IMPROVING THE UNDERSTANDING AND METHODOLOGIES OF GROUND MOTION VARIABILITY

(Deliverable D2-111)

AUTHORS			REVIEW			APPROVAL		
NOM	DATE	VISA	NOM	DATE	VISA	NOM	DATE	VISA
A. Rodriguez Marek (Virginia Tech)			F. Scherbaum (Potsdam University)			G. Senfaute		
			J. Douglas (BRGM)					

DISSEMINATION: Authors; Steering Committee; Work Package leaders, Scientific Committee, Archiving.

Final Report

DRAFT

Scientific Cooperation between EDF and Virginia Tech in the Field of Seismic Hazard Assessment

Co-operation contract No 3000-5910098949

May 13, 2014

Prepared by

Adrian Rodriguez-Marek and Haitham Dawood

Abstract

This report presents the results of the research conducted at Virginia Tech as part of the SIGMA project. The most salient contributions of this research include a database of uniformly processed response spectral accelerations for over 100000 records from the KiKnet array. These data was used to evaluate values of single-station sigma. The computed values are slightly higher than values computed in previous studies. Probably the most interesting observation is that magnitude dependence is not evident in the single-station standard deviation, although this observation is tempered by the lack of recordings at magnitudes above 7.0. The site-to-site variability of single-station phi was evaluated. The largest factor contributing to this variability is the poor sampling across various source-to-site azimuths, which implies that stations that record low values of $\phi_{ss,s}$ are likely sampling single-paths, thus these values have to be used with caution. Finally, a model to constrained path effects is presented and evaluated using a subset of the KiKnet database. The advantage of this approach is that the proposed approach allows for the estimate of "path terms" that can be used in forward prediction.

Executive Summary

This report presents the research conducted at Virginia Tech as part of the sigma project. The main contributions of this research include: a) the compilation of a comprehensive database, including metadata for each station (which, among other parameters, included the computation of kappa), of ground motions from the KiKnet network, b) an analysis of the single-station sigma from these data, including analysis of the station-to-station variability in single-station sigma, c) the development of a new approach for computing single-station, single-path sigma, and d) guidelines for the implementation of partially non-ergodic PSHA.

This study was built upon previous work done by the PI on analyses of single-station sigma for various datasets. These studies were extended with the analyses of a large dataset of ground motion compiled as part of this project. A total of about 157000 records from over 12000 earthquakes with M_{JMA} were compiled. An automated ground motion protocol was developed to apply baseline correction and to identify noisy records. The protocol applies multiple (and sometimes redundant) checks to ensure that the useable bandwidth of the records is identified. Earthquake metadata, including moment magnitude, was obtained from the F-Net network. Each event was classified using the Garcia et al. (2012) classification system to differentiate crustal and subduction events. Declustering algorithms were run to identify dependent and independent events. Station metadata was obtained directly from KiKnet. The automated protocol is described in detail in Dawood et al. (2014; see also Appendix I). To augment the station metadata, the site kappa was also computed for each station using various methods.

The KiKnet data was used to develop a ground motion prediction equation (GMPE) for crustal earthquakes. The development of the GMPE used the functional form of Abrahamson et al. (2013; part of the NGA West 2 project). The regression was conducted on 13735 six-component GM records from 679 earthquakes recorded at 643 stations. A combination of linear and nonlinear mixed-effects regression analysis using SAS is conducted for this analysis. The regression analysis is conducted in several consecutive steps. In each step a certain coefficient is smoothed along all spectral periods and then fixed for the next steps. Borehole and surface records are simultaneously regressed for and the same magnitude and distance scaling coefficients are set for both types of records. On the other hand, the coefficients that represent the site effect in the GMPE are different for surface and borehole records. We also note that a GMPE for subduction earthquakes was developed as part of a parallel project using the KiKnet data supplemented with data from other regions.

The GMPE for crustal earthquakes was used to study the standard deviation components of ground motion parameters (pseudo-spectral accelerations at 5% damping), with a particular focus on the standard deviation of station- and event-corrected residuals (e.g., ϕ_{ss}). The values of ϕ_{ss} averaged over all magnitudes and distances are, on average, about 10% higher than those computed by Rodriguez-Marek et al. (2013; RM13) for the PRP project. For longer periods, the values are similar. One feature of the RM13 model was distance dependence. The distance dependence is strong when looking across all magnitudes; however, this dependence is controlled by small magnitude earthquakes. The RM13 model had a strong distance dependence which is not observed in the KiKnet data. This is possibly due to the limited magnitude range of the data. Distance- and Magnitude-dependent models for ϕ_{ss} are given to illustrate the observed dependencies on these parameters. Single station sigma values for subduction data were in

general period independent with an average value of 0.45. These values are somewhat smaller at short periods and higher at long periods, than those computed for crustal earthquakes, but in general the values are remarkably similar.

One objective of this study was to find a correlation between the single-station phi at a given station (we will denote this as $\phi_{ss,s}$) and site- or event-parameters such that it would be possible to determine a-priori those stations that have a low value of $\phi_{ss,s}$. One clear result of this study is that the measurement of single-station sigma at a station depends strongly on the degree of variation in source-to-site azimuths of recordings at a station: stations that sample only a narrow range of azimuths tend to have low values of $\phi_{ss,s}$. This observation points to the importance to account for path effects in the computation of single-station sigma. A novel approach is proposed where corrections to the anelastic attenuation are computed for regions within a geographical grid. The advantage of this approach (as opposed to the approach of Lin et al., 2013) is that the proposed approach allows for the estimate of "path terms" that can be used in forward prediction.

Other parameters were study to determine if there is a correlation with $\phi_{ss,s}$. Topography was observed to have a relatively strong bias using two sets of California data (the NGA West 2 dataset and the small and medium magnitude dataset of Chiou and Youngs, 2010). For these data, a simple parameterization of topography indicated clear bias in the site term for stations located in valleys (and to a lesser degree those in hills), and a relatively strong effect on single-station phi at periods up to about 0.5 seconds. The same results, however, were not observed in Japanese stations. Further study is underway to determine if the correlations seen in California data are indeed topographic-related or the parameter used compensates in some other way for the deficiencies in parameterization of site response. Another parameter that was investigated was the Horizontal to Vertical Horizontal Ratio. In particular, we tested the hypothesis that stations where the HVSR have a strong azimuthal dependency would have stronger variability than stations where HVSR are independent of azimuth. The results, however, do not support this hypothesis: there is only a mild correlation between measures of azimuthal dependency of HVSR and $\phi_{ss,s}$. The effect of nonlinearity on single-station sigma has also been study, but only preliminary results are available at this time.

Kappa was not included as part of the regression model; however, preliminary analyses of residuals indicate that the introduction of kappa can somewhat reduce the bias in the prediction of the site term, in particular for sites with high kappa. Somewhat surprisingly, there is also a trend with kappa in the values of $\phi_{ss,s}$ at a given station, with lower $\phi_{ss,s}$ for higher kappas. This implies that sites with higher kappa have less within-site variability. This empirical observation should serve as a motivating factor to model the effect of varying site conditions on the within-site variability of ground motions.

An important corollary to the research presented here are the necessary requirements for conducting a partially non-ergodic probabilistic seismic hazard analysis (e.g., a PSHA using single-station sigma). These requirements include: 1) The median value of the site term must be properly estimated for the site under analysis, 2) the epistemic uncertainty on the value of the site term must be fully accounted for, and 3) the epistemic uncertainty on the single-station sigma must be accounted for. All three of these requirements must be met.

Scientific Cooperation between EDF and Virginia Tech in the Field of Seismic Hazard Assessment

1   Report Organization

This report presents the research conducted at Virginia Tech as part of the sigma project. The main contributions of this research include: a) the compilation of a comprehensive database, including metadata for each station (which, among other parameters, included the computation of kappa), of ground motions from the KiKnet network, b) an analysis of the single-station sigma from these data, including analysis of the station-to-station variability in single-station sigma, c) the development of a new approach for computing single-station, single-path sigma, and d) guidelines for the implementation of partially non-ergodic PSHA.

This study was built upon previous work done by the PI on analyses of single-station sigma for various datasets. These studies were extended with the analyses of a large dataset of ground motion compiled as part of this project. This report summarizes the main results of this research; more details are presented in Appendices (each Appendix is a paper that has been either submitted or accepted for publication). The main body of the report starts by presenting background into the concept of single-station sigma. This is followed by a discussion on the elements that are necessary for conducting partially non-ergodic seismic hazard analyses (e.g., PSHA using single-station sigma). We then present a summary of the scientific contributions of this study.

2   Objectives of Project

The objectives of this project, identified at the project's outset are:

a)   Identification of the stations for which variability is larger than ergodic values. Investigation on the physical causes (e.g. geology, geophysics parameters) of such a large variability will be carried out using available seismic data and information. This work will include exploring GIS-based parameterization of surface topography to determine whether topographical effects lead to systematic changes in the single-station standard deviation.

b)   Guidelines for PSHA application. PSHA includes the integration across aleatoric uncertainty and the inclusion of epistemic uncertainty through logic trees. The use of single-station sigma implies eliminating a component of uncertainty associated with site-to-site variability. However, the estimation of the site term, in most cases, is coupled with a degree of epistemic uncertainty. This uncertainty must be put back into the PSHA analysis.

c)   Effects of nonlinearity. Soil non-linearity can have significant effects on the site term and potentially on the single-station standard deviation. The onset of nonlinearity can also lead to site terms that are not constant across magnitude, making the estimate of single-station sigma more difficult. Finally, a procedure must be established to incorporate nonlinearity into non-ergodic PSHA.

The first two objectives were accomplished and are described below. The third objective is still under study. In addition to these objectives, a large effort of this project has been dedicated

to the compilation of an extensive database of strong ground motions. A flatfile describing these motions will be made publicly available.

3   Background on the single-station sigma concept

In applying a GMPE to the assessment of seismic hazard at a specific location, the interest is in the variation of motions at this site due to different earthquakes that could occur over time. Since it is rare to have recording from the location under study, and even in the few cases where such recordings exist they will cover at most a few decades, PSHA generally invokes what is referred to as the ergodic assumption (Anderson and Brune, 1999). The ergodic assumption essentially states that variability over space can be used as a substitute for variation over time, and it is invoked in practice since the sigma values calculated from regression analyses to develop GMPEs represent the variability across many different sites (and sometimes many regions). Where there are multiple recordings from individual sites, they display lower variability than indicated by the sigma values of GMPEs (e.g., Atkinson, 2006), the reason being that there are components of the behavior at an individual site that are repeated in all cases and therefore do not contribute to variability. The effect is even more pronounced when there are multiple recordings from a single site of earthquakes associated with a single seismogenic source because in that case there are repeatable effects of both the path and the site.

If the repeatable contributions to the seismic motion at the site of interest can be modeled through an appropriate adjustment to the median predictions, then the sigma value can be reduced to a value referred to as 'single-station sigma' (Atkinson, 2006). The reduction is proportional to an amount that reflects the variability of the site term. The single-station sigma concept can also be invoked if epistemic uncertainty in the site term is incorporated into the logic-tree for the PSHA. In such a case, the net effect on the mean hazard is expected to be zero since the increased epistemic uncertainty balances the decreased aleatory variability, but it does mean that the division between randomness and uncertainty is being applied more completely. The ergodic assumption effectively folds the epistemic uncertainty regarding individual site terms into the sigma value of the GMPE, representing it as aleatory variability.

In this report, we follow the notation of Al Atik *et al.* (2010). Total residuals (Δ_{es}) are defined as the difference between recorded ground motions and the values predicted by a GMPE (in natural log units). Total residuals are separated into a between-event term (δB_e) and a within event term (δW_{es}):

$$\Delta_{es} = \delta B_e + \delta W_{es} \tag{1.1}$$

where the subscripts denote an observation for event e at station s. The between-event and the within-event residuals have standard deviations τ and ϕ, respectively, and are assumed to be uncorrelated. The within-event residuals can in turn be separated into:

$$\delta W_{es} = \delta S2S_s + \delta WS_{es} \tag{1.2}$$

where $\delta S2S_s$ represents the systematic deviation of the observed ground motion at site s (*e.g.*, the "site term") from the median event-corrected ground motion predicted by the GMPE, and δWS_{es} is the site- and event-corrected residual. The standard deviation of the $\delta S2S_s$ and δWS_{es} terms are denoted by ϕ_{S2S} and ϕ_{ss}, respectively. Table 1.1 lists the components of the total residual, their respective standard deviations, and the terminology used for each standard deviation component.

Table 1.1. Terminology used for residual components and their standard deviations. $SD(\cdot)$ denotes the standard deviation operator.

Residual Component	Residual Notation	Standard Deviation component	Definition of standard deviation component
Total residual	Δ_{es}	Total or ergodic standard deviation	$\sigma_{ergodic} = SD(\Delta_{es})$
Event term	δB_e	Between-event (or inter-event) standard deviation (τ)	$\tau = SD(\delta B_e)$
Event-corrected residual	δW_{es}	Within-event (or intra-event) standard deviation (phi)	$\phi = SD(\delta W_{es})$
Site term	$\delta S2S_s$	Site-to-site variability	$\phi_{S2S} = SD(\delta S2S_s)$
Site- and event-corrected residual	δWS_{es}	Event-corrected single-station standard deviation (single-station phi)	$\phi_{ss} = SD(\delta WS_{es})$

In traditional (*e.g.*, ergodic) PSHA, all of the residuals components are considered as part of the aleatory variability, such that:

$$\sigma_{ergodic} = \sqrt{\tau^2 + \phi_{ss}^2 + \phi_{S2S}^2} \qquad (1.3)$$

In the partially non-ergodic approach, the site term ($\delta S2S_s$) is assumed to be known (or knowable) and hence its standard deviation (ϕ_{S2S}) is excluded from Eq. 1.3. In this case, the standard deviation is known as the *single-station standard deviation* and is given by:

$$\sigma_{ss} = \sqrt{\tau^2 + \phi_{ss}^2} \qquad (1.4)$$

The principal motivation to adopt a single-station sigma for hazard projects is to avoid double counting uncertainty. This double counting would result if the site-to-site variability (ϕ_{S2S}) is included in the total sigma (see Eq. 1.3) and in addition the site term is assigned an epistemic uncertainty.

An additional motivation for the adoption of a single-station sigma approach in hazard projects is that the value of single-station phi (ϕ_{SS}) has proven to be relatively constant across different regions and tectonic environments (Rodriguez-Marek et al., 2013). The lack of regional dependence of the single-station phi implies that they are more readily "exportable" to different regions, and that global datasets can be used to estimate their values.

The topic of single-station sigma has garnered considerable attention in the last few years. The idea of including site terms in the regression for ground motion prediction equations was first proposed by Joyner and Boore (1981), and the concept of partially non-ergodic PSHA and the associated single-station sigma was first formulated by Anderson and Brune (1999). However, only the recent increase in ground motion data has allowed for the computation of single station sigma and the development of models for this component of variability (e.g., Chen and Tsai, 2002; Atkinson, 2006; Morikawa et al., 2008; Lin et al., 2011; Rodriguez-Marek et al., 2011; Ornthammarath et al., 2011; Chen and Faccioli, 2013; Luzi et al, 2014). The models developed for the Pegasos Refinement Project (PRP; Renault et al., 2010) in Switzerland, and the model developed for the Thyspunt Nuclear Siting Project (TNSP, Bommer et al., 2014) in South Africa (Rodriguez-Marek et al., 2014) are the only models (known to the authors) that have postulated actual values for use in hazard analyses. The work of this proposal is an evolution from these two models.

4   Requirements for application of single-station sigma

The basic requirements to apply a partially non-ergodic PSHA are:

1. The median value of the site term ($\delta S2S_s$) must be properly estimated for the site under analysis.

2. The epistemic uncertainty on the value of the site term must be fully accounted for.

3. The epistemic uncertainty on the single-station sigma must be accounted for.

These requirements reflect both the results of this study (which highlight the station-to-station variability in single-station sigma), and the experience of the PI in several Nuclear Power Plant projects. In particular, both in the Thyspunt Nuclear Siting Project in South Africa (Bommer et al., 2014; Rodriguez-Marek et al., 2014) and in an ongoing project for the Columbia Generating Station in Washington State, USA, the PI is member of the ground motion characterization Technical Integration team. In both these projects, the guidelines outlined above were considered essential requirements for the implementation of a partially non-ergodic PSHA.

The first requirement was discussed while deriving Equation 1.4. The only reason why single-station sigma can be invoked is because the site term is assumed to be known (i.e., a deterministic quantity with zero standard deviation) or "knowable". The second requirement is because if the site term is not known with exactitude (which will be essentially in all cases), the site term has an uncertainty that must be accounted for, but since this uncertainty is epistemic in nature, it is not accounted as part of the aleatory variability (e.g., the "sigma"). It is important to note that the uncertainty in the site term comes both from the response of shallow soil deposits as well as from the response at a larger scale (e.g., upper few kilometers). This is evident in the correlation of site terms at the surface and at depth in the KiKnet data (Rodriguez-Marek et al., 2011). From a practical stand point, this implies that the site term (and its uncertainty) must be accounted for through site response analyses that account for the effects of shallow

surface deposits, and analyses that quantify regional differences in the response down to a few kilometers (what is known as Vs-kappa corrections, e.g., Van Houte et al., 2011; Al Atik et al., 2013). Both these computations have inherent uncertainties, both in the input parameters and in the modeling processes, and these uncertainties must be accounted for either in the logic tree or in the site response analyses (e.g., Bazurro and Cornell, 2004).

The final requirement is not as readily apparent as the two previous ones, but it is equally important. Observations from large databases (Rodriguez-Marek et al., 2011 and 2013) have shown that there are significant site-to-site variations in the standard deviation of the site- and event-corrected residuals at a single station (a term that will be referred to as $\phi_{ss,s}$). While over all the stations this standard deviation has an average that is close to ϕ_{ss}, at any given station this standard deviation has an uncertainty. This uncertainty is partially due to measurement uncertainty (Rodriguez-Marek et al., 2013), but it also responds to physical processes. At any given site, factors such as topography or subsurface layering could introduce more (or less) variability than at other sites.

5   Scientific Contributions

The scientific contributions of this project are summarized below.

5.1   Flatfile compilation

KiK-net is one of several seismic networks established in Japan following the devastating Kobe earthquake (January 17, 1995) to better monitor the seismic activity around the country (Okada et al., 2004). As of December, 2011, the KiK-net network consisted of 692 stations. Each KiK-net station consists of two strong GM seismographs, one at the ground surface and the other in a borehole, and each instrument records three components of motion. The seismic velocity profile for 655 of these stations is reported in the KiK-net website (http://www.kyoshin.bosai.go.jp/). The velocity profiles at these stations were obtained from downhole PS logging (Oth et al., 2011). Details regarding the KiK-net network and the specifications of the instruments are given in Aoi et al. (2011) and Okada et al. (2004).

We processed data from the Japanese KiKnet network using an automated processing protocol (ground motion processing was started prior to the initiation of this project, but was concluded with funding from this project). Data was downloaded from the KiKnet website for events up to December 2011. In addition to the information contained in the KiKnet records, data from the F-Net seismic catalog (Okada et al., 2004) was used to compute associated metadata. 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. The catalog includes 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**, moment tensor solution, variance reduction and number of stations used).

The automated processing protocol applies first a a zero[th] order baseline correction to the record by first subtracting the mean of the first 100 points from the whole 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 automatically detected arrival time for the six components. The records are first tapered and zero-padded and then an 8th order acausal Butterworth filter is applied.

The low-cut frequency is selected iteratively to satisfy several criteria to ensure zero displacement and zero velocity at the end of the time history. The minimum useable frequency of the records is taken to be 70% the low-cut corner frequency of the filter (f_c). The maximum frequency is defined by a built-in anti-aliasing filter in the instruments and is 30 Hz (Aoi et al., 2011). A signal-to-noise ratio (SNR) check is also applied to all records. A noise window is defined using the last $2/f_c$ second of the record. The FAS for the whole record and the noise window are calculated, smoothed, and the ratio between both is calculated and defined as the SNR. Frequencies outside the range where the SNR is above three are not used.

Both crustal and subduction records were processed. We compared common records processed using the automated protocol and the NGA-W2 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-W2 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 dataset.

Most of the earthquake metadata was obtained from the F-Net catalog. Earthquakes identified in the KiK-net record files were matched to earthquakes in the F-Net catalog using spatial and temporal windows. The moment magnitude and hypocentral location from the F-Net catalog were used for each earthquakes. Not all earthquakes in the KiK-net records could be matched to an event in the F-Net catalog, and those records without a matching event were discarded. A notable exception to the source of moment magnitude was the 11 March 2011 Tohoku earthquake. For this earthquake, the moment magnitude reported in the F-Net catalog is 8.7. We used instead a moment magnitude of 9.0 as reported by the USGS (Hayes, 2011).

The identification of dependent and independent events was done through declustering the F-net catalog. The F-net seismic catalog was declustered using the algorithms by Gardner and Knopoff (1974) and Reasenberg (1985). The Gardner and Knopff algorithm was implemented using three different sets of input parameters (see Dawood, 2014, for more information). None of the application of the declustering algorithms resulted in a Poissonian catalog. The algorithm results using the typical input parameters of Gardner and Knopoff's algorithm were used.

The database compiled by Dawood et al. (2014) includes crustal and subduction earthquakes. The algorithms by Allen et al. (2008) and Garcia et al. (2012) were used to classify the earthquakes in the F-net catalog into tectonic categories. The latter algorithm was validated by automatically classifying a catalog of earthquakes that also were manually classified. For most event types, the validation showed a considerable improvement in the number of correctly classified events using this algorithm in comparison to the Allen et al. (2008) algorithm. However, for intraslab events the number of misclassified events increased. This was attributed to the lack of slab models for about half of the misclassified earthquakes. Additional details are given in Garcia et al. (2012). The results of applying both algorithms to the F-net catalog are shown in Table 5.1.1. The 951 earthquakes that were not classified using the algorithm by Allen et al. (2008) are events with $\mathbf{M} \leq 7.7$ and depth greater than 50 km. Allen et al. (2008) do not provide a classification for events that fall within these depth-magnitude combinations. The number of events classified as interface events using both algorithms are very different. This is consistent with the

Garcia et al. (2012) finding that the Allen et al. (2008) algorithm misclassified about 54% of the interface events. For this project we use the Garcia et al. (2012) classification.

Table 5.1.1. Classification of F-net earthquakes in the Dawood et al. (2014) database using the algorithms by Allen et al. (2008) and Garcia et al. (2012)

Class[a]	Number of earthquakes classified by Allen et al. (2008)	Number of earthquakes classified by Garcia et al. (2012)
SZ intraslab	1161	1123
SZ interface	5	873
SZ outer	-	16
Shallow active crustal	1093	-
ACR deep	-	112
ACR shallow	-	1083
OBR	-	3
Not classified	951	-
Total	3210	3210

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

The epicentral and hypocentral distances were computed by Dawood et al. (2014). For events with published finite fault solutions, the published finite-fault was used to compute the closest distance to the rupture plane (R_{rup}). For other events, the method of Chiou and Youngs (2008b, Appendix B) was used. This method uses the published hypocenter location and the two fault plane solutions from the F-Net catalog to simulate fault planes from which R_{rup} is computed. For interface events, the plane with the shallower dip was used; for intraslab events, distances were computed for the two planes and the geometric mean from both planes was used. For additional information, see Appendix I.

An additional parameter that was computed for all the KiKnet stations is the site kappa. While not included the initial proposal, this work was considered important because the introduction of kappa can lead to reduction in aleatoric variability in ground motion prediction equations, hence a better constrain on single-station sigma. Towards achieving this goal, a Ph.D. student has traveled to Paris to work with Dr. Fabian Bonilla of the LCPC. This was possible in part thanks to financial collaboration from the LCPC. Results of this effort (which is work in progress) are summarized in Appendix III.

5.2 Ground Motion Prediction Equation development and study of single-station within-event standard deviation (φ_{SS})

We developed a ground motion prediction equation (GMPE) using motions from active crustal earthquakes from the database described in Section 5.1. The regression analyses were conducted using a mixed-effects model with both site and event terms. The site- and event-corrected residuals were then used to study the dependence of ϕ_{ss} on magnitude and distance.

5.2.1 Ground motion prediction equation (GMPE)

In this study we developed a site-specific GMPE using a functional form similar to the Abrahamson et al. (2014) relationships from the NGA West 2 project. The database used in the regression analyses consisted of a total of 13735 six-components strong GM records from 679 active crustal earthquakes (shallow and deep) recorded at 643 KiK-net stations.

Regression analyses were conducted using a combination of linear and nonlinear mixed-effects regression using SAS. The regression analyses were conducted in several consecutive steps. In each step a certain coefficient is smoothed along all spectral periods and then fixed for the next steps. Borehole and surface records were included simultaneously in the regressions, hence the magnitude and distance scaling was constrained using both sets of records. On the other hand, the coefficients that represent the site effect in the GMPE are different for surface and borehole records. Details of the GMPE development are included in Appendix II.

5.2.2 Single-station within-event standard deviation (φ_{SS}) models

Models for single station phi (ϕ_{ss}) must correctly account for magnitude and distance dependencies existing in the data. The models developed for the PRP project (Rodriguez-Marek et al., 2013) include constant, magnitude dependent, distance dependent, and magnitude and distance dependent models. These dependencies relied on relatively sparse data at large magnitudes and short distances. The model used for the Thyspunt Nuclear Siting Project (Rodriguez-Marek et al. 2014) used only the magnitude dependence because the distance dependence was controlled by small magnitude events that did not contribute to hazard for this project. As part of this project, we use the database and GMPE developed as part of this project to study the magnitude and distance dependence of ϕ_{ss}.

Rodriguez-Marek et al. (2011) estimated single-station phi values for Japan, and Rodriguez-Marek et al. (2013) estimated values using motions from different regions. Figure 5.2.1 shows a comparison by the $\boldsymbol{\phi_{ss}}$ values estimated in this stud with those of the previous studies. The $\boldsymbol{\phi_{ss}}$ estimated in the current study are slightly higher than the values estimated in previous studies. Values of ϕ from this study also tend to be higher than the values from previous studies at short periods, but lower at long periods (Figure 5.2.2). Plots of ϕ_{ss} and ϕ_{ss} including only large magnitude (M>5) events are also indluded in Figures 5.2.1 and 5.2.1. The same trends (higher values) are observed for these magnitude ranges. Further studies are currently being conducted to try to understand these differences.
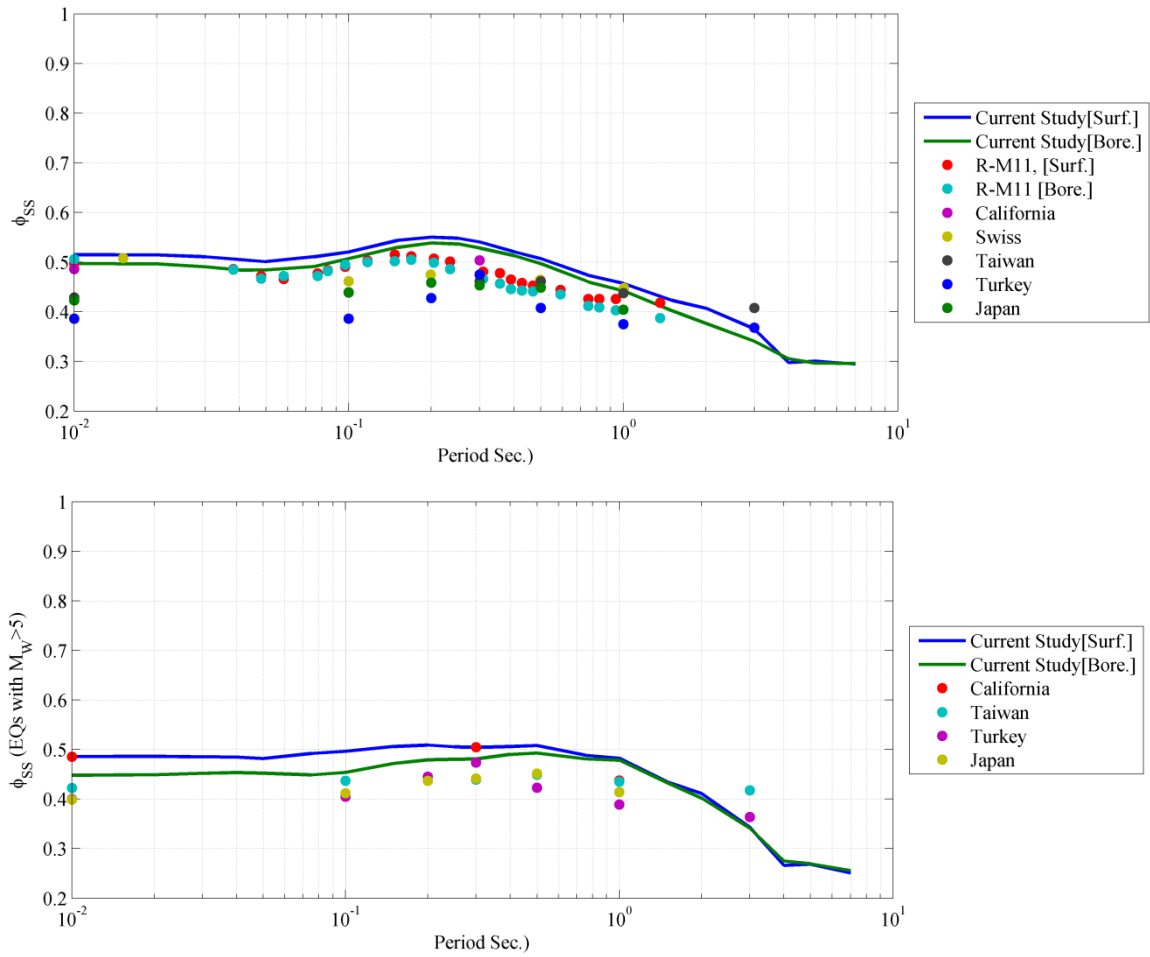
Figure 5.2.1. Comparison of ϕ_{ss} values from this study and Rodriguez-Marek et al. (2011; R-M11) for Japan, and Rodriguez-Marek et al. (2013) for various tectonic regions. PGA is plotted at T=0.001 seconds.
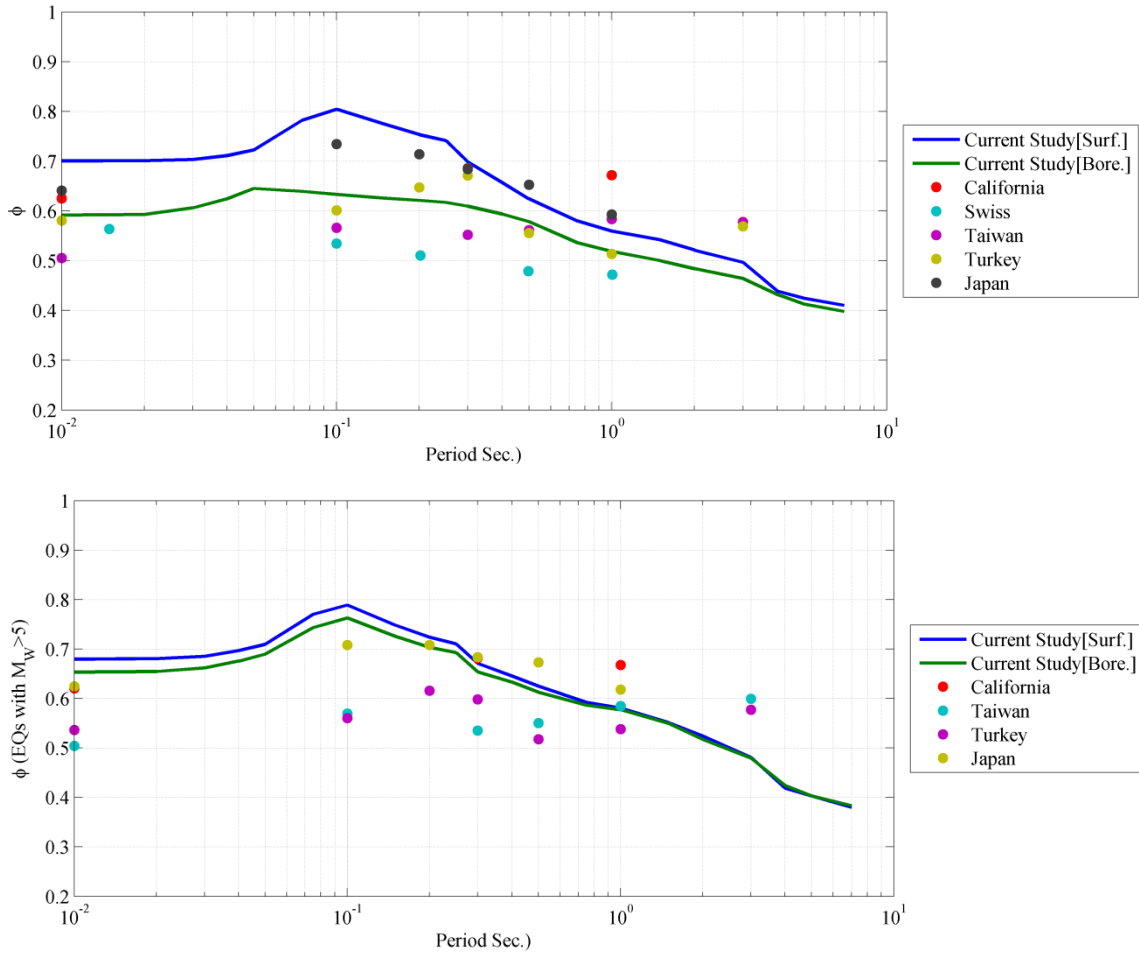
Figure 5.2.2. Comparison of ϕ_{ss} values from this study and Rodriguez-Marek et al. (2013) for various tectonic regions. PGA is plotted at T=0.001 seconds.

In addition to computing average ϕ_{ss} values, an attempt was made at developing magnitude- and distance- dependency into models for ϕ_{ss}. For both magnitude- and distance-dependence, bilinear models were fit to the data. The magnitude and distance at which there is a break in the linear relationship was smoothed across period. Figure 5.2.3 shows the magnitude dependent model for four different periods. The magnitude-dependence is not strong in the data, and is poorly constrained for large oscillator periods. While the models presented herein are not intended as models for direct use in hazard application, it appears from the data that a model including only magnitude dependency is not superior to a magnitude-independent model. Figure 5.2.4 shows the distance dependent models. Distance dependence is clear for low oscillator periods, but for long oscillator periods, the distance dependence is not well constrained and the constant model appears to performed better. While there is the possibility that distance dependence is due to poor constrain of the metadata at small distances (in particular, hypocentral depth), seismological simulations have also shown larger scatter at small distances (Cotton, personal communication).

The magnitude- and distance-dependent models are shown in Figure 5.2.5. At short periods, these models show the same characteristics or previous models, namely thas ϕ_{ss} values at larger for small magnitudes and short distances, and that distance dependence disappears for larger magnitudes. At an

14

oscillator period of 1 second, the magnitude-dependence is reversed (this is also seen in Figure 5.2.3).and at 7 seconds the distance dependency is reversed.

Additional discussion of the analysis of residuals is presented in Appendix II.
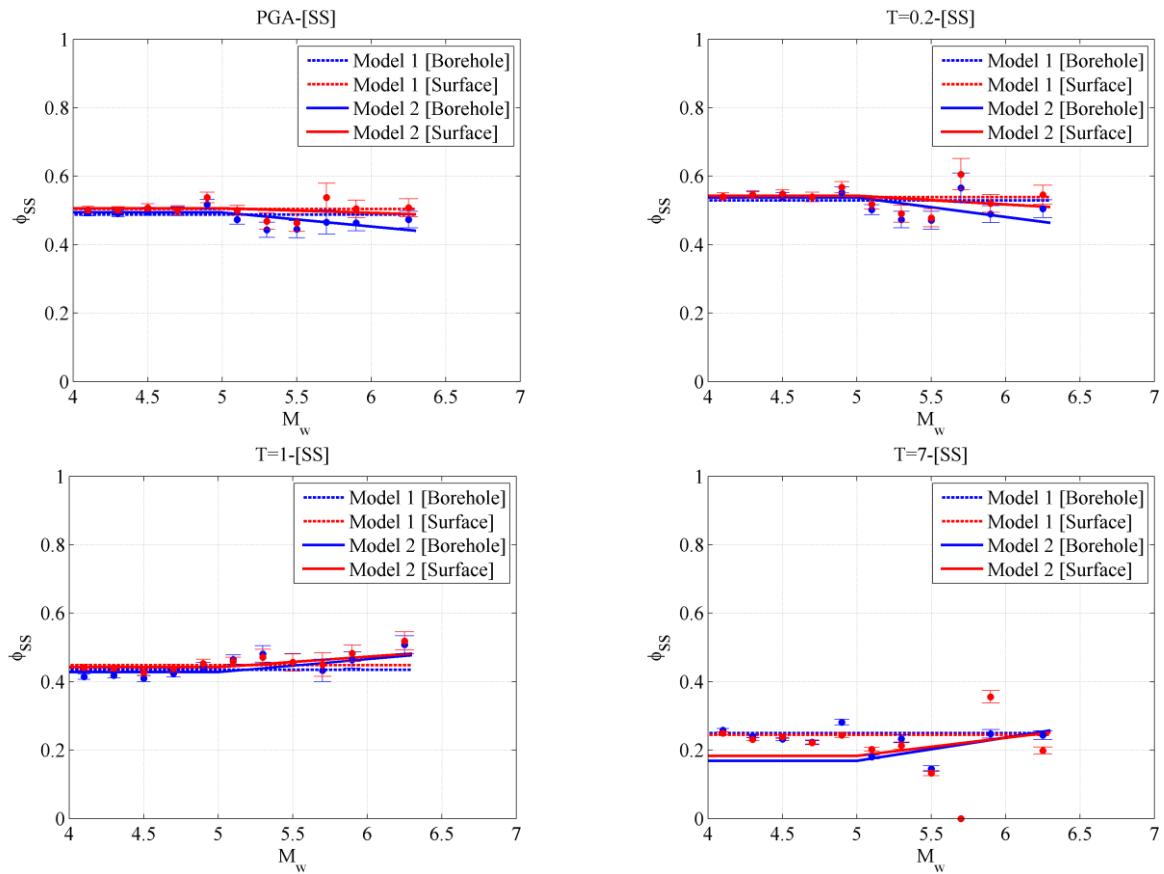
Figure 5.2.3. Models for ϕ_{ss}. Model 1 is a distance- and magnitude-independent model. Model 2 is a magnitude-dependent model. The latter model is poorly constrained for T=7 seconds. The circles show the standard deviation of site and event corrected residuals within a magnitude bin.
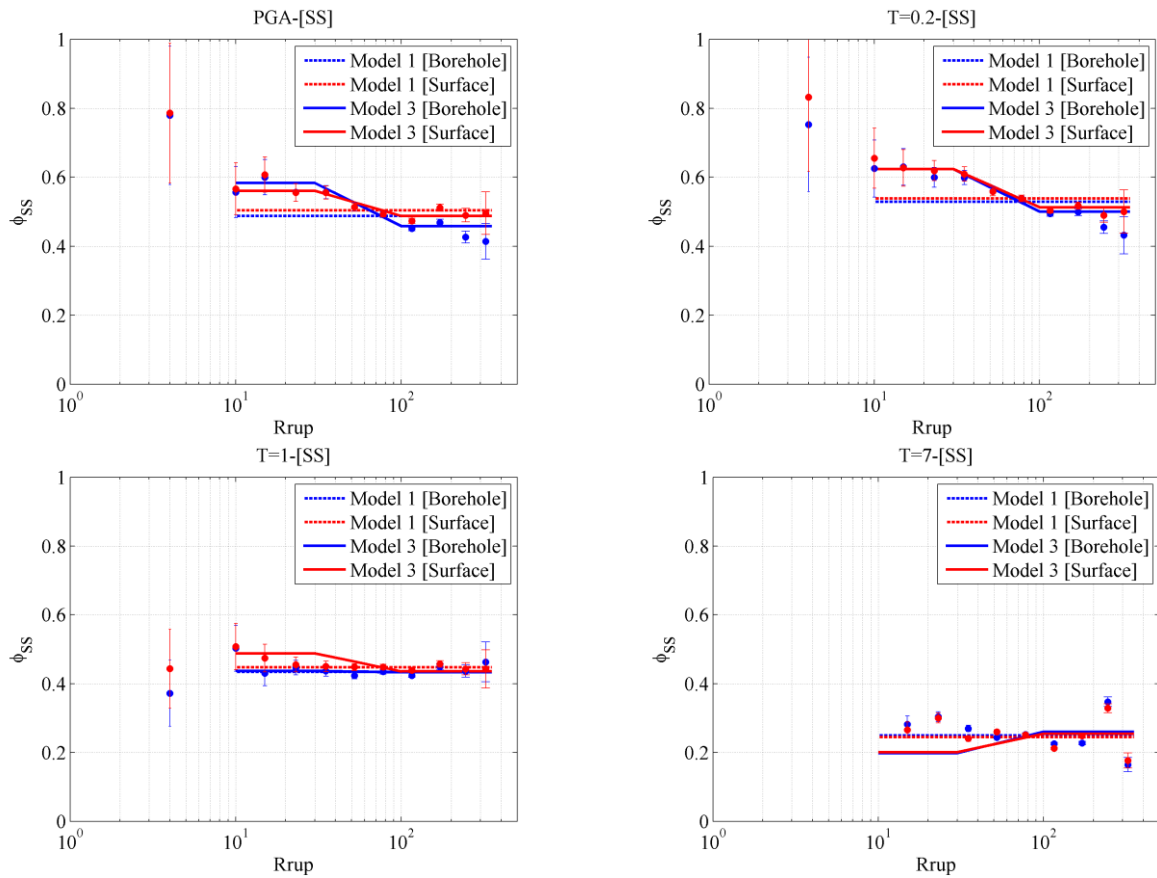
Figure 5.2.4. Models for ϕ_{ss}. Model 1 is a distance- and magnitude-independent model. Model 3 is a distnace-dependent model. The latter model is poorly constrained for T=7 seconds. The circles show the standard deviation of site and event corrected residuals within a distnace bin.
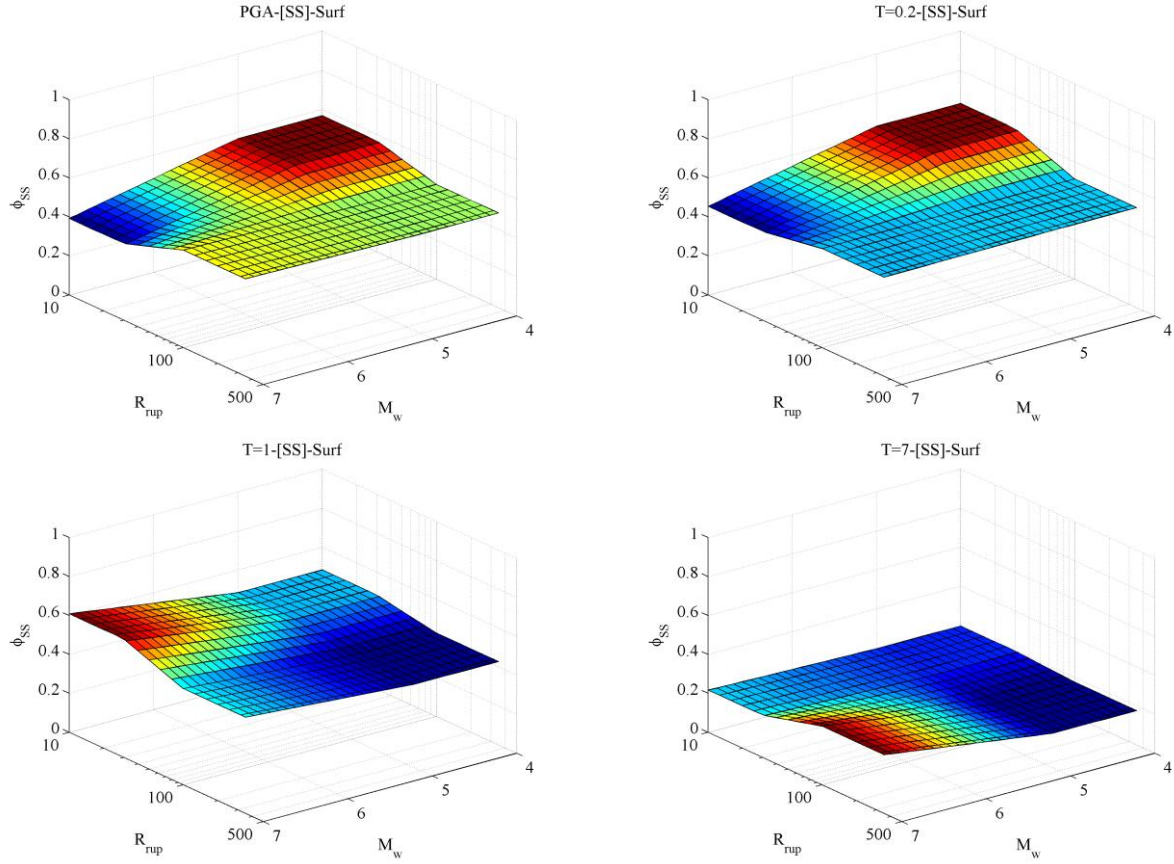
Figure 5.2.5. Models for ϕ_{ss}. Model 1 is a distance- and magnitude-independent model. Model 2 is a magnitude-dependent model. The latter model is poorly constrained for T=7 seconds.

5.3 Identification of stations with large single-station phi ($\phi_{ss,s}$) values

The standard deviation of the site- and event-corrected residuals (δW_{es}^0) measured from records at a single site is referred to as $\phi_{ss,s}$. The difference between φ_{ss} and $\varphi_{ss,s}$ is that the first represent the observed variability for all stations that recorded motions in the database, while the later represent the variability at a specific station. The average of these values across the entire database converges to the value of ϕ_{ss}, but for any given station, these values can change, and determining a-priori which sites could have large $\phi_{ss,s}$ is important for future hazard studies. As part of this project, we investigate potential causes for high $\phi_{ss,s}$ values.

5.3.1 Azimuth dependence

It was observed that $\phi_{ss,s}$ can be either much higher or much lower than the $\phi_{SS}^{G/B}$ (where superscripts represent Ground or Borehole) estimated from the models for some stations. This is important because the adopted value of $\phi_{ss,s}$ can have a major impact on the hazard curves obtained from PSHA. Figure 5.3.1 shows the histograms of the $\varphi_{ss,s}$ for stations that recorded 5 motions or more. The figure also presents the mean, mean±standard deviation of $\phi_{SS,S}^G$ and the distance and magnitude independent ϕ_{ss} computed

17

from the entire dataset. As expected, the average $\phi_{ss,s}$ for all stations is nearly equal to ϕ_{ss}. The standard deviation of $\phi_{ss,s}$ ranges between 0.11 and 0.13 for PGA, 0.3s, and 1.0s. These values are reduced when the residuals are corrected for azimuth dependency (as discussed later).

In this section we particularly focus on stations with either high of low φ_{ss} values to better understand why there is such a large variability. We plotted plots of δW_{es}^0 with distance and azimuth for each station. Based on these plots the following observations were made:

- In general, the site-specific within-event residual component (δW_{es}^0) do not show strong dependence on distance. On the other hand, in some cases $\delta W_{esb/es}^0$ from motions recorded from certain distance ranges seem to have strong bias. We found that in most of these cases, these biased motions also share the same azimuth range (i.e., the motions originate from the same source.

- The site-specific within-event residual component (δW_{es}^0) is azimuth dependent for many stations characterized with high $\varphi_{ss,s}$ For example, Figure 5.3.2a shows residuals for station AKTH17; it is clear from the figure that δW_{es}^0 estimated from motions with an azimuth between 150 and 180 degrees are strongly biased towards negative values. These motions also share the same distance range (55-80 km). Hence, they originate from the same seismic region. This seismic region is close to active volcanoes. The magma chambers associated with these active volcanoes might be the cause of such a high bias (e.g., Dawood and Rodriguez-Marek, 2013). Another plausible cause of this bias is that the seismic source have certain characteristics that makes it generate weaker motions compared to the other seismic sources in Japan. These observations show the importance of including the single path and source effects (whenever possible) to improve the predictions obtained from GMPEs.

- For stations with high $\phi_{ss,s}$ values, the azimuth dependency of δW_{es}^0 was removed and $\phi_{ss,s}$ was re-estimated. This is similar (but not exactly the same) to considering the path- and source-specific effects to obtain the fully non-ergodic variability. To remove this dependency, we sorted all motions recorded at each station in eight azimuth regions (i.e., motions with azimuth between 0-45 degrees, 45-90 degrees, etc.). If more than three motions were available in a region, we corrected the residual so as to remove the bias at different azimuth ranges. $\phi_{ss,s}$ is then recalculated for each station using these modified residuals. On average, the modified ϕ_{ss} was 13% less than its original estimated value for motions recorded at ground surface. The reduction ranged from 0% to 50% at the different stations. The 0% represents an extreme case where the station recorded less than 3 motions in a certain azimuth bin and no azimuthal correction was made (station AOMH07 (PGA-Ground surface motion); Figure 5.3.2-b).

- Japan is characterized by a very complex tectonic setting that might result in major biases for some source-path combinations compared to the median predictions (e.g., active volcanoes, faults, large mountains). Probably as a result of this complexity, some stations that recorded many earthquakes show a $\phi_{ss,s}$ that is well above ϕ_{ss} (e.g., station FKSH14; Figure 5.3.2-c). In this case, the recorded motions cover almost all possible azimuth and distance ranges, yet ϕ_{ss} is about 0.66. On the other hand, the residual plots show some clear biases for some azimuth-distance bins that led to such a high $\phi_{ss,s}$.

- In general, stations characterized with low ϕ_{ss} are associated motions with relatively small δW_{es}^0 values with no or limited azimuth dependence. In many cases, the recordings at these stations cover narrow azimuth ranges (e.g., station FKOH03, Figure 5.3.2d), or narrow azimuth and distance ranges (e.g., station HDKH02, Figure 5.3.2e) ranges. This means that the single-path and/or source effects are included in the estimated $\phi_{ss,s}$, which might explain the observed low values at these stations. In this case, any systematic bias due to the path and/or source effects will be transferred to $\delta B2B_{sb}$ and $\delta S2S_s$. This shows that that calculating $\phi_{ss,s}$ using motions recorded from a mainshock and its subsequent aftershocks may result in an under estimation of $\phi_{ss,s}$. If there are other sources surrounding that specific station, then the estimated $\phi_{ss,s}$ might underestimating the real variability at the station.

- Based on the observations summarized in this section, we believe that calculating $\phi_{ss,s}$ using recorded motions should be conducted with care. This particularly true because of the source and path contribution to $\phi_{ss,s}$. Hence, it seems that there is no specific number of motions that could be considered enough to estimate a reliable value for $\phi_{ss,s}$. Instead, the relative location of sources of the recorded motion with respect to the location of the seismic sources around the site of interest.

Based on these observations, we conducted additional analysis to check the effect of azimuth on $\varphi_{ss,s}$ for all stations. To this end we adopted the following procedure: 1) divide the area around each station into eight 45 degree regions; 2) count the number of motions recorded at each station from every region; 3) rotate the axes so that the region that has the maximum number of recordings points towards 0 degrees; 4) normalize the number of records at each of the eight region by the largest number of motions in any of them. These steps result in a single curve (azimuth on the horizontal axis versus "relative number" of motions in each region on the vertical axis) for each station at every spectral period. If this curve plots perfectly horizontally, this means that the station recorded the same number of motions from all directions; while a curve with a peak at a relative azimuth of zero implies that most motions arrive from a narrow range of azimuths. We superimposed these curves from all stations and calculated the average and median of the "relative number" of records from each region (Figure 5.3.3). We distinguished between stations characterized by "High", "Neutral" and "Low" $\phi_{ss,s}$ values. It is clear from the figure that on average "Low" $\phi_{ss,s}$ stations have curves that plot lower than the "Neutral" which in turn has lower curves compared to the "High" stations. This observation is constantly seen over all periods for most azimuth regions. This observation points out that on average stations with "Low" $\phi_{ss,s}$ recorded motions from a narrower azimuthal range compared to the "Neutral" and "High" stations.
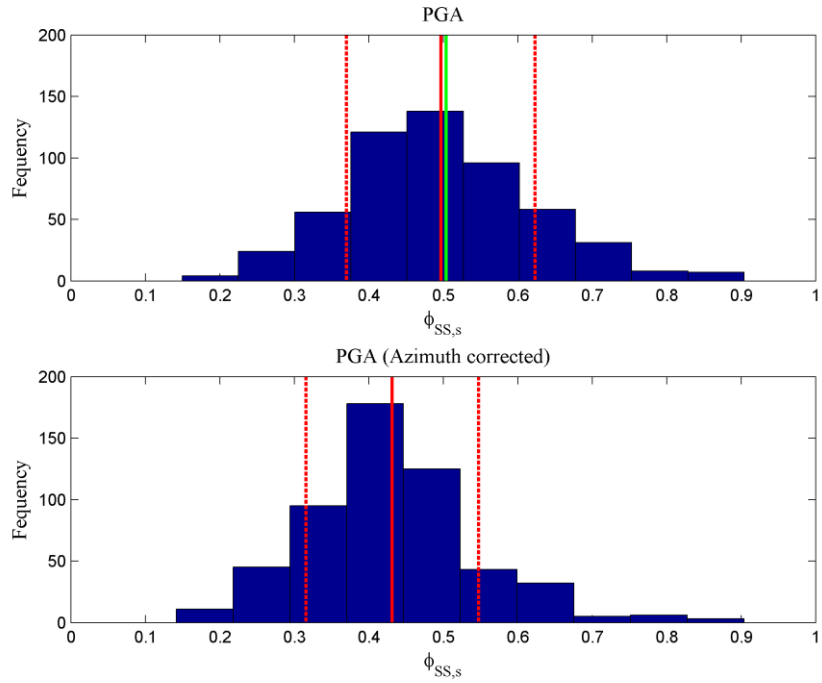
Figure 5.3.1a. Histogram of $\phi_{ss,s}$ values. Top: non-corrected dat. Bottom: residuals are corrected for azimuthal dependency. The red lines show the average and one standard deviation of $\phi_{ss,s}$; the green line is the value of ϕ_{ss}. Data shown for PGA, but similar resutls are observed at other oscillator periods.

[Bore]-V4-ModelVersusPSA-AKTH17-High $\phi_{SS,s}$=0.71413

(a)

(b)
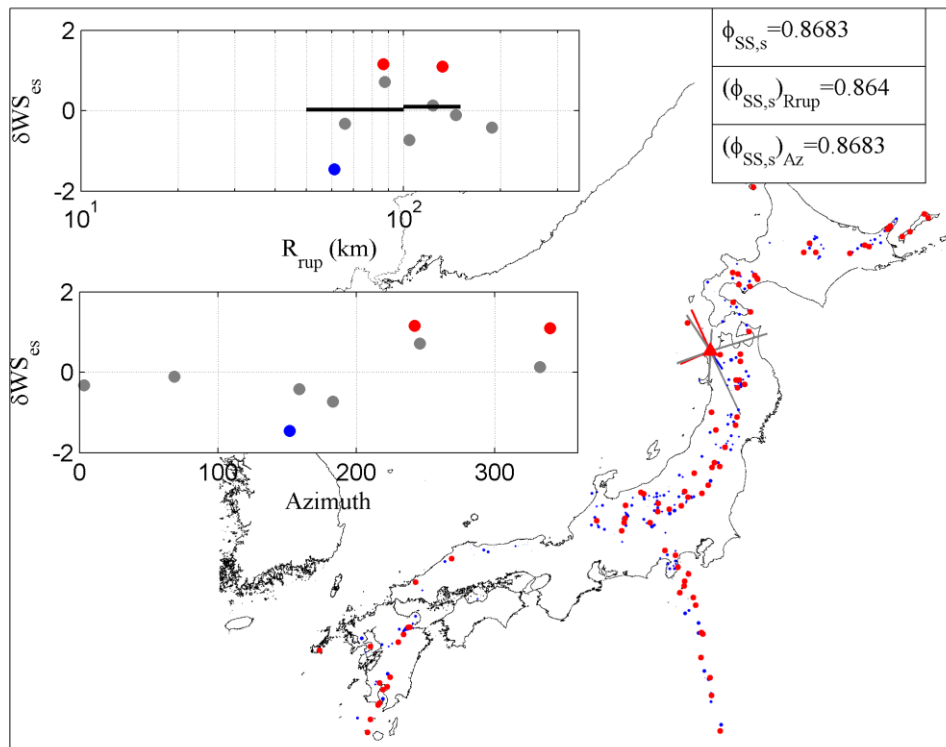
[Bore]-V4-ModelVersusPSA-FKSH14-High $\phi_{SS,s}$=0.65942

(c)

[Bore]-V4-ModelVersusPSA-FKOH03-Low $\phi_{SS,s}$=0.26699

(d)

[Bore]-V4-ModelVersusPSA-HDKH02-Low $\phi_{SS,s}$=0.28558

$\phi_{SS,s}$=0.2856

$(\phi_{SS,s})_{Rrup}$=0.2995

$(\phi_{SS,s})_{Az}$=0.3053
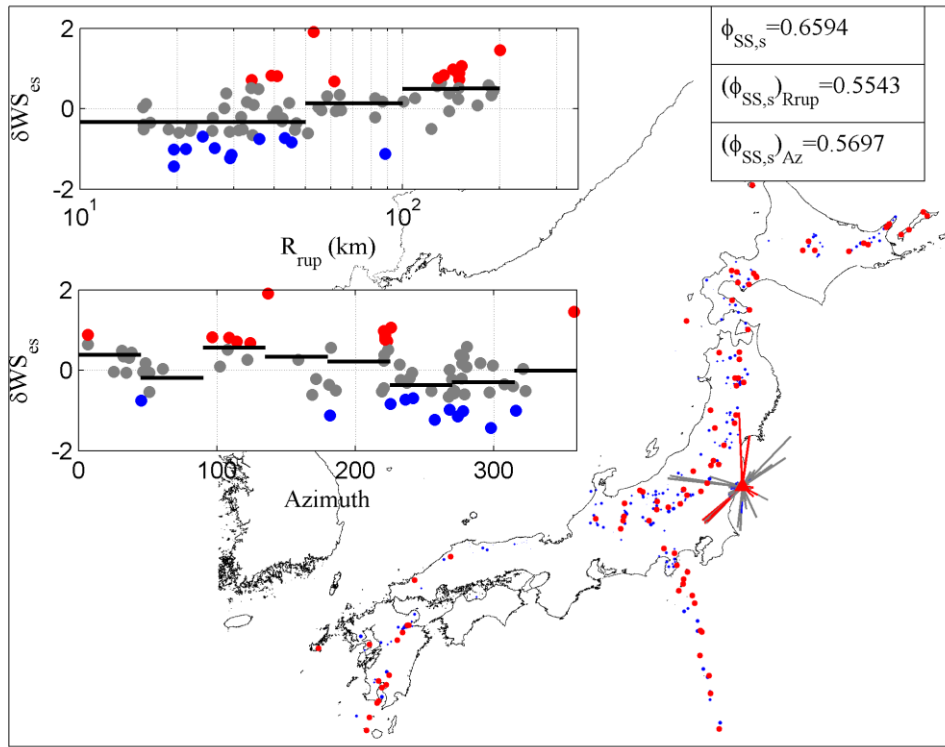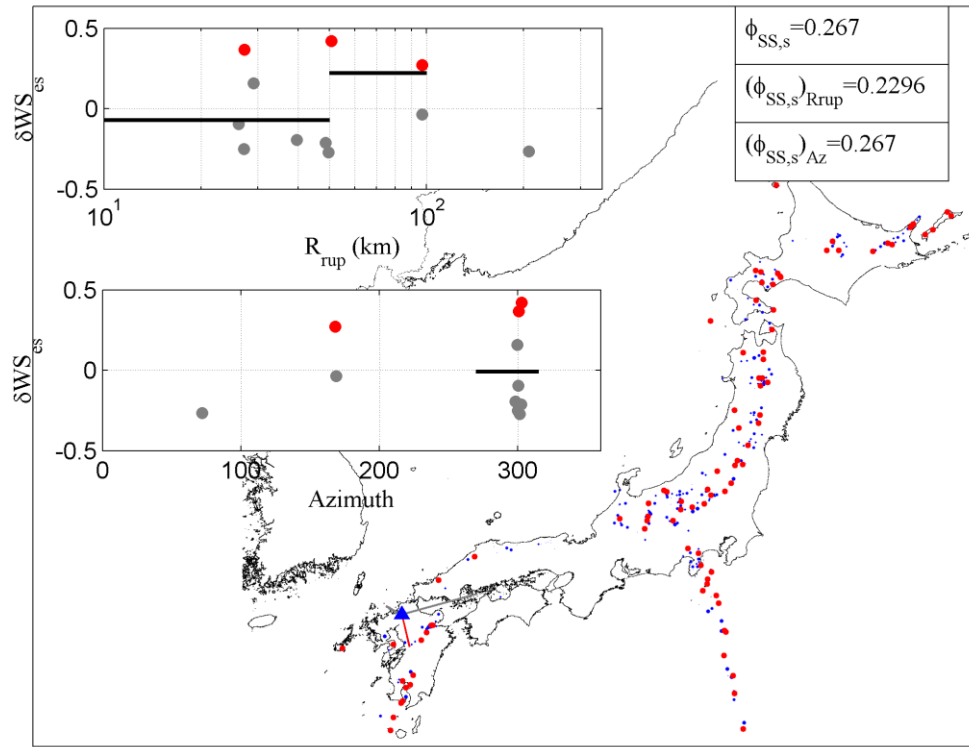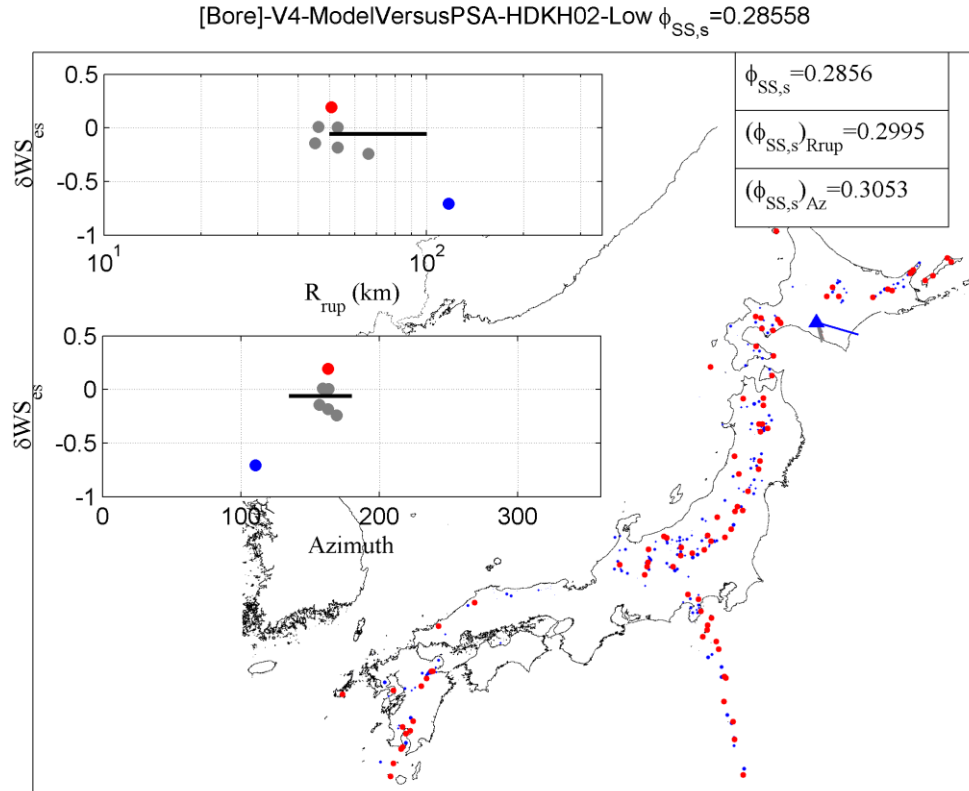
(e)

Figure 5.3.2. **Each panel shows the location of a recording station on the Japanese map. Lines showing the different source to site paths travelled by all motions recorded at that station are added. The two subplots in each panel show the δW_{es}^0 [estimated for a specific spectral period at either surface or borehole] versus R_{rup} and azimuth. The panels show the following stations:**

a) **AKTH17-PGA-Borehole: example of a station where the bias of δW_{esb}^0 versus azimuth is clear. The plots show that records from a specific region is particularly biased;**

b) **AOMH07-PGA-Surface: example of a case where it was not possible to remove the δW_{es}^0 bias with azimuth because very few motions were recorded at each azimuth bin;**

c) **FKSH14-PGA-Borehole: example of a station with many motions from a wide range of azimuth and Rrup, yet it has a high $\varphi_{SS,S}^{G/B}$ due to the complexity of the tectonic environment in Japan;**

d) **FKOH03-PGA-Borehole: example of station that recorded motions from limited azimuths. This could be a potential reason as of why the $\varphi_{SS,S}^{B}$ value at this station is low; and**

e) **HDKH02-PGA-Borehole: example of a station that recorded motions from a limited range of Rrup and azimuths. In this case all motions are from the same source except for only one motion. $\varphi_{SS,S}^{B}$ is believed to be small for this station because it recorded almost all motions from a single seismic source.**

25
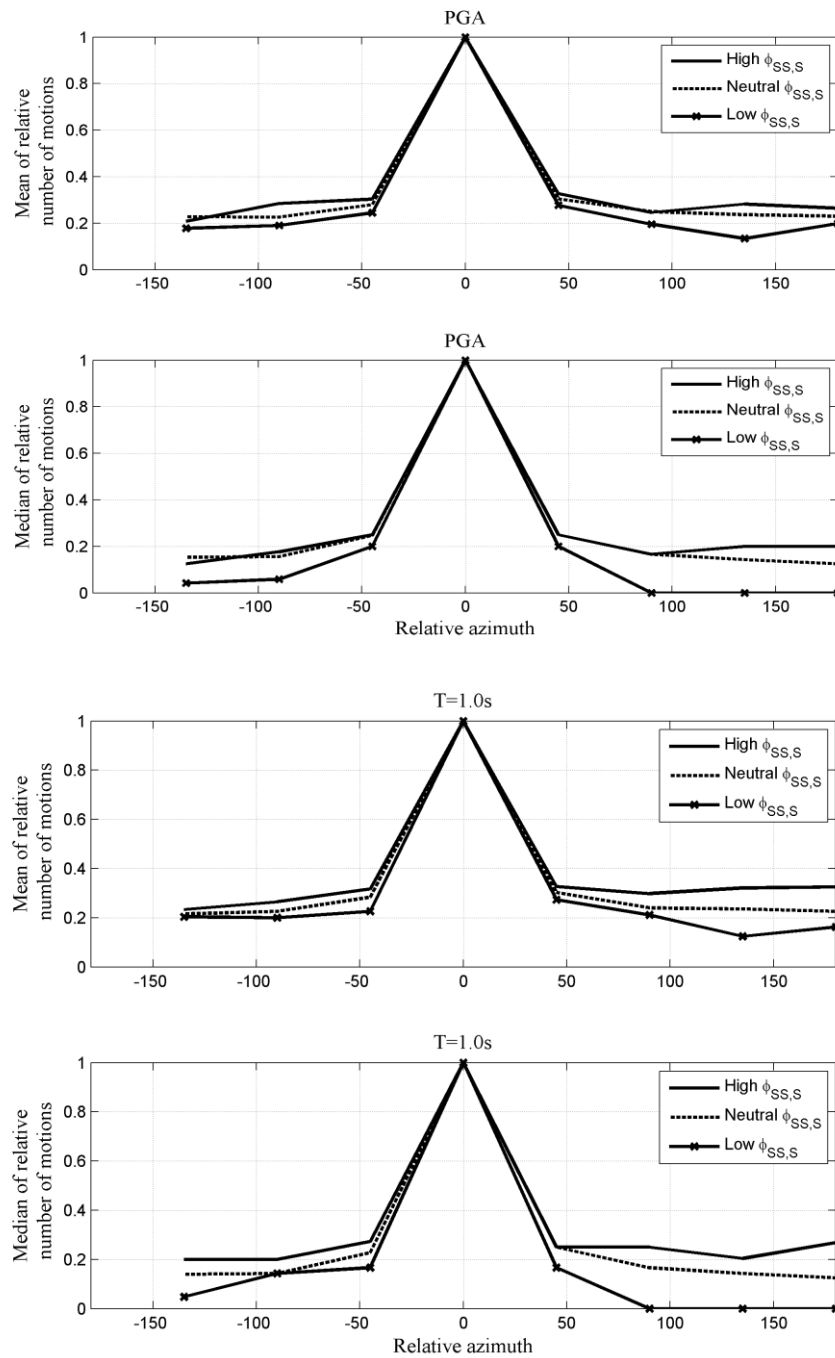
Figure 5.3.2. **Each line represents the relative number of records within a given azimuth range. The azimuth range is measured from the azimuth bin at which there is the largest number of records. Separate lines are plotted as averages of stations with high, neutral, and low values of $\phi_{ss,s}$. For each oscillator period, top is for surface and bottom is for borehole**

5.3.2 Dependence on Topography

One hypothesis that was postulated to explain stations with high $\phi_{ss,s}$ is that those stations are affected by topography. Sites located at topographic accidents will respond differently to waves coming at different incidence angles, hence leading to an event-dependent site amplification, hence to large within-site variability. A comprehensive study of a California strong motion database had been previously conducted by the research team of the PI. These results are summarized in Appendix IV. The most relevant results, computed using the NGA West 2 database, are shown in Figures 5.4 to 5.6. An elevation-based topographic parameterization (see Appendix IV) was used to classify stations as either high, intermediate, or low. The parameter used is simple to compute, but matches very well with curvature. Maufroy et al. (2012) had shown using simulations that curvature is a strong predictor of topographic effects. More details on the parameterization are given in Appendix IV. Figure 5.4 shows that there is a strong bias in site terms for stations that are plotted low, in particular at long oscillator periods. For some periods, these differences are shown to be statistically significant (see Appendix IV). Topographic effects have also a strong impact on station-to-station variability of the site term (quantified by the standard deviation of site terms, of ϕ_{S2S}, Figure 5.5). Stations that are located on topographic accidents, either on ridges (high) or valleys (low) have more variability across stations than those located on flat (intermediate) areas. The most relevant result for this research, however, is the observation that ϕ_{ss} tends to be, on average, higher for stations that classify as low, and to a lesser degree for stations that classify as high (Figure 5.6).

The data compiled in this study was also used to study these effects. However, the same effects are not observed on Japanese data. A larger study is currently underway at VT to address this issue.
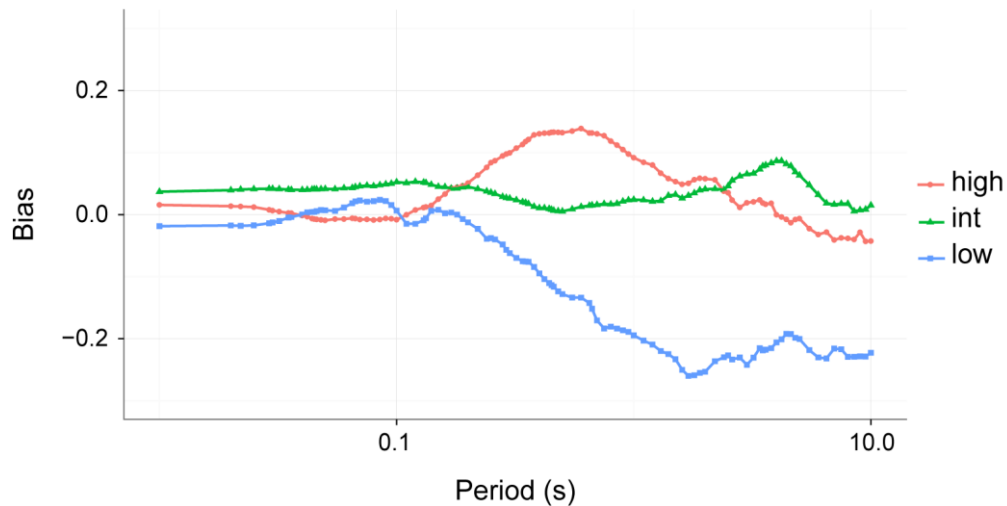
Figure 5.4. Bias or the mean site residual for high, low and intermediate sites are shown for different periods. A scale of 1500 meters and a threshold of 20 meters were used for the classification (see Appendix IV).
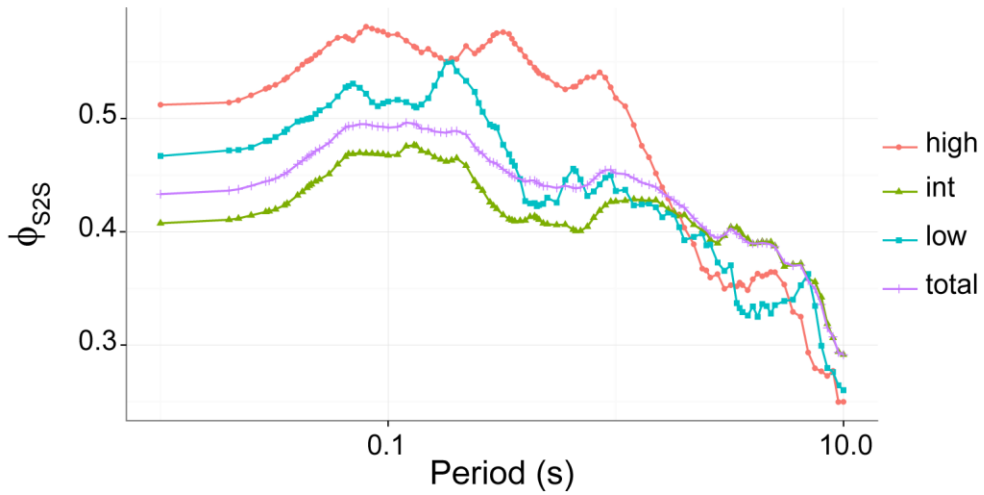
Figure 5.5. ϕ_{s2s} **values are shown for each topographic class at each period. Also shown is the total** ϕ_{s2s} **of the data.**
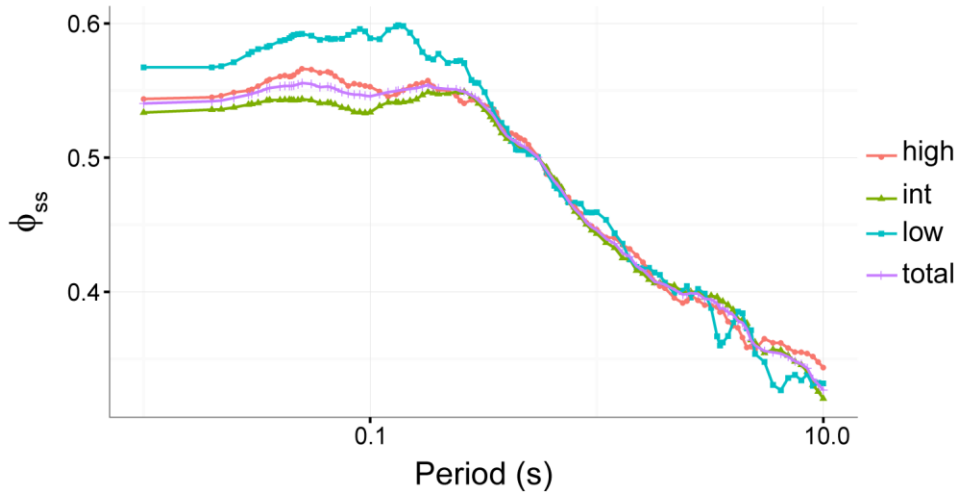
Figure 5.6. ϕ_{ss} **values are shown for each topographic class at each period. Also shown is the total** ϕ_{ss} **of the data.**

5.3.3   Effects of site conditions

Another potential effect on $\phi_{ss,s}$ are site conditions. In principle, the site term accounts for any local site effects; however, some local effects can Figure 5.3.7 shows the scatter plot of the φ_{ss} versus V_{s30}. The scatter plots show that $\varphi_{ss,s}$ tend to decrease for large V_{s30}. It also show substantial scatter at low V_{s30}

values. These trends are expected since hard rock sites are less heterogeneous than softer sites. This heterogeneity results in substantial scatter that inflate $\varphi_{ss,s}$ at softer sites. The same data is plotted (Figure 5.3.8) as a function of the kappa value implicit in the borehole-to-surface transfer function (see Appendix III). The same trends are observed as for Vs30, which is no surprise since kappa and Vs30 are correlated. For longer periods, the correlation with kappa is stronger than that for Vs30.
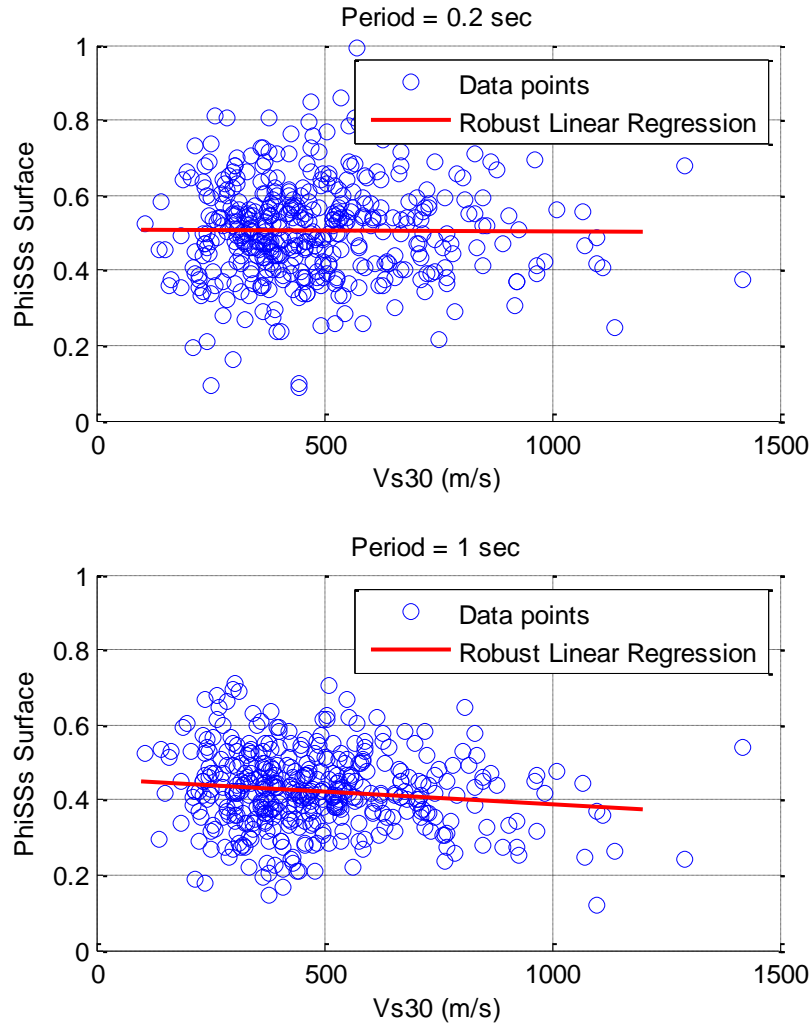
Figure 5.3.7. Plots of $\boldsymbol{\phi_{ss,s}}$ as a function of Vs30.
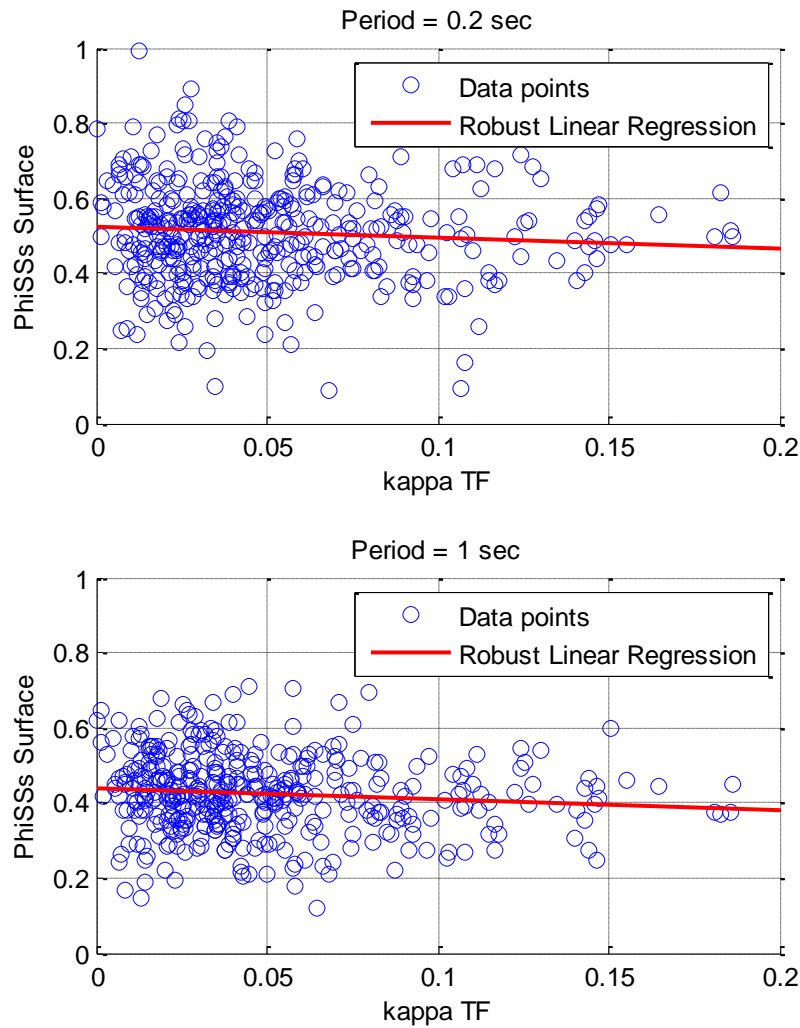
Figure 5.3.8. Plots of $\boldsymbol{\phi}_{ss,s}$ as a function of site the value of kappa implicit in the surface to borehole transfer function.

5.3.4 Use of Horizontal-to-Vertical Ratios to identify stations with high $\boldsymbol{\phi}_{ss,s}$

We consider the hypothesis that stations that have a uniform site response for waves with various incidence angles would have lower values of $\phi_{ss,s}$. To test this, we examined horizontal-to-vertical Fourier spectral ratios (HVSR) averaged for multiple motions across different azimuths. These ratios could be considered a proxy for site effects. We then used various measures of dispersion to quantify the azimuthal variability of these ratios, and plotted these dispersion measures versus $\phi_{ss,s}$. Results for one period are shown in Figure 5.3.9, where the measure of dispersion is the maximum over the average HVSR. The Figure shows results at T=1second, results were similar for other periods. While there is a mild trend of increasing $\phi_{ss,s}$ with increasing dispersion, the trend is not strong and is not statistically significant.
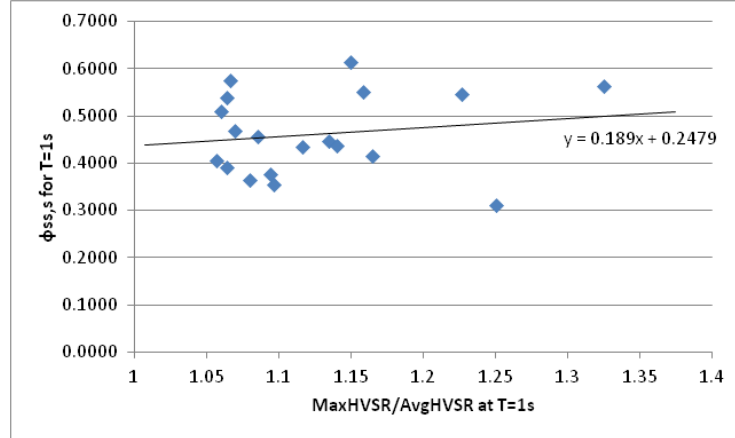
Figure 5.3.9. Values of $\phi_{ss,s}$ for selected stations plotted versus a measure of azimuthal variability.

5.4 Proposed method to account for the single-path effects

The previous discussion has highlighted that the value of $\phi_{ss,s}$ depends on the range source-to-site azimuths that are sampled as a station. This is largely due to path effects; stations that are recording only a narrow azimuthal range are likely recording single-station and single path standard deviations. The anelastic attenuation rate could potentially be different along each source-to-site path. This difference can be attributed to differences in travel path geology or the presence of faults. Surface topography can also have an effect in the attenuation of surface waves. Despite these effects, current GMPEs use an average anelastic attenuation rate independent of the source-to-site travel path of seismic waves. This is generally necessary because very large and well sampled databases are necessary to constrain regional differences in attenuation rates.

The main objective of this component of the study is to present a methodology to account for regional dependences of anelastic attenuation rates within the format used by GMPEs used in engineering applications to reduce the scatter in predicting ground-motion parameters. The constraint on regional dependent attenuation allows for constraining path terms for nonergodic PSHA. Several studies had been conducted to quantify path effects to reduce the scatter that result from the different travel paths of the waves through the Earth's crust. These studies covered a wide range of regions around the world (e.g., Central China, Colombia, France, United Kingdom, and Middle East). According to Phillips (1999), path terms are generally constrained either by investigating the correlations with path characteristics (e.g., topography, crustal depth, among others) or by removing the known effects from the records (e.g., event characteristics, site characteristics, and distance), then the path effects are constrained from the residuals. The approach we adopt in this study falls within the second category as we remove the effects of site response, geometric attenuation, and source, while at the same time constraining path effects using a mixed-effects analysis. An important characteristic of the approach we adopt in this study is that the analysis output has the form of GMPEs used for engineering PSHA studies.

5.4.1 Methodology

The methodology consists of dividing the region that contains all source-to-site paths of all records into equal squares (from this point on, these squares will be referred to as elements). The travel distance through each of the elements is calculated for each recording assuming that the path of the waves from the source's epicenter to the site is a straight line. The attenuation rate of each element becomes a term in the GMPE functional form and is regressed for its value. The elements to be used in this study are set to 25 km squares. The element size has to be small enough such that particularities in the attenuation rate can be captured, but large enough so that a significant amount of source-to-site paths pass through the elements. The data was divided into three magnitude bins (i.e., $4 \leq M_{JMA} < 5$; $5 \leq M_{JMA} < 6$; and $6 \leq M_{JMA} < 7$). A linear mixed-effects regression analysis was conducted on each magnitude bin separately.

The geometric mean of the pseudo-spectral accelerations (5% damping) of the as-recorded components of the motions recorded by the downhole instruments are the GM parameters used in this study. The choice of downhole records over the surface records is to reduce the effects of shallow site response on the ground motions. The general form of the GMPE model is given by

$$\ln(y_{es}) = \ln(\mu_{es}) + \delta B2B_{sb} + \delta B_e + \delta W^0_{esbl} \qquad \text{Equation 5.4.1}$$

where y_{es} is the ground motion parameter for the recorded motions in units of gravity (given as a ratio of 9.81 cm/s^2) for event e at site s; μ_{es} is the median prediction; and $\delta B2B_{sb}$, δB_e, and δW^0_{esbl} are defined in Equations 2.2 and 2.6. The functional form of the median prediction is given by:

$$\ln(\mu_{es}) = C_1 + C_2 \ln(R_{epi}) + C_3 D + C_4 \ln(V_{s30}/V_{ref}) + \sum_{i=1}^{n} \delta_i P_{ies} \qquad \text{Equation 5.4.2}$$

where R_{epi} is the epicentral distance in km; D is the focal depth of the earthquake in km; V_{s30} is the average shear wave velocity over the top 30 meters at the recording site; V_{ref} is a reference shear wave velocity taken here as 760 m/s; P_{ies} is the distance (in km) through element i for a straight line path originating at the epicenter of source e to site s; δ_i is the attenuation per kilometer of travel in the ith element; and n is the total number of elements in which the region under study is divided in. The term $C_2 ln(R)$ represents geometric spreading. For simplicity geometrical spreading is assumed to follow a theoretical model and the coefficient C_2 is assumed to be -1 and is not regressed for.

5.4.2 Validation

We used 117 aftershocks that followed the magnitude 9.0 Tohoku earthquake in 2011 to validate the methodology. The hypothesis that the attenuation rate through elements that contain volcanoes is statistically higher than elements that doesn't contain volcanoes is tested to validate the proposed methodology. This is because in Japan and other subduction regions the presence of active volcanoes with magma chambers can result in significantly larger attenuation rates through volcanic regions compared to other regions (Cousins et al. 1999, Ojeda and Ottemoller 2002, McVerry et al. 2006, and Zhao 2010).

5.4.3 Results

The attenuation rate of the elements with volcanoes was found to be higher than the elements without volcanoes at a 95% confidence level (i.e., the null hypothesis that volcanic and non-volcanic elements

have the same average attenuation rate is rejected), for most oscillator periods. At the same confidence level, the attenuation in the forearc and back-arc regions was found to be similar for most periods. Since these conclusions are confirmed by previous studies conducted in subduction zone earthquakes, the methodology to estimate path specific attenuation rates is considered valid.

The single-station within-event standard deviations obtained in the present study are consistently and significantly lower than the values estimated by Rodriguez-Marek et al. (2011) except for the first magnitude bin at longer periods. This represents the reduction from single-station to single-station-single-path standard deviations (Figure 5.4.1). For additional information regarding the methodology and the results the reader is referred to Dawood and Rodriguez-Marek (2013, also available as Appendix V)
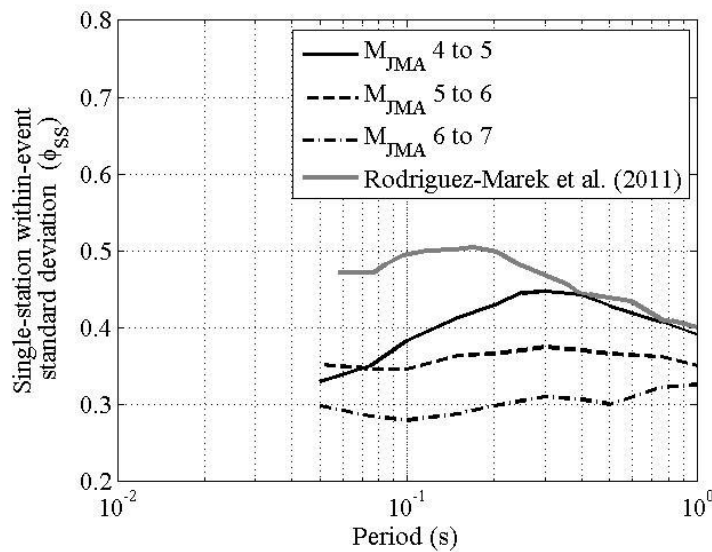
Figure 5.4.1. Single-station within-event standard deviation that accounts for path effects obtained in this study for three magnitude bins compared to the φ_{SS} values from Rodriguez-Marek et al. (2011)

6 Conclusions

This report has presented the results of the research conducted at Virginia Tech as part of the SIGMA project. The most salient contributions of this research include a database of uniformly processed response spectral accelerations for over 100000 records from the KiKnet array. These data was used to evaluate values of single-station sigma. The computed values are slightly higher than values computed in previous studies. Probably the most interesting observation is that magnitude dependence is not evident in the single-station standard deviation, although this observation is tempered by the lack of recordings at magnitudes above 7.0. The site-to-site variability of single-station phi was evaluated. The largest factor contributing to this variability is the poor sampling across various source-to-site azimuths, which implies that stations that record low values of $\phi_{ss,s}$ are likely sampling single-paths, thus these values have to be used with caution. Finally, a model to constrained path effects is presented and evaluated using a subset of the KiKnet database. The advantage of this approach (as opposed to the approach of Lin et al., 2013) is that the proposed approach allows for the estimate of "path terms" that can be used in forward prediction.

7 Acknowledgments

8 References

Abrahamson NA, WJ Silva, and R Kamai. 2014. Update of the AS08 Ground-Motion Prediction Equations Based on the NGA-West2 Data Set. *Earthquake Spectra* (In Press). DOI: http://dx.doi.org/10.1193/070913EQS198M

Al Atik, L, A Kottke, N Abrahamson, and J Hollenback. 2013. Kappa (κ) Scaling of Ground-Motion Prediction Equations Using an Inverse Random Vibration Theory Approach. *Bull. Seism. Soc. Am.* 104(1), 336-346.

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

Allen TI, DJ Wald, AJ Hotovec, K Lin, PS Earle, and KD Marano. 2008. An Atlas of ShakeMaps for Selected Global Earthquakes, U.S. Geological Survey Open-File Report 2008-1236, U.S. Geological Survey, Reston, Virginia.

Ancheta TD, RB Darragh JP Stewart, E Seyhan, WJ Silva, BSJ Chiou, KE Wooddell, RW Graves, AR Kottke, DM Boore, T Kishida, and JL Donahue. 2013. *PEER NGA-West2 Database*, PEER Report No 2013/03, PEER Center, University of California, Berkeley.

Anderson, J. G. and J. N. Brune (1999). Probabilistic seismic hazard assessment without the ergodic assumption, *Seism. Res. Lett*. **70**, 19-28.

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

Atkinson, G. M. (2006). Single-station sigma, *Bull. Seismol. Soc. Am*. **96**, 446-455.

Bommer, J.J., Coppersmith, K.J., Coppersmith, R.T., Hanson, K.L., Mangongolo, A., Neveling, J., Rathje, E.M., Rodriguez-Marek, A., Scherbaum, F., Shelembe, R., Stafford, P., and Strasser, F.O. (2014). "A SSHAC Level 3 probabilistic seismic hazard analysis for a new-build nuclear site in South Africa." *Earthquake Spectra*, in press (doi: http://dx.doi.org/10.1193/060913EQS145M)

Chen L and E Faccioli. 2013. Single-station standard deviation analysis of 2010–2012 strong-motion data from the Canterbury region, New Zealand. *Bulletin of Earthquake Engineering* 11:1617−1632. DOI: 10.1007/s10518-013-9454-3.

Chen Y-H and C-CP Tsai. 2002. A new method for estimation of the attenuation relationship with variance components. *Bulletin of the Seismological Society of America* 92:1984–1991.

Chiou BS-J and RR Youngs. 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 sec*. PEER Report, Pacific Earthquake Engineering Center, Berkeley, California.

Cousins,W. J., J. X. Zhao, and N. D. Perrin (1999). A model for the attenuation of peak ground acceleration in New Zealand earthquakes based on seismograph and accelerograph data, Bull. New Zeal. Soc. Earthquake Eng. **32,** 193–220.

Dawood HM. 2014. Partitioning uncertainty for non-ergodic probabilistic seismic hazard analyses. PhD dissertation, Virginia Polytechnic Institute and State University, USA.

Dawood, H. and Rodriguez-Marek. A. (2013). "A New Approach to Include Path Effects in Ground Motion Prediction Equations Using the Mw 9.0 Tohoku Earthquake Aftershocks." *BSSA* 103, 1360-1732, DOI: 10.1785/0120120125.

Dawood, H. M., and A., Rodriguez-Marek (2013). A Method for Including Path Effects in Ground-Motion Prediction Equations: An Example Using the magnitude 9.0 Tohoku Earthquake Aftershocks, Bull. Seism. Soc. Am. 103 1360-1372.

Garcia D, DJ Wald, and MG Hearne. 2012. A global earthquake discrimination scheme to optimize ground-motion prediction equation selection. *Bulletin of the Seismological Society of America* 102:185−203.

Gardner JK and L Knopoff. 1974. Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bulletin of the Seismological Society of America*. 64(5)1363–1367.

Hayes G. 2011. Finite Fault Model. Updated Result of the Mar 11, 2011 Mw 9.0 Earthquake Offshore Honshu, Japan. Accessed at http://earthquake.usgs.gov/earthquakes/eqinthenews/2011/usc0001xgp/finite_fault.php.

Lin, P.-S., B. S.-J. Chiou, N. A. Abrahamson, M. Walling, C.-T. Lee, and C.-T. Cheng (2011). Repeatable source, site, and path effects on the standard deviation for empirical ground-motion prediction models, *Bull. Seismol. Soc. Am*. **101**, 2281-2295.

Luzi L, D Bindi, R Puglia, F Pacor, and A Oth. 2014. Single station sigma for Italian strong motion stations. *Bulletin of the Seismological Society of America* 104(1):467−483. DOI: 10.1785/0120130089.

Maufroy, E. Cruz-Atienza, VM, Gaffet, S. (2012). A Robust Method for Assessing 3-D Topographic Site Effects: A Case Study at the LSBB Underground Laboratory, France. *Earthquake Spectra* 28(3):1097-1115.

McVerry, G. H., J. X. Zhao, N. A. Abrahamson, and G. H. Somerville (2006). Crustal and subduction zone attenuation relations for New Zealand earthquakes, Bull. New Zeal. Soc. Earthquake Eng. **39,** 1–58.

Morikawa N, T Kanno, A Narita, H Fujiwara, T Okumura, Y Fukushima, and A Guerpinar. 2008. Strong motion uncertainty determined from observed records by dense network in Japan. *Journal of Seismology* 12:529−546.

Ojeda, A., and L. Ottemoller (2002). Q_{Lg} tomography in Colombia, Phys. Earth Planet In. **130,** 253-270.

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

Ornthammarath T, J Douglas, R Sigbjörnsson, and CG Lai. 2011. Assessment of ground motion variability and its effects on seismic hazard analysis: a case study for Iceland. *Bulletin of Earthquake Engineering* 9:931−953.

Oth A, S Parolai, and D Bindi. 2011. Spectral analysis of K-NET and KiK-net data in Japan, Part I: Database compilation and peculiarities. *Bulletin of the Seismological Society of America* 101:652–666.

Phillips, W.S. (1999). Empirical Path Corrections for Regional-phase Amplitudes, Bull. Seismol. Soc. Am. **89**, 384–393.

Reasenberg P. 1985. Second-order moment of central California seismicity, 1969-82. *Journal of Geophysical Research* 90:5479−5495.

Renault, P., S. Heuberger and N. A. Abrahamson (2010). PEGASOS Refinement Project: An improved PSHA for Swiss nuclear power plants. *Proceedings of the 14th European Conference of Earthquake Engineering*, Ohrid, Republic of Macedonia, Paper ID 991.

Rodriguez-Marek, A., Cotton, F., Abrahamson, N., Akkar, S., Al Atik, L., Edwards, B., Montalva, G.A., and Dawood, H.M. (2013). "A model for single-station standard deviation using data from various tectonic regions." *BSSA* 103, 3149-3163.

Rodriguez-Marek, A., Rathje, E.M., Bommer, J.J., Scherbaum, F., and Stafford, P.J. (2014). "Application of single-station sigma and site response characterization in a probabilistic seismic hazard analysis for a new nuclear site." Accepted for Publication, *BSSA* 104(4).

Rodriguez-Marek. A., G. A. Montalva, F. Cotton, and F. Bonilla (2011). Analysis of Single-Station Standard Deviation Using the KiK-net Data, Bull. Seismol. Soc. Am. **101,** 1242–1258.

Van Houtte, C., Drouet, S. & Cotton, F. (2011). Analysis of the origins of κ (kappa) to compute hard rock to rock adjustment factors for GMPEs. *Bulletin of the Seismological Society of America* 101(6), 2926-2941.

Zhao, J. X. (2010). Geometric Spreading Functions and Modeling of Volcanic Zones for Strong-Motion Attenuation Models Derived from Records in Japan, Bull. Seismol. Soc. Am. **100,** 712–732.

Appendix I

PROCESSING THE KIK-NET STRONG GROUND MOTION DATABASE AND
COMPILATION OF METADATA FOR GMPE DEVELOPMENT

By

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

[Submission prepared for Earthquake Spectra]

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

Appendix II

AN EMPIRICAL GROUND MOTION PREDICTION EQUATION FOR ACTIVE CRUSTAL EARTHQUAKES USING THE JAPANESE KIK-NET DATABASE (ERGODIC AND SITE-SPECIFIC FORMULATIONS)

By

Haitham M. Dawood and Adrian Rodriguez-Marek[a]

[Submission prepared for Earthquake Spectra]

a) The Charles E. Via, Jr. Department of civil and environmental engineering, Virginia Tech, 200 Patton Hall, Blacksburg, VA 24061

Appendix III

COMPUTATION OF KAPPA FOR THE KIKNET STATIONS
INTERNAL REPORT, VIRGINIA TECH

By

Ashly M. Cabas and A. Rodriguez-Marek

Appendix IV

SUMMARY OF RESEARCH CONDUCTED AT VIRGINIA TECH ON THE
PARAMETERIZATION OF TOPOGRAPHIC EFFECTS

Internal Report, Virginia Tech

By

Manisha Rai and A. Rodriguez-Marek

Appendix V

A METHOD FOR INCLUDING PATH EFFECTS IN GROUND-MOTION PREDICTION EQUATIONS: AN EXAMPLE USING THE Mw 9.0 TOHOKU EARTHQUAKE AFTERSHOCKS

By

Haitham M. Dawood and Adrian Rodriguez-Marek[a]

a) The Charles E. Via, Jr. Department of civil and environmental engineering, Virginia Tech, 200 Patton Hall, Blacksburg, VA 24061

PROCESSING THE KIK-NET STRONG GROUND MOTION DATABASE AND COMPILATION OF METADATA FOR GMPE DEVELOPMENT

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

The Kiban-Kyoshin network (KiK-net) database is an important resource for ground motion (GM) studies. Processing these records, including baseline correction and filtering to determine bandwidth with appropriate signal to noise ratio, is necessary to use these records 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 corner frequencies for band-pass filtering. 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 given.

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. In particular, the data needed for the development of ground motion prediction equations (GMPEs) for ground motion parameters of engineering interest must include a comprehensive compilation of event and site metadata. The collection of these data, also including 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 first Next Generation Attenuation (NGA W1) project; Akkar et al. 2010 for Turkey; Arango et al. 2011-a for Central American subduction zone; Arango et al. 2011-b for Peru-Chile subduction zone;

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

Pacor et al. 2011 for the ITACA database in Italy; and most recently Akkar et al. 2014 for the RESORCE database in Europe). Table 1 summarizes some of the characteristics of the databases used in each of these studies. 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 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 adjust the raw records such that the final velocity of the record is zero and the final displacement is compatible with tectonic displacements (e.g., baseline correction). Depending on the application and on the shape of the instrument's response, the instrument response must also 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 78840 and 34456 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 [in review]), 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.

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. 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 following the devastating Kobe earthquake (January 17, 1995) to better monitor the seismic activity around the country (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. Nearly all the borehole instruments are located at depths ranging between 100 and 250 meters. The borehole instrument is deeper than 500 meters at only 31 stations. 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).

KIK-NET RECORDS

The un-processed strong records recorded at 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 M_w 9.0 Tohoku earthquake, which was associated with a large number of foreshocks and aftershocks. Table 2 shows the number of records for which the geometric mean of the PGA (two horizontal components) exceeds different acceleration thresholds. Note that the majority of the records have recorded 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

4

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 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 $M_w>=4$). The figures show 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.

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 follow a heuristics to select the 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.   <u>*Baseline correction*</u>

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

5

noise window using an automated algorithm to detect the first arrival. The first arrival is defined as the first 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):

Length of zero pad (sec) $= 1.5\ N/f_c$ (1)

Where, N is the order of the Butterworth filter, and f_c is the high-pass corner frequency (Hz)

Step IV.   *Filtering*

A high-pass, acausal, 4th 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 band 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 criteria listed below:

a. The final displacement in the displacement time series (obtained using 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.

b. The final velocity in the velocity time series must be less than 0.001 and 0.005 cm/sec for records from events with $M_{JMA} <7.0$ and $M_{JMA} \geq 7.0$, respectively.

The threshold values of 0.001 cm and 0.001 cm/s for criteria "a" and "b", respectively, were found to be excessively strict for large magnitude earthquakes. 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 criteria "a" and "b" 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 in using automatically processed motions recorded close to the fault.

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

d. A linear regression is applied to the trailing portion of the displacement 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/sec.

e. A linear regression is applied to the trailing portion of the velocity time series. The slope of the fitted line must be less than 0.001 cm/sec^2.

f. 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 through only 5 points in the smoothed FAS which still show some irregularities, unlike the theoretical smooth FAS that is linear in the low frequency range. The slope check was skipped for records from earthquakes with $M_{JMA} \geq 6.0$ because the corner frequency (f_c) for large earthquakes is expected to be outside the range of frequencies within which we search for f_c.

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 the code returns to Step IV. If all criteria are fulfilled for all components, the code proceeds to the next step. 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 $2/f_c$ second of the record. The FAS for the whole record component (obtained from step I) and the noise window are calculated, smoothed, and the ratio between both is calculated (from this point on, this ratio will be denoted as 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. In this study we chose to set the maximum usable spectral period to be 50% of the inverse of the corner frequency (the same criteria was used by Abrahamson and Silva 1997 and Spudich et al. 1999). 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.

Step VIII. *Saving acceleration time series and figures*

The acceleration time series of the filtered components are saved. The zero pads applied during processing were not removed before saving the acceleration time series. Stripping off the zero pads might compromise the filtered record especially at longer periods (Boore 2005, Boore et al. 2012).

Step IX. *Additional checks*

An additional check that was conducted within this study was to automatically flag the records that possibly contain multiple wave trains or sub-events (see Dawood 2014 for additional information).

All the instruments in the KiK-net network are broadband digital 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-W2 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-W2 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 dataset.

PART III: METADATA

This section presents the different information collected for each record. This information is known as the metadata. The collected metadata includes the information related to each earthquake, differentiating between dependent and independent earthquakes, the class of each

earthquake (e.g., shallow active crustal, interface subduction, etc), the different source-to-site distance measurements, 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). KiK-net provides a subset of these parameters (M_{JMA}, hypocenter location, and time of occurrence) 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 datasets to match records in the KiK-net database to the corresponding earthquake in the F-net catalog. According to the F-net website, the inversion is automatically conducted by a full-wave inversion code using about 70 different broadband waveforms (Fukuyama et al. 1998, Fukuyama and Dreger 2000). The 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 datafiles. 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 3). 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 4943 earthquakes with $M_{JMA} \geq 4.0$ were identified in the KiK-net database. Table 4 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 date, time and location of the event. The 4534 earthquakes (categories A to D) from the KiK-net database mapped into 3210 earthquakes in the F-net seismic catalog (2776 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 October, 11 1997 at 14:44 (JST), was reported inside 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 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. 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 (mainshocks) 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-W1 database that were recorded from aftershocks were found to behave differently than the records from mainshocks. For this reason, the records from aftershocks were either totally excluded while 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 2008-a). 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 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 Stated 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 5. None of these analyses resulted in a Poissonian catalog. Attempts to change the input parameters of the Gardner and Knopff's algorithm to enforce obtaining a Poissonian declustered catalog resulted in nonphysical (much lower than 1) Gutenberg-Richter b-value 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 5) 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 classification dividing earthquakes into their tectonic environment such as: *active crustal* (shallow and deep); *subduction zone* (SZ) (interface and intraslab); and *stable continental region* earthquakes is widely used in the development of GMPEs. Given the availability of enough records, 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 6 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 showed a considerable improvement in the number of correctly classified events using this algorithm in comparison to the Allen et al. (2008) algorithm. However, for intraslab events the number of misclassified events 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 7. 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. Allen et al. (2008) do not provide a classification for events that fall within these depth-magnitude combinations. The number of events classified as interface events using both algorithms are very different. This is consistent with the Garcia et al. (2012) finding 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, 1616, 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. (2011). The faulting style is determined using the P-, B- and T- axes plunges of each earthquake. These values were calculated from fault plane (dip, strike and rake) provided by the F-net catalog using a program named FOCMEC (Snoke 2003).

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 8 provides a list of these earthquakes. For the records from the earthquakes listed in Table 8, 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 8 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 8 shows the distribution of M_w versus the different distance measures for the events listed in Table 8.

The finite fault models of the following three 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.

For events for which a finite fault model is currently unavailable we simulate the rupture parameters following the method of Chiou and Youngs (2008-b, Appendix B). 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.

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) classifications to determine the appropriate magnitude-area relationship. For the events in Table 8, 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 9 compares the M_w-area relationships for the events in Table 8 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 magnitude.

The second step for simulating the faults is to estimate the aspect ratio (AR). Chiou and Youngs (2008-b) developed equations for AR as a function of M_w from the NGA-W1 database.

Figure 10 (a) compares this relationship to the finite fault models in our database. The finite fault models in our database are sufficiently similar to the previously published equations and so we use the Chiou and Youngs (2008-b) 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 10 (b) for reverse events:

$$\log(\text{AR}) = 0.0321(M_w - 4)^{1.33}, \sigma_{\log} = 0.35. \tag{3}$$

The third step is to simulate the location of the hypocenter on the fault plane. To do this we 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 distributions are given in Figure 11. 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.

To assess the accuracy of the simulated distances, Figure 12 plots the base 10 logarithmic distance residual (using the above rules for determining a single value) versus the hypocentral distance for the events in Table 8 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. So 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.

RECORDING STATIONS

We calculated several parameters that characterize the soil profile at 655 KiK-net stations. Table 9 lists these parameters along with a brief description. We also included the fundamental frequency (f_0) for 537 stations calculated by Cadet (2007). For details about the approach used to calculate f_0 the reader is referred to Cadet (2007). The histograms that show the NEHRP site class (BSSC 2001), $V_{S,30}$. and the depth of the borehole instrument at all stations are shown in Figure 13. Most of the stations fall into the NEHRP site classes C [58%] and D [28%]. The median, mode and mean of the $V_{S,30}$ of all stations falls in the range between 445 and 500 m/s. The depth of the borehole instruments vary between 99 and 2008 meters.

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 motion for events between October 1997 and December 2011. In addition, 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 band-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.

We also declustered the F-net seismic catalog to differentiate between dependent and independent events for the purpose of GMPE development. Declustering seismic catalogs is associated with subjective choices that have a strong impact on the results. When applied to the F-net catalog, the Gardner and Knopoff (1974) and the Reasenberg (1985) algorithms did not result in a Poissonian catalog. On the other hand, when a Poissonian catalog was obtained by changing the window parameters in the Gardner and Knopoff (1974) algorithm, the number of independent earthquakes (mainshocks) was considerably low and the "b" value for the Guttenberg-Richter recurrence relationship was unrealistically low (around 0.5). Hence, care should be taken when using the output of the declustering algorithms.

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.

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 (2014). 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

17

from the fault plane (dip, strike and rake) to the stress axes (lower hemisphere trend a plunge of T, P, and B).

ACKNOWLEDGMENTS

REFERENCES

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

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

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., Sisi, A.Azari, 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 ShakeMaps for Selected Global Earthquakes, U.S. Geol. Surv. Open-File Rept. 2008-1236.

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.

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

Arango, M. C., Strasser, F. O., Bommer, J. J., Hernández, D. A., and Cepeda, J. M., 2011-a. 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., 2011-b. A strong-motion database from the Peru–Chile subduction zone, J. Seismol. 15 19-41.

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.

Bird, P., 2003. An updated digital model of plate boundaries, G^3 34 1027

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).

Cadet, H., 2007. Utilisation combine des methods basees sur le bruit de fond dans le cadre du microzonage sismique. PhD thesis, Universite Joseph Fourier, France.

Campbell K. W., and Bozorgnia, Y., 2008. NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s, Earthquake Spectra. 24 139–171.

Chiou, B. S.-J., and Youngs, R. R., 2008-a. 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., 2008-b. 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 Center, Berkeley, California.

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

Converse, A.M., 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 accelero-metric data and point-sources stochastic simulations,Bull. Seismol. Soc. Am.98,no. 5, 2244–2261.

Dawood, H. M., 2014. Partitioning uncertainty for non-ergodic probabilistic seismic hazard analyses. PhD dissertation, Virginia Polytechnic Institute and State University, USA.

Dawood, H. M., Rodriguez-Marek, A., Bayless, J., Goulet, C., and Thompson, E., 2014. Processing the KiK-net strong ground motion database and compilation of metadata for GMPE development, 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 predisction models with M^2 terms, Bull. Seism. Soc. Am. 86 329-336.

Fukuyama, E., Ishida, M., Dreger, D., 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.

Goulet, C.A and Bayless, J., 2014. Evaluation of partitioned standard deviation terms for the integration of site effects into seismic hazard analyses, Bull. Seism. Soc. Am. (in review)

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.

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.

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.

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

Kobayashi, T., Tobita, M., Koarai, M., Okatani, T., Suzuki, A., Noguchi, Y., Yamanaka, M., and Miyahara, B., 2013. InSAR-derived crustal deformation and fault models of normal faulting earthquake (Mj 7.0) in the Fukushima-Hamadori area, Earth Planets Space 64, 1209-1221.

Kanno, T., Narita, A., Morikawa, N., Fujiwara, H., 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 (5866), 1076-1079.

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

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.

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 cariations 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.

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

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 September 30, 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.

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.

Pousse, G. (2005). *Analyse des données accélérométriques de K-net et KiK-net implications puor 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,

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.

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

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.

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 (Editors), Academic Press, San Diego, Chapter 85.12.

Spudich P, Joyner, W. B., Lindh, A. G., Boore, D. M., Margaris, B. M., Fletcher, J. B., 1999. SEA99: a revised ground motion prediction relation for use in extensional tectonic regimes. Bull Seismol Soc Am 89 1156–70.

Strasser, F. O., Arango, M. C., and Bommer, J. J., 2012. 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 intraslab 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.

Uhrhammer, R., 1986, Characteristics of Northern and Central California seismicity, Earthquake Notes, 57 , 21.9

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

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.

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.

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.

TABLES

Table 1. Examples of databases compiled and processed for different parts of the world

Region / Database name	NGA-W1	NGA-W2	Turkey	Central American subduction zone	Peru-Chile subduction zone	ITACA	RESORCE	Japan (KiK-net & K-net)
Number of GM records	3551	21539	4607	554	98	2008	5882	78840
Number of seismic events	173	600	2996	80	15	562	1814	2201
M_w range	4.2 - 7.9	2.99-7.9	-	5.0 - 7.7	6.3-8.4	3.0-6.9	2.8 - 7.8	-
Temporal window	1935-2003	1935-2011	1976-2007	1976-2006	1966-2007	1972-2009	1967-2012	K-NET 05/1996-02/2009 KiK-net 10/1997-02/2009
Number of stations	1456	4160	-	124	59	399	1540	1037 K-NET and 688 KiK-net
Reference	Chiou et al. 2008	--	Akkar et al. (2010)	Arango et al. (2011-a)	Arango et al. (2011-b)	Pacor et al. (2011)	Akkar et al. (2014)	Oth et al. (2011)
Notes	-Shallow active crustal earthquakes	-Shallow active crustal earthquakes	--	Of the 80 earthquakes in this database 22 are interface and 58 are intraslab events	All earthquakes are subduction type earthquakes	-Strong GM records from Italy. -ITACA database includes a total of 3955 records. The statistics presented here is for events with $M_w \geq 3.0$	-Strong GM records from several regions in Europe -72 of these records misses one component of motion.	Additionally 34456 GM records at borehole were considered reliable by the authors M_{JMA} range between 2.7 and 8

1 **Table 2.** Summary of the number of KiK-net records with geometric mean of PGA that
2   exceed specific values

		Borehole		Surface	
		$R_{epi}{\leq}300$km	$R_{epi}{>}300$km	$R_{epi}{\leq}300$km	$R_{epi}{>}300$km
	0.01g	3081	150	22454	1195
Number of	0.05g	304	14	3025	97
records with	0.10g	107	4	1055	34
PGA$^{*}\geq$	0.20g	14	1	343	14
	0.30g	2	0	167	5

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

3

4

5

6

7

8

9

10

11

12

13

14

15

16

1   **Table 3.** The error margins allowed during the cross matching of events from the KiK-net and F-
2   net databasess for each match category

Maximum error in:	Category			
	(A)	(B)	(C)	(D & E)
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	-

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

1   **Table 4.** The number of earthquakes and records associated with the five match categories

	Number of earthquakes matched from the KiK-net database	Number of GM records
Category (A)	4393	139273
Category (B)	73	2807
Category (C)	46	2861
Category (D)	28	3238
Category (E)	403	8778
Total	**4943**	**156957**

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

1 **Table 5.** The output of the different declustering trials on the F-net seismic catalog

Algorithm	Parameters	Number of clusters	Number of mainshocks (Independent earthquakes)	Gutenberg-Richter b-Value	Is the declustered catalog Poissonian?
Reasenberg (1985)	τ_{min}=2 days, τ_{max}=10 days, p_1=0.99, x_k=0.50, and r_{fact}=10	881	17093	0.8963	No
Gardner and Knopoff (1974)	-Gardner and Knopoff (1974)	2054	6376	0.7624	No
	-Gruenthal (personal Comm.)	1739	3688	0.6908	No
	-Uhrhammer (1986)	1286	11165	0.8710	No

2

3

4

5

6

7

8

9

10

11

12

13

14

Table 6. 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 coll./trans. • Hotspot • REVISION[*]

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

Table 7. 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	1161	1123
SZ interface	5	873
SZ outer	-	16
Shallow active crustal	1093	-
ACR deep	-	112
ACR shallow	-	1083
OBR	-	3
Not classified	951	-
Total	3210	3210

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

Table 8. 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] R$_{Rup}$	R$_{JB}$	Classification Garcia et al (2012)	Reference	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. (2013)

Table 9. Parameters calculated at each KiK-net site

Parameter	Description
$V_{s,X}$	Average shear wave velocity from ground surface up to a depth of X meters (e.g., $V_{s,30}$)
$V_{s,max}$	Maximum shear wave velocity in the profile
$V_{s,min}$	Minimum shear wave velocity in the profile
$V_{s,0}$	Shear wave velocity at the ground surface
$V_{s,mean}$	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
$V_{s,h800}$	Average shear wave velocity from ground surface up to the depth h_{800}
$V_{s,borehole}$	Shear wave velocity at the location of the borehole instrument
Hole depth	Depth of the borehole instrument
NEHRP site class	A if $V_{s,30} > 1500$; B if $1500 \geq V_{s,30} > 760$; C if $760 \geq V_{s,30} > 360$; D if $360 \geq V_{s,30} \geq 180$; and E if $180 > V_{s,30}$

FIGURES
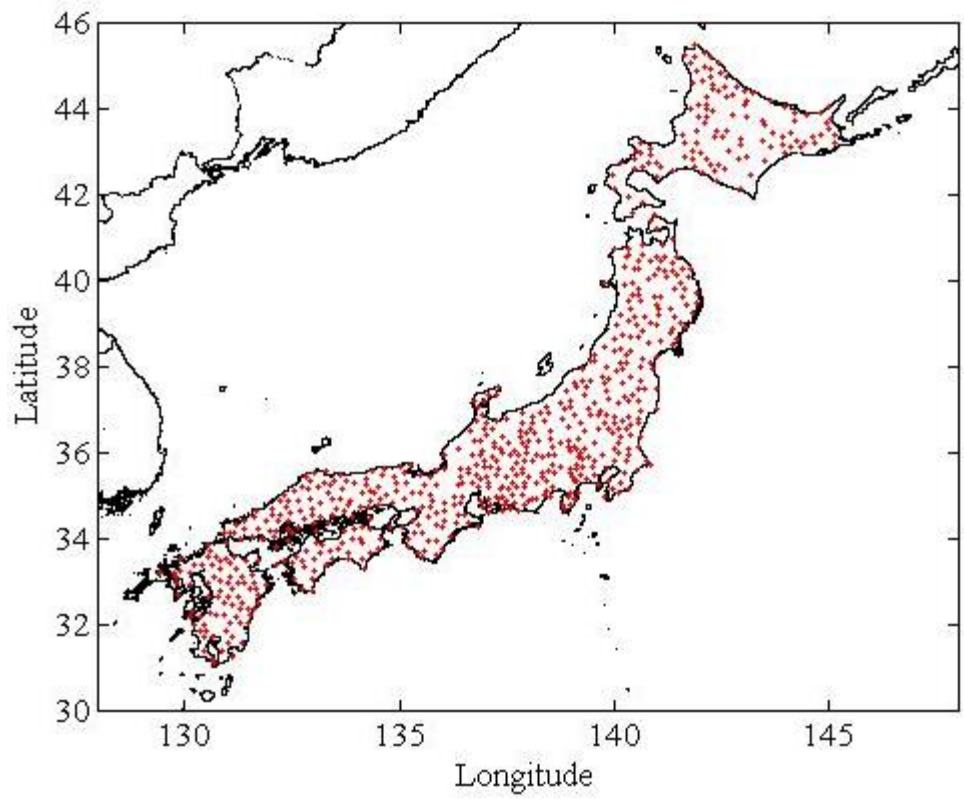
Figure 1. Location of the KiK-net stations (red dots) on the Japanese map
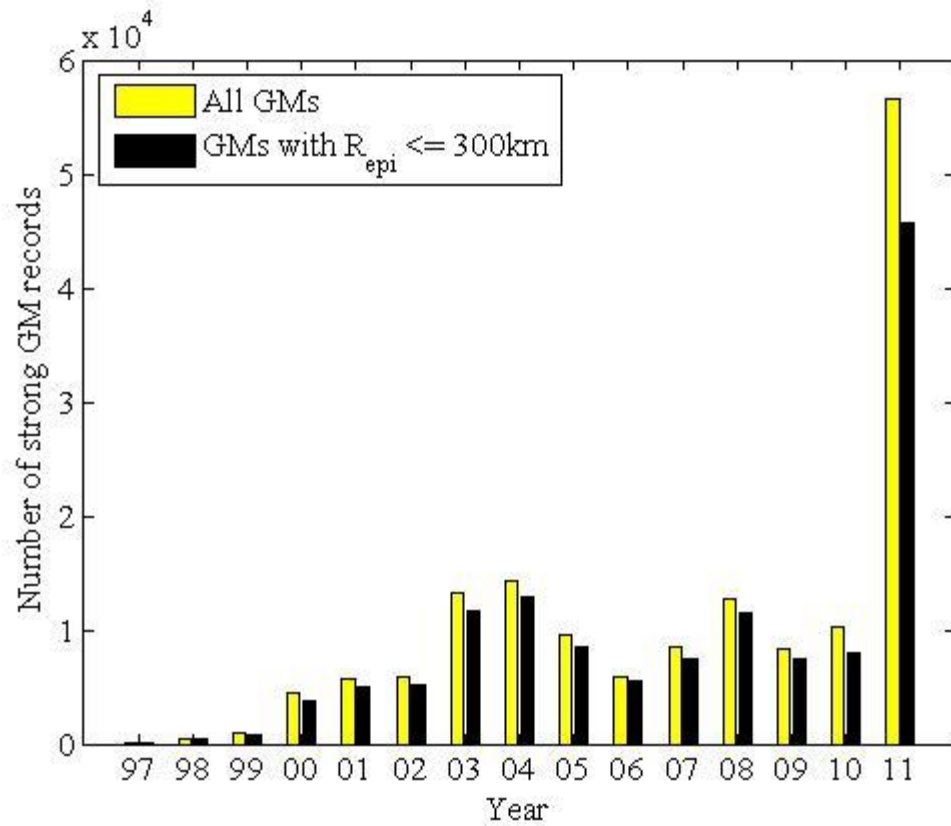
Figure 2. Temporal distribution of the ground motion (GM) records in the KiK-net database. Each GM consists of six components (three components at borehole and three at the ground surface).
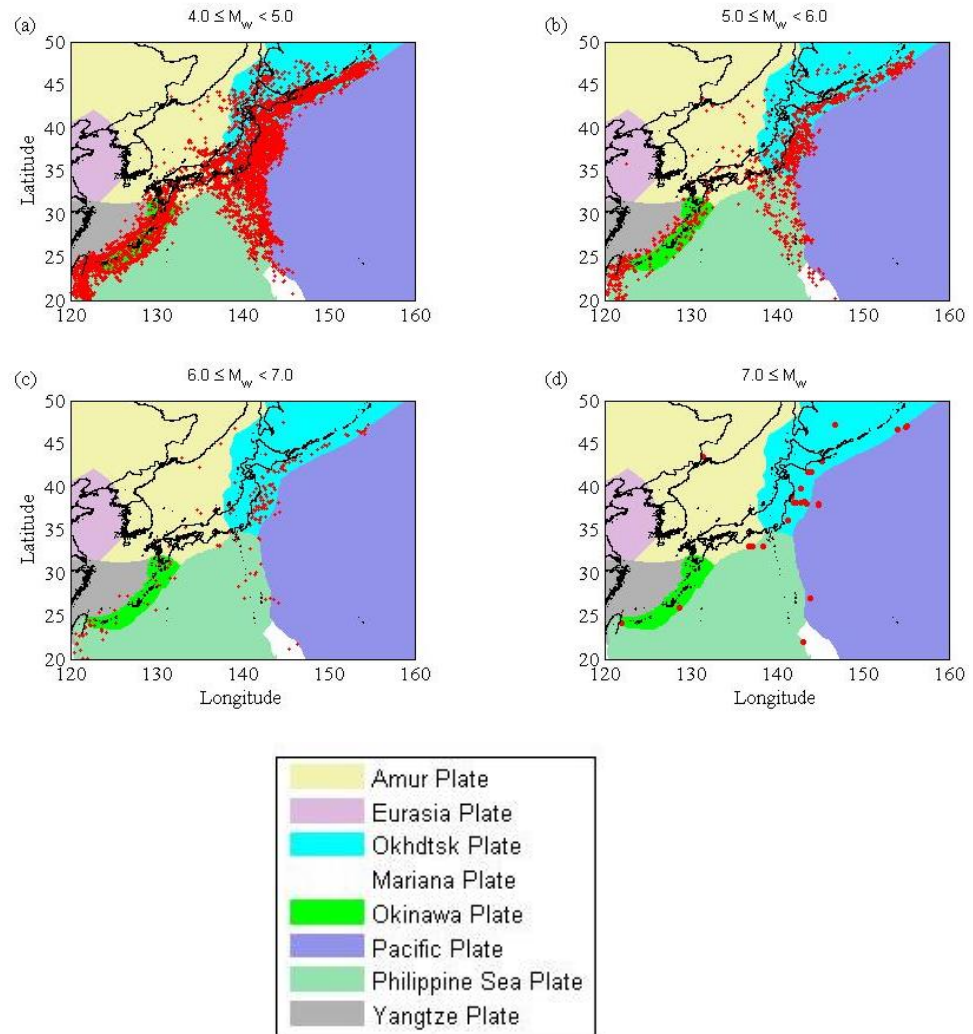
Figure 3. Spatial distribution of the earthquakes in the F-net catalog (between 1997 and 2011) for: (a) $4.0 \leq M_w < 5.0$, (b) $5.0 \leq M_w < 6.0$, (c) $6.0 \leq M_w < 7.0$, and (d) $7.0 \leq M_w$ [The different tectonic plates around Japan are assigned different colors (Plate boundaries obtained from Bird 2003).]
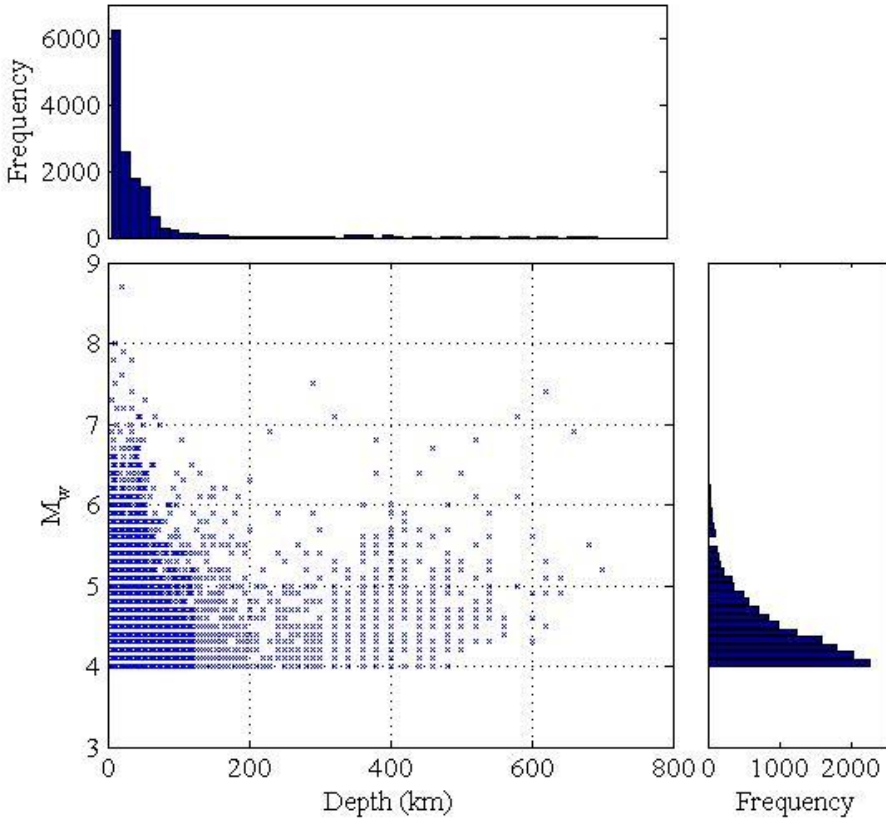
Figure 4. A scatter plot for M_W versus depth (km) for the earthquakes in the F-net catalog (bottom-left subplot) and two histograms that show the number of earthquakes in depth bins (top-left subplot) and the number of earthquakes in different M_W bins [bottom-right subplot].
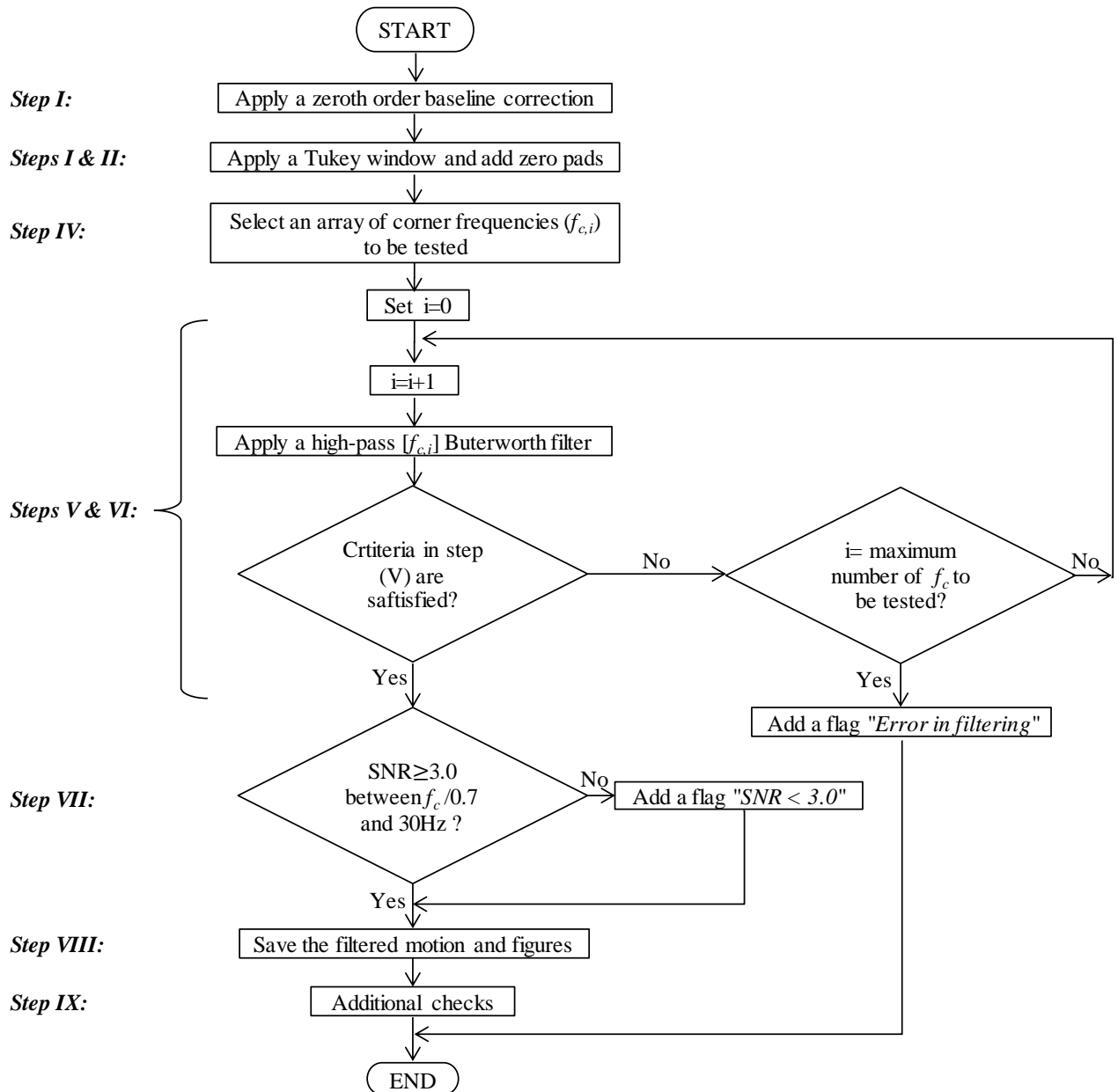
Figure 5. Flow chart diagram showing the steps of the processing protocol
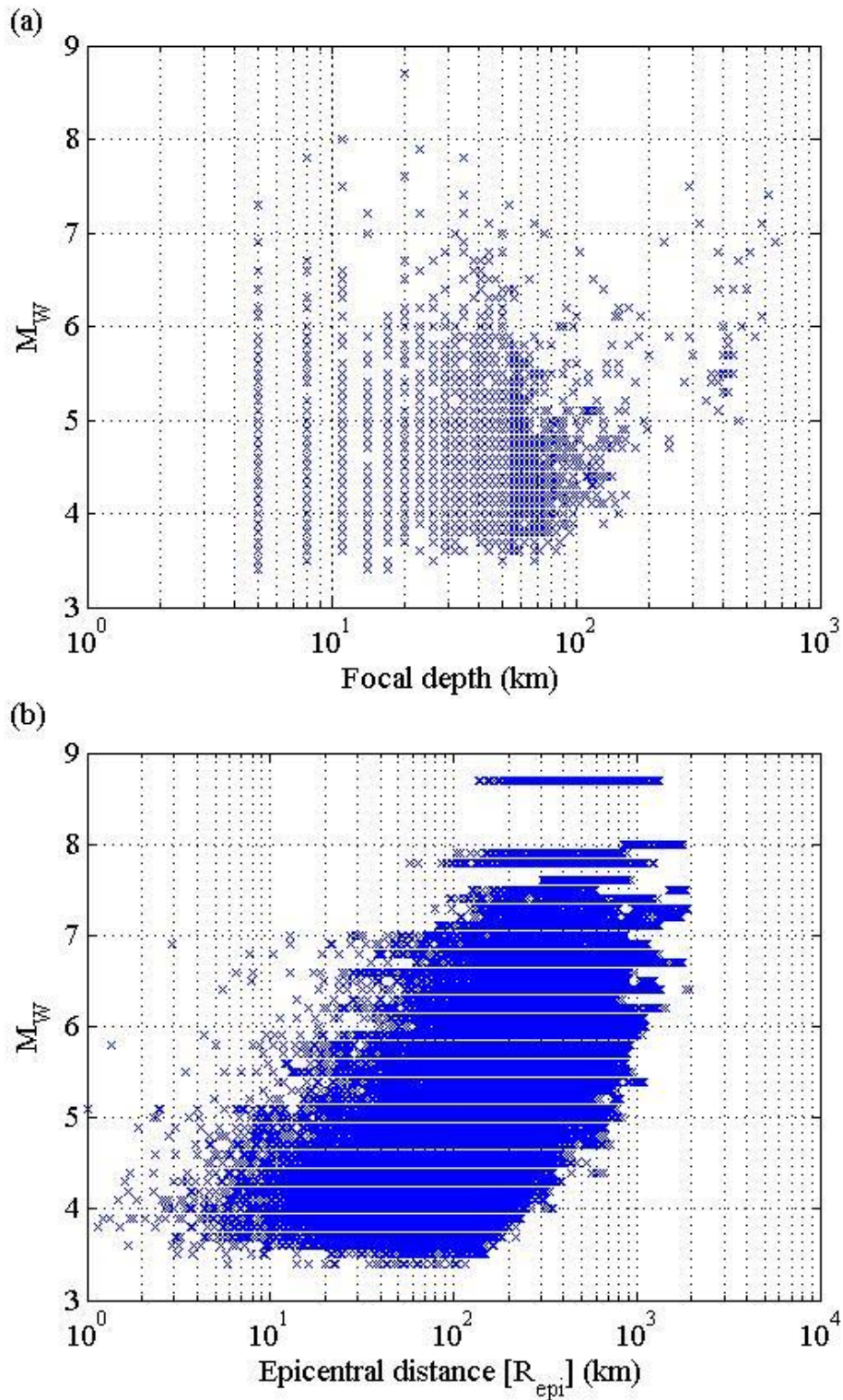
Figure 6. Plots of M_w magnitude for the events that fall within the matching categories (A) through (D) versus: (a) the focal depth and (b) epicentral distance for each GM record.
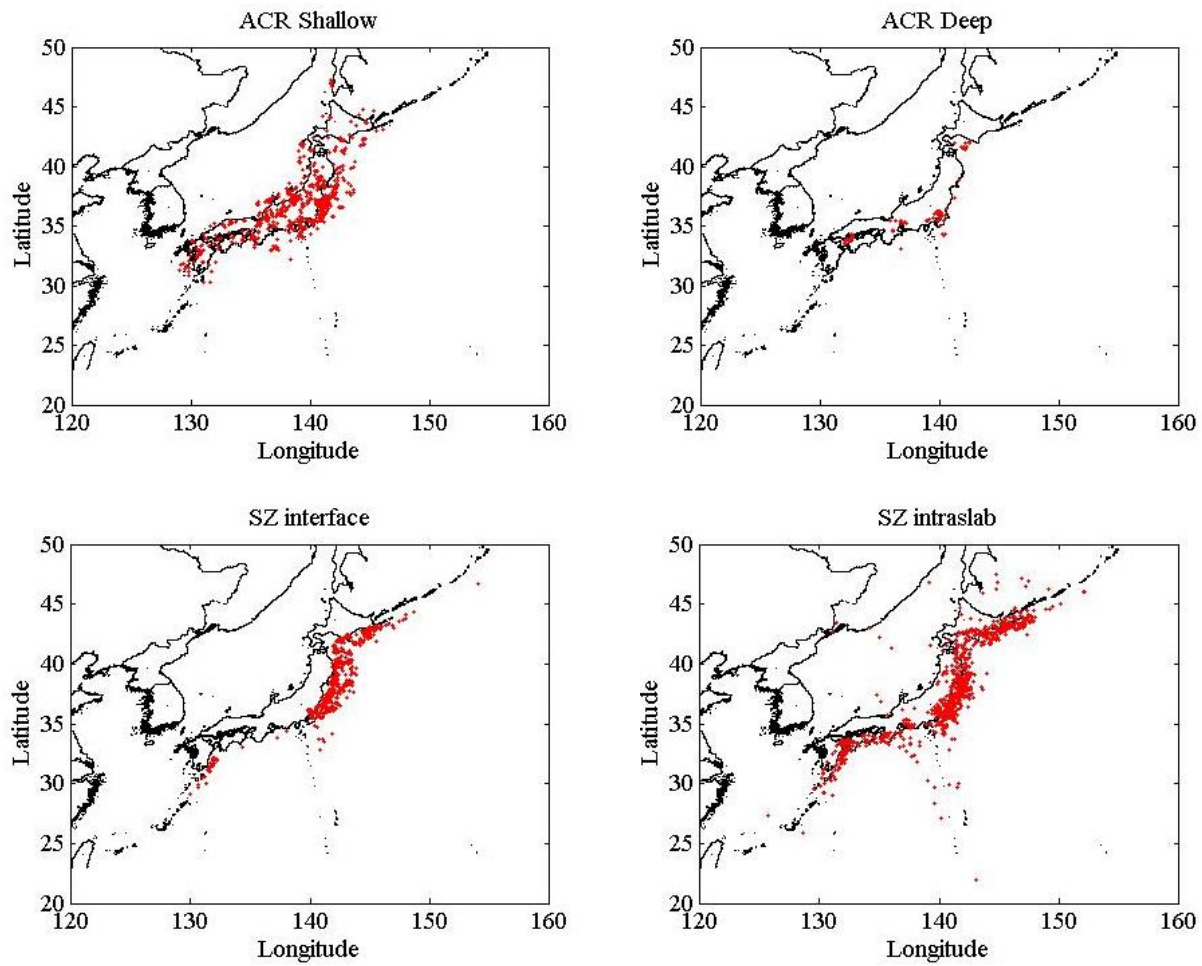
Figure 7. The location of the earthquakes classified as ACR shallow (top left); ACR deep (top right); SZ interface (bottom left); and SZ intraslab (bottom right) on the Japanese map. The classification is made using the algorithm by Garcia et al. (2012)

44
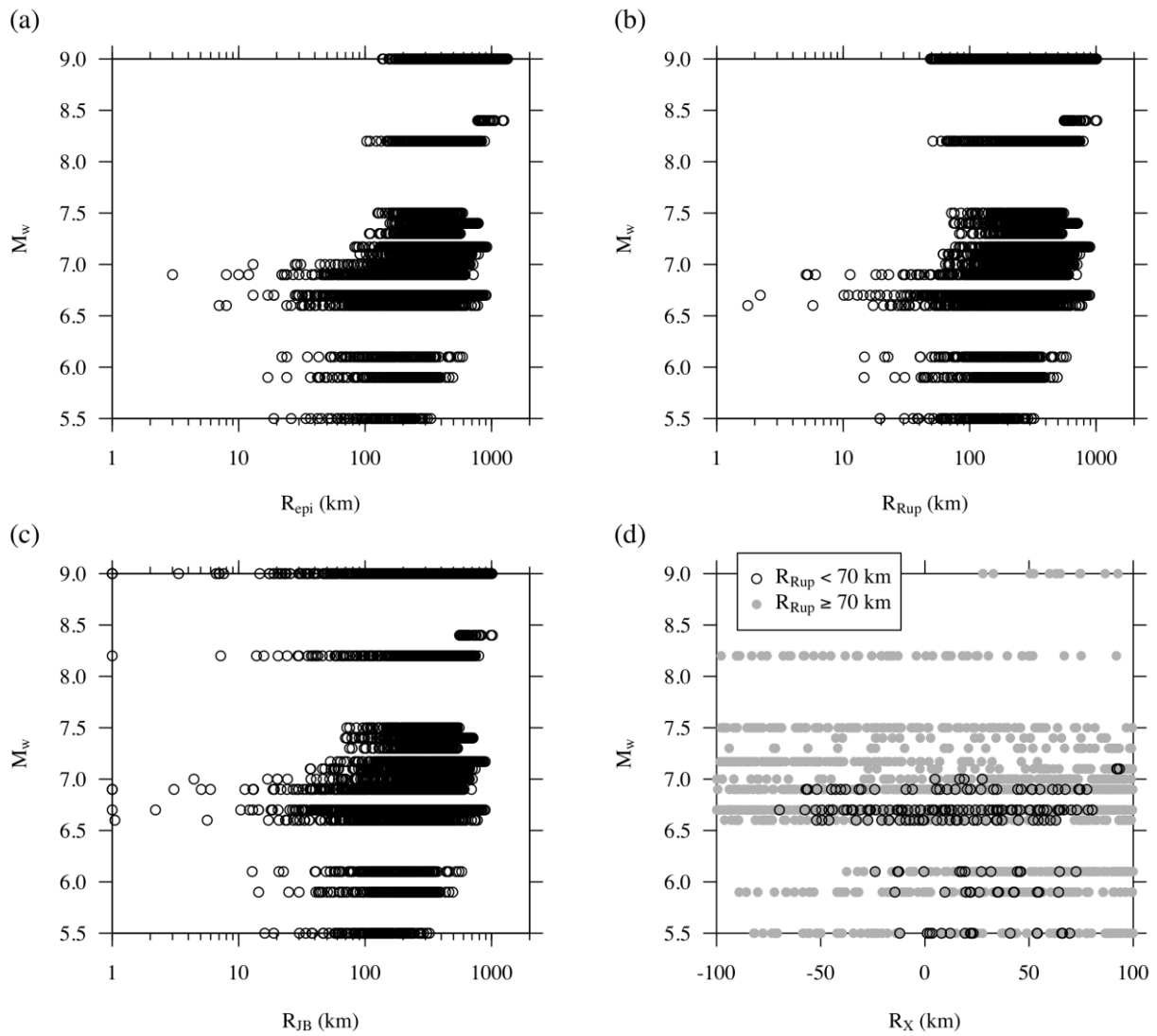
Figure 8. The distribution of M_w versus: (a) R_{epi}; (b) R_{Rup}; (c) R_{JB}; and (d) R_x, for the earthquakes shown in Table 9. For log-axes, distances less than 1 km were rounded up to 1 km.
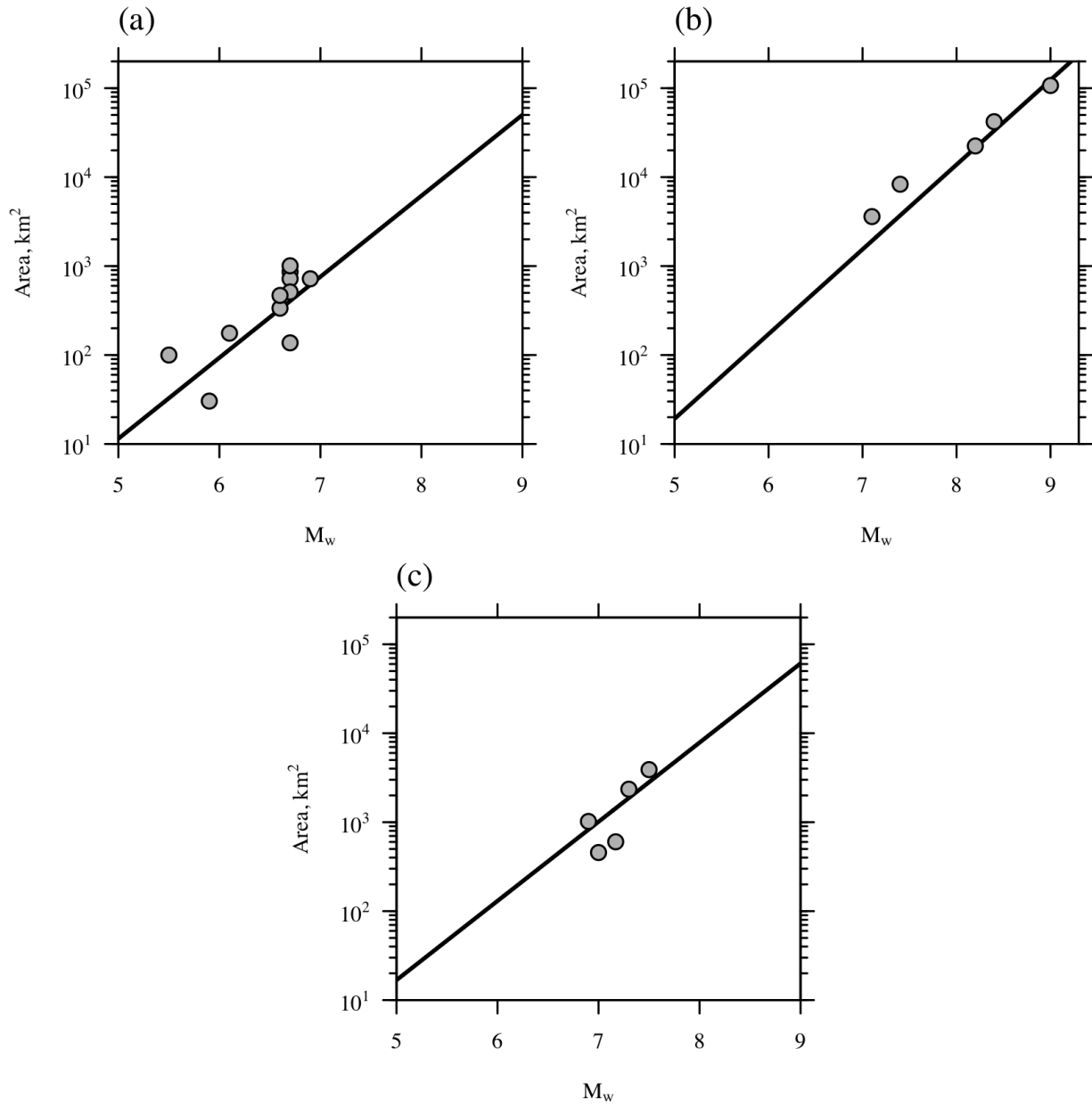
Figure 9. M_w versus rupture area for the finite fault models in Table 9; (a) compares the Wells and Coppersmith (1994) relationship for active crustal events, and (b) and (c) compare the Strasser et al. (2010) relationships to the finite fault models in Table 9 for subduction zone interface and intraslab events, respectively.
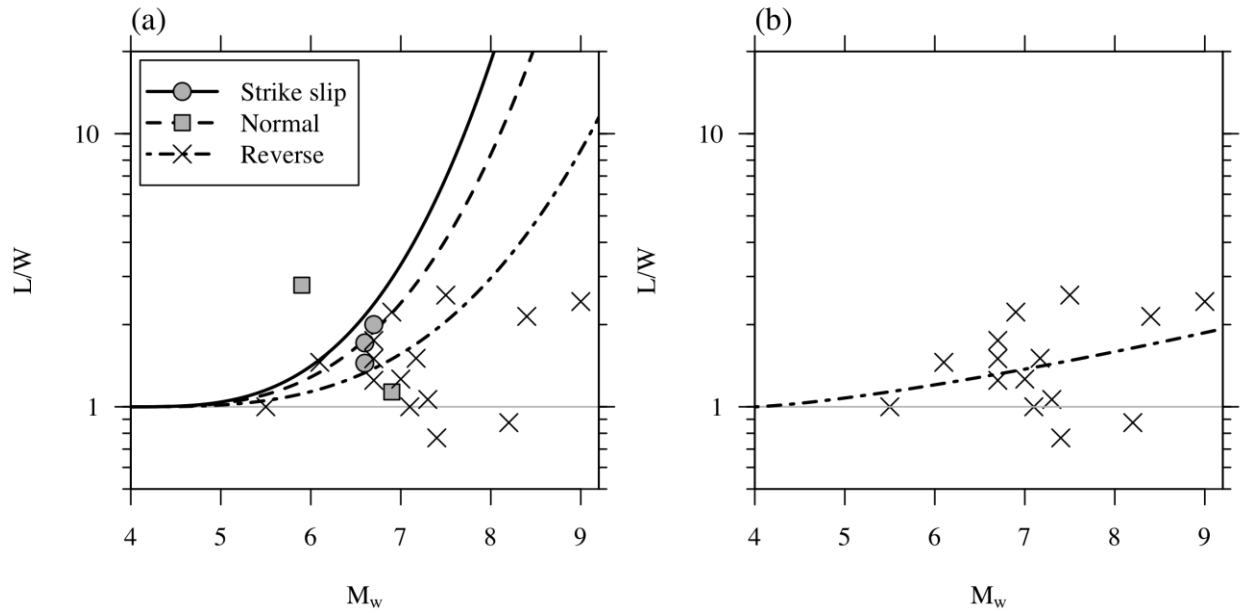
Figure 10. M_w versus AR from the finite fault models in Table 9 compared to (a) the Chiou and Youngs (2008) relationships for different fault types, and (b) a curve fit to the reverse faults in Table 9. .
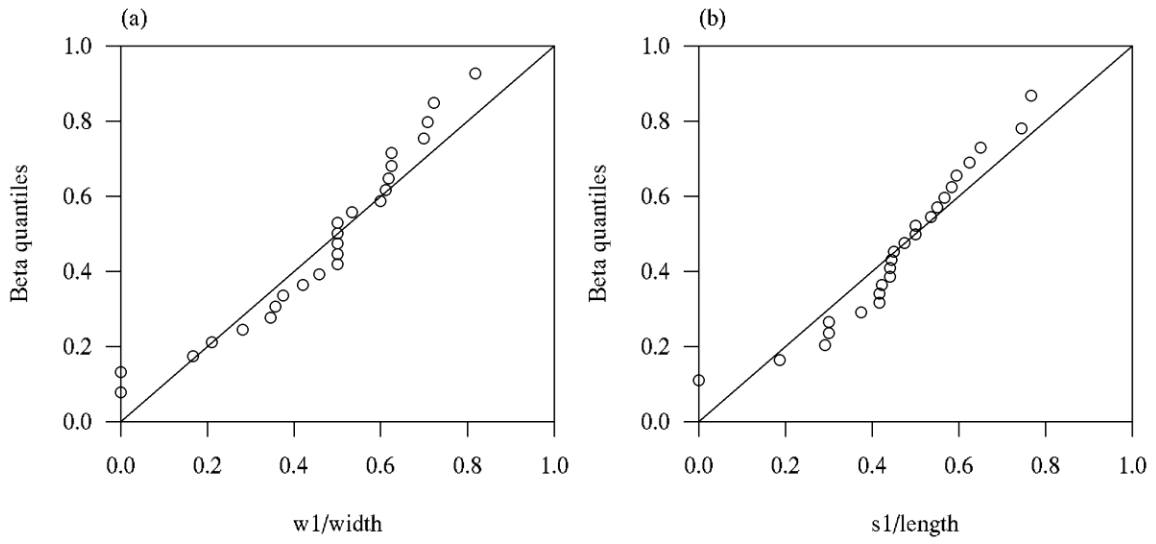
Figure 11. QQ plots for w1/width (a), and (b) s1/length.
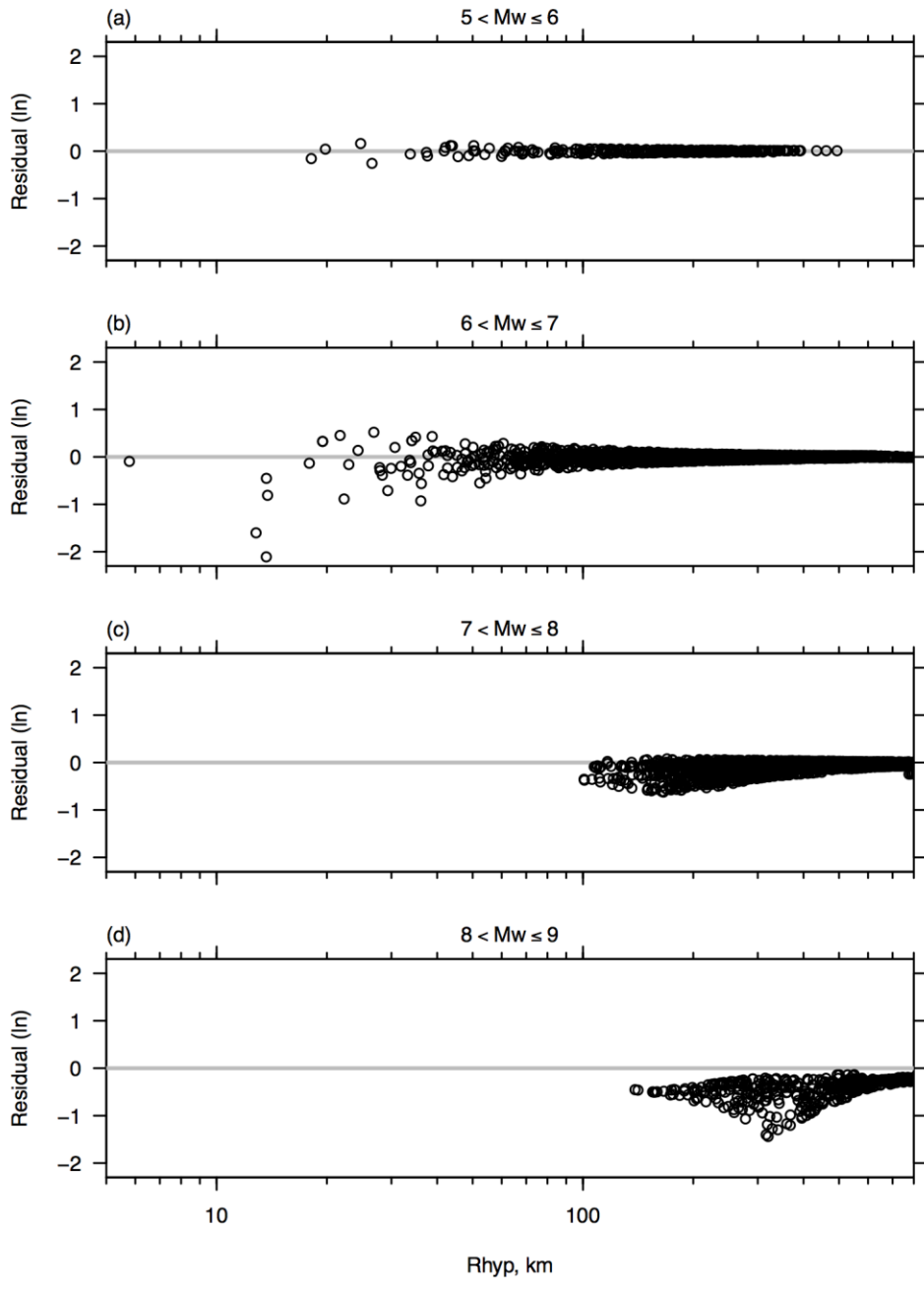
Figure 12. Comparison of the preferred simulated R_{Rup} and those computed from finite fault models for different magnitude ranges.

50
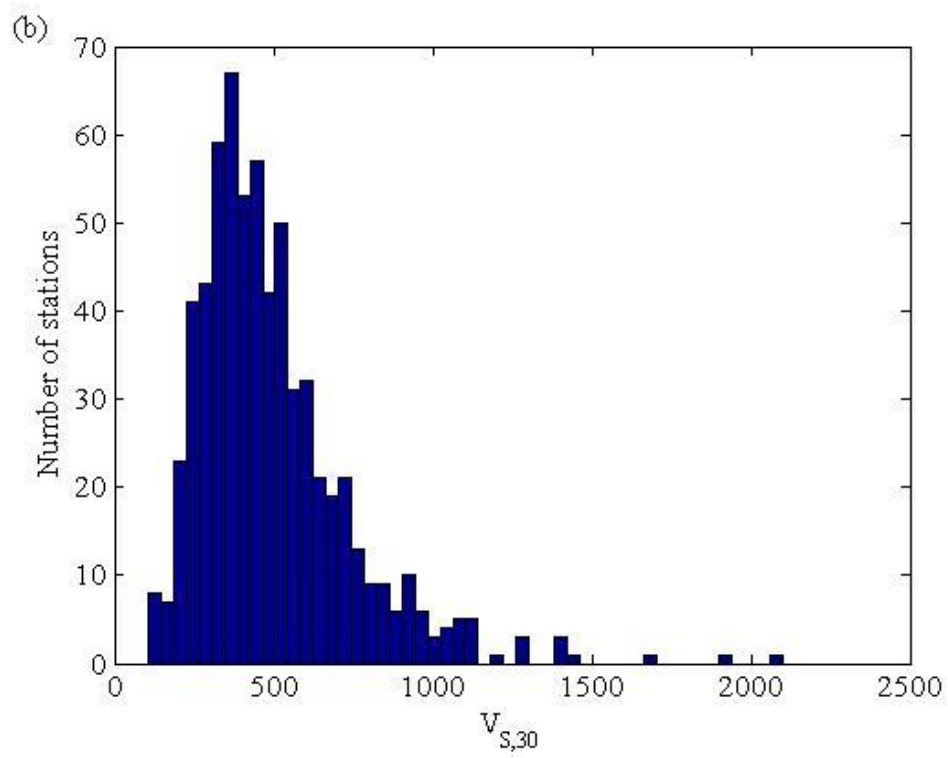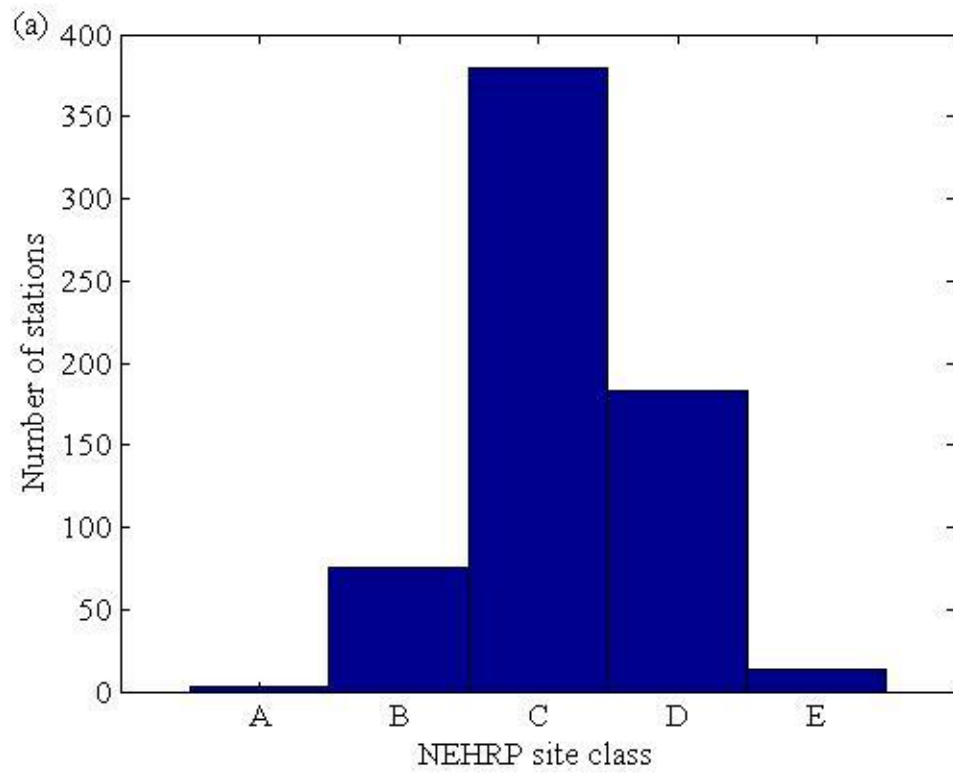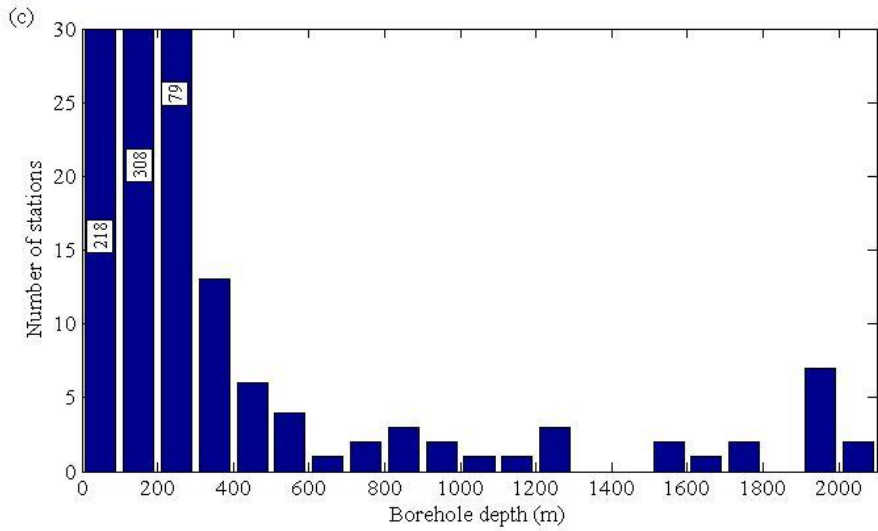
Figure 13. Histograms showing: (a) the NEHRP site class for KiK-net stations; (b) the $V_{S,30}$ for KiK-net stations; and (c) the depth of the borehole instrument for KiK-net stations

An empirical Ground Motion Prediction Equation for Active Crustal Earthquakes Using the Japanese KiK-net Database (Ergodic and Site-Specific Formulations)

Haitham M. Dawood[1] and Adrian Rodriguez-Marek

ABSTRACT

In this manuscript, we present an empirical ground motion prediction equation (GMPE) for active crustal earthquakes. The database consists of 13735 six-components strong ground motions, from 679 active crustal earthquakes recorded at 643 KiK-net stations in Japan. We used a functional form similar to the form adopted by Abrahamson et al. (2013; ASK13) in the NGA-West 2 project to develop an ergodic and a site-specific GMPE. We used mixed effects regression analyses to obtain the regression coefficients for both GMPEs. We conducted a comprehensive residual analysis to capture any potential biases in the residuals. Then, we propose different models to predict the components of variability as a function of different parameters. We then conduct a comparison between the ergodic and site-specific GMPEs to investigate the adequacy of using the residuals from the ergodic GMPE to obtain site-specific residual components. We investigate stations that show considerably higher or lower variability than the average value.

[1] Corresponding Author:
hdawood@vt.edu
200 Patton Hall, Virginia Tech
Blacksburg, VA 24061

INTRODUCTION

Empirical ground motion prediction equations (GMPE) are widely used in seismic hazard analyses to predict the median and standard deviation of specific ground motion parameters (e.g., peak ground acceleration). Larger standard deviations can have a large impact on hazard estimates, in particular for critical facilities that are designed for long return periods (Bommer and Abrahamson 2006). Hence, a better quantification of the ground motion uncertainty is essential. The number of independent parameters in the GMPEs' functional forms as well as the number of motions used to regress for its coefficients considerably increased over the past 40 years. This might have improved the predictions, but the total standard deviation did not reduce during these 40 years (Strasser et al. 2009). This lack of reduction in the standard deviation could be attributed to using the ergodic assumption in developing these GMPEs. The ergodic assumption implies that the standard deviation applicable to a specific site-path-source combination is equal to the standard deviation estimated using the whole database (Anderson and Brune 1999). It was found that a better quantification (in most cases reduction) of the ground motion uncertainty could be achieved by releasing the ergodic assumption. On the other hand, the removal of the ergodic assumption requires dense instrumental networks operating in seismically active zones so that a sufficient number of recordings are made. Such networks were built just recently, and the KiK-net network is a prime example.

In this study we used a large database of strong ground motions to develop an ergodic and site-specific GMPE using the same database. We then propose different models to predict the different components of variability obtained from both GMPEs. Some recording stations show considerably higher or lower variability than the average value observed across all stations. Hence, we present a discussion about the potential causes of such extreme values of variability

and its implications in hazard analyses. In most hazard analyses that release the ergodic assumptions, the residuals obtained from ergodic GMPEs were split using random effects analyses. Hence, these studies assumed that the median prediction is similar for ergodic and non-ergodic GMPEs. This should be the case if the database in hand is well balanced. In this manuscript we explore the differences in the median predictions of the ergodic and non-ergodic GMPEs. This study expands the work previously done by Rodriguez-Marek et al. (2011) by using a larger database; using a different GMPE functional form; extending the maximum usable spectral period to 7.0 seconds; comparing the ergodic and site-specific formulations; developing models to predict the different components of variability; and studying more in depth the stations characterized with small or large variability.This manuscript primarily addresses developers of GMPEs and researchers in the field of non-ergodic seismic hazard analyses.

DATABASE

The database used in this manuscript consists of 13,735 six-components strong ground motions, from 679 active crustal earthquakes recorded at 643 KiK-net stations. The earthquakes were classified s active crustal based on an automated algorithm proposed by Garcia et al. (2012). Figure 1 shows the distribution of the earthquakes and recording stations on the Japanese map. Figure 2 shows the shows the distribution of moment magnitude (M_W) versus hypocentral depth, the distribution of M_w versus closest distance (R_{rup}), and R_{rup} versus v_{s30} for the motions used in this study.

The motions were processed using an automated processing protocol presented in Dawood et al. (2014; Dea14). The metadata for most earthquakes was obtained from the F-net seismic catalog. An exception to that are 10 earthquakes for which a finite fault models were found in the literature (Table 1). For these earthquakes, the metadata was obtained from the

3

published references. The R_{rup} for motions of these 10 earthquakes were calculated using the published fault models. For all other motions, R_{rup} were calculated using the methodology described in Chiou and Youngs (2008). For additional details about the metadata collection, the reader is referred to Dea14.

The database used in this study is a subset of the database described in details in Dea14. The subset was chosen based on a set of criteria summarized as follows:

- Remove motions from earthquakes with $M_w \leq 4.0$.

- Remove motions from earthquakes with questionable metadata.

- Remove motions from earthquakes that are not classified as active crustal earthquakes according to the algorithm by Garcia et al. (2012).

- Remove motions for which the signal-to-noise ratio drops below 3 in the frequency range between $2f_c$ and 25Hz in any of the four horizontal components. Where f_c is the corner frequency of the acausal filter used to baseline correct the motion (Dea14).

- Remove motions with R_{rup} greater than 350 km.

- Remove motions that were found to contain multiple wave trains (Dawood 2014). The only exception in this case was motions from earthquakes with $M_w \geq 5.0$. In that case, the wave trains from the fore/afterchocks were trimmed from the motion and the remaining portion of the motion was processed and included in the current database.

- Remove motions recorded at stations without a measured shear wave velocity profiles.

- Remove motions from earthquakes that were recorded at less than 5 stations.

In addition to the criteria listed above, two earthquakes classified as "active crustal" by the algorithm by Garcia et al. (2012) were removed since we found that these two earthquakes (Kii Peninsula earthquakes, 2004/09/05) were classified by Park and Mori (2005) as subduction earthquakes. For more details about these adopted criteria the reader is referred to Dawood (2014).

The maximum usable spectral period of the ground motions was found to be half of the inverse of the corner frequency of the filter used to filter the motion (Dawood 2014). Hence, the number of usable motions is reduced as the spectral period of interest increases. Figure 3 shows the number of earthquakes and motions used in regression analysis at each spectral period. The large drop in the number of earthquakes and motions used in the analysis for spectral periods larger than 1.0s is obvious. It is also worth noting that only about 25% of the motions in this database were recorded from earthquakes with M_w greater than or equal to 5.0.

FORMULATION OF THE GMPES

GMPEs can be categorized as ergodic, partially non-ergodic (e.g., site-specific or path- and site-specific), or fully non-ergodic (Al Atik et al. 2010). In the current study, we develop GMPEs with different levels of non-ergodicity. In this paper, we present the analysis of the ergodic and site-specific GMPEs, while the path- and site-specific and fully non-ergodic GMPEs are addressed in Dawood and Rodriguez-Marek (2014; hereinafter referred to as companion paper). The key difference between the different levels of non-ergodicity is way the residuals are split between the different sources of variability (i.e., source, source-to-site travel path, site and

random error). In this section we summarize the formulation of the ergodic and site-specific GMPEs.

The formulation of the ergodic GMPE is as follows:

$$ln\ (y_{es}^{G/B}) = ln(\mu_{es}^{G/B}) + \delta B_e + \delta W_{es}^{G/B} \qquad \text{Equation (1)}$$

where $ln\ (y_{es}^{G/B})$ is the natural logarithm of an observed ground motion parameter; $ln(\mu_{es}^{G/B})$ is the median prediction of a ground motion parameter (pseudo-spectral acceleration in this study); δB_e is the inter-event residual component; and $\delta W_{es}^{G/B}$ intra-event residual component. The subscripts e and s denote earthquake and station, respectively. The superscripts G and B denote motions recorded at the ground surface or at the borehole instrument (within motion), respectively. In this case, δB_e represents the observed shift from the median for motions from earthquake e. Since δB_e is an event-specific residual component, it should be the same whether the recorded motion is on ground surface or in borehole. Assuming that the residual components are uncorrelated, the total standard deviation at the ground surface and the borehole instruments can be defined as:

$$\sigma_{tot}^{G/B} = \sqrt{\tau^2 + (\varphi^{G/B})^2} \qquad \text{Equation (2)}$$

where τ and $\varphi^{G/B}$ are the standard deviations of δB_e and $\delta W_{es}^{G/B}$, respectively.

In the case of a site-specific (partially non-ergodic) GMPE, $\delta W_{es}^{G/B}$ can be split to include a site-specific component:

$$\delta W_{es}^B = \delta B2B_{sb} + \delta W_{esb}^0 \qquad \text{Equation (3)}$$

$$\delta W_{es}^G = \delta S2S_s + \delta W_{es}^0 \qquad \text{Equation (4)}$$

where $\delta B2B_{sb}$ is a borehole-to-borehole residual component which represents the observed shift from the median for motions from this specific borehole that underlies site s; $\delta S2S_s$ is a surface-to-surface residual component which represents the observed shift from the median for motions from site s; and δW_{esb}^0 and δW_{es}^0 represent the remaining residual components after accounting for the site-specific residual components. Developing this type of GMPEs requires a database that contains multiple strong motions recorded at the same station. Assuming that the residual components are uncorrelated, the total standard deviations can be defined as:

$$\sigma_{tot}^G = \sqrt{\tau^2 + \varphi_{S2S}^2 + \varphi_{G,0}^2} \qquad\qquad \text{Equation (5)}$$

$$\sigma_{tot}^B = \sqrt{\tau^2 + \varphi_{B2B}^2 + \varphi_{B,0}^2} \qquad\qquad \text{Equation (6)}$$

Since the KiK-net network has two co-located instruments on ground surface and below surface, a site-specific empirical amplification residual component ($\delta S2S_s^{AMP}$) and remaining amplification component (δAMP_{esb}^0) can be defined as:

$$\delta S2S_s^{AMP} = \delta S2S_s \text{ - } \delta B2B_{sb} \qquad\qquad \text{Equation (7)}$$

$$\delta AMP_{esb}^0 = \delta W_{es}^0 \text{ - } \delta W_{esb}^0 \qquad\qquad \text{Equation (8)}$$

Hence, δW_{es}^G can be re-written as:

$$\delta W_{es}^G \quad = \delta W_{es}^B + \delta S2S_s^{AMP} + \delta AMP_{esb}^0 \qquad\qquad \text{Equation (9)}$$

FUNCTIONAL FORM OF THE MEDIAN PREDICTION

In this manuscript, we adopted the functional form used by ASK13 within the NGA-W2 project for the median prediction for the pseudo-spectral accelerations. Whenever necessary we

altered this functional form to better fit the nature of the database used in this study. A key difference between the NGA-W2 database and the current database is that we performed the regression analysis using the surface and borehole motions simultaneously. A similar approach was adopted by Rodriguez-Marek et al. (2011).

The median prediction is given by:

$$ln(\mu_{es}^{G/B}) = \mathrm{F_{Surf}}\ a1_S + (1-\mathrm{F_{Surf}})\ a1_B + f_1\,(\mathrm{M_w,R_{rup}}) + \ \mathrm{F_N}\,f_8(\mathrm{M_w}) + f_5{}^{G/B}\,(\mathrm{Sa_{1100},V_{S30}}) + f_6(\mathrm{Z_{TOR}}) +$$

$$f_{10}{}^{G/B}\,(\mathrm{Z_1,V_{S30}}) + f_{11}{}^{G/B}\,(\mathrm{V_{S,Borehole}}) \hspace{3cm} \text{Equation (10)}$$

where $ln(\mu_{es}^{G/B})$ is the natural logarithm of the predicted pseudo spectral acceleration in units of gravity (given as a ratio of 9.81m/s^2). The super-script *(G)* denotes motions recorded at ground surface and *(B)* denotes borehole motions. The different parameters used in this model are presented in Table 2.

The functional forms of $f_1, f_8, f_5{}^{G/B}, f_6,$ and $f_{10}{}^{G/B}$ are as shown below:

Abrahamson et al. (2013) adopted a three segment model for f_1. One of the segments was applicable for motions from earthquakes with $\mathrm{M_w}$ greater than 6.75 to 7.25 (depending on the spectral period on interest). Since the largest earthquake in the database is $\mathrm{M_w}$ 6.9, we only used two segments for f_1. Based on preliminary regression analyses; we found that an $\mathrm{M_w}$ of 5.0 was a suitable boundary between the two segments. We used the same magnitude and distance scaling coefficients for surface and borehole motions. Hence, the functional form of f_1 was chosen to be:

$f_1\,(\mathrm{M_w,R_{rup}}) =$

$$\begin{cases} a5\,(M_w - M_1) + a8\,(8.5 - M_w)^2 + [a2 + a3(M_w - M_1)]ln(R) + a17R_{rup} & for\ M_w \geq M_1 \\ a6\,(M_w - M_2) + a8\,(8.5 - M_2)^2 + [a2 + a3(M_2 - M_1)]ln(R) + a17R_{rup} & for\ M_w < M_1 \end{cases}$$

<div align="right">Equation (11)</div>

where:

$$R = \sqrt{R_{rup}^2 + c_{4M}^2} \qquad\qquad\qquad Equation\ (12)$$

and

$$c_{4M} = \begin{cases} c4 & for\ 5 < M_w \\ c4 - (c4 - 1)(5 - M_w) & for\ 4 < M_w \leq 5 \\ 1 & for\quad M_w \leq 4 \end{cases} \qquad Equation\ (13)$$

The median prediction for motions from normal faults is corrected by f_8:

$$f_8(M_w) = \begin{cases} a12 & for\quad M_w > 5 \\ a12(M_w - 4) & for\ 4 \leq M_w \leq 5 \\ 0 & for\quad M_w < 4 \end{cases} \qquad Equation\ (14)$$

We adopted the same site response model presented in ASK13. Since we use borehole and surface motions in the regression analysis, we adopted different linear site response coefficients for surface and borehole motions (i.e., $a10_S$ and $a10_B$). Hence the functional form of f_5 used is the regression analysis was chosen to be:

$f_5{}^{G/B}$ (Sa$_{1100}$,V$_{S30}$) =

$$
\begin{cases}
\begin{aligned}
&F_{Surf}\left[a10_S \ln\left(\tfrac{V_{S,30}^*}{V_{Lin}}\right) - b\ln\left(Sa_{1100,Surf} + c\right)\right] + \\
&(1 - F_{Surf})\left[a10_B \ln\left(\tfrac{V_{S,30}^*}{V_{Lin}}\right) - b\ln\left(Sa_{1100,Bore} + c\right)\right] + \\
&b\ln\left(F_{Surf}Sa_{1100,Surf} + (1 - F_{Surf})Sa_{1100,Bore} + c\left(\tfrac{V_{S,30}^*}{V_{Lin}}\right)^n\right) \quad for\ V_{S,30} < V_{Lin}
\end{aligned} \\[1em]
\left[F_{Surf}\ a10_S + (1 - F_{Surf})\ a10_B + bn\right]\ln\left(\tfrac{V_{S,30}^*}{V_{Lin}}\right) \qquad\qquad for\ V_{S,30} \geq V_{Lin}
\end{cases}
$$

Equation (15)

where:

$$
V_{S,30}^* = \begin{cases} V_{S,30} & for\ V_{S,30} < V_1 \\ V_1 & for\ V_{S,30} \geq V_1 \end{cases}
$$

Equation (16)

Abrahamson et al. (2013) adopted a model where the depth to top of rupture (Z$_{TOR}$) scaling function (f_6) is constant for Z$_{TOR}$ deeper than 20 km because their data is sparse beyond 20km. The current database contains a considerable amount of earthquakes with Z$_{TOR}$ deeper than 20km (a total of 233 earthquakes). Preliminary regression analyses suggested that the model used by ASK13 was not suitable for this database. Hence, we modified the Z$_{TOR}$ model to be:

$$
f_6\ (Z_{TOR}) = \begin{cases}
a_{15a}\dfrac{Z_{TOR}-3}{10-3} & for\qquad Z_{tor} < 10km \\
a_{15a} & for\ 10 \leq Z_{tor} < 30km \\
a_{15a} + (a_{15b} - a_{15a})\dfrac{Z_{TOR}-30}{50-30} & for\ 30 \leq Z_{tor} < 50km \\
a_{15b} & for\ 50 \leq Z_{tor}
\end{cases}
$$

Equation(17)

We used the same model adopted by ASK13for the depth to a V$_s$=1.0 km/s (Z$_1$). The only modification we did was that we regressed for different coefficients for surface and borehole motions. Hence, we modified the Z$_{TOR}$ model to be:

$f_{10}{}^{G/B}$ (Z$_{1.0}$,V$_{S30}$)

$$
\begin{cases}
\text{F}_{Surf}\, a43_S\, ln\left(\dfrac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) + (1 - \text{F}_{Surf})\, a43_B\, ln\left(\dfrac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) & for \qquad V_{S,30} \leq 200m/s \\[2mm]
\text{F}_{Surf}\, a44_S\, ln\left(\dfrac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) + (1 - \text{F}_{Surf})\, a44_B\, ln\left(\dfrac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) & for\ 200 < V_{S,30} \leq 300m/s \\[2mm]
\text{F}_{Surf}\, a45_S\, ln\left(\dfrac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) + (1 - \text{F}_{Surf})\, a45_B\, ln\left(\dfrac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) & for\ 300 < V_{S,30} \leq 500m/s \\[2mm]
\text{F}_{Surf}\, a46_S\, ln\left(\dfrac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) + (1 - \text{F}_{Surf})\, a46_B\, ln\left(\dfrac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) & for\ 500 < V_{S,30}
\end{cases}
$$

Equation (18)

Preliminary residual analysis showed that the residuals of the borehole records are biased with respect to the shear wave velocity of the layer where the borehole instrument is installed. Hence, we included an additional term to the GMPE functional form to remove the bis in the residuals as follows:

$$
f_{11}{}^{G/B}\, (V_{S,\text{Borehole}}) = (1 - \text{F}_{Surf})a_{Borehole}\, ln\left(\frac{V_{S,Borehole}}{1600}\right) \qquad\qquad \text{Equation (19)}
$$

We did not consider the hanging wall effect on the median predictions to simplify the regression analysis.

REGRESSION ANALYSIS

We conduct two sets of regression analyses to obtain an ergodic and a site-specific GMPE. The regression analyses were conducted using multiple steps. After each step, one of more of the regression coefficients were smoothed with respect to spectral periods and fixed for the subsequent steps. The regression analysis was conducted using a combination of linear and

nonlinear mixed effects regression analyses incorporated in commercial statistical software named SAS.

Similar to ASK13, we first included a normal faulting flag (F_N) and a reverse faulting flag in the regression analysis. The coefficient associated with of the reverse faulting flag was smoothed to zero across all periods. This suggests that the motions from reverse faults and strike-slip faults in the current database are similar. Hence, the reverse faulting component was removed from the functional form of the median prediction (Equation 10). It is worth noting that ASK13 smoothed that coefficient to a value of "zero" across all spectral periods as well.

Since the information required to adopt the aftershock model used by ASK13 was not available for the current database, we chose to adopt the aftershock model adopted by Abrahamson and Silva (2008). We found that the coefficient associated with the aftershock flag is not significantly different than zero. Hence, we smoothed it to zero across all spectral periods. Hence, we removed the aftershock flag component from the median prediction (Equation 10).

For details about the step-by-step analyses adopted, and the comparisons between the regression coefficients before and after smoothing, the reader is referred to Dawood (2014). Some of the coefficients (i.e., V_1, V_{LIN}, b, c, and n) were not regressed for in this study as they were adopted directly from ASK13 (all except V_1 are based on a study by Kamai et al. (2013)). Tables 3 through 6 present the values of the regression coefficients for the ergodic and site-specific GMPEs.

COMPONENTS OF VARIABILITY

ERGODIC GMPE

The total variability is split into an inter-event component (τ) and an intra-event component ($\varphi^{G/B}$) (see Equation 2). Abrahamson et al. (2013) parameterized the two components as a function of magnitude, except for φ^{G} for the Japanese data which was parameterized as a function of distance. Preliminary analysis showed that τ does not show M_w dependence while $\varphi^{G/B}$ shows M_w and R_{rup} dependence. Hence, we proposed four different models (i.e., constant, M_w-dependent, R_{rup}-dependent, and M_w-R_{rup}-dependent models) to parameterize $\varphi^{G/B}$ and one (i.e., constant model) to parameterize τ. All models are summarized in Table 7. We used Maximum likelihood estimation to regress for the coefficients of these models (Table 8). As a first step, the M_w and R_{rup} at which the different branches of the models are applicable were regressed for. These break points were then smoothed across all spectral periods and fixed in the final analysis. Figure 4 shows $\varphi^{G/B}$ and τ obtained from the different models at PGA, 0.3s, 1.0s, and 7.0s. The figure also shows the standard deviations calculated from the current data at different M_w and R_{rup} bins to show the suitableness of the proposed models. The following observations can be drawn from these plots:

τ models:

- The τ and $\varphi^{G/B}$ models were developed simultaneously in the same analysis. Hence, three different constant models for τ that were developed with the first three $\varphi^{G/B}$ models. As expected, the three models resulted in very close estimates of τ across all spectral periods. The three values were reported for completeness (Table 8 and Figure 4).

$\varphi^{G/B}$ models:

- We recommend using the distance dependent models for distances between 10 and 350 km because of the scarcity of the data we have beyond these limits. Also, the M_w-dependent models should be used for M_w between 4 and 6.5.

- In general, the models for φ^G and φ^B converge at longer spectral period. This could be attributed to the fact that φ^G is inflated at short periods due to the heterogeneities in the upper soil layers. The effect of these heterogeneities is minimal at long periods.

- The within event standard deviation ($\varphi^{G/B}$) obtained from the M_w-dependent model was found to decrease with magnitude for small periods and increase for large periods. NGA models observed similar trends with M_w up to a certain M_w. Beyond that threshold (ranges between 5.5 and 6.5 for the different models), they report a constant φ^G. The current data does not support the presence of this cap. However, this could be attributed to the limited number of motions from large earthquakes.

- Abrahamson et al. (2013) developed R_{rup}-dependent models for the Japanese motions in their database. Their φ^G model was assumed to be constant for distances less than 30km and larger than 80km. The current data does not support their model since we observe a clear reduction in $\varphi^{G/B}$ with distance. Hence, we adopted a sloping model up to a distance of 350km.

- Another difference between the current R_{rup}-depdendet model and the model by ASK13 is that they reported higher φ^G at larger distanced for small periods and lower φ^G at larger distanced for long periods, while we observed the opposite

14

behavior. The M_w-R_{rup}-dependent models show the reason for this discrepancy at short periods. The M_w-R_{rup}-dependent models show the tendency of φ^G to: a) increase with distance for large magnitudes and b) decrease with distance for small magnitudes at short periods. Hence, our R_{rup}-dependent models were probably biased by the considerably larger number of motions from small magnitude events to show a decreasing φ^G with distance. On the other hand, ASK13 used mainly large events from Japan (6.1 to 6.9) which made them capture the tendency of increase in φ^G with distance. This shows the importance of developing models that are M_w-R_{rup}-dependent.

SITE-SPECIFIC (SS) GMPE

Similar to what we carried out in the previous section, we developed models to predict $\varphi_{SS}^{G/B}$ (also known as $\varphi_{G/B,0}$) as a function of R_{rup} and M_w and constant τ. Table 9 shows the functional forms of the developed models. Table 10 shows the regression coefficients. Similar to the Ergodic analysis, we obtained the break points of the models from preliminary regression analysis, we smoothed the values across the different periods and rerun the regression with these break points fixed. The only difference is that the first break point in the R_{rup}-dependent model (i.e., 30km) was manually fixed. Figure 5 shows the behavior of the proposed models at four spectral periods (PGA, 0.2s, 1.0s, and 7.0s). The plots also show the standard deviation of δW_{esb}^0, δW_{es}^0, and δB_e (Equations 1, 3 and 4) within selected R_{rup} or M_w bins to show the suitableness of the developed models. The following observations can be drawn from these plots:

$\varphi_{SS}^{G/B}$ models:

15

- Similar to the $\varphi^{G/B}$ models, we recommend using the distance dependent models for distances between 10 and 350 km and the M_w-dependent models between 4 and 6.5 for the $\varphi_{SS}^{G/B}$ models.

- The current database suggests that the two breaks of the Rrup-dependent model should be at 30 and 100 km instead of at 16 and 36 km adopted by R-M13.

- Rodriguez-Marek et al. (2013) had two breaks at magnitudes of 5 and 7 for the Mw-dependent φ_{SS}^{G} model. The current database supported using a similar model with a break at a magnitude of 5, but we had no motions from earthquakes larger than 7. Hence, we could not investigate the behavior beyond a magnitude of 7.

- The constant φ_{SS}^{G} model from this study predicted values that are 10% higher than those from the study by R-M13 at PGA and 0.2s. At 1.0s, both studies predicted the same φ_{SS}^{G}.

- The Mw-dependent model for $\varphi_{SS}^{G/B}$ does not show much improvement over the constant model especially for motions recorded at ground surface.

- This study and R-M13 predict fairly similar values of φ_{SS}^{G} for earthquakes with M_w less than 5. On the other hand, the magnitude dependence observed by R-M13 was much stronger than the one we found in this study. Hence, R-M13 predicts fairly smaller (about 27% smaller) values for φ_{SS}^{G} compared to our study.

- As observed by R-M13, $\varphi_{SS}^{G/B}$ for the current database show a strong R_{rup} dependency. The M_w-R_{rup}-models showed that this distance dependence is stronger for small magnitude earthquakes at short periods and for large magnitude earthquakes at long periods.

- An important observation was that φ_{SS} seem to be fairly similar at the ground surface and borehole. This is a significant observation since it might prove that

16

the large variability observed at the surface site conditions (e.g., site amplifications, and topographic effects) were properly accounted for in the φ_{S2S} and φ_{B2B} components of variability.

Figure 6 shows the values obtained for φ_{S2S} *and* φ_{B2B} from the site-specific GMPE. The site-to-site variability observed at the ground surface is considerably higher at the ground surface that at the borehole. This observation was expected because the borehole motions are less affected by the near surface heterogeneities.

SITE-SPECIFIC WITHIN-EVENT VARIABILITY AT A SPECIFIC SITE ($\varphi_{SS,S}^{G/B}$)

Site-specific within-event variability at a specific site ($\varphi_{SS,S}^{G/B}$) is an important component of variability in performing site-specific PSHA. This variability represents the standard deviation of either δW_{esb}^0 or δW_{es}^0 for all motions recorded at the site of interest. The difference between $\varphi_{SS}^{G/B}$ and $\varphi_{SS,S}^{G/B}$ is that the first represent the observed variability for all stations that recorded motions in the database, while the later represent the variability at a specific station. Hence, $\varphi_{SS}^{G/B}$ represent an average variability observed over a large number of stations. Models to predict $\varphi_{SS}^{G/B}$ as a function of M_w and R_{rup} are important for stations where there is lack of sufficient recorded strong motions to reliably estimate $\varphi_{SS,S}^{G/B}$.

It was observed that $\varphi_{SS,S}^{G/B}$ can be either much higher or much lower than the $\varphi_{SS}^{G/B}$ estimated from the models for some stations. This is important because the adopted value of $\varphi_{SS,S}^{G/B}$ can have a major impact on the hazard curves obtained from PSHA. Figure 7 (see top panels) shows the histograms of the $\varphi_{SS,S}^G$ for stations that recorded 5 motions or more in the current database. The figure also presents the mean, mean±standard deviation of $\varphi_{SS,S}^G$ and the

17

moment and magnitude independent φ_{SS}^G obtained from the models (Model 1 in Table 10) presented in the previous section. As expected, the average $\varphi_{SS,S}^G$ for all stations almost equals φ_{SS}^G. The standard deviation of $\varphi_{SS,S}^G$ ranges between 0.11 and 0.13 for PGA, 0.3s, and 1.0s. The correlation coefficients between $\varphi_{SS,S}^G$ and $\varphi_{SS,S}^B$ are 0.70, 0.84, 0.40, and 0.98 for PGA, 0.2s, 1.0s and 7.0s, respectively. These relatively high correlation coefficients between the variability at borehole and ground surface might reflect the fact that a substantial portion of $\varphi_{SS,S}^{G/B}$ is reflecting the variability in the source and travel path which are in common between borehole and surface motions. Moreover, the correlation coefficient seems to increase with spectral periods and reaches almost 1.00 at a spectral period of 7.0s. At longer periods, the heterogeneities in the properties of the near surface layers have a weaker effect on the motions. On the other hand, we have just 10 stations with 5 or more usable motions at a spectral period of 7.0s. Hence, additional data is needed to confirm that such high correlation coefficient is expected at very long periods.

In this section we particularly focus on stations with either high of low $\varphi_{SS,S}^{G/B}$ values to better understand why there is such a large variability. We plotted plots of $\delta W_{esb/es}^0$ with R_{rup} and azimuth for each station. Based on these plots the following observations were made:

- In general, $\delta W_{esb/es}^0$ (see Equations 3 and 4) do not show strong dependence on R_{rup}. On the other hand, in some cases $\delta W_{esb/es}^0$ from motions recorded from

18

certain R_{rup} ranges seem to have strong bias. We found that in most of these cases, these biased motions also share the same azimuth range (i.e., the motions originate from the same source.

- *The* site-specific within-event residual component ($\delta W^0_{esb/es}$) is azimuth dependent for many stations characterized with high $\varphi^{G/B}_{SS,S}$ (e.g., station AKTH17 (PGA-Borehole motion); Figure 8-a). It is clear from the figure that $\delta W^0_{esb/es}$ estimated from motions with an azimuth between 150 and 180 degrees are strongly biased towards negative values. These motions also share the same R_{rup} range (55-80 km). Hence, they originate from the same seismic region. This seismic region is close to active volcanoes. The magma chambers associated with these active volcanoes might be the cause of such a high bias (e.g., Dawood and R-M13). Another plausible cause of this bias, is that the seismic source have certain characteristics that makes it generate weaker motions compared to the other seismic sources in Japan. These observations show the importance of including the single path and source effects (whenever possible) to improve the predictions obtained from GMPEs.

- For stations with high $\varphi^{G/B}_{SS,S}$ values, we tried to remove the azimuth dependency of $\delta W^0_{esb/es}$ and re-estimate $\varphi^{G/B}_{SS,S}$. This is similar (but not exactly the same) to considering the path- and source-specific effects to obtain the fully non-ergodic variability. To remove this dependency, we sorted all motions recorded at each station in eight azimuth regions (i.e., motions with azimuth between 0-45 degrees, 45-90 degrees,….etc). If more than three motions were available in a region, we subtracted the mean of $\delta W^0_{esb/es}$ of these motions from each $\delta W^0_{esb/es}$. By doing

so, we remove the bias of residuals at the different azimuth ranges. $\varphi_{SS,S}^{G/B}$ is then recalculated for each station using these modified residuals. On average, the modified $\varphi_{SS,S}^{G/B}$ was 13% less than its original estimated value for motions recorded at ground surface (Figure 7). The reduction ranged from 0% to 50% at the different stations. The 0% represents an extreme case where the station recorded less than 3 motions in a certain azimuth bin. Hence, the the average of the residuals was not subtracted from the residuals at any region and $\varphi_{SS,S}^{G/B}$ was kept as is (e.g., station AOMH07 (PGA-Ground surface motion); Figure 8-b). In this case, δW_{es}^0 seems to be biased low between 150 and 180 degrees, and biased high around 240 and 340 degrees. Since there are not many motions in each azimuth region, our algorithm did not modify δW_{es}^0. Hence, $\varphi_{SS,S}^G$ was kept unchanged for this station.

- Japan is characterized by a very complex tectonic setting that might result in major biases for come source-path combinations compared to the median predictions (e.g., active volcanoes, faults, large mountains). Hence, some stations that recorded many earthquakes, still show a $\varphi_{SS,S}^{G/B}$ that is well above $\varphi_{SS}^{G/B}$ (station FKSH14 (PGA-Borehole motion); Figure 8-c). In this case, the recorded motions cover almost all possible azimuth and R_{rup} ranges, yet $\varphi_{SS,S}^B$ is about 0.66. On the other hand, the residual plots show some clear biases for some azimuth-R_{rup} bins that led to such a high $\varphi_{SS,S}^{G/B}$.

- In general, stations characterized with low $\varphi_{SS,S}^{G/B}$ are associated motions with relatively small $\delta W_{esb/es}^0$ values with no or limited azimuth dependence. In many

cases, the recordings at these stations cover narrow azimuth ranges (e.g., station FKOH03 (PGA-Borehole motion); Figure 8-d), or narrow azimuth and distance ranges (e.g., station HDKH02 (PGA-Borehole motion); Figure 8-e) ranges. This means that the single-path and/or source effects are included in the estimated $\varphi_{SS,S}^{G/B}$, which might explain the observed low $\varphi_{SS,S}^{G/B}$ at these stations. In this case, any systematic bias due to the path and/or source effects will be transferred to $\delta B2B_{sb}$ and $\delta S2S_s$. This shows that that calculating $\varphi_{SS,S}^{G/B}$ using motions recorded from a mainshock and its subsequent aftershocks may result in an under estimation of $\varphi_{SS,S}^{G/B}$ unless this is the only seismic source to be considered in the PSHA at that site. If there are other sources surrounding that specific station, then the estimated $\varphi_{SS,S}^{G/B}$ might underestimating the real variability at the station.

- Based on the observations summarized in this section, we believe that calculating $\varphi_{SS,S}^{G/B}$ using recorded motions should be conducted with care. This particularly true because of the source and path contribution in $\varphi_{SS,S}^{G/B}$. Hence, it seems that there is no specific number of motions that could be considered enough to estimate a reliable value for $\varphi_{SS,S}^{G/B}$. Instead, the relative location of sources of the recorded motion with respect to the location of the seismic sources around the site of interest.

Based on these observations, we conducted additional analysis to check the effect of azimuth on $\varphi_{SS,S}^{G/B}$ for the whole network. To this end we did the followings: 1) divide the area around each station into eight 45 degree regions; 2) count the number of motions recorded at

21

each station from every region; 3) rotate the axes so that the region that has the maximum number of recordings points towards 0 degrees; 4) normalize the number of records at each of the eight region by the largest number of motions in any of them. These steps result in a single curve (azimuth on the horizontal axis versus "relative number" of motions in each region) for each station at every spectral period. If this curve plots perfectly horizontally, this means that the station recorded the same number of motions from all directions. While the closest this curve from being vertical, the narrower is the window from where the motions travelled to the site. Obviously the curve is in general neither of these two hypothetical cases. We superimposed these curves from all stations and calculated the average and median of the "relative number" of records from each region (Figure 9). We distinguished between stations characterized by "High", "Neutral" and "Low" $\varphi_{SS,S}^{G}$ values. It is clear from the figure that on average "Low" $\varphi_{SS,S}^{G}$ stations have lower curve compared to the "Neutral" which in turn has lower curves compared to the "High" stations. This observation is constantly seen over all periods for most azimuth regions. This observation points out that on average stations with "Low" $\varphi_{SS,S}^{G}$ recorded s motions from a narrower azimuthal window compared to the "Neutral" and "High" stations.

Figure 10 shows the scatter plot of the $\varphi_{SS,S}^{G}$ versus V$_{s30}$. The scatter plots show that $\varphi_{SS,S}^{G}$ tend to decrease for large V$_{s30}$. It also show substantial scatter at low V$_{s30}$ values. These trends are expected since hard rock sites are less heterogeneous than softer sites. This heterogeneity results in substantial scatter that inflate $\varphi_{SS,S}^{G}$ at softer sites.

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 and November 2012, respectively. The F-net seismic catalog was also provided by NIED at (http://www.fnet.bosai.go.jp). It was last accessed in October 2012.

REFERENCES

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.

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

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.

Dawood, H. M., 2014. Partitioning uncertainty for non-ergodic probabilistic seismic hazard analyses. PhD dissertation, Virginia Polytechnic Institute and State University, USA.

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.

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.

Kamai, R., Abrahamson, N.A., and Silva W.J., 2013. Nonlinear horizontal site response for the NGA-West2 project, PEER Report No. 2013/12, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Kobayashi, T., Tobita, M., Koarai, M., Okatani, T., Suzuki, A., Noguchi, Y., Yamanaka, M., and Miyahara, B., 2013. InSAR-derived crustal deformation and fault models of

normal faulting earthquake (Mj 7.0) in the Fukushima-Hamadori area, Earth Planets Space 64, 1209-1221.

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.

Rodriguez-Marek, A., Montalva, G. A., Cotton, F., and Bonilla, F., 2011. Analysis of single-station standard deviation using the KiK-net data, Bull. Seismol. Soc. Am., Vol. 101, pp. 1242–1258.

Rodriguez-Marek, A., Cotton, F., Abrahamson, N., Akkar, S., Al Atik, L., Edwards, B., Montalva, G. A., Dawood, H. M. 2013. A model for single-station standard deviation using data from various tectonic regions. Bull. Seismol. Soc. Am., Vol. 103, pp. 3149-3163.

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.

TABLES

Table 1. List of earthquakes with finite fault models used in this analysis (adopted from Dea14)

Earthquake name	Date (Y/M/D)	M_w	Reference
Geiyo [F-net #4055]	2001/03/24	6.7	Yagi and Kikuchi (2001)
Miyagi-ken Hokobu [mainshock] [F-net#6730]	2003/07/26	6.1	Hikima and Koketsu (2004)
Miyagi-ken Hokobu [foreshock] [F-net#6725]	2003/07/26	5.5	Hikima and Koketsu (2004)
West Off Fukuoka Prefecture [F-net# 9638]	2005/03/20	6.6	Asano and Iwata (2006)
2007 Notohanto [F-net# 12669]	2007/03/25	--	Pulido et al. (2008)
Fukushima-Hamadori [F-net# 20807]	2011/04/11	6.7	Kobayashi et al (2013)
Iwate-Miyagi Nairiku [F-net# 14479]	2008/06/13	6.9	Suzuki et al. (2010)
Niigata Prefecture Chuetsu Earthquake/Mid Niigata Prefecture [F-net# 8880]	2004/10/23	6.7	Honda et al. (2005)
Niigata-ken Chuetsu-oki [F-net# 13215]	2007/07/16	6.7	Aoi et al. (2008)
Western Tottori [F-net# 3575]	2000/10/06	6.6	Peyrat and Olsen (2004)

Table 2. Description of the parameters used in this manuscript

Parameter	Definition
M_w	Moment magnitude
R_{rup}	Closest distances to the fault [the reader is referred to Dea14 for details about how R_{rup} was calculated for each motion]
V_{S30}	Average shear wave velocity of the top 30m at the recording station.
Z_1	Depth to a Vs=1.0km/s measured in km
Z_{TOR}	Depth to the top of rapture in km
F_N	Flag for normal faulting events [normal faulting definition used in this study is the same one used in NGA-W2 project]. This flag equals 1 for records from normal faulting events and 0 otherwise.
F_{Surf}	Flag for surface records. This flag equals 1 for surface records and 0 for borehole records.
Sa_{1100}	Median pseudo-spectral acceleration for Vs=1100m/s

Table 3. The period independent coefficients [common between ergodic and site-specific GMPEs]

a4	a5	M_1	M_2	c	n
-	1.0909	5	5	2.4	1.5

Table 4. Period dependent regression coefficients common between the ergodic and site-specific

GMPEs

Period (Sec)	V_{Lin}	b	V1	a3	c4
PGA	660	-1.470	1500	0.125	4.4500
0.01	660	-1.470	1500	0.125	4.4500
0.02	680	-1.460	1500	0.125	4.4500
0.03	770	-1.390	1500	0.125	4.4500
0.04	853	-1.297	1500	0.125	4.4500
0.05	914	-1.220	1500	0.120	4.4500
0.075	962	-1.150	1500	0.083	4.4500
0.10	913	-1.230	1500	0.052	4.4500
0.15	740	-1.590	1500	0.015	4.0826
0.20	590	-2.010	1500	0.010	3.8219
0.25	495	-2.410	1500	0.010	3.6197
0.30	430	-2.760	1500	0.010	3.4544
0.40	360	-3.280	1500	0.010	3.1937
0.50	340	-3.600	1500	0.010	2.9915
0.75	330	-3.800	1302	0.016	2.6241
1.00	330	-3.500	1177	0.031	2.3634
1.50	330	-2.400	1021	0.055	1.9960
2.00	330	-1.000	923	0.060	1.7353
3.00	330	0.000	800	0.060	1.3678
4.00	330	0.000	800	0.060	1.1071
5.00	330	0.000	800	0.060	0.9049
7.00	330	0.000	800	0.060	0.6000

Table 5a. Period dependent regression coefficients for the ergodic GMPE

Period (Sec)	$a1_B$	$a1_S$	$a15_a$	$a15_b$	$a8$	$a12$	$a2$	$a6$	$a17$	$a_{Borehole}$
PGA	-1.0813	0.4109	0.7	1.2	-0.006	0.28	-1.2362	1.1809	-0.0066	-0.201
0.01	-1.0825	0.4096	0.7	1.2	-0.006	0.28	-1.2362	1.1814	-0.0066	-0.2012
0.02	-1.0617	0.429	0.7	1.2	-0.006	0.28	-1.2362	1.1791	-0.0067	-0.1945
0.03	-0.9025	0.576	0.7	1.2	-0.006	0.28	-1.2362	1.1504	-0.0075	-0.1001
0.04	-0.6618	0.8209	0.7	1.2	-0.006	0.28	-1.2362	1.1033	-0.0085	-0.0415
0.05	-0.448	1.109	0.7	1.2	-0.0111	0.28	-1.2362	1.069	-0.0088	-0.0569
0.075	-0.076	1.46	0.7	1.2	-0.0267	0.28	-1.2362	1.0544	-0.0085	-0.1387
0.10	0.1188	1.65	0.7	1.2	-0.0377	0.28	-1.2362	1.1178	-0.0076	-0.2147
0.15	0.2152	1.78	0.7	1.2	-0.0532	0.27	-1.2362	1.2645	-0.0057	-0.3288
0.20	0.2145	1.82	0.69	1.15	-0.0643	0.245	-1.2362	1.3688	-0.0044	-0.36
0.25	0.1672	1.83	0.67	1.0368	-0.0728	0.2252	-1.2362	1.459	-0.0032	-0.376
0.30	0.155	1.73	0.64	0.9035	-0.0798	0.2056	-1.2362	1.541	-0.0025	-0.376
0.40	0.1225	1.4958	0.5731	0.6931	-0.0908	0.1747	-1.2362	1.6923	-0.0014	-0.376
0.50	0.0557	1.3696	0.5128	0.5299	-0.0994	0.1508	-1.2362	1.8109	-0.0006	-0.376
0.75	-0.0841	1.0917	0.4033	0.2334	-0.1149	0.1073	-1.2362	2.0169	-0.0001	-0.376
1.00	-0.2027	0.8337	0.3256	0.023	-0.126	0.0764	-1.2362	2.1703	0	-0.376
1.50	-0.45	0.4622	0.2161	-0.2735	-0.1415	0.0329	-1.2362	2.4	0	-0.376
2.00	-0.7153	-0.0451	0.1384	-0.4839	-0.1525	0.002	-1.1884	2.5	0	-0.376
3.00	-1.5767	-1.164	0.0288	-0.7804	-0.1681	-0.0415	-1.0417	2.5222	0	-0.376
4.00	-2.2219	-1.9302	-0.0489	-0.9908	-0.1791	-0.0723	-0.9377	2.5222	0	-0.376
5.00	-2.9412	-2.6643	-0.1091	-1.1539	-0.1877	-0.0963	-0.857	2.5222	0	-0.376
7.00	-4.1577	-3.8739	-0.2	-1.4	-0.2006	-0.1324	-0.7353	2.5222	0	-0.376

Table 5b. Period dependent regression coefficients for the ergodic GMPE

Period (Sec)	$a10_B$	$a10_S$	$a43_B$	$a43_S$	$a44_B$	$a44_S$	$a45_B$	$a45_S$	$a46_B$	$a46_S$
PGA	2.0109	1.9408	0.14	-0.082	0.101	-0.1159	0.1191	-0.22	0.05	-0.217
0.01	2.0109	1.9405	0.14	-0.082	0.101	-0.1159	0.1191	-0.22	0.05	-0.217
0.02	1.9949	1.935	0.14	-0.082	0.101	-0.1159	0.1191	-0.22	0.05	-0.217
0.03	1.8688	1.9036	0.14	-0.082	0.101	-0.1159	0.1191	-0.22	0.05	-0.217
0.04	1.7412	1.8624	0.14	-0.082	0.101	-0.1159	0.1191	-0.22	0.05	-0.217
0.05	1.6259	1.838	0.14	-0.082	0.101	-0.1159	0.1191	-0.22	0.05	-0.217
0.075	1.5456	1.7895	0.14	-0.082	0.101	-0.1159	0.1191	-0.22	0.06	-0.217
0.10	1.687	1.6671	0.14	-0.082	0.101	-0.1159	0.1191	-0.22	0.0712	-0.217
0.15	2.2122	1.8584	0.17	-0.082	0.101	-0.1159	0.1191	-0.22	0.0887	-0.19
0.20	2.8379	2.2489	0.2031	-0.082	0.101	-0.1159	0.1191	-0.1941	0.1012	-0.12
0.25	3.4168	2.7189	0.2444	-0.082	0.101	-0.1159	0.1191	-0.1617	0.1108	-0.04
0.30	3.9545	3.1474	0.2713	-0.082	0.101	-0.1	0.1191	-0.0982	0.1187	0.04
0.40	4.7631	3.8499	0.3268	-0.082	0.101	-0.0468	0.1191	0.0371	0.1312	0.1615
0.50	5.2573	4.3159	0.3681	-0.082	0.101	0.0069	0.1191	0.1522	0.1408	0.2148
0.75	5.5992	4.6036	0.4432	-0.05	0.101	0.1043	0.1191	0.2907	0.1583	0.2692
1.00	5.1016	4.1561	0.4965	0.0167	0.13	0.1734	0.1362	0.3405	0.1708	0.2859
1.50	3.2241	2.4748	0.5716	0.3532	0.2095	0.2709	0.2349	0.4173	0.1883	0.291
2.00	1.033	0.3676	0.6249	0.44	0.2864	0.34	0.3049	0.45	0.2055	0.28
3.00	-0.643	-1.1134	0.7307	0.48	0.3949	0.3824	0.4036	0.47	0.2113	0.2602
4.00	-0.6291	-1.0095	0.7056	0.45	0.4069	0.3879	0.425	0.4586	0.17	0.2299
5.00	-0.5156	-0.8916	0.65	0.4	0.3487	0.3363	0.3492	0.3866	0.1386	0.2002
7.00	-0.2391	-0.652	0.57	0.28	0.2755	0.2807	0.2334	0.2649	0.0671	0.106

Table 6a. Period dependent regression coefficients for the site-specific GMPE

Period (Sec)	$a1_B$	$a1_S$	$a15_a$	$a15_b$	$a8$	$a12$	$a2$	$a6$	$a17$	$a_{Borehole}$
PGA	-0.8376	0.522	0.67	1.2	-0.01	0.2	-1.3	1.2292	-0.006	-0.2072
0.01	-0.8387	0.5209	0.67	1.2	-0.01	0.2	-1.3	1.2297	-0.006	-0.2065
0.02	-0.8192	0.5349	0.67	1.2	-0.01	0.2	-1.3	1.2273	-0.0061	-0.1976
0.03	-0.6781	0.6476	0.67	1.2	-0.01	0.2	-1.3	1.1971	-0.0069	-0.1278
0.04	-0.4807	0.8467	0.67	1.2	-0.01	0.2	-1.3	1.1568	-0.0078	-0.0804
0.05	-0.3063	1.0767	0.67	1.2	-0.012	0.2	-1.3	1.1236	-0.0082	-0.0682
0.075	0.0559	1.5	0.67	1.2	-0.025	0.2	-1.3	1.1204	-0.008	-0.1339
0.1	0.2631	1.65	0.67	1.2	-0.0356	0.2	-1.3	1.1844	-0.0071	-0.197
0.15	0.3799	1.8	0.67	1.2	-0.0506	0.2	-1.3	1.3194	-0.0051	-0.3134
0.2	0.3714	1.85	0.67	1.2	-0.0612	0.2	-1.3	1.4244	-0.0038	-0.3606
0.25	0.3408	1.87	0.67	1.0556	-0.0694	0.2	-1.3	1.512	-0.0027	-0.3606
0.3	0.3326	1.81	0.6268	0.9377	-0.0762	0.2	-1.3	1.5809	-0.002	-0.3606
0.4	0.2949	1.6648	0.5586	0.7516	-0.0868	0.2	-1.3	1.728	-0.001	-0.3606
0.5	0.2281	1.5414	0.5057	0.6072	-0.095	0.195	-1.3	1.8439	-0.0002	-0.3606
0.75	0.122	1.2889	0.4095	0.3449	-0.11	0.18	-1.3	2.0504	0	-0.3606
1	0.0009	1.0342	0.3413	0.1588	-0.1206	0.1627	-1.3	2.2133	0	-0.3606
1.5	-0.1482	0.632	0.2452	-0.1035	-0.1356	0.11	-1.3	2.4996	0	-0.3606
2	-0.6046	0.0784	0.177	-0.2896	-0.1462	0.0727	-1.25	2.5733	0	-0.3606
3	-1.5622	-1.1221	0.0809	-0.5519	-0.1612	0.02	-1.0896	2.5733	0	-0.3606
4	-2.3887	-2.0776	0.0127	-0.738	-0.1718	-0.0173	-0.9404	2.5733	0	-0.3606
5	-3.2455	-2.9597	-0.0402	-0.8823	-0.1801	-0.0463	-0.8246	2.5733	0	-0.3606
7	-4.6665	-4.4262	-0.12	-1.1	-0.1925	-0.09	-0.65	2.5733	0	-0.3606

Table 6b. Regression coefficients for the site-specific GMPE

Period (Sec)	a10_B	a10_S	a43_B	a43_S	a44_B	a44_S	a45_B	a45_S	a46_B	a46_S
PGA	2.0033	1.7706	0.1472	-0.0295	0.1353	-0.08	0.0815	-0.1878	0.0379	-0.3125
0.01	2.0025	1.7706	0.1472	-0.0295	0.1353	-0.08	0.0815	-0.1878	0.0379	-0.3125
0.02	1.9848	1.7649	0.1472	-0.0295	0.1353	-0.08	0.0815	-0.1878	0.0379	-0.3125
0.03	1.8609	1.7145	0.1472	-0.0295	0.1353	-0.08	0.0815	-0.1878	0.0379	-0.3125
0.04	1.7137	1.6532	0.1472	-0.0295	0.1353	-0.08	0.0815	-0.1878	0.0379	-0.3125
0.05	1.5928	1.615	0.1472	-0.0295	0.1353	-0.08	0.0815	-0.1878	0.0379	-0.3125
0.075	1.5074	1.5516	0.1472	-0.0295	0.1353	-0.08	0.0815	-0.1878	0.0379	-0.3125
0.1	1.6486	1.4949	0.16	-0.0295	0.1353	-0.08	0.0815	-0.1878	0.0465	-0.29
0.15	2.2065	1.759	0.2045	-0.0295	0.1353	-0.08	0.0815	-0.1878	0.0776	-0.2
0.2	2.8288	2.2251	0.2452	-0.0295	0.1353	-0.08	0.0815	-0.154	0.1	-0.0707
0.25	3.4217	2.7243	0.2768	-0.0295	0.1353	-0.08	0.0815	-0.1049	0.113	0.0354
0.3	3.9698	3.1686	0.3026	-0.0295	0.1353	-0.0688	0.0815	-0.0251	0.125	0.1405
0.4	4.7777	3.8737	0.3432	-0.0295	0.1353	-0.0278	0.0815	0.1038	0.142	0.2664
0.5	5.2589	4.3306	0.3748	-0.0295	0.1353	0.0334	0.0815	0.1926	0.15	0.3054
0.75	5.5945	4.6177	0.4321	-0.0295	0.1353	0.1697	0.0815	0.303	0.156	0.3339
1	5.1137	4.175	0.4728	-0.01	0.1472	0.2482	0.097	0.36	0.157	0.3429
1.5	3.2844	2.5014	0.5301	0.3512	0.2747	0.3988	0.16	0.4	0.1605	0.351
2	1.1124	0.3992	0.5708	0.41	0.33	0.4298	0.2014	0.4109	0.1601	0.33
3	-0.5592	-1.0721	0.6282	0.4953	0.4067	0.4629	0.2967	0.4258	0.1601	0.2829
4	-0.5748	-0.9652	0.6199	0.3439	0.4081	0.3933	0.34	0.4399	0.1373	0.2504
5	-0.4643	-0.8436	0.4676	0.258	0.3575	0.3477	0.2907	0.3729	0.12	0.2236
7	-0.2502	-0.6029	-	-	0.271	0.2771	0.2176	0.2567	0.0757	0.1157

Table 7. Models to estimate τ and $\varphi^{G/B}$ from the ergodic GMPE as functions of M_w and R_{rup}

Model number	Compo-nents	Functional form
1	$\varphi^{G/B}$	$\varphi^{G/B}(T) = s_0^{G/B}(T)$
1	τ	$\tau(T) = s_1(T)$
2	$\varphi^{G/B}$	$\varphi^{G/B}(T, M_w) =$ $\begin{cases} s_2^{G/B}(T) & for\ \ 4 \leq M_w \leq 4. \\[2mm] s_2^{G/B}(T) - \left(s_2^{G/B}(T) - s_3^{G/B}(T)\right)\frac{M_W - 4.7}{2.3} & for\ \ 4.7 > M_w \leq 7 \end{cases}$
2	τ	$\tau(T, M_w) = s_4(T)$
3	$\varphi^{G/B}$	$\varphi^{G/B}(T, R_{rup}) = \begin{cases} s_8^{G/B}(T) & for\ \ R_{rup} \leq 30 \\[2mm] s_8^{G/B}(T) - \left(s_8^{G/B}(T) - s_9^{G/B}(T)\right)\frac{log(R_{rup}) - (30)}{log(350) - log(30)} & for\ \ 30 > R_{rup} \leq 350 \end{cases}$
3	τ	$\tau(T, M_w) = s_{10}(T)$
4	$\varphi^{G/B}$	$\varphi^{G/B}(T, M_w, R_{rup}) =$ $\begin{cases} M1R1^{G/B}(T) & for\ \ \ R_{rup} \leq 30\ and\ 4 \leq M_w \leq 4.7 \\[3mm] M1R1^{G/B}(T) - (M1R1^{G/B}(T) - M1R3^{G/B}(T))\frac{log(R_{rup})}{log(350)} & for\ \ 30 > R_{rup} \leq 350\ and\ 4 \leq M_w \leq 4.7 \\[3mm] M1R1^{G/B}(T) + (M2R1^{G/B}(T) - M1R1^{G/B}(T))\frac{M_W - 4.7}{2.3} & for\ \ \ R_{rup} \leq 30\ and\ 4.7 < M_w \leq 7 \\[3mm] (M1R1^{G/B}(T) - (M1R1^{G/B}(T) - M1R3^{G/B}(T))\frac{log(R_{rup})}{log(350)})(1 - \frac{M_W - 4.7}{2.3}) + \\[3mm] (M2R1^{G/B}(T) - (M2R1^{G/B}(T) - M2R3^{G/B}(T))\frac{log(R_{rup})}{log(350)})\frac{M_W - 4.7}{2.3} & for\ \ 30 > R_{rup} \leq 350\ and\ 4.7 < M_w \leq 7 \end{cases}$

Table 8. The regression coefficients used in the models shown in Table 7

T (s.)	Model (1)			Model (2)				
	s_0^G	s_0^B	s_1	s_2^G	s_2^B	s_3^G	s_3^B	s_4
PGA	0.6826 [0.0042]	0.5812 [0.0036]	0.4507 [0.0132]	0.6921 [0.0047]	0.5975 [0.0040]	0.6055 [0.0163]	0.4348 [0.0125]	0.4503 [0.0132]
0.01	0.6826 [0.0042]	0.5812 [0.0036]	0.4508 [0.0132]	0.6921 [0.0047]	0.5975 [0.0040]	0.6055 [0.0163]	0.4349 [0.0125]	0.4504 [0.0132]
0.02	0.6819 [0.0042]	0.5823 [0.0036]	0.4541 [0.0133]	0.6914 [0.0047]	0.5986 [0.0040]	0.6050 [0.0163]	0.4358 [0.0125]	0.4536 [0.0133]
0.03	0.6770 [0.0041]	0.5931 [0.0036]	0.4701 [0.0137]	0.6865 [0.0046]	0.6103 [0.0041]	0.5997 [0.0161]	0.4391 [0.0126]	0.4697 [0.0137]
0.04	0.6749 [0.0041]	0.6046 [0.0037]	0.4989 [0.0145]	0.6842 [0.0046]	0.6216 [0.0042]	0.5989 [0.0160]	0.4537 [0.0129]	0.4984 [0.0145]
0.05	0.6758 [0.0041]	0.6198 [0.0038]	0.5131 [0.0149]	0.6845 [0.0046]	0.6369 [0.0043]	0.6044 [0.0161]	0.4675 [0.0131]	0.5125 [0.0149]
0.075	0.7462 [0.0047]	0.6183 [0.0039]	0.5274 [0.0153]	0.7608 [0.0053]	0.6361 [0.0044]	0.6249 [0.0168]	0.4652 [0.0130]	0.5270 [0.0153]
0.1	0.7690 [0.0047]	0.6083 [0.0037]	0.4951 [0.0145]	0.7815 [0.0053]	0.6249 [0.0042]	0.6637 [0.0174]	0.4640 [0.0129]	0.4946 [0.0145]
0.15	0.7759 [0.0048]	0.6163 [0.0038]	0.4558 [0.0135]	0.7854 [0.0053]	0.6304 [0.0043]	0.6984 [0.0186]	0.4920 [0.0138]	0.4555 [0.0136]
0.2	0.7749 [0.0048]	0.6218 [0.0039]	0.4555 [0.0136]	0.7830 [0.0054]	0.6345 [0.0044]	0.7123 [0.0193]	0.5096 [0.0146]	0.4555 [0.0136]
0.25	0.7658 [0.0048]	0.6223 [0.0039]	0.4475 [0.0133]	0.7723 [0.0054]	0.6334 [0.0044]	0.7182 [0.0197]	0.5248 [0.0152]	0.4474 [0.0134]
0.3	0.7130 [0.0044]	0.6094 [0.0038]	0.4300 [0.0128]	0.7182 [0.0049]	0.6181 [0.0042]	0.6725 [0.0180]	0.5339 [0.0148]	0.4298 [0.0128]
0.4	0.6592 [0.0040]	0.5874 [0.0036]	0.4105 [0.0122]	0.6636 [0.0045]	0.5920 [0.0041]	0.6227 [0.0164]	0.5485 [0.0145]	0.4102 [0.0122]
0.5	0.6197 [0.0038]	0.5714 [0.0035]	0.4063 [0.0120]	0.6199 [0.0042]	0.5719 [0.0039]	0.6185 [0.0161]	0.5670 [0.0149]	0.4063 [0.0120]
0.75	0.5742 [0.0036]	0.5262 [0.0033]	0.3898 [0.0116]	0.5679 [0.0041]	0.5208 [0.0037]	0.6214 [0.0161]	0.5675 [0.0149]	0.3901 [0.0116]
1	0.5570 [0.0035]	0.5105 [0.0032]	0.3751 [0.0111]	0.5467 [0.0039]	0.5004 [0.0036]	0.6343 [0.0162]	0.5867 [0.0153]	0.3754 [0.0111]
1.5	0.5411 [0.0045]	0.5038 [0.0042]	0.3740 [0.0131]	0.5155 [0.0052]	0.4809 [0.0049]	0.6629 [0.0180]	0.6127 [0.0171]	0.3745 [0.0130]
2	0.5181 [0.0042]	0.4855 [0.0040]	0.3466 [0.0122]	0.4995 [0.0050]	0.4684 [0.0047]	0.6077 [0.0164]	0.5679 [0.0158]	0.3468 [0.0122]
3	0.4868 [0.0066]	0.4574 [0.0062]	0.3503 [0.0184]	0.4465 [0.0087]	0.4163 [0.0083]	0.5853 [0.0197]	0.5572 [0.0191]	0.3514 [0.0183]
4	0.4413 [0.0103]	0.4330 [0.0102]	0.3453 [0.0317]	0.3493 [0.0153]	0.3522 [0.0153]	0.5407 [0.0224]	0.5209 [0.0218]	0.3469 [0.0314]
5	0.4254 [0.0100]	0.4163 [0.0099]	0.3168 [0.0298]	0.3401 [0.0152]	0.3316 [0.015]	0.5159 [0.0216]	0.5060 [0.0213]	0.3184 [0.0295]
7	0.4018 [0.0134]	0.3949 [0.0132]	0.3281 [0.0459]	0.2407 [0.0289]	0.2128 [0.0258]	0.4895 [0.0259]	0.4913 [0.0248]	0.3303 [0.0453]

Table 8 (Cont'd). The regression coefficients used in the models shown in Table 7

T (s.)	Model (3)				
	s_8^G	s_8^B	s_9^G	s_9^B	s_{10}
PGA	0.7213 [0.0102]	0.6935 [0.0091]	0.6363 [0.0119]	0.4364 [0.0096]	0.448 [0.0132]
0.01	0.7214 [0.0102]	0.6936 [0.0091]	0.6362 [0.0119]	0.4363 [0.0096]	0.448 [0.0132]
0.02	0.7201 [0.0102]	0.6969 [0.0091]	0.6363 [0.0119]	0.4345 [0.0096]	0.451 [0.0133]
0.03	0.7124 [0.0102]	0.7169 [0.0092]	0.635 [0.0118]	0.433 [0.0096]	0.4663 [0.0137]
0.04	0.7067 [0.0101]	0.7263 [0.0094]	0.6368 [0.0117]	0.4479 [0.0098]	0.4947 [0.0144]
0.05	0.7023 [0.01]	0.7441 [0.0097]	0.6439 [0.0117]	0.4602 [0.0102]	0.5087 [0.0148]
0.075	0.7458 [0.0113]	0.725 [0.0097]	0.747 [0.0135]	0.4828 [0.0103]	0.5235 [0.0152]
0.1	0.7912 [0.0115]	0.7053 [0.0094]	0.7412 [0.0135]	0.4855 [0.0101]	0.4929 [0.0144]
0.15	0.835 [0.0115]	0.7012 [0.0095]	0.7027 [0.0132]	0.5084 [0.0105]	0.4545 [0.0135]
0.2	0.8383 [0.0114]	0.6903 [0.0096]	0.6964 [0.013]	0.5346 [0.0108]	0.4542 [0.0136]
0.25	0.8222 [0.0111]	0.6839 [0.0095]	0.6964 [0.0128]	0.5441 [0.0108]	0.4463 [0.0133]
0.3	0.7788 [0.0104]	0.665 [0.0092]	0.6307 [0.0117]	0.5395 [0.0106]	0.4287 [0.0128]
0.4	0.7298 [0.0096]	0.6246 [0.0087]	0.57 [0.0107]	0.541 [0.01]	0.4104 [0.0122]
0.5	0.6739 [0.0089]	0.5973 [0.0083]	0.5516 [0.01]	0.5391 [0.0097]	0.4063 [0.012]
0.75	0.5999 [0.0082]	0.528 [0.0077]	0.5422 [0.0094]	0.524 [0.0092]	0.3899 [0.0116]
1	0.5736 [0.0078]	0.4999 [0.0074]	0.5363 [0.0091]	0.5236 [0.0089]	0.3752 [0.0111]
1.5	0.5326 [0.0093]	0.4766 [0.0084]	0.5506 [0.0109]	0.5368 [0.0104]	0.3732 [0.013]
2	0.5204 [0.0088]	0.4784 [0.0084]	0.5154 [0.0101]	0.494 [0.0098]	0.3466 [0.0122]
3	0.4657 [0.0128]	0.4386 [0.0121]	0.5115 [0.0153]	0.48 [0.0146]	0.3503 [0.0184]
4	0.4021 [0.0182]	0.4076 [0.018]	0.4858 [0.0225]	0.4626 [0.0217]	0.3462 [0.0317]
5	0.4036 [0.0182]	0.3947 [0.0178]	0.4503 [0.0216]	0.4414 [0.0213]	0.3169 [0.0298]
7	0.3268 [0.0272]	0.3204 [0.0256]	0.4718 [0.031]	0.4647 [0.0295]	0.3294 [0.0458]

Table 8 (Cont'd). The regression coefficients used in the models shown in Table 7

T (s.)	Model (4)							
	M1R1^G	M1R3^G	M2R1^G	M2R3^G	M1R1^B	M1R3^B	M2R1^B	M2R3^B
PGA	0.7330 [0.0116]	0.6378 [0.0148]	0.4984 [0.0417]	0.6924 [0.0342]	0.7105 [0.0104]	0.4357 [0.0125]	0.3800 [0.0351]	0.5225 [0.0282]
0.01	0.7329 [0.0116]	0.6378 [0.0148]	0.4988 [0.0418]	0.6921 [0.0342]	0.7105 [0.0104]	0.4355 [0.0125]	0.3801 [0.0351]	0.5225 [0.0282]
0.02	0.7311 [0.0115]	0.6385 [0.0148]	0.5013 [0.0420]	0.6893 [0.0342]	0.7144 [0.0104]	0.4327 [0.0125]	0.3817 [0.0354]	0.5239 [0.0282]
0.03	0.7199 [0.0114]	0.6423 [0.0147]	0.5214 [0.0432]	0.6640 [0.0343]	0.7398 [0.0107]	0.4241 [0.0127]	0.3611 [0.0353]	0.5511 [0.0287]
0.04	0.7119 [0.0112]	0.6470 [0.0146]	0.5412 [0.0437]	0.6481 [0.0341]	0.7482 [0.0109]	0.4405 [0.0129]	0.3941 [0.0361]	0.5502 [0.0286]
0.05	0.7104 [0.0112]	0.6499 [0.0145]	0.5145 [0.0424]	0.6757 [0.0342]	0.7603 [0.0112]	0.4615 [0.0132]	0.4541 [0.0384]	0.5253 [0.0283]
0.075	0.7364 [0.0122]	0.7945 [0.0164]	0.6016 [0.0496]	0.6267 [0.0366]	0.7330 [0.0109]	0.4980 [0.0132]	0.4597 [0.0378]	0.5088 [0.0286]
0.1	0.7934 [0.0127]	0.7650 [0.0167]	0.5826 [0.0493]	0.7220 [0.0386]	0.7140 [0.0106]	0.4985 [0.0129]	0.4439 [0.0376]	0.5175 [0.0290]
0.15	0.8546 [0.0133]	0.6897 [0.0168]	0.5724 [0.0480]	0.8146 [0.0399]	0.7013 [0.0106]	0.5290 [0.0132]	0.5441 [0.0440]	0.4876 [0.0313]
0.2	0.8614 [0.0131]	0.6738 [0.0165]	0.5866 [0.0465]	0.8331 [0.0396]	0.6873 [0.0105]	0.5586 [0.0134]	0.5748 [0.0469]	0.4879 [0.0323]
0.25	0.8504 [0.0128]	0.6639 [0.0162]	0.5457 [0.0446]	0.8677 [0.0395]	0.6789 [0.0104]	0.5675 [0.0133]	0.6144 [0.0508]	0.4836 [0.0335]
0.3	0.7996 [0.0120]	0.6038 [0.0150]	0.6007 [0.0450]	0.7579 [0.0368]	0.6586 [0.0101]	0.5598 [0.0130]	0.6471 [0.0496]	0.4750 [0.0328]
0.4	0.7438 [0.0112]	0.5491 [0.0138]	0.6341 [0.0456]	0.6535 [0.0334]	0.6244 [0.0097]	0.5460 [0.0125]	0.5964 [0.0446]	0.5319 [0.0318]
0.5	0.6904 [0.0104]	0.5192 [0.0129]	0.6055 [0.0441]	0.6643 [0.0332]	0.5976 [0.0093]	0.5348 [0.0121]	0.6175 [0.0447]	0.5463 [0.0321]
0.75	0.6087 [0.0094]	0.5092 [0.0120]	0.6402 [0.0438]	0.6326 [0.0324]	0.5155 [0.0085]	0.5267 [0.0113]	0.7132 [0.0454]	0.4651 [0.0299]
1	0.5753 [0.0089]	0.5050 [0.0116]	0.7072 [0.0454]	0.6026 [0.0327]	0.4799 [0.0080]	0.5274 [0.0109]	0.7872 [0.0470]	0.4357 [0.0302]
1.5	0.5249 [0.0107]	0.4994 [0.0145]	0.7464 [0.0493]	0.6160 [0.0346]	0.4579 [0.0095]	0.5131 [0.0136]	0.7234 [0.0476]	0.5234 [0.0330]
2	0.5224 [0.0103]	0.4645 [0.0137]	0.6346 [0.0435]	0.6052 [0.0322]	0.4656 [0.0098]	0.4703 [0.0133]	0.6558 [0.0425]	0.5084 [0.0293]
3	0.4736 [0.0170]	0.4061 [0.0231]	0.5274 [0.0463]	0.6391 [0.0399]	0.4344 [0.0164]	0.3872 [0.0227]	0.5373 [0.0462]	0.5797 [0.0375]
4	0.4017 [0.0270]	0.2502 [0.0401]	0.4646 [0.0471]	0.6265 [0.0469]	0.4129 [0.0279]	0.2341 [0.0402]	0.4631 [0.0455]	0.5926 [0.0437]
5	0.4086 [0.0274]	0.2079 [0.0369]	0.4507 [0.0482]	0.5976 [0.0454]	0.3824 [0.0272]	0.2341 [0.0374]	0.4704 [0.0478]	0.5519 [0.0425]
7	0.2560 [0.0564]	0.2364 [0.0878]	0.4238 [0.0641]	0.5371 [0.0590]	0.2216 [0.0528]	0.2158 [0.0811]	0.4354 [0.0599]	0.5314 [0.0547]

Table 9. Models to estimate τ and $\varphi_{SS}^{G/B}$ from the site-specific GMPE as functions of M_w and R_{rup}

Model number	Components	Functional form
1	$\varphi_{SS}^{G/B}$ or $\varphi_{G/B,0}$	$\varphi_{SS}^{G/B}(T) = s_0^{G/B}(T)$
2	$\varphi_{SS}^{G/B}$ or $\varphi_{G/B,0}$	$\varphi_{SS}^{G/B}(T, M_w) =$ $$\begin{cases} s_2^{G/B}(T) & for \quad 4 \le M_w \le 5 \\ s_2^{G/B}(T) - \left(s_2^{G/B}(T) - s_3^{G/B}(T)\right)\frac{M_W-5}{2} & for \quad 5 > M_w \le 7 \end{cases}$$
3	$\varphi_{SS}^{G/B}$ or $\varphi_{G/B,0}$	$\varphi_{SS}^{G/B}(T, R_{rup}) = \begin{cases} s_7^{G/B}(T) & for \quad R_{rup} \le 30 \\ s_7^{G/B}(T) - \left(s_7^{G/B}(T) - s_8^{G/B}(T)\right)\frac{log(R_{rup})-log(30)}{log(100)-log(30)} & for \quad 30 > R_{rup} \le 100 \\ s_8^{G/B}(T) & for \quad 100 > R_{rup} \le 350 \end{cases}$
4	$\varphi_{SS}^{G/B}$ or $\varphi_{G/B,0}$	$\varphi_{SS}^{G/B}(T, M_w, R_{rup}) =$ $$\begin{cases} M1R1^{G/B}(T) & for \quad R_{rup} \le 30 \text{ and } 4 \le M_w \le 5 \\ M1R1^{G/B}(T) - (M1R1^{G/B}(T) - M1R3^{G/B}(T))\frac{log(R_{rup})-log(30)}{log(100)-log(30)} & for \quad 30 > R_{rup} \le 100 \text{ and } 4 \le M_w \le 5 \\ M1R3^{G/B}(T) & for \quad 100 > R_{rup} \le 350 \text{ and } 4 \le M_w \le 5 \\ M1R1^{G/B}(T) + (M2R1^{G/B}(T) - M1R1^{G/B}(T))\frac{M_W-5}{2} & for \quad R_{rup} \le 30 \text{ and } 5 < M_w \le 7 \\ (M1R1^{G/B}(T) - (M1R1^{G/B}(T) - M1R3^{G/B}(T))\frac{log(R_{rup})-log(30)}{log(100)-log(30)})(1 - \frac{M_W-5}{2}) + \\ \quad (M2R1^{G/B}(T) - (M2R1^{G/B}(T) - M2R3^{G/B}(T))\frac{log(R_{rup})-log(30)}{log(100)-log(30)})\frac{M_W-5}{2} & for \quad 30 > R_{rup} \le 100 \text{ and } 5 < M_w \le 7 \\ M1R3^{G/B}(T) + (M2R3^{G/B}(T) - M1R3^{G/B}(T))\frac{M_W-5}{2} & for \quad 100 > R_{rup} \le 350 \text{ and } 5 < M_w \le 7 \end{cases}$$
-	τ	$\tau(T) = s_1(T)$

Table 10. The regression coefficients used in the models shown in Table 9

T (s.)	Model (1)		Tau	Model (2)			
	s_0^G	s_0^B	s_1	s_2^G	s_2^B	s_3^G	s_3^B
PGA	0.5041 [0.0030]	0.4877 [0.0029]	0.4039 [0.0017]	0.5062 [0.0033]	0.4939 [0.0032]	0.4794 [0.0135]	0.4119 [0.0120]
0.01	0.5041 [0.0030]	0.4877 [0.0029]	0.4041 [0.0017]	0.5062 [0.0033]	0.4939 [0.0032]	0.4794 [0.0135]	0.4120 [0.0120]
0.02	0.5037 [0.0030]	0.4874 [0.0029]	0.4081 [0.0017]	0.5058 [0.0033]	0.4935 [0.0032]	0.4796 [0.0135]	0.4129 [0.0120]
0.03	0.5002 [0.0030]	0.4812 [0.0029]	0.4159 [0.0018]	0.5023 [0.0032]	0.4862 [0.0031]	0.4751 [0.0134]	0.4195 [0.0122]
0.04	0.4944 [0.0030]	0.4743 [0.0029]	0.4308 [0.0018]	0.4958 [0.0032]	0.4782 [0.0031]	0.4782 [0.0135]	0.4265 [0.0123]
0.05	0.4903 [0.0030]	0.4743 [0.0029]	0.4396 [0.0019]	0.4911 [0.0032]	0.4781 [0.0031]	0.4811 [0.0135]	0.4281 [0.0123]
0.075	0.4997 [0.0030]	0.4813 [0.0029]	0.4453 [0.0019]	0.5002 [0.0032]	0.4869 [0.0031]	0.4930 [0.0138]	0.4138 [0.0121]
0.1	0.5091 [0.0031]	0.4970 [0.0030]	0.4326 [0.0018]	0.5103 [0.0033]	0.5042 [0.0032]	0.4952 [0.0139]	0.4072 [0.0120]
0.15	0.5320 [0.0032]	0.5198 [0.0031]	0.409 [0.0017]	0.5352 [0.0035]	0.5277 [0.0034]	0.4942 [0.0140]	0.4224 [0.0125]
0.2	0.5387 [0.0033]	0.5291 [0.0032]	0.4086 [0.0017]	0.5426 [0.0035]	0.5376 [0.0035]	0.4921 [0.0140]	0.4238 [0.0126]
0.25	0.5370 [0.0032]	0.5270 [0.0032]	0.3992 [0.0017]	0.5411 [0.0035]	0.5349 [0.0034]	0.4873 [0.0139]	0.4298 [0.0127]
0.3	0.5296 [0.0032]	0.5189 [0.0031]	0.3952 [0.0017]	0.5327 [0.0034]	0.5249 [0.0034]	0.4926 [0.0140]	0.4459 [0.0130]
0.4	0.5104 [0.0031]	0.5039 [0.0030]	0.3916 [0.0017]	0.5112 [0.0033]	0.5073 [0.0033]	0.5009 [0.0143]	0.4628 [0.0134]
0.5	0.4974 [0.0030]	0.4880 [0.0029]	0.3856 [0.0016]	0.4961 [0.0032]	0.4882 [0.0032]	0.5122 [0.0145]	0.4858 [0.0138]
0.75	0.4635 [0.0029]	0.4512 [0.0028]	0.3628 [0.0016]	0.4608 [0.0031]	0.4478 [0.0030]	0.4927 [0.0138]	0.4876 [0.0135]
1	0.4478 [0.0028]	0.4341 [0.0027]	0.351 [0.0015]	0.4425 [0.0030]	0.4276 [0.0029]	0.5037 [0.0139]	0.5031 [0.0139]
1.5	0.4097 [0.0033]	0.3905 [0.0031]	0.3517 [0.0017]	0.4019 [0.0036]	0.3771 [0.0034]	0.4635 [0.0134]	0.4804 [0.0137]
2	0.3941 [0.0032]	0.3655 [0.0029]	0.3326 [0.0016]	0.3892 [0.0035]	0.3544 [0.0032]	0.4291 [0.0128]	0.4410 [0.0129]
3	0.3410 [0.0046]	0.3199 [0.0043]	0.3315 [0.0021]	0.3349 [0.0055]	0.3045 [0.0051]	0.3625 [0.0132]	0.3731 [0.0137]
4	0.2578 [0.0059]	0.2652 [0.0061]	0.3499 [0.0035]	0.2262 [0.0080]	0.2344 [0.0083]	0.3033 [0.0131]	0.3100 [0.0135]
5	0.2606 [0.0061]	0.2591 [0.0060]	0.3281 [0.0034]	0.2336 [0.0084]	0.2287 [0.0082]	0.2987 [0.0130]	0.3018 [0.0131]
7	0.2448 [0.0081]	0.2488 [0.0082]	0.3429 [0.0045]	0.1828 [0.0159]	0.1685 [0.0146]	0.2885 [0.0161]	0.3041 [0.0161]

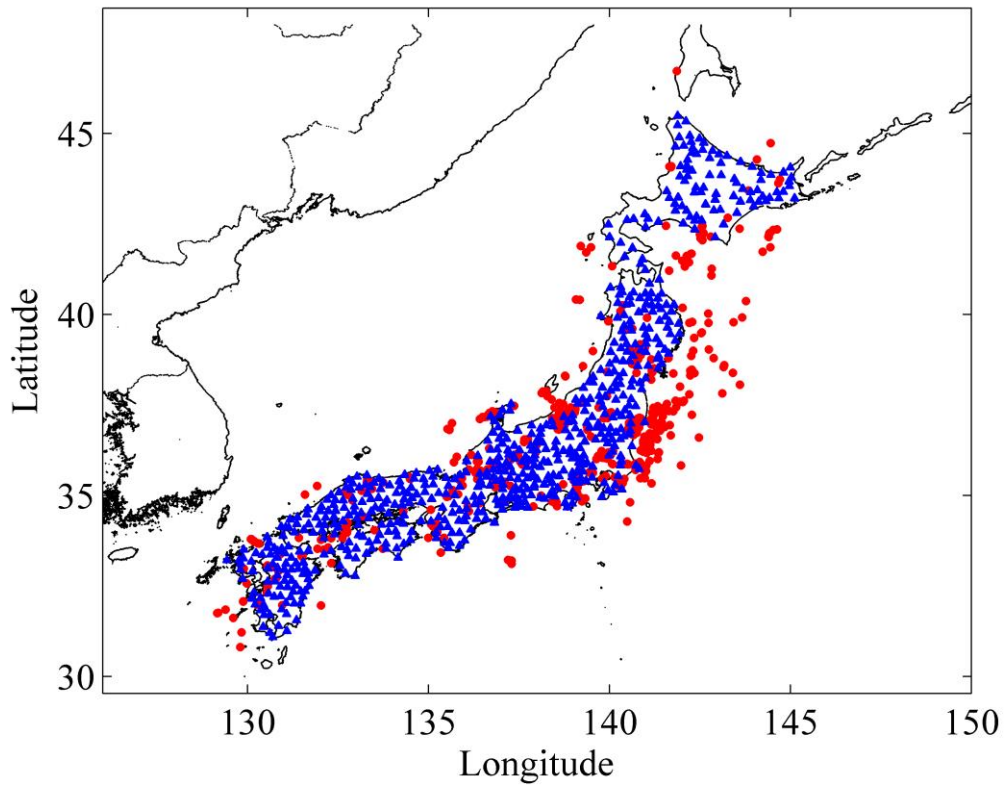**Table 10 (Cont'd).** The regression coefficients used in the models shown in Table 9

T (s.)	Model (3)			
	s_7^G	s_7^B	s_8^G	s_8^B
PGA	0.5602 [0.0087]	0.5829 [0.0090]	0.4872 [0.0036]	0.4583 [0.0035]
0.01	0.5602 [0.0087]	0.5830 [0.0090]	0.4872 [0.0036]	0.4583 [0.0035]
0.02	0.5588 [0.0087]	0.5825 [0.0090]	0.4872 [0.0036]	0.4579 [0.0035]
0.03	0.5484 [0.0085]	0.5714 [0.0087]	0.4857 [0.0036]	0.4532 [0.0034]
0.04	0.5354 [0.0083]	0.5585 [0.0085]	0.4822 [0.0036]	0.4482 [0.0034]
0.05	0.5263 [0.0081]	0.5558 [0.0085]	0.4796 [0.0035]	0.4491 [0.0034]
0.075	0.5388 [0.0084]	0.5651 [0.0087]	0.4880 [0.0036]	0.4556 [0.0034]
0.1	0.5613 [0.0087]	0.5877 [0.0091]	0.4934 [0.0037]	0.4690 [0.0035]
0.15	0.6163 [0.0095]	0.6180 [0.0096]	0.5063 [0.0038]	0.4896 [0.0037]
0.2	0.6237 [0.0097]	0.6239 [0.0098]	0.5127 [0.0039]	0.5001 [0.0038]
0.25	0.6181 [0.0096]	0.6092 [0.0096]	0.5122 [0.0039]	0.5020 [0.0038]
0.3	0.6068 [0.0095]	0.5878 [0.0093]	0.5061 [0.0038]	0.4981 [0.0038]
0.4	0.5793 [0.009]	0.5600 [0.0088]	0.4895 [0.0037]	0.4870 [0.0036]
0.5	0.5526 [0.0087]	0.5303 [0.0084]	0.4808 [0.0036]	0.4755 [0.0036]
0.75	0.509 [0.0082]	0.4687 [0.0077]	0.4498 [0.0035]	0.4460 [0.0034]
1	0.4878 [0.0079]	0.4365 [0.0071]	0.4358 [0.0034]	0.4334 [0.0033]
1.5	0.4301 [0.0086]	0.3840 [0.0076]	0.4031 [0.0040]	0.3925 [0.0039]
2	0.4282 [0.0085]	0.3652 [0.0073]	0.3830 [0.0038]	0.3655 [0.0036]
3	0.3653 [0.0114]	0.3259 [0.0104]	0.3325 [0.0055]	0.3179 [0.0053]
4	0.2866 [0.0149]	0.3007 [0.0157]	0.2476 [0.0070]	0.2527 [0.0072]
5	0.295 [0.0156]	0.2822 [0.0154]	0.2486 [0.0071]	0.2512 [0.0072]
7	0.2006 [0.0187]	0.1979 [0.0189]	0.2545 [0.0097]	0.2600 [0.0099]

**Table 10 (Cont'd).** The regression coefficients used in the models shown in Table 9

T (s.)	Model (4)							
	M1R1^G	M1R3^G	M2R1^G	M2R3^G	M1R1^B	M1R3^B	M2R1^B	M2R3^B
PGA	0.5684 [0.0091]	0.4855 [0.0040]	0.3932 [0.0384]	0.5001 [0.0151]	0.5926 [0.0094]	0.4606 [0.0038]	0.3239 [0.0366]	0.4364 [0.0136]
0.01	0.5684 [0.0091]	0.4855 [0.0040]	0.3934 [0.0384]	0.5001 [0.0151]	0.5927 [0.0094]	0.4605 [0.0038]	0.3248 [0.0365]	0.4364 [0.0136]
0.02	0.5669 [0.0091]	0.4855 [0.0040]	0.3942 [0.0385]	0.5001 [0.0151]	0.5923 [0.0094]	0.4600 [0.0038]	0.3243 [0.0365]	0.4373 [0.0136]
0.03	0.5555 [0.0089]	0.4847 [0.0040]	0.3995 [0.0389]	0.4931 [0.0150]	0.5825 [0.0092]	0.4535 [0.0037]	0.2961 [0.0346]	0.4488 [0.0139]
0.04	0.5418 [0.0087]	0.4807 [0.0039]	0.4070 [0.0392]	0.4946 [0.0150]	0.5695 [0.0090]	0.4472 [0.0037]	0.3003 [0.0352]	0.4555 [0.0140]
0.05	0.5338 [0.0085]	0.4770 [0.0039]	0.3818 [0.0373]	0.5006 [0.0150]	0.5666 [0.0089]	0.4483 [0.0037]	0.3022 [0.0365]	0.4554 [0.0139]
0.075	0.5456 [0.0088]	0.4853 [0.0039]	0.4078 [0.0407]	0.5108 [0.0154]	0.5735 [0.0091]	0.4579 [0.0038]	0.3465 [0.0383]	0.4331 [0.0135]
0.1	0.5684 [0.0091]	0.4911 [0.0040]	0.4254 [0.0425]	0.5122 [0.0154]	0.5959 [0.0096]	0.4736 [0.0039]	0.3504 [0.0383]	0.4254 [0.0134]
0.15	0.6234 [0.0100]	0.5058 [0.0042]	0.4709 [0.0459]	0.5085 [0.0155]	0.6247 [0.0101]	0.4952 [0.0041]	0.4057 [0.0417]	0.4370 [0.0140]
0.2	0.6313 [0.0102]	0.5130 [0.0042]	0.4557 [0.0461]	0.5079 [0.0155]	0.6319 [0.0102]	0.5062 [0.0042]	0.3782 [0.0397]	0.4414 [0.0141]
0.25	0.6263 [0.0101]	0.5128 [0.0042]	0.4312 [0.0446]	0.5053 [0.0155]	0.6172 [0.0100]	0.5076 [0.0042]	0.3618 [0.0400]	0.4490 [0.0143]
0.3	0.6119 [0.0099]	0.5064 [0.0042]	0.4944 [0.0482]	0.5024 [0.0155]	0.5937 [0.0097]	0.5023 [0.0041]	0.4151 [0.0424]	0.4587 [0.0145]
0.4	0.5825 [0.0094]	0.4875 [0.0040]	0.5313 [0.0491]	0.5066 [0.0157]	0.5644 [0.0092]	0.4886 [0.0040]	0.4488 [0.0454]	0.4719 [0.0148]
0.5	0.5583 [0.0091]	0.4754 [0.0039]	0.4742 [0.0472]	0.5270 [0.0162]	0.5334 [0.0088]	0.4735 [0.0039]	0.4777 [0.0476]	0.4930 [0.0153]
0.75	0.5086 [0.0086]	0.4449 [0.0038]	0.5534 [0.0501]	0.4908 [0.0152]	0.4638 [0.0080]	0.4426 [0.0037]	0.5867 [0.0531]	0.4754 [0.0147]
1	0.4839 [0.0083]	0.4288 [0.0037]	0.6096 [0.0535]	0.4939 [0.0152]	0.4257 [0.0074]	0.4281 [0.0036]	0.6610 [0.0565]	0.4801 [0.0148]
1.5	0.4213 [0.0090]	0.3946 [0.0046]	0.5832 [0.0531]	0.4488 [0.0146]	0.3690 [0.0079]	0.3799 [0.0044]	0.6191 [0.0559]	0.4605 [0.0146]
2	0.4222 [0.0090]	0.3764 [0.0044]	0.5402 [0.0505]	0.4185 [0.0138]	0.3481 [0.0076]	0.3563 [0.0041]	0.6199 [0.0561]	0.4162 [0.0136]
3	0.3611 [0.0128]	0.3225 [0.0071]	0.4121 [0.0467]	0.3587 [0.0146]	0.3070 [0.0115]	0.3021 [0.0067]	0.4834 [0.0553]	0.3583 [0.0147]
4	0.2568 [0.0167]	0.2059 [0.0103]	0.3912 [0.0461]	0.2921 [0.0145]	0.2684 [0.0181]	0.2116 [0.0108]	0.4159 [0.0497]	0.2960 [0.0148]
5	0.2868 [0.0186]	0.1983 [0.0100]	0.3438 [0.0478]	0.3002 [0.0145]	0.2614 [0.0181]	0.2075 [0.0104]	0.3712 [0.0499]	0.2949 [0.0144]
7	0.1956 [0.0295]	0.1770 [0.0226]	0.2177 [0.0414]	0.3014 [0.0197]	0.1567 [0.0270]	0.1774 [0.0208]	0.2634 [0.0459]	0.3073 [0.0187]

FIGURES

Figure 1. The distribution of the seismic events (red circles) in the current database and KiK-net recording stations (red triangles) on the Japanese map.
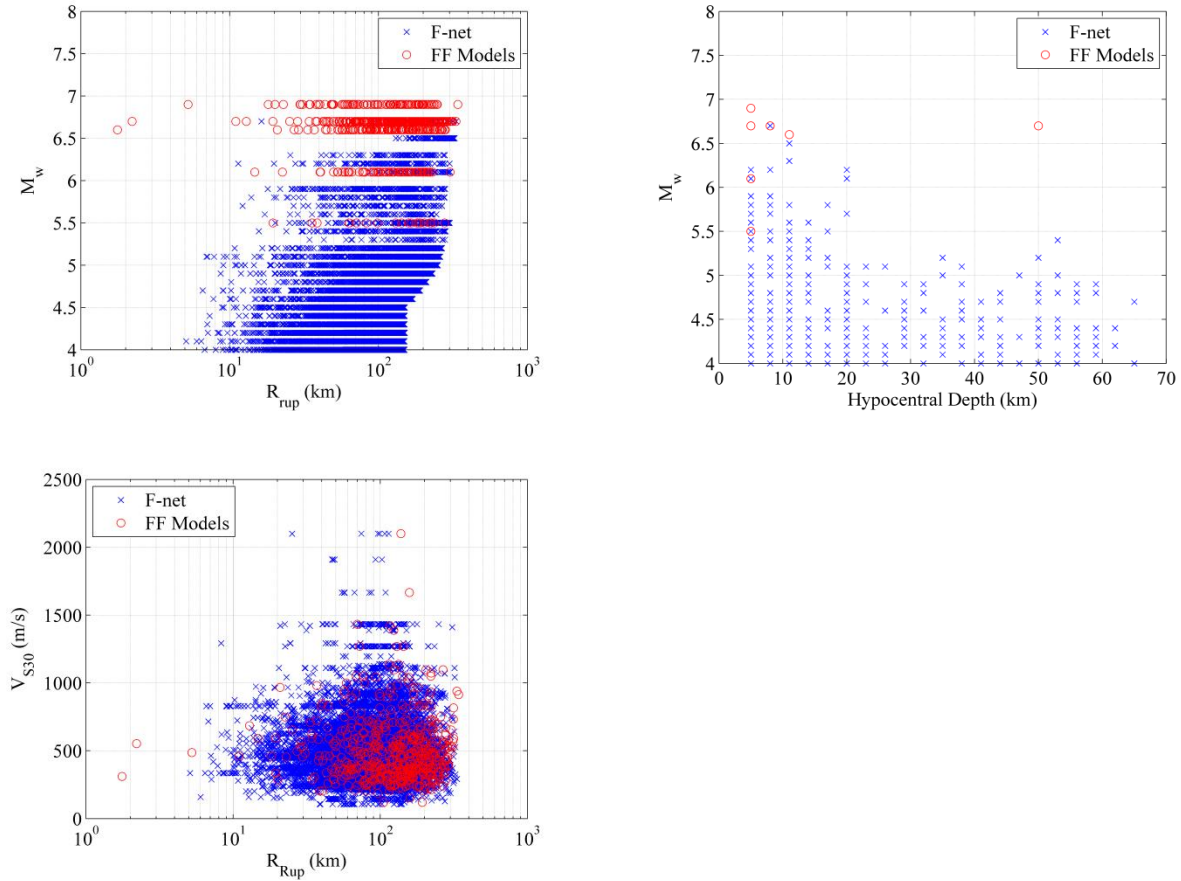
Figure 2. Scatter plots for M_w versus R_{rup} (top left panel), M_w versus hypocentral depth (top right panel), and R_{rup} versus V_{s30} (bottom left panel) for the database used in this study. Two different symbols had been used to differentiate between M_w, and R_{rup} obtained from the F-net catalog (blue crosses) and previously published finite fault source models (red circles) (see Dea14 for additional details)
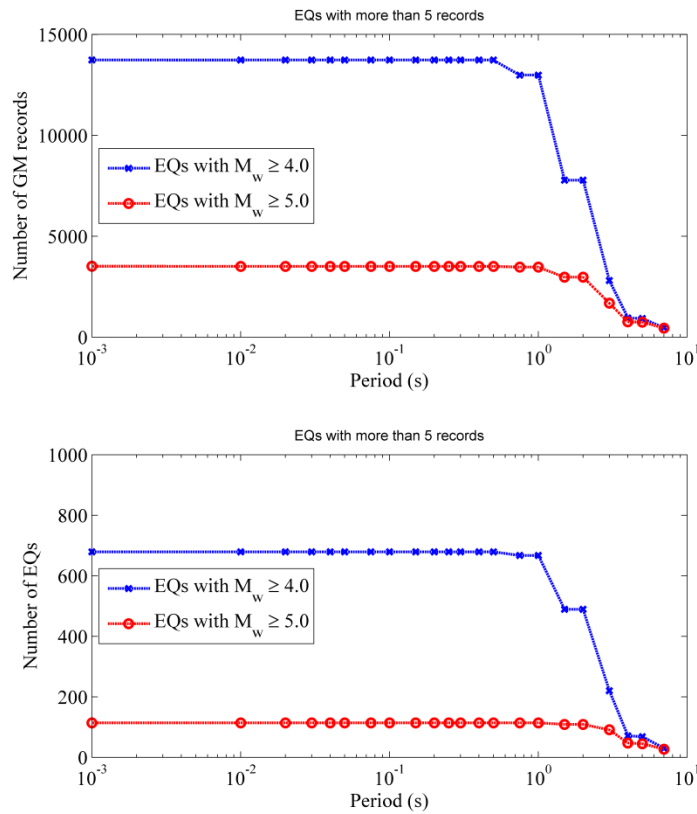
Figure 3. The number of motions (top panel) and
earthquakes (bottom panel) used in the regression analysis
at each spectral period. Two different lines are plotted in
each panel to show the number of earthquakes and motions
from all earthquakes (blue lines) and from earthquakes with
M_w greater than or equal to 5.0 (red lines)

47
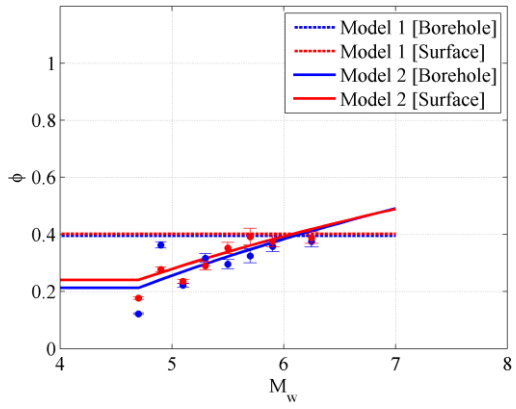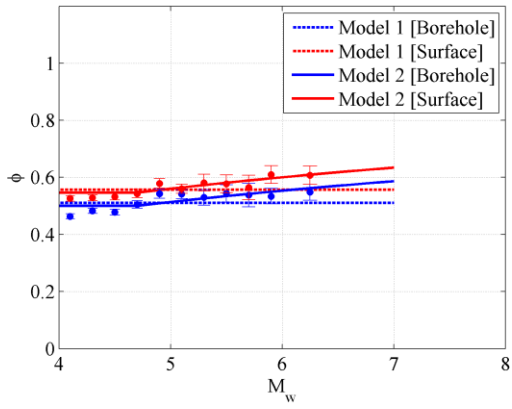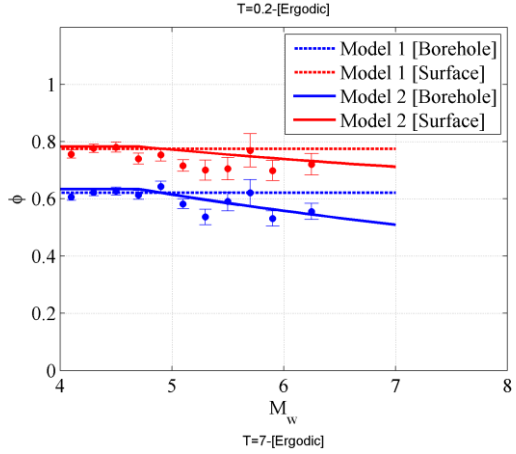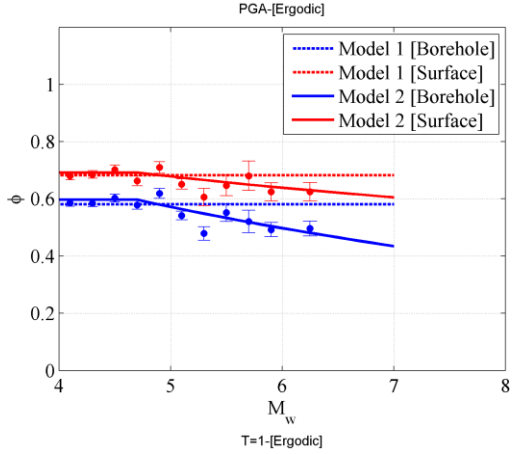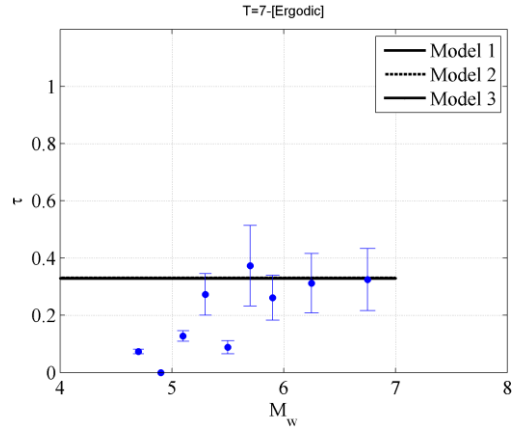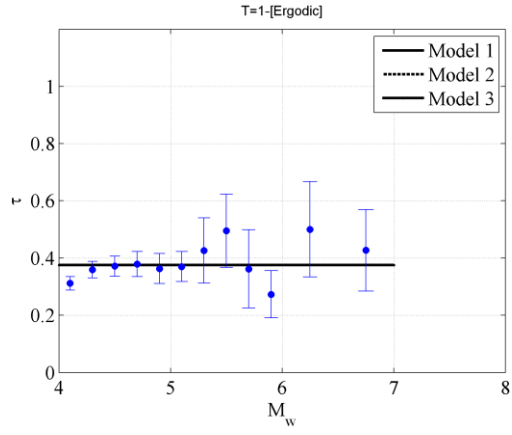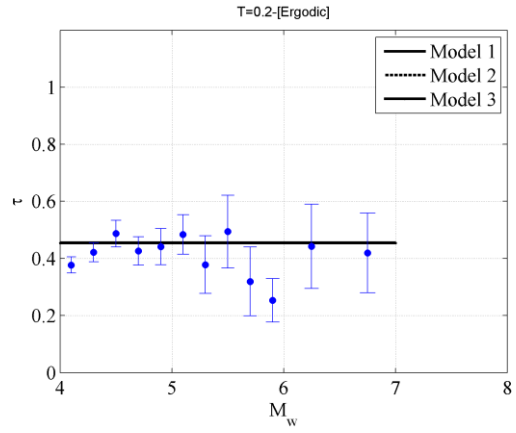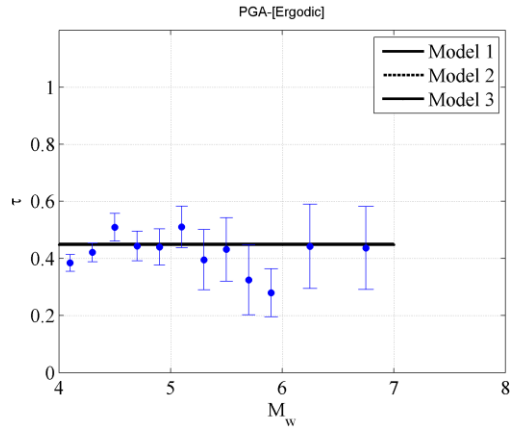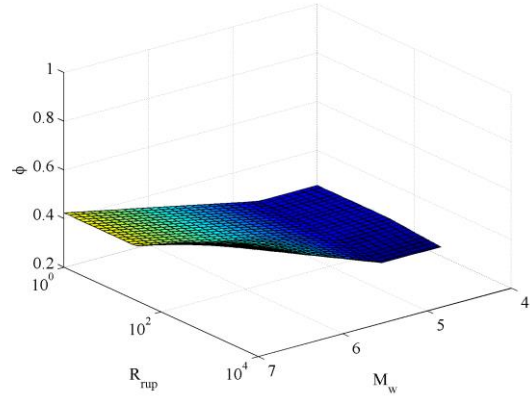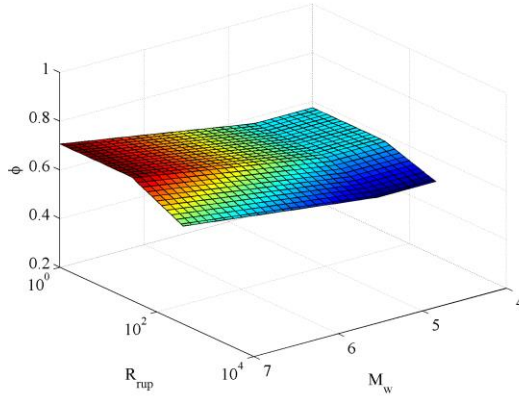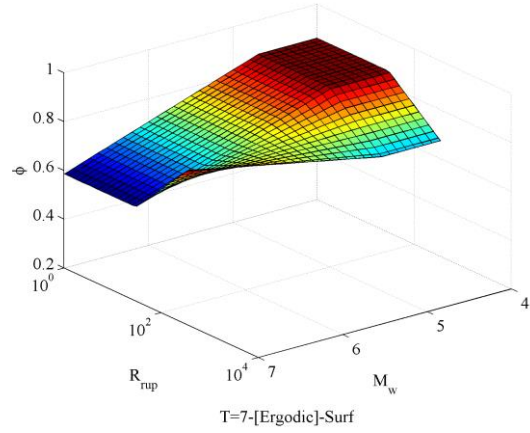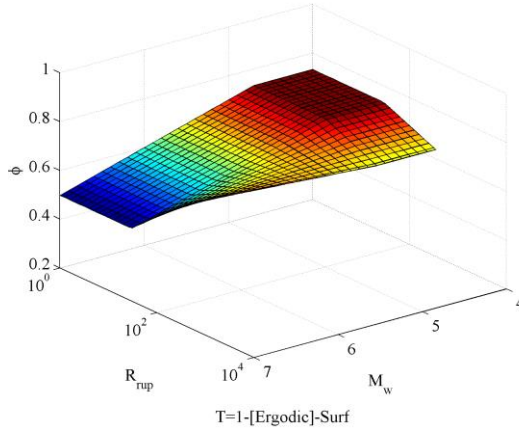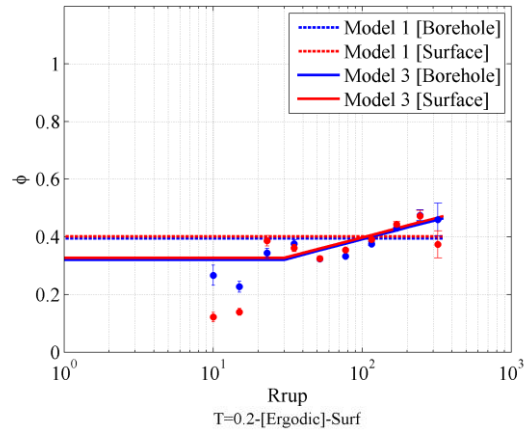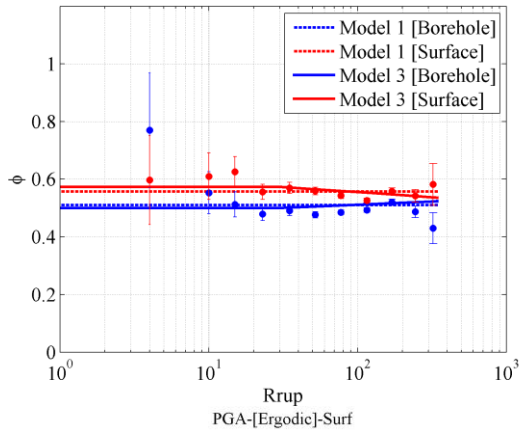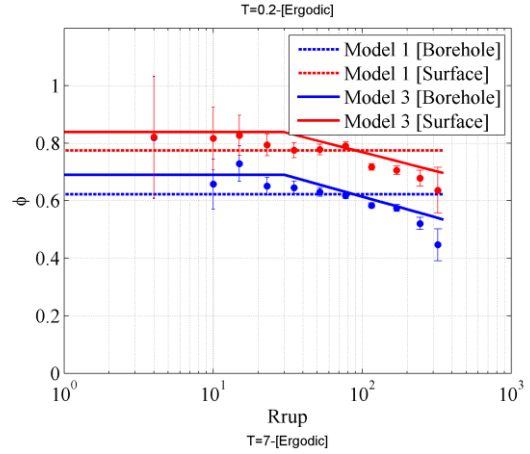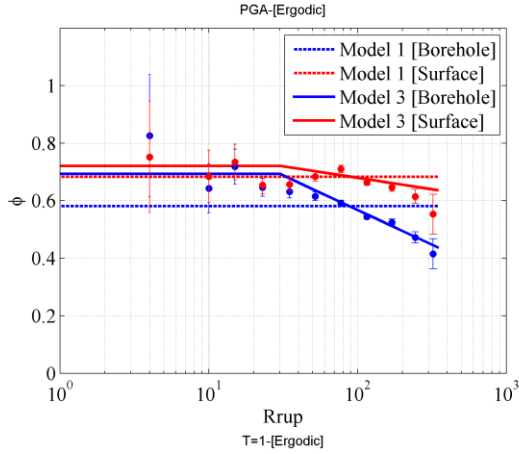
Figure 4. The behavior of the proposed models to predict τ and $\varphi^{G/B}$ of the Ergodic GMPE at four different spectral periods (PGA, 0.2s, 1.0s and 7.0s) for:
1) τ models [constant, top 4 panels],
2) $\varphi^{G/B}$ models [magnitude-dependent and constant, 5th to 8th panels],
3) $\varphi^{G/B}$ models [distance-dependent and constant, 9th to 12th panels], and
4) $\varphi^{G/B}$ models [distance and magnitude-dependent, bottom 4 panels]

The plots also show the standard deviation of $\delta W_{es}^{S/B}$, and δB_e [Equation 1] within selected R_{rup} or M_w bins to show the suitableness of the proposed models to predict τ and $\varphi^{G/B}$.
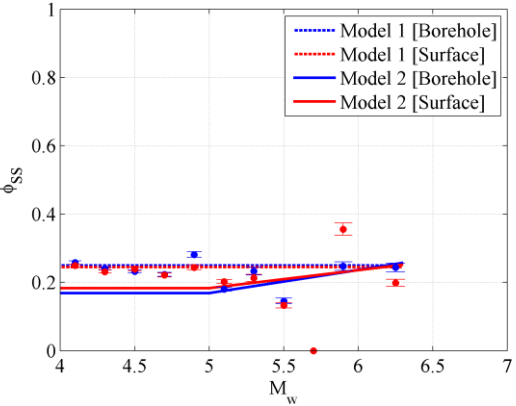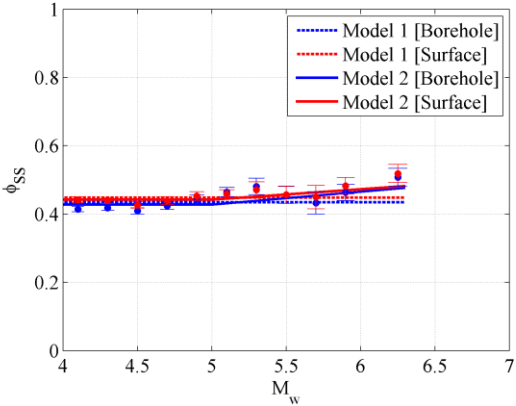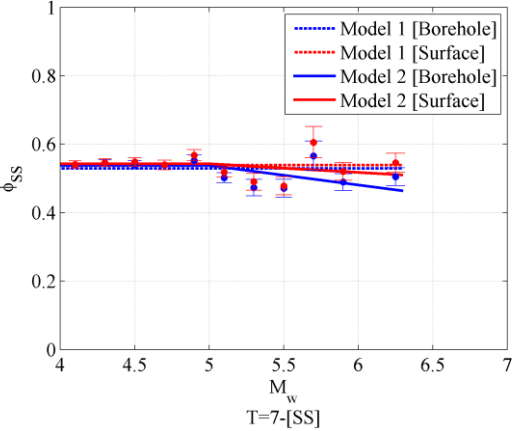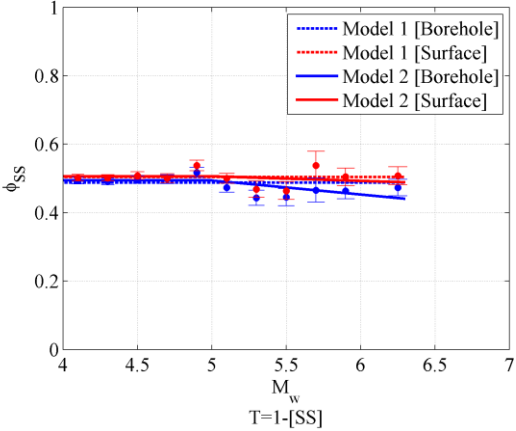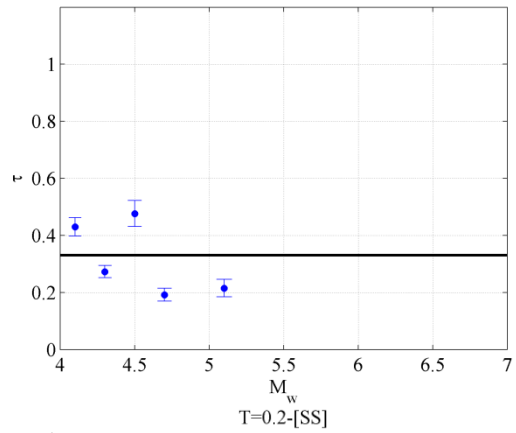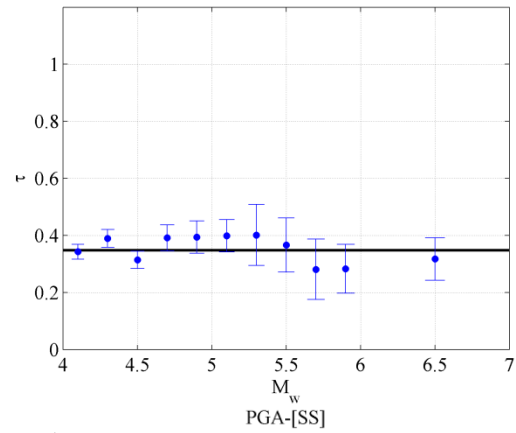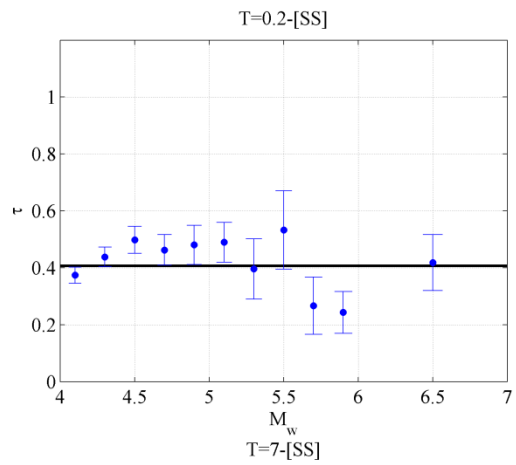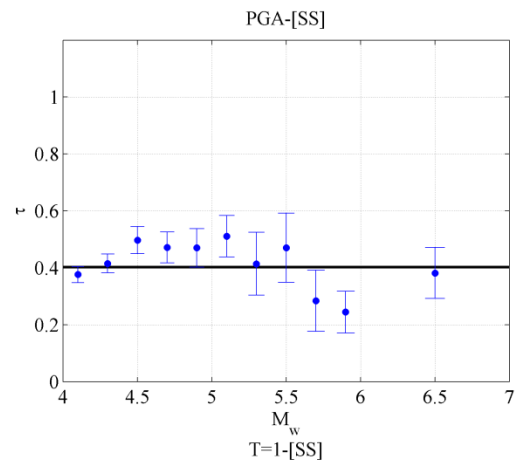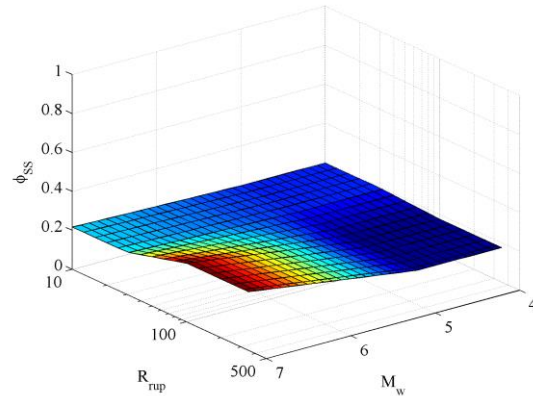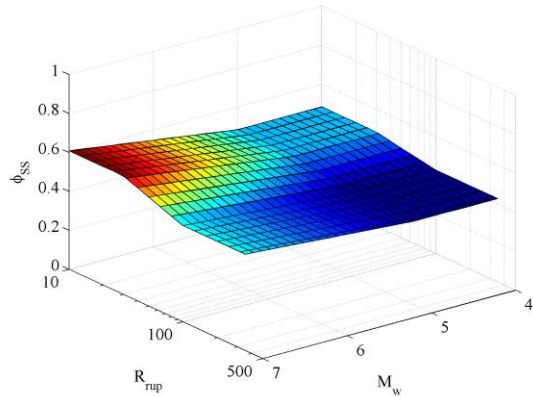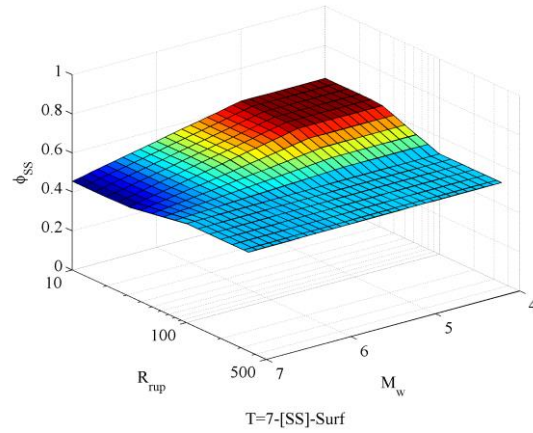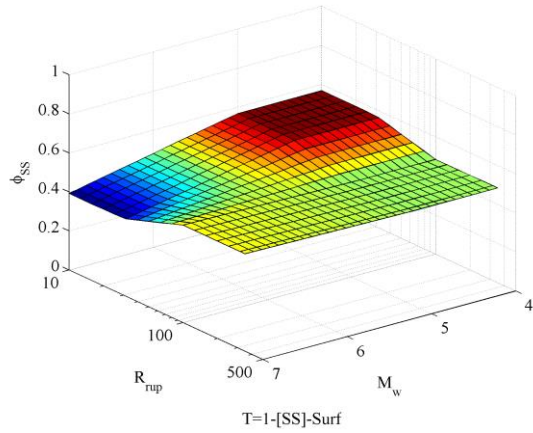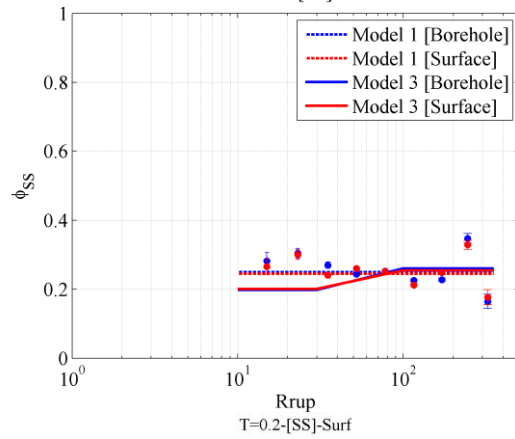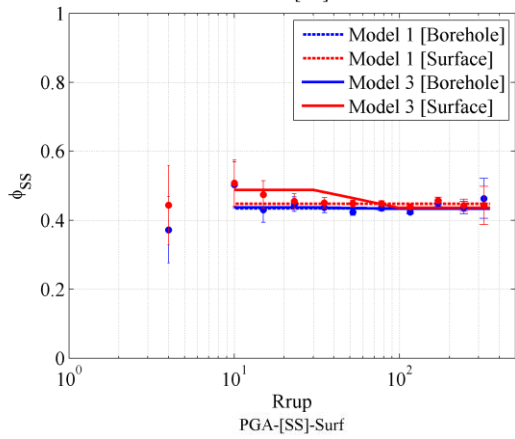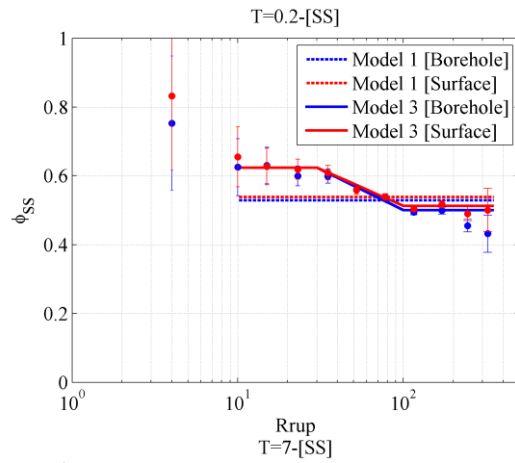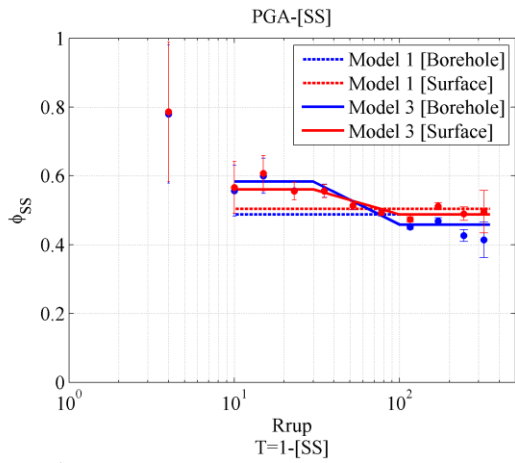
Figure 5. The behavior of the proposed models to predict τ and $\varphi_{SS}^{G/B}$ of the site-specific GMPE at four different spectral periods (PGA, 0.2s, 1.0s and 7.0s) for:

1) τ models [constant, top 4 panels],
2) $\varphi_{SS}^{G/B}$ models [magnitude-dependent and constant, 5th to 8th panels],
3) $\varphi_{SS}^{G/B}$ models [distance-dependent and constant, 9th to 12th panels], and
4) $\varphi_{SS}^{G/B}$ models [distance and magnitude-dependent, bottom 4 panels]

The plots also show the standard deviation of δW_{esb}^0, δW_{es}^0, and δB_e [Equations 1, 3 and 4] within selected R_{rup} or M_w bins to show the suitableness of the developed models to predict τ and $\varphi_{SS}^{G/B}$.
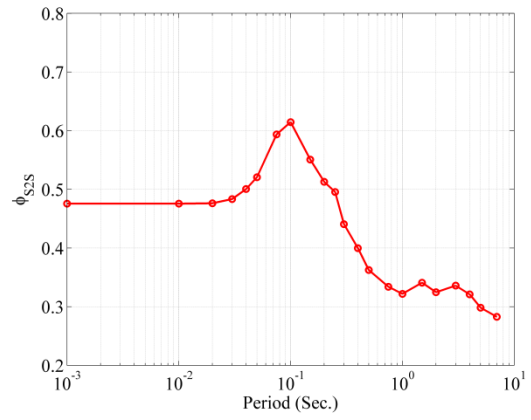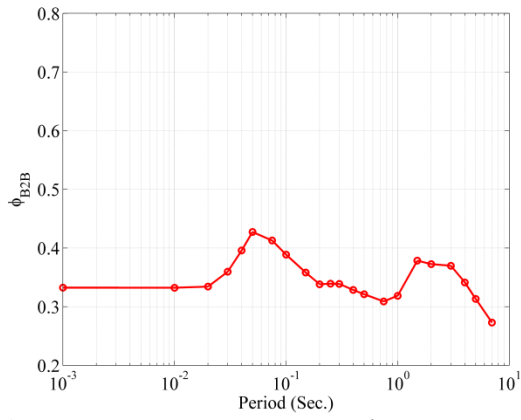
Figure 6. The values φ_{S2S} *and* φ_{B2B} obtained from the site-specific GMPE [not smoothed]
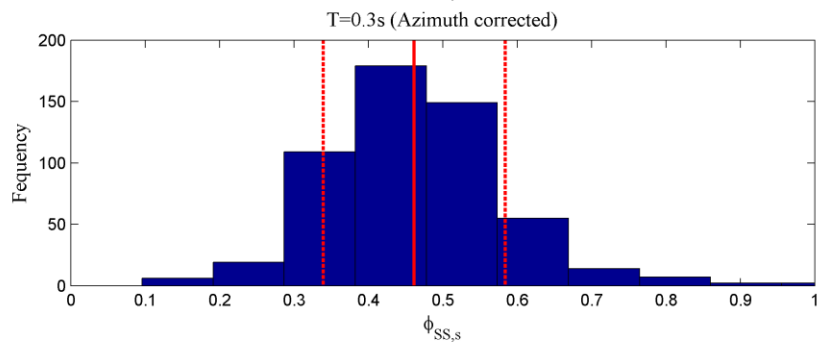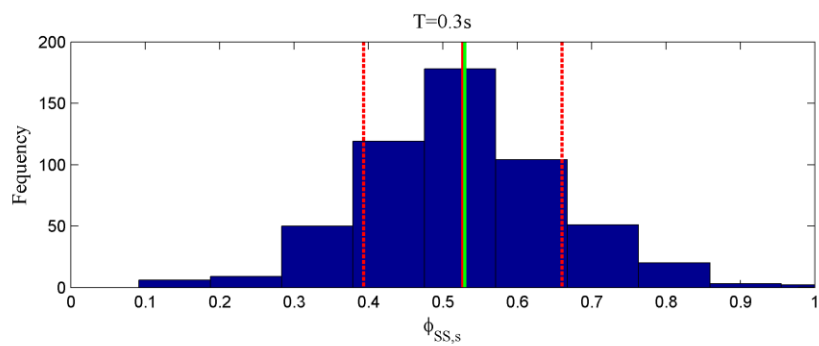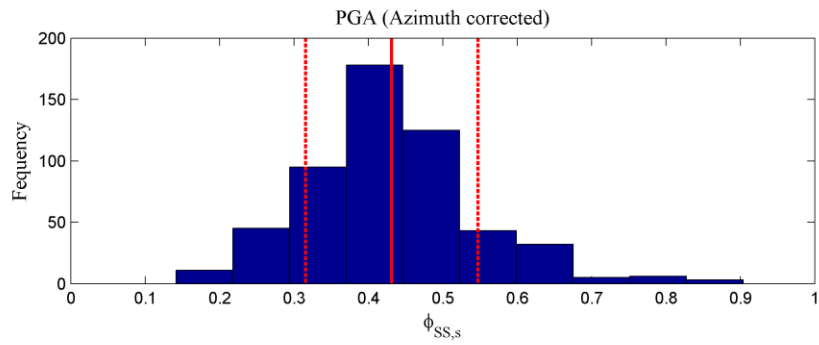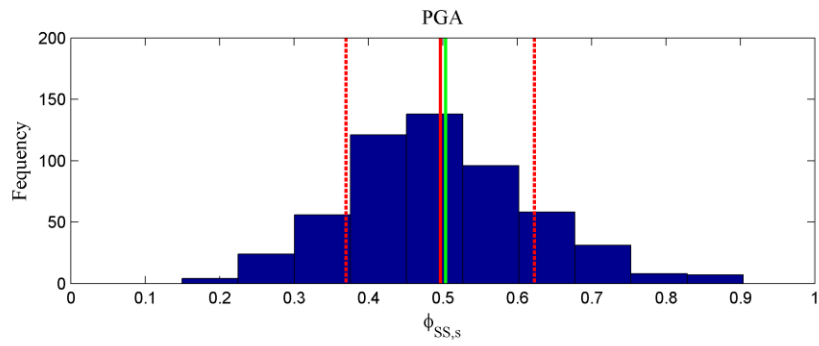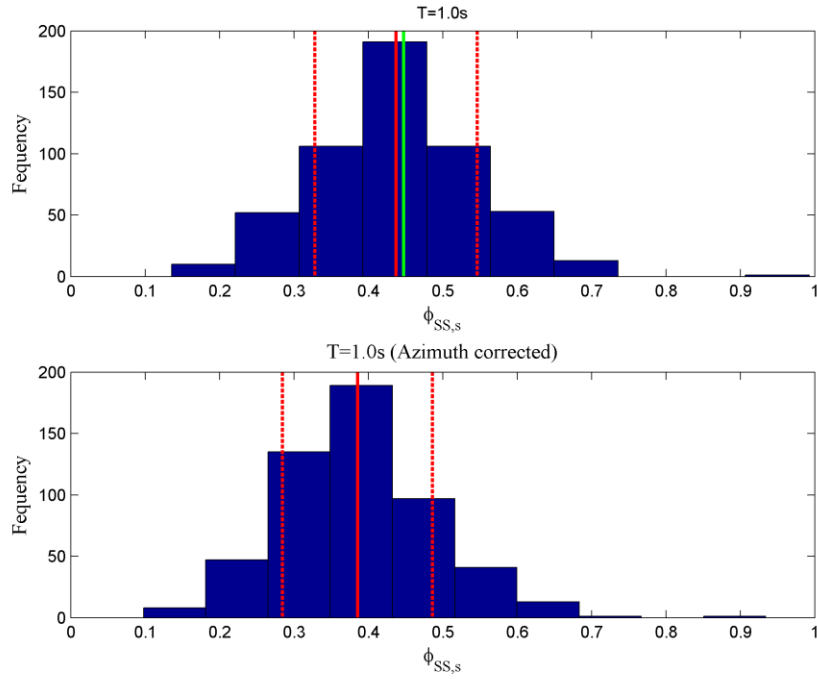
Figure 7. The histogram plot of the $\varphi_{SS,s}$ for the different stations in the database. The red continuous and dashed lines represent the average $\varphi_{SS,s}$ and average $\varphi_{SS,s} \pm$ standard deviation of $\varphi_{SS,s}$. The green continuous line represent φ_{SS} estimated from the constant model (Table 9)

(a)

56
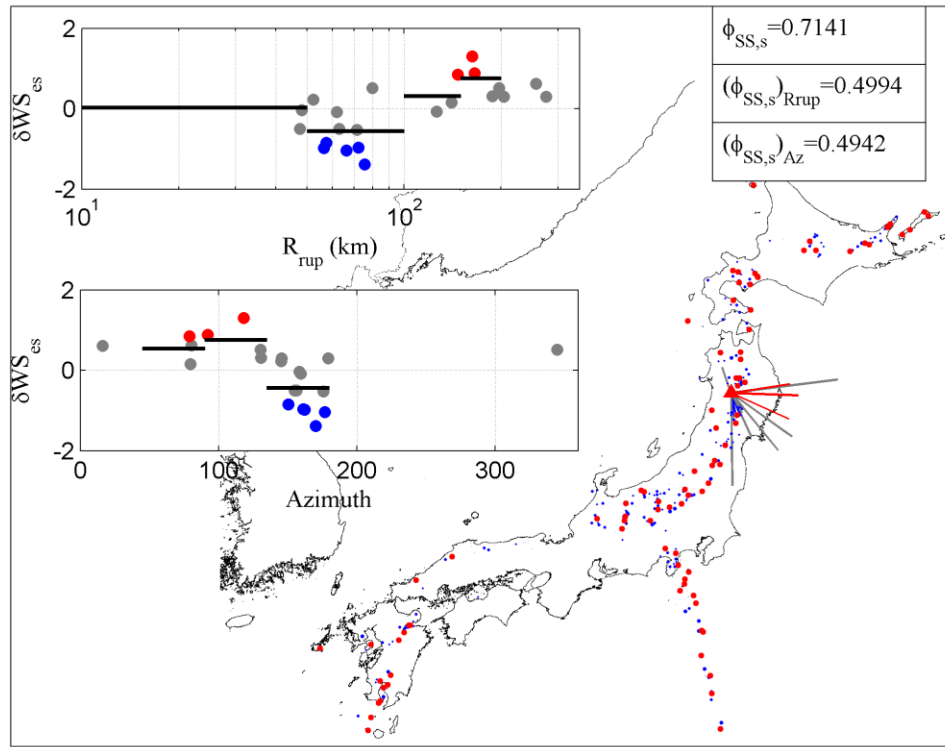
[Surf]-V4-ModelVersusPSA-AOMH07-High $\phi_{SS,s}$=0.86826

$\phi_{SS,s}$=0.8683

$(\phi_{SS,s})_{Rrup}$=0.864

$(\phi_{SS,s})_{Az}$=0.8683

(b)
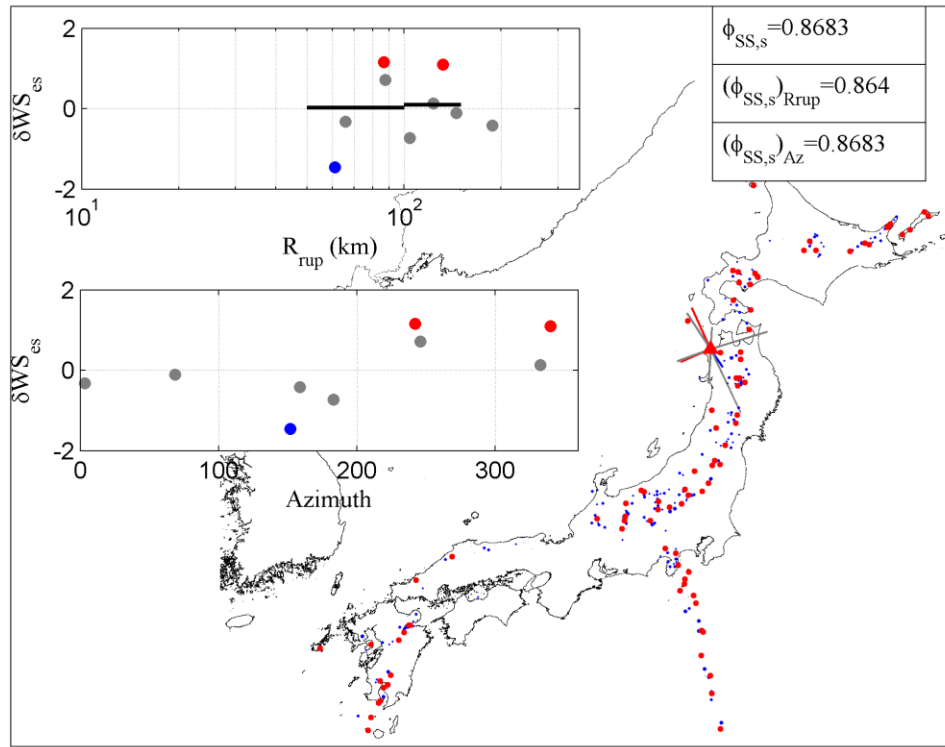
[Bore]-V4-ModelVersusPSA-FKSH14-High $\phi_{SS,s}$=0.65942
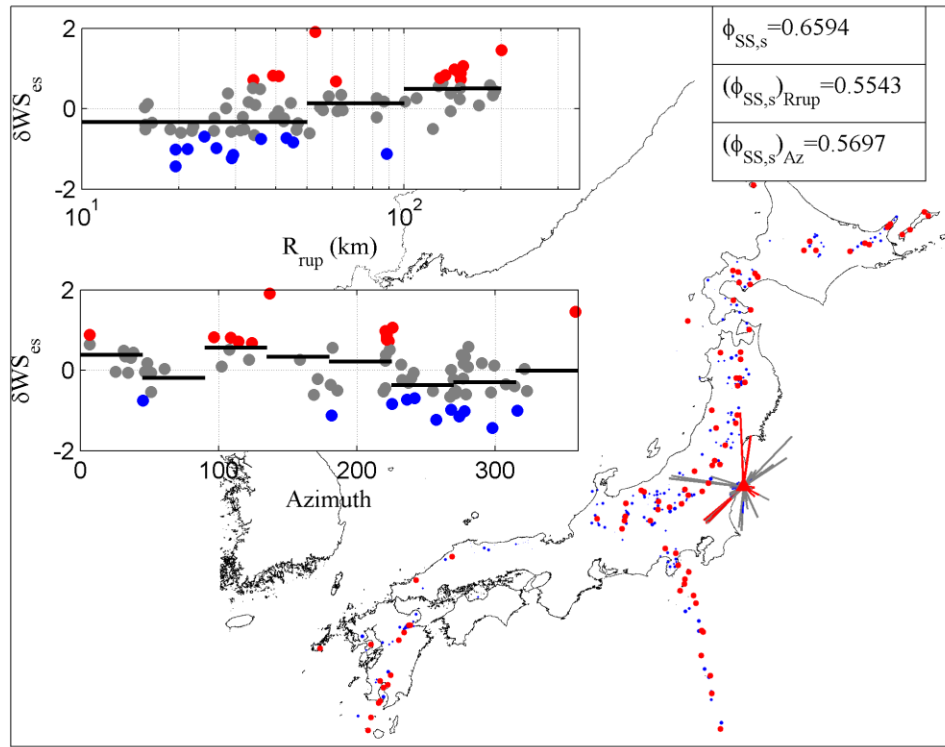
(c)

6

[Bore]-V4-ModelVersusPSA-FKOH03-Low $\phi_{SS,s}$=0.26699
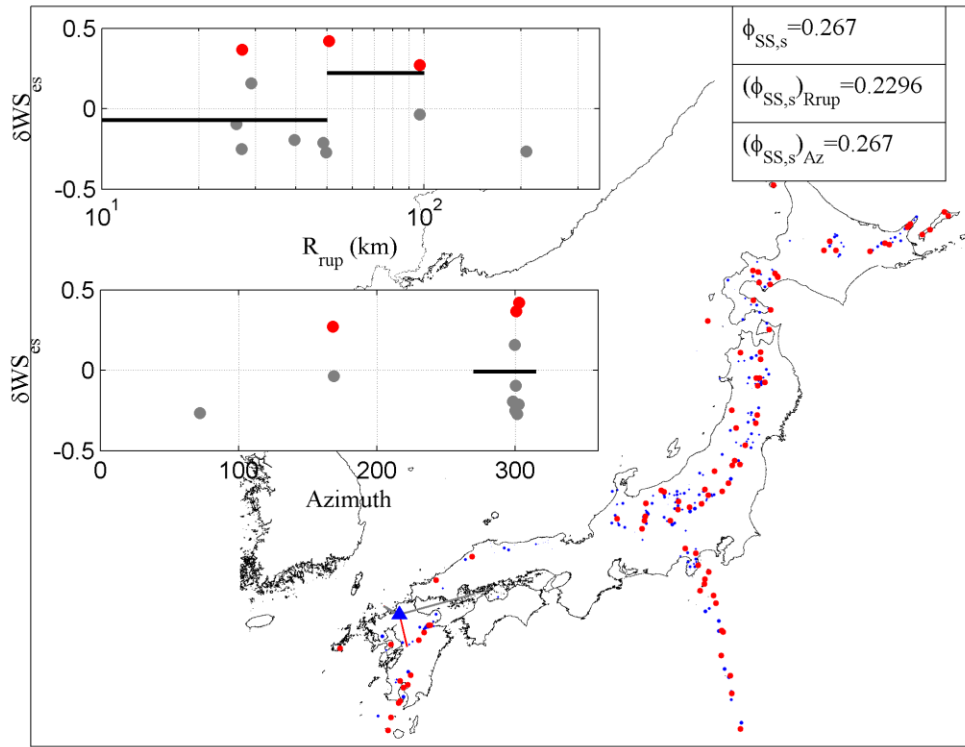
(d)
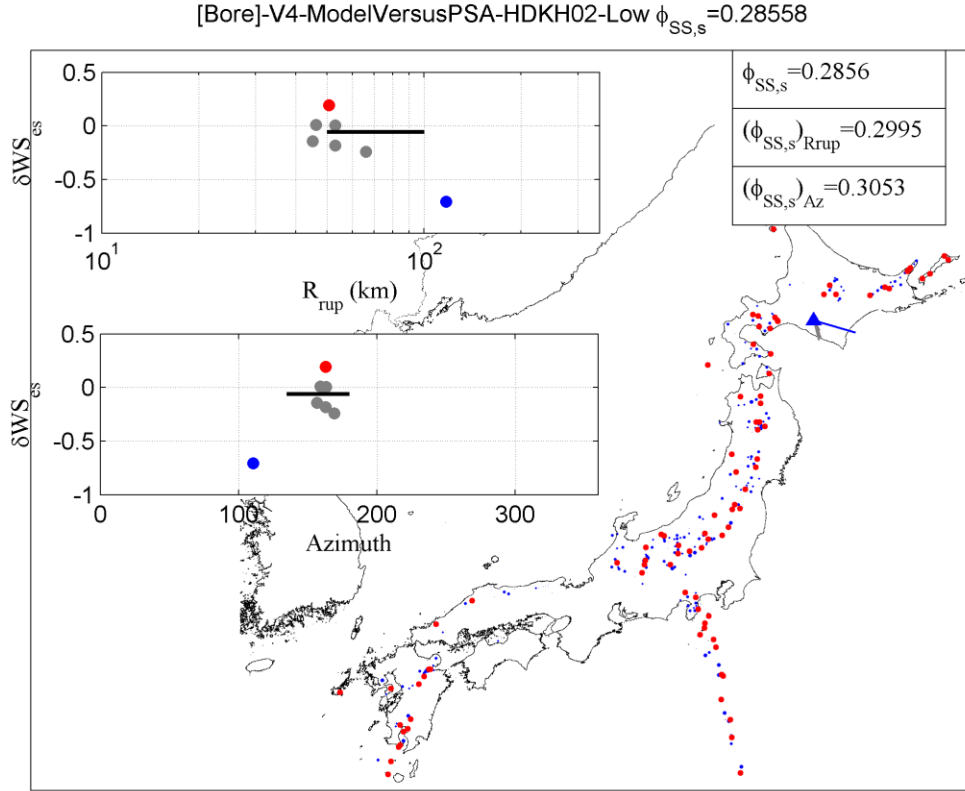
(e)

Figure 8. Each panel shows the location of a recording station on the Japanese map. Lines showing the different source to site paths travelled by all motions recorded at that station are added. The two subplots in each panel show the $\delta W^0_{esb/es}$ [estimated for a specific spectral period at either surface or borehole] versus R_{rup} and azimuth. The panels show the following stations:

  a) AKTH17-PGA-Borehole: example of a station where the bias of δW^0_{esb} versus azimuth is clear. The plots show that records from a specific region is particularly biased;

  b) AOMH07-PGA-Surface: example of a case where it was not possible to remove the δW^0_{es} bias with azimuth because very few motions were recorded at each azimuth bin;

  c) FKSH14-PGA-Borehole: example of a station with many motions from a wide range of azimuth and Rrup, yet it has a high $\varphi^{G/B}_{SS,S}$ due to the complexity of the tectonic environment in Japan;

  d) FKOH03-PGA-Borehole: example of station that recorded motions from limited azimuths. This could be a potential reason as of why the $\varphi^B_{SS,S}$ value at this station is low; and

  e) HDKH02-PGA-Borehole: example of a station that recorded motions from a limited range of Rrup and azimuths. In this case all motions are from the same source except for only one motion. $\varphi^B_{SS,S}$ is believed to be small for this station because it recorded almost all motions from a single seismic source.

60

Figure 9. Plots that show the mean and median relative number of motions recorded at different azimuths. Each subplot shows three curves that represent stations characterized with high, neutral and high $\varphi_{SS,s}$ values

PGA
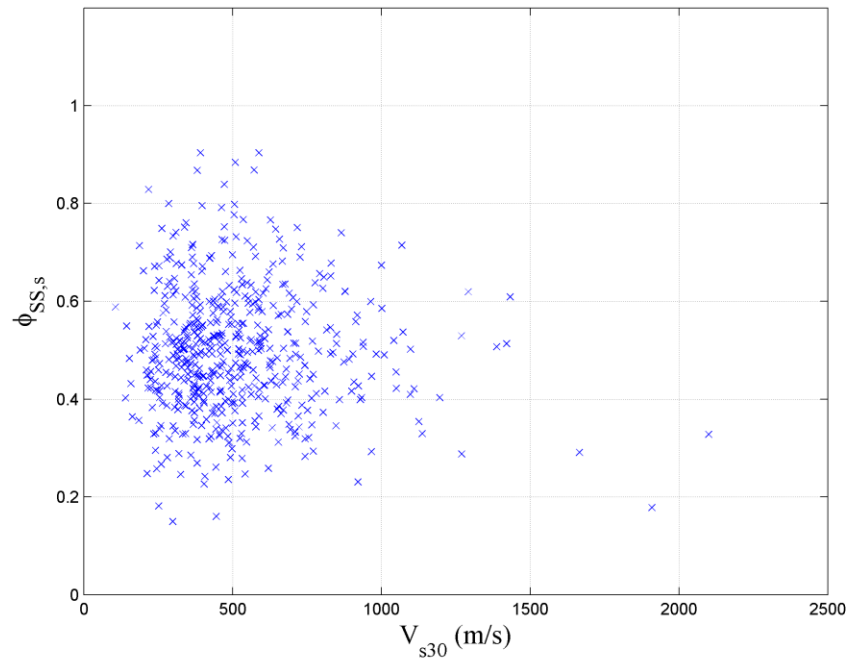
T=0.3s
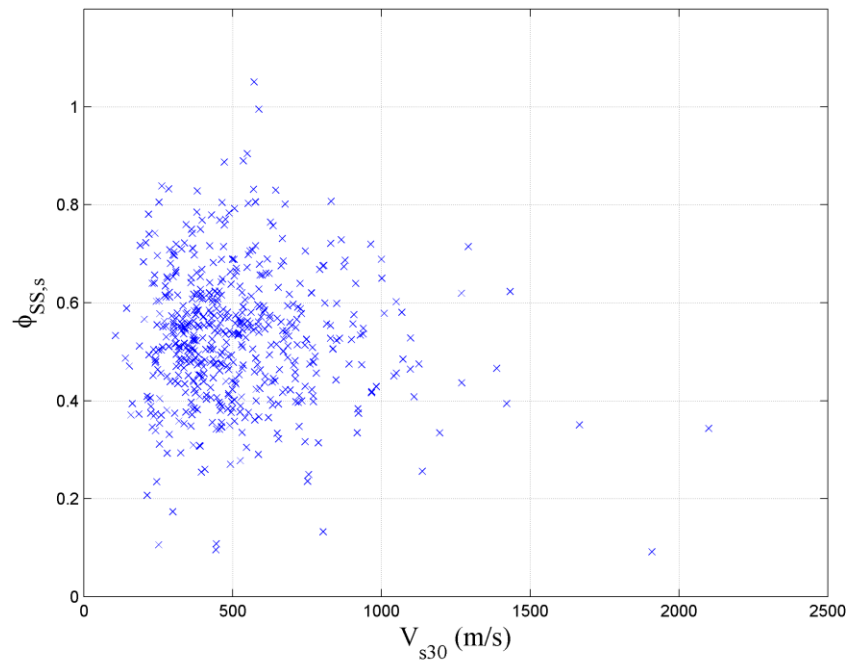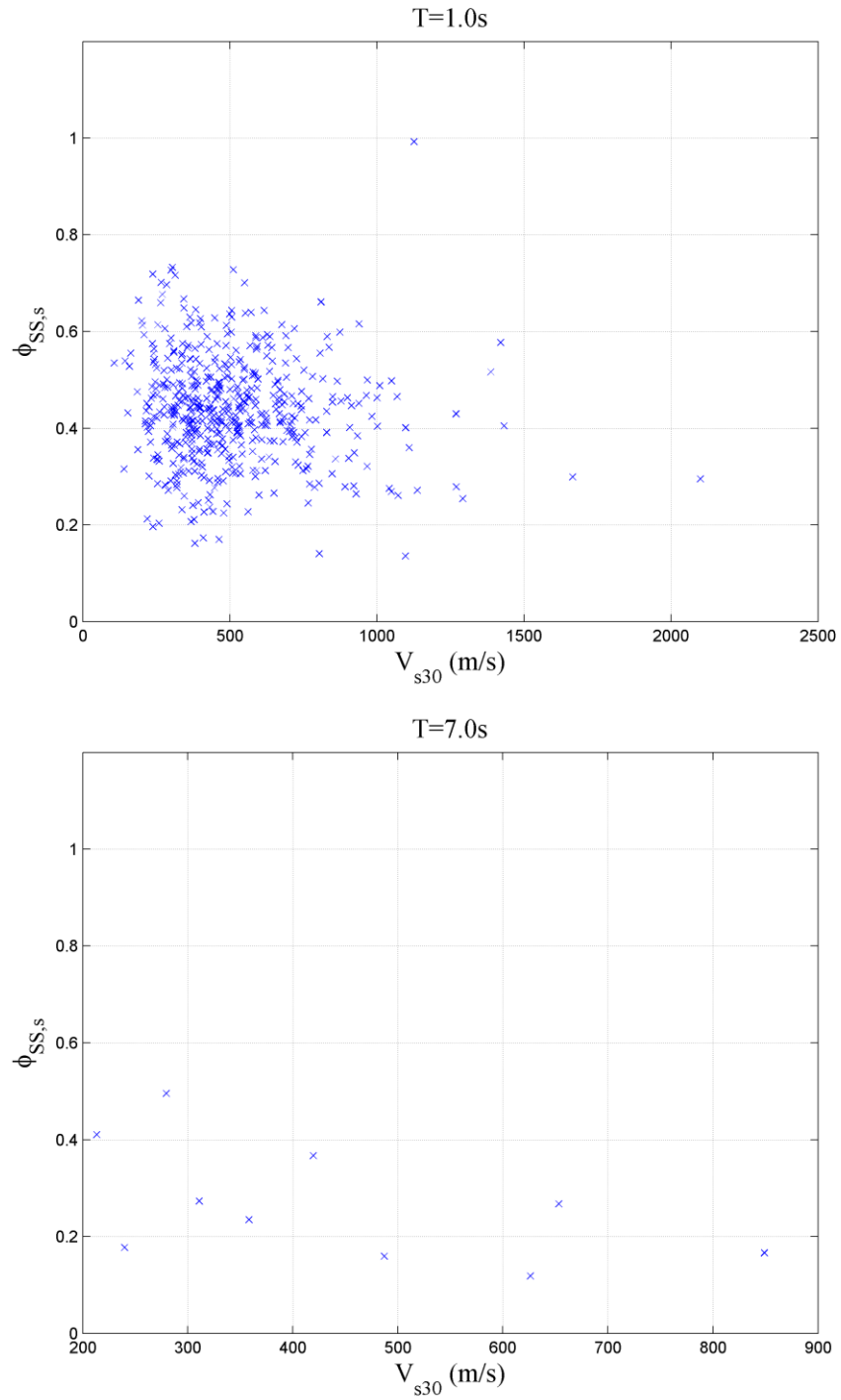
Figure 10. Scatter plots of $\varphi_{SS,s}$ versus V_{s30} for the different stations in the KiK-net network

<div align="center">

Computation of Kappa for the KiKnet Stations

Internal Report, Virginia Tech

By Ashly M. Cabas and A. Rodriguez-Marek

May 5, 2014

</div>

Background info:

It has been shown that path terms based solely on the seismic quality factor (Q) do not completely account for the observed attenuation of high frequencies at a site (Boore, 1983 and 2003). Additionally, site effects in rock and stiff soil sites are not always appropriately constrained by an average shear wave velocity computed within the upper 30 meters ($V_{s,30}$). In fact, the latter mainly characterizes the shallower velocity structure of the profile and fails to capture the influence of shallow crustal attenuation (Laurendeau et al., 2013). Hence, the exploration of alternative site parameterizations has led to the investigation of the high frequency decay parameter "kappa" as originally introduced by Anderson and Hough (1984).

Equation 1 shows the attenuation model as introduced by Anderson and Hough (1984):

$$A(f) = A_o \, e^{-\pi \kappa (r) f} \quad\quad for \, f > f_E \quad\quad\quad (1)$$

Where A_o is a source-and path-dependent parameter, f_E is the frequency above which the spectral decay of log $A(f)$ versus frequency can be considered approximately linear, r is the epicentral distance and $\kappa(r)$ corresponds to the slope of the high-frequency decay of the acceleration Fourier amplitude spectrum in a log–linear space (Anderson and Hough, 1984). A site-specific kappa value (typically known as κ_0) can be obtained by extrapolating the $\kappa(r)$ trend to zero epicentral distance (i.e. for r = 0). It is important to note that equation 1 is only valid under the assumption that Q is frequency-independent (Anderson and Hough, 1984; Ktenidou et al., 2013).

Kappa has become a key input parameter to create and calibrate GMPEs for regions lacking a comprehensive catalog of strong ground motions (e.g. it is used as a high-frequency filter in stochastic models for the generation of synthetic motions).

Different methodologies to compute kappa values or estimate them based on other site parameters such as $V_{s,30}$ has populated the literature recently (e.g. Silva and Darragh, 1995; Chandler et al. 2006; Drouet et al., 2010; Van Houtte et al., 2011). Consequently, as suggested by Ktenidou et al., (2013) a subscript will be used to indicate the specific methodology applied in this study.

Work of Drouet et al. (2010) on the computation of kappa from the transfer function

Drouet et al. (2010) measured a site-specific (independent on distance) kappa value directly on the high frequency portion (i.e. frequencies higher than 10 Hz) of the site's transfer function. Source-path-site inversions using weak to moderate earthquakes recorded by the French Accelerometric Network were used to estimate the transfer function at 76 stations. Kappa values derived from this methodology are called κ_{0_TF} (Ktenidou, et al., 2013).

In this study, linear regressions are used to compute the slope corresponding to the decay of the empirical (linear) borehole transfer function at each station from the Japanese KiK-net database. The selection of an appropriate range of high frequencies for each case imposes a challenge given the high variability that is found from site to site. Therefore, as part of the scope of this study, three iterations have been made in the calculation of κ_{0_TF} for each station and a "taxonomy" to identify similar shapes in the transfer functions has been developed. The defined "taxonomy" serves as a criteria created to select the range of frequencies where kappa should be computed so that more consistency in the results can be achieved.

KiK-net database

Given that the KiK-net database has more than 600 pairs of surface-downhole recording stations it provides a unique opportunity to assess the variability in the estimation of kappa in a systematic way. As pointed out by Drouet et al. (2010) when referring to the French Accelerometric Network, the advantage of this kind of databases is the homogeneous coverage of a unique region which allows the recovery of path and site terms. Thus, the possibility of obtaining transfer functions empirically will help decipher kappa's origins and its correlation to other site parameters such as the shear wave velocity.

The empirical linear site response at each station was computed in terms of borehole Fourier spectral ratios (BFSR) for weak motion. The details concerning the estimation of the mean BFSR for each site can be found in Régnier (2013).

Methodology

Two methods were evaluated in order to perform linear regressions in the high frequency portion of the empirical transfer functions. Initially, a robust linear regression (which is less sensitive to outliers) was used as suggested by Ktenidou et al. (2013) and then, the traditional linear regression was also tested. The standard deviation of the residuals computed for both types of regression were compared and it was found that the ones corresponding to the robust linear regression were smaller. Hence, this approach is believed to be best suited for our analyses.

Taxonomy

Providing consistency to the computation of kappa values is a very important but rather challenging task. Douglas et al. (2010) used the insight of three different analysts to pick frequency ranges for the computation of kappa values. Thus, they were able to define error bars to resolve cases where kappa was sensitive to the frequency range selected. In this study, a taxonomy was defined in order to identify repeatable shapes in the empirical transfer functions corresponding to the KiK-net stations, so that frequency ranges to compute κ_{0_TF} were selected with more consistency. Seven characteristics shapes identified after careful evaluation of all the empirical BFSR are illustrated in Figure 1. Additionally, actions required and suggested frequency ranges are indicated for each case. The example figures provided are actual BFSR used in this study and the corresponding $V_{s,30}$ (in m/s) is also indicated in the figure.

Category	Action required	Suggested frequency range	Example Figure
1- Continuously increasing curve	Do not report a kappa value	N/A	ABSH01 - Vs30 = 822.53 — NO — k = 0.058
2- Second peak at very high frequencies	Report kappa value measured after the second peak	Drouet et al. (2010) compute kappa on the decay of the second peak.	MIEH04 - Vs30 = 490.28 — k = 0.029
3- Flat curves	Do not report a kappa value	N/A	ABSH13 - Vs30 = 463.6 — NO — k = 0.03
4-Dominant peak with secondary bump occurring afterwards (defining a diff slope)	Compare size and significance of the bump relative to the peak	if amplification defined by the bump ≤ 2, compute kappa on the decay portion after the main peak. For less clear cases, select a wider range of frequencies.	AOMH05 - Vs30 = 238.3 — k = 0.073; ABSH12 - Vs30 = 268.54 — k = 0,098
5- Long clear decay at high frequencies	Report kappa value	Select a wide range of frequencies.	FKSH18 - Vs30 = 307.18 — k = 0.027 (0 5 10 15 20 25 30)
6- Clear drop over a limited frequency range	Report kappa value	Identify the decay portion and select the corresponding frequency range.	ABSH10 - Vs30 = 609.68 — k = 0.121
7-Bumpy curves (presence of small consecutive bumps)	Report kappa value	Identify the linear decay "trend" and select the frequency range accordingly.	AOMH14 - Vs30 = 360.65 — k = 0.049

Figure 1: Taxonomy defined for the empirical transfer functions corresponding to the recording stations in the kik-net database.

When the transfer function is characterized by a continuously increasing curve (i.e. case 1 - usually for rock sites) it might be reflecting an amplification at very high frequencies resulting from some heterogeneities within the rock mass (e.g. zones of weakness adjacent to more sound or unweathered rock) that may be generating strong impedance contrasts. Drouet et al. (2010) also observed this issue when computing κ_{0_TF} from his database and they decided that no kappa value

would be reported for these cases (i.e. positive slopes – negative kappa values). Similarly, κ_{0_TF} values for such stations were not reported in this study either.

For some rock sites, values of kappa are very hard to define because their empirical transfer functions show very little amplification if any, which results in very flat BFSR curves (i.e. case 3). In these cases, it is best not to report κ_{0_TF} values unless there is a distinctive sign of deamplification at high frequencies.

The remaining cases where κ_{0_TF} values must be reported still possess a degree of subjectivity regarding the range of frequencies selected for the linear regression. Besides, there are BFSR curves where more than one of the characteristic shapes described in Figure 1 seem to be applicable. This results in greater uncertainty when trying to choose the appropriate boundaries for the computation of the slope at such stations. Nonetheless, it is believed that the implementation of the aforementioned classification system (i.e. taxonomy) still assists in the achievement of more consistent results and might represent the first attempt towards more rigorous methodologies to reduce the uncertainty in kappa calculations.

Results

κ_{0_TF} values were estimated for 668 KiK-net recording stations. However, only the values corresponding to 523 stations are shown in Table 1 along with their corresponding $V_{s,30}$. Problematic stations, which typically corresponded to cases 1 and 3 (see Figure 1) were not included. Likewise, sites were the definition of an appropriate frequency range imposed a significant amount of uncertainty or subjectivity were not included either.

In addition to the computation of κ_{0_TF} values for the Japanese data, a comparison with other estimations of kappa (for the same database but using a different methodology) was believed to assist in the understanding of the variability and fundamental implications associated with this parameter. Van Houtte et al. (2011) has presented estimations of kappa values also for some of the KiK-net recording stations. Their calculations for κ_0 from the surface and downhole recordings were based on the traditional Anderson and Hough (1984) approach. They found that the stratigraphy at shallower depths influence κ_0 most significantly, but a remaining component with a deep origin was also identified. In their study some difficulties arose when trying to estimate kappa at stations with strong site effects (i.e. spikes in the spectra caused by site amplifications made it difficult to select the appropriate slope).

Laurendeau et al. (2013) highlight that the kappa values provided by Van Houtte et al. (2011) were obtained without considering the influence of the instrument response. They neglected to account for a low-pass filter in the Kik-net array instruments, which may have led to overestimation of the rock site attenuation due to inclusion of the instrument response effects in κ. "Data should preferably only be used after correcting or, at least, assessing the instrument response" (Ktenidou et al., 2013).

Considering that the kappa values computed in this study are based on the empirical transfer functions at the stations (i.e. κ_{0_TF}) and not on particular ground motion recordings (i.e. κ_{0_AS}

according to Ktenidou et al. notation), the instrument response becomes less of an issue. However, special consideration was taken not to use frequencies above 30 Hz which was the cut-off frequency of the filter used for the processed records in the Japanese database.

In this framework, the comparison between κ_{0_TF} and κ_{0_AS} (from Van Houtte et al., 2011) as illustrated in Figure 2 should be interpreted with caution. As a first hypothesis, the contribution of the shallower layers (between the surface and downhole instruments) to the attenuation of seismic waves can be characterized by the difference between κ_{0_AS} estimated using recordings at the surface and at depth (i.e. κ_{0_SURF} - κ_{0_DOWN}). Intuitively, this same contribution should be captured when computing κ_{0_TF} from the empirical borehole transfer functions at each station. It turns out that this relationship, as shown in Figure 2a, is obscured by the scatter and uncertainty embedded in each of the methods for the estimation of kappa. The tested hypothesis is shown in green and it can be noted that some of the stations provide a good agreement with the expected trend. However, a large portion of the station (even when only considering non-problematic stations) still falls far from such trend. Further investigation regarding the characteristics of those sites that provide a good match should provide more lights in the understanding of the extent of the influence of kappa values in the characterization of the attenuation in surficial layers. Figures 2b and 2c are also presented for completeness and they reflect a greater scatter and hence a lower correlation between κ_{0_TF}, κ_{0_SURF} and κ_{0_DOWN}.
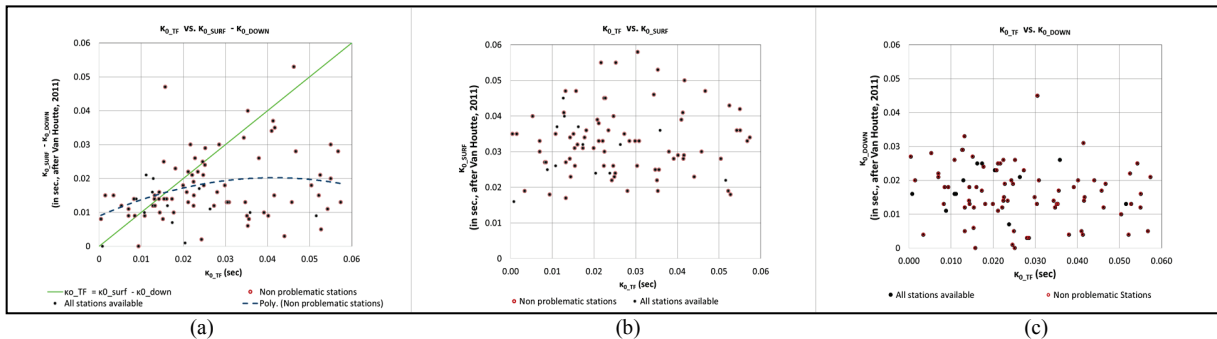
Figure 2: from left to right, relationships between κ_{0_TF} and κ_{0_SURF} - κ_{0_DOWN}, κ_{0_TF} and κ_{0_SURF}, and κ_{0_TF}, κ_{0_DOWN}. Values corresponding to κ_{0_SURF} - κ_{0_DOWN} were taken from Van Houtte et al. (2011). A trend dashed line has been included in (a) for comparison purposes. The greater scatter shown in figures (b) and (c) suggest a lower correlation between the kappas evaluated in each case. Further study is needed to fully characterize the slightly better correlation observed in figure (a).

Drouet et al. (2010) also compared their κ_{0_TF} with kappa values computed by Douglas et al., (2010) for the same French database when using the traditional approach introduced by Anderson and Hough (1984). They indicated their estimations were predominantly lower than those from Douglas et al. (2010) but also commented on the existence of a "clear correlation between the kappa values from both methods". Unlike Drouet et al. (2010), our computations for κ_{0_TF} were not always lower than the κ_{0_AS} estimated by Van Houtte et al. (2011) as Figures 2b and 2c show. In fact, a potential correlation was found between κ_{0_TF} and κ_{0_SURF} - κ_{0_DOWN} (i.e. difference between κ_{0_AS} computed at the surface and the downhole instruments).

Given that $V_{s,30}$ values were also available at the original 668 stations, the correlation with κ_{0_TF} values was investigated. The $V_{s,30}$-kappa relationships derived from the computed values from this

study and those pertaining to Van Houtte et al. (2011) are shown in Figure 3 for non-problematic stations.
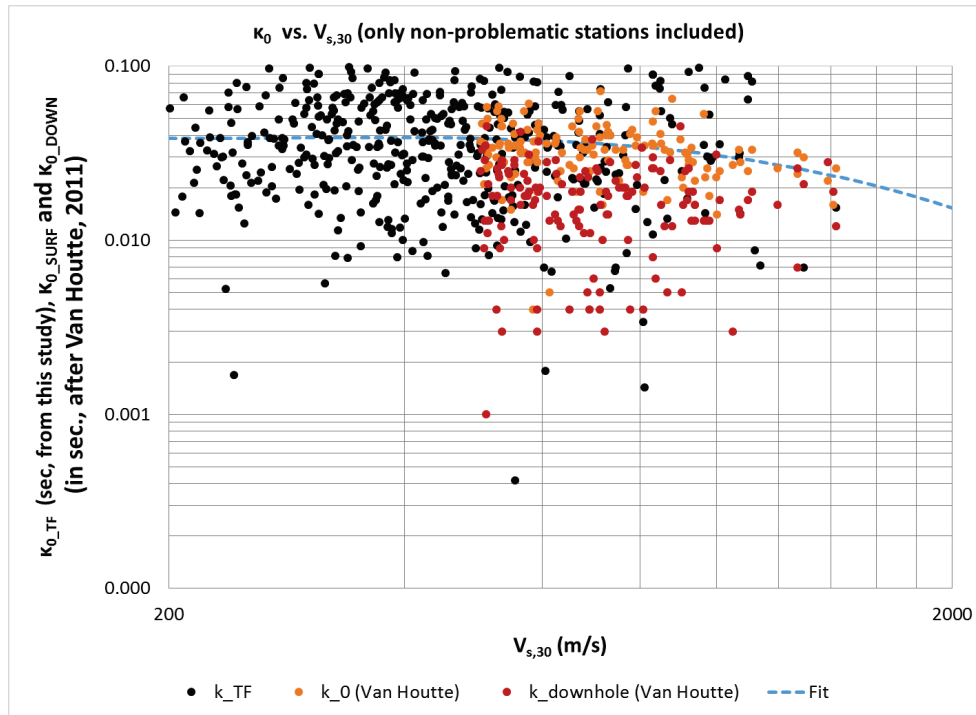
Figure 3: $V_{s,30}$-kappa relationship when using values obtained from this study and the ones from Van Houtte et al. (2011). The amount of scatter is consistent with what other studies have found (Laurendeau et al., 2013; Ktenidou et al., 2013; Chandler et al., 2006)

The lack of a strong correlation in the data suggests, as also indicated by Drouet et al. (2010), that kappa values may be influenced by deeper geologic structures at rock sites. This would make them less dependent on a parameter that mostly describes surficial properties of the site (namely $V_{s,30}$). Analogously, Campbell (2009) has investigated the effect of the sediment thickness on kappa for soft sites.

Conclusions

Site-specific kappa values have been successfully computed for 668 stations from the KiK-net database. A simple classification system for the selection of appropriate frequency ranges for the calculation of κ_{0_TF} has also been proposed in order to achieve more consistent results. The relationship between κ_{0_TF} and other site–specific kappa values from the literature (κ_{0_AS}) has been explored; and it seems like the influence of the material in between the surface and downhole instruments from the Japanese data can be isolated and characterized by different estimations of κ_0. The addition of kappa values to obtain the net attenuation at a site (i.e. $\kappa_{0\_SURF} = \kappa_{0\_DOWN} + \kappa_{0\_TF}$) still needs further study. In fact, the following section explains how this issue will be addressed in the future. Likewise, relationships between $V_{s,30}$ and kappa remain a subject of dispute in the literature and the corresponding relationship derived using this study's estimations of κ_{0_TF} proved to be consistent with the scatter seen by other researchers.

Future work

κ_{0_AS} values will be computed for the KiK-net stations following the definition of Anderson and Hough (1984). The recordings will be corrected to account for the instrument response by using a Butterworth filter with f_{cut} = 30 Hz and n=3, as recommended by Laurendeau et al. (2013). Kappa values dependent on distance (i.e. κ_r) will be obtained from different recordings at each station so that a graph with kappa values associated with different epicentral distances can be generated. Site-specific kappa values (i.e. κ_0) can then be obtained by extrapolating the κ_r to zero epicentral distance. A linear regression using the simple model defined by equation 2 will be used to find the intersection with the y-axis (i.e. κ_0).

$$\kappa = \kappa_0 + \kappa_r \times r \qquad\qquad (2)$$

where r refers to the epicentral distance. κ_r can be obtained by estimating the slope of a linear regression considering all the kappa values estimated for different ground motion recordings at the same station (i.e. κ). An example of the computation of κ_0 and κ_r is shown in Figure 4, where the red line represents the aforementioned linear regression and the blue circles are assumed to be individual estimations of kappa for different recordings at the same station.
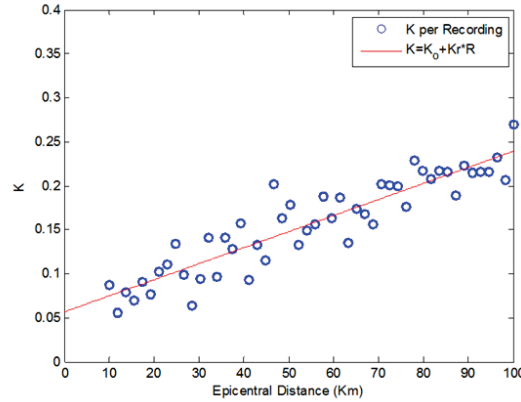
Figure 4: Vs,30-kappa relationship when using values obtained

The ultimate goal will be to compare the site-specific kappa values obtained from two different approaches (κ_{0_AS} vs. κ_{0_TF}) so that more insights can be unveiled on the relationships between different methodologies as well as decipher the true origins of the variability observed in reported values of kappa in the literature.

In addition, future work will be oriented towards the investigation of the reasons behind κ_{TF} not being as close as expected to $\kappa_{0\_SURF} - \kappa_{0\_BOREHOLE}$. This hypothesis has been initially tested in this study (using Van Houtte et al. (2011) estimates) but more systematic analyses will be performed with new computations of $\kappa_{0\_SURF} - \kappa_{0\_DOWN}$ in order to identify specific patterns in the data.

References

Anderson, J., and S. Hough (1984). A model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies, Bulletin of the Seismological Society of America 74, no. 5, 1969–1993.

Boore, D. M. (1983). Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra, Bulletin of the Seismological Society of America 73, 1865–1894.

Boore, D. (2003). Simulation of ground motion using the stochastic method, Pure Applied Geophysics 160, no. 3, 635–676.

Campbell, K. W. (2009). Estimates of shear-wave Q and $\kappa 0$ for unconsolidated and semiconsolidated sediments in eastern North America, Bulletin of the Seismological Society of America 99, 2365–2392.

Van Houtte, C., S. Drouet, and F. Cotton (2011). Analysis of the origins of κ (Kappa) to compute hard rock to rock adjustment factors for GMPEs, Bulletin of the Seismological Society of America 101, no. 6, 2926–2941.

Drouet S., F. Cotton and P. Gueguen (2010). VS30, κ, regional attenuation and Mw from accelerograms: application to magnitude 3–5 French earthquakes. Geophysical Journal International, 182, 880–89.

Douglas,J., P.Gehl, L.Bonilla, and C. Gélis (2010). A κ model for main land France, Pure Applied Geophysics 167, no. 11, 1303–1315.

Ktenidou, O.-J., F. Cotton, N. Abrahamson, and J. Anderson (2013). Taxonomy of kappa: A review of definitions and estimation approaches targeted to applications, Seismological Research Letters. 85, no. 1.

Ktenidou, O.-J., C. Gélis, and L.-F. Bonilla (2013). A study on the variability of kappa (κ) in a borehole: Implications of the computation process, Bulletin of the Seismological Society of America 103, no. 2A, 1048–1068.

Laurendeau, A., F. Cotton, O.-J. Ktenidou, L.-F. Bonilla, and F. Hollende (2013). Rock and Stiff-Soil Site Amplification: Dependency on V_{S30} and Kappa (κ_0). Bulletin of the Seismological Society of America, Vol. 103, No. 6, pp. –, December 2013.

Régnier, J. (2013). Variabilité de la réponse sismique: de la classification des sites au comportement non-linéaire des sols. Doctoral Thesis.

Silva, W. J., I. G. Wong, and R. B. Darragh (1995). Engineering characterization of earthquake strong ground motions in the Pacific Northwest, in Assessing Earthquake Hazards and Reducing Risk in the Pacific Northwest, U.S. Geol. Surv. Profess. Pap. 1560.

Silva, W., and R. Darragh (1995). Engineering characterization of earthquake strong ground motion recorded at rock sites, Electric Power Research Institute, Report TR-102261, Palo Alto, California.

Summary of Research Conducted at Virginia Tech on the Parameterization of Topographic Effects

Internal Report, Virginia Tech

By Manisha Rai and A. Rodriguez-Marek

May 5, 2014

Topographic irregularities such as hills, valleys and ridges have been shown to significantly affect the intensity and frequency characteristics of strong ground motion during earthquakes. However, these effects are not accounted for in ground motion prediction equations that are used for seismic hazard assessment. We performed an empirical study to test if simple parameters derived solely from the elevation data can be used to capture systematic bias in GMPE residuals from a global database. The dataset selected for this study is the Chiou and Youngs (2014) NGAWest2 dataset. Topographic parameters are computed at each ground motion recording station using the elevation data at the station location. Site residuals, representing average error in prediction from ground motion model at a given ground motion recording station (after accounting for the earthquake effects), are computed using residuals from Chiou and Youngs (2014) ground motion model. To get a good estimate of the site residuals at each station, only stations that have at least three recording on them are included in the study. A statistical analysis is performed to test which parameters can predict the bias in the ground motion data. The results from the analysis are used to propose a model to correct the biases.

The geometric parameter used to quantify topography is 'Relative elevation'. Relative elevation (also referred to as the Topographic Position Index) is a measure of how high or low a point on the surface is with respect to the average in the neighborhood. Some researchers have used it in the past for terrain classification in watershed studies. Relative elevation is obtained for each cell of the digital elevation model (DEM) by taking the difference of elevation at the cell and the mean elevation of the cells within a neighborhood of the cell. The size of the neighborhood is referred to as the scale-parameter. In this study, a circular neighborhood is considered. The diameter of the circle is referred to as the scale parameter (d) and the relative elevation computed at a scale d is referred to as H_d. A positive H_d means that the elevation of the cell is higher than the mean in the neighborhood; a negative value means that the elevation of the cell is lower than the mean in the neighborhood and a value of zero means that the station is located in a region of uniform slope. For this reason, H is effective in highlighting features such as ridges ($H_d > 0$), valleys ($H_d < 0$) and plains/slopes ($H_d = 0$). The value of scale parameter considered range from 250 m to 3000 m. To classify a terrain into discrete classes based on H_d, boundaries are set for each class using the threshold-parameter, where only positive values are allowed. The boundaries for zones are set at both a negative threshold and a positive threshold, which allows the continuous values of H_d of a terrain to be separated into three discrete classes: 'high', where H_d is greater than the positive threshold; 'intermediate', where H_d is between the positive and the negative thresholds; and, 'low", where H_d is less than the negative threshold.

As topographic effects are not modeled by the GMPEs (and if the effects are repeatable and observable) the residuals from the GMPE should have a systematic relationship with topography. The residual component that is of interest for this analysis is the site residual. Site residual is the error in prediction at a site, after accounting for the effects of earthquake. As topographic effects are site effects, the bias resulting from them is expected be visible in the site residuals. To obtain the site terms, intra-event residuals are partitioned using mixed effects regression as follows:

$$\delta W_{es} = \delta S2S_s + \delta WS_{es} \tag{1}$$

where δW_{es} is the intra-event residual, $\delta S2S_s$ is the site residual and δWS_{es} is the site and event corrected residual. Both $\delta S2S_s$ and δWS_{es} are assumed to be normally distributed with a mean of zero and a standard deviation of ϕ_{s2s} (or the site to site variability) and ϕ_{ss} (or within site variability). To test if topography systematically affects the residuals, we used equation 1 to split the residuals in each topographic class to compute class specific $\delta S2S$, ϕ_{s2s} and ϕ_{ss}, using random effects regression. The value of scale and threshold parameter used for classification is 1500 m and 20m, respectively. The resulting mean site residuals for each topographic class at each period are shown in Figure 1. The figure shows that the average site residuals for high lying sites are greater than the average site residual in the intermediate class for periods in the range of 0.3 – 1s. The average site residuals for low lying sites are lower than the average site residuals in the intermediate class for periods between 0.3 and 10 s. Higher mean site residuals imply that the recorded ground motion on the sites were on average higher than the predicted value and vice versa. As most of the stations in the dataset are member of the intermediate class, the average site residual of the intermediate class represents the overall mean residual in the dataset. Thus the average site residual for intermediate sites is close to zero as the GMPE was carefully fitted to keep the mean residual of the complete dataset close to zero. Note that the residuals are in the log scale, the average amplification and de-amplification factors for each class can be obtained by taking the exponent of observed mean site residual for that class. The analysis of variance (ANOVA) and Tukey's test are performed to find periods for which the difference in mean site residuals for the topographic class are statistically significant. Significant higher biases are observed for the high-lying sites for a period range of 0.38-0.75 sec and a significant lower bias for low-lying stations for periods of 0.5 s to10 s with respect to intermediate sites (Figure 2). The magnitude of these effects is period-dependent with amplifications peaking at about 0.4 s for high lying sites. Similar trends in mean site residuals were also observed for other values of scales and threshold parameters.

Site to site variabilities (ϕ_{s2s}) are calculated for the three topographic classes and are plotted with period in Figure 3. The total ϕ_{s2s} of the data is also shown. The figure shows a higher ϕ_{s2s} value for the high and low classes, as compared to the intermediate class, for periods up to about 1 s. For periods greater than 1 s the ϕ_{s2s} values drop lower than the intermediate. We typically expect to see a higher variance in the site terms for high and low lying sites, compared to the intermediate sites, in the period range where topography significantly affects ground motions. The reason being that the classification scheme groups all different kinds of ridges together and ignores other complicated aspects of ridge geometry, such as the direction of the ridge, dimensions of the ridge and such. These unaccounted factors can lead to a higher variability in site response

from one site to another. Similar reasoning applies to the low sites. Intermediate sites are often flat terrains with relatively smaller slope and thus the sites classified as intermediate are relatively less likely to show the same variability in their site response. In the case of NGAWest2 data, the number of usable data starts dropping after 1 s and so it is difficult to provide an explanation for the lower variability in site terms for high and low class at higher periods. Moreover, high lying sites are only affected by topography in a narrow periods range from 0.38 – 0.75, and so the aforementioned behavior for high sites can only be expected in this period range.

Within site variabilities (ϕ_{ss}) for the three topographic classes are shown in Figure 4. Within site variabilities represent the variability in the site and event corrected residuals. ϕ_{ss} values for high and low class are higher than intermediate class from periods between 0.01 to about 0.2 seconds. After 0.2 seconds, the difference becomes very small. Just like in the case of ϕ_{s2s}, we typically expect to see a higher ϕ_{ss} values for sites with topography (high, low class) than sites that are flat (intermediate).

The same analysis was repeated for the KikNet dataset and the results are shown in Figure 5-8. Figure 5 shows the mean site residuals for the three topographic classes plotted with period. Figure 6 shows the significance values from ANOVA. The figure shows that the differences in mean site residuals for the three classes are not statistically significant at any period. We do not know the real reason for this behavior. However, a possible reason could be that the elevation data used for parameter computations for Japanese stations are not as accurate as the elevation data used for California stations. This inaccuracy can introduce random errors into the parameters that can obscure the trends.
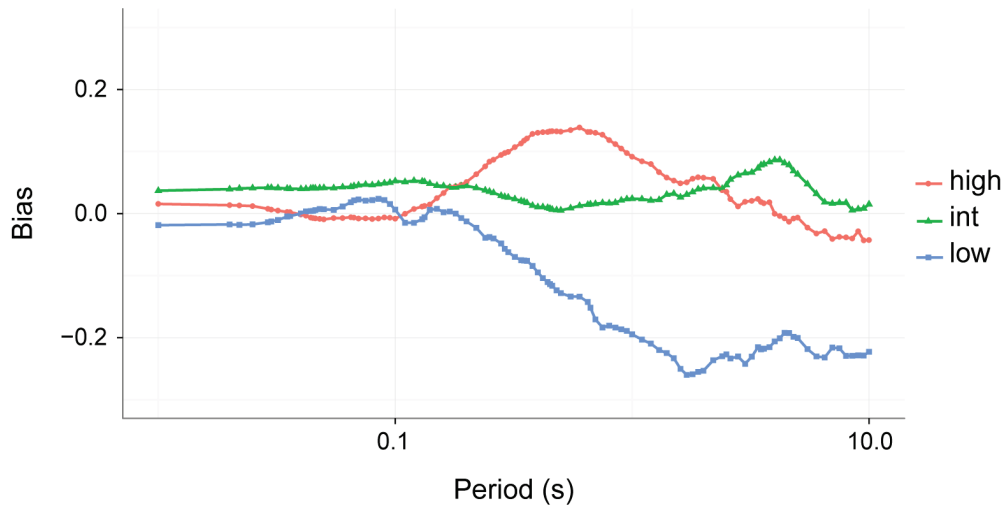
Figure 1: Bias or the mean site residual for high, low and intermediate sites are shown for different periods. A scale of 1500 meters and a threshold of 20 meters were used for the classification.
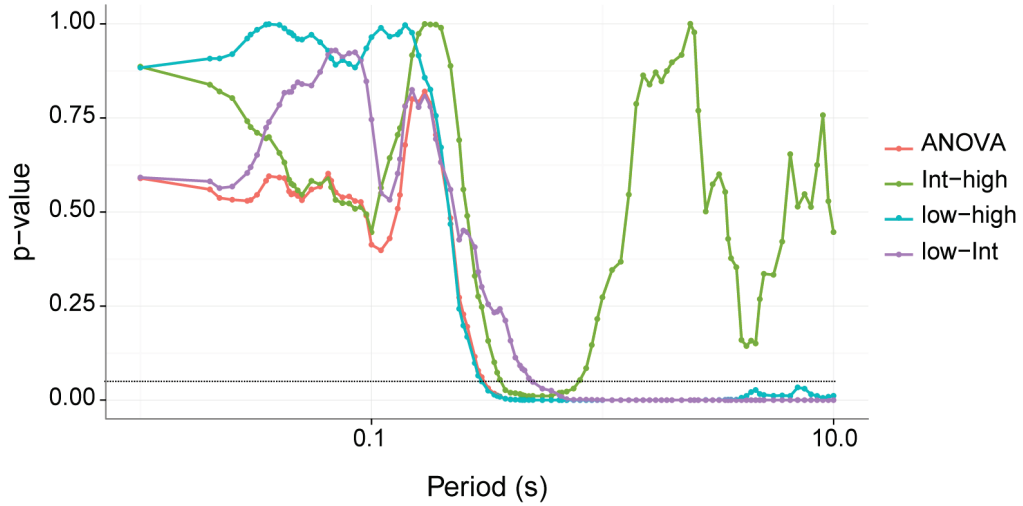
Figure 2: p-values from ANOVA and pairwise Tukey's test are shown with periods. A p = 0.05 line is also shown in grey. P-value less than 0.05 signify that the differences in mean site residuals of the classes are statistically significant.
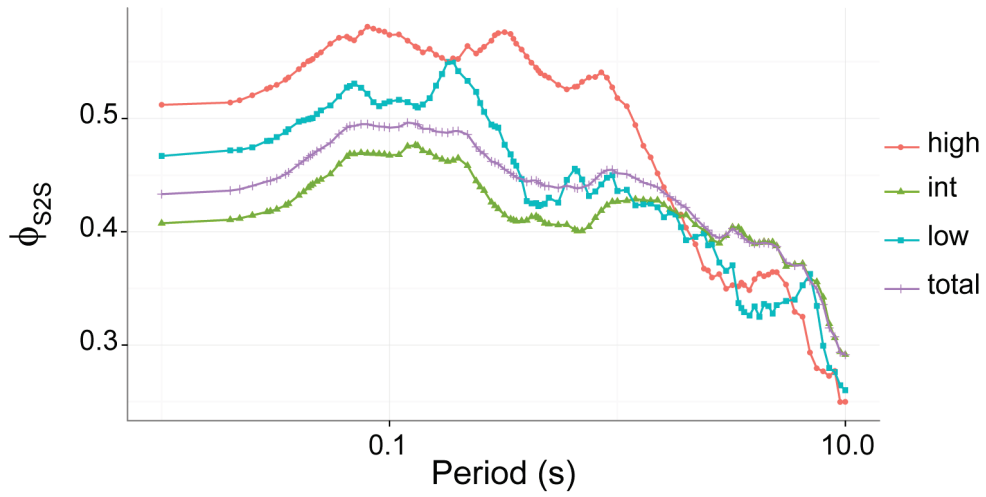
Figure 3: ϕ_{s2s} values are shown for each topographic class at each period. Also shown is the total ϕ_{s2s} of the data.
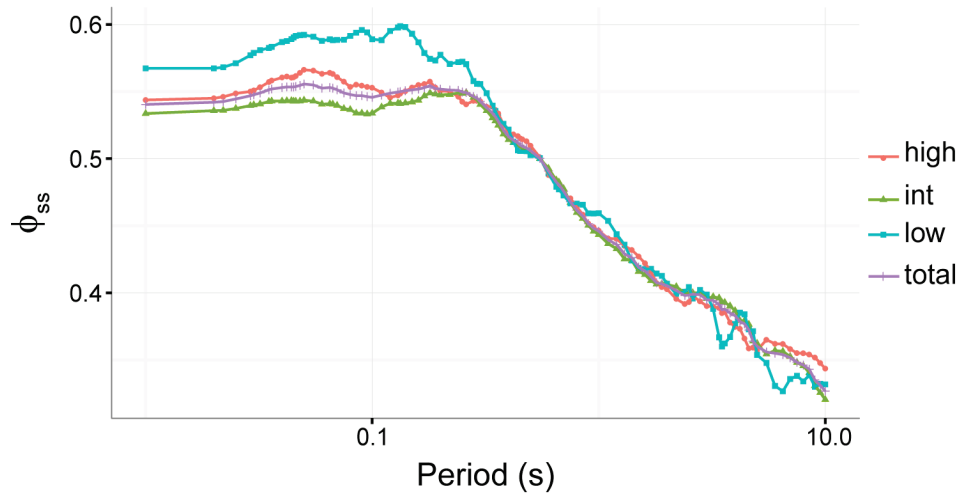
Figure 4: ϕ_{ss} values are shown for each topographic class at each period. Also shown is the total ϕ_{ss} of the data.
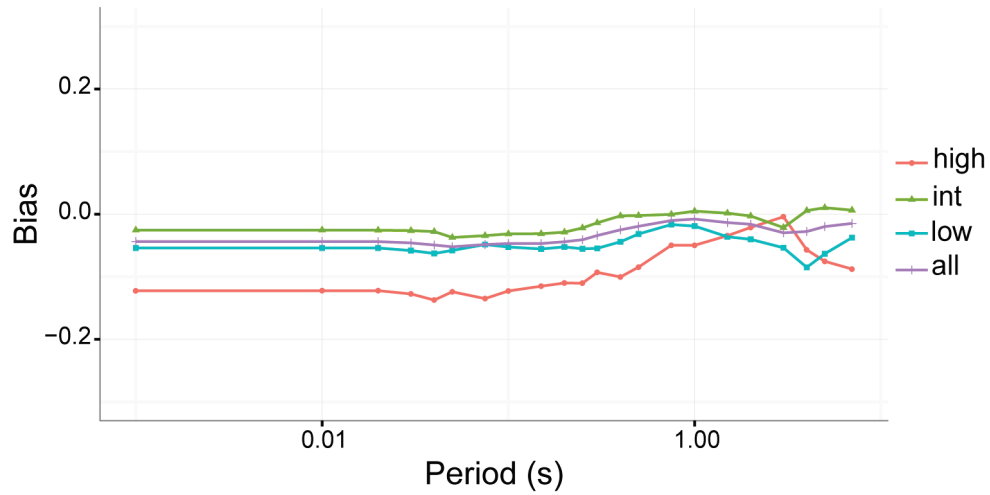
Figure 5: Bias or the mean site residual for high, low and intermediate sites are shown for different periods for the KikNet dataset. The total bias in the dataset is also shown. A scale of 1500 meters and a threshold of 20 meters were used for the classification.
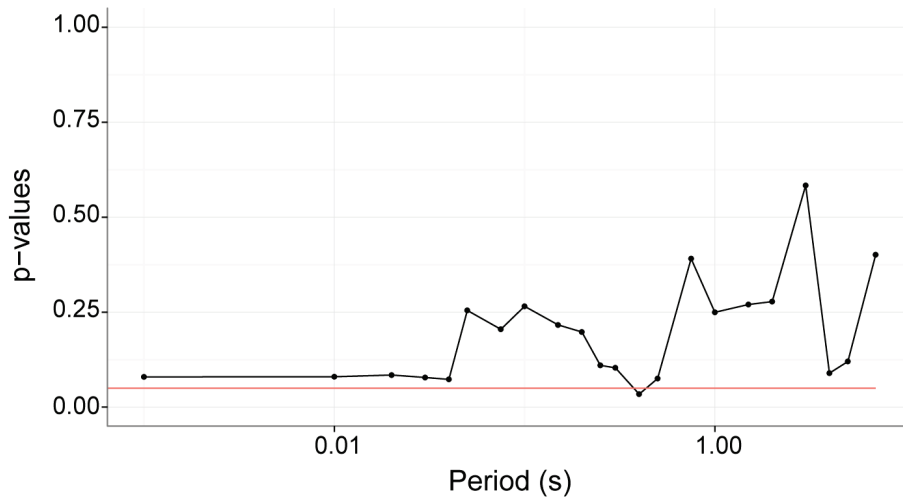
Figure 6: p-values from ANOVA and pairwise Tukey's test are shown with periods. A p = 0.05 line is also shown in red. P-value less than 0.05 signify that the differences in mean site residuals of the classes are statistically significant.
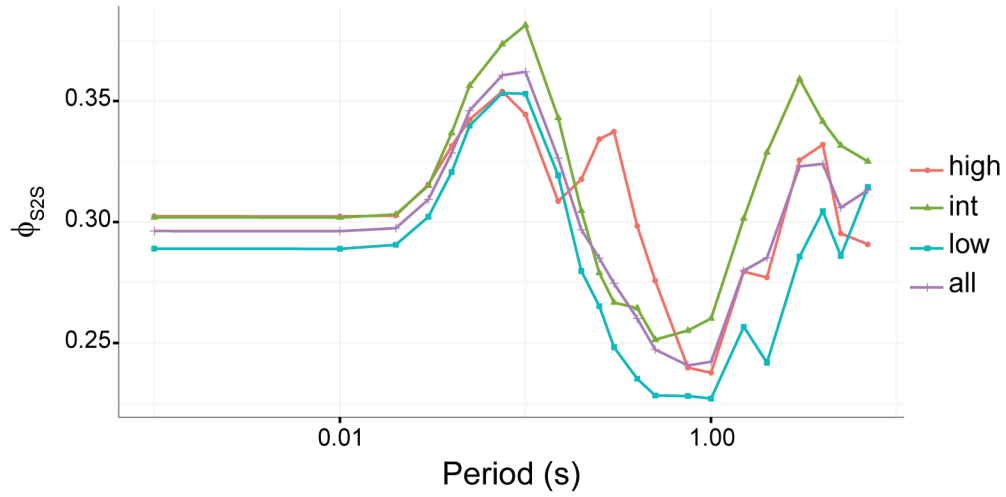
Figure 7: ϕ_{s2s} values are shown for each topographic class at each period. Also shown is the total ϕ_{s2s} of the data.
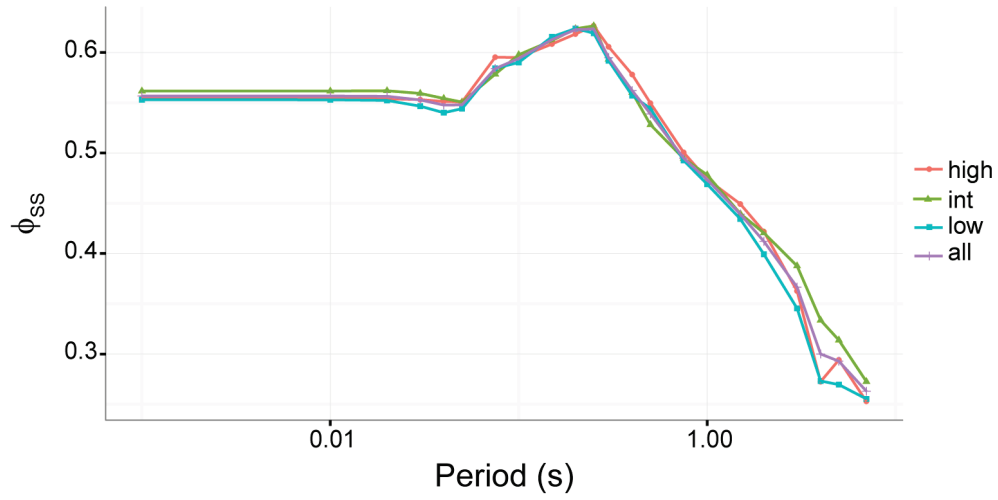
Figure 8: ϕ_{ss} values are shown for each topographic class at each period. Also shown is the total ϕ_{ss} of the data.

A Method for Including Path Effects in Ground-Motion Prediction Equations: An Example Using the M_w 9.0 Tohoku Earthquake Aftershocks

by Haitham M. Dawood and Adrian Rodriguez-Marek

Abstract Past studies in tectonic regions dominated by subduction zones that result in the creation of volcanic belts, such as New Zealand and Japan, have pointed to higher attenuation rates across the volcanic regions. This study uses the downhole motions recorded at KiK-net stations from 117 aftershocks that hit Japan after the great M_w 9.0 Tohoku earthquake to quantify region-dependent strong-motion attenuation rates. To this end, an approach to include path effects in the development of ground-motion prediction equations (GMPEs) is presented. In this approach, regional path terms are constrained using the strong-motion data. The constraint on path terms also makes this methodology suitable for the development of GMPEs that permit the removal of the ergodic assumption on path. The analysis results indicate the viability of the proposed methodology for constraining regional path terms and provide an estimate of single-station, single-path standard deviations. In addition, results confirm that the attenuation rate in volcanic regions is significantly higher than in nonvolcanic regions. Finally, a moderate correlation coefficient was found between the attenuation rate for weak and strong ground motions.

Introduction

The ergodic assumption is used in the development of most of the currently available ground-motion prediction equations (GMPEs). The ergodic assumption implies that the standard deviation applicable to a specific site-path-source combination is equal to the standard deviation estimated using the whole database (Anderson and Brune, 1999). The use of the ergodic assumption can result in an over-prediction of the standard deviation for cases in which repeatable site, source, or path effects can be measured or estimated. Larger standard deviations can have a large impact on hazard estimates, in particular for critical facilities that are designed for long return periods (Bommer and Abrahamson, 2006). A better quantification of the ground-motion uncertainty could be achieved by releasing the ergodic assumption. In a nonergodic probabilistic seismic hazard analysis (PSHA), source, site, and path terms are constrained and their uncertainty is treated as epistemic, with a consequent reduction of the aleatoric standard deviation (Al Atik *et al.*, 2010). Al Atik *et al.* (2010) also differentiated between fully and partially nonergodic PSHA, depending whether all (source, site, and path) or only some (e.g., site) components are constrained. The quantification of source, path, and/or site terms requires dense instrumental networks operating in seismically active zones so that a sufficient number of recordings are made. Only recently, with the advent of networks such as the Japanese KiK-net network (Okada *et al.*, 2004) or the Shake-Map network in California, has this become possible. Recent publications reflect the trend toward the development of GMPEs applicable to nonergodic PSHA (e.g., Bindi *et al.*, 2000; Chen and Tsai, 2002; Atkinson, 2006; Morikawa *et al.*, 2008; Anderson and Uchiyama, 2011; Lin *et al.*, 2011; Rodriguez-Marek *et al.*, 2011).

The main objective of this study is to present a methodology to account for regional dependences of anelastic attenuation rates within the format used by GMPEs used in engineering applications to reduce the scatter in predicting ground-motion parameters. The constraint on regional-dependent attenuation allows for constraining path terms for nonergodic PSHA. Several studies had been conducted to quantify path effects to reduce the scatter that result from the different travel paths of the waves through the Earth's crust (e.g., Campillo and Plantet, 1991; Zhang and Lay, 1994; Rodgers *et al.*, 1997, 1999; Fan and Lay, 1998; Phillips *et al.*, 1998, 2001; Phillips, 1999; Fan *et al.*, 2002; Ojeda and Ottemoller, 2002; Taylor *et al.*, 2002; Edwards *et al.*, 2008; Pasyanos, Matzel, *et al.*, 2009; Pasyanos and Walter, 2009; Pasyanos and Walter, *et al.*, 2009; Ford *et al.*, 2010). These studies covered a wide range of regions around the world (e.g., Central China, Phillips *et al.*, 1998 and Phillips, 1999; Colombia, Ojeda and Ottemoller, 2002; France,

Campillo and Plantet, 1991; United Kingdom, Edwards *et al.*, 2008; Middle East, Rodgers *et al.*, 1997 and Pasyanos, Matzel, *et al.*, 2009). A wide variety of approaches had been adopted in these studies. According to Phillips (1999), path terms are generally constrained either by investigating the correlations with path characteristics (e.g., topography, crustal depth, among others) or by removing the known effects from the records (e.g., event characteristics, site characteristics, and distance), then the path effects are constrained from the residuals. The approach we adopt in this study falls within the second category as we remove the effects of site response, geometric attenuation, and source, while at the same time constraining path effects using a mixed-effects analysis. An important characteristic of the approach we adopt in this study is that the analysis output has the form of GMPEs used for engineering PSHA studies.

The methodology presented in this study is applied to the aftershocks from the M_w 9.0 Tohoku earthquake; the density of aftershocks coupled with the density of stations in the KiK-net network creates a large enough dataset to constrain site and path terms in a nonergodic GMPE. Of particular importance for this study is the volcanic arc that crosses Japan. Past studies observed a higher attenuation rate across volcanic arcs for Japan (Kanno *et al.*, 2006; Zhao, 2010), New Zealand (Cousins *et al.*, 1999; McVerry *et al.*, 2006), and across volcanic regions in Colombia (Ojeda and Ottemoller, 2002). The GMPE developed in this study uses a simple functional form intended only for constraining site and path terms. The results show the viability of the method in developing GMPEs for use in PSHA. However, the GMPE developed in this study is not intended for ground-motion prediction in future events, because it uses a simplified functional form that does not include moment magnitude or closest distance to the fault rupture (part of the criteria listed by Bommer *et al.* [2010] for GMPEs to be used in PSHA), and is developed with earthquakes from a single source zone. Moreover, the GMPE is developed with an emphasis on interpolation of existing data, and no control is placed on the functional form to extrapolate to regions of sparse data that might be important for engineering applications. Finally, this GMPE is developed for downhole records, which have limited use in engineering applications.

The manuscript starts by summarizing the database used and the processing protocol implemented on the raw data downloaded from the KiK-net website. A description of the proposed methodology to account for path effects is then introduced, followed by a description of the selected GMPE functional form and the results of the regression analysis. The resulting path attenuation rates for the volcanic regions are compared with the rates for nonvolcanic regions to test if the differences in attenuation rates are statistically significant. We also compute an estimate of single-station, single-path standard deviation. Finally, we present a discussion of the correlation coefficient between the attenuation rates for small, moderate, and large magnitudes and the possible implications of this correlation.

Database

The KiK-net stations consist of two co-located strong-motion seismographs at the surface and at depth. The depth of the downhole instrument varies, but it is generally located at either 100 m or 200 m depth. The two instruments are triggered once a threshold acceleration is exceeded in the downhole instrument (Okada *et al.*, 2004). The cutoff frequency of the instruments is 30 Hz (Aoi *et al.*, 2004). The raw ground-motion data was downloaded from the KiK-net website. A simple processing protocol was applied to each strong-motion record as follows:

- The mean of all data points was first removed from the entire record.
- A fourth-order high-pass Butterworth acausal filter with a corner frequency of 0.1 Hz was applied (Kanno *et al.*, 2006). Several approaches could be applied to choose an appropriate corner frequency for each record (e.g., Boore and Bommer, 2005), but we found that the effect of the corner frequency of 0.10 Hz on spectral accelerations at 5% damping at oscillator periods of 1.0 and below was minimal. The high-pass filter serves as a baseline correction to remove low-frequency noise (Boore and Bommer, 2005).
- All the acceleration time histories were visually inspected and motions with multiple wave trains were removed.
- A noise window was manually picked for each motion. The noise window is defined as the length of the background noise recorded by the instrument before the first *P*-wave arrival. In the cases where no background noise was available, the last 10 s of the motion that followed the coda motion were used as the noise window.
- The Fourier amplitude spectra of the whole motion and the noise window were then computed. A Konno–Ohmachi (Konno and Ohmachi, 1998) smoothing window (with $b = 40$) was applied to the two Fourier amplitude spectra. The ratio between the whole-motion and the noise-window smoothed Fourier spectra was then calculated. Motions for which the signal-to-noise ratio dropped below 3 (applied to smoothed Fourier amplitude spectra) in a frequency range of 0.7–20 Hz were removed from the database. We chose 20 Hz as the maximum usable frequency because many records do not fulfill the signal-to-noise ratio criterion above this frequency.
- Pseudoacceleration response spectra at 5% damping were computed for each time history using the piecewise exact method (Nigam and Jennings, 1969) and the geometric mean of the two recorded downhole horizontal components were used in the development of the GMPE. We used the downhole components to minimize the effect of soil nonlinearity on the analysis. This is especially because of the wide range of event's magnitudes used in this study.

Figure 1 shows the location of the epicenters for earthquakes used in this study and the strong-motion stations for
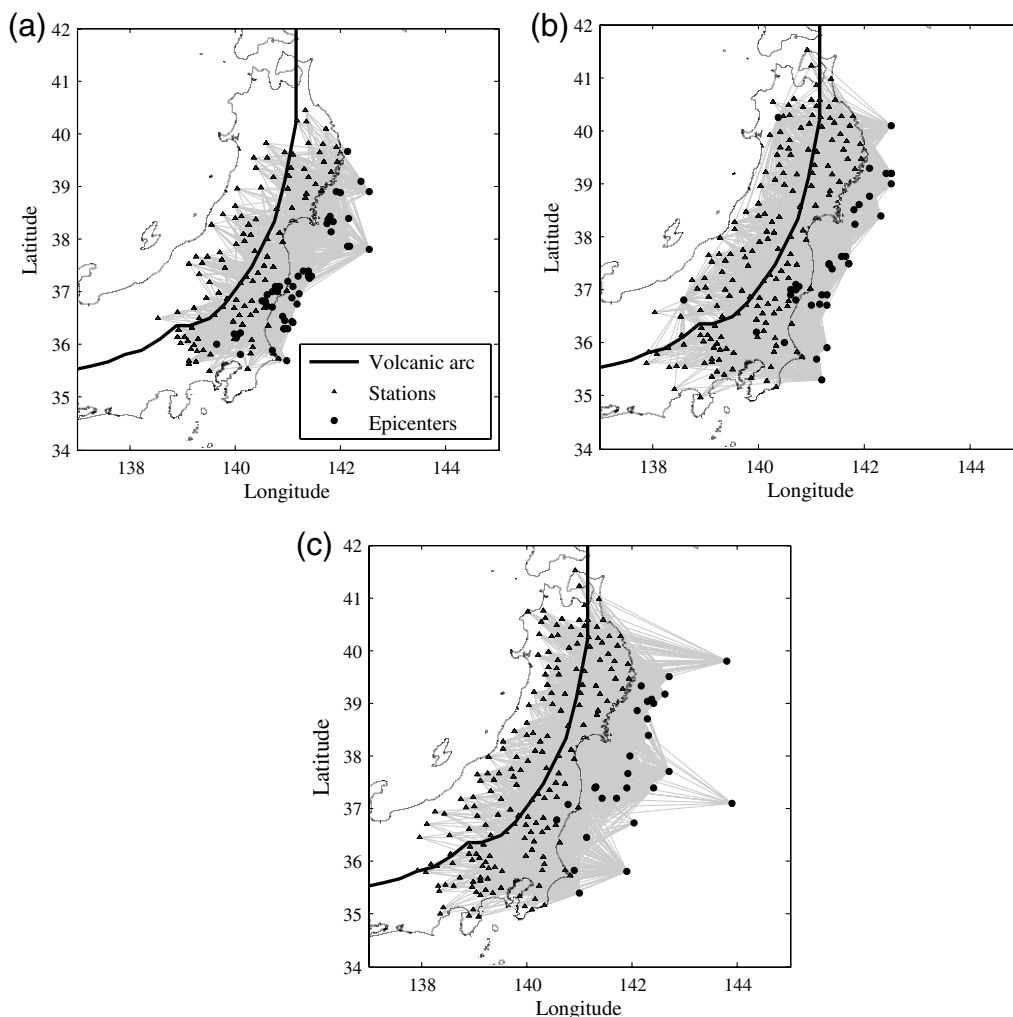
Figure 1.   Map of Japan showing the earthquake epicenters (circles), recording stations (triangles), and travel paths used in the analyses. Data is separated in three magnitude bins: (a) $4 \leq M_{JMA} < 5$, (b) $5 \leq M_{JMA} < 6$, and (c) $6 \leq M_{JMA} < 7$.

three magnitude bins (i.e., $4 \leq M_{JMA} < 5$, $5 \leq M_{JMA} < 6$, and $6 \leq M_{JMA} < 7$). In this study we look at attenuation differences between volcanic and nonvolcanic regions. Figure 2 shows the location of the active volcanoes in the region investigated in this manuscript along with the Japanese volcanic arc. The locations of active volcanoes were obtained from the National Institute of Advanced Industrial Science and Technology (AIST) website (see Data and Resources). Figure 3 shows the distribution of magnitude (M_{JMA}) versus focal depth for the events included in the study and the distribution of magnitude and epicentral distance for all the records included in the analysis. The maximum epicentral distance used was 300 km. Stations that recorded less than four motions in a magnitude bin were not used in the analysis of that magnitude bin. Moreover, events with less than four records were also removed. A total of 7242 ground-motion records recorded from 117 aftershocks fulfilled these criteria. Table 1 presents a summary of the database used in this study.

Proposed Methodology to Account for the Path Effects

The anelastic attenuation rate is expected to be different along each source-to-site path. This difference can be attributed to differences in travel path geology or the presence of faults. Surface topography can also have an effect in attenuation of surface waves. In particular, in Japan and other subduction regions the presence of active volcanoes with magma chambers can result in significantly larger attenuation rates through volcanic regions compared to other regions (Cousins et al., 1999; Ojeda and Ottemoller, 2002; McVerry et al., 2006; Zhao, 2010). Despite these effects, current GMPEs use an average anelastic attenuation rate independent of the source-to-site travel path of seismic waves. This is generally necessary because very large and well sampled databases are necessary to constrain regional differences in attenuation rates. Moreover, GMPEs developed for a specific region are generally also used in other regions with similar tectonic
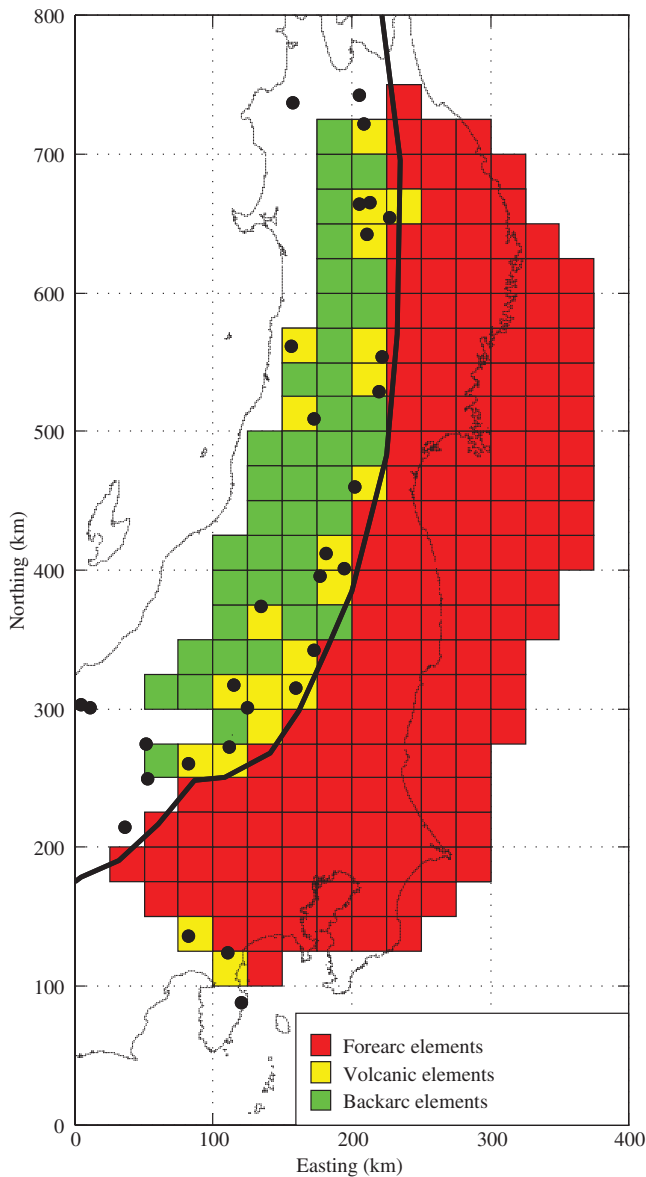
Figure 2. Map of Japan showing the locations of the active volcanoes (black circles), the volcanic belt (black line), and the different elements used in the analysis. Elements categorized as forearc, volcanic, and backarc are colored in red, yellow, and green, respectively. (The origin of the axes is located at 34.2113 N and 137.7016 E.)

Figure 3. Plots of M_{JMA} magnitude for the events used in this study versus: (a) the focal depth and (b) epicentral distance for each recorded ground motion.

environments. For example the GMPE developed by Zhao *et al.* (2006) for subduction earthquakes in Japan was adopted by USGS in the seismic hazard analysis of the Cascadia region (Petersen *et al.*, 2008).

In this study an approach is proposed to develop GMPEs that account for the effects of different attenuation rates for each source-to-site path. This approach consists of dividing the region that contains all source-to-site paths of all records into equal squares (from this point on, these squares will be referred to as elements). The travel distance through each of the elements is calculated for each recording assuming that the path of the waves from the source's epicenter to the site is

a straight line. The path attenuation of each element is then considered separately and becomes a term in the GMPE functional form. This approach can capture the anomalous attenuation behavior of some regions and has the potential to constrain the standard deviation needed for nonergodic seismic hazard analysis. This approach requires a large number of earthquakes recorded by a dense array of strong ground-motion stations to be able to constrain the attenuation rates for each element; the Tohoku earthquake aftershock sequence provides an ideal dataset to test the validity of the proposed methodology.

The elements used in this study were set to 25 km squares. The element size has to be small enough such that particularities in the attenuation rate can be captured but large enough so that a significant amount of source-to-site paths pass through the elements. For example, in a first iteration of this study we used 100 km square. At this scale, the effect of the volcanoes on the attenuation rates was not captured well. On the other hand, using elements with size smaller than 25 km resulted in convergence and computational issues. The

Table 1
Summary of the Database Used in This Study

	All Bins	$4 \leq M_{JMA} < 5$	$5 \leq M_{JMA} < 6$	$6 \leq M_{JMA} < 7$
Number of records	7242	2259	2948	2035
Number of stations	187	109	167	182
Number of events	117	54	36	27

final choice reflected a balance of available data and the desired resolution considering the size of the region of interest.

To test the hypothesis that the attenuation rate through volcanic regions is higher than in other regions, elements that contain volcanoes were identified (see Fig. 2). In the regres-

sion analyses, the events were sorted into three magnitude bins ($4 \leq M_{JMA} < 5$, $5 \leq M_{JMA} < 6$, and $6 \leq M_{JMA} < 7$). Figure 4 shows the elements used for each magnitude bin and the number of passes for each element. The elements with the maximum number of passes have 656, 662, and 314 passes for the three magnitude bins, respectively.
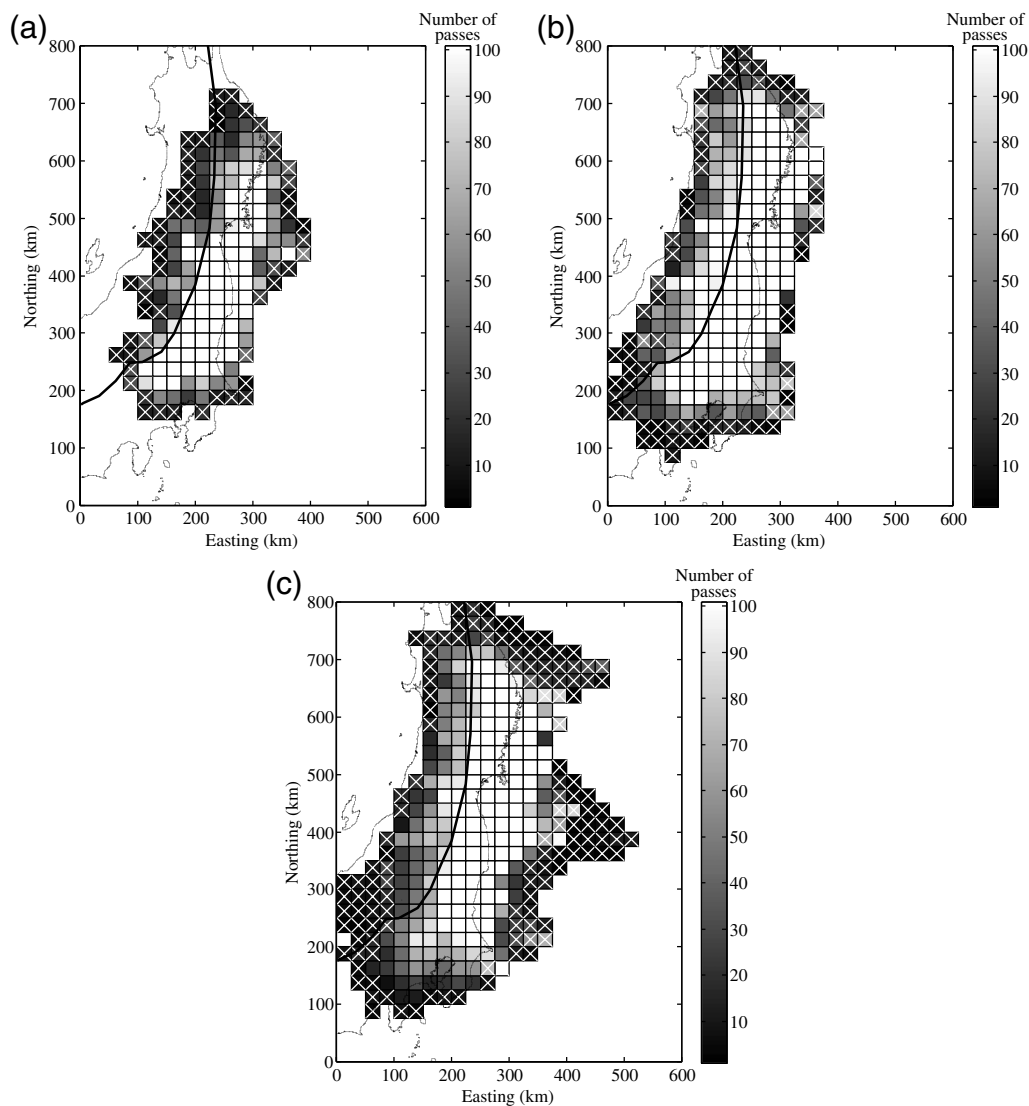
Figure 4. Map of Japan with the elements used in the analysis of each magnitude bin. Data is separated in three magnitude bins. (a) $4 \leq M_{JMA} < 5$, (b) $5 \leq M_{JMA} < 6$, and (c) $6 \leq M_{JMA} < 7$. (The element color represents the number of passes for each element. White elements are elements with more than 100 passes. The crossed elements are those with less than seven passes or those that are badly constrained. These elements were assigned an average attenuation value during the analysis. [The origin of the axes is located at 34.2113 N and 137.7016 E.])

The ground-motion parameters used in this study are the geometric mean of the pseudospectral accelerations (5% damping) of the as-recorded components of the downhole instruments. The records were processed as described in the Database section. The choice of downhole records over the surface records was made to reduce the effects of shallow site response on the ground motions. The general form of the GMPE model is given by

$$\ln(y_{\mathrm{es}}) = \ln(\mu_{\mathrm{es}}) + \delta S2S_{\mathrm{s}} + \delta B_{\mathrm{e}} + \delta WS_{\mathrm{es}}, \qquad (1)$$

where y_{es} is the ground-motion parameter for the recorded motions in units of gravity (given as a ratio of 9.81 cm/s^2) for event e at site s, μ_{es} is the median prediction; $\delta S2S_{\mathrm{s}}$ is the site term and represents the average deviation from the median prediction of the recorded ground motions at site s, δB_{e} is the between-event residual and represents the observed shift from the median for event e, and δWS_{es} is the within-event single-station residual that corresponds to the difference between the recorded and the predicted spectral acceleration (accounting for event and site terms) for each record in the database. The residual terms ($\delta S2S_{\mathrm{s}}$, δB_{e}, and δWS_{es}) are assumed to be zero-mean random variables with standard deviations equal to ϕ_{S2S}, τ, and ϕ_{SS}, respectively. The notation follows that proposed by Al Atik *et al.*, (2010).

The functional form of the median prediction used in this study is given by

$$\ln(\mu_{\mathrm{es}}) = C_1 + C_2 \ln(R) + C_3 D + C_4 \ln(V_{S30}/V_{\mathrm{ref}})$$
$$+ \sum_{i=1}^{n} \delta_i P_{i\mathrm{es}}, \qquad (2)$$

where R is the epicentral distance in km, D is the focal depth of the earthquake in km, V_{S30} is the average shear-wave velocity over the top 30 m at the recording site, V_{ref} is a reference shear-wave velocity taken here as 760 m/s, $P_{i\mathrm{es}}$ is the distance (in km) through element i for a straight line path originating at the epicenter of source e to site s, δ_i is the attenuation per kilometer of travel in the ith element, and n is the total number of elements into which the region under study was divided. Both V_{S30} and V_{ref} should have the same units. The term $C_2 \ln(R)$ represents geometric spreading. For simplicity, geometrical spreading was assumed to follow a theoretical model and the coefficient C_2 was assumed to be -1 and was not regressed for. This assumption may introduce errors because the theoretical value ($C_2 = -1$) applies to Fourier coefficients for a point source and not to spectral amplitudes for finite sources. Moreover, stochastic simulations have shown that the value of the coefficient changes with distance to accommodate Moho reflections. The last term in equation (2) is generally referred to as the "anelastic attenuation term"; however, it serves to capture all attenuation effects not accounted for by the geometric attenuation term. This includes both the effects of true anelastic attenuation and the effects of scattering.

Although the proposed GMPE is developed using downhole records, V_{S30} is included in the parameterization. Rodriguez-Marek *et al.* (2011) found that V_{S30} has an effect on the ground motions even at the downhole level due to waves reflected from the surface. On the other hand, the effect is less relevant for downhole motions compared with surface motions. A magnitude term was not included in the predictive terms of the GMPE, except for the magnitude grouping previously indicated (e.g., different sets of parameters were developed for different magnitude ranges). This choice was dictated in part by the unavailability of moment magnitude estimates for most of the earthquakes in the database and in part due to a desire to simplify the regression analyses. Note, however, that any magnitude dependency of the attenuation terms is captured by the magnitude binning, and magnitude scaling is captured in the event terms (δB_{e}). This implies that event-corrected residuals should not have any magnitude bias.

A mixed-effects regression analysis was carried out to estimate the coefficients for different spectral periods. Commercial statistical analysis software (SAS) was used to accomplish this task. C_1, C_3, C_4, and δ_i were the fixed effects, while $\delta S2S_{\mathrm{s}}$ and δB_{e} were the random effects in the mixed effects analysis. The maximum likelihood estimation method was used. The analysis was carried following an iterative procedure. On the first iteration, the average attenuation for the volcanic and nonvolcanic elements is calculated. For subsequent iterations, the attenuation rates for all elements with less than seven passes were replaced with the average attenuation rate of all elements computed in the first iteration, and these elements were excluded from the regression. The elements on the edges of the region under study were badly constrained and resulted in poorly constrained attenuation rates. In order to minimize the effect of such poorly constrained elements on the regression analysis, the average attenuation was assigned to them and these elements were also removed from the regression analysis. Each excluded element was assigned the average value of the group to which it belongs (volcanic or nonvolcanic). Additional iterations were conducted until the assumed average attenuation assigned to the elements excluded from the regression matched the average attenuation of the elements that were included in the regression. A total of 55, 75, and 135 elements for the three magnitude bins were assigned the average attenuation. The total number of elements that enclosed all paths for the three magnitude bins are 184, 266, and 340, respectively (Fig. 4). Table 2 summarizes the values of the parameters calculated from the mixed-effects analysis. Figure 5 shows the path-attenuation terms for the nonvolcanic elements at spectral periods of 0.3 and 1.0 s. Figure 6 shows the estimated attenuation rates through the different elements at a spectral period of 0.3 s for the three magnitude bins. Although not shown for brevity, the site terms ($\delta S2S_{\mathrm{s}}$) and the site-corrected within-event residuals (δWS_{es}) had zero mean and were unbiased with respect to the predictive variables. The event terms (δB_{e}) had zero mean, but there

Table 2
Model Parameters from the Mixed-Effects Analysis

Period (s)	Bin 1 ($4 \leq M_{JMA} < 5$)			Bin 2 ($5 \leq M_{JMA} < 6$)			Bin 3 ($6 \leq M_{JMA} < 7$)		
	C_1	C_3	C_4	C_1	C_3	C_4	C_1	C_3	C_4
0.05	−1.6138	0.0131	−0.0579	−0.1552	0.0005	−0.1442	0.8436	−0.0051	−0.2730
0.075	−1.5189	0.0138	−0.1050	0.0636	0.0004	−0.1558	1.0551	−0.0034	−0.2846
0.1	−1.6031	0.0132	−0.1332	0.1674	−0.0013	−0.1989	1.1684	−0.0040	−0.3124
0.15	−1.8663	0.0096	−0.2597	0.0795	−0.0010	−0.3134	1.1869	−0.0055	−0.3830
0.2	−2.1183	0.0086	−0.2573	−0.0462	−0.0023	−0.3340	1.1043	−0.0058	−0.3954
0.25	−2.3412	0.0078	−0.2672	−0.2521	−0.0021	−0.3696	1.0079	−0.0066	−0.4129
0.3	−2.5842	0.0078	−0.2754	−0.3666	−0.0024	−0.3716	0.8589	−0.0061	−0.4101
0.4	−2.9233	0.0070	−0.2889	−0.6651	−0.0008	−0.3397	0.6841	−0.0060	−0.3486
0.5	−3.2089	0.0059	−0.2405	−0.8803	−0.0023	−0.3147	0.5600	−0.0087	−0.3213
0.75	−3.8622	0.0035	−0.1721	−1.3021	−0.0055	−0.2228	0.2073	−0.0070	−0.2327
1	−4.3906	0.0023	−0.2223	−1.7566	−0.0062	−0.3052	−0.1470	−0.0058	−0.2576

was a linear trend with respect to M_{JMA}. This bias should not affect our conclusions.

We checked the adequacy of the size of elements and the database distribution by replacing the estimated attenuation rates through the different elements by a randomly generated set with a mean and standard deviation that matches the mean and standard deviation of the attenuation rates estimated using the real database. A set of normally distributed random numbers with zero mean and standard deviation of 0.037 was used as the event- and site-corrected residuals. We then ran the regression analysis as described in the manuscript to estimate the attenuation structure. The resulting attenuation structure was nearly identical to the randomly generated one.

Results

This section discusses the results of this study. The discussion is separated into three subsections. First, the standard deviations resulting from the regression analysis are discussed. Then, inferences on different regional attenuation rates are presented based on the location relative to the volcanic arc. Finally, the correlation of the path attenuation and site terms between different magnitude ranges is discussed.

Standard Deviations for Nonergodic PSHA

The GMPE described by equations (1) and (2) includes both site terms ($\delta S2S_s$) and event terms (δB_e), and allows for a breakdown of uncertainty into its components (e.g., the
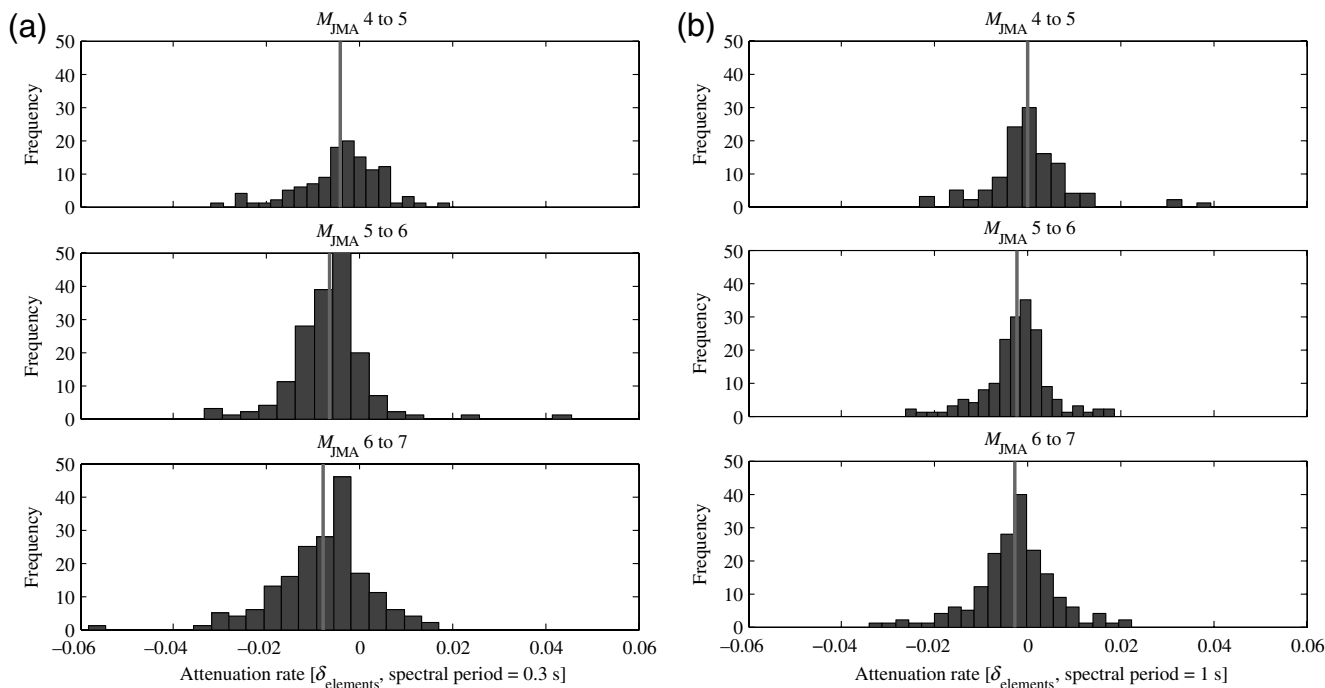
Figure 5. Histograms for the estimated attenuation rate ($\delta_{elements}$) through the nonvolcanic elements for spectral periods (a) 0.3 s and (b) 1.0 s. The vertical lines represent the average attenuation rate for all the elements shown in each histogram.
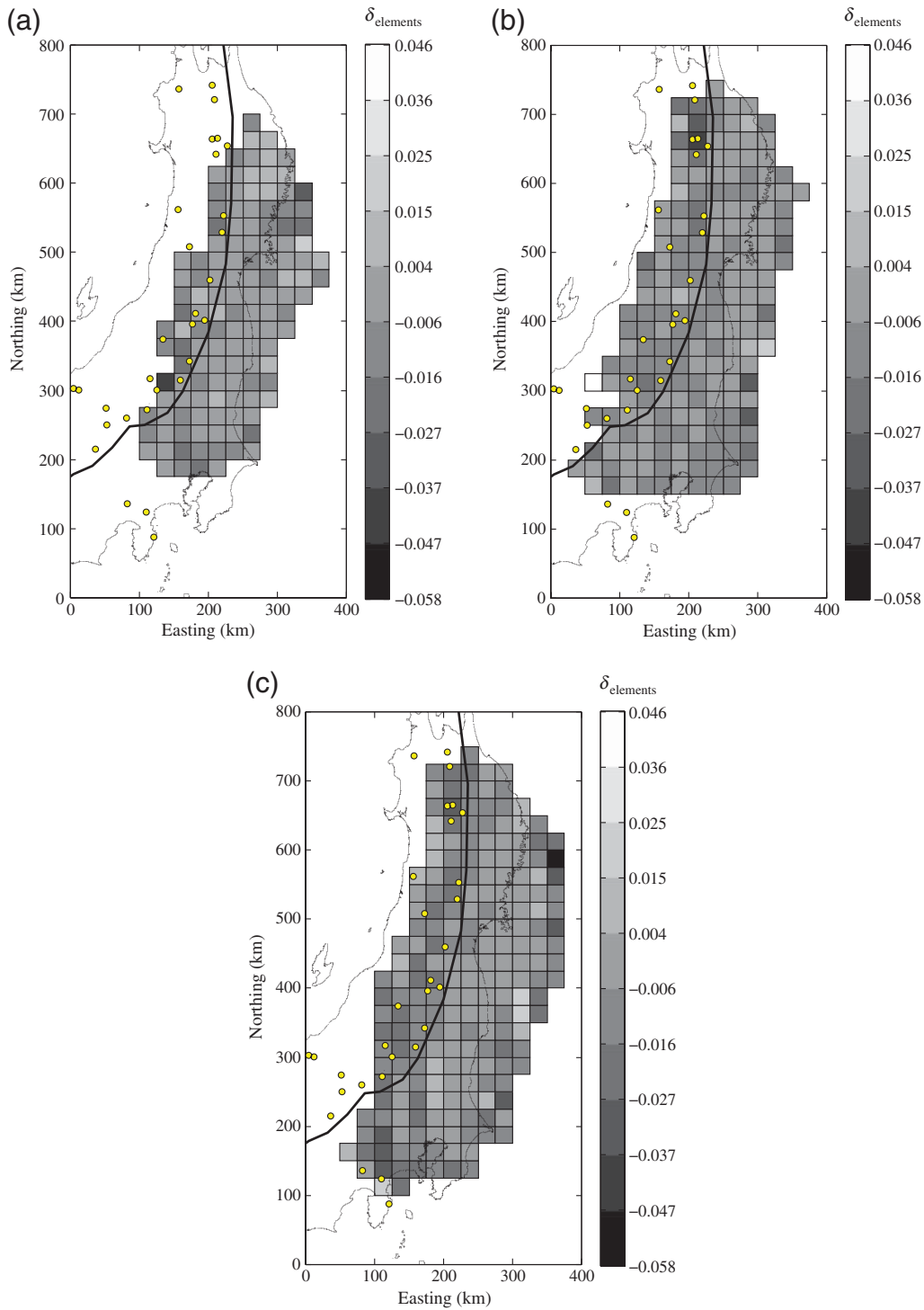
Figure 6. Map of Japan with the elements used in the analysis. The element color represents the attenuation rate (δ_{Elements}) estimated for each element at a spectral period of 0.30 s for: (a) $4 \le M_{\text{JMA}} < 5$, (b) $5 \le M_{\text{JMA}} < 6$, and (c) $6 \le M_{\text{JMA}} < 7$. The active volcanoes are shown as circles. (The origin of the axes is located at 34.2113 N and 137.7016 E.)

between-event standard deviation, τ; the site-to-site standard deviation, ϕ_{S2S}; and the residual or single-station, within-event standard deviation, ϕ_{SS}. In the exercise presented in this paper, magnitude was not introduced in the regression analysis. The magnitude dependence that is not captured by the coarse magnitude binning used is captured in the event term. As a consequence, the standard deviation of the event terms (τ) is artificially inflated and should not be compared with those resulting from more elaborate regression analyses. On the other hand, the site-to-site standard deviation (ϕ_{S2S})

and the single-station, within-event standard deviation (ϕ_{SS}) should be reliable estimates of the variability for the data used in this study. Figure 7 shows these standard deviations for the three magnitude bins used in this study. The site-to-site standard deviation does not show a clearly defined magnitude dependency. On the other hand, the single-station, within-event standard deviation does show a stronger magnitude dependency, in particular for long periods. A similar dependency was observed for worldwide data by Rodriguez-Marek and Cotton (2011).

The introduction of regional path terms decreased the site and residual standard deviation components with respect to an analysis where the anelastic attenuation term is unique for the entire region. The reduction ranged from 4% to 8% for M_{JMA} between 4 and 5, and from 8% to 19% for M_{JMA} between 6 and 7. These reductions (expressed as ratios for each standard deviation component from before to after adding the regional path terms) are shown in Figure 8. This reduction can be attributed to using path-specific anelastic
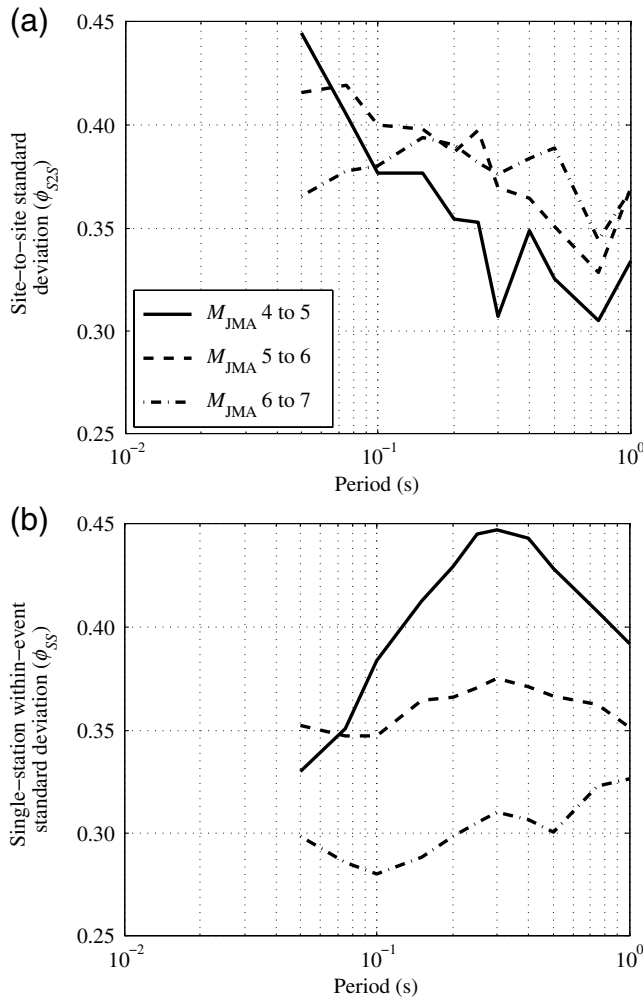
Figure 7. Components of standard deviation calculated from the three magnitude bins for: (a) the site-to-site standard deviation (ϕ_{S2S}); and (b) the single-station, within-event standard deviation (ϕ_{SS}).
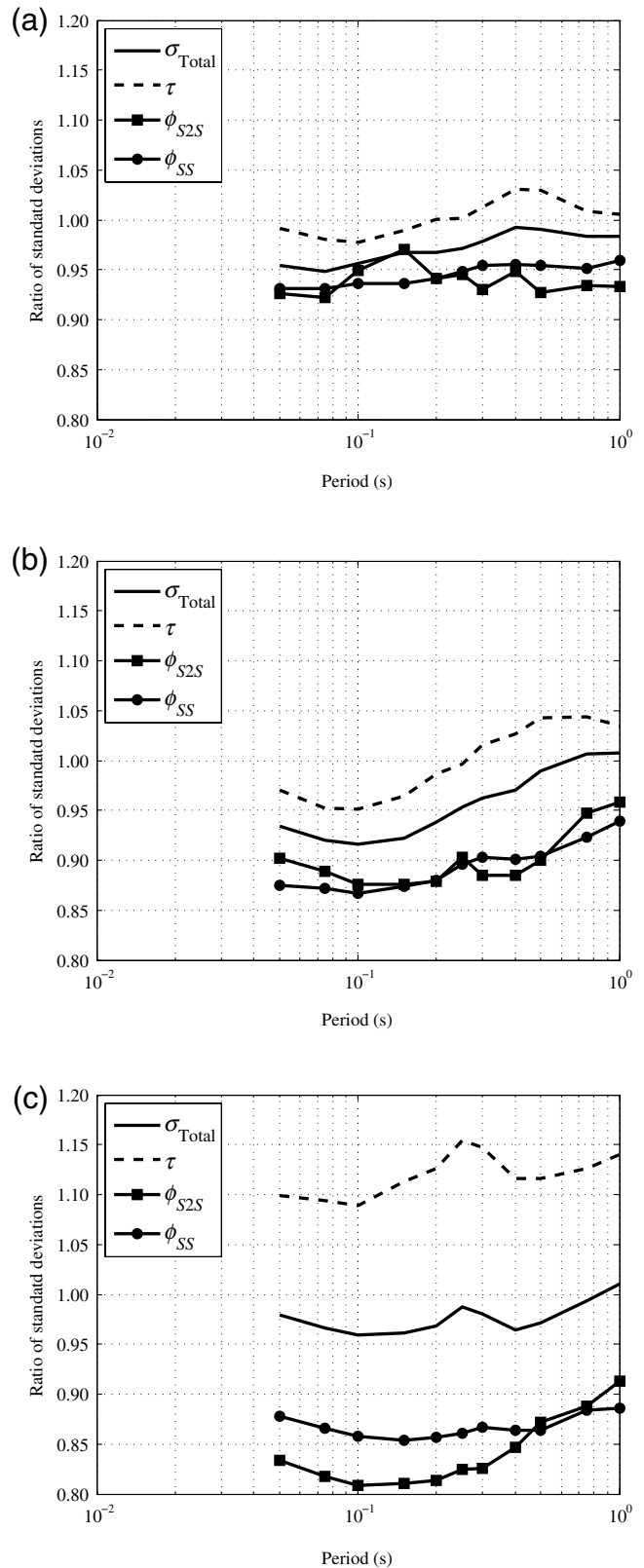
Figure 8. Ratio of the total (σ_{Total}), site-to-site (ϕ_{S2S}), event (τ), and the single-station, within-event(ϕ_{SS}) standard deviations for an analysis including the path specific attenuation terms to their counterpart standard deviations calculated from an analysis that does not include the path-specific attenuation terms for: (a) $4 \leq M_{JMA} < 5$, (b) $5 \leq M_{JMA} < 6$, and (c) $6 \leq M_{JMA} < 7$.

attenuation for each source-to-site combination instead of an average anelastic attenuation for the entire region. On the other hand, the between-event standard deviation increases when the path terms are introduced in equation (2). This increase is possibly due to a better constraint of the event terms.

The use of the regional path terms implies that anelastic attenuation varies for different source–site combinations. Under the assumption that the path terms are constant from event to event, this would be equivalent to a deterministic constraint on path terms for prediction of future earthquakes. This also implies that the resulting standard deviation would represent a single-station, single-path standard deviation. This is illustrated in Figure 9, where the resulting residual (or single-station, single-path within-event) standard deviations are compared with those obtained by Rodriguez-Marek *et al.* (2011) for a study on the KiK-net database (single-station within-event only). Note that the standard deviations obtained in the present study are consistently and significantly lower than those estimated by Rodriguez-Marek *et al.* (2011) except for the first magnitude bin at longer periods, possibly representing the reduction from single-station to single-station, single-path standard deviations.

Forearc Versus Backarc Rate of Attenuation

A flag was used to differentiate between the forearc elements (east of the volcanic arc), the elements that contain volcanoes, and the backarc elements (west of the volcanic arc; Fig. 2). The average attenuation rate for each subgroup for the three M_{JMA} bins is shown in Figure 10. The figure shows that elements that contain volcanoes have a larger attenuation compared with the elements without volcanoes. A *t*-test was performed for the estimates of the attenuation rate. The attenuation rate of the elements with volcanoes was

found to be higher than the elements without volcanoes at a 95% confidence level (i.e., the null hypothesis that volcanic and nonvolcanic elements have the same average attenuation
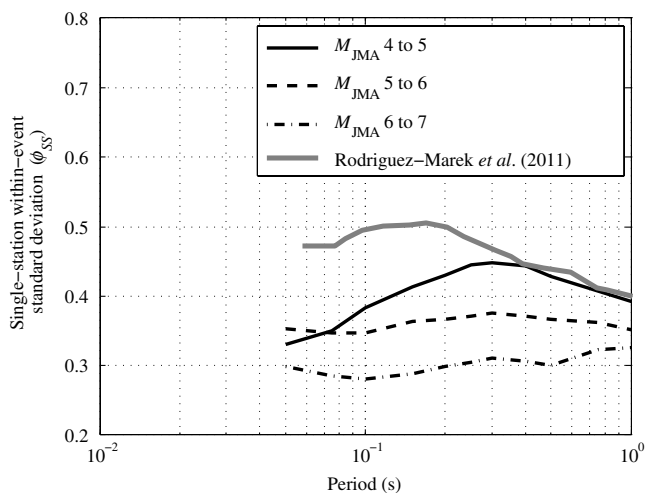
Figure 9. Single-station within-event standard deviation that accounts for path effects (ϕ_{SS}) obtained in this study for three magnitude bins compared to the ϕ_{SS} values from Rodriguez-Marek *et al.* (2011).
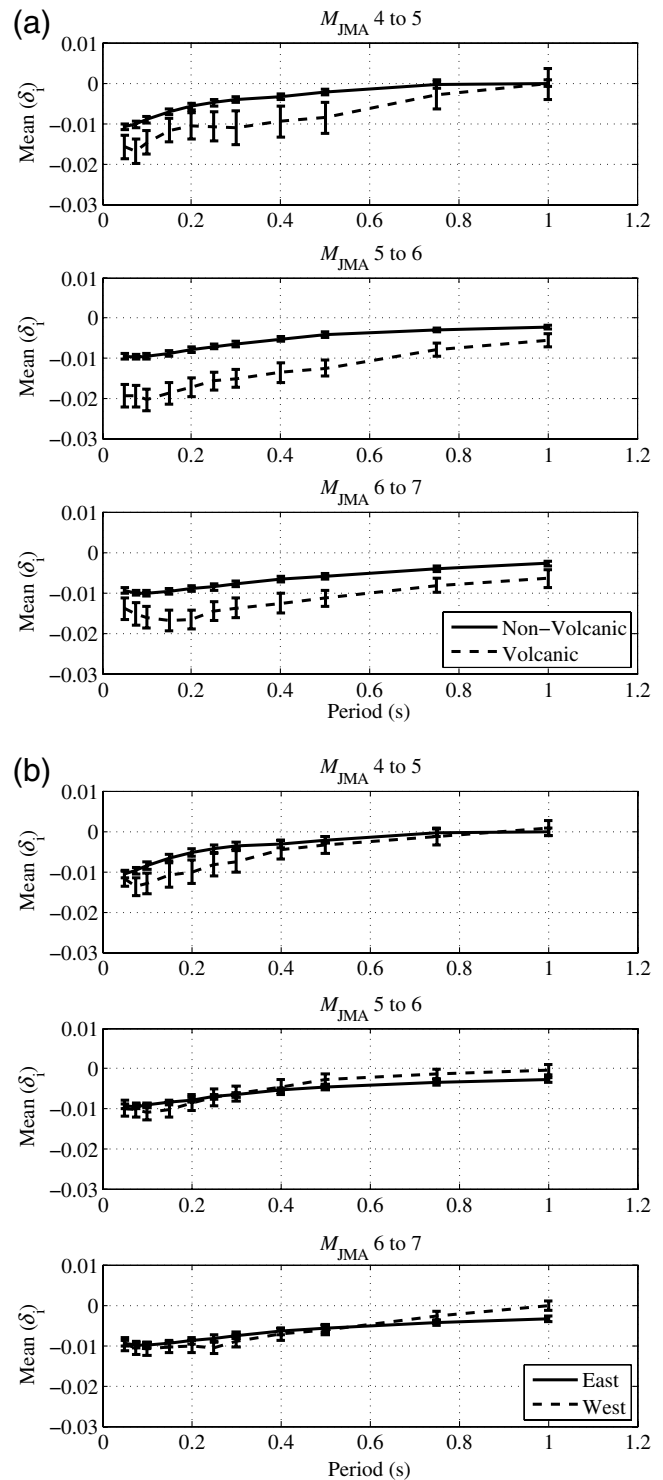
Figure 10. Average attenuation rates for the three magnitude bin at all periods for: (a) volcanic and nonvolcanic elements and (b) forearc and backarc elements. The error bars represent the standard error of the estimate of the attenuation rate for each subset of elements.
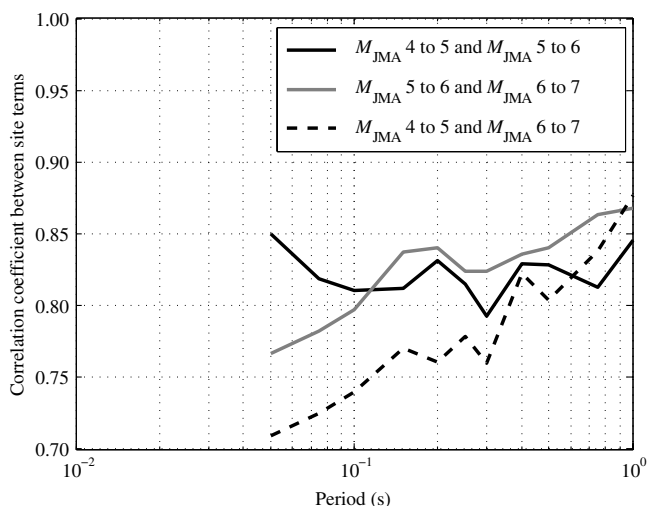
Figure 11. Correlation coefficients between site terms (δ_{S2S}) estimated for different magnitude bins.

Figure 12. Correlation coefficients between attenuation rates in the elements ($\delta_{elements}$) estimated from different magnitude bins.

rate was rejected), for all oscillator periods of the three magnitude bins except at 0.05, 0.15, 0.20, 0.75, and 1.00 s for the first magnitude bin ($4 \leq M_{JMA} < 5$) and 1 s for the third bin ($6 \leq M_{JMA} < 7$) (i.e., the null hypothesis that volcanic and nonvolcanic elements have the same average attenuation rate could not be rejected). The fact that the difference in attenuation rate is not significant at long periods could be attributed to a lower sensitivity of longer wavelengths to lateral heterogeneities across the volcanic chambers. Cousins *et al.* (1999) found that the high attenuation rate through the volcanic region in New Zealand is clearer for frequencies 1 Hz and above (Fourier domain). In the current study, the analysis of the first bin (lower magnitudes) shows that the difference in attenuation is not significant at some of the shorter periods. This could be attributed to the small number of volcanic elements involved in the analysis (11 elements), and the relatively lower number of passes beyond the volcanic arc (Fig. 1).

At the same confidence level, the attenuation in the forearc and backarc regions was found to be similar for all periods except 1 s for the two larger magnitude bins. Cousins *et al.* (1999), McVerry *et al.* (2006), and Zhao (2010) also used the same anelastic attenuation for forearc and backarc regions, while observing differences in attenuation for paths across the volcanic regions. The findings also confirm that the database used in this study was sufficiently large to constrain the attenuation rates for each element. Figure 10 also shows that the attenuation rate decreases at large spectral periods. This decrease is consistent with the current understanding of anelastic attenuation.

As discussed earlier in the manuscript, assuming a theoretical geometric spreading with a single slope might not strictly follow strong-motion observations. Various studies have characterized geometric spreading using two different slopes and a horizontal plateau that results from the Moho reflection. To test if using the theoretical geometric spreading

is responsible for the resulting attenuation rates, we also used the geometrical spreading reported in Atkinson and Mereu (1992). In their study, the slope of the geometric term is about −1 for distances below 70 km; between 70 and 130 km a horizontal plateau is used; and above 130 km the slope is −0.5. We tested this model for three different periods (0.05, 0.3, and 0.75) for the three magnitude bins. The attenuation rates that resulted from these analyses were strongly correlated with their counterparts that resulted from using a single slope for geometric spreading. The attenuation rate for forearc and backarc elements was not statistically different for the analysis using the Atkinson and Mereu (1992) decay coefficients. This matches the results obtained using a single decay coefficient. The only difference is that the attenuation rate through the volcanic elements was not statistically lower than the nonvolcanic elements for spectral periods of 0.05 and 0.75 s for the larger magnitude bin, while they were statistically lower when theoretical single geometric spreading coefficient was used. Based on these observations, the authors do not believe that assuming theoretical geometric spreading coefficient of −1 affects the main conclusions of the study.
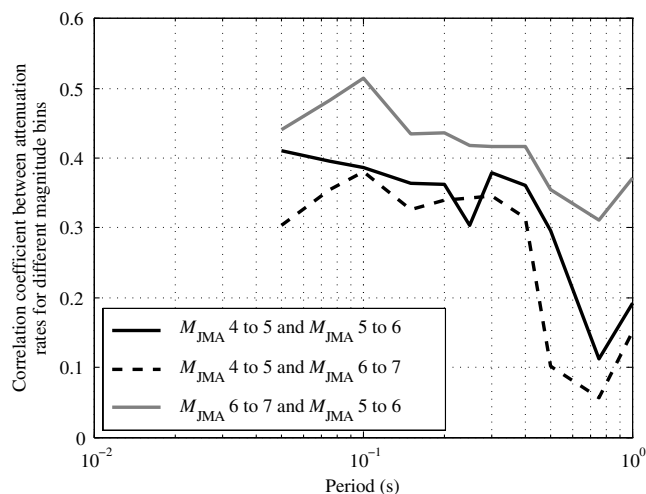
Correlation Coefficients between Attenuation Rates for Different Magnitude Ranges

The estimation of site or path terms for nonergodic PSHA requires extensive and well sampled datasets. For this reason, previous studies that constrain event terms make use of a wider range of magnitudes than those used in GMPEs for engineering applications (e.g., Rodriguez-Marek and Cotton, 2011). The implicit assumption is that site terms estimated from small magnitudes are applicable at large magnitudes (Rodriguez-Marek *et al.*, 2011). This hypothesis is tested in this study for the site and path terms. Figures 11 and 12 show the correlation coefficient between the different combinations of the magnitude bins ($4 \leq M_{JMA} < 5$,

$5 \leq M_{JMA} < 6$, and $6 \leq M_{JMA} < 7$), both for the site term (Fig. 11) and the elements' attenuation terms (Fig. 12). For site terms, the correlation coefficients are relatively strong at all periods (0.71–0.88). These high correlation coefficients favor the hypothesis that, in the absence of nonlinearity, site terms constrained using data from small magnitudes are applicable at large magnitudes. For the path terms, the correlation coefficients for periods below 0.4 s are between 0.30 and 0.51, which indicates a medium-to-low correlation. The correlation coefficients fall rapidly at periods larger than 0.40 s. The relatively low correlation values imply that careful accounting of the magnitude dependence of path-specific attenuation must be taken into account when extrapolating results from small magnitudes to large magnitudes.

Conclusions

In the current study, an approach to include the path effects in GMPEs is presented. In this approach, the region of interest is divided into equal squares (here named elements) and the travel distances in each of these elements is calculated for each source-to-site path. A mixed-effects regression analysis was used to constrain the attenuation rate for each element. The ground motions from 117 aftershocks from the Tohoku earthquake were used in the regression analysis. The most relevant conclusions from this study are:

- The attenuation rate in the elements that contained volcanoes was found to be higher than the elements without volcanoes. This difference is statistically significant at a 95% confidence level for most magnitudes and oscillator periods.
- The attenuation rate for nonvolcanic elements (forearc and backarc regions) was found to be the same at a 95% confidence level.
- The GMPE developed in this study resulted in single-station within-event standard deviations smaller than those of Rodriguez-Marek *et al.* (2011), which did not incorporate regional variations in path attenuation. This can be interpreted as indicating a reduction from single-station to single-path, single-station standard deviations.
- Strong correlation coefficients were found between the site terms estimated for low-magnitude events and large-magnitude events. Moreover, medium to low correlation coefficients were found between the attenuation rates of the elements estimated for low-magnitude events and large-magnitude events. This shows that small-magnitude events can provide valuable information about site terms for large-magnitude events to be used in nonergodic PSHA (in the absence of soil nonlinearity). More care must be used in extrapolating information on path terms.

Path attenuation was assumed to be independent of the source-to-site azimuth. In the dataset used in this study, most of the travel paths are east-to-west, as most of the sources are located offshore. This may have biased the estimates of path attenuation. However, most of the GMPE in subduction

regions were developed using a database with similar biases, hence this limitation is not significant. The GMPE developed in this study uses a simple functional form that does not account for finite-source effects, does not include a moment-magnitude term, is derived for downhole motions, and does not make use of an elaborate accounting of source effects. Further study is needed to account for finite sources, where the travel path from source to site is nonunique. This may not be the case when local geology presents preferential travel paths for seismic waves. On the other hand, the proposed methodology can be applied to more elaborate functional forms, where the path terms will represent "residual" attenuation left over after accounting for a global rate of geometric and anelastic attenuation. GMPEs developed in such a way can be used to remove the ergodic assumption on path.

Data and Resources

The KiK-net strong motions used in this study were provided by National Research Institute for Earth Science and Disaster Prevention (NIED) at (www.kik.bosai.go.jp). Data was last accessed in August 2011. The location of active volcanoes on the Japanese islands was extracted from The National Institute of Advanced Industrial Science and Technology (AIST) at the following link (http://riodb02 .ibase.aist.go.jp/strata/VOL_JP/EN/active_v.htm). Data was last accessed in July 2012. We used SAS software to conduct the mixed-effects statistical analysis.

Acknowledgments

References

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

Anderson, J. G., and J. Brune (1999). Probabilistic seismic hazard analysis without the ergodic assumption, *Seismol. Res. Lett.* **70,** 19–28.

Anderson, J. G., and Y. Uchiyama (2011). A methodology to improve ground-motion prediction equations by including path corrections, *Bull. Seismol. Soc. Am.* **101,** 1822–1846.

Aoi, S., T. Kunugi, and H. Fujiwara (2004). Strong-motion seismograph network operated by NIED: K-NET and KiK-net, *J. Japan Assoc. Earthq. Eng.* **4,** 65–74.

Atkinson, G. M. (2006). Single-station sigma, *Bull. Seismol. Soc. Am.* **96,** 446–455.

Atkinson, G. M., and R. F. Mereu (1992). The shape of ground-motion attenuation curves in southeastern Canada, *Bull. Seismol. Soc. Am.* **82,** 2014–2031.

Bindi, D., S. Parolai, D. Spallarossa, and M. Catteneo (2000). Site effects by H/V ratio: Comparison of two different procedures, *J. Earthq. Eng.* **4,** 97–113.

Bommer, J. J., and N. A. Abrahamson (2006). Why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates?, *Bull. Seismol. Soc. Am.* **96,** 1967–1977.

Bommer, J. J., J. Douglas, F. Scherbaum, F. Cotton, H. Bungum, and D. Fäh (2010). On the selection of ground-motion prediction equations for seismic hazard analysis, *Seismol. Res. Lett.* **81,** 783–793.

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

Campillo, M., and J. L. Plantet (1991). Frequency dependence and spatial distribution of seismic attenuation in France: Experimental results and possible interpretations, *Phys. Earth Planet. In.* **67,** 48–64.

Chen, Y.-H., and C.-C. P. Tsai (2002). A stable algorithm for regression analyses using the random effects model, *Bull. Seismol. Soc. Am.* **92,** 1984–1991.

Cousins, W. J., J. X. Zhao, and N. D. Perrin (1999). A model for the attenuation of peak ground acceleration in New Zealand earthquakes based on seismograph and accelerograph data, *Bull. New Zeal. Natl. Soc. Earthq. Eng.* **32,** 193–220.

Edwards, B., A. R. Rietbrock, J. J. Bommer, and B. Baptie (2008). The acquisition of source, path, and site effects from microearthquake recordings using Q tomography: Application to the United Kingdom, *Bull. Seismol. Soc. Am.* **98,** 1915–1935.

Fan, G.-W., and T. Lay (1998). Regionalized versus single-station waveguide effects on seismic discriminants in western China, *Bull. Seismol. Soc. Am.* **88,** 1260–1274.

Fan, G.-W., T. Lay, and S. Bottone (2002). Path corrections for source discriminants: A case study at two international seismic monitoring stations, *Pure Appl. Geophys.* **159,** 651–678.

Ford, S. R., W. S. Phillips, W. R. Walter, M. E. Pasyanos, K. Mayeda, and D. S. Dreger (2010). Attenuation tomography of the Yellow Sea/Korean Peninsula from Coda-source normalized and direct Lg amplitudes, *Pure Appl. Geophys.* **167,** 1163–1170.

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

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

Lin, P., B. Chiou, N. Abrahamson, M. Walling, C.-T. Lee, and C. Cheng (2011). Repeatable source, site, and path effects on the standard deviation for empirical ground-motion prediction models, *Bull. Seismol. Soc. Am.* **101,** 2281–2295.

McVerry, G. H., J. X. Zhao, N. A. Abrahamson, and G. H. Somerville (2006). Crustal and subduction zone attenuation relations for New Zealand earthquakes, *Bull. New Zeal. Natl. Soc. Earthq. Eng.* **39,** 1–58.

Morikawa, N., T. Kanno, A. Narita, H. Fujiwara, T. Okumura, Y. Fukushima, and A. Guerpinar (2008). Strong motion uncertainty determined from observed records by dense network in Japan, *J. Seismol.* **12,** 529–546.

Nigam, N. C., and P. C. Jennings (1969). Calculation of response spectra from strong-motion earthquake records, *Bull. Seismol. Soc. Am.* **59,** 909–922.

Ojeda, A., and L. Ottemoller (2002). Q_{Lg} tomography in Colombia, *Phys. Earth Planet. In.* **130,** 253–270.

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

Pasyanos, M. E., and W. R. Walter (2009). Improvements to regional explosion identification using attenuation models of the lithosphere, *Geophys. Res. Lett.* **36,** L14304.

Pasyanos, M. E., E. M. Matzel, W. R. Walter, and A. J. Rodgers (2009). Broad-band Lg attenuation modelling in the Middle East, *Geophys. J. Int.* **117,** 1166–1176

Pasyanos, M. E., W. R. Walter, and E. M. Matzel (2009). A Simultaneous multiphase approach to determine P-wave and S-wave attenuation of the crust and upper mantle, *Bull. Seismol. Soc. Am.* **99,** 3314–3325.

Petersen, M. D., A. D. Frankel, S. C. Harmsen, C. S. Mueller, K. M. Haller, R. L. Wheeler, R. L. Wesson, Y. Zeng, O. S. Boyd, D. M. Perkins, N. Luco, E. H. Field, C. J. Wills, and K. S. Rukstales (2008). Documentation for the 2008 update of the United States National Seismic Hazard Maps, *U.S. Geol. Surv. Open-File Rept. 2008-1128,* 60 pp.

Phillips, W. S. (1999). Empirical path corrections for regional-phase amplitudes, *Bull. Seismol. Soc. Am.* **89,** 384–393.

Phillips, W. S., H. E. Hartse, S. R. Taylor, A. A. Velasco, and G. E. Randall (2001). Application of Regional Phase Amplitude Tomography to Seismic Verification, *Pure Appl. Geophys.* **158,** 1189–1206.

Phillips, W. S., G. E. Randall, and S. R. Taylor (1998). Regional phase path effects in central China, *Geophys. Res. Lett.* **25,** 2729–2732.

Rodgers, A., J. Ni, and T. Hearn (1997). Propagation characteristics of short-period Sn and Lg in the Middle East, *Bull. Seismol. Soc. Am.* **87,** 396–413.

Rodgers, A. J., W. R. Walter, C. A. Schultz, S. C. Myers, and T. Lay (1999). A comparison of methodologies for representing path effects on regional P/S discriminants, *Bull. Seismol. Soc. Am.* **89,** 394–408.

Rodriguez-Marek, A., and F. Cotton (2011). "Draft final report: single station sigma project" Pegasos Refinement Project, report EXT-TB-1058, 79 pp.

Rodriguez-Marek, A., G. A. Montalva, F. Cotton, and F. Bonilla (2011). Analysis of single-station standard deviation using the KiK-net data, *Bull. Seismol. Soc. Am.* **101,** 1242–1258.

Taylor, S. R., A. A. Velasco, H. E. Hartse, W. S. Phillips, W. R. Walter, and A. J. Rodgers (2002). Amplitude corrections for regional seismic discriminants, *Pure Appl. Geophys.* **159,** 623–650.

Zhang, T.-R., and T. Lay (1994). Analysis of short-period regional phase path effects associated with topography in Eurasia, *Bull. Seismol. Soc. Am.* **84,** 119–132.

Zhao, J. X. (2010). Geometric spreading functions and modeling of volcanic zones for strong-motion attenuation models derived from records in Japan, *Bull. Seismol. Soc. Am.* **100,** 712–732.

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

The Charles Edward Via,
Jr. Department of Civil and Environmental Engineering
Virginia Tech
200 Patton Hall
Blacksburg, Virginia 24061
hdawood@vt.edu

Review of 'Improving the understanding and methodologies of ground motion variability' (report actually entitled 'Scientific cooperation between EDF and Virginia Tech in the field of seismic hazard assessment') by A. Rodriguez-Marek and H. Dawood (SIGMA deliverable D2-111)

The authors present an analysis of strong-motion data from the Japanese KiK-Net with the aim of proposing improved models of the aleatory variability to be used with ground motion prediction equations (GMPEs) within partially non-ergodic probabilistic seismic hazard assessments (PSHAs). The topic is an important one for high-level PSHAs, where slight changes in sigma can have a large impact on the hazard results, and hence many studies on this subject have been conducted in the past decade.

This report is clearly the fruit of a considerable effort in order to manipulate a much larger database than has been used for such studies in the past. The philosophy being: the larger the database the more robust the results. In general, I agree with this approach but it should be checked that the use of an automatic procedure to process the data and the inclusion of records with limited engineering significance (small earthquakes and/or large distance) does not bias the results. It could be better to focus on a smaller but carefully processed set of records so that noise in the results coming from errors in the data or metadata is reduced.

The large database compiled for this work will be a valuable resource for future work on this topic (and others) and the authors are encouraged to make it publicly available. The analysis presented is detailed and appears to have been carefully conducted. The consideration of the impact of azimuth and topography on the variability is a considerable advance. In addition, the results are generally well presented, although the presentation could be improved (e.g. there are a number of typographic errors, formatting issues and some figures could be made clearer and their style unified). Despite my positive opinion of this study (and the previous work by Prof. Rodriguez-Marek and his group on this topic), I do have some minor technical and editorial comments, which are listed below in the order that they occur in the text. I only reviewed in detail the main report, which provides a summary of the work undertaken and the principal results, and not the appendices because of their length and as they are mainly papers that have either been published or submitted. I encourage the authors to publish the findings of the main report too once they are finalised and the presentation improved.

Technical remarks and questions:

1. Throughout: As the focus of SIGMA is on seismic hazard assessment for mainland France, an area of low to moderate seismicity, the applicability of the results based on analysis of data from Japan, an area of high seismicity, should be discussed. It is noted on p. 8 that single-station φ shows weak regional dependency but, I believe, this has not yet been clearly demonstrated for areas of low seismicity but only for moderate and high hazard areas (e.g. Rodriguez-Marek et al., 2013).
2. P. 6: The authors appear to suggest that the reduction in aleatory variability when using a single-station sigma is completely balanced by the increase in epistemic uncertainty due to not knowing the site term for the location of interest. Is this always the case for a well-studied site for which many strong-motion records are available?
3. P. 10: Was the approach of NGA West 2 (by Wooddell & Abrahamson) attempted to classify the events into mainshocks and aftershocks?
4. P. 13, Figure 5.2.1: Are the very low values (about 0.3) at long periods in the current study realistic (or due to a lack of data or problems with the data processing)? What is the physical explanation for such low values?
5. P. 14, 'are not intended as models for direct use in hazard application [sic]': Why? How does a 'model for direct use in hazard applications' differ from a model for research purposes?

6. P. 14 and Figure 5.2.4: How well constrained are the estimates of φ_{SS} at short distances? For example, at r_{rup}=5km there are probably few records available to compute this variability and hence it is uncertain since it is based on data from only a few stations and earthquakes.

7. Figure 5.2.5: Douglas and Jousset (Seismological Research Letters, 2011) argue that variations in kappa between stations can lead to magnitude-dependent aleatory variability. Could this help explain the magnitude dependency of this model?

8. P. 18, 'for all stations is nearly equal to φ_{SS}': Why is it not exactly equal? I thought that φ_{SS} is defined as the average of the $\varphi_{SS,s}$.

9. P. 25, caption of Figure 5.3.2, 'yet it has high $\varphi_{SS,s}$ due to the complexity of the tectonic environment in Japan': This explanation could apply to all stations in Japan (and particularly those close to this station, which would have recorded similar events). Why would this station be particularly affected?

10. Section 5.4: For their interest, the authors may wish to study the articles of Douglas et al. (Bulletin of Earthquake Engineering, 2004 and 2007) on a different approach to include path effects within GMPEs.

Minor editorial comments:

1. P. 8: The authors write that '[t]he idea of including site terms in the regression for ground motion prediction equations was first proposed by Joyner and Boore (1981)'. This is not correct as Trifunac (Bulletin of the Seismological Society of America, 1976) included such terms. Perhaps the authors meant to cite Joyner and Boore (Bulletin of the Seismological Society of America, 1993), who were the first to evaluate the inter-site variability, as far as I know.

2. P. 8: The authors write that the aleatory-variability models derived during the PEGASOS Refinement Project and the Thyspunt Nuclear Siting Project are the only ones to have 'postulated actual values for use in hazard analyses'. Why cannot the other studies cited in the same paragraph be used in hazard analyses or those by Douglas et al. (Bulletin of the Seismological Society of America, 2013)?

3. P. 11: For crustal earthquakes without known fault planes, how were the two fault plane solutions handled (details are only given for interface and intraslab events)?

4. P. 11 (and elsewhere): LCPC is known since 2011 as 'Institut français des sciences et technologies des transports, de l'aménagement et des réseaux' (IFSTTAR).

5. P. 12, 'As part of this project … as part of this project …': This sentence is not clear.

6. P. 12, 'indluded': Please spell check before submission of the final version since there are a few spelling errors (including this one).

7. Figures 5.2.1 and 5.2.2: Should the solid lines on these figures be the same (as they are both labelled as 'Current Study')?

8. P. 19, 'calculating $\varphi_{SS,s}$ using recorded motions': Is there another way to calculate it?

9. Figure 5.3.2: I would recommend plotting the scatter plots with respect to azimuthal on radial axes so that azimuths near 0 and 360° are plotted close to each other.

10. P. 32, 'doesn't': Contractions should be avoided in formal text.

11. Figure 5.4.1: The new model could be added to this figure.

12. P. 34 (and elsewhere): The authors state that 'over 100000 records' have been used when according to p. 12 only 13735 records were used.

John Douglas

BRGM, Orléans

21st May 2014

Comments on the report of May 13, 2014 regarding the

Scientific cooperation between EDF and Virginia Tech in the field of seismic hazard assessment

Ref Contract No. 30000-5910098949

Frank Scherbaum
University of Potsdam
Potsdam, June 3, 2014

General remarks

The report by Adrian Rodriguez-Marek and Haitham Dawod presents results of research conducted at Virginia Tech as part of the sigma project and includes:

a) the compilation of a comprehensive database, including metadata for each station (which, among other parameters, included the computation of kappa), of ground motions from the KiKnet network,
b) an analysis of the single-station sigma from these data, including analysis of the station-to-station variability in single-station sigma,
c) the development of a new approach for computing single-station, single-path sigma, and
d) guidelines for the implementation of partially non-ergodic PSHA.

It consists of a collection of reports, journal manuscripts (both drafts and a reprint), and an executive summary. The list of topics above and the sheer volume of the report (185 pages of material) already demonstrate that a huge amount of work was done in the framework of this study. Below I am commenting specifically on individual parts.

Specific comments on *the main report*
by A. Rodriguez-Marek and Haitham Dawood

Reading the main report left me with rather mixed feelings, because the obviously high quality of the work done in the framework of the contract, as judged for example from reading the appendices, is not properly conveyed in the presentation of the results in the main report.

The main report would have benefitted considerably from a final check for typos, nomenclature inconsistencies, inconsistencies in the use of capital and small greek letters, indexing inconsistencies (indeces sometimes appear as subscripts, sometimes not) as well as typos in formulas (e. g. square roots missing). This is especially unfortunate,

since the understanding of the contributions of the different components to the overall ground motion variability becomes rather difficult without a clear and unique notation (e. g. Al Atik et al., 2010).

One the other hand, I appreciate the fact that a main report was produced at all and the deliverable did not only consist only of the collection of the individual papers and manuscripts. In particular, the paragraph on the original objectives as identified at the outset of the project was helpful as a starting point for the evaluation of the results.

I also very much liked the way the background of the single-station sigma concept is discussed because it nicely explains not only the rationale for its use but also the requirements for its application, something which I find missing in other discussions of the subject in the literature.

In conclusion, the main report is „selling the results below value". It should undergo a thorough revision to iron out the inconsistencies mentioned above which severally hamper the readability of the document (at least for me).

Specific comments on Appendix I: *Processing the KIKnet strong ground motion database and compilation of metadata for GMPE development*
by H. M. Dawood, A. Rodriguez-Marek, J. Bayless, C. Goulet, and E. Thompson

This manuscript, prepared for submission to *Earthquake Spectra*, describes the compilation of a comprehensive database, including metadata for each station (which, among other parameters, included the computation of kappa), of ground motions from the KiKnet network.

I have really nothing to criticise regarding this work. On the contrary, the results of this effort are an extremely valuable resource for ground-motion model development for engineering purposes. All the procedures are up-to-date and the protocol used to process the ground-motion data is well described.

Specific comments on Appendix II: An empirical ground motion prediction equation for active crustal earthquakes using the Japanese Kik-net database (ergodic and site specific formulations)
by H. M. Dawood, A. Rodriguez-Marek,

In this manuscript, the authors report on the development of a ground-motion prediction equation which contains an ergodic and a non-ergodic site specific formulation. This is a novel approach which directly allows the application of the GMPE for site conditions where the site term is well known and the requirements for single station sigma are met.

I find this an very good paper which, besides providing a novel type of GMPE, contains a wealth of information regarding the properties of the different ground motion variability components. In particular the results on the distance and magnitude dependence of intra event variability but also the reported effects of azimuth on the single station sigma demonstrate that there is still a lot to learn.

What I find unfortunate, however, is that the authors use a notation which differs from the Al Atik et al. (2010) notation, for example in using small greek letters for Phi etc. This probably also contributes to the issues which I discussed in the context with the main report. I strongly urge the authors to follow the suggested notation of Al Atik et al. (2010), if they don´t want to add confusion to this already complicated subject.

Specific comments on on Appendix III: *Computation of Kappa for the KIKnet stations*
by A. M. Cabas and A. Rodriguez-Marek

The determination of the high frequency slope of surface-to-borehole Fourier spectral ratios is particularly interesting since it has the potential to shed more light on origin of high frequency filter effects on the seismic spectrum. In contrast to authors, however, I don´t see the resulting κ values as κ_{0_TF} in the sense of Drouet et al. (2010). If I understand it correctly, κ_{0_TF} is obtained by comparing an observed spectrum with a model spectrum and interpreting the spectral ratio as indicative of κ. In this case, errors in the model will map onto the resulting high frequency slope and subsequently into the interpretation. As a consequence, it is not clear if the resulting κ values can be attributed to a particular depth range, which may also be partially responsible for the discrepancy between different kappa estimates seen in Figure 2. In the case of the surface-to-borehole Fourier spectral ratios, however, the spatial origin of the „filter effect" should be well constrained. Therefore, I suggest to use yet another term such as κ_{S2DSR}.

I find the example figures in Figure 1 of the manuscript to be quite illustrative but also disturbing, since they suggest that one might bias (actually censor) the perception of the phenomenon responsible for what is referred to as κ if one does not report a kappa value if the slope is positive (category 1) or flat (category 3). If a physical phenomenon (even if we don´t understand it) exists that may cause an increase in the high frequency part of the seismic spectrum between borehole and surface (e. g. a focusing effect in a 3D heterogeneous situation) it may also be present to some degree in situations where the slope of the spectral ratio is still negative. As I see it, κ is simply a descriptive parameter for the high frequency slope of a Fourier spectrum, not a physical parameter which for physical reasons has to be positive. I would therefore suggest to also report negative slopes.

I would also suggest to use the proper kappa notation in Figure 1. If the authors believe what they write, namely that proper taxonomy is important, they might as well use it (:-).

With respect to the interpretation of Figure 3, one could, in contrast to the interpretation of the authors „*that kappa values may be influenced by deeper geologic structures at rock sites*" argue that V_{s30} is simply a poor parameter to capture the filter effects seen in κ_{S2DSR}.

Specific comments on Appendix IV: Summary of research conducted at Virginia Tech on the parametrisation of topographic effects
by M. Rai and A. Rodriguez-Marek

In this study, the authors investigate the potential influence of the site topography on GMPE residuals. Since these effects are not usually included in GMPE development, little is known regarding their characteristics.

Although the methods used seem sensible and results are interesting in their own right, in particular when comparing the results from the global database with the results for the Kiknet dataset, it does not really become clear what the punchline of the study is. In this context it would have been quite helpful if the authors would have made a suggestion what to actually do with them.

Specific comments on Appendix V: *A method for including path effects in Ground-Motion Prediction Equations: An example using the Mw 9.0 Tohoku earthquake aftershocks*
by H. M. Dawood and A. Rodriguez-Marek

The authors present a method to accommodate path effects in the developments of GMPEs which in my opinion is problematic because it applies a conceptual model for how attenuation acts on seismic waves expressed in the Fourier domain without modification to the response spectral domain. Because the relationship between Fourier spectra and response spectra is non-linear, there is no general one-to-one relationship between these two domains which holds for the whole range of oscillator frequencies. One can theoretically show that there is an approximate proportionality between Fourier spectral amplitudes and response spectral amplitudes for low oscillator frequencies (where the attenuation effects are usually small) but not for high oscillator frequencies (roughly above the peak of the acceleration response spectrum) and certainly not for PGA which is affected by the energy within the whole signal passband.

What I am missing and would have liked to see if I would have been a reviewer of the paper is that the authors take into account the fundamental differences between Fourier and response spectra. Plots like Fig. 6 for the whole range of oscillator frequencies would have enabled to judge if the spatial distribution of the „attenuation" is really independent of the source signals (which it should if it is a spatial property) and if the frequency dependence of the values is consistent with knowledge about attenuation mechanisms.

Concluding remarks

Overall, the deliverable reports on a large amount of very good and innovative work. However, it seems to have been put together in a hurry which at places makes it a rather laborious read. The main reason for this is the large amount of notational inconsistencies regarding the taxonomy of the different components of ground motion variability as well as of kappa as discussed in detail above which, however, could be easily corrected through a editorial revision of the manuscript.

References

Alatik, L., Abrahamson, N., Cotton, F., Scherbaum, F., Bommer, J. J. and Kuehn, N. M. (2010). The Variability of Ground-Motion Prediction Models and its Components. Seismological Research Letters, 81 (5), 794-801.

Drouet S., F. Cotton and P. Gueguen (2010). VS30, κ, regional attenuation and Mw from accelerograms: application to magnitude 3–5 French earthquakes. Geophysical Journal International, 182, 880–89.