A Flatfile for the KiK-net Database Processed Using an Automated Protocol

Article *in* Earthquake Spectra · September 2015

DOI: 10.1193/071214EQS106

CITATIONS
47

READS
1,605

5 authors, including:

Haitham Dawood
Fugro Consultants, Inc.
16 PUBLICATIONS **583** CITATIONS
SEE PROFILE

Adrian Rodriguez-Marek
Virginia Polytechnic Institute and State University
117 PUBLICATIONS **2,344** CITATIONS
SEE PROFILE

Jeff Bayless
AECOM
9 PUBLICATIONS **90** CITATIONS
SEE PROFILE

Eric M. Thompson
United States Geological Survey
78 PUBLICATIONS **1,491** CITATIONS
SEE PROFILE

DATA PAPER

A Flatfile for the KiK-net Database Processed Using an Automated Protocol

Haitham M. Dawood,[a] Adrian Rodriguez-Marek,[b] Jeff Bayless,[c] Christine Goulet,[d] and Eric Thompson[e]

The Kiban-Kyoshin network (KiK-net) database is an important resource for ground motion (GM) studies. The processing of the KiK-net records is a necessary first step to enable their use in engineering applications. In this manuscript we present a step-by-step automated protocol used to systematically process about 157,000 KiK-net strong ground motion records. The automated protocol includes the selection of the corner frequency for high-pass filtering. In addition, a comprehensive set of metadata was compiled for each record. As a part of the metadata collection, two algorithms were used to identify dependent and independent earthquakes. Earthquakes are also classified into active crustal or subduction type events; most of the GM records correspond to subduction type earthquakes. A flatfile with all the metadata and the spectral acceleration of the processed records is uploaded to NEEShub (https://nees.org/resources/7849, Dawood et al. 2014). [DOI: 10.1193/071214EQS106]

INTRODUCTION

The use of strong ground motion (GM) records for engineering purposes requires that strong motion data be processed in a rigorous and uniform manner. When these data are used for the development of ground motion prediction equations (GMPEs) comprehensive sets of event and site metadata are needed, in addition to the GM data. The collection of these data, along with other parameters of interest, is usually summarized in a flatfile. Several researchers have processed such databases for different parts of the world (e.g., Chiou et al. 2008 for shallow crustal earthquakes as part of the NGA-West 1 project; Akkar et al. 2010 for Turkey; Arango et al. 2011a for the Central American subduction zone; Arango et al. 2011b for the Peru-Chile subduction zone; Pacor et al. 2011 for the ITACA database in Italy; and, most recently, Akkar et al. 2014 for the RESORCE database in Europe). The availability of processed databases is a key factor for advancing research in earthquake engineering in general, and for the development of GMPEs in particular. The objective of this paper is to introduce a

[a] Fugro Consultants, Inc., 101 West Main St., Norfolk, VA 23510. Formerly at The Charles E. Via, Jr. Department of civil and environmental engineering, Virginia Tech, 200 Patton Hall, Blacksburg, VA 24061
[b] The Charles E. Via, Jr. Department of civil and environmental engineering, Virginia Tech, 200 Patton Hall, Blacksburg, VA 24061
[c] URS Corporation, 915 Wilshire Boulevard, Suite 700, Los Angeles, CA 90017
[d] Pacific Earthquake Engineering Research Center, 325 Davis Hall, University of California, Berkeley, CA 94720
[e] Geological Sciences, San Diego State University, 500 Campanile Dr., San Diego, CA 92182-1020

Earthquake Spectra, Volume 32, No. 2, pages 1281–1302, May 2016;

protocol for automated processing of the Kiban-Kyoshin network (KiK-net) GM database and to produce a flatfile from these data for use in GMPE development.

The objective of strong motion processing is to remove the high and low frequency noise from the raw records and to adjust the records such that the final velocity of the record is zero and the final displacement is compatible with tectonic displacements (e.g., baseline correction). The processing has as its goal to obtain realistic and accurate ground motion intensity measurements. The instrument response may also need to be deconvolved from the record. In addition, GM processing determines the bandwidth over which the signal can be used for engineering purposes. Within this usable bandwidth, the earthquake signal should be strong enough such that the noise does not compromise the characteristics of the signal.

The Japanese strong GM networks, KiK-net and Kyoshin network (K-NET), have recorded a large number of strong motion data and have been used extensively by researchers around the world. Pousse (2005) developed a flatfile for KiK-net records up to 2004 which was used by various researchers (e.g., Cotton et al. 2008 and Rodriguez-Marek et al. 2011). Oth et al. (2011) studied the two databases and extracted a subset (2201 seismic events, with 78,840 and 34,456 acceleration time series recorded at surface and borehole, respectively) that they considered reliable. In addition, various GMPE developers used subsets of the KiK-net database (e.g., Kanno et al. 2006, Zhao et al. 2006, Dawood and Rodriguez-Marek 2013, and Goulet and Bayless n.d.), but do not provide the resulting flatfile and do not include a detailed documentation of the ground motion processing protocol. In this manuscript we present a processing protocol for the KiK-net records, describe in detail the procedures adopted to prepare the metadata for these records up to December 2011, and provide a flatfile with the processed records and the relevant metadata (Dawood et al. 2014; also see the *Data and Resources* section).

This manuscript consists of three main parts: Part I provides a background on the KiK-net network and the Full Range Seismograph Network of Japan (F-net) seismic catalog. Part I also describes the KiK-net strong ground motion database used in this study. Part II provides a detailed description of the automated record processing protocol applied to the KiK-net database. Part III describes the approach adopted to obtain the earthquake, distance, and site metadata for each record. Additional information, not included in the paper due to length limitations, is included in an electronic supplement. The manuscript ends with some remarks and conclusions. The flatfile with the metadata and the spectral accelerations of the processed records is given in Dawood et al. (2014).

PART I: KIK-NET DATABASE AND F-NET CATALOG

KIK-NET

KiK-net is one of several seismic networks established in Japan to better monitor the seismic activity around the country following the devastating Kobe earthquake (17 January 1995; Okada et al. 2004). As of December 2011, the KiK-net network consisted of 692 stations (Figure 1). Each KiK-net station consists of two strong GM seismographs, one at the ground surface and the other in a borehole. With the exception of 31 instruments, where the instrument depth is larger than 500 m, nearly all the borehole instruments are located at depths ranging between 100 m and 250 m. The sampling frequency of the instruments

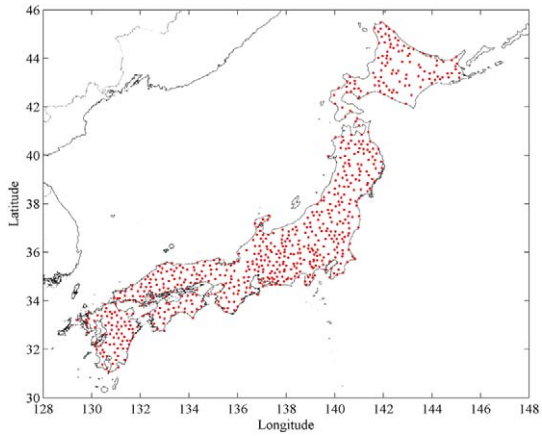

Figure 1. Location of the KiK-net stations (red dots) on the Japanese map.

are either 100 Hz or 200 Hz. Since each instrument records three components of motion, each KiK-net record has a total of six components. In the context of this paper, the six components of the motion are referred to as a record. The seismic velocity profile for 655 of these stations was obtained from the KiK-net website. The velocity profiles at these stations were obtained from downhole PS-logging (Oth et al. 2011). For more details regarding the KiK-net network and the specifications of the instruments the reader is referred to Aoi et al. (2011) and Okada et al. (2004).

The un-processed records from the KiK-net stations can be downloaded from the National Research Institute for Earth Science and Disaster Prevention (NIED) website at http://www.kyoshin.bosai.go.jp/. We use about 157,000 GM time series recorded between October 1997 and December 2011 from earthquakes with $M_{JMA} \geq 4.0$. About 134,000 of these records were recorded at an epicentral distance (R_{epi}) less than or equal to 300 km. The temporal distribution of the records is shown in Figure 2. As shown in the figure, the number of recorded motions in 2011 was considerably higher than the previous years due to the occurrence of the moment magnitude (M_W) 9.0 Tohoku earthquake, which was associated with a large number of foreshocks and aftershocks. Table 1 shows the number of records for which the geometric mean of the peak ground acceleration (PGA) of the two horizontal components exceeds different acceleration thresholds. Note that the majority of the records have very low PGAs.

F-NET SEISMIC CATALOG

F-net is a broadband seismograph network installed in Japan (Okada et al. 2004). The F-net website (http://www.fnet.bosai.go.jp/) provides a searchable database of earthquakes recorded by the F-net network. For each earthquake, the F-net catalog provides the origin time, location (latitude and longitude), M_{JMA} magnitude, JMA depth, region, and mechanism from the NIED moment tensor solution (strike, dip, rake, seismic moment, M_W, moment tensor solution, variance reduction, and number of stations used). A total of 25,212 earthquakes with M_W between 3.1 and 8.7 (M_{JMA} between 2.1 and 9.0) were recorded between

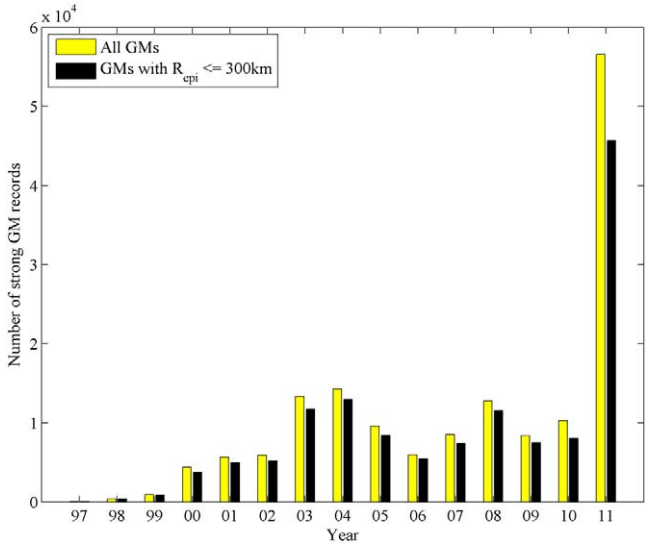

Figure 2. Temporal distribution of the ground motion (GM) records in the KiK-net database. Each GM consists of six components (three components at the borehole and three at the ground surface).

January 1997 and December 2011. The earthquakes' epicenters were located between 20° and 49° North and 120° and 156° East.

The spatial distribution of earthquakes in the F-net catalog is shown in Figure 3 for different magnitude bins (a total of 14,395 events with $M_W \geq 4$). Figure 3 shows that earthquakes are mainly concentrated along plate boundaries. Figure 4 shows the scatter plot of M_W versus depth along with the M_W and depth histograms of the earthquakes. More than 50% of the earthquakes occurred at depths shallower than 30 km.

Table 1. Summary of the number of KiK-net records with PGA (geometric mean) that exceed specified values

		Borehole		Surface	
		$R_{epi} \leq$ 300 km	$R_{epi} >$ 300 km	$R_{epi} \leq$ 300 km	$R_{epi} >$ 300 km
Number of records with PGA* $\geq$	0.01 g	3,081	150	22,454	1,195
	0.05 g	304	14	3,025	97
	0.10 g	107	4	1,055	34
	0.20 g	14	1	343	14
	0.30 g	2	0	167	5

* PGA for this table is defined as the geometric mean of the peak acceleration of the two horizontal components of the GM record

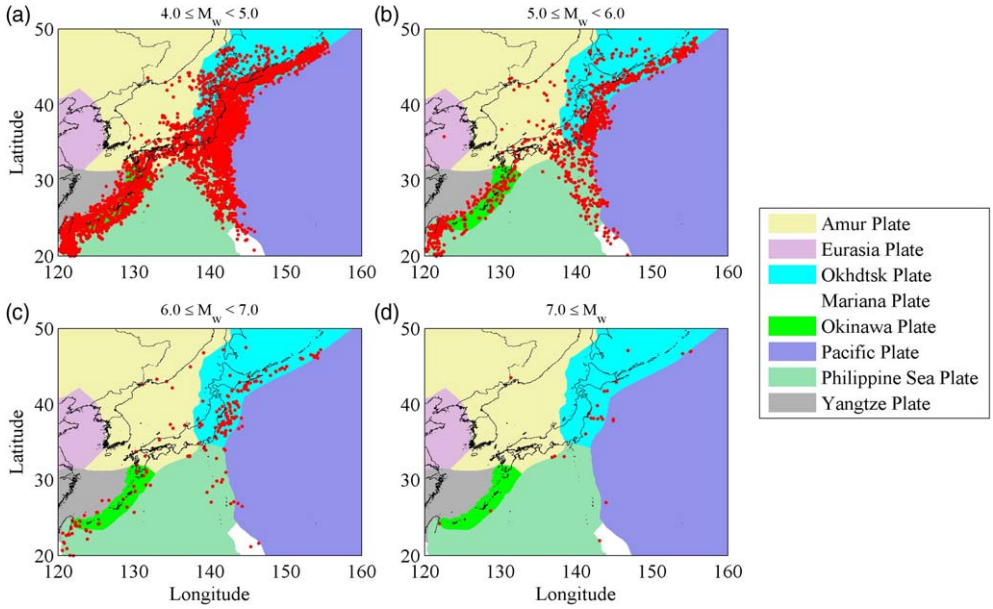

Figure 3. Spatial distribution of the earthquakes in the F-net catalog (between 1997 and 2011) for: (a) $4.0 \leq M_W < 5.0$, (b) $5.0 \leq M_W < 6.0$, (c) $6.0 \leq M_W < 7.0$, and (d) $M_W \geq 7.0$. The different tectonic plates around Japan are assigned different colors. (Plate boundaries obtained from Bird 2003.)

PART II: AUTOMATED GM PROCESSING PROTOCOL

Due to the large size of the KiK-net database, a fully automated processing protocol is desirable. Figure 5 is a flowchart that provides an overview of the processing protocol. In general, we relied on the guidelines provided by Boore et al. (2002), Boore (2005), and Boore and Bommer (2005) to develop the GM processing protocol used in this manuscript. The idea behind automation is to use a heuristic approach to select parameters for ground motion processing and use these to process a record. The processed record is then checked against several preset criteria. If these criteria are not fulfilled, another set of parameters is assumed and another processing loop is initiated. This loop is repeated until the processed record fulfills the criteria. The following paragraphs present a detailed description of the processing protocol along with a justification to the adopted criteria. Unless otherwise specified, the processing steps are applied on each of the six components of a record.

Step I. Baseline correction. A zeroth order baseline correction is performed by first subtracting the mean of the first 100 points from the whole acceleration time series (to account for a shift in the recorded acceleration time series) and then subtracting the mean of the pre-event noise window using an automated algorithm to detect the first arrival. The first arrival is defined as the first

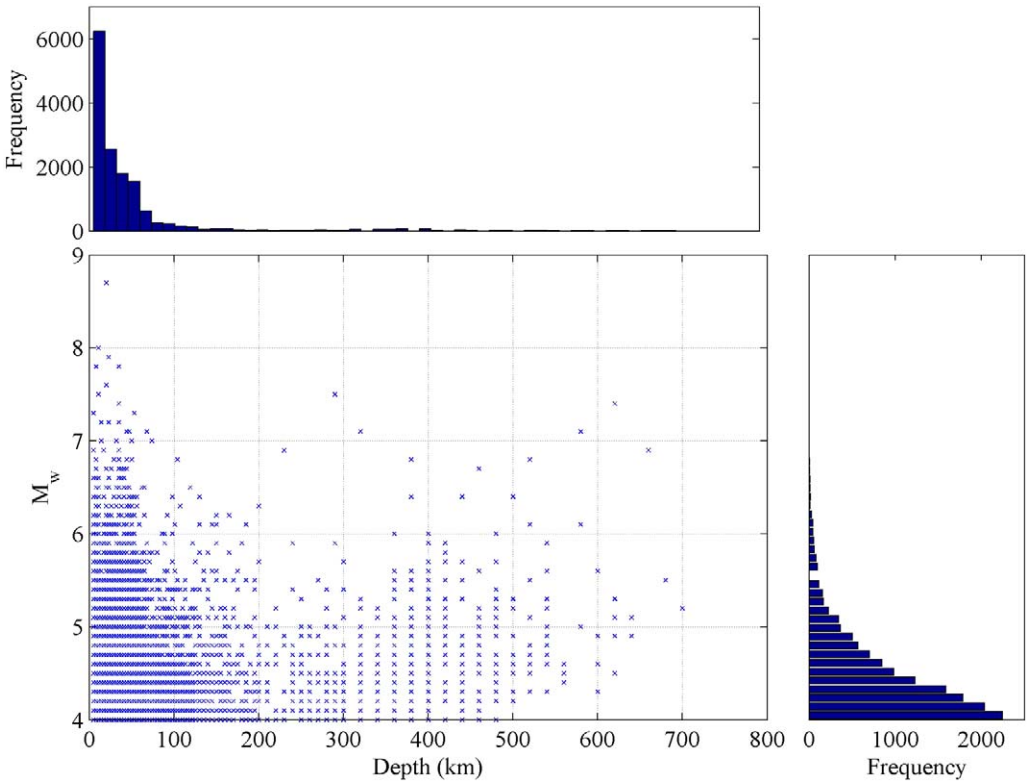

Figure 4. A scatter plot for M_W versus depth (km) for the earthquakes in the F-net catalog (bottom-left subplot) and two histograms that show the number of earthquakes in depth bins (top-left subplot) and the number of earthquakes in different M_W bins (bottom-right subplot).

automatically detected arrival time for the six components (see Dawood 2014 for details).

Step II. *Record tapering.* A tapered cosine (Tukey) window is applied to both ends of the acceleration time series to ensure a gradual transition to zero. The window length of the cosine window is set to be 5% of the total record length.

Step III. *Zero padding.* Zero pads are added before and after the acceleration time series. The length of the pads follows the recommendations by Converse and Brady (1992):

$$\text{Length of zero pad (sec)} = 1.5\,\text{N}/f_c \tag{1}$$

where N is the order of the Butterworth filter and f_c is the high-pass corner frequency (in Hz). These zero pads are added before and after the motions to avoid distortion in the filtered motion (the reader is referred to Boore 2005 for additional information). The computation of the Fast Fourier Transform may require additional zero padding to insure that the record length is a power of 2.

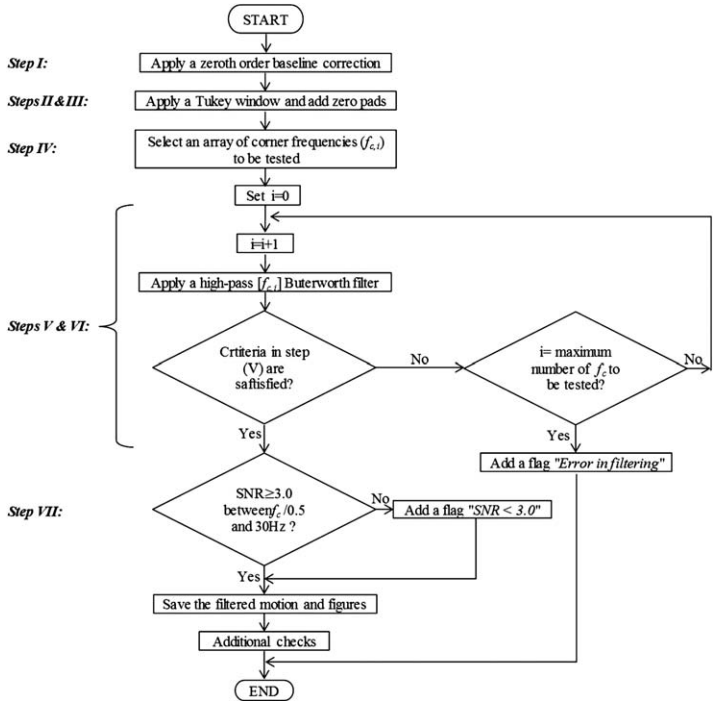

Figure 5. Flow chart diagram showing the steps of the processing protocol.

Step IV. Filtering. A high-pass, acausal, fourth-order Butterworth filter is used to filter the record. A set of high-pass corner frequencies is pre-selected (0.07, 0.09, 0.14, 0.17, 0.22, 0.35, 0.46, and 0.70 Hz). The lowest of these frequencies is first used to build the high-pass Butterworth filter. Each component is then filtered and checked against different pre-set criteria (see Step V). If the criteria are not satisfied, the next high-pass frequency is selected and this step is repeated.

Step V. Testing a high-pass corner frequency (f_c). The suitability of the selected high-pass corner frequency (Step IV) is determined by checking if the filtered components of the record satisfy the following criteria:

a. The final displacement in the displacement time series (obtained from numerical integration of the acceleration time series in the time domain) must be less than 0.005 cm and 0.025 cm for records from events with $M_{JMA} < 7.0$ and $M_{JMA} \geq 7.0$, respectively; and the final velocity in the velocity time series must be less than 0.001 and 0.005 cm/s for records from events with $M_{JMA} < 7.0$ and $M_{JMA} \geq 7.0$, respectively. The lower threshold values (0.005 cm and 0.001 cm/s for final displacement and velocity, respectively) were first applied to motions from large magnitude earthquakes, but were

found to be excessively strict for these motions. This observation is based on visual inspection of a set of randomly selected records from the database. Hence, we multiplied these two criteria by a factor of 5.0 for the records from earthquakes with $M_{JMA} \geq 7.0$. We note that using these criteria to choose the optimal f_c value in cases where major residual displacements are expected (e.g., records affected by permanent tectonic displacements) might remove part or all of the observed surface displacement; hence, caution should be taken when using automatically processed motions recorded close to the fault.

b. The ratio between the final and the maximum displacements must be less than 0.2.

c. A linear regression is applied to the trailing portion of the displacement and velocity time series. We assume the length of the trailing portion to be the last 10% of the recorded motion plus the length of the trailing zero pad (Equation 1). The slope of the best fit line must be less than 0.001 cm/s for the displacement time series and 0.001 cm/s^2 for the velocity time series.

d. The smoothed Fourier amplitude spectra (FAS) of acceleration (smoothed using the Konno-Ohmachi window with $b = 40$, Konno and Ohmachi 1998) should follow an f^2 decay at the low frequency end (Boore and Bommer 2005). To check for this criterion, a line is fit between frequency and the smoothed FAS in log-log space using the smallest five frequencies greater than the corner frequency (selected in Step VI). If the slope of the line is found to range between 1.0 and 3.0 for all components of the record, then the FAS decay criteria is fulfilled. This check is not applied to records from earthquakes with $M_{JMA} \geq 6.0$. The Brune (1970) source model predicts a slope of 2 for the FAS decay at the low frequency end of the spectrum. Our criteria is more flexible (e.g., accepts a slope that falls between 1.0 and 3.0) because the slope is calculated by fitting a line only through 5 points in the smoothed FAS. The slope check was skipped for records from earthquakes with $M_{JMA} \geq 6.0$ because the corner frequency for large earthquakes is expected to be outside the range of frequencies within which we search for f_c for the Butterworth filter.

Step VI. *Iterative search for a suitable high-pass corner frequency.* If some of the criteria described in Step V are not satisfied for any component of the record, a larger value for the high-pass corner frequency of the filter is chosen and steps IV and V are repeated. If all criteria are fulfilled for all components for the same tested corner frequency, the code proceeds to the next step (i.e., the same corner frequency is acceptable for the six components). If all the pre-selected frequencies (in step VI) are tested and neither resulted in a filtered record that fulfill all the criteria, the record is flagged with an "*Error in Filtering*" flag. Flagged records could then be evaluated manually (this was not done as a part of this study and these records were simply rejected).

Step VII. *Signal-to-noise ratio (SNR) check.* A noise window is defined using the last $2/f_c$ seconds of the record. The FAS for the whole record component (obtained from Step I) and for the noise window are computed and smoothed, and the ratio between both is calculated (from this point on, this ratio will be denoted as the SNR). If the SNR drops below 3 within the frequency band between

f_c/0.5 and 30 Hz, the record is flagged with an "SNR < 3.0" flag (Boore and Bommer 2005). In this study we chose to set the maximum usable spectral period to be 50% of the inverse of the corner frequency. This was based on a comparison of the motions in common between the database described herein and the NGA-West 2 database (Ancheta et al. 2013). It is worth mentioning that the same recommendation for acausal Butterworth filters had been suggested by Dr W. J. Silva through a written communication with the authors of Akkar and Bommer (2006). Hence, f_c/0.5 was chosen as the minimum frequency for the SNR check. The SNR was computed only up to 30 Hz as the response characteristic of the low-pass filter applied to the records is almost flat up to 30 Hz (Aoi et al. 2011). The same filter is applied to the six components of the record. Hence, the usable bandwidth of all components is the same. We note that due to the automated nature of the search for the high-pass corner frequency, the proposed algorithm is not likely to work well for pulse-type motions, such as those that may result from forward directivity effects.

The acceleration time series of the filtered components, including the zero-pads applied during processing, are saved. Stripping off the zero pads might compromise the filtered record especially at longer periods (Boore 2005, Boore et al. 2012). An additional check for records with multiple wave trains or sub-events was also applied and is presented in an electronic supplement to this manuscript. The KiK-net network has only broadband digital instruments and the processing protocol presented in this manuscript is suitable only for this type of instruments. Additional steps and checks might be necessary for analog instruments. For additional information about processing analog records the reader is referred to Douglas (2003) and Boore and Bommer (2005).

We conducted a comparison between the common records in the database presented in this manuscript and the NGA-West2 database (Ancheta et al. 2013). The comparison showed that, in most cases, the automated protocol resulted in records with a narrower usable frequency bandwidth compared to NGA-West2 records. This is the result of applying multiple (and possibly redundant) checks in the automated protocol with conservative threshold values. These checks were necessary, in the absence of a record-by-record processing with manual input, to obtain a set of high quality records from the automated protocol. The loss of usable frequency bandwidth is a trade-off that we accepted for the benefit of automating the processing of such a large data set. Additional details on this comparison are given in Dawood (2014).

The application of the Tukey window on motions that don't have a long enough noise window (e.g., motions that were triggered by the S-wave component) could affect the spectral accelerations of the processed motions, especially at high frequencies. An indication of this potential error is the difference between the PGA computed for the processed motions and the PGA reported in the KiK-net files. This difference (reported as a percentage) is included in the flatfile for the six components of each record. We leave it up to the users of this database to choose the suitable threshold or percentage error for the motions they use in their studies.

PART III: METADATA

This section presents the metadata collected for each record. The information includes event information, various source-to-site distance measures, and station information.

EARTHQUAKE INFORMATION

For the engineering use of strong motion records, it is important to associate each record with an earthquake and to have an accurate parameterization of the source (e.g., M_W, focal mechanism, rupture plane location and extent). The KiK-net data files provide a subset of these parameters (M_{JMA}, hypocenter location, and date and time of occurrence in the following format YYYY/MM/DD HH:MM:SS.SS, but SS.SS is always 00.00) and the full set of parameters are available in the F-net seismic catalog. We use the common parameters between the KiK-net and F-net data sets to match records in the KiK-net database to the corresponding earthquake in the F-net catalog. According to the F-net website, the source parameters are computed from a full-wave inversion code that uses about 70 different broadband waveforms (Fukuyama et al. 1998, Fukuyama and Dreger 2000). The automatic solution is then improved by manual operations. Hence, we consider the hypocentral location obtained from the F-net catalog more reliable than the values reported in the KiK-net data files. Moreover, we use the M_W reported in the F-net catalog since we believe it is more accurate than obtaining M_W from M_{JMA} through empirical correlations.

Oth et al. (2011) report that the location provided in the KiK-net website has a horizontal resolution of 0.1°. Hence, a perfect match between the KiK-net earthquake locations and the F-net catalog location is unlikely. We match KiK-net records to F-net earthquakes and classify the match into one of five categories (A through E) that allow different error margins (Table 2). Category A represents the strictest error margin, Category D contains earthquakes that were manually matched, and Category E contains earthquakes for which no match was found. A total of 4,943 earthquakes with $M_{JMA} \geq 4.0$ were identified in the KiK-net database. Table 2 provides the number of earthquakes matched at each category and the number of records associated with each. About 89% of the matched earthquakes and records fall in Category A. The manually-matched events (Category D) are mainly matched using the

Table 2. The error margins allowed during the cross matching of events from the KiK-net and F-net databases for each match category and the number of earthquakes and records associated with the five match categories

Category		A	B	C	D	E
Maximum error in:	Latitude (°)	0.1	0.1	0.15	–	–
	Longitude (°)	0.1	0.1	0.15	–	–
	Time (minutes)	1	60	60	–	–
	M_{JMA}	0.0	0.2	0.3	–	–
	Depth (km)	2.0	5.0	20	–	–
Number of earthquakes matched from the KiK-net database		4,393	73	46	28	403
Number of GM records		139,273	2,807	2,861	3,238	8,778

date, time, and location of the event. The 4,534 earthquakes (Categories A to D) from the KiK-net database mapped into 3,210 earthquakes in the F-net seismic catalog (2,776 of these F-net events have $M_W \geq 4$). This is due to the fact that the information of some earthquakes is documented slightly different in some KiK-net data files than others. For example, a 4.9 M_{JMA} earthquake occurred in 11 October 1997 at 14:44 (JST), was reported in the KiK-net data files using two slightly different locations (0.4 km apart). Hence, both of these events were mapped into a single event in the F-net catalog. Figure 6 shows the M_W versus depth and M_W versus epicentral distances for the earthquakes that fall into

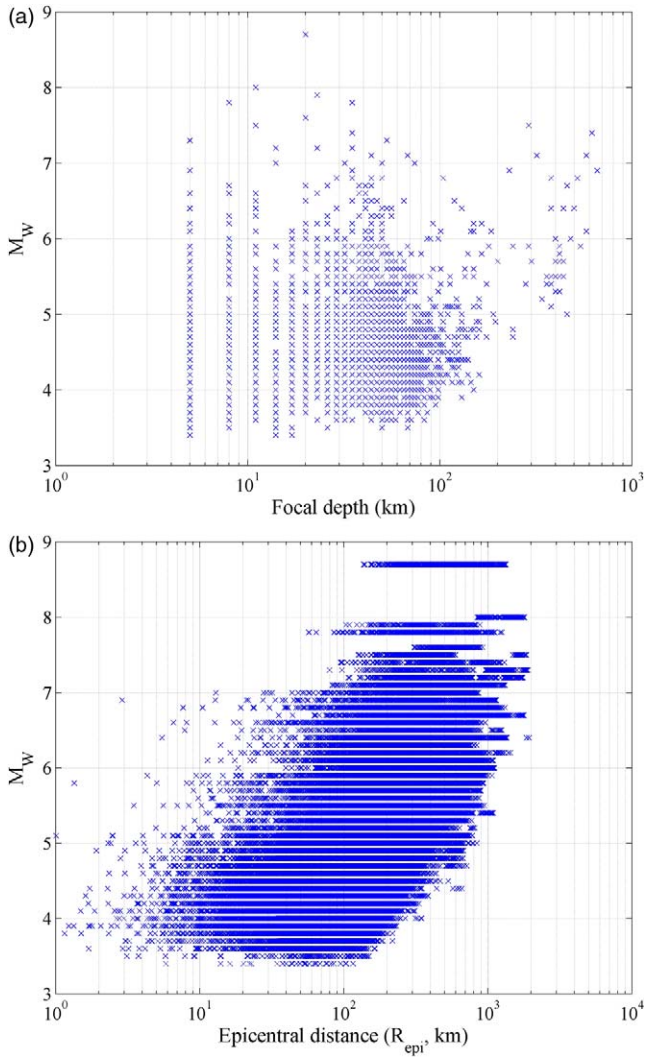

Figure 6. Plots of M_W magnitude for the events that fall within the matching Categories A through D versus: (a) the focal depth and (b) epicentral distance for each GM record.

the matching Categories A through D. Note that the M_W 9.0 Tohoku earthquake is shown as an M_W 8.7 earthquake in the F-net catalog and subsequently in our flatfile (to maintain consistency with the earthquake magnitudes reported in the F-net catalog). Care should be taken in using the metadata for earthquakes other than those in Category A because the data may not correspond to the correct event.

DECLUSTERING THE F-NET CATALOG

Earthquakes in seismic catalogs are classified as independent events (main shocks) or dependent events (foreshocks and aftershocks). Declustering is a process in which the dependent events are removed from the catalog. Identifying whether the events are independent or not is important for the development of GMPEs because source scaling may be different for these two types of events. For example, the strong ground motions in the NGA-West 1 database (Chiou et al. 2008) that were recorded from aftershocks were found to behave differently than the records from main shocks. For this reason, the records from aftershocks were either totally excluded from the data used in developing the GMPEs (e.g., Campbell and Bozorgnia 2008 and Boore and Atkinson 2008), or treated differently by including specific terms for aftershocks (e.g., Abrahamson and Silva 2008 and Chiou and Youngs 2008a). Nevertheless, in most GMPEs the difference between dependent and independent earthquakes is simply neglected.

Because different subsets of events in a catalog can be Poissonian, the declustering process has no unique solution and the different declustering algorithms can give different results (Stiphout et al. 2012). Many declustering algorithms had been developed over the years (e.g., Knopoff and Gardner 1972, Gardner and Knopoff 1974, Reasenberg 1985, Molchan and Dmitrieva 1992, Zhuang et al. 2002, Marsan and Lengline 2008, and Zaliapin et al. 2008). Stiphout et al. (2012) provide an overview of each of the previously mentioned declustering algorithms.

The F-net seismic catalog was declustered using the algorithms by Gardner and Knopoff (1974), as implemented in a Matlab script included in the software ZMAP, and by Reasenberg (1985), as implemented in a FORTRAN script coded by Dr. Norm Abrahamson (see *Data and Resources* section). These two algorithms are widely used. For example, the Gardner and Knopoff's (1974) algorithm was used to decluster the seismic catalogs used to develop the United States national seismic hazard maps (Petersen et al. 2008) and the Southeast Asia seismic hazard maps (Petersen et al. 2007). Reasenberg's algorithm was used to decluster catalogs used by Özturk and Bayrak (2012) and Vipin et al. (2013).

The Gardner and Knopff algorithm was implemented using three different sets of input parameters (see Stiphout et al. 2012 for more details). Summary of the analyses results are shown in Table 3. None of these analyses resulted in a Poissonian catalog. Attempts to change the input parameters of Gardner and Knopff's algorithm to force a Poissonian declustered catalog resulted in nonphysical (much lower than 1) Gutenberg-Richter *b*-values which might indicate that the declustering criteria is excessively strict.

Even if it is typical practice to assume that a declustered catalog follows a Poissonian model, this assumption is often debated (e.g., Matthews et al. 2002). Hence, it was decided to use the typical input parameters of Gardner and Knopoff's algorithm (see Table 3)

Table 3. The output of the different declustering trials on the F-net seismic catalog

Algorithm	Parameters	Number of clusters	Number of main shocks (independent earthquakes)	Gutenberg-Richter *b*-value	Is the declustered catalog Poissonian?
Reasenberg (1985)	$\tau_{min} = 2$ days, $\tau_{max} = 10$ days, $p_1 = 0.99$, $x_k = 0.50$, and $r_{fact} = 10$*	881	17,093	0.90	No
Gardner and Knopoff (1974)	Gardner and Knopoff (1974)	2,054	6,376	0.76	No
	Gruenthal (Personal Comm.)	1,739	3,688	0.69	No
	Uhrhammer (1986)	1,286	11,165	0.87	No

* τ_{min}: Minimum value of the look-ahead time for building clusters when the first event is not clustered**
τ_{max}: Maximum value of the look-ahead time for building clusters**
p_1: Probability of detecting the next clustered event used to compute the look-ahead time τ**
x_k: The increase of the lower cut-off magnitude during clusters**
r_{fact}: The number of crack radii surrounding each earthquake within new events considered to be part of the cluster**
** As defined by Stiphout et al. (2012). The reader is referred to Reasenberg (1985) for more details.

despite the fact that the resulting catalog is non-Poissonian. For this reason, we recommend against using the output of the declustering analysis from this study in source characterization studies for seismic hazard analyses, where the Poissonian assumption is often invoked. On the other hand, using the output of this analysis to differentiate between dependent and independent events in developing GMPEs should be valid, since the assumption of a Poissonian distribution for earthquake occurrence is not used in GMPE development. Additional information regarding declustering the F-net catalog is presented in Dawood (2014).

EARTHQUAKE CLASSIFICATION

Earthquakes can be classified into different categories based on parameters such as location, depth, and style of faulting. A widely used classification in the development of GMPEs is based on tectonic environment and divides earthquakes into *active crustal* (shallow and deep); *subduction zone* (interface and intraslab); and *stable continental region* earthquakes. If enough records are available, GMPEs are generally developed using records from a specific tectonic regime.

We use the algorithms by Allen et al. (2008) and Garcia et al. (2012) to classify the earthquakes in the F-net catalog. The latter algorithm was proposed to overcome several shortcomings in the algorithm by Allen et al. (2008). Table 4 shows a summary of the input parameters and the output earthquake categories for both algorithms. It is clear from the table the higher level of complexity of the Garcia et al. (2012) algorithm. This algorithm was validated by automatically classifying a catalog of earthquakes that also were manually classified. For most event types, the validation exercise showed considerable improvement in the number of correctly classified events when the Garcia et al. (2012) algorithm was used in place of the Allen et al. (2008) algorithm. However, for intraslab events the

Table 4. Comparison between two algorithms to classify earthquakes (Allen et al. 2008 and Garcia et al. 2012)

Algorithm	Allen et al. (2008)	Garcia et al. (2012)
Input parameters	• Hypocentral depth • Magnitude • Region classification (active tectonic or stable continental).	• Earthquake location • Finn-Engdahl geographic regions (Young et al. 1996) • Hypocentral depth • Focal mechanism • Information about trench and subduction slab interface models (Hayes and Wald 2009, Hayes et al. 2009, and 2012). [only for events in subduction regions]
Earthquake classes	• SZ* intraslab • SZ interface • Shallow active crustal • Stable continental	• SZ* outer • SZ intraslab • SZ interface • ACR* deep • ACR shallow SCR* • SOR* • OBR* ridge • OBR collision/transform • Hotspot • REVISION*

* SZ: subduction zone; ACR: active crustal region; SCR: stable continental region; SOR: stable oceanic region; OBR: oceanic boundary region; and REVISION: manual classification is needed for that specific event.

number of misclassified events using the Garcia et al. (2012) algorithm increased. This was attributed to the lack of slab models for about half of the misclassified earthquakes. The reader is referred to Garcia et al. (2012) for additional details. The results of applying both algorithms to the F-net catalog are shown in Table 5. The 951 earthquakes that were not classified using the algorithm by Allen et al. (2008) are events with $M_W \leq 7.7$ and depth greater than 50 km because Allen et al. (2008) do not provide a classification for events that fall within this depth-magnitude combination. The number of events classified as interface events using both algorithms are very different. This is consistent with the finding by Garcia et al. (2012) that the Allen et al. (2008) algorithm misclassified about 54% of the interface events. Figure 7 shows the location of the events in the four main categories as classified by Garcia et al. (2012).

Earthquakes were also classified based on faulting mechanism. A total of 281, 457, 1,616, and 856 earthquakes originated from strike slip, normal, reverse and unknown type faults, respectively. The determination of the faulting style was carried out using Equation (1) in Garcia et al. (2012). The faulting style is determined using the P-, B-, and T- axes plunges of each earthquake. These values were calculated from the fault plane parameters (dip, strike and rake) provided by the F-net catalog using the software FOCMEC (Snoke 2003).

Table 5. Classification of F-net earthquakes in the current database using the algorithms by Allen et al. (2008) and Garcia et al. (2012)

Class	Number of earthquakes classified by Allen et al. (2008)	Number of earthquakes classified by Garcia et al. (2012)
SZ intraslab	1,161	1,123
SZ interface	5	873
SZ outer	–	16
Shallow active crustal	1,093	–
ACR deep	–	112
ACR shallow	–	1,083
OBR	–	3
Not classified	951	–
Total	3,210	3,210

SZ: subduction zone; ACR: active crustal region; and OBR: oceanic boundary region.

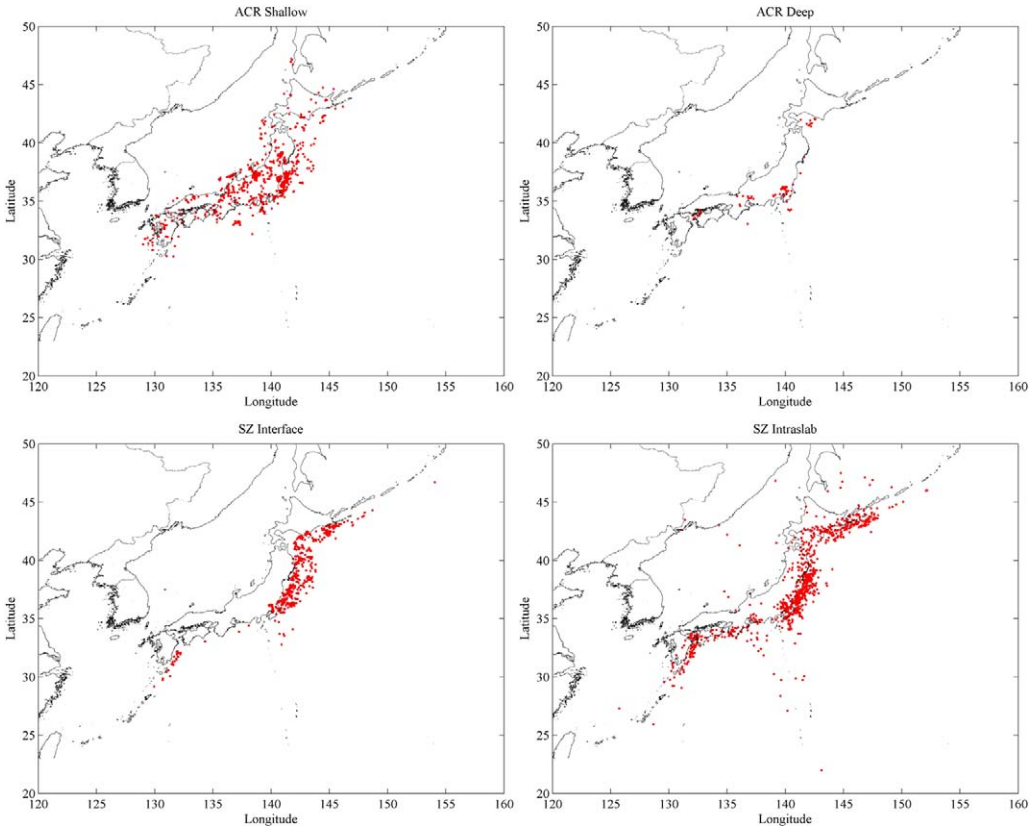

Figure 7. Location of the earthquakes classified as ACR shallow (top left); ACR deep (top right); SZ interface (bottom left); and SZ intraslab (bottom right). The classification is made using the algorithm by Garcia et al. (2012).

DISTANCE MEASUREMENTS

The epicentral distance, hypocentral distance, azimuth, and finite-source distance measures (closest distance to the fault, R_{rup}; closest distance to the surface projection of the fault, R_{JB}; and the horizontal distance from the top of rupture measured perpendicular to the fault strike, R_x) were computed for all the records from events for which there was a match in the F-net catalog. Finite-source rupture models for 21 earthquakes were found in the literature; these models were used to compute finite-source distance measures for 5,734 records associated with these earthquakes. For events for which a finite fault model is currently unavailable we simulate the rupture parameters following the method of Chiou and Youngs (2008b, Appendix B of that paper). We used the F-net hypocenter location, M_W, and focal mechanism. However, we could not determine which strike/dip pair was the actual fault plane for all events and so we compute the distances from 100 simulated faults for each plane separately. For each of the simulated faults, we compute the median R_{rup}, R_{JB}, R_x, and azimuth. Detailed information regarding the earthquakes with published finite-source rupture models and the methodology adopted to simulate the faults for the other earthquakes are presented in an electronic supplement to this manuscript.

RECORDING STATIONS

We calculated several parameters that characterize the soil profile at 655 KiK-net stations. Table 6 lists these parameters along with a brief description. The histograms that show the NEHRP site class (BSSC 2001), V_{S30} and the depth of the borehole instrument at all stations are shown in Figure 8. Most of the stations fall into the NEHRP site classes C [58%] and D [28%]. The median, mode and mean of the V_{S30} of all stations falls in the

Table 6. Parameters calculated at each KiK-net site

Parameter*	Description
V_{SX}	Average shear wave velocity from ground surface up to a depth of X meters (e.g., V_{S30})
$V_{S\,max}$	Maximum shear wave velocity in the profile
$V_{S\,min}$	Minimum shear wave velocity in the profile
V_{S0}	Shear wave velocity at the ground surface
V_{Smean}	Average shear wave velocity from ground surface up to the location of the borehole instrument
h_{800}	Depth at which the shear wave velocity exceeds 800 m/s
V_{Sh800}	Average shear wave velocity from ground surface up to the depth h_{800}
$V_{Sborehole}$	Shear wave velocity at the location of the borehole instrument
Hole depth	Depth of the borehole instrument
NEHRP site class	A if $V_{S30} > 1{,}500$ m/s; B if $1{,}500\ \text{m/s} \geq V_{S30} > 760$ m/s; C if $760\ \text{m/s} \geq V_{S30} > 360$ m/s D if $360\ \text{m/s} \geq V_{S30} \geq 180$ m/s; and E if $180\ \text{m/s} > V_{S30}$

* Shear wave velocities reported in the flatfile are based on measured shear wave velocities as reported in the KiK-net website

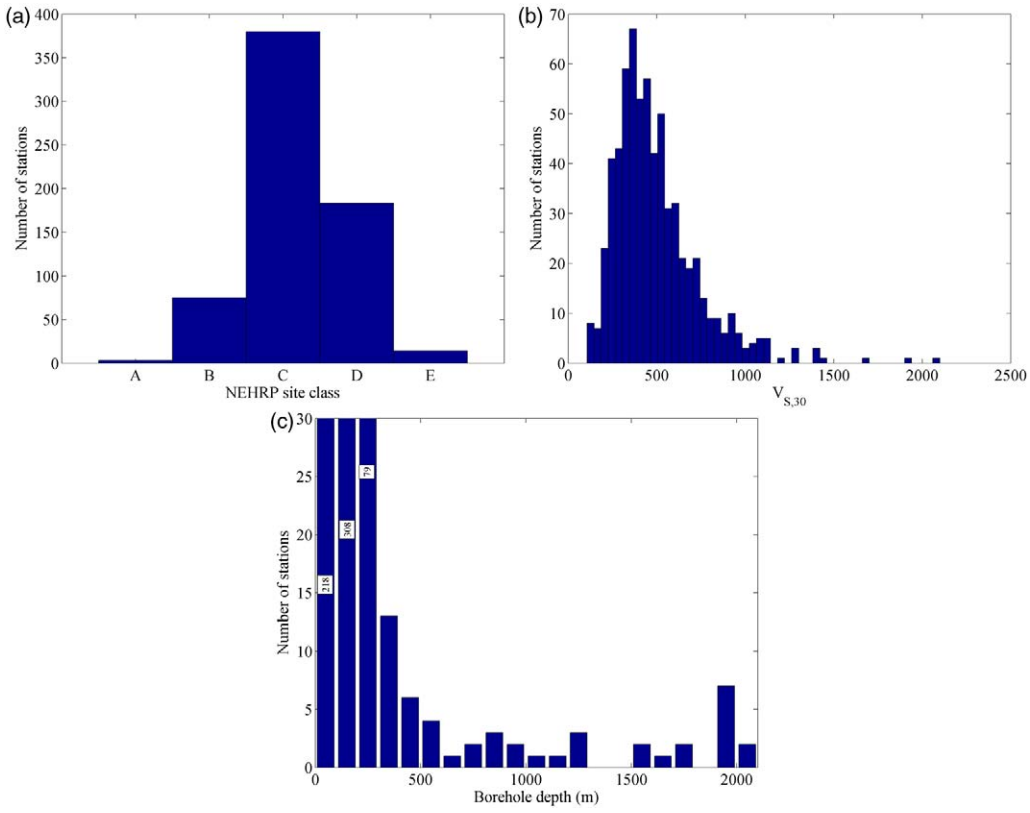

Figure 8. Histograms showing: (a) NEHRP site classes for KiK-net stations, (b) V_{S30} for KiK-net stations; and (c) depth of the borehole instrument for KiK-net stations.

range between 445 m/s and 500 m/s. The depth of the borehole instruments vary between 99 m and 2,008 m.

SUMMARY AND CONCLUSIONS

We presented an automated ground motion processing protocol and applied it to the KiK-net strong motion database to obtain corrected ground motions for events between October 1997 and December 2011. We used the F-net seismic catalog to obtain the earthquake metadata for each ground motion recording. This manuscript describes in detail the approach used to process the records and obtain the necessary metadata. Essential steps of the ground motion processing protocol include high-pass filtering of the record with a variable high-pass corner frequency such that baseline errors are minimized and the usable bandwidth is optimized. In addition, SNR checks were performed to ensure that the usable bandwidth is not contaminated by noise. An additional check was conducted to flag records that potentially have multiple wave trains.

Event metadata (magnitude and location) was obtained from the F-net catalog. A perfect match between earthquake parameters present in the KiK-net data files and the earthquakes in the F-net catalog was not always possible. Hence, we attach to each event a classification indicating the reliability of the match. For earthquakes with poor match, care should be taken if the earthquake parameters from the F-net catalog are to be used for further analyses. Declustering algorithms were used to differentiate between dependent and independent events for the purpose of GMPE development. Earthquakes were also classified by earthquake type. Two different earthquake classification algorithms were used in this study. Both algorithms showed that the earthquakes recorded by the KiK-net network are mainly active shallow crustal or subduction (interface and intraslab) type earthquakes. The flatfile containing all the data described in this article, in addition to the 5% damped response spectra, are publically available (Dawood et al. 2014).

DATA AND RESOURCES

The KiK-net strong-motions and shear wave velocity profiles used in this study were provided by National Research Institute for Earth Science and Disaster Prevention (NIED) at (www.kik.bosai.go.jp). The strong-motions data and the shear wave velocity profiles were last accessed in October 2012. The F-net seismic catalog was also provided by NIED at http://www.fnet.bosai.go.jp. It was last accessed in October 2012.

The automated protocol is based on the Matlab scripts by Goulet and Bayless (n.d.). We used the Matlab scripts included in the software package ZMAP (http://www.earthquake.ethz.ch/software/zmap) that are programmed to decluster the F-net seismic catalog using the Gardner and Knopoff algorithms. We used a FORTRAN script coded by Dr Norm Abrahamson to decluster the F-net seismic catalog using the Reasenberg (1985) algorithm. We also used the FOCMEC software (http://www.geol.vt.edu/outreach/vtso/focmec/focmec-doc/readme_focmec.html) to convert from the fault plane (dip, strike and rake) to the stress axes (lower hemisphere trend a plunge of T, P, and B).

The flatfile described in this manuscript was uploaded to the NEEShub website and is accessible through https://nees.org/resources/7849/.

ACKNOWLEDGMENTS

This research was partially supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number G11AP200049 and the SeIsmic Ground Motion Assessment (SIGMA) project under grant No. 3000-5910098949. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. government. Additional support was provided by Virginia Tech. The Pacific Earthquake Engineering Research Center (PEER) provided funding for some of the co-authors of this manuscript.

The authors would like to acknowledge the National Research Institute for Earth Science and Disaster Prevention (NIED) in Japan for providing the data used in this manuscript. The authors would like to acknowledge Dr. Daniel Garcia for classifying the different earthquakes in the F-net catalog on their behalf using the STREC code. The authors would also like to acknowledge several researchers who reviewed and discussed parts of the procedures

adopted in this manuscript: (in alphabetical order) Dr. Martin Chapman, Dr. Fabrice Cotton, Dr. Adrien Oth, Dr Frank Scherbaum, and Dr. Peter Stafford. Their comments improved the quality of this manuscript.

APPENDIX

Please refer to the online version of this paper to access the supplementary material provided in the appendix.

REFERENCES

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

Abrahamson, N. A., Silva, W. J., and Kamai, R., 2013. *Update of the AS08 Ground-Motion Prediction Equations Based on the NGA-West2 Data Set*, PEER Report No. 2013/04, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Akkar, S., and Bommer, J. J., 2006. Influence of long-period filter cut-off on elastic spectral displacement, *Earthquake Engng. Struct. Dyn.* **35**, 1145–1165.

Akkar, S., Çagnan, Z., Yenier, E., Erdogan, Ö., Sandıkkaya, M. A., and Gülkan, P., 2010. The recently compiled Turkish strong motion database: preliminary investigation for seismological parameters, *J. Seismol.* **14**, 457–479.

Akkar, S., Sandıkkaya, M. A., Şenyurt, M., Azari Sisi, A., Ay, B. Ö., Traversa, P., Douglas, J., Cotton, F., Luzi, L., Hernandez, B., and Godey, S., 2014. Reference database for seismic ground-motion in Europe (RESORCE), *Bull. Earthquake Eng.* **12**, 311–339.

Allen, T. I., Wald, D. J., Hotovec, A. J., Lin, K., Earle, P. S., and Marano, K. D., 2008. *An Atlas of Shake Maps for Selected Global Earthquakes*, U.S. Geol. Surv. Open-File Rept. 2008–1236.

Ancheta, T. D., Darragh, R. B., Stewart, J. P., 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.

Aoi, S., Kunugi, T., Nakamura, H., and Fujiwara, H., 2011. Deployment of new strong motion seismographs of K-NET and KiK-net, in *Engineering Seismology* (S. Akkar, P. Gülkan and T. Van Eck, eds.), Geotechnical, Geological, and Earthquake Engineering, Springer.

Arango, M. C., Strasser, F. O., Bommer, J. J., Hernández, D. A., and Cepeda, J. M., 2011a. A strong-motion database from the Central American subduction zone, *J. Seismol.* **15**, 261–294.

Arango, M. C., Strasser, F. O., Bommer, J. J., Boroschek, R., Comte, D., and Tavera, H., 2011b. A strong-motion database from the Peru-Chile subduction zone, *J.Seismol.* **15**, 19–41.

Bird, P., 2003. An updated digital model of plate boundaries, G^3 **4**, 1027, doi:10.1029/2001GC000252.

Boore, D. M., Stephens, C. D., and Joyner, W. B., 2002. Comments on baseline correction of digital strong-motion data: Examples from the 1999 Hector Mine, California, earthquake, *Bull. Seism. Soc. Am.* **92**, 1543–1560.

Boore, D. M., 2005. On pads and filters: Processing strong-motion data, *Bull. Seism. Soc. Am.* **95** 745–750.

Boore, D. M., Sisi, A. A., and Akkar, S., 2012. Using pad-stripped acausally filtered strong-motion data, *Bull. Seism. Soc. Am.* **102**, 751–760.

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.

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

Brune, J. N., 1970. Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res.* **75**, 4997–5002.

Building Seismic Safety Council (BSSC), 2001. *2000 Edition, NEHRP Recommended Provisions for Seismic Regulations for New Buildings and Other Structures*, Part1: Provisions (FEMA-368).

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 from0.01 to 10 s, *Earthquake Spectra* **24**, 139–171.

Chiou, B. S.-J., and Youngs, R. R., 2008a. An NGA model for the average horizontal component of peak ground motion and response spectra, *Earthquake Spectra* **24**, 173–215.

Chiou, B. S.-J., and Youngs, R. R., 2008b. *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 to 10 Seconds*, PEER Report, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

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

Converse, A. M., and Brady, A. G., 1992. *BAP: Basic Strong-Motion Accelerogram Processing Software, Version 1.0*, U.S. Geol. Surv., Open-File Report 92–296A.

Cotton, F., Pousse, G., Bonilla, F., and Scherbaum, F., 2008. On the discrepancy of recent European ground-motion observations and predictions from empirical models: Analysis of KiK-net accelerometric data and point-sources stochastic simulations, *Bull. Seismol. Soc. Am.* **98**, 2244–2261.

Dawood, H. M., 2014. Partitioning Uncertainty for Non-Ergodic Probabilistic Seismic Hazard Analyses, Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Dawood, H. M., Rodriguez-Marek, A., Bayless, J., Goulet, C., and Thompson, E., 2014. A Flatfile for the KiK-net Database Processed Using an Automated Protocol, https://nees.org/resources/7849.

Dawood, H. M., and Rodriguez-Marek, A., 2013. A method for including path effects in ground-motion prediction equations: An example using the M_W 9.0 Tohoku earthquake aftershocks, *Bull. Seism. Soc. Am.* **103**, 1360–1372.

Douglas, J., 2003. What is a poor quality strong-motion record?, *Bull. Earthquake Eng.* **1**, 141–156.

Fukushima, Y., 1996. Scaling relations for strong ground motion prediction models with M^2 terms, *Bull. Seism. Soc. Am.* **86**, 329–336.

Fukuyama, E., Ishida, S., Dreger, D.S., and Kawai, H., 1998. Automated seismic moment tensor determination by using on-line broadband seismic waveforms, *Zishin* **51**, 149–156 (in Japanese with English abstract).

Fukuyama, E., and Dreger, D. 2000. Performance test of an automated moment tensor determination system for the future "Tokai" earthquake, *Earth Planets Space* **52**, 383–392.

Garcia, D., Wald, D. J., and Hearne, M. G., 2012. A global earthquake discrimination scheme to optimize ground-motion prediction equation selection, *Bull. Seism. Soc. Am.* **102**, 185–203.

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

Goulet, C. A., and Bayless, J., n.d. Evaluation of partitioned standard deviation terms for the integration of site effects into seismic hazard analyses, *Bull. Seism. Soc. Am.*, in review.

Hayes, G. P., and Wald, D. J., 2009. Developing framework to constrain the geometry of the seismic rupture plane on subduction interfaces a priori - a probabilistic approach, *Geophys. J. Int.* **176**, 951–964.

Hayes, G. P., Wald, D. J., and Keranen, K. 2009. Advancing techniques to constrain the geometry of the seismic rupture plane on subduction interfaces a priori: Higher-order functional fits, G^3 **10**, Q09006.

Hayes, G. P., Wald, D. J., and Johnson, R. L., 2012. Slab1.0: A three-dimensional model of global subduction zone geometries, *J. Geophys. Res.* **117** B01302.

Knopoff, L., and Gardner, J., 1972. Higher seismic activity during local night on the raw world-wide earthquake catalogue, *Geophys. J. Roy. Astron. Soc.* **28**, 311–313.

Kanno, T., Narita, A., Morikawa, N., Fujiwara, H., and Fukushima, Y., 2006. A new attenuation relation for strong ground motion in Japan based on recorded data, *Bull. Seism. Soc. Am.* **96**, 879–897.

Konno, K., and Ohmachi, T., 1998. Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor, *Bull. Seism. Soc. Am.* **88**, 228–241.

Marsan, D., and Lengline, O., 2008. Extending earthquakes' reach through cascading, *Science* **319**, 1076–1079.

Molchan, G., and Dmitrieva, O., 1992. Aftershock identification: methods and new approaches, *Geophys. J. Int.* **109**, 501–516.

Okada, Y., Kasahara, K., Hori, S., Obara, K., Sekiguchi, S., Fujiwara, H., and Yamamoto, A., 2004. Recent progress of seismic observation networks in Japan: Hi-net, F-net, K-NET and KiK-net, *Earth Planets Space* **56**, 15–28.

Oth, A., Parolai, S., and Bindi, D., 2011. Spectral analysis of K-NET and KiK-net data in Japan, Part I: Database compilation and peculiarities, *Bull. Seism. Soc. Am.* **101**, 652–666.

Özturk, S., and Bayrak, Y., 2012. Spatial variations of precursory seismic quiescence observed in recent years in the eastern part of Turkey, *Acta Geophys Pol.* **60**, 92–118.

Pacor, F., Paolucci, R., Ameri, G., Massa, M., and Puglia, R., 2011. Italian strong motion records in ITACA: overview and record processing, *Bull. Earthquake Eng.* **9**, 1741–1759.

Petersen, M., Harmsen, S., Mueller, C., Haller, K., Dewey, J., Luco, N., Crone, A., Lidke, D., and Rukstales, K., 2007. *Documentation for the Southeast Asia Seismic Hazard Maps*, U.S. Geol. Surv., Administrative Report, 30 September 2007.

Petersen, M. D., Frankel, A. D., Harmsen, S. C., Mueller, C. S., Haller, K. M., Wheeler, R. L., Wesson, R. L., Zeng, Y., Boyd, O. S., Perkins, D. M., Luco, N., Field, E. H., Wills, C. J., and Rukstales, K. S., 2008. Documentation for the 2008 Update of the United States National Seismic Hazard Maps, U.S. Geol. Surv., Open-File Report 2008–1128.

Pousse, G., 2005. Analyse des données accélérométriques de K-net et KiK-net implications pour la prédiction du mouvement sismique 'accélérogrammes et spectres de réponse' et la prise en compte des effets de site non linéaires, Ph.D. Thesis, Université Joseph Fourier, France.

Reasenberg, P., 1985. Second-order moment of central California seismicity, 1969–82, *J. Geophys. Res.* **90**, 5479–5495.

Snoke, J. A., 2003. FOCMEC: FOCal MEChanism determinations, in *International Handbook of Earthquake and Engineering Seismology* (W. H. K. Lee, H. Kanamori, P. C. Jennings and C. Kisslinger, eds.), Academic Press, San Diego, Chapter 85.12.

Uhrhammer, R., 1986. Characteristics of Northern and Central California seismicity, *Earthquake Notes* **57**, 21.

van Stiphout, T., Zhuang, J., and Marsan, D., 2012. Seismicity declustering, Community Online Resource for Statistical Seismicity Analysis, doi: 10.5078/corssa-52382934, available at http://www.corssa.org.

Vipin, K. S., Sitharam, T. G., and Kolathayar, S., 2013. Assessment of seismic hazard and liquefaction of Gujarat based on probabilistic approaches, *Nat. Hazards* **65**, 1179–1195.

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

Young, J. B., Presgrave, B. W., Aichele, H., Wiens, D. A., and Flinn, E. A., 1996. The Flinn-Engdahl regionalisation scheme: The 1995 revision, *Phys. Earth Planet. In.* **96**, 221–297.

Zaliapin, I., Gabrielov, A., Keilis-Borok, V., and Wong, H., 2008. Clustering analysis of seismicity and aftershock identification, *Phys. Rev. Lett.* **101**, 1–4.

Zhao, J. X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H. K., Somerville, P. G., Fukushima, Y., and Fukushima, Y., 2006. Attenuation relations of strong ground motion in Japan using site classification based on predominant period, *Bull. Seismol. Soc. Am.* **96**, 898–913.

Zhuang, J., Harte, D., Werner, M. J., Hainzl, S., and Zhou, S., 2012. Basic models of seismicity: temporal models, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-79905851, available at http://www.corssa.org.

Zhuang, J., Ogata, Y., and Vere-Jones, D., 2002. Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.* **97**, 369–380.

(Received 12 July 2014; accepted 9 July 2015)

APPENDIX

MULTIPLE WAVE TRAINS

The presence of multiple wave trains in a single data file is a problem that limits the usability of the GM record. In some cases, especially after the occurrence of large earthquakes, a strong GM instrument can record two seismic wave trains (or more) from different earthquakes and include both in the same file assuming they are from a single earthquake (Figure A1). Strasser and Bommer (2005) reported that using such records could result in distortion of the high frequency portion of the response spectra. These authors report that they were not able to calibrate two different methods to automatically separate these sub events because of the large variations in the magnitude and distances available in their study. Hence, they conducted a manual separation of a subset of their database. The size of the database used in the current study complicates the process of manually localizing and separating such motions. Hence, we developed a simple algorithm to automatically detect the motions that potentially contain multiple wave trains (MWT).

The concept underlying the algorithm is to automatically detect if the acceleration time history starts to considerably increase after it has decreased. This detection is difficult because there are no definite criteria to detect the end of a main shock and the arrival of an aftershock. Hence, we decided to test different criteria within the algorithm and count the number of cases where the different criteria reported a motion as containing MWT. The output of the algorithm is used to classify the motion into three classes: unlikely, possibly, or likely containing MWT. It is then up to the user to choose which subset to use in their analyses. Hence, this algorithm is not intended to provide a definitive answer about whether the motion contains, or not, MWT, but is intended to provide an indicator to that effect. The algorithm can be summarized as follows:

I. The acceleration time histories for the two horizontal components at the ground surface are divided into different segments. Each segment is 5 s wide.
II. The maximum of the absolute acceleration and its time of occurrence in each segment are localized.
III. We chose three different ratios that are used as preset criteria to pick motions with MWT. In the current algorithm we selected Criterion $1 = 0.05$, Criterion $2 = 0.10$, and Criterion $3 = 0.20$. The algorithm loops through the different criteria. For each Criterion:

- All segments that follow the one that contains the PGA are looped through. The first segment that has a maximum acceleration smaller than or equal to Criterion 1 multiplied by PGA is localized. This segment (S_{end}) is considered to be the end of the motion from the main shock.
- If the peak acceleration in any segment following that S_{end} has a peak acceleration that exceeds 1.5 multiplied by Criterion 1 and PGA, the motion component is flagged as MWT based on Criterion 1. In that case, we assume that the motion reinitiated again after it was ending. Hence, there is a possibility that there is a new wave train from an aftershock that reached the station.
- The algorithm loops the three criteria for the two horizontal components.

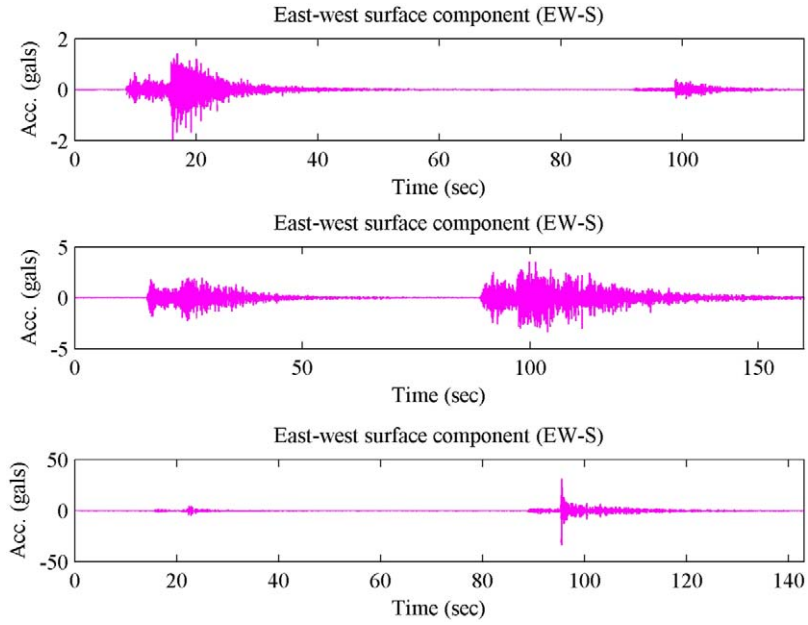

Figure A1. The East-west component recorded at the ground surface of three different GM records that have multiple wave trains (from top to bottom: GMs# 140150, 141404, and 142133).

IV. The same loops are performed for the segments that precede the one that contains the PGA. Similar flags are assigned for each criterion and component combinations to flag motions that contain foreshocks.

V. Each record is classified as "likely," "possibly," or "unlikely" MWT as follows:

- Likely MWT: If the two horizontal components were classified two times or more as aftershock MWT or foreshock MWT.
- Unlikely MWT: If the twelve checks (two horizontal components, three different criteria, and foreshock/aftershock checks) resulted in a non-MWT.
- Possibly MWT: The motions that cannot be either classified as "likely" or "unlikely" MWT.

Figure A2 shows another example of a MWT record. The top plots show the acceleration time histories. This particular record shows three different wave trains. PGA occurs in the first wave train. The second row of plots show the maximum accelerations in each segment (5 s in width) normalized by the PGA of each component of the motion. The titles of the plots in the second row show the result of the different checks. The algorithm identified an after-shock using the three criteria on the two components (i.e., an aftershock MWT was detected 6 times). Hence, the algorithm classified this motion as "Likely MWT" which is confirmed by the visual inspection. The second line in the titles of the lower row of plots shows the flags for the foreshock MWT check. Since the PGA is in the first wave trains, the algorithm did not detect any foreshocks and the six flags resulted in a non-foreshock-MWT flag. The vertical lines are plotted to show the automatically detected arrivals of the different aftershocks using

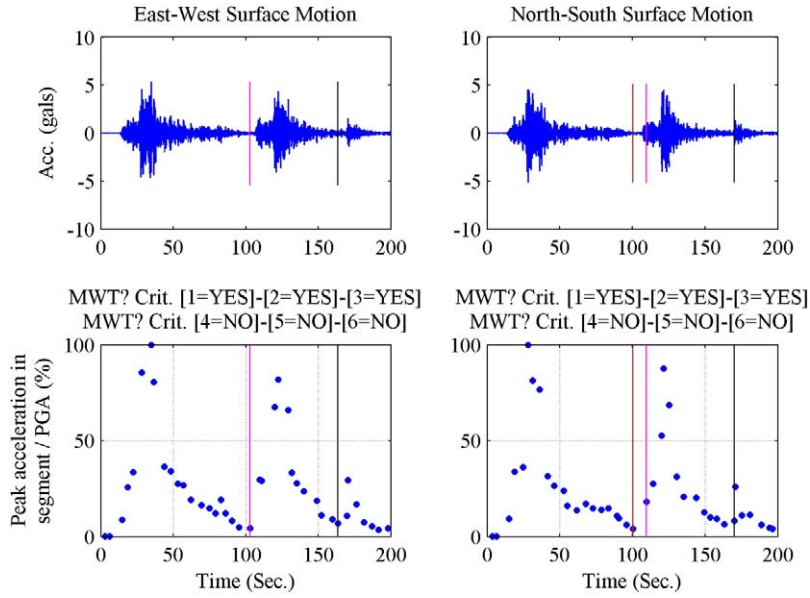

Figure A2. An example of a motion that contains multiple wave trains (GM# 59351). The plots to the left are for the East-West Surface component of the motion while the plots on the right are for the North-South Surface component. The top plots show the acceleration time histories and the vertical lines show the automatically picked arrivals of the aftershocks using the proposed algorithm. The bottom plots show the ratio between the peak acceleration in each segment (5 s in width) and the PGA, in percent. The automatically detected arrivals of the aftershocks are based on these ratios.

the different criteria. It is clear from the acceleration time history plots that the algorithm was able to localize the presence of the multiple wave trains effectively.

ADDITIONAL DETAILS ON THE PROCESSING PROTOCOL

Figures A3 and A4 present a visual example of the loops in the processing protocol used to pick up a suitable f_c value. Figure A3 shows the acceleration, velocity and displacement time histories of one component of a recorded time history after zeroth order baseline correction (Step I). It is clear from the velocity and displacement time histories that the motion needs to be baseline corrected. Figure A3 also shows the smoothed FAS. The FAS is almost horizontal at very low frequencies after it started to decay below about 0.8 Hz. This confirms the presence of long period noise that needs to be filtered out.

Figure A4 shows the time histories of the same component shown in Figure A3 after applying a filter with a corner frequency of 0.07 Hz (first iteration through Step IV). The velocity and displacement time histories visually look physical (i.e., they return to zero at the end of the motion). Another subplot is included to show the result of the five checks conducted using the different preset criteria (see Step V in the main text). The resulting motion fulfills the first five criteria. Also, four additional lines were added to the

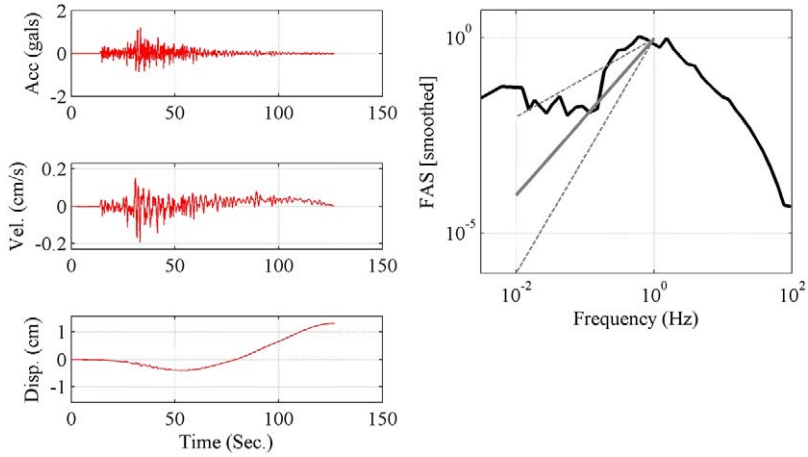

Figure A3. Acceleration, velocity, and displacement time histories of one component of a recorded time history (East-West borehole component of GM# 2351) after zeroth order baseline correction (left panels). The plot on the right shows the smoothed acceleration FAS of the same component.

smoothed FAS. The gray solid line represent the theoretical slope of the FAS decay with frequency (slope of 2), and the gray dashed lines represent the limits of the range of slopes that we considered acceptable while automatically processing the motions (slopes of 1 and 3). The black solid line represents the actual slope of the FAS of this component of the motion. Since this solid black line falls between the two gray dashed ones, this component passes the sixth criteria. The code checks the remaining five components of the motion to check if this f_c value is also suitable for these components. If this is not the case, the algorithm will select the next f_c value and repeat Steps IV through V.

DISTANCE MEASUREMENTS

The epicentral distance, hypocentral distance, and azimuth were computed for all the records from events for which there was a match in the F-net catalog. Finite-source rupture models for 21 earthquakes were found in the literature. These earthquakes are associated with 5,734 records in the database. Table A1 provides a list of these earthquakes. For the records from the earthquakes listed in Table A1, we calculated the closest distances to the fault (R_{rup}), closest distance to surface projection of the fault (Joyner-Boore distance, R_{JB}), and the horizontal distance from top of rupture measured perpendicular to fault strike (R_x). Table A1 includes the range of these distances for the corresponding stations, the magnitude reported by the finite fault reference, the Garcia et al. (2012) classification, and the classification by the finite fault reference (for those references where the authors indicated a classification). Figure A5 shows the distribution of M_W versus the different distance measures for the events listed in Table A1. When multiple source models were available, a single model for the distance calculations was chosen. This choice was based on two factors: (1) the most recently published models were preferred to older models, and (2) the precision with which the model

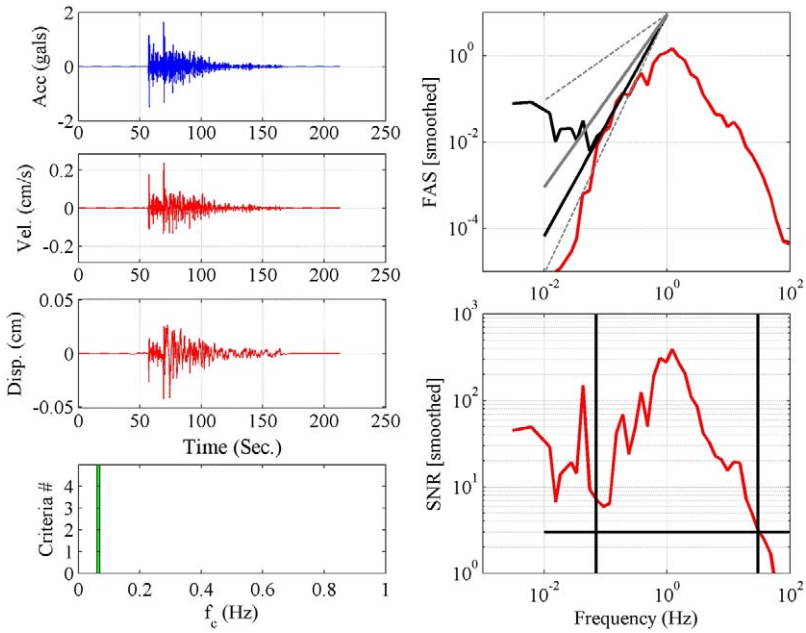

Figure A4. Acceleration, velocity and displacement time histories of the same component of motion shown in Figure A3 after applying a filter with a high-pass corner frequency of 0.07 Hz. The fourth panel on the left shows that the five preset criteria were satisfied after this filter was applied (shown as green rectangles). Four additional lines were added to the panel that shows the smoothed FAS (right top panel). The gray solid line represents the theoretical slope of the FAS decay with frequency, and the gray dashed lines represent limits of the range of slopes that we considered acceptable while processing the motions (slopes of 1 and 3). The black solid line represents the actual slope of the FAS of this component of the motion. The SNR panel (bottom right) shows the signal-to-noise ratio curve (in red), a horizontal black line that represents the threshold of 3, and two vertical black lines that represent the range of frequencies between which the SNR is checked.

is described. There were examples where the location and/or geometry of the most recent finite source model was ambiguous, in those cases we selected an older model.

The finite fault models of the following earthquakes consisted of multiple segments: 2003 Miyagi-ken Oki, 2003 Miyagi-Ken Hokobu, 2008 Northern Iwate, and 2011 Fukushima-Hamadori. For the purpose of computing R_x, the multi-segment models were simplified to a single segment. This was done subjectively, where the simplified fault parameters were selected to fit the shape of the multiple segments as accurately as possible. Thus, we expect there to be additional uncertainty in the values of R_x values that we report for multi-segment models.

For events for which a finite fault model is currently unavailable we simulate the rupture parameters following the method of Chiou and Youngs (2008b, Appendix B). We used the F-net hypocenter location, M_W, and focal mechanism. However, we could not determine

Table A1. List of earthquakes for which a finite-source rupture model was found in the literature

EQ name	Date	# of Rec.	M_W	Distance Range [km]		Classification		Reference
				R_{rup}	R_{JB}	Garcia et al. (2012)	Reference	
Western Tottori	2000/10/06	203	6.6	1.76–625	1.05–625	ACR shallow		Peyrat and Olsen (2004)
Geiyo	2001/03/24	217	6.7	45.2–564	12.2–563	ACR deep		Yagi and Kikuchi (2001)
Miyagi-ken Oki	2003/05/26	321	7.0	64.4–709	4.43–705	SZ intraslab	SZ intraslab	Wu and Takeo (2004)
Miyagi-Ken Hokobu Foreshock	2003/07/26	132	5.5	19.5–322	16.1–332	ACR shallow		Hikima and Koketsu (2004)
Miyagi-Ken Hokobu	2003/07/26	165	6.1	14.7–580	12.8–580	ACR shallow		Hikima and Koketsu (2004)
Tokachi-Oki	2003/09/25	264	8.2	51.2–792	0–792	SZ interface	SZ interface	Honda et al. (2004)
Kii Peninsula (1)	2004/09/05	303	7.3	83.1–604	75.3–6.3	ACR shallow	SZ intraslab	Park and Mori (2005)
Kii Peninsula (2b)	2004/09/05	323	7.5	71.9–571	71.7–571	ACR shallow	SZ intraslab	Park and Mori (2005)
Mid Niigata Prefecture	2004/10/23	287	6.7	2.21–406	0–406	ACR shallow		Honda et al. (2005)
West Off Fukuoka Pref.	2005/03/20	191	6.6	27.2–772	27.1–772	ACR shallow	ACR shallow	Asano and Iwata (2006)
Miyagi-Oki	2005/08/16	327	7.1	61.7–743	36.8–741	SZ interface		Wu et al. (2008)
Kuril	2006/11/15	25	8.4	559–1016	558–1015	SZ interface	SZ interface	Baba et al. (2009)
Notohanto	2007/03/25	276	–	16.3–633	14.4–633	ACR shallow		Pulido et al. (2008)
Niigata-ken Chuetsu-oki	2007/07/16	260	6.7	24.3–429	15.2–429	ACR shallow		Aoi et al. (2008)
Iwate-Miyagi Nairiku	2008/06/13	260	6.9	5.10–606	0–606	ACR shallow	ACR shallow	Suzuki et al. (2010)
Northern Iwate	2008/07/23	287	6.9	93.0–701	0–694	SZ intraslab	SZ intraslab	Suzuki et al. (2009)
Sanriku-Oki	2011/03/09	351	7.4	74.9–715	69.9–714	SZ interface		Shao et al. (2011)
Tohoku	2011/03/11	525	9.0	49.0–1009	0–1008	SZ interface	SZ interface	Suzuki et al. (2011)
Fukushima-Hamadori 1	2011/03/22	184	5.9	14.6–493	14.4–493	ACR shallow	ACR shallow	Kobayashi et al. (2013)
Tohoku Aftershock	2011/04/07	419	7.17	77.3–892	53.2–891	SZ intraslab	SZ intraslab	Ohta et al. (2011)
Fukushima-Hamadori 2	2011/04/11	461	6.7	14.4–889	14.4–889	ACR shallow	ACR shallow	Kobayashi et al. (2012)

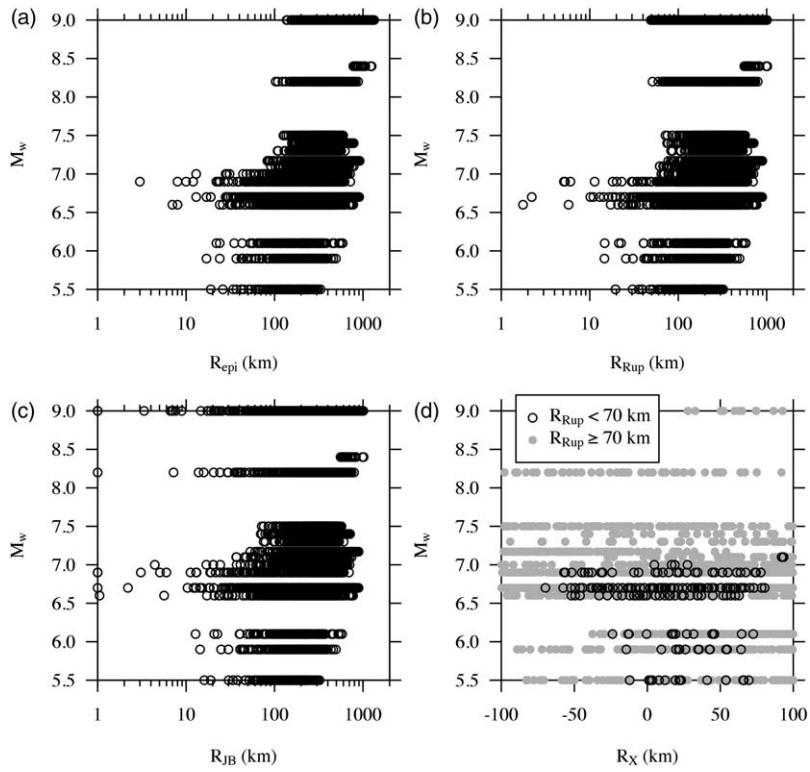

Figure A5. The distribution of M_W versus: (a) R_{epi}; (b) R_{rup}; (c) R_{JB}; and (d) R_x, for the earthquakes listed in Table A1. Distances less than 1 km were rounded up to 1 km.

which strike/dip pair was the actual fault plane for all events and so we compute the distances from 100 simulated faults for each plane separately. For each of the simulated faults, we compute the median R_{rup}, R_{JB}, R_x, and source-to-site azimuth.

The first step in the simulation process is to compute the rupture area for a given M_W. We use the M_W-area relationship by Wells and Coppersmith (1994) for active crustal events, and two M_W-area relationships by Strasser et al. (2010) for subduction zone events. We use the Garcia et al. (2012) classification to determine the appropriate magnitude-area relationship. For the events in Table A1, two of the Garcia et al (2012) classifications do not match the earthquake classification in the published literature (the two Kii Peninsula events); for these events, we used the classification given in the literature. Figure A6 compares the M_W-area relationships for the events in Table A1 with the models in the literature. The events for which we have a finite fault model are consistent with these previously published relationships, and so we use the Wells and Coppersmith (1994) and Strasser et al. (2010) equations for simulating the rupture area from a known magnitude.

The second step for simulating the faults is to estimate the aspect ratio (AR) of the ruptured area. Chiou and Youngs (2008b) developed equations for AR as a function of M_W from the NGA-W1 database. Figure A7a compares this relationship to the finite fault models in our

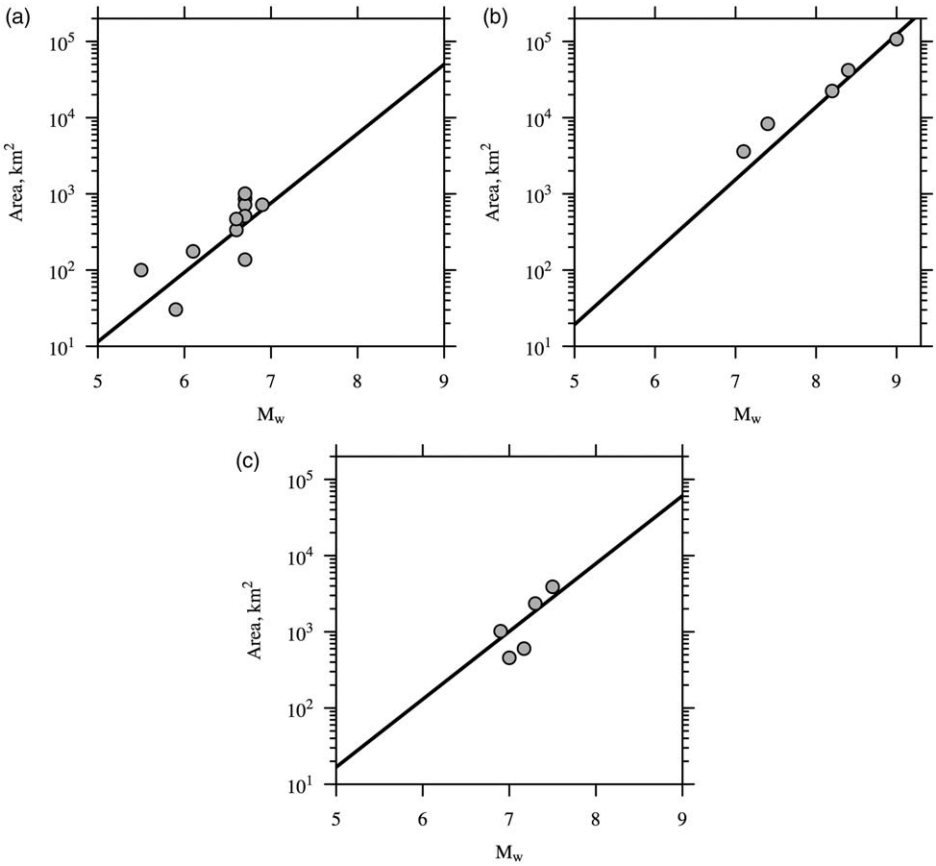

Figure A6. M_W versus rupture area for the finite fault models in Table A1; (a) compares the Wells and Coppersmith (1994) relationship for active crustal events, and (b) and (c) compare the Strasser et al. (2010) relationships to the finite fault models in Table A1 for subduction zone interface and intraslab events, respectively.

database. The finite fault models in our database are sufficiently similar to the previously published equations and so we use the Chiou and Youngs (2008b) equations for normal and strike slip events. However, the AR relationship for reverse faults tends to give values that are too large, particularly at larger magnitudes. So we estimate the coefficients from the data in Figure A7b for reverse events:

$$\log(AR) = 0.0321(M_W - 4)^{1.33}, \sigma_{\log} = 0.35 \quad \text{(A1)}$$

The third step is to simulate the location of the hypocenter on the fault plane. To do this we use the finite-fault models in our database to compute (a) the ratio of the distance along strike to the hypocenter (s1) to the fault length, and (b) the ratio of distance down-dip (w1) to the hypocenter to the fault width. Since these values are both bounded from zero to one, we fit a beta distribution to the samples. The QQ plots showing the fit of the data to the selected beta

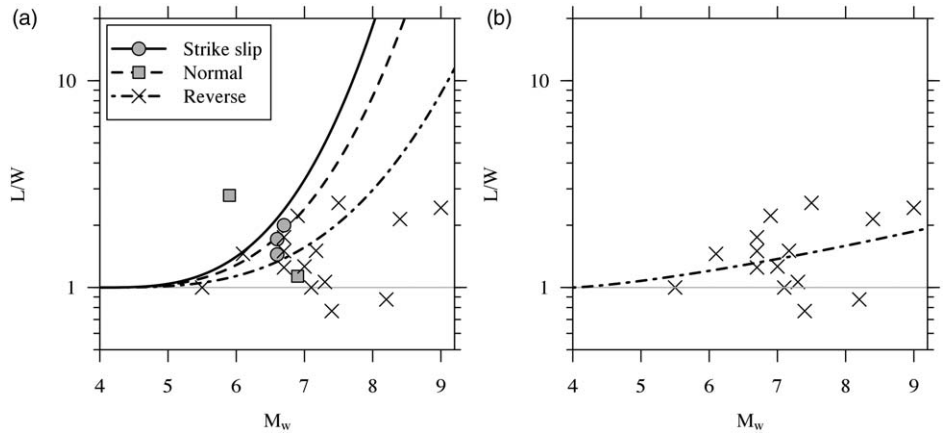

Figure A7. M_W versus AR from the finite fault models in Table A1 compared to (a) the Chiou and Youngs (2008) relationships for different fault types and (b) a curve fit to the reverse faults in Table A1.

distributions are given in Figure A8. If the depth to the top of the rupture is negative, then the value of w1 is also adjusted so that the depth to the top of the rupture is zero.

To determine a single distance value from the two planes for R_x, R_{JB}, and R_{rup}, we recommend the following rules:

- Use the plane with the shallower dip for interface events.
- For non-interface events, use the geometric mean of the two distances for R_{rup} and R_{JB}.
- For R_x, if the sign is the same for the two fault planes, then use the geometric mean of the two values.

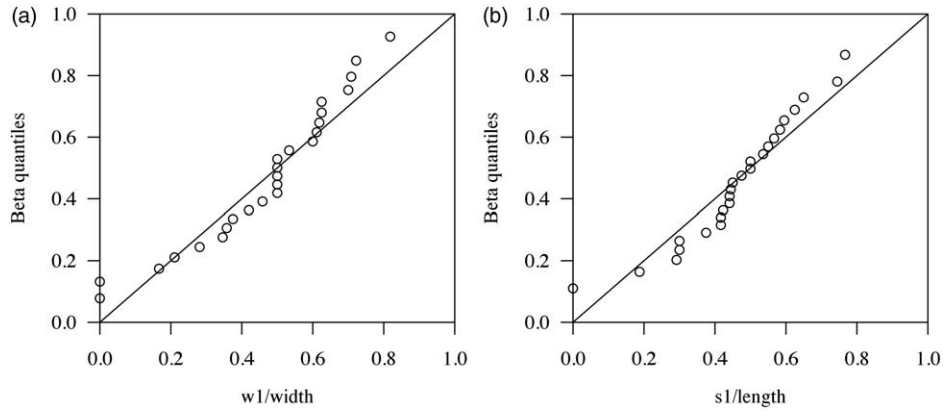

Figure A8. QQ plots for (a) w1/width, and (b) s1/length.

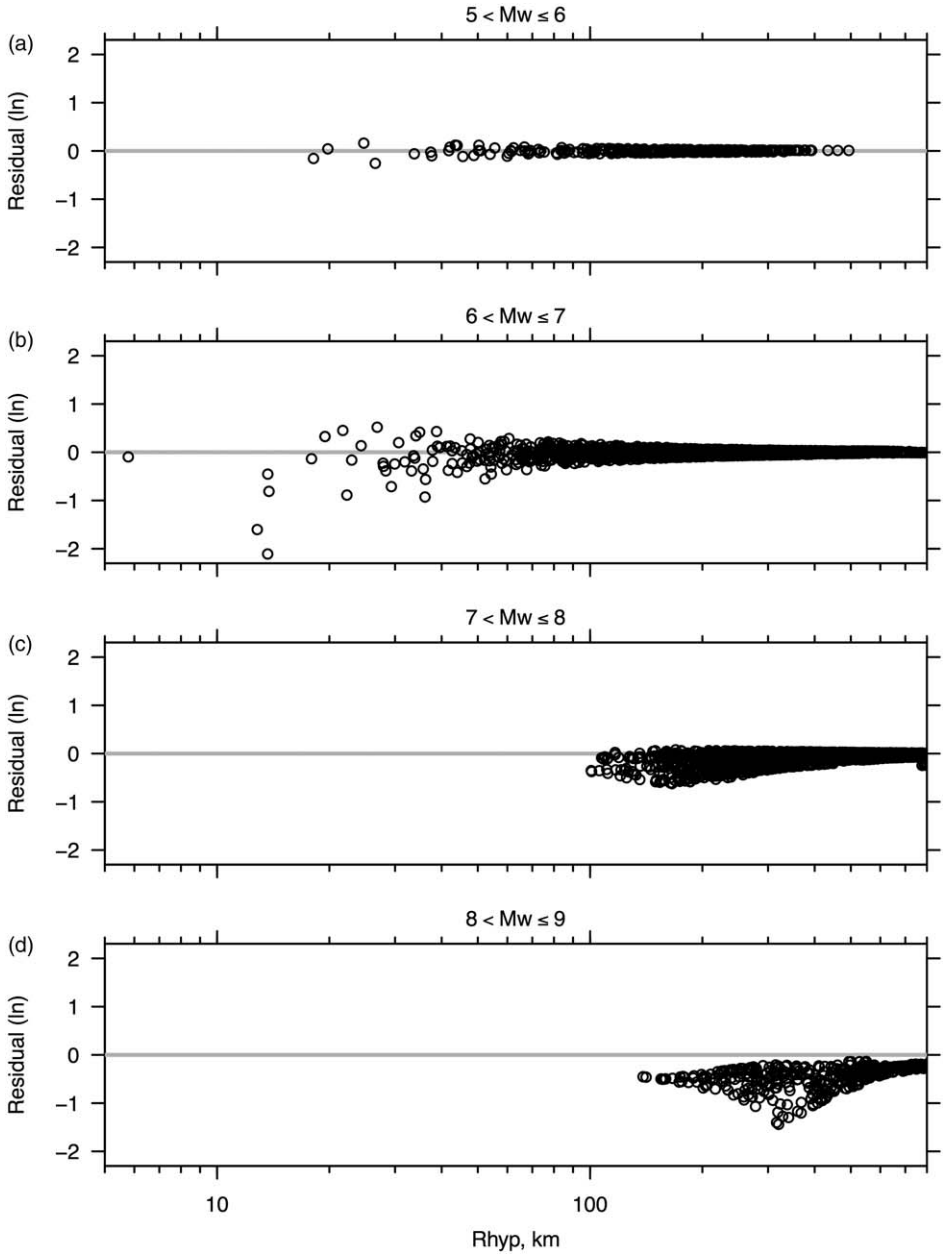

Figure A9. Comparison of the preferred simulated R_{rup} and those computed from finite fault models for different magnitude ranges.

To assess the accuracy of the simulated distances, Figure A9 plots the base ten logarithmic distance residual (using the above rules for determining a single value) versus the hypocentral distance for the events in Table A1 for four different M_W ranges. This figure shows that the simulated distances are very accurate for all distances if the $M_W < 6$. For M_W between 6 and 7 the scatter increases for hypocentral distances less than 20–30 km The scatter for all events in the 7–8 bin is very small, but this is likely because there are not any stations at hypocentral distances less than 100 km. Even where the scatter increases for events with magnitudes less than 8, the bias is relatively small. Hence, these distances may be used for situations where the additional uncertainty is acceptable. For $M_W > 8$, there is significant bias and so we do not recommend using the simulated values in this range.

REFERENCES

Aoi, S., Sekiguchi, H., Morikawa, N., and Kunugi, T., 2008. Source process of the 2007 Niigata-ken Chuetsu-oki earthquake derived from near-fault strong motion data, *Earth Planets Space* **60**, 1131–1135.

Asano, K., and Iwata, T., 2006. Source process and near-source ground motions of the 2005 West Off Fukuoka Prefecture earthquake, *Earth Planets Space* **58**, 93–98.

Baba, T., Cummins, P. R., Thio, H. K., and Tsushima, H., 2009. Validation and joint inversion of teleseismic waveforms for earthquake source models using deep ocean bottom pressure records: A case study of the 2006 Kuril megathrust earthquake, *Pure and Applied Geophysics* **166**, 55–76.

Honda, R., Aoi, S., Morikawa, N., Sekiguchi, H., Kunugi, T., and Fujiwara, H., 2004. Ground motion and rupture process of the 2003 Tokachi-Oki earthquake obtained from strong motion data of K-net and KiK-net, *Earth Planets Space* **56**, 317–322.

Honda, R., Aoi, S., Morikawa, N., Sekiguchi, H., Kunugi, T., and Fujiwara, H., 2005. Ground motion and rupture process of the 2004 Mid Niigata Prefecture earthquake obtained from strong motion data of K-net and KiK-net, *Earth Planets Space* **57**, 527–532.

Kobayashi, T., Tobita, M., Koarai, M., Okatani, T., Suzuki, A., Noguchi, Y., Yamanaka, M., and Miyahara, B., 2012. InSAR-derived crustal deformation and fault models of normal faulting earthquake (M_j 7.0) in the Fukushima-Hamadori area, *Earth Planets Space* **64**, 1209–1221.

Ohta, Y., Miura, S., Ohzono, M., Kita, S., Iinuma, T., Demachi, T., Tachibana, K., Nakayama, T., Hirahara, S., Suzuku, S., Sato, T., Uchida, N., Hasegawa, A., and Umino, N., 2011. Large intraslab earthquake (2011 April 7, M 7.1) after the 2011 off the Pacific coast of Tohoku Earthquake (M 9.0): Coseismic fault model based on the dense GPS network data, *Earth Planets Space* **63**, 1207–1211.

Park, S.-C., and Mori, J., 2005. The 2004 sequence of triggered earthquakes off the Kii peninsula, Japan, *Earth Planets Space* **57**, 315–320.

Peyrat, S., and Olsen, K. B., 2004. Nonlinear dynamic rupture inversion of the 2000 Western Tottori, Japan, earthquake, *Geophys. Res. Lett.* **31**, L05604, doi:10.1029/2003GL019058.

Pulido, N., Aoi, S., and Fujiwara, H., 2008. Rupture process of the 2007 Notohanto earthquake by using isochrones back-projection method and K-net/KiK-net data, *Earth Planets Space* **60**, 1035–1040.

Shao, G., Ji, C., and Zhao, D. 2011. Rupture process of the 9 March, 2011 Mw 7.4 Sanriku-Oki, Japan earthquake constrained by jointly inverting teleseismic waveforms, strong motion data and GPS observations, *Geophys. Res. Lett.* **38**, L00G20.

Strasser, F. O., and Bommer, J. J., 2005. *Analysis of Intra-Event Ground-Motion Residuals from K-NET and KiK-NET Data*, Research report number: 05-003-SM, Imperial College London.

Strasser, F. O., Arango, M. C., and Bommer, J. J., 2010. Scaling of the source dimensions of interface and intraslab subduction-zone earthquakes with moment magnitude, *Seismol. Res. Lett.* **81**, 941–950.

Suzuki, W., Aoi, S., and Sekiguchi, H., 2009. Rupture process of the 2008 Northern Iwate intra-slab earthquake derived from strong-motion records, *Bull. Seism. Soc. Am.* **99**, 2825–2835.

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

Suzuki, W., Aoi, S., Sekiguchi, H., and Kunugi, T. 2011. Rupture process of the 2011 Tohoku-Oki mega-thrust earthquake (M9.0) inverted from strong-motion data, *Geophys. Res. Lett.* **38**, L00G16.

Wu, C., and Takeo, M., 2004. An intermediate deep earthquake rupturing on a dip-bending fault: Waveform analysis of the 2003 Miyagi-ken Oki earthquake, *Geophys. Res. Lett.* **31**, L24619, doi:10.1029/2004GL021228.

Wu, C., Koketsu, K., and Miyake, H., 2008. Source processes of the 1978 and 2005 Miyagi-Oki, Japan, earthquakes: Repeated rupture of asperities over successive large earthquakes, *J. Geophys. Res.* **113**, B08316, doi:10.1029/2007JB005189.