Partitioning Uncertainty for Non-Ergodic Probabilistic Seismic Hazard Analyses

Haitham Mohamed Mahmoud Mousad Dawood

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Civil Engineering

Adrian Rodriguez-Marek, Committee Chair
Martin C. Chapman
Russell A. Green
James Martin, II

September 29, 2014
Blacksburg, VA

Keywords: Non-Ergodic, Probabilistic Seismic Hazard Analysis, Ground Motion Prediction Equation, Ground Motion Processing, KiK-net Network.

Partitioning Uncertainty for Non-Ergodic Probabilistic Seismic Hazard Analyses

Haitham Mohamed Mahmoud Mousad Dawood

ABSTRACT

Properly accounting for the uncertainties in predicting ground motion parameters is critical for Probabilistic Seismic Hazard Analyses (PSHA). This is particularly important for critical facilities that are designed for long return period motions. Non-ergodic PSHA is a framework that allows for this proper accounting of uncertainties. This, in turn, allows for more informed decisions by designers, owners and regulating agencies.

The ergodic assumption implies that the standard deviation applicable to a specific source-path-site combination is equal to the standard deviation estimated using a database with multiple source-path-site combinations. The removal of the ergodic assumption 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 has this become possible. This study contributes to the state of the art in earthquake engineering and engineering seismology in general and in non-ergodic seismic hazard analysis in particular. The study is divided in four parts. First, an automated protocol was developed and implemented to process a large database of strong ground motions for GMPE development. A comparison was conducted between the common records in the database processed within this study and other studies. The comparison showed the viability of using the automated algorithm to process strong ground motions. On the other hand, the automated algorithm resulted in narrower usable frequency bandwidths because of the strict criteria adopted for processing the data. Second, an approach to include path-specific attenuation rates in GMPEs was proposed. This approach was applied to a subset of the KiK-net database. The attenuation rates across regions that contains volcanoes was found to be higher than

other regions which is in line with the observations of other researchers. Moreover, accounting for the path-specific attenuation rates reduced the aleatoric variability associated with predicting pseudo-spectral accelerations. Third, two GMPEs were developed for active crustal earthquakes in Japan. The two GMPEs followed the ergodic and site-specific formulations, respectively. Finally, a comprehensive residual analysis was conducted to find potential biases in the residuals and propose models to predict some components of variability as a function of some input parameters.

DEDICATION

For Eman, Noor, and Gannah

ACKNOWLEDGEMENTS

I'm grateful to Dr. Adrian Rodriguez-Marek, my advisor and committee chair, for his continuous guidance, encouragement and support throughout the past four years. I would like to thank my committee members: Dr. Russell Green, Dr. Martin Chapman, and Dr James Martin. I would also like to that Dr Guney Olgun. I want to acknowledge the contribution of many researchers who reviewed reports and/or papers prepared within this study. Their valuable suggestions and insights improved the quality of this work. In particular, I would like to mention Dr. Fabrice Cotton. I also want to acknowledge the faculty in Virginia Tech in general, and the faculty in the Geotechnical Engineering Group of the Civil and Environmental Engineering department in particular for their passion and commitment in teaching their students how to be better engineers and researchers. Also, I would like to thank the staff in the Civil Engineering department for their continuous help and support.

I would not have been where I'm today without the immense support of my family. I would like to thank my mother (Gihan), my father (Mohamed) and my sweet sisters (Hend and Heba) for surrounding me with their love and kindness throughout this journey even if we were thousands of miles away.

I would like to express my deepest appreciation to my wife (Eman) for her love, patient and faithful encouragement and helping me for setting up a good study atmosphere in order to finish my degree. The past four years were full of challenges and hard times, but she was always there to support me and encourage me to carry on.

I would like to thank my son Noor and daughter Gannah. They are the joy of my life. Whenever I was tired or stressed out, their smile was enough to get me back on track. I wish I spent more time with them during the past years, but I'm sure they understand my reasons and forgive me.

I would also like to thank all my friends from Virginia Tech. Your friendship and support helped me a lot during these years. In particular, I would like to mention Sherif Abdel Aziz. I would also like to thank my dear friends Ahmed El-Hadidy and Ahmed Awed.

ATTRIBUTION

All the work I performed for my dissertation was done under the advising of Professor Adrian Rodriguez-Marek. Accordingly, he is my co-author on all articles. Other than my advisor, I have three additional co-author on the article presented in Chapter 3. Mr Jeff Bayless and Dr Christine Goulet developed the first version of the processing protocol. This "beta" version was then expanded and fully automated during the course of this study. Dr Eric Thompson worked on estimating the different distance measurements (presented in Appendix A) used in the analysis of Chapter 5.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 5: Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations

Chapter 6: Residual Analysis of Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations

Appendix A: Electronic Supplement For "A Flatfile For The KiK-Net Database Processed Using An Automated Protocol"

Appendix B: A Flatfile For The KiK-Net Database Processed Using An Automated Protocol: Supplementary Information

Appendix C: Comparison between the Records Processed Using the Automated Algorithm by Dawood et al. (2014), and the Motions Used in the NGA-W2 and BC-Hydro Projects

Appendix D: Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations: Supplementary Information

LIST OF FIGURES

Chapter 5: Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations

Chapter 6: Residual Analysis of Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations

Appendix A: Electronic Supplement For "A Flatfile For The KiK-Net Database Processed Using An Automated Protocol"

Appendix B: A Flatfile For The KiK-Net Database Processed Using An Automated Protocol: Supplementary Information

Appendix C: Comparison between the Records Processed Using the Automated Algorithm by Dawood et al. (2014), and the Motions Used in the NGA-W2 and BC-Hydro Projects

Appendix D: Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations: Supplementary Information

Chapter 1: Introduction and Objectives

The ground shaking that results from earthquakes is the result from the interaction of very complex phenomena. These phenomena have been studied by researchers from a variety of fields over the years resulting in significant advances in the understanding of earthquake induced ground shaking. However, the complexity of these phenomena combined with the massive destruction potential from earthquakes serve as a motivation for researchers around the world to continue working on understanding earthquakes and their consequences.

Predicting the size and nature of the strong ground motions generated by an earthquake is a major research area explored by engineers, seismologists and statisticians, among others. Predicting seismic waves at a specific location necessitates the understanding of the nature and characteristics of the seismic sources (source), the location at which the predictions are made (site), and the travel path through which the seismic waves travel (path). The relationships that make use of the source, path, and site information to predict parameters of strong ground motions are known as ground motion prediction equations (GMPEs). The input parameters of GMPEs (e.g., moment magnitude and closest distance from the site to ruptured fault) are not deterministic. Moreover, strong ground motions are inherently variable. One approach to get over this problem is to build probability density functions (PDFs) for the key input parameters and for the estimates of ground motion parameters. These PDFs along with a GMPE functional form are used to perform what is known as probabilistic seismic hazard analysis (PSHA) to obtain probabilistic predictions of the ground motion parameters. PSHA is a mathematical formulation that incorporates the uncertainties associated with the different input parameters and the uncertainties associated with the predictions from GMPEs to get rational estimates of ground motion parameters and their probability of exceedance.

Most the GMPEs that are currently used in PSHA make use of the ergodic assumption. Anderson and Brune (1999) define the ergodic assumption as assuming that spatial uncertainty of ground motions and the uncertainty over time at a single point are interchangeable. The ergodic assumption implies that the standard deviation applicable to a specific source-path-site combination is equal to the standard deviation estimated using a database with multiple source-path-site combinations. In non-ergodic GMPEs the total standard deviation is split into four

components (however, this split requires a very complete database). The first three components are associated with source, path, and site while the fourth is the random error that represents the randomness of the earthquake phenomena.

INTRODUCTION

The research presented in this dissertation aims to advance the state of the art in earthquake engineering and engineering seismology. Over the past few years, several researchers around the world explored the possibility of removing the ergodic assumption when performing seismic hazard analysis. Much progress had been accomplished over the years. However, some questions related to removing the ergodic assumption have not yet been answered. This research study intends to explore and find answers to some of these questions.

In order to conduct non-ergodic PSHA analyses, GMPEs must be framed within the non-ergodic framework. In order to develop non-ergodic GMPEs, large and well-balanced databases of strong ground motion records are needed. The strong GM records recorded by the Kiban Kyoshin (KiK-net) network were chosen for this study. The KiK-net database was chosen for several reasons:

- It consists of a large number of strong GM records from many earthquakes that originated from different tectonic environments;

- Two collocated instruments at ground surface and at depth are installed at each station which can be used to empirically estimate the site response; and

- The shear wave velocity profiles were measured at most stations within the network.

The KiK-net website provides raw GMs that need to be processed before being used. Hence, processing the KiK-net database and develop the associated metadata is an integral component of this study. A sample of several studies that were conducted to process and compile metadata for different strong GM databases around the world is shown in Table 1. Chapter 3 of

this dissertation presents a methodology for automatic processing and generation of metadata for the KiK-net database.

The processed GM data from the KiK-net website is used to develop a new GMPE for Japanese active crustal earthquakes. This GMPE is developed in Chapter 5 using both an ergodic and non-ergodic formulation. These formulations are compared and the residuals from these GMPEs are analyzed to find potential biases in the residuals and propose models to predict the dependence of some of the components of variability on parameters such as magnitude and distance (Chapter 6).

Ground motion prediction equations include models for the attenuation of seismic waves with distance. This attenuation has geometric and anelastic components. As seismic waves travel through the earth's crust, their energy is gradually absorbed and converted into heat. This phenomena is known as anelastic attenuation of seismic waves. The rate at which the seismic energy is absorbed is known as anelastic attenuation rate and it depends on the nature of the medium through which the waves travel. In general, the attenuation rates are lower through hard and continuous mediums. The earth's crust is heterogeneous in nature; hence, the anelastic attenuation rates are also spatially heterogeneous. If not accounted for properly, this heterogeneity artificially increases the scatter associated with predicting GM parameters. 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), Fan and Lay (1998), Phillips et al. (1998), Phillips (1999), Rodgers et al. (1999), Phillips et al. (2001), Fan et al. (2002), Ojeda and Ottemoller (2002), Taylor et al. (2002), Edwards et al. (2008), Ford et al. (2010), Pasyanos and Walter (2009), Pasyanos et al. (2009-a) and Pasyanos et al. (2009-b)). Yet, most of the GMPEs used in seismic hazard analyses assume homogenous attenuation rates. Chapter 4 in this dissertation presents a methodology to account for regional dependences of anelastic attenuations rates within the format used by GMPEs used in engineering applications to reduce the scatter in predicting GM parameters.

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

SIGNIFICANCE

Properly accounting for the uncertainties in predicting ground motion parameters is critical when performing PSHA. This is particularly important for critical facilities that are designed for long return period motions. Non-ergodic PSHA is a framework that allows for this proper accounting of uncertainties. This, in turn, allows for more informed decisions by designers, owners and regulating agencies.

The uncertainties in a PSHA analyses are partitioned into aleatoric variability and epistemic uncertainty. The aleatoric variability is generally taken as the standard deviation associated with predicting a certain ground motion parameter from a GMPE. The epistemic uncertainty is subjectively assigned to reflect the lack of knowledge regarding any input parameter of the PSHA (including, source, path, and site information) and the models used in the predictions. In the current state of practice of PSHA, the uncertainties associated with the site may be double counted (epistemic uncertainty is assumed for some of the site parameters while the aleatoric variability contains a part of this variability as well). In non-ergodic PSHA the quantification of uncertainties is improved by using an aleatoric variability that does not contain the contribution of the site and treating it as epistemic uncertainties only.

The field of non-ergodic PSHA is relatively young and is in active development. To perform a fully non-ergodic PSHA: 1) the source, path and site terms should be estimated and treated as epistemic uncertainty; and 2) the error component of variability should be estimated while developing the fully non-ergodic GMPE. Constraining path terms in a form that can be used for forward predictions of ground motion parameters is not a simple task to achieve. Moreover, most researchers do not constrain the different residual components while developing the GMPEs. They rather split the total residuals from an ergodic GMPE to obtain the different residual components. In doing this, they inherently assume that the median predictions from ergodic and non-ergodic GMPEs are the same.

This dissertation mainly contributes to the state of the art and practice of PSHA by: 1) developing an alternative approach to estimate the path-specific residual components for forward

predictions, and 2) investigate the validity of assuming that the ergodic and non-ergodic GMPEs result in the same median predictions.

OBJECTIVES

The objectives of this study are:

- Process the database of strong GM records recorded by the KiK-net network, and compile ground motion parameters and site-, path- and event-metadata in a single spreadsheet.

- Develop a methodology to account for the path-specific attenuation that can be incorporated in a GMPE framework.

- Develop ergodic and non-ergodic GMPEs for active crustal earthquakes.

- Investigate the dependency of the different residual components and their standard deviations on several parameters including M_w, PGA, azimuth, and distance; and identify conditions under which residual components are large.

ORGANIZATION AND CONTENT OF THIS DISSERTATION

This dissertation contains seven chapters and four appendices. Some of the chapters (Chapters 3, 4 and 5) are manuscripts that at the time of submission of this thesis have been either published or submitted for publication. Chapter 6 is a stand-alone chapter. Chapter 2 presents a summary of previous studies conducted in areas relevant to this study. Chapter 3 introduces the strong GM database processed within this study and provides a detailed description of an automated protocol developed to process the raw data downloaded from the KiK-net website, along with the approaches carried out to prepare a comprehensive set of metadata for each GM record. A subset of the ground motion data presented in Chapter 3 is used to develop a methodology to account for path-specific attenuation rates in GMPEs (Chapter 4). Another subset (e.g., ground motion data from active crustal earthquakes) is used to develop a GMPE using both an ergodic and a non-ergodic approach (Chapter 5). Chapter 6 presents an analysis of the residuals from the GMPE presented in Chapter 5. Finally, Chapter 7 summarizes

the main conclusions from this study. Since most of the Chapters in this theses were written as journal manuscripts, they were develop within the length constraints of the journals where they were published. Additional details that are complementary to these publications are presented as Appendices to this thesis. Table 2 summarizes the content of the appendices.

Table 2. Summary of the content of the appendices of this dissertation

Appendix A	Is an electronic supplement that has been submitted with the manuscript presented in Chapter 3. This supplement provides additional details about the approaches we adopted to get the different distance measurements for the different GM records.
Appendix B	Presents additional details about the analysis presented in Chapter 3. In particular, it provides additional details and explanations about the automated processing protocol as well as the attempts conducted to de-cluster the F-net seismic catalog.
Appendix C	Presents a detailed comparison between the processed GM records in common between the current study and two different studies that previously processed portions of the KiK-net database (i.e., Zhao et al. (2006) (this subset was used in the BC-Hydro project) and the NGA-W2 project. The pseudo-spectral accelerations obtained from each algorithm were compared. Also, the metadata (e.g., the different distance measurements) we obtained in this study was compared to the metadata adopted by the NGA-W2 project.
Appendix D	Presents additional details about the analysis presented in Chapter 5 of this dissertation. It presents step by step description of the steps adopted to obtain the subset of data used in this study, the steps adopted to regress and smooth the different regression coefficients and a detailed residual analysis (among other details).

REFERENCES

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.

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

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.

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.

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

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. Seism. Soe. Am. **98**, 1915-1935.

Fan, G.-W and T. Lay (1998). Regionalized versus single-station waveguide effects on seismic discriminants in western China, Bull. Seism. Soe. 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.

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

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.

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

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., W.R. Walter, and E.M. Matzel (2009b). A Simultaneous Multiphase Approach to Determine P-Wave and S-Wave Attenuation of the Crust and Upper Mantle, Bull. Seism. Soc. Am., **99**, 3314–3325.

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., 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. Seism. Soc. Am., **89**, 394–408.

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

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 (1994a). Analysis of short-period regional phase path effects associated with topography in Eurasia, Bull. Seism. Soc. Am. **84**, 119-132.

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.

Chapter 2: Literature Review

INTRODUCTION

This chapter presents a summary of previous studies that were conducted in areas that are relevant to the research presented in this dissertation. These areas are:

- Processing of strong ground motion (GM) records
- Declustering of seismic catalogs.
- Automatic classification of earthquakes.
- Ground motion prediction equations (GMPEs) that remove the ergodic assumption.

STRONG GM PROCESSING

In general, strong GM records are contaminated with noise. The purpose of GM processing is to adjust the raw GM records and identify the bandwidth that can be reliably used for each record. The most common adjustments of GM records include removing the effect of long period noise and correcting for the instrument response. Boore and Bommer (2005) present a comprehensive summary of the different techniques adopted to process strong ground motions.

The long period noise often results in unphysical velocity and displacement time histories. This is known as baseline error. This error can be corrected by either: (1) using high-pass filters to filter out the long period noise; or (2) fitting polynomials of different orders to the velocity and/or displacement time histories, then subtracting the fitted equation from the time history. According to Boore and Bommer (2005), the first approach is the most widely used and less subjective approach to process strong GM records. The second approach is more appropriate in cases where residual displacements are expected. The two previous approaches can be combined to process GM records (e.g., Chiu 1997). This section focuses on the first approach because it is used in the current study.

Filtering ground motions is the process of reducing the amplitude of the motion at a specific frequency range (s). This reduction is generally gradual and its shape depends on the type of the applied filter. A high-pass filter is a filter that does not affect the amplitudes at

frequencies higher than the corner frequency of the filter. Hence, applying a high-pass filter to the motions will cut long period noise. Yet, choosing the corner frequency is critical because we need to choose a corner frequency high enough so we ensure that the noise was removed, but low as much as possible to minimize its impact of the filter on the frequency range of interest in the motions. The three different approaches that can be used to choose the high-pass corner frequency of the filter were listed by Boore and Bommer (2005) as follows:

- The corner frequency is chosen to be the frequency above which the signal-to-noise ratio (SNR) is above 3. In this case SNR is defined as the ratio between the FAS of the motion (including signal and noise) and the FAS of the noise window.

- According to the point-source model (Brune 1970), the FAS of acceleration should decay according to f^2 at low frequencies. The point at which the FAS deviates from this decay slope is picked as the corner frequency.

- Visually inspect the velocity and displacement time histories of the filtered motions. The selected corner frequency is considered appropriate if the velocity and displacement time histories are judged to be physical.

Instrument response tends to distort the amplitude and phase of the motions at high frequencies. This correction is not generally recommended for digital instruments (Boore and Bommer 2005), and is recommended for motions recorded by analog instruments forat stiff sites from a relatively close distance.

There is no standard approach to process strong GM records (Boore and Bommer 2005). Different processing procedures had been proposed in several studies to process strong GM databases. The Next Generation Attenuation (NGA) project (Power et al. 2008) makes use of the strong motion processing protocol used by the Pacific Earthquake Engineering Research (PEER) center (Darragh et al. 2004). This is considered to be one of the most comprehensive processing protocols available up to date. While it would be desirable to adopt this protocol in the current study, several factors make its applicability difficult. These include:

- The flow chart of the protocol presented in Darragh et al. (2004) is not detailed enough (e.g., the protocol includes 'plotting the acceleration, velocity and displacement time histories to review'. It is not clear how this review process is conducted. If it conducted by visual inspection of the different plots, then it cannot be used for very large databases. On the other hand, if it is conducted automatically, there is not sufficient information to replicate it).

- There is no access to the software used in processing the PEER database. Boore et al. (2012) reported that they were not able to get the software from the contractor who processed the database.

For the reasons indicated above, an automated ground motion protocol is developed to obtain processed GM time histories.

DECLUSTERING OF SEISMIC CATALOGS

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 identified and, in some cases, removed from the catalog. Declustering seismic catalogs is important, especially when the catalog is used for characterizing seismic sources for probabilistic seismic hazard analyses (PSHA). This is because temporal occurrence of earthquakes is assumed to follow a stationary Poisson process when performing most of PSHA. Hence, it is important to insure that the declustered catalog is Poissonian. Moreover, the rate of seismicity for each source calculated from the declustered catalog is important for PSHA. Rate of seismicity can be different according to the declustering algorithm used. For use in GMPEs, identifying independent and dependent events is important because GM parameters may have different scaling with source parameters for these two types of events. For example, the strong ground motions in the NGA database that were recorded from aftershocks were found to be lower than motions from main shocks with similar magnitudes at short periods. 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).

Nevertheless, in most GMPEs the difference between dependent and independent earthquakes is simply neglected.

It is typical practice to assume that a declustered catalog follows a Poissonian model (Zhuang et al., 2012), although this assumption is often debated (e.g., Matthews et al. 2002). Because different subsets of events in a catalog can be Poissonian, the declustering process has no unique solution and the different declustering algorithms can result in different results (Stiphout et al., 2012). Many declustering algorithms had been developed over the years (e.g., 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.

Researchers have used different declustering algorithms in their studies. Since, the objective of the current study is to develop non-ergodic GMPEs, a detailed comparison between the different declustering algorithms is beyond the scope of this study. The algorithms by Gardner and Knopoff (1974) and Reasenberg (1985) were selected for the declustering of the seismic catalog used in this study because both 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).

AUTOMATED CLASSIFICATION OF EARTHQUAKES

Earthquakes can be classified into many categories based on various parameters such as location, depth, and style of faulting (among others). A classification dividing earthquakes into *active crustal* earthquakes (shallow and deep); *subduction zone* (SZ) earthquakes (interface and intraslab); and *stable continental region* earthquakes is widely used in seismic hazard analysis. Given the availability of enough records, GMPEs are generally developed using GM records from a specific category. The different travel paths and source characteristics result in differences in the attenuation characteristics of the GM records from each of these categories (Kanno et al. 2006).

Allen et al. (2008) and Garcia et al. (2012) have developed two algorithms to classify the seismic events in near real time. Table 1 presents a summary of the input parameters used in both algorithms. It also shows the different earthquake classifications that can be obtained from each algorithm. It is clear from the table the elevated level of complexity associated with the later algorithm compared to the former one. Both algorithms are used to automatically classify the earthquakes used in this study.

Table 1. Comparison between two algorithms to classify earthquakes by 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). Info only for events in subduction regions.
Earthquake classes	• SZ intraslab • SZ interface • Shallow active crustal • Stable continental	• SZ* outer • SZ intraslab • SZ interface • ACR* deep • ACR shallow • SCR* • SOR* • OBR* ridge • OBR collision/transform • Hotspot • REVISION*

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

GMPES WITHOUT THE ERGODIC ASSUMPTION

GMPEs are widely used in seismic hazard analyses to predict the median of specific ground motion parameters (e.g., PGA) along with the standard deviations associated with the prediction. The number of available GMPEs is increasing rapidly with the increase of available strong ground motion databases. Douglas (2003) provided a comparison between the different GMPEs available at that time in terms of the data selection techniques, GM processing techniques, the characterization of source, path and site and the different regression techniques. Douglas (2011) provided a summary of the 289 GMPEs developed for PGA and 188 GMPEs developed for spectral ordinates. Gunberg and Green (2011) coded MATLAB scripts and Excel spread sheets for about 45 GMPEs. Nine of these were developed for ground motion parameters other than the pseudo-spectral accelerations (e.g., Arias intensity and duration).

This section starts by introducing the non-ergodic concept. The different components in which the total residual can be split into will be then introduced. A summary of previous studies that investigated non-ergodic GMPEs will then be introduced. It is worth noting that a full review of the existing GMPEs is beyond the scope of this study since the focus is on studies working on the non-ergodic concept.

The non-ergodic concept

Joyner and Boore (1993) postulated that the total residual should be split into three components (site-specific, earthquake-specific, and random error). They did not estimate the site-specific residual components due to lack of records. Hence, traditionally, the standard deviation of a GMPE was computed assuming that the variability associated with a specific site-path-source combination is equal to the standard deviation computed from an entire database with various site-path-source combinations. This assumption has been referred to as the ergodic assumption (Anderson and Brune 1999). More recently, it was postulated that if a ground motion dataset is sufficiently large such that the repeatable contributions of site and/or path can be computed, then the standard deviations can be computed without recourse to the ergodic assumption; these standard deviations are referred to as non-ergodic standard deviations (Al-Atik et al. 2010). The amount of strong ground motion data available now for researchers is considerably larger than what was available in the past. Hence, some databases (e.g., KiK-net

database) has enough motions to allow for the removal of the ergodic assumption and estimate the different residual components. Several recent seismic hazard analyses project have used a non-ergodic approach, highlighting the importance of evaluating how non-erdogic GMPEs are computed.

Anderson and Brune (1999) proposed that PSHA can be conducted without the ergodic assumption. A key in adopting non-ergodic PSHA is to distinguish clearly between “aleatoric variability” and “epistemic uncertainty” as they are treated differently within PSHA. Toro et al. (1997) define epistemic uncertainty as *"uncertainty that is due to incomplete knowledge and data about the physics of the earthquake process. In principle, epistemic uncertainty can be reduced by the collection of additional information"* and aleatoric variability as *"uncertainty that is inherent to the unpredictable nature of future events. Aleatory uncertainty cannot be reduced by collection of additional information"*. In PSHA, the aleatoric variability is accounted for by including it in the hazard integral, while the epistemic uncertainty is accounted for by the use of logic trees. Even if, the theoretical distinction is clear, in practice, the distinction between aleatoric and epistemic uncertainty is not trivial and depends on the nature of GM database and the parameterization of the GMPE.

While developing non-ergodic GMPEs, the total standard deviation can be split into up to four components (depending on the nature of the strong ground motion database).Three of these components are associated with source, path, and station and are treated as epistemic uncertainty in non-ergodic PSHA. The fourth is the random error and is treated as an aleatoric variability. Al-Atik et al. (2010) also differentiated between fully and partially non-ergodic PSHA, depending whether all (source, site, and path) or only some (e.g., site) components are constrained. This implies that a considerable portion of the total standard deviation that is conventionally assumed to be aleatoric in conventional PSHA is treated as epistemic in non-ergodic PSHA. Hence, upon the availability of information about a specific source-path-site combination, non-ergodic PSHA can be used to obtain more accurate estimates of hazard.

Components of the total residual

The number and type of the different components in which the total residual can be split into is decided upon the nature of the available database. In general, GMPEs take the following form:

$$ln\,(y_{es}^{G/B}) = ln(\mu_{es}^{G/B}) + \Delta_{es}^{G/B} \quad (1)$$

where the superscripts *G* and *B* denote motions recorded at ground surface, and bedrock (within motion), respectively. The subscripts *e* and *s* denote event and station, respectively. $ln\,(y_{es}^{G/B})$ are the natural logarithm of an observed ground motion parameter at ground surface and borehole; $ln(\mu_{es}^{G/B})$ are the median predictions; and $\Delta_{es}^{G/B}$ are the total residuals.

In ergodic GMPEs, $\Delta_{es}^{G/B}$ are generally split into an inter-event (δB_e) and intra-event ($\delta W_{es}^{G/B}$) residuals components (Equation 2). In this case, δB_e represents the observed shift from the median for records from event *e*. Hence, the GMPE takes the following form:

$$ln\,(y_{es}^{G/B}) = ln(\mu_{es}^{G/B}) + \delta B_e + \delta W_{es}^{G/B} \quad (2)$$

Figure 1 shows this concept of inter- and intra-event residual components. Since δB_e is an event-specific residual component, it should be the same whether the recorded motion is on ground surface or in bedrock.

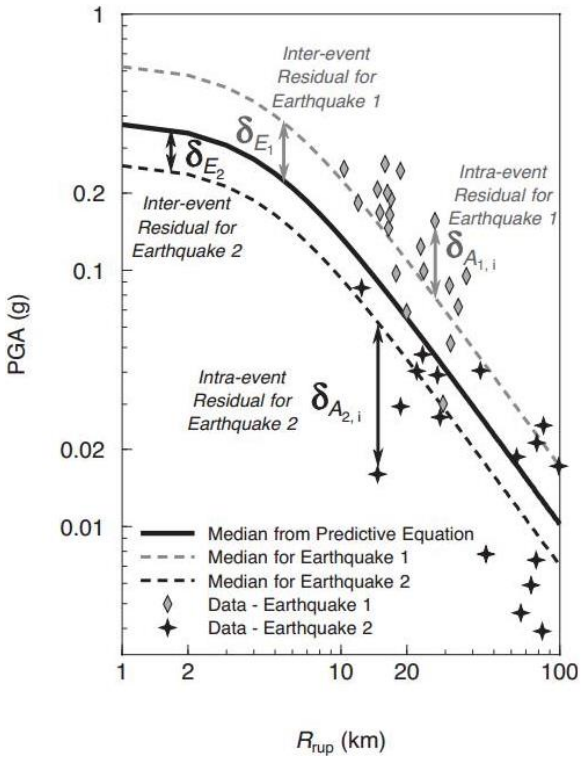


Figure 1. Inter-event and intra-event components of ground-motion variability. The figure shows two earthquakes for which the median of its records is shifted from the median of the GMPE (figure adopted from Strasser et al. 2009)

Upon the availability of more robust databases, δB_e and $\delta W_{es}^{G/B}$ can be split into additional components. As discussed in the previous section, the total residual can be split to include location- (source), path- and site-specific components. If the available database includes records from multiple earthquakes that originated from the same source, the location-specific residual component could be constrained. Since the location component is directly related to the earthquakes, δB_e is split into a location-to-location and remaining residual components:

$$\delta B_e = \delta L2L_l + \delta B_{el}^0 \quad (3)$$

where $\delta L2L_l$ is a location-to-location residual component which represents the observed shift from the median for records from events from this specific location *"l"*, and δB_{el}^{0} represent the remaining residual component for event *e* after removing the location-specific component.

If the available database contains multiple strong GM records recorded at the same station, $\delta W_{es}^{G/B}$ can be split to include a site-specific component:

$$\delta W_{es}^{B} = \delta S2S_{S}^{B} + \delta W_{es,0}^{B} \quad (4)$$

$$\delta W_{es}^{G} = \delta S2S_{S}^{G} + \delta W_{es,0}^{G} \quad (5)$$

where $\delta S2S_{S}^{B}$ is a bedrock-to-bedrock residual component which represents the observed shift from the median for records from this specific bedrock that underlies site *s*; $\delta S2S_{S}^{G}$ is a surface-to-surface residual component which represents the observed shift from the median for records from site *s*; and $\delta W_{es,0}^{\mathrm{B}}$ and $\delta W_{es,0}^{\mathrm{G}}$ represent the remaining residual components after accounting for the site-specific residual components. A GMPE that includes the partition of residuals as described in Equations 4 and 5 is known as a site-specific GMPE.

If the database with multiple records at the same station also has records that crossed the same path multiple times, $\delta W_{es}^{G/B}$ can be split to include site- and path-specific components:

$$\delta W_{es}^{B} = \delta S2S_{S}^{B} + \delta P2P_{sl} + \delta W_{esl,0}^{B} \quad (6)$$

$$\delta W_{es}^{G} = \delta S2S_{S}^{G} + \delta P2P_{sl} + \delta W_{esl,0}^{G} \quad (7)$$

where $\delta P2P_{sl}$ is a path-to-path residual component which represents the observed shift from the median for motions that travelled through this specific source-to-site path; and $\delta \mathrm{W}_{\mathrm{esl,0}}^{\mathrm{B}}$ and $\delta \mathrm{W}_{\mathrm{esl,0}}^{\mathrm{G}}$ represent the remaining residual components after accounting for the path- and site-specific residual components. In this case, the path-to-path component should be independent of whether the record is taken on bedrock or ground surface. Hence, the path-specific component is common between bedrock and ground surface records.

If the database is recorded by a network with co-located instruments on the ground surface and below the surface (such as the KiK-net network), the site-specific empirical amplifications can be written as follows:

$$\delta S2S_S^{AMP} = \delta S2S_S^G - \delta S2S_S^B \quad (8)$$

$$\delta AMP_{esb,0} = \delta W_{es,0}^G - \delta W_{es,0}^B \quad (9)$$

$$\delta AMP_{esbl,0} = \delta W_{esl,0}^G - \delta W_{esl,0}^B \quad (10)$$

where $\delta S2S_S^{AMP}$ is the site-specific empirical amplification residual component and δAMP_{esb}^0 *or* δAMP_{esbl}^0 are the remaining amplification components Hence, Equations 5 and 7 can be re-written as:

$$\delta W_{es}^G = \delta W_{es}^B + \delta S2S_S^{AMP} + \delta AMP_{esb,0} \quad (11)$$

$$\delta W_{es}^G = \delta W_{es}^B + \delta S2S_S^{AMP} + \delta AMP_{esbl,0} \quad (12)$$

All residual components are assumed to be zero mean normally distributed random variables. Al-Atik et al. (2010) assumed that the different residual components are uncorrelated; hence, the total variance is simply the sum of the variances of the different residual components. However, Rodriguez-Marek et al. (2011) found non-zero correlation coefficients between some of the residual components they obtained from their analysis. The notation used in this manuscript for the residual components and corresponding standard deviations are slightly modified from the notations used by Al-Atik et al. (2010) and Rodriguez-Marek et al. (2011).

Previous research on non-ergodic GMPEs

Recent publications reflect the trend towards splitting the total variability in predicting ground motion parameters to be implement non-ergodic PSHA (e.g., Bindi et al. 2000; Chen and Tsai 2002; Atkinson 2006; Morikawa et al. 2008; Bindi et al. 2009; Anderson and Uchiyama 2011; Lin et al. 2011; Rodriguez-Marek et al. 2011; Dawood and Rodriguez-Marek 2013, Rodriguez-Marek et al. 2013). The different studies implement different approaches to split the residuals. Most of these studies start with the residuals obtained from an ergodic GMPE to obtain

the different components of variability as applicable to the non-ergodic GMPEs. This section summarizes these studies and list the key findings obtained from each.

Morikawa et al. (2008) used 7753 strong GM records from 50 earthquakes recorded at the K-NET and KiK-net stations in Japan. The data was selected based on strict selection criteria (e.g., these 50 earthquakes were generated from six different source areas with an area less than 50 km by 50 km such that the earthquakes from each area share the same source mechanism). They used the GMPE developed by Kanno et al. (2006) for Japan to calculate the residuals used in the study. The average residual for each site-source combinations was calculated and subtracted from the total residual to obtain the corrected residual. This correction reduced the source-, path- and site-specific effects; hence, the ergodic assumption was removed over the source-, path-, and site-components. The standard deviation of the corrected residual was substantially smaller than the pre-correction one (e.g., standard deviation of the corrected residual is 54% smaller than standard deviation of the pre-correction residual for PGA).

Bindi et al. (2009) developed a GMPE using 27 earthquakes recorded at 146 stations at distances (R_{JB}) up to 200 km from the ITACA database (Pacor et al. 2011). The residual was split in three steps. In the first step, the total residual was split into an inter-event and intra-event components (similar to Equation 2). In the second step the total residual was split into an inter-station and intra-station components. In the third step, the random error was calculated by subtracting the inter-event and inter-station components from the total residual. This approach to split the residual is different than the approach described by Al-Atik et al. (2010). They investigated the earthquakes and sites for which the inter-event and inter-station residual component was considerably high or low. Uncertainties in the magnitude, uncertainties in location of some events, deficiency in the energy emitted by the source, and region specific attenuation were considered the reasons behind the considerably high or low inter-event residuals observed for some earthquakes. The anomalous inter-station residuals were associated with region specific attenuation, presence of major structures close to the recording stations, and site-specific amplification characteristics that had been documented by other researchers. These observations show the importance of accounting for the path- and source-specific effects as well.

Anderson and Uchiyama (2011) proposed a methodology to correct a single-station GMPE to also account for the effect of single-path. The methodology was validated by developing a GMPE that follows the functional form used by Si and Midorikawa (1999) using PGA and PGV of strong GM records from Japan. They also developed a GMPE (new functional from) using the Guerrero accelerograph network in Mexico. The spatial distribution of the event- and site-corrected residual component is used to obtain contours around each station in the network. These contours represent the path correction and should be developed for each station. On the other hand, there should be correlations between the contours for close stations. Most notable in this study is the fact that it was implied that the contours a specific site can be obtained from background seismicity. This implies that the contours that result from large earthquakes (used to validate this methodology) should match the contours that results from small magnitude earthquakes.

Lin et al. (2011) developed a GMPE with a functional form slightly different than the one adopted by Chiou and Youngs (2008) using 4756 strong GM records from 64 earthquakes recorded at 285 stations in Taiwan. They use what they name a regionless approach to obtain the standard deviations of the path- and site-corrected residual component. In this approach they define a closeness index (*CI*) between two paths from two sources to a single site as the ratio between the distance between the two earthquakes and the average source-to-site distances. They describe a procedure through which they estimate the standard deviation of the path- and site-corrected residual component using *CI*. The standard deviation at $CI=0$ was considered to be the standard deviation of the path- and site-corrected residual component. The limitation of this approach (which is clear from their plots) is that there is no data at very low values of *CI*. This indicates that at $CI=0$, the path-specific corrected component of standard deviation is just an extrapolation that is not really constrained by data. The total standard deviation was reduced by 9%-14% when the site-specific standard deviation was constrained. The reduction was between 39% and 47% when the single-path was constrained as well. The later reduction is equivalent to constraining the site-, path- and source- terms simultaneously.

Rodriguez-Marek et al. (2011) used a subset of the KiK-net database up to 2004 to develop a site-specific GMPE using the functional form of Boore and Atkinson (2008). The

standard deviation of $\delta W_{es,0}^{B}$ and $\delta W_{es,0}^{G}$ (Equations 4 and 5) were considerably reduced compared to the standard deviation of $\delta W_{es}^{G/B}$(Equation 2). On the other hand the standard deviation of $\delta W_{es,0}^{B}$ and $\delta W_{es,0}^{G}$ from records at a single station was in some cases larger than the standard deviation of $\delta W_{es}^{G/B}$. These stations were found to be station with a wide range of station-to-event azimuth. This shows the importance of accounting the single-path as well. A clear trend was not found between V_{s30} and the standard deviation of $\delta W_{es,0}^{B}$ and $\delta W_{es,0}^{G}$.

Rodriguez-Marek et al. (2013) used residuals from five different regions around the world (i.e., California, USA; Japan; Taiwan; Turkey and Switzerland) to investigate the standard deviation of $\delta W_{es,0}^{G}$ (φ_{ss}). φ_{ss} was found to be relatively constant across the five different regions. They proposed models to estimate φ_{ss} as a function of magnitude or distance. This study also provided guidelines on how to use its results in non-ergodic PSHA.

The most notable observation from these studies is the similarity between the ergodic components of standard deviations even if these studies used different databases and GMPE functional forms. In most cases the authors did not provide guidance on the applicability of their results in non-ergodic PSHA. This study aims at addressing some of the gaps found in the previous studies. These gaps include but are not limited to the following:

- Lin et al. (2011), and Anderson and Uchiyama (2011) explicitly proposed methodologies to include the single-path effect in GMPEs. The former study used available data to fit a curve and extrapolated it to obtain the standard deviation of the path- and site-corrected residual component. Hence, the standard deviation of the path- and site-corrected residual component that they obtained might be questionable due to this extrapolation. In the later study, it was assumed that the contours used to correct for the path-specific effects are the same for large earthquakes and background seismicity. Such assumption is not yet validated, which limits the practical usage of this approach. Hence, another approach that overcomes the limitations of the previous approaches is needed.

- Most of the previous studies did not explicitly study all levels of non-ergodicity on the same database.
- The effect of the distribution of azimuth and distance of the records at different stations on the estimates of the site terms and site-specific variability at these stations is not addressed in any of these studies.

REFERENCES

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

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. Earthquake Engr. **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. Earthquake 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 predictinon equations for seismic hazard analysis. Seism. Res. Lett., **81**, 783-793.

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

Brune, J. N., 1970. Tectonic stress and the spectra of seismic shear waves from earthquakes, J. Geephys. Res., 75, 4997-5009.

Campbell K. W., and Y. Bozorgnia, 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.

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.

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

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

Chiu, H.-C., 1997. Stable baseline correction of digital strong-motion data, Bull. Seismol. Soc. Am. 87, 932–944.

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.

Darragh, B., W. Silva, and N. Gregor, 2004 . Strong motion record processing for the PEER center, Pacific Engineering, 311 Pomona Avenue, El Cerrito, California 94530

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. Seism. Soe. Am. **98**, 1915-1935.

Fan, G.-W and T. Lay, 1998. Regionalized versus single-station waveguide effects on seismic discriminants in western China, Bull. Seism. Soe. 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.

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

Joyner, W. B., and D. M. Boore, 1993. Methods for Regression Analysis of Strong-Motion Data, Bull. Seismol. Soc. Am. **83,** 469-487.

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.

Marsan, D., and O. Lengline, 2008. Extending earthquakes' reach through cascading, Science, 319 (5866, 1076-1079.

Matthews, M. V., W. L. Ellsworth, and P. A. Reasenberg, 2002, A Brownian model for recurrent earthquakes, Bull. Seismol. Soc. Am. **92**, 2233-2250.

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.

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

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.

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.

Özturk, S., and Y. Bayrak, 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., R. Paolucci, G. Ameri, M. Massa, and R. Puglia, 2011. Italian strong motion records in ITACA: overview and record processing, Bull. Earthquake Eng. 9, 1741-1759.

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

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., W.R. Walter, and E.M. Matzel, 2009b. A Simultaneous Multiphase Approach to Determine P-Wave and S-Wave Attenuation of the Crust and Upper Mantle, Bull. Seism. Soc. Am., **99**, 3314–3325.

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

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. Geological Survey Open-File Report 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.

Power, M., B. Chiou, N. Abrahamson, Y. Bozorgnia, T. Shantz, and C. Roblee, 2008. An Overview of the NGA Project, Earthquake Spectra. 24 3–21.

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

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. Seism. Soc. Am., **89**, 394–408.

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

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.

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

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

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

Zhang, T.-R. and T. Lay, 1994a. Analysis of short-period regional phase path effects associated with topography in Eurasia, Bull. Seism. 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.

Zhuang, J., D. Harte, M. J. Werner, S. Hainzl, and S. Zhou, 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., Y. Ogata, and D. Vere-Jones, 2002. Stochastic declustering of space-time earthquake occurrences, J. Am. Stat. Assoc., 97, 369-380.

Chapter 3: A Flatfile For The KiK-Net Database Processed Using An Automated Protocol

Submitted to EERI for publication in EARTHQUAKE SPECTRA

A Flatfile For The KiK-Net Database Processed Using An Automated Protocol

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

ABSTRACT

The Kiban-Kyoshin network (KiK-net) database is an important resource for ground motion (GM) studies. The processing of these records is a necessary first step to enable their use in engineering applications. In this manuscript we present a step-by-step automated protocol used to systematically process about 157,000 KiK-net strong ground motion records. The automated protocol includes the selection of corner frequencies for band-pass filtering. In addition, a comprehensive set of metadata was compiled for each record. As a part of the metadata collection, two algorithms were used to identify dependent and independent earthquakes. Earthquakes are also classified into active crustal or subduction type events; most of the GM records correspond to subduction type earthquakes. A flatfile with all the metadata and the spectral acceleration of the processed records is 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. When these data are used for the development of ground motion prediction equations (GMPEs) comprehensive sets of event and site metadata are needed, in addition to the GM data. The collection of these data, 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

[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

al. 2010 for Turkey; Arango et al. 2011-a for Central American subduction zone; Arango et al. 2011-b for Peru-Chile subduction zone; Pacor et al. 2011 for the ITACA database in Italy; and most recently Akkar et al. 2014 for the RESORCE database in Europe). The availability of processed databases is a key factor for advancing research in earthquake engineering in general, and for the development of GMPEs in particular. The objective of this paper is to introduce a 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. Additional information, not included in the paper due to length limitations, is included in an electronic supplement. The manuscript ends with some remarks and conclusions. The flatfile with the metadata and the spectral accelerations of the processed records is given in Dawood et al. (2014).

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

KiK-net

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

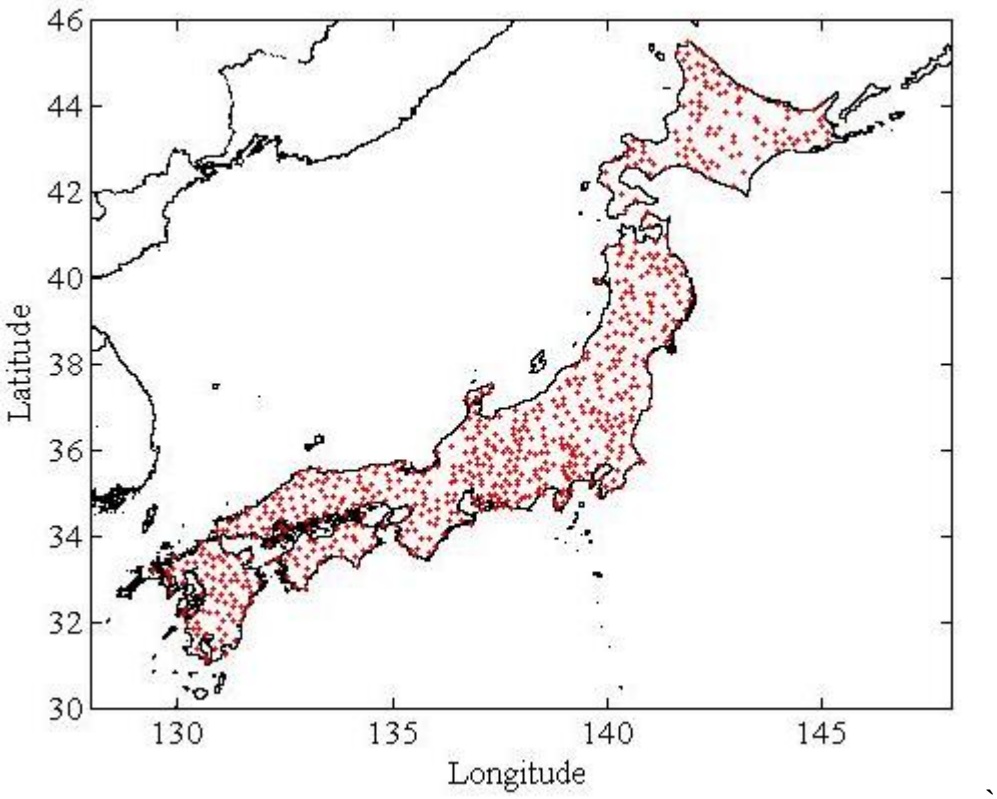


Figure 1. Location of the KiK-net stations (red dots) on the Japanese map

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

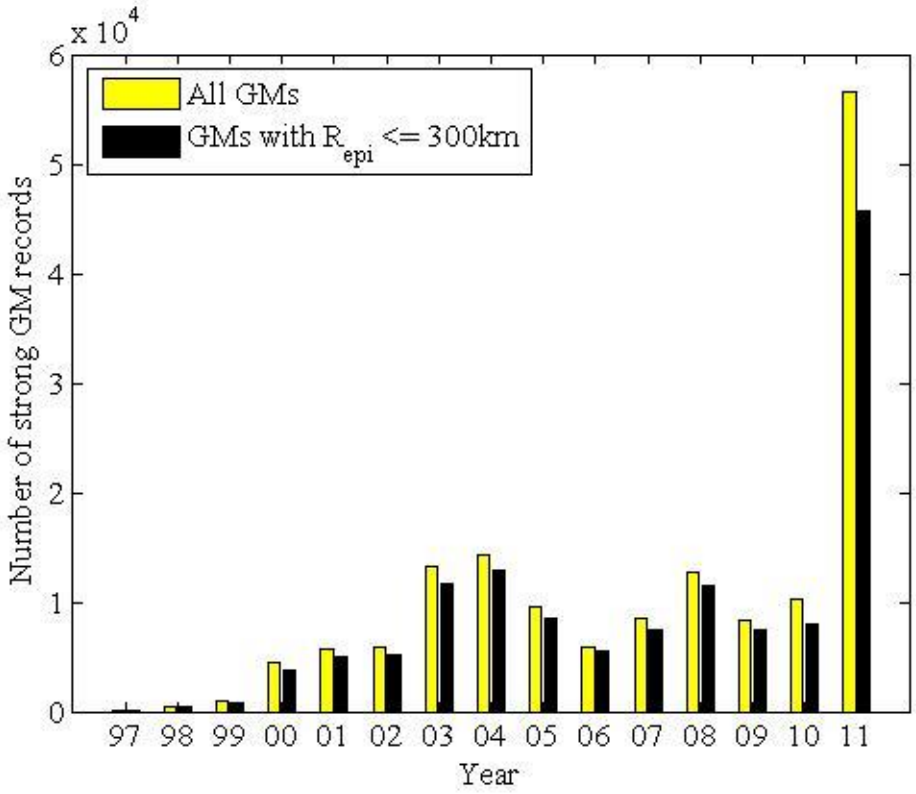


Figure 2. Temporal distribution of the ground motion (GM) records in the KiK-net database. Each GM consists of six components (three components at borehole and three at the ground surface).

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

		Borehole		Surface	
		$R_{epi}\leq 300$km	$R_{epi}>300$km	$R_{epi}\leq 300$km	$R_{epi}>300$km
Number of records with PGA* ≥	0.01g	3081	150	22454	1195
	0.05g	304	14	3025	97
	0.10g	107	4	1055	34
	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

F-net Seismic Catalog

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

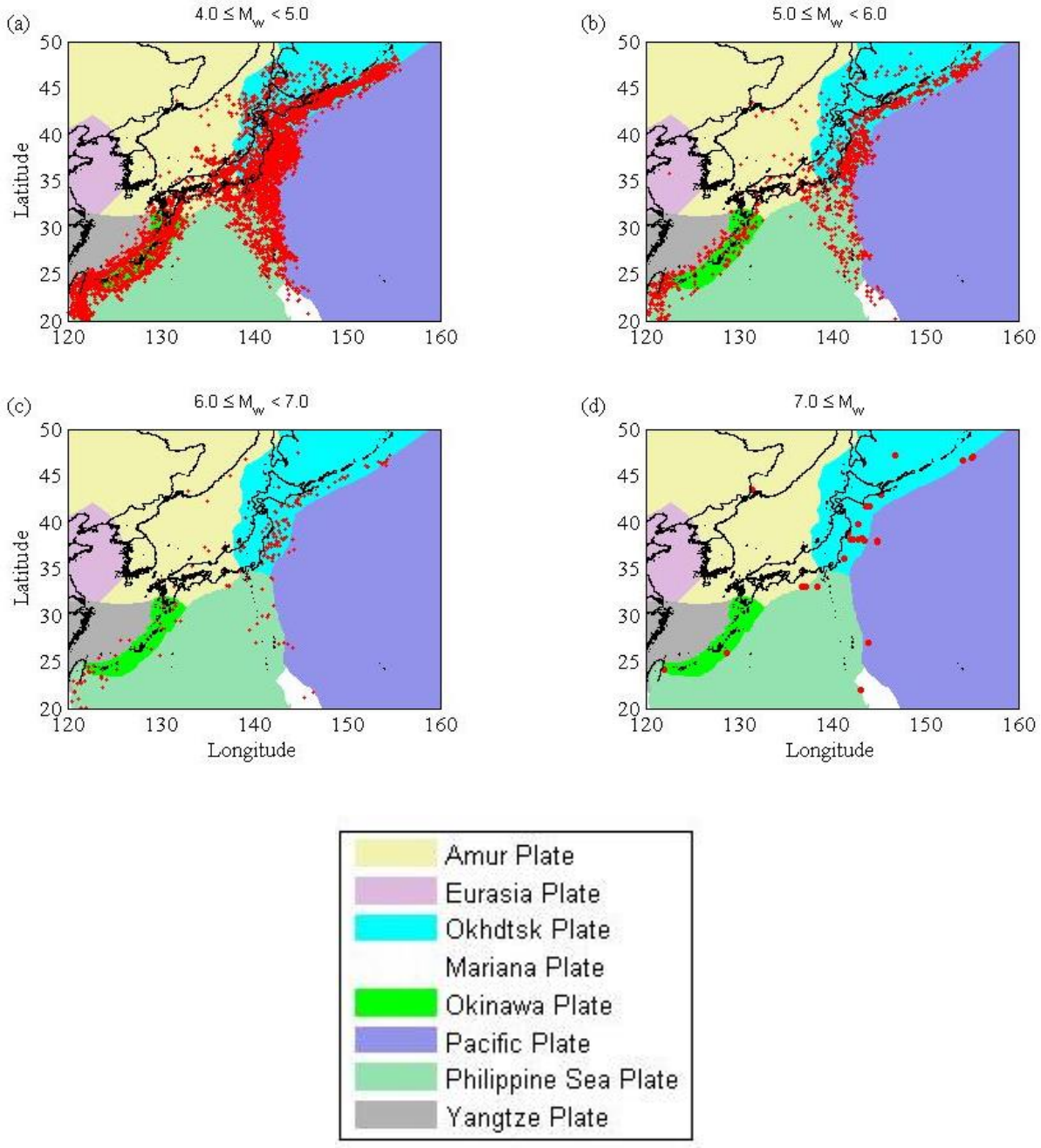


Figure 3. Spatial distribution of the earthquakes in the F-net catalog (between 1997 and 2011) for: (a) $4.0 \leq M_w < 5.0$, (b) $5.0 \leq M_w < 6.0$, (c) $6.0 \leq M_w < 7.0$, and (d) $M_w \geq 7.0$ (The different tectonic plates around Japan are assigned different colors (Plate boundaries obtained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

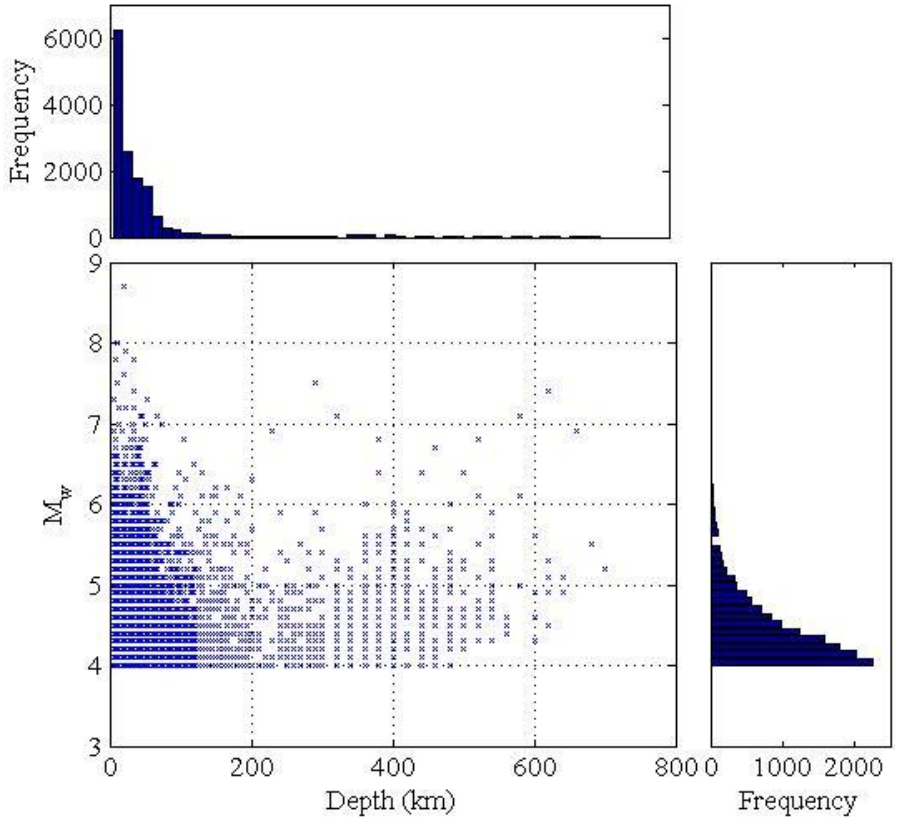


Figure 4. A scatter plot for M_w versus depth (km) for the earthquakes in the F-net catalog (bottom-left subplot) and two histograms that show the number of earthquakes in depth bins (top-left subplot) and the number of earthquakes in different M_w bins (bottom-right subplot).

PART II: AUTOMATED GM PROCESSING PROTOCOL

Due to the large size of the KiK-net database, a fully automated processing protocol is desirable. Figure 5 is a flowchart that provides an overview of the processing protocol. In general, we relied on the guidelines provided by Boore et al. (2002), Boore (2005) and Boore and Bommer (2005) to develop the GM processing protocol used in this manuscript. The idea behind automation is to use a heuristic approach to select parameters for ground motion processing and use these to process a record. The processed record is then checked against

several preset criteria. If these criteria are not fulfilled, another set of parameters is assumed and another processing loop is initiated. This loop is repeated until the processed record fulfills the criteria. The following paragraphs present a detailed description of the processing protocol along with a justification to the adopted criteria. Unless otherwise specified, the processing steps are applied on each of the six components of a record.

Step I. *Baseline correction.* A zeroth order baseline co^{rr}ection is performed by first subtracting the mean of the first 100 points from the whole acceleration time series (to account for a shift in the recorded acceleration time series) and then subtracting the mean of the pre-event noise window using an automated algorithm to detect the first arrival. The first arrival is defined as the first automatically detected arrival time for the six components (see Dawood 2014 for details).

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

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

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

where N is the order of the Butterworth filter and f_c is the high-pass corner frequency (in 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 following criteria:

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, and 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.
b. The ratio between the final and the maximum displacements must be less than 0.2.
c. A linear regression is applied to the trailing portion of the displacement and velocity time series. We assume the length of the trailing portion to be the last 10% of the recorded motion plus the length of the trailing zero pad (Equation 1). The slope of the best fit line must be less than 0.001 cm/sec for the displacement time series and 0.001 cm/sec^2 for the velocity time series.
d. The smoothed Fourier amplitude spectra (FAS) of acceleration (smoothed using the Konno-Ohmachi window with b=40, Konno and Ohmachi 1998) should follow an f^2 decay at the low frequency end (Boore and Bommer 2005). To check for this criterion, a line is fit between frequency and the smoothed FAS in log-log space using the smallest five frequencies greater than the corner frequency (selected in Step VI). If the slope of the line is found to range between 1.0 and 3.0 for all components of the record, then the FAS decay criteria is fulfilled. This check is not applied to records from earthquakes with $M_{JMA} \geq 6.0$.

The Brune (1970) source model predicts a slope of 2 for the FAS decay at the low frequency end of the spectrum. Our criteria is more flexible (e.g., accepts a slope that falls between 1.0 and 3.0) because the slope is calculated by fitting a line 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/fc second of the record. The FAS for the whole record component (obtained from Step I) and for 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 fc/0.5 and 30 Hz, the record is flagged with an "SNR < 3.0" (Boore and Bommer 2005) 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, fc/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.

The acceleration time series of the filtered components, including the zero-pads applied during processing, are saved. Stripping off the zero pads might compromise the filtered record especially at longer periods (Boore 2005, Boore et al. 2012). An additional check for records

with multiple wave trains or sub-events was also applied (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).

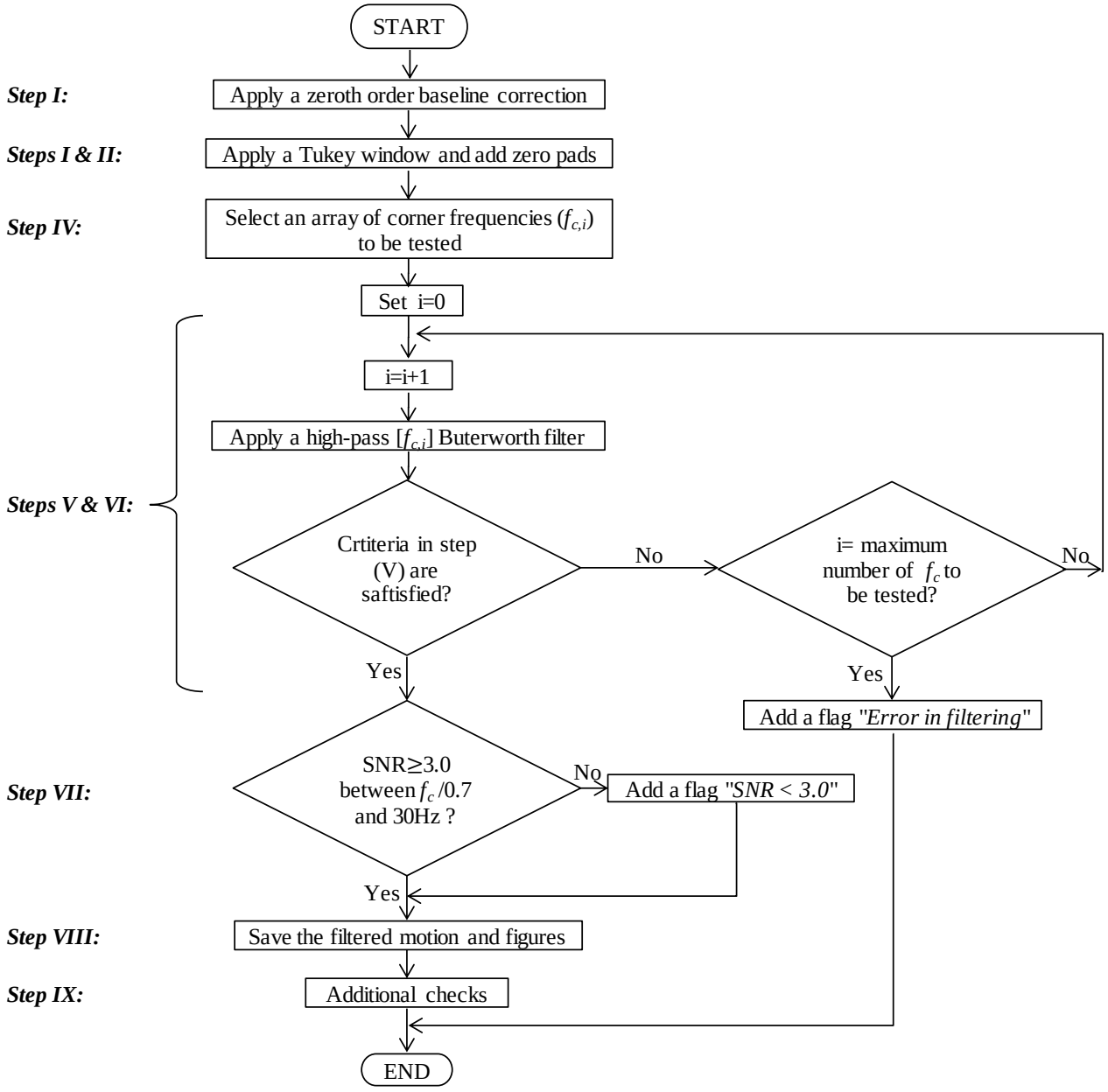


Figure 5. Flow chart diagram showing the steps of the processing protocol

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. Additional details on this comparison are given in Dawood (2014).

PART III: METADATA

This section presents the metadata collected for each record. The information includes event information (whether the event is a dependent event (e.g., an aftershock), earthquake type (e.g., shallow active crustal, interface subduction, etc.)), various source-to-site distance measures, and station information.

Earthquake information

For the engineering use of strong motion records, it is important to associate each record with an earthquake and to have an accurate parameterization of the source (e.g., $M_{w,}$ focal mechanism, rupture plane location and extent). The KiK-net data files provide a subset of these parameters (M_{JMA}, hypocenter location, and 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 source parameters are computed from a full-wave inversion code that uses about 70 different broadband waveforms (Fukuyama et al. 1998, Fukuyama and Dreger 2000). The automatic solution is then improved by manual operations. Hence, we consider the hypocentral location obtained from the F-net catalog more reliable than the values reported in the KiK-net 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 2). Category (A) represents the strictest error margin, category (D) contains earthquakes that were

manually matched, and category (E) contains earthquakes for which no match was found. A total of 4943 earthquakes with $M_{JMA} \geq 4.0$ were identified in the KiK-net database. Table 2 provides the number of earthquakes matched at each category and the number of records associated with each. About 89% of the matched earthquakes and records fall in category (A). The manually-matched events (category D) are mainly matched using the 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.

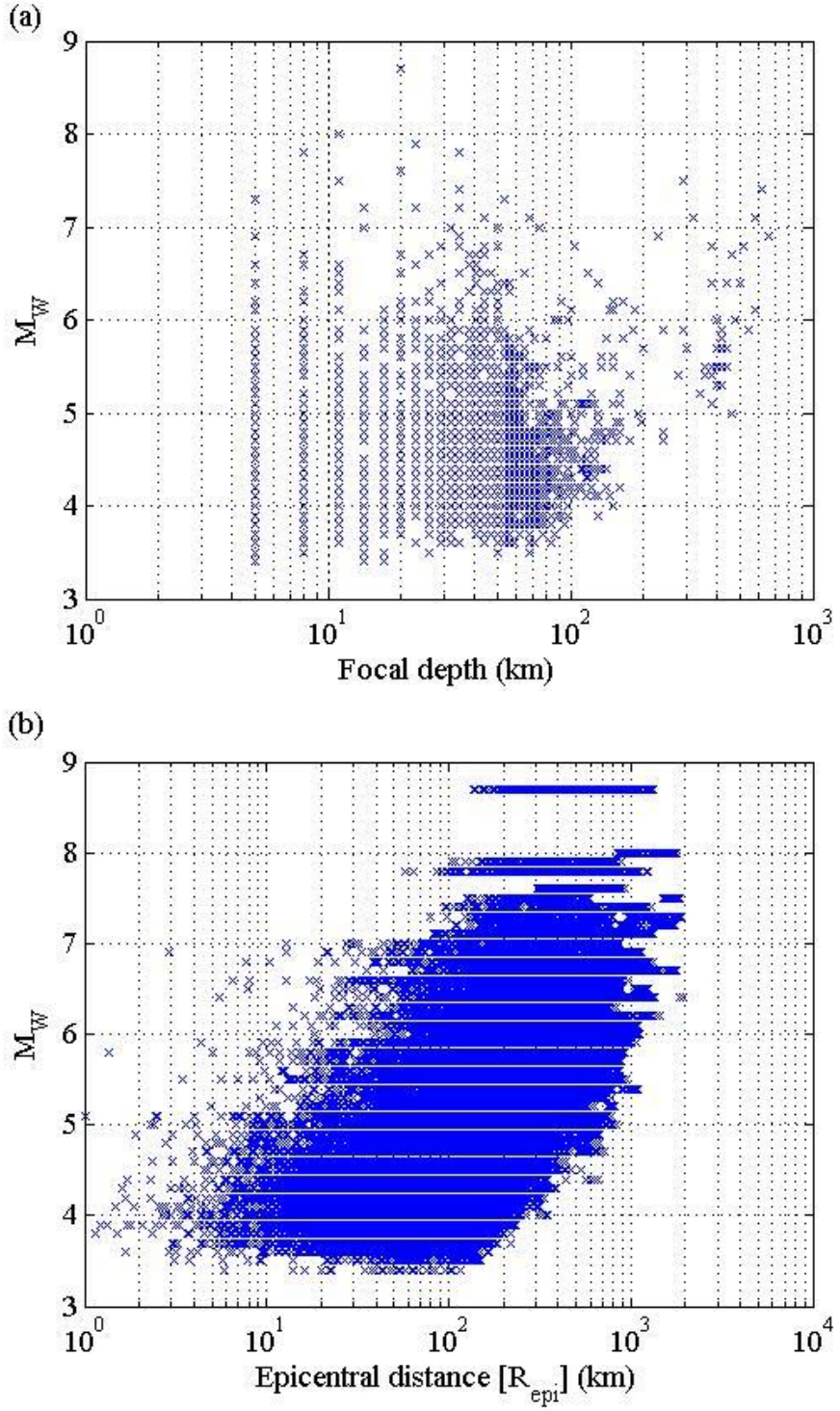


Figure 6. Plots of M_w magnitude for the events that fall within the matching categories (A) through (D) versus: (a) the focal depth and (b) epicentral distance for each GM record.

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

Category		(A)	(B)	(C)	(D)	(E)
Maximum error in:	Latitude (°)	0.1	0.1	0.15	-	-
	Longitude (°)	0.1	0.1	0.15	-	-
	Time (minutes)	1	60	60	-	-
	M_{JMA}	0.0	0.2	0.3	-	-
	Depth (km)	2.0	5.0	20	-	-
Number of earthquakes matched from the KiK-net database		4393	73	46	28	403
Number of GM records		139273	2807	2861	3238	8778

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 from the data used in developing the GMPEs (e.g., Campbell and Bozorgnia 2008 and Boore and Atkinson 2008), or treated differently by including specific terms for aftershocks (e.g., Abrahamson and Silva 2008 and Chiou and Youngs 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 3. 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.

Table 3. 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

Even if it is typical practice to assume that a declustered catalog follows a Poissonian model, this assumption is often debated (e.g., Matthews et al. 2002). Hence, it was decided to use the typical input parameters of Gardner and Knopoff's algorithm (see Table 3) 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 4 shows a summary of the input parameters and the output earthquake categories for both algorithms. It is clear from the table the higher level of complexity of the Garcia et al. (2012) algorithm. This algorithm was validated by automatically classifying a catalog of earthquakes that also were manually classified. For most event types, the validation showed considerable improvement in the number of correctly classified events using this algorithm over 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 5. The 951 earthquakes that were not classified using the algorithm by Allen et al. (2008) are events with $M_w \leq 7.7$ and depth greater than 50 km. 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 finding by Garcia et al. (2012) that the Allen et al. (2008) algorithm misclassified about 54% of the interface events. Figure 7 shows the location of the events in the four main categories as classified by Garcia et al. 2012.

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

Algorithm	Allen et al. (2008)	Garcia et al. (2012)
Input parameters	• Hypocentral depth • Magnitude • Region classification (active tectonic or stable continental).	• Earthquake location • Finn-Engdahl geographic regions (Young et al. 1996) • Hypocentral depth • Focal mechanism • Information about trench and subduction slab interface models (Hayes and Wald 2009, Hayes et al. 2009, and 2012). Info 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*

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

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

Class	Number of earthquakes classified by Allen et al. (2008)	Number of earthquakes classified by Garcia et al. (2012)
SZ intraslab	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.

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 the fault plane parameters (dip, strike and rake) provided by the F-net catalog using the software FOCMEC (Snoke 2003).

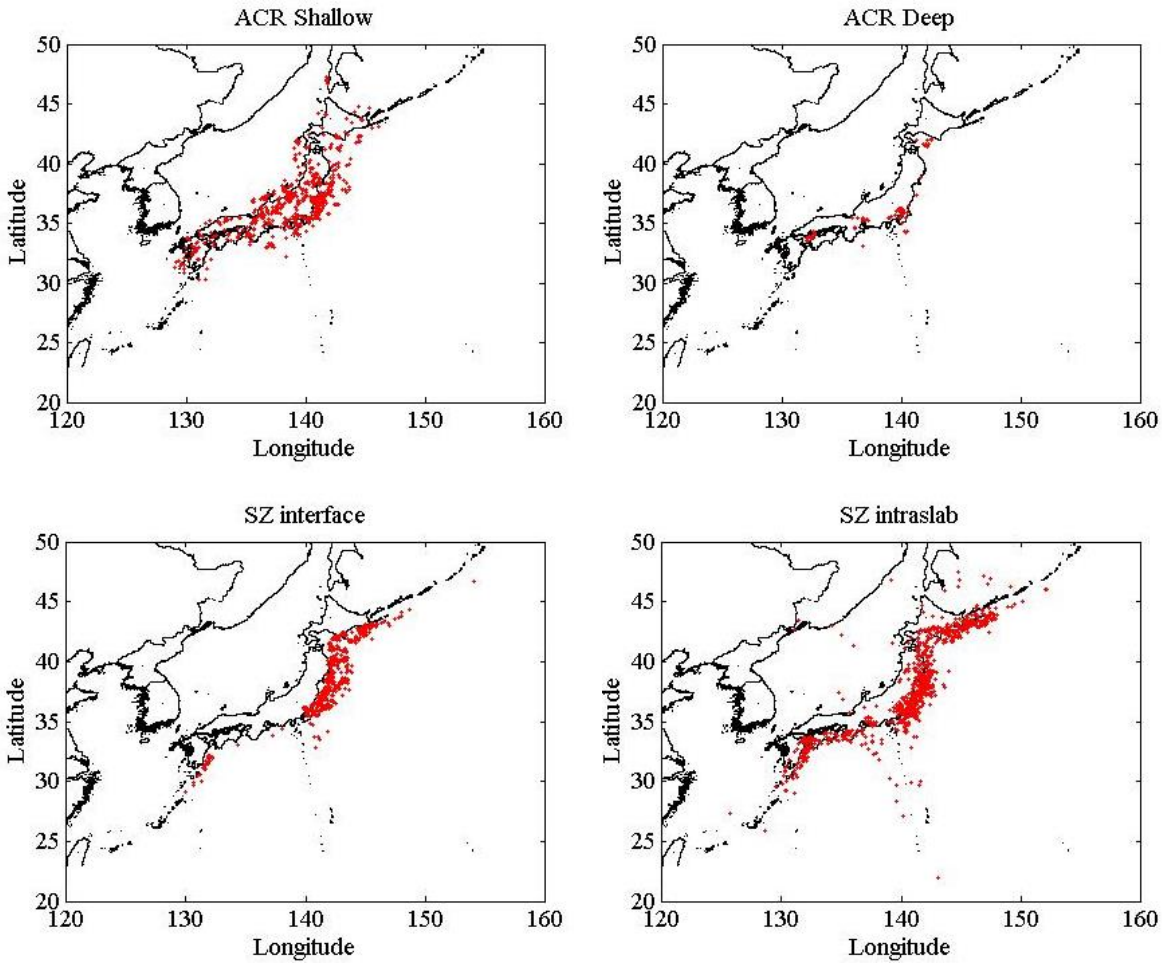


Figure 7. The location of the earthquakes classified as ACR shallow (top left); ACR deep (top right); SZ interface (bottom left); and SZ intraslab (bottom right). The classification is made using the algorithm by Garcia et al. (2012)

Distance Measurements

The epicentral distance, hypocentral distance, azimuth, and finite-source distance measures (closest distance to the fault, R_{rup}; closest distance to the surface projection of the fault, R_{JB}; and the horizontal distance from the top of rupture measured perpendicular to the fault strike, R_x) were computed for all the records from events for which there was a match in the F-net catalog. Finite-source rupture models for 21 earthquakes were found in the literature; these models were used to compute finite-source distance measures for 5,734 records associated with these earthquakes. For events for which a finite fault model is currently unavailable we simulate the rupture parameters following the method of Chiou and Youngs (2008-b, Appendix B). We

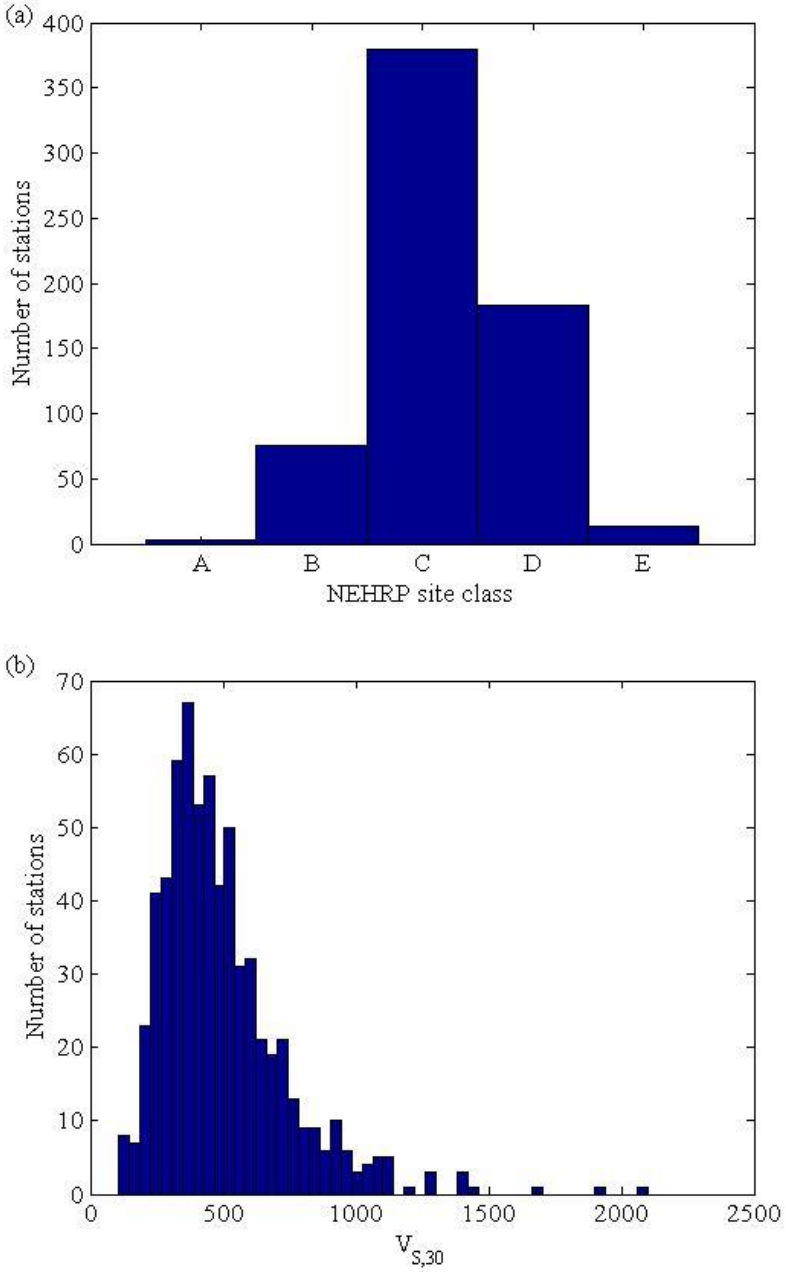
used the F-net hypocenter location, M_w, and focal mechanism. However, we could not determine which strike/dip pair was the actual fault plane for all events and so we compute the distances from 100 simulated faults for each plane separately. For each of the simulated faults, we compute the median R_{rup}, R_{JB}, R_x, and azimuth. Detailed information regarding the earthquakes with published finite-source rupture models and the methodology adopted to simulate the faults for the other earthquakes are presented in an electronic supplement to this manuscript.

Recording Stations

We calculated several parameters that characterize the soil profile at 655 KiK-net stations. Table 6 lists these parameters along with a brief description. The histograms that show the NEHRP site class (BSSC 2001), V_{s30}. and the depth of the borehole instrument at all stations are shown in Figure 8. Most of the stations fall into the NEHRP site classes C (58%) and D (28%). The median, mode and mean of the V_{s30} of all stations falls in the range between 445 and 500 m/s. The depth of the borehole instruments vary between 99 and 2008 meters.

Table 6. 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_{s30})
$V_{s,max}$	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 $V_{s30} > 1500$; B if $1500 \geq V_{s30} > 760$; C if $760 \geq V_{s30} > 360$; D if $360 \geq V_{s30} \geq 180$; and E if $180 > V_{s30}$


(a)
Number of stations
NEHRP site class
A
B
C
D
E
(b)
Number of stations
$V_{S,30}$

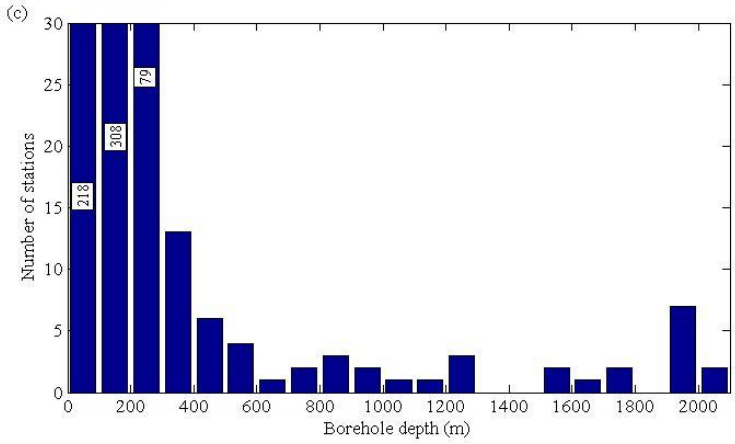


Figure 8. Histograms showing: (a) the NEHRP site class for KiK-net stations; (b) the V_{s30} for KiK-net stations; and (c) the depth of the borehole instrument for KiK-net stations

SUMMARY AND CONCLUSIONS

We presented an automated ground motion processing protocol and applied it to the KiK-net strong motion database to obtain corrected ground motions for events between October 1997 and December 2011. We used the F-net seismic catalog to obtain the earthquake metadata for each ground motion recording. This manuscript describes in detail the approach used to process the records and obtain the necessary metadata. Essential steps of the ground motion processing protocol include 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. Declustering algorithms were used to differentiate between dependent and independent events

for the purpose of GMPE development. Earthquakes were also classified by earthquake type. Two different earthquake classification algorithms were used in this study. Both algorithms showed that the earthquakes recorded by the KiK-net network are mainly active shallow crustal or subduction (interface and intraslab) type earthquakes. The flatfile containing all the data described in this article, in addition to the 5% damped response spectra, are publically available (Dawood et al. 2014).

DATA AND RESOURCES

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

The automated protocol is based on the Matlab scripts by Goulet and Bayless (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 from the fault plane (dip, strike and rake) to the stress axes (lower hemisphere trend a plunge of T, P, and B).

ACKNOWLEDGMENTS

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

Engineering Research Center (PEER) provided funding for some of the co-authors of this manuscript.

The authors would like to acknowledge the National Research Institute for Earth Science and Disaster Prevention (NIED) in Japan for providing the data used in this manuscript. The authors would like to acknowledge Dr. Daniel Garcia for classifying the different earthquakes in the F-net catalog on their behalf using the STREC code. The authors would also like to acknowledge several researchers who reviewed and discussed parts of the procedures adopted in this manuscript. In alphabetical order, Dr. Martin Chapman, Dr. Fabrice Cotton, Dr. Adrien Oth, Dr Frank Scherbaum, Dr. Peter Stafford. Their comments improved the quality of this manuscript.

REFERENCES

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

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., 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.

Bird, P., 2003. An updated digital model of plate boundaries, G3 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).

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 M2 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, G3 10 Q09006.

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

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

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.

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.

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.

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,

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

Snoke, J. A., 2003. FOCMEC: FOCal MEChanism determinations, in International Handbook of Earthquake and Engineering Seismology, W. H. K. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger (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.

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

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, 369-38

Chapter 4: A method for Including Path Effects in Ground Motion Prediction Equations: An Example Using the M_w 9.0 Tohoku Earthquake Aftershocks

Published in the Bulletin of the Seismological Society of America

Citation: 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." Bulletin of the Seismological Society of America, Vol. 103, No. 2B, pp. 1360–1372.

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

Haitham M. Dawood[1] 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 ground motion prediction equations 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 non-volcanic 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

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

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 non-ergodic 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 non-ergodic 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 ShakeMap network in California, has this become possible. Recent publications reflect the trend towards the development of GMPEs applicable to non-ergodic 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; and Rodriguez-Marek et al. 2011).

The main objective of this study is to present a methodology to account for regional dependences of anelastic attenuations 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 non-ergodic 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), Fan and Lay (1998), Phillips et al. (1998), Phillips (1999), Rodgers et al. (1999), Phillips et al. (2001), Fan et al. (2002), Ojeda and Ottemoller (2002), Taylor et al. (2002), Edwards et al. (2008), Ford et al. (2010), Pasyanos and Walter (2009), Pasyanos et al. (2009-a) and Pasyanos et al. (2009-b)]. 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 et al. 2009-a)].

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 fall 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 non-ergodic 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 and Zhao 2010), New Zealand (Cousins et al. 1999 and 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 non-volcanic 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 4^{th} 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., see 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 s and below was minimal. The high-pass filter serves as a base-line 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 was 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 0.7 to 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.

- Pseudo-acceleration 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 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 non-volcanic 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 section). 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 4 motions in a magnitude bin were not used in the analysis of that magnitude bin. Moreover, events with less than 4 records were also removed. A total of 7242 ground motion records recorded from 117 aftershocks fulfilled these criteria. Table 1 present a summary of the database used in this study.

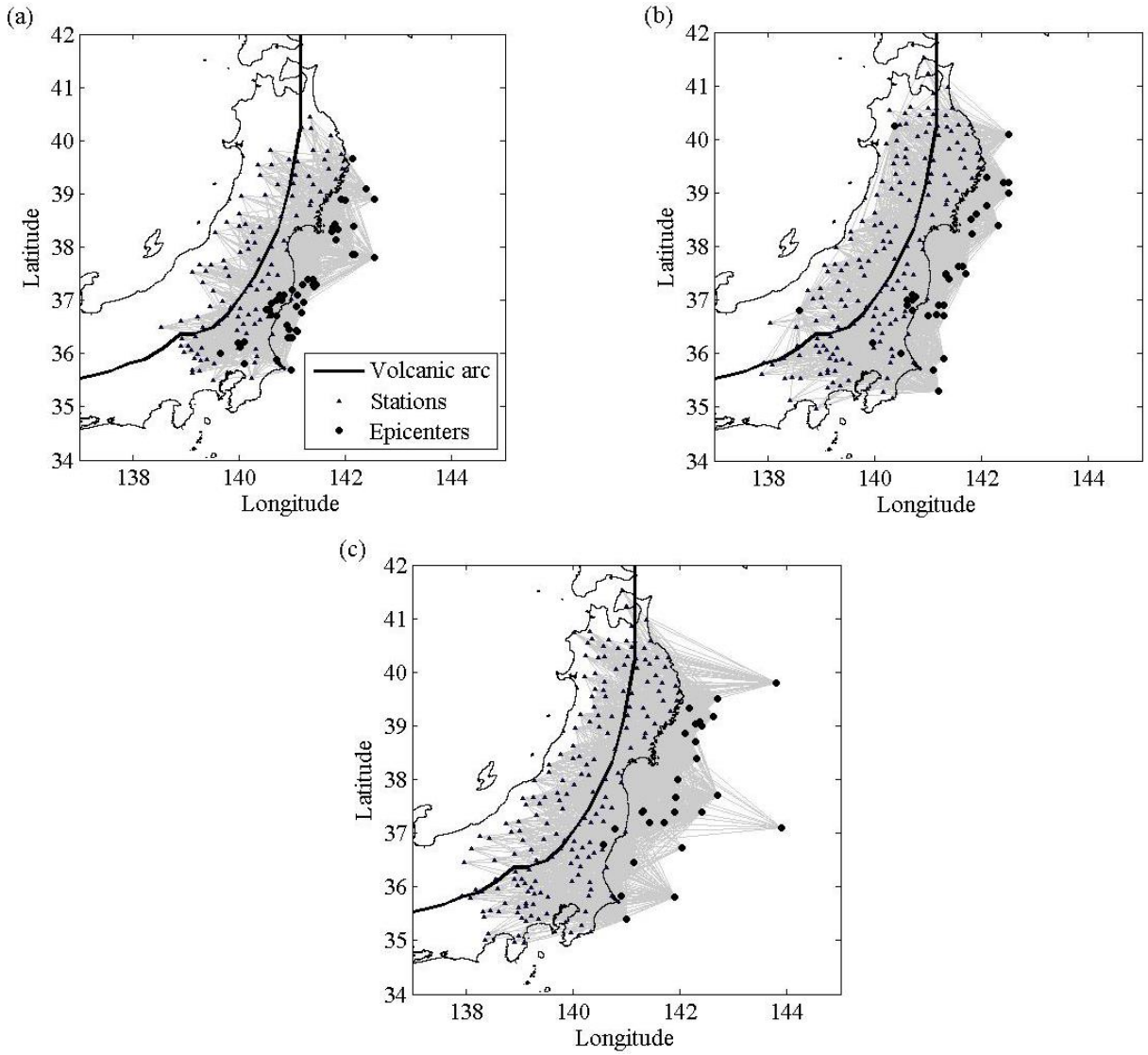


Figure 1. Map of Japan showing the earthquake epicenters (circles), recording stations (triangles), and travel paths used in the analyses. Data is separated in three magnitude bins. (a) $4 \leq M_{JMA} < 5$; (b) $5 \leq M_{JMA} < 6$; and (c) $6 \leq M_{JMA} < 7$.

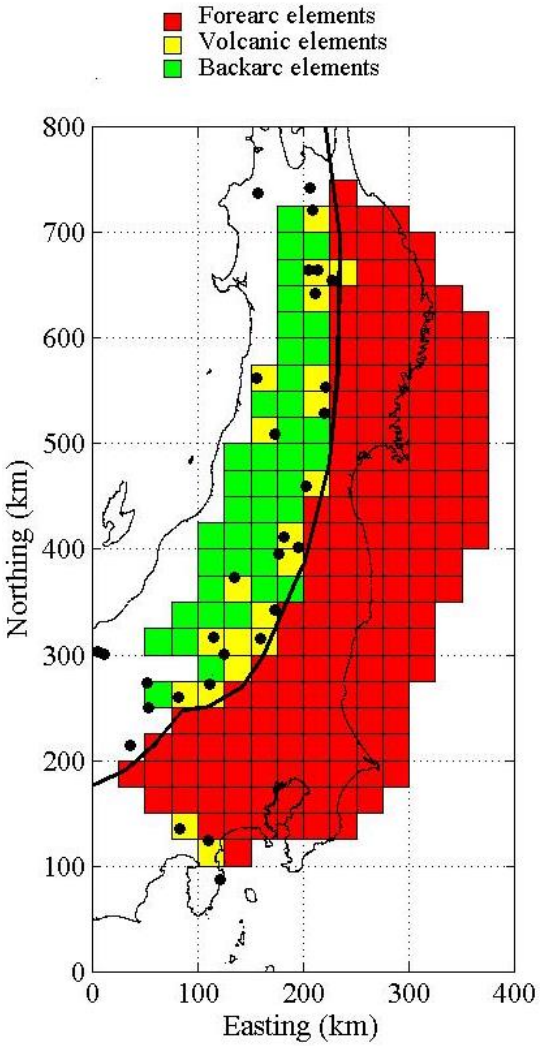


Figure 2. Map of Japan showing the locations of the active volcanoes (black circles), the volcanic belt (black line), and the different elements used in the analysis. Elements categorized as forearc, volcanic, and back-arc are colored in red, yellow, and green, respectively. [The origin of the axes is located at 34.2113N and 137.7016E].

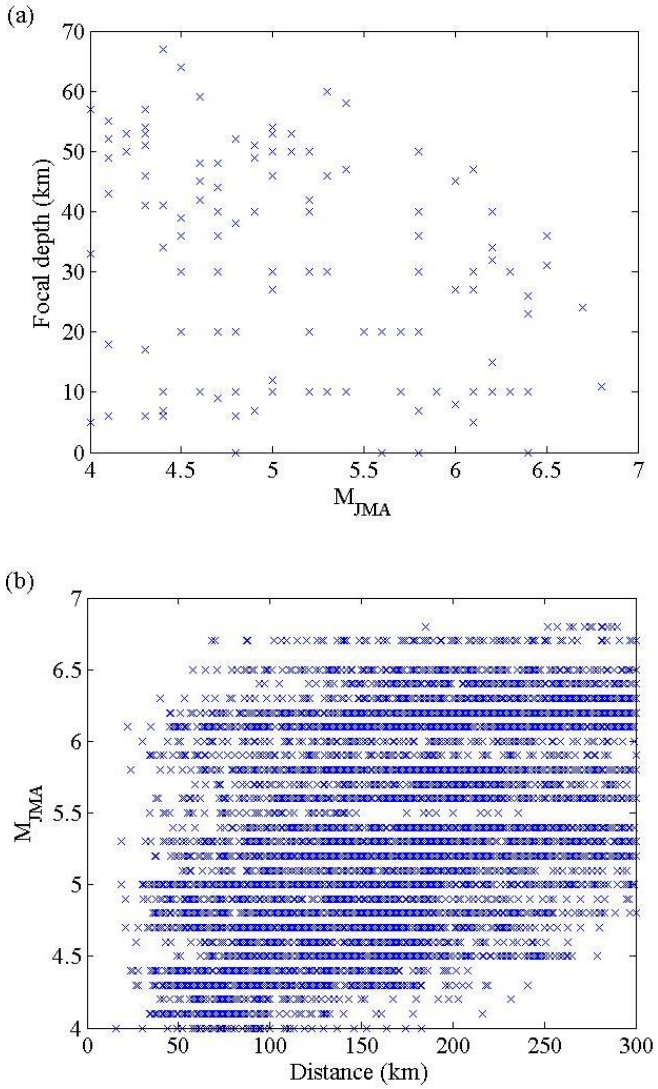


Figure 3. Plots of M_{JMA} magnitude for the events used in this study versus: (a) the focal depth; and (b) epicentral distance for each recorded ground motion.

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
Nember of stations	187	109	167	182
Nember of events	117	54	36	27

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, and 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 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 non-ergodic 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 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 Figure 2). In the regression 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.

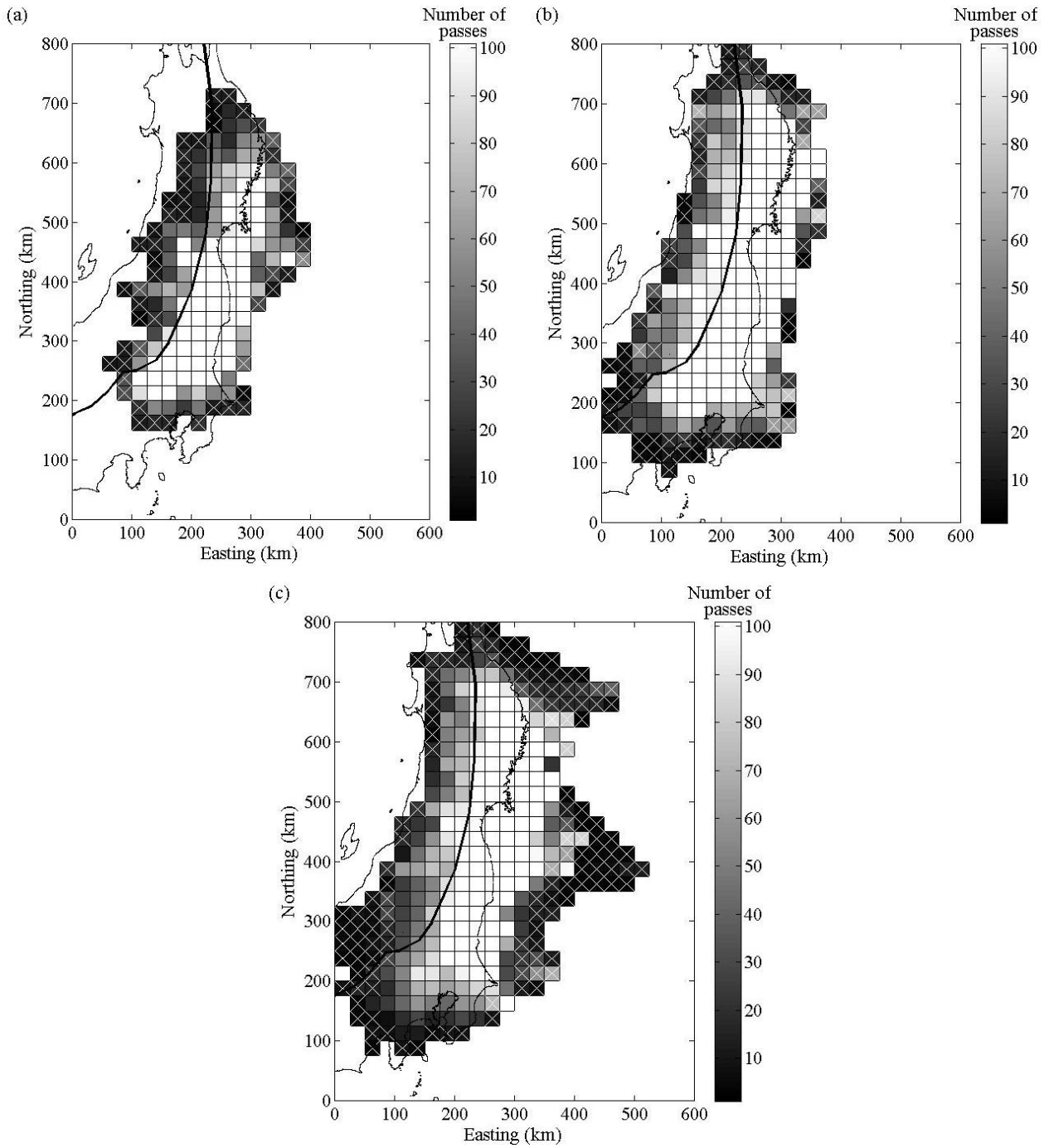


Figure 4. Map of Japan with the elements used in the analysis of each magnitude bin. Data is separated in three magnitude bins. (a) $4 \leq M_{JMA} < 5$; (b) $5 \leq M_{JMA} < 6$; and (c) $6 \leq M_{JMA} < 7$. (The element color represents the number of passes for each element. White elements are elements 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 34.2113N and 137.7016E]).

The ground motion parameters used in this study are the geometric mean of the pseudo-spectral 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

$$ln\ (yes) = ln(\mu es) + \delta S2Ss + \delta Be + \delta WSes \quad (1)$$

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; $\delta S2S_s$ 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 ($\delta S2S_s$, 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 notations follow that proposed by Al Atik et al. (2010).

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

$$ln(\mu_{es}) = C_1 + C_2\ ln\ (R) + C_3\ D + C_4\ ln(V_{s30}/V_{ref}) + \sum_{i=1}^{n} \delta_i P_{ies} \quad (2)$$

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 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 i^{th} element; and n is the total number of elements in which the region under study was divided in. 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 set 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. A 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 $\delta S2S_s$ 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 non-volcanic 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 (Figure 4). Table 2 summarizes the values of the parameters calculated from the mixed effects analysis. Figure 5 shows the path-attenuation terms for the non-volcanic 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 ($\delta S2S_s$) 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 was a linear trend with respect to M_{JMA}. This bias should not affect our conclusions.

Table 2. Model parameters from the mixed effects analysis

Period	Bin 1 ($4 \leq M_{JMA} < 5$)			Bin 2 ($5 \leq M_{JMA} < 6$)			Bin 3 ($6 \leq M_{JMA} < 7$)		
(s)	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

We check 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 run 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.

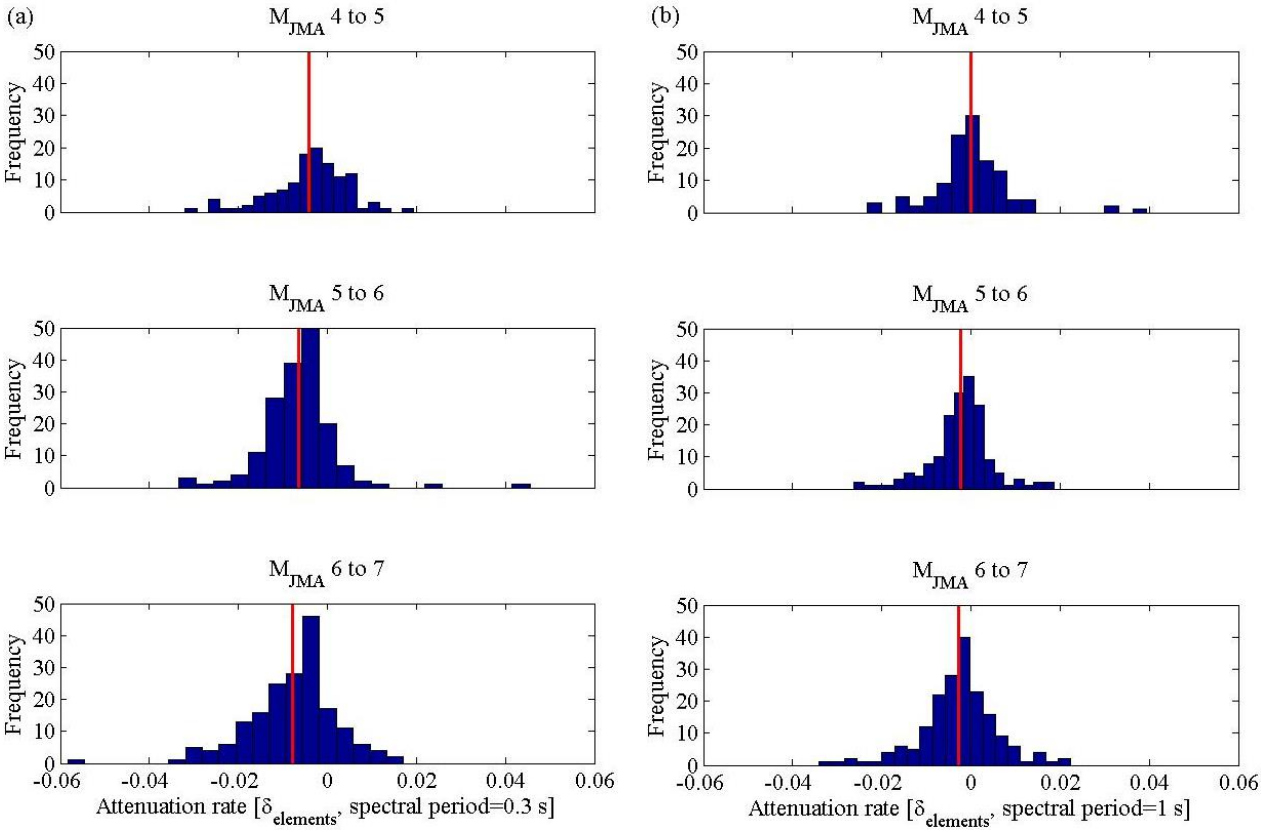


Figure 5. Histograms for the estimated attenuation rate ($\delta_{elements}$) through the non-volcanic elements for spectral periods: (a) 0.3 s.; and (b) 1.0 s. The vertical lines (in red) represent the average attenuation rate for all the elements shown in each histogram.

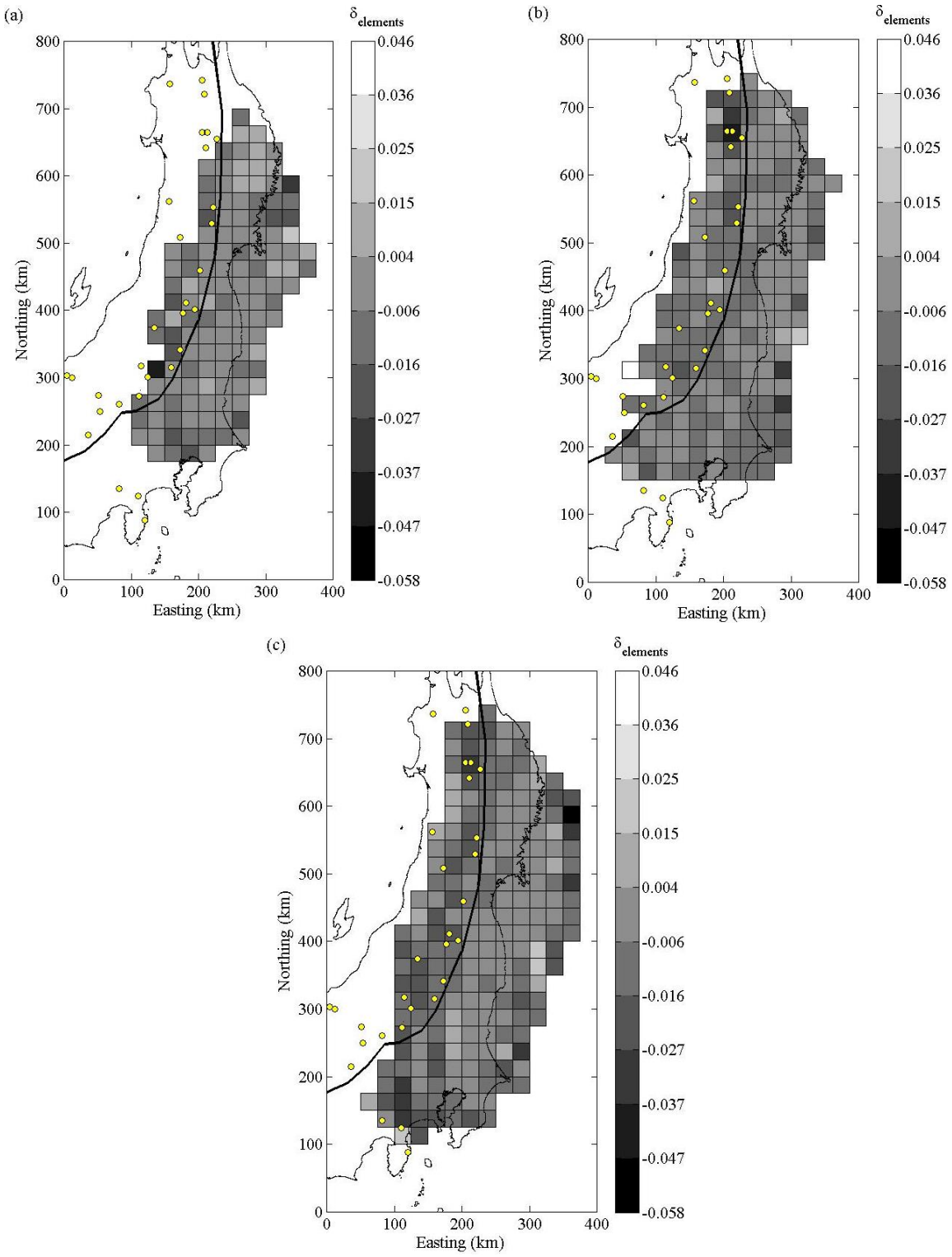


Figure 6. Map of Japan with the elements used in the analysis. The element color represents the attenuation rate ($\delta_{elements}$) estimated for each element at a spectral period of 0.30 s for: (a) $4 \leq M_{JMA} < 5$; (b) $5 \leq M_{JMA} < 6$; and (c) $6 \leq M_{JMA} < 7$. The active volcanoes are shown as yellow circles. [The origin of the axes is located at 34.2113N and 137.7016E].

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 Non-ergodic 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 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 world-wide data by Rodriguez-Marek and Cotton (2011).

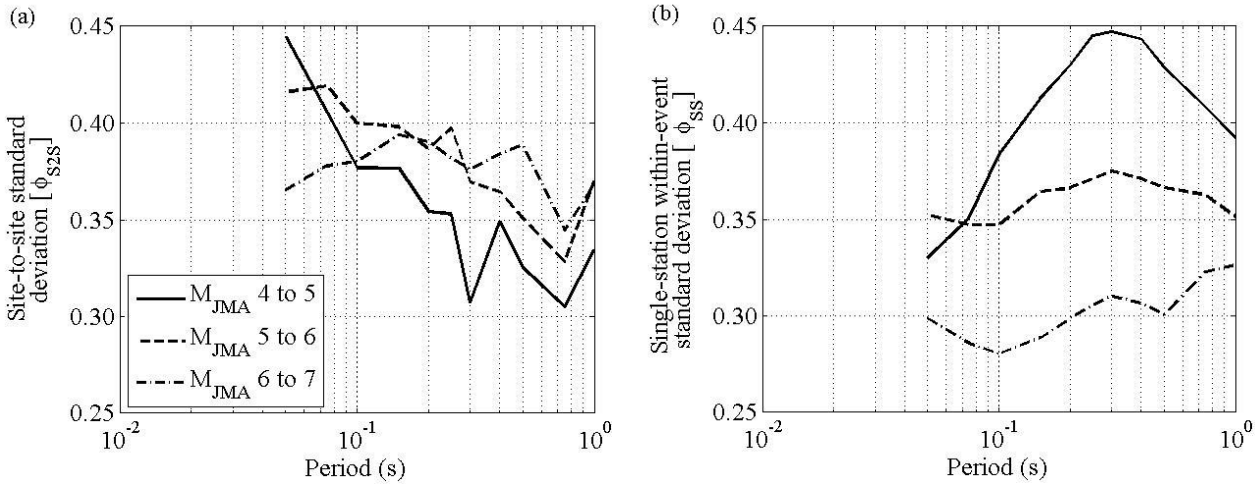


Figure 7. Components of standard deviation calculated from the three magnitude bins for: (a) the site-to-site standard deviation (φ_{S2S}); and (b) the single-station, within-event standard deviation (φ_{SS}).

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 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.

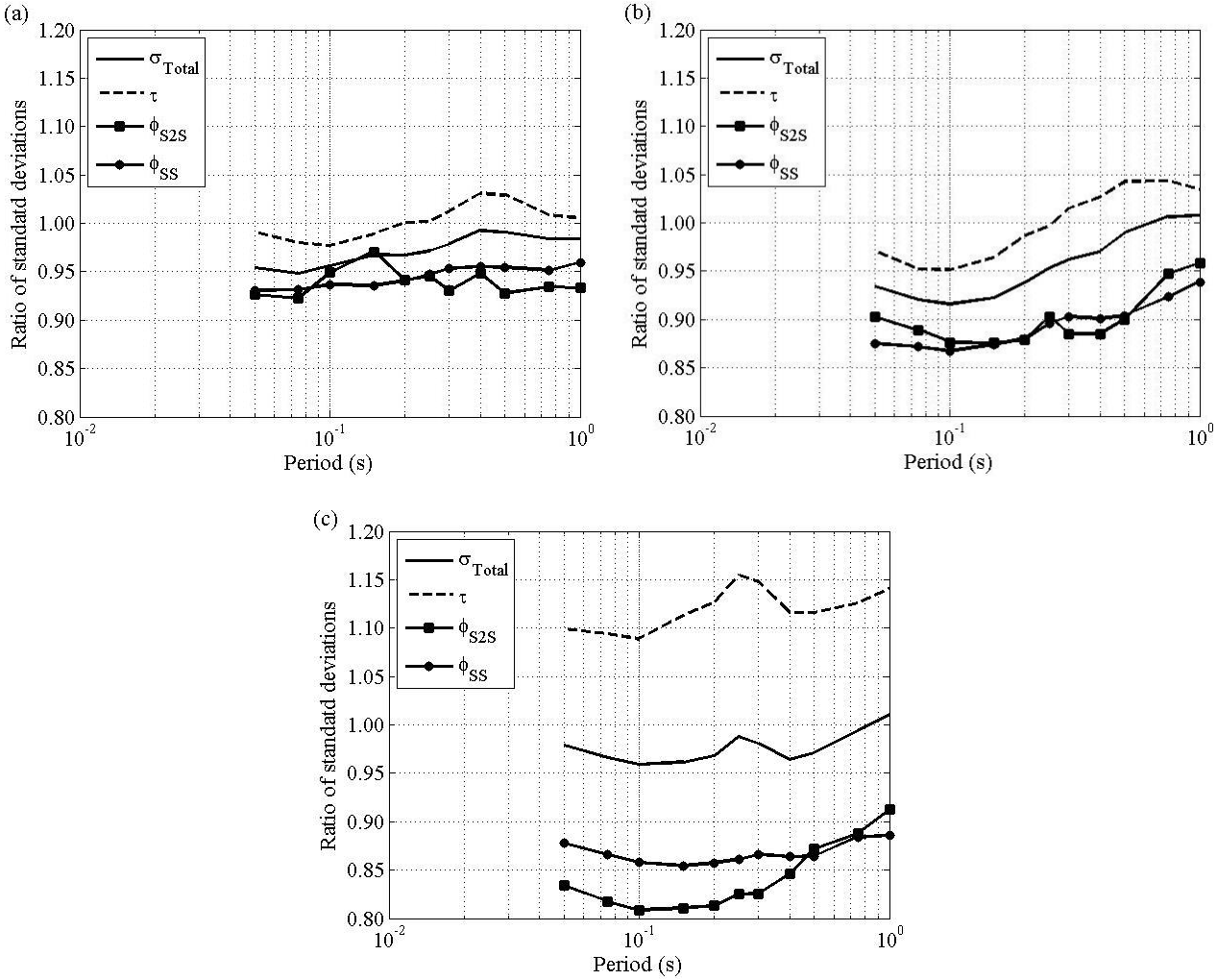


Figure 8. Ratio of the total (σ_{Total}), site-to-site (φ_{S2S}), event (τ), and the single-station, within-event (φ_{SS}) standard deviations for an analysis including the path specific attenuation terms to their counterpart standard deviations calculated from an analysis that does not include the path specific attenuation terms for:(a) $4 \leq M_{JMA} < 5$; (b) $5 \leq M_{JMA} < 6$; and (c) $6 \leq M_{JMA} < 7$.

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.

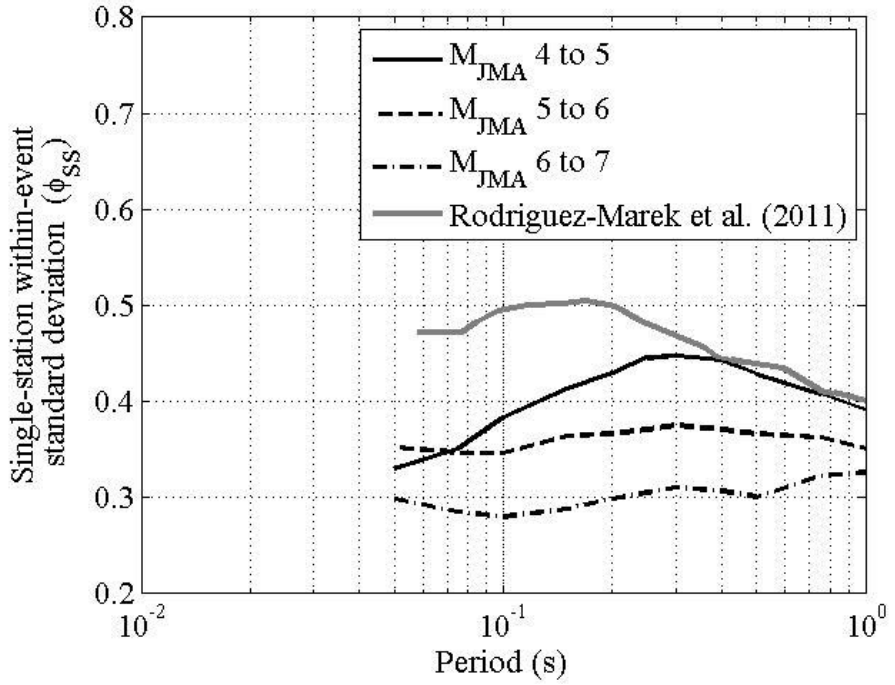


Figure 9. Single-station within-event standard deviation that accounts for path effects (φ_{SS}) obtained in this study for three magnitude bins compared to the φ_{SS} values from Rodriguez-Marek et al. (2011).

Forearc Versus Back-Arc 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 back-arc elements (west of the volcanic arc) (Figure 2). The average attenuation rate for each sub-group 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 non-volcanic elements have the same average attenuation 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 1s for the third bin ($6 \leq M_{JMA} < 7$) (i.e., the null hypothesis that volcanic and non-volcanic elements have the same average attenuation rate couldn't 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 wave lengths 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 (Figure 1).

At the same confidence level, the attenuation in the forearc and back-arc 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 back-arc 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 back-arc elements was not statistically different for the analysis using the Atkinson and Mereu 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 non-volcanic 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.

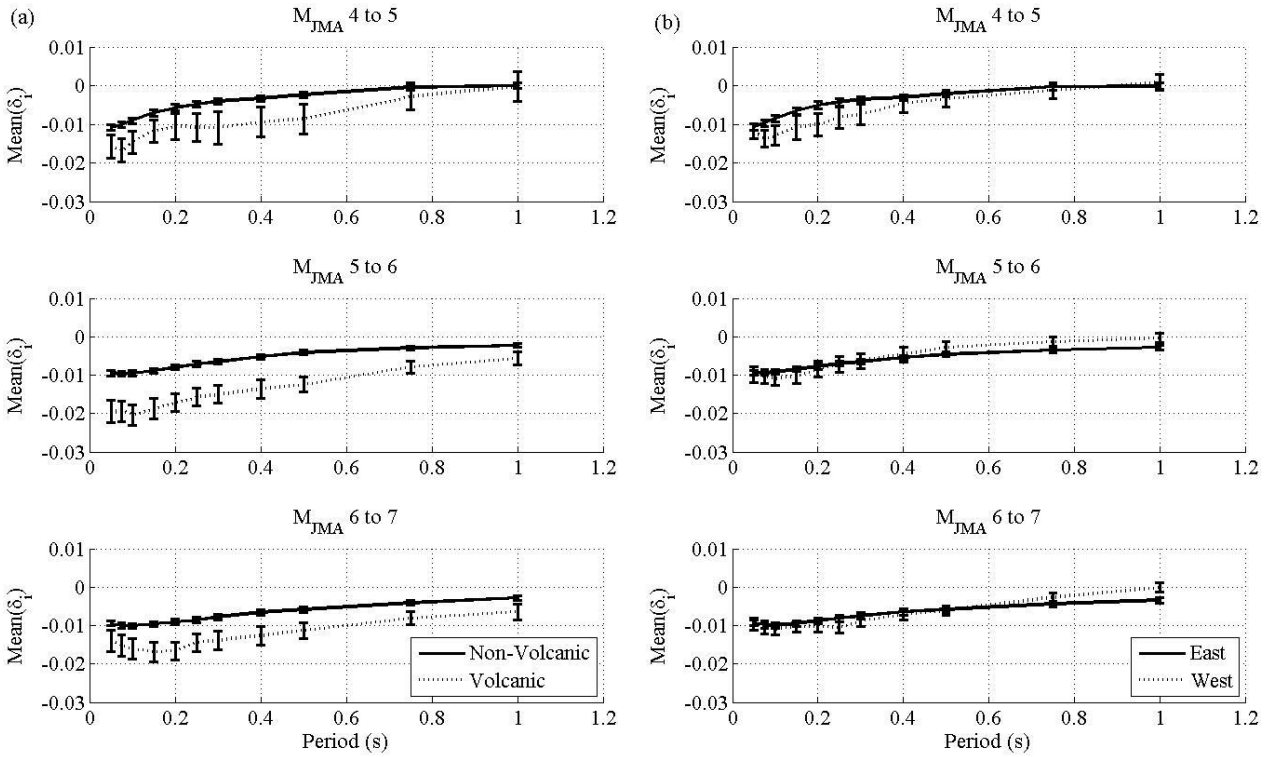


Figure 10. Average attenuation rates for the three magnitude bin at all periods for: (a) volcanic and non-volcanic elements; and (b) forearc and back-arc elements. The error bars represent the standard error of the estimate of the attenuation rate for each subset of elements.

Correlation Coefficients between Attenuation Rates for Different Magnitude Ranges

The estimation of site or path terms for non-ergodic 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., see 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 (Figure 11) and the elements' attenuation terms (Figure 12). For site terms, the correlation coefficients are relatively strong at

all periods (0.71 to 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.

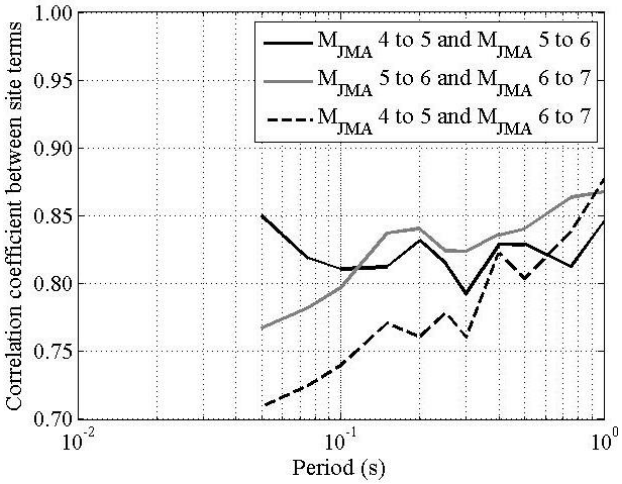


Figure 11. Correlation coefficients between site terms (δS2S) estimated for different magnitude bins.

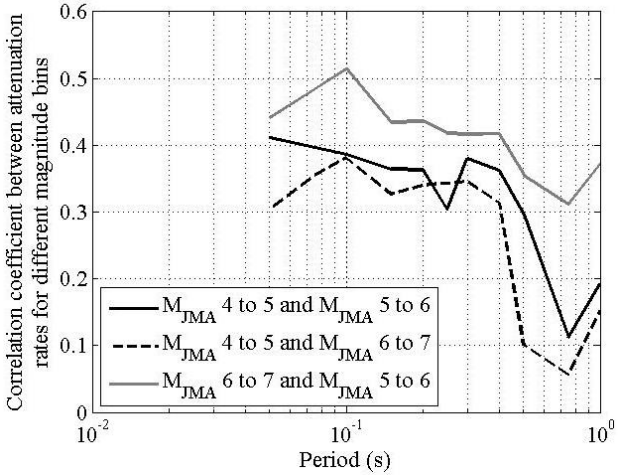


Figure 12. Correlation coefficients between attenuation rates in the elements ($\delta_{elements}$) estimated from different magnitude bins.

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 non-volcanic elements (forearc and back-arc 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 non-ergodic 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 off-shore. This may have biased the estimates of path attenuation. However, most of the GMPE in subduction regions are 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, it 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 non-unique. 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.

ACKNOWLEDGEMENTS

This research was partially supported by the U. S. Geological Survey (USGS), Department of the Interior, under USGS award number G11AP200049. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U. S. government. Additional support was provided by Virginia Tech.

The authors would like to acknowledge the National Research Institute for Earth Science and Disaster Prevention (NIED) in Japan for providing the data for this analysis.

The authors would also like to acknowledge Dr. Arthur Rodgers, Dr. Justin Rubinstein and an anonymous reviewer for their constructive comments. Their comments improved the quality of this manuscript.

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. Earthquake Engr. **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. Earthquake 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 predictinon equations for seismic hazard analysis. Seism. Res. Lett., **81**, 783-793.

Boore, D.M., and Bommer, J.J. (2005). Processing of strong-motion accelerograms: needs, options and consequences, Soil. Dyn. Earthquake 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. Soc. Earthquake 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. Seism. Soe. Am. **98**, 1915-1935.

Fan, G.-W and T. Lay (1998). Regionalized versus single-station waveguide effects on seismic discriminants in western China, Bull. Seism. Soe. 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. 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.

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

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., W.R. Walter, and E.M. Matzel (2009b). A Simultaneous Multiphase Approach to Determine P-Wave and S-Wave Attenuation of the Crust and Upper Mantle, Bull. Seism. 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. Geological Survey Open-File Report 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., 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. Seism. Soc. Am., **89**, 394–408.

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

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 (1994a). Analysis of short-period regional phase path effects associated with topography in Eurasia, Bull. Seism. 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.

Chapter 5: Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations

To be submitted for publication

Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations

Haitham M. Dawood[2] and Adrian Rodriguez-Marek

ABSTRACT

This manuscript presents the development of empirical ground motion prediction equations (GMPEs) for active crustal earthquakes using the Japanese KiK-net database. GMPEs are developed using an ergodic and a site-specific formulation. The database used in developing the GMPE consists of 13735 six-components strong ground motions from 679 active crustal earthquakes recorded at 643 KiK-net stations. Mixed effects regression analyses were used to obtain the regression coefficients for both GMPEs. The median predictions and the different component of variability obtained from both GMPEs were compared with previous GMPEs developed for Japan and western United States. The observed differences between the different GMPEs were a result of using strong ground motion databases that are different in nature and size. A comprehensive residual analysis was conducted to capture any potential biases in the residuals. Based on this residual analysis, the model was found to be applicable to moment magnitudes 4 - 6.5, depth to top of rupture 0 - 45km, distances 30 - 250 km, average shear wave velocity of the top 30 meters of the sites 180 - 1500 m/sec and spectral periods of 0 - 7 sec.

INTRODUCTION

Empirical ground motion prediction equations (GMPEs) are widely used in seismic hazard analyses to predict the median and standard deviation of ground motion parameters. While median predictions are important, the standard deviation can have a large impact on

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

hazard estimates, in particular for critical facilities that are designed for long return periods (Bommer and Abrahamson 2006). Traditionally, the standard deviation of a GMPE was computed assuming that the variability associated with a specific site-path-source combination is equal to the standard deviation computed from an entire database with various site-path-source combinations. This assumption has been referred to as the ergodic assumption (Anderson and Brune 1999). If a ground motion dataset is sufficiently large such that the repeatable contributions of site and/or path can be computed, then the standard deviations can be computed without recourse to the ergodic assumption; these standard deviations are referred to as single-station or single-path standard deviation. Several recent seismic hazard analyses project have used a non-ergodic approach, highlighting the importance of evaluating how non-erdogic GMPEs are computed. In this study we use a large database of strong ground motions recorded by the KiK-net array in Japan to develop two separate GMPEs for spectral acceleration (5% damping): one in which the traditional assumption of ergodicity is used and another in which the repeatable site terms are computed along with the single-station standard deviation. We refer to the latter as a site-specific GMPE.

The estimates of non-ergodic standard deviations are usually obtained from analyses of residuals from an ergodic GMPE. However, the median estimate from ergodic and non-ergodic GMPEs are not necessarily the same (Stafford 2014). In this paper, the differences between the ergodic and site-specific formulations, both in the median and the standard deviation, are explored. The paper also presents a comparison of these GMPEs with existing ones for Japan and elsewhere. The proposed GMPE is used in a companion paper to develop a model for single-station standard deviation.

DATABASE

The database used in this manuscript consists of 13,735 six-components (three at the surface, three at a borehole station) strong ground motions records from 679 active crustal earthquakes recorded at 643 KiK-net stations. The “active crustal” classification is based on the automated algorithm proposed by Garcia et al. (2012). Most of the borehole records are located at a depth of 100 or 200 meters, but some are located at larger depths. Figure 1 shows the distribution of the earthquakes and recording stations on the Japanese map and Figure 2 shows

the distribution of earthquake and site parameters for the records used in this study. Observe that only about 25% of the motions in this database were recorded from earthquakes with moment magnitude (M_w) greater than or equal to 5.0 (Figure 2).

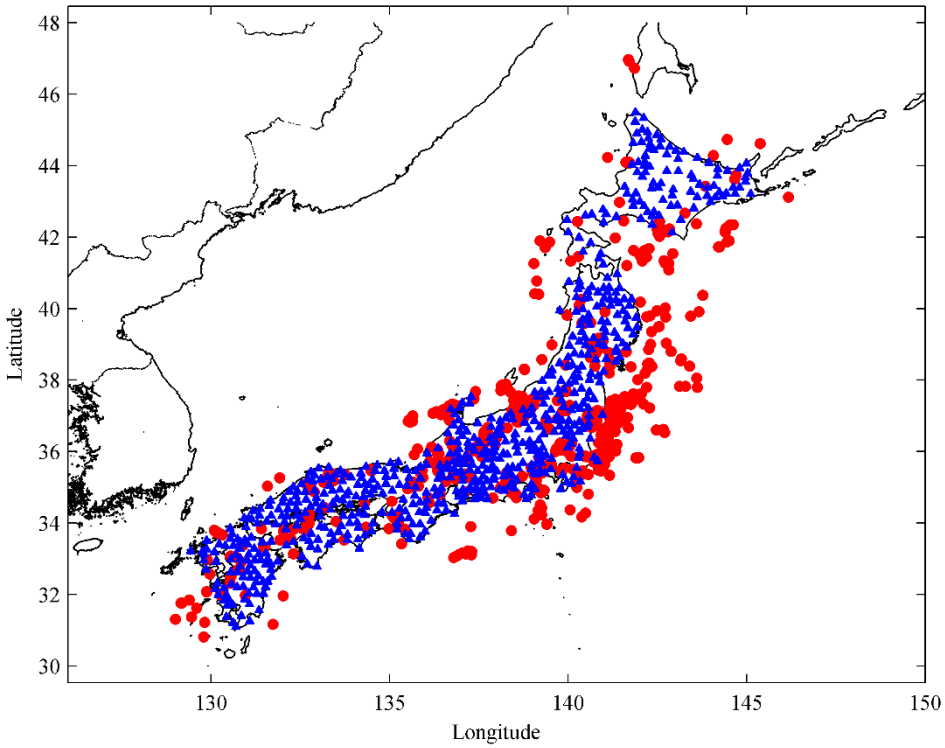


Figure 1. The distribution of the earthquakes (red circles) in the current database and KiK-net recording stations (red triangles) on the Japanese map.

The records were processed using an automated protocol (Dawood et al. 2014). Event metadata was obtained from the information provided in the F-net seismic catalog for all earthquakes except for 10 earthquakes for which published finite fault models were used (see Dawood et al. 2014). Closest distances to the ruptured fault (R_{rup}) were computed using either the published finite fault models or the methodology of Chiou and Youngs (2008). Site metadata was obtained from the information available in the KiK-net website. For additional details about the metadata collection, the reader is referred to Dawood et al. (2014).

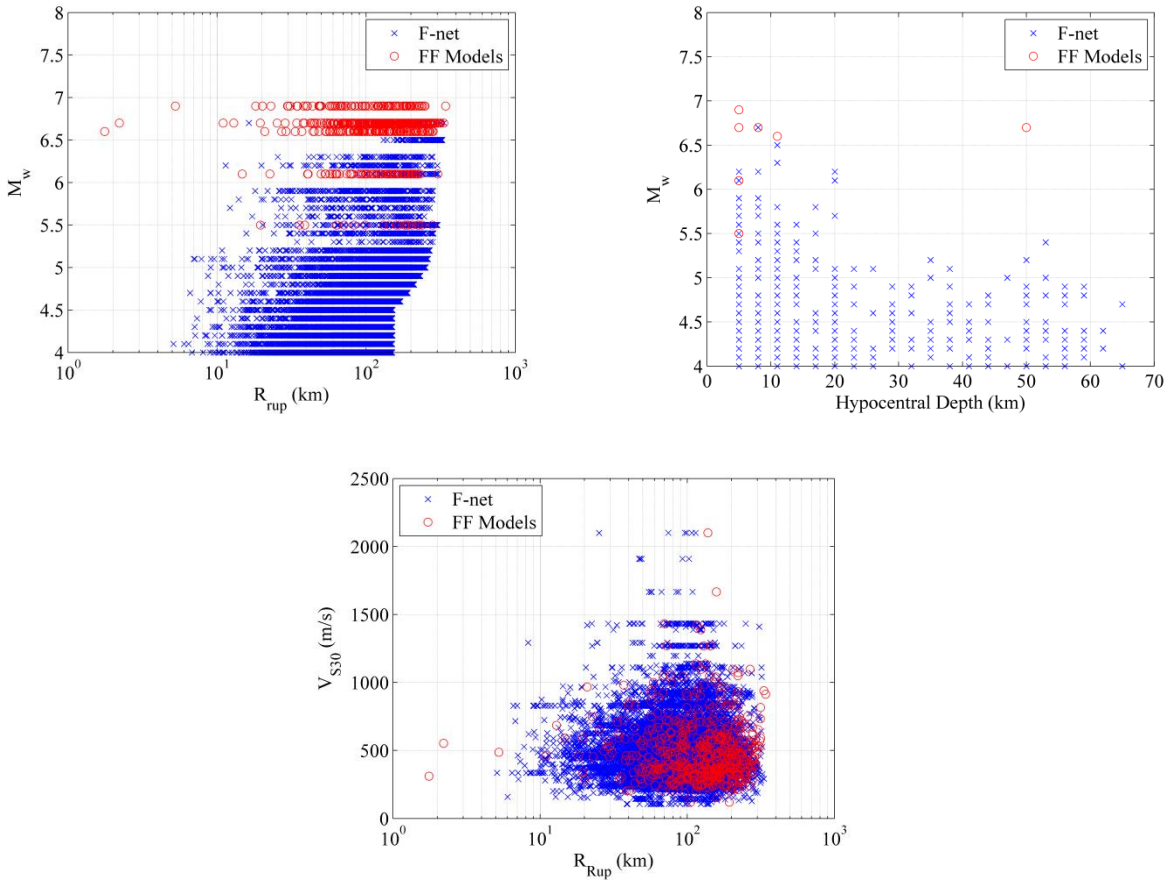


Figure 2. Scatter plots for the moment magnitude (M_w) versus the closest distances to the ruptured fault (R_{rup}; top left panel), M_w versus hypocentral depth (top right panel), and R_{rup} versus the average shear wave velocity of the top 30m at the recording station (V_{s30}; bottom panel) for the 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 Dawood et al. 2014 for additional details)

The database used in this study is described in detail in Dawood et al. (2014); however, only a subset of these records was used. This subset was created by eliminating records using the following criteria:

- Remove records from earthquakes with $M_w \leq 4.0$.
- Remove records from earthquakes with questionable metadata.
- Remove records from earthquakes that are not classified as *active crustal* according to the algorithm by Garcia et al. (2012).
- Remove records for which the signal-to-noise ratio drops below 3 in the frequency range between $2f_c$ (where f_c is the corner frequency of the acausal filter used for ground motion processing) and 25Hz in any of the four horizontal components.
- Remove records with R_{rup} greater than 350 km.
- For earthquakes with $M_w<5.0$, remove records that were found to contain multiple wave trains (Dawood 2014).
- Remove records recorded at stations without a measured shear wave velocity profiles.
- Remove records 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 as subduction earthquakes by Park and Mori (2005). For more details about these criteria the reader is referred to Dawood (2014).

The maximum usable spectral period for each record was set at half of the inverse of the corner frequency of the filter used in the processing protocol. This choice was made by comparing records processed with the automated algorithm with records processed by the NGA West 2 database (Ancheta et al. 2013; for details about the comparison see Dawood 2014). Hence, the number of usable records is reduced as the spectral period of interest increases, as shown in Figure 3. Note the large drop in the number of earthquakes and records used in the analysis for spectral periods larger than 1.0 s.

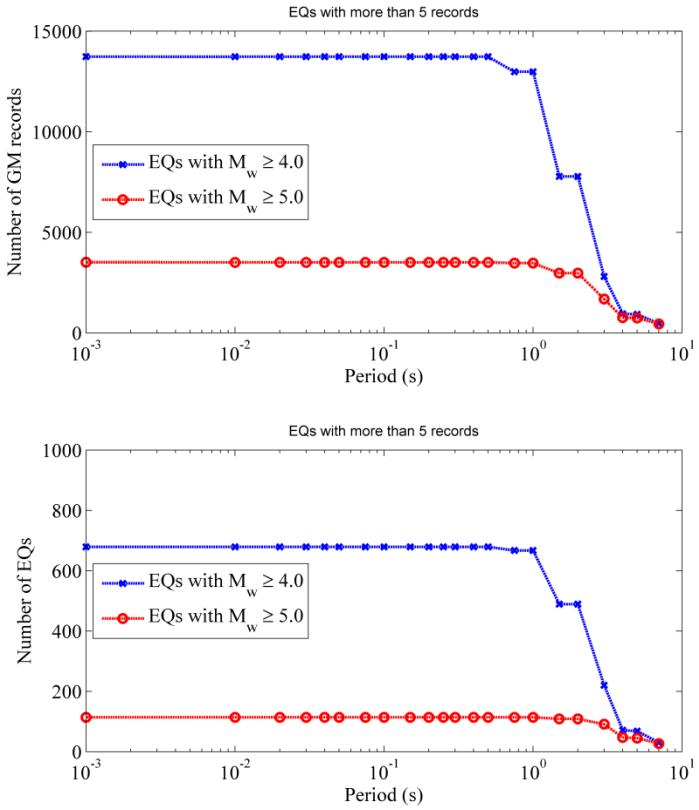


Figure 3. The number of motions (top panel) and earthquakes (bottom panel) used in the regression analysis at each spectral period. Two different lines are plotted in each panel to show the number of earthquakes and motions from all earthquakes (blue lines) and from earthquakes with $M_w \geq 5$ (red lines)

FORMULATION OF THE GMPES

The traditional formulation of a GMPE (heretofore referred as an ergodic formulation) is:

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

where the subscripts *e* and *s* denote earthquake and station, respectively and the superscripts *G* and *B* denote motions recorded at the ground surface or at the borehole instrument (within motion), respectively. $ln\ (y_{es}^{G/B})$ are the natural logarithm of an observed ground motion

parameter (5% damped pseudo-spectral acceleration) at ground surface and borehole; $ln(\mu_{es}^{G/B})$ are the median prediction; δB_e is the event-term (also referred to as between-event or inter-event residual); and $\delta W_{es}^{G/B}$ is the within-event (or intra-event) residual. The event-term (δB_e) represents the average shift from the median for records from earthquake *e*; since this is an event-specific residual component we assume that it is the same for surface and borehole records. Assuming that the residual components are uncorrelated, the total standard deviation at the ground surface and the borehole instruments are defined as:

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

where τ and $\varphi^{G/B}$ are the standard deviations of δB_e and $\delta W_{es}^{G/B}$, respectively. Alternatively, in a site-specific (partially non-ergodic) formulation the within-event residual ($\delta W_{es}^{G/B}$) is divided into two components:

$$\delta W_{es}^{G/B} = \delta S2S_s^{G/B} + \delta W_{es,o}^{G/B} \tag{3}$$

where $\delta S2S_s^{G/B}$ are the site-terms for station s for the borehole and the surface, depending on the superscript (also referred to as the site-to-site residual component) and $\delta W_{es,o}^{G/B}$ represent the remaining residual components after accounting for the site-terms. The site terms represent the average shift from the median for records from the specific site; hence, they capture repeatable effects at the site. Developing a regression that accounts for this partition of residuals requires a database that contains multiple strong motions recorded at the same station. Assuming that the residual components are uncorrelated, the total standard deviations are defined as:

$$\sigma_{tot}^{G/B} = \sqrt{\tau^2 + \left(\varphi_{S2S}^{G/B}\right)^2 + \left(\varphi_{ss}^{G/B}\right)^2} \tag{4}$$

where $\varphi_{S2S}^{G/B}$ are the standard deviations of the $\delta S2S_s^{G/B}$ terms and $\varphi_{ss}^{G/B}$ are the standard deviations of the $\delta W_{es,o}^{G/B}$ residual components.

FUNCTIONAL FORM OF THE MEDIAN MODEL

We adopted the functional form used by Abrahamson et al. (2013), with some modifications, for the median model. The availability of borehole records for the KiK-net database allows for the simultaneous regression for surface and borehole records, which was not possible for the Abrahamson et al. (2013; ASK13) study, which used the NGA W2 database (Ancheta et al. 2013).

The functional form for the median is given by:

$$ln(\mu_{es}^{G/B}) = F_{Surf}\, a1_S + (1-F_{Surf})\, a1_B + f_1(M_w, R_{rup}) + F_N\, f_8(M_w) + f_5^{G/B}(Sa_{1100}, V_{s30}) + f_6(Z_{TOR}) + f_{10}^{G/B}(Z_1, V_{s30}) + f_{11}^{G/B}(V_{s,Borehole}) \quad (5)$$

where $ln(\mu_{es}^{G/B})$ are the natural logarithm of the predicted pseudo spectral acceleration in units of gravity (given as a ratio of 9.81m/s^2) at the ground surface and borehole level; a_i are the model coefficients; and f_1, f_8, $f_5^{G/B}$, f_6, and $f_{10}^{G/B}$ are functions discussed in the following paragraphs. The other parameters (M_w, R_{rup}, V_{s30}, Z_1, Z_{TOR}, F_N, F_{Surf}, $V_{s,Borehole}$ and Sa_{1100}) are explained in Table 1.

Table 1. 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 Dawood et al. (2014) 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 V_s=1.0km/s measured in km
Z_{TOR}	Depth to the top of rupture in km
F_N	Flag for normal faulting events (using the normal faulting definition from the 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.
$V_{s,Borehole}$	The shear wave velocity of the layer where the borehole instrument is installed
Sa_{1100}	Median pseudo-spectral acceleration for Vs_{30}=1100m/s

The function f_1 is the magnitude-dependent attenuation term, which was assumed to be the same for surface and borehole records. Abrahamson et al. (2013) adopted a three segment model for this term. However, for the more limited range of magnitudes of the database used in this study (M_w 6.9) a two-segment model is sufficient:

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

$$\begin{cases} a5\,(M_w - M_1) + a8\,(8.5 - M_w)^2 + [a2 + a3(M_w - M_1)]ln(R) + a17R_{rup} & for\ M_w \geq M_1 \\ a6\,(M_w - M_2) + a8\,(8.5 - M_2)^2 + [a2 + a3(M_2 - M_1)]ln(R) + a17R_{rup} & for\ M_w < M_1 \end{cases} \quad (6)$$

where M_1 is the boundary between both terms and is set at 5.0; R is given by:

$$R = \sqrt{R_{rup}^2 + c_{4M}^2} \quad (7)$$

and c_{4M} is given by:

$$c_{4M} = \begin{cases} c4 & for\ 5 < M_w \\ c4 - (c4 - 1)(5 - M_w) & for\ 4 < M_w \leq 5 \\ 1 & for\ M_w \leq 4 \end{cases} \quad (8)$$

The function f_8 is the *style of faulting* term. An analysis of residuals indicated no difference between records from reverse and strike slip earthquakes; hence, this term includes a correction only for normal faulting:

$$f_8(\mathrm{M_w}) = \begin{cases} a12 & for\ M_w > 5 \\ a12(M_w - 4) & for\ 4 \leq M_w \leq 5 \\ 0 & for\ M_w < 4 \end{cases} \quad (9)$$

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 The data used by ASK13 is sparse beyond 20km; however, 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:

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

To capture site response, we use both a V_{s30} scaling term (f_5) and a term that scales with Z_1 (f_{10}), where Z_1 is the depth to V_s=1.0 km/s. For the V_{s30} scaling term, the model used in Kamai et al. (2013) is used with different linear site response coefficients for surface and borehole records (i.e., $a10_S$ and $a10_B$). The same non-linear decay is used for surface and borehole. Hence, the term f_5 is given by:

$f_5{}^{G/B}(Sa_{1100}, V_{s30}) =$

$$\begin{cases} F_{Surf}\left[a10_S\, ln\left(\frac{V_{S30}^*}{V_{Lin}}\right) - b\, ln\left(Sa_{1100,Surf} + c\right)\right] + \\ \left(1 - F_{Surf}\right)\left[a10_B\, ln\left(\frac{V_{S30}^*}{V_{Lin}}\right) - b\, ln\left(Sa_{1100,Bore} + c\right)\right] + \\ b\, ln\left(F_{Surf} Sa_{1100,Surf} + \left(1 - F_{Surf}\right) Sa_{1100,Bore} + c\left(\frac{V_{S30}^*}{V_{Lin}}\right)^n\right) & for\ V_{S30} < V_{Lin} \\ \left[F_{Surf}\, a10_S + \left(1 - F_{Surf}\right) a10_B + bn\right] ln\left(\frac{V_{S30}^*}{V_{Lin}}\right) & for\ V_{S30} \geq V_{Lin} \end{cases} \quad (11)$$

where:

$$V_{S30}^* = \begin{cases} V_{S30} & for\ V_{S30} < V_1 \\ V_1 & for\ V_{S30} \geq V_1 \end{cases} \quad (12)$$

and V_{Lin} and V_1 are coefficients adopted from Kamai et al. (2013)

We used the same model adopted by ASK13 for the f_{10} term, with the exception that we regressed for different coefficients for surface and borehole records:

$f_{10}{}^{G/B}(Z_{1.0}, V_{s30})$

$$\begin{cases} F_{Surf}\, a43_S\, ln\left(\frac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) + \left(1 - F_{Surf}\right) a43_B\, ln\left(\frac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) & for \quad V_{S30} \leq 200m/s \\ F_{Surf}\, a44_S\, ln\left(\frac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) + \left(1 - F_{Surf}\right) a44_B\, ln\left(\frac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) & for\ 200 < V_{s30} \leq 300m/s \\ F_{Surf}\, a45_S\, ln\left(\frac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) + \left(1 - F_{Surf}\right) a45_B\, ln\left(\frac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) & for\ 300 < V_{s30} \leq 500m/s \\ F_{Surf}\, a46_S\, ln\left(\frac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) + \left(1 - F_{Surf}\right) a46_B\, ln\left(\frac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) & for\ 500 < V_{s30} \end{cases} \quad (13)$$

where $Z_{1,ref}$ is the reference depth to a shear wave velocity of 1 km/s and is calculated as a function of V_{s30} as proposed by Chiou and Youngs (2013) for Japan. The adopted functional form is as follows:

$$Z_{1,ref}\ \frac{1}{1000}\ exp\left(\frac{-5.23}{2} ln\left(\frac{V_{S30}^2 + 412^2}{1360^2 + 412^2}\right)\right) \quad (14)$$

Z_1 was assumed to equal $Z_{1,ref}$ at stations where the shear wave velocity does not exceed 1000 m/s within the measured shear wave velocity profiles.

Preliminary residual analysis showed that the residuals of the borehole records are biased with respect to $V_{s,Borehole}$. Hence, an additional term is added to the GMPE functional form to remove the bias in the residuals:

$$f_{11}{}^{G/B}(V_{s,Borehole}) = (1 - F_{Surf})a_{Borehole}\, ln\left(\frac{V_{S,Borehole}}{1600}\right) \quad (15)$$

Hanging wall effects on the median predictions were not considered due to the paucity of records located in the hanging wall.

REGRESSION ANALYSES

We conduct two sets of regression analyses to obtain an ergodic and a site-specific GMPE. The regression analyses were conducted in 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 analyses were conducted using a combination of linear and nonlinear mixed effects regression analyses incorporated in the commercial statistical software SAS 9.3 (SAS Institute, Cary, NC, USA).

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 records 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 5). 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). However, regression analyses indicated that the coefficient associated with the aftershock flag is not significantly different than zero. Hence, the proposed model (Equation 5) does not include an aftershock flag.

Some of the coefficients in Equations 10 and 11 (i.e., V_{LIN}, b, c, and n) were adopted directly from Kamai et al. (2013). V_1 was reported in Abrahamson and Silva (2008) and adopted by AS13. All other coefficients were obtained through the regression analyses. All of the coefficients are given in Tables 2 through 5 for the ergodic and site-specific GMPEs. 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 Appendix D of this dissertation

Following the recommendation given by Abrahamson et al. (2014; ASK-2014), for forward predictions we recommend assuming that the values obtained from the different branches for f_{10} (Equation 13) are at the center of each bin and then linearly interpolate to get the values of this coefficient at the V_{s30} value or interest.

Table 2. 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 3. 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 4a. 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 4b. 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 5a. 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 5b. 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

COMPARISON OF ERGODIC AND SITE-SPECIFIC REGRESSIONS

In this section, the models developed using the ergodic and non-ergodic formulations are evaluated and compared. As discussed previously, the difference between these formulations lies in representing the site term ($\delta S2S_S^{G/B}$) as a zero-mean normal random variable in the site-specific formulation (Equation 3). Including this normal random variable should –theoretically– have no effect on the resulting median prediction if the database on hand is well-balanced.

However, this was not the case for the models obtained in this study. Stafford (2014) had similar results for analyses of the NGA-W1 database (Chiou et al. 2008).

There are two potential sources for this discrepancy. The first source is artificial and results from smoothing the regression coefficients across the different spectral periods. Some coefficients can change considerably in the smoothing process, which might affect the predictions obtained from both formulations if the smoothing process is not conducted in a consistent way. Hence, the smoothing process should be handled with care. Discrepancies can also result from the unbalance in the strong motion database used in the regression. Such heterogeneities lead to different regression models from both formulations even before starting the smoothing process. Hence, to evaluate the differences in ergodic and non-ergodic regressions we compare the median predictions before any smoothing. Hence, we use the regression coefficients obtained from one of the first steps adopted to conduct the regression analysis (Dawood 2014).

To compare the ergodic and site-specific GMPEs we generated two sets of 1280 response spectra assuming different combinations of input parameters (i.e. M_w =5, 5.5, 6, and 6.5; R_{rup} =20, 50, 100, 150, 200, 250, 300, and 350 km; V_{s30} =200, 300, 500, and 700 m/s); and 0.3 km; Z_{TOR} =0, 10, 20, 30, and 50 km; and F_{nm} =0 and 1). The average and average plus/minus standard deviation of the difference between the natural log of the predicted PSA using both GMPEs were calculated at all spectral periods (Figure 4). The average difference (on a natural log scale) at a specific spectral period ranged between 0.0642 (at T=0.03s, PSA from ergodic is 6.6% higher) and -0.07799 (at T=2.0s, PSA from ergodic is 7.5% lower) for surface motions and 0.05207 (at T=7.0s, PSA from ergodic is 5.3% higher) and -0.04675 (at T=2.0s, PSA from ergodic is 4.6% lower) for borehole records. Even if the average difference does not seem significant, it is important to note that the standard deviations associated with these averages are quite large (Figure 4). The database used in this study is quite large; if this exercise was conducted on a smaller database, the discrepancies could potentially be more significant. Moreover, the discrepancies shown in Figure 4 are purely from the formulation used in the regression analysis, if the effect of smoothing is added the differences can be substantial.

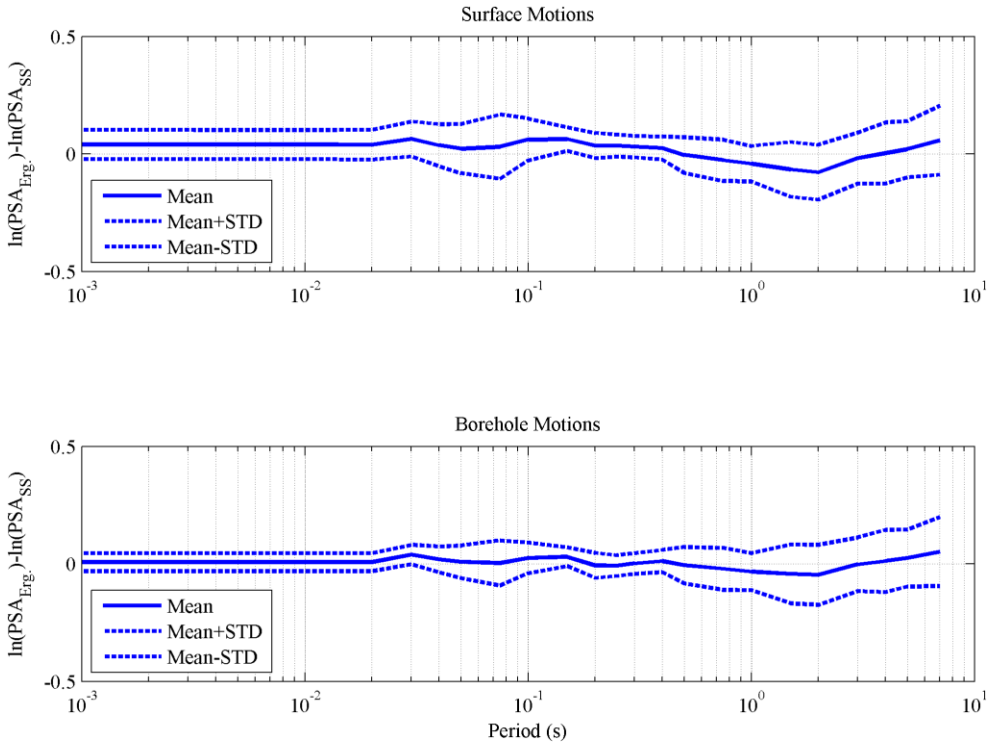


Figure 4. The mean, and mean±standard deviation of the difference in the predicted ln(PSA) using the ergodic (Erg.) and site-specific (SS) GMPEs across the different spectral periods for surface motions (top panel) and borehole motions (bottom panel)

We also estimated and compared the variability components of the two regressions. In order to compare the within event component of variability, we calculated $\varphi^{G/B}$ from the site-specific formulation as follows:

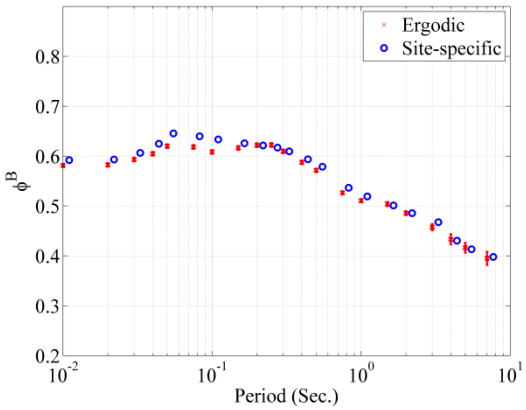
$$\varphi^{G/B} = \sqrt{\left(\varphi_{S2S}^{G/B}\right)^2 + \left(\varphi_{ss}^{G/B}\right)^2} \tag{16}$$

The estimated values of within- and between event variability ($\varphi^{G/B}$ and τ, respectively) obtained from the ergodic and site-specific formulations are very similar (Figure 5). The intra-event component of variability at ground surface (φ^{G}) has a peak at around 0.15 seconds. This

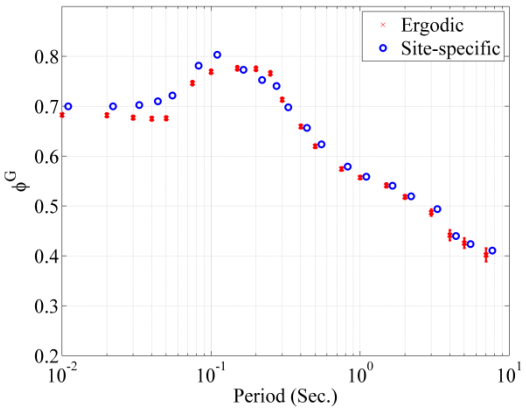
peak is less obvious in the borehole component (φ^B), suggesting that the increased variability is due to near surface effects. Also observe that the peak in the φ_{ss}^G plot is less prominent than the one in the φ^Gplot. Since φ_{ss} does not incorporate site-to-site variability, it can be inferred that the observed peak in φ^G is due to site-to-site variability of the near-surface deposits that is not captured by the site effects parameterization (e.g., V_{s30} and Z_1). Moreover, φ_{SS}^B shows a peak around 0.06 seconds, which is at smaller period compared to the peak observed at φ_{SS}^G and φ^G.

The site corrected within event component of variability ($\varphi_{ss}^{G/B}$) obtained from the site-specific formulation is, as expected, considerably lower than $\varphi^{G/B}$ from the ergodic formulation (Figure 5). Figure 5 shows a comparison between $\varphi_{ss}^{G/B}$ obtained directly from the site-specific formulation and the $\varphi_{ss}^{G/B}$ obtained from the residuals of the ergodic formulation. The $\varphi_{ss}^{G/B}$ estimates from both approaches are very close, but there are some differences at some periods.

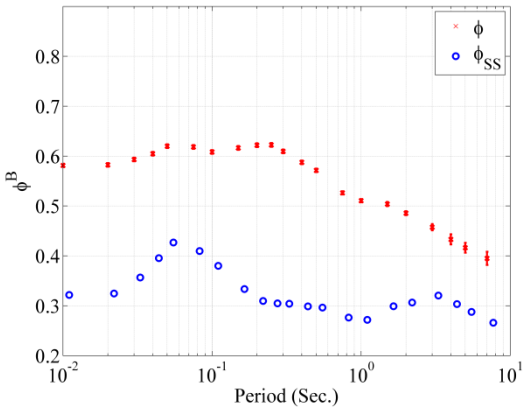
We also explore whether the event terms (δB_e) obtained from both formulations are similar. For this purpose, we compute the correlation coefficient between δB_e obtained from both formulations for selected periods. This correlation is above 0.92 for PGA, 0.2s, 1.0s, and 7.0s. Recall that only earthquakes recorded at 5 or more stations are used, this explains in part why δB_e is stable using both approaches.


Ergodic
Site-specific
ϕ^B
Period (Sec.)

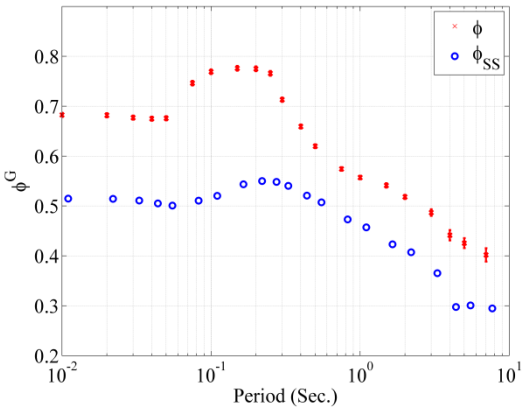
(a)


Ergodic
Site-specific
ϕ^G
Period (Sec.)

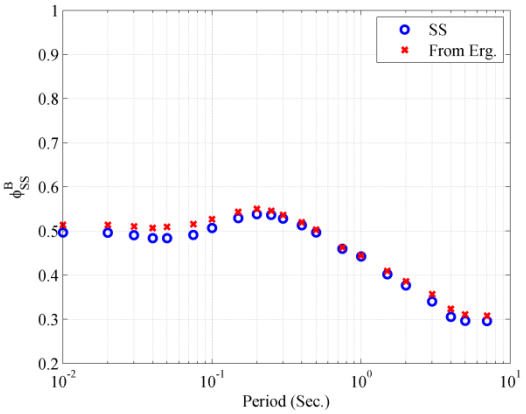
(b)


ϕ
ϕ_{SS}
ϕ^B
Period (Sec.)

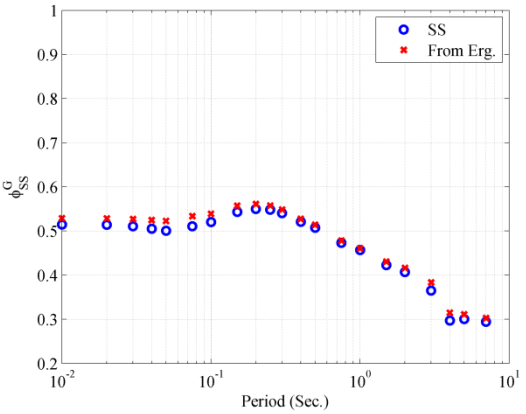
(c)


ϕ
ϕ_{SS}
ϕ^G
Period (Sec.)

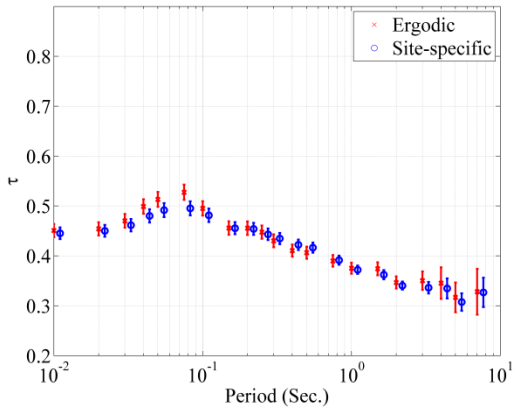
(d)


SS
From Erg.
ϕ^B_{SS}
Period (Sec.)

(e)


SS
From Erg.
ϕ^G_{SS}
Period (Sec.)

(f)



(g)

Figure 5. The comparison between the different components of variability: a) φ^{B} obtained from the ergodic and site-specific GMPEs; b) φ^{G} obtained from the ergodic and site-specific GMPEs; c) φ^{B} obtained from the ergodic GMPE with φ_{ss}^{B}(Equations 2 and 4); d) φ^{G} obtained from the ergodic GMPE with φ_{ss}^{G}(Equations 2 and 4); e) φ_{ss}^{B} obtained directly from the site-specific formulation and φ_{ss}^{B} obtained from the residuals of the ergodic formulation; f) φ_{ss}^{G} obtained directly from the site-specific formulation and φ_{ss}^{G} obtained from the residuals of the ergodic formulation; and g) τ obtained from the ergodic and site-specific formulations

RESIDUALS

A comprehensive set of plots for the different residual components as a function of various parameters are presented in Appendix D of this dissertation. Examples of these plots are presented in this manuscript along with brief discussions of the major observations. The figures presented in this section show the residual components as gray open circles. The residuals are binned with respect to the parameter shown in the horizontal axis. The mean and mean $\pm$ standard deviations of the residuals in each bin are also shown.

Figure 6 shows the inter-event residual component obtained from the Ergodic and the site-specific GMPEs at spectral periods of 0.2 s and 1.0 s plotted against M_w, Z_{TOR}, and rake. The inter-event residuals show some bias for earthquakes with M_w between 6.5 and 7.0 (Figure 6a). The limited number of large earthquakes in the current database could be the cause of this observed bias. The plots also show a positive bias for earthquakes with a depth to top of rupture

larger than 45km at 1.0s (Figure 6b). On the other hand, the residuals do not show any clear bias with rake (Figure 6c). Figures 6d, 6e and 6f differentiate between the residuals from main shocks and fore/aftershocks. These plots do not show any clear differences in the residuals from either class of earthquakes, which is consistent with regression results that indicated that the fore/aftershock coefficient was close to zero.

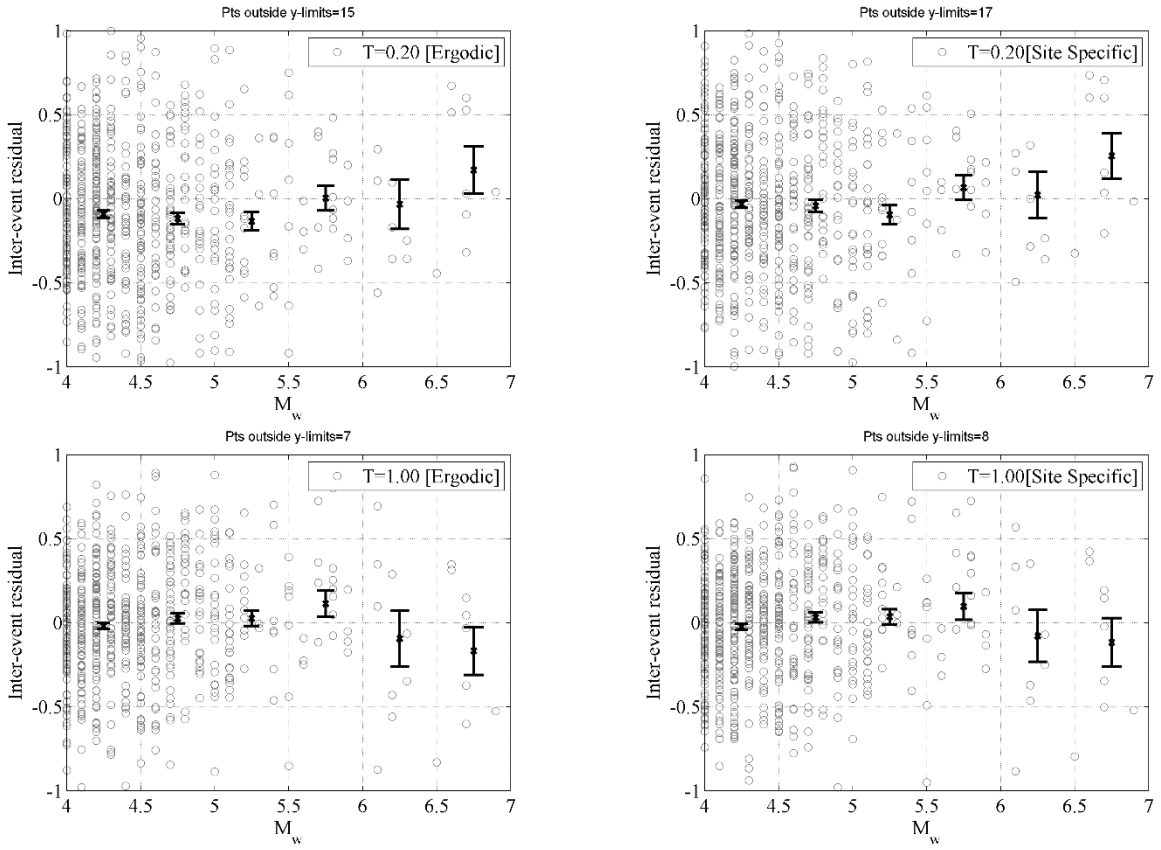


Figure 6a. Inter-event residual component obtained from the Ergodic GMPE (left column) and the site-specific GMPE (right column) at spectral periods of 0.2s and 1.0s plotted against M_w.

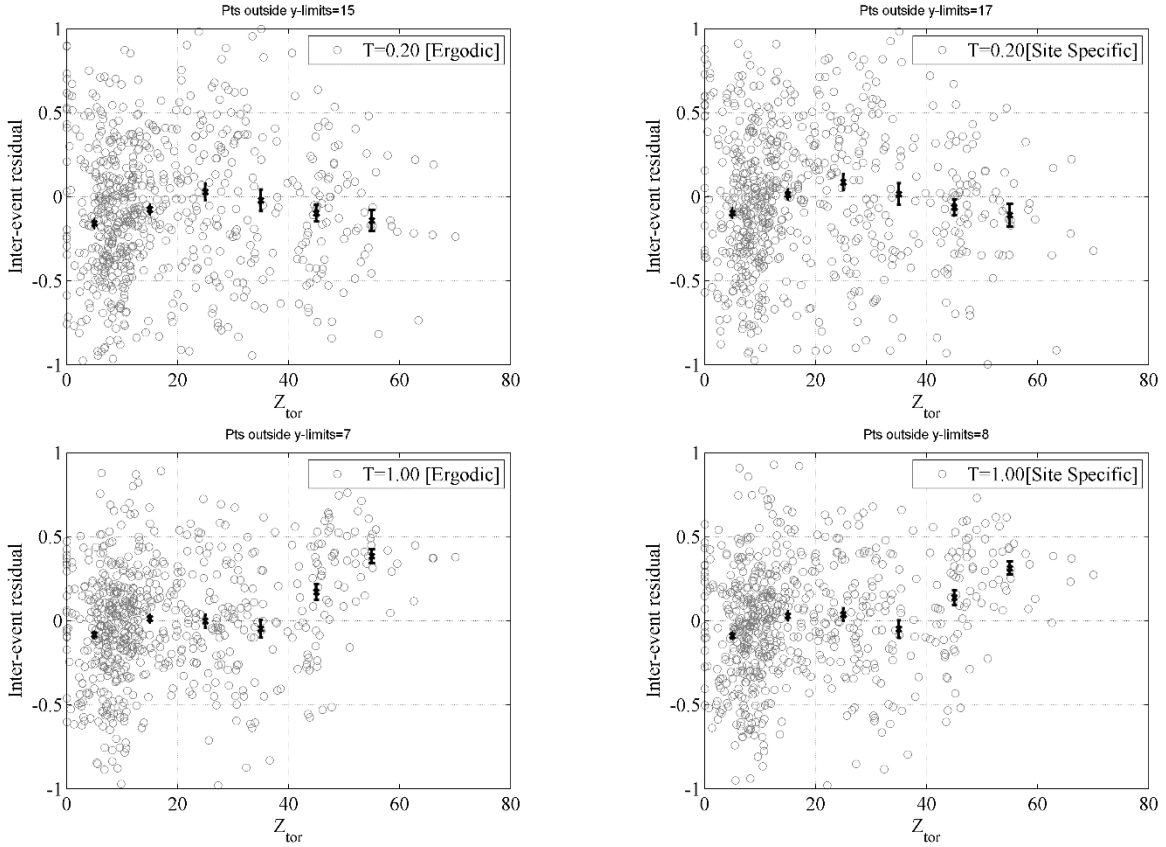
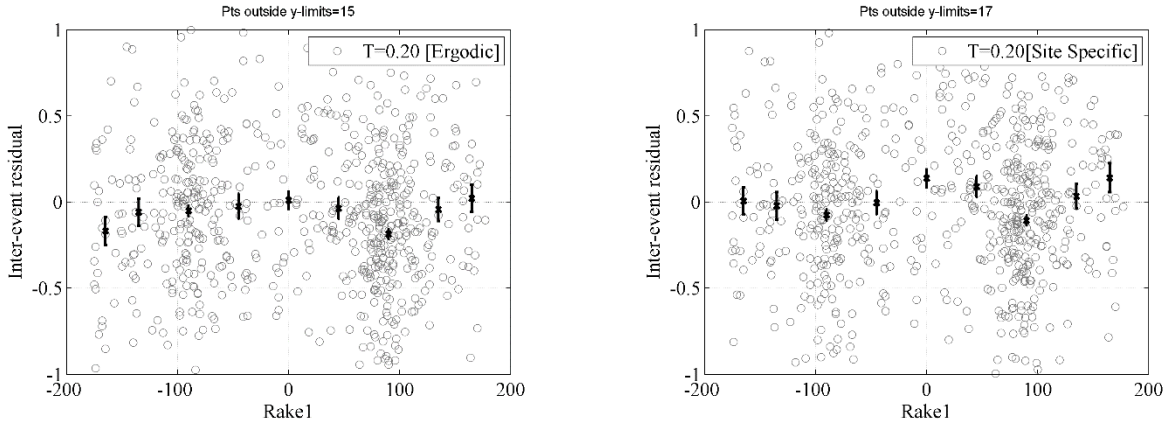


Figure 6b. Inter-event residual component obtained from the Ergodic GMPE (left column) and the site-specific GMPE (right column) at spectral periods of 0.2s and 1.0s plotted against Z_{TOR}.



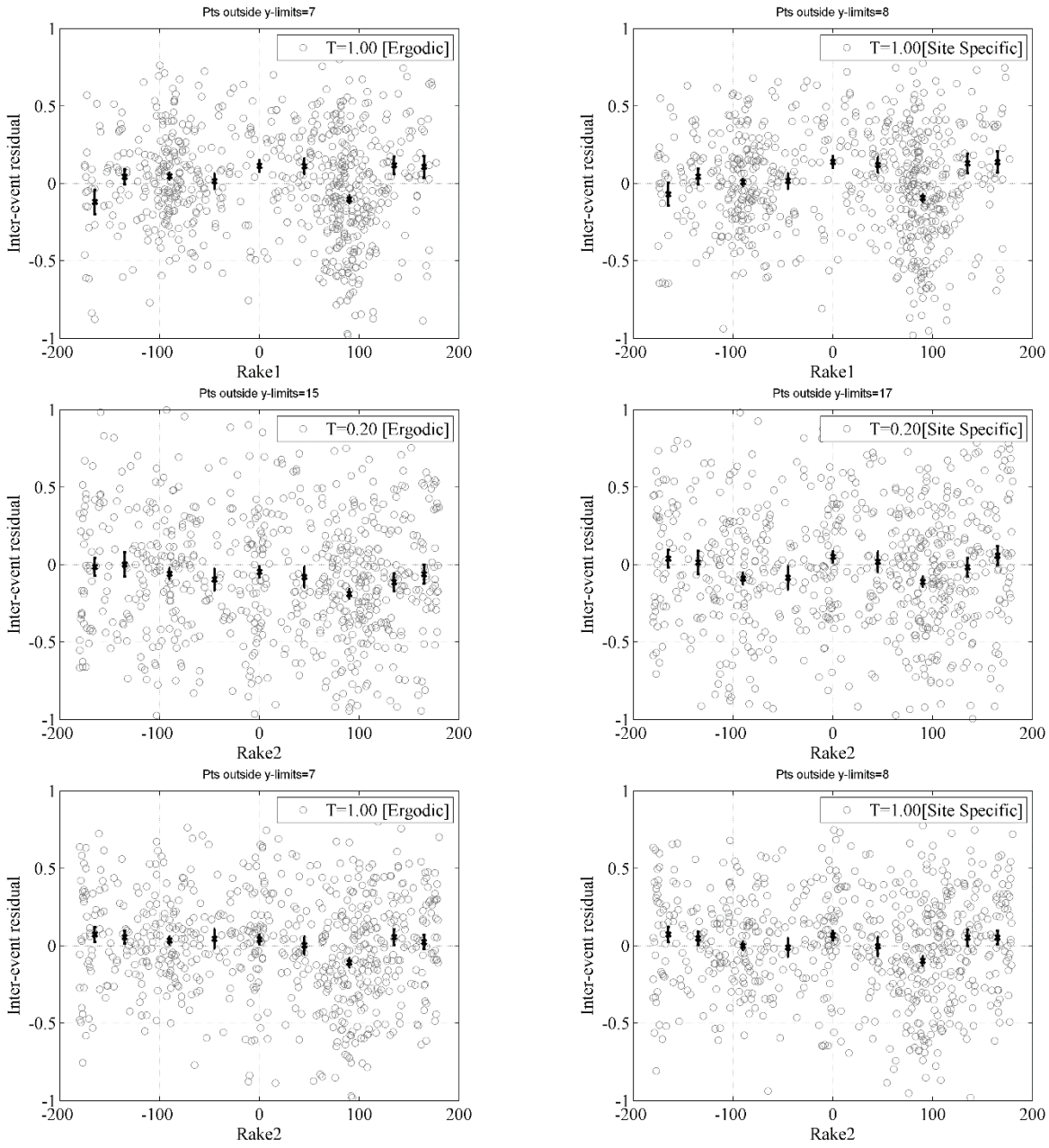


Figure 6c. Inter-event residual component obtained from the Ergodic GMPE (left column) and the site-specific GMPE (right column) at spectral periods of 0.2s and 1.0s plotted against rake.

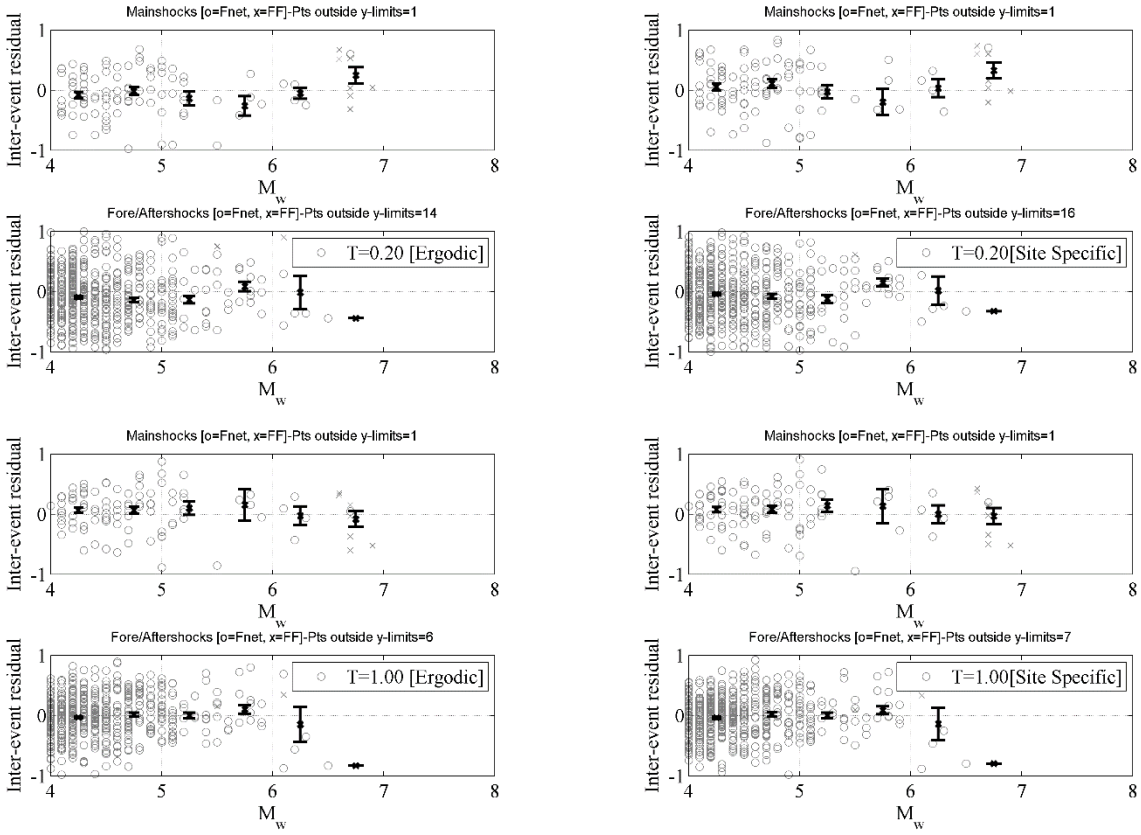


Figure 6d. Inter-event residual component obtained from the Ergodic GMPE (left column) and the site-specific GMPE (right column) at spectral periods of 0.2s and 1.0s plotted against M_w. The residuals from main shocks and fore/aftershocks are presented in different sub-plots.

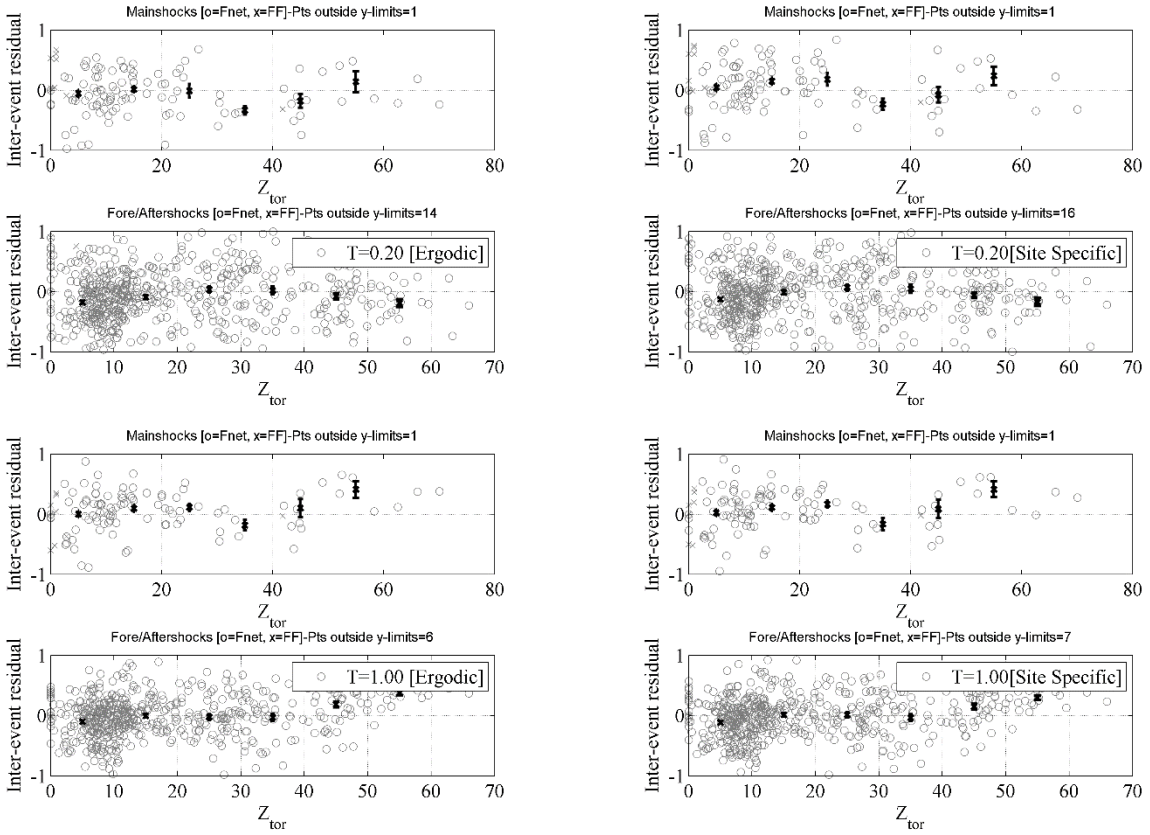
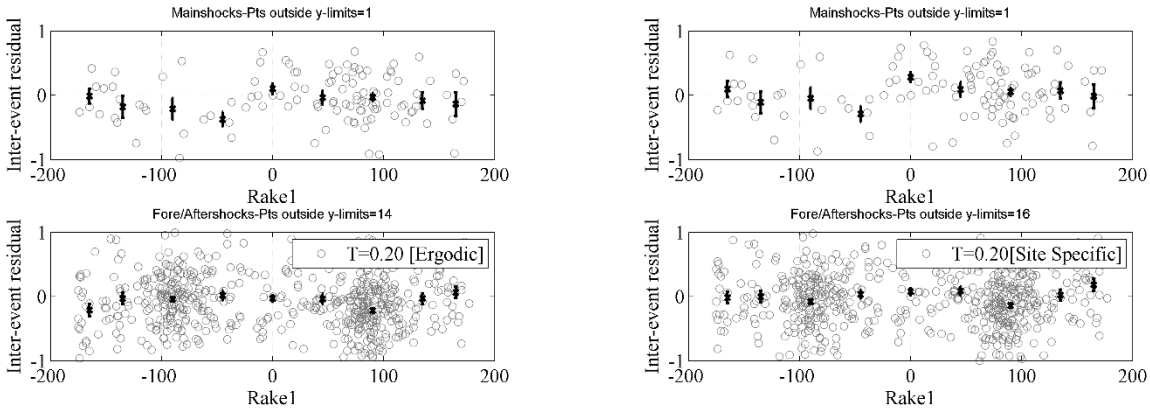


Figure 6e. Inter-event residual component obtained from the Ergodic GMPE (left column) and the site-specific GMPE (right column) at spectral periods of 0.2s and 1.0s plotted against Z_{TOR}. The residuals from main shocks and fore/aftershocks are presented in different sub-plots.



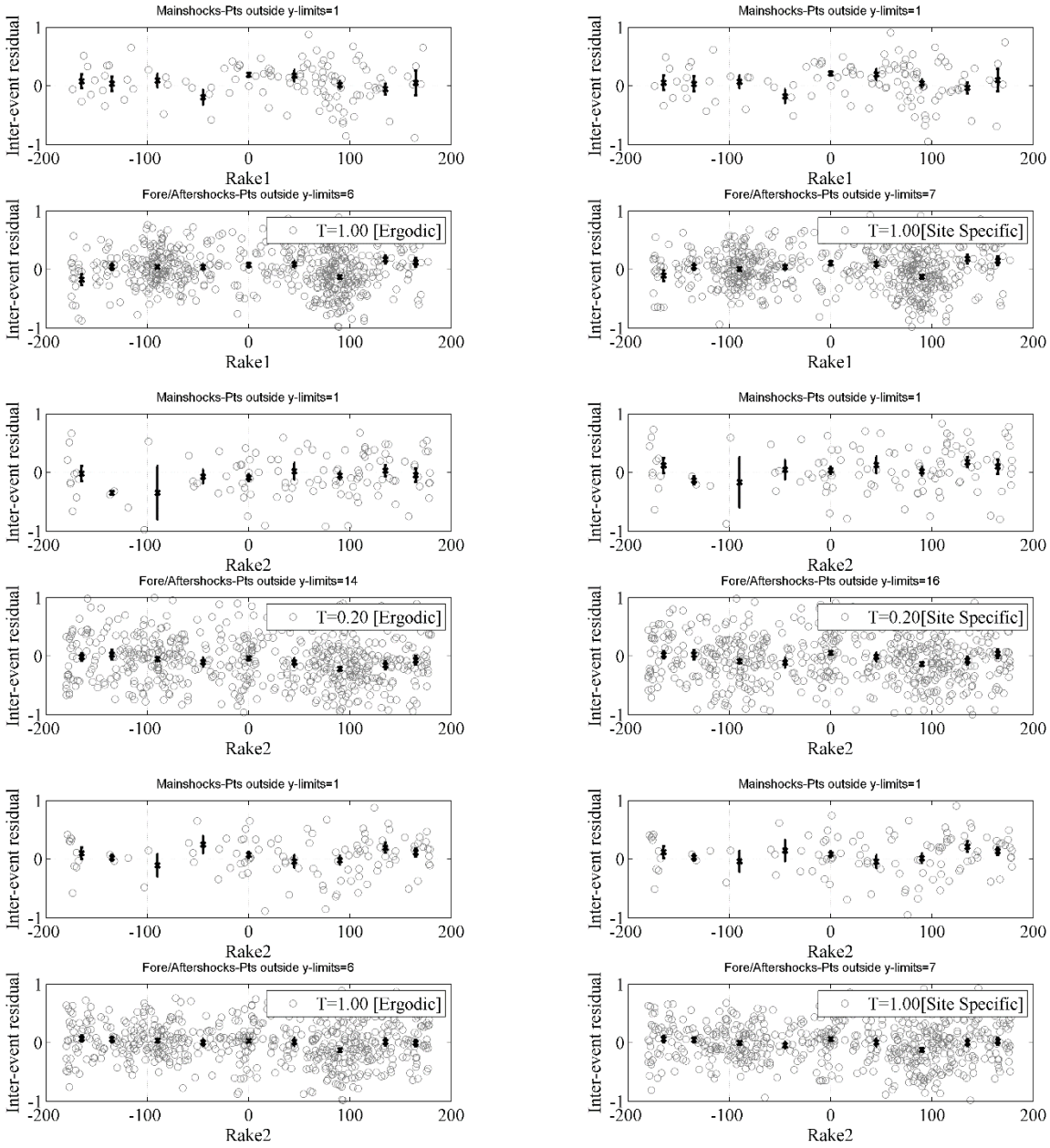


Figure 6f. Inter-event residual component obtained from the Ergodic GMPE (left column) and the site-specific GMPE (right column) at spectral periods of 0.2s and 1.0s plotted against rake. The residuals from main shocks and fore/aftershocks are presented in different sub-plots.

Figure 7 shows the intra-event residual component obtained from the ergodic GMPE and the intra-event site corrected residual component (δWS_{es}) from the site-specific GMPE at

spectral periods of 0.2s and 1.0s plotted against R_{rup} and V_{s30}. The residuals show some bias with R_{rup} at short and long distances (Figure 7a). For the ergodic GMPE, the average of the residuals from records recorded at soft (NEHRP Site Class E) and hard rock sites (NEHRP Site Class A) are biased (Figure 7b). This bias is expected since the site response model that we adopted in the proposed models (Kamai el al. 2013) was constrained using simulations for site with V_{s30} between 190 and 900 m/s. Hence, similar to the recommendation by the developers of the ASK-2014 model, we recommend using the proposed ergodic GMPE for sites with V_{s30} between 180 and 1000 m/s only. On the other hand, the residuals from the site-specific GMPE show weaker bias at soft and hard rock sites compared with the ergodic GMPE. This is expected since the site-specific residual component partially corrected this bias at these stations (Figure 8).

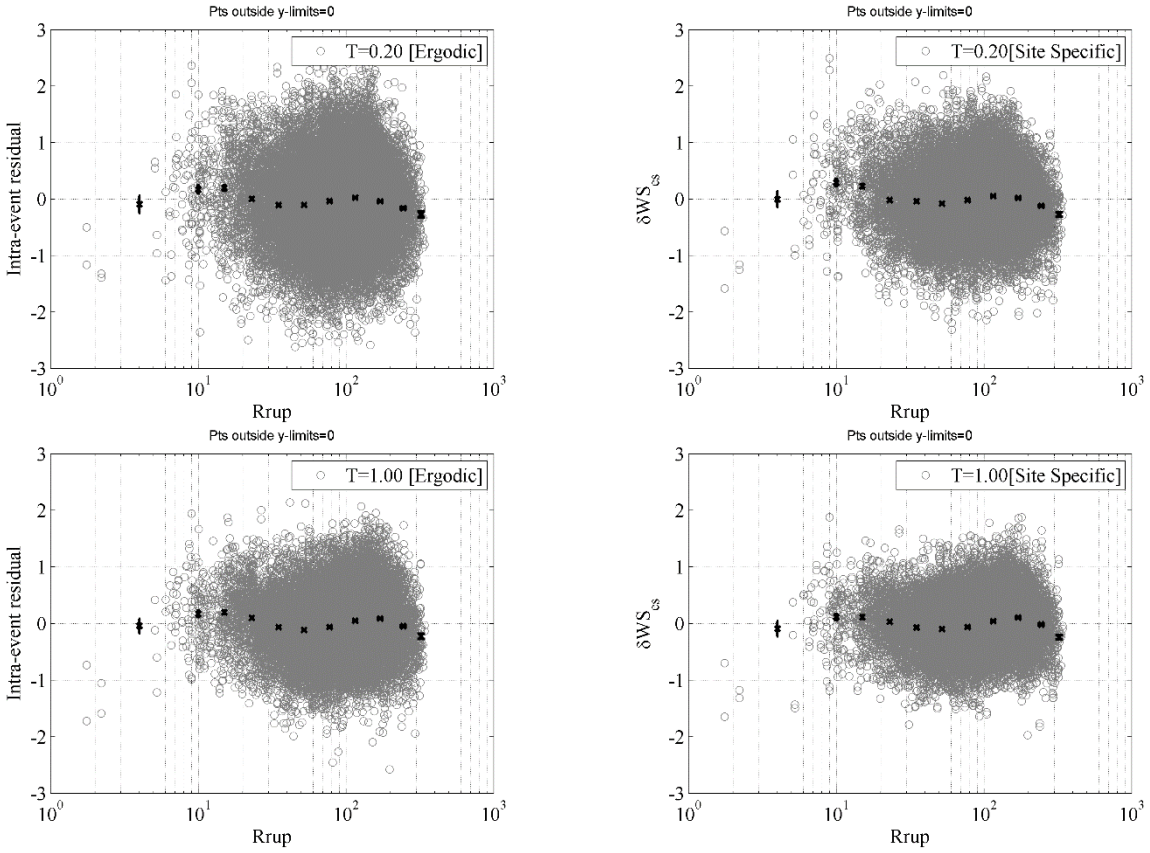


Figure 7a. Intra-event residual component obtained from the Ergodic GMPE (left column) and the intra-event site corrected residual component (δWS_{es}) from the site-specific GMPE (right column) at spectral periods of 0.2s and 1.0s plotted against R_{rup}.

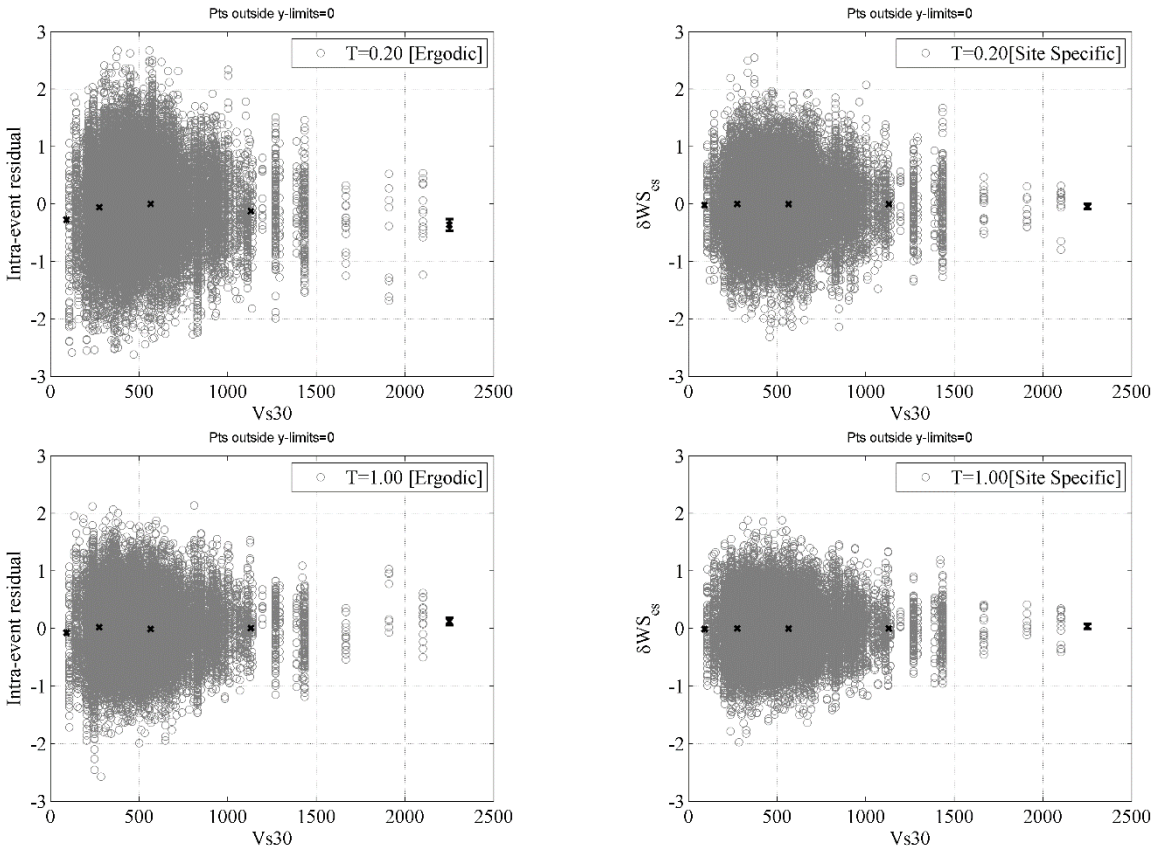


Figure 7b. Intra-event residual component obtained from the Ergodic GMPE (left column) and the intra-event site corrected residual component (δWS_{es}) from the site-specific GMPE (right column) at spectral periods of 0.2s and 1.0s plotted against V_{s30}.

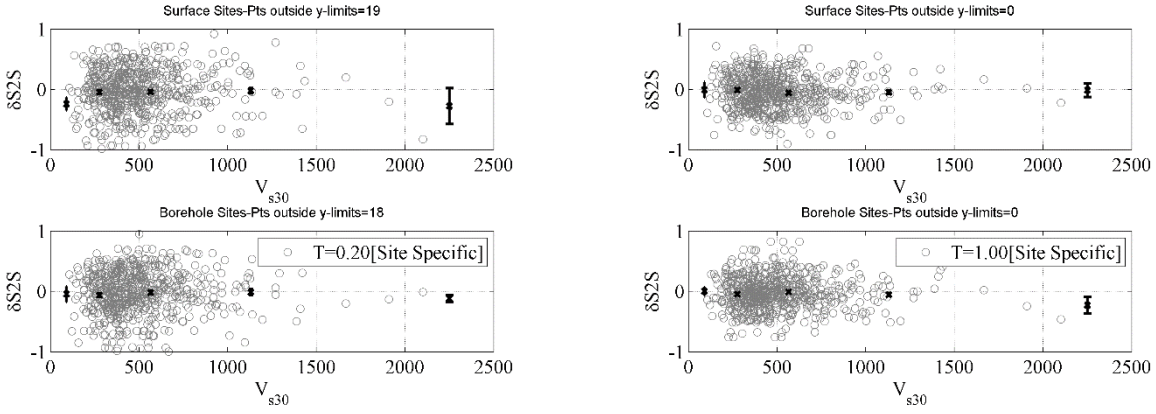


Figure 8. The site-specific residual component at the ground surface and borehole obtained from the site-specific GMPE at spectral periods of 0.2s and 1.0s plotted against V_{s30}

COMPARISONS WITH OTHER GMPES

In this section we present a comparison between the GMPEs developed in this study and two sets of GMPEs. The first set was developed mainly for Japan (Zhao et al. 2006, which will be referred to as Zea-2006; Kanno et al. 2006, Kea-2006; and Rodriguez-Marek et al. 2011, R-M-2011) and the second set was developed mainly for western United States (NGA models by Abrahamson and Silva 2008, AS-2008, and ASK-2014). We conducted the comparison between spectral accelerations predicted for R_{rup} values of 30 and 100km and for V_{s30} of 270 and 760 m/s. The predictions of the proposed models were made for strike slip faults assuming that $Z_{1.0}$ equals 0.1 km, Z_{TOR} equals 2 km, and $V_{s,hole}$ equals twice the V_{s30} used. The other GMPEs used some additional input parameters that are not used in the proposed models. We assumed these additional parameters as follows: focal depth equals 1.25 of Z_{TOR}, depth to a shear wave velocity of 800 m/s equals 70 m, R_{JB} and R_x equals R_{rup}, and dip equals 90 degrees. We compared the proposed GMPEs with the models for Japan in terms of the median predictions of spectral accelerations (Figure 9), magnitude and distance scaling (Figures 10 and 11), while the comparison with the NGA models was in terms of median predictions only (Figure 12).

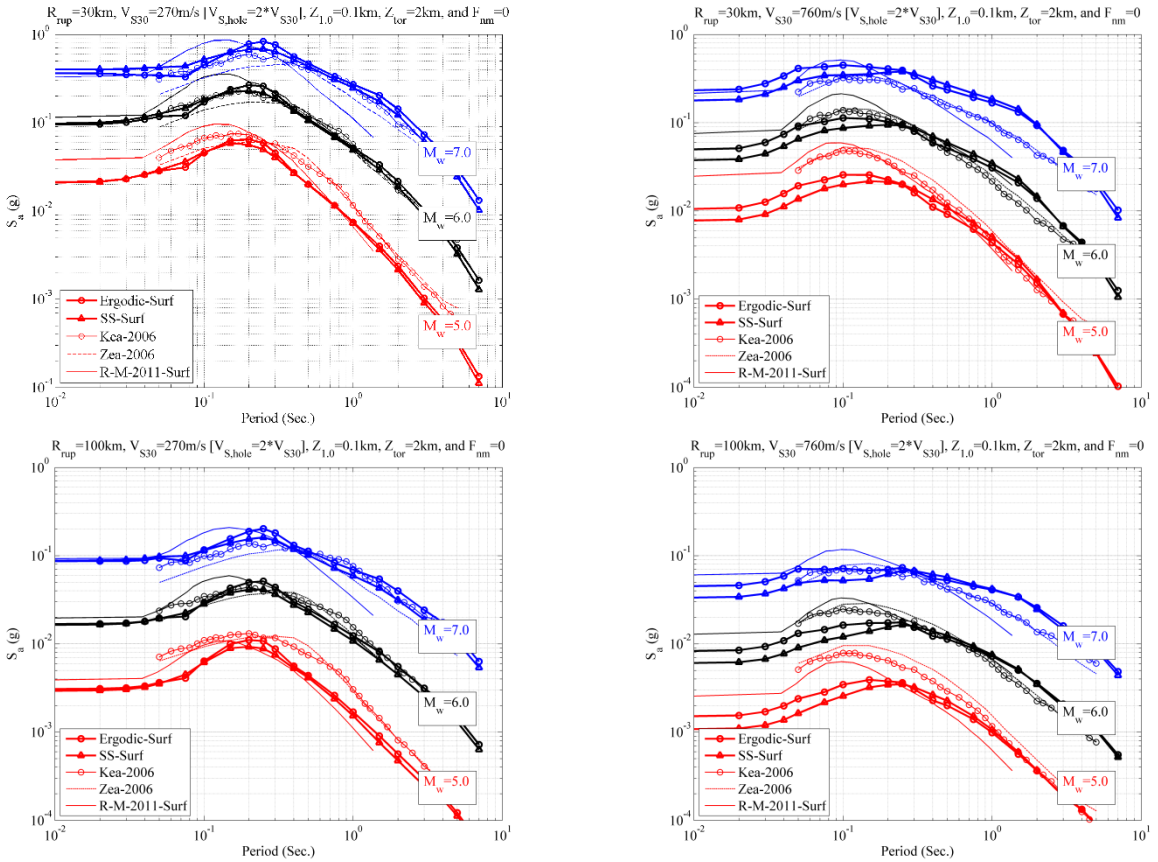
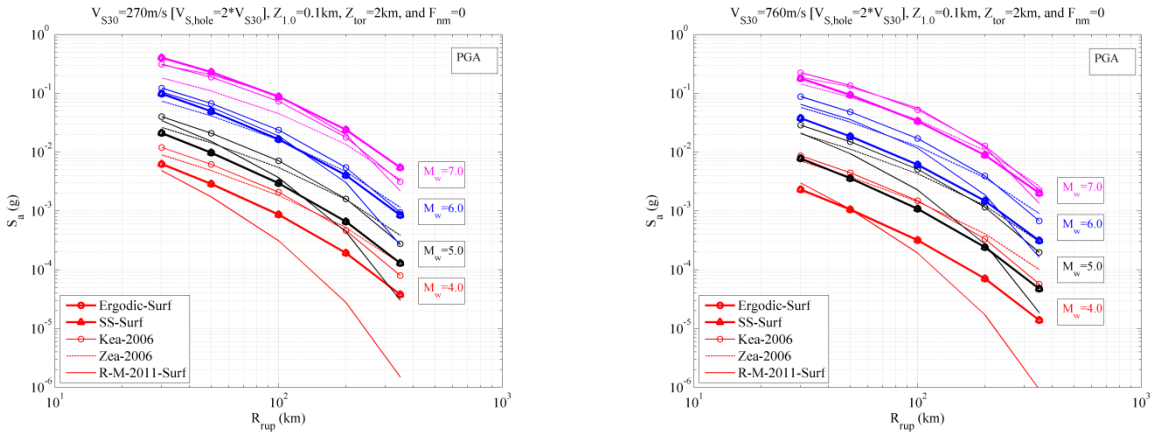


Figure 9. Comparison between the spectral accelerations obtained from the current study (Ergodic-Surf and SS-Surf) and the predictions from Zhao et al. (2006; Zea-2006), Kanno et al. (2006; Kea-2006) and Rodriguez-Marek et al. (2011; R-M-2011-Surf).



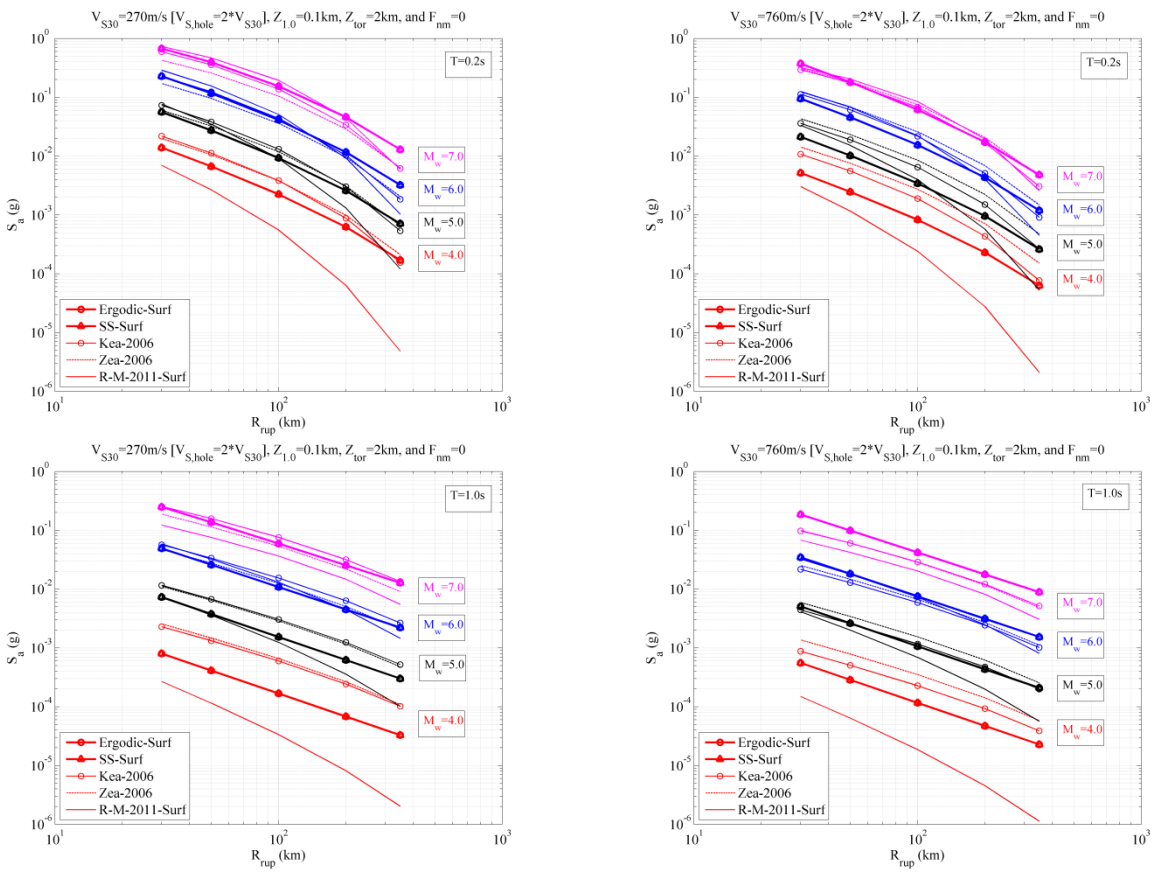


Figure 10. Comparison of the distance scaling obtained from the current study (Ergodic-Surf and SS-Surf) and the predictions from Zhao et al. (2006; Zea-2006), Kanno et al. (2006; Kea-2006) and Rodriguez-Marek et al. (2011; R-M-2011-Surf). V_{s30} equals 270m/s for the left panels and 760 for the right panels

For soft sites (e.g., V_{s30}=270m/s), the proposed models predict lower spectral accelerations from small magnitude earthquakes compared to the other models except for the model by R-M-2011 at longer periods (Figure 9). There is a better agreement between the proposed models and the models by Zea-2006 and Kea-2006 for larger magnitude earthquakes (i.e., 6 and 7), especially at long periods. The model by R-M-2011 predicts higher spectral accelerations at short periods for earthquakes larger than 6, and lower spectral accelerations at long periods for earthquakes larger than 7. For rock sites (i.e., V_{s30} =760m/s), the proposed models predicts higher spectral accelerations at long periods especially for large earthquakes. At short periods, the proposed models tend to predict lower values compared with the other models.

In general, the different models show similar attenuation rates with distance except for smaller magnitudes from the R-M-2011 model (Figure 10). This could be attributed to the distance censoring applied to the data used by R-M-2011 (about 50 and 100 km for M_w of 4 and 5, respectively). Hence, this model is not applicable beyond these limits. The models by Zea-2006 and Kea-2006 used GMs from earthquakes with M_w larger than 5.0 and 5.5, respectively. Hence, the observed differences between the predictions of these two models and the proposed models for small earthquakes are expected. Also, these two models show a higher decay with distance beyond 200km at short periods compared with the proposed models. All models predict lower spectral accelerations at long periods for rock sites compared to the proposed models (Figure 10).

The predictions by R-M-2011 show a non-linear decrease of spectral accelerations at small magnitudes, while the other models show an almost linear relationship (Figure 11). Hence, R-M-2011 predicts lower spectral accelerations from small earthquakes at the distances, V_{s30} values, and spectral periods included in the comparisons. In most cases, the proposed models predict lower spectral accelerations at short periods and higher spectral accelerations at long periods compared with Zea-2006 and Kea-2006. The gaps in the predictions seem to increase as the distance increases and the magnitude decreases (Figure 11).

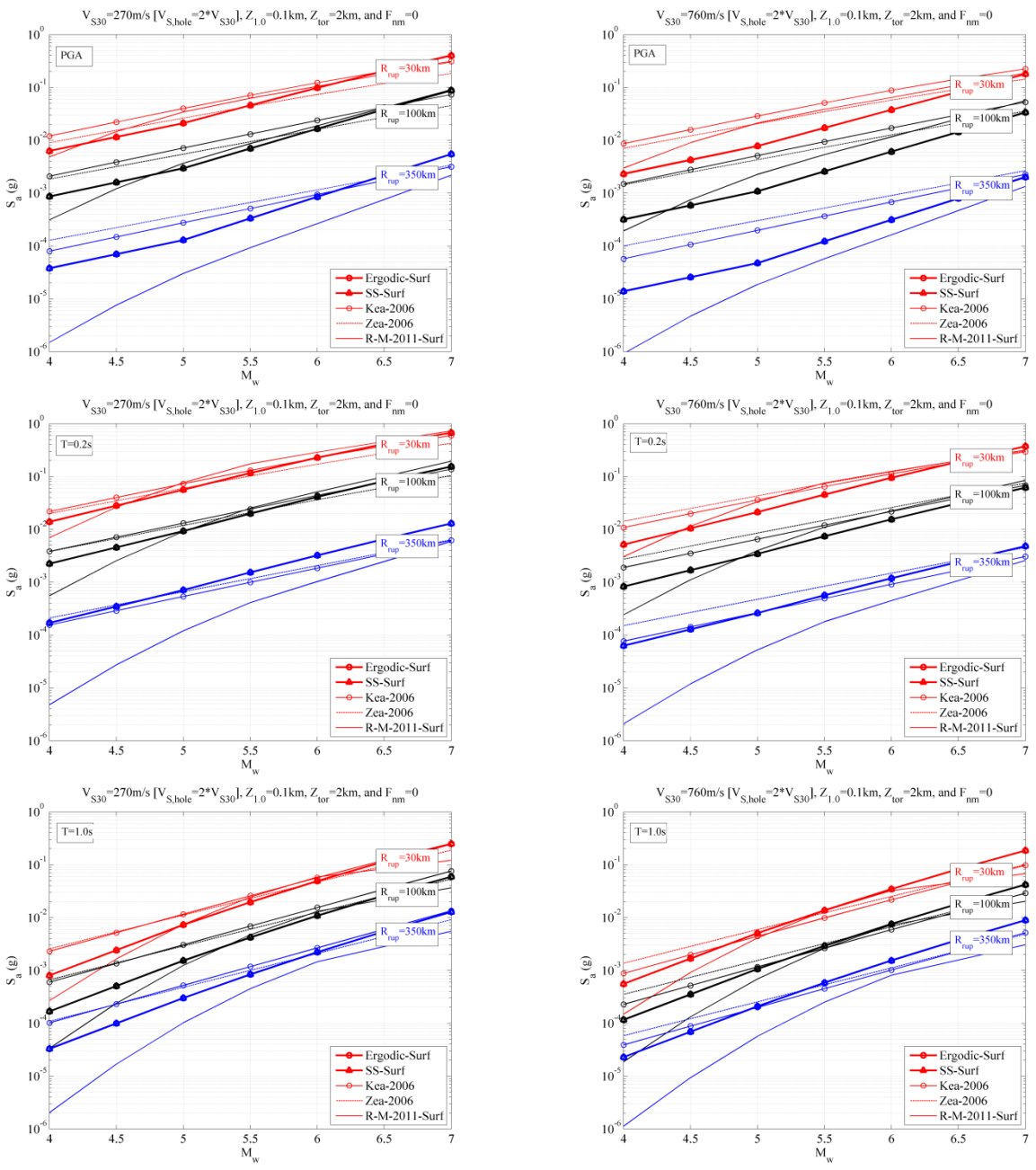


Figure 11. Comparison of the magnitude scaling obtained from the current study (Ergodic-Surf and SS-Surf) and the predictions from Abrahamson et al. (2008; AS-2008) and Abrahamson et al. (2014; ASK-2014). V_{s30} equals 270m/s for the lest panels and 760 for the right panels

Figure 12 shows a comparison between the proposed models and the models by AS-2008 and ASK-2014. The model by ASK-2014 made use of some Japanese data and some parameters

that has Japanese-specific parameters, which were used in this comparison. We present the comparison with AS-2008 to show some of the differences between Japan and western United States. It is clear from the figure that in general the ASK-2014 model – that has some coefficients regressed for Japan in particular – does not result in better predictions compared with AS-2008 (that does not have coefficients for Japan). This could be explained by the fact that most of the data used to develop the ASK-2014 model was not from Japan. Hence, the predictions might be a little biased towards regions that contributed more to their database. Moreover, just a few coefficients were specific for Japan while the majority was common between the different regions. The predicted spectral accelerations in Japan are higher than western United States spectral accelerations for large earthquakes (i.e., M_w of 7), and lower for medium earthquakes (i.e., M_w of 5). The predictions converge at long periods for medium earthquakes (i.e., M_w of 5).

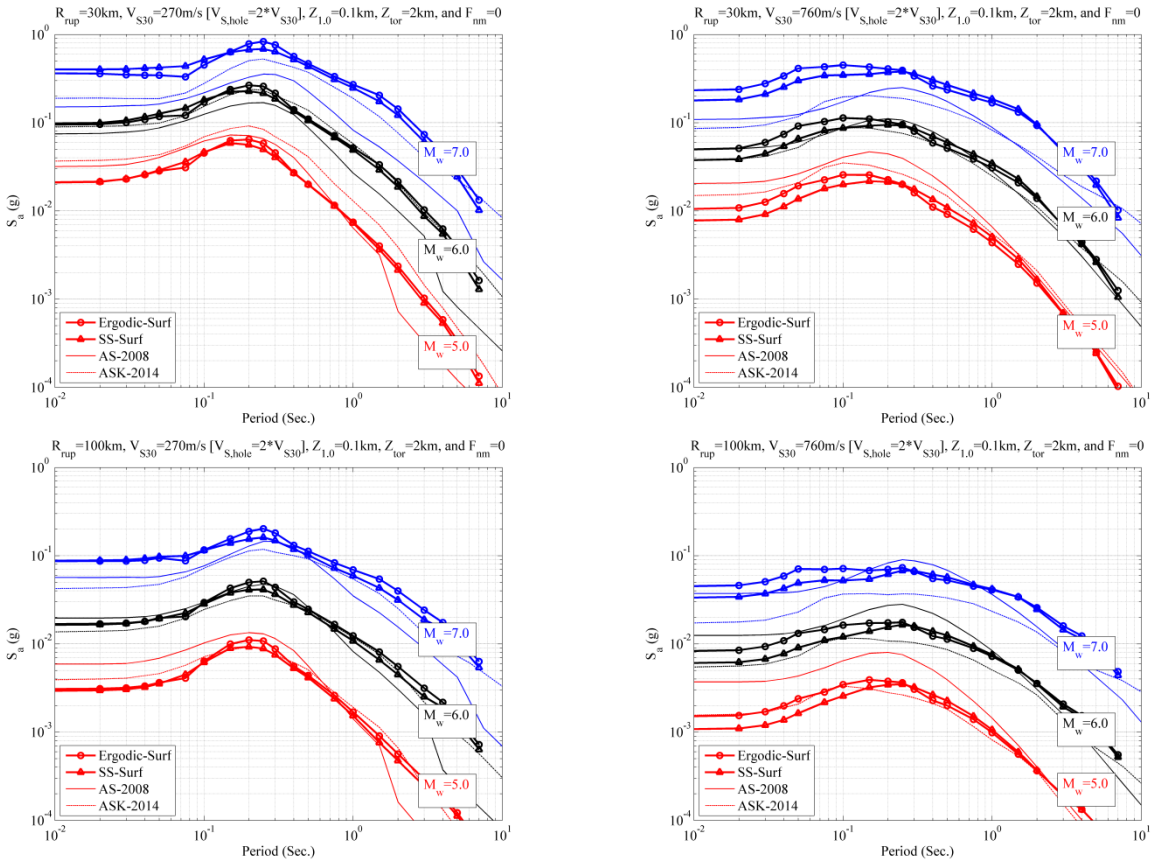


Figure 12. Comparison between the spectral accelerations obtained from the current study (Ergodic-Surf and SS-Surf) and the predictions from Abrahamson et al. (2008; AS-2008) and Abrahamson et al. (2014; ASK-2014)

Rodriguez-Marek et al. (2011) estimated single-station phi (φ_{ss}) values for Japan, and Rodriguez-Marek et al. (2013) estimated values using records from different regions. Figure 13 shows a comparison by the φ_{ss} values estimated in this study with those of the previous studies. The φ_{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 14). Plots of φ_{ss} and φ including only large magnitude (M_w>5) events are also included in Figures 13 and 14. The same trends (higher values) are observed for these magnitude ranges.

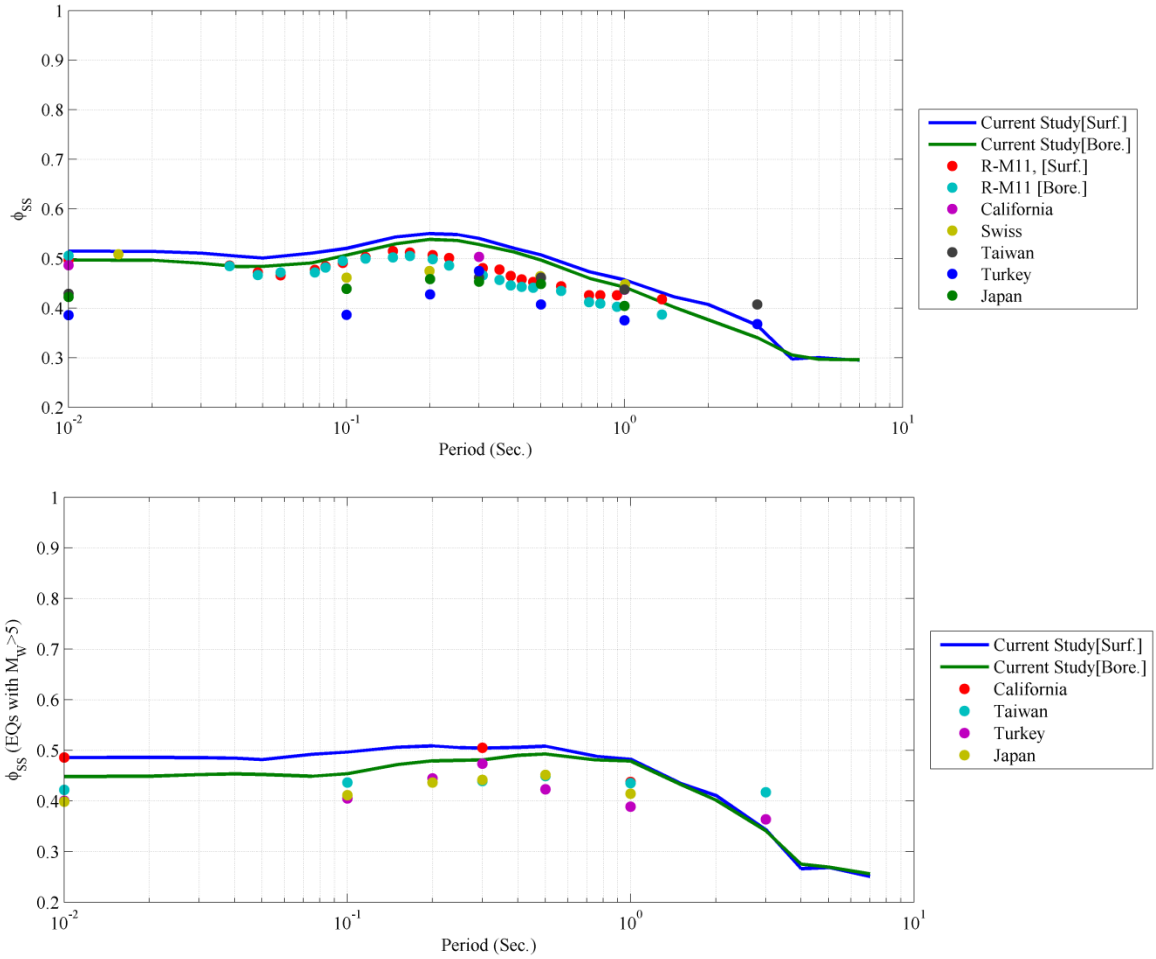


Figure 13. Comparison of φ_{ss} values from this study and Rodriguez-Marek et al. (2011; R-M11) for Japan, and Rodriguez-Marek et al. (2013) for various tectonic regions. PGA is plotted at T=0.01 seconds.

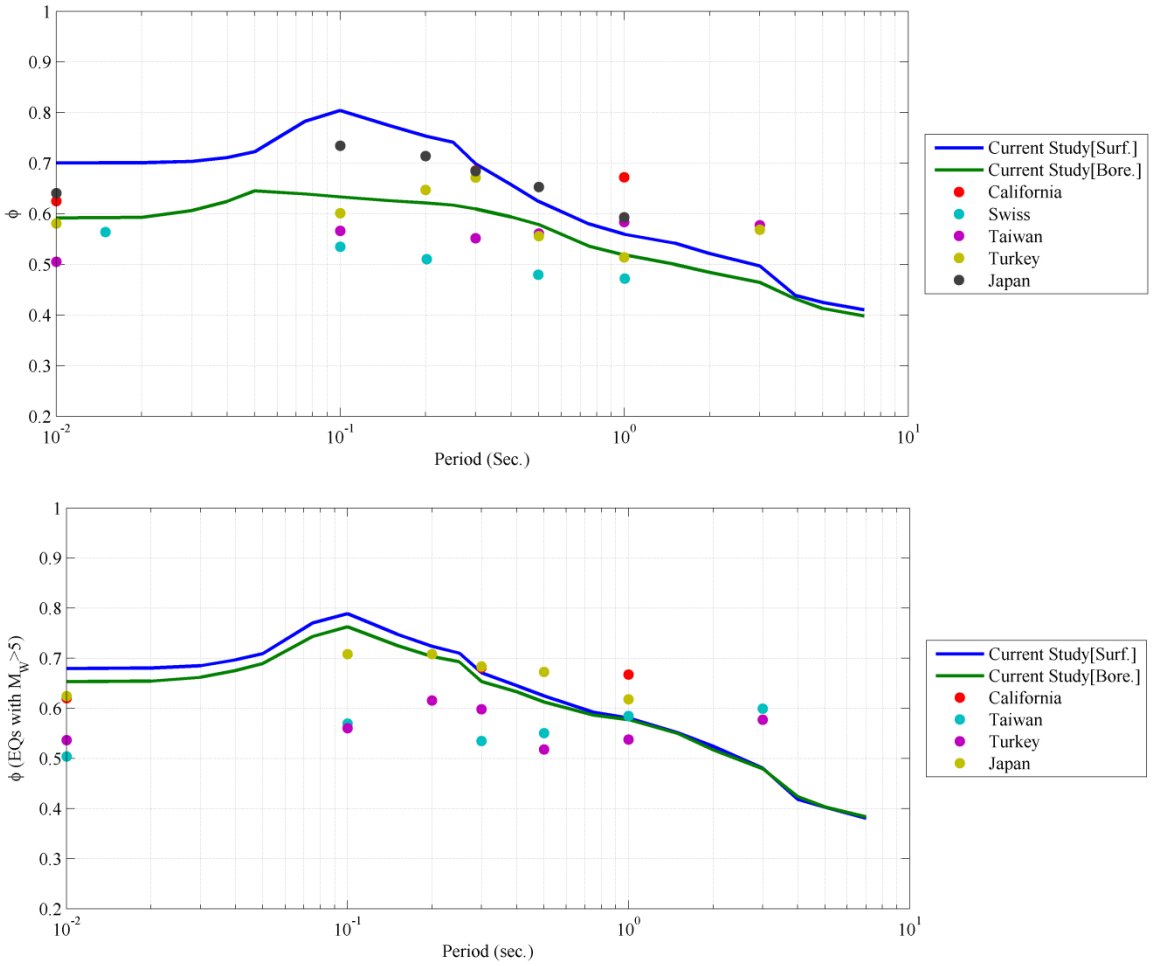


Figure 14. Comparison of φ_{ss} values from this study and Rodriguez-Marek et al. (2013) for various tectonic regions. PGA is plotted at T=0.01 seconds.

Figure 15 shows the comparison between τ obtained from the proposed models, τ reported by R-M-2011, and the average τ obtained from four GMPEs developed within the NGA-W1 project (i.e., Abrahamson et al. 2014, Boore et al. 2014, Campbell and Bozorgnia 2014, and Chiou and Youngs 2014). The τ values estimated by R-M-2011 – which used KiK-net records – for M_w larger than 6.5 are higher than the values obtained from the proposed models at short periods, but lower at long periods. The τ from R-M-2011 for M_w smaller than 5 agrees with the proposed models at short periods, but is considerably higher at long periods. The reduction in τ from the proposed models compared with R-M-2011 could be a result of using the better metadata collected by Dawood et al. (2014) or the considerably larger database used in this

study. At short periods, the average τ estimated from the four NGA-W1 models for M_w of 4 and 5 are fairly similar to the τ from the proposed models (for all earthquakes with M_w between 4 and 7, but the majority of the records are from smaller events). At long periods, the τ values predicted by the NGA-W1 models for M_w of 6 and 7 are closer to τ from the proposed models (the majority of the records at long periods are from larger earthquakes in the proposed models). This shows that the τ values estimated for western United States and Japan are fairly similar.

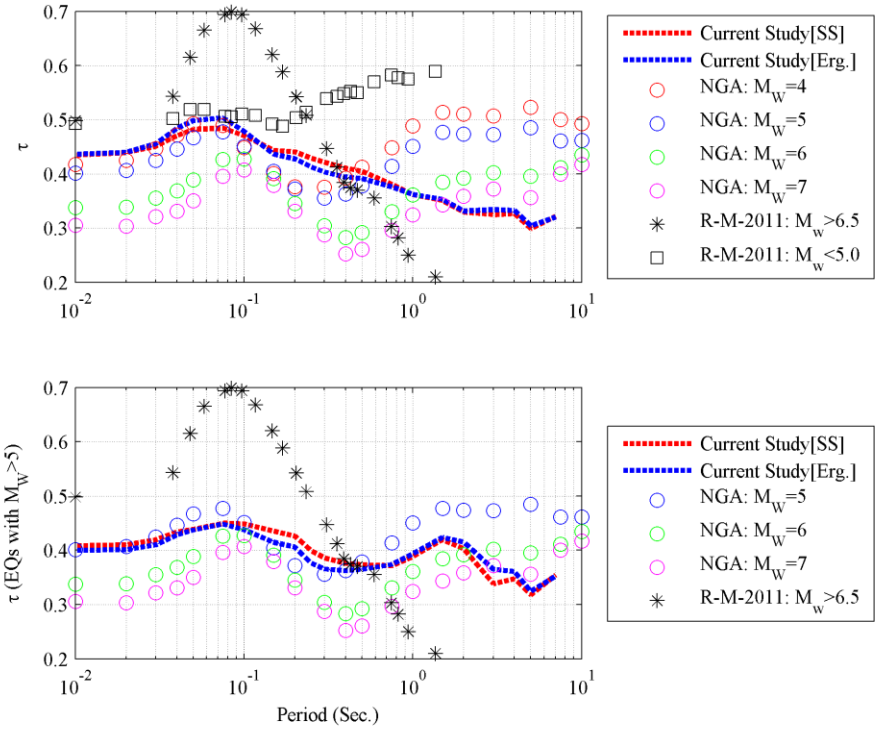


Figure 15. Comparison of τ values from this study, Next Generation Attenuation models at different magnitudes, and Rodriguez-Marek et al. (2011). PGA is plotted at T=0.01 seconds. NGA denotes the average τ obtained from four GMPEs developed within the NGA-W2 project (Abrahamson et al. 2014, Boore et al. 2014, Campbell and Bozorgnia 2014, and Chiou and Youngs 2014).

CONCLUSIONS

In this study we presented two GMPEs (ergodic and site-specific) developed using KiK-net records from active crustal earthquakes. The GMPEs are developed for surface and borehole

motions simultaneously. We adopted a modified form of the functional form from Abrahamson et al. (2013). A mixed effects regression analyses was conducted in 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 ergodic and site-specific GMPEs were compared and the different residual components were studied to investigate potential biases that can limit the applicability of the models. We also compared the GMPEs developed in the current study with other GMPEs developed for Japan and western United States.

The most relevant conclusions from this study are:

- The ergodic and site-specific regressions result in different median predictions. Differences in the median indicate that, despite the considerable size of the database, there is an interplay between event- and site-terms. Regressions that include the site term should be used whenever possible, in particular when using poorly balanced databases.
- The $\varphi^{G/B}$, $\varphi_{ss}^{G/B}$and τ values obtained from the ergodic and site-specific formulations are very similar; however, at some period ranges there are some differences. This shows that estimates of single-station standard deviations should be obtained from non-ergodic regressions.
- The intra-event component of variability at the ground surface (φ^{G}) has a peak at around 0.15 seconds. This peak is less obvious in the borehole component (φ^{B}), suggesting that the increased variability is due to near surface effects.
- The peak in the φ_{ss}^{G} plot is less prominent than the one in the φ^{G} plot. Since φ_{ss} does not account for site-to-site variability, it can be inferred that the observed peak in φ^{G} is due to site-to-site variability of the near-surface deposits that is not captured by the site effects parameterization (e.g., V_{s30} and Z_1).
- We observe some bias in the inter-event residuals for large magnitudes. The limited number of large magnitude earthquakes in the current database could be the cause of this observed bias. Hence, we consider that the proposed models are applicable only for earthquakes with M_w between 4 and 6.5.
- Even if we modified the model that accounts for the depth to top of rupture from the model proposed by ASK13, the inter-event residuals still show bias for earthquakes with

a focal depth larger than 45km. Hence, we discourage using the proposed GMPEs for earthquakes with a top of rupture deeper than 45 km.

- The intra-event residuals show some bias at short and long distances. Hence, the range of applicability of the proposed models is from 30 to 250 km.
- The intra-event residuals from records recorded at soft ($V_{s30} \leq 180$ m/s) and hard rock sites ($V_{s30} > 1500$ m/s) are biased. Hence, the proposed GMPEs are applicable only for predictions at sites with NEHRP Site Class B through D.
- The comparisons between the predictions from the proposed GMPEs and three GMPEs developed for Japan showed some discrepancies. This was to be expected because of the major differences in the nature of the databases used in each study and the parameters used in the regression analyses. For example, the databases used by Zhao et al. (2006) and Kanno et al (2006) contained motions from crustal and subduction earthquakes. Both databases included some motions from regions other than Japan. The database by Rodriguez-Marek et al. (2011) was considerably smaller than the one used in the proposed study.

The database used in this study is considerably large and only includes motions from Japanese active crustal earthquakes. This provided a unique opportunity to apply the ergodic and site-specific formulations on such a large database. The residuals obtained from this study are comprehensively studied in a companion paper.

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.

The mixed effects regression analysis performed for this paper was generated using SAS software, Version 9.3 of the SAS System for Windows. Copyright © 2001-2010 by SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

We used Matlab scripts for the Kea-2006 and Zea-2006 GMPEs coded by Kathryn A. Gunberg (last modified on 5/19/2009 and 5/22/2009, respectively), the AS-2008 GMPE coded by Yoshifumi Yamamoto (last modified on 5/6/10), and the ASK-2014 GMPE coded by Ronnie Kamai (last modified on 07/10/2013).

ACKNOWLEDGEMENTS

This research was partially supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number G11AP200049 and the SIGMA (SeIsmic Ground Motion Assessment) project under grant No. 3000-5910098949. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U. S. government. Additional support was provided by Virginia Tech. The authors would like to acknowledge the National Research Institute for Earth Science and Disaster Prevention (NIED) in Japan for providing the data used in this manuscript.

REFERENCES

Abrahamson N.A., and Silva W.J., 2008. Summary of the Abrahamson & Silva NGA Ground-Motion Relations, Earthquake Spectra. 24 67–97

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

Abrahamson N.A., Silva W.J. and Kamai R., 2014 (In press). Summary of the ASK14 Ground-Motion Relation for Active Crustal Regions. Earthquake Spectra.

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.

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

Aoi, S., 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.

Boore, D. M., Stewart, J. P., Seyhan, E., and Atkinson, G. M., 2014 (In press). NGA-West 2 Equations for Predicting PGA, PGV, and 5%-Damped PSA for Shallow Crustal Earthquakes, Earthquake Spectra.

Campbell K. W., and Bozorgnia, Y., 2014 (In press). NGA-West2 Ground Motion Model for the Average Horizontal Components of PGA, PGV, and 5%-Damped Linear Acceleration Response Spectra, Earthquake Spectra.

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

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

Chiou B., Youngs, R.R., 2013. Update to the Chiou and Youngs 2008 GMPE, PEER Report No. 2013/07, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Chiou B., Youngs, R.R., 2014 (In press). Update of the Chiou and Youngs NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra, Earthquake Spectra.

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.

Dawood, H. M., and Rodriguez-Marek, A., Bayless, J., Goulet, C., and Thompson, E., (in review). A Flatfile For The KiK-Net Database Processed Using An Automated Protocol. Earthquake Spectra

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.

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.

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.

Stafford, P. J., 2014. Crossed and Nested Mixed-Effects Approaches for Enhanced Model Development and Removal of the Ergodic Assumption in Empirical Ground-Motion Models, Bull. Seism. Soc. Am. 104, 702-719.

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.

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.

Chapter 6: Residual Analysis of Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations

Residual Analysis of Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations

INTRODUCTION

This chapter presents the analysis conducted on the residuals and the different components of variability obtained from the GMPEs presented in Chapter 5 of this dissertation. The residuals are studied to: 1) capture its biases with different parameters, and 2) explore the possibility of developing models to be used to predict the different components of variability to be used in seismic hazard analyses. This chapter is divided into two sections. The first presents the investigation of the different residual and its variability, while the second presents the investigation of the empirical amplification factors and its variability estimated at the different KiK-net stations. The residual analyses are preceded by a discussion of the effects of smoothing the regression coefficients on the predictions of the ergodic and site-specific GMPEs formulated in Chapter 5.

EFFECT OF SMOOTHING THE REGRESSION COEFFICIENTS

When developing a GMPE, it is typical practice to smooth the regression coefficients so the predicted response spectra is smooth. Smoothing the coefficients is important from a practical point of view because the smoothing process implies hazard estimates that vary smoothly from period to period. On the other hand, smoothing may lead to biased estimates at some periods, and, in some cases, it may obscure trends in residuals. Hence, it is important to study GMPE residuals prior to smoothing the coefficients in order to have a better understanding of the behavior of these residuals.

In order to evaluate the GMPE residuals without smoothing we repeat the regression presented in Chapter 5 of this dissertation without none of the smoothing steps (as a result, the only coefficients that are smoothed are a3 and c4, see Chapter 5 for more details). From this point on, we refer to the residuals from the GMPE with smoothed coefficients as “smoothed

residuals" and the residuals from the GMPEs with unsmoothed coefficients as "unsmoothed residuals". We conducted the regression twice using the ergodic and site-specific formulations. The ergodic is denoted in this manuscript as "V3", and the site-specific as "V4". The within-event residuals from the ergodic formulation are split into site-specific and site- and event-corrected components. This split is accomplished using Maximum Likelihood Estimations. These residual components are denoted as "V5" in this manuscript (Table 1). The correlation coefficients between the residual components obtained from the GMPEs with smoothed and unsmoothed regression coefficients are generally very high (Figure 1). The correlation coefficients are slightly reduced at long periods because the number of usable motions considerably decreases as the period increase. Figure 2 show examples of predicted pseudo-spectral accelerations from the GMPEs with smoothed and unsmoothed regression coefficients. The plots show some discrepancies that result from smoothing some of the coefficients. I particular, there is a clear discrepancy at short periods which is more prominent for smaller magnitudes. Also, the curves are irregular for the prediction from the unsmoothed GMPE especially in regions of limited data (e.g., long periods for small magnitude earthquakes). Moreover, the surface predictions from the ergodic GMPE show a clear discrepancy in the spectral periods between 0.07 and 0.3s.

Table 1. The different versions of the residual components presented in this study

Version	Description
V3	Residuals from the ergodic GMPE
V4	Residuals from the site-specific GMPE
V5	Residuals obtained from splitting the within-event residuals from the ergodic GMPE

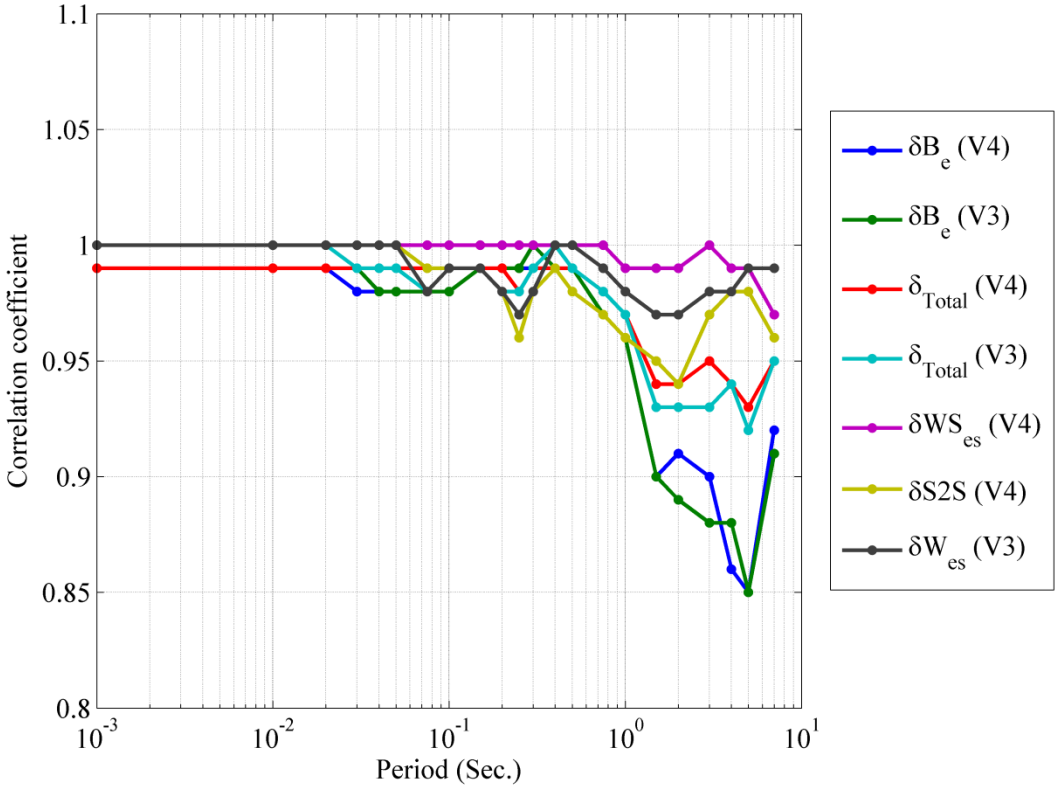


Figure 1. The correlation coefficient between residual components obtained from GMPEs with smoothed and unsmoothed regression coefficients

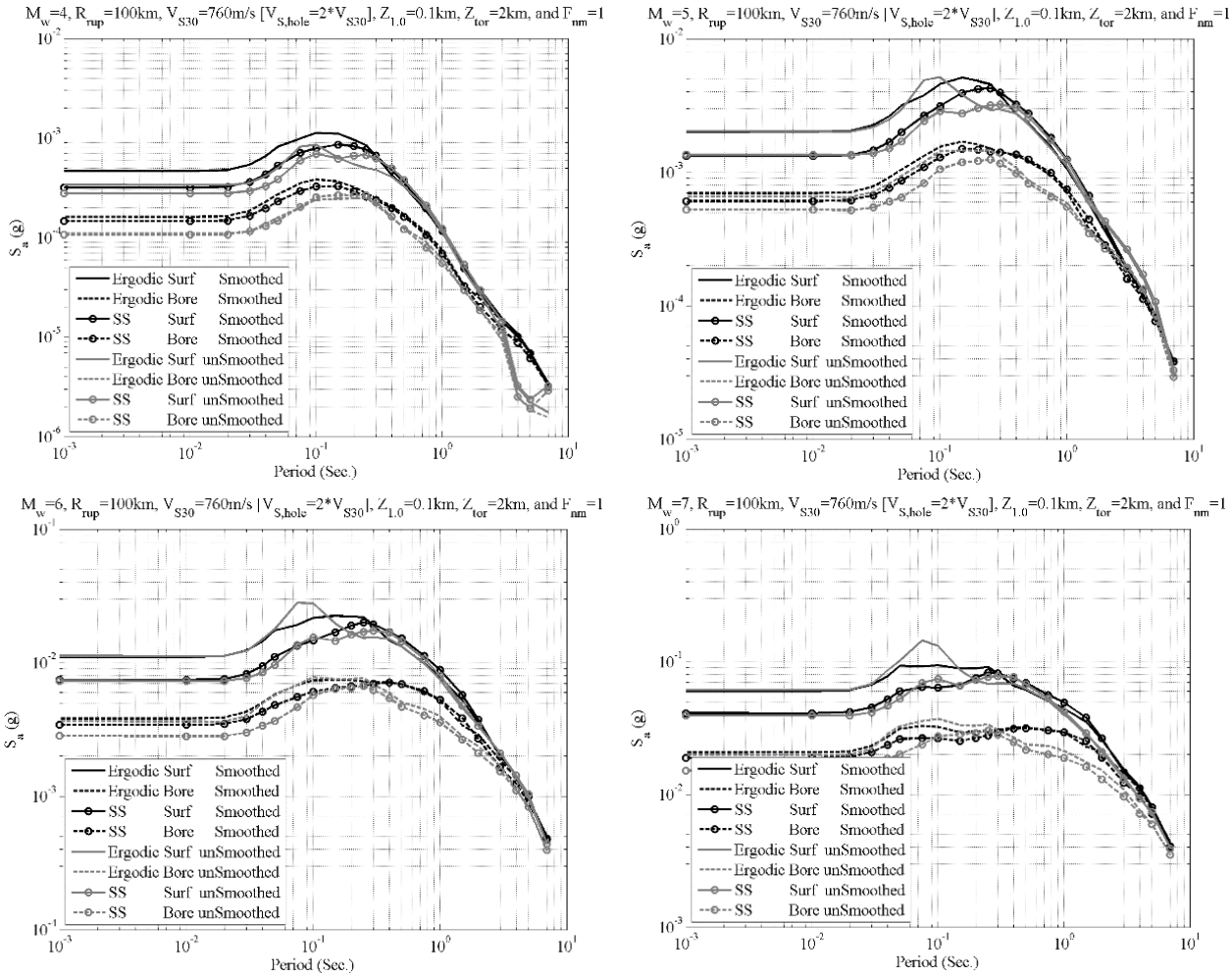


Figure 2. Examples of predicted pseudo-spectral accelerations from GMPEs with smoothed and unsmoothed regression coefficients. The plots show some bias that is a result of smoothing some of the coefficients. Also, the curves are irregular for the unsmoothed GMPE especially in regions of limited data (e.g., long periods for small magnitude earthquakes)

ANALYSIS OF RESIDUAL AND VARIABILITY COMPONENTS

Residual components were defined in Chapter 5 when presenting the formulation of ergodic and site-specific formulations. We investigated the dependence of $\delta W_{es}^{G/B}$, $\delta WS_{es}^{G/B}$, $\varphi^{G/B}$, $\varphi_{SS}^{G/B}$, and $\varphi_{ss,s}^{G/B}$ on event and site parameters. We compared the scatter plots of the "smoothed residuals" and the "unsmoothed residuals" in a preliminary analysis. Both sets of

residuals resulted in very similar scatter plots. This was expected given the very high correlation coefficients between the different residual components (Figure 3). Hence, we did not differentiate between the "smoothed residuals" and "unsmoothed residuals" in all the figures shown subsequently. The following discussion focuses only on parameters for which a clear dependency was observed. The discussion is preceded by a review of the partition of residuals into its ergodic and site-specific components.

Partitioning of Residuals

The number and type of the different components in which the total residual can be split into depends on the nature of the available database. In general, GMPEs take the following form:

$$ln\,(y_{es}^{G/B}) = ln(\mu_{es}^{G/B}) + \Delta_{es}^{G/B} \quad (1)$$

where the superscripts *G* and *B* denote motions recorded at ground surface, and bedrock (within motion), respectively. The subscripts *e* and *s* denote event and station, respectively. $ln\,(y_{es}^{G/B})$ are the natural logarithm of an observed ground motion parameter at ground surface and borehole; $ln(\mu_{es}^{G/B})$ are the median predictions; and $\Delta_{es}^{G/B}$ are the total residuals.

In ergodic GMPEs, $\Delta_{es}^{G/B}$ are generally split into an inter-event (δB_e) and intra-event ($\delta W_{es}^{G/B}$) residuals components. Hence, the GMPE takes the following form:

$$ln\,(y_{es}^{G/B}) = ln(\mu_{es}^{G/B}) + \delta B_e + \delta W_{es}^{G/B} \quad (2)$$

where, δB_e represents the observed shift from the median for records from event *e*.

Figure 3 shows this concept of inter- and intra-event residual components. Since δB_e is an event-specific residual component, it should be the same whether the recorded motion is on ground surface or in bedrock.

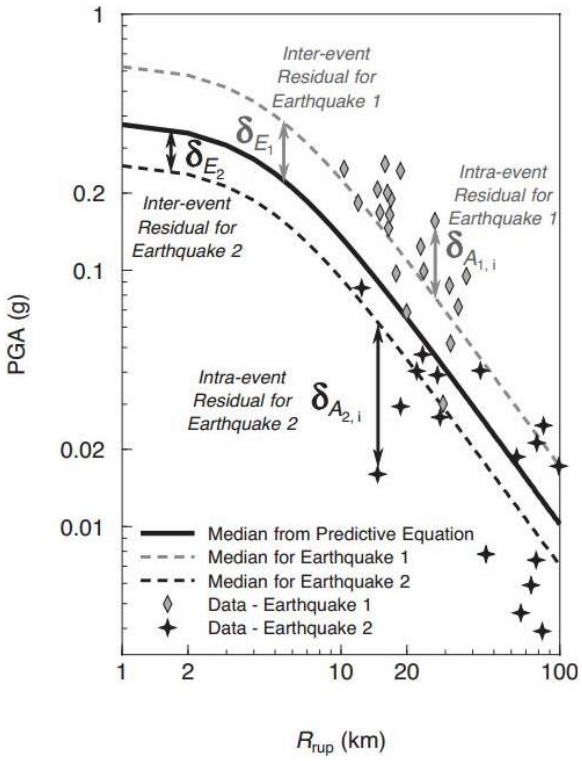


Figure 3. Inter-event and intra-event components of ground-motion variability. The figure shows two earthquakes for which the median of its records is shifted from the median of the GMPE (figure adopted from Strasser et al. 2009)

If the available database contains multiple strong GM records recorded at the same station, $\delta W_{es}^{G/B}$ can be split to include a site-specific component:

$$\delta W_{es}^{B} = \delta S2S_S^B + \delta W_{es,0}^{B} \quad (3)$$

$$\delta W_{es}^{G} = \delta S2S_S^G + \delta W_{es,0}^{G} \quad (4)$$

where $\delta S2S_S^B$ is a site-to-site residual component for the borehole motions and which represents the average shift from the median for borehole records at site *s*; $\delta S2S_S^G$ is a surface-to-surface residual component which represents the observed shift from the median for records from site *s*; and $\delta W_{es,0}^{B}$ and $\delta W_{es,0}^{G}$ represent the remaining residual components after accounting for the site-

specific residual components. A GMPE that includes the partition of residuals as described in Equations 3 and 4 is known as a site-specific GMPE.

Assuming that the residual components are uncorrelated, the total standard deviation at the ground surface and the borehole instruments for an ergodic GMPE are defined as:

$$\sigma_{tot}^{G/B} = \sqrt{\tau^2 + (\varphi^{G/B})^2} \quad (5)$$

where τ and $\varphi^{G/B}$ are the standard deviations of δB_e and $\delta W_{es}^{G/B}$, respectively. Alternatively, in a site-specific (partially non-ergodic) formulation, the total standard deviations are defined as:

$$\sigma_{tot}^{G/B} = \sqrt{\tau^2 + \left(\varphi_{S2S}^{G/B}\right)^2 + \left(\varphi_{ss}^{G/B}\right)^2} \quad (6)$$

where $\varphi_{S2S}^{G/B}$ are the standard deviations of the $\delta S2S_s^{G/B}$ terms and $\varphi_{ss}^{G/B}$ are the standard deviations of the $\delta W_{es,o}^{G/B}$ residual components.

If *m* ground motions were recorded at station *s*, then the event-corrected single-station standard deviation at station *s* ($\varphi_{ss,s}$) can be defined as follows:

$$\varphi_{ss,s}^{B/G} = stdev\left(\left[\left(\delta W_{es,0}^{G/B}\right)_1, \left(\delta W_{es,0}^{G/B}\right)_2, \ldots, \left(\delta W_{es,0}^{G/B}\right)_m\right]\right) \quad (7)$$

where $\left[\left(\delta W_{es,0}^{G/B}\right)_1, \left(\delta W_{es,0}^{G/B}\right)_2, \ldots, \left(\delta W_{es,0}^{G/B}\right)_m\right]$ is a vector of the remaining residual components (Equations 3 and 4) from all motions recorded at station *s*.

Magnitude and Distance Dependence

As part of this study, we use the residuals from the GMPEs presented in Chapter 5 of this dissertation to study the magnitude and distance dependence of the within-event variability ($\varphi^{G/B}$) and single station phi ($\varphi_{ss}^{G/B}$). Based on the observed trends, we developed models to predict $\varphi^{G/B}$ and $\varphi_{ss}^{G/B}$ as functions of magnitude and distance. The φ_{ss}^{G} models developed for the PEGASOS Refinement 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 that project.

Rodriguez-Marek et al. (2011) estimated $\varphi_{ss}^{G/B}$ values for Japan, and Rodriguez-Marek et al. (2013) estimated φ_{ss}^{G} using motions from different regions. Figure 4 shows a comparison of the $\varphi_{ss}^{G/B}$ values estimated in this study with φ_{ss}^{G} from previous studies. The $\varphi_{ss}^{G/B}$ estimated in the current study are slightly higher. Values of $\varphi^{G/B}$ from this study also tend to be higher than the values from previous studies at short periods, but lower at long periods (Figure 5). Plots of $\varphi_{ss}^{G/B}$ and $\varphi^{G/B}$ including only large magnitude (M>5) events are also included in Figures 4 and 5. The same trends (higher values) are observed for these magnitude ranges as well.

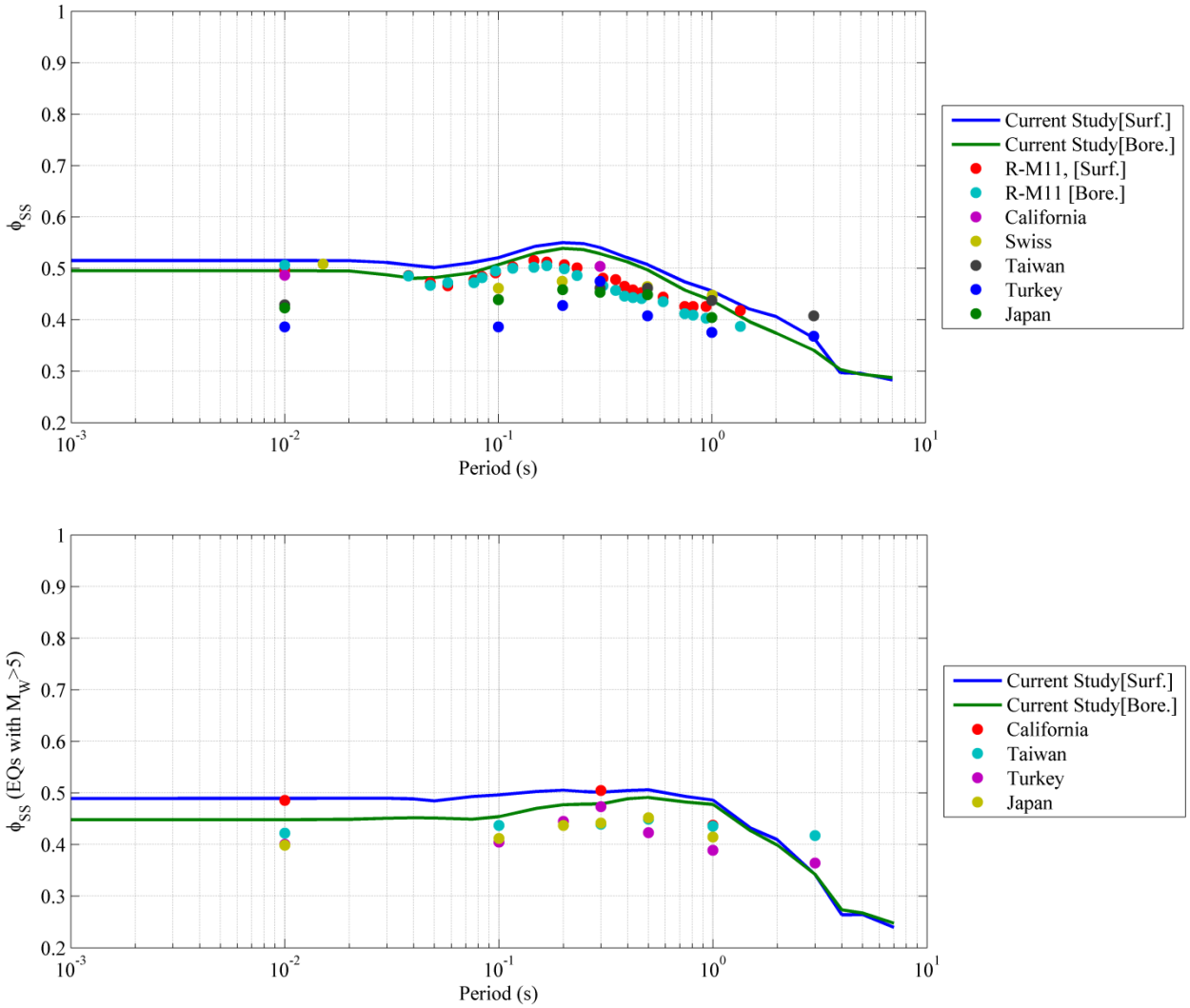


Figure 4. Comparison of $\varphi_{ss}^{G/B}$ values from this study and Rodriguez-Marek et al. (2011; R-M11) for Japan, and φ_{ss}^{G} from Rodriguez-Marek et al. (2013) for various tectonic regions. PGA is plotted at T=0.01 seconds.

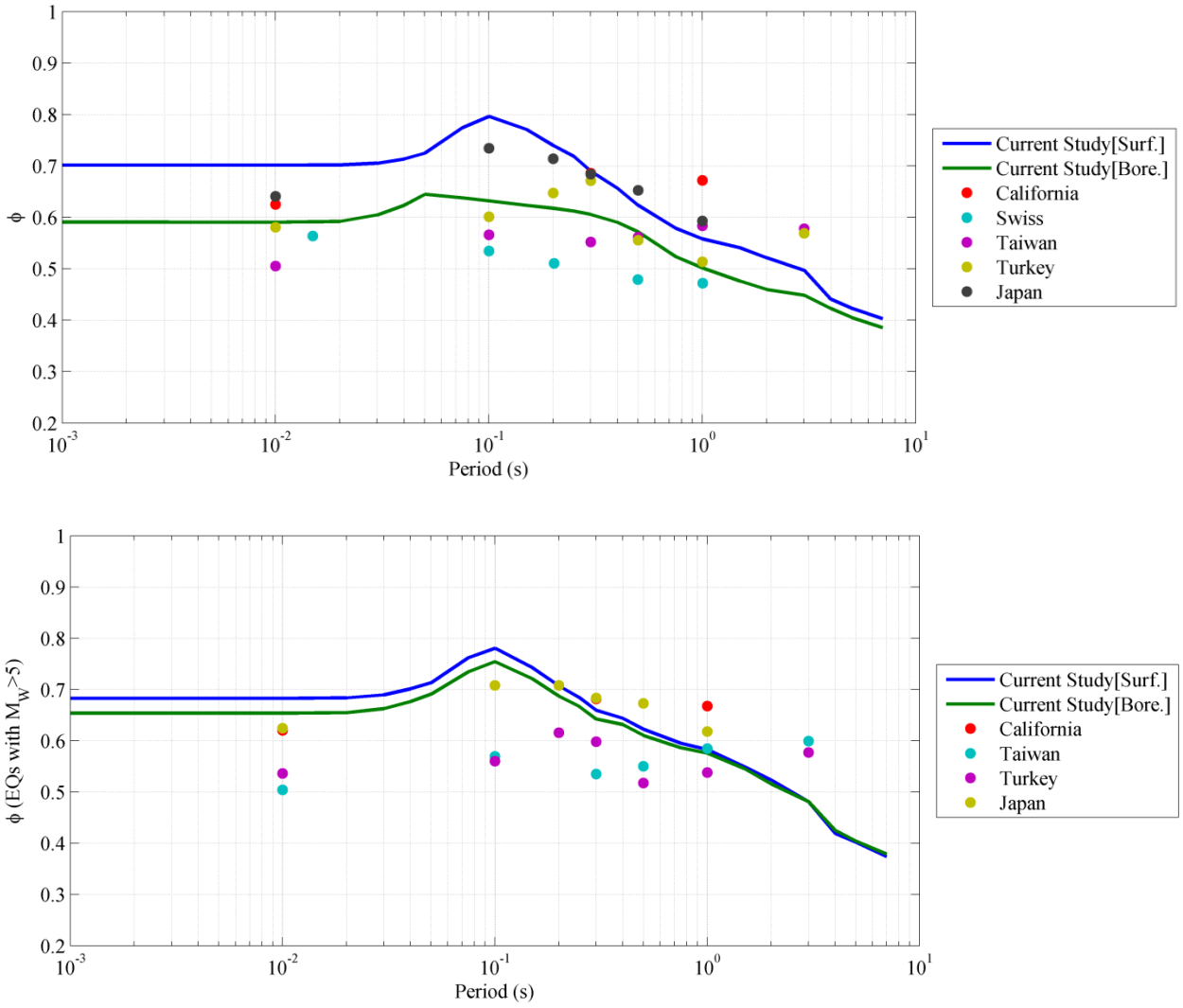


Figure 5. Comparison of $\varphi^{G/B}$ values from this study and φ^{G} from Rodriguez-Marek et al. (2013) for various tectonic regions. PGA is plotted at T=0.01 seconds.

In addition to computing average $\varphi^{G/B}$ and $\varphi_{ss}^{G/B}$ values, an attempt was made at developing magnitude- and distance- dependency into models for $\varphi^{G/B}$ and $\varphi_{ss}^{G/B}$. For both magnitude- and distance-dependence, multi-linear models were fit to the data. These models are fit separately for the Ergodic GMPE and the site-specific GMPE.

Ergodic GMPE

The total variability is split into an inter-event component (τ) and an intra-event component ($\varphi^{G/B}$). 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. The form of the different adopted models is given by

$\varphi^{G/B}(T) = s_0^{G/B}\ (T)$ (8)

$\tau\ \ (T) = s_1\ \ (T)$ (9)

$$\varphi^{G/B}(T, M_w) = \begin{cases} s_2^{G/B}(T) & for\ \ 4 \le M_w \le 4.7 \\ s_2^{G/B}(T) - \left(s_2^{G/B}(T) - s_3^{G/B}(T)\right)\frac{M_W - 4.7}{2.3} & for\ \ 4.7 > M_w \le 7 \end{cases} \quad (10)$$

$\tau\ (T, M_w) = s_4(T)$ (11)

$$\varphi^{G/B}\ (T, R_{rup}) = \begin{cases} s_8^{G/B}(T) & for\ \ R_{rup} \le 30 \\ s_8^{G/B}(T) - \left(s_8^{G/B}(T) - s_9^{G/B}(T)\right)\frac{log(R_{rup}) - (30)}{log(350) - log(30)} & for 30 > R_{rup} \le 350 \end{cases} \quad (12)$$

$\tau\ (T, M_w) = s_{10}(T)$ (13)

$\varphi^{G/B}$ (T,M_w, R_{rup})=

$$\begin{cases} M1R1^{G/B}(T) & for \quad R_{rup} \leq 30 \; and \; 4 \leq M_w \leq 4.7 \\ M1R1^{G/B}(T) - (M1R1^{G/B}(T) - M1R3^{G/B}(T))\frac{log(R_{rup})}{log(350)} & for \quad 30 > R_{rup} \leq 350 \; and \; 4 \leq M_w \leq 4.7 \\ M1R1^{G/B}(T) + (M2R1^{G/B}(T) - M1R1^{G/B}(T))\frac{M_W - 4.7}{2.3} & for \quad R_{rup} \leq 30 \; and \; 4.7 < M_w \leq 7 \\ (M1R1^{G/B}(T) - (M1R1^{G/B}(T) - M1R3^{G/B}(T))\frac{log(R_{rup})}{log(350)})(1 - \frac{M_W - 4.7}{2.3}) + & \\ (M2R1^{G/B}(T) - (M2R1^{G/B}(T) - M2R3^{G/B}(T))\frac{log(R_{rup})}{log(350)})\frac{M_W - 4.7}{2.3} & for \quad 30 > R_{rup} \leq 350 \; and \; 4.7 < M_w \leq 7 \end{cases} \quad (14)$$

where Equations 8 and 9 represent Model (1) where $\varphi^{G/B}$ and τ are only period dependent. Equations 10 and 11 represent Model (2) where $\varphi^{G/B}$ is period and M_w dependent. Equations 12 and 13 represent Model (3) where $\varphi^{G/B}$ is period and R_{rup} dependent. Equation 14 represent Model (4) where $\varphi^{G/B}$ is period, M_w and R_{rup} dependent

We used Maximum likelihood estimation to regress for the coefficients of these models (Table 2). 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. Figures 6 through 9 show $\varphi^{G/B}$ and τ obtained from the different models at PGA, 0.3s, 1.0s, and 7.0s. The figures also show 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:

Table 2. The regression coefficients used in the models presented in Equations 8 through 14

T (sec.)	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]

T (sec.)	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
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 2 (Cont'd). The regression coefficients used in the models shown in Equations 8 through 14

T (Sec.)	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]

T (Sec.)	Model (3)				
	s_8^G	s_8^B	s_9^G	s_9^B	s_{10}
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 2 (Cont'd). The regression coefficients used in the models shown in Equations 8 through 14

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]

T (s.)	Model (4)							
	$M1R1^{G}$	$M1R3^{G}$	$M2R1^{G}$	$M2R3^{G}$	$M1R1^{B}$	$M1R3^{B}$	$M2R1^{B}$	$M2R3^{B}$
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]

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 2 and Figure 6).

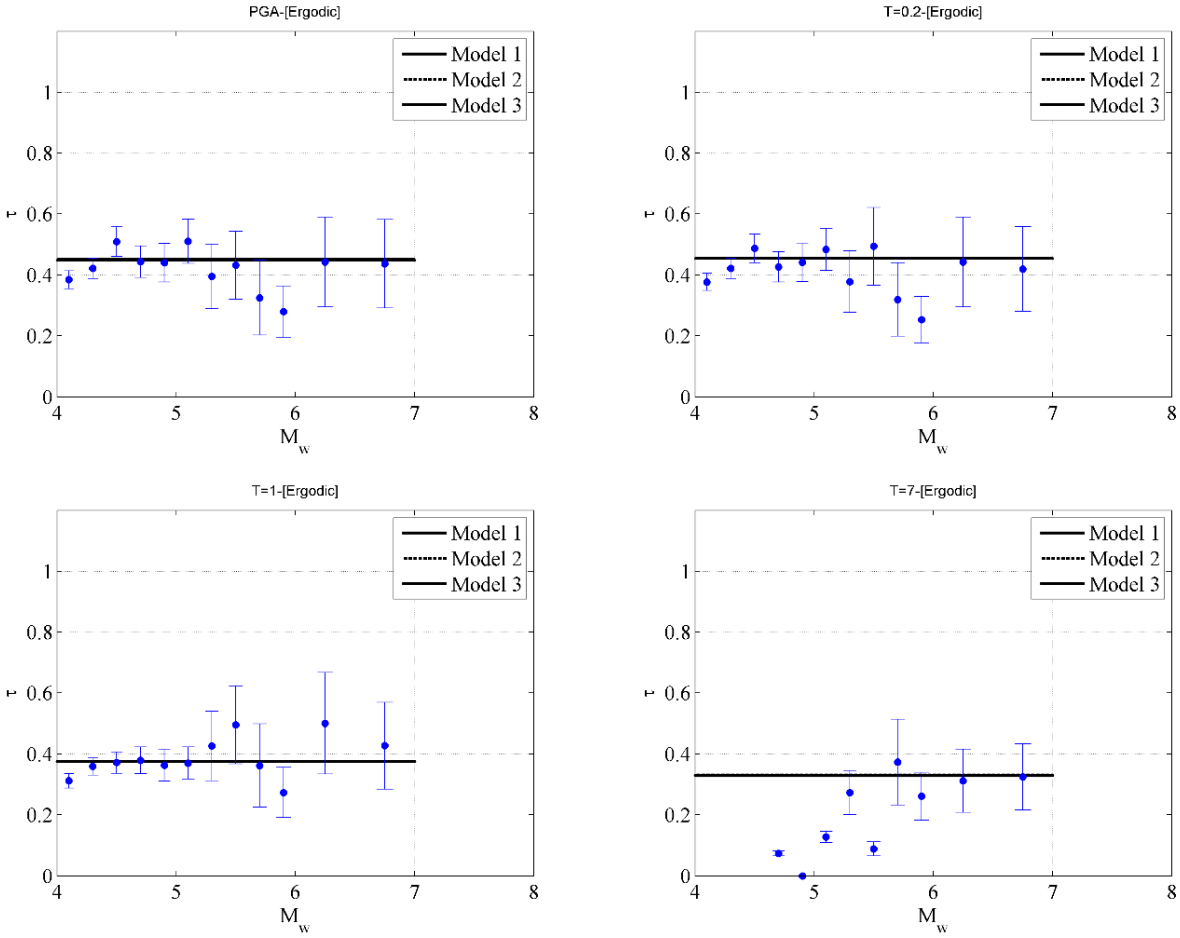


Figure 6. Models for τ from the ergodic formulation. The three models are magnitude and distance independent, but each was developed in conjunction with a specific φ model. The three models result in almost the same values. The three models are poorly constrained for T=7 seconds

$\varphi^{G/B}$ 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 only for M_w between 4 and 6.5. In general, the models for φ^G and φ^B are different at short periods and 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^Gmodel 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 constant slope model up to a distance of 350km. Another difference between the current R_{rup}-dependent model and the model by Abrahamson et al. (2013) is that they reported higher φ^Gat larger distanced for small periods and lower φ^Gat larger distances for long periods, while we observed the opposite 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 and 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, Abrahamson et al. (2013) used mainly large events from Japan (6.1 to 6.9) which made them capture the tendency of φ^G to increase with distance. This shows the importance of developing models that are M_w-R_{rup}-dependent.

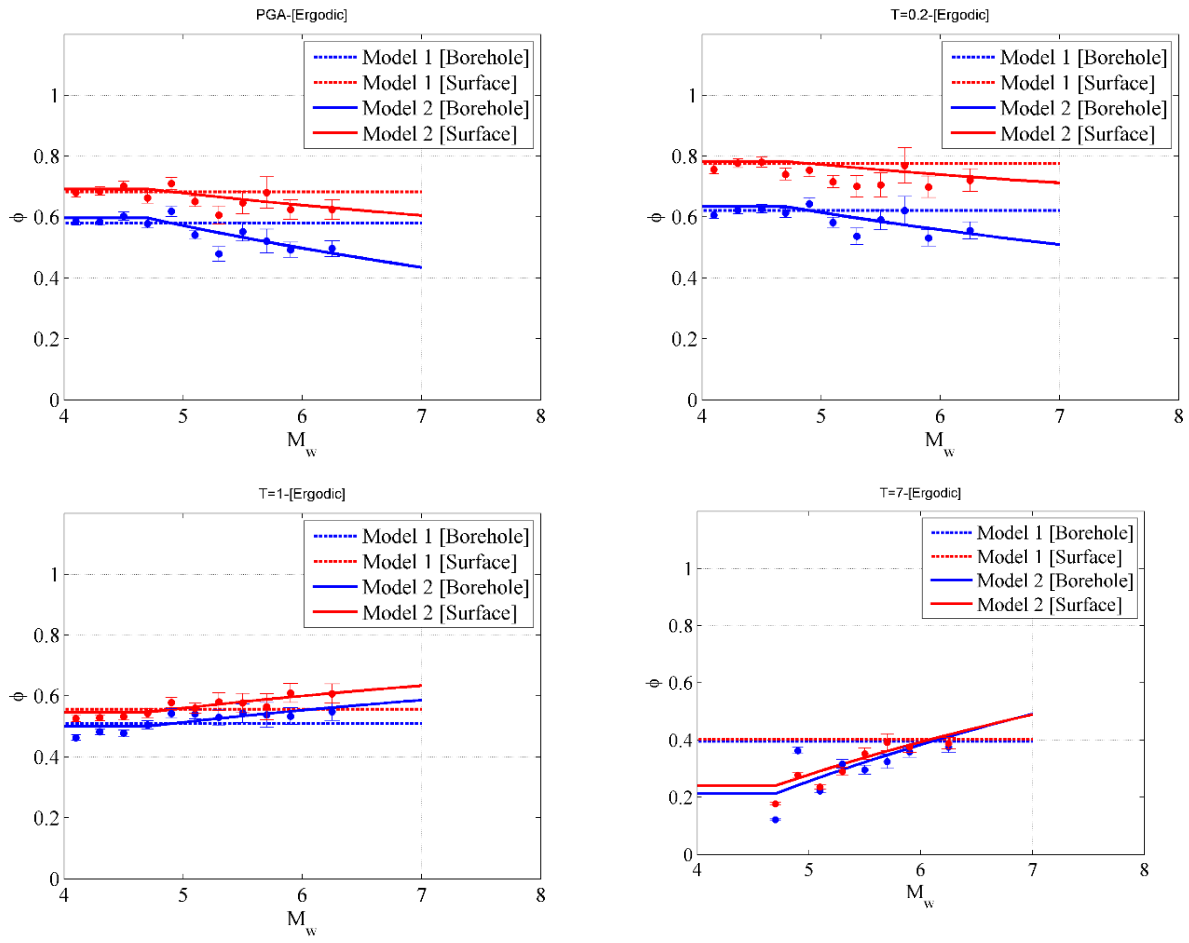


Figure 7. Models for φ. Model 1 is a distance- and magnitude-independent model. Model 2 is a magnitude-dependent model. The latter model is poorly constrained for T=7 seconds. The circles show the standard deviation of site and event corrected residuals within a magnitude bin.

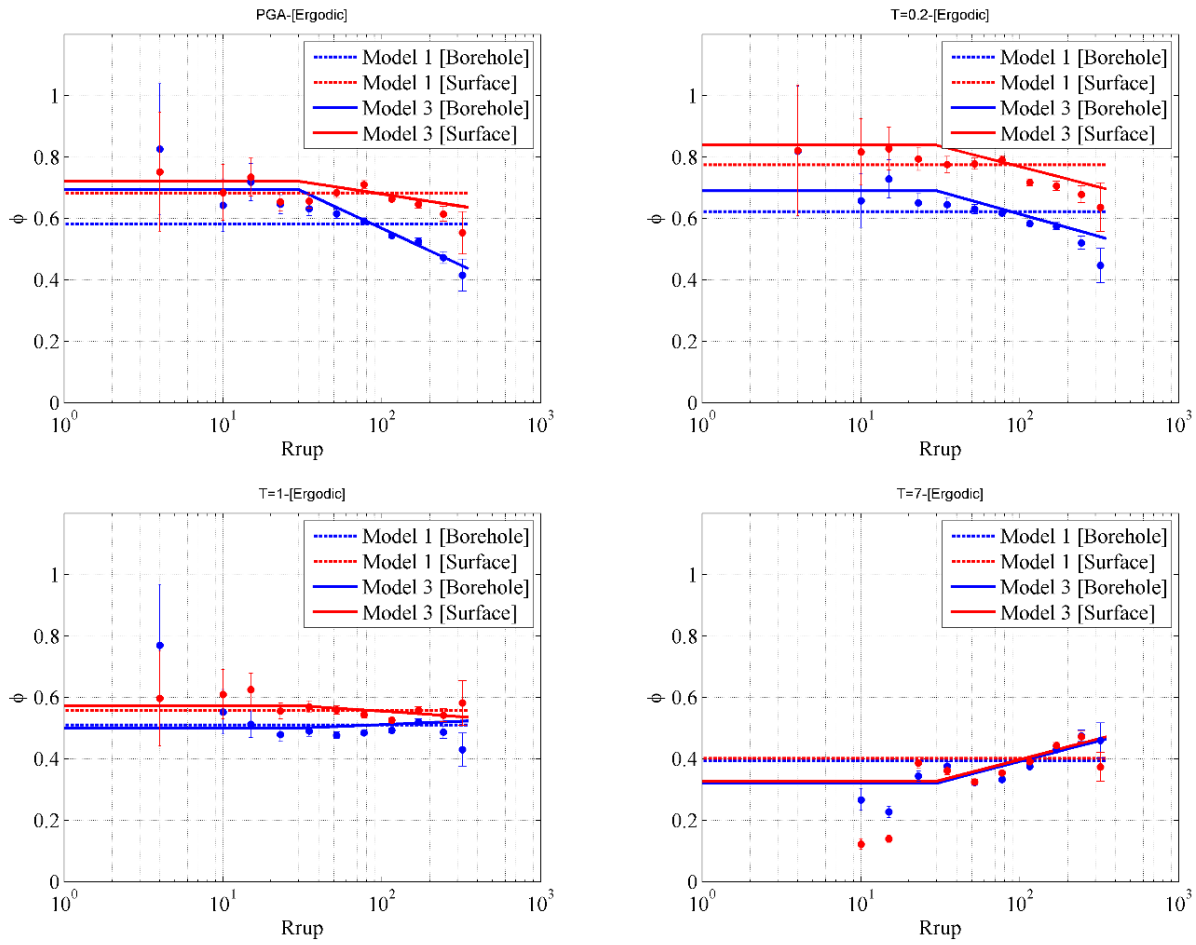


Figure 8. Models for φ. Model 1 is a distance- and magnitude-independent model. Model 3 is a distnace-dependent model. The latter model is poorly constrained for T=7 seconds. The circles show the standard deviation of site and event corrected residuals within a distance bin.

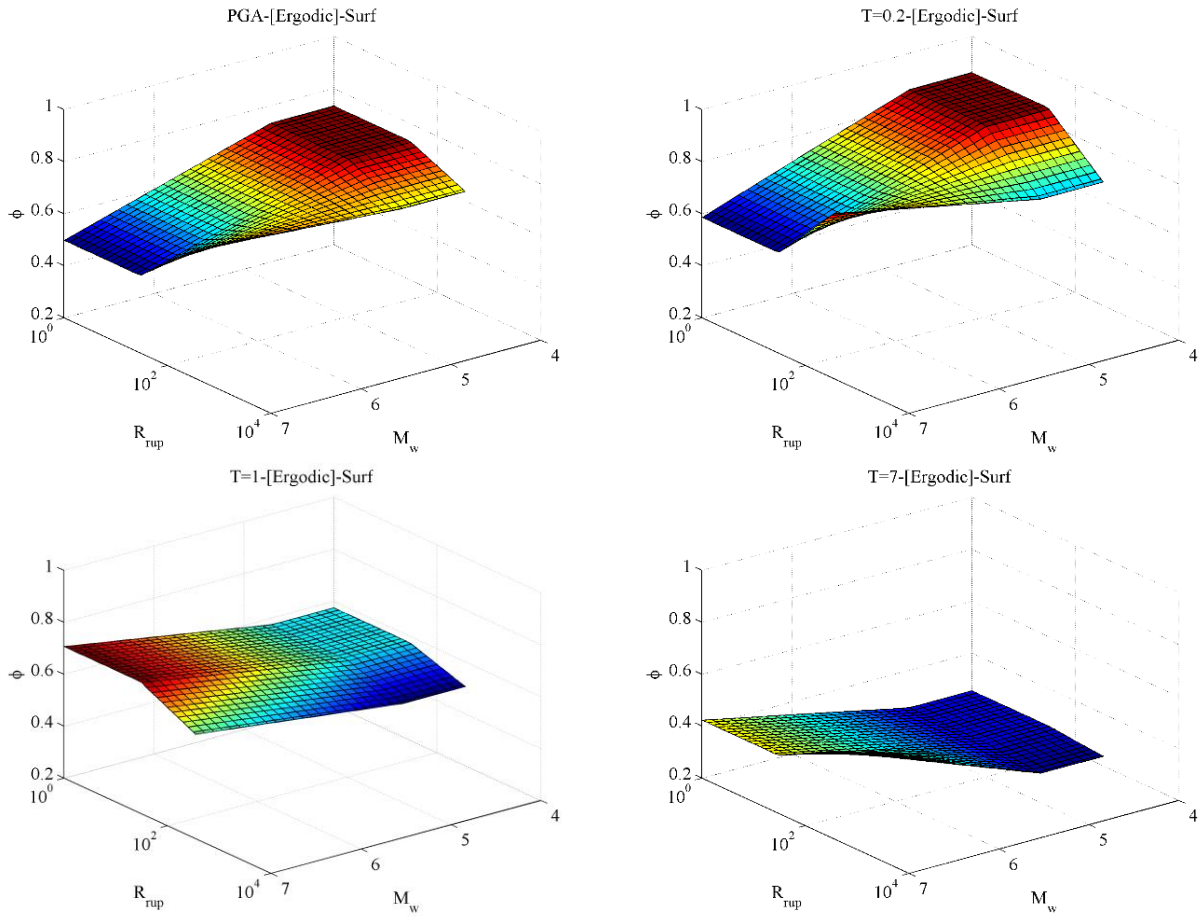


Figure 9. Model 3 for φ which is a distance- and magnitude-dependent model.

Site-Specific GMPE

Similar to the analyses presented in the previous section, we developed models to predict $\varphi_{SS}^{G/B}$ as a function of R_{rup} and M_w, as well as a homoskedastic model for τ. The form of the different models is given by

$$\varphi_{SS}^{G/B}(T) = s_0^{G/B}(T) \tag{15}$$

$$\varphi_{SS}^{G/B}(T, M_w) = \begin{cases} s_2^{G/B}(T) & for \quad 4 \leq M_w \leq 5 \\ s_2^{G/B}(T) - \left(s_2^{G/B}(T) - s_3^{G/B}(T)\right)\frac{M_W - 5}{2} & for \quad 5 > M_w \leq 7 \end{cases} \tag{16}$$

$$\varphi_{SS}^{G/B}(T, R_{rup}) = \begin{cases} s_7^{G/B}(T) & for \quad R_{rup} \leq 30 \\ s_7^{G/B}(T) - \left(s_7^{G/B}(T) - s_8^{G/B}(T)\right)\frac{log(R_{rup}) - log(30)}{log(100) - log(30)} & for \quad 30 > R_{rup} \leq 100 \\ s_8^{G/B}(T) & for \quad 100 > R_{rup} \leq 350 \end{cases} \tag{17}$$

$\varphi_{SS}^{G/B}$ (T, M_w, R_{rup}) =

$$\begin{cases} M1R1^{G/B}(T) & for \quad R_{rup} \le 30 \; and \; 4 \le M_w \le 5 \\ M1R1^{G/B}(T) - (M1R1^{G/B}(T) - M1R3^{G/B}(T)) \frac{log(R_{rup}) - log(30)}{log(100) - log(30)} & for \quad 30 > R_{rup} \le 100 \; and \; 4 \le M_w \le 5 \\ M1R3^{G/B}(T) & for \quad 100 > R_{rup} \le 350 \; and \; 4 \le M_w \le 5 \\ M1R1^{G/B}(T) + (M2R1^{G/B}(T) - M1R1^{G/B}(T)) \frac{M_W - 5}{2} & for \quad R_{rup} \le 30 \; and \; 5 < M_w \le 7 \\ (M1R1^{G/B}(T) - (M1R1^{G/B}(T) - M1R3^{G/B}(T)) \frac{log(R_{rup}) - log(30)}{log(100) - log(30)})(1 - \frac{M_W - 5}{2}) + & for \quad 30 > R_{rup} \le 100 \; and \; 5 < M_w \le 7 \\ \quad (M2R1^{G/B}(T) - (M2R1^{G/B}(T) - M2R3^{G/B}(T)) \frac{log(R_{rup}) - log(30)}{log(100) - log(30)}) \frac{M_W - 5}{2} & \\ M1R3^{G/B}(T) + (M2R3^{G/B}(T) - M1R3^{G/B}(T)) \frac{M_W - 5}{2} & for \; 100 > R_{rup} \le 350 \; and \; 5 < M_w \le 7 \end{cases} \quad (18)$$

τ (T)= s_1 (T) (19)

where Equation 15 represent Model (1) where $\varphi_{SS}^{G/B}$ is only period dependent. Equation 16 represent Model (2) where $\varphi_{SS}^{G/B}$ is period and M_w dependent. Equation 17 represent Model (3) where $\varphi_{SS}^{G/B}$is period and R_{rup} dependent. Equation 18 represent Model (4) where $\varphi_{SS}^{G/B}$ is period, M_w and R_{rup} dependent. Equation 19 represent the homoskedastic model for τ.

Table 3 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 10 through 13 show the behavior of the proposed models at four spectral periods (PGA, 0.2s, 1.0s, and 7.0s). The figures also show 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:

$\varphi_{SS}^{G/B}$ models: 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 R_{rup}-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 M_w-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 M_w-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 for large magnitude earthquakes (M_w=7). 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 the large variability observed at the surface site conditions (e.g., site amplifications, and topographic effects) were accounted for in the $\varphi_{S2S}^{G/B}$ components of variability.

Table 3. Regression coefficients used in the models shown in Equations 15 through 19

T (s.)	Model (1)		Tau Model	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	0.4512	0.3628	0.4608	0.4478	0.4927	0.4876

T (s.)	Model (1)		Tau Model	Model (2)			
	s_0^G	s_0^B	s_1	s_2^G	s_2^B	s_3^G	s_3^B
	[0.0029]	[0.0028]	[0.0016]	[0.0031]	[0.0030]	[0.0138]	[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 3 (Cont'd). Regression coefficients used in the models shown in Equations 15 through 19

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	0.6239	0.5127	0.5001

T (s.)	Model (3)			
	s_7^G	s_7^B	s_8^G	s_8^B
	[0.0097]	[0.0098]	[0.0039]	[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 3 (Cont'd). Regression coefficients used in the models shown in Equations 15 through 19

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]

T (s.)	Model (4)							
	$M1R1^G$	$M1R3^G$	$M2R1^G$	$M2R3^G$	$M1R1^B$	$M1R3^B$	$M2R1^B$	$M2R3^B$
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]

The magnitude and distance at which there is a break in the linear relationship was smoothed across the different periods. Figures 10 and 11 shows the magnitude dependent models at four different periods for τ and $\varphi_{SS}^{G/B}$, respectively. 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 12 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 perform 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).

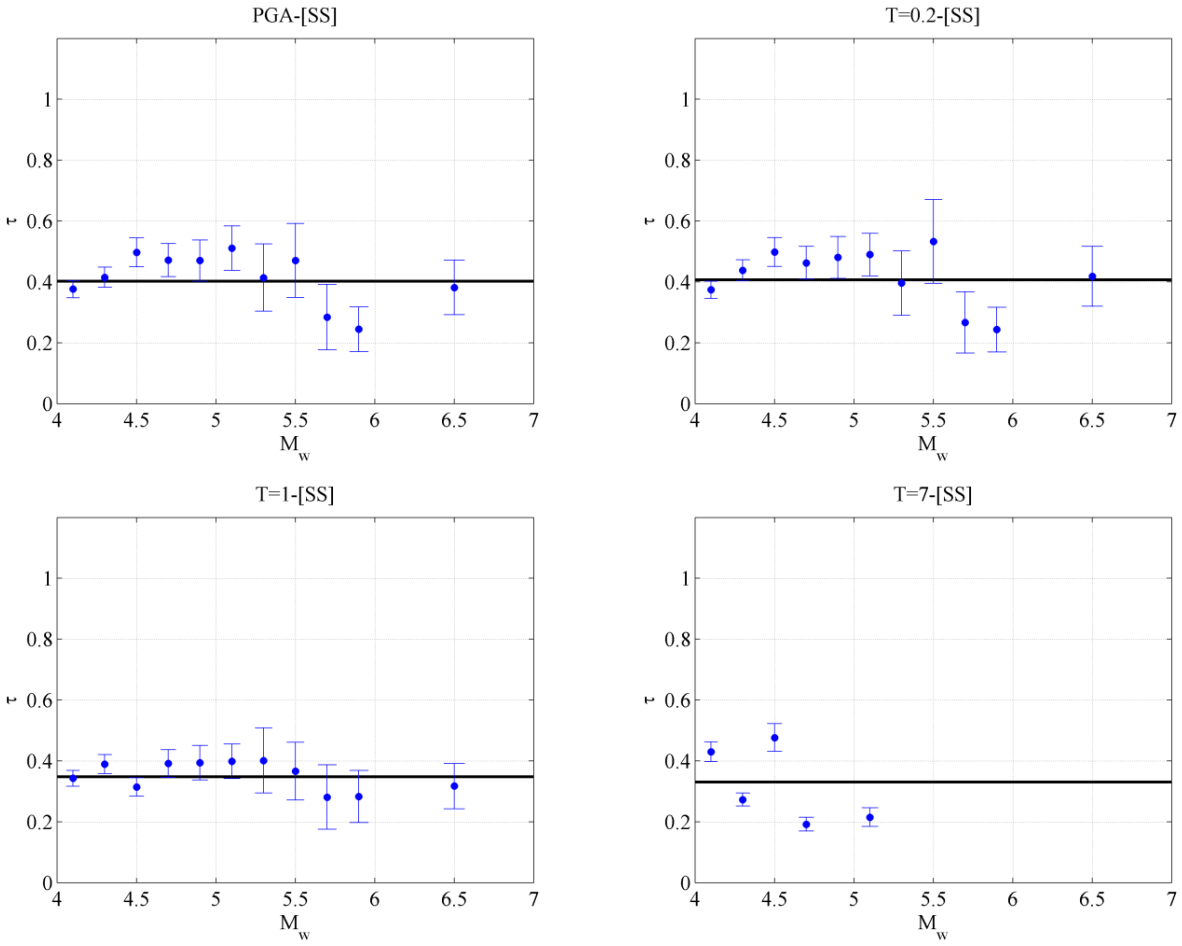


Figure 10. Models for τ from the site-specific formulation. The three models are magnitude and distance independent, but each was developed in conjunction with a specific φ model. The three models result in almost the same values. The three models are poorly constrained for T=7 seconds

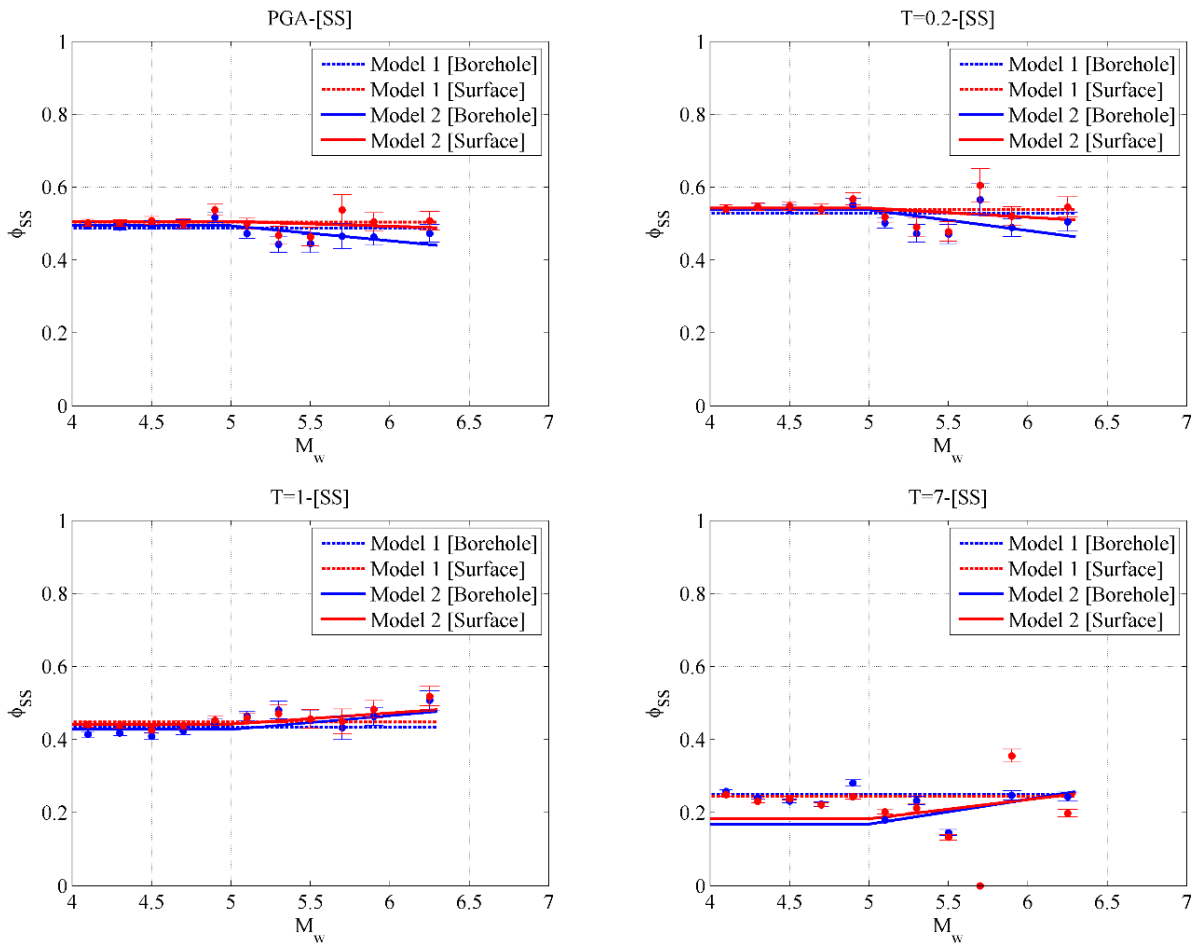


Figure 11. Models for $\varphi_{SS}^{G/B}$ Model 1 is a distance- and magnitude-independent model. Model 2 is a magnitude-dependent model. The latter model is poorly constrained for T=7 seconds. The circles show the standard deviation of site and event corrected residuals within a magnitude bin.

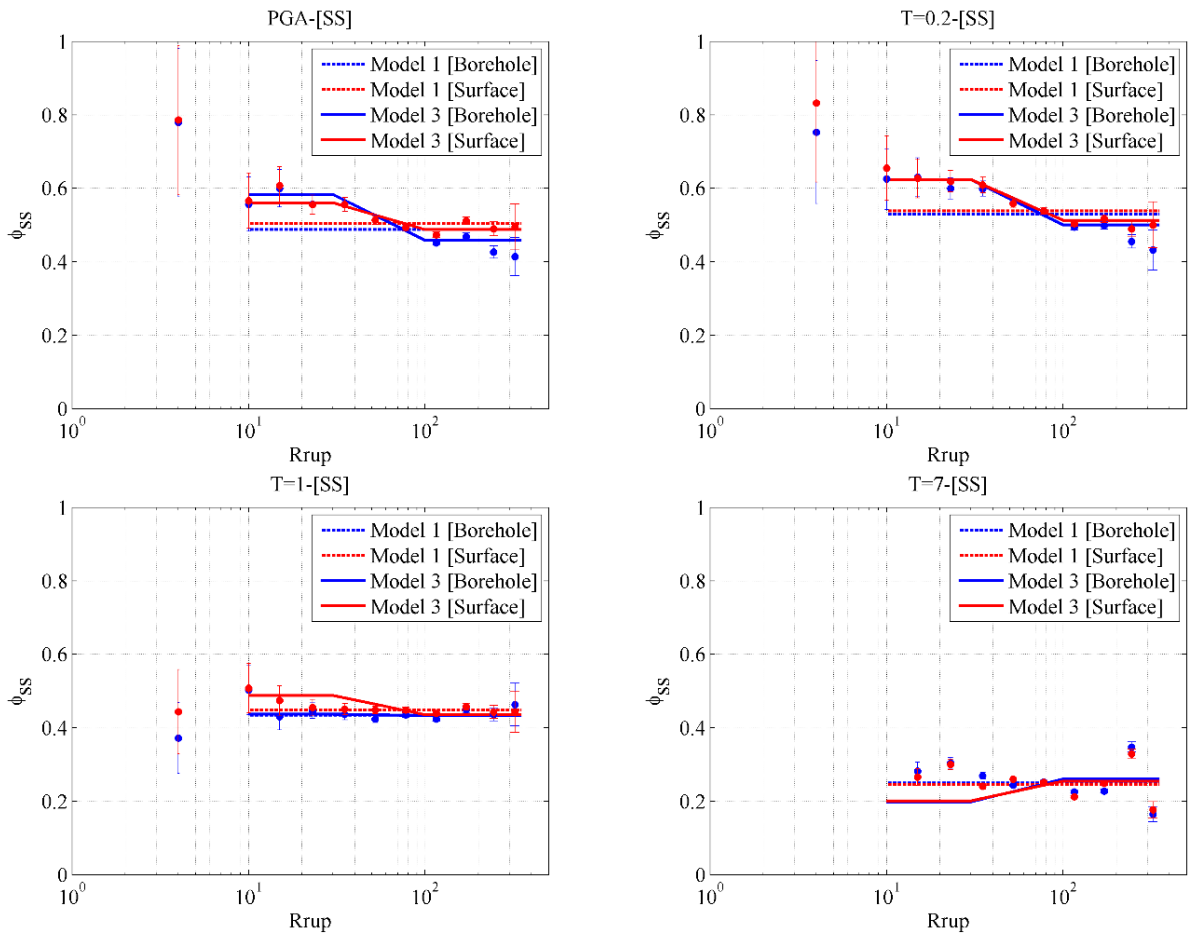


Figure 12. Models for $\varphi_{SS}^{G/B}$ Model 1 is a distance- and magnitude-independent model. Model 3 is a distnace-dependent model. The latter model is poorly constrained for T=7 seconds. The circles show the standard deviation of site and event corrected residuals within a distance bin.

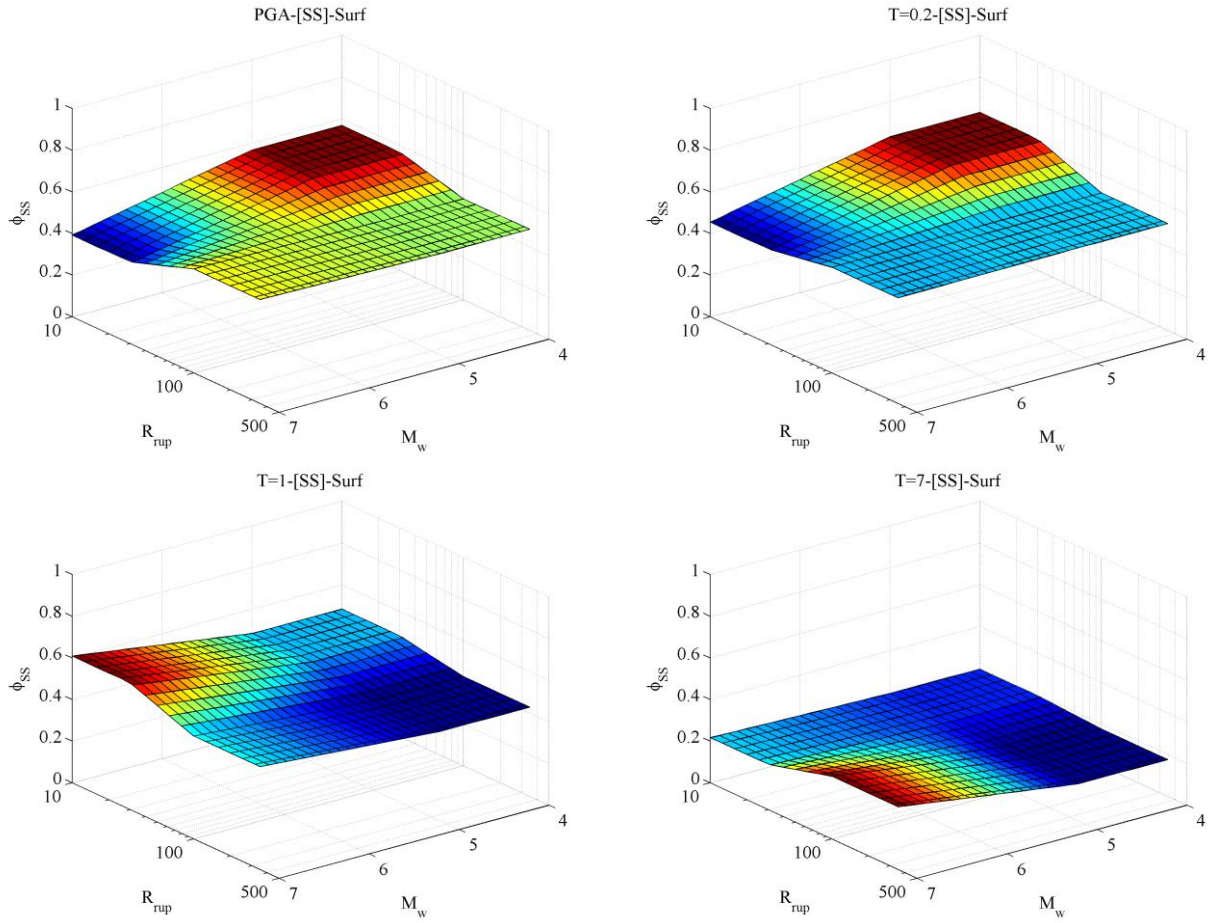


Figure 13. Model 3 for $\varphi_{SS}^{G/B}$ which is a distance- and magnitude-dependent model.

Azimuth Dependence

The values of $\varphi_{ss,s}^{G/B}$ (Equation 7) at some stations are much higher or lower than the $\varphi_{SS}^{G/B}$ estimated from the models. This is important because the value of $\varphi_{ss,s}^{G/B}$ adopted for a site-specific PSHA can have a major impact on the resulting hazard curves. Figure 14 shows the histograms of $\varphi_{ss,s}^{G}$ for stations that recorded 5 motions or more. The figure also presents the mean and mean$\pm$standard deviation of $\varphi_{SS,S}^{G}$ and the distance and magnitude independent φ_{ss}^{G} computed from the entire dataset. As expected, the average $\varphi_{ss,s}^{G}$ for all stations nearly equals to 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. These values are reduced when the residuals are corrected for azimuth dependency (as discussed later).

In this section we focus the discussion on stations with either high or low $\varphi_{ss,s}^{G/B}$ values to better understand why there is such a large variability. The discussion is presented as observations inferred from selected plots:

- The site-specific within-event residual component ($\delta W_{es,0}^{G/B}$) is azimuth dependent for many stations characterized by high $\varphi_{ss,s}^{G/B}$ For example, Figure 15-a shows residuals for station AKTH17; it is clear from the figure that $\delta W_{es,0}^{G}$ 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 sources have certain characteristics that make 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 $\varphi_{ss,s}^{G/B}$ values, the azimuth dependency of $\delta W_{es,0}^{G/B}$ was removed and $\phi_{ss,s}$ was re-estimated. This is similar (but not exactly the same) to considering path- and source 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 each of the azimuth ranges. $\varphi_{ss,s}^{G/B}$ is then recalculated for each station using these modified residuals. On average, the modified $\varphi_{ss,s}^{G}$ was 13% less than its original estimated value for motions recorded at the 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 applied (e.g., station AOMH07; Figure 15-b).

- Japan is characterized by a very complex tectonic setting that might result in significant biases for some source-path combinations compared to the median predictions (e.g., active volcanoes, faults, and large mountains). Probably as a result of this complexity, some stations that recorded many earthquakes show a $\varphi_{ss,s}^{G/B}$ that is larger than $\varphi_{ss}^{G/B}$ (e.g., station FKSH14; Figure 15-c). In this case, the recorded motions cover almost all possible azimuth and distance ranges, yet $\varphi_{ss,s}^{B}$ is relatively high. On the other hand, the residual plots show clear biases for some azimuth-distance bins (Figure 15-c); these biases led to the high $\varphi_{ss,s}^{G}$.

- In many cases, the recordings at stations characterized with low $\varphi_{ss,s}^{B/G}$ cover narrow azimuth ranges (e.g., station YMGH08, Figure 15-d), or narrow azimuth and distance ranges (e.g., station HDKH02, Figure 15-e) ranges. This means that the single-path and/or source effects are included in the estimated $\varphi_{ss,s}^{B/G}$, 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 $\delta S2S_S^{B/G}$. This shows that that estimating $\varphi_{ss,s}^{B/G}$ using motions recorded from a mainshock and its subsequent aftershocks may result in an under estimation of $\varphi_{ss,s}^{B/G}$.

The observations in this section point to the fact that estimates of $\varphi_{ss,s}^{B/G}$ could be contaminated by single-path and single-source measurements. Hence, the relative location of the sources and the distribution of source-to-site azimuths should be considered when estimating $\varphi_{ss,s}^{B/G}$ at a station.

We conducted additional analyses to check the effect of azimuth on $\varphi_{ss,s}^{B/G}$ 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 is horizontal, 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 16). We distinguished between stations characterized by “High”, “Neutral” and “Low” $\varphi^{G}_{ss,s}$ values. It is clear from the figure that on average “Low” $\varphi^{G}_{ss,s}$ 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” $\varphi^{G}_{ss,s}$ recorded motions from a narrower azimuthal range compared to the “Neutral” and “High” stations.

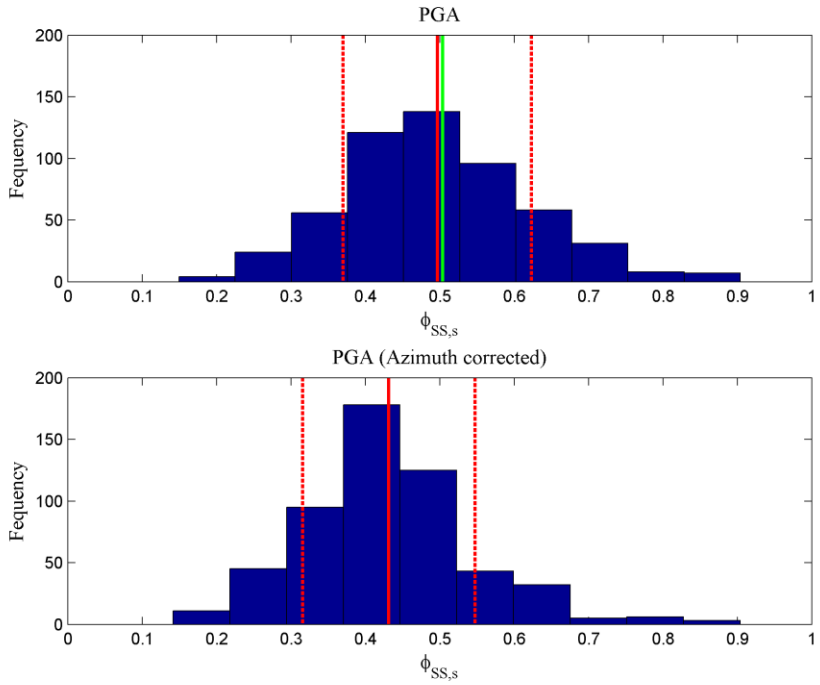
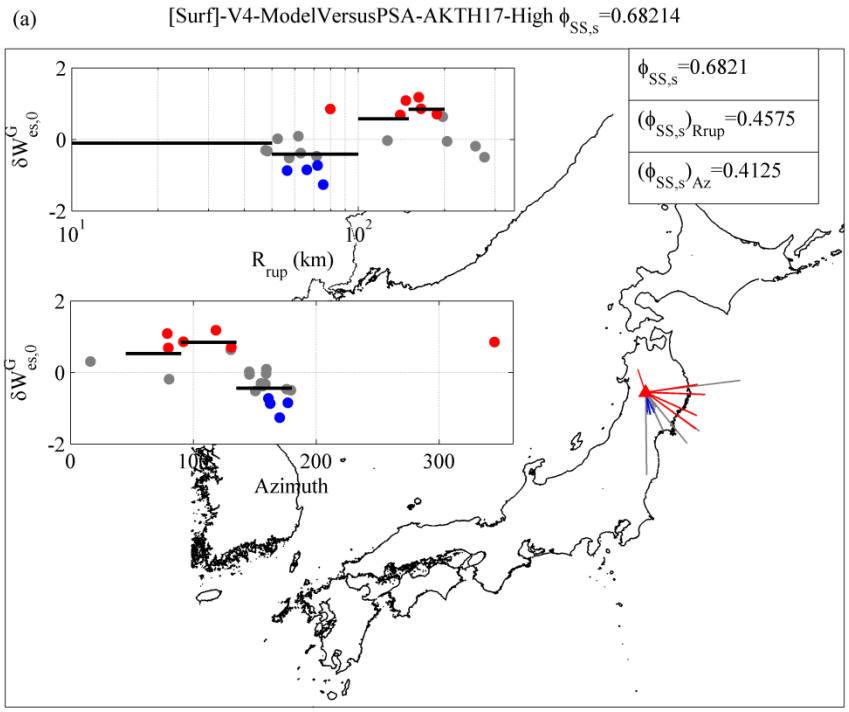
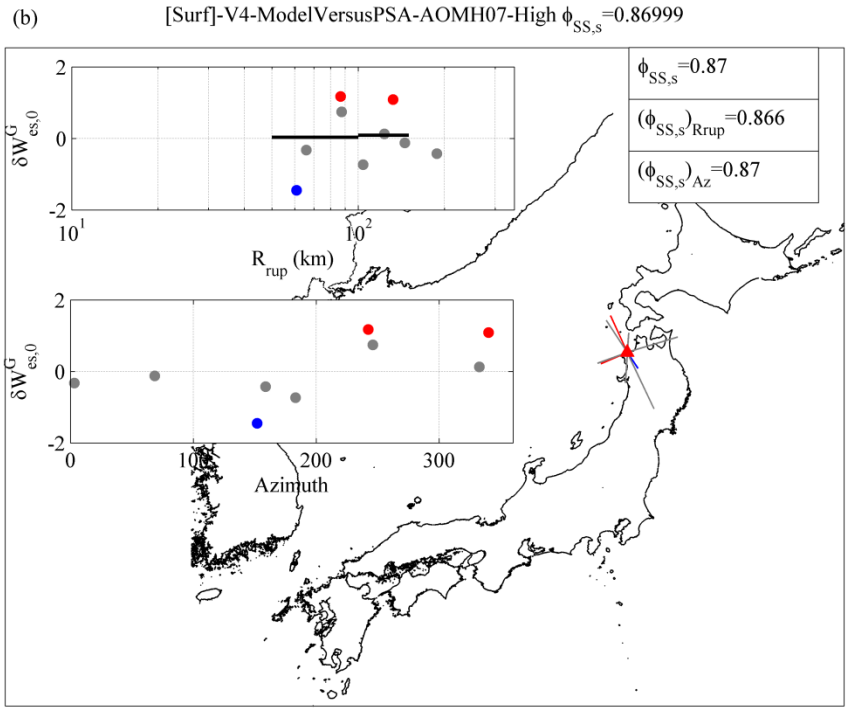


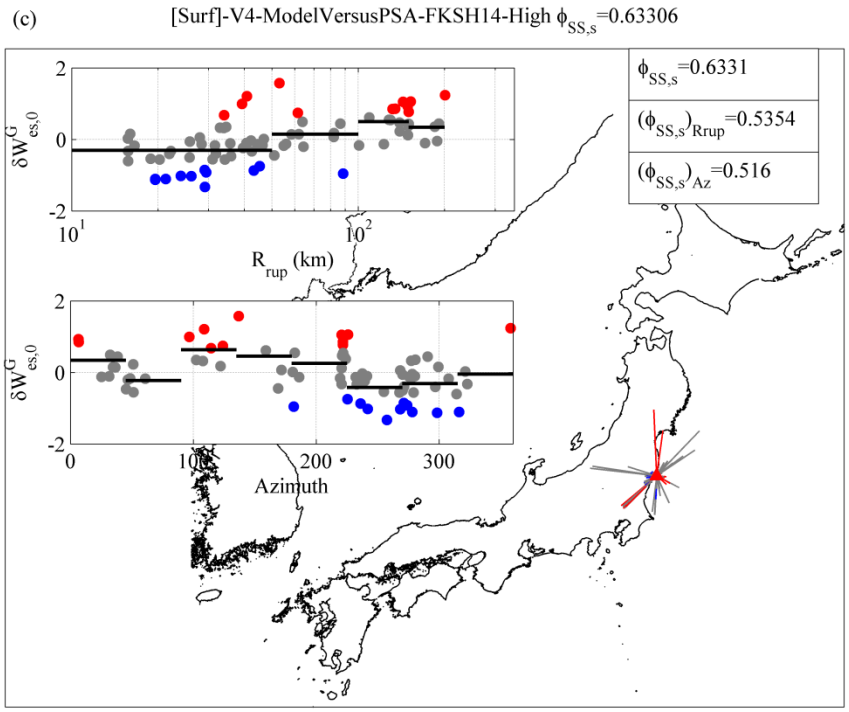
Figure 14. Histogram of $\varphi^{G}_{ss,s}$ values. Top: non-corrected data. Bottom: residuals are corrected for azimuthal dependency. The red lines show the average and one standard deviation of $\varphi^{G}_{ss,s}$; the green line is the value of φ^{G}_{ss}. Data shown for PGA; similar results are observed at other oscillator periods.



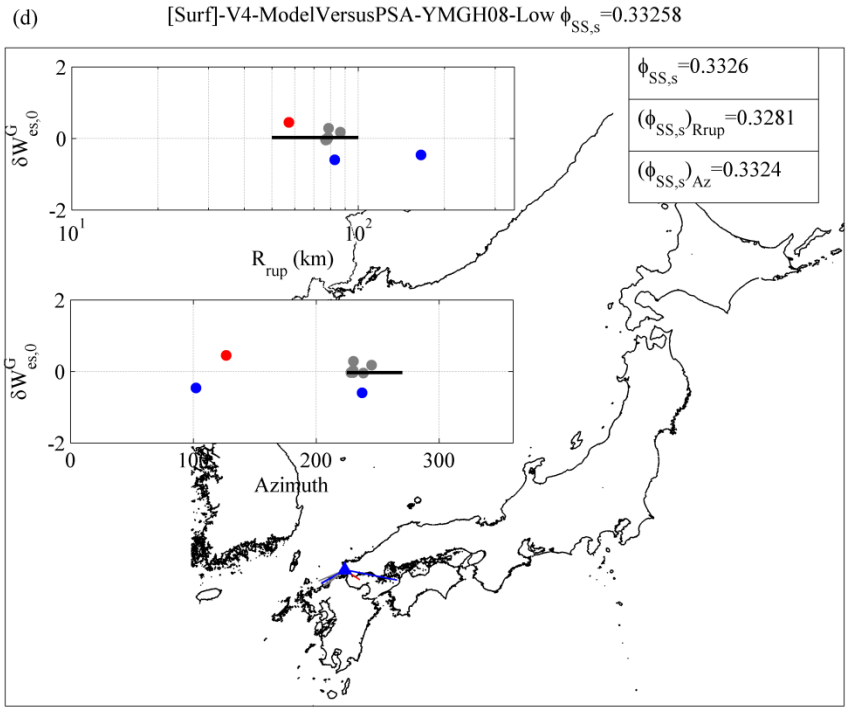
AKTH17-PGA-Surface: example of a station where the bias of $\delta W^{G}_{es,0}$ versus azimuth is clear. The plots show that records from a specific region is particularly biased.



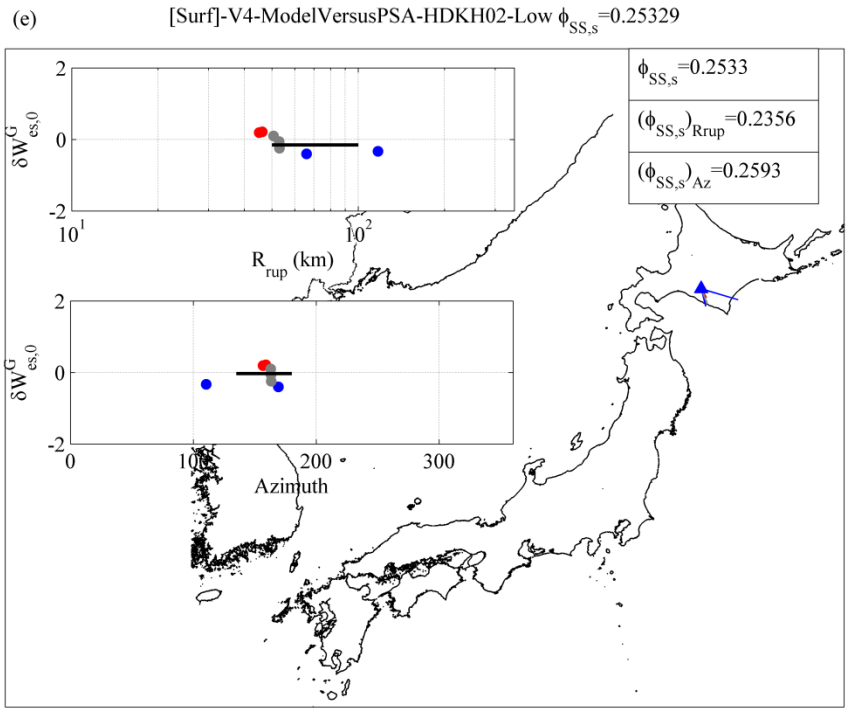
AOMH07-PGA-Surface: example of a case where it was not possible to remove the $\delta W^{G}_{es,0}$ bias with azimuth because very few motions were recorded at each azimuth bin.



FKSH14-PGA-Surface: example of a station with many motions from a wide range of azimuth and R_{rup}, yet it has a high $\varphi^{G}_{SS,S}$ due to the complexity of the tectonic environment in Japan.



YMGH08-PGA-Surface: example of station that recorded motions from limited azimuths. This could be a potential reason as of why the $\varphi^{G}_{SS,S}$ value at this station is low.



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

Figure 15. Each panel shows the location of a recording station on the Japanese map. Lines showing the different source to site paths travelled by the motions recorded at that station are added. The two subplots in each panel show the $\delta W^{G}_{es,0}$ or $\delta W^{B}_{es,0}$ (estimated for a specific spectral period at either surface or borehole) versus R_{rup} and azimuth. The solid lines are average values for bins of azimuth or distance.

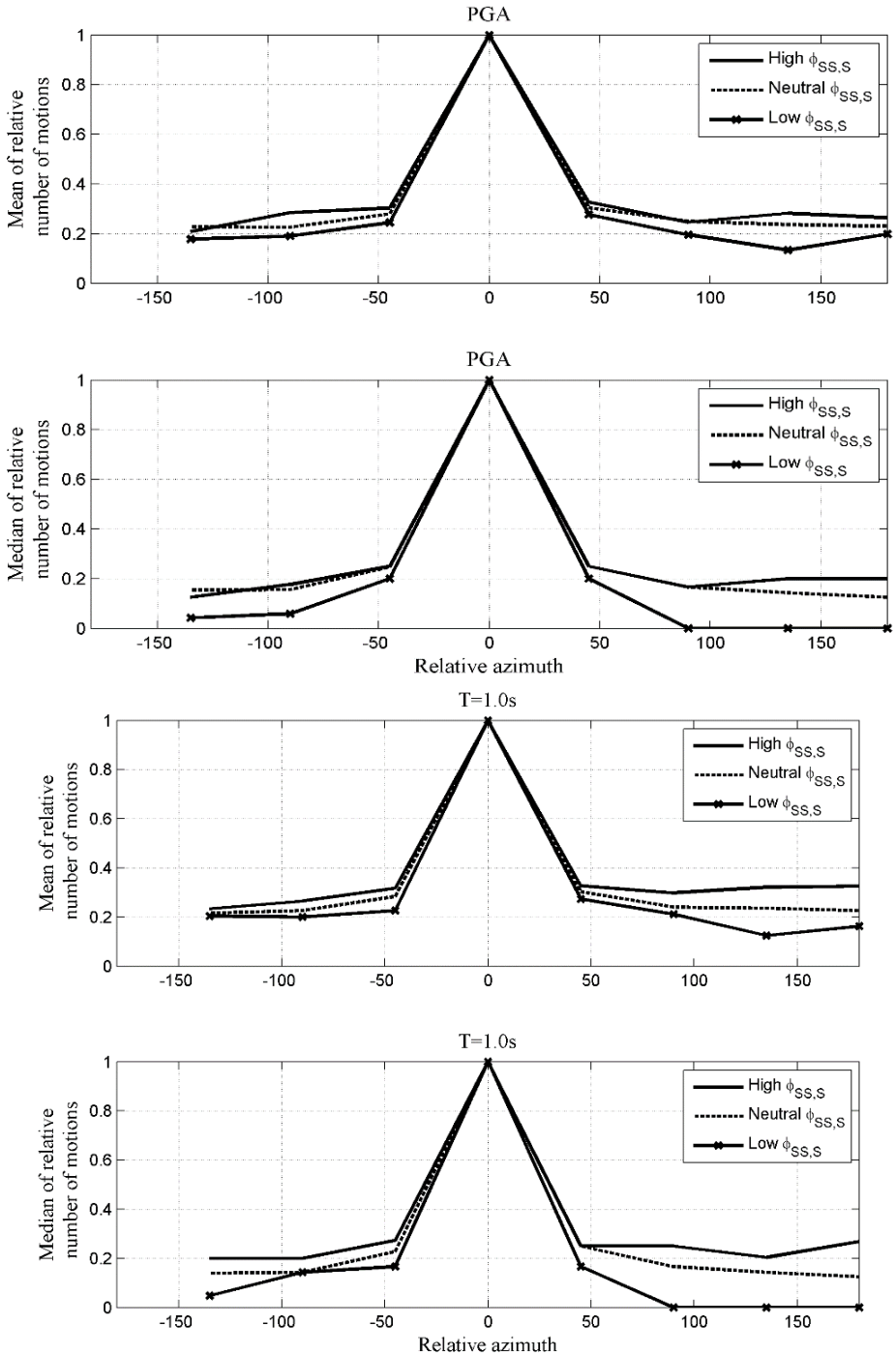


Figure 16. Plots of relative azimuth versus a normalized number of motions for stations with High, Medium, and Low $\varphi^{G}_{ss,s}$ values. Each line represents the relative number of records within a given azimuth range. The azimuth range is measured from the azimuth bin at which there is the largest number of records. Separate lines are plotted as averages of stations with high, neutral, and low values of $\varphi^{G}_{ss,s}$. For each oscillator period, top is for the average and bottom is for median number of relative motions

Shear Wave Velocity and Kappa Transfer Function Dependence

In principle, the site term should account for any local site effects; however, some local effects can affect the value of $\varphi_{ss,s}^{G}$ as well. The scatter plots of $\varphi_{ss,s}^{G}$ and azimuth corrected $\varphi_{ss,s}^{G}$ versus V_{s30} are shown in Figures 17a and 17b, respectively. 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. We also evaluate the trends of $\varphi_{ss,s}^{G}$ with respect to the high-frequency attenuation implicit in the borehole-to-surface transfer function. This parameter, referred to as transfer-function kappa (κ_{TF}) was computed from the high-frequency slope of the transfer function (Cabas, personal communication). The values of $\varphi_{ss,s}^{G}$ at various stations are plotted as a function of the kappa transfer function values in Figure 18. The same, yet weaker, trends are observed as for V_{s30}, which is no surprise since kappa transfer function and V_{s30} are correlated.

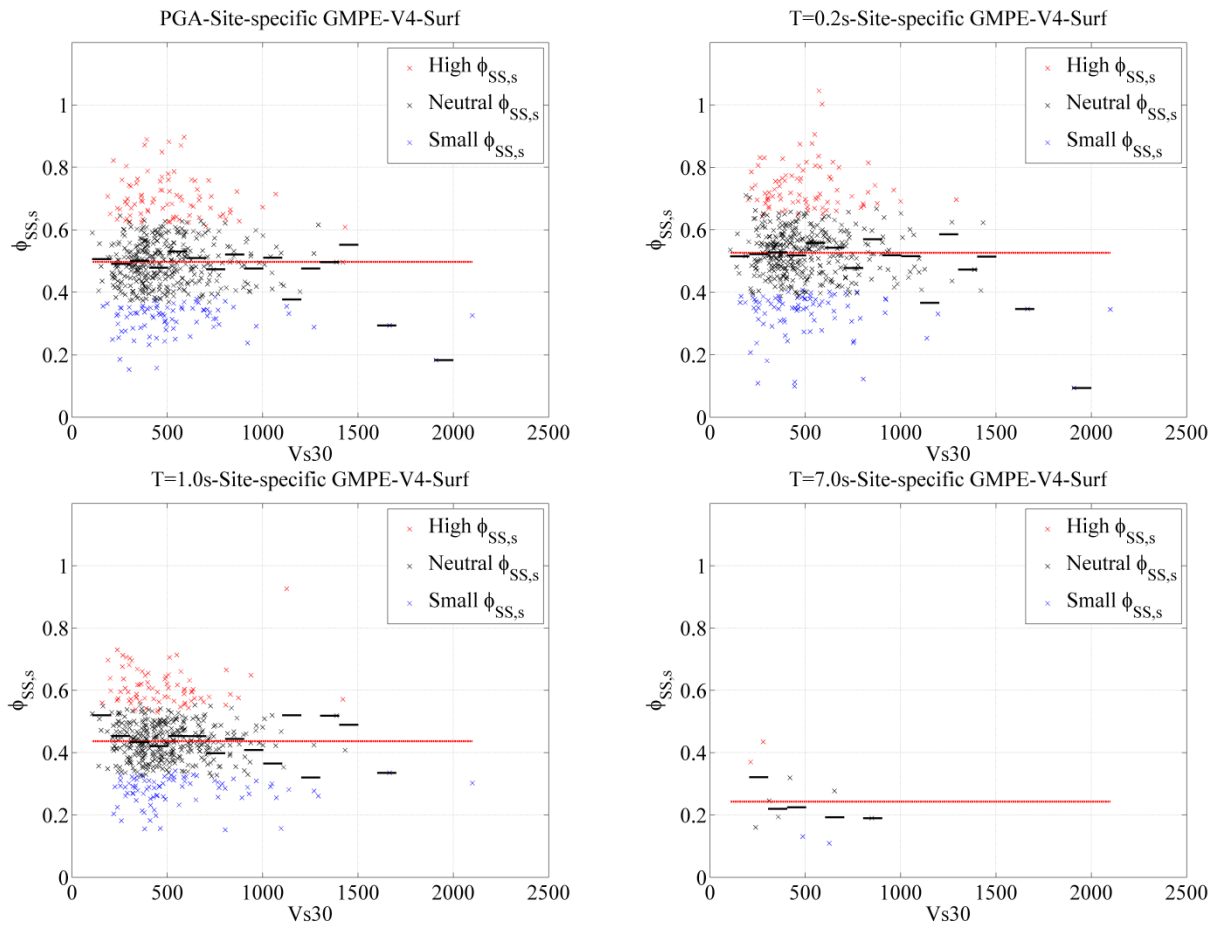


Figure 17a. Scatter plots of $\varphi^{G}_{ss,s}$ versus V_{s30} for the different stations in the KiK-net network. The red line is the average value of all data points. The black lines are the average values at specific V_{s30} bins

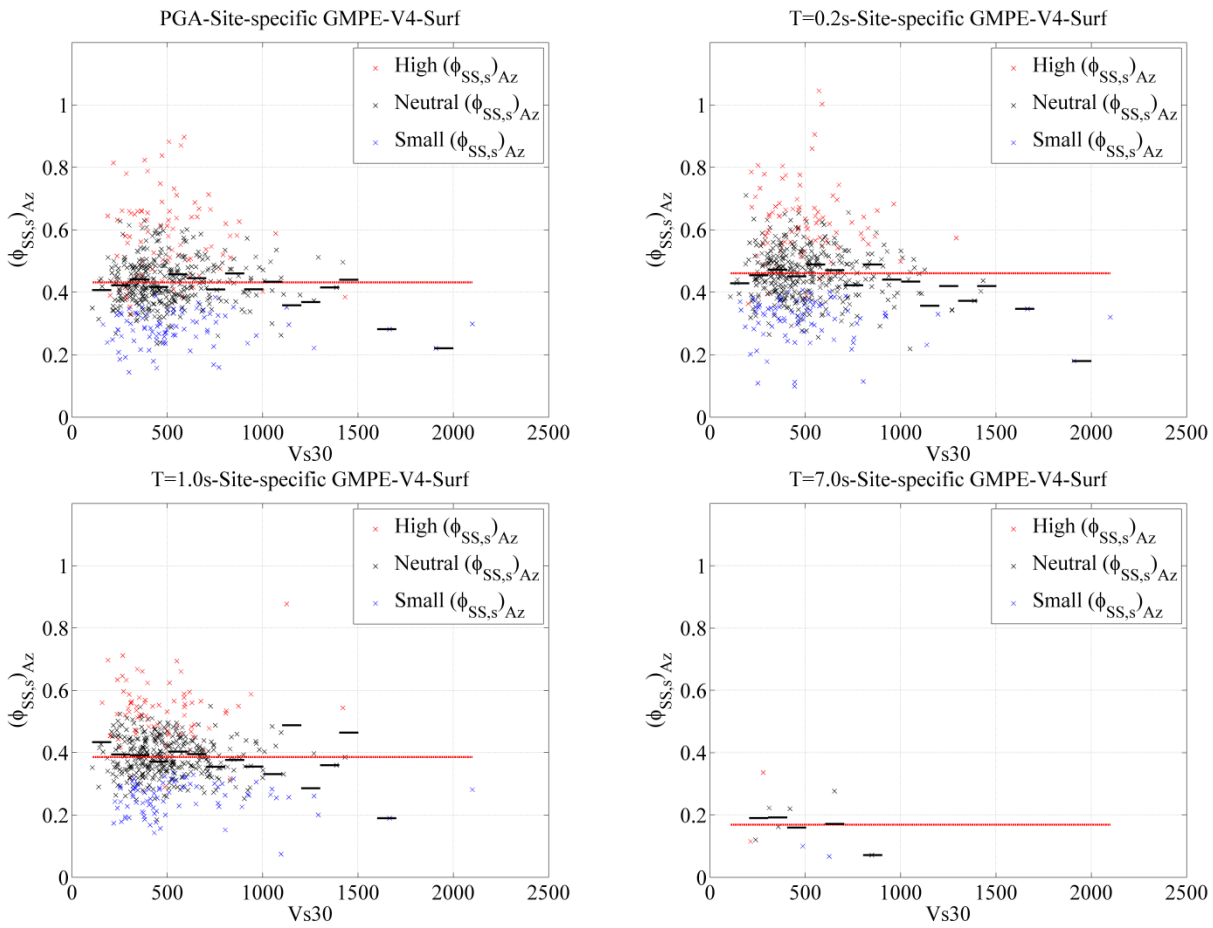


Figure 17b. Scatter plots of $\varphi^{G}_{ss,s}$ corrected for azimuth versus V_{s30} for the different stations in the KiK-net network. The red line is the average value of all data points. The black lines are the average values at specific V_{s30} bins

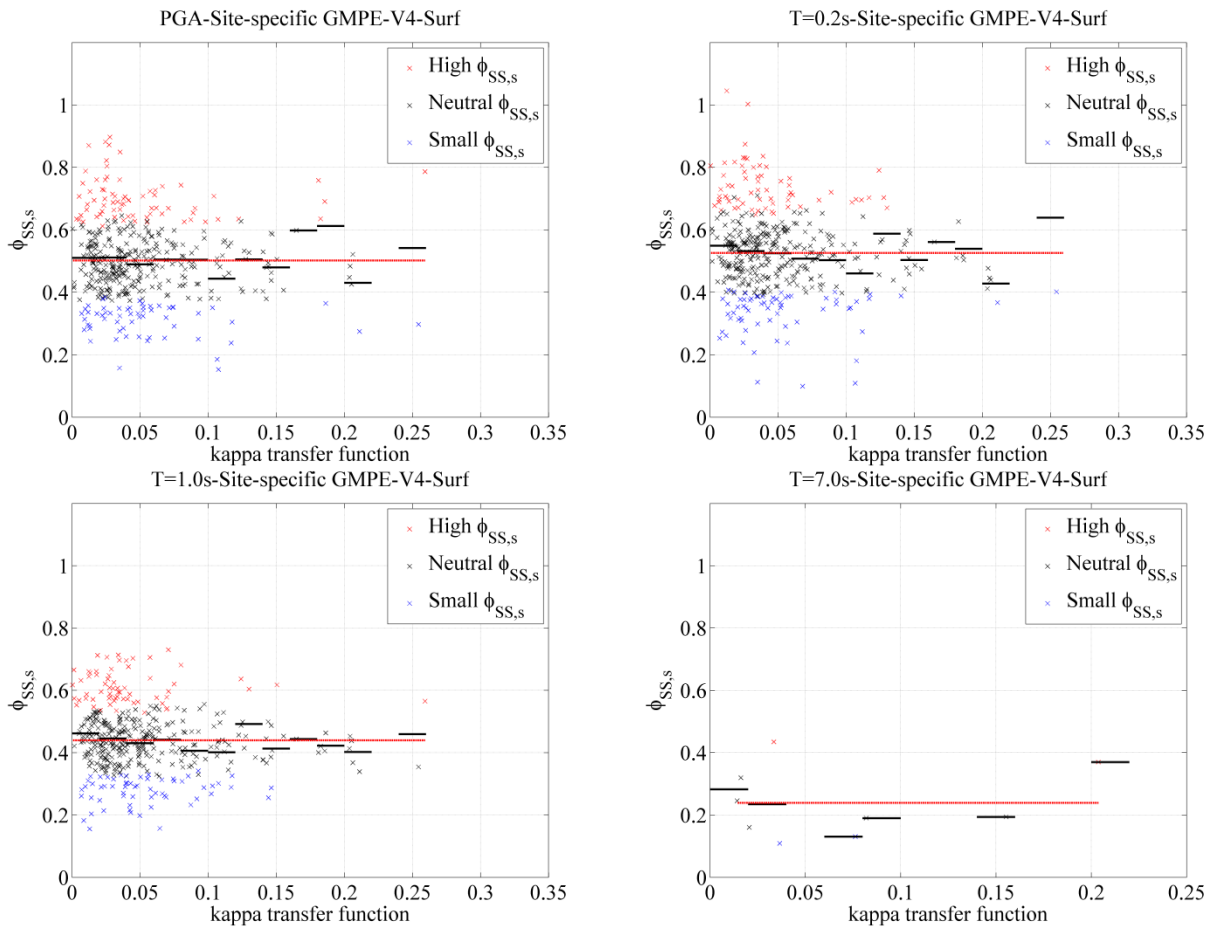


Figure 18a. Scatter plots of $\varphi^{G}_{ss,s}$ versus transfer function kappa for the different stations in the KiK-net network. The red line is the average value of all data points. The black lines are the average values at specific V_{s30} bins

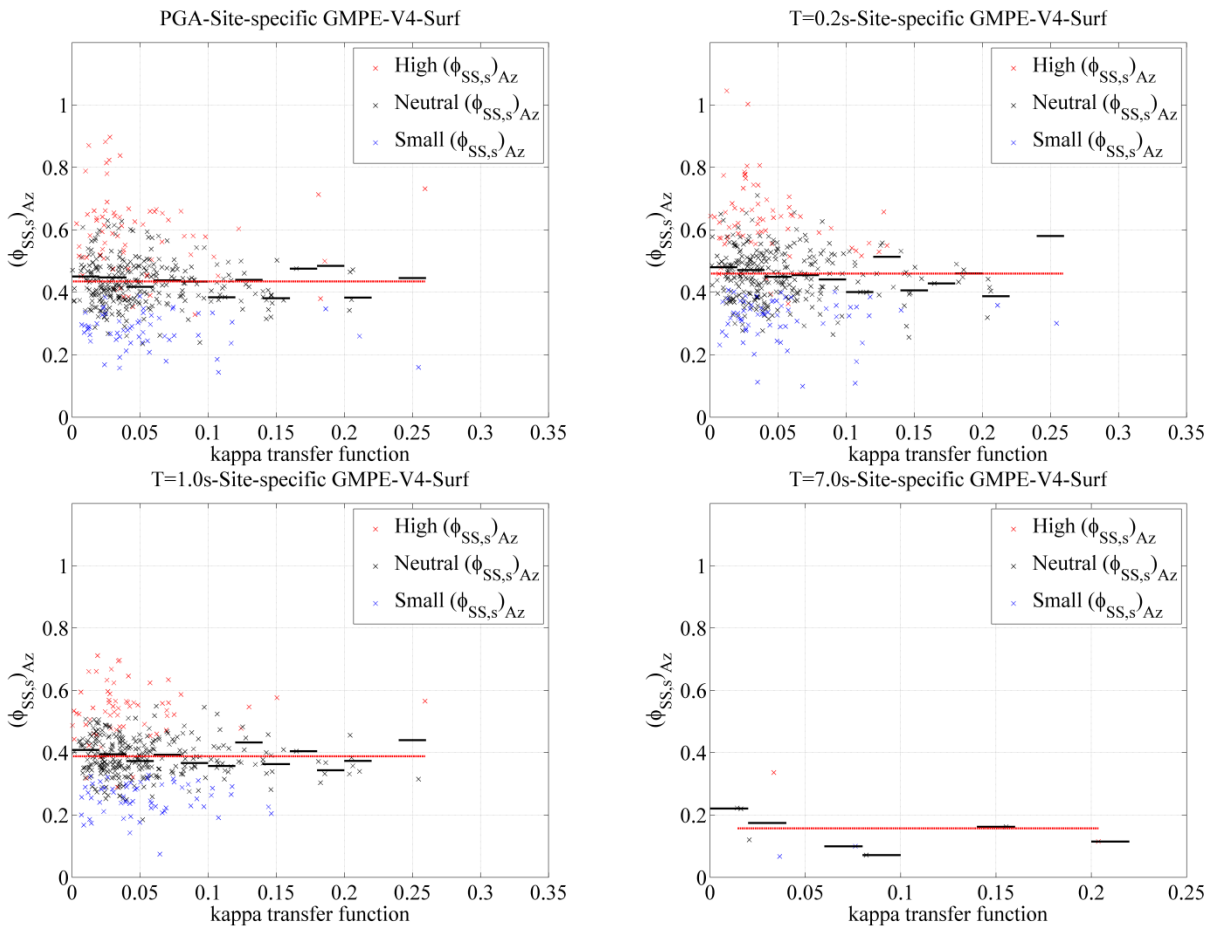


Figure 18b. Scatter plots of $\varphi_{ss,s}^{G}$ corrected for azimuth versus transfer function kappa for the different stations in the KiK-net network. The red line is the average value of all data points. The black lines are the average values at specific V_{s30} bins

Earthquake Metadata Dependence

We investigated the dependency of δB_e, $\delta W_{es}^{G/B}$, and $\delta W_{es,0}^{G/B}$ on different earthquake metadata. The δB_e versus M_{JMA} depth scatter plots showed the same bias shown in the δB_e versus the moment tensor depth scatter plots. Even if dip, M_{JMA}, and strike were not included in the functional form adopted by Dawood and Rodriguez-Marek (2014), the residuals did not show any clear trends versus δB_e, $\delta W_{es}^{G/B}$, or $\delta W_{es,0}^{G/B}$.

ANALYSIS OF RESIDUALS CONSIDERING THE EFFECTS OF BOREHOLE TO SURFACE AMPLIFICATION

In this section we consider a partition of residuals into repeatable site terms at the borehole, repeatable borehole-to-surface amplification terms, and the respective residuals. This partition is first formally introduced, and then the dependencies of different residuals components on various parameters (such as V_{s30} and kappa transfer function) are presented.

Partitioning of residuals

Each KiK-net station has two co-located instruments on the ground surface and below the surface. Hence, the site-specific empirical amplifications can be written as follows:

$$\delta S2S_S^{AMP} = \delta S2S_S^G - \delta S2S_S^B \quad (20)$$

$$\delta AMP_{esb,0} = \delta W_{es,0}^G - \delta W_{es,0}^B \quad (21)$$

where $\delta S2S_S^{AMP}$ is the site-specific empirical amplification residual component and δAMP_{esb}^0 is the remaining amplification components Hence, Equation 4 can be re-written as:

$$\delta W_{es}^G = \delta W_{es}^B + \delta S2S_S^{AMP} + \delta AMP_{esb,0} \quad (22)$$

Assuming that the residual components are uncorrelated, the total standard deviation at the ground surface can be defined as:

$$\sigma_{tot}^G = \sqrt{\tau^2 + (\varphi^B)^2 + (\varphi_{S2S}^{AMP})^2 + (\varphi^{AMP})^2} \quad (23)$$

where φ^B is the standard deviation of the δW_{es}^B term, φ_{S2S}^{AMP} is the standard deviation of the $\delta S2S_S^{AMP}$ term, and φ^{AMP} is the standard deviation of the $\delta AMP_{esb,0}$ term.

Shear Wave Velocity Dependence

The scatter plots of $\delta S2S_S^{AMP}$ versus V_{s30} at different spectral periods (Figure 19) show that the variability in $\delta S2S_S^{AMP}$ ($\varphi_{S2S}^{B/G}$) decreases as V_{s30} increases. This implies that there is less site-to-site variability in stiff rock sites than at softer soil sites. Also, $\delta S2S_S^{AMP}$ at the individual sites show a considerable decrease at long periods (Figure 19). Moreover, $\delta S2S_S^{AMP}$ seems to be biased low for high V_{s30} at short periods. φ^{AMP} does not show any clear trend with V_{s30} except at

a spectral period of 1.0 Sec. (Figure 20). On average, φ^{AMP} tend to decrease as spectral periods increases.

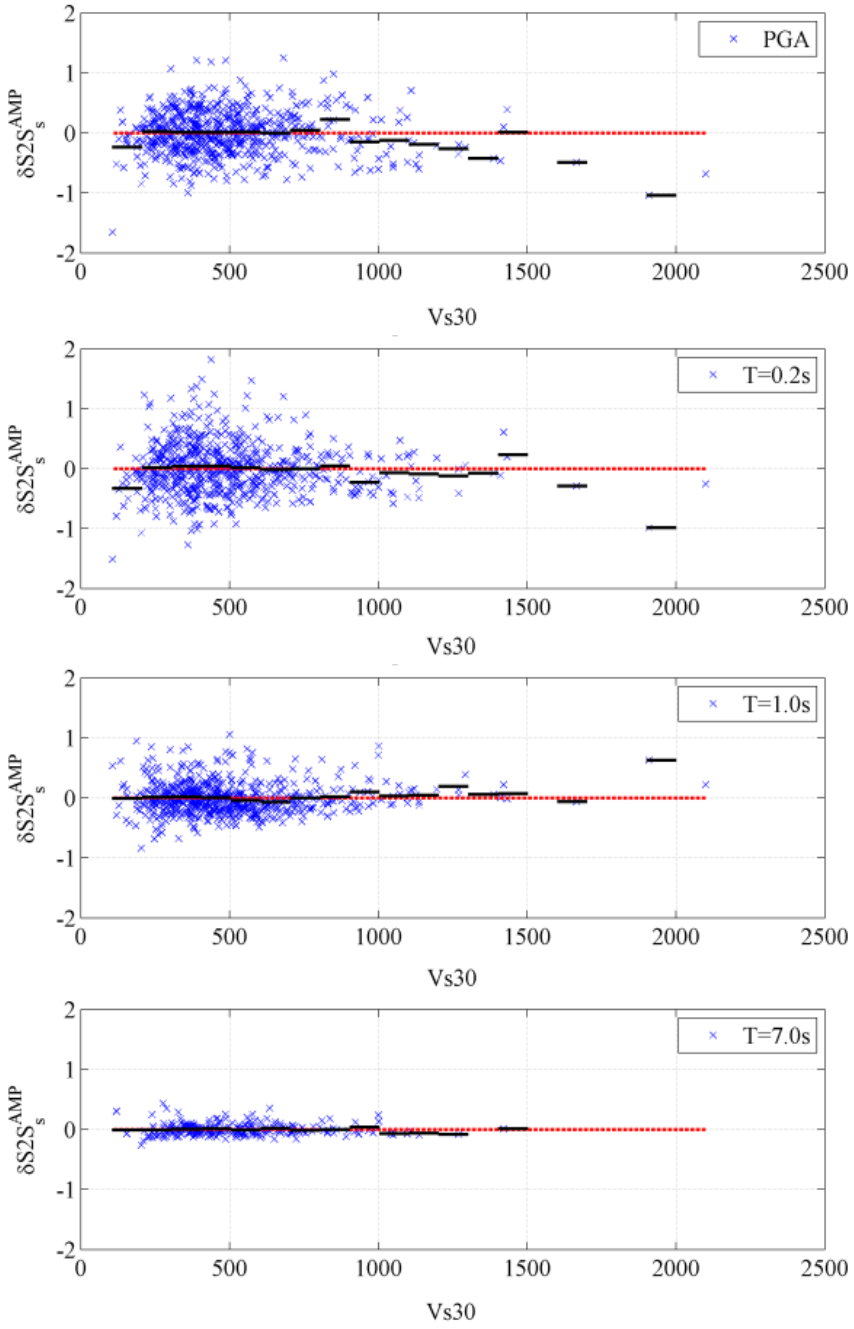


Figure 19. Scatter plots of $\delta S2S_s^{AMP}$ versus ground elevation for the different stations in the KiK-net network. The red line is the average value of all data points. The black lines are the average values at specific V_{s30} bins

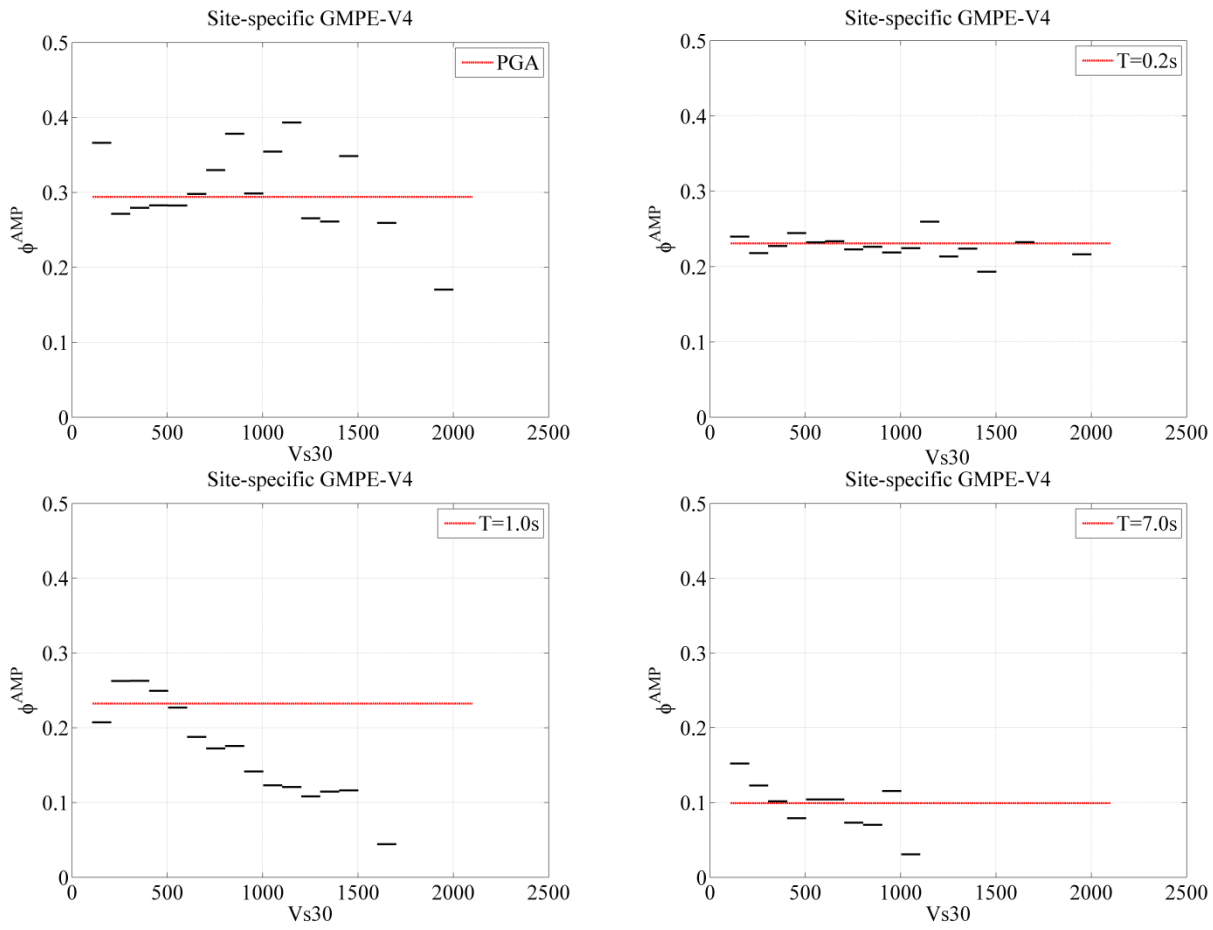


Figure 20. φ^{AMP} calculated at different V_{s30} bins (black lines) at different spectral periods. The red line is the average value of all data points. The standard deviation at a bin was calculated if there were at least 5 observations within that bin.

Kappa Transfer Function Dependence

In general, $\delta S2S_S^{AMP}$ and φ_{S2S}^{AMP} do not show clear bias with kappa transfer function (Figures 21 and 22) except at high kappa values for short spectral periods where a positive bias is observed. This bias could be attributed to the scarcity of stations at such high kappa values rather than a physical reason.

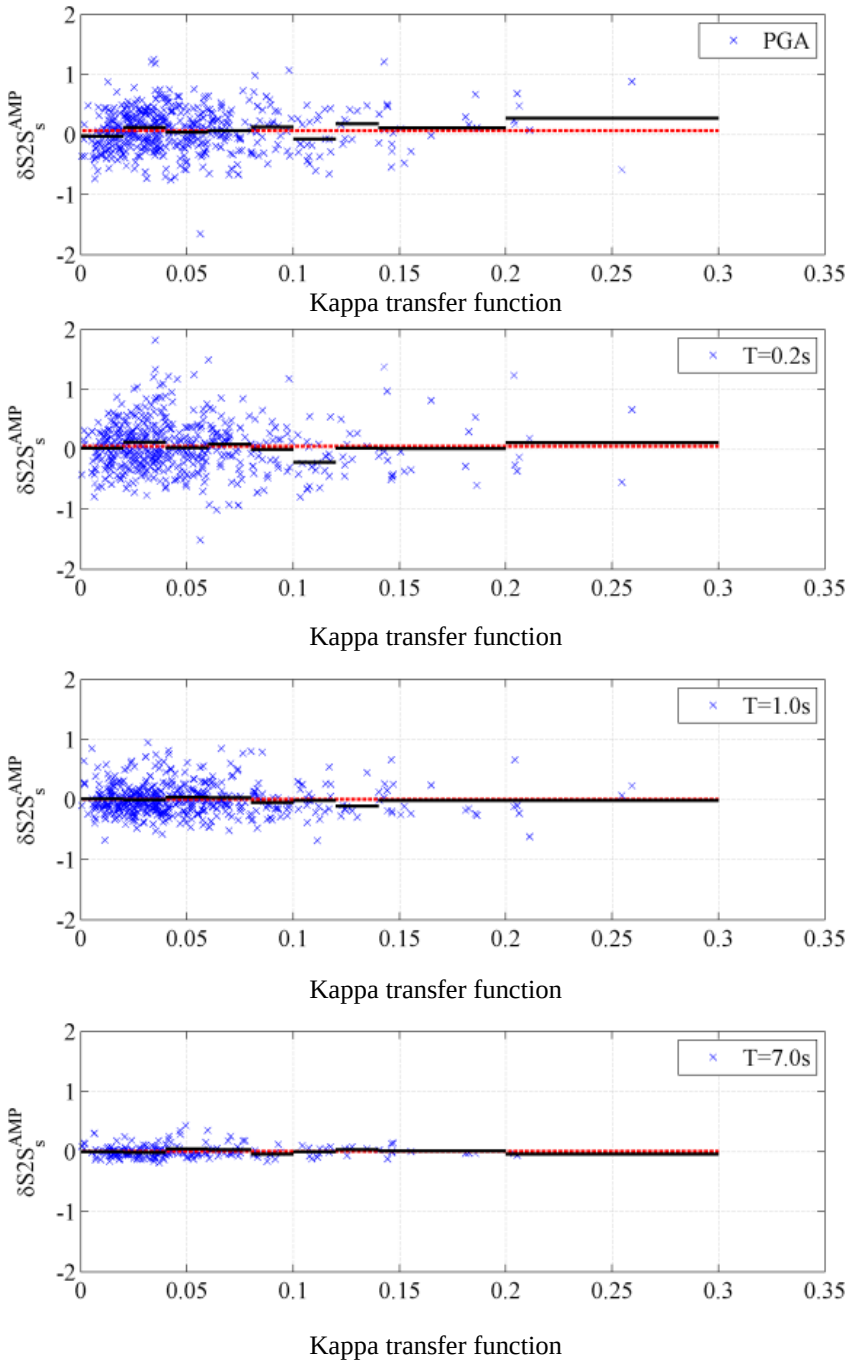


Figure 21. Scatter plots of $\delta S2S_s^{AMP}$ versus kappa transfer function of the different stations in the KiK-net network. The red line is the average value of all data points. The black lines are the average values at specific kappa bins

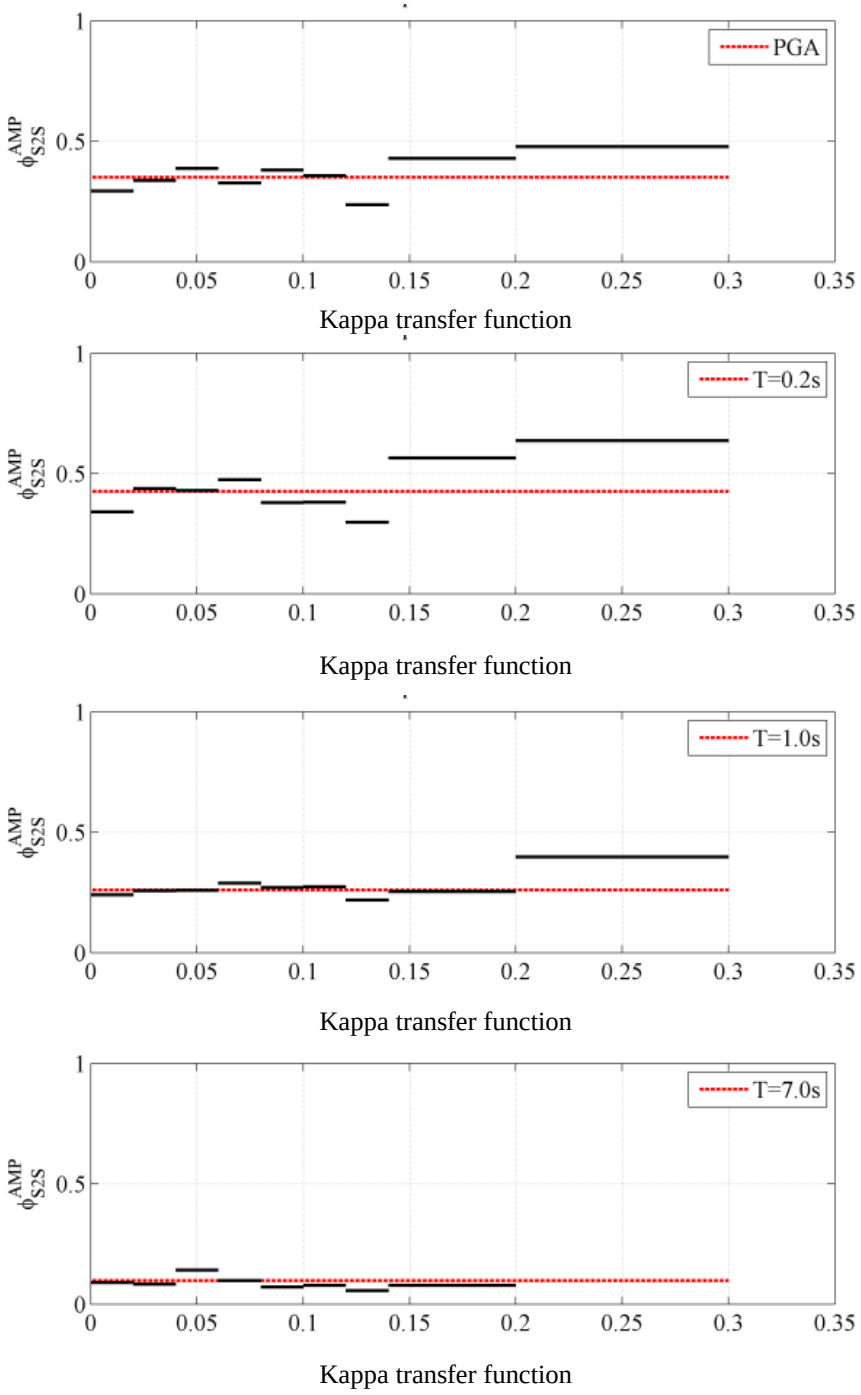


Figure 22. Scatter plots of φ_{S2S}^{AMP} versus kappa transfer function of the different stations in the KiK-net network. The red line is the average value of all data points. The black lines are the average values at specific kappa bins

FINDINGS AND CONCLUSIONS

This chapter presents the analysis conducted on the residuals and the different components of variability obtained from the GMPEs presented in Chapter 5 of this dissertation. We investigated the effect of smoothing the regression coefficients across spectral periods while developing GMPEs. We also proposed different models to predict the within-event and single station components of variability as functions of different parameters. The most relevant findings and conclusions from the investigations presented in this chapter are:

- Strong correlation coefficients were found between the residual components obtained from the GMPEs with smoothed and unsmoothed regression coefficients.
- The ergodic and site-specific regressions presented in Chapter 5 resulted in different median predictions. Differences in the median indicate that, despite the considerable size of the database, there is an interplay between event- and site-terms. Hence, regressions that include the site term should be used whenever possible, in particular when using poorly balanced databases.
- estimating $\varphi_{ss,s}^{B/G}$ using motions recorded from a mainshock and its subsequent aftershocks may result in an under estimation of $\varphi_{ss,s}^{B/G}$
- The relative location of the sources and the distribution of source-to-site azimuths should be considered when estimating $\varphi_{ss,s}^{B/G}$ at a station.
- On average, stations with "Low" $\varphi_{ss,s}^{G}$ recorded motions from a narrower azimuthal range compared to stations characterized with "High"$\varphi_{ss,s}^{G}$. Hence, the reduction in $\varphi_{ss,s}^{G}$ could be a result of single-path and/or single-source effects.

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.

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.

Dawood, H. M., and Rodriguez-Marek, A., 2014. Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations, to be submitted to Bull. Seism. Soc. Am.

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.

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.

Chapter 7: Findings and Conclusions

Findings and Conclusions

SUMMARY

The research presented in this dissertation aims to advance the state of the art in earthquake engineering and engineering seismology in general and non-ergodic PSHA in particular. To this end, four different problems were addressed in the different chapters of this dissertation. Each is summarized below.

Compilation of a Strong GM Database

The release of the ergodic assumption necessitates the availability of a large database of strong GM records, in particular one where many sources generated multiple earthquakes, many sites recorded multiple motions, and several motions travelled through the same source-to-site paths. Having sources, sites, and paths that are associated with multiple records allows for the computation of the repeatable terms that are a necessary component in developing non-ergodic GMPEs. The KiK-net database fits this model, but its large size makes processing the raw motions to be used in engineering applications a major challenge.

In this study, an automated ground motion processing protocol was developed and applied to the KiK-net strong motion database to obtain corrected ground motions for events between October 1997 and December 2011. The F-net seismic catalog was used to obtain the earthquake metadata for each ground motion recording. The approach used to process the records and obtain the necessary metadata was described in details in Chapter 3 of this dissertation. 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, signal-to-noise 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. Declustering algorithms were used to differentiate between dependent and independent events for the purpose of GMPE development. Earthquakes were also classified by earthquake type. Two different earthquake classification algorithms were used in this study.

Develop a Methodology to Develop Path-Specific GMPEs

An approach to include the path effects in GMPEs was presented in Chapter 4 of this dissertation. In this approach the region of interest is divided into 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 to validate the proposed methodology.

Develop a GMPE for Active Crustal Earthquakes

A subset of the KiK-net database records from active crustal (shallow and deep) earthquakes was used to develop ergodic and non-ergodic GMPEs. The development of these GMPEs is presented in Chapter 5. The functional form of a GMPE developed within the NGA-W2 project by Abrahamson et al. (2013) was used. The final subset of the KiK-net database chosen for this analysis consists of 11,574 six-components strong GM records from 790 active crustal earthquakes recorded at 624 stations. A nonlinear mixed-effects regression analysis using SAS was conducted for this analysis. The regression analysis was conducted in multiple consecutive steps. In each step a certain coefficient(s) was smoothed across all spectral periods and then fixed for the subsequent steps. Borehole and surface records were simultaneously regressed for and the same magnitude and distance scaling coefficients were set for both types of records. On the other hand, the coefficient that represent the linear site response term in the GMPE was assumed to be different for surface and borehole records.

A comparison between the output of the ergodic and non-ergodic regressions was conducted. The predictions of the developed GMPEs were compared with the predictions obtained from previously developed GMPEs.

Residual Analysis

A comprehensive residual analysis of the residuals obtained from Chapter 5 of this dissertation was conducted. This residual analysis aimed to investigate any potential biases in the different residual components and components of variability. Models to predict some of the components of variability as functions of different parameters were proposed. Also, stations

characterized by high a low variability were studied to understand why they are characterized with such high or low variability.

FINDINGS AND CONCLUSIONS

The major findings obtained in this study can be summarized as follows:

- The pseudo-spectral accelerations (PSA) obtained from the automated protocol presented in this study were compared with the PSA reported by the NGA-W2 project for the GM records in common between the two studies. This comparison showed a good agreement which proved the viability of using the proposed automated protocol to process the KiK-net GM records.

- The corner frequencies obtained from the automated algorithm were found to be larger than the corner frequencies chosen by the NGA-W2 project in most cases. This means that the usable frequency bandwidth of the automatically processed motions is narrower than the manually processed motions (by NGA-W2). Hence, in processing a very large database of strong GM records there is a tradeoff between manually processing the motions one by one (which is a tedious process) to get a larger usable bandwidth and automatically processing it (a faster process) but getting a narrower usable frequency bandwidth.

- The two earthquake classification algorithms showed that the earthquakes recorded by the KiK-net network are mainly active shallow crustal or subduction (interface and intraslab) type earthquakes.

- The different declustering algorithms adopted on the KiK-net database resulted in non-Poissonian declustered catalogs. When the resulting catalog was forced to be Poissonian by altering the spatial and temporal windows used to flag the different events, the catalog was found to be incomplete (i.e., b-value was much smaller than 1.0). This means that there is a conflict –at least for the F-net catalog– between obtaining a Poissonian and a complete catalog.

- The attenuation rate in the regions that contained volcanoes was found to be higher than the regions without volcanoes. This difference is statistically significant at a 95% confidence level for most magnitudes and oscillator periods. This finding confirms findings from previous studies and follows our understanding of the anelastic attenuation phenomena.

- The attenuation rate for non-volcanic regions (forearc and back-arc regions) was found to be the same at a 95% confidence level.

- The GMPE developed adopting the methodology presented in Chapter 4 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.

- By analyzing the residuals obtained from the study presented in Chapter 4, 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 non-ergodic PSHA (in the absence of soil nonlinearity). More care must be used in extrapolating information on path terms.

- The ergodic and site-specific regressions presented in Chapter 5 resulted in different median predictions. Differences in the median indicate that, despite the considerable size of the database, there is an interplay between event- and site-terms. Hence, regressions that include the site term should be used whenever possible, in particular when using poorly balanced databases.

- The values of the different components of variability obtained from the ergodic and site-specific formulations are very similar; however, at some period ranges there are some differences.

- The intra-event component of variability at the ground surface (φ^{G}) has a peak at around 0.15 seconds. This peak is less obvious in the borehole component (φ^{B}), suggesting that the increased variability is due to near surface effects.

Appendix A: Electronic Supplement For “A Flatfile For The KiK-Net Database Processed Using An Automated Protocol”

Submitted to EERI for publication in EARTHQUAKE SPECTRA

Appendix A

Electronic supplement

For

A Flatfile For The KiK-Net Database Processed Using An Automated Protocol

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

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 A.1 provides a list of these earthquakes. For the records from the earthquakes listed in Table A.1, 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 A.1 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 A.1 shows the distribution of M_w versus the different distance measures for the events listed in Table A.1.

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-

[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

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 A.1, 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 A.2 compares the M_w-area relationships for the events in Table A.1 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 A.3 (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 A.3 (b) for reverse events:

$$\log(\mathrm{AR}) = 0.0321(\mathrm{M_w} - 4)^{1.33}, \sigma_{\log} = 0.35. \quad \text{(A.1)}$$

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 A.4. 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 A.5 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 A.1 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.

REFERENCES

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

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

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

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

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

Kobayashi, T., Tobita, M., Koarai, M., Okatani, T., Suzuki, A., Noguchi, Y., Yamanaka, M., and Miyahara, B., 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.

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

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

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

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

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

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.

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

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

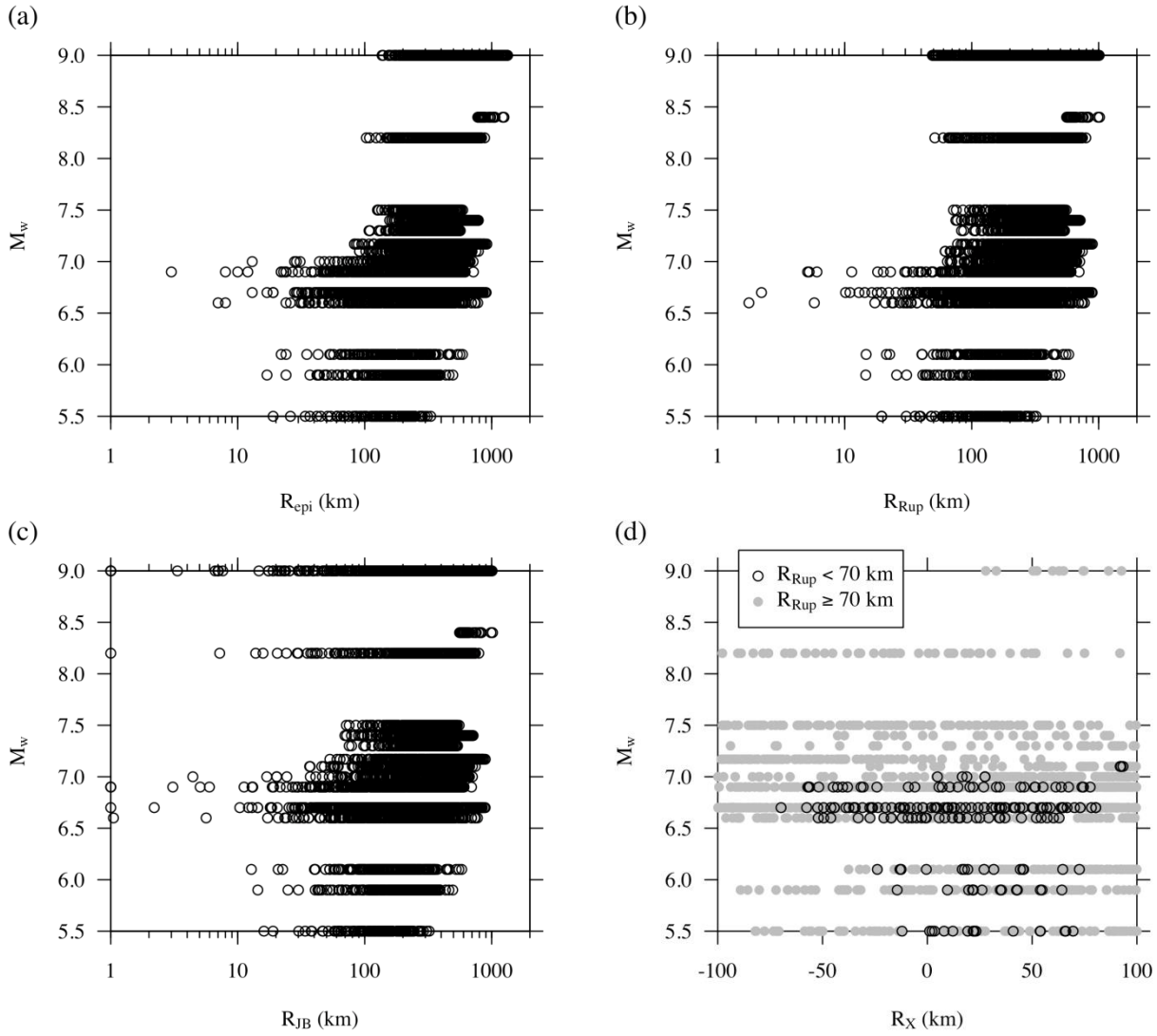


Figure A.1. The distribution of M_w versus: (a) R_{epi}; (b) R_{rup}; (c) R_{JB}; and (d) R_x, for the earthquakes shown in Table A.1. For log-axes, distances less than 1 km were rounded up to 1 km.

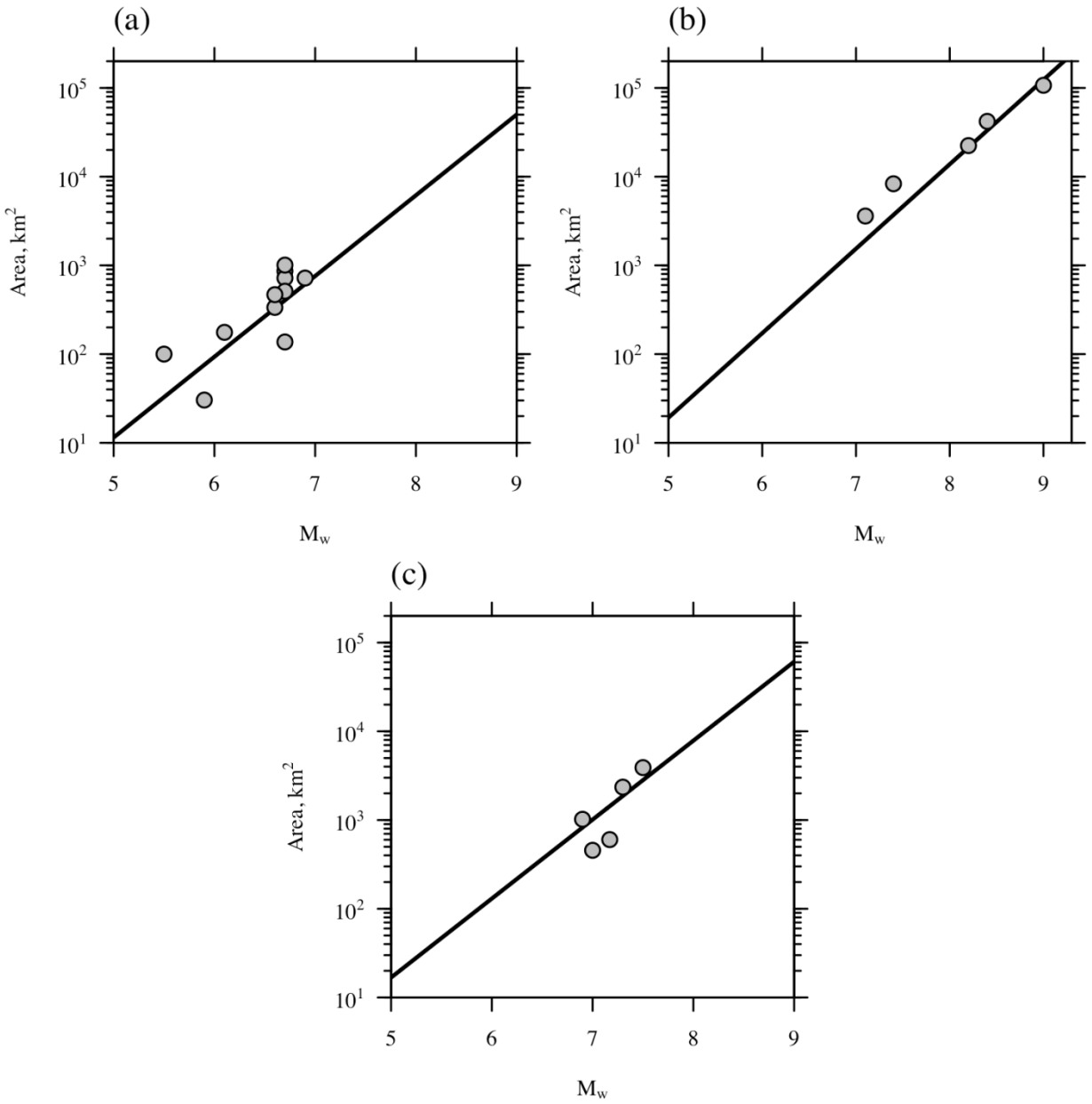


Figure A.2. M_w versus rupture area for the finite fault models in Table A.1; (a) compares the Wells and Coppersmith (1994) relationship for active crustal events, and (b) and (c) compare the Strasser et al. (2010) relationships to the finite fault models in Table A.1 for subduction zone interface and intraslab events, respectively.

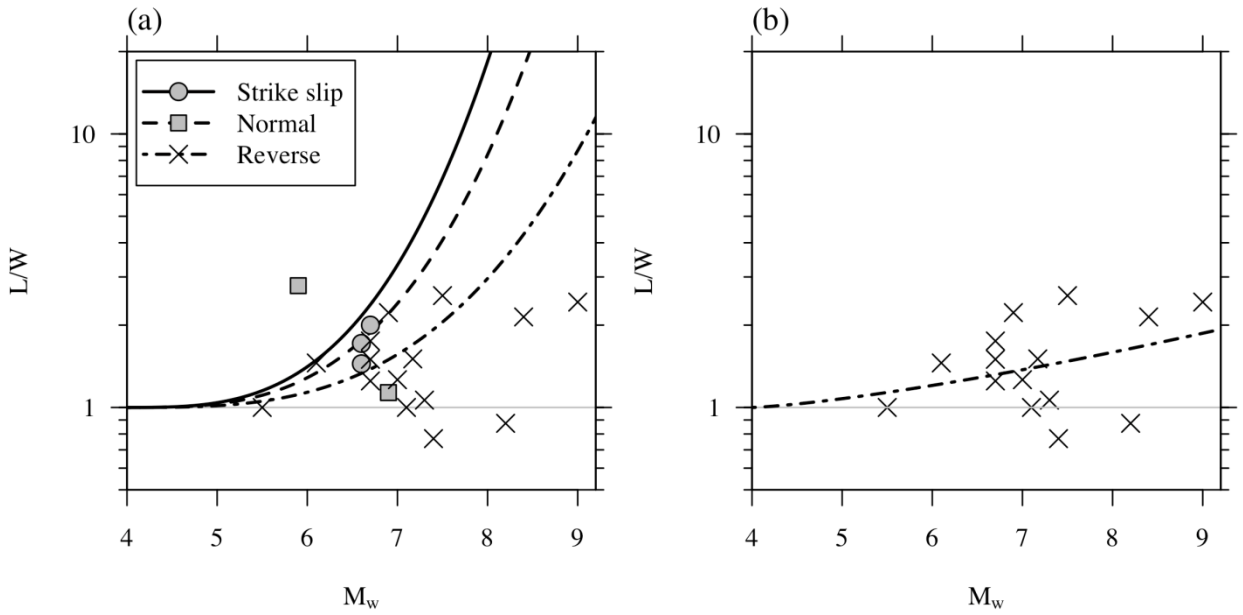


Figure A.3. M_w versus AR from the finite fault models in Table A.1 compared to (a) the Chiou and Youngs (2008) relationships for different fault types, and (b) a curve fit to the reverse faults in Table A.1.

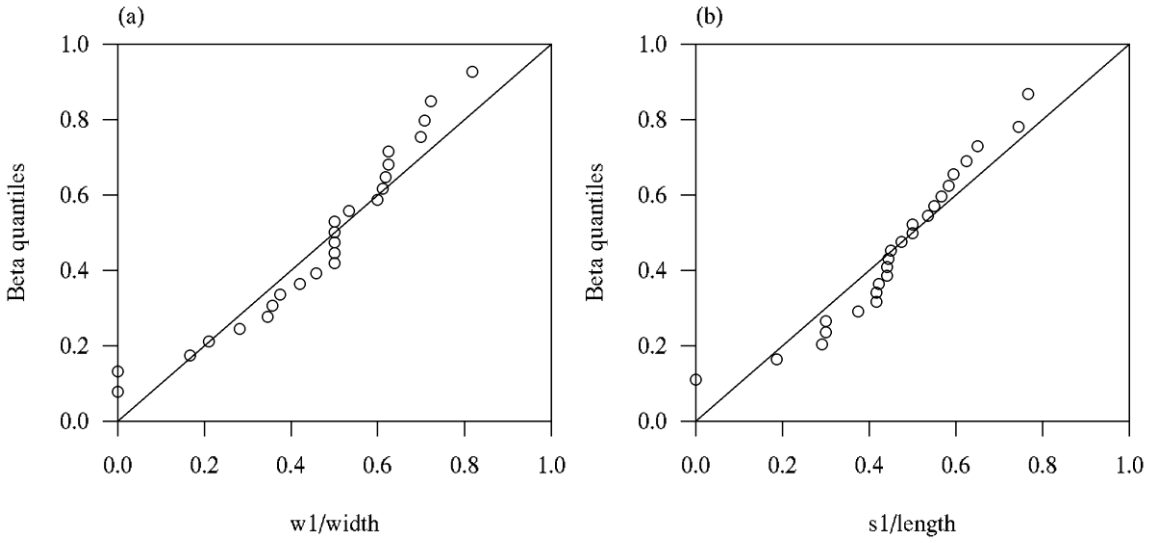


Figure A.4. QQ plots for (a) w1/width, and (b) s1/length.

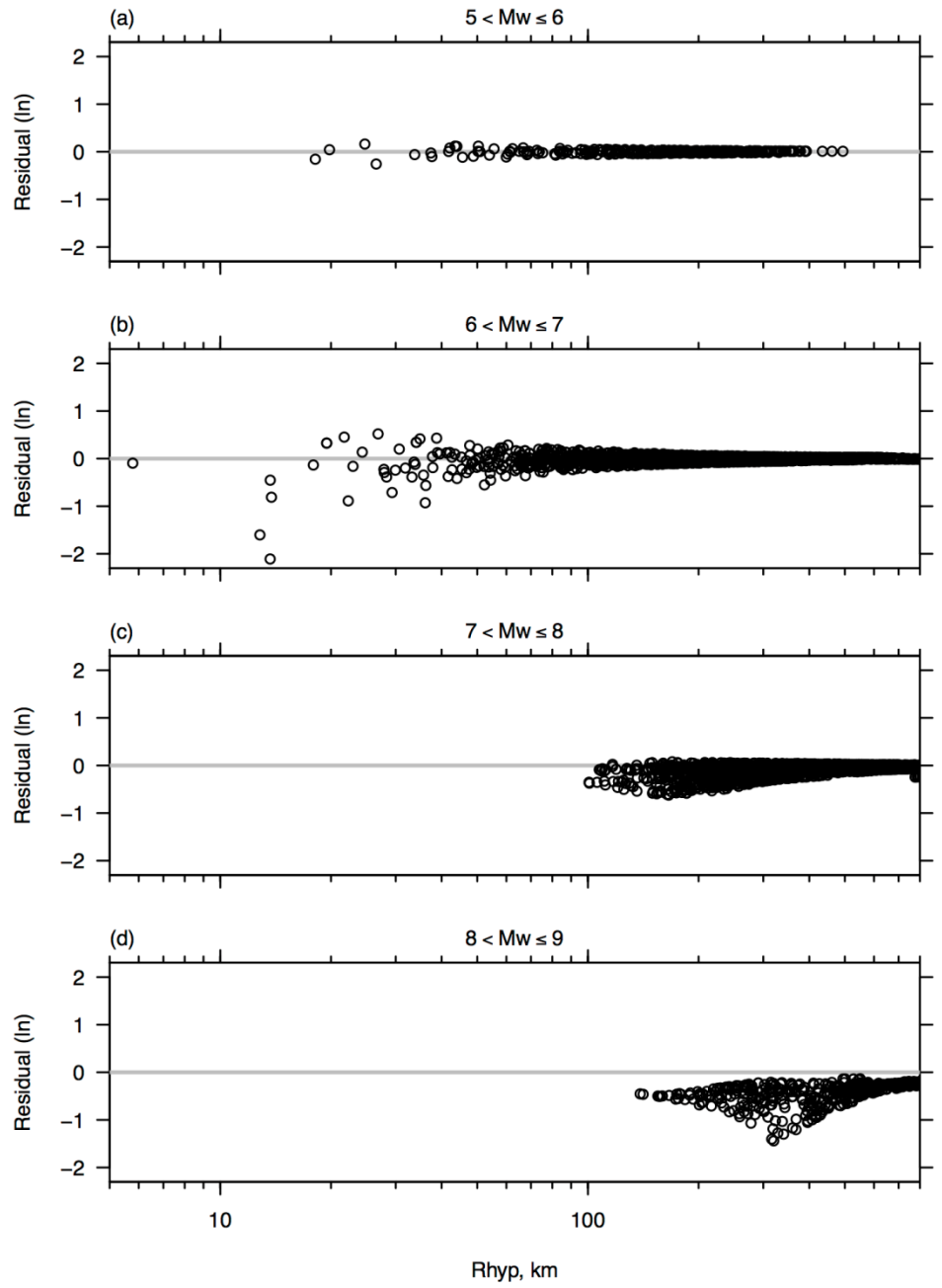


Figure A.5. Comparison of the preferred simulated R_{rup} and those computed from finite fault models for different magnitude ranges.

Table A.1. List of earthquakes for which a finite-source rupture model was found in the literature

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

Appendix B: A Flatfile For The KiK-Net Database Processed Using An Automated Protocol: Supplementary Information

Appendix B

A Flatfile For The KiK-Net Database Processed Using An Automated Protocol: Supplementary Information

This appendix is intended to provide additional details regarding the analyses presented in Chapter 3 of the dissertation. The title of each section in this appendix is named after the section in Chapter 3 that is being addressed.

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

Part I of Chapter 3 introduces the database. Some additional information was not included in Chapter 3 because of space limitation, this information is included below. Table B.1 shows the temporal distribution of the earthquakes present in the F-net between 1997 and 2011 with $M_w \geq 4.0$ (total of 14,395 events). Table B.2 shows the number of these earthquakes that falls in the different depth ranges.

Table B.1. Temporal distribution of the earthquakes the F-net catalog between 1997 and 2011 with $M_w \geq 4.0$

Year	Number of earthquakes in a magnitude range				Total
	$4.0 \leq M_w <$ 5.0	$5.0 \leq M_w <$ 6.0	$6.0 \leq M_w <$ 7.0	$7.0 \leq M_w$	
1997	145	38	5	3	191
1998	298	42	1	3	344
1999	323	44	3	0	370
2000	1192	132	12	3	1339
2001	473	71	10	2	556
2002	646	88	7	0	741
2003	710	117	11	1	839
2004	919	127	14	0	1060
2005	804	104	14	2	924
2006	787	110	14	0	911
2007	834	112	16	0	962
2008	805	111	17	2	935
2009	769	83	13	0	865
2010	890	106	13	2	1011
2011	2899	393	48	7	3347
Total	**12494**	**1678**	**198**	**25**	**14395**

Table B.2. The number of earthquakes that fall within specific depth ranges

Depth range (km)	Number and percentage of earthquakes
Depth $\leq$ 30	8177 (56.80 %)
30 $<$ Depth $\leq$ 50	3146 (21.85%)
35 $<$ Depth $\leq$ 70	1269 (8.82%)
70 $<$ Depth $\leq$ 100	616 (4.28%)
100 $<$ Depth	1187 (8.25%)

PART II: STRONG GM PROCESSING

Refinement of the Zeroth Order Baseline Correction:

This section provides additional details on the approach adopted to refine the zeroth order baseline correction conducted in *Step I* of the ground (GM) processing protocol.

The background noise window is assumed to be from the start of the GM record to the estimated start of the strong GM minus 2 seconds. Since the start of the strong GM is automatically picked, the noise window is shortened by 2 seconds to minimize the chances that a part of the strong GM record falls within the picked noise window.

The start of the ground motion is defined as the time where the ratio between the acceleration and the first non-zero acceleration in the record is greater than an arbitrarily selected ratio. Three ratios were tested (i.e., 5, 10 and 15) which result in three starting points. We pick the first starting point that is greater than 22 seconds as the start point of a specific component of the GM record. In case all the arbitrary ratios result in a starting time less than 22 seconds, the starting time that resulted from the largest ratio (i.e., 15) is chosen. In case the three selected starting points were found to be less than 7 seconds (i.e., noise window less than 5 seconds), it is assumed that a starting point could not be automatically found for that GM component. The choice of 22 seconds as the search criteria of the start of strong GM record and 5 seconds as the

smallest allowed noise window width are arbitrary and aim to ensure that the noise window has enough point so a reliable zeroth order baseline correction is conducted.

The mean of all points within the background noise window is subtracted from the originally recorded motion. When the start of the strong GM shaking was not automatically picked for a specific component, the mean of all the points in the component is subtracted to apply a zeroth order baseline correction.

Visual Inspection

During the automatic GM processing, several plots are saved for each GM record. These plots can then be used to visually inspect the motion for potential problems (e.g., Figure B.1) and assess the suitableness of the f_c picked by the automated process if needed. The saved plots are as follows:

- *PLOT (I)* shows the acceleration time histories of the six GM components after applying the refined zeroth order baseline correction (*step I*, AUTOMATED GM PROCESSING PROTOCOL section, in chapter 3). Figure B.2 shows an example of *PLOT (I)* saved during the database analysis.

- *PLOT (II)* presents an overview of the iterations conducted while processing the GM record to find a value for the corner frequency (f_c) that result in a filtered motion that satisfies the pre-set criteria (*step V*, AUTOMATED GM PROCESSING PROTOCOL section, in chapter 3). The figure shows the different criteria and specifies which criteria were not fulfilled for which component of motion at a tested f_c value. Figure B.3 shows an example of *PLOT (II)* saved for a GM from an M_{JMA} 5.8 (M_w 5.6) earthquake recorded at 149 km. The five subplots represent the first five pre-set criteria (*step V*, AUTOMATED GM PROCESSING PROTOCOL section, in chapter 3). The vertical axis in each subplot denotes the six GM components while the horizontal axis shows the tested f_c values (in this case the tested f_c are 0.07, 0.09, 0.14, 0.17 and 0.22 Hz). A green rectangle denotes that a specific criteria is fulfilled for a specific GM component at that f_c value. A yellow rectangle means that if the criteria is relaxed by 50% (i.e., multiplied by 1.50), the tested f_c would have been valid. A red

rectangle means that the criteria should be relaxed by more than 50% in order to be satisfied by that tested f_c value. For the case shown in Figure B.3, when f_c is set to be 0.14, 0.17 and 0.22 Hz the filtered GM record fulfilled the first five criteria. On the other hand, the slope of the FAS (sixth criteria in *step V*, AUTOMATED GM PROCESSING PROTOCOL section, in chapter 3) was not fulfilled. Hence, the code iterated up to an f_c of 0.22 Hz which is the first tested f_c that fulfills the six criteria for the six components simultaneously.

- *PLOT (III)* presents three different subplots (e.g., Figure B.4). The first subplot presents the smoothed FAS of acceleration for the six components before and after filtering the records as dashed and continuous lines, respectively. The second and third subplots present the velocity and displacement time histories for the filtered records. The title of the first subplot contains info about the date, time, and M_{JMA} of the event as well as the site at which the motion was recorded and epicentral distance of that site from the source.

- *PLOT (IV)* has two subplots (e.g., Figure B.5). The first subplot shows the smoothed FAS of acceleration for the six components of the whole record and the noise window (see *Step VII*, AUTOMATED GM PROCESSING PROTOCOL section, in chapter 3 for the definition of the noise window used here) as bold and thin lines, respectively. The second subplot shows the SNR for the six GM components along with three lines (bold, red lines) that represent the minimum SNR threshold of three and the frequency bandwidth within which we need to check if the SNR is constantly above the threshold (*step VII*, AUTOMATED GM PROCESSING PROTOCOL section, in chapter 3).

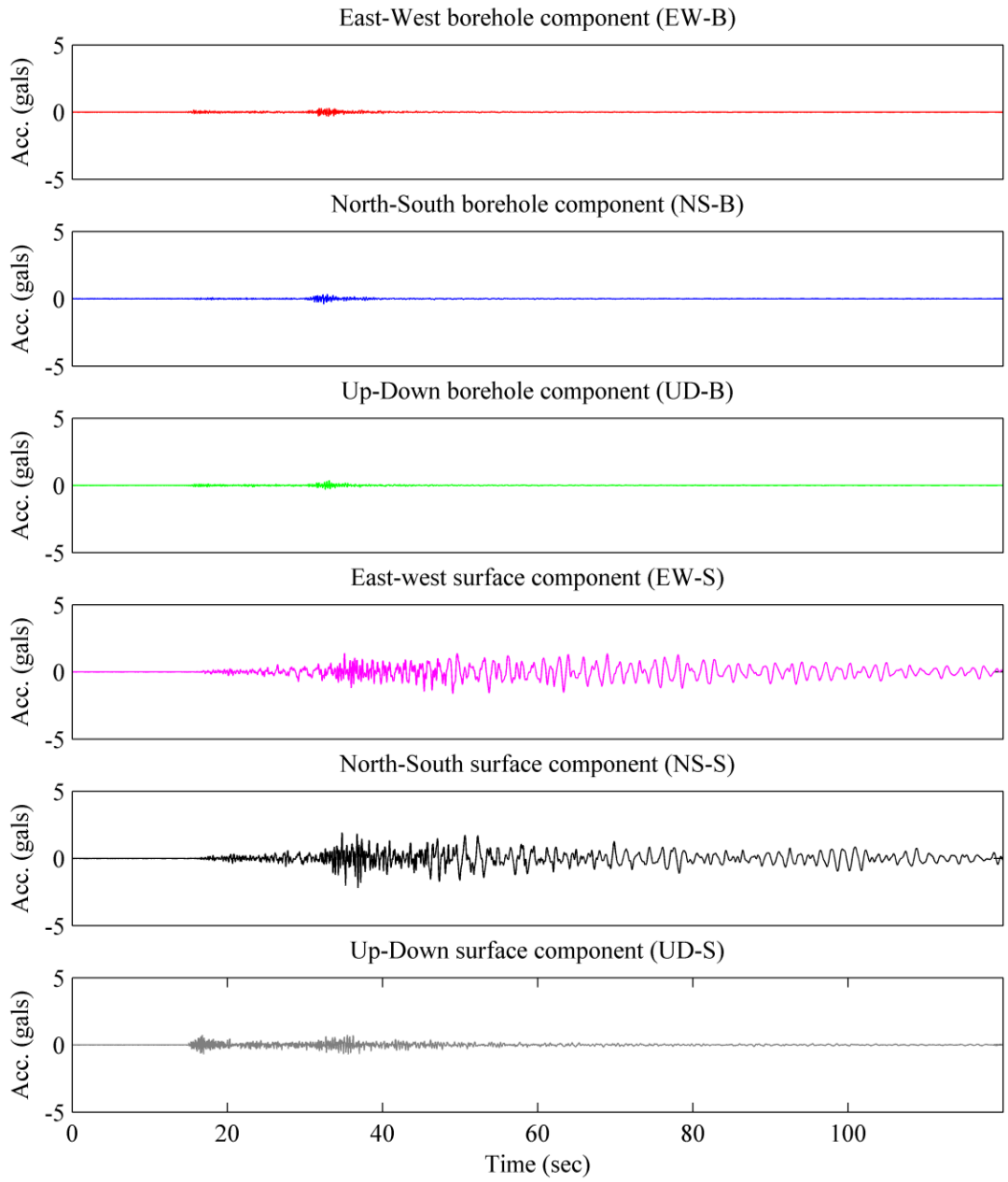


Figure B.1. Acceleration time history for a GM recorded at an epicentral distance of 142 km from an M_{JMA}5.1 (M_w 5.0) earthquake. The surface East-West and North-South components of this GM show large amplitude, long period harmonic wave

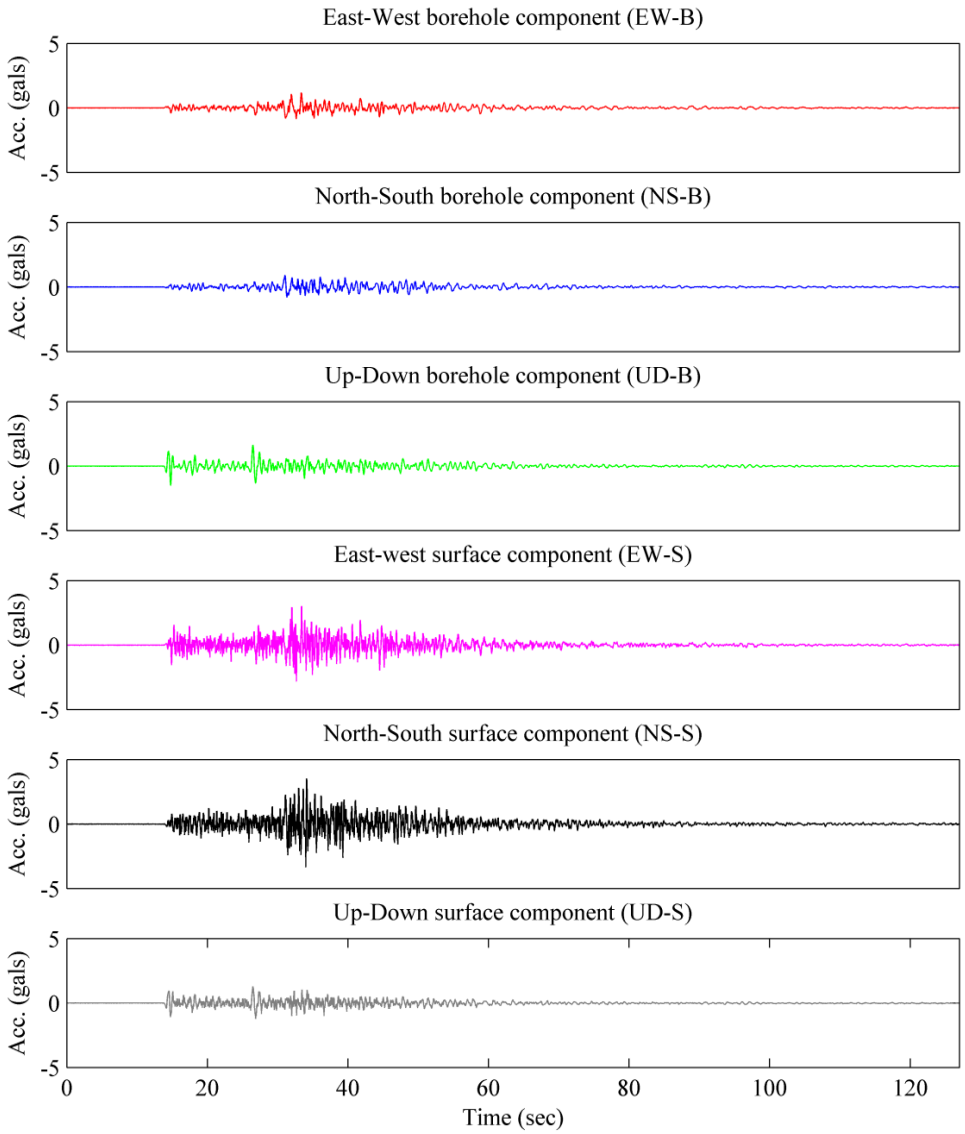


Figure B.2. An example for *PLOT (I)* saved during the processing of the GM records. The six subplots show the acceleration time histories for the different components of the GM record.

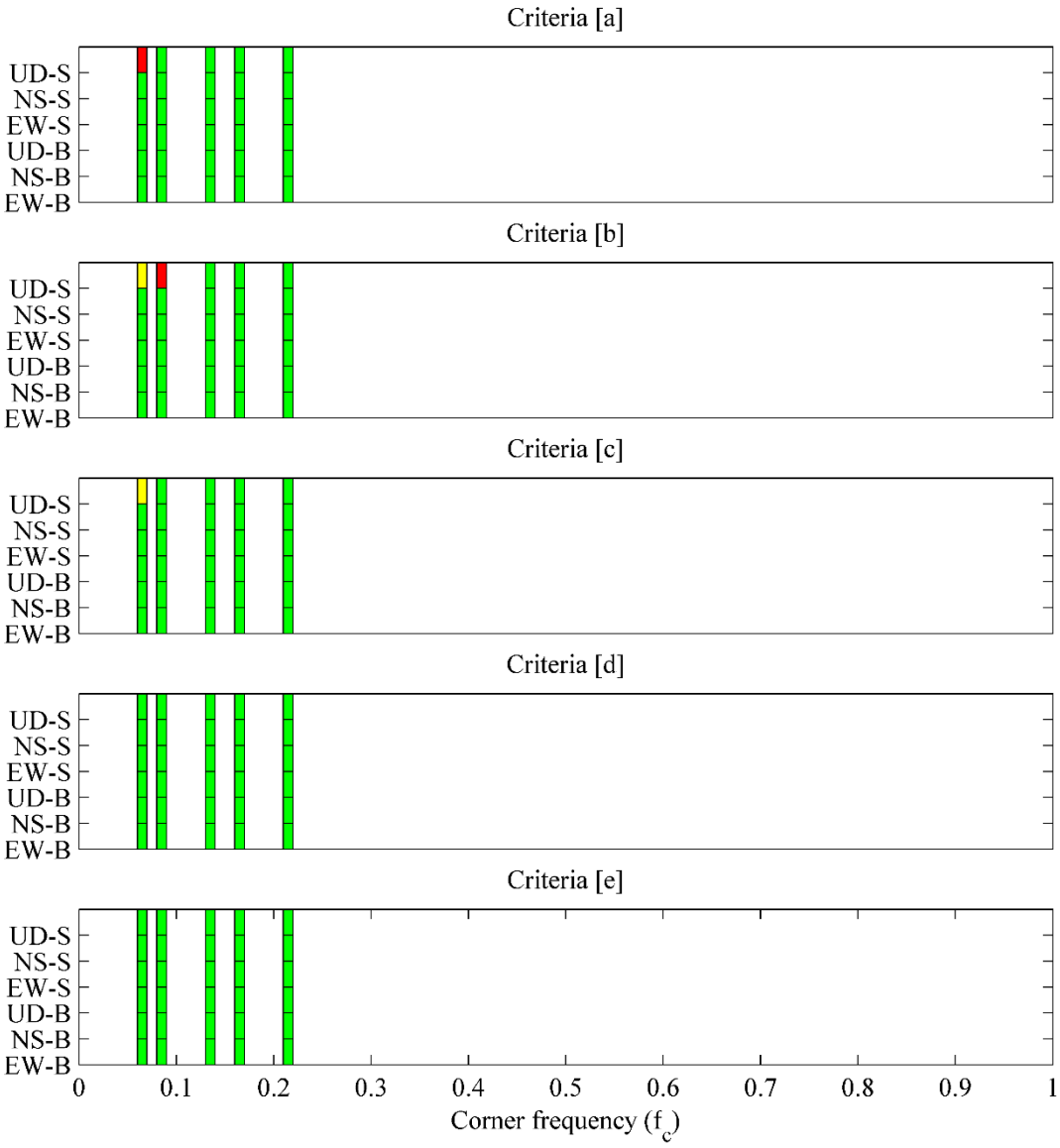


Figure B.3. An example for *PLOT (II)* saved during the processing of the GM records. Each subplot represents one of the first five pre-set criteria (*step V*, AUTOMATED GM PROCESSING PROTOCOL section, in chapter 3). The horizontal axes represent the values of f_c tested, while the vertical axes in each subplot represent the six GM components (EW denotes the East-West component, NS denotes the North-South component, UD denotes the vertical component, S denotes the motions recorded at the ground surface, and B denotes the motions recorded at the borehole). A green rectangle denotes that a specific criteria is fulfilled for a specific GM component at that tested f_c. A yellow rectangle means that if the criteria is relaxed by 50%, the tested f_c value would have been valid. A red rectangle means that the criteria should be relaxed by more than 50% in order to accept that specific f_c value

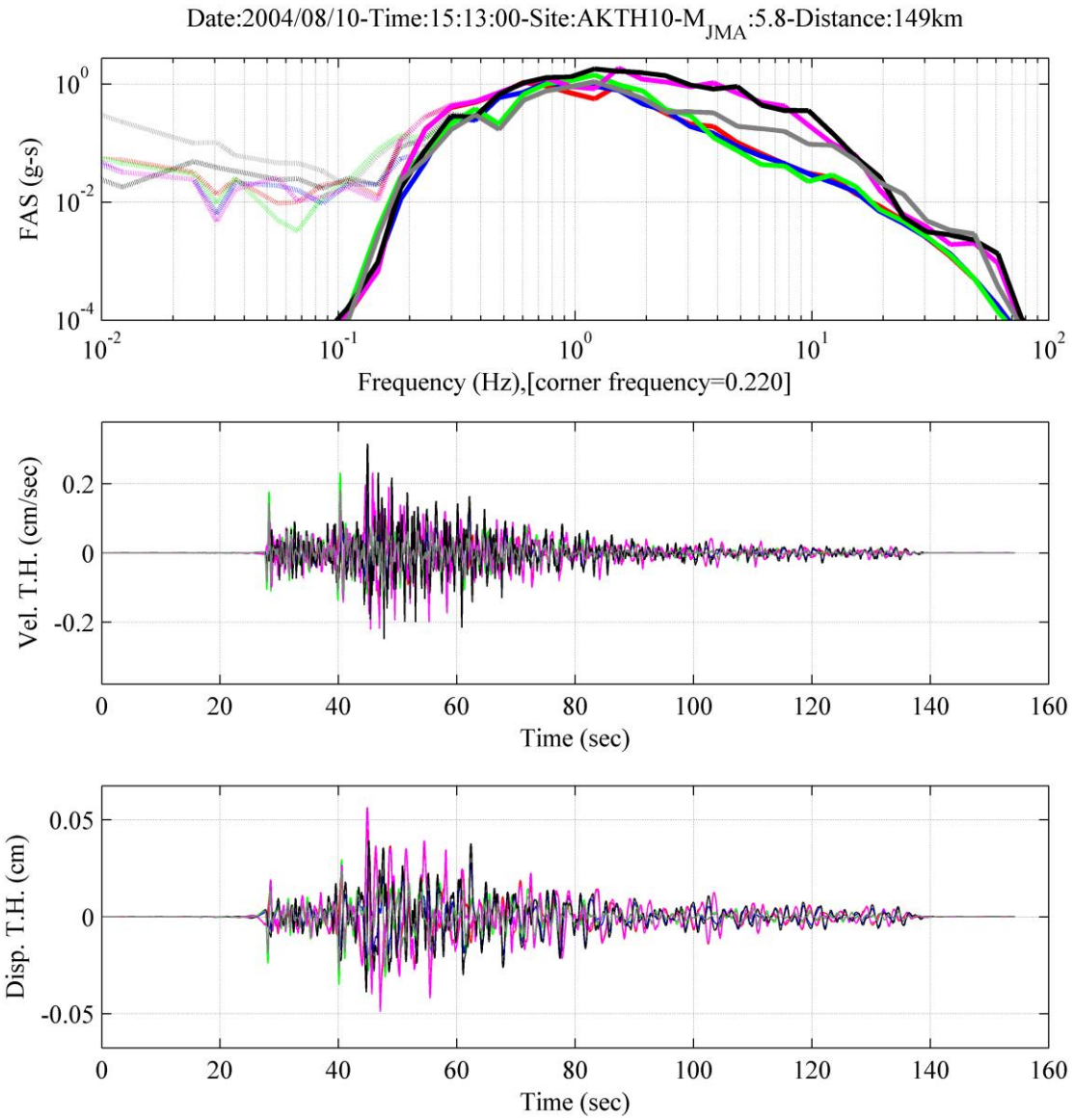


Figure B.4. An example for *PLOT (III)* saved during the processing of the GM records. The first subplot (from top) presents the smoothed FAS of acceleration for the six components before and after filtering the motion as dashed and continuous lines, respectively. The second and third subplots present the velocity and displacement time histories for the filtered motions. The title of the first subplot contains info about the date, time, and M_{JMA} of the event as well as the site at which the motion is recorded and epicentral distance of that site from the source.

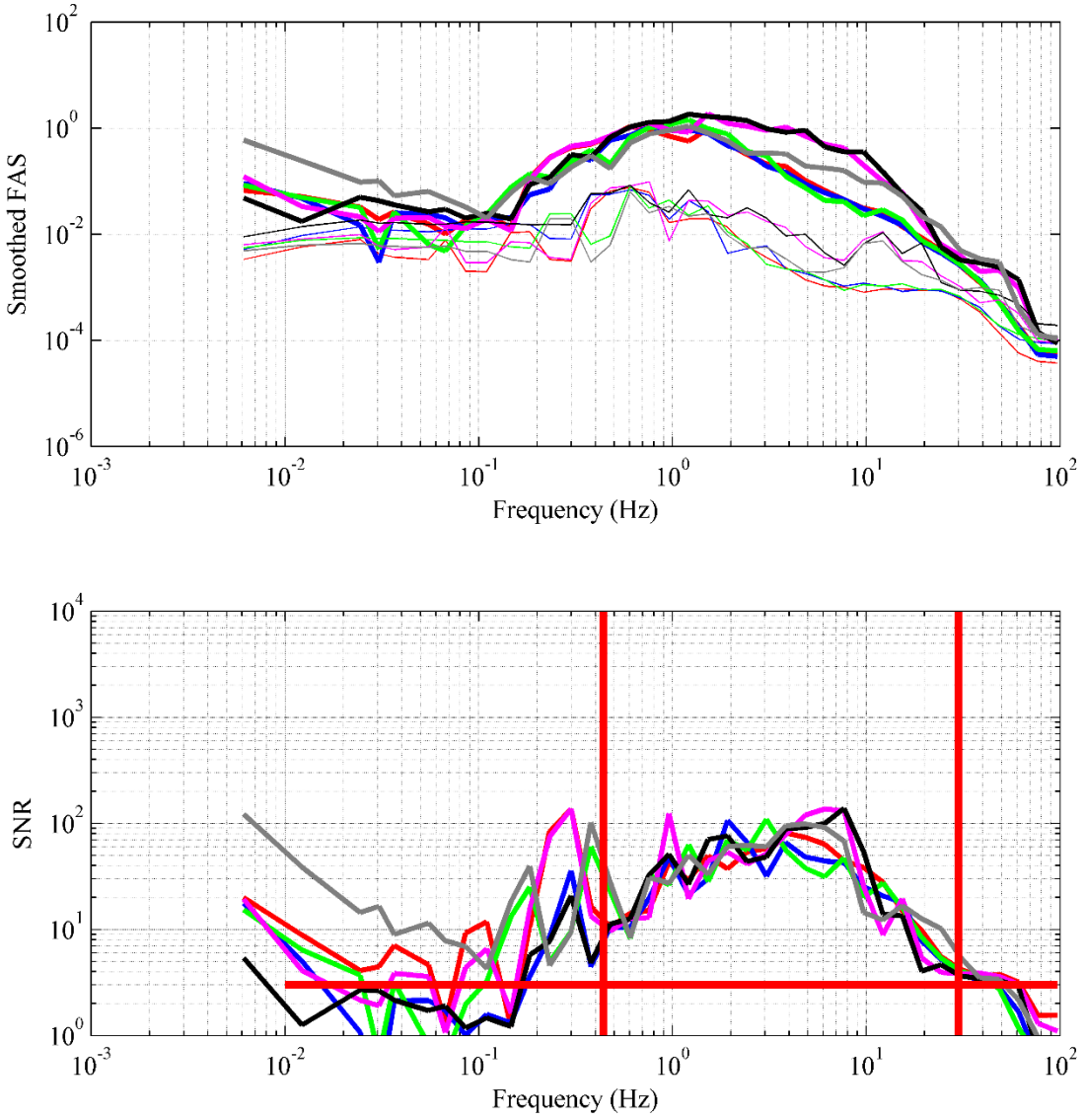


Figure B.5. An example for *PLOT (IV)* saved during the processing of the GM records. The first subplot (from top) shows the smoothed FAS of acceleration for the six components of the whole record and the noise window as bold and thin lines, respectively. The second subplot shows the SNR for the six GM components along with three lines (bold, red lines) that represent the minimum SNR threshold of three and the frequency bandwidth within which we need to check if the SNR is constantly above the threshold.

The presence of multiple wave trains in a single data file is a problem that limits the usability of the GM record. In some cases, especially after the occurrence of large earthquakes, a strong GM instrument can record two seismic wave (or more) trains from different earthquakes and include both in the same file assuming they are from a single earthquake (Figure B.6). Strasser and Bommer (2005) reported that using such records could result in distortion of the high frequency portion of the response spectra. They mentioned that they were not able to calibrate two different methods to automatically separate these sub events because of the large variations in the magnitude and distances available in their study. Hence, they conducted a manual separation of a subset of their database. In this study, the GM records with multiple wave trains are flagged during the visual inspection. In the current study we developed an automated algorithm to detect GM records that contain multiple wave trains (see Appendix C in this dissertation for more details)

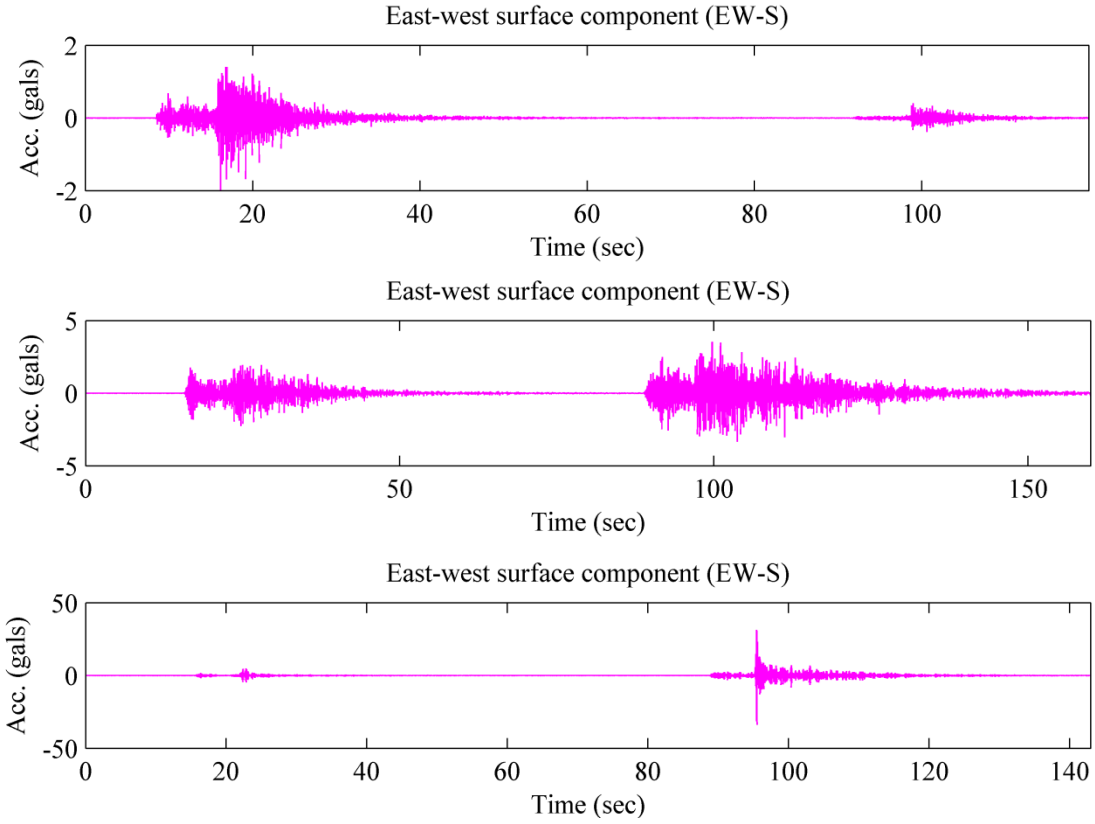


Figure B.6. The East-west component recorded at the ground surface of three different GM records that have multiple wave trains

Processing Sequence

Figures B.7 and .8 present a visual example of the different loops through which the processing protocol go through to pick up a suitable f_c value. These plots are presented for clarification purposes only. Figure B.7 shows the acceleration, velocity and displacement time histories of one component of a recorded time history after zeroth order baseline correction. It is clear from the velocity and displacement time histories that the motion needs to be baseline corrected. Figure B.7 also shows the smoothed FAS. The FAS is almost horizontal at low frequencies after it started to decay. This confirms the presence of long period noise that needs to be filtered out.

Figure B.8 shows the time histories of the same component shown in Figure B.7 but after applying a filter with a corner frequency of 0.07Hz. The velocity and displacement time histories visually look physical. Another subplot is included to show the result of the five checks conducted using the different preset criteria (see *Step V*, in *Part II* of Chapter 3 of this dissertation). The resulting motion fulfills the first five criteria. Also, four additional lines were added to the smoothed FAS. The gray solid line represent the theoretical slope of the FAS decay with frequency (slope of 2), and the gray dashed lines represent the limits of the range of slopes that we considered acceptable while processing the motions (slopes of 1 and 3). The black solid line represents the actual slope of the FAS of this component of the motion. Since this solid black line falls between the two gray dashed ones, this component passes the sixth criteria. Hence, the code checks the remaining 5 components to check if this f_c value results in acceptable components similar to this one.

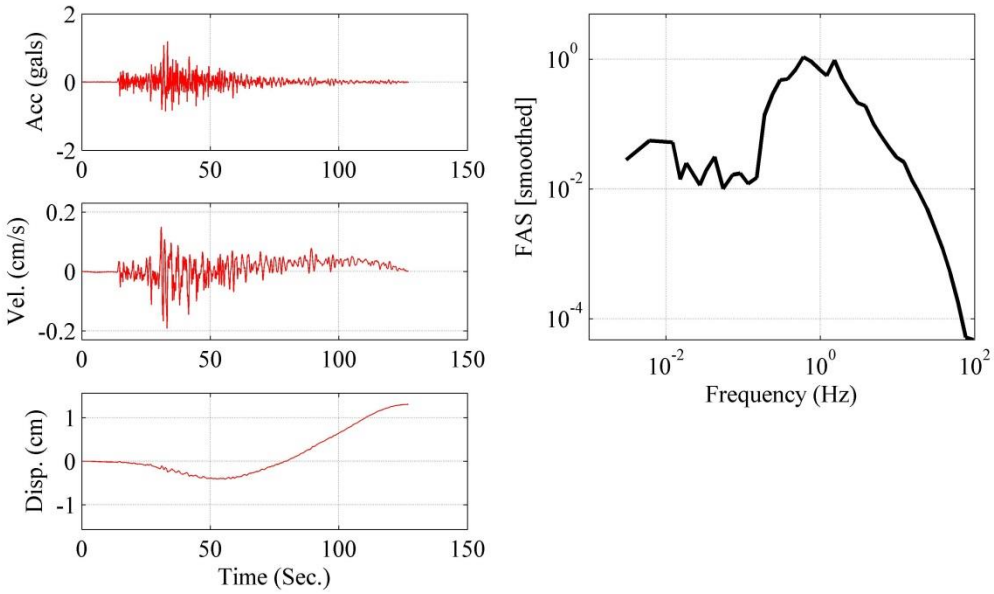


Figure B.7. Acceleration, velocity and displacement time histories of one component of a recorded time history after zeroth order baseline correction (left panels). It also shows the smoothed FAS of the same component (right panel)

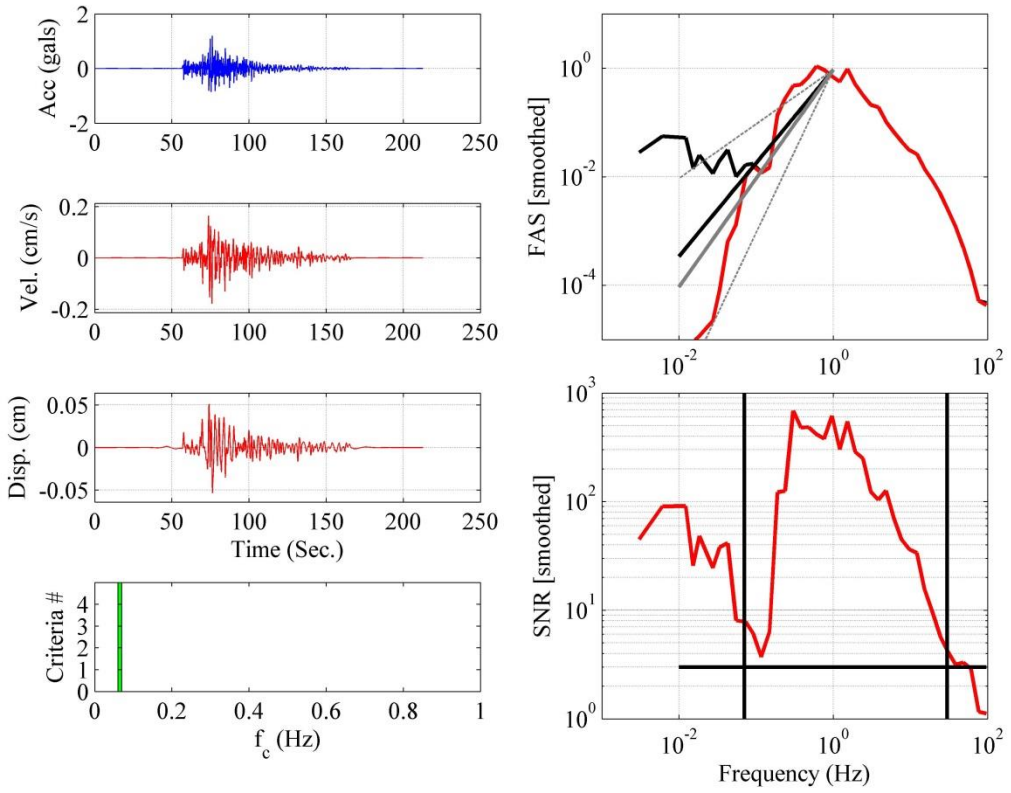


Figure B.8. Acceleration, velocity and displacement time histories of the same component of motion shown in Figure B.7 after applying a filter with a high-pass corner frequency of 7Hz. The fourth panel to the left shows that the five preset criteria were satisfied after this filter was applied. Four additional lines were added to the panel that shows the smoothed FAS. The gray solid lines represent the theoretical slope of the FAS decay with frequency, and the gray dashed lines represent limits of the range of slopes that we considered acceptable while processing the motions (slopes of 1 and 3). The black solid line represents the actual slope of the FAS of this component of the motion.

PART III: METADATA

Part III of Chapter 3 describes the metadata collected for the different ground motion records that is included in the final *Flatfile*. This section presents some additional information that was not included in Chapter 3 because of space limitation.

Earthquake Information

Table B.3 shows the range of errors in the matching parameters for events that fall in Category D described in Chapter 3. Some of these errors are quite large. This is why we discourage users to use any motions from earthquakes that have a matching category other than "A" unless the user re-check the match in person.

Table B.3. Error range in the matching parameters for earthquakes in category (D)

	Minimum Error	Maximum Error
Latitude (°)	0	0.429
Longitude (°)	0	2.822
Date	0	0
Hours	0	0
Minutes	0	19
M_{JMA}	0	2.4
Depth (km)	0	34

Declustering the F-net Catalog

Declustering of the earthquakes in the KiKnet database is conducted as part of the development of metadata for the database. This section includes additional details on how the declustering was conducted. In order to test if the declustered catalog is Poissonian. The following steps were conducted:

- The period of time through which the earthquake catalog is collected is divided into equal width temporal windows (from here on named as bins).

- The number of seismic events that occurred in each bin is counted.
- The number of bins that contain a certain number of seismic events is then calculated. These numbers are then tested using a Chi-square goodness-of-fit test to check if it Poissonian.

We were not able to find any reference that describes the reasons behind using a specific temporal width for the bins. We believe that the choice of this width is arbitrary. Hence, we test several widths that range from 10 to 50 days on the four declustered catalogs (i.e., obtained by applying the algorithms by Reasenberg 1985 and Gardner and Knopoff 1974). In all cases the null hypothesis that the occurrence of earthquakes in the declustered catalog follows a Poissonian distribution can be rejected at the 5% significance level.

Hence, we attempted to alter the spatial and temporal windows suggested by Gruenthal (personal communication with J. Woessner) so as to obtain a Poissonian catalog. We obtained, the windows shown in Equations B.1 and B.2 that resulted in a declustered catalog that follows a Poissonian distribution when the bin width was assumed to be above 22 days (Figure B.9). Since the catalog was not found to be Poissonian if the width of the bin was assumed to be less than 22 days, we decided to conduct an additional check.

As an additional check, we developed 100 synthetic catalogs with the same number of earthquakes and the same average time between the consecutive earthquakes as the declustered catalog. These synthetic catalogs are developed using an exponential random variable for the occurrence time of the events. We then check if the catalog is Poissonian using the different bin widths. Figure B.9-b shows the number of synthetic catalogs (out of 100) that were found Poissonian at each tested binning temporal window width. This shows that the chosen binning temporal window width can affect the output of the goodness of fit test. Hence, by comparing the two plots in Figure B.9, we conclude that declustered catalog is Poissonian. Figure B.10 shows the distribution of the mainshocks on the Japanese territory. The F-net earthquakes that were found in the KiK-net datafiles we used in this study are shown as rec circles.

$$\text{Spatial window size} = e^{\left(1.77+\sqrt{0.037+1.02M_w}\right)} \quad \text{km} \qquad \text{(B.1)}$$

$$\text{Temporal window size} = \begin{cases} \left| e^{(-5.135+\sqrt{1.05+29.27M_w})} \right|, & M_w < 6.5 \\ 10^{(3.36+0.029M_w)}, & M_w \geq 6.5 \end{cases} \quad \text{days} \qquad \text{(B.2)}$$

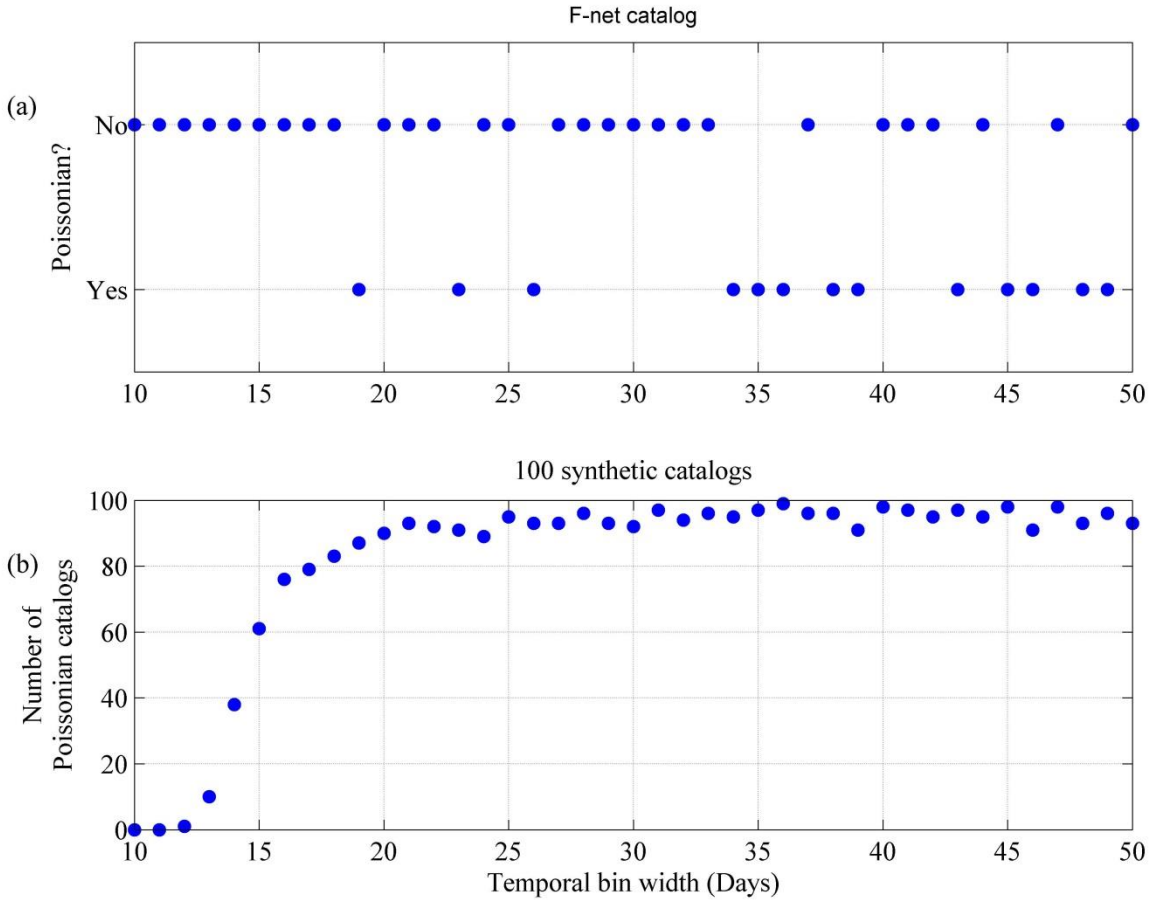


Figure B.9. A figure that shows: (a) The output of the chi-squared goodness of fit test for the declustered F-net catalog when assuming different binning temporal window widths. This declustered catalog resulted from implementing the window sizes calculated using Equations B.1 and B.2 in the Gardner and Knopoff's algorithm; and (b) the number of synthetic catalogs (out of 100) that were found to be Poissonian assuming different binning temporal window widths. These synthetic catalogs had been developed so that the time of occurrence of the earthquakes follows an exponential distribution.

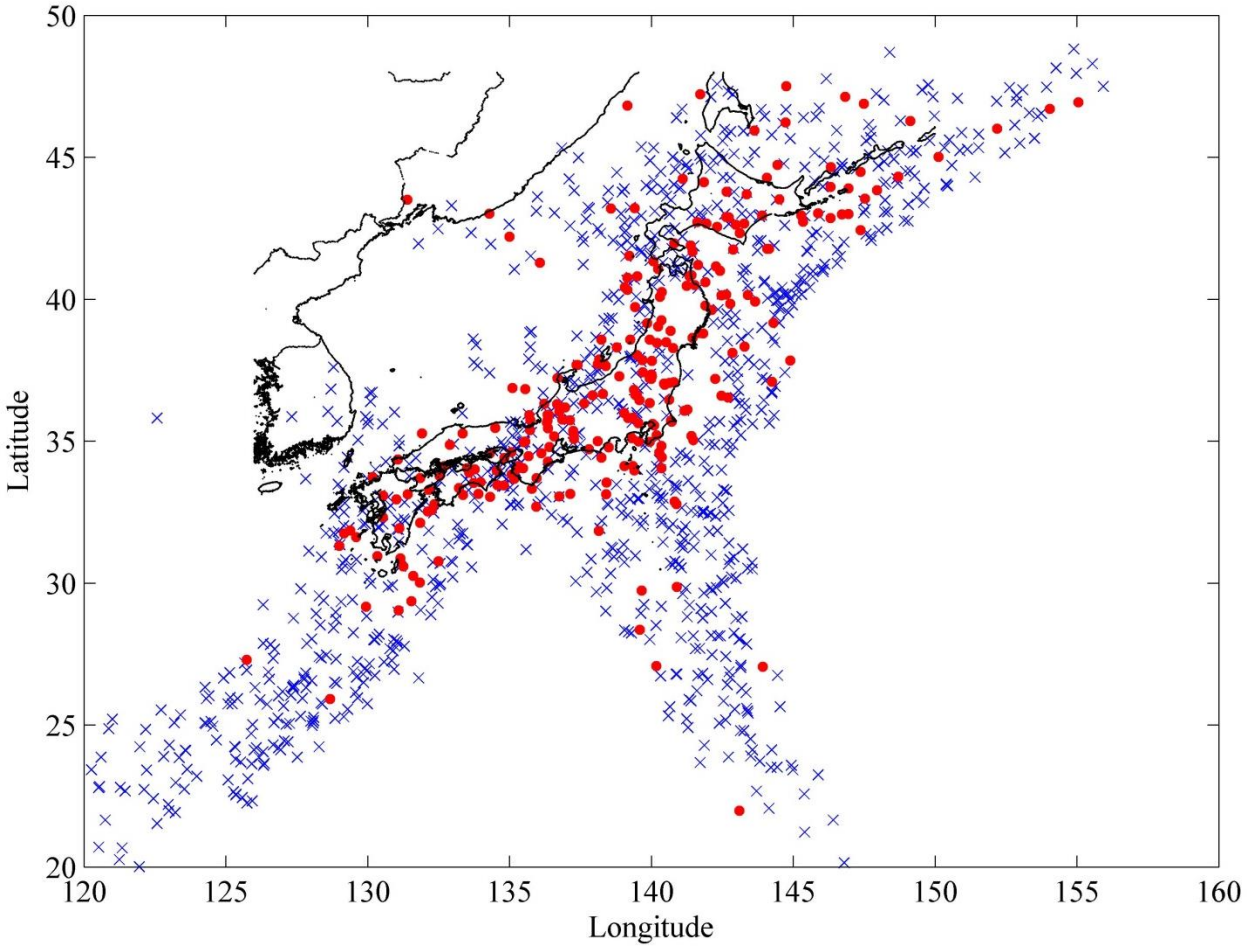


Figure B.10. The distribution of the mainshocks on the Japanese map. The circles and crosses are earthquakes that were or were not matched to GM records from the KiK-net database, respectively.

Figure B.11 shows the Gutenberg–Richter recurrence relationship for the catalog before and after declustering. If the Gutenberg–Richter is expressed as:

$$\log_{10} N = a - b\, M_w \qquad \text{(B.3)}$$

where N is the number of earthquakes with a moment magnitude greater than or equals to M_w. Then, the coefficients "a" and "b" were found to be "7.7951" and "0.9241" before declustering and "5.1497" and "0.5510" after declustering. The b-value is found to be very close to one for complete catalogs. Hence, even if this declustered catalog is believed to be Poissonian, it does not seem to be complete. This means that there is a conflict –at least for this catalog– between

obtaining a Poissonian and a complete catalog. In the Flatfile associated with this study, we report the output of the standard declustering algorithms. Hence, care should be taken if a user needs to use the current databases in applications where a Poissonian catalog is needed and/or recurrence rate is important.

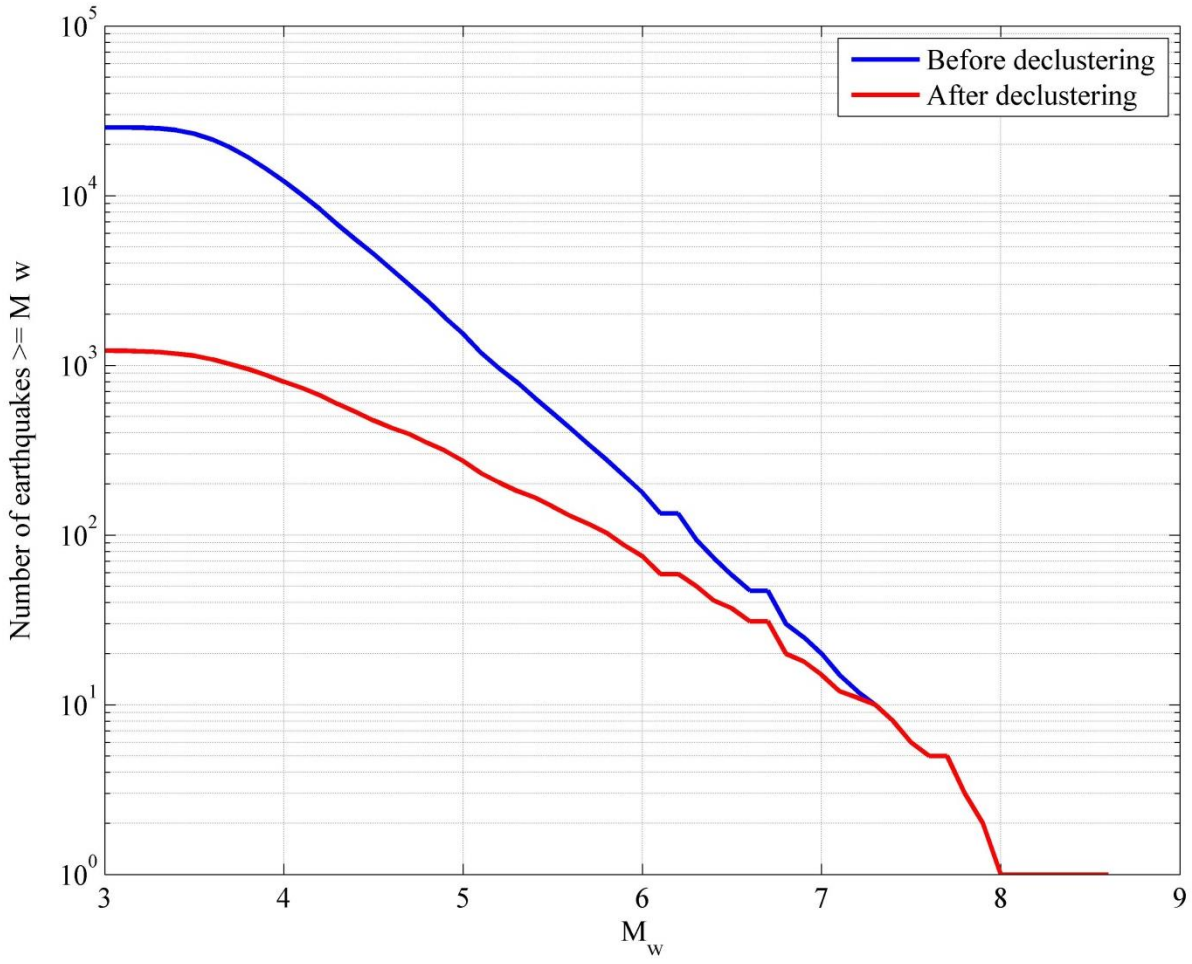


Figure B.11. The Gutenberg–Richter recurrence relationship for the catalog before and after declustering. The b-value was found to be 0.9241 and 0.5510 before and after declusteing, respectively

GROUND MOTION PARAMETERS

For the six components of each record in the database, the peak ground acceleration (PGA), peak ground velocity (PGV), peak ground displacement (PGD), and pseudo-spectral

accelerations (5% damping) at multiple spectral periods were calculated. The pseudo-spectral accelerations were calculated using the piece-wise exact method (Nigam and Jennings 1969). The "*maximum PGA*" and "*maximum PGV*" for the borehole and surface horizontal motions were also calculated. If we define the resultant acceleration "$ACC_R\ (t)$" at a point in time "*t*" using Equation B.4. Then the "*maximum PGA*" is defined as the maximum value of ACC_R in the whole acceleration time history. Same for "*maximum PGV*"

$$\mathrm{ACC_R}\ (t) = \sqrt{ACC(t)^2_{N-S} + ACC(t)^2_{E-W}} \tag{B.4}$$

where, $\mathrm{ACC(t)_{N\text{-}S}}$ and $\mathrm{ACC(t)_{E\text{-}W}}$ are two orthogonal components of the horizontal acceleration time history. We set the minimum usable spectral period to be greater that the inverse of the instrument's corner frequency (i.e., 0.035 seconds). Moreover, the filtering process affects the FAS of the records even above the high pass corner frequency (f_c) of the filter, because the Butterworth filter is not sharp. Hence, to avoid the underestimation of the pseudo spectral accelerations at long periods, the maximum usable spectral period should be less than the inverse of f_c of the filter. In general, the maximum usable spectral period of a GM record is assumed to be fraction of ($1/f_c$). Several researchers have used different values of this fraction over the years (see Boore and Bommer 2005, and Akkar and Bommer 2006). In this study we chose to set the maximum usable spectral period to be 50% of ($1/f_c$). Based on this, the number of pseudo spectral accelerations calculated for each GM record changes as the f_c value changes. For example, we calculated the pseudo spectral accelerations at 92 spectral periods (the same periods adopted from the NGA project, Chiou et al. 2008) between 0.035 and 7 seconds for GM records that have a maximum usable spectral period of 7 seconds.

REFERENCES

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

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.

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

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

Appendix C: Comparison between the Records Processed Using the Automated Algorithm by Dawood et al. (2014), and the Motions Used in the NGA-W2 and BC-Hydro Projects

Appendix C

Comparison between the Records Processed Using the Automated Algorithm by Dawood et al. (2014), and the Motions Used in the NGA-W2 and BC-Hydro Projects

INTRODUCTION

Chapter 3 of this dissertation, which is submitted for publication as Dawood et al. (2014; heretofore referred to as Dea14), presented an automated ground motion (GM) processing protocol that was applied to strong GMs recorded at KiK-net stations.Two different subsets of the KiK-net records had been previously processed by Zhao et al. (2006; Zea06) (this subset was used in the BC-Hydro project) and the NGA-W2 project. In the cases of Zea06 and the NGA-W2 project, the corner frequency (f_c) of the high-pass filter was picked one-by-one by visually inspecting the strong GM records. On the other hand, Dawood et al. (2014; Dea14) automatically pick the f_c that result in a processed record that fulfills some preset criteria. The objective of this appendix is to compare the pseudo spectral acceleration (PSA) obtained from the automatically processed motions with the PSA from the other databases for the common motions. This could help identify potential problems in the Flatfile developed by Dea14. A total of 150 and 168 records were used for the comparison with Zea06 and NGA-W2, respectively.

PROCEDURE

For each GM record a figure similar to Figure C.1 is prepared. The figure presents the PSA (geometric mean of the two horizontal components at the ground surface) obtained from the automated processing protocol and the other source in the same panel. The figure also presents the time histories (acceleration, velocity, and displacement), smoothed Fourier amplitude spectra (FAS) and signal-to-noise ratio of the two horizontal components of the motions recorded at the ground surface. This figure is used to visually search for any differences in the PSA of the GM records processed by Dea14 and either Zea06 or NGA-W2. It also presents the time histories,

FAS and SNR to ensure that the automated protocol picked a suitable corner frequency for the high-pass filter.

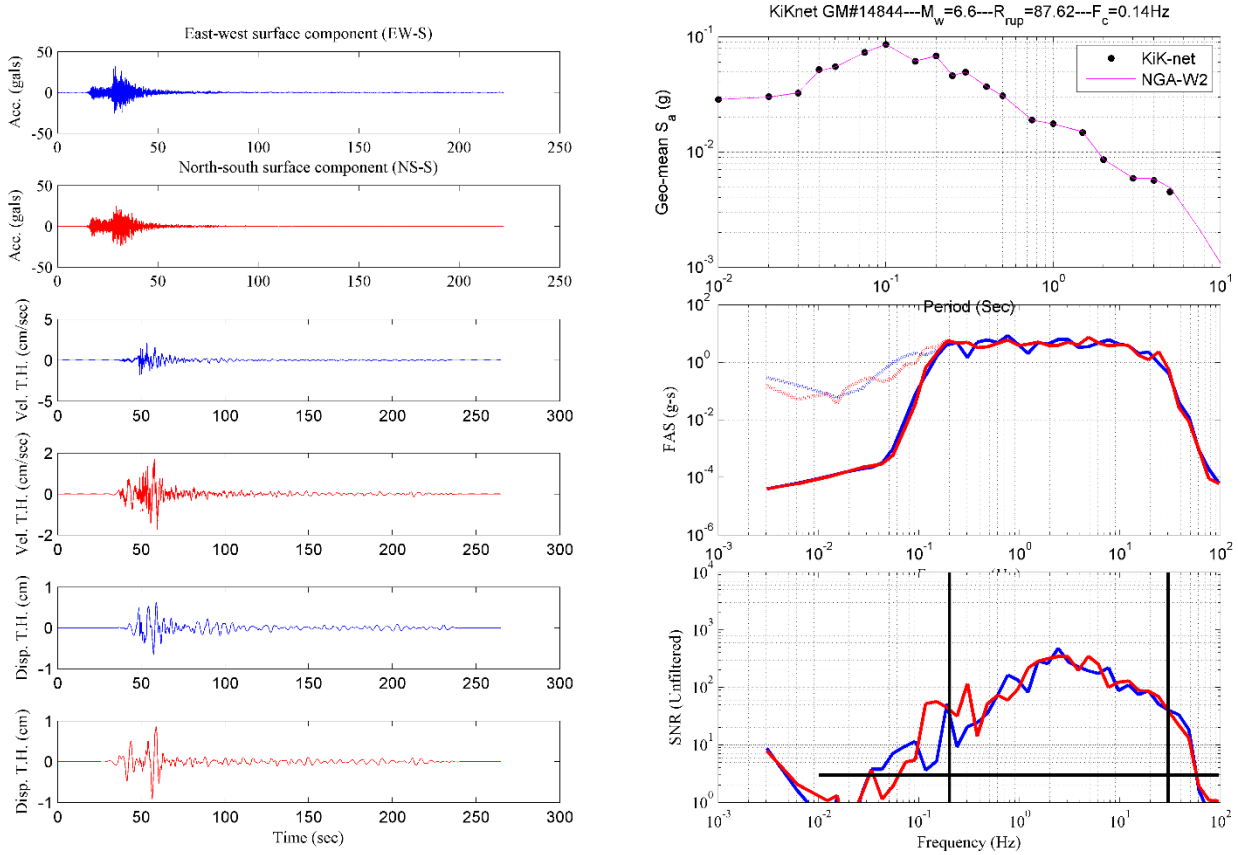


Figure C.1. An example of a figure prepared for each GM record that presents the following:

I. **<u>Left panels</u>: the time histories (acceleration, velocity, and displacement) for the automatically processed ground motions by Dea14 (two horizontal components at ground surface).**

II. **<u>Top right panel</u>: PSA (geometric mean) of the automatically processed record (black closed circles) along with the PSA of the same motion obtained from either Zea06 or the NGA-W2 databases (plotted in magenta).**

III. **<u>Middle right panel</u>: the smoothed FAS of the ground motion before (dotted lines) and after applying the Butterworth filter (continuous lines) for the two horizontal components of motion at ground surface.**

IV. **<u>Bottom right panel</u>: the signal-to noise ratio for the two horizontal surface components of the motion.**

COMPARISON WITH NGA-W2 MOTIONS

Figure C.2 compares the f_c of the high-pass filter obtained from the automated protocol and the NGA-W2 processing. In general, the automatically picked f_c of the filter is higher than the NGA-W2 corner frequency (only 3 GM records out of 168 are an exception to that; GMs# 14879, 48287, and 2911). This means that the automated protocol is probably strict in picking the f_c value. Hence, the usable bandwidth from the automated protocol is narrower than the bandwidth by NGA-W2. On the other hand, this also means that the usable frequency bandwidth obtained from the automated protocol can be reliably used in different applications. It is also important to note that the size of the database processed by Dea14 is very large to be processed on a one-by-one basis. Hence, obtaining a narrower usable bandwidth might be a fair price to pay in exchange of having such a large database of strong GM records.

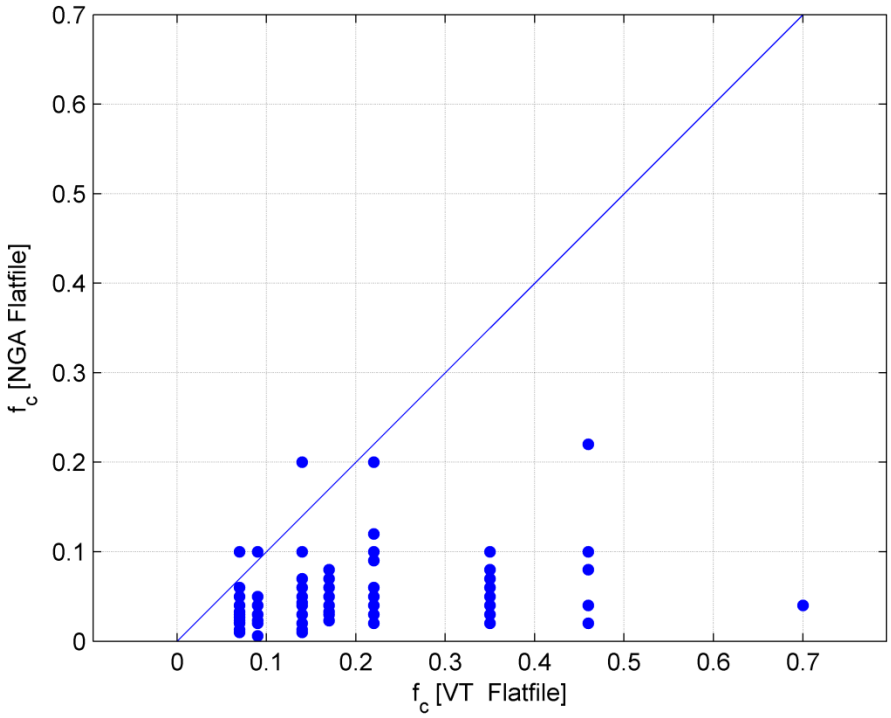


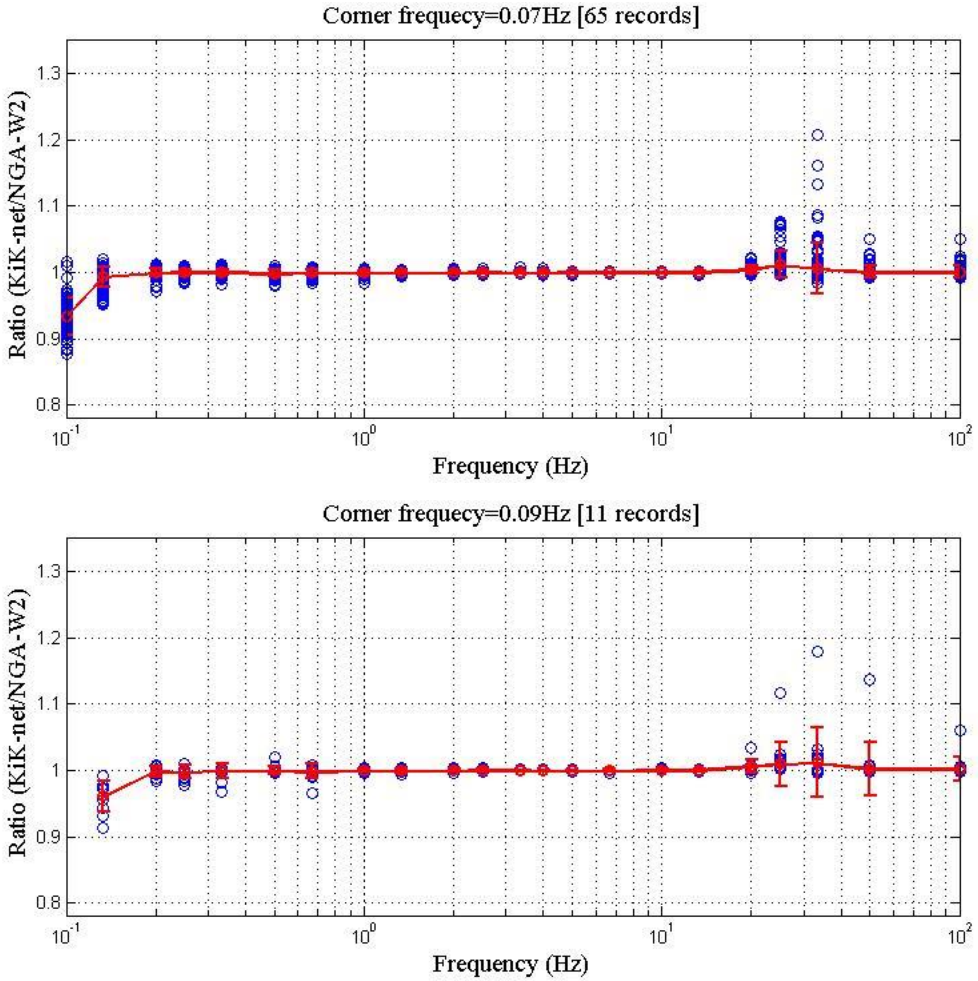
Figure C.2. Comparison between the high-pass corner frequency of the filter used by NGA-W2 and Dea14

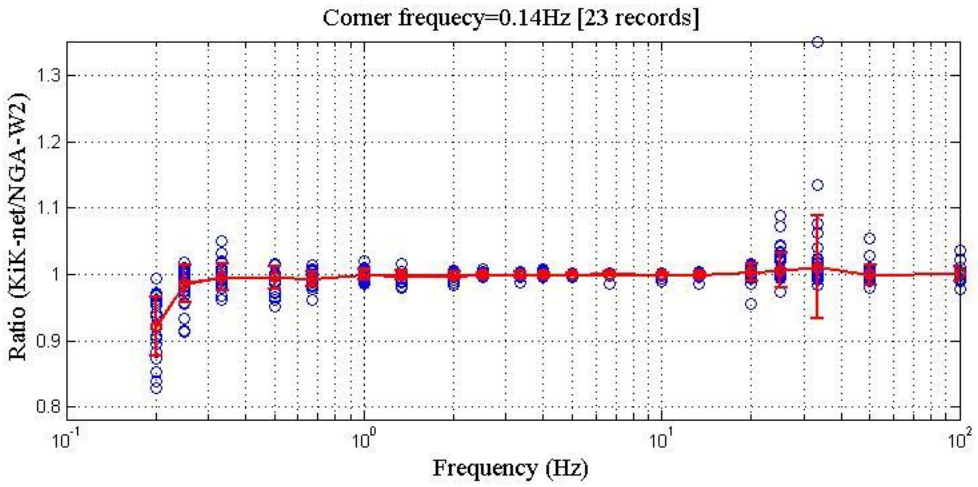
Figure C.3 shows the ratio between the PSA obtained from the automated processing protocol to the PSA from the NGA-W2 databases as a scatter plot. Each subplot contains all

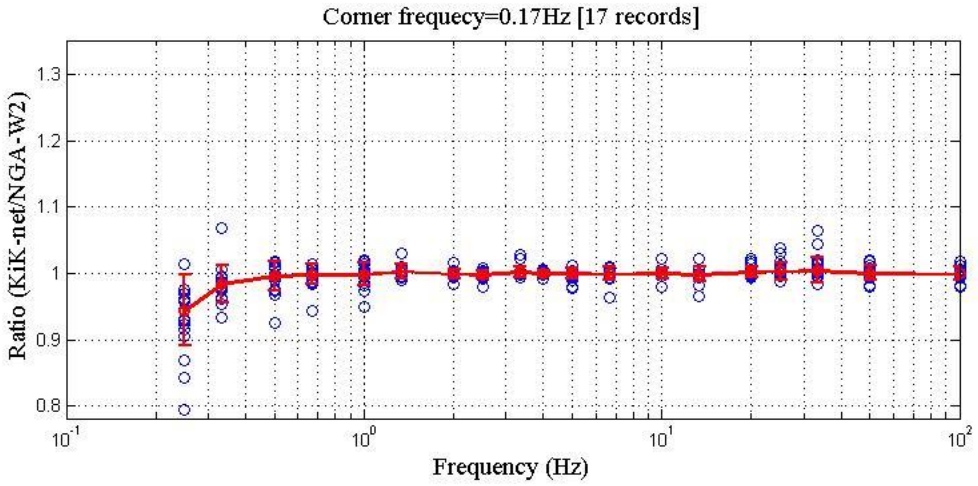
records that had been filtered with a high-pass filter with the same f_c (for the Dea14 records). The median and error bars (representing one standard deviation above and below the median) are added to the figures (red lines). The scatter plots show that there are some differences in some GM records between the PSA calculated from the automatically processed motions and the motions processed within the NGA-W2 project at high frequencies (around 30 Hz) and/or low frequencies. The figures similar to Figure C.1 for the records that are characterized with these differences were inspected. The three sources that are believed to be the source of the inconsistency of the results are:

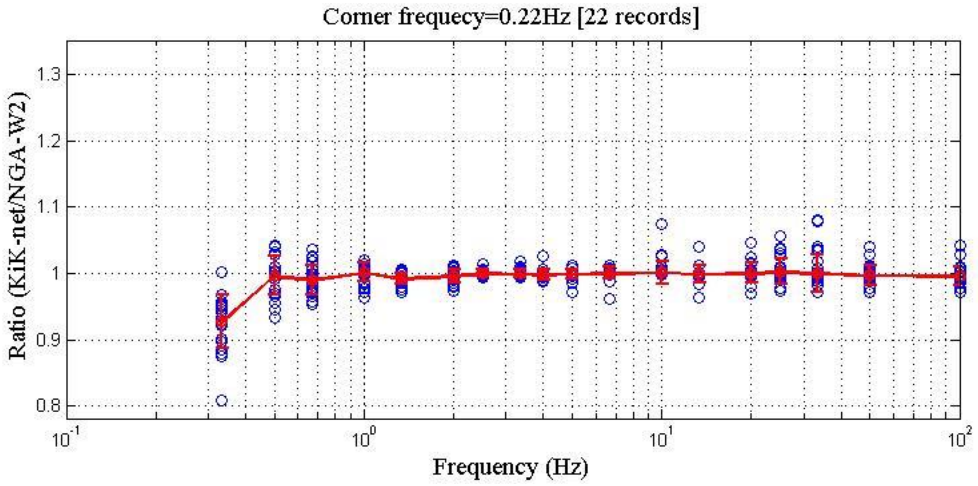
- Inconsistency in high-cut filtering
- Maximum usable spectral period for the automatically processed GM records
- Multiple wave trains

These items are described in more detail below.




Corner frequecy=0.14Hz [23 records]
Ratio (KiK-net/NGA-W2)
Frequency (Hz)


Corner frequecy=0.17Hz [17 records]
Ratio (KiK-net/NGA-W2)
Frequency (Hz)


Corner frequecy=0.22Hz [22 records]
Ratio (KiK-net/NGA-W2)
Frequency (Hz)

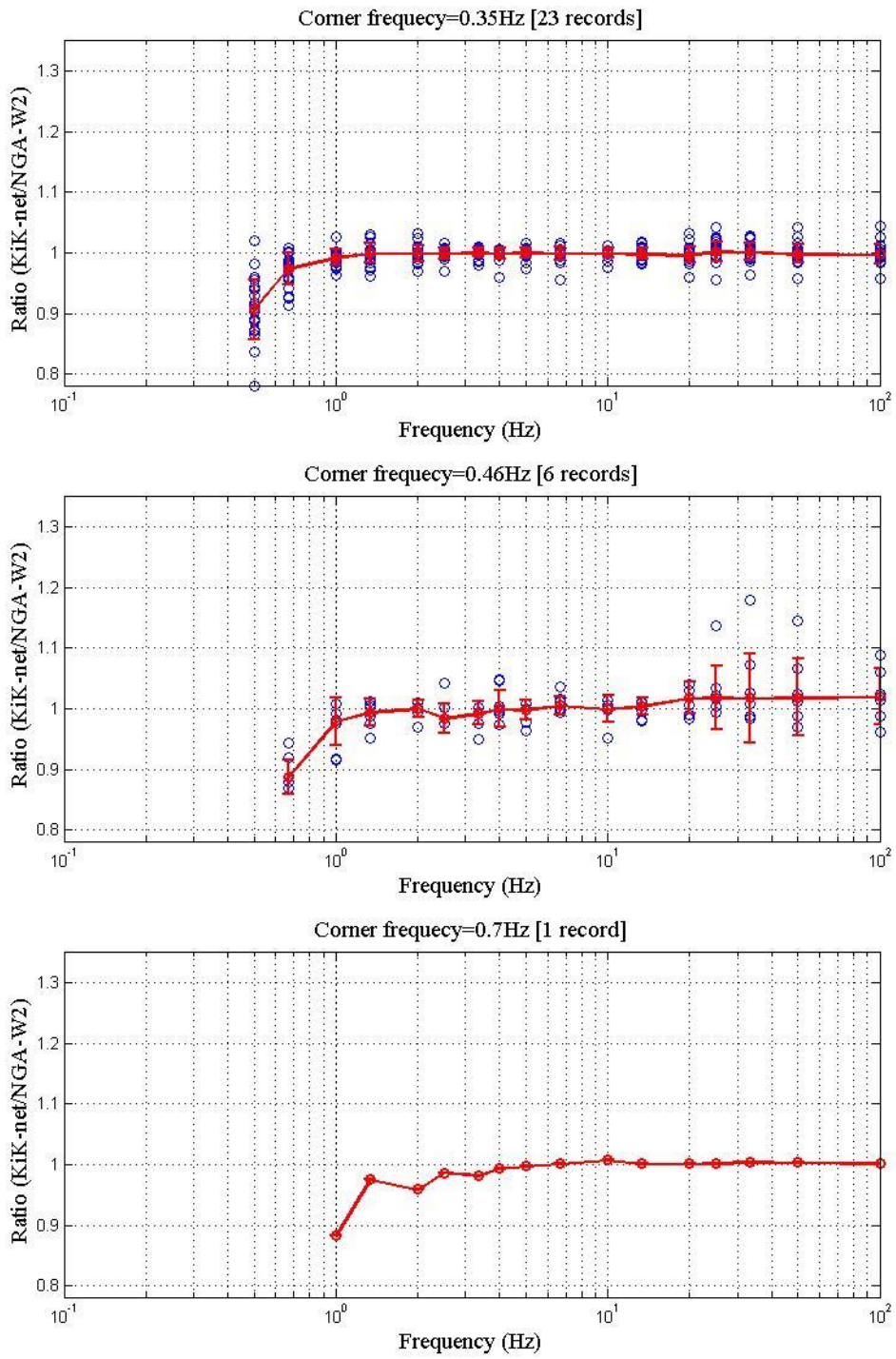


Figure C.3. Scatter plots of the ratio between the PSA ordinates calculated from the automatically processed motions and the same motions obtained from the NGA-W2 database. The median and error bars (one standard deviation above and below the median) of these ratios are also shown (in red). Each subplot contains the records with the same high-pass f_c (as selected by Dea14). The high-pass f_c of the filter is shown in the title of each subplot. The number of records in each plot is also shown.

Inconsistency in High-Cut Filtering

According to the NGA-W2 Flatfile, a high-cut filter was applied to the GM records. This high pass filter is the cause of the inconsistency with the automatically recorded GMs. The instrument response of the KiK-net instruments is flat with a high-cut filter at 30Hz. Hence, both records should show a loss of the energy beyond about 20Hz. However, the ratio between the smoothed FAS of the motions automatically processed and the smoothed FAS of the motions processed within the NGA-W2 project shows differences at high frequencies (Figure C.4). It is clear from that figure that an additional high-cut filter was applied to the GM records within the NGA-W2 processing process. A plot similar to the one presented in Figure C.4 was prepared for records that didn't show a bias in the PSA around 30 Hz showed that a high-cut filter was applied to these records as well. The reason why the usage of the high-cut filter on these records didn't show major effects on the PSA is that the FAS energy at high frequencies is small for these records (GMs recorded far away from the source and/or recorded at soft sites with relatively low V_s).

The bias shown in Figure C.3 around a spectral frequency of 30Hz suggests that care should be taken when using the PSA from Dea14 and the NGA-W2 Flatfile in developing GMPEs at such high frequencies. Moreover, the ratio between the two PSA was plotted versus V_{s30} and epicentral distance to find a potential Magnitude-distance threshold beyond which using the PSA at such high frequencies can be biased. Even if the scatter plots show that at low distances and/or high V_{s30} the ratio tends to be high, the scatter was substantial. This led us to the conclusion that having such criteria is not currently possible.

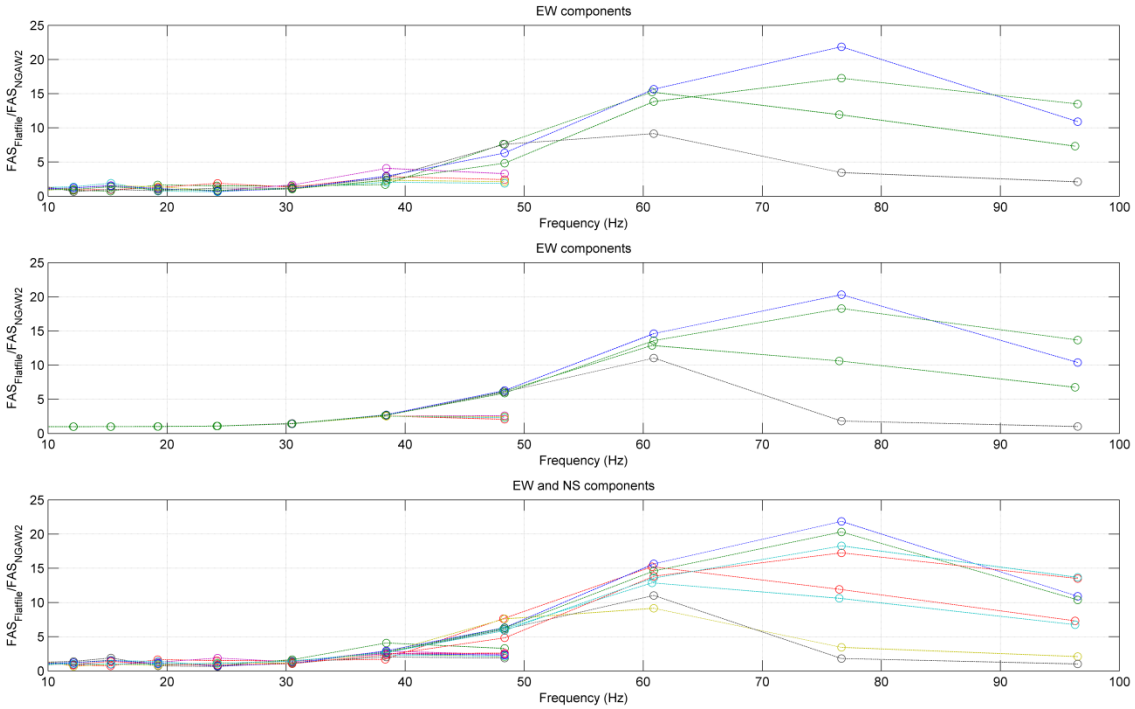


Figure C.4. The ratio between the smoothed FAS obtained from the automatically processed GM records and the smoothed FAS obtained from the NGA-W2 database

Maximum Usable Spectral Period for the Automatically Processed Records

Figure C.5 presents a comparison between the NGA-W2 and Dea14 processing protocols for a common record. The figure compares the acceleration time histories and the smoothed FAS. From the figure it is clear that the automatically applied high-pass filter cuts a larger portion of the FAS at low frequencies compared to the filter applied within the NGA-W2 project. Hence, the bias observed in the PSA at low frequencies (Figure C.3) is likely to be from the result of the effect of the Butterworth filter applied to the motions in the automated protocol.

Figure C.3 shows that if we consider the NGA West 2 records as a guide for selection, the minimum usable spectral frequency is around twice the f_c of the high-pass Butterworth filter (See Table C.1). This was based on the assumption that the PSA from the automatically processed records are acceptable if the median of the ratios ranges between 0.95→1.05 at a

certain spectral frequency. It is important to note that if the acceptance criteria are changed, the minimum usable frequency will change accordingly.

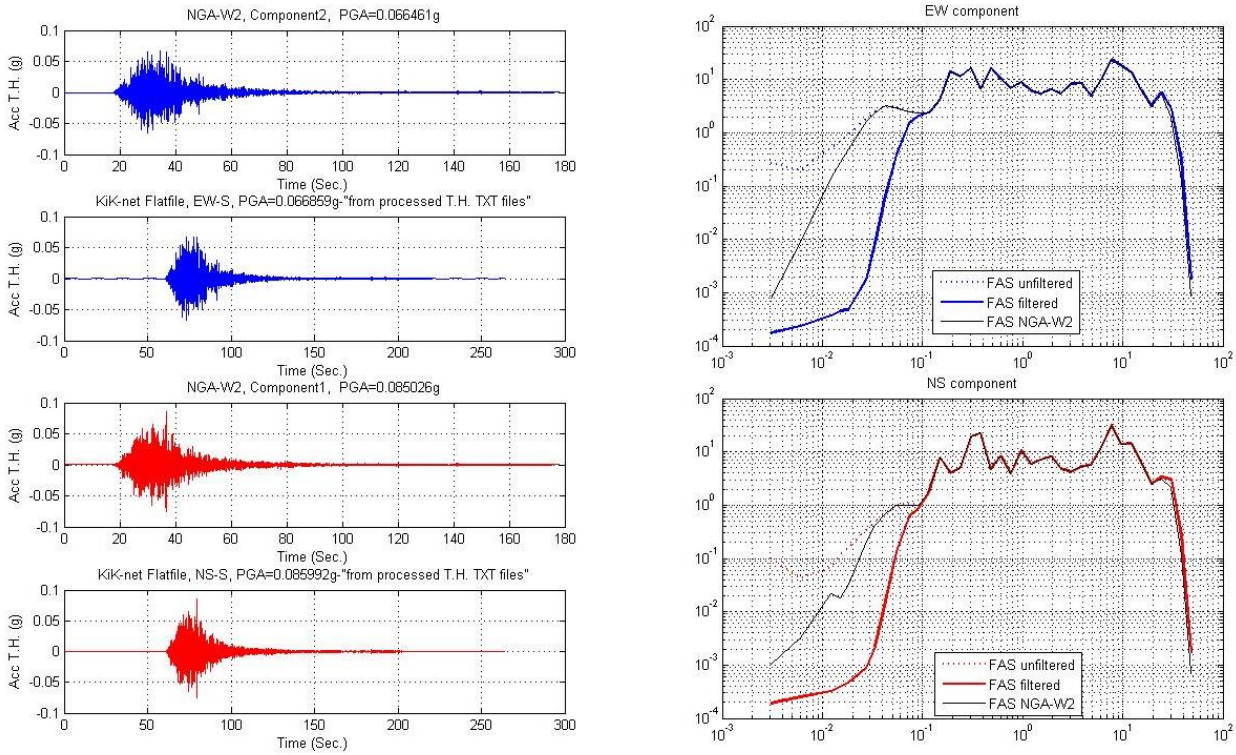


Figure C.5. The figure shows the following for the KiK-net GM # 2042:

I. **<u>Left panels</u>: the acceleration time histories for the automatically processed ground motions (panels 2 and 4 from top) and the records from the NGA-W2 database (panels 1 and 3 from top).**

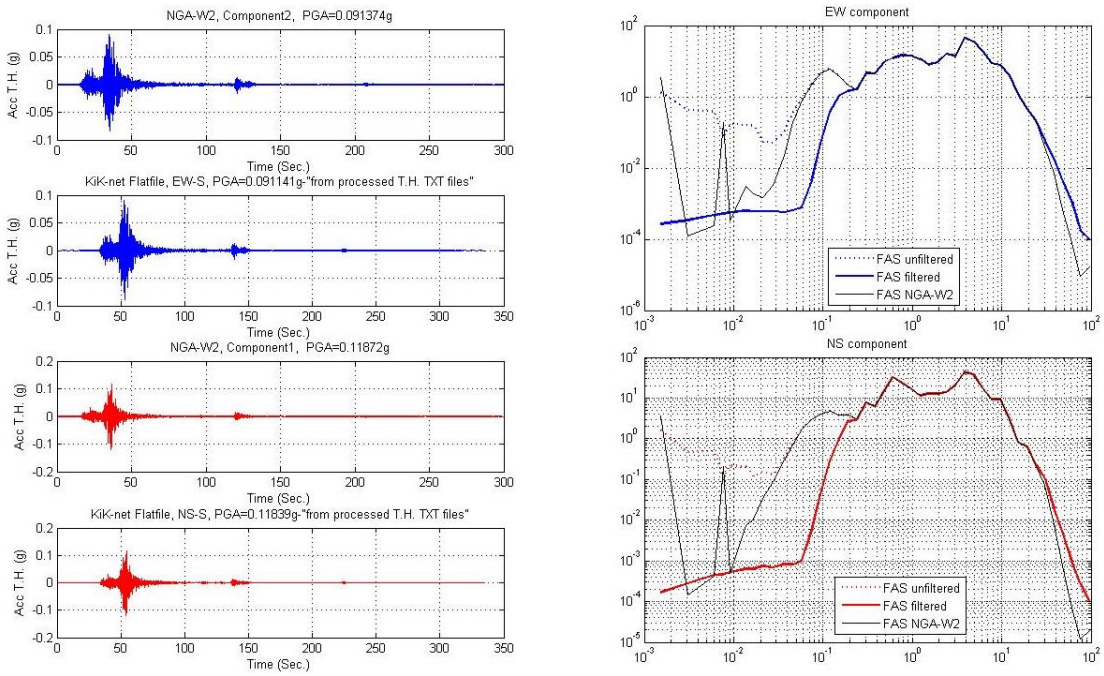
II. **<u>Right panels</u>: the smoothed FAS of the two surface components of motion. Each panel shows the smoothed FAS of the: a) unprocessed record; b) the motion automatically processed; and c) the motion from the NGA-W2 database. These panels show that the automatically picked f_c of the high-pass filter is higher than the one picked within the NGA-W2 project. This is likely the reason of the observed bias at low frequencies (Figure C.3).**

Table C.1. The minimum usable spectral frequencies as a function of the high-pass f_c of the Butterworth filter used to filter the records. This info is obtained by a visual inspection of the scatter plot shown in Figure C.3.

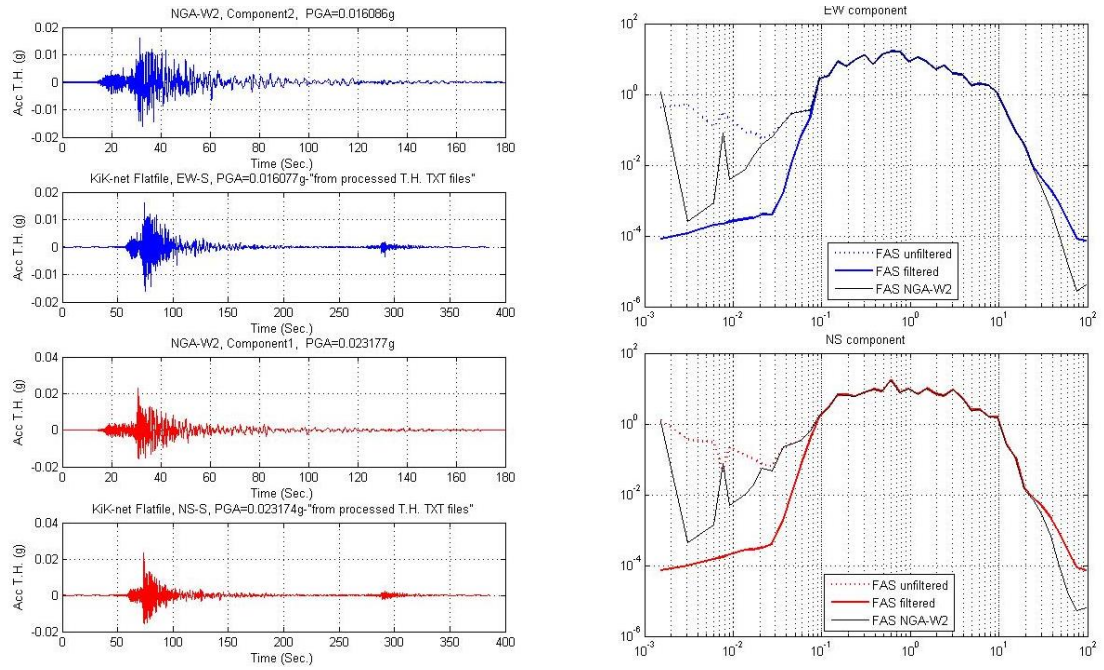
f_c (Hz)	f_{min} (assumed)	f_{min} (scatter plot)	$f_{min,(scatter\ plot)}/f_c$
0.07	0.1	0.133	1.92
0.09	0.133	0.133	1.50
0.14	0.2	0.25	1.79
0.17	0.25	0.33	1.96
0.22	0.33	0.5	2.27
0.35	0.5	0.67	1.92
0.46	0.67	0.98	2.13
0.70	1.0	1.33	1.92

Multiple Wave Trains

Several motions are characterized by multiple wave trains. Using the records with multiple wave trains might result in the distortion of the PSA especially at high frequencies (Strasser and Bommer 2005). In the motions for the NGA-W2 project, the small wave trains were not trimmed for all motions (Figure C.6), which might bias the GMPEs at high frequencies. In the Dea14 protocol, an automated algorithm was developed to flag records with multiple wave trains. Testing of the potential biases due to multiple wave trains is beyond the scope of this report.



(a)



(b)

Figure C.6. Acceleration time histories (left panels) and the smoothed FAS (right panel) of the unfiltered and processed records in the NGA-W2 database and the Dea14 database for: a) KiK-net GM #20449: The two wave trains following the mainshock were not removed while processing the GM record for the NGA-W2 project; and b) KiK-net GM#39844: The second wave train was removed while processing the GM records for the NGA-W2 project.

COMPARISON WITH BC-HYDRO MOTIONS

Scatter plots similar to the ones presented in Figures C.1 and C.3 were plotted for the common records between the databases of Dea14 and Zea06 (used within the BC-Hydro project). A few records have very large differences. These records are listed in Table (2) along with the potential reason for the differences. Figure C.7 shows the ratios of the PSA of some of the records listed in Table C.2. Figure C.8 shows the time histories and PSA of some of these records. The following list of observations was compiled from a more in-depth inspection of the records:

- The high-pass f_c of the filter used by Zea06 is in most cases considerably higher than the ones automatically picked within the protocol by Dea14. The opposite was observed with the NGA-W2 processed records.

- Zea06 mentioned the followings statement "*If a record had relatively large values at the beginning (P-wave) and the end of the record, the record was mirrored and tapered for 5 sec at each end of the record*". An example for such case wasn't presented by Zea06. However, applying such a tapper might be possibly causing the large differences in PSA observed at both ends of the spectra for records with remark "A" in Table C.2. See, for example, Figure C.8-a.

- The Tukey window applied in the automatic algorithm trimmed a large portion of the EW component from KiK-net GM# 67935 frequencies (records with remark "B" in Table C.2) because the noise window was very small for this component. See, for example, Figure C.8-b. It is believed that this problem will not affect the overall quality of the Deal14 Flatfile since this was the only case found among hundreds of records investigated.

- For several records, the PSA of the automatically processed GM records was higher than the PSA from Zea06 at low frequencies (records with remark "C" in Table C.2). See, for example, Figure C.8-c. It is possible that the bias in low frequencies is caused by the high-pass filter applied by Zea06. Even if the bias at low frequencies is observed for almost all records, it is not yet clear why this is very prominent for this small subset of records. It is important to note that BC-Hydro project used 125% of the f_c of the high-pass filter as the minimum usable spectral frequency. This validity of using 125% of f_c at the minimum usable frequency was tested. The calculated ratios between the PSA from Dea14 and Zea06 show that this assumption is in general valid. It is important to note that for some records the ratio was beyond 1.1 at low frequencies.

- In some cases it was not possible to find potential reasons as of why there is a difference in the PSA frequencies (records with remark “D” in Table C.2). See, for example, Figure C.8-d.

- If the time histories prepared by Zea06 were available, the conclusions in this section might have been more specific.

Table C.2. The list of records characterized by large differences in the PSA calculated for the motions automatically processed and processed by Zea06

KiK-net GM#	BC-Hydro GM#	f_c (Hz) (Automatically picked)	f_c (Hz) (Zea06)	Large ratios at high (H) or low (L) frequencies?	Remarks
1677	11972	0.07	0.14	H	A
2878	11987	0.07	0.16	H	A
24604	12020	0.07	0.10	L	C
26282	12026	0.07	0.14	L	C
34462	12030	0.07	0.10	L	C
35798	12227	0.07	0.10	L	C
58606	12243	0.07	0.10	L	C
1581	11971	0.09	0.10	H&L	A
1757	11973	0.09	0.10	L	C
23613	12016	0.09	0.10	L	C
25675	12024	0.09	0.14	L	C
25373	12214	0.09	0.14	L	C
67935	12215	0.09	0.13	H&L	B
26000	12216	0.09	0.13	L	C
34487	12222	0.09	0.13	L	C
31789	13222	0.09	0.30	L	C
24815	12021	0.14	0.18	H&L	D
34065	12029	0.14	0.16	L	C
21410	12196	0.14	0.40	H&L	D
30799	12920	0.14	0.10	L	C
2668	11985	0.17	0.14	H	A

List of remarks listed in Table C.2:

A: Maybe a portion of the P-wave was trimmed by Zea06 (Figure C.8-a).

B: The Tukey window applied in the automatic algorithm trimmed a large portion of the EW component because the noise window was very small for this component (Figure C.8-b).

C: The PSA at long periods from Zea06 is found to be lower than the PSA from the automated protocol for most spectral periods. This might be the effect of the high-pass filter used by Zea06 (Figure C.8-c).

D: It is not clear why there is such a difference in the PSA (Figure C.8-d).

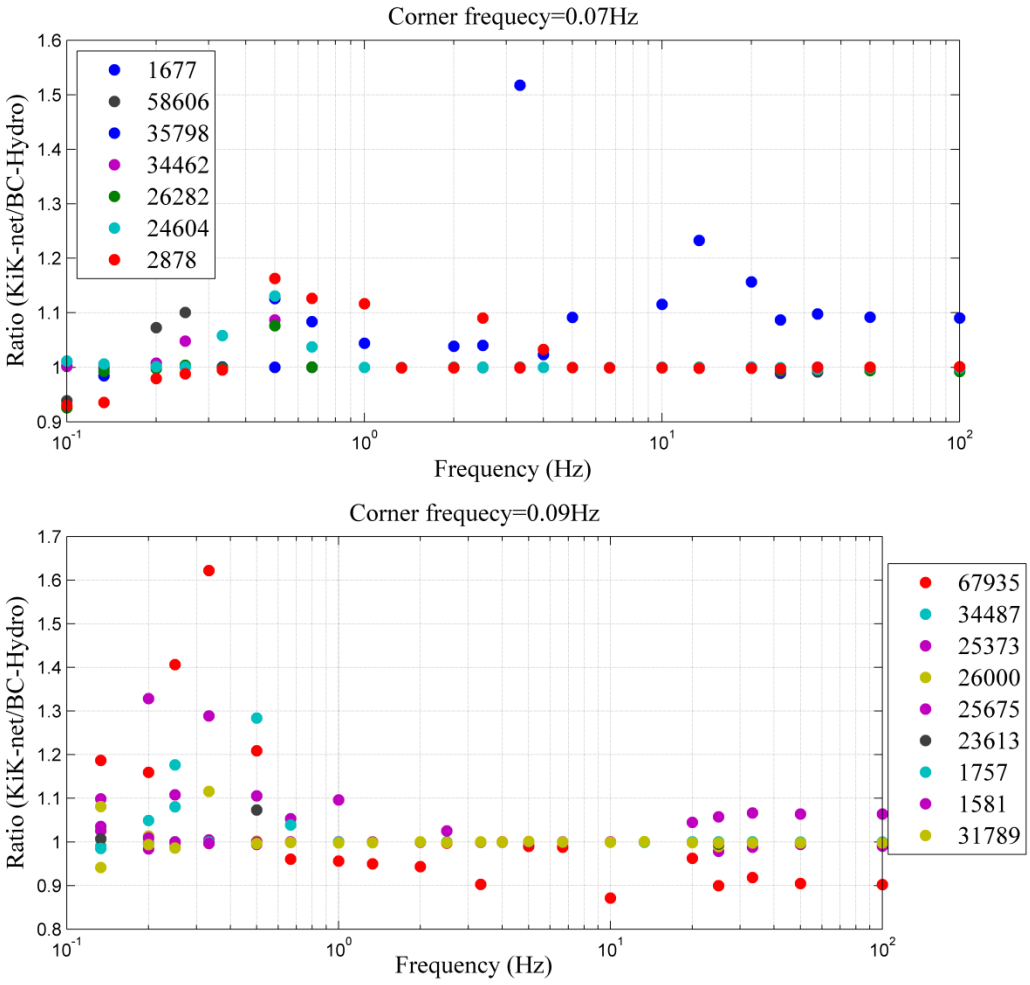
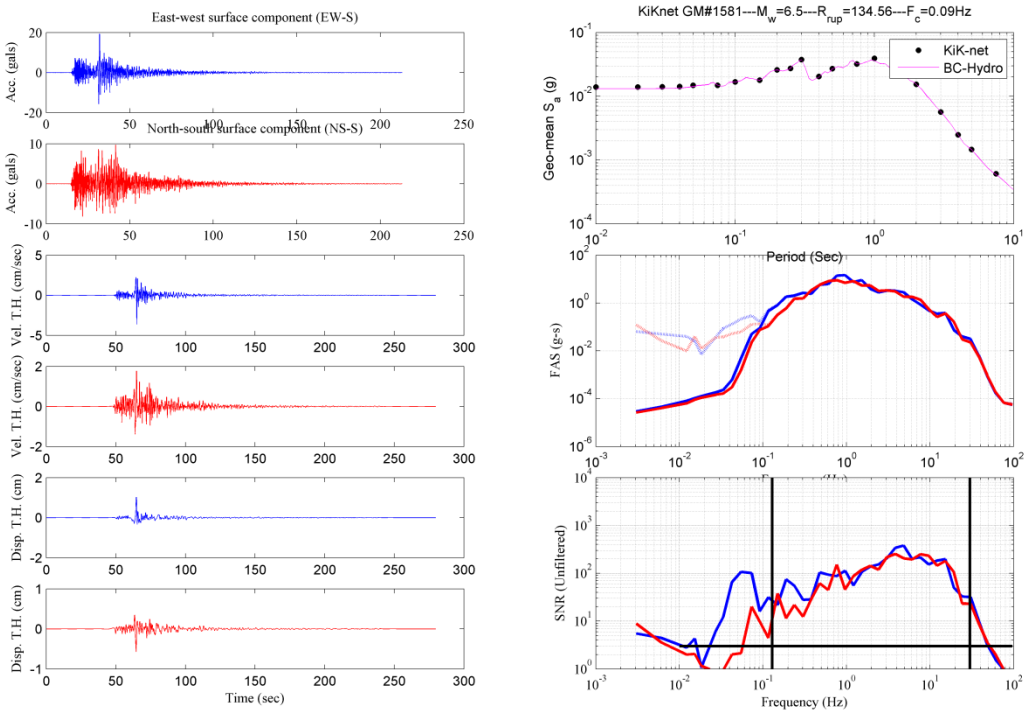
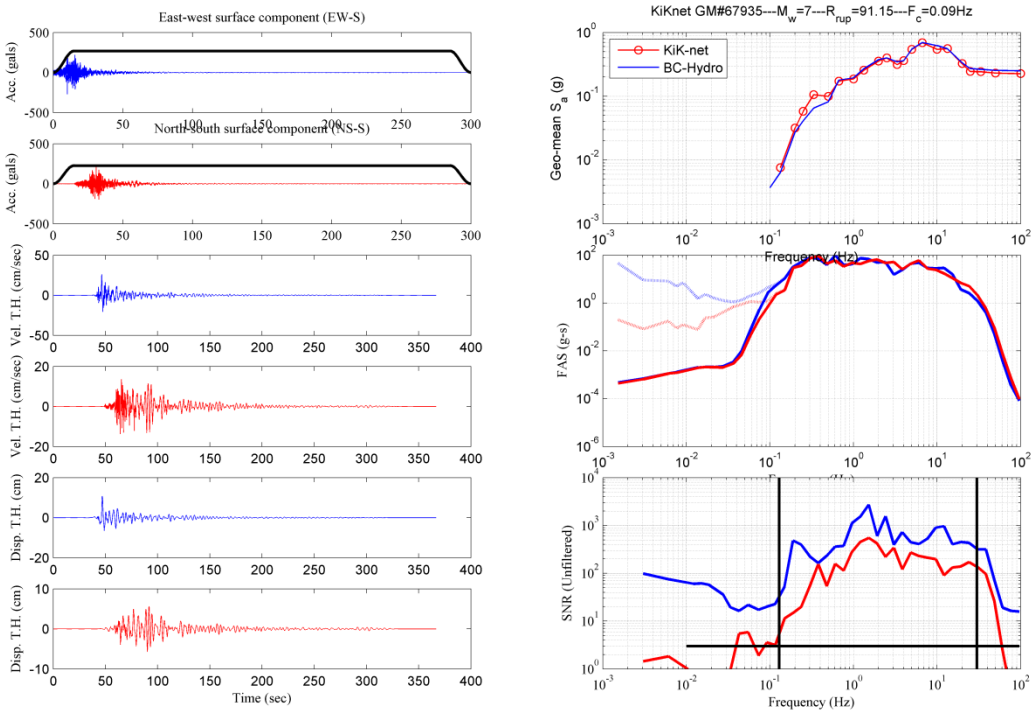


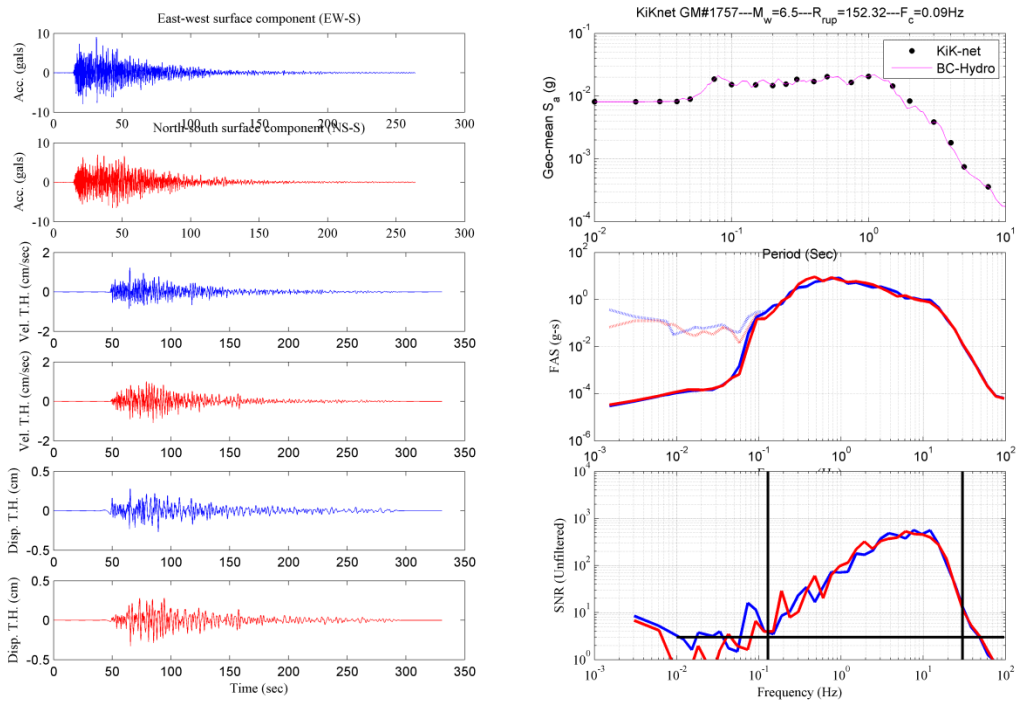
Figure C.7. The scatter plots of the ratio between the PSA from the automatically processed records and the same records processed by Zea06. This figure includes some of the records listed in Table C.2.


East-west surface component (EW-S)
North-south surface component (NS-S)
Acc. (gals)
Vel. T.H. (cm/sec)
Disp. T.H. (cm)
Time (sec)
KiKnet GM#1581---M_w=6.5---R_rup=134.56---F_c=0.09Hz
KiK-net
BC-Hydro
Geo-mean S_a (g)
Period (Sec)
FAS (g-s)
SNR (Unfiltered)
Frequency (Hz)

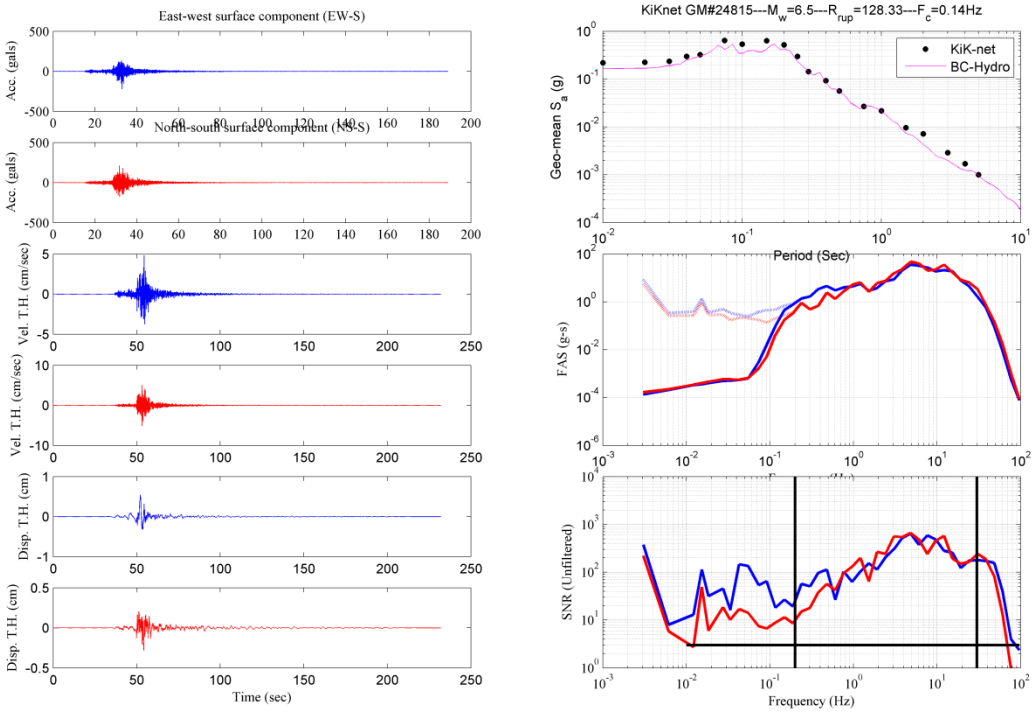
(a)


East-west surface component (EW-S)
North-south surface component (NS-S)
Acc. (gals)
Vel. T.H. (cm/sec)
Disp. T.H. (cm)
Time (sec)
KiKnet GM#67935---M_w=7---R_rup=91.15---F_c=0.09Hz
KiK-net
BC-Hydro
Geo-mean S_a (g)
Frequency (Hz)
FAS (g-s)
SNR (Unfiltered)

(b)



(c)

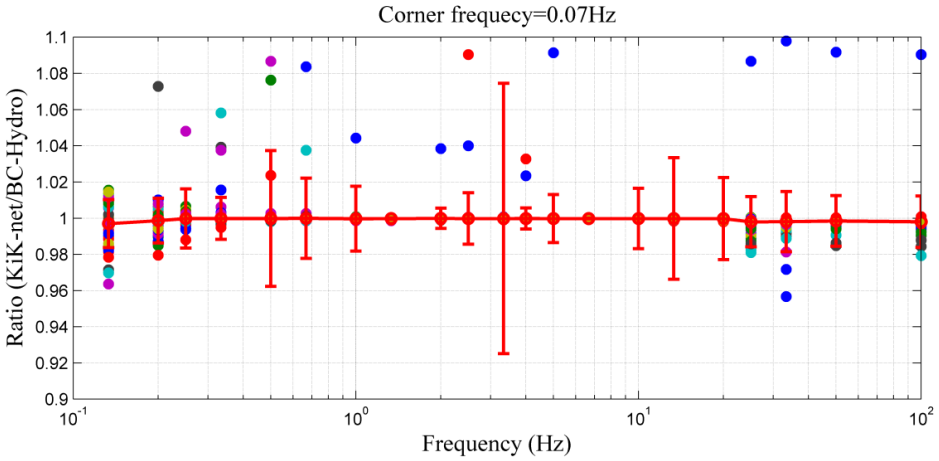


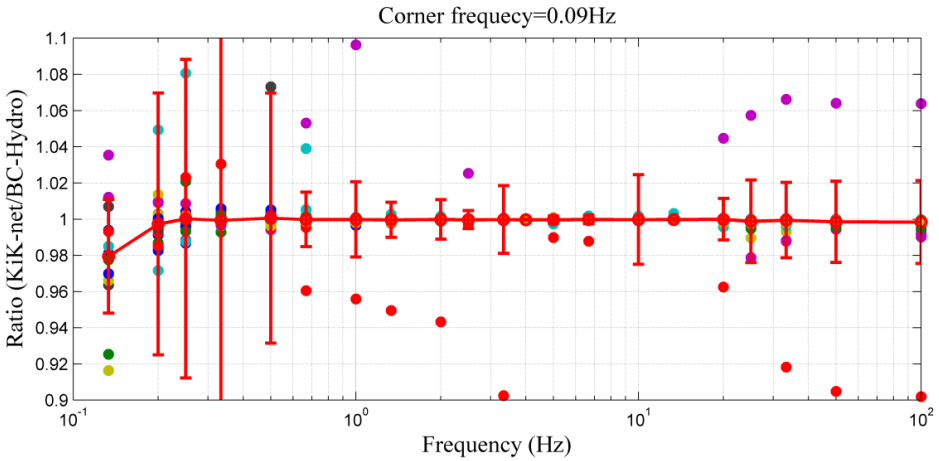
(d)

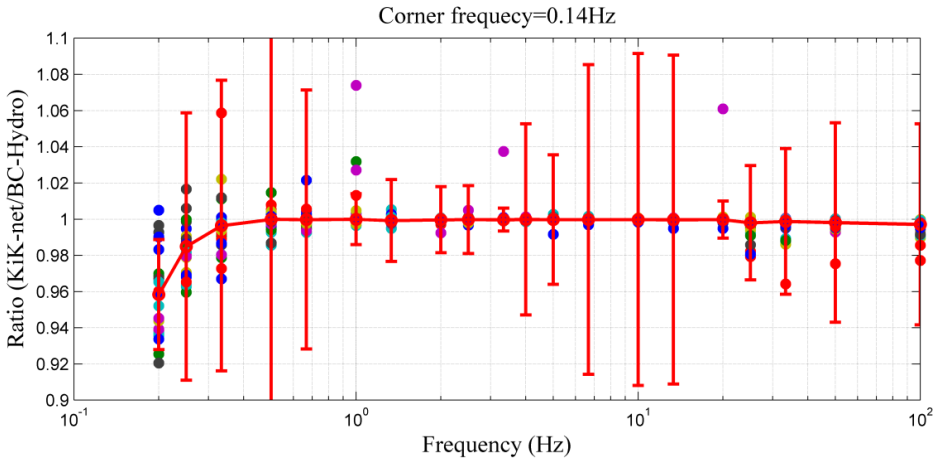
Figure C.8. Time histories, PSA, FAS and SNR for some of the motions listed in Table C2. In particular the four plots serve as examples for the remarks A, B, C, and D, respectively (details of the remarks are given in Table C.2).

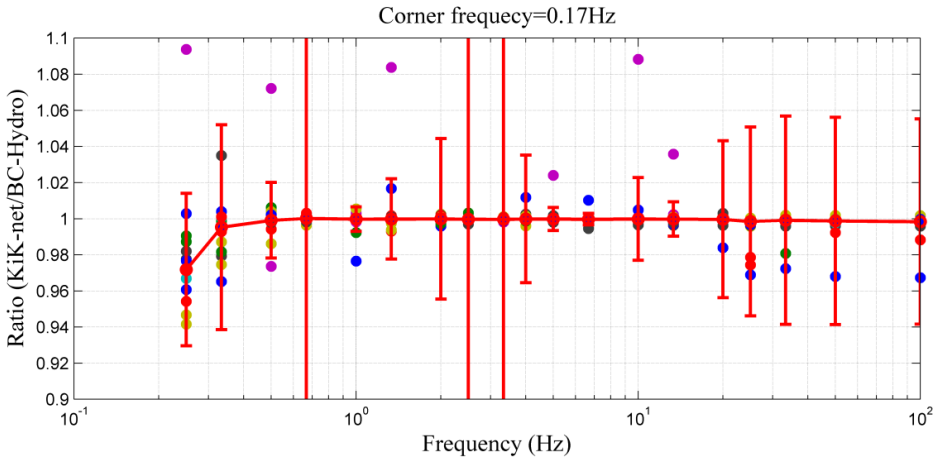
Figure C.9 shows the ratio between the PSA obtained from Dea14 database to the PSA from the Zea06 database as a scatter plot (note that the y-axis is limited to range between 1.1 and 0.9 to make the bias clearer). Each subplot contains all records that had been filtered with a high-pass filter with the same f_c (for the Dea14 records). Even if the erratic records listed in Table C.2 are included in the scatter plots, the median bias is minimal except at low frequencies. The bias at low frequency is slightly towards a ratio less than one. While in the comparison to the NGA-W2 database (section 2) the ratio for the individual records tended to be less than one, in the comparison with the Zea06 many of the records have ratios greater than one (Section 2 of this report). This could be attributed to the relatively large high-pass f_c used by Zea06 to filter the records. Table C.3 suggests that the maximum usable spectral frequency is about 143% of the f_c of the high-pass filter. This info is based on a visual inspection of the scatter plot shown in Figure C.9. However, it is also true that there is evidence that the relatively high f_c values of the filter adopted by Zea06 could potentially bias the PSA low at low spectral frequencies. This is especially true because in most cases its corner frequency is higher than the one automatically picked.

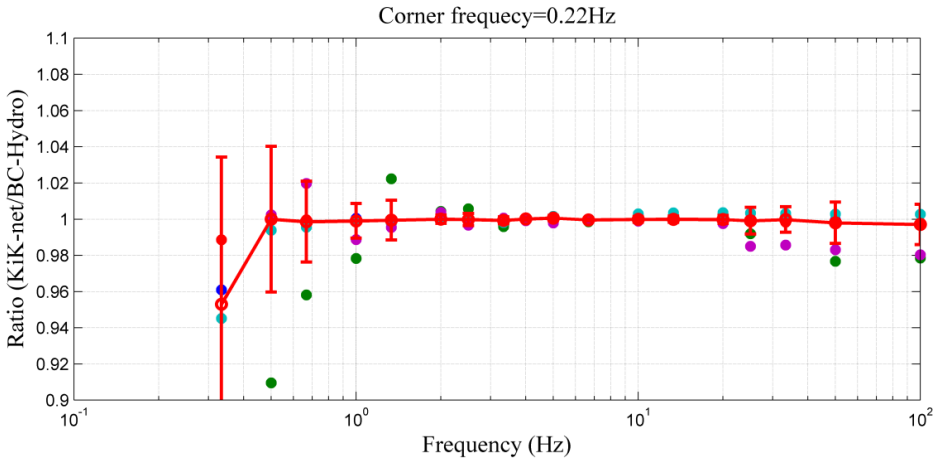
The comparisons with the GM processing algorithms in the NGAW2 and Zea06 can help the determination of the ideal ratio of high-pass filter frequency to use as a lower bound of the "usable bandwidth". Due to the considerations listed in the paragraph above, we consider that the comparisons with the NGA-W2 database are more reliable and conservative, hence we adopt a definition of 2*f_c for the lowest usable frequency for spectral acceleration for the automatically prepared PSA.

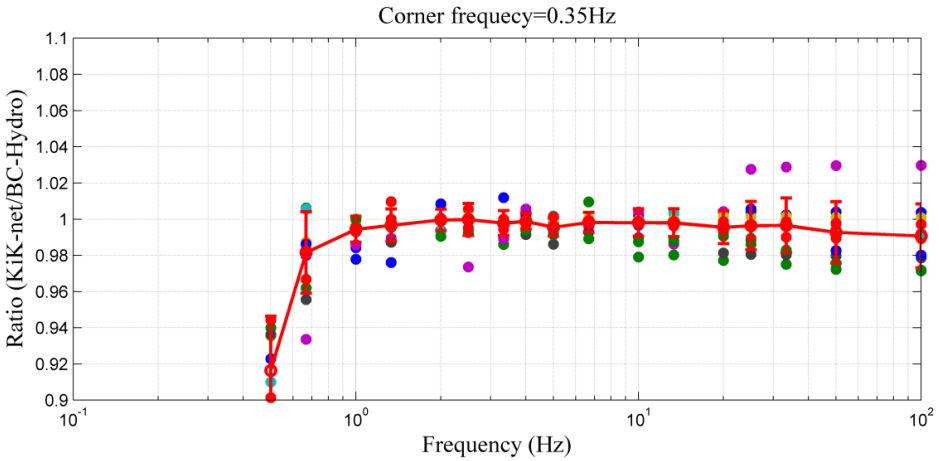


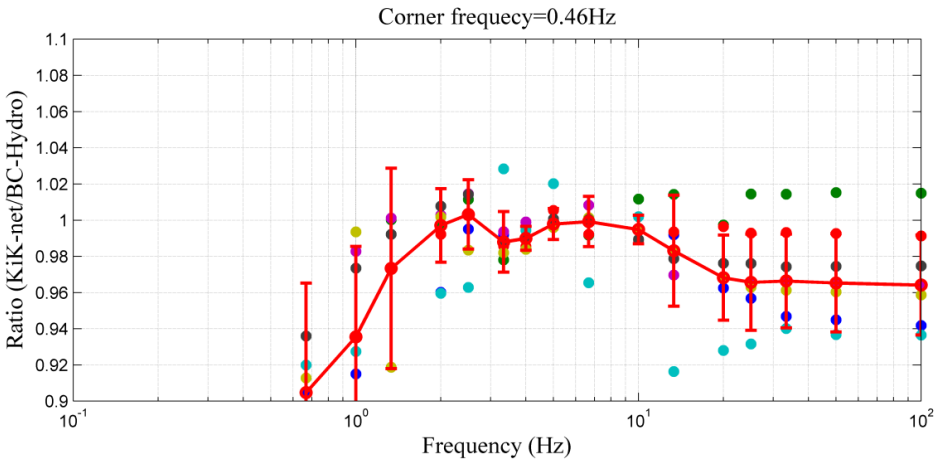

Corner frequecy=0.09Hz
Ratio (KiK-net/BC-Hydro)
Frequency (Hz)


Corner frequecy=0.14Hz
Ratio (KiK-net/BC-Hydro)
Frequency (Hz)


Corner frequecy=0.17Hz
Ratio (KiK-net/BC-Hydro)
Frequency (Hz)


Corner frequecy=0.22Hz
Ratio (KiK-net/BC-Hydro)
Frequency (Hz)


Corner frequecy=0.35Hz
Ratio (KiK-net/BC-Hydro)
Frequency (Hz)


Corner frequecy=0.46Hz
Ratio (KiK-net/BC-Hydro)
Frequency (Hz)

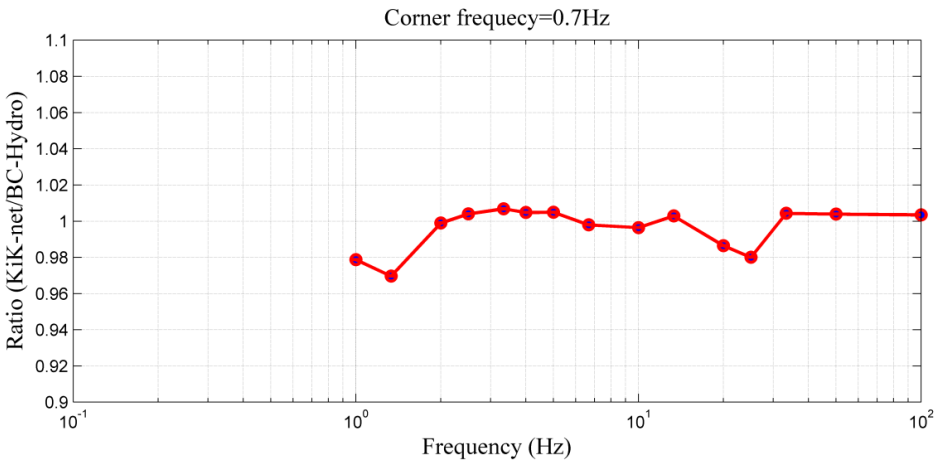


Figure C.9. Scatter plot of the ratio between the PSA ordinates calculated from the automatically processed motions and the same motions obtained from the BC-Hydro database. The median and error bars (one standard deviation above and below the median) of these ratios are shown (in red). Each subplot contains the records filtered using a filter with the same high-pass corner frequency (in the automated processing protocol). The high-pass corner frequency of the filter is shown in the title of each subplot

Table C.3. The minimum usable spectral frequencies as a function of the f_c of the high-pass Butterworth filter used to filter the records. This info is obtained by a visual inspection of the scatter plot shown in Figure C.9

f_c (Hz)	f_{min} (assumed)	f_{min} (scatter plot)	$f_{min,(scatter\ plot)}$ /f_c
0.07	0.10	0.10	1.43
0.09	0.13	0.13	1.43
0.14	0.20	0.20	1.43
0.17	0.25	0.25	1.43
0.22	0.33	0.33	1.43
0.35	0.50	0.67	1.92
0.46	0.67	1.33	2.86
0.70	1.00	1.00	1.43

COMPARISON WITH NGA-W2 METADATA

The Flatfile developed by Dea14 includes a set of metadata associated with each GM record. This section of the report is intended to present a comparison of the metadata compiled by Dea14 with the metadata by NGA-W2. Figures C.10 toC.13 show the values compiled in both databases for several important parameters.

Figure C.10 compares the different earthquake parameters obtained from the two databases. Only four events are common to both databases. The M_w presented in the first subplots are the M_w from the NGA-W2 flatfile and the M_w compiled by Dea14 from readily published finite source models for these four seismic events. The comparison shows that the difference in M_w is in general less than 0.1. The second subplot shows that in general the depth of the hypocenter reported by NGA-W2 is larger than the depth compiled by Dea14. In this case, the depth reported by Dea14 is obtained from the F-net seismic catalog. The differences in depth are less than 2 km for three events, while it is about 6 km for the fourth one. The third and fourth subplots compare the depth to the top-of-rupture (Z_{TOR}). The comparison is almost perfect for three events (the exception is F-net EQ# 8880) when Z_{TOR} compiled from published finite source models (3^{rd} subplot). On the other hand, a clear bias is noticed when the simulated rupture surfaces (simulations conducted by Dea14) was used (4^{th} subplot).

Figure C.11 compares the reported V_{s30} for common station between the two databases. The match is almost perfect except for few sites. Table C.4 shows the information of the five sites with the largest differences. The V_{s30} reported by NGA-W2 for four of these stations didn't match the profile reported in KiK-net (see screenshots in Table C.4). On the other hand, the V_{s30} reported for NGNH37 in Dea14 is not correct. From a closer investigation, it was found that the profile currently available from the KiK-net website is different from the one downloaded in 2011 for that station.

Figure C.12 compares the different distance measurements (R_{rup}, R_{JB}, and R_x) presented in the NGA-W2 and Dea14. The scatter plots show that in general there is a good agreement between the two databases.

Figure C.13 compares the rake, strike and dip of the fault. The rake and dip seem to match well even though the match of the strikes is not as good.

Table C.4: A list of stations for which the reported V_{s30} is different between NGA-W2 and Dea14

Station	V_{s30} (NGA-W2)	V_{s30} (Dea14)	Screenshot from the geotech info provided by the KiK-net website
OKYH13	598.28	1042	No Thickness (m) Depth (m) Vp (m/s) Vs (m/s) 1, 2.00, 2.00, 300.00, 160.00 2, 8.00, 10.00, 2800.00, 950.00 3, -----, -----, 5230.00, 2540.00
NGNH37	338.37	378	No Thickness (m) Depth (m) Vp (m/s) Vs (m/s) 1, 2.00, 2.00, 230.00, 120.00 2, 8.00, 10.00, 450.00, 210.00 3, 95.00, 105.00, 2000.00, 590.00 4, 125.00, 230.00, 2500.00, 930.00 5, 280.00, 510.00, 2300.00, 930.00 6, 60.00, 570.00, 2700.00, 1400.00 7, -----, -----, 3200.00, 1700.00
NGNH28	518.1	586.5	No Thickness (m) Depth (m) Vp (m/s) Vs (m/s) 1, 2.00, 2.00, 440.00, 120.00 2, 5.00, 7.00, 1430.00, 560.00 3, -----, -----, 2130.00, 900.00
NIGH06	493.27	336.1	No Thickness (m) Depth (m) Vp (m/s) Vs (m/s) 1, 4.00, 4.00, 400.00, 100.00 2, 6.00, 10.00, 1500.00, 270.00 3, -----, -----, 1870.00, 740.00
NGNH27	562.18	381.2	No Thickness (m) Depth (m) Vp (m/s) Vs (m/s) 1, 2.00, 2.00, 600.00, 170.00 2, 8.00, 10.00, 1070.00, 300.00 3, 10.00, 20.00, 1430.00, 410.00 4, -----, -----, 2130.00, 630.00

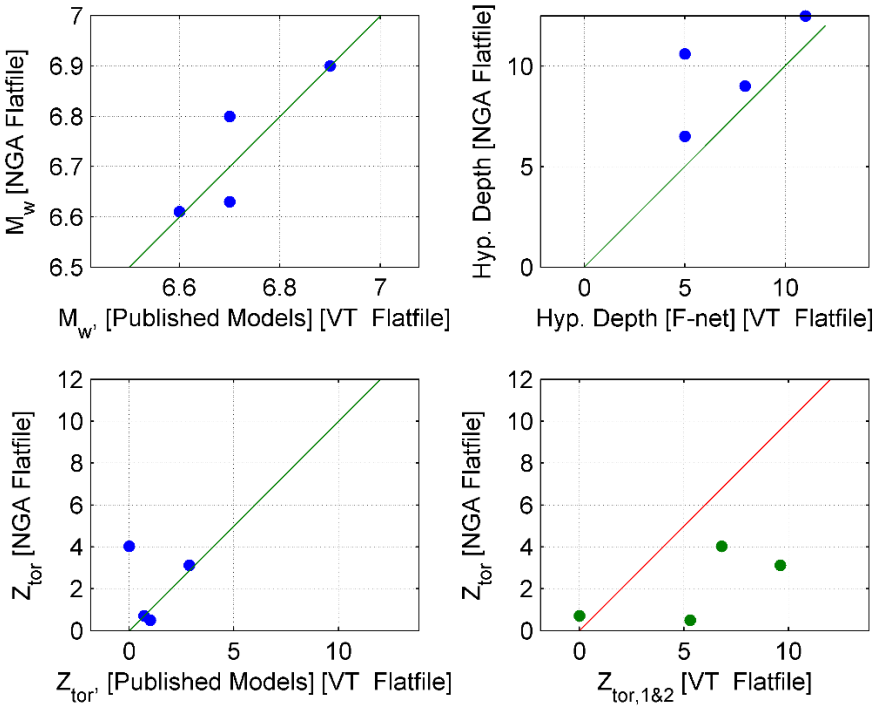


Figure C.10. Comparison between the values reported in NGA-W2 and Dea14 for: M_w (top left panel); depth of hypocenter (top right panel); and depth to top-of-rupture (bottom panels)

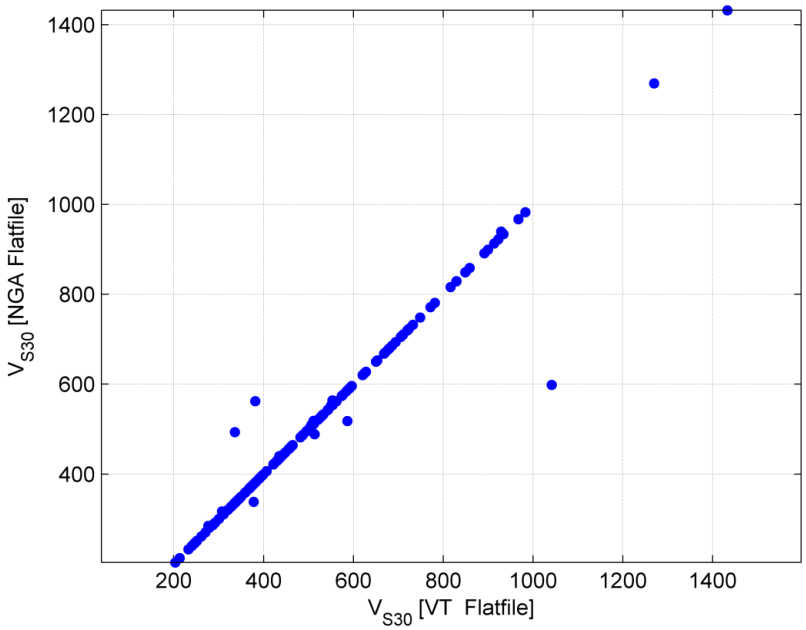


Figure C.11. Comparison between the V_{s30} reported in NGA-W2 and Dea14. The stations whete the values were significantly different are listed in Table (4)

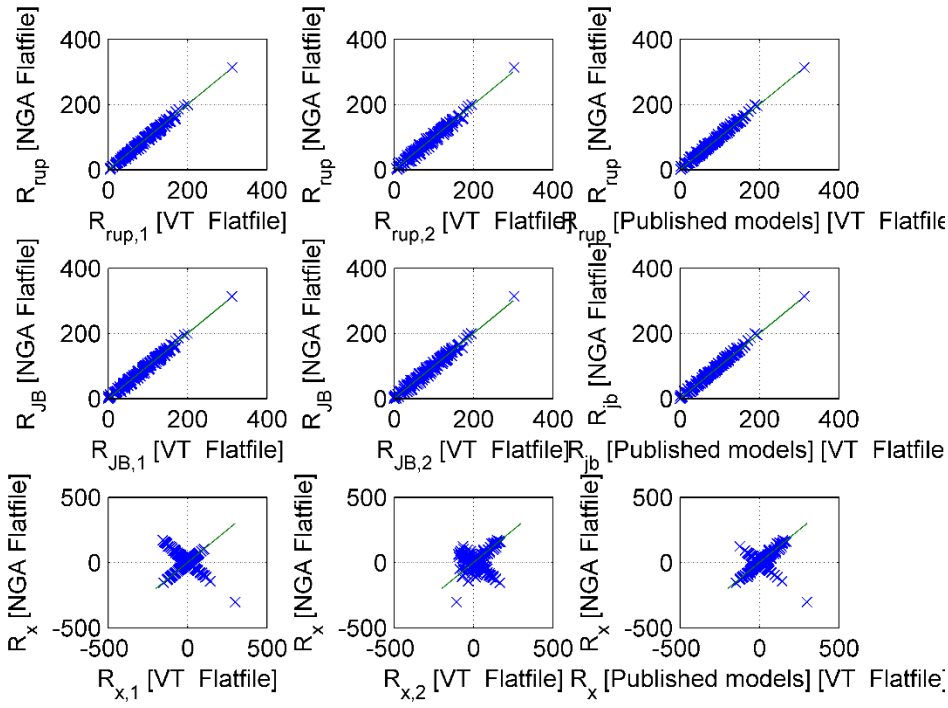


Figure C.12. Comparison between the different distance measurements (R_{rup}, R_{JB} and R_x) reported in NGA-W2 and Dea14.

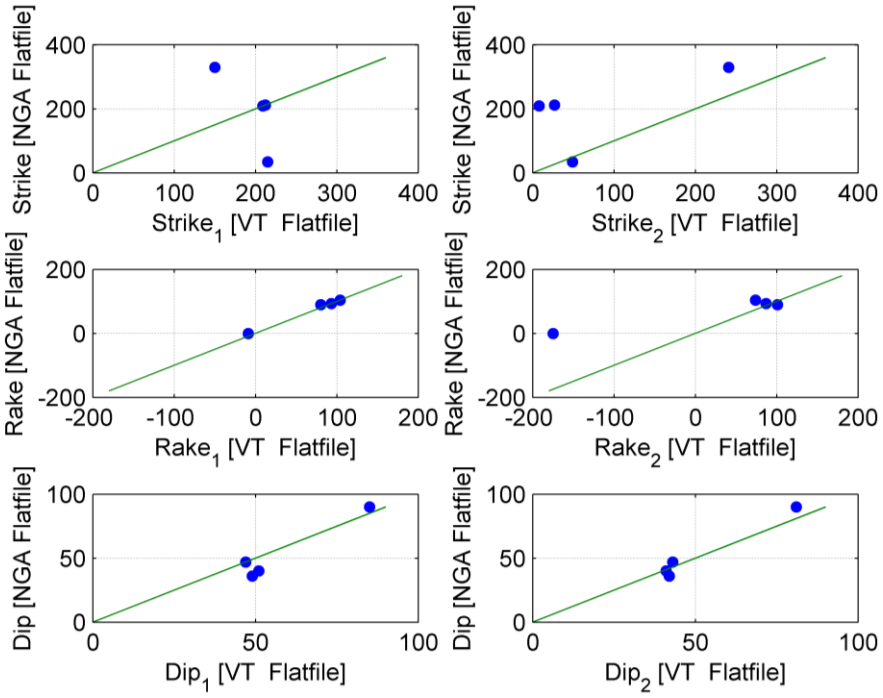


Figure C.13. Comparison between the stikes, rakes, and dips reported in NGA-W2 and Dea14.

CONCLUSIONS

The following are some conclusions based on the comparisons shown in this appendix:

Automatically processed records:

- Zea06 trimmed a part of the P-wave arrival when it was strong compared with the S-wave. This wasn't included in the automated protocol. This can potentially influence the resulting PSA ordinates.

- The Tukey window used in the automated processing protocol can affect the time history especially for cases where the background noise window is limited. However, this problem doesn't seem to be prominent as most of the GM records have a wide enough background noise compared to the portion of the record affected by the Tukey window.

NGA-W2:

- The records with multiple wave trains were not systematically excluded or dealt with. This might lead to distortions in the PSA obtained from such records especially at high frequencies.

- A high-cut filter was applied to the records. This reduce the PSA around 30Hz for some records. Using PSA around 30Hz calculated from the processed records can bias the regression coefficients of the GMPEs at such frequencies.

BC-Hydro:

- The BC-Hydro project used spectral frequencies up to 125% of the f_c of the high-pass filter. This ratio seems quite high for some records based on the scatter observed for the records processed by Zea06. On the other hand, the median of the ratios doesn't seem to be highly biased.

- The PSA of some records processed by Zea06 didn't match the PSA obtained from the automated processing protocol. A possibility is that we are comparing two different records. Other reasons for this discrepancy are not clear.

- The conclusions based on comparing records processed by Zea06 with the automatically processed motions are not final. This is because of the absence of the acceleration time histories for the records processed by Zea06.

- Figure C.9 shows that the ratio between the PSA is less than unit at high frequencies. This might be reflecting the fact that Zea06 mentioned that they corrected for the instrument response. On the other hand, they mentioned also that they applied a high-cut filter. This issue should be investigated to get a final answer.

Appendix D: Empirical Ground Motion Prediction Equations for Active Crustal Earthquakes Using the Japanese KiK-net Database: Ergodic and Site-Specific Formulations: Supplementary Information

Appendix D

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

INTRODUCTION

This appendix provides additional information for the analyses presented in Chapter 5 of this dissertation (Published in Dawood and Rodriguez-Marek 2014). The appendix starts with describing the different steps carried out to choose the subset of the motions used in the regression analysis. Then, a brief summary about some of the reasons behind the changes made to the functional form and regression steps presented by Abrahamson et al. (2013). The different steps carried out to regress for the different coefficients are presented. Plots that show the different regression coefficients before and after smoothing as well as plots that compare the coefficients obtained in this study with the ones reported by Abrahamson et al. (2013) are then shown. We then present histograms of the differences between the predictions from the ergodic and site-specific formulations, scatter plots of the different residual components. We end this appendix by discussing the fat tails observed in some residual components.

STEPS FOR DATA SELECTION

The database used in the analysis presented in Chapter 5 of this dissertation is a subset of the database described in Chapter 3. Table D.1 shows the different preset criteria applied to choose the final subset and the number of ground motion records and earthquakes in the subset after each step is applied.

In most cases the steps (criteria) presented in Table D.1 are self-explanatory. Hence, only steps that need additional details are the ones discussed as follows:

Step 4:

In this step we remove the motions for which the peak ground acceleration (PGA) of the filtered motion is more than 2% different than the PGA reported in the KiK-net data files. This step ensures that we only use motions that were not compromised by the processing protocol.

Step 7:

In this step, the motions with a match category (see Chapter 3) other than "*A*" with an F-net earthquake are eliminated. Except for 487 motions associated with 12 events with M_w between 5.0 and 6.8 that have a matching category other than "*A*". The suitableness of the match of these 12 earthquakes was done one-by-one to ensure that these records are associated with the correct event. This was done due to the importance of the motions associated with these large events for the regression analysis. The following are the events with which the records were associated and the number of records that were kept in the database used in Chapter 3 between parentheses: F-net EQID 4055 (122), 8629(94), 19782 (8), 18927 (7), 6730 (75), 18778 (48), 6725 (22), 17912 (22), 6742 (19), 19250 (26), 19500 (16), and 17607 (28).

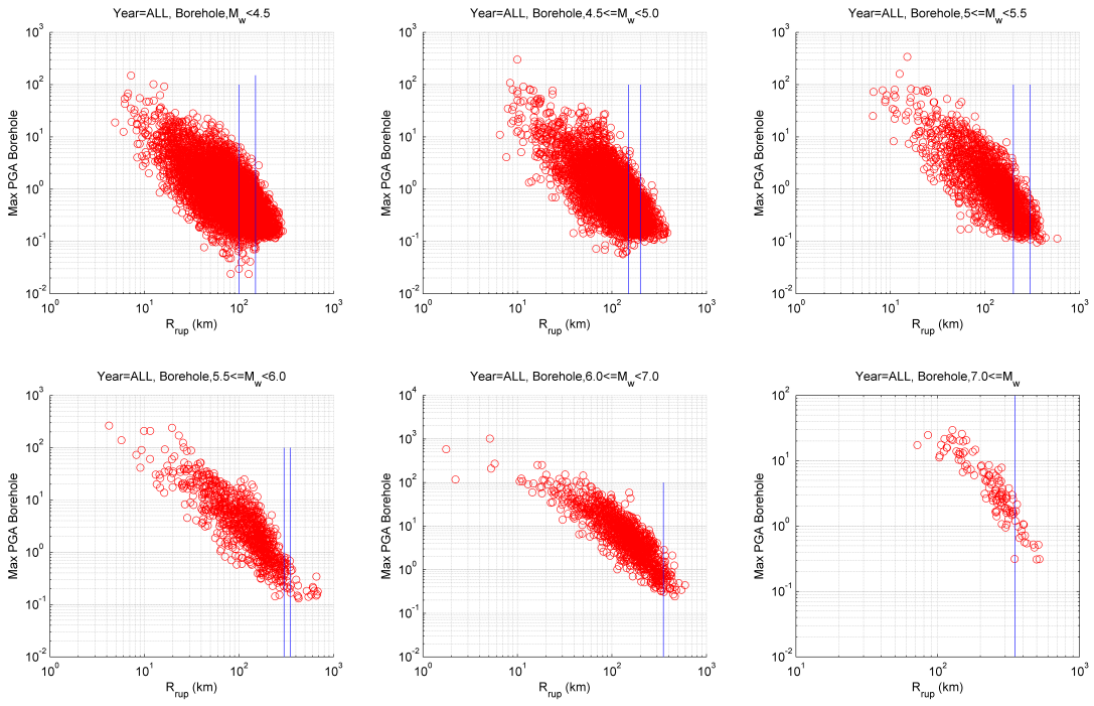
Step 8:

Censoring was conducted in a similar manner as described in Abrahamson et al. (2013). Two different censoring models were developed based on visual inspection of the scatter plots of the maximum PGA for different magnitude bins (Figure D.1). The models were obtained based on the maximum of the two horizontal PGA components recorded by the borehole, and surface instruments. Figure D.2 shows the maximum R_{rup} versus M_w obtained from visually inspecting the plots in Figure D.1. We chose to use the one based on surface motions. Figure D.3 shows the scatter plots of R_{rup} versus M_w for the records kept in the database and the censored records. Applying this censoring model to the database resulted in removing a total of 2046 motions.

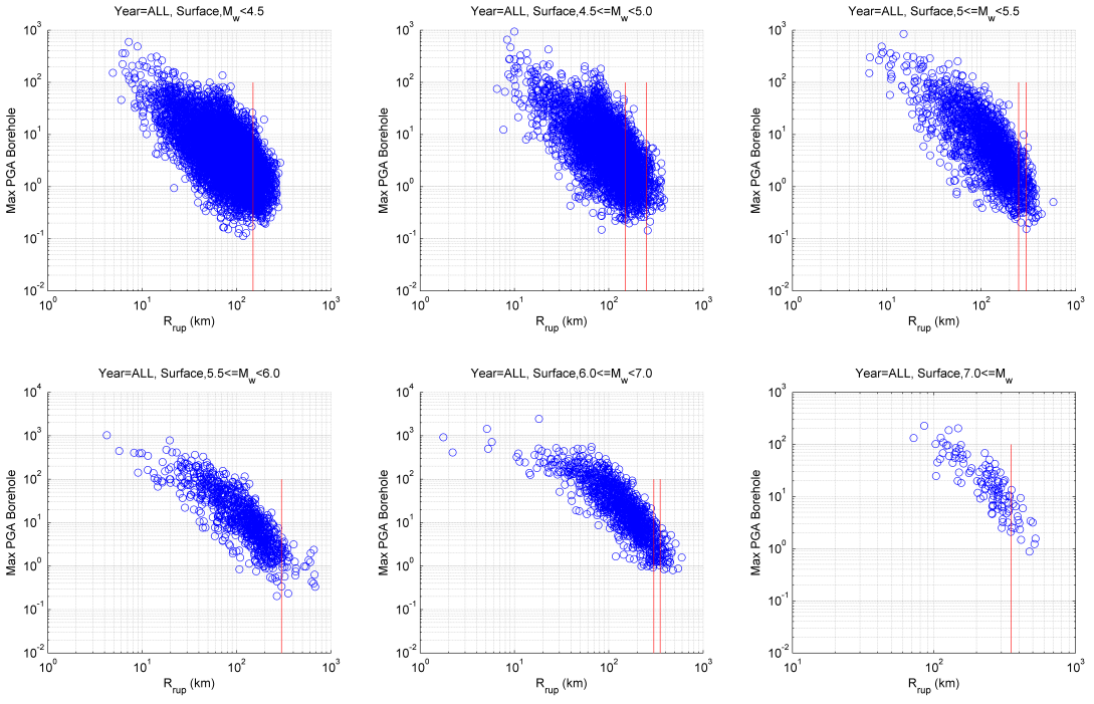
Table D.1. The subsequent steps adopted to obtain the final database used in the analysis presented in Chapter 5

Criteria (Step)	Description of the criteria applied progressively to select the final set of motions from the Flatfile	Number of records (number of F-net earthquakes)	Number of records from $M_w \geq 5.0$ (number of F-net Earthquakes)	Number of records from $M_w \geq 5$, $R_{rup} \leq 350$km (number of F-net Earthquakes)	Number of records from $M_w \geq 5$, $R_{rup} \leq 100$km (number of F-net Earthquakes)
0	Full database described in Chapter 3	183322	--	--	--
1	Remove motions from earthquakes with M_w (from F-net catalog) < 4.0	138341 (2776)	66706 (728)	50507 (678)	4916 (392)
2	Remove motions from earthquakes that are not classified as active crustal earthquakes according to the algorithm by Garcia et al. (2012).	45013 (922)	16255 (170)	14803 (170)	2712 (113)
3	Remove motions recorded at stations for which the shear wave velocity profile was not available	42937 (903)	14808 (162)	14088 (162)	2594 (109)
4	Remove motions for which the error in PGA is larger than 2%	39079 (901)	12534 (161)	11721 (161)	2477 (108)
5	Remove motions if one or more of its four horizontal components did not pass the signal-to-noise ratio check	28855 (897)	6968 (158)	6822 (158)	1937 (105)
6	Remove motions that were not possibly automatically processed using the pre-set criteria	19124 (858)	4366 (143)	4268 (142)	1387 (102)

7	Remove motions from earthquakes with a match category other than "A" or matched one-by-one with an F-net event	18525 (837)	4253 (140)	4158 (139)	1326 (99)
8	Remove motions recorded at distances beyond the censoring model adopted in this study.	16479 (830)	4006 (138)	4006 (138)	1326 (99)
9	Flag motions classified as multiple wave trains (U→ Unlikely; P→ Possibly; and L→Likely)	U:13345 (806) P:1444 (524) L:1690 (424)	U:2854 (132) P:479 (104) L:673 (80)	U:2854 (132) P:479 (104) L:673 (80)	U: 884 (91) P:123 (51) L:319 (45)
10	Re-process the motions (from events with $M_w \geq 5.0$) that were found to be MWT from the visual inspection and add them to list of usable motions	14141 (810)	3650 (136)	3650 (136)	1128 (96)
11	Remove the motions recorded from two events because they were reported as subduction events in the literature (Park and Mori 2005)	14041 (808)	3550 (134)	3550 (134)	1126 (94)
12	Remove motions from earthquakes that were recorded at less than 5 stations	13735 (679)	3510 (114)	3510 (114)	1109 (90)



(a)



(b)

Figure D.1. The scatter plots of the maximum PGA versus R_{rup} for the different magnitude bins recorded at: a) borehole; and b) ground surface.

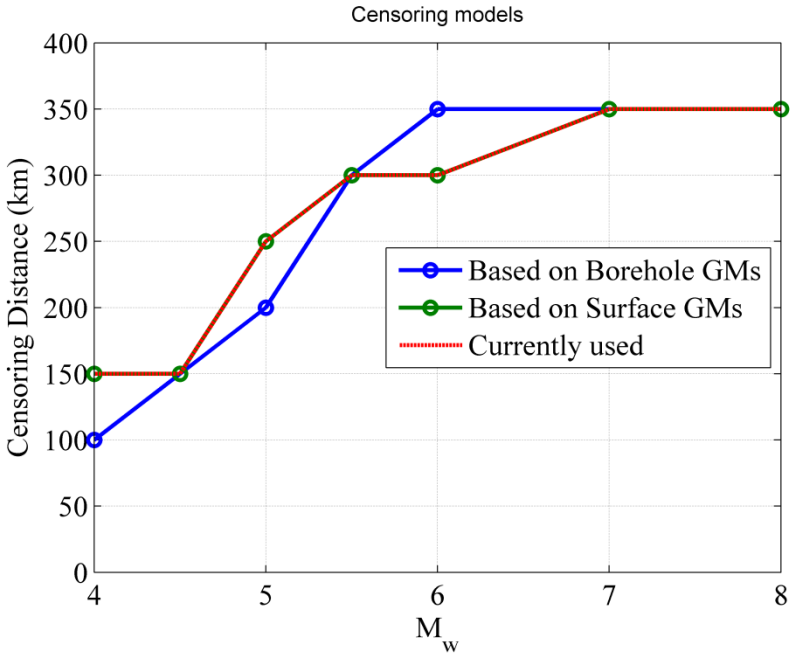


Figure D.2. The maximum R_{rup} versus M_w models obtained from borehole and surface motions. The default one chosen in this study is the model obtained from the surface records

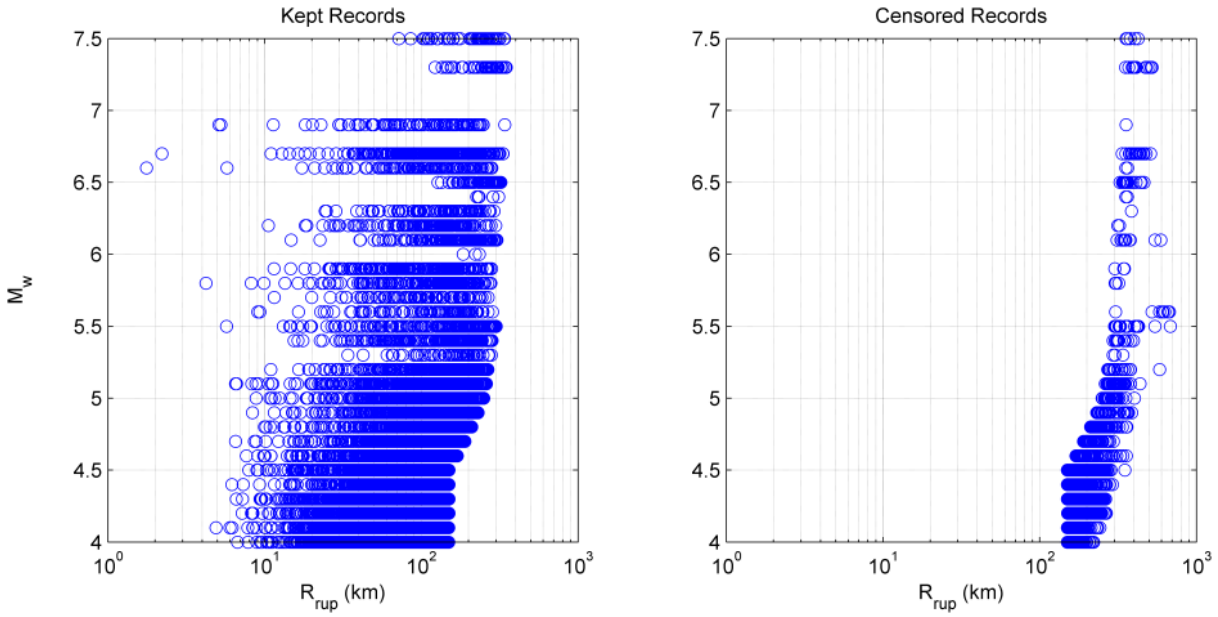


Figure D.3. The scatter plots of R_{rup} versus M_w for the records kept in the database (left panel), and the censored records (right panel).

Step 9:

The presence of multiple wave trains in a single data file is a problem that limits the usability of the GM record. In some cases, especially after the occurrence of large earthquakes, a strong ground motion instrument can record two seismic wave trains from different earthquakes and include both in the same file assuming they are from a single earthquake (Figure D.4). Strasser and Bommer (2005) reported that using such records could result in distortion of the high frequency portion of the response spectra. They attempted to calibrate two different methods to automatically separate these sub events. Their attempts failed because of the large variations in the magnitude and distances available in their study. Hence, they conducted a manual separation of a subset of their database. The size of the database used in the current study complicates the process of manually localizing and separating such motions. Hence, we developed an algorithm to automatically detect the motions that potentially contain multiple wave trains (MWT). The idea of the algorithm is to automatically detect if the acceleration time history started to considerably increase after it decreased. The limitation of preparing such algorithm is that there are no cutting edge criteria to be used to detect the ends of mainshocks and arrival of afterchocks' seismic waves. Hence, we decided to test different criteria within the algorithm and count the number of the cases where the different criteria reported a motion as MWT. Then the output of applying these different algorithm is used to classify the motion into three classes (i.e., unlikely, possible, or likely contains MWT). It is then up to the user to choose which subset to use in their analyses. Hence, this algorithm is not intended to provide a clear definite answer about whether the motion contains MWT or not, but is intended to provide a good indication about that. The algorithm can be summarized as follows:

I. The acceleration time histories for the two horizontal components at the ground surface are divided into different segments. Each segment is 5 seconds wide.

II. The maximum of the absolute acceleration and its time of occurrence in each segment are localized.

III. We chose three different ratios that are used as preset criteria to pick motions with MWT. In the current algorithm we selected Criteria(1)=0.05, Criteria(2)=0.10 and

Criteria(3)=0.20. The algorithm loops through the different criteria. For each Criteria(i):

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

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

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

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

The automated algorithm was applied to a total of 13345 motions. The algorithm flagged 1690 motions as "likely multiple wave trains (MWT)"; 1444 as possibly MWT; and 13345 motions as unlikely MWT.

Figure D.4 shows an example of a MWT motion. The top plots show the acceleration time histories. This particular motion shows three different wave trains. PGA is in the first wave train. The second row of plots show the maximum accelerations in each segment (5 seconds in width) normalized by the PGA of each component of the motion. The titles of the plots in the second row show the result of the different checks. The algorithm localized an aftershock using the three criteria on the two components (i.e., an aftershock MWT was detected 6 times). Hence, the algorithm classified this motion as "Likely MWT" which is confirmed by the visual inspection. The second line in the titles of the lower row of plots shows the flags for the foreshock MWT check. Since the PGA is in the first wave trains, the algorithm did not detect any foreshocks and the six flags resulted in a non-foreshock-MWT flag. The vertical lines are plotted to show the automatically detected arrivals of the different aftershocks using the different criteria. It is clear from the acceleration time history plots that the algorithm was able to localize the arrivals effectively.

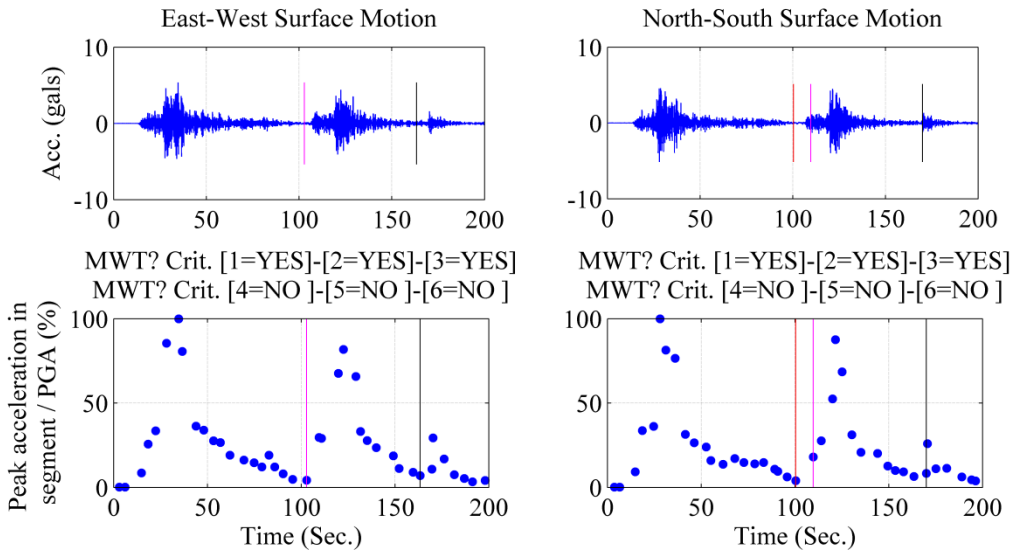


Figure D.4. An example of a motion that contains multiple wave trains (GM# 59351). The plots to the left are for the East-West Surface component of the motion while the left plots are for the North-South Surface component. The top plots show the acceleration time histories and the vertical lines show the automatically picked arrivals of the aftershocks by the current algorithm. The bottom plots show the ratio between the peak acceleration in each segment (5 seconds in width) and the PGA in percent. The automatically detected arrivals of the aftershocks are based on these ratios.

Step 10: All records that are from events with M_w<5.0 and were flagged as "likely" or "possibly" MWT, were removed from the database (1982 motions). The MWT check (*Step 9*) showed that about 29% (1152 motions) of the motions from events with $M_w \geq 5.0$ were either *possibly* or *likely* MWT. This means that if these motions were removed, a substantial fraction of the motions from large events would have been lost. Hence, a visual inspection was conducted on the 1152 motions (479 possibly MWT; and 673 likely MWT) to check if they are really MWT and decide whether the smaller wave trains can be trimmed and the main train re-processed again. The visual inspection resulted in the following:

- 408 motions were not found MWT.

- 614 motions that contain MWT can be re-processed after trimming the small waver trains. These motions were reprocessed and kept in the database.

- 130 motions that contain MWT were judged to be difficult to be re-processed. The motions in this category contain multiple wave trains that are very close in amplitude (Figure D.4). In this case, it is difficult to decide which wave train is from the mainshock. A subjective decision of which wave train to keep in such case, could potentially artificially inflate the variability associated with the GMPE predictions. It was decided neither to re-process these motions nor to keep them in the database.

In this step, 614 motions were re-processed. Re-processing resulted in flagging 124 motions of this list as "Error in Filtering". These 124 motions were not investigated any further. In reprocessing these motions, the result of the signal-to-noise (SNR) check was neglected for two reasons. First, all these records already passed the SNR check before the wave trains from fore/after-shocks were trimmed. Second, after trimming the after-shock wave trains, the trailing portion of the record that is used in the SNR check will be contaminated with seismic waves and will not be only the background noise. Figure D.5 shows an example of a motion with MWT that was reprocessed. The figure shows the East-West Surface component of the motion. The waves from the potential aftershock earthquake were automatically trimmed and re-processed. The automatically detected arrival of the aftershock motion is used as the point at which the motion is trimmed. After that 102 motions were removed as they were violating the maximum allowable error in PGA (2%) in the filtered motion.

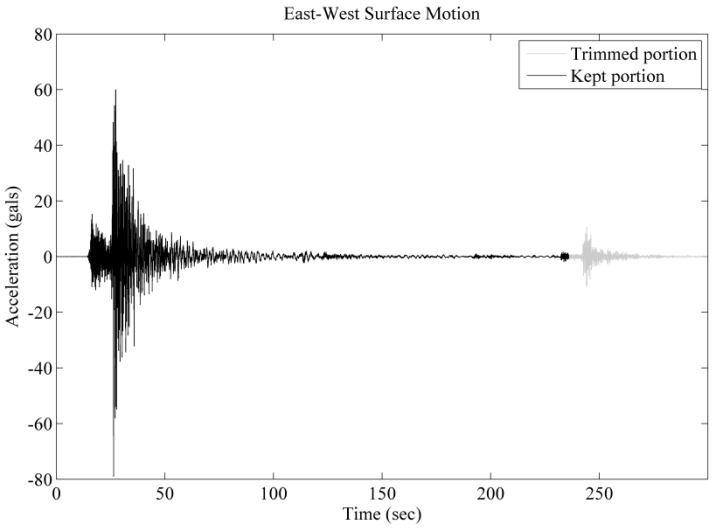


Figure D.5. An example of a motion that contains multiple wave trains where the aftershock was trimmed and the motion re-processed (GM# 7386). The trimmed portion is shown in light gray and the kept portion of the motion is shown in black. The arrival time of the aftershock waves obtained from the automated algorithm was used to automatically trim this time history.

Step 11: All motions recorded from the Kii Peninsula earthquakes occurred on September 5, 2004 (F-net# 8592: M_w=7.5 and F-net# 8590: M_w=7.3) in the database were removed because they were reported as intraslab subduction type events in the literature (Park and Mori 2005).

REGRESSION ANALYSIS

Regression Steps

Following the footsteps of Abrahamson et al. (2013), we conducted the regression analysis over multiple steps to reach the final regression coefficients for the ergodic and site-specific GMPEs. After each step one or more of the regression coefficients is smoothed across

different spectral periods and fixed in the subsequent steps. Table D.2 shows a summary of the different steps. Following are some notes regarding the regression analysis:

- The regression analysis was conducted using earthquakes with 5 or more records in all steps.

- During preliminary analyses, we tested using a different magnitude and distance (M_w-R_{rup}) ranges in the different regression steps. The nature of the database on hand for this study led to a choice of an M_w-R_{rup} ranges different that the one adopted by Abrahamson et al. (2013).

- In the current study, we include surface and borehole records in the regression analysis. Hence, we differentiate between the constant coefficient (a1) and the site response coefficient (a10) obtained from surface and borehole records as these coefficients should be different due to the site response.

- In the study by Abrahamson et al. (2013), the distance coefficient (*a17*) wasn't regressed for in the initial steps because the maximum R_{rup} they used in these steps was 80km. In the case of our study, we go up to 350 km in the first steps. Hence, we decided to include *a17* in the first regression steps.

- The regression in *Step 01a* was conducted for PGA only. The value of a5 is then fixed for subsequent steps and was used for all spectral periods.

- M_1 and M_2 were not regressed for. Instead they were picked by inspecting the residuals obtained from preliminary analyses and were kept unchanged for all steps.

- While running *Step01d* we tested the aftershock flags obtained from different declustering algorithms (see Dawood et al. 2014, also in Chapter 3 of this dissertation, for additional information).We conducted this test to check if using

any of these different flags will result in an estimation of the aftershock coefficient that is different than zero. The coefficient a14 was not significantly different than zero when the different flags were tested. Hence, a14 was set to be zero over the different spectral periods.

- In preliminary analysis, we introduced several coefficients in the regression analysis during *Step 02a* (i.e., *a43S, a44S, a45S, a46S, a43B, a44B, a45B, a46B a6* and *a7* "see Abrahamson et al. 2013 for more information about *a7*"). We found that *a6* and *a7* were highly correlated (0.96 at PGA). Hence, we decided to remove *a7* and keep only *a6* in the final regression analysis.

- Since all the motions used in this study are Japanese records, we did not include some of the Japanese specific coefficients presented in Abrahamson et al. (2013) to correct for the effect of including Japanese data with data from other regions around the world (i.e., a36S, a37S, a38S, a39S, a40S, a41S, a36B, a37B, a38B, a39B, a40B, a41B).

- We adopted the same V1, VLin, and b values used by Abrahamson et al. (2013).

Table D.2. Table showing the different steps conducted to regress for the different coefficients in the GMPE

Step number	Dataset used in the regression	Estimated parameters	Parameters Smoothed after the step	Type of analysis
01a	Motions from earthquakes with $M_w \geq 5.0$ only	a1s, a1b, a2, a3, a5, a10s, a10b, a11, a12, a14, $a15_a$,$a15_b$, a17, c4	a5 (obtained using PGA, assumed constant across all periods)	Nonlinear mixed effects
01b	Same	a1s, a1b, a2, a3, a10s, a10b, a11, a12, a14, $a15_a$, $a15_b$, a17, c4, **a8**	a3, c4	Nonlinear mixed effects
01c	Same	a1s, a1b, a2, a10s, a10b, a11, a12, a14, $a15_a$, $a15_b$, a17, a8	$a15_a$, $a15_b$	Nonlinear and linear mixed effects
01d	Same	a1s, a1b, a2, a10s, a10b, a11, a12, a14, a17, a8	a14, a11	*Nonlinear and linear mixed effects*
01e	Same	a1s, a1b, a2, a10s, a10b, a12, a17, a8	a8, a12	Nonlinear and linear mixed effects
02a	Full database ($M_w \geq 4.0$)	a1s, a1b, a2, a10s, a10b, a17, **a43S**, **a44S**, **a45S**, **a46S**, **a43B**, **a44B**, **a45B**, **a46B**, **a6**	a2, a6	Nonlinear and linear mixed effects
02b	Same	a1s, a1b, a10s, a10b, a17, a43S, a44S, a45S, a46S,	a17	Nonlinear and linear mixed effects

		a43B, a44B, a45B, a46B		
02c	Same	a1s, a1b, a10s, a10b, a43S, a44S, a45S, a46S, a43B, a44B, a45B, a46B	a43S, a44S, a45S, a46S, a43B, a44B, a45B, a46B	Nonlinear and linear mixed effects
02d	Same	a1s, a1b, a10s, a10b, **$a_{borehole}$**	$a_{borehole}$	Nonlinear and linear mixed effects
02e	Same	a1s, a1b, a10s, a10b	a1s, a1b, a10s, a10b	Nonlinear and linear mixed effects
02f	Same	Residuals	Tau and Sigma models	

The study by Rodriguez-Marek et al. (2011) found that the downward moving waves reflected from the soil layers affect the borehole motion; for this reason, we shoes to include effects of site response at the borehole instrument. However, it is unclear whether the nonlinear site response term should be used for borehole motions. In the final model, we decided to keep the nonlinear component for the borehole predictions. However, the data available allows for an estimate of the errors that would be introduced in the prediction of borehole motions if nonlinearity was ignored. To this effect, we regressed for site response amplification for the borehole records as well. In this analysis we run three different regression analyses (*Step 01a*) assuming three different nonlinear site response models as follows:

Version 0: This is the version that was adopted in the GMPEs presented in Chpter 5 of this dissertation. In this model, the non-linear model was assumed to apply to the borehole as well as the surface motions:

$$f_5^{\,G/B}(Sa_{1100}, V_{s30}) =$$

$$\begin{cases} F_{Surf}\left[a10_S\, ln\left(\frac{V^*_{S30}}{V_{Lin}}\right) - b\, ln\left(Sa_{1100,Surf} + c\right)\right] + \\ \left(1 - F_{Surf}\right)\left[a10_B\, ln\left(\frac{V^*_{S30}}{V_{Lin}}\right) - b\, ln\left(Sa_{1100,Bore} + c\right)\right] + \\ b\, ln\left(F_{Surf} Sa_{1100,Surf} + \left(1 - F_{Surf}\right) Sa_{1100,Bore} + c\left(\frac{V^*_{S30}}{V_{Lin}}\right)^n\right) & for\ V_{S30} < V_{Lin} \\ \left[F_{Surf}\, a10_S + \left(1 - F_{Surf}\right) a10_B + bn\right] ln\left(\frac{V^*_{S30}}{V_{Lin}}\right) & for\ V_{S30} \geq V_{Lin} \end{cases} \quad (D.1)$$

Version 1: In this version, the non-linear model was assumed to apply only to the surface motions. Hence, "$b\, ln\left(Sa_{1100,Bore} + c\right)$" and "$\left(1 - F_{Surf}\right) Sa_{1100,Bore}$" were removed from Equation D.1 for the borehole motions as follows:

$$f_5^{\,G/B}(Sa_{1100}, V_{s30}) =$$

$$\begin{cases} F_{Surf}\left[a10_S\, ln\left(\frac{V^*_{S30}}{V_{Lin}}\right) - b\, ln\left(Sa_{1100,Surf} + c\right)\right] + \\ \left(1 - F_{Surf}\right)\left[a10_B\, ln\left(\frac{V^*_{S30}}{V_{Lin}}\right)\right] + \\ b\, ln\left(F_{Surf} Sa_{1100,Surf} + c\left(\frac{V^*_{S30}}{V_{Lin}}\right)^n\right) & for\ V_{S30} < V_{Lin} \\ \left[F_{Surf}\, a10_S + \left(1 - F_{Surf}\right) a10_B + bn\right] ln\left(\frac{V^*_{S30}}{V_{Lin}}\right) & for\ V_{S30} \geq V_{Lin} \end{cases} \quad (D.2)$$

Version 2: In this version, the non-linear model was assumed to apply only to the surface motions. Hence, "$b\, ln\left(Sa_{1100,Bore} + c\right)$", "$\left(1 - F_{Surf}\right) Sa_{1100,Bore}$", "$c\left(\frac{V^*_{S30}}{V_{Lin}}\right)^n$" and "$bn$" were removed from Equation D.1 for borehole motions as follows:

$$f_5^{\,G/B}(Sa_{1100}, V_{s30}) =$$

$$\begin{cases} F_{Surf}\left[a10_S\, ln\left(\frac{V^*_{S30}}{V_{Lin}}\right) - b\, ln\left(Sa_{1100,Surf} + c\right)\right] + \\ \left(1 - F_{Surf}\right)\left[a10_B\, ln\left(\frac{V^*_{S30}}{V_{Lin}}\right)\right] + \\ b\, ln\left(F_{Surf} Sa_{1100,Surf} + \mathbf{F_{Surf}} c\left(\frac{V^*_{S30}}{V_{Lin}}\right)^n\right) & for\ V_{S30} < V_{Lin} \\ \left[F_{Surf}\, a10_S + \left(1 - F_{Surf}\right) a10_B + \mathbf{F_{Surf}} bn\right] ln\left(\frac{V^*_{S30}}{V_{Lin}}\right) & for\ V_{S30} \geq V_{Lin} \end{cases} \quad (D.3)$$

Figure D.6 present the comparison between the predictions of PGA obtained from the aforementioned three versions. These plots show that including the nonlinear site response model in the prediction of the borehole motions did not affect the predictions in most cases. On the other hand, for some specific cases, including the nonlinear site response model had a major impact on the predictions (i.e., a ratio around 0.77 Hence, it is important to know the limits beyond which including of excluding the nonlinear site response component from the borehole predictions can be significant. Figure D.7 shows V_{s30} versus Sa_{1100} for two different ratio bins (i.e., 0.7→0.95 and 0.95→0.98). The figure shows also the best fit line between the different data point. The equations of the best fit lines are:

PGA (Version 0 / Version 1) for ratios 0.70→0.95: $V_{s30} = 49 \ *\ln(Sa_{1100})+ 387$ (D.4)

PGA (Version 0 / Version 1) for ratios 0.95→0.98: $V_{s30} = 101*\ln(Sa_{1100})+ 635$ (D.5)

If the user of the GMPE would like to neglect the effect of nonlinearity on the borehole records, he/she can use these aforementioned equations to either find limits beyond which the GMPE predictions deviates from the predictions he/she needs (i.e., predictions at borehole level that is not affected by soil nonlinearity), or correct the predictions obtained from the GMPE to remove the effects of non-linear site response from the borehole records.

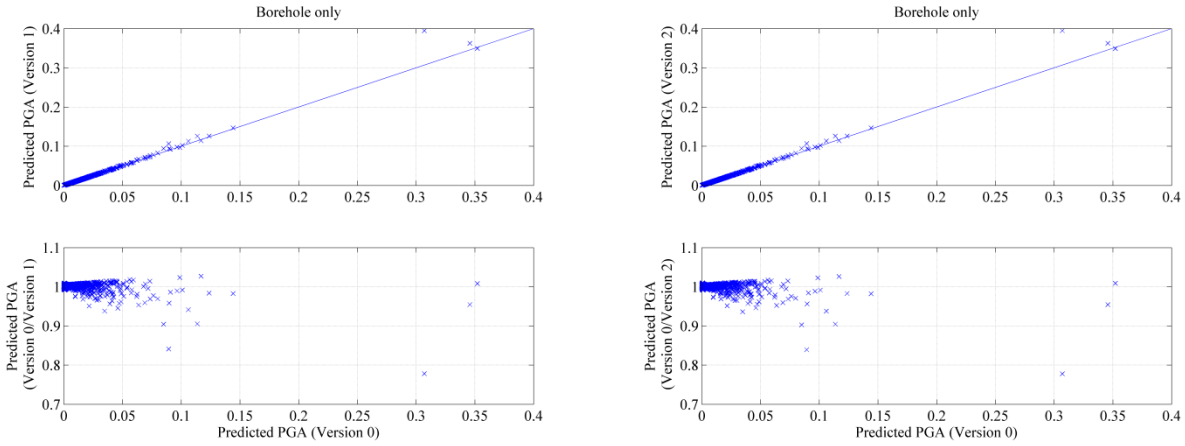


Figure D.6. Scatter plots that compare: 1) the predicted PGA from the different versions (top plots); and 2) the ratios between the different versions versus the predictions of Version 0 (bottom plots)

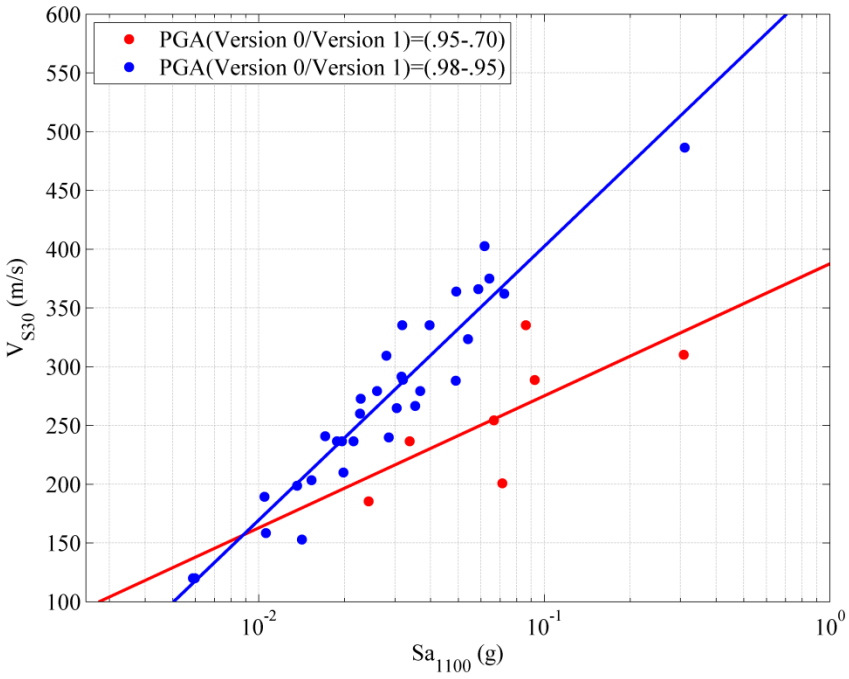
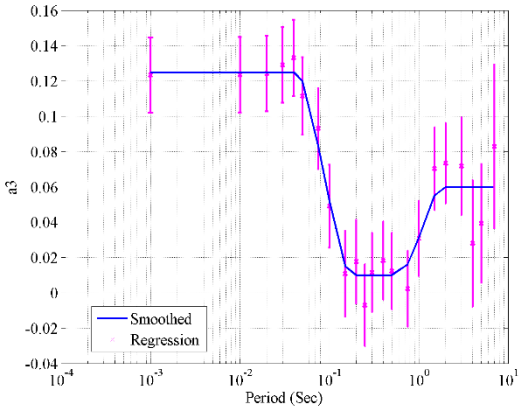
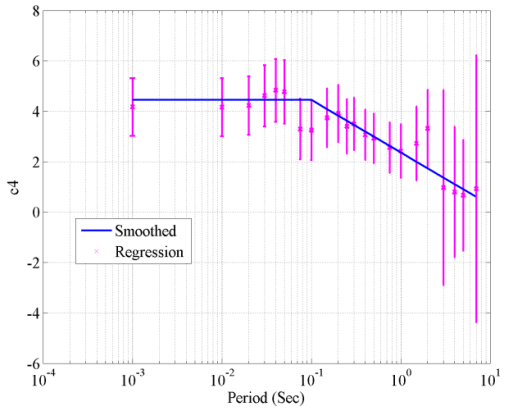


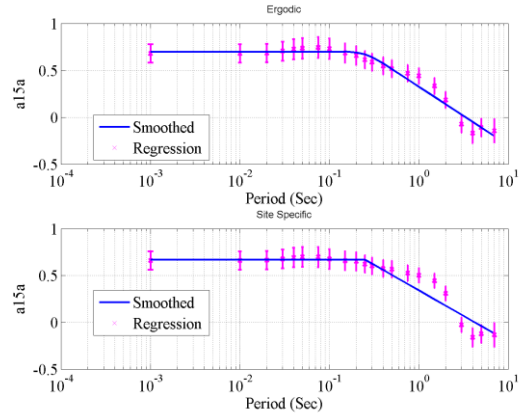
Figure D.7. Scatter plots of V_{s30} versus Sa1100 for motions that resulted in a ratio between the predicted PGA from Version 0 and 1 for: 1) Blue data points: 0.98→0.95, and 2) Red data points : 0.95→0.70. The best fit lines between the two sets of scatter plots are shown as continuous lines with the same color of the data points.

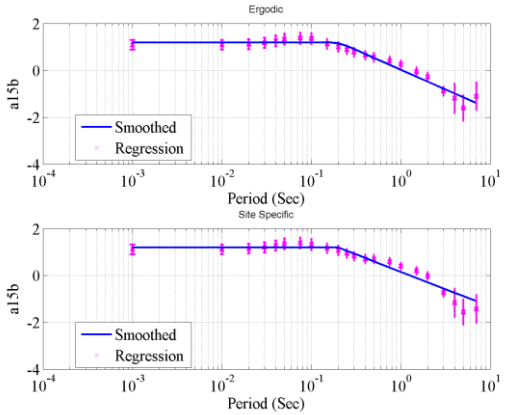
Smoothing the Regression Coefficients

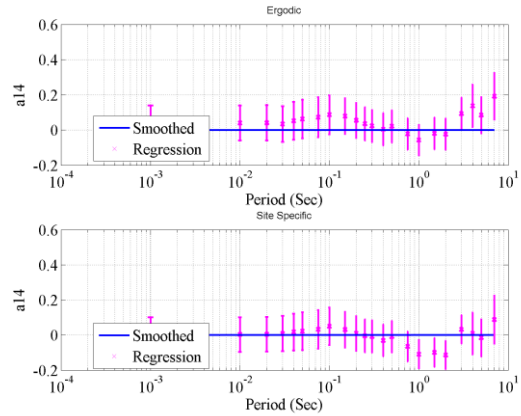
Figure D.8 shows the regression coefficients obtained from the regression analysis and the smoothed values that are presented in Dawood and Rodriguez-Marek (2014; also in Chapter 5 of this dissertation). Wherever there are two subplots, the top subplot is for the coefficients of the ergodic formulation and the bottom one is for the site-specific formulation.

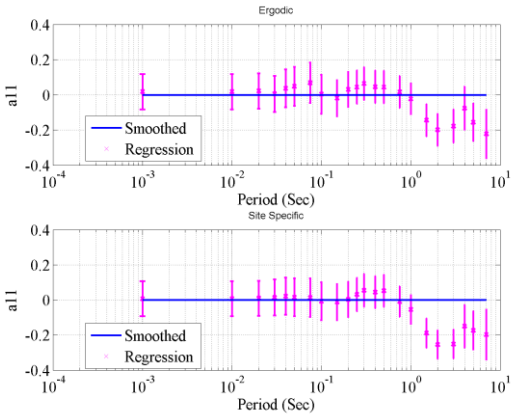

a3
Smoothed
Regression
Period (Sec)

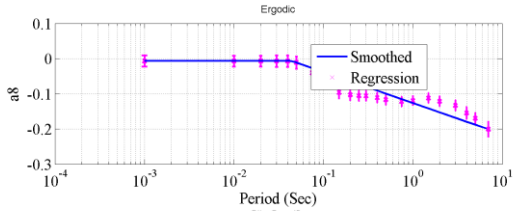

c4
Smoothed
Regression
Period (Sec)

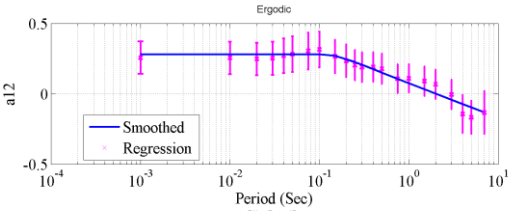

Ergodic
a15a
Smoothed
Regression
Period (Sec)
Site Specific
a15a
Smoothed
Regression
Period (Sec)

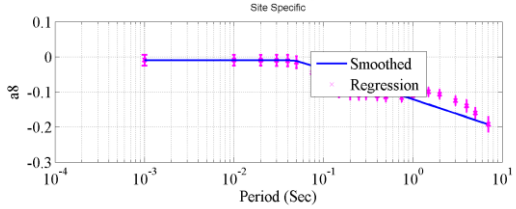

Ergodic
a15b
Smoothed
Regression
Period (Sec)
Site Specific
a15b
Smoothed
Regression
Period (Sec)

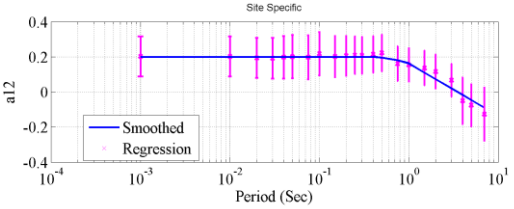

Ergodic
a14
Smoothed
Regression
Period (Sec)
Site Specific
a14
Smoothed
Regression
Period (Sec)

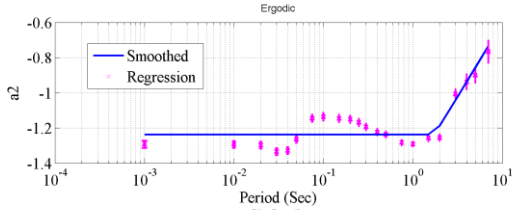

Ergodic
a11
Smoothed
Regression
Period (Sec)
Site Specific
a11
Smoothed
Regression
Period (Sec)

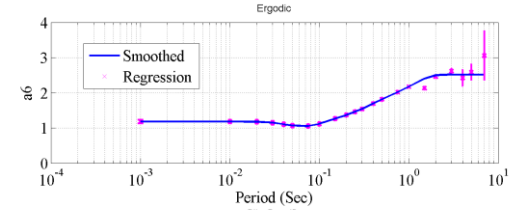

Ergodic
a8
Smoothed
Regression
Period (Sec)

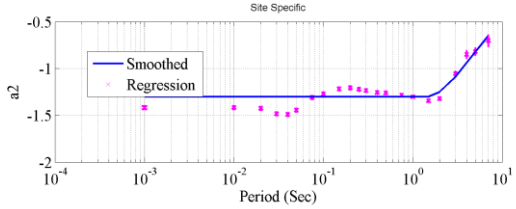

Ergodic
a12
Smoothed
Regression
Period (Sec)

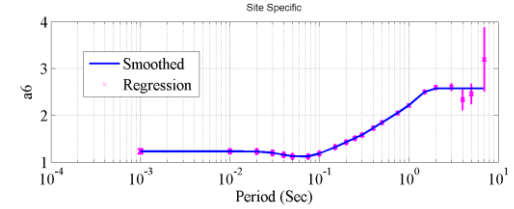

Site Specific
a8
Smoothed
Regression
Period (Sec)

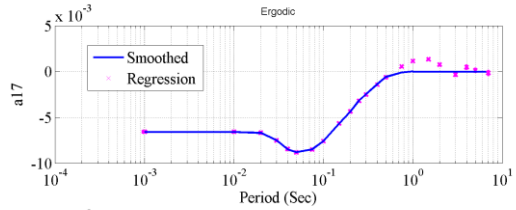

Site Specific
a12
Smoothed
Regression
Period (Sec)

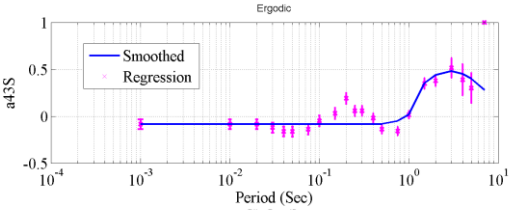

Ergodic
a2
Smoothed
Regression
Period (Sec)

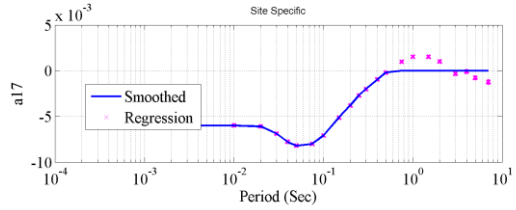

Ergodic
a6
Smoothed
Regression
Period (Sec)

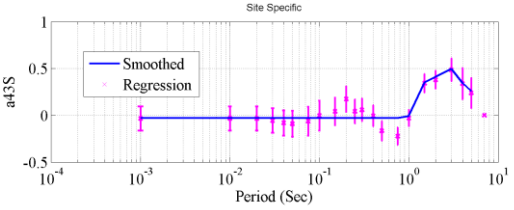

Site Specific
a2
Smoothed
Regression
Period (Sec)

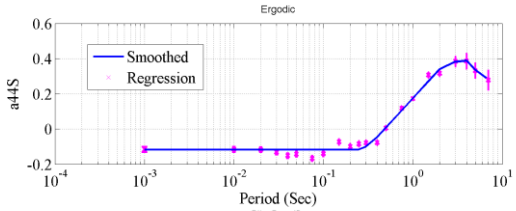

Site Specific
a6
Smoothed
Regression
Period (Sec)

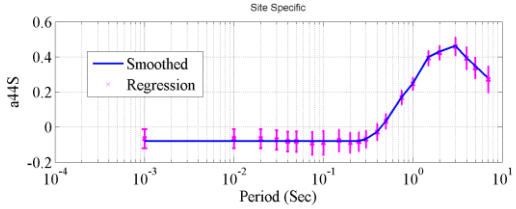

Ergodic
x 10^-3
a17
Smoothed
Regression
Period (Sec)

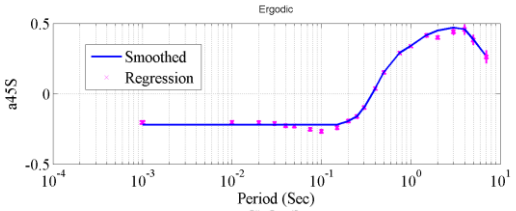

Ergodic
a43S
Smoothed
Regression
Period (Sec)

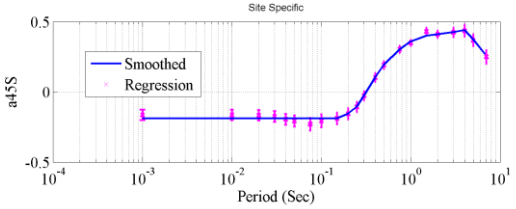

Site Specific
x 10^-3
a17
Smoothed
Regression
Period (Sec)

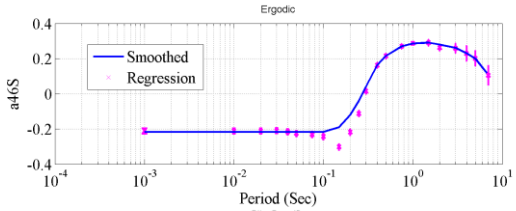

Site Specific
a43S
Smoothed
Regression
Period (Sec)

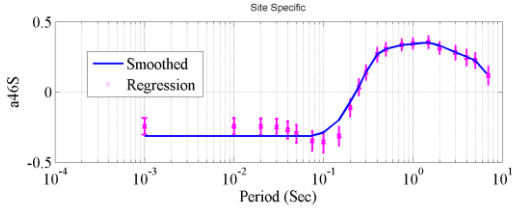

Ergodic
Smoothed
Regression
a44S
Period (Sec)

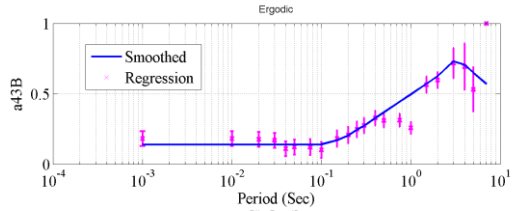

Site Specific
Smoothed
Regression
a44S
Period (Sec)

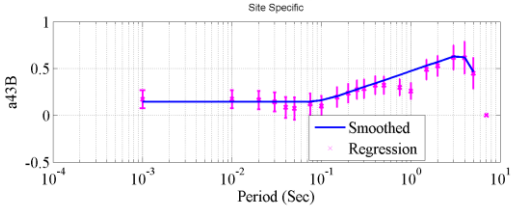

Ergodic
Smoothed
Regression
a45S
Period (Sec)

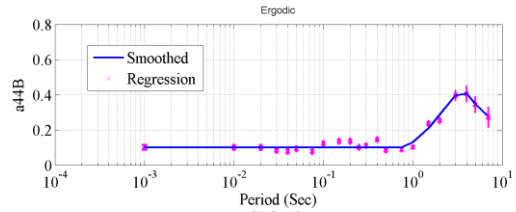

Site Specific
Smoothed
Regression
a45S
Period (Sec)

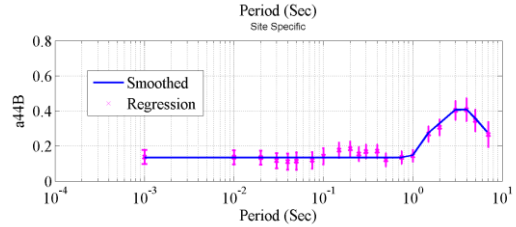

Ergodic
Smoothed
Regression
a46S
Period (Sec)

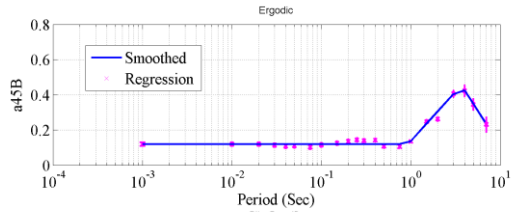

Site Specific
Smoothed
Regression
a46S
Period (Sec)

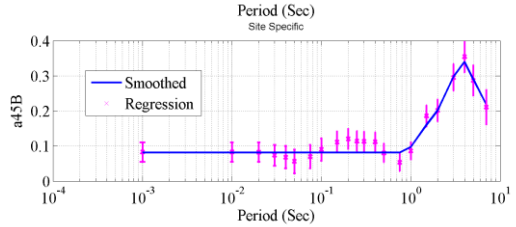

Ergodic
Smoothed
Regression
a43B
Period (Sec)


Site Specific
Smoothed
Regression
a43B
Period (Sec)


Ergodic
Smoothed
Regression
a44B
Period (Sec)


Site Specific
Smoothed
Regression
a44B
Period (Sec)


Ergodic
Smoothed
Regression
a45B
Period (Sec)


Site Specific
Smoothed
Regression
a45B
Period (Sec)

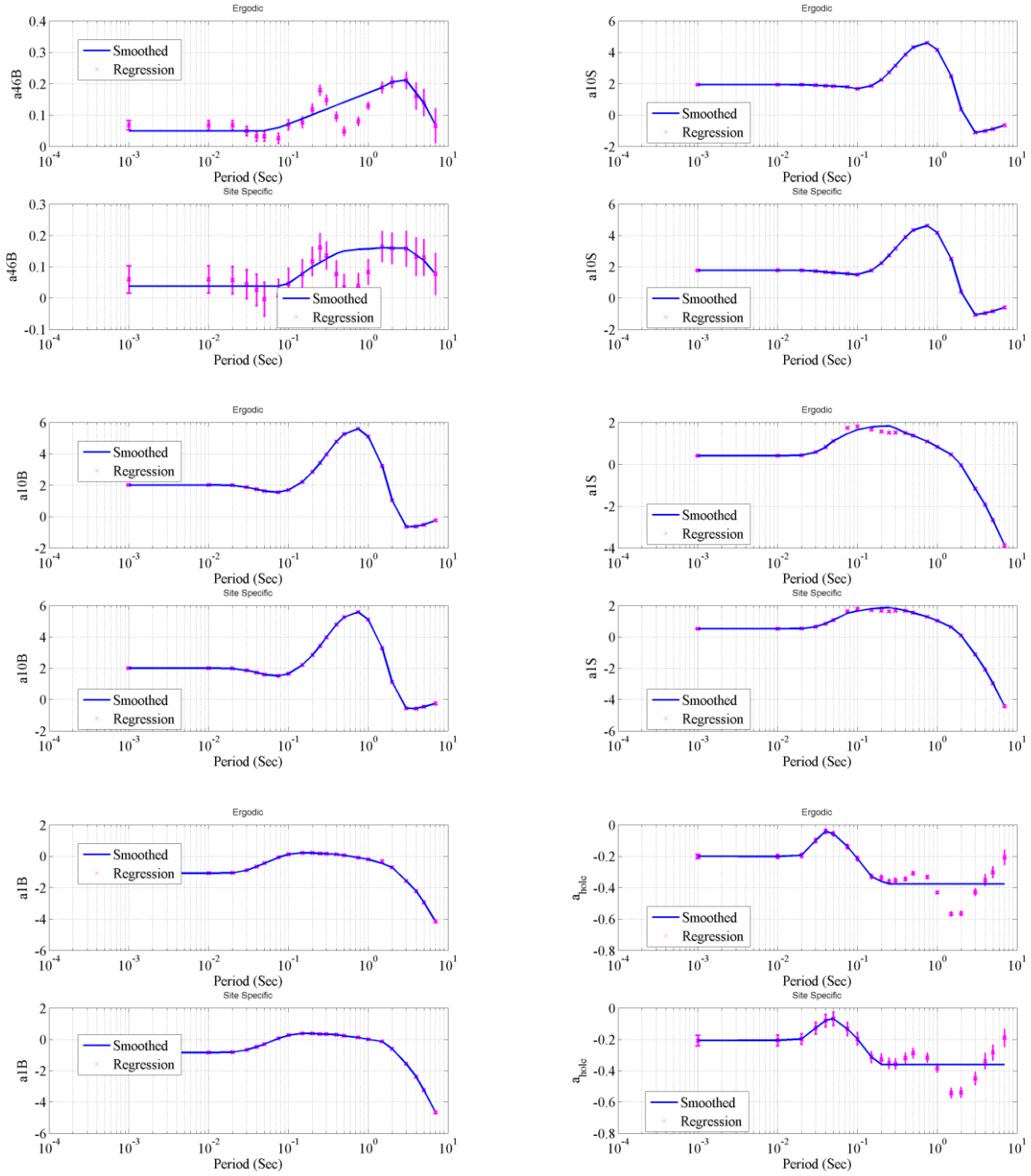
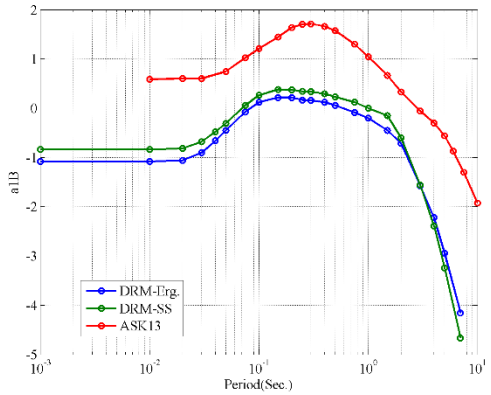
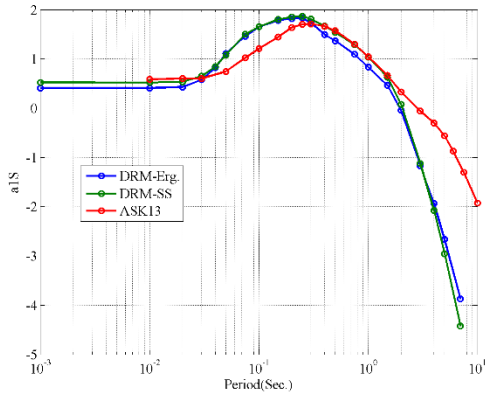
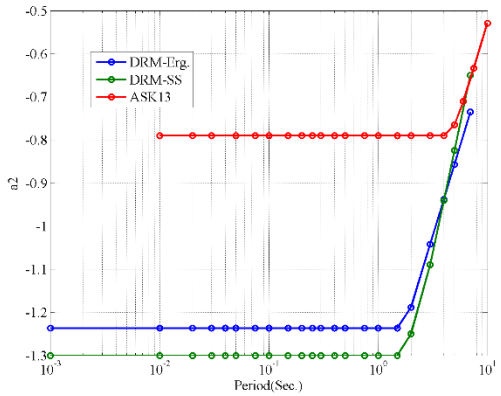


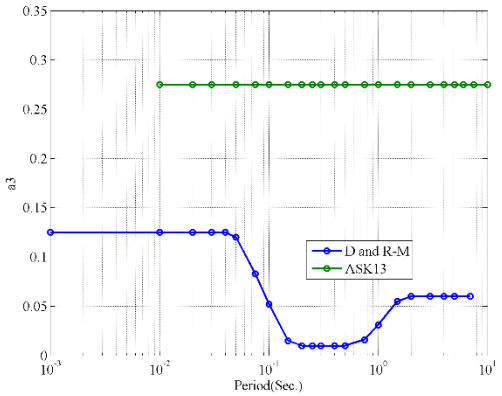
Figure D.8. The regression coefficients obtained from the regression analysis (shown as crosses with error bars) and the smoothed coefficients that are presented in Dawood and Rodriguez-Marek (2014; also in chapter 5 of this dissertation) shown as continuous lines. Wherever there are two subplots, the top subplot is for the coefficients of the ergodic formulation and the bottom one is for the site-specific formulation.

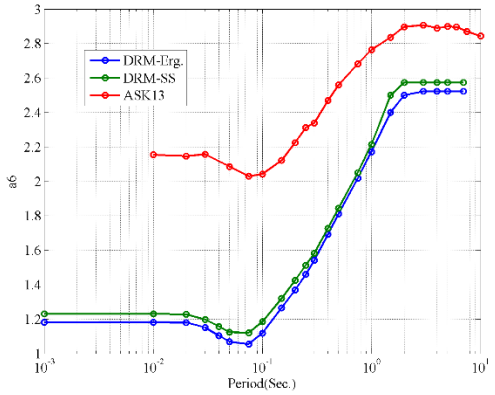
Comparing the Regression Coefficients with the Coefficients from ASK13

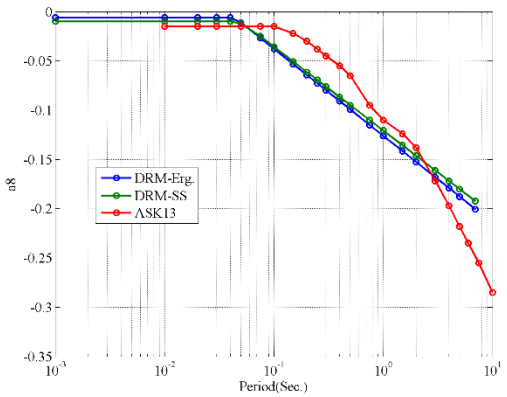


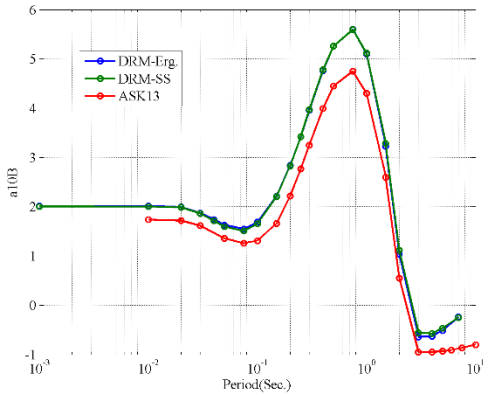


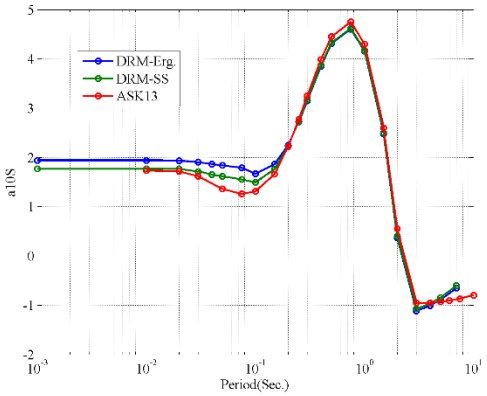


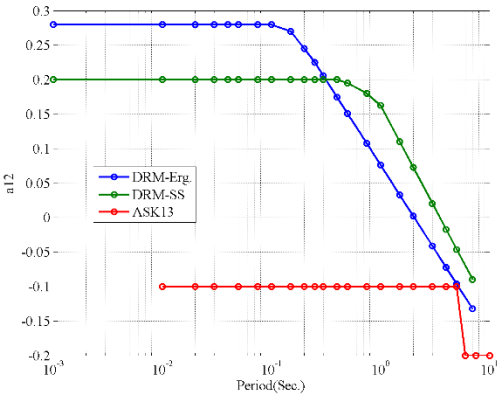


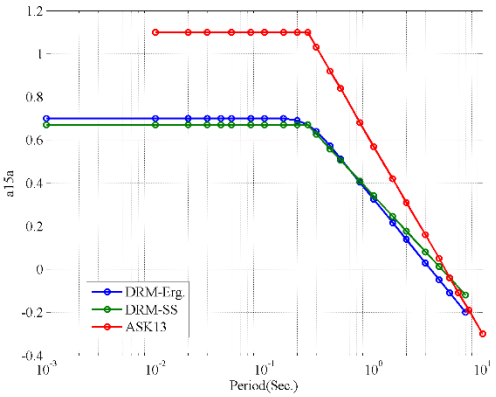


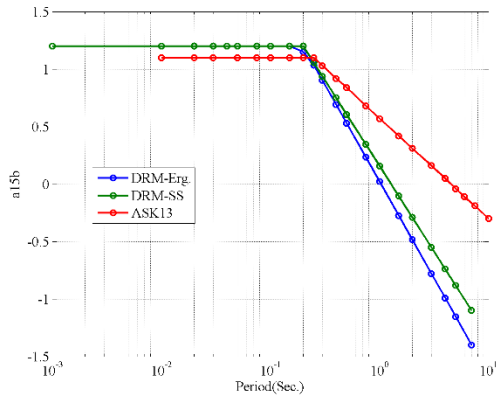


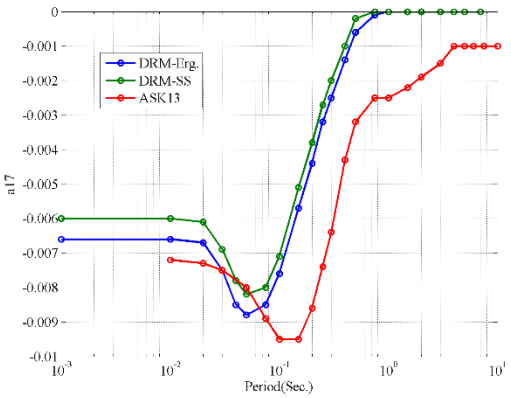

DRM-Erg.
DRM-SS
ASK13
a10B
Period(Sec.)

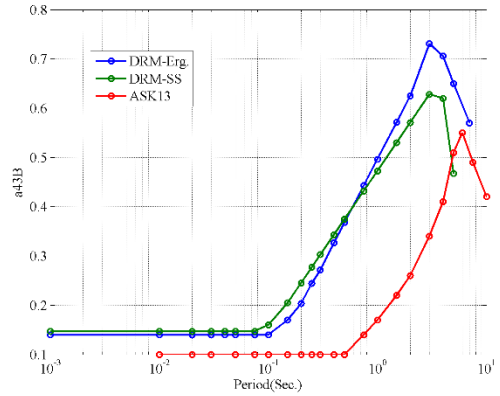

DRM-Erg.
DRM-SS
ASK13
a10S
Period(Sec.)

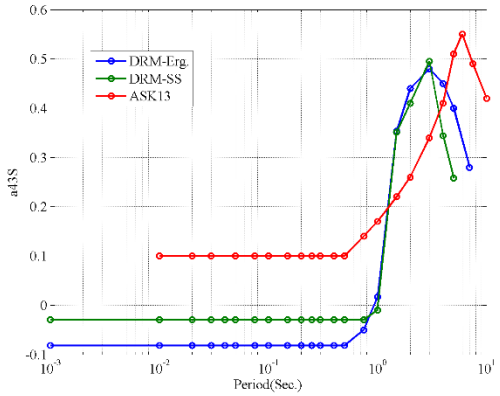

DRM-Erg.
DRM-SS
ASK13
a12
Period(Sec.)

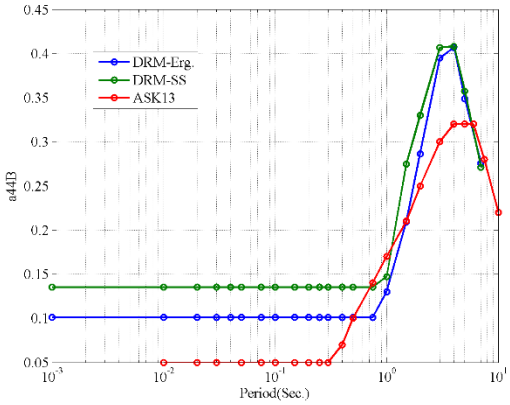

DRM-Erg.
DRM-SS
ASK13
a15a
Period(Sec.)

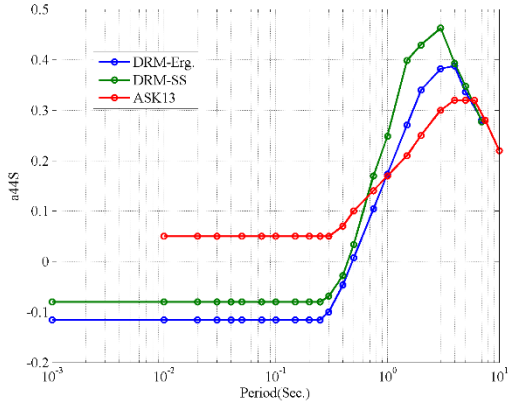

DRM-Erg.
DRM-SS
ASK13
a15b
Period(Sec.)

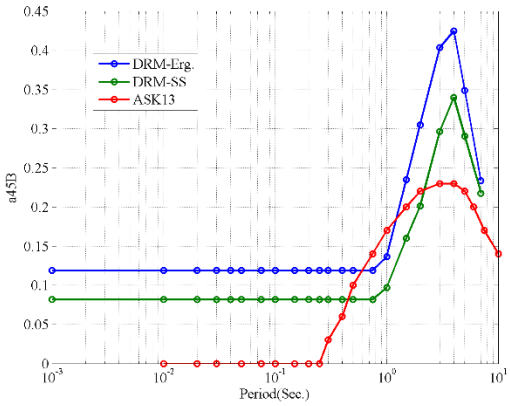

DRM-Erg.
DRM-SS
ASK13
a17
Period(Sec.)

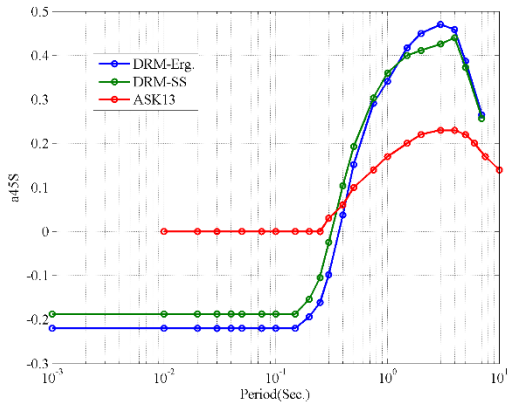

DRM-Erg.
DRM-SS
ASK13
a43B
Period(Sec.)


DRM-Erg.
DRM-SS
ASK13
a43S
Period(Sec.)


DRM-Erg.
DRM-SS
ASK13
a44B
Period(Sec.)


DRM-Erg.
DRM-SS
ASK13
a44S
Period(Sec.)


DRM-Erg.
DRM-SS
ASK13
a45B
Period(Sec.)


DRM-Erg.
DRM-SS
ASK13
a45S
Period(Sec.)

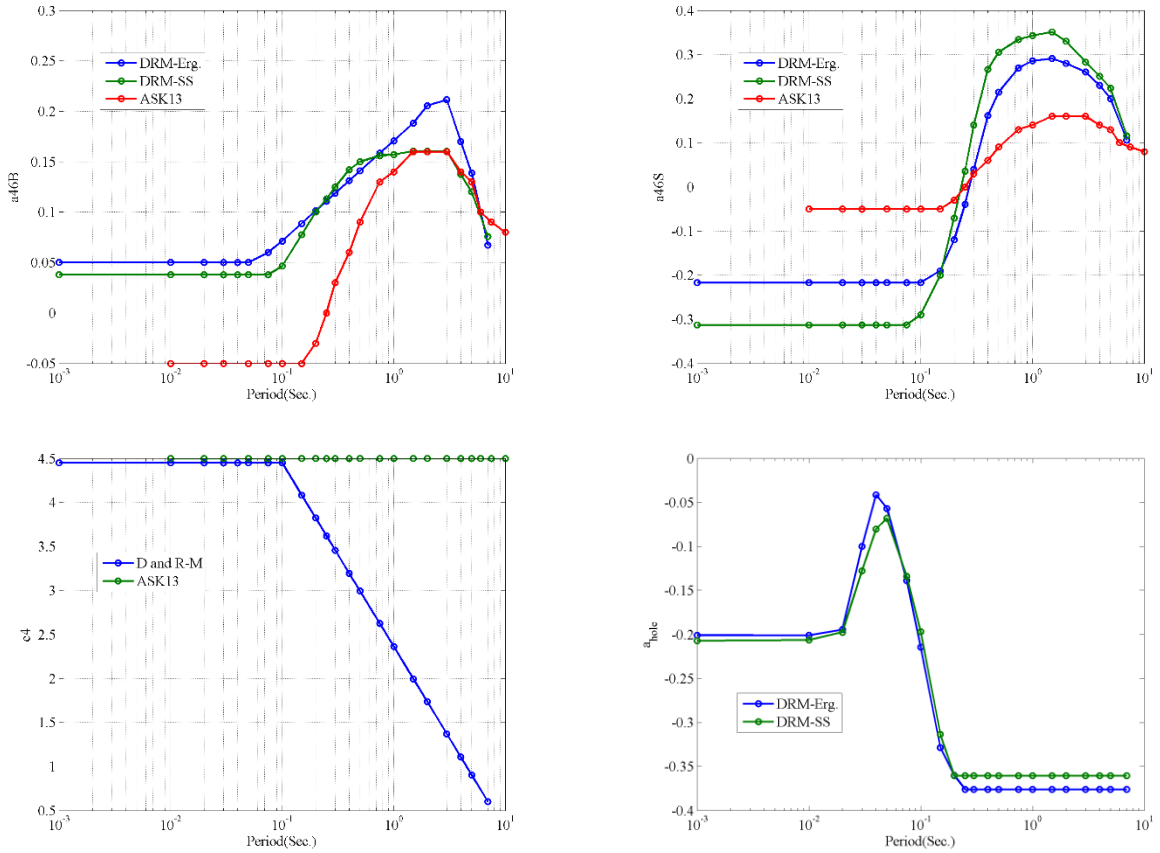


Figure D.9. Comparison between the regression coefficients reported by Abrahamosn et al. (2013; ASK13) and the smoothed regression coefficients obtained in this study from the ergodic formulation (DRM-Erg.) and the site-specific formulation (DRM-SS). The coefficients C4 and a3 have common values for the ergodic and site-specific GMPEs. Hence, they were reported as D and R-M in these plots.

Comparing the Predictions from the Ergodic and Site-Specific Formulations

Figure D.10 shows the histograms of the difference between the predictions obtained from the ergodic and site-specific formulations at different spectral accelerations. We used the final (with smoothed regression coefficients) ergodic and site-specific GMPEs to generated two sets of 7680 response spectra assuming different combinations of input parameters (i.e. M_w =4.0, 4.5, 5.0, 5.5, 6, and 6.5; R_{rup} =20, 50, 100, 150, 200, 250, 300, and 350 km; V_{s30} =200, 300, 500, and 700 m/s); and 0.3 km; $Z_{1.0}$ = 0.05, 0.1, 0.2 and 0.3km; Z_{TOR} =0, 10, 20, 30, and 50 km; and F_{nm} =0 and 1).The presented plots show that the borehole predictions match slightly better

compared with the surface predictions. At PGA (ground surface) and 7 seconds (ground surface and borehole), the differences in the predictions seem to be bimodal. This causes the standard deviations of the differences to inflate at these periods.

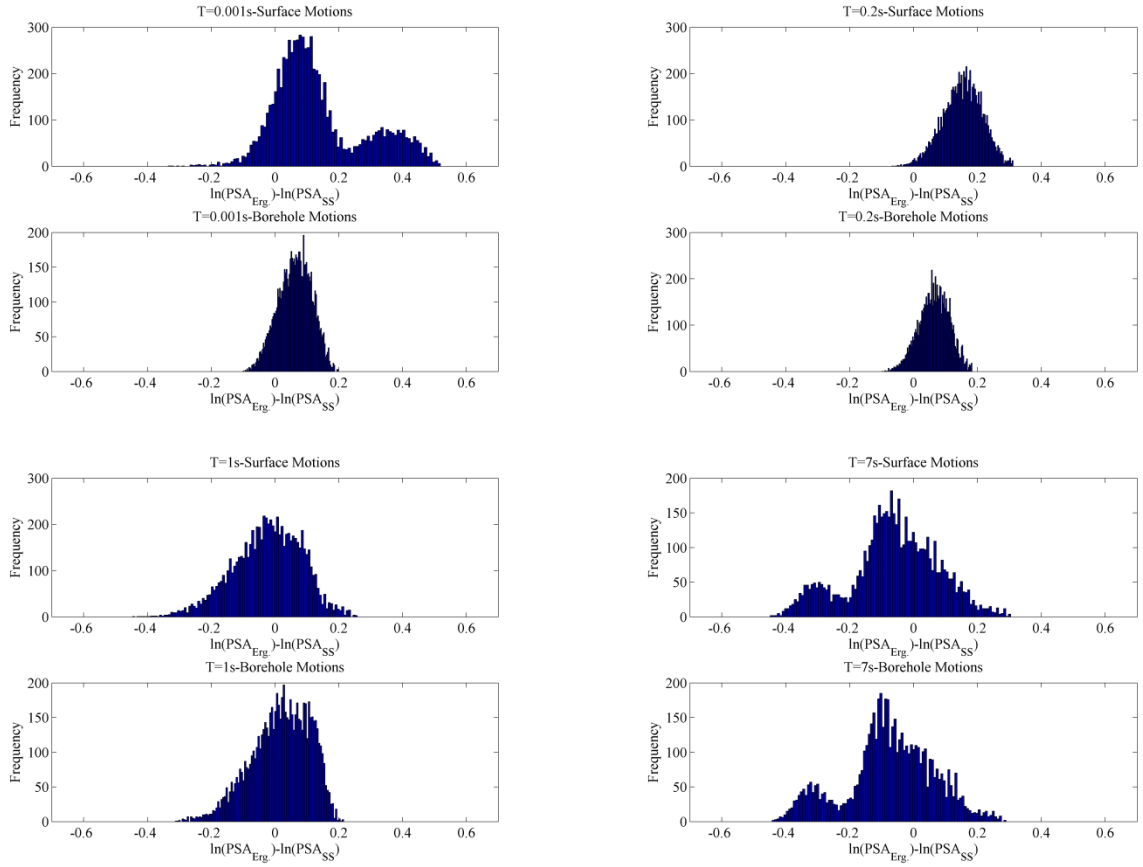
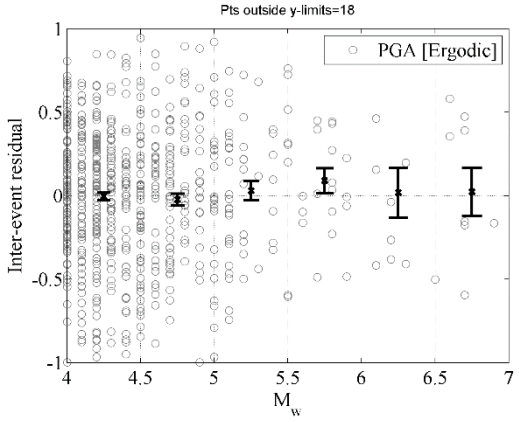
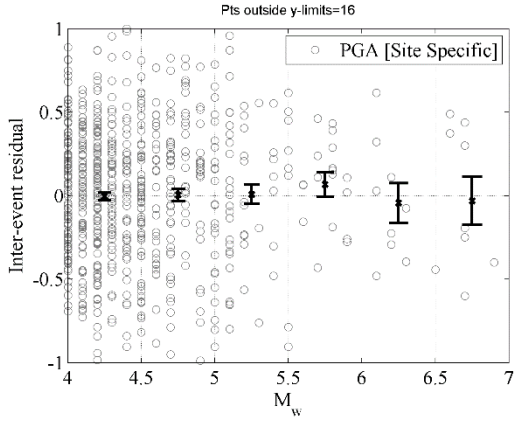


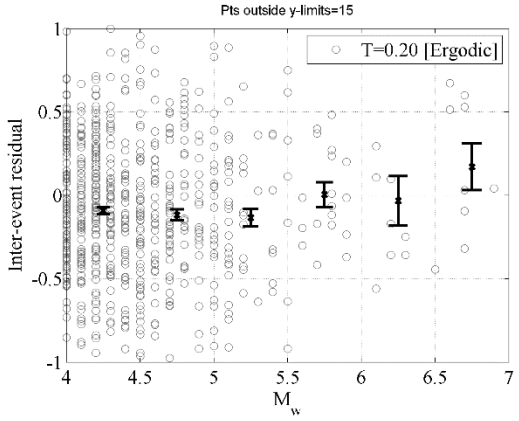
Figure D.10. Histograms showing the differences between the predictions from the ergodic and site-specific GMPEs at PGA (0.001s), 0.2s, 1.0s, and 7.0s. The histograms at a spectral period of 7.0s shows a bimodal behavior that inflates the standard deviation of the differences between the two formulatios

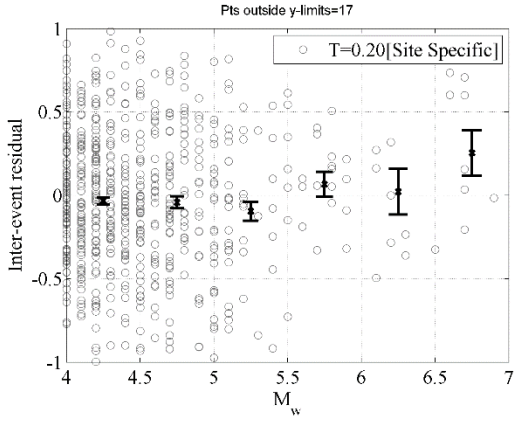
Residuals

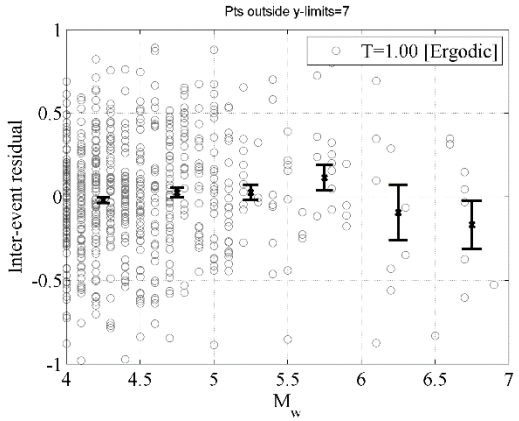
Figure D.11 shows a large set of scatter plots for the different residual components against different parameters at different spectral periods. The plots to the right are for the residuals obtained from the site-specific formulation, while the left column contains plots for residuals from the ergodic formulation. A subset of these plots is shown is presented in Dawood and Rodriguez-Marek (2014; chapter 5 in this dissertation)

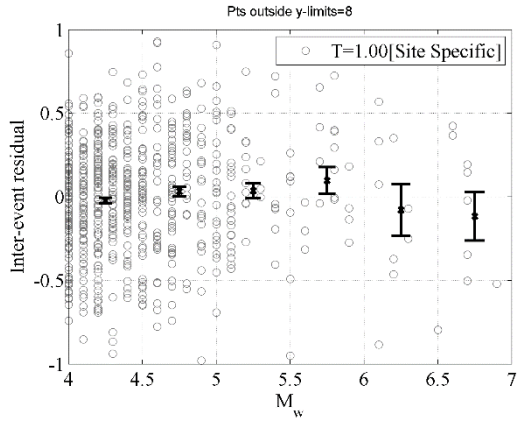

Pts outside y-limits=18
PGA [Ergodic]
Inter-event residual
M_w

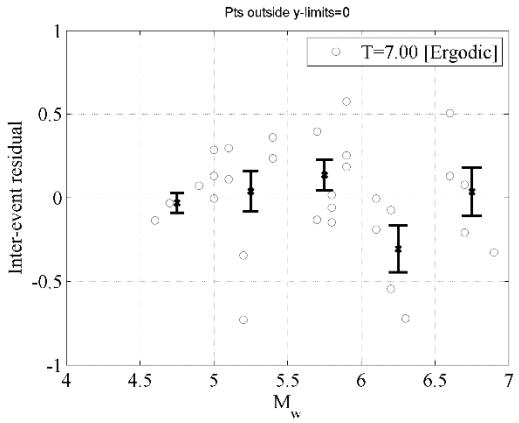

Pts outside y-limits=16
PGA [Site Specific]
Inter-event residual
M_w

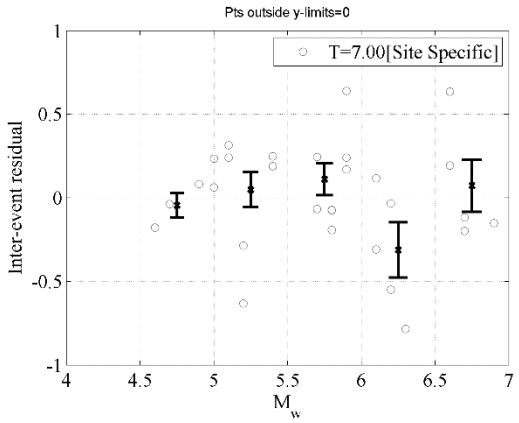

Pts outside y-limits=15
T=0.20 [Ergodic]
Inter-event residual
M_w

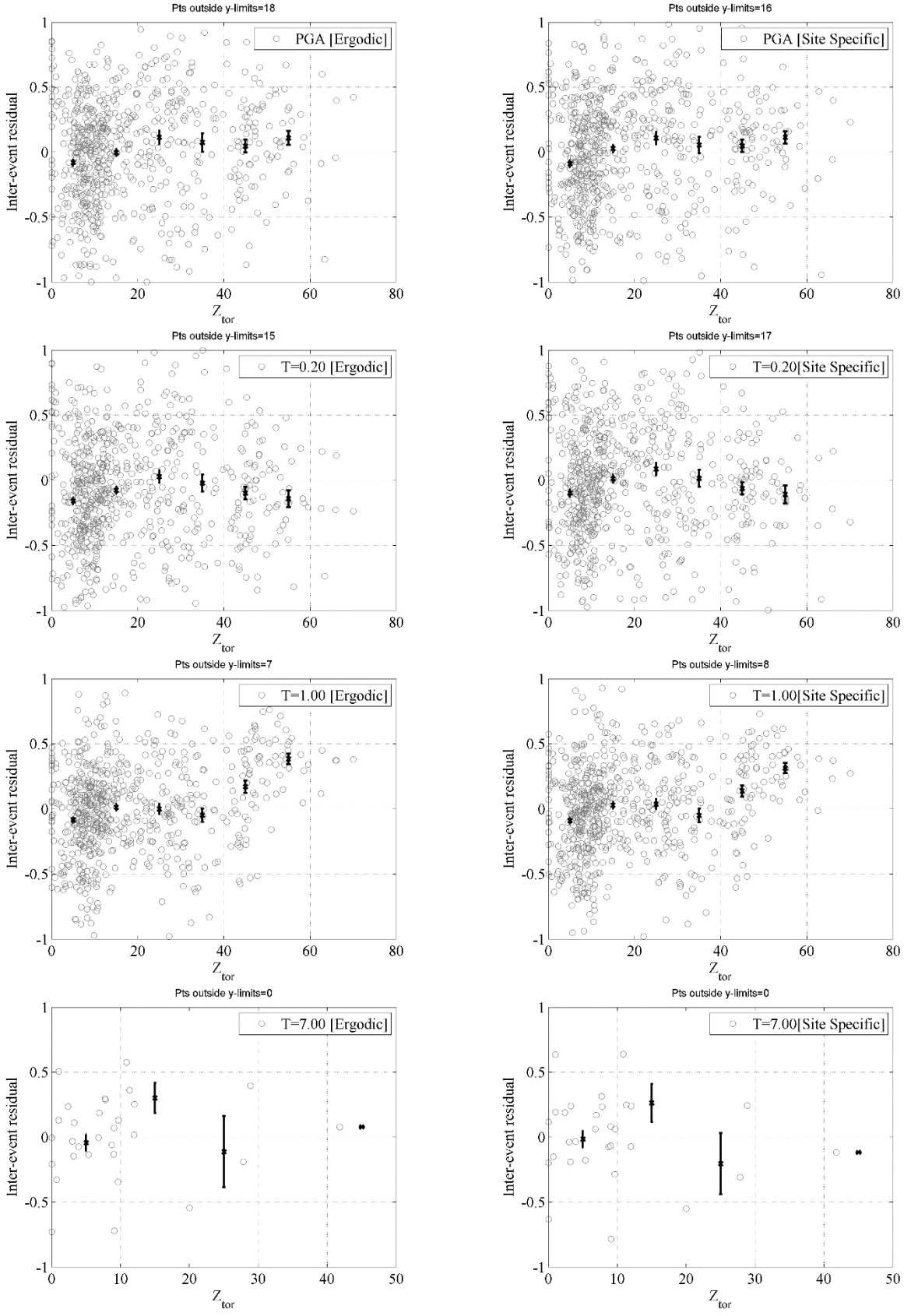

Pts outside y-limits=17
T=0.20[Site Specific]
Inter-event residual
M_w

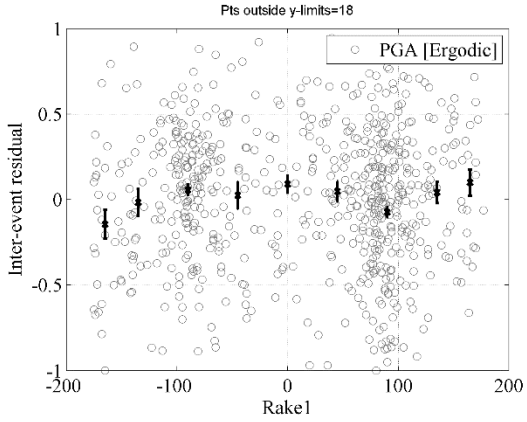

Pts outside y-limits=7
T=1.00 [Ergodic]
Inter-event residual
M_w

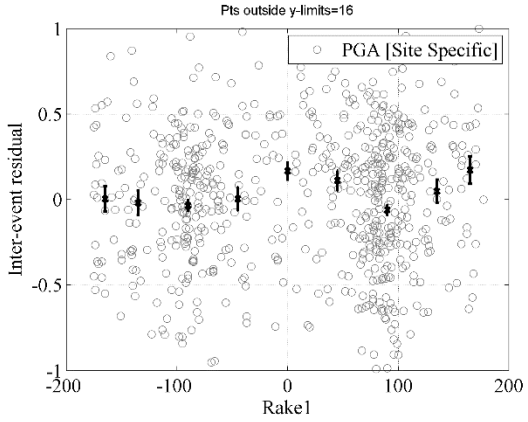

Pts outside y-limits=8
T=1.00[Site Specific]
Inter-event residual
M_w

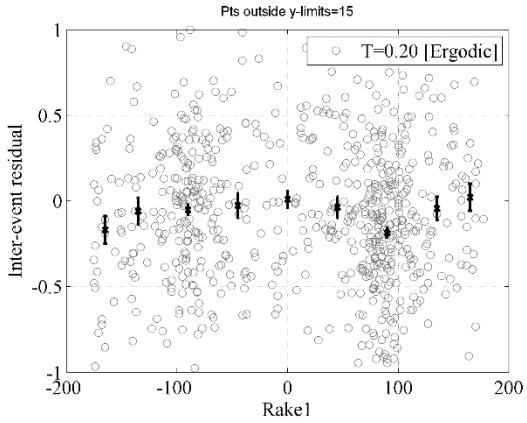

Pts outside y-limits=0
T=7.00 [Ergodic]
Inter-event residual
M_w

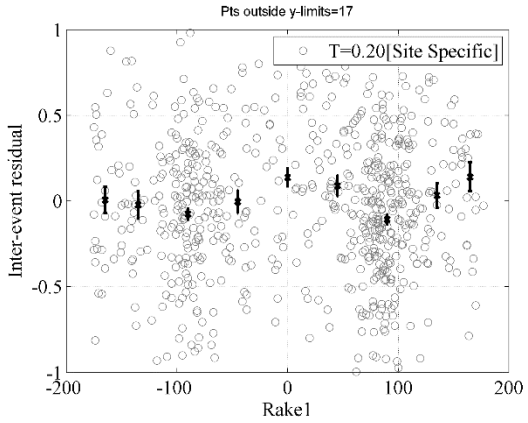

Pts outside y-limits=0
T=7.00[Site Specific]
Inter-event residual
M_w

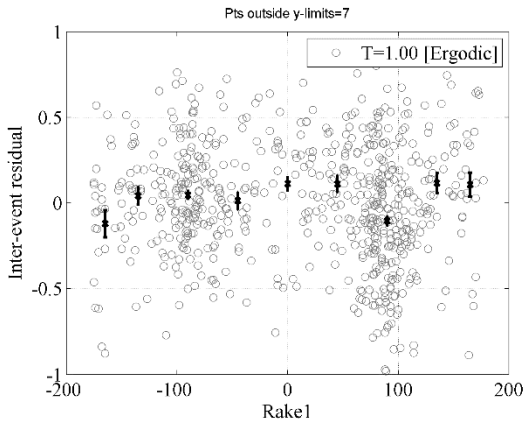

Pts outside y-limits=18
PGA [Ergodic]
Pts outside y-limits=16
PGA [Site Specific]
Pts outside y-limits=15
T=0.20 [Ergodic]
Pts outside y-limits=17
T=0.20[Site Specific]
Pts outside y-limits=7
T=1.00 [Ergodic]
Pts outside y-limits=8
T=1.00[Site Specific]
Pts outside y-limits=0
T=7.00 [Ergodic]
Pts outside y-limits=0
T=7.00[Site Specific]
Inter-event residual
Z_tor

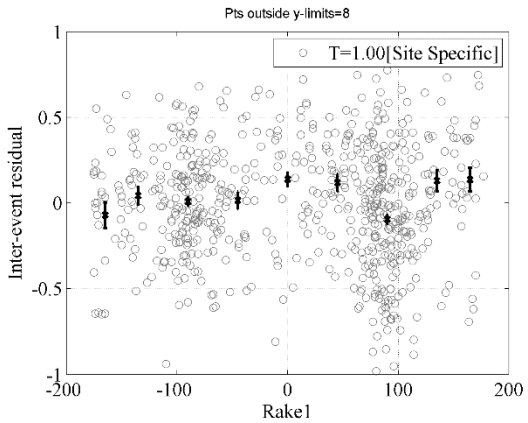

Pts outside y-limits=18
PGA [Ergodic]
Inter-event residual
Rake1

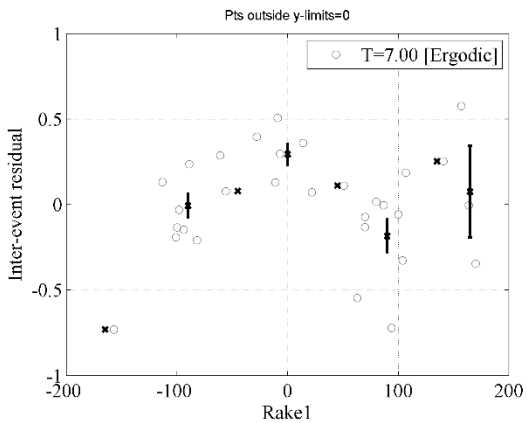

Pts outside y-limits=16
PGA [Site Specific]
Inter-event residual
Rake1

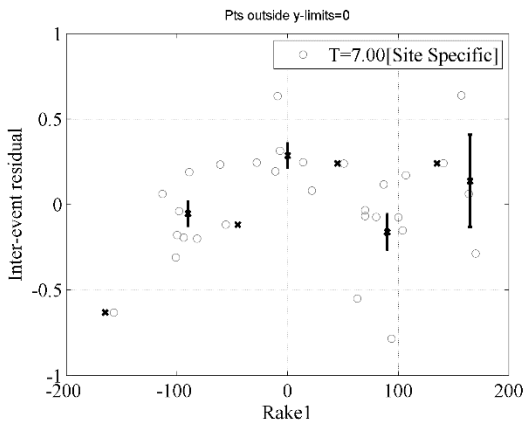

Pts outside y-limits=15
T=0.20 [Ergodic]
Inter-event residual
Rake1

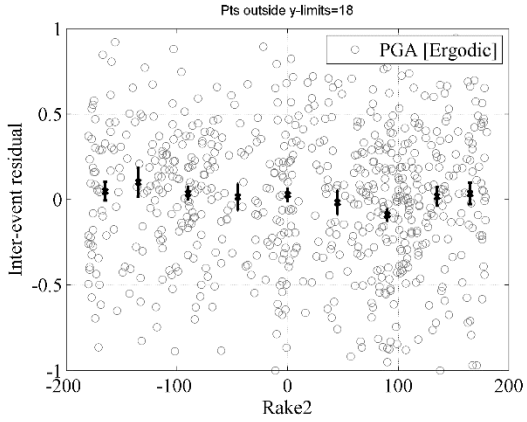

Pts outside y-limits=17
T=0.20[Site Specific]
Inter-event residual
Rake1

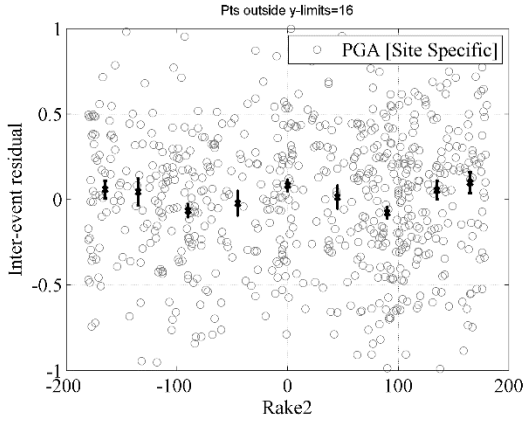

Pts outside y-limits=7
T=1.00 [Ergodic]
Inter-event residual
Rake1

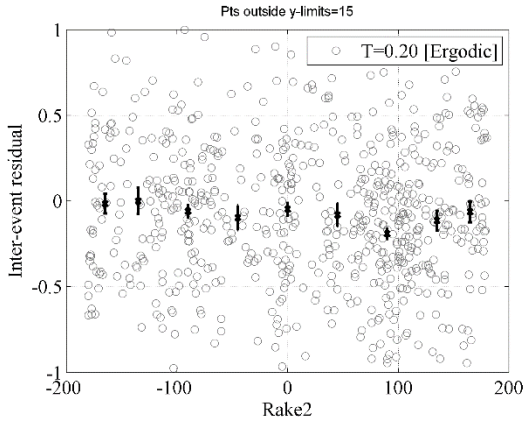

Pts outside y-limits=8
T=1.00[Site Specific]
Inter-event residual
Rake1

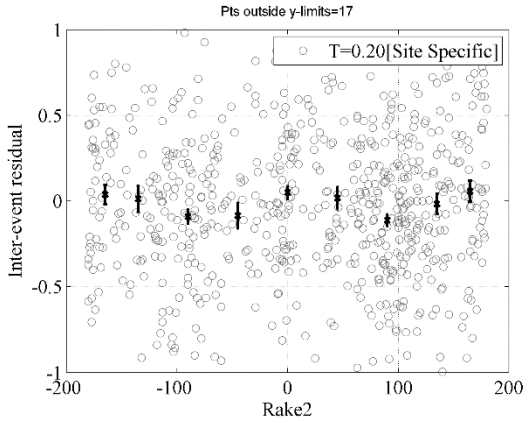

Pts outside y-limits=0
T=7.00 [Ergodic]
Inter-event residual
Rake1

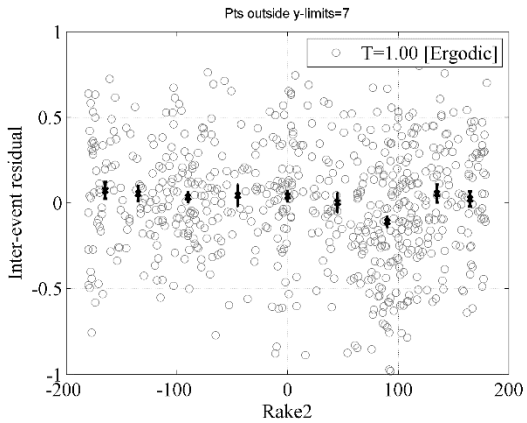

Pts outside y-limits=0
T=7.00[Site Specific]
Inter-event residual
Rake1

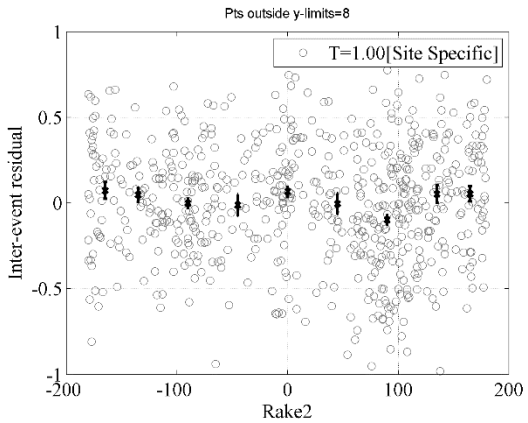

Pts outside y-limits=18
PGA [Ergodic]
Inter-event residual
Rake2

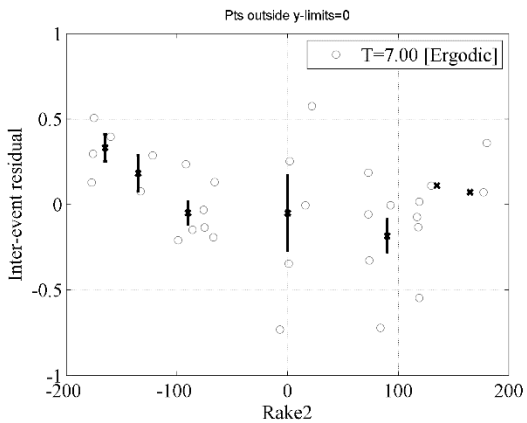

Pts outside y-limits=16
PGA [Site Specific]
Inter-event residual
Rake2

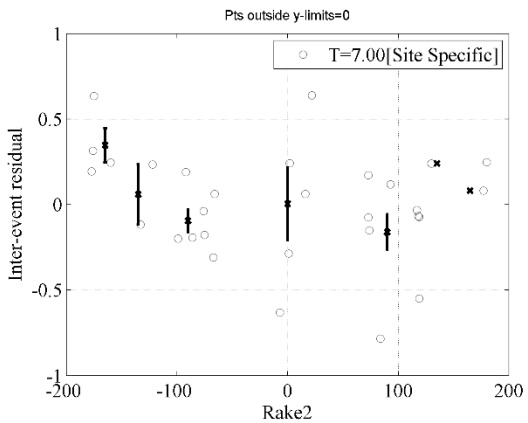

Pts outside y-limits=15
T=0.20 [Ergodic]
Inter-event residual
Rake2

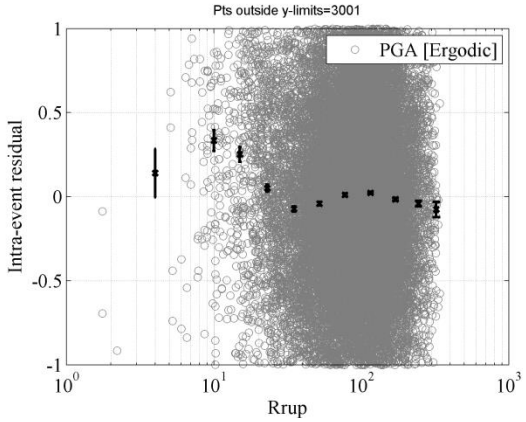

Pts outside y-limits=17
T=0.20[Site Specific]
Inter-event residual
Rake2

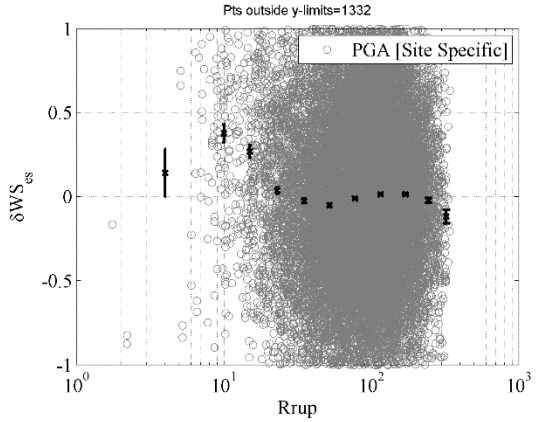

Pts outside y-limits=7
T=1.00 [Ergodic]
Inter-event residual
Rake2

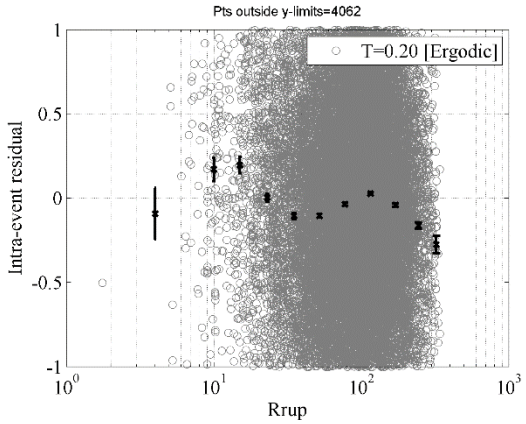

Pts outside y-limits=8
T=1.00[Site Specific]
Inter-event residual
Rake2

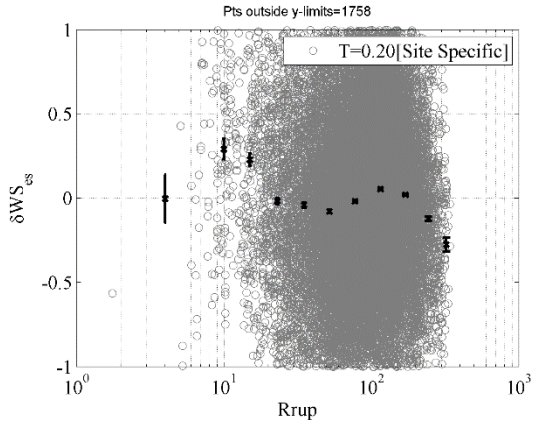

Pts outside y-limits=0
T=7.00 [Ergodic]
Inter-event residual
Rake2

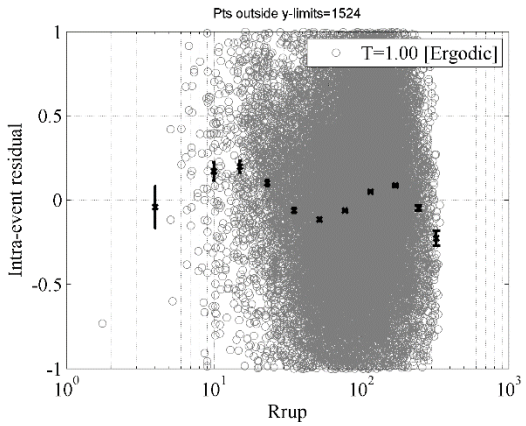

Pts outside y-limits=0
T=7.00[Site Specific]
Inter-event residual
Rake2

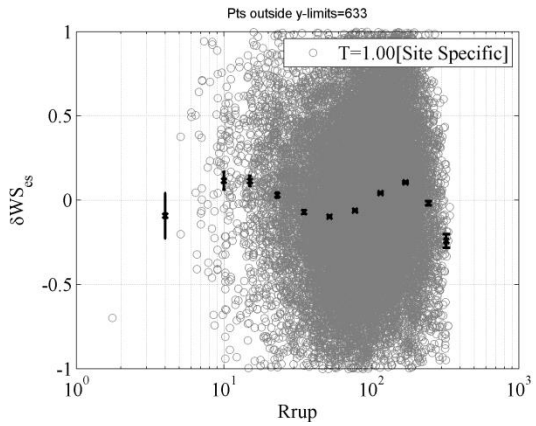

Pts outside y-limits=3001
PGA [Ergodic]
Intra-event residual
Rrup

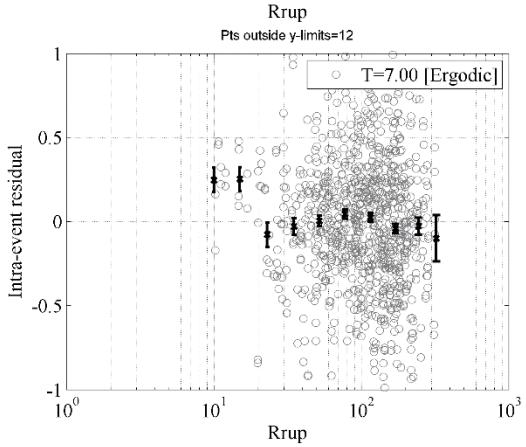

Pts outside y-limits=1332
PGA [Site Specific]
δWS_es
Rrup

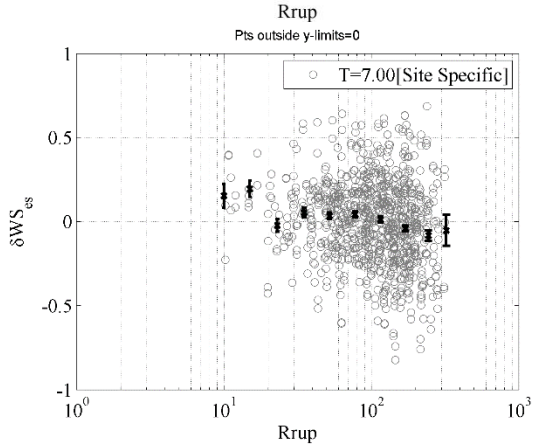

Pts outside y-limits=4062
T=0.20 [Ergodic]
Intra-event residual
Rrup

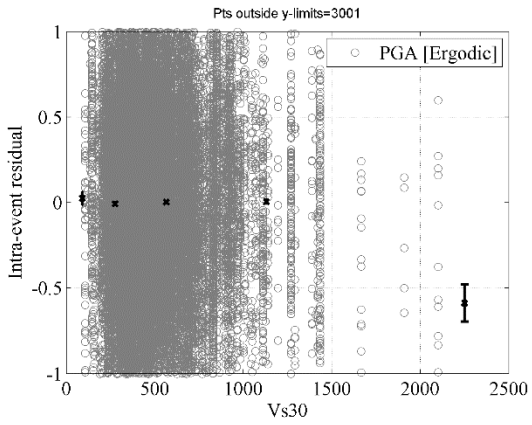

Pts outside y-limits=1758
T=0.20[Site Specific]
δWS_es
Rrup

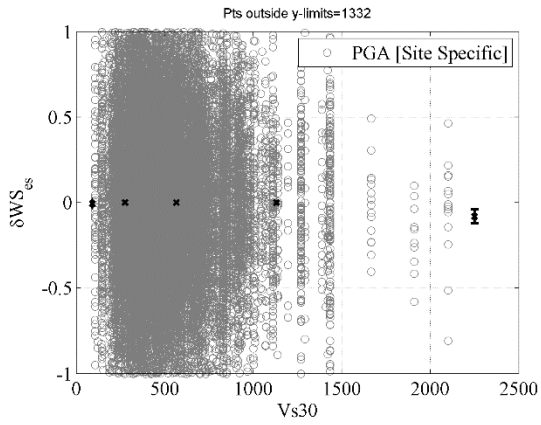

Pts outside y-limits=1524
T=1.00 [Ergodic]
Intra-event residual
Rrup

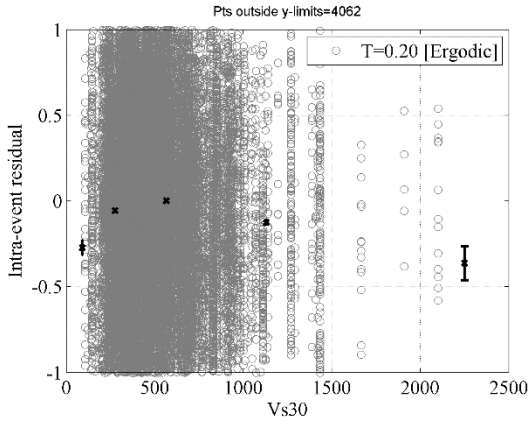

Pts outside y-limits=633
T=1.00[Site Specific]
δWS_es
Rrup

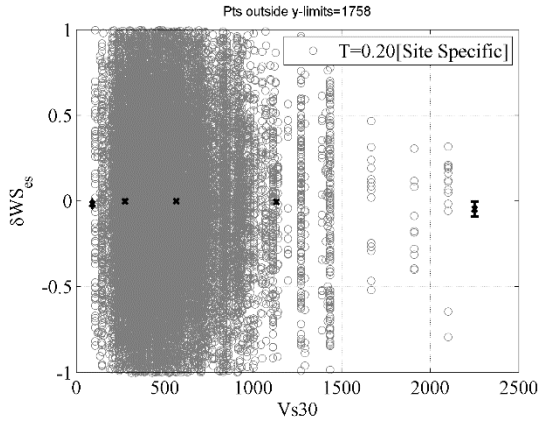

Pts outside y-limits=12
T=7.00 [Ergodic]
Intra-event residual
Rrup

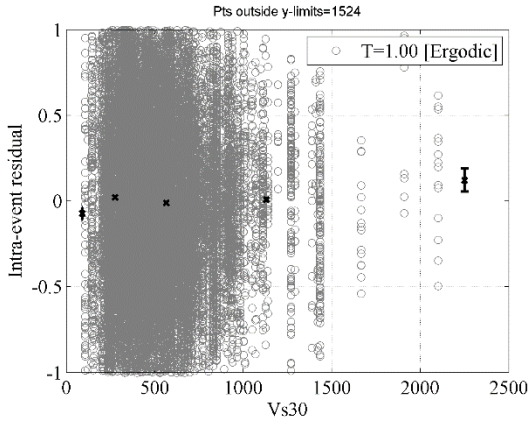

Pts outside y-limits=0
T=7.00[Site Specific]
δWS_es
Rrup

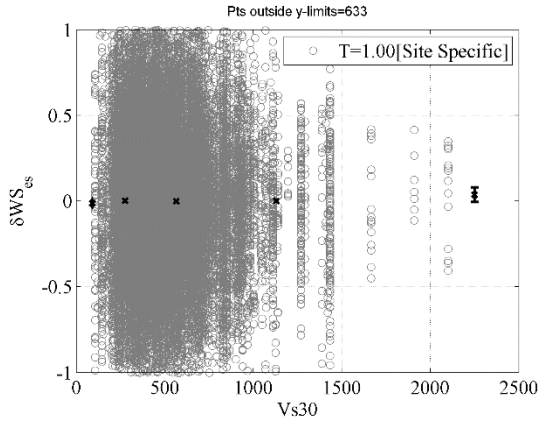

Pts outside y-limits=3001
PGA [Ergodic]
Intra-event residual
Vs30

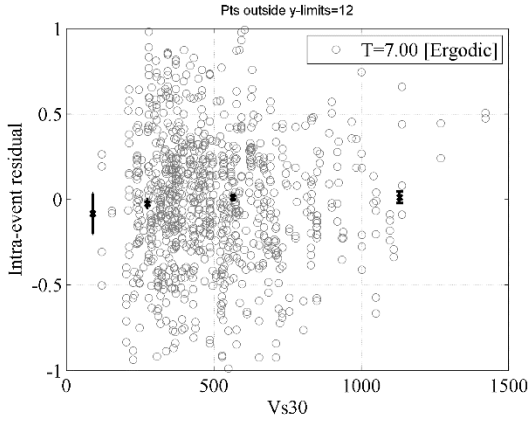

Pts outside y-limits=1332
PGA [Site Specific]
δWS_es
Vs30

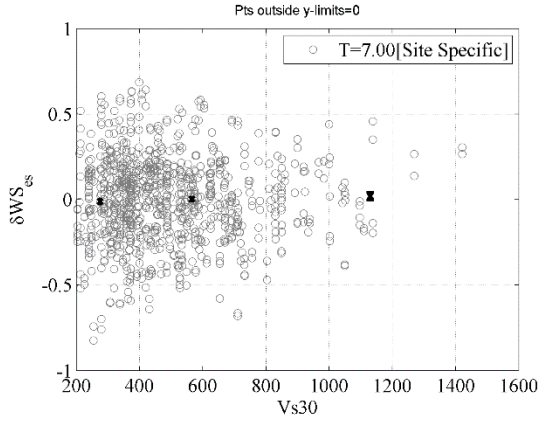

Pts outside y-limits=4062
T=0.20 [Ergodic]
Intra-event residual
Vs30

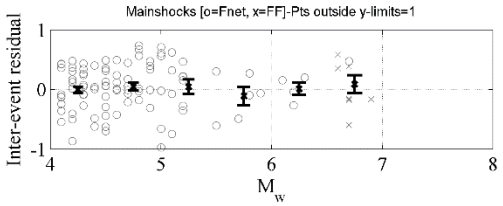

Pts outside y-limits=1758
T=0.20[Site Specific]
δWS_es
Vs30

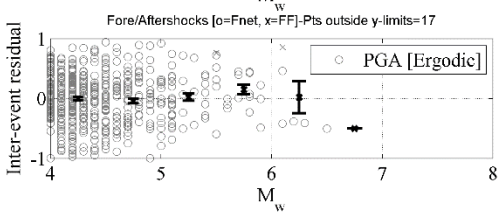

Pts outside y-limits=1524
T=1.00 [Ergodic]
Intra-event residual
Vs30

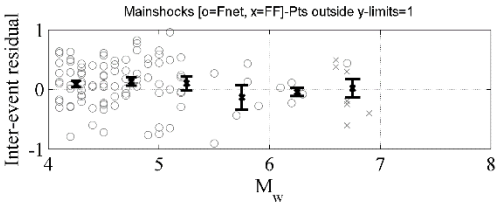

Pts outside y-limits=633
T=1.00[Site Specific]
δWS_es
Vs30

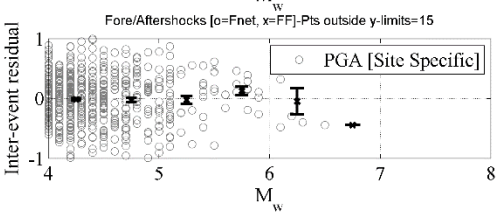

Pts outside y-limits=12
T=7.00 [Ergodic]
Intra-event residual
Vs30

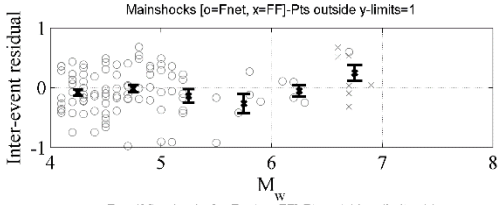

Pts outside y-limits=0
T=7.00[Site Specific]
δWS_es
Vs30

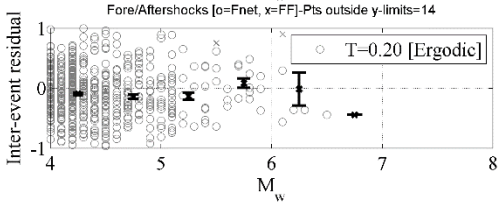

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
M_w

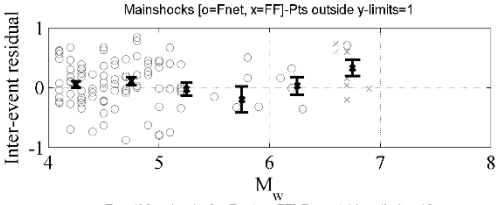

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=17
PGA [Ergodic]
Inter-event residual
M_w

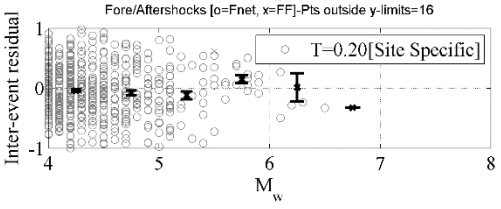

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
M_w

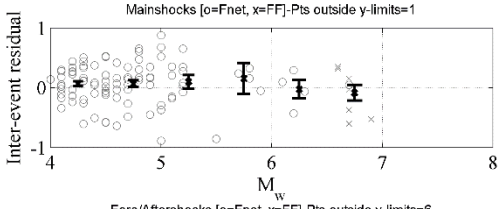

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=15
PGA [Site Specific]
Inter-event residual
M_w

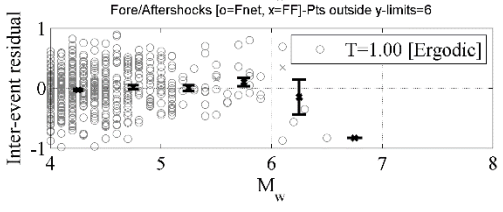

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
M_w

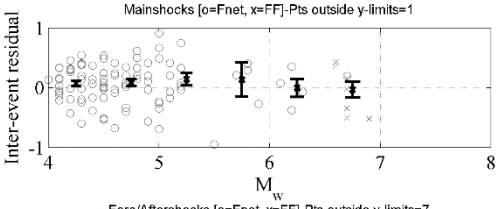

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=14
T=0.20 [Ergodic]
Inter-event residual
M_w

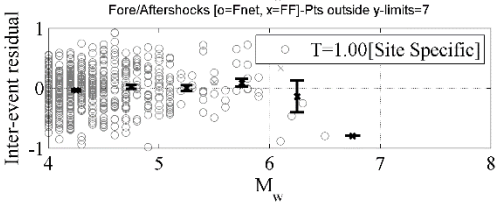

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
M_w

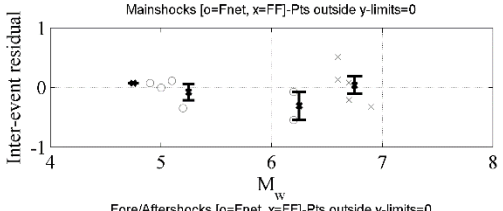

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=16
T=0.20[Site Specific]
Inter-event residual
M_w

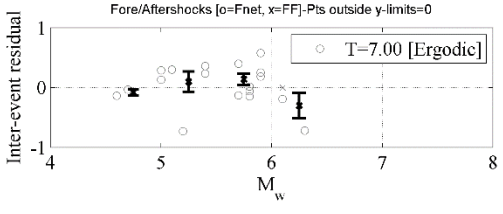

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
M_w

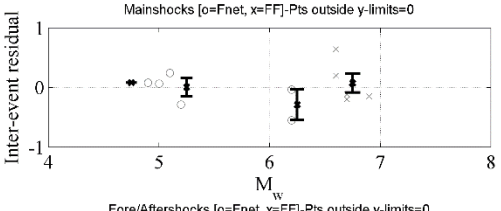

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=6
T=1.00 [Ergodic]
Inter-event residual
M_w

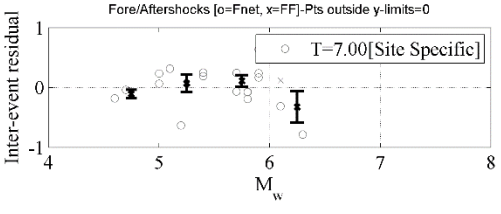

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
M_w

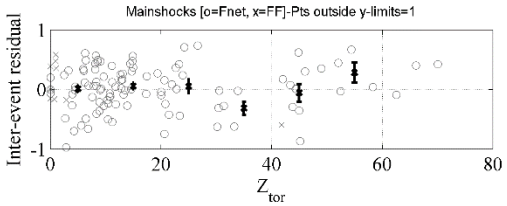

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=7
T=1.00[Site Specific]
Inter-event residual
M_w

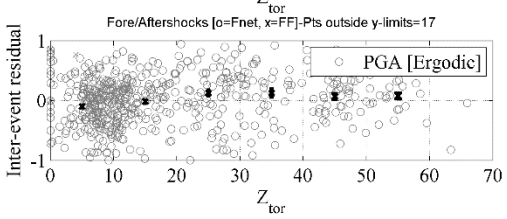

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=0
Inter-event residual
M_w

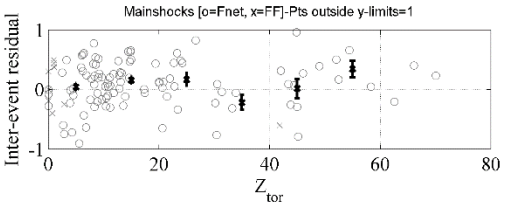

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=0
T=7.00 [Ergodic]
Inter-event residual
M_w

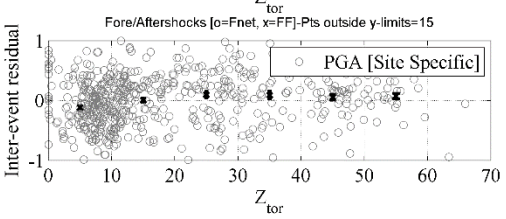

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=0
Inter-event residual
M_w

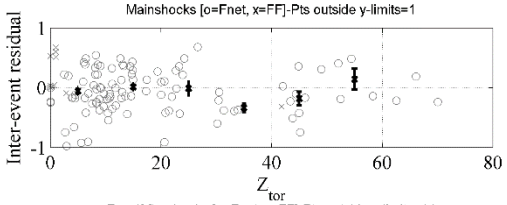

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=0
T=7.00[Site Specific]
Inter-event residual
M_w

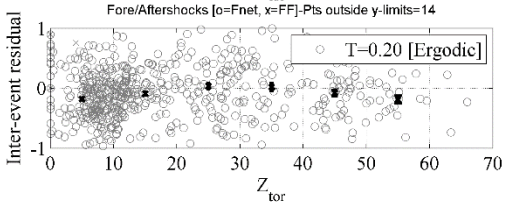

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
Z_tor

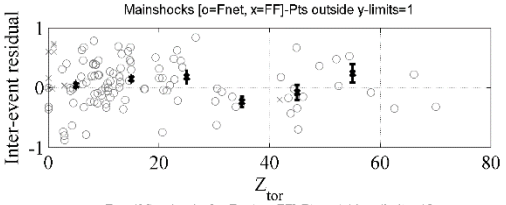

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=17
PGA [Ergodic]
Inter-event residual
Z_tor

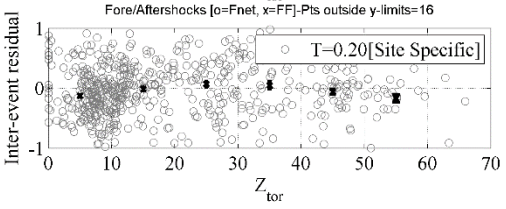

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
Z_tor

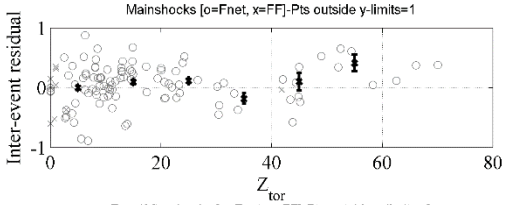

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=15
PGA [Site Specific]
Inter-event residual
Z_tor

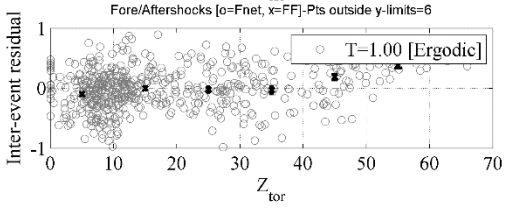

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
Z_tor

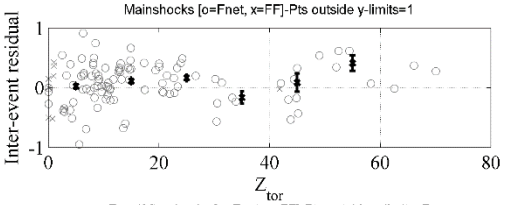

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=14
T=0.20 [Ergodic]
Inter-event residual
Z_tor

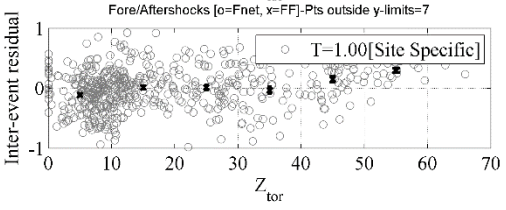

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
Z_tor

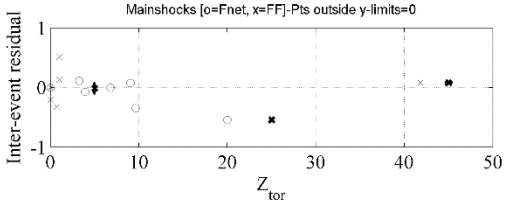

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=16
T=0.20[Site Specific]
Inter-event residual
Z_tor

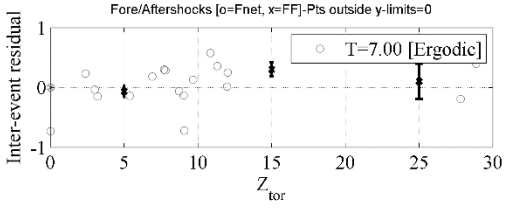

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
Z_tor

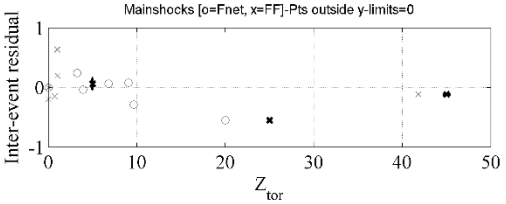

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=6
T=1.00 [Ergodic]
Inter-event residual
Z_tor

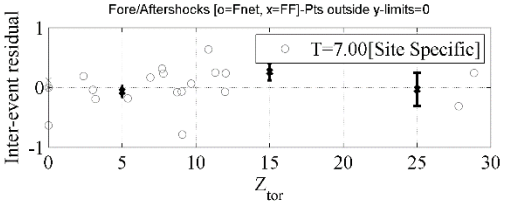

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=1
Inter-event residual
Z_tor

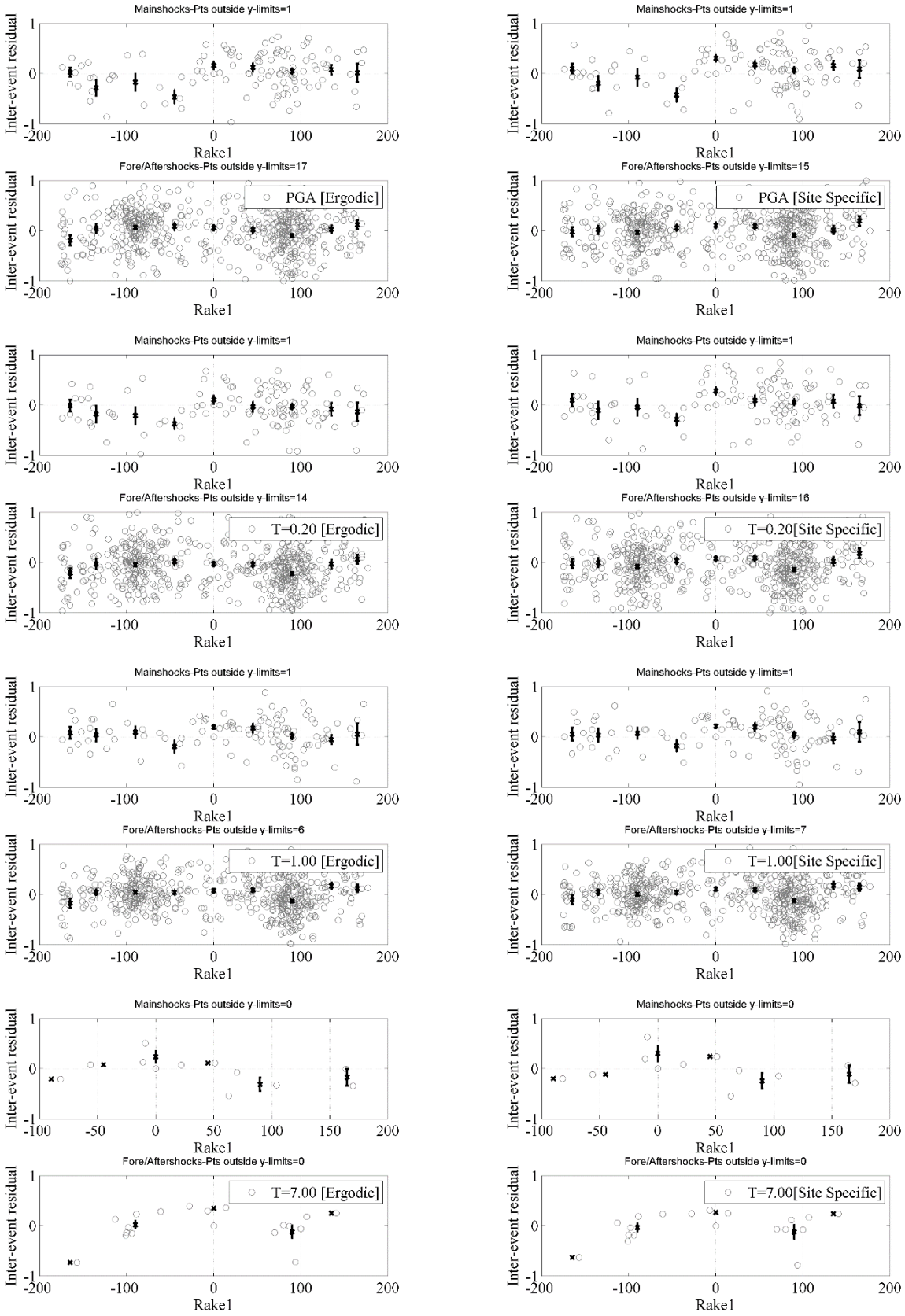

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=7
T=1.00[Site Specific]
Inter-event residual
Z_tor

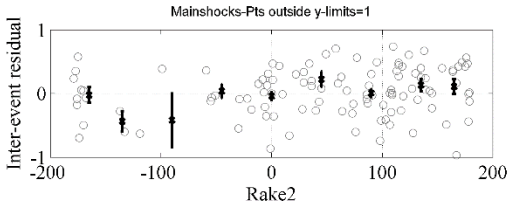

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=0
Inter-event residual
Z_tor

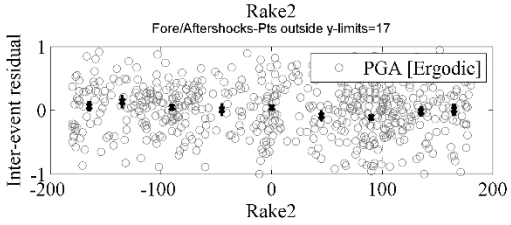

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=0
T=7.00 [Ergodic]
Inter-event residual
Z_tor

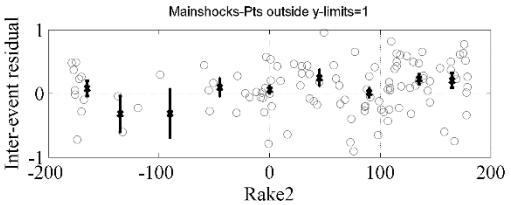

Mainshocks [o=Fnet, x=FF]-Pts outside y-limits=0
Inter-event residual
Z_tor

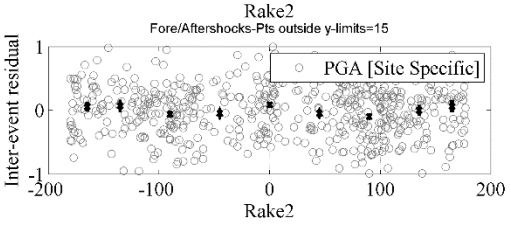

Fore/Aftershocks [o=Fnet, x=FF]-Pts outside y-limits=0
T=7.00[Site Specific]
Inter-event residual
Z_tor

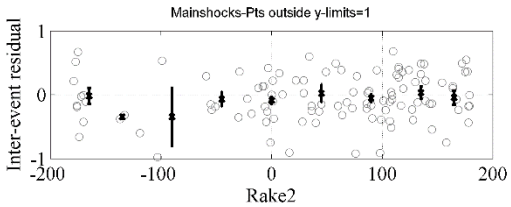

Mainshocks-Pts outside y-limits=1
Inter-event residual
Rake1
Fore/Aftershocks-Pts outside y-limits=17
PGA [Ergodic]
Mainshocks-Pts outside y-limits=1
Fore/Aftershocks-Pts outside y-limits=15
PGA [Site Specific]
Mainshocks-Pts outside y-limits=1
Fore/Aftershocks-Pts outside y-limits=14
T=0.20 [Ergodic]
Mainshocks-Pts outside y-limits=1
Fore/Aftershocks-Pts outside y-limits=16
T=0.20[Site Specific]
Mainshocks-Pts outside y-limits=1
Fore/Aftershocks-Pts outside y-limits=6
T=1.00 [Ergodic]
Mainshocks-Pts outside y-limits=1
Fore/Aftershocks-Pts outside y-limits=7
T=1.00[Site Specific]
Mainshocks-Pts outside y-limits=0
Fore/Aftershocks-Pts outside y-limits=0
T=7.00 [Ergodic]
Mainshocks-Pts outside y-limits=0
Fore/Aftershocks-Pts outside y-limits=0
T=7.00[Site Specific]

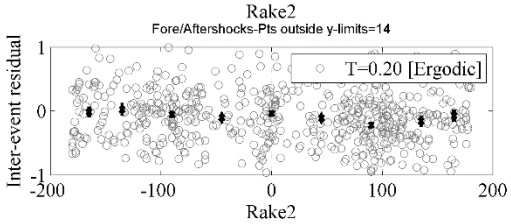

Mainshocks-Pts outside y-limits=1
Inter-event residual
Rake2

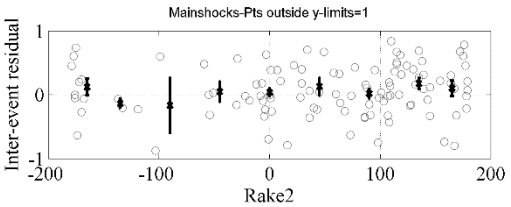

Fore/Aftershocks-Pts outside y-limits=17
PGA [Ergodic]
Inter-event residual
Rake2

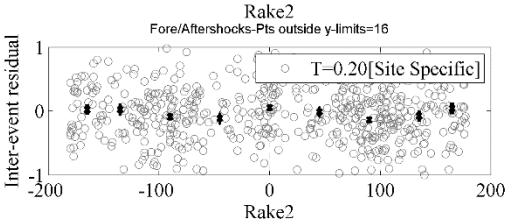

Mainshocks-Pts outside y-limits=1
Inter-event residual
Rake2

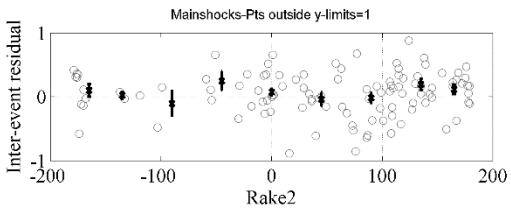

Fore/Aftershocks-Pts outside y-limits=15
PGA [Site Specific]
Inter-event residual
Rake2

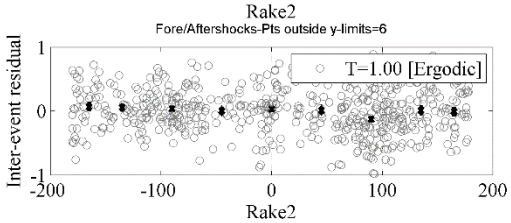

Mainshocks-Pts outside y-limits=1
Inter-event residual
Rake2

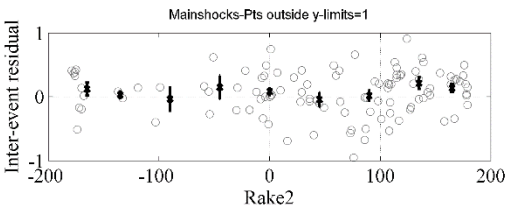

Fore/Aftershocks-Pts outside y-limits=14
T=0.20 [Ergodic]
Inter-event residual
Rake2

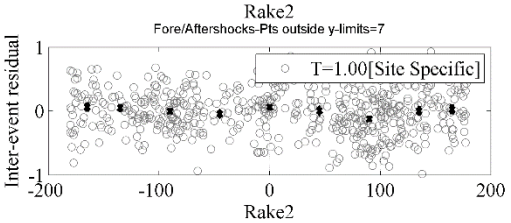

Mainshocks-Pts outside y-limits=1
Inter-event residual
Rake2

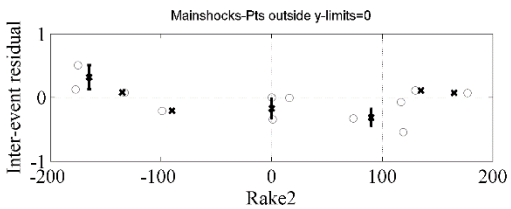

Fore/Aftershocks-Pts outside y-limits=16
T=0.20[Site Specific]
Inter-event residual
Rake2

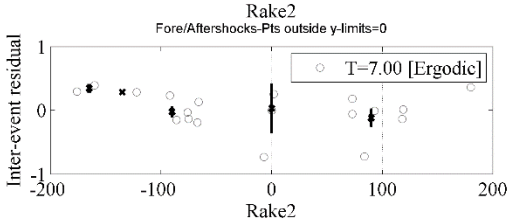

Mainshocks-Pts outside y-limits=1
Inter-event residual
Rake2

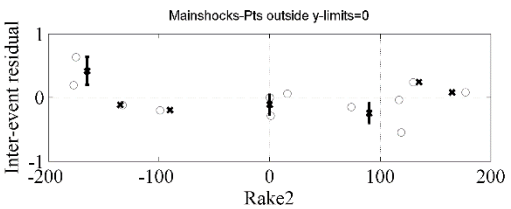

Fore/Aftershocks-Pts outside y-limits=6
T=1.00 [Ergodic]
Inter-event residual
Rake2

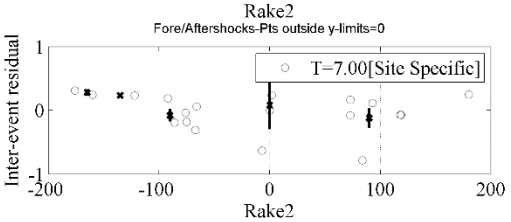

Mainshocks-Pts outside y-limits=1
Inter-event residual
Rake2

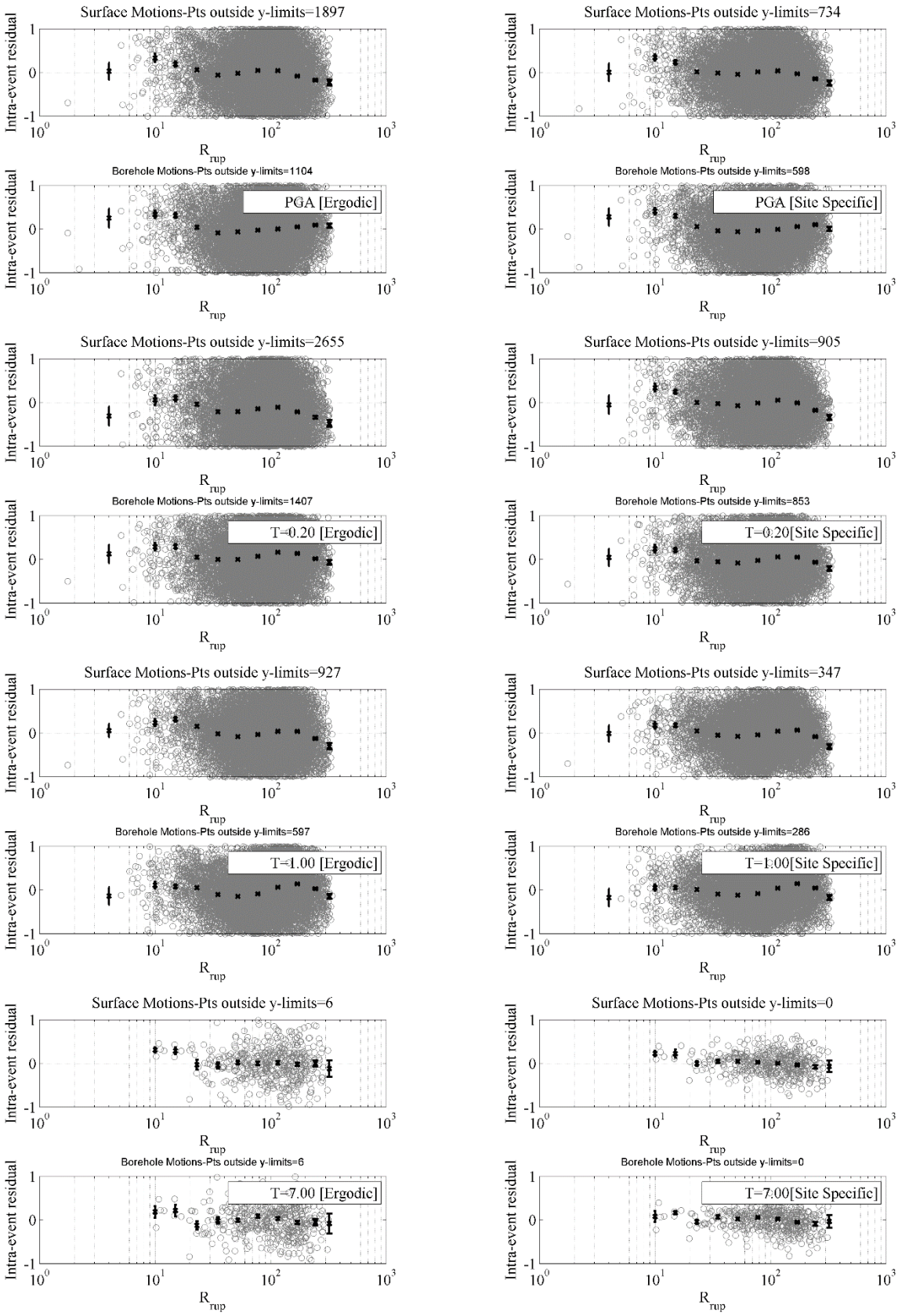

Fore/Aftershocks-Pts outside y-limits=7
T=1.00[Site Specific]
Inter-event residual
Rake2

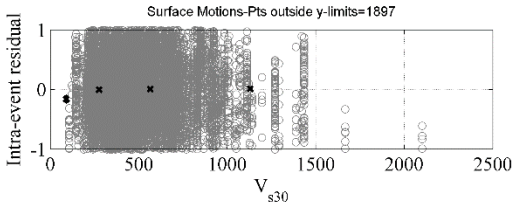

Mainshocks-Pts outside y-limits=0
Inter-event residual
Rake2

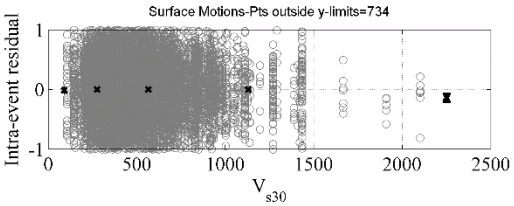

Fore/Aftershocks-Pts outside y-limits=0
T=7.00 [Ergodic]
Inter-event residual
Rake2

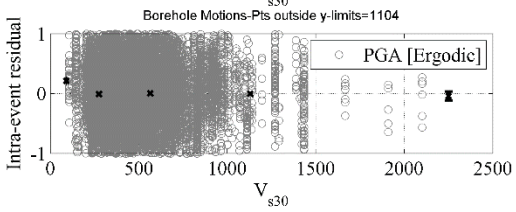

Mainshocks-Pts outside y-limits=0
Inter-event residual
Rake2

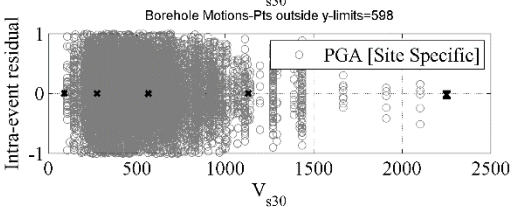

Fore/Aftershocks-Pts outside y-limits=0
T=7.00[Site Specific]
Inter-event residual
Rake2

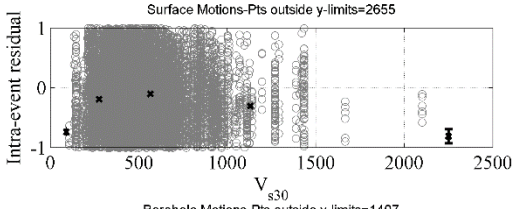

Surface Motions-Pts outside y-limits=1897
Borehole Motions-Pts outside y-limits=1104
PGA [Ergodic]
Surface Motions-Pts outside y-limits=734
Borehole Motions-Pts outside y-limits=598
PGA [Site Specific]
Surface Motions-Pts outside y-limits=2655
Borehole Motions-Pts outside y-limits=1407
T=0.20 [Ergodic]
Surface Motions-Pts outside y-limits=905
Borehole Motions-Pts outside y-limits=853
T=0.20[Site Specific]
Surface Motions-Pts outside y-limits=927
Borehole Motions-Pts outside y-limits=597
T=1.00 [Ergodic]
Surface Motions-Pts outside y-limits=347
Borehole Motions-Pts outside y-limits=286
T=1.00[Site Specific]
Surface Motions-Pts outside y-limits=6
Borehole Motions-Pts outside y-limits=6
T=7.00 [Ergodic]
Surface Motions-Pts outside y-limits=0
Borehole Motions-Pts outside y-limits=0
T=7.00[Site Specific]
Intra-event residual
R rup

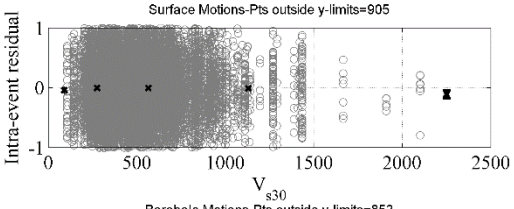

Surface Motions-Pts outside y-limits=1897
Intra-event residual
V s30

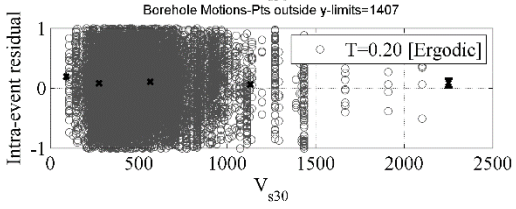

Surface Motions-Pts outside y-limits=734
Intra-event residual
V s30

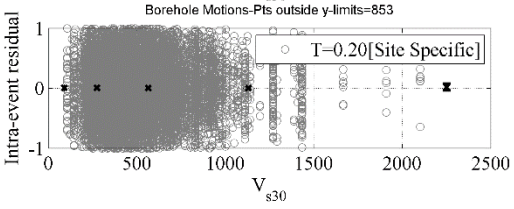

Borehole Motions-Pts outside y-limits=1104
PGA [Ergodic]
Intra-event residual
V s30

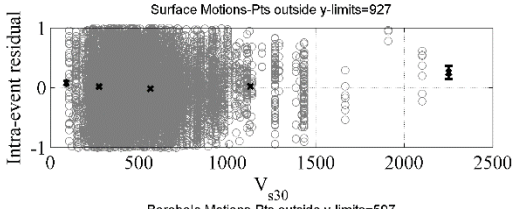

Borehole Motions-Pts outside y-limits=598
PGA [Site Specific]
Intra-event residual
V s30

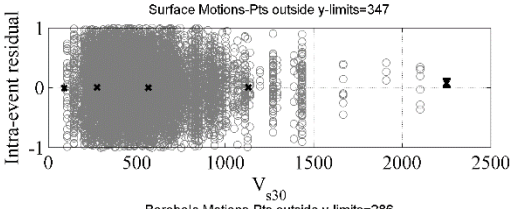

Surface Motions-Pts outside y-limits=2655
Intra-event residual
V s30

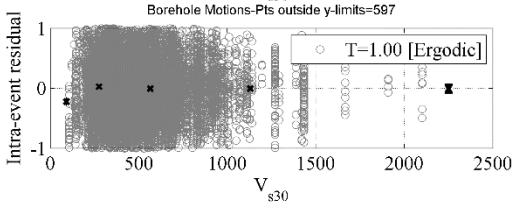

Surface Motions-Pts outside y-limits=905
Intra-event residual
V s30

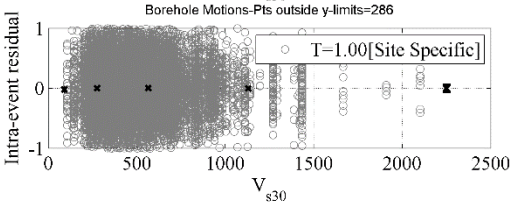

Borehole Motions-Pts outside y-limits=1407
T=0.20 [Ergodic]
Intra-event residual
V s30

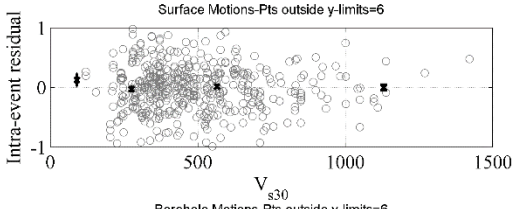

Borehole Motions-Pts outside y-limits=853
T=0.20[Site Specific]
Intra-event residual
V s30

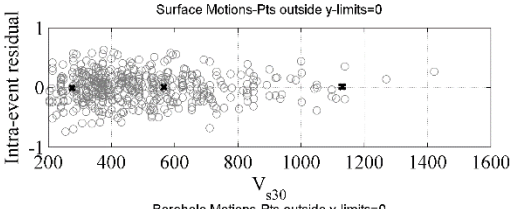

Surface Motions-Pts outside y-limits=927
Intra-event residual
V s30

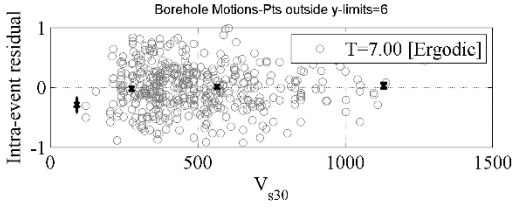

Surface Motions-Pts outside y-limits=347
Intra-event residual
V s30

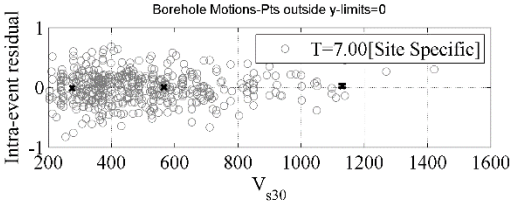

Borehole Motions-Pts outside y-limits=597
T=1.00 [Ergodic]
Intra-event residual
V s30


Borehole Motions-Pts outside y-limits=286
T=1.00[Site Specific]
Intra-event residual
V s30

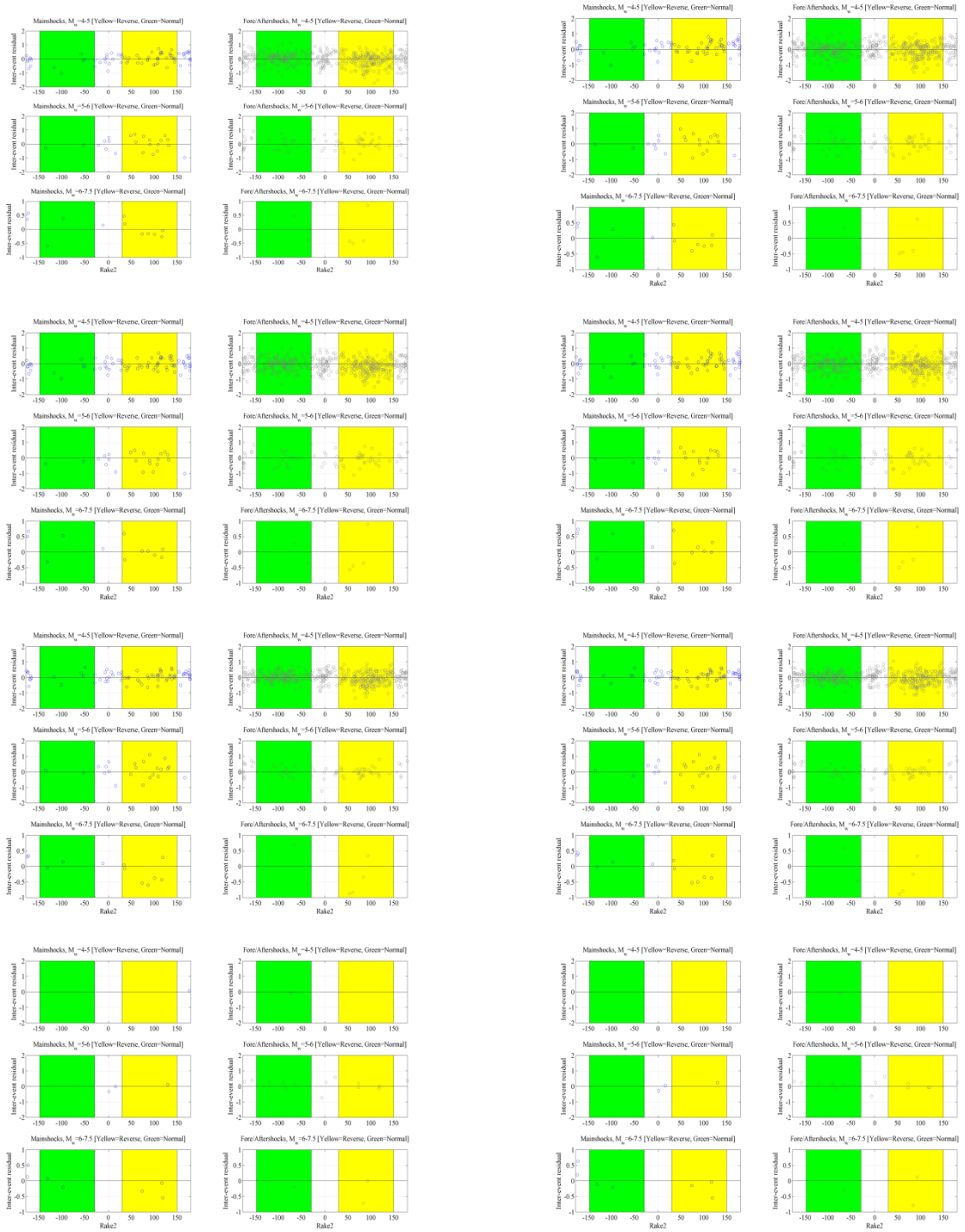

Surface Motions-Pts outside y-limits=6
Intra-event residual
V s30


Surface Motions-Pts outside y-limits=0
Intra-event residual
V s30


Borehole Motions-Pts outside y-limits=6
T=7.00 [Ergodic]
Intra-event residual
V s30


Borehole Motions-Pts outside y-limits=0
T=7.00[Site Specific]
Intra-event residual
V s30


Mainshocks, Mw=4-5 [Yellow=Reverse, Green=Normal]
Fore/Aftershocks, Mw=4-5 [Yellow=Reverse, Green=Normal]
Mainshocks, Mw=5-6 [Yellow=Reverse, Green=Normal]
Fore/Aftershocks, Mw=5-6 [Yellow=Reverse, Green=Normal]
Mainshocks, Mw=6-7.5 [Yellow=Reverse, Green=Normal]
Fore/Aftershocks, Mw=6-7.5 [Yellow=Reverse, Green=Normal]
Inter-event residual
Rake1


Mainshocks, Mw=4-5 [Yellow=Reverse, Green=Normal]
Fore/Aftershocks, Mw=4-5 [Yellow=Reverse, Green=Normal]
Mainshocks, Mw=5-6 [Yellow=Reverse, Green=Normal]
Fore/Aftershocks, Mw=5-6 [Yellow=Reverse, Green=Normal]
Mainshocks, Mw=6-7.5 [Yellow=Reverse, Green=Normal]
Fore/Aftershocks, Mw=6-7.5 [Yellow=Reverse, Green=Normal]
Inter-event residual
Rake2

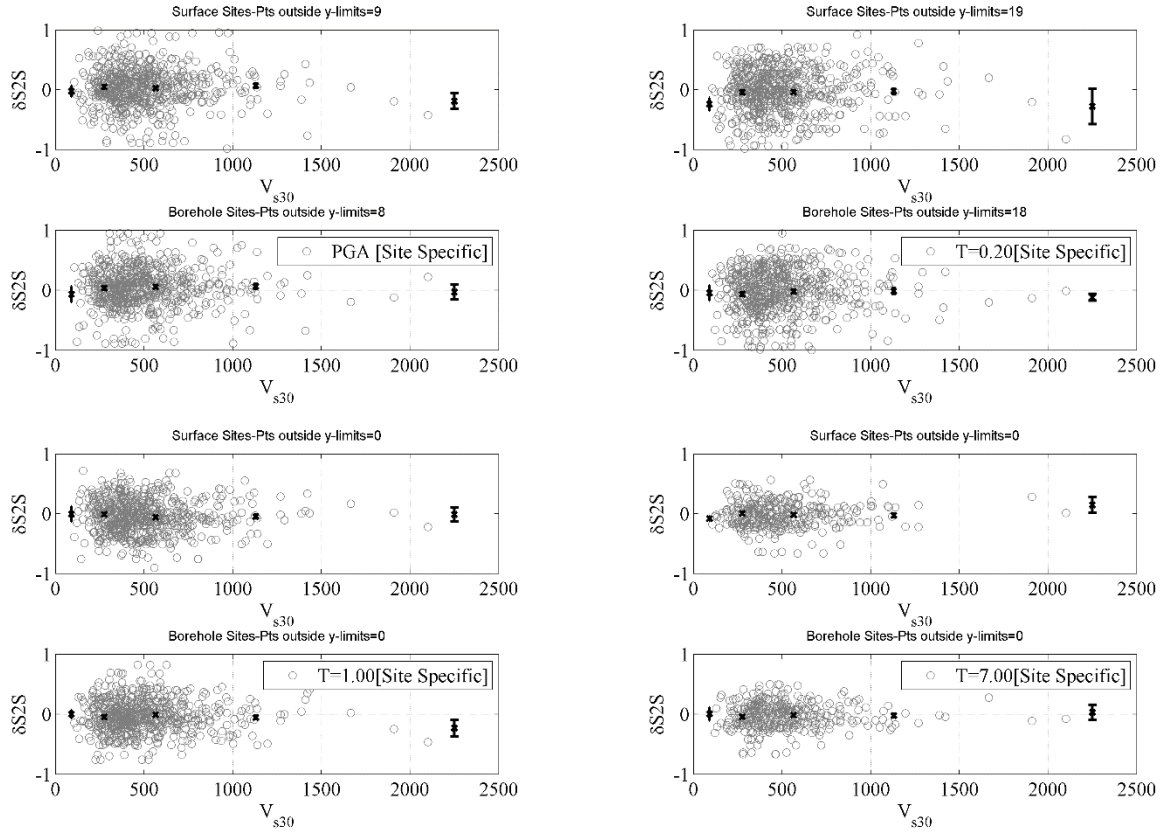


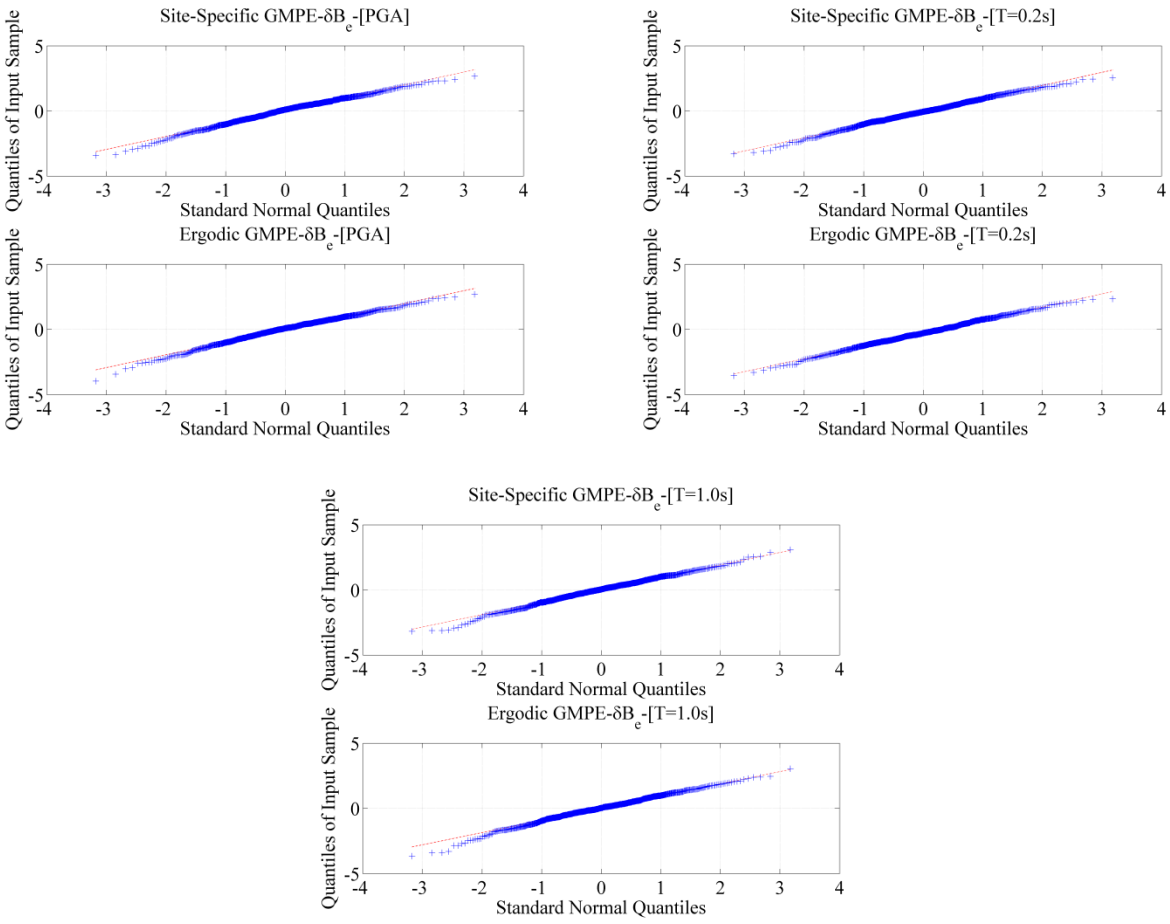
Figure D.11. Scatter plots for the different residual components against different parameters at different spectral periods. The plots to the right are for the residuals obtained from the site-specific formulation, while the left column contains plots for residuals from the ergodic formulation (except the two bottom rows are residuals from sit-especific components). The residuals are presented as gray open circles, while the average of the residuals at different bins is presented as a black cross with error bars that represent a $\pm$ standard deviations from the mean.

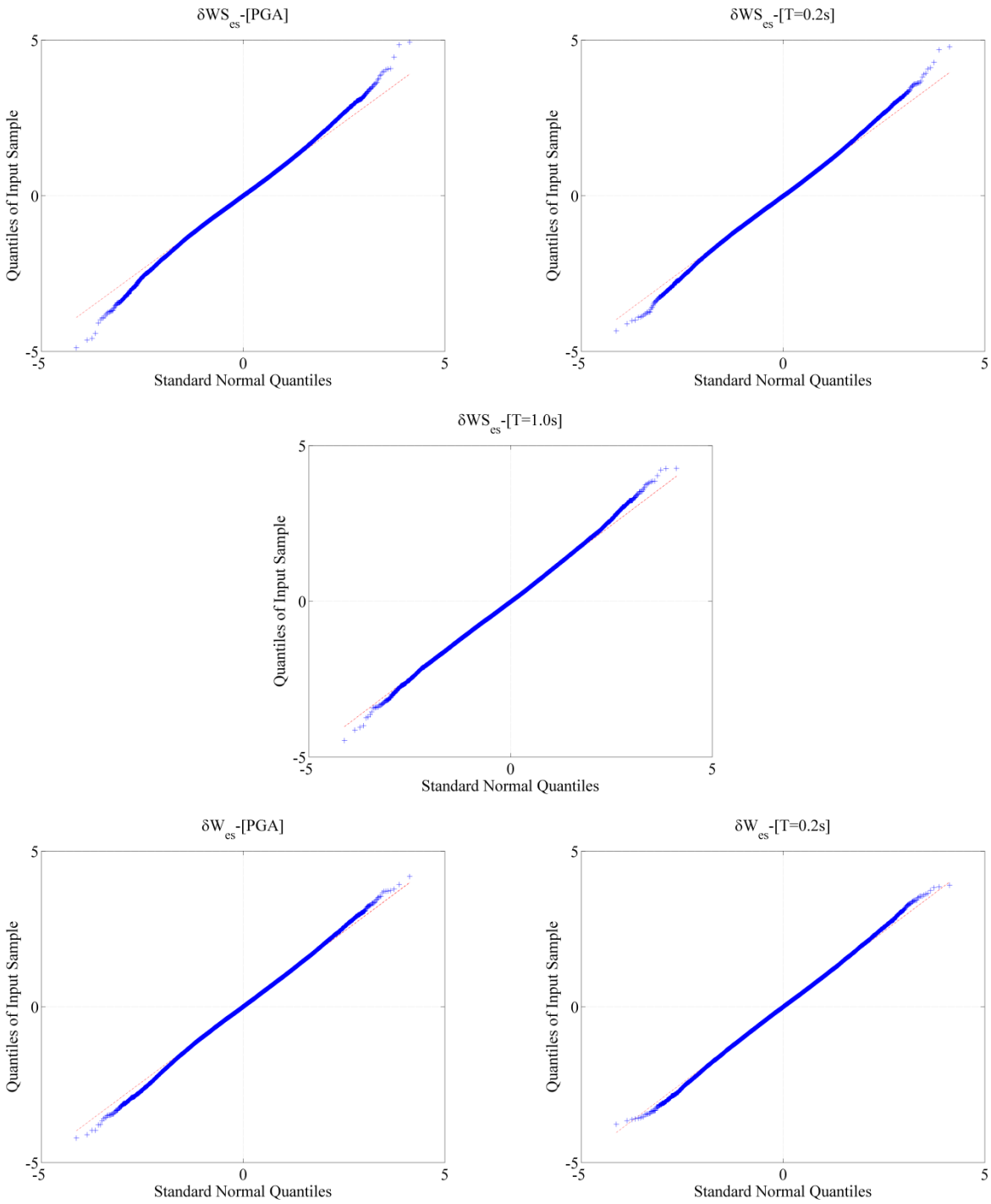
Fat tails

A set of observations is described as fat-tailed if these observations result in higher probabilities towards the tails of the distribution compared with the distribution that it is assumed to follow (e.g., normal distribution). Using a certain PDF (e.g., normal distribution) to predict probability of exceedance of very high values can result in major underestimation of the probabilities if the data is fat-tailed. Hence, care should be taken in estimating probabilities for fat-tailed data to prevent under-estimation of the risks associated with extreme events. For this

reason, it is important to investigate whether the different components of variability obtained from the GMPEs are fat tiled.

Figure D.12 presents the QQ plots at three different spectral periods (i.e., PGA, 0.2s and 1.0s) for δW_{es} , δWS_{es} and δB_e obtained from the regression analysis. There is some deviation from the log-normal distribution especially for δWS_{es}. The between event residual component (δB_e) is negatively skewed. The within-event site-corrected (δWS_{es}) seems to be fat tailed. The within-event residual component (δWS_{es}) is fat as well, but weaker than δWS_{es}.




δWS_es-[PGA]
δWS_es-[T=0.2s]
δWS_es-[T=1.0s]
δW_es-[PGA]
δW_es-[T=0.2s]
Quantiles of Input Sample
Standard Normal Quantiles
5
0
-5

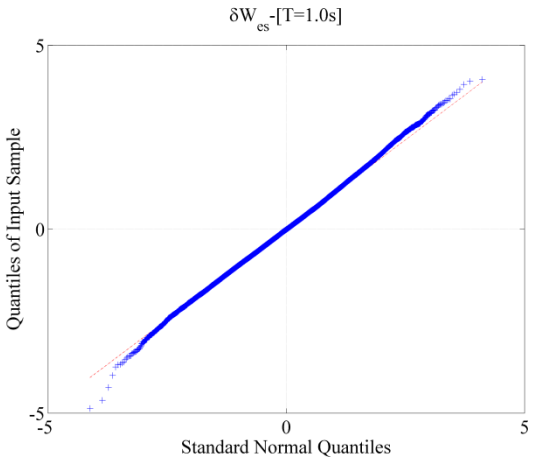


Figure D.12. The QQ plots at three different spectral periods (i.e., PGA, 0.2s and 1.0s) for δW_{es} , δWS_{es} and δB_e obtained from the regression analysis.

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.

Dawood, H. M., Rodriguez-Marek, A., Bayless, J., Goulet, C., and Thompson, E., (2014-In review). Processing the Kik-Net strong ground motion database and compilation of metadata for GMPE development, Earthquake Spectra

Dawood, H. M., and Rodriguez-Marek, A., (2014-In review). An empirical Ground Motion Prediction Equation for Active Crustal Earthquakes Using the Japanese KiK-net Database (Ergodic and Site-Specific Formulations), Earthquake Spectra

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.

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