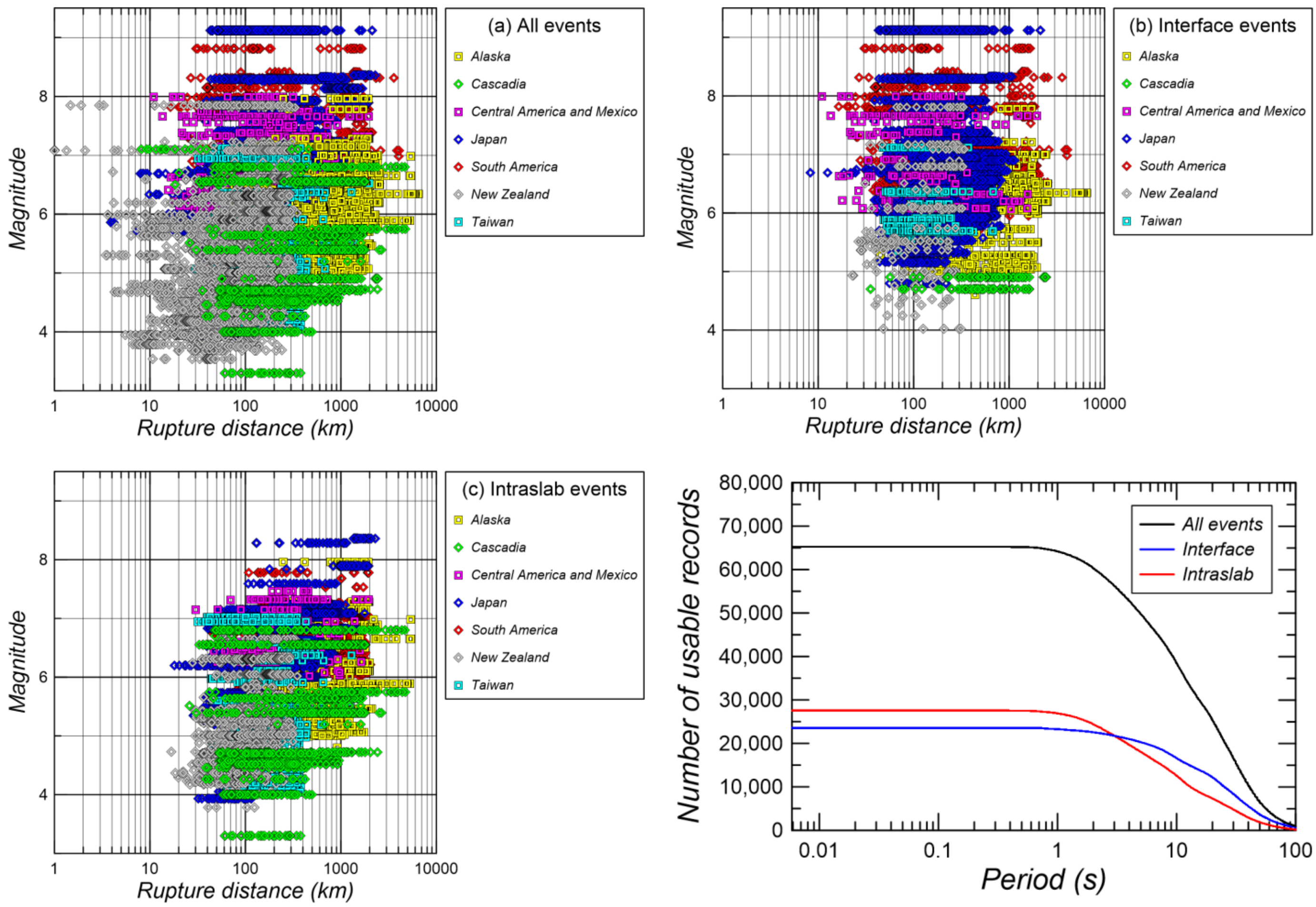

Fig. 7 – Distribution in magnitude-rupture distance space of recordings from (a) 976 events , (b) 360 interface events and (c) 383 intraslab events that pass screening criteria described in Section 3. (d) Fall-off of number of usable records from 976 screened events as oscillator period increases.

5. Acknowledgements

Core support for the NGA-Sub research project was provided by FM Global, the US Geological Survey, the California Department of Transportation, and the Pacific Gas & Electric Company. The support of these organizations is gratefully appreciated.

6. Copyrights

7. References

[1] Ahdi SK, Ancheta TD, Contreras V, Kishida T, Kwak DY, Kwok AO, Mazzoni S, Parker GA, Bozorgnia Y, Stewart JP (2020): NGA-Subduction Site Database. *Proc. 17th World Conference on Earthquake Engineering (17WCEE)*, Sendai, Japan.

[2] Power M, Chiou BS-J, Abrahamson NA, Bozorgnia Y, Shantz T, Roblee C (2008): An overview of the NGA Project. *Earthquake Spectra*, **24**, 3–21.

[3] Bozorgnia Y, Abrahamson NA, Al Atik L, Ancheta TD, Atkinson GM, Baker JW, Baltay A, Boore DM, Campbell KW, Chiou BS-J, Darragh R, Day S, Donahue J, Graves RW, Gregor N, Hanks T, Idriss IM, Kamai R, Kishida T, Kottke A, Mahin SA, Rezaeian S, Rowshandel B, Seyhan E, Shahi S, Shantz T, Silva WJ, Spudich P, Stewart JP, Watson-Lamprey J, Wooddell K, Youngs RR (2014): NGA-West 2 research project. *Earthquake Spectra*, **30**, 973-987.

[4] Goulet CA, Bozorgnia Y, Abrahamson NA, Kuehn N, Al Atik L, Youngs RR, Graves RW, Atkinson GM (2018): Central and Eastern North America Ground Motion Characterization: NGA-East Final Report. *PEER Report 2018/08,* Pacific Earthquake Engineering Research Center, UC Berkeley, CA.

[5] Si et al. (2020): Report in preparation.

[6] Youngs et al. (2020). Report in preparation.

[7] Chiou BS-J, Darragh RB, Gregor N, Silva WJ (2008): NGA project strong-motion database. *Earthquake Spectra*, **24**, 23-44.

[8] Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BS-J, Wooddell KE, Graves RW, Kottke AR, Boore DM, Kishida T, Donahue JL (2014): NGA-West2 database. *Earthquake Spectra*, **30**, 989-1005.

[9] Mazzoni S, Kishida T, Ahdi SK, Contreras V, Darragh RB, Chiou BS-J, Kuehn N, Bozorgnia Y, Stewart JP (2020): Chapter 2: Relational Database, in Data Resources for NGA-Subduction Project. *PEER Report 2020/xx*, JP Stewart (editor), Pacific Earthquake Engineering Research Center, UC Berkeley (headquarters).

[10] Bird P. (2003): An updated digital model of plate boundaries. *Geochemistry, Geophysics, Geosystems*, **4(3)**, 1-52.

[11] Contreras V, Stewart JP, Kishida T, Darragh RB, Chiou BS-J, Mazzoni S, Kuehn N, Ahdi SK, Wooddell K, Youngs RR, Bozorgnia Y, Boroschek R, Rojas F, Órdenes J (2020): Chapter 4: Source and Path Metadata, in Data Resources for NGA-Subduction Project. *PEER Report 2020/xx*, JP Stewart (editor), Pacific Earthquake Engineering Research Center, UC Berkeley (headquarters).

[12] Ancheta T.D., Darragh R.B., Stewart J.P., Seyhan E., Silva W.J., Chiou B.S.J., Woodell K.E., Graves R.W., Kottke A.R., Boore D.M., Kishida T, Donahue JL (2013). PEER NGA-West 2 database, PEER Report 2013/03, Pacific Earthquake Engineering Research Center, Berkeley, CA.

[13] Goulet C.A., Kishida T., Cramer C.H., Darragh R.B., Silva W.J., Hashash Y.M.A., Harmon J., Stewart J.P., Wooddell K., Youngs R.R. (2014). PEER NGA-East database, PEER Report 2014/17, Pacific Earthquake Engineering Research Center, Berkeley, CA.

[14] Boore D.M. (2010). Orientation-independent, non geometric-mean measures of seismic intensity from two horizontal components of motion, Bull. Seismol. Soc. Am. 100, 1830–1835.

[15] Wang, P., Stewart J.P., Bozorgnia Y., Boore D.M., Kishida T. (2017). "R" package for computation of earthquake ground-motion response spectra" Report No. 2017/09, Pacific Earthquake Engineering Research Center, UC Berkeley.

[16] Boore D.M. (2020). TSPP---A Collection of FORTRAN Programs for Processing and Manipulating Time Series, U.S. Geological Survey Open-File Report 2008-1111 (Revised 10 January 2020).

[17] Kishida T, Darragh RB, Chiou BS-J., Bozorgnia Y, Mazzoni S, Contreras V, Boroschek R, Rojas F, Stewart JP (2020): Chapter 3: Ground Motions and Intensity Measures, in Data Resources for NGA-Subduction Project. *PEER Report 2020/xx*, JP Stewart (editor), Pacific Earthquake Engineering Research Center, UC Berkeley (headquarters)

