Title

Ground Motion and Seismic Site Amplification in Central and Eastern North America and Regional Subduction Zones

Author
Parker, Grace Alexandra

Publication Date
2018

UNIVERSITY OF CALIFORNIA

Los Angeles

Ground Motion and Seismic Site Amplification in Central and Eastern North America and Regional Subduction Zones

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Civil Engineering

by

Grace Alexandra Parker

2018

ABSTRACT OF THE DISSERTATION

Ground Motion and Seismic Site Amplification

in Central and Eastern North America

and Regional Subduction Zones

by

Grace Alexandra Parker

Doctor of Philosophy in Civil Engineering

University of California, Los Angeles, 2018

Professor Jonathan Paul Stewart, Chair

Ground motion intensity measures are used to represent various components of earthquake shaking intensity and frequency content in the form of simple parameters; examples include peak ground acceleration, Arias intensity, and pseudo-spectral acceleration (PSA). Ground motion models (GMMs) are developed to predict these intensity measures as a function of earthquake source, wave propagation path, and local geotechnical site conditions. GMMs are formulated to capture the underlying physics of source processes, wave propagation, and site response, with individual model parameters set based on various combinations of empirical ground motion data analysis and physics-based ground motion simulations. The majority of GMMs are conditioned for hard rock reference sites, with shear wave velocity (V_S) = 3000 m/s, or with a time-averaged shear wave velocity in the upper 30 meters of the crust (V_{S30}) = 760 m/s. Additional site amplification models are necessary in order to estimate GMIMs for other site conditions, including weathered rock and

soil sites. As shear waves propagate vertically in the near-surface, the conservation of energy dictates that the wave amplitude must increase as the seismic velocity of the medium decreases. This amplification, or the so-called linear site effect, is usually parameterized using V_{S30}, and sometimes site fundamental frequency or depth to bedrock, if available.

This thesis has two parts, according to subject matter. The first part of this thesis, consisting of Chapters 2, 3, and 4, focuses on seismic site characterization and site amplification in central and eastern North America (CENA) in the context of the Next Generation Attenuation-East (NGA-East) project. Chapter 2 presents a hybrid geology-slope approach for V_{S30} estimation that utilized a new and expanded shear-wave velocity (V_S) measurement database for CENA. The proxy is conditioned on geologic category from newly considered large-scale geologic maps, the extent of Wisconsin glaciation, sedimentary basin structure, and 30 arc-sec topographic gradient. Nonglaciated sites were found to have a modest natural log dispersion of V_{S30} ($\sigma_{\ln V}$= 0.36) relative to glaciated sites ($\sigma_{\ln V}$= 0.66), indicating better predictability of V_{S30} for the former. These findings were used estimate the mean and standard deviation of V_{S30} for NGA-East recording stations when measurements were not available. Chapter 3 presents empirical linear site amplification models conditioned on time-averaged shear wave velocity in the upper 30 m (V_{S30}) for CENA, developed using a combination of least-squares, mixed effects, and Bayesian techniques. Site amplification is found to scale with V_{S30} for intermediate to stiff site conditions (V_{S30} > 300 m/s) in a weaker manner than for active tectonic regions. For stiff sites (> 800 m/s), I find differences in amplification for previously glaciated and non-glaciated regions, with non-glaciated sites having lower amplification. The models account for predictor uncertainty, which does not affect the median model, but decreases model dispersion. Lastly, Chapter 4 presents recommendations for modeling of ergodic site amplification in CENA, based primarily on results from the literature (including the model in Chapter 3), for application in the U.S. Geological Survey national seismic hazard maps. Previously, the maps have used site factors developed using data and simulations for active tectonic regions; however, results from NGA-East demonstrate different levels of site amplification in CENA. The recommended model has three terms, two of which describe linear site amplification: an empirically constrained V_{S30}-scaling term relative to a 760 m/s reference, and a simulation-based term to adjust site amplification from the 760 m/s to the CENA reference of $V_S = 3000$ m/s.

The second part of this thesis, consisting of Chapters 5 and 6, focuses on the development of a global GMM and site amplification model with regional adjustment factors for subduction zone regions as a part of the Next Generation Attenuation-Subduction (NGA-Sub) project. Chapter 5 presents global subduction zone GMMs for interface and intraslab events, with regionalized terms for Alaska, Cascadia, Central America. Mexico, Japan, South America, and Taiwan. The near-source saturation model, magnitude-dependent geometrical spreading, and magnitude-scaling break point are constrained using simulations and fault geometry, and the anelastic attenuation, magnitude scaling, and depth scaling terms are constrained empirically. The model is regionalized in the constant, anelastic attenuation, and depth-scaling terms, and the magnitude break-point. When applying the model to a region not considered in the study, we recommend using an appropriate range of epistemic uncertainty that captures regional variation. Chapter 6 presents a subduction-specific site amplification model, meant to be paired with the reference-rock GMM of Chapter 5. This site amplification model for subduction regions accounts for regional differences in V_{S30}-scaling, and re-calibrates a widely used nonlinear site term for active tectonic regions.

The dissertation of Grace Alexandra Parker is approved.

Scott Joseph Brandenberg

Yousef Bozorgnia

Robert Kayen

An Yin

Jonathan Paul Stewart, Committee Chair

University of California, Los Angeles

2018

In memory of

Thomas Campbell Parker

December 3, 1991 – October 2, 2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SUPPLEMENTARY MATERIALS

Table E1. Coefficients for the linear site amplification model relative to 3000m/s recommended in Chapter 4 for use in CENA.

Table E2. Coefficients for Next Generation Attenuation – Subduction interface ground motion model for (a) global subduction zones, (b) Alaska and the Aleutian Islands, (c) Cascadia, (d) Central America and Mexico, (e) Japan, (f) South America, and (g) Taiwan.

Table E3. Coefficients for Next Generation Attenuation – Subduction intraslab ground motion model for (a) global subduction zones, (b) Alaska and the Aleutian Islands, (c) Cascadia, (d) Central America and Mexico, (e) Japan, (f) South America, and (g) Taiwan.

ACKNOWLEDGEMENTS

The projects in this dissertation were sponsored by U.S. Geological Survey contracts G14AP00103, G16AP00005, and G16AP0018, and the PEER Center. Any opinions, findings, conclusions or recommendations expressed in this dissertation are those of the author and do not necessarily reflect those of the USGS or PEER. The UCLA Civil & Environmental Engineering Department also provided support in the form of teaching assistant positions and Graduate Division fellowships.

First and foremost, I would like to thank my PhD advisor Jonathan Stewart for his support, guidance and encouragement throughout my time in graduate school. I would also like to thank my PhD committee members Professor Scott Brandenberg, Professor Yousef Bozorgnia, Professor Robert Kayen, and Professor An Yin for their insightful feedback on this work.

I would like to thank my collaborators in the Next Generation Attenuation East Geotechnical Working Group Youssef Hashash, Ellen Rathje, Robert Darragh, Walter Silva, Joseph Harmon, and Okan Ilhan. I would like to thank the following people who helped substantially at various stages in NGA-East station and profile database development as presented in Chapter 2: Rajaa Al-Rayes, Caio Vilar, Christine Goulet, Robert Kayen, John Adams, Stephen Halchuk. Alan Yong, David Wald, Vince Quitoriano, and Eric Thompson. I would also like to thank the many individuals who have worked with the GWG and contributed feedback to database and model development presented in Chapters 2 and 3, including Justin Hollenback, Byungmin Kim, Lian Fan, Cheryl Moss, and Sissy Nikolaou. I sincerely thank Robb Moss for his guidance on the Bayesian analysis used in Chapter 3, and for sharing his MatLab codes.

I thank Gail Atkinson, Dave Boore, Youssef Hashash, Walter Silva, Robert Darragh, Mark Petersen, Arthur Frankel, Peter Powers, Behzad Hassani, Joseph Harmon, Okan Ilhan, Yousef

Bozorgnia, Linda Al Atik, and Christine Goulet for their contributions to the recommendations presented in Chapter 4 of this dissertation.

I would like to thank the researchers involved in the NGA-Subduction project database development for their work: Silvia Mazzoni, Tadahiro Kishida, Victor Contreras, Sean Ahdi, Robert Darragh, and Katie Wooddell, as well as other NGA-Subduction model developers for their helpful input: Norman Abrahamson, Yousef Bozorgnia, Ken Campbell, Brian Chiou and Nico Kuehn.

I want to thank my colleagues and friends during my time at UCLA: Peter Haproff, Sean Ahdi, Yi Tyan Tsai, Victor Contreras, Paolo Zimmaro, Maria Giovanna Durante, Pengfei Wang, Kioumars Afshari, Yang Yang, Ebuka Nweke, and Dong Youp Kwak. Lastly, I want to thank my family for their much appreciated love and support.

VITA

2018	Hired, Mendenhall Postdoctoral Fellow *United States Geological Survey*
2017	C. Phil, Civil Engineering *University of California, Los Angeles*
2016 - 2018	Teaching Assistant *University of California, Los Angeles*
2014	B.S., Applied Geophysics *University of California, Los Angeles*

SELECTED PUBLICATIONS

Parker, G.A., J.P. Stewart, Y.M.A. Hashash, E.M. Rathje, K.W. Campbell, and W.J. Silva (2018). Empirical Linear Seismic Site Amplification in Central and Eastern North America. *Earthq. Spectra*, *preprint.*

Abrahamson, N.A., Kuehn, N., Gulerce, Z., Gregor, N., Bozorgnia, Y., **Parker, G.A.**, Stewart, J.P., Chiou, B., Idriss, I.M., Campbell, K.W., and Youngs, R. (2018). Update of the BC Hydro Subduction Ground-Motion Model using the NGA- Subduction Dataset. *PEER Report 2018/02*. PEER, Berkeley, CA.

Parker G.A., Stewart J.P., Hassani B., Atkinson G.M., and Boore D.M. (2018). Preliminary NGA-Subduction Global Ground Motion Model with Regional Path Effects. *Proceedings of the 11th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Los Angeles, CA.

Hashash, Y.M.A., J. A. Harmon, O. Ilhan, **G.A. Parker**, and J.P. Stewart (2017). Recommendations for Ergodic Nonlinear Site Amplification in Central and Eastern North America. *PEER Report 2017/05*. PEER, Berkeley, CA.

Stewart, J.P., **G.A. Parker**, J.A. Harmon, G.M. Atkinson, D.M. Boore, R.B. Darragh, W.J. Silva, and Y.M.A. Hashash (2017). Expert Panel Recommendations for Ergodic Site Amplification in Central and Eastern North America. *PEER Report 2017/04*. PEER, Berkeley, CA.

Parker, G.A., J.A. Harmon, J.P. Stewart, Y.M. A. Hashash, A.R. Kottke, E.M. Rathje, W.J. Silva, and K.W. Campbell (2017). Proxy-Based VS30 Estimation in Central and Eastern North America. *Bull. Seismol. Soc. Am.*, 107(1), 117-131.

1 INTRODUCTION

Although full acceleration time series can be used to represent seismic demand, there is a significant need for simple representations of the demands applied by strong ground motion for a number of engineering applications, including probabilistic and deterministic seismic hazard analyses (PSHA and DSHA, respectively). This is achieved using so-called ground motion intensity measures (GMIMs) that represent various components of earthquake shaking intensity and frequency content as single parameters or vectors of parameters; examples include peak ground acceleration (PGA), Arias intensity (AI), and pseudo-spectral accelerations (PSA) for different oscillator periods.

Regional ground motion models (GMMs) are developed to predict these GMIMs as a function of earthquake source, wave propagation path, and local geotechnical site conditions (e.g. Boore et al., (2014) and Campbell and Bozorgnia (2014) for active tectonic regions such as the western United States (WUS), Japan and Taiwan; Yenier and Atkinson (2015) for central and eastern North America (CENA); Abrahamson et al., (2016) for global subduction zone earthquakes; and Skarlatoudis et al. (2013) for Greece). Earthquake source parameters used in GMMs can include moment magnitude (**M**), depth parameters such as hypocentral depth (d_{hyp}) and depth to the top of the rupture plane (Z_{tor}), and focal mechanism (e.g. normal, thrust or strike-slip). Path effects are parameterized using site-to-source distance, typically the closest distance between the site and the fault plane (R_{rup}), with terms that encompass geometric spreading and anelastic attenuation. The majority of GMMs are conditioned for hard rock reference sites with shear wave velocity (V_S) = 3000 m/s (Hashash et al., 2014) or the NEHRP B/C boundary condition with a time-averaged shear wave velocity in the upper 30 meters of the crust (V_{S30}) = 760 m/s (Frankel et al. 1996). Site amplification models used in combination with GMMs are therefore necessary to estimate GMIMs for conditions other than the reference such as weathered rock or soil sites. As shear waves propagate vertically in the near-surface, the conservation of energy dictates that the wave amplitude must increase as the seismic velocity of the medium decreases. This amplification, or the so-called linear site effect, is usually parameterized using V_{S30}, and sometimes site fundamental frequency (f_{peak}) or depth to bedrock, if available.

The first part of this dissertation focuses on work undertaken as part of the Next Generation Attenuation-East (NGA-East) Project organized by the Pacific Earthquake Engineering Research center (PEER, 2015). This work was undertaken with the NGA-East Geotechnical Working Group (Members: Youssef Hashash, Kenneth Campbell, Ellen Rathje, Walter Silva, and Jonathan Stewart), and included estimation of V_{S30} for sites in CENA in the absence of seismic velocity measurements, as well as empirical characterization of linear site effects parameterized on V_{S30}. Additionally, this includes a closely related activity in which expert panel recommendations were provided to the United States Geological Survey (USGS) on site amplification in CENA, for use in the USGS national seismic hazard model (Petersen et al., 2015).

The second and part of this dissertation describes work done as part of the Next Generation Attenuation – Subduction Project (NGA-Sub) organized by PEER (e.g., Kishida et al., 2017). This includes the development of semi-empirical ground motion models for global subduction zones for a 760m/s reference rock condition, as well as an accompanying empirical site amplification model. These models encompass the Cascadia region (Northern California, Oregon, Washington and Bristish Columbia), Alaska and the Aleutian Islands, Japan, Taiwan, Mexico, Central America, and South America, and are applicable to both interface and intraslab earthquakes.

Lastly, it should be noted that Chapter 2.0 of this document has been published as an article in the Bulletin of the Seismological Society of America (Parker et al. 2017), Chapter 3.0 is in press as an article preprint in Earthquake Spectra (Parker et al. 2019), and parts of Chapter 4.0 have been previously published as PEER Report 2017/04.

2 PROXY-BASED V_{S30} ESTIMATION IN CENTRAL AND EASTERN NORTH AMERICA

2.1 INTRODUCTION

The Next Generation Attenuation East (NGA-East) Project developed a series of semi-empirical ground motion models (GMMs) for predicting ground motion intensity measures in central and eastern North America (CENA). These GMMs include models for earthquake source and path effects, and are conditional on certain site conditions (PEER 2015). All such GMMs were required to provide predictions for a reference site condition consisting of a relatively uniform shear-wave velocity (V_S) profile of 3000 m/s near the ground surface (Hashash et al., 2014). For softer site conditions, various site factors can be used that are based at least in part on the time-averaged shear wave velocity in the upper 30 m of the site (V_{S30}) (PEER 2015; Parker et al., 2016; Harmon et al., 2016). Some NGA-East GMMs do not provide a recommended site factor, but nonetheless utilize V_{S30} in connection with the definition of an alternate, softer, reference site condition of V_{S30} = 760 m/s (which is the reference value for USGS national seismic hazard maps; Petersen et al., 2015) (Yenier and Atkinson 2015; PEER 2015). A challenge faced in the development and application of these GMMs and associated site amplification models is the lack of measured V_{S30} values at a large majority ground motion recording stations.

When no measurement of V_{S30} is available, which is the case for 94% of recording sites in the NGA-East database flatfile (Goulet et al., 2014), it becomes necessary to provide an estimate. Although it is possible to estimate site information from interpretation of recordings (Kim et al., 2016; Hassani and Atkinson, 2016a), such estimates are currently possible for a relatively small number of stations due in part to requirements of multiple recordings at the same site. Moreover, a consensus has not yet emerged on the appropriateness of estimating site parameters from attributes of recordings, when the performance of the resulting GMMs are then judged against those same recordings. For these reasons, it is often necessary to estimate the logarithmic mean and standard deviation of V_{S30} via proxy methods. In the current NGA-East flatfile (Goulet et al., 2014), the considered proxies were associated with small-scale (1:2,000,000 to 1:5,000,000) geologic map categories specific to CENA (Kottke et al., 2012; hereafter Kea12), a hybrid slope-geology proxy also derived from small-scale geologic maps for CENA (Thompson and Silva, 2013; hereafter TS13); geomorphology-based terrain categories related to V_{S30} based on data from California (Yong et al., 2012; hereafter Yea12), and a topographic gradient-V_{S30} relation developed using limited data from Memphis and Australia (Wald and Allen, 2007; hereafter WA07).

The present work was motivated by our general discomfort with the adequacy of the available proxies used to assign V_{S30} values in the development of the preliminary station database presented in Goulet et al. (2014), which was associated with the aforementioned issues of map scale and the 'borrowing' of proxies from other regions. We were also concerned with the size of the V_S data set used to evaluate proxy performance of that prior work, which was based only on measurements from ground motion stations (34 sites). With regard to geology-based proxies, we anticipate that geologic conditions identified from larger-scale maps will be more reliable, and that consideration of Wisconsin glaciation and the presence of sites in basins may influence V_{S30}. We describe below a database of sites in CENA with measured V_S, including sites with and without ground motion recording stations. We compiled geologic and terrain-based information for V_S measurement sites and query the data to develop proxy-based V_{S30} relationships. The V_{S30} assignment protocols are then updated in consideration of these results and an updated station database for NGA-East is provided as an electronic supplement (available as Table S1 in the electronic supplement to Parker et al. 2017).

2.2 CENA V_{S30} DATABASE FROM MEASUREMENTS

We have compiled a database of 2755 V_{S30} values from seismic velocity measurements in CENA. We consider sites having both V_S profiles as a function of depth, and sites with only a reported V_{S30} value from measurements. The data are derived from 82 source documents including research reports, microzonation studies, and professional engineering reports for project sites (including nuclear power plants). The present database updates the earlier CENA profile database of Kea12, which had 1930 entries derived from seven source documents, and includes many V_S profiles compiled for use in Hashash et al. (2014). A variety of measurement methods were used in developing these profiles, including downhole logging, suspension logging, and surface wave techniques. In many cases, we lacked the level of documentation required to render opinions on the relative reliability of data from different providers, and have not attempted to screen the data on this basis. Table S2, available in the electronic supplement to Parker et al. (2017), presents summary information on each entry in the database. Figure 1 shows the spatial distribution of measurement sites along with strong motion sites in the NGA-East database flatfile.

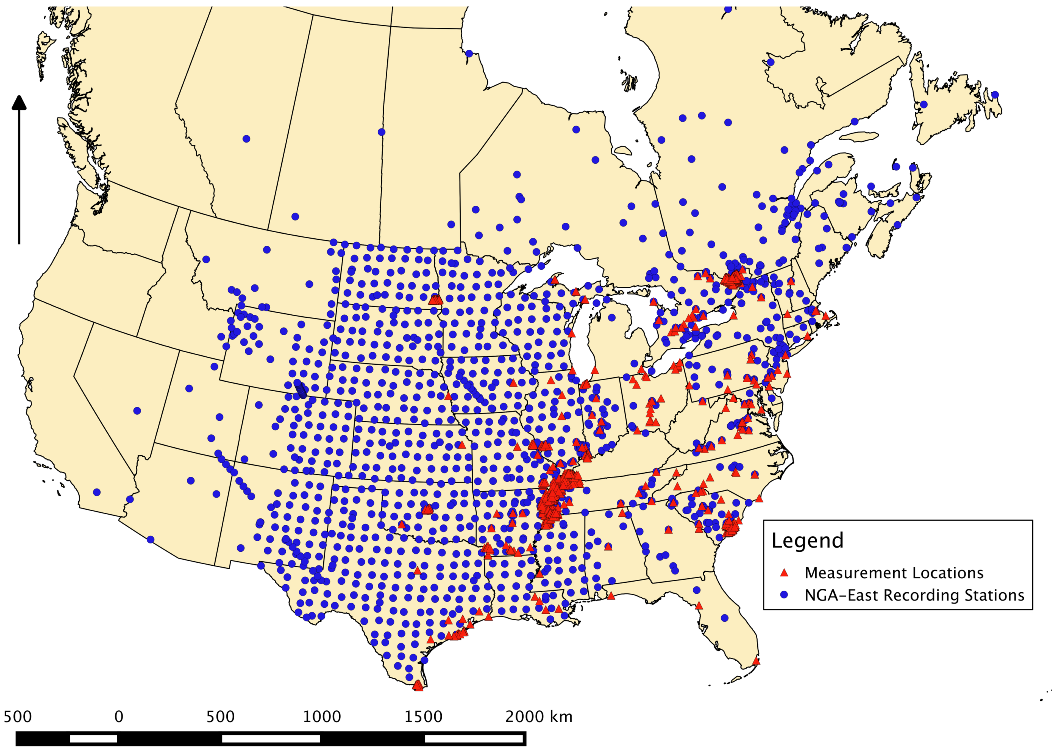

Figure 2.1 Locations of V_s measurements in CENA included in the measurement database.

For each site in the database, we report location (latitude, longitude), measurement type, V_{S30}, and the data source (Table S2, available in the electronic supplement to Parker et al. 2017). Figure 2 is a histogram of measured V_{S30} values from the database and shows that a plurality of the data sample low velocity sites (V_{S30} < 450 m/s). Concentrations of data are present in Ottawa (Canada), Charleston South Carolina, and Mississippi Embayment (1230, 326, and 535 measurements, respectively). In the earlier version of the database (Kea12), nearly all of the measurements above 450 m/s were from Ottawa (Crow et al. 2007). The present version has 583 profiles with V_{S30} > 450 m/s, 213 of which are outside of Ottawa. Due to the spatial nonuniformity of the dataset, we have considered the possibility of regional bias in V_{S30} values from areas with clustered profiles, as described below.

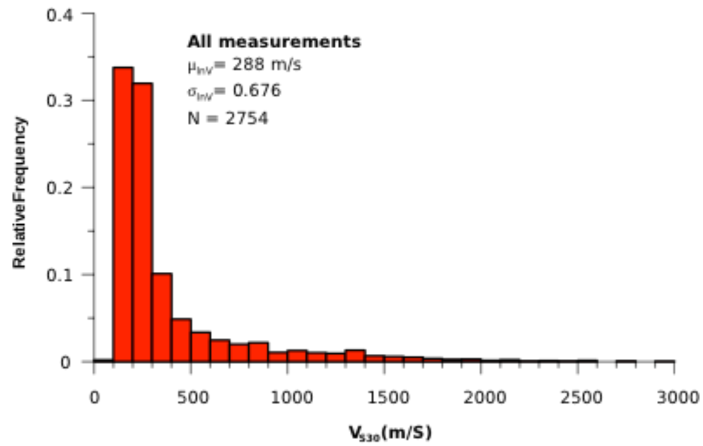

Figure 2.2 Histogram of all measured V_{S30} values in the CENA *VS* measurement database.

2.3 GEOLOGY- AND GEOMORPHOLOGY-BASED PROXIES

Information compiled for measurement sites as part of the present work includes geologic site conditions as indicated from geologic maps at larger scales than used for this application previously, an indicator of whether the profile is within the region of CENA that was overlain by the Wisconsin ice sheet during the last glaciation, indicators of whether the profiles are in various mapped basins, and indicators of whether or not the profile is in an area of data concentration (Ottawa, Charleston, or the Mississippi Embayment). In addition, geomorphology-related parameters were compiled from digital elevation models (DEMs) at 3- and 30 arc-s resolution. The 30 arc-s DEM consists of raster files from USGS (2011); parameters compiled include geomorphic terrain categories based on procedures in Iwahashi and Pike (2007) and topographic gradient in the manner used by WA07. The 3 arc-s DEM is drawn from the NHDPlusV2 dataset, a geospatial, hydrologic framework dataset developed with support from the Environmental Protection Agency Office of Water and the USGS. The data is available only for the contiguous U.S., and is corrected for the canopy effect. We extract topographic gradient from this DEM.

Geologic conditions were taken from geologic maps ranging in scale from 1:24,000 to 1:500,000 for locations in the United States and Canada, and from Crow et al. (2007) metadata files for locations in Ottawa, Canada. The map scale from these sources is much larger than has been used previously for proxy development (i.e., Kea12; TS13), which used 1:5,000,000 and 1:2,000,000 scale maps for the United States (Soller et al. 2009; Fullerton et al. 2003) and a 1:5,000,000 scale map for Canada (Fulton 1996). The larger scale of the maps used in the present work is expected to reduce, although not to eliminate, potential surface geology misclassifications.

Geologic maps used in this study were primarily sourced from the United States Geological Survey (USGS) national geologic map database (NGMDB). For areas not covered by the USGS NGMDB, digital state geologic maps compiled by the USGS Division of Mineral Resources (DMR) were used. For Canada, we adopt geologic classifications from Crow et al. (2007) for Ottawa; elsewhere in Ontario we utilize the Ontario Geological Survey spatial dataset 14 (Ontario Geological Survey, 2000); and in Québec we use an online interactive map from the Sysème d'information géominière of Québec (SIGÉOM). Table S2, available in the electronic supplement to Parker et al. (2017), provides specific map sources for each measurement site. The geologic maps show the extent of and contacts between rock and sedimentary units, and include structural features and measurements. Site-specific information compiled in the V_{S30} database (Table S2, available in the electronic supplement to Parker et al. 2017) includes descriptions of geologic age, geologic group, formation, and unit names where applicable, and lithologic information.

The extent of the Wisconsin glaciation was taken from Reed and Bush (2005). Any measured profile north of the extent of glaciation was given a flag of 1. Other sites to the south of the glacial limit have a flag of 0. This information was compiled because we expected glaciation to impact geologic conditions and seismic velocities in a number of ways: (1) potential overconsolidation of sediments, (2) removal of soil and weathered rock due to glacial scouring, and (3) the deposition of glacial and post-glacial sediments. We also considered the use of earlier, more extensive glacial limits (Reed and Bush 2005), but these limits affected a small number of additional sites and did not improve the predictive ability of the model.

The locations of known sedimentary basins of any age were taken from the electronic supplement of Coleman and Cahan (2012) and are listed in Table 2.1. The CENA V_{S30} database includes a column for basin name, where applicable. This information was compiled to enable studies of possible basin-specific biases of seismic velocities.

Table 2.1. Sedimentary basins, as defined by Coleman and Cahan (2012), containing measured V_S measurements in CENA.

Basin Name	Number of Measurements
Appalachian Basin	34
Arkoma Basin - Ouachita Thrust Belt	2
Buried Newark Group Basins	8
Exposed Newark Group Basins	3
Forest City Basin	1
Fort Worth Basin	5
Great Smoky Mountains Rift Basin	2
Gulf of Mexico Basin	49
Illinois Basin	70
Michigan Basin	10
Midcontinent Rift	2
Mississippi Embayment	175
Reelfoot Rift	17
Rough Creek Graben	12
West Atlantic Basin	87

2.4 PROXY DEVELOPMENT METHOD

2.4.1 Grouping

The 2755 locations with measurement-based V_{S30} values were grouped by attributes to identify features that produce distinct mean V_{S30} values (taken as the exponent of the natural log mean, and denoted μ_{lnV}, which has units of m/s), standard deviations (σ_{lnV}, dimensionless), and trends with 30 arc-s topographic gradient. We use the natural log of velocities because the data distribution is visually better approximated by a log-normal distribution than other distributions such as normal or beta. Attributes considered in the grouping process include geologic age, lithology, glaciation history, and location relative to known basins. Because three regions (Ottawa, Charleston, and Mississippi Embayment) have large data concentrations, we investigated statistics for these regions separately from those of otherwise similar geology to identify potentially distinct regional features.

Age was first examined by geologic era and then broken down into further subdivisions by geologic period and epoch when possible. Cenozoic was divided into Quaternary and Tertiary periods, and Quaternary was further divided into epochs: Holocene, Pleistocene, or undivided when mapped as such (undivided indicates that the age is known to be Quaternary, but the epoch is unknown).

Well populated age bins were further broken down by lithology. This was considered for the Holocene, Pleistocene, Quaternary undivided, and Paleozoic groups. Holocene lithology bins were initially investigated for all well-populated categories (e.g. alluvial, deltaic, estuarine, eolian, marine, lacustrine, fluvial, and organic deposits); many lithology-based bins were then combined on the basis of similar statistical attributes (i.e., μ_{lnV}, σ_{lnV}, and trend with topographic gradient) when possible.

The presence of Wisconsin glaciation (Reed and Bush, 2005) was investigated separately from age and lithology. Sites flagged as glaciated include locations with Holocene geology; in such cases the Holocene sediments themselves can be a product of glacial runoff, but are not subject to the overconsolidation effects of glacial unloading. By separating these sites from non-glaciated Holocene sites with similar lithology, we are in essence investigating whether the glacially derived sediments have unique features and possible impacts on V_{S30} of older, potentially over-consolidated layers at depth. As we look at groups that were previously glaciated compared to those that were not, we observe a significant increase in σ_{lnV}. This divergence of dispersion values was a motivating factor for considering glaciation in the formation of proxy groups along with the mean V_{S30}.

The location of a site in one of the sedimentary basins listed in Table 2.1 was examined to evaluate whether V_{S30} statistics for particular basin structures are distinct from otherwise similar conditions (age, lithology, glaciation).

With the many factors considered in the proxy development process, we required a systematic approach for deciding when groups or bins of V_{S30} values were statistically distinct. For this purpose, we used two types of F-tests (Snedecor and Cochran, 1989), which compare the statistical performance of submodels with that of a full model for a common data set. For example, if a full model applies to Holocene sediments, a pair of submodels could comprise glaciated and non-glaciated groups. One type of F-test uses the residual sum of squares (based on misfits from median model predictions) for the submodels (RSS_1 and RSS_2) and the full model (RSS_f). The relative performance of submodels and the full model is quantified using the difference $RSS_f -$ (RSS_1+RSS_2). If this difference is "small," then the submodels and full model fit the data about equally well, suggesting that data segregation in submodel groups is not justified. For normally distributed sets of residuals, this is interpreted using the one-way analysis of variance (ANOVA) F-statistic, which can be written as (adapted from Snedecor and Cochran, 1989; specific form used here is from Stewart et al., 2003):

$$F_1 = \frac{\left(RSS_f - (RSS_1 + RSS_2)\right) / \left((df_1 + df_2) - df_f\right)}{\hat{\sigma}^2} \tag{2.1}$$

where df_i refers to the degree of freedom for model or submodel i (one if the model consists of a simple mean, two if the model includes a slope gradient term), and

$$\hat{\sigma}^2 = \frac{RSS_1 + RSS_2}{N_f - (df_1 + df_2)} \tag{2.2}$$

where N_f is the number of data points in the full model. This F-statistic can be compared with the F distribution to evaluate significance level (p) for the test. Large values of p (> 0.05) are often taken to imply that the submodels are not distinct. One shortcoming of the F_1-statistic is that it does not effectively distinguish data groups having similar means but differing dispersion. For this reason, we also compute a second F-statistic (Snedecor and Cochran, 1989):

$$F_2 = \frac{\sigma_1^2}{\sigma_2^2} \tag{2.3}$$

for the null hypothesis that two normal populations from which samples are drawn have the same variance. As before, this statistic is compared with the F distribution and a p value is computed, which is interpreted as before (values < 0.05 indicate the sub-groups have distinct variances). If either F_1 or F_2 have p values < 0.05, the sub-groups are considered distinct.

To meet the requirement of normally populated data populations, both F-tests were performed on residuals in natural logarithmic units because V_{S30} has generally been found to be approximately log-normal (e.g., Wills and Clahan, 2006). Within each age category, alternative strategies for binning V_{S30} data were tested, with the resulting distinct sub-groups listed in Table 2.2. In one case (Groups 14 and 16), one of the p-values is 0.06, thus not strictly meeting the < 0.05 criteria, but are retained as distinct based on judgment driven by the different geological conditions and different means (the Group 14 mean has high uncertainty due to sparse data). At the bottom of Table 2.2, we also provide examples of F-test results for submodel groups that were not distinct and hence are not reflected in our recommended V_{S30} estimation procedure. Details regarding the selected groupings and the interpretation of test results are given in the *Results* section below.

Table 2.2. Results of F-tests performed on binned V_{S30} groups (Eqs. 2.1-2.3).

Groups*	F_1	p_1	F_2	p_2	Distinct (0) or Non-Distinct (1) Groups
1+2	4.6	<0.05	0.61	<0.05	0
1+3	6.7	<0.05	0.12	<0.05	0
1+4	54.6	<0.05	0.10	<0.05	0
2+3	0.99	0.51	0.19	<0.05	0
2+4	24.6	<0.05	0.16	<0.05	0
3+4	8.4	<0.05	0.88	0.48	0
5+6	473	<0.05	0.39	< 0.05	0
5+7	30.5	<0.05	0.30	< 0.05	0
5+8	73.2	< 0.05	0.30	< 0.05	0
6+7	55.4	< 0.05	0.77	0.24	0
6+8	33.1	< 0.05	0.77	0.25	0
7+8	2.18	< 0.05	1.00	0.98	0
9 +10	1.61	<0.05	2.05	< 0.05	0
9 +11	35.5	<0.05	1.97	< 0.05	0
10 + 11	43.6	<0.05	0.95	0.81	0
14+15	1.08	0.37	0.13	0.06	1
14+16	3.4	< 0.05	0.08	< 0.05	1
15+16	11.2	<0.05	0.64	< 0.05	0
p€† glac. + p€ non-Glac.	0.0017	1.0	1.20	0.68	1
P‡ shale + P limestone	0.021	1.0	0.86	0.43	1

* See Table 3 for definition of groups
† Precambrian
‡ Paleozoic

2.4.2 Trends with topographic gradient

Within the various groups identified in the previous section, trends of V_{S30} with 30 arc-s topographic slope gradients (s) were investigated using semi-log, and log-log regressions:

$$ln(V_{S30}) = c_0 + c_1 s \quad (2.4)$$

$$ln(V_{S30}) = c_2 + c_3 ln(s) \quad (2.5)$$

where V_{S30} is in m/s and slope gradient s is expressed as a decimal (meters per meter). Expressions similar to Eq. (2.5) have been used by Thompson et al. (2014), among others.

Values of either c_1 or c_3 having zero outside the range of their 95% confidence intervals indicate statistically significant effects of gradient. When the trend with gradient is significant, either the semi-log or log-log model is selected based on visual inspection of the fit and which of the models produces lower standard deviation of residuals. In this case, the V_{S30} log-mean estimate is calculated using either Eq. (2.4) or (2.5) and the standard deviation of the fit residuals is taken as σ_{lnV}. When the trend with gradient is not significant, a gradient-independent mean is selected (μ_{lnV}).

2.5 RESULTS

Table 2.3 summarizes the proposed hybrid geology-slope proxy procedure for V_{S30} estimation. Rows in Table 2.3 are differentiated first by geologic age and the flag for Wisconsin glaciation. Within age and glaciation groups additional sub-groups are recommended in some cases based on lithology, location, or presence within certain basins. For each sub-group, either a natural log mean and standard deviation are given or a gradient-dependent relation is given for the mean along with σ_{lnV}. Aside from geologic age, the presence (or not) of Wisconsin glaciation has the strongest effect on V_{S30} distributions, generally increasing both means and standard deviations relative to otherwise similar non-glaciated conditions. Our interpretation of the physical explanations for these trends is provided in the *Proxy Performance* section below. Additionally, V_{S30} values calculated from profiles in Ottawa differ significantly from the rest of the data across all age groups. When possible, data from Ottawa within an age group were used to create a separate recommended V_{S30} value for that region. Profiles in the Charleston and Mississippi Embayment regions were analyzed separately for comparison against the remaining data set. However, they did not differ in a statistically significant manner from otherwise similar sites, and are not considered as a separate category. Results and recommendations are described in more detail below for geologic age groups.

Table 2.3. Summary of proposed V_{S30} estimation procedures based on large-scale geologic maps, Wisconsin glaciation, location of site in a basin, and topographic gradient*

Category							Group Moments		Gradient Relationship				
Group	Era	Period	Epoch	Wisconsin Glaciation?	Other Criteria	N	μ_{lnV} (m/s)	σ_{lnV}	Semi-Log		Log-Log		σ_{lnV}
									c_0	c_1	c_2	c_3	
1	C	Q	H	No	Alluvium, fluvial, & deltaic	308	210	0.23					
2	C	Q	H	No	All other lithology	183	221	0.29					
3	C	Q	H	Yes	In Ottawa, Canada	981	232	0.67	5.38	9.30			0.67
4	C	Q	H	Yes	Not in Ottawa, Canada	51	308	0.72			7.47	0.295	0.67
5	C	Q	Pli	No		284	271	0.36	5.47	31.4			0.31
6	C	Q	Pli	Yes	Till in Ottawa, Canada	104	777	0.57	6.51	22.4			056
7	C	Q	Pli	Yes	Other in Ottawa, Canada	60	377	0.65			7.20	0.22	0.63
8	C	Q	Pli	Yes	Not in Ottawa, Canada	63	448	0.65					
9	C	Q	U	No	Not in sedimentary basin	154	296	0.43			6.21	0.096	0.41
10	C	Q	U	No	In sedimentary basin	151	280	0.29					
11	C	Q	U	Yes	Not in sedimentary basin†	66	209	0.31	5.28	24.7			0.29
12	C	T				111	315	0.31			6.07	0.059	0.30
13	M					20	822	0.68					
14	P				In the Illinois Basin‡	5	513	0.23					
15	P			No	Not in the Illinois Basin	96	684	0.61					
16	P			Yes	Not in the Illinois Basin	76	972	0.77					
17	p€					37	699	0.85					
18	p€				Site visit- hard rock confirmation		2000						

*C for Cenozoic; Q for Quaternary; H for Holocene; Pli for Pleistocene; U for undivided; T for Tertiary; M for Mesozoic; P for Paleozoic, p€ for Precambrian.

† Can be applied for within-basin sites with increased epistemic uncertainty (category unpopulated).

‡ Because of the small population size, the mean and standard deviation carry a larger degree of epistemic uncertainty than for other group

2.5.1 Holocene

Of 2,755 V_{S30} values from measurement, 1,523 are classified as Holocene. The Holocene sites were subdivided into previously glaciated and not previously glaciated bins. The Holocene non-glaciated bin was subdivided further on the basis of lithology, with one group consisting of alluvial, fluvial, and deltaic deposits, and a second group consisting of all other lithologies (Groups 1 and 2, respectively). Figures 2.3a-b show that the V_{S30} histograms for these groups have non-similar means and standard deviations, which are confirmed as statistically distinct by the F-test results in Table 2.2. Figures 2.4a-b show that these groups exhibit no trend with gradient, so the recommended V_{S30} for each was taken as μ_{lnV} of the binned V_{S30} values.

The Holocene previously glaciated bin is subdivided based on location (Ottawa vs. other locations – Groups 3 and 4, respectively). Histograms for Groups 3 and 4 (Figures 2.3c-d) show much higher dispersion than those for Groups 1 and 2 (Figures 2.3a-b) and slower velocities in Ottawa (Figure 2.3c) than non-Ottawa locations (Figure 2.3d). Both Groups 3 and 4 have a statistically significant trend with gradient (Figures 2.4c-d). The high dispersion in glaciated groups is a persistent feature of the data, the interpretation of which is given in the *Proxy Performance* section below.

Factors found to not be impactful for the Holocene age group included the presence of sites in sedimentary basins (Table 2.1) and location within Charleston or the Mississippi Embayment. These factors are considered for all other age groups as well, and are only commented on below when bins are well populated and a dependence was identified.

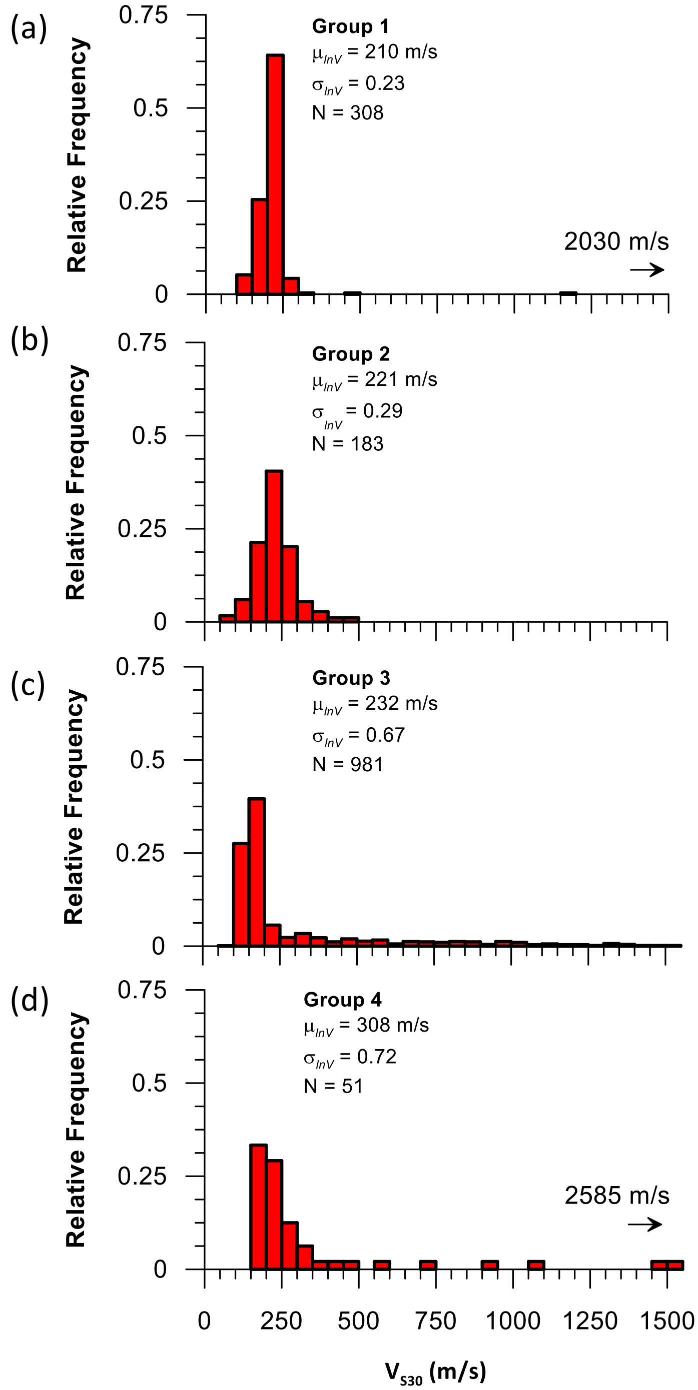

Figure 2.3. Histograms of V_{S30} values for Groups 1 through 4 (see Table 2.3).

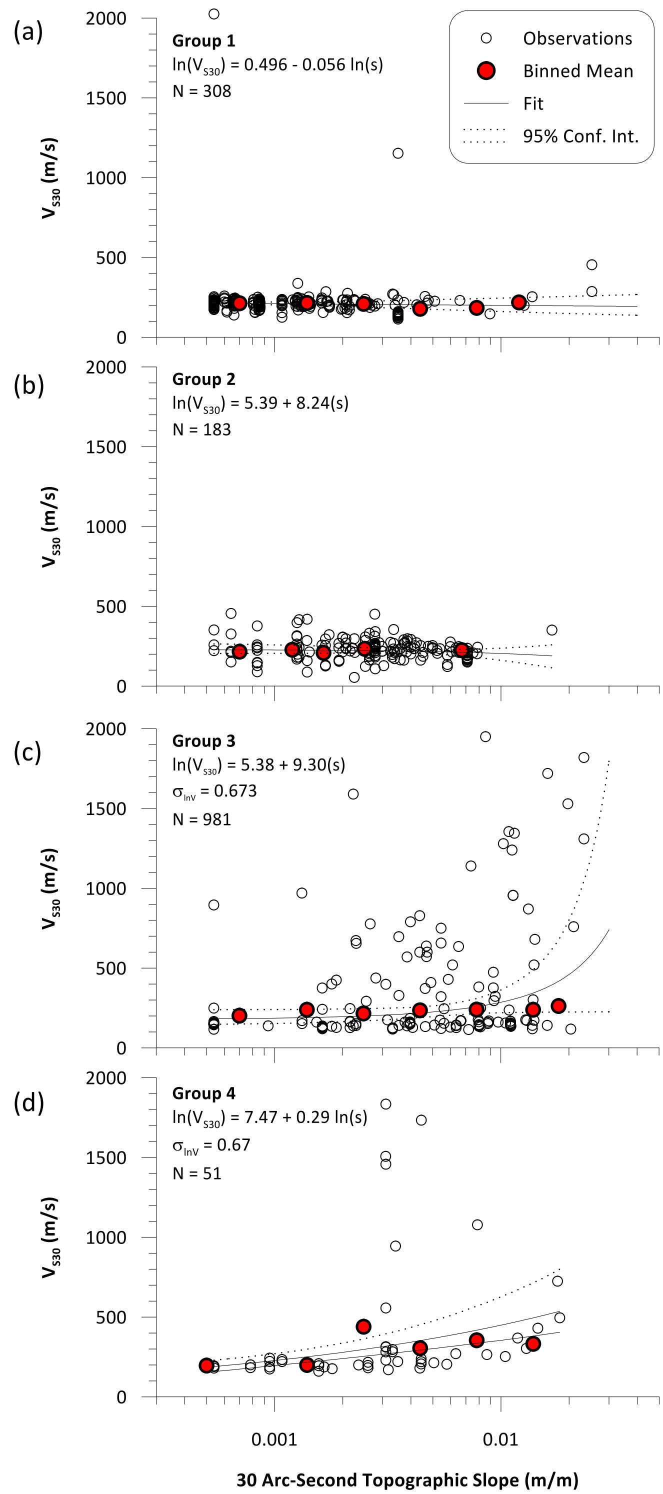

Figure 2.4 V_{S30} as a function of 30 arc-s topographic gradient for Groups 1 through 4 (see Table 2.3)

2.5.2 Pleistocene

The Pleistocene age bin contains 511 V_{S30} values from measurements, and is subdivided into previously non-glaciated locations (Group 5), locations in Ottawa (all previously glaciated, Groups 6-7), and glaciated locations outside of Ottawa (Group 8). The measurements in Ottawa were further divided by lithology, with Group 6 being for measurements on till, and Group 7 encompassing all other lithologies. Figures 2.5a-d show histograms for these groups, with the glaciated groups clearly having higher dispersions. Figures 2.6a-d show V_{S30} trends with gradient, which are not significant for Group 8, but are for the other three Pleistocene groups. This relationship is described using Eq. (2.4) for Groups 5 and 6, and Eq. (2.5) for Group 7.

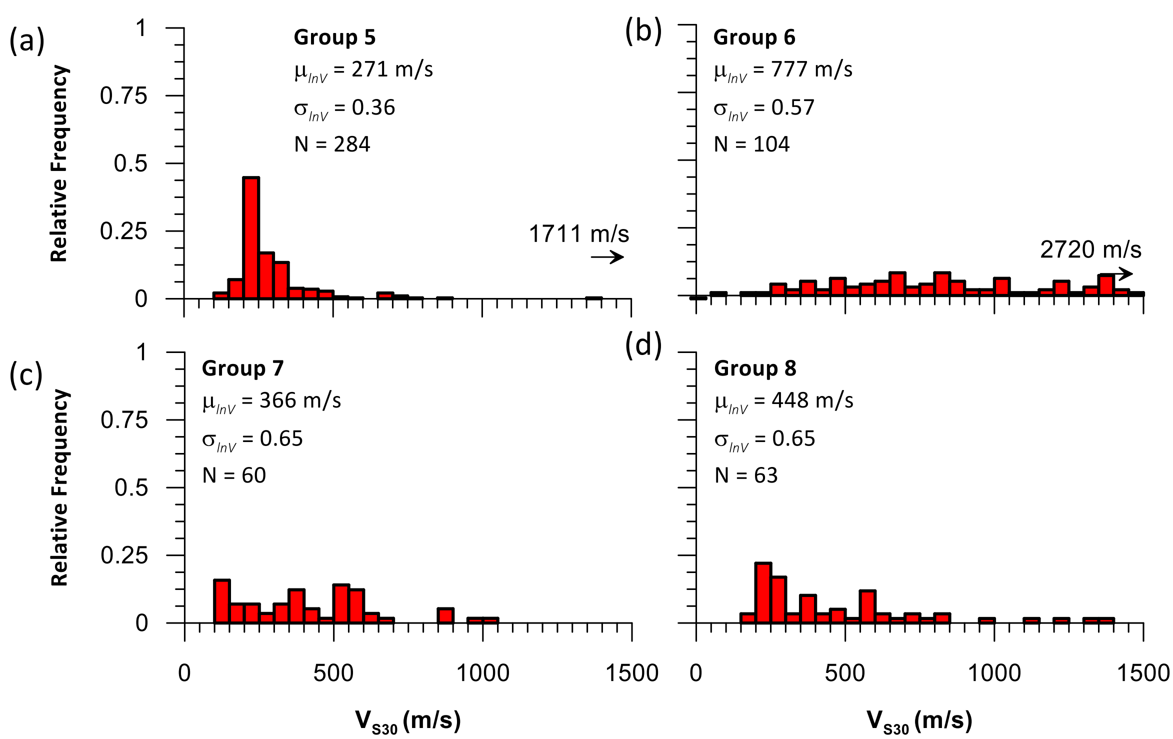

Figure 2.5 Histograms of V_{S30} values for Groups 5 through 8 (see Table 2.3).

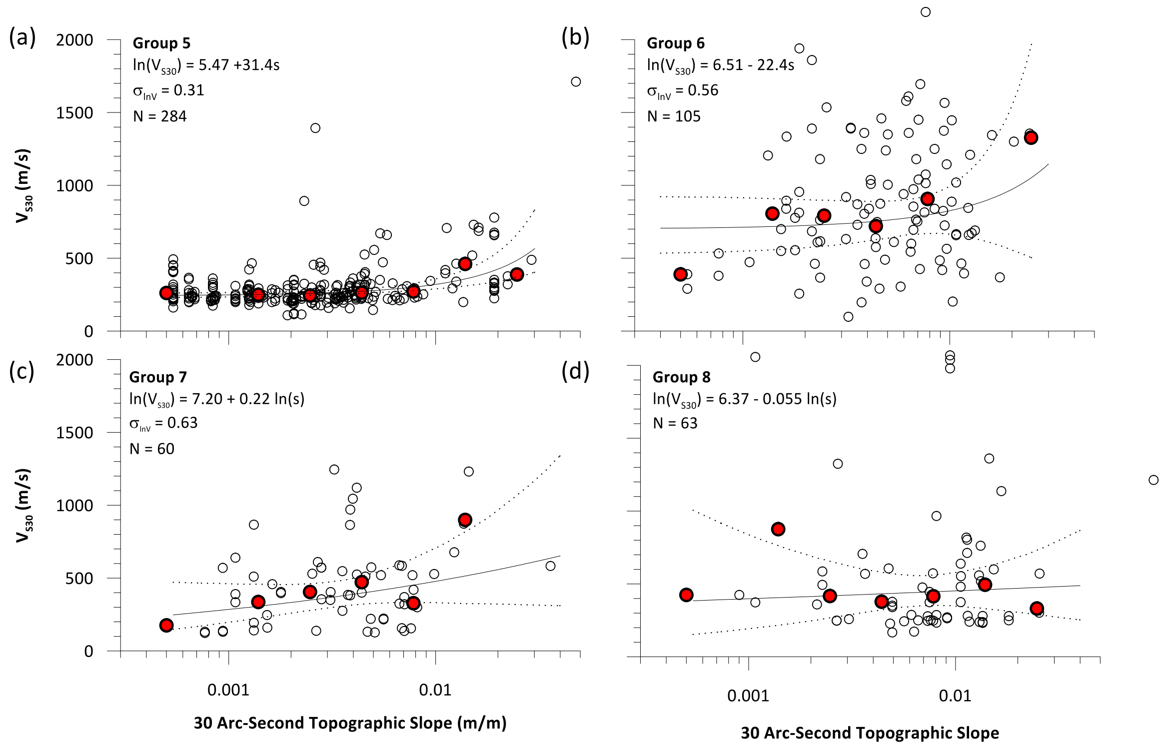

Figure 2.6. V_{S30} values as a function of 30 arc-s topographic gradient for Groups 5 through 8 (see Table 2.3). Legend from Figure 2.4 applies.

2.5.3 Quaternary undivided

The Quaternary undivided (QU) age bin contains 371 V_{S30} values. This age bin was subdivided into groups based on previous glaciation, and whether or not the profile was measured in a mapped basin (Table 2.1). Of the four possible bins, one is not populated (previously glaciated and in-basin). Figures 2.7a, c and e show V_{S30} histograms for Groups 9 – 11, and Figures 2.7b, d and f show the gradient relationships for the same groups. Group 10 did not display a significant gradient relationship, whereas the gradient relationships in Groups 9 and 11 were fit using Eq. (2.5) and Eq. (2.4), respectively. For application purposes we recommend using the Group 11 estimates for previously glaciated basin sites.

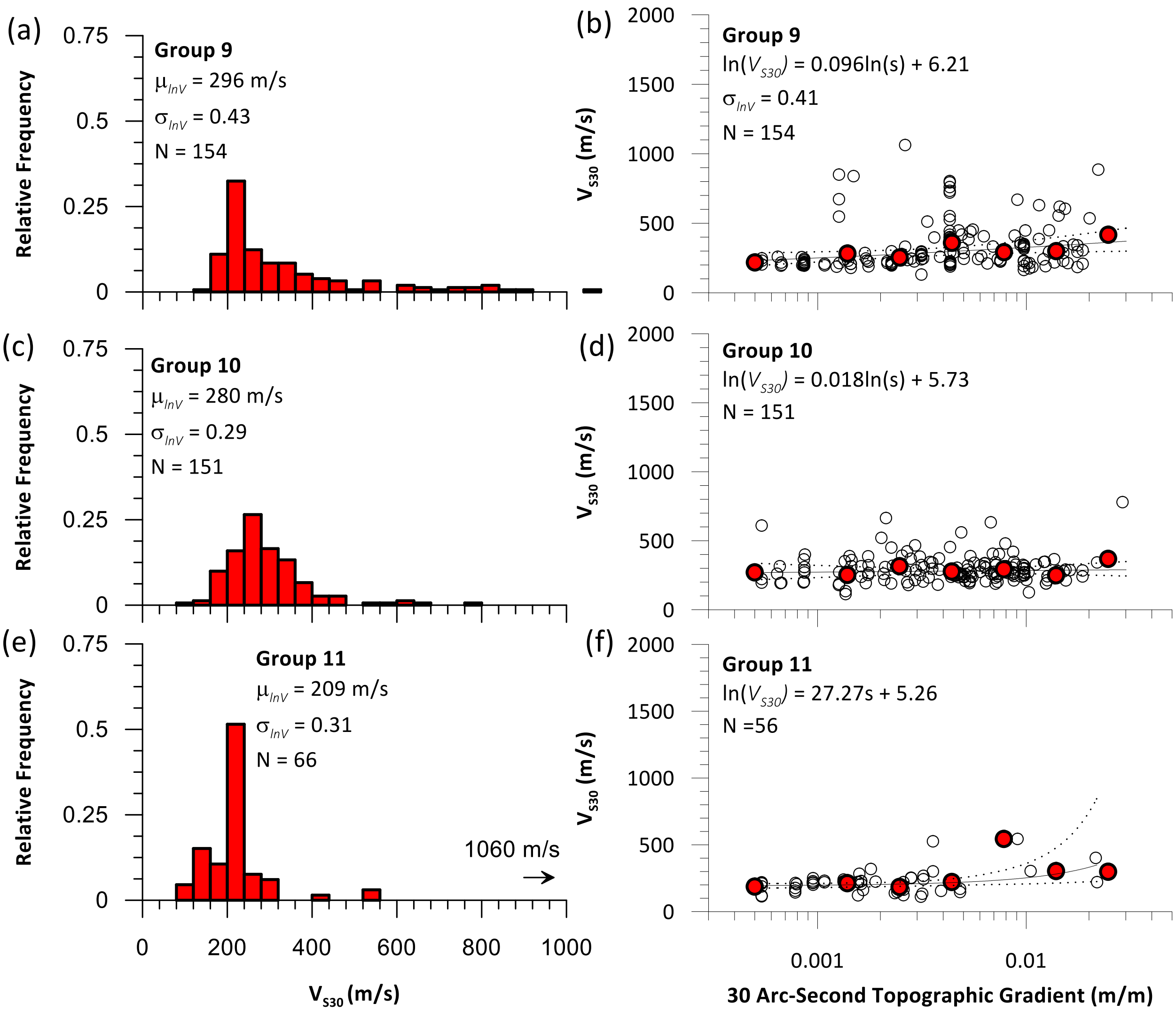

Figure 2.7. (a,c,e) Histogram of V_{S30} values for Groups 9 through 11 (Table 2.3), and (b,d,f) V_{S30}values as a function of 30 arc-s gradient for Groups 9 through 11 (Table 2.3), with binned means shown as filled circles. Legend from Figure 2.4 applies.

2.5.4 Tertiary and Mesozoic

The Tertiary age group (Group 12) contains 111 V_{S30} values, and is not subdivided further because sub-groups would be too sparsely populated. Figure 2.8a shows the V_{S30} histogram for Group 12, and Figure 2.8b shows the gradient-dependence, which is fit using the log-log relation (Eq. 2.5). The CENA category statistics for Tertiary ($\square_{lnV}$ = 315 m/s and $\square_{lnV}$ = 0.31) indicate slightly lower velocities than multiple Tertiary categories in California (Wills and Clahan, 2006), and similar dispersion levels to those in California. The Tertiary sites in our CENA V_{S30} from database are not glaciated, and the modest dispersion in this case appears to result from deep weathering profiles that avoids the presence of thin soft layers over firm deeper layers, which accentuates data variability.

Figure 2.9a shows a histogram for the Mesozoic age group (Group 13), which contains only 20 V_{S30} values from measurements. The CENA statistics for Mesozoic μ_{lnV} = 822 m/s and σ_{lnV} = 0.68) indicate faster velocities with more dispersion than Mesozoic sites in active tectonic regions (e.g., the Franciscan complex in California has μ_{lnV} = 710 m/s and σ_{lnV}= 0.43 (Wills and Clahan, 2006); Mesozoic sites in Greece have μ_{lnV} = 590 m/s and σ_{lnV}= 0.38; (Stewart et al., 2014). While Mesozoic sites in our database are not glaciated, the relatively large dispersion appears to be associated with thin, soft surficial layer effects that occur within this category (further discussion in *Proxy Performance*, below).

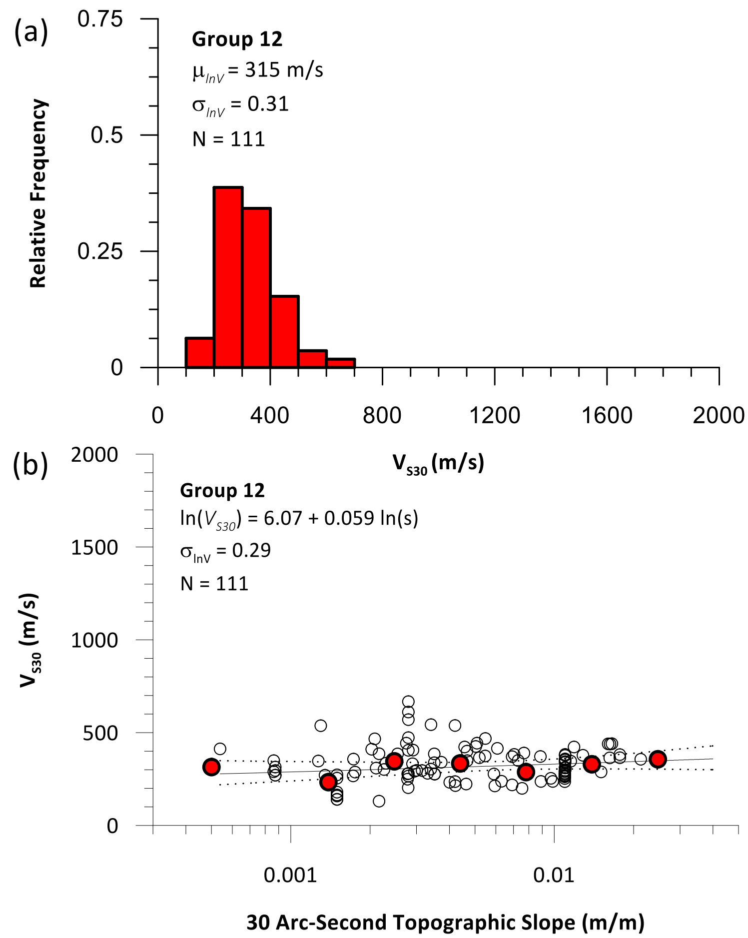

Figure 2.8. (a) Histogram of V_{S30} values in Group 12 (Table 2.3), and (b) V_{S30} as a function of slope for profiles in Group 12 (Table 2.3). Legend from Figure 2.4 applies.

2.5.5 Paleozoic

The Paleozoic age bin contains 177 V_{S30} values, and is subdivided into three groups (Groups 14-16) as shown in Figures 2.9b-d. Group 14 (Figure 2.9b) consists of Paleozoic sites in the Illinois Basin (Coleman and Cahan, 2012) and is only populated by 5 measurements. However, the log mean V_{S30} for this group is significantly lower than that of Groups 15-16 (μ_{lnV} = 513 m/s) and thus is retained as a separate group. Because Group 14 is so poorly populated, there is large epistemic uncertainty in its category mean and standard deviation. Groups 15-16 are divided in accordance with glaciation status (Figures 2.9c-d), and have μ_{lnV} = 684 m/s and μ_{lnV} = 972 m/s, respectively. Other basin structures (besides the Illinois basin) were not found to affect Paleozoic bin statistics.

Trends with slope gradient are not significant for Paleozoic sites and hence the recommended models are reported in Table 2.3 as μ_{lnV} and σ_{lnV} values only.

We suspect that the velocities in Groups 14-16 are affected by a number of issues, as we would expect intact Paleozoic bedrock to have a higher shear wave velocity than those reported in Table 2.3. One explanation is that Paleozoic residuum, or bedrock that has weathered in place, was included in this category (as mapped by Palmer 2006). Additionally in some cases, the geologic mapping may not be recognizing a thin layer of younger, softer sediments overlying the Paleozoic materials that is affecting the value of V_{S30}. Nonetheless, we do not remove these sites from our statistical analyses for two reasons: (1) we do not have independent confirmation of the presence of non-Paleozoic sediments at these sites and (2) such potential misclassifications are inherent to the use of geologic maps (and other proxies as well), and because such misclassifications are also unavoidable for forward application, they need to be reflected in group statistics until more refined geologic site classifications become available.

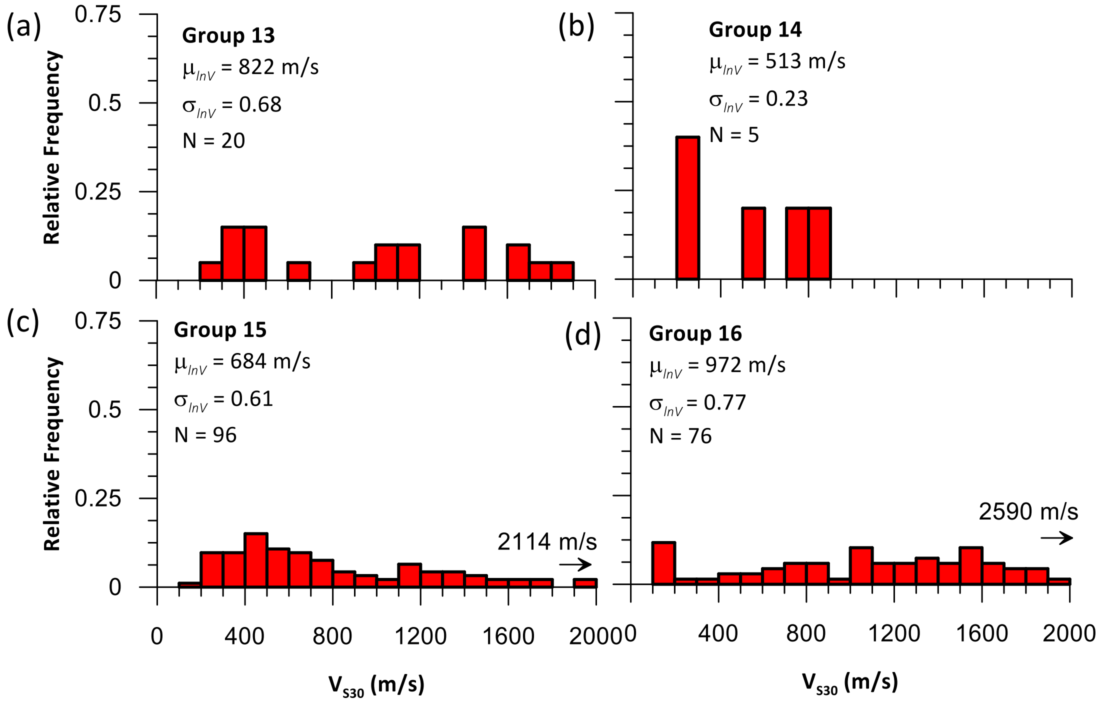

Figure 2.9. Histograms of V_{S30} values for (a) Group 13, (b) Group 14, (c) Group 15, and (d) Group 16.

2.5.6 Precambrian

The Precambrian age bin contains 37 V_{S30} values (Group 17). Figure 2.10 shows the histogram of V_{S30} obtained at Precambrian sites. The glaciated and non-glaciated measurements within Group 17 were determined to be non-distinct and hence were kept as a single group (Tables 2.2 and 2.3). We suggest using Group 17 when the location in question is mapped as Precambrian bedrock, with no site visit by a geologist having taken place. If a site visit has taken place, and the mapped Precambrian bedrock is confirmed to be outcropping at the site, we suggest using a V_{S30} of 2000 m/s (Group 18, Table 2.3), which is based on measurements at sites with geologic conditions of this type in Ottawa city and Quebec Province (Assatourians, personal communication, 2011; based on Atkinson and Mereu, 1992).

For Group 17, there are some V_{S30} values that do not seem physically reasonable (e.g. V_{S30} < 300 m/s). This is a consequence of using mapped geology as a proxy for V_{S30} from measurement, as discussed in the previous section. The effects of these complexities are reflected in the large natural log standard deviation associated with the proxy estimates σ_{lnV} = 0.85), the causes of which are discussed further in the next section.

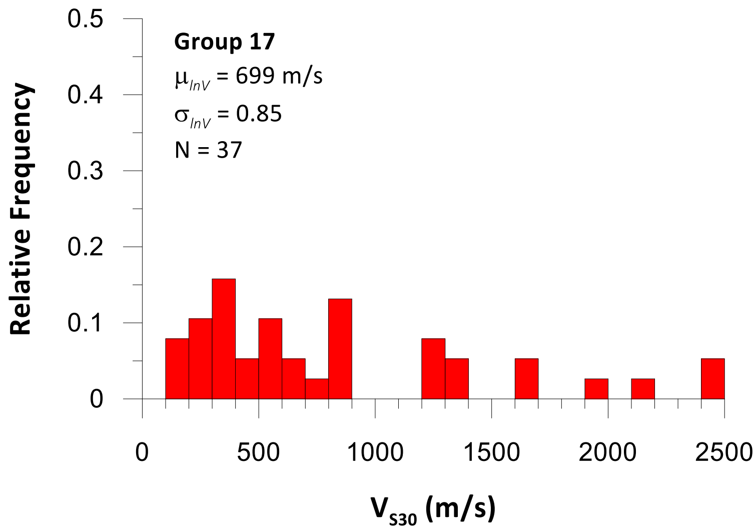

Figure 2.10. Histogram of V_{S30} values from Group 17 measurements.

2.6 PROXY PERFORMANCE

Proxy-based estimates of V_{S30} were assigned to the 2755 profiles in the database using the protocols summarized in Table 2.3. Residuals in natural log units were calculated as:

$$R_i = \ln(V_{S30})_i - \overline{\ln(V_{S30})_i} \tag{2.6}$$

where $\ln(V_{S30})_i$ is the natural log of the V_{S30} calculated from the velocity profile i, and $\overline{\ln(V_{S30})_i}$ is the proxy-based estimate for profile i (the overbar indicates that the mean is taken in natural logarithm units). Means and log standard deviations of the residuals can be computed for particular geologic conditions or for the data set as a whole; in the present case the means are expected to be near zero because the performance is evaluated using the same data set used in model development. Hence, our primary interest is in the standard deviation, σ_{lnV}.

Figure 2.11 shows histograms of the residuals for all profiles, previously glaciated profiles, and non-glaciated profiles. The metrics for overall proxy performance are $\overline{R}$ = 0.0016 and σ_{lnV} = 0.533 (comparisons to results of other proxies are given in the next section). An important outcome of the present work is quantification of the effect of glaciation on dispersion. Non-glaciated sites have relatively modest overall dispersion (0.357) that is significantly lower than has been found previously for CENA, but which is comparable to overall proxy dispersions for active tectonic regions (Seyhan et al., 2014). The σ_{lnV} for glaciated regions is much higher at about 0.656. Hence, the predictability of V_{S30} is better for non-glaciated than for glaciated sites. We suspect that the relative dispersion levels are caused by large impedance contrasts within the upper 30 m of glaciated sites, as seen in numerous V_S profiles measured in CENA. These sites presumably have had weathered geologic materials removed by glacial scour, with the remaining material being relatively competent and comprising the portions of the profiles below a strong impedance contrast. The relatively soft materials above the contrast have likely been laid down during or after glaciation. For sites of this type, V_{S30} is strongly correlated to the depth of materials above the impedance contrast, and because these depths are highly variable, the V_{S30} values too are strongly variable. In the absence of glaciation, sites are less likely to have these strong impedance contrasts, which could explain why the CENA proxy dispersions are comparable to those found in non-glaciated active tectonic regions. Moreover, among the non-glaciated sites, dispersion increases with age from about 0.23 for Holocene to 0.31 for Tertiary.

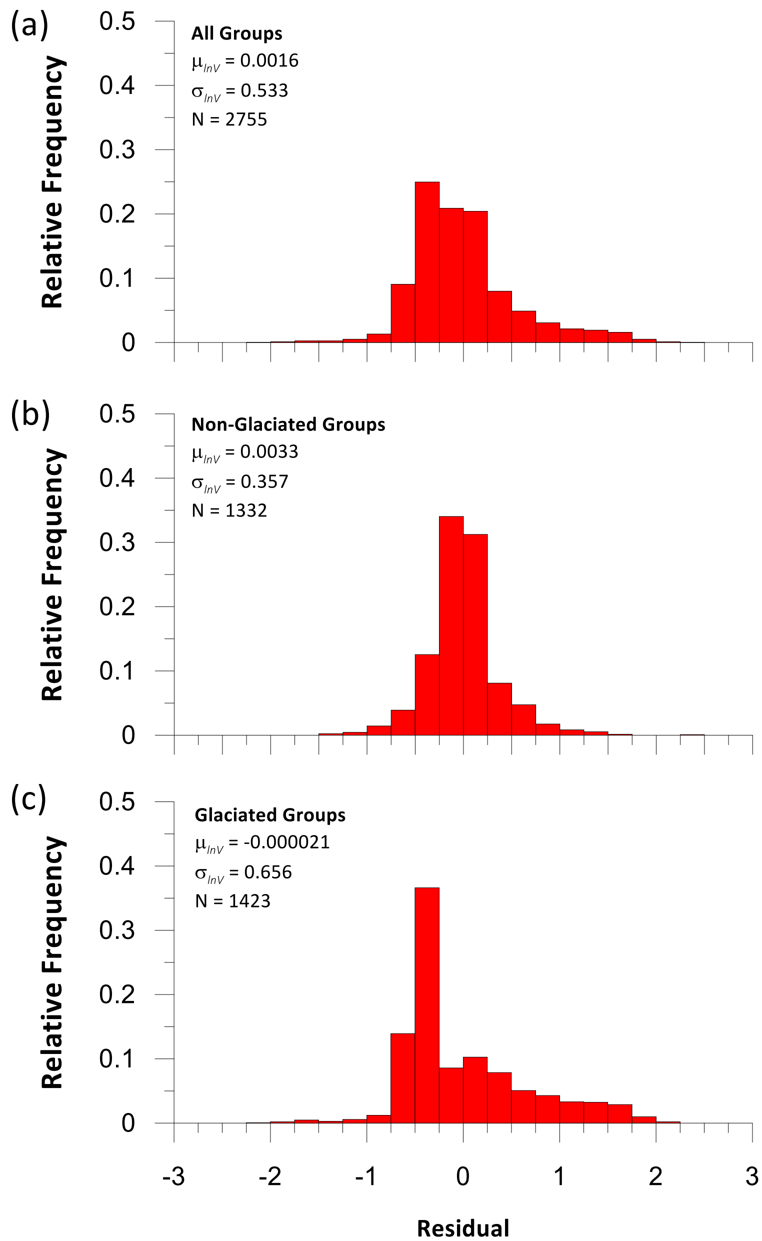

Figure 2.11. Comparison of residuals of V_{S30} for (a) all groups, (b) non-glaciated groups, and (c) previously glaciated groups.

Figure 2.12 plots residuals against 30 and 3 arc-s topographic gradients. The 3 arc-s gradient data capture higher resolution topography and thus include larger values of topographic slope. Results for 30 arc-s (Figure 2.12a) show minimal trends for gradients $\geq 3\times10^{-3}$ m/m, which is expected because 30 arc-s gradient was considered in model development. Residuals for 3 arc-s (Figure 12b) are comparable to those for 30 arc-s, with little bias. Plots similar to those in Figure 2.12b, but using data only for specific categories that exhibit a significant gradient effect when using 30 arc-s DEM (not shown) generally exhibit no residual trends. Hence, we conclude that our proposed hybrid slope-geology proxy captures gradient effects at either 30 or 3 arc-s resolution, and that the 3 arc-s gradients do not provide more predictive power than 30 arc-s gradients.

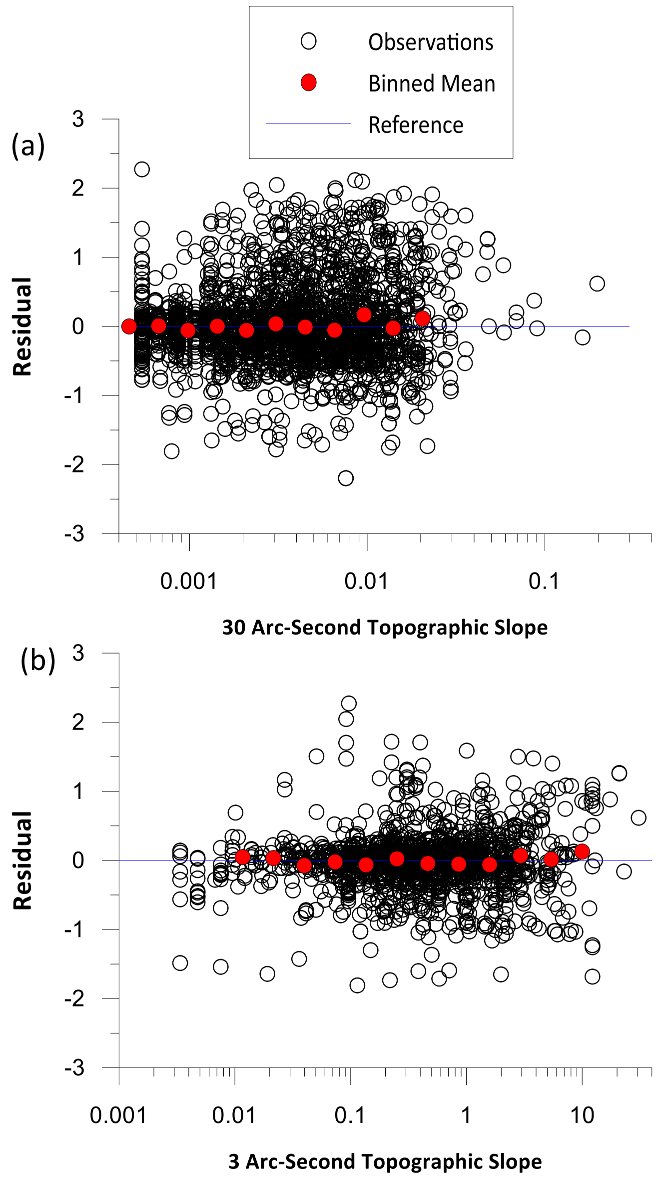

Figure 2.12. Proxy residuals as a function of (a) 30 arc-s topographic gradient for all measurements, and (b) 3 arc-s topographic gradient for measurements in the US, showing the binned mean of residuals as filled circles and a reference line at 0.

2.7 COMPARISON TO PRIOR WORK

As described previously, several proxy-based V_{S30} estimation procedures pre-date this work. To provide a consistent basis for comparing the proposed approach with prior relationships, we compute residuals using Eq. (2.6) for database sites, with the prior proxy relationships applied as published. As described in the *Geology- and Morphology-Based Proxies* section, information required to exercise each of these proxies is provided as metadata in the measurement database. The proxy relationships used in these analyses were:

a) WA07, which uses topographic gradient at 30 arc-s resolution.

b) Yea12, which uses terrain classes (site look-ups provided by A. Yong, 2012, pers. communication).

c) TS13 hybrid slope-geology using small-scale geologic maps (predicted V_{S30} values provided by E. Thompson, 2014, pers. communication).

d) Kea12 small-scale geology.

Model bias is estimated from the mean of the residuals ($\overline{R}$) and dispersion from the standard deviation of residuals (σ_{lnV}), which are evaluated over the entire set of residuals. The best-performing proxies will have relatively small biases (low $\overline{R}$) and low standard deviations. We should note here that our model was developed to best fit the dataset used for comparisons, whereas the other proxies (a-d) were not.

Figure 2.13 shows values of $\overline{R}$ and σ_{lnV} for each proxy, including the proposed approach. All four of the previous proxy relationships have a negative bias, indicating that they overpredict the measured V_{S30} values. The Kea12 surface-geology based proxy has the lowest σ_{lnV} of 0.592, but a bias of -0.282. The WA07 ground slope-based proxy has the lowest bias of -0.064, but a relatively large σ_{lnV} of 0.677. The proposed approach is unbiased (as expected), and has an overall σ_{lnV} of 0.533, which is modestly reduced from the lowest σ_{lnV} found from earlier proxies (0.592 for Kea12). The level of dispersion reduction is greater for the other proxy relations.

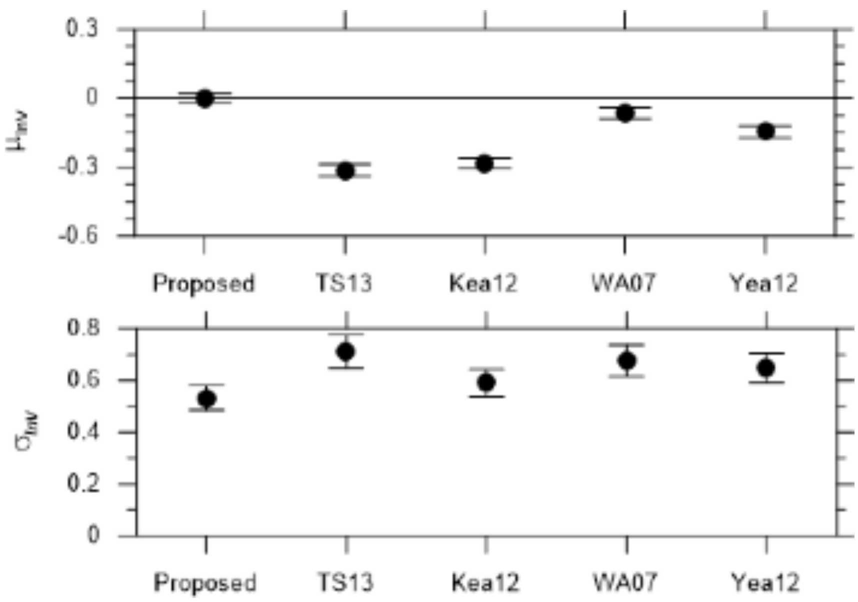

Figure 2.13. Comparison of log mean (top) and standard deviation (bottom) of residuals with 95% confidence intervals for existing and proposed VS30 proxies (TS13 = Thompson and Silva (20143); Kea12 = Kottke et al. (2012); WA07 = Wald and Allen (2007); Yea12 = Yong et al. (2012).

We also examined the residuals for a CENA-specific Yea12 proxy, taking the V_{S30} applied to each terrain category as the lognormal mean V_{S30} from the measurements in our database for that terrain category. This approach yielded $\overline{R}$ = -0.0003 and σ_{lnV}= 0.588, lower than that of proxies a-d. This is expected as it was developed to best fit the dataset used for comparisons.

The similarity of the σ_{lnV} values for the Kea12, CENA-specific Yea12, and proposed approaches suggest that the Kea12 and CENA-specific Yea12 approaches could be applied in forward applications. However, we propose that our method should replace these proxies because it better distinguishes between the effects of glaciation and non-glaciation (rather than glacially-derived sediments, which can be deposited outward of glacial limits), which as discussed previously has a significant impact on σ_{lnV}. Moreover, we have more confidence in the present larger-scale surface geology assignments that we have made than in previous assignments from small-scale maps in Kea12. To illustrate the significance of the geologic mapping source, we show in Figure 2.14a a plot of V_{S30} against topographic gradient for seemingly similar geologic categories that are well populated: the major unit of young non-glacial sediments (YN) from Kea12 [which includes alluvium (YNa), colluvium (YNc), loess (YNl), lacustrine, marine and marsh (YNm) and beach, dune, and sheet sands (YNs)], and the Holocene non-glaciated (HNG) category in the present work (encompassing all observed lithologies, Groups 1 and 2 in Table 2.3). The YN category in Kea12 encompasses non-glaciated sediments from late Pleistocene and younger, whereas the HNG category in this work excludes Pleistocene conditions, only including sites with geology 11,000 years and younger. The Kea12 bin has a wide range of V_{S30} (100 to 1000 m/s) for gradients ranging from 0 to 0.1. In contrast, the HNG category in the present work has narrower ranges of V_{S30} (100 to 500 m/s) and gradient (0 to 0.02). The differences in the data are such that a strong trend of V_{S30} with gradient is present in the Kea12 category, but no trend is observed using the presently defined HNG category. Figure 2.14b shows trends of data residuals from both groups (computed using Eq. 2.6) against topographic gradient; in the case of Kea12, YNc and YNl sites are excluded from the residuals calculations due to lack of estimated mean velocities, which removes many of the highest velocity sites. The Kea12 residuals in Figure 2.14b show a trend with topographic gradient that is not present for HNG. Moreover, the dispersion (σ_{lnV}) is lower using the present approach (0.25 as compared to 0.30 from Kea12). Our conclusion is that in this case, as in others not shown for brevity, the proposed approach based on larger-scale geologic maps better differentiates V_{S30} as represented by within-category μ_{lnV}, σ_{lnV}, and trend with gradient.

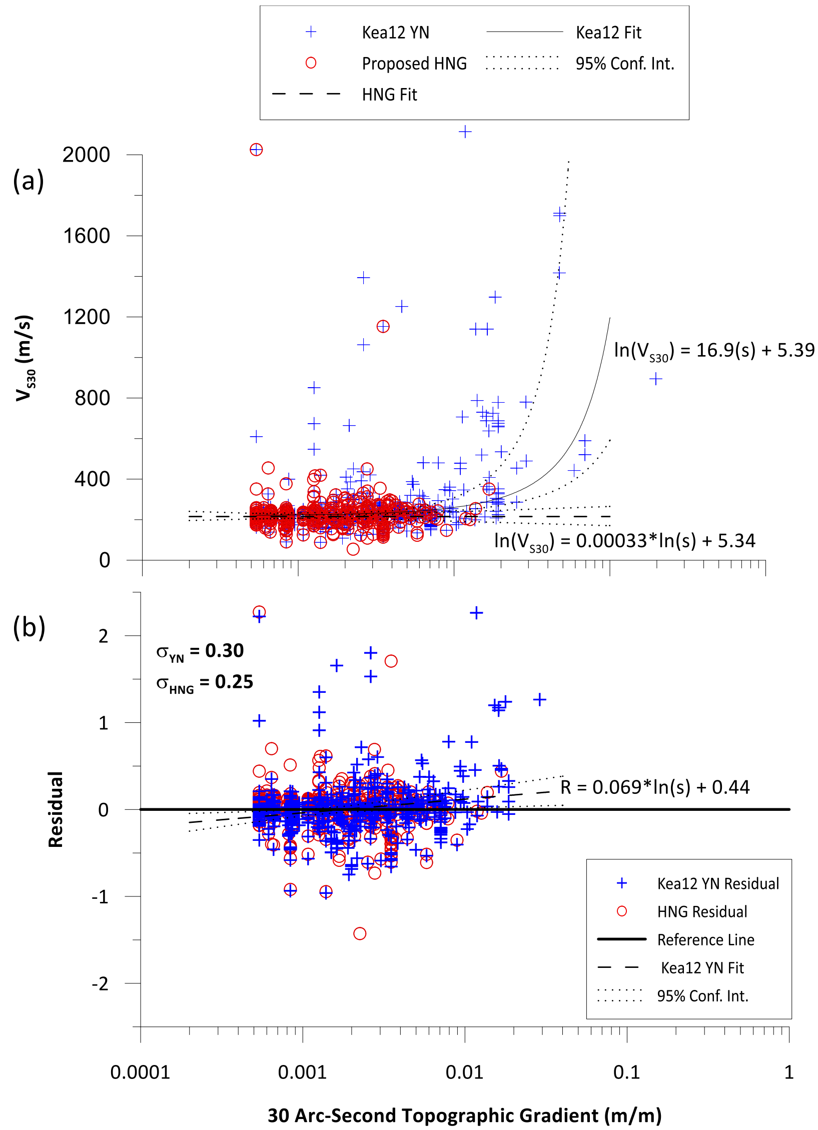

Figure 2.14. (a) Comparison of V_{S30} as a function of topographic gradient for the Kea12 young non-glacial (YN) category (including all sub-categories), and the Holocene non-glaciated (HNG, encompassing Groups 1 and 2) categories from the present work. Lines of best fit (Eqs. 2.4, 2.5) are shown for both groups. (b) Comparison of residuals as a function of topographic gradient for the Kea12 YNa, YNm, and YNs categories, and the Holocene non-glaciated (HNG, encompassing Groups 1 and 2) categories from the present work.

2.8 IMPLEMENTATION

Best practices in site characterization are to develop full V_S profiles (extending to rock) derived from geophysical data. When it is necessary to estimate V_{S30} for sites lacking such data, we recommend applying the P-wave seismogram method for V_{S30} estimation (Kim et al., 2016) when sufficient ground motion recordings are available (relationships between the frequency of the peak in horizontal-to-vertical spectral ratios and V_{S30} are an alternate approach, but have not been applied here; Hassani and Atkinson, 2016a), and otherwise we recommend application of the proxy relationships in this chapter. For application to V_{S30} assignments in the NGA-East station database (available in Table S2 of the electronic supplement to Parker et al. 2017), we applied the protocols below (listed in order of preference), which update those given in Section 5.5 of Goulet et al. (2014):

0. Assign mean V_{S30} from measured V_S profiles. Standard deviation taken as σ_{lnV} = 0.1 per Seyhan et al. (2014).
1. Assign mean V_{S30} from known site conditions and geology based on measurements of V_S profiles at different location but the same geological condition. This assignment is only used based on a recommendation or site visit from a geologist. Standard deviation taken as σ_{lnV} = 0.3, as per Goulet et al. (2014).
2. Estimate mean V_{S30} by P-wave seismogram method (Kim et al., 2016) for sites having multiple ground motion recordings and corresponding V_{S30} values from measurements. Standard deviation is taken as σ_{lnV} = 0.456.
3. Estimate by hybrid slope-geology proxy developed in this chapter. Mean and standard deviation taken from Table 2.3.

The numbers in the above list are codes provided in the station database. Of the 445 sites in the flatfile recommended by the NGA-East Technical Integration team for GMM development, 53 (12%) are Code 0 (on-site V_S profile measurement), 77 (17%) are Code 1 (V_{S30} assigned after a site visit by a geologist), 10 (2%) are Code 2 (from P-wave seismogram method), and 305 (69%) are Code 3 (assigned based on the protocols in this work).

2.9 CONCLUSIONS

Because the overwhelming majority of seismic recording stations in CENA lack measured V_S profiles, the estimation of the site parameter V_{S30} is critical for the application of strong motion data during GMM development and in the ongoing process of developing site factors. Preliminary estimates of V_{S30} were provided in the NGA-East data report (Goulet et al., 2014), which are updated herein.

We compiled a database of V_{S30} values obtained from measured V_S profiles that was not utilized in the preliminary NGA-East V_{S30} assignments. When predictions from pre-existing proxy relationships are compared to the V_{S30} values in this database, significant bias and large dispersion is found, which partly motivated the present work. We compiled geologic information from larger-scale geologic maps maps, supplemented by mapping that indicates glaciation/non-glaciation and the presence of sedimentary basins, which forms the basis for the present recommendations. None of this information was utilized in the development of previous proxy relations (Kea12, TS13, WA07, Yea12).

Table 2.3 presents coefficients needed to apply the recommended proxy relationship. Some geologic categories take the mean V_{S30} as a simple category natural log mean, whereas others take the mean from a topographic gradient based model specific to the category using Eqs. (2.4) or (2.5). Values of σ_{lnV} to accompany each mean estimate are given in Table 2.3. These estimates are used when more reliable, site-specific information is unavailable, as given by the implementation procedures in the previous section.

An important outcome of the present work is quantification of the effect of glaciation on dispersion. Non-glaciated groups have dispersions that are significantly lower than has been found previously for CENA (0.357), but which is comparable to proxy dispersions for active tectonic regions (Seyhan et al., 2014). The σ_{lnV} for glaciated regions is higher at 0.656. Hence, the predictability of V_{S30} is better for non-glaciated than for glaciated groups, which should be taken into consideration in the weighting of ground motion data from the two site types during GMM development. A disadvantage of the mapping approach adopted herein is that we cannot create a map like Figure 7 of Kea12 or Figure 5 of TS13 that shows the geologic conditions across CENA. Both Kea12 and TS13 started with one continuous map source and assigned a V_{S30} value to each map unit or combinations of map units. However, this is not practical with present resources using the larger-scale maps because they are not continuous across CENA and map units are not consistently defined across map resources. Moreover, the majority of mapped geologic units are not available as shape files that can be imported to geographic information system (GIS)-based mapping software.

3 EMPIRICAL LINEAR SEISMIC SITE AMPLIFICATION IN CENTRAL AND EASTERN NORTH AMERICA

3.1 INTRODUCTION

The Next Generation Attenuation – East (NGA-East) Project, coordinated by the Pacific Earthquake Engineering Research Center (PEER), resulted in the development of 21 ground motion models (GMMs) applicable to very hard rock reference site conditions (shear wave velocity, V_S = 3,000 m/s; Hashash et al. 2014), in central and eastern North America (CENA) (PEER, 2015a; Goulet et al. 2017). Therefore, models for seismic site amplification are needed to predict ground motion intensity measures for other site conditions, including weathered rock and soil (i.e. for site conditions with time averaged shear wave velocity in the upper thirty meters, V_{S30}, less than 3000 m/s; this includes essentially all sites in CENA except for facilities founded on hard rock below the weathered zone).

As part of NGA-East, the Geotechnical Working Group (GWG) was formed in part to develop site amplification models. Other GWG tasks included defining the CENA reference site condition (Hashash et al., 2014) and providing information on site conditions at CENA recording stations, including the use of proxy-based V_{S30} prediction models where needed (Parker et al. 2017). The concept behind the GWG site amplification model development is broadly similar to the approach in NGA-West1 (Power et al., 2008) and NGA-West2 (Bozorgnia et al., 2014). In those projects, site amplification models were developed in consideration of both empirical data and simulation results. Whereas site amplification terms related to the linear scaling of ground motion with V_{S30} were empirically derived, simulations were used to support development of nonlinear terms (using one-dimensional ground response simulations; Walling et al. 2008, Kamai et al. 2014) and basin amplification terms (using three-dimensional simulations; Day et al. 2008).

In a similar manner, the GWG approach was to develop modular site response models derived using a combination of empirical data analysis and simulations. In this context, *modular* refers to model components that can be combined to compute total site amplification. As described further by Harmon et al. (2019b), model components include linear (shaking amplitude-independent) terms conditioned on V_{S30}, sediment depth, and site period, as well as nonlinear terms. The simulation component of GWG work is described by Harmon et al. (2019a, 2019b) and produced versions of each model component. Here we present an empirically-derived model for the linear component of site amplification conditional on V_{S30}. Our use of V_{S30} as the primary site parameter is motivated in part by practical reasons (e.g., its widespread use in seismic codes and ground motion models, Bozorgnia et al. 2014; PEER 2015; Dobry et al. 2000) which provides benefits such as facilitating comparisons to other regions/models (e.g. NGA-West2). Moreover, in most regions V_{S30} has been found to be an effective first order site parameter for predicting site response over a broad period range, even though it cannot predict site-specific features such as resonance at a site fundamental period. We do not investigate nonlinear site response effects nor sediment depth effects because the available ground motion recordings in CENA are generally of low amplitude and information on basin depth is not available for CENA ground motion stations. The modularity of the GWG models is such that the V_{S30}-scaling terms presented here can be combined with simulation-based terms for nonlinearity (further details in Harmon et al. 2019b).

In the following chapter, we review relevant prior work including NGA-East GMMs and CENA site response studies, describe the model development process (data considered, equations, regression procedures), illustrate model performance including residuals analysis and comparisons to prior results for active tectonic regions, and provide recommendations for practical applications. The model presented here is for linear site amplification conditioned on V_{S30} and applies for the intensity measures of peak ground velocity (PGV), and 5% damped 50th percentile rotated pseudo-spectral acceleration (RotD50 PSA) for oscillator periods between 0.065 to 8.0s. Peak ground acceleration (PGA) is excluded due to inadequate time sampling (i.e., too-large time step) of NGA-East ground motions to resolve portions of the data with frequencies higher than the Nyquist frequency, which influences PGA (e.g., Boore and Goulet 2014).

3.2 PRIOR WORK

3.2.1 NGA-East Ground-Motion Models

The development of NGA-East ground motion models (GMMs) began with a series of candidate models (PEER 2015a), a subset of which were selected and then adjusted to correct for various distance scaling issues (PEER 2015b). The models from the PEER (2015a) report were used as seed models for the generation of a range of GMMs, described in Goulet et al. (2017), that are intended to capture, in aggregate, epistemic uncertainties in ground motions from source and path effects following a Sammon's map approach (e.g., Scherbaum et al. 2010). Our work in model development spanned the time period from seed model development to practical completion, and as a result, models from several stages of this GMM development process are considered.

Table 3.1 summarizes some of the principal attributes of ten NGA-East candidate GMMs (PEER 2015a). Three of the models (Boore 2015a; Darragh et al. 2015; and Yenier and Atkinson 2015) are based on the point-source simulation methodology. Parameters included in the simulations, especially the stress parameter and path attenuation terms, are set based on comparisons to NGA-East data or prior compilations of CENA data. Two of the models (Pezeshk et al. 2015, Shahjouei and Pezeshk 2015) use the hybrid empirical approach of Campbell (2003), in which GMMs for active tectonic regions (from NGA-West2) are modified for CENA using ratios of simulated ground motions. One model uses a conceptually-similar referenced empirical approach in which a GMM for active tectonic regions is adjusted through residuals analysis using NGA-East data (Hassani and Atkinson 2015). Three of the models are based on direct regression of NGA-East data to develop GMMs (Al Noman and Cramer 2015; Graizer 2015; and Hollenback et al. 2015a). Yenier and Atkinson 2015 also uses direct regression to calibrate the regionally-adjustable parameters of the generic point-source model. Due to the limited parameter space covered by the data, some developers used additional information during model building, including intensity data (Al Noman and Cramer 2015), or simulations (Graizer 2015; Hollenback et al. 2015a). Finally, one GMM consists of an inventory of finite-fault simulation results (Frankel 2015).

All of the GMMs in Table 3.1 provide ground motion estimates for either a reference site condition of V_S = 3000 m/s or V_S = 760 m/s, and site damping parameter κ_0 = 0.006 s (Hashash et al. 2014). Five of the models contain no site term and provide ground motion estimates only for the reference site condition. Five models contain a V_{S30}-based site term that is intended to capture the effects of V_{S30} on the linear site amplification. Some models used a variety of site corrections during development, even if the models themselves do not contain a site term. As a result, there are a number of site amplification models, reflecting various approaches in their development, within the documentation for the ten NGA-East candidate GMMs.

As shown in Table 3.1, the alternative approaches for estimating site amplification that were used during NGA-East GMM development included:

1. Adopting models for active tectonic regions, specifically the Seyhan and Stewart (2014) model developed for NGA-West2. This model was used as the site term in NGA-East

models by Yenier and Atkinson (2015) and by Hassani and Atkinson (2015), and to support model development by Pezeshk et al. (2015) and Shahjouei and Peszehk (2015).

2. Regression of data to develop a linear V_{S30}-scaling model (Al Noman and Cramer 2015; Hollenback et al. 2015a).
3. Ground response analysis simulations, typically using viscoelastic soil conditions (Darragh et al. 2015; Graizer 2015).

We note at this stage that the V_{S30} values at ground motion stations used in these prior analyses were relatively rough preliminary estimates provided in Chapter 5 of Goulet et al. (2014), which have since been updated as given by Parker et al. (2017).

As shown in this article, site amplification in CENA is found to differ from that in active regions. Hence there is an issue of incompatibility in the site terms used in GMM development relative to those recommended for application. We comment on this issue subsequently.

Table 3.1. Summary of attributes of NGA-East median candidate GMMs (PEER 2015a). AB06, AB11 = Atkinson and Boore (2006, 2011); BT15=Boore and Thompson (2015); SS14 = Seyhan and Stewart (2014); GRA=ground response analysis. All point source simulations utilize parameters calibrated against NGA-East data.

PEER (2015a) Chapter	Author	Approach	Distance type	Distance range (km)	M range	Site term & parameter	Site correction: V_{S30} to 760	Site correction: 760 to 3000
2	DM Boore	Point source simulations	R_{ps}	0–1200	4–8	No	N/A	Boore (2015b)
3	RB Darragh, NA Abrahamson, WJ Silva, N Gregor	Point source simulations	R_{JB}	0–1000	4.5–8.5	No	1D GRA transfer functions for NEHRP Cats; goes from V_{S30} to 4.68 km/s	
4	E Yenier and GM Atkinson	Point source simulations	R_{ps}	0–600	3–8	Yes (V_{S30})	SS14	AB06 BC crustal amp (Atkinson 2012)
5	S Pezeshk, A Zandieh, KW Campbell, B Tavakoli	Hybrid empirical	R_{RUP}	0–1000	3–8	No	SS14 (for validation only)	BT15
6	AD Frankel	Finite fault simulations	R_{RUP}	2–1000	4.5–8	No	N/A	Frankel et al. (1996)
7	A Shahjouei and S Pezeshk	Hybrid empirical	R_{JB}	2–1000	5–8	No	SS14 (used for validation only)	AB06 and BT15;
8	N Al Noman and CH Cramer	Empirical with intensity data	R_{RUP}	<10–2000	2.5–7.7	Yes (V_{S30})	Set by regression, parameter d_1	NA
9	V Graizer	Empirical	R_{RUP}	0–1000	4–8.2	Yes (V_{S30})	GRA-based: Eq. 9.6	GRA-based: similar to AB06, AB11
10	B Hassani and GM Atkinson	Referenced empirical	R_{JB}	0–400	3–8	Yes (V_{S30})	SS14	AB06 BC crustal amp (Atkinson 2012)
11	J Hollenback, N Kuehn, CA Goulet, NA Abrahamson	Empirical with finite fault simulations	R_{RUP}	0–1200	4–8.2	Yes (V_{S30})	Set by regression, parameter c_8	Boore (2015b)

3.2.2 CENA Site Amplification

Empirical site amplification studies, while numerous in active tectonic regions, are relatively rare in stable continental regions like CENA. Khaheshi Banab et al. (2012) showed that for a soft soil site in eastern Canada, weak motions were amplified near the site period by more than a factor of 10 with respect to a nearby hard-rock reference site. Atkinson et al. (2015) used ground-motion regression to develop a GMM for southern Ontario in which site amplification was determined for each soil site with respect to motions on hard-rock sites. Hassani and Atkinson (2016a) derived the frequency of peaks in H/V spectral ratios using CENA data, and used those peak frequencies as predictive parameters for analysis of site effects. They find that the data-derived peak frequencies are more effective than V_{S30} at predicting site effects in the CENA data. Building on that result, Hassani and Atkinson (2017) developed an empirical amplification model for CENA using the NGA-East database and selected GMMs. This model provides linear site amplification components as a function of site frequency and V_{S30}, with site frequency being considered the primary site parameter if available. A recent GMM for Oklahoma (Yenier et al. 2017) includes empirical site terms referenced to the regional average site condition of NEHRP C (V_{S30}=360-760 m/s; Dobry et al. 2000).

The limited previous studies of empirical site response in CENA, especially prior to the NGA-East project, is due to a number of factors, including a lack of V_{S30} information at seismographic sites. A major component of the GWG scope of work was assembling these estimates (Parker et al. 2017), which enabled the work described here and by others. The majority of past seismic site amplification work in CENA has focused on simulation-based approaches. A review of that prior literature is provided in Harmon et al. (2019b).

3.3 MODEL DEVELOPMENT

We adopt a non-reference site approach (Field and Jacob, 1995) to investigate site amplification in CENA. Following procedures widely used in active tectonic regions (e.g., Stewart et al. 2003, Sandıkkaya et al. 2013, Seyhan and Stewart 2014), site amplification is taken as the within-event rock residual (ε_R) computed relative to GMMs conditioned at a reference rock site condition:

$$\varepsilon_{Rij} = ln(Y_{ij}) - \left(\mu_{ln}(\mathbf{M}_i, R_{ij}, V_{ref}) + \eta_{E,i}\right) \tag{3.1}$$

where Y_{ij} is the intensity measure from a recorded ground motion for event i at station j, $\mu_{ln}(\mathbf{M}_i, R_{ij}, V_{ref})$ is the median GMM prediction for the appropriate magnitude ($\mathbf{M}_i$) and distance (R_{ij}) for the recording, and the reference V_{S30} site condition (V_{ref}) for the GMM, and $\eta_{\square,i}$ is the event term for event i. For well recorded earthquakes, the event term, $\eta_{E,i}$, represents approximately the mean bias of recordings from event i relative to a GMM. As described further below, the GMMs used for this purpose are Yenier and Atkinson (2015) (YA15), Hassani and Atkinson (2015) (HA15), and Boore (2015a).

By using within-event residuals, we remove event-specific bias of the GMM that would otherwise add scatter and potential bias to estimates of site amplification. Within-event residuals include the effects of between-site and between-path variability; the operating assumption of the non-reference site approach is that the between-path variability averages to zero over a large population (if the reference GMM is well behaved), so that the mean of the remaining variability represents mean site effects. The resulting site response can be considered as linear when the ground motion amplitudes are predominantly small, which is the case here. This process of residual partitioning, the data set, and the GMMs that were used in our approach are discussed further in the following sections. To our knowledge, this work is distinct from previous site amplification work performed prior to and during NGA-East as a result of the following two aspects: (1) we use only empirical data in lieu of simulations to infer site amplification and (2) site amplification is related to a truly independent variable, V_{S30}, rather than a predictor variable derived from ground motions at the stations (i.e., H/V spectral ratios).

3.3.1 Ground Motion Database

The NGA-East ground motion database (available as an electronic supplement to Goulet et al. 2014) was used in model development, with some modifications. The V_{S30} values for recording stations were updated using the recommended procedure in Parker et al. (2017). In order of preference, this procedure assigns V_{S30} using in situ measurements, estimates derived from measurements at sites having similar geologic conditions as confirmed by a geologist (hard rock sites only), estimates derived from on-site recordings using the P-wave seismogram method (Kim et al. 2016), and a hybrid slope-geology proxy (Parker et al. 2017).

Spectral ordinates in the database were only used between their lowest and highest useable periods as defined and given in Goulet et al. (2014). We also screen data by the same criteria applied in the development of the GMM used to compute the median term in Eq. (3.1). For YA15 this includes: (1) $\mathbf{M} \geq 3$, (2) $R_{rup} \leq 600$ km, (3) events with at least 3 recordings, and (4) events with an estimate of hypocentral depth. For HA15 this includes: (1) $\mathbf{M} \geq 3$, (2) $R_{rup} \leq 400$ km, (3) events with at least 3 recordings. Lastly, events and recordings from the Gulf Coast region (as defined in Goulet et al. 2014) were used, but only when both the event and recording were in the Gulf Coast region (matching criteria used in NGA-East GMM development). As shown in Figure 3.1, these screening criteria affect the number of events and recordings used as a function of the PSA oscillator period. The data used for PGV is shown in magnitude-distance space in Figure 3.2.

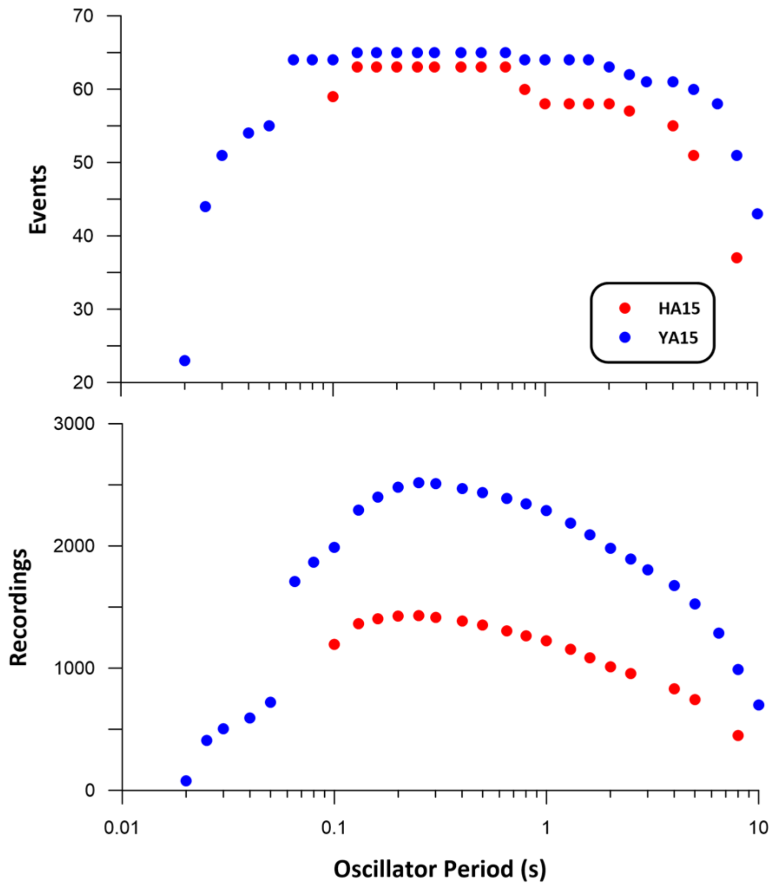

Figure 3.1. The number of events (top) and recordings (bottom) as a function of oscillator period due to the data screening criteria used for the YA15 (blue) and HA15 (red) GMMs.

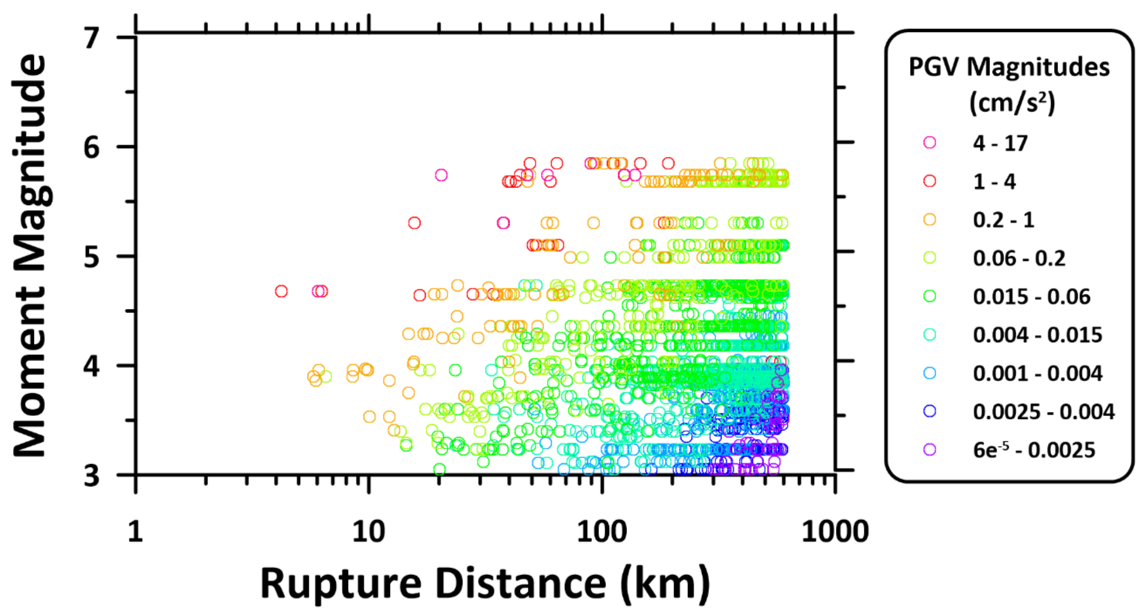

Figure 3.2. The PGV data from the NGA-East database used for site amplification model development shown in magnitude-distance space.

3.3.2 Rock Conditioned Ground Motion Models

As described in the *Prior Work* section, we considered reference site GMMs developed at different stages of the NGA-East project. Initially, we selected the rock-conditioned GMMs of YA15 and HA15. Our two principle considerations were that the models were developed using different and (in our judgment) credible approaches and that they were published in peer-reviewed journals. Following subsequent analysis in which these and other seed models were plotted in Sammon's map space (Goulet et al. 2017), we selected a third model (termed B_SGD02, from Boore 2015a) so as to cover a range of model attributes as indicated by their diverse positions on the Sammon's map.

Sammon's maps illustrate the multi-dimensional differences between GMM predictions in two-dimensional space. Figure 3.3, modified from Goulet et al. (2017), shows the Sammon's map for 1.0 s PSA, where the distance between points represents differences between models, and the map is centered on the mean of all predictions (labelled 'mix' in Figure 3.3). Distances between models can be measured in different azimuths within the space, which represent different GMM attributes. For example, in Figure 3.3 the azimuth approximately 12 degrees counter-clockwise from the direction of the ordinate represents magnitude scaling, with points labelled M+ and M- indicating faster and slower, respectively, increases in ground motion with magnitude. For the present application, the GMM attribute of greatest interest is distance scaling, because source scaling effects are immaterial due to the use of event terms η_i in Eq. (3.1). The distance scaling of the GMMs is represented along the southwest to northeast diagonal and marked with the R+ and R- symbols, indicating slower and faster attenuation, respectively.

In Figure 3.3, the red points represent the locations of selected models. The YA15 GMM is near the global mean, whereas the HA15 GMM (near the R++ reference point) has slower distance attenuation. We added the B_SGD02 model (Boore 2015a), which has faster distance attenuation (steeper slope) in the moderate magnitude range (4-5), to span the range of NGA-East models with respect to their distance attenuation attributes. As described further below, our proposed site amplification model is based jointly on the YA15 and HA15 models, with the B_SGD02 model used for validation.

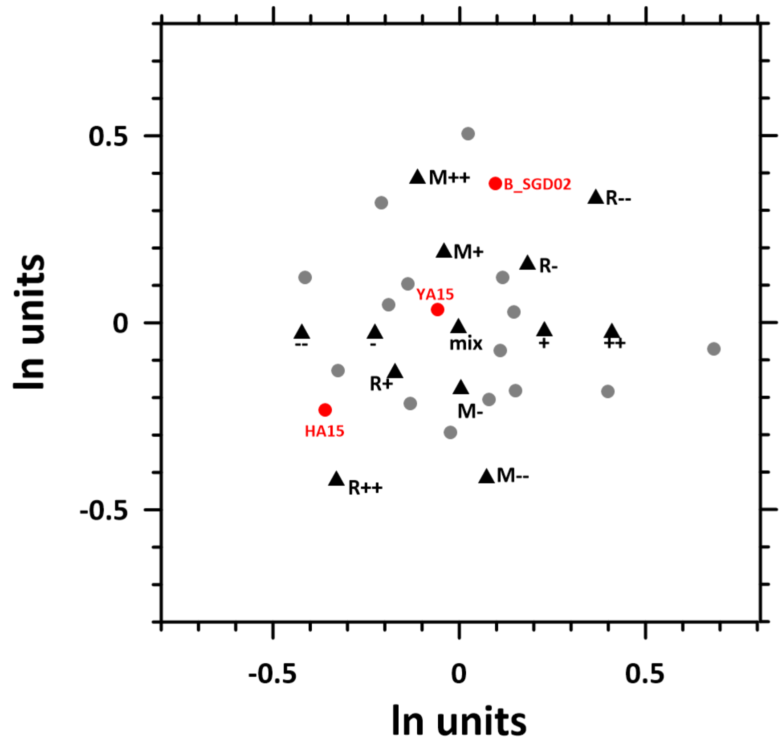

Figure 3.3. Sammon's Map space representation of NGA-East seed GMMs at 1.0s PSA (PEER 2015). The results shown here are modified from a figure given by Goulet et al. (2017). Circular symbols represent seed models, with GMMs used in this study shown in red (HA15 = Hassani and Atkinson 2015; YA15 = Yenier and Atkinson 2015, and B_SGD02 = Boore 2015a). Triangular symbols represent reference points, with the map centered on the mean of all predictions (labeled 'mix'). The triangles labeled with +/- represent scaled versions of the average model. The triangles labeled with M represent the averaged model with changed magnitude scaling: mix + β(M-6), with β = -0.4, -0.2, 0.2 and 0.4. Lastly, the triangles labeled with R represent the average model with modified distance scaling: mix + β(lnR-100), with β = -0.5, -0.25, 0.25, and 0.5.

Our goal in considering multiple models was to investigate possible sensitivities in site response to the selected GMMs. We note that B_SGD02 is not recommended for forward use by Boore (2015a) (The recommended models in Boore 2015s are B_AB95, B_BCA10D, and

B_BS11. These recommendations have been updated in Boore 2018). The attributes of the model that cause it to occupy the targeted portion of Sammon's maps space are the same features that cause it not to be recommended by Boore (2015a). Nonetheless, we consider the model to be suitable for the present sensitivity study.

The YA15 GMM is based on an equivalent point-source simulation framework with scaling characteristics calibrated using ground motion data from California. This generic GMM is then applied to CENA by using regression of the NGA-East data to derive the regionally-adjustable parameters (stress drop, anelastic attenuation coefficient, and calibration constant). The GMM provides median predictions of PGA, PGV and 5%-damped RotD50 PSA at oscillator periods up to 10s, for **M** 3-8 earthquakes, rupture distances (R_{rup}) ≤ 600 km, and V_{S30} = 760 m/s.

The HA15 GMM is based on a referenced empirical approach (Atkinson 2008). The NGA-West2 Boore et al. (2014) GMM, developed for active tectonic regions, was compared to ground motion data from the NGA-East database and other resources. The Boore et al. GMM was then calibrated to CENA based on the ratio of observed ground motions to GMM predictions. The GMM provides median predictions of PGA, PGV and 5%-damped PSA at oscillator periods from 0.05-10s, for **M** 3-6 earthquakes, rupture distances (R_{rup}) ≤ 400 km, and for V_{S30} = 760 m/s.

The B_SGD02 GMM consists of a set of stochastic point-source simulation results for **M** 4-8 earthquakes, R_{rup} = 2-1200km, and a V_{S30} = 3000 m/s reference site condition. The intensity measures provided are PGA, PGV, and 5% damped PSA from 0.01-10s. The stress parameters used in the simulations were derived from inversion of PSA data at 0.1 and 0.2s from nine earthquakes in eastern North America (Boore 2015a), and the attenuation model (Q and geometric spreading) are from Silva et al. (2002). Predictions of this model were adjusted to a 760 m/s reference condition using the V_{S30} 760/3000 m/s adjustment factors recommended in Stewart et al.

(2017a) (which, in turn, are based on models by Campbell and Boore 2017, Harmon et al. 2019b, and Darragh et al. 2015).

A comparison of the distance scaling of the three GMMs for **M** = 5 and 6.5 is shown in Figure 3.4a, and a comparison of the magnitude scaling of the three GMMs for rupture distances of 30 and 300 km is shown in Figure 3.4b. The faster distance attenuation of B_SGD02 is evident from Figure 3.4a, as is steeper magnitude scaling (Figure 3.4b). These differences are most pronounced for long-period PSA and combine to produce appreciably lower ground motions from B_SGD02 relative to YA15 and HA15 for low magnitudes and large distances (e.g., factor of 3 for R_{rup} = 300 km and **M** = 5). Such features are also present to a lesser degree for short-period PSA and PGV.

The NGA-East project documentation provides adjustments to GMMs for events in the Gulf Coast region to account for faster distance attenuation. Hollenback et al. (2015b) present two sets of adjustment factors based on empirical data analysis and calibrated simulations. We use the average adjustment provided by the two models, which is applied to events in the Gulf Coast that were recorded in the Gulf Coast (i.e., recordings for which the path did not cross into or out of the Gulf Coast as defined by Goulet et al. 2014).

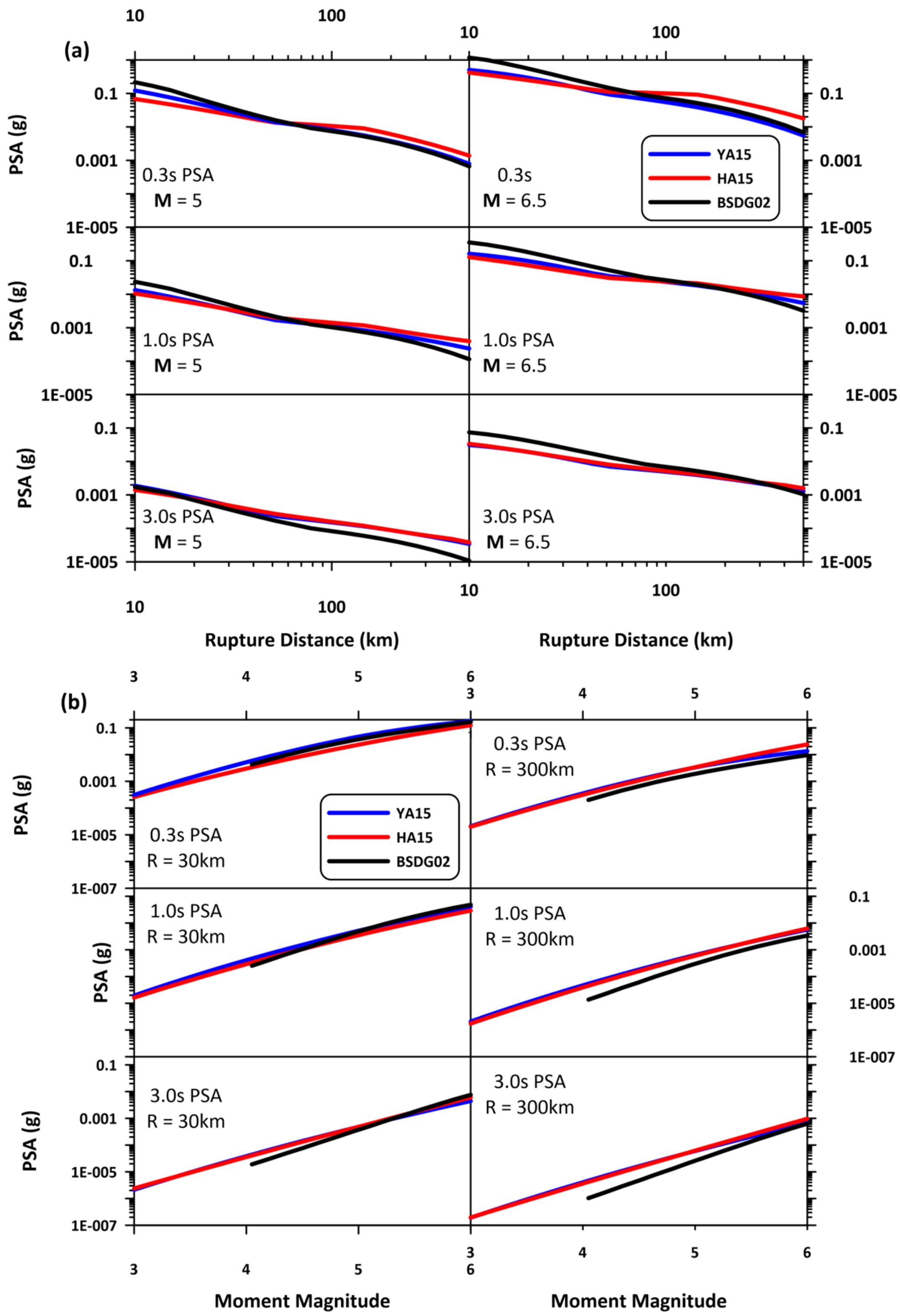

Figure 3.4. Trellis plots comparing median predictions from Hassani and Atkinson 2015 (HA15), Yenier and Atkinson 2015 (YA15), and Boore 2015a (B_SGD02) with regard to (a) distance attenuation for **M** 5 and 6.5 events and (b) magnitude scaling for rupture distances of 30 and 300km. Results shown for 0.3, 1.0, and 3.0s PSA. Boore 2015a results not shown below **M**4 as they are not provided below that magnitude.

3.3.3 Iterative Partitioning of Residuals

The analysis of within-event rock residuals in Eq. (3.1) was iterative, because event terms η_i depend on the site amplification model, which in turn is derived from the ε_{Rij} values. We begin by calculating total residuals for the selected data subset as:

$$R_{ij} = ln(Y_{ij}) - [\mu_{ln}(\mathbf{M}_i, R_{ij}, 760) + F_S(V_{S30j})] \tag{3.2}$$

where F_S is a V_{S30}-dependent site term, initially taken from a model applicable to active tectonic regions (Seyhan and Stewart, 2014), and other terms are as defined previously.

Next, we partition R_{ij} into between- and within-event components using mixed-effects analyses via the *lme4* package in *R* (Bates et al., 2015; R Development Core Team, 2008), as follows:

$$R_{ij} = c_k + \eta_{E,i} + \delta W_{ij} \tag{3.3}$$

This process provides the mean GMM misfit, c_k, as well as event terms $\eta_{E,i}$. The remaining residual is δW_{ij}, which represents the within-event component of variability.

As shown in Eq. (3.1), within-event rock residuals (ε_{Rij}) are calculated using selected GMMs with a reference site condition (indicated by V_{ref}) modified by event terms $\eta_{E,i}$. Due to the removal of event terms in Eq. (3.1), rock residual ε_{Rij} is presumed to have contributions from GMM misfits to the data associated with site and path effects. An implicit assumption in the non-reference site approach is that path misfits are randomly distributed about zero, such that they contribute scatter but not bias. Because the GMM is for rock, whereas the data is for various site conditions differing from rock, the site components are expected to produce both scatter and bias. It is this bias that we use in model development, as shown in subsequent sections.

Before proceeding further, it should be pointed out that the bias c_k could be subtracted from $\ln(Y_{ij})$ in Eq. (3.1), which amounts to a shift of the data up or down in natural log ground motion space. However, because we apply a shift to force the models through an amplification of unity at the selected reference site condition (V_{S30} = 760 m/s), as explained further in the next section, there is no need to apply the c_k adjustment at this stage.

The iterative process operates by deriving interim models for site amplification F_S using the CENA data, as explained in the next section. The first such interim model is used in Eqs. (3.2-3.3) to produce new residuals and event terms, followed by rock residuals (Eq. 3.1) from which F_S models are again derived. These iterations are repeated until the regressed coefficients of the F_S model remain constant to three significant figures between iterations. This stabilization usually occurs after about eight iterations.

3.3.4 Model Development Using Least-Squares Regression

Figure 3.5 shows within-event rock residuals, averaged between the results from YA15 and HA15, plotted against V_{S30} for 1.0 s PSA data. Based on visual inspection, we developed a functional form to represent the trends, which consists of a flat region at slow V_{S30}, followed by a negatively sloping region through the center of the data range, followed by another approximately flat region at fast V_{S30}. Similar trends are observed for other intensity measures. The negative slope observed in the central part of the plot is referred to as V_{S30}-scaling, and reflects stronger ground motions on soft sediments as compared to weaker ground motions on stiffer sites for linear site response (e.g., Borcherdt and Gibbs 1976, Seed et al. 1976, Idriss 1990, Borcherdt and Glassmoyer 1994). It is worth emphasizing at this stage that the flat region at slow V_{S30} is a feature of the CENA data that has not generally been observed in active regions, although an exception is Campbell and Bozorgnia (2014), who flattened V_{S30} scaling for slow sites in their regional model for Japan.

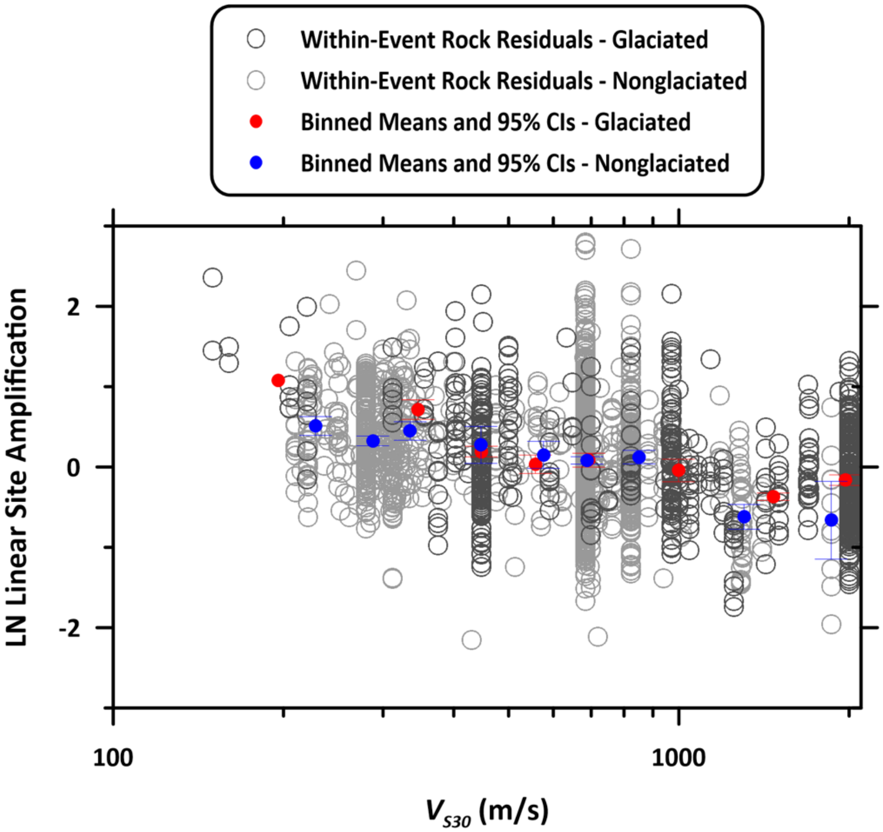

Figure 3.5. Within-event rock residuals for 1.0s PSA shown as a function of V_{S30} for sites with and without prior glaciation per Reed and Bush (2005).

We describe these trends with a piecewise trilinear function, as follows:

$$F_{lin} = \begin{cases} c\ln\left(\frac{V_1}{V_{ref}}\right) & for\ V_{S30} \leq V_1 \\ c\ln\left(\frac{V_{S30}}{V_{ref}}\right) & for\ V_1 < V_{S30} \leq V_2 \\ c\ln\left(\frac{V_2}{V_{ref}}\right) & for\ V_{S30} > V_2 \end{cases} \tag{3.4}$$

where V_1 represents the upper limit of the flat region at slow V_{S30}, V_2 represents the lower limit of the flat region at fast V_{S30}, c is the slope parameter representing V_{S30}-scaling at intermediate velocities, and V_{ref} indicates the reference velocity where F_{lin}=0. The model building process, using Eq. (3.4), began by setting preliminary values of V_1 and V_2 by visual examination. A regression was then performed, which set slope parameter c and reference velocity V_{ref}. The resulting model controls amplification levels over the full range of V_{S30}, including for sites slower than V_1 and faster than V_2. The V_1 and V_2 parameters were then adjusted to improve the fit for the central portion of the data ($V_{S30} \sim$ 300 to 1300 m/s), with re-regression applied with each adjustment.

The model fit developed in this manner does not have reference velocity V_{ref} = 760 m/s, which is needed for compatibility with the underlying GMMs. This occurs because when coupled with F_{lin} from Eq. (3.4) the GMMs are biased with respect to the data, which is not surprising because those GMMs were developed using different site amplification models. As the F_{lin} model is intended to capture changes in ground motion between site conditions, any reference condition could be selected and the ability of the model to capture these changes is unaffected. Accordingly, we shift ε_{Rij} values uniformly (in the vertical direction in Figure 3.5) such that the model is forced to pass through null at 760 m/s (thereby setting V_{ref} = 760 m/s). This adjustment is conceptually similar to subtracting the c_k term in Eq. (3.1) (which was not done), but is applied in the manner described here so as to ensure a reference velocity of 760 m/s.

Figure 3.5 shows that high-velocity sites without prior glaciation have lower average amplification levels than those with prior glaciation. The differences in amplification are likely due to the differences in velocity gradient. Glaciated sites presumably have had weathered geologic materials removed by glacial scour, which leaves relatively competent material overlain by soft materials deposited during or after glaciation. We capture different data trends for glaciated and non-glaciated high-velocity sites by regressing V_2 separately for the two groups while constraining c and V_1 to values obtained from the combined data set. This provides V_{2G} for the previously glaciated sites and V_{2NG} for non-glaciated sites. The regression produced V_{2NG} > 2000 m/s, which exceeds the data range, and hence was set at 2000 m/s. For the considered intensity measures, velocity V_{2G} ranges from 760 to about 1200 m/s.

The above least-squares regression process was performed for amplification calculated using the YA15 and HA15 GMMs, resulting in two amplification models. The two models are identical in terms of data trends and provide coefficients that are statistically indistinguishable. Accordingly, we developed a combined model from ε_{Rij} values averaged from those computed using the YA15- and HA15-based models, and regressed in the manner described above.

Coefficients obtained through the above process were smoothed with respect to PSA oscillator period. A moving weighted average across periods was taken of coefficient c and corner velocities V_1 and V_{2G} using a 5-point triangular window that gives the center point the most weight. Using this scheme, coefficients at the upper and lower end of the period range remain unchanged (Figure 3.6). V_{2NG} is period-independent and was not smoothed.

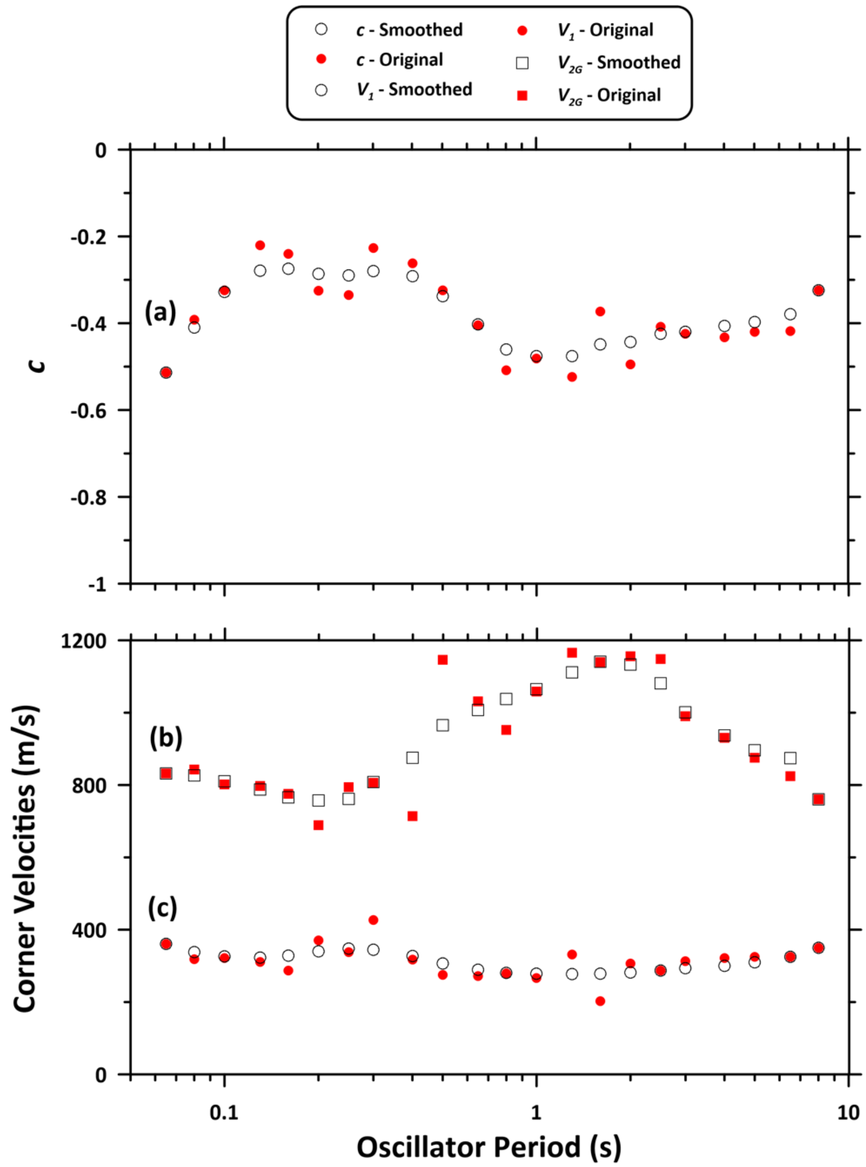

Figure 3.6. An illustration of the coefficient smoothing process for (a) *c*, (b) V_{2G}, and (c) V_1. The original coefficients are shown as red dots, and the final smoothed coefficients, in some cases the result of multiple rounds of smoothing, are shown as open black circles.

3.3.5 Incorporation of Predictor Uncertainty Using Bayesian Analysis

During model development, we were concerned with the impact of uncertainty in V_{S30} values (denoted σ_{lnV}) on the resulting median model and its dispersion. Some stations have large σ_{lnV} – up to 0.6-0.8 in natural log units for glaciated sites with V_{S30} assigned by proxy methods (Parker et al. 2017). There is also a large range of assigned σ_{lnV}; stations with V_{S30} from measurements have assigned σ_{lnV} = 0.1, whereas the range for proxy-based estimates is 0.23-0.85.

Least-squares regression does not account for uncertainty in independent variables, which in this case are the assigned V_{S30} values. To evaluate the significance of this effect, we considered applying orthogonal regression, however this approach requires an assumption of equal measurement uncertainty for all data points, which does not apply in this case. Moss (2011) applied a Bayesian framework to re-fit the NGA-West1 Chiou and Youngs (2008) GMM to the data from which it had originally been derived, while accounting for V_{S30} uncertainty at recording stations. We applied a similar approach.

Application of the Bayesian framework is regression in the sense that a minimized error is found between data and a best-fit model. Unlike least-squares regression, however, the independent variables are treated as inexact, possessing uncertainty due to measurement and V_{S30} proxy model error. The model takes the general form:

$$Z(x,\Theta) = \hat{z}(x,\Theta) + \varepsilon' \tag{3.5}$$

where $\hat{z}$ is the selected model functional form (e.g. Eq. 3.4), x is a vector containing the independent variable (e.g. V_{S30}), Θ is a vector of model parameters (e.g. c, V_1, V_2), and ε' is an error term capturing the imperfect fit of model to data. Error in the independent variable can then be incorporated as $x_i = \hat{x}_i + e_{ix}$, where $\hat{x}_i$ is the assigned value of the variable and e_{ix} is its error (randomly distributed about zero).

The Bayesian framework used to estimate unknown model parameters uses Bayes rule expressed as:

$$f(\Theta) = kL(\Theta) \cdot p(\Theta) \tag{3.6}$$

where $f(\Theta)$ is the posterior distribution, representing the updated state of knowledge about parameters Θ, $L(\Theta)$ is the likelihood function, containing information from observations of x, $p(\Theta)$ is the prior, containing *a priori* information about Θ, and k is a normalizing constant. Bayesian updating involves formulating the likelihood function, selecting a prior distribution, calculating the normalizing constant, and then calculating posterior statistics.

The analysis is performed using a Bayesian analysis code written for Mathworks MatLab included in the appendix to Moss (2009). We modified the original scripts to accommodate our site amplification function (Eq. 3.4) and Robb Moss modified the code to accommodate large datasets via a jack-knife resampling procedure (Tukey 1958). This entails systematically excluding each observation from the dataset, calculating a parameter estimate, and then taking the final parameter estimate as the average of these calculations. We consider uncertainty in site amplifications using within-event standard deviations, taken as the standard deviation of ε values from Eq. (3.3). Standard deviations of V_{S30} values are variable between sites, and are taken from the site database of Parker et al. (2017).

We use an uninformative prior to allow the data to provide the main influence on the posterior estimates. Importance sampling was used to perform integrations over the Bayesian kernel, with the joint lognormal distribution used for the sampling distribution (i.e., coefficients assumed to have log normal distributions). The trial estimates of the sampling distributions took the means as the model coefficients from least-squares regression, and the standard deviations as about equal in magnitude to the means. The analysis was run using these distributions, and then the resulting posterior distribution was used as the initial guess for the next iteration. This was repeated, approximately five times, until the coefficients of variation for the posterior mean estimates were as small as possible, usually between 0.1-0.3. The coefficients were then smoothed in the same manner as described for the coefficients resulting from the least-squares analysis. Figure 3.7 shows the values of model parameters c, V_1, and V_{2G} as a function of period for both the least-squares and Bayesian regressions. The results are similar, which is also what was found in prior work by Moss (2011).

Table 3.2. Model coefficients and the intra-event standard deviation as a function of oscillator period (-1 = PGV).

Period (s)	c	V_{ref} (m/s)	V_1 (m/s)	V_{2NG} (m/s)	V_{2G} (m/s)	$\phi_{\ln,m}$
-1	-0.4486	760	331	2000	760	0.528
0.065	-0.5141	760	361	2000	832	0.703
0.08	-0.41	760	338	2000	826	0.708
0.1	-0.3281	760	326	2000	810	0.678
0.13	-0.2788	760	323	2000	787	0.655
0.16	-0.2748	760	328	2000	766	0.646
0.2	-0.286	760	340	2000	757	0.656
0.25	-0.2901	760	348	2000	762	0.664
0.3	-0.2803	760	344	2000	808	0.657
0.4	-0.2918	760	327	2000	875	0.633
0.5	-0.3377	760	306	2000	965	0.602
0.65	-0.4024	760	289	2000	1007	0.569
0.8	-0.4605	760	281	2000	1038	0.533
1	-0.4762	760	278	2000	1065	0.496
1.3	-0.4758	760	277	2000	1111	0.468
1.6	-0.4493	760	278	2000	1141	0.455
2	-0.4436	760	282	2000	1133	0.458
2.5	-0.4247	760	287	2000	1081	0.464
3	-0.4199	760	294	2000	1001	0.469
4	-0.4061	760	300	2000	937	0.469
5	-0.3971	760	310	2000	896	0.474
6.5	-0.3797	760	325	2000	874	0.482
8	-0.3244	760	350	2000	760	0.482

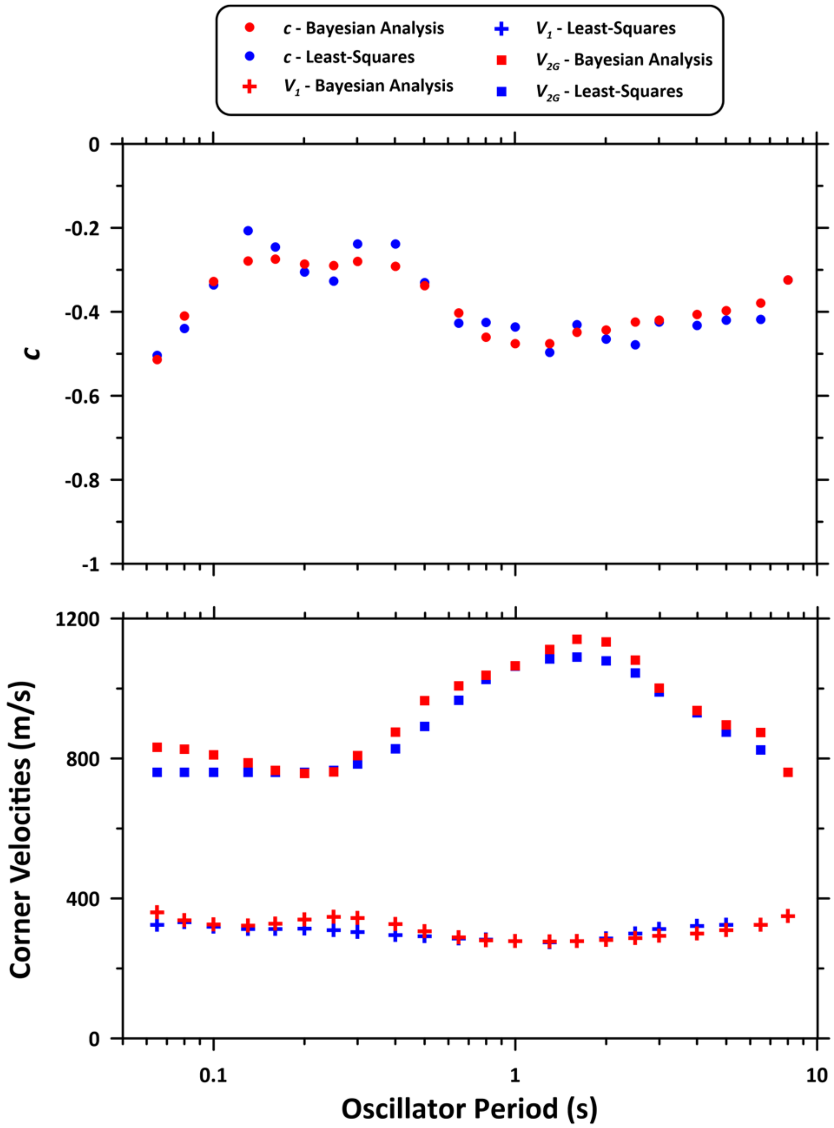

Figure 3.7. Model coefficients c, V_1, and V_{2G} (Eq. 3.5) determined via least-squares regression (blue) and Bayesian analysis (red) and smoothed. Coefficients V_{ref} and V_{2NG} are constants at 760 and 2000 m/s, respectively, and are not shown.

3.4 RESULTS

3.4.1 Model Summary and Attributes

The recommended model for linear site amplification consists of Eq. (3.4) with coefficients listed in Table 3.2. The recommended coefficients are those from Bayesian analysis after smoothing, although there is little difference from the coefficients obtained using least squares regression. The recommended model is shown against the within-event rock residuals data in Figure 3.8 for 0.1, 0.2, 1 and 2.0s PSA.

The within-event standard deviation ($\phi_{\mathrm{ln,m}}$) of the model is an output of Bayesian analysis, and is also given in Table 3.2. By incorporating the V_{S30} errors into the analysis, $\phi_{\mathrm{ln,m}}$ is reduced (up to 23%) from values provided by taking the standard deviation of within-event residuals (denoted ϕ_{ln}). The two standard deviation values are shown in Figure 3.9. The smaller value of $\phi_{\mathrm{ln,m}}$ represents model error if independent variable V_{S30} lacks uncertainty. Uncertainty in V_{S30} (σ_{lnV}) can be combined with $\phi_{\mathrm{ln,m}}$ to estimate the total uncertainty ϕ_{ln} as follows:

$$\phi_{ln}^2 = \phi_{ln,m}^2 + c^2\sigma_{lnV}^2 \tag{3.7}$$

where *c* is the slope from Eq. (3.4) (given in Table 3.2 for velocities between V_1 and V_2, 0 outside of that range), and σ_{lnV} is the uncertainty associated with the V_{S30} value (details on estimation of this uncertainty for NGA-East are given in Parker et al. 2017).

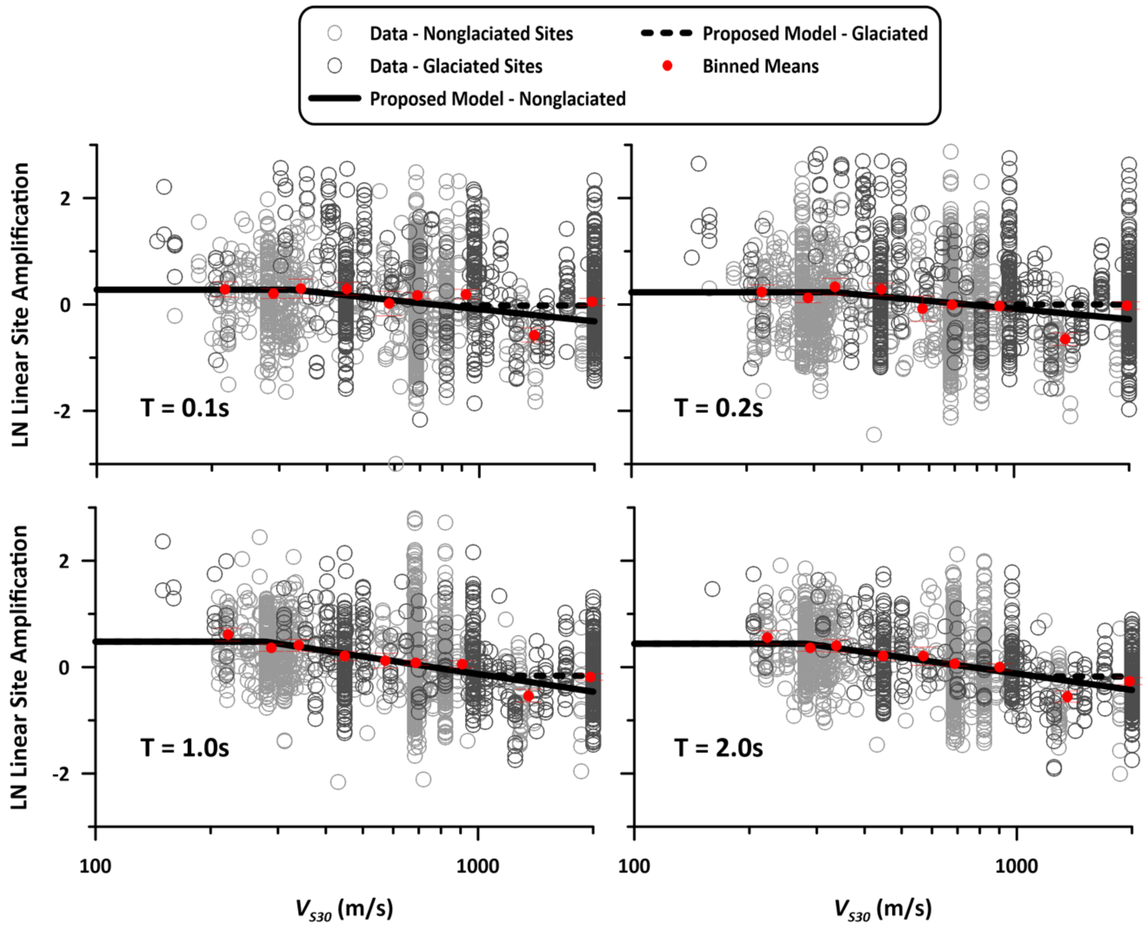

Figure 3.8. Recommended model (black) shown against the computed within-event rock residuals (gray), with binned means and 95% confidence intervals (red) for 0.1, 0.2 1.0 and 2.0s PSA. The models for glaciated (dashed) and nonglaciated (solid) sites deviate at high V_{S30}.

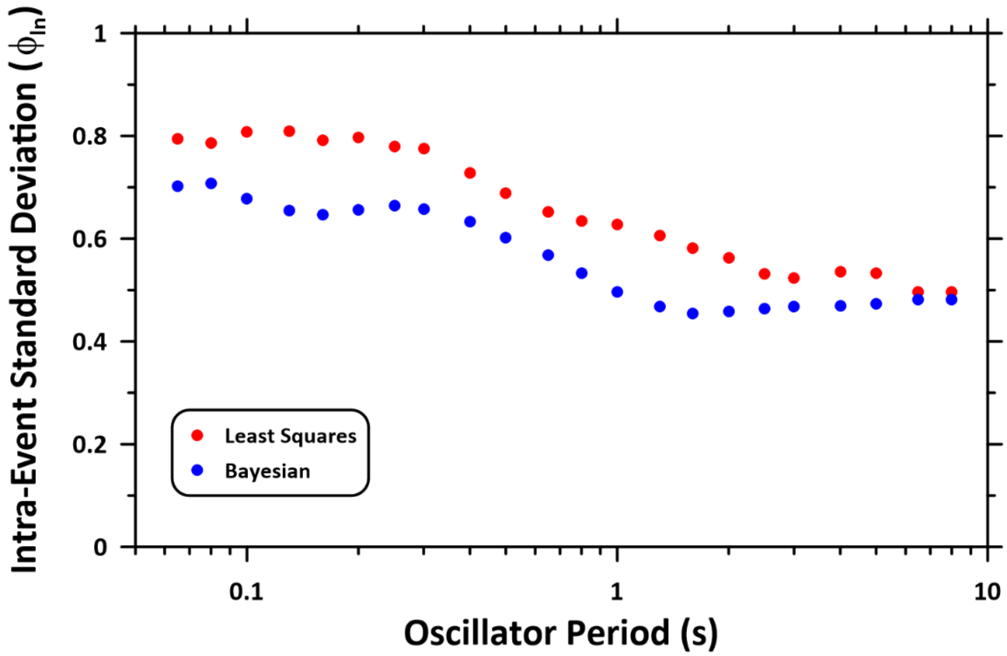

Figure 3.9. The within-event standard deviation, ϕ_{ln}, resulting from the Bayesian analysis as compared to that resulting from least-squares regression.

3.4.2 Comparison to Other Models

Figure 3.10 compares the V_{S30}-scaling coefficient c from the present study for CENA with the corresponding parameter for active tectonic regions by Seyhan and Stewart (2014). Also shown in Figure 3.10 is the approximate range of the c parameter across the major regions contributing NGA-West2 data (Japan, California, Taiwan, China). The differences are significant, with CENA results generally falling outside the range for active regions and having weaker V_{S30}-scaling and relatively little variation as a function of oscillator period. The coefficients estimated using data from CENA are closest to the upper regional bound for active regions, representing the Mediterranean and China (Figure 3.10); however, CENA exhibits shallower scaling than even these regions at PSA oscillator periods 0.15 s and longer.

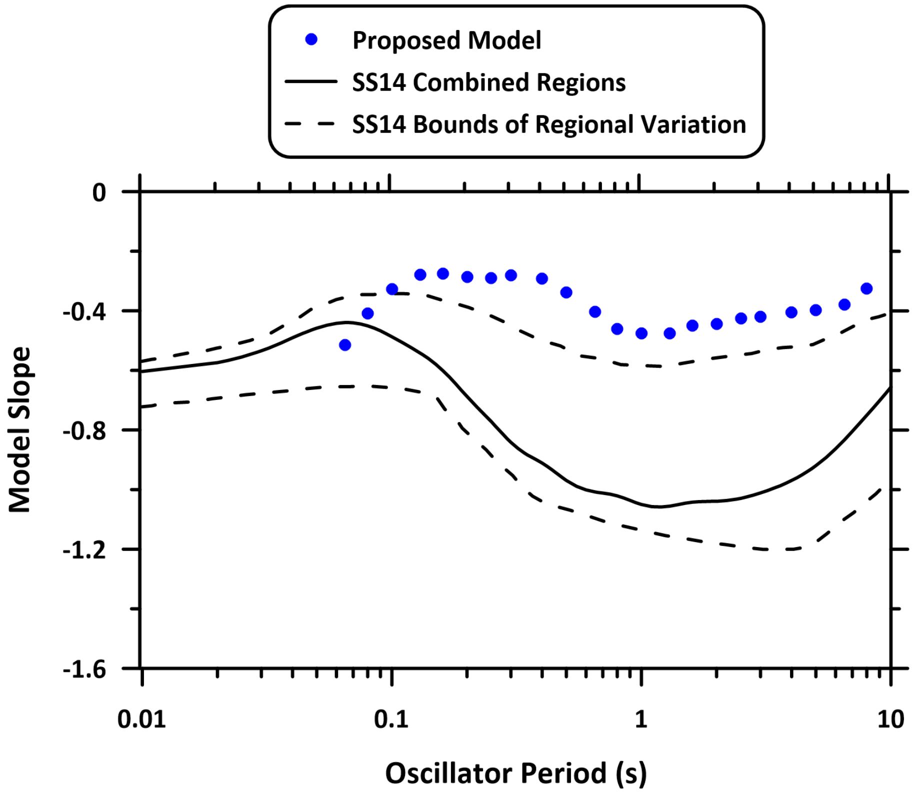

Figure 3.10. Comparison of V_{S30}-scaling coefficient, c (Table 3.2, shown here in blue), with the coefficient from Seyhan and Stewart 2014 (SS14) as a function of oscillator period. The SS14 coefficients are modified from Figure 4 of SS14.

The F_{lin} functional forms for the two models are not the same (lack of flat portion of amplification function for low V_{S30} values in Seyhan and Stewart 2014), so the coefficient comparison in Figure 3.10 does not fully explain the differences in site amplification for these two tectonic regimes. Figure 3.11 compares median amplification for the two regimes (using F_{lin} from Seyhan and Stewart, 2014) for V_{S30} = 200 m/s, V_{S30} = 400 m/s, and V_{S30} = 1000 m/s. CENA results are shown over the applicable period range 0.065 to 8s. The differences in amplification in Figure 3.11 reflect the differences in scaling from Figure 3.10 For soft to moderate soil (V_{S30} = 200 and 400 m/s), CENA amplification is smaller, which is expected given the large offset from the reference condition and the smaller (in an absolute value sense) values of c in CENA. Differences in amplification are smaller for stiffer sites (V_{S30} = 1000 m/s) closer to the 760 m/s reference condition. For all site conditions, the relatively period-independent V_{S30}-scaling in CENA is evident versus stronger scaling features in the 0.3-7.0 s period range for active regions.

Comparisons have also been made between the V_{S30}-scaling model presented herein and the Harmon et al. (2018) simulation-based V_{S30}-scaling model for CENA. There are significant differences at short periods and low V_{S30} values. These comparisons are presented and discussed in an expert panel report to the U.S. Geological Survey (Stewart et al. 2017a).

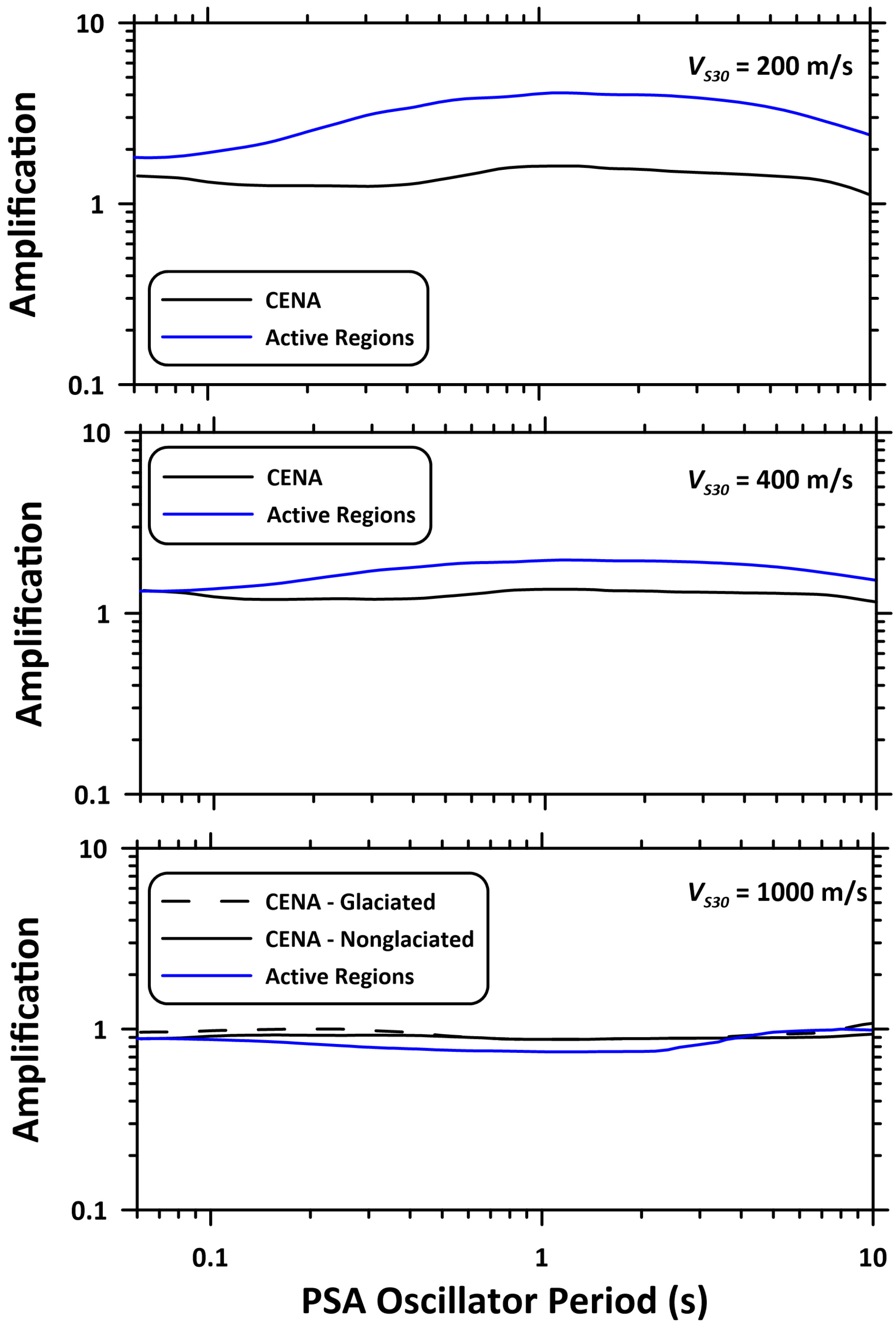

Figure 3.11. Comparison of the predicted amplification from the proposed model for CENA to a model for active tectonic regions (Seyhan and Stewart 2014, SS14). Amplification for V_{S30} = 200 m/s (top), V_{S30} = 400 m/s (middle), and V_{S30} = 1000 m/s (bottom).

3.4.3 Model Residuals

We consider two types of residuals. The first are within-event residuals of the proposed model (Eq. (3.4) with coefficients from Table 3.2) relative to the data, computed as:

$$\delta W_{ij} = \varepsilon_{Rij} - F_{lin} \tag{3.8}$$

These residuals are plotted as a function of V_{S30} in Figure 3.12 for T = 0.1, 0.2, 1.0 and 2.0s PSA using the full data set. The binned means for the model residuals show no appreciable trends with V_{S30}, however the residuals have a negative bias for some individual bins of V_{S30} at low oscillator periods. This results from the coefficient smoothing process. We also considered residuals for several major basins in CENA, including the Mississippi Embayment, the West Atlantic Basin, and the Gulf of Mexico Basin (Coleman and Cahan 2012). Figure 3.13a shows the locations of these basins and Figure 3.13b shows binned mean within-event residuals for each along with results for sites outside of known basin structures. In general, there is minimal bias in the residuals, especially in the central range of oscillator period (0.3-3.0s) where the dataset is most populated (Figure 3.1). Sites in the West Atlantic Basin and the Gulf of Mexico Basin show similar trends, especially at short periods where the binned mean residuals are negative. However, the Atlantic Coast Basin shows the largest bias and uncertainty at long periods, most likely due to scarcity of data (< 20 sites for periods > 3s). Sites in the Mississippi Embayment show a small positive bias across the majority of periods. As expected, sites located outside of mapped basins as defined by Coleman and Cahan (2012) show minimal trend and bias, with small uncertainty due to the large dataset. We recognize that there could be trends of residuals with depth within these basins, but we are unable to assess such features because basin depths are not available for accelerograph sites that provided data used in this analysis.

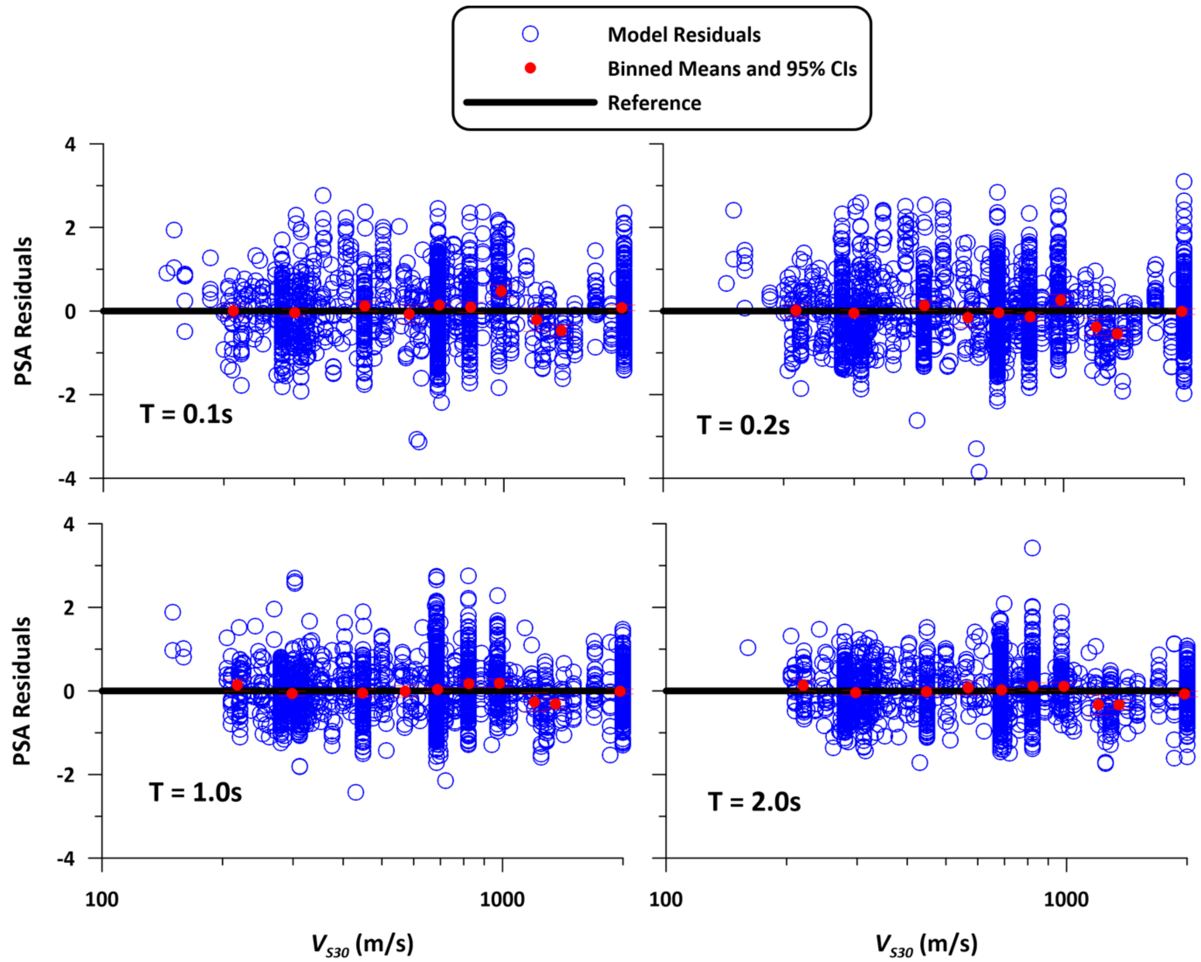

Figure 3.12. Model within-event residuals (open circles) from entire dataset for 0.1, 0.2, 1.0 and 2.0s PSA. Binned means and 95% confidence intervals are shown as closed circles, and a reference line at zero.

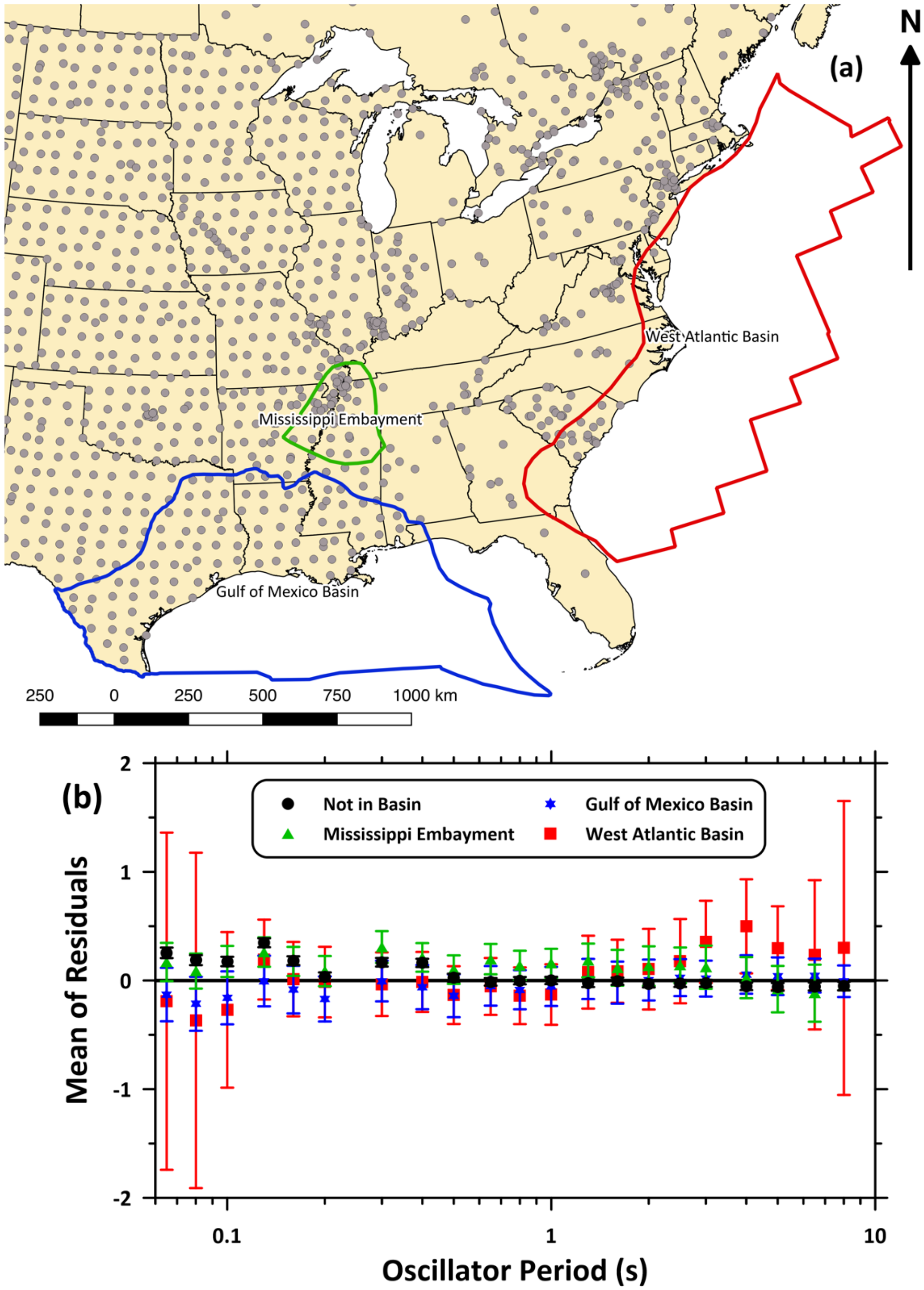

Figure 3.13. (a) Locations of the Mississippi embayment (green), the West Atlantic Basin (red), and the Gulf of Mexico Basin (blue) from Coleman and Cahan (2012), with ground motion recording stations used in this study shown in gray. (b) Binned mean and 95% confidence intervals for sites located in the above basins as well as all sites outside of known basins (Coleman and Cahan 2012) shown as a function of oscillator period.

The second set of residuals is derived using GMMs combined with the proposed site terms (Eq. 3.4). These residuals are computed using Eq. (3.2) and then partitioned using Eq. (3.3). The between-event residuals ($\eta_{E,i}$) for PGV are shown as a function of **M** in Figure 3.14 and show no discernable trend, which is consistent with results presented in the documentation of the GMMs. Figure 3.15 shows that within-event PGV residuals (δW_{ij}) are unbiased and have minimal trend with rupture distance (R_{RUP}). An exception is bias at close distance (< 10 km), which is a feature of the GMMs that is not vital for the present application for estimating site effects using a non-reference site approach, due to the small number of observations that are affected. This check is particularly important to demonstrate that modification of the site term (relative to what was considered in GMM development) does not appreciably affect the ability of the modified GMM to fit the geometric spreading and anelastic attenuation effects implied by the data. Similar trends, not shown for brevity, were encountered for other intensity measures.

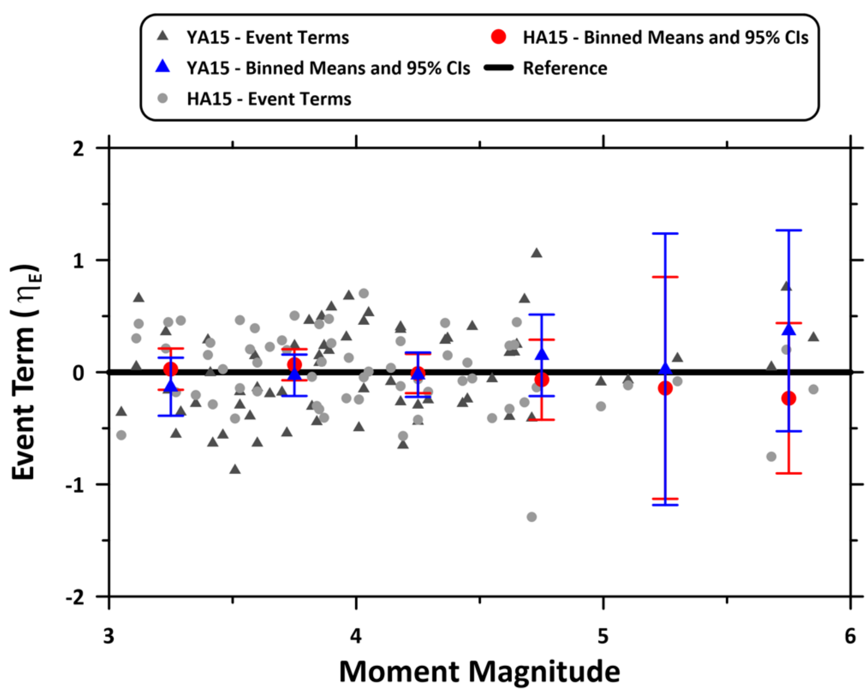

Figure 3.14. Peak ground velocity (PGV) event terms ($\eta_{E,i}$) for Yenier and Atkinson 2015 (YA15) and Hassani and Atkinson 2015 (HA15) shown as a function of moment magnitude (**M**), with binned means and 95% confidence intervals on the mean.

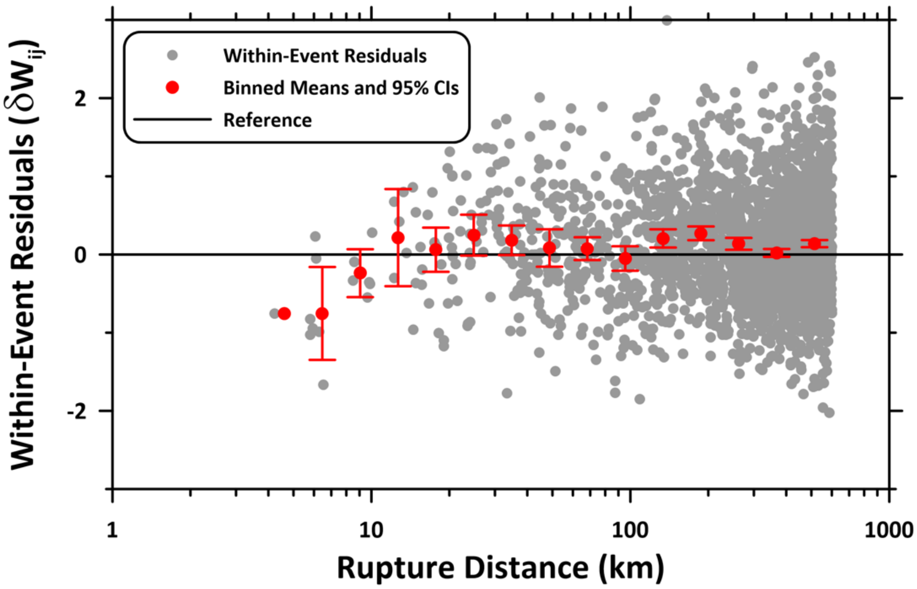

Figure 3.15. Within-event residuals (δW_{ij}) as a function of rupture distance for peak ground velocity (PGV).

The within-event model residuals (δW_{ij}) were partitioned using mixed-effects analysis into site terms ($\eta_{S,j}$) and the remaining within-event single station residual (ε_{ij}) according to Eq. (3.9). The standard deviations of site terms represent site-to-site variability (ϕ_{S2S}) and the standard deviation of the remaining residuals is the single station within-event variability (ϕ_{SS}).

$$\delta W_{ij} = \eta_{S,j} + \varepsilon_{ij} \tag{9}$$

The ϕ_{S2S} values for sites in CENA, computed using YA15, the F_{lin} model presented herein, and the screened dataset from the *Ground Motion Database* section with an additional criterion that each site has greater than three recordings, are plotted as a function of oscillator period in Figure 3.16a. Also shown in Figure 3.16a are ϕ_{S2S} terms computed by Goulet et al. (2017) using the NGA-West2 data (Ancheta et al. 2014) and Japanese data (Dawood and Rodriguez-Marek 2016) for magnitudes less than 5. The ϕ_{SS} values for CENA computed in this study are shown in Figure 3.16b to be comparable to the CENA-specific, magnitude-independent ϕ_{SS} model presented in Goulet et al. (2017).

The results in Figure 3.16a indicate that all three regions have comparable ϕ_{S2S} values at small magnitudes. This was not necessarily expected, as the weaker V_{S30}-scaling in CENA indicates that the model captures less of the site response than in other regions, and could be expected to produce higher ϕ_{S2S} values. The lack of significant differences in ϕ_{S2S} suggests that the effectiveness of V_{S30}-scaling models for the respective regions in describing site-to-site variability in ground motions are comparable. For CENA and elsewhere, such variability can be reduced by adding resonant peaks to the amplification function near a site frequency (e.g., Hassani and Atkinson 2017; Kwak et al. 2017), which was not considered in the present study due to lack of independent information on peak frequencies at recordings sites.

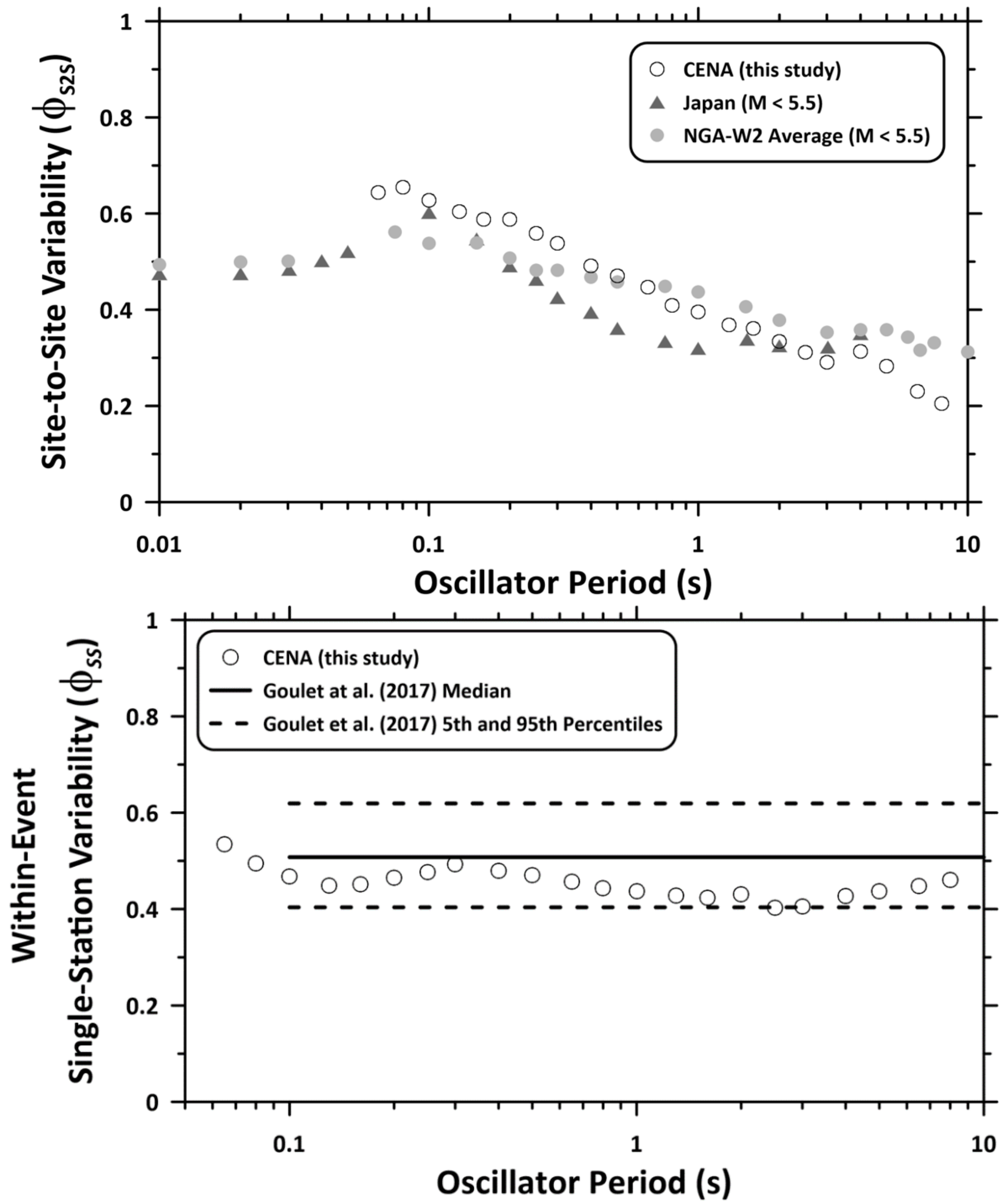

Figure 3.16. (a) Comparison of site-to-site variability, ϕ_{S2S}, for small magnitude events from CENA (this study), the NGA-W2 dataset (using the BSSA14 GMM), and Japanese data (Dawood and Rodriguez-Marek 2016). Values of ϕ_{S2S} for the latter two regions are taken from Goulet et al. (2017); (b) a comparison of the within-event single-station variability (ϕ_{SS}) computed in this study with the NGA-East period and magnitude-independent ϕ_{SS} model (Goulet et al. 2017).

Figure 3.17 shows c_k as a function of period, which represents GMM bias. The bias terms are modest but non-zero for the YA15 and HA15 GMMs, which has implications for GMM utilization as described further in *Summary and Recommendations*.

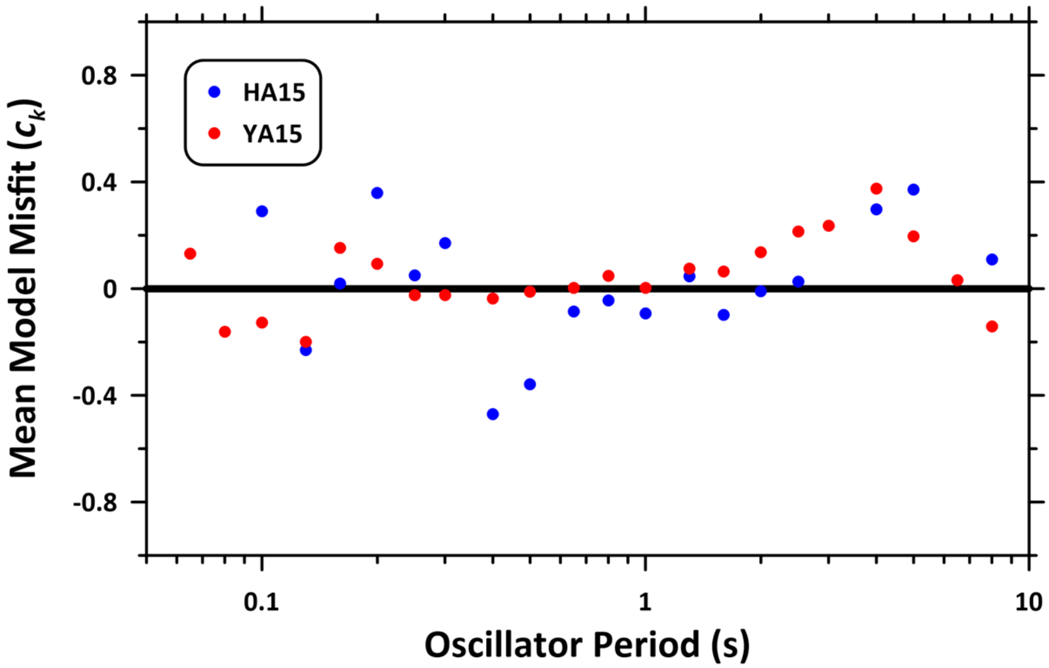

Figure 3.17. Ground motion model bias, c_k, as a function of oscillator period.

3.4.4 Model Validation Using B_SGD02 GMM

The model validation using the B_SGD02 GMM from Boore (2015a) (as described in *Model Development*) consisted of:

1. Computing a new set of within-event rock residuals ($\square_{Rij}$) derived using reference ground motions from B_SGD02, and a dataset screened using criteria consistent with the applicability of B_SGD02,
2. Plotting the resulting residuals against V_{S30}, as shown in Figure 3.18,
3. Checking the performance of the model (derived using different GMMs) against the new observations.

Based on the results in Figure 3.18 for 1.0s PSA, we find misfits in some individual velocity bins, but the model globally provides a visually reasonable fit to the data. This finding, along with the similarity of coefficients derived using the HA15 and YA15 GMMs, suggests that reasonable variations in distance attenuation functions do not appreciably change the resulting site amplification model.

Figure 3.17 does not show c_k values for B_SGD02 as Boore (2015a) does not recommend using this GMM for forward prediction, and therefore the mean misfit of the GMM with an updated site term is not pertinent.

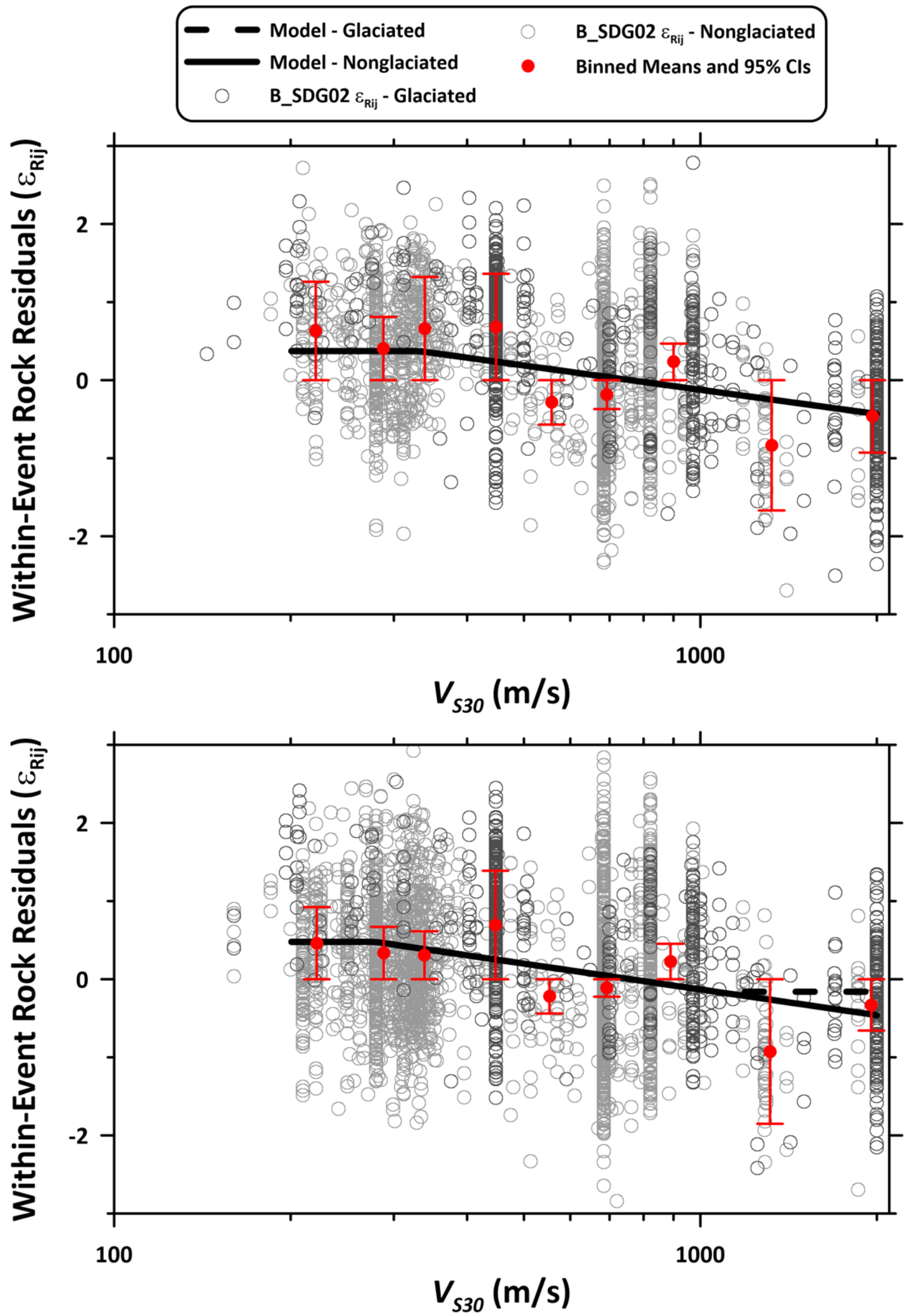

Figure 3.18. Validation of the proposed model using within-event rock residuals computed with Boore 2015a (B_SDG02). Within-event rock residuals (ε_{Rij}) for PGV (top) and 1.0s PSA (bottom).

3.4.5 Model Limitations

The model presented here applies for small-strain, linear conditions associated with weak ground motions. As such, it does not account for soil nonlinear effects that would be significant for soft sites and at short oscillator periods (< ~ 1.0 s). Moreover, this model does not include the effects of sediment depth or site period, which can significantly affect observed amplification levels (Hassani and Atkinson 2016a, 2017) and the results of simulations for CENA site conditions (Harmon et al. 2019b). As described in the *Introduction*, the model presented here can be combined (in a natural log additive sense) with simulation-based models for nonlinearity. The model can also likely be used with models for resonance effects near a site frequency, although we have not tested this.

The present model is ergodic, and as such cannot account for site-specific features related to specific geologic conditions that may appreciably affect ground motions at a particular site. As a result, for important projects we encourage non-ergodic site response modeling (e.g., Stewart et al., 2017b) that can account for these effects. These effects are not provided by the suite of GWG models, of which this chapter presents one component.

The proposed model is applicable for PGV and PSA oscillator periods from 0.065 to 8s. The range of V_{S30} is 200 m/s to 2000 m/s. When used with GMMs that provide ground motions at a reference site condition of 3000 m/s, an adjustment for the 760 to 3000 m/s reference conditions is needed – recommended models for this are provided in Stewart et al. (2017a).

3.5 SUMMARY AND RECOMMENDATIONS

Although there has been significant previous work on rock-conditioned GMMs for stable continental regions including CENA, models of site amplification for soil and weathered rock site conditions are limited. Most previously available models are simulation-based using CENA site conditions or are empirically derived for active regions and adopted for CENA with limited validation. We present linear site amplification models intended for use with GMMs developed in the NGA-East project for reference rock site conditions corresponding to CENA hard rock (V_S = 3000 m/s; Hashash et al. 2014) or the NEHRP B/C boundary (V_{S30} = 760 m/s). Our models apply over the V_{S30} range 200-2000 m/s and for the intensity measures of PGV and PSA at oscillator periods between 0.065 and 8 seconds. The proposed model is given in Eq. (3.4) with the coefficients in Table 3.2. The value of $\phi_{\mathrm{ln,m}}$ in Table 3.2 from Bayesian regression does not include effects of independent parameter uncertainty, which can be incorporated for evaluation of ϕ_{lln} using Eq. (3.7). The model can be combined with models for the nonlinear component of site response derived from simulations (Harmon et al. 2019b), and potentially with additional terms that account for resonance effects near a site frequency (Harmon et al. 2019b, Hassani and Atkinson 2017). Neither of these model combinations has been tested using CENA data, although the use of V_{S30}-scaling and nonlinear models is a well-established (and validated) practice in active regions (e.g., Seyhan and Stewart, 2014).

The proposed model demonstrates distinct site amplification features from active regions, with weaker V_{S30}-scaling and less variation in amplification with period. For this reason, we recommend against applying active region models in CENA, which to this point has been a common practice.

There is an important caveat to the joint use of the GWG site amplification models (the model presented, along with Harmon et al. 2019b for effects other than V_{S30}-scaling) with NGA-East GMMs, which is that the GMMs were derived with different site models. When those GMMs are used with the present model, bias relative to observation is encountered. Such effects can be accounted for through modification of the constant term in the reference rock GMMs. The approximate magnitude of this adjustment is represented by the bias terms in Figure 3.17 for the YA15 and HA15 models, although there is considerable epistemic uncertainty regarding this term for the strong shaking conditions that will typically control hazard. The c_k values presented for YA15 and HA15 do not necessarily apply for other GMMs. Accordingly, this potential bias should be checked and incorporated into model predictions.

4 EXPERT PANEL RECOMMENDATIONS FOR ERGODIC SITE AMPLIFICATION IN CENTRAL AND EASTERN NORTH AMERICA

4.1 INTRODUCTION

The Next Generation Attenuation – East (NGA-East) project produced ground motion models (GMMs) for central and eastern North America (CENA) (PEER 2015a, b, and Goulet et al., 2017). The majority of these models provide ground motion intensity measure predictions as a function of earthquake source and wave propagation path for sites with a hard-rock reference condition defined as shear-wave velocity V_s = 3000 m/s and site decay parameter κ_0 = 0.006s (Hashash et al. 2014). Some of those models also provide ground motions for the National Earthquake Hazards Reduction Program (NEHRP) B/C boundary condition of V_{S30} = 760 m/s, where V_{S30} is the time-averaged shear-wave velocity in the upper 30 m of the site.

The United States Geological Survey (USGS) National Seismic Hazard Maps present ground motion intensity measures with specified probabilities of exceedance over a 50-year time period (Petersen et al. 2015). As of this writing, a major update of these maps is underway that is utilizing NGA-East GMMs for the CENA region (Petersen et al. 2018). A special consideration for this update is that maps are being produced for a variety of site conditions (represented by a range of V_{S30}) and periods, as a result of recommendations from Project 17 [M. Petersen, *pers. communication*, July 2016]. This is a departure from past practice in which the maps were produced for the NEHRP B/C boundary site condition (V_{S30}=760 m/s) and the ground motion measures of peak acceleration and 5% damped pseudo-spectral acceleration at oscillator periods of 0.2 and 1.0s.

An expert panel (comprised of Jonathan Stewart, Gail Atkinson, David Boore, Youssef Hashash, Walter Silva, and Robert Darragh) was convened in 2016 with a charge to review alternate site amplification models for CENA and to provide recommendations to the USGS and other interested parties regarding estimation of median site effects and their epistemic uncertainties. This work required that recommended models be based on V_{S30} as the sole predictive variable for site response, for compatibility with the NEHRP site categories A-E used in current practice (which are defined for ranges of V_{S30}). The consideration of models conditioned on alternative or additional parameters such as depth or dominant site period was beyond our scope; all authors agreed that such alternative models should be included for CENA in future code updates. The panel developed initial recommended procedures that were presented in two reports in June 2017 (Stewart et al. 2017a; Hashash et al. 2017). As the USGS implemented these models, feedback was provided to the panel (both from USGS scientists and via public comment), which resulted in several adjustments. This chapter presents models ultimately recommended by the panel and implemented for the national maps by USGS, including adjustments since June 2017. We explain the reasoning behind the model formulation and the definition of uncertainties. The emphasis here is on the linear components of the model, which presented the principle technical challenges. The nonlinear component of the model and its uncertainty are given in a companion paper (Hashash et al. 201x), which updates a prior report (Hashash et al. 2017).

4.2 PRIOR WORK

4.2.1 Empirical Site Amplification Studies

Empirical site amplification studies, while numerous and well-established in some active tectonic regions, are a relatively recent development in stable continental regions like CENA. This is due to a number of factors, including a lack of V_{S30} information at seismographic sites in CENA (addressed in NGA-East by the development of a regional, proxy-based V_{S30}-prediction model; Parker et al. 2017). Parker et al. (2019) present an empirical linear site amplification model, conditioned on V_{S30}, that was developed by the NGA-East Geotechnical Working Group (GWG). Hassani and Atkinson (2016a) derived the frequency of peaks in H/V spectral ratios using CENA data, and used those peak frequencies as predictive parameters for analysis of site effects. They find that the data-derived peak frequencies are more effective than V_{S30} at predicting site effects in the CENA data. Additional literature review on CENA empirical site amplification is presented by Parker et al. (2019), Chapter 3.0 herein. The panel considered the Parker et al. (2019) and Hassani and Atkinson (2016a) empirical models.

4.2.2 Simulation-Based Site Amplification

As a result of limited empirical site amplification studies for the reasons described above, previous work in CENA has largely investigated site amplification using simulations of wave propagation through shallow sediments. The panel considered three studies (or collections of studies) for CENA. The first was by Hwang et al. (1997) and was targeted at the CENA region generally. They computed site coefficients akin to those for the NEHRP Provisions for CENA using equivalent-linear ground response simulations with simulated input motions generated using the method described in Hwang and Huo (1994). They considered five representative profiles for NEHRP site classes A-E (profiles shown in Lin et al. 1996). Their results for site classes A and B (rock sites) match those in the 1992 NEHRP Provisions. Site factors for Classes C-E are generally higher. Figure 4.1(a) shows their recommended site amplification for Classes C-E for a rock peak acceleration level of 0.3*g*, and Figure 4.1(b) shows the variation of Class D amplification with shaking amplitude.

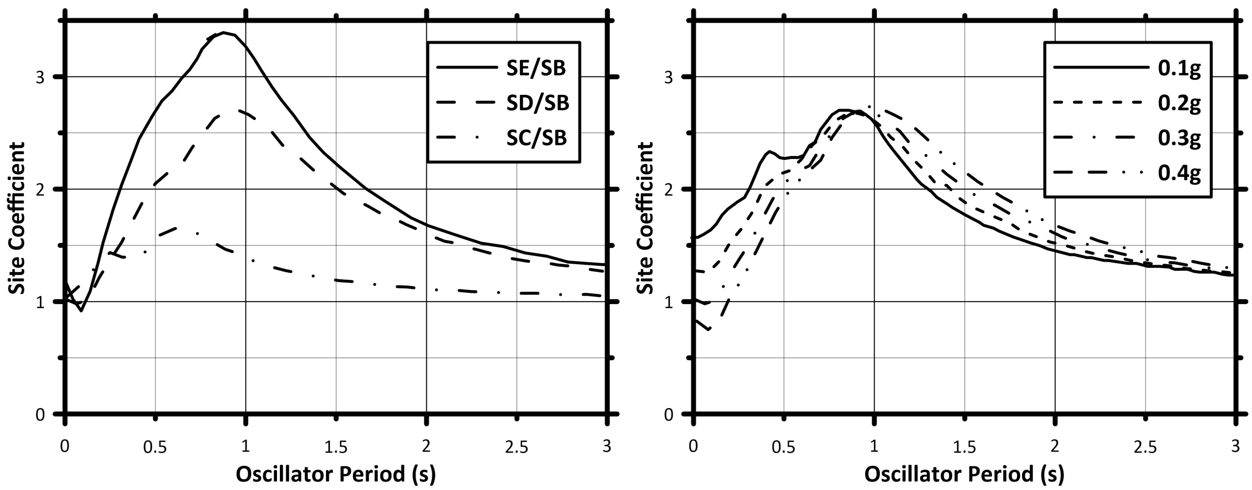

Figure 4.1. (a) Computed CENA site amplification by Hwang et al. (1997) for NEHRP classes C, D, and E relative to a site class B condition for rock peak acceleration 0.3g; (b) Dependence of computed amplification for class D on rock peak acceleration.

The second group of studies evaluated site effects for the Mississippi embayment region (Hashash and Park 2001; Romero and Rix 2001; Park and Hashash 2005a; Park and Hashash 2005b; and Hashash et al. 2008). The literature for this region is substantial and has arguably been supplanted by more recent work by the NGA-East GWG as presented in Harmon et al. (2019a,b). The GWG study considered a wide variety of site conditions and input motions, and used fully nonlinear ground response simulations. Models were provided for V_{S30}-scaling, other linear effects, and nonlinearity.

The third CENA study is from Aboye et al. (2015), who developed site factors for Charleston, South Carolina. They developed a series of reference V_S profiles assuming different Quaternary layer thicknesses and taking layer velocities from measurements in Quaternary and Tertiary units. After introducing V_S profile variability, they adopt 56 profiles, placed over a half-space with V_S = 700 m/s. They simulated input motions and both equivalent linear and nonlinear ground response simulation methods. Figure 4.2 shows representative results for amplification of 0.2s PSA.

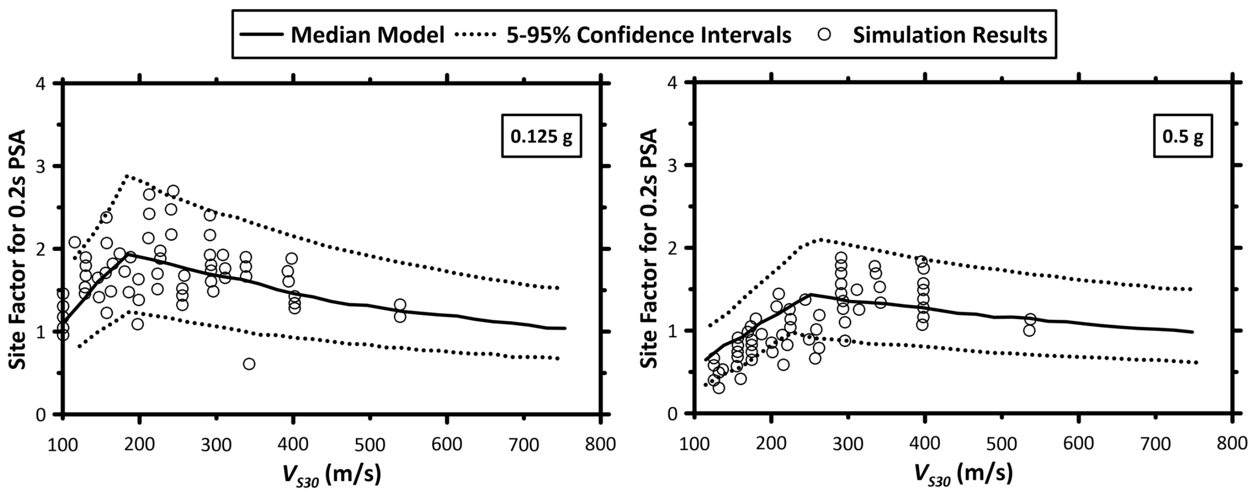

Figure 4.2. Computed amplification of 0.2s PSA for Charleston, South Carolina by Aboye et al. (2015) for input ground motion intensity for rock of 0.2s PSA = (a) 0.125g and (b) 0.5g.

4.3 RECOMMENDED MODEL

4.3.1 Approach

Site amplification relative to a V_S = 3000 m/s reference condition is denoted F_S and is provided in natural log units. The recommended model has three additive components representing: (*i*) V_{S30}–scaling relative to V_{S30}=760 m/s, (*ii*) amplification at the V_{S30}=760 m/s site condition relative to 3000 m/s, and (*iii*) nonlinear effects. The first two components are independent of shaking amplitude, and hence are described as linear and are denoted F_{lin}. The nonlinear component is denoted F_{nl} and is also in natural log units. The total amplification is the sum:

$$F_S = F_{lin} + F_{nl} \tag{4.1}$$

where

$$F_{lin} = F_V(V_{S30}, T) + F_{760}(V_{S30}, T) \tag{4.2}$$

where F_V is the V_{S30}-scaling term and F_{760} represents amplification at the V_{S30} = 760 m/s site condition relative to a 3000 m/s reference condition. Recommended median models for F_V and F_{760} are given in the following sub-sections along with their epistemic uncertainties. Justification for the selection of these models is given in later sections of this chapter. Hashash et al. (201x) present the model for nonlinear effects and related uncertainties. Equation 4.2 is suitable for use with a GMM having a reference condition of V_S = 3000 m/s. It can be extended to a reference condition of V_{S30}=760 m/s by dropping the F_{760} term.

For the F_V term the recommended model is largely controlled by empirical observations (NGA-East ground motion data), and for the F_{760} and F_{nl} terms it is controlled by simulations. The rationale for this approach is discussed in the *Summary and Discussion* section of this chapter.

4.3.2 V_{S30}-Scaling Model

The V_{S30}-scaling model is quad-linear in log-log space, as given below:

$$
F_V = \begin{cases} c\ln\left(\frac{V_1}{V_{ref}}\right) & V_{S30} \le V_1 \\ c\ln\left(\frac{V_{Ss30}}{V_{ref}}\right) & V_1 < V_{S30} \le V_2 \\ c\ln\left(\frac{V_2}{V_{ref}}\right) & V_2 < V_{S30} \le V_u \\ c\ln\left(\frac{V_2}{V_{ref}}\right) - \left[c\ln\left(\frac{V_2}{V_{ref}}\right) + F_{760}\right] ^{ln\left(\frac{V_{S30}}{V_u}\right)} \Big/ {ln\left(\frac{3000}{V_u}\right)} & V_u < V_{S30} \le 3000 m/s \end{cases} \quad (4.3)
$$

The third term is constant at the amplification for 2000 m/s. The model form is shown in Figure 4.3. Term c represents the slope in log-log space for the central portion between corner velocities V_1 and V_2. Velocities V_ℓ and V_u represent the approximate lower and upper limits of the range constrained by observations. Velocity V_{ref} is taken as 760 m/s; its physical meaning is the velocity at which $F_V = 0$. The model is flat (constant F_V) for $V_{S30} < V_1$ and $V_2 < V_{S30} < V_u$. The last interval of the model represents interpolation between constrained amplification levels at V_u and 3000 m/s, the latter being $-F_{760}$ as shown in Figure 4.3. Model coefficients c, V_1, and V_2 are oscillator period-dependent. The coefficients are plotted as a function of period in Figure 4.4 and are tabulated in the electronic supplement. The basis for the proposed V_{S30}-scaling model is described in the *F_V Model Development* section below.

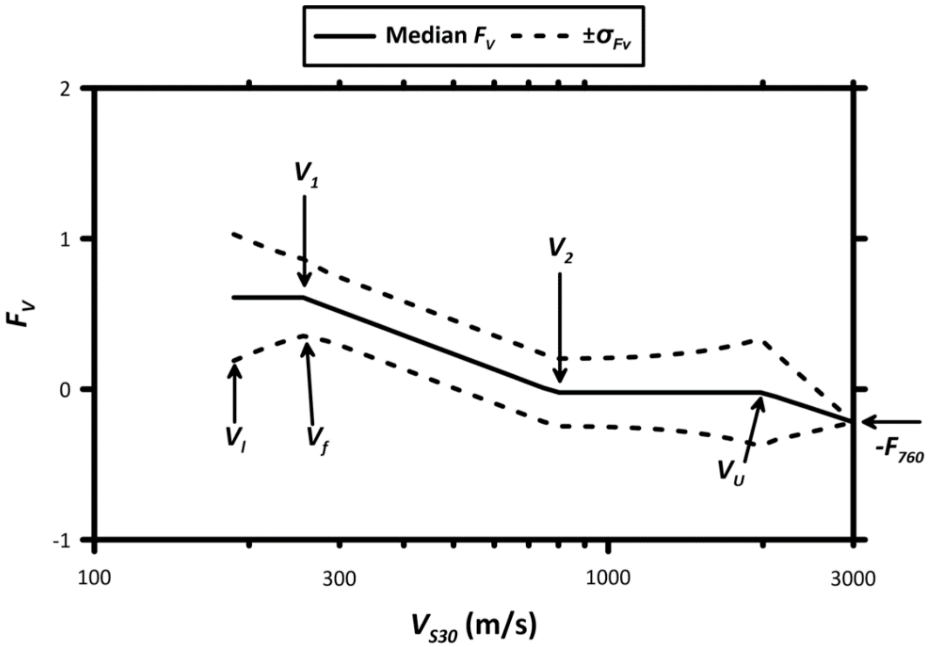

Figure 4.3. Form of recommended median V_{S30}-scaling model (Eq. 4.3) and the associated uncertainty (Eq. 4.4) for 1.0s oscillator period. Coefficients in electronic supplement Table E1.

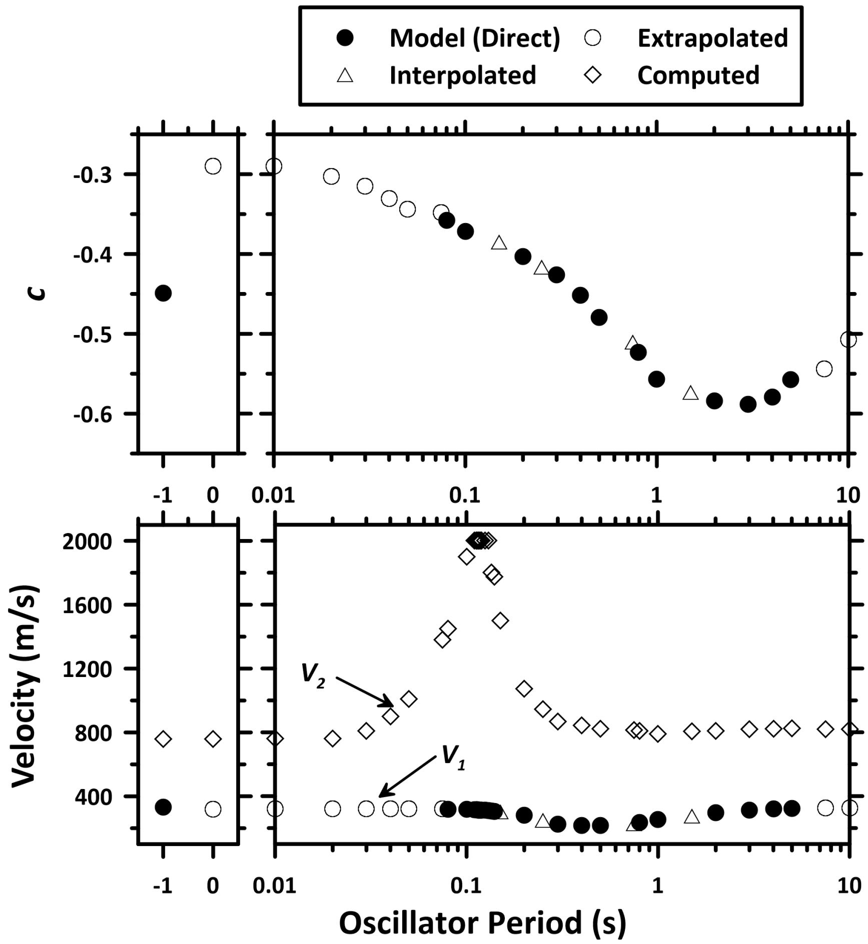

Figure 4.4 Period-dependence of coefficients in F_V model. Coefficients that are interpolated, extrapolated, and computed using simulations as a guide are indicated separately from those developed from data and model inferences

The epistemic uncertainty associated with the model is given by a log-normal standard deviation σ_v that is constant over the middle portion of the V_{S30} range (between V_f and V_2) and increases at the low- and high-velocity limits of the model, as shown in Figure 4.3. σ_v is given by:

$$\sigma_v = \begin{cases} \sigma_\ell - 2(\sigma_\ell - \sigma_{vc})\frac{V_{s30}-V_\ell}{V_f-V_\ell} + (\sigma_\ell - \sigma_{vc})\left(\frac{V_{s30}-V_\ell}{V_f-V_\ell}\right)^2 & V_\ell < V_{s30} < V_f \\ \sigma_{vc} & V_f \le V_{s30} \le V_2 \\ \sigma_{vc} + (\sigma_u - \sigma_{vc})\left(\frac{V_{s30}-V_2}{V_u-V_2}\right)^2 & V_2 < V_{s30} < V_u \\ \sigma_u\left(1 - \frac{ln\left(V_{s30}/V_u\right)}{ln\left(3000/V_u\right)}\right) & V_u < V_{s30} < 3000m/s \end{cases} \quad (4.4)$$

The coefficients for the uncertainty model are the uncertainty in the central portion of the velocity range (σ_{vc}), the increased uncertainty ($\sigma_\ell - \sigma_{vc}$) at the lower-limit velocity for the model (V_ℓ), and the increased uncertainty ($\sigma_u - \sigma_{vc}$) at the upper-limit velocity (V_u). Velocity V_f is specific to the uncertainty model and velocities V_2 and V_u are the same as for the median model. These and other coefficients are given in the electronic supplement

4.3.3 F_{760} Model

The F_{760} model modifies ground motion intensity measures from the NGA-East reference condition of $V_S = 3000$ m/s to $V_{S30} = 760$ m/s as a function of oscillator period. The recommended model is a weighted combination of two models derived from simulations using two groups of velocity profiles (each with $V_{S30} = 760$ m/s) characterized by large impedance contrasts and velocity gradients. The resulting amplification models are denoted F_{760}^{imp} and F_{760}^{gr}, respectively. Figure 4.5 shows the median models and their epistemic uncertainties, σ_{lnF760}.

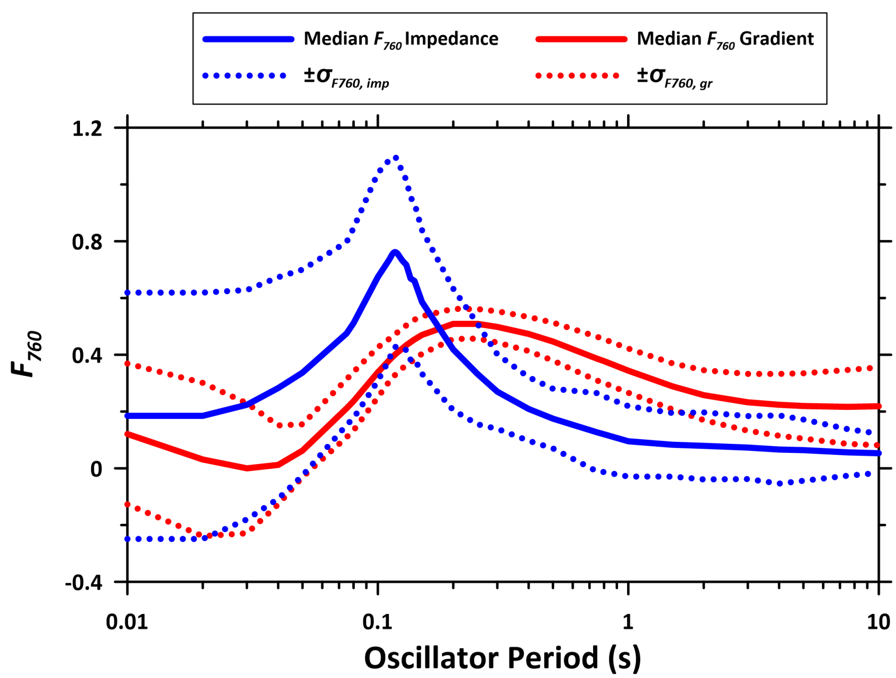

Figure 4.5. Reference condition site factors, F_{760}, for impedance and gradient conditions, and the associated uncertainties as a function of oscillator period.

The recommended model for F_{760} is given as:

$$F_{760}(V_{s30}, T) = w_{imp}(V_{s30}) F_{760}^{imp}(T) + w_{gr}(V_{s30}) F_{760}^{gr}(T) \tag{4.5}$$

The weights are a function of V_{S30}. Sites with a $V_{S30} \geq V_{w1}$ receive a high weight (w_{imp}) to the F_{760}^{imp} model, and sites with $V_{S30} < V_{w2}$ receive a high weight (w_{gr}) to the F_{760}^{gr} model. The weights taper between the models for velocities between V_{w1} and V_{w2},

$$w_{imp}(V_{s30}) = \begin{cases} w_1 & for\ V_{s30} \geq V_{w1} \\ 1.97 \cdot ln\left(\frac{V_{s30}}{V_{w2}}\right) + w_2 & for\ V_{w2} \leq V_{s30} < V_{w1} \\ w_2 & for\ V_{s30} < V_{w2} \end{cases} \tag{4.6}$$

$$w_{gr} = 1 - w_{imp} \tag{4.7}$$

At each value of V_{S30}, weights w_{imp} and w_{gr} sum to 1.0. Coefficients tabulated in the electronic supplement include the median models (F_{760}^{imp} and F_{760}^{gr}), standard deviations σ_{lnF760}, weight transition velocities V_{w1} and V_{w2}, and weights w_1 and w_2. Justification for the proposed model is given in the F_{760} *Model Development* section of this chapter.

4.4 F_V MODEL DEVELOPMENT

4.4.1 Models Considered

The proposed model for V_{S30}-scaling (F_V) is based upon results from prior research. Here we describe how results for selected models were adapted for the model-to-model comparisons and explain why certain models were not selected.

We consider two empirical models: (1) a model relating site amplification to peak frequency (f_{peak}) from horizontal to vertical spectral rations using NGA-East data for CENA (Hassani and Atkinson 2016a); and (2) an empirical V_{S30}-scaling model developed by the NGA-

East Geotechnical Working Group (referred to subsequently as GWG-E; Parker et al. 2019). Additional empirical models that were considered but ultimately not used are Hollenback et al. (2015), Al Noman and Cramer (2015), and Graizer (2015). The site effects model for two Hollenback et al. (2015) GMMs was developed in Fourier amplitude space, which is not readily applicable to response spectral ratios. The GMMs developed by Al Noman and Cramer (2015) and Grazier (2015) were not considered ready to be used as seed models over a wide frequency range (Goulet, personal communication, 2017), and hence were not used. Upon the completion of the panel's analysis work, a new model was published (Hassani and Atkinson, 2017). Because the V_{S30}-scaling in that model is similar to GWG-E, a renewal of panel activity to formally consider the Hassani and Atkinson (2017) model was considered unnecessary.

The Hassani and Atkinson (2016a) model provides a variation of amplification that is peaked at site peak frequency f_{peak} (i.e, amplification tapers down for frequencies lower and higher than f_{peak}). To apply this model, we convert f_{peak} to V_{S30} using a relationship between these site parameters as given by Hassani and Atkinson (2016b). Values of f_{peak} corresponding to four values of V_{S30} (one in each NEHRP category D-A) were derived as follows: 270 m/s – 2.33 Hz, 560 m/s – 7.41 Hz, 1170 m/s – 23.8 Hz, and 2032 m/s – 57.3 Hz. Tabulated amplification values (provided by B Hassani, personal communication, 2016) were then used to estimate the site term for each approximate V_{S30}. The Hassani and Atkinson results were shifted vertically so that the average between classes C and B passes through unity (zero in ln units) at 760 m/s. The GWG-E model was used without modification.

We also considered four simulation-based models: (1) Darragh et al. (2015) [also referred to as Pacific Engineering and Analysis, i.e. PEA]; (2) a simulation-based V_{S30}-scaling model developed by the NGA-East Geotechnical Working Group (referred to subsequently as GWG-S;

Harmon et al. 2019b); (3) Hwang et al. (1997); and (4) Aboye et al. (2015). Models (1), (3), and (4) were introduced in the *Prior Work* section. The PEA model uses a reference condition of V_S = 3000 m/s. To apply this model, we adjusted amplification values to a reference condition of V_{S30} = 760 m/s by dividing by F_{760} values given in Darragh et al. (2015). Hwang et al. (1997) present tabulated amplification values for 0.2 and 1.0s PSA for NEHRP categories A-D, which we plot at category mid-velocities (V_{S30} = 1868, 1052, 498, and 243 m/s). The Hwang et al. (1997) results were adjusted to an amplification of 1.0 at V_{S30} = 760 m/s; original results were at 1.0 for class B. We applied the median model from Aboye et al. (2015) as shown in Figure 4.2 for 0.2 and 1.0s PSA. The GWG-S model was provided by J. Harmon (personal communication, 2016) in a form that was already corrected to the 760 m/s reference rock condition.

4.4.2 Model Comparisons and Recommended Median Model

Figure 4.6 present the considered CENA site amplification models for periods of 0.1 and 1.0s. Also shown for comparison is the Seyhan and Stewart (2014) model for active tectonic regions (all periods) and the site factors in the NEHRP provisions for periods of 0.2 and 1.0 s.

One notable feature in the plots is that the GWG-S and Aboye et al. (2015) simulation-based models have downward curvature in the V_{S30}-scaling at short periods ($T \leq 0.3$ s), which is not present in the PEA model. One explanation for the difference in simulation results is different small-strain soil damping formulations. The PEA model is based on equivalent-linear simulations that used strain-dependent 'Peninsular Range' modulus reduction and damping curves (Silva et al. 1997) as well as a subset of the EPRI (1993) curves in the upper 150 m with visco-elastic soil below. At greater depth, the visco-elastic damping was limited so as to not allow the site damping parameter (κ_0) to exceed 0.04s. The linear viscous-elastic simulations in Harmon et al. (2019b) used the small-strain damping ratio (D_{min}) from the Campbell [2009] Q-V_S Model 1 without constraining it according to the resulting surface κ_0. As a result, the GWG-S simulations often have higher levels of profile damping than those of PEA. The physics of wave propagation require increased damping to decrease ground motion, particularly at high frequencies. The panel elected to not incorporate the downward curvature feature in V_{S30}-scaling into the recommended median model, due to this feature not being evident in the GWG-E empirical data.

The Hassani and Atkinson (2016a) model exhibits peaked behavior in amplification-V_{S30} space at the V_{S30} value corresponding to the PSA oscillator period being plotted. For example, in Figure 4.6 (oscillator response for T = 0.1 s, corresponding to f_{peak} = 10 Hz) the model peaks at ~600 m/s. The peaks occur at slower velocities as period increases. This behavior is a consequence of f_{peak} being the sole site parameter in the Hassani and Atkinson (2016a) model; in the implementation of the model for this study, V_{S30} is used as a proxy-measure for f_{peak}, in which stiffer sites (higher V_{S30}) have higher peak frequencies.

The GWG-E model demonstrates relatively flat scaling at both slow ($V_{S30} < V_1$) and fast ($V_{S30} > V_2$) velocities. Both trends are generally supported by the simulation-based models and have different physical explanations. At slow V_{S30} and short periods, the reduction of scaling is likely due at least in part to the effects of soil damping. For longer periods, the cause of the flat scaling at slow V_{S30}, especially as compared to western models (SS14), may be attributable to averaging the effects of peaked response curves over profiles having different average soil depths, which peak near different periods. While sediment depth information at seismograph sites is generally unknown, Parker et al. (2019) investigated bias in the GWG-E model for sites in particular basins, and found no systematic features that would justify adjustment to the model. At fast V_{S30}, the reduction of scaling is thought to be caused by the reduced predictive power of V_{S30} as a site parameter for stiff sites with relatively long wavelengths (compared to slower sites with shorter wavelengths). Overall the best agreement between GWG-E and simulation-based models are at V_{S30} > ~400 m/s and T > 0.2s.

The model for active tectonic regions (Seyhan and Stewart 2014) provides a poor match to the CENA results for most periods. Some particular areas of divergence are:

- The SS14 model does not show flattening of the V_{S30}-scaling at slow velocities
- For the central range of V_{S30} (approximately between V_1 and V_2), the active region V_{S30}-scaling is steeper than that for CENA models.

Because the NEHRP site factors follow the Seyhan and Stewart model, just as the CENA results reject SS14, they also reject the current NEHRP factors (in CENA).

The panel based the median model largely on GWG-E. Referring to Equation (4.3), the zero gradient for $V_{S30} < V_1$ and slope c for $V_1 < V_{S30} < V_2$ are taken from GWG-E. The third and fourth elements of the recommended model (i.e., the segments for $V_{S30} > V_2$) depart from GWG-E. Those elements of the model in Eq. (4.3) were constrained by simulations as described further below (*Fast Velocity Model Elements* section). A second exception is that at slow velocities and oscillator periods of 0.3–0.8s, we decrease V_1 from GWG-E values, which raises the amplification. This change was motivated by the GWG-E amplification being lower than other models for soft soils in this period range.

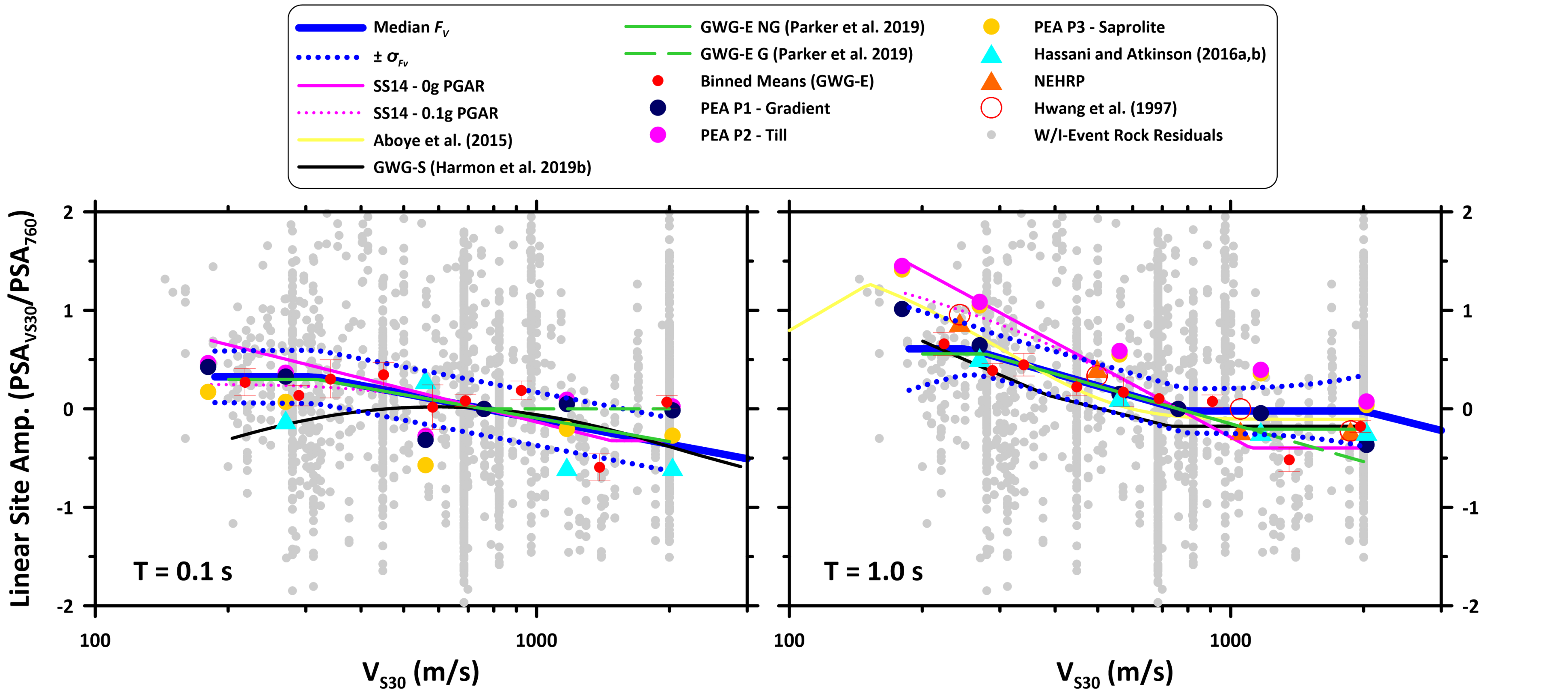

Figure 4.6. Scaling of site amplification with VS30 at oscillator periods of 0.1 and 1.0s, for CENA region from alternate models, and for a reference model for active tectonic regions. SS14 = Seyhan and Stewart (2014) for active regions, for PGAr = 0 (linear site amplification only) and for PGAr = 0.1g (as used for developing current NEHRP site factors). GWG-E G and GWG-E NG = Geotechnical Working Group empirically-based model for glaciated and nonglaciated regions, respectively (Parker et al. 2019). GWG-S = Geotechnical Working Group simulation based model (Harmon et al. 2019b). Hassani and Atkinson (2016a,b) = fpeak-based model for CENA adjusted to unity at 760 m/sec. PEA = Darragh et al. (2015) simulation-based model, adjusted to a reference condition of 760 m/s using three simulation-based factors for representative VS profiles (Profile 1 – Gradient, Profile 2 – Till, and Profile 3 – Piedmont Region Saprolite). Within-event rock residuals and their binned means represent the empirical data considered in GWG-E.

4.4.3 Period Interpolation and Extrapolation

The original work of the panel was constrained by the useable period range of NGA-East data, which is approximately 0.08 to 5.0s.At the request of USGS, the panel estimated coefficients for a wider range of periods and for a few periods inside of the originally considered range but for which plots such as in Figure 4.6 had not been developed. Intensity measures for which these estimates are provided are indicated in Figure 4.4 (both interpolated and extrapolated). Parameter V_2 is not obtained by interpolation or extrapolation, but rather by procedures described in the next section.

Interpolated periods are 0.15, 0.25, 0.75 and 1.5s. Coefficients other than V_2 in Eq. (4.3-4.4) were obtained by log-linear interpolation of the nearest-neighbors.

In the case of extrapolated short period coefficients (0.01, 0.02, 0.03, 0.04, 0.05, and 0.075 sec), we considered the trend of coefficients with period as provided by simulations (Harmon et al. 2019a,b). In the simulation results, coefficient c decreases modestly for periods less than 0.1s before increasing to a local peak at 0.015s, and then saturates to match the values for PGA at about 0.007 s. These features are shown in Figure 4.4. We use a coefficient at the 0.015s peak that is 20% lower than that at 0.1s (-0.28) which is motivated by this same shift in simulation-based coefficients. For V_1, values derived from data increase as the period shortens (Figure 4.4), which is consistent with features in the simulation-based model for periods under about 0.1 s. We follow this pattern, using the V_1 at 0.1 s for shorter periods.

In the case of extrapolated long period coefficients (7.5 and 10s), we project values of c using the slope computed between existing coefficients at 4.0 and 5.0s (Figure 4.4). This pattern matches the general trend of models for active regions. We consider the use of empirical model trends to be preferred to guidance from simulations due to difficulties in modeling site response with 1D models at long periods (e.g., Stewart et al. 2017b). For V_1, we maintain the value at 5.0s for longer periods.

4.4.4 Fast Velocity Model Elements

The empirical data in Figure 4.6 provide relatively weak constraint to the F_V model for fast sites (V_{S30} > ~ 1000 m/s). In order to provide sensible variations of site amplification in this range, we considered simulation results for amplification at sites with V_{S30} = 2000 m/s by Boore (2015). Boore performed computations using the square-root-impedance method, also known as the quarter-wavelength method (Boore 2013). These simulations used velocity profiles with V_{S30} = 2000 m/s and 3000 m/s, which were modified from the very hard rock crustal model of Boore and Joyner (1997). The site damping parameter κ_0 was taken as 0.006s for both profiles. Figure 4.7 shows the site amplification at 2000 m/s relative to the 3000 m/s reference as interpreted from these simulations.

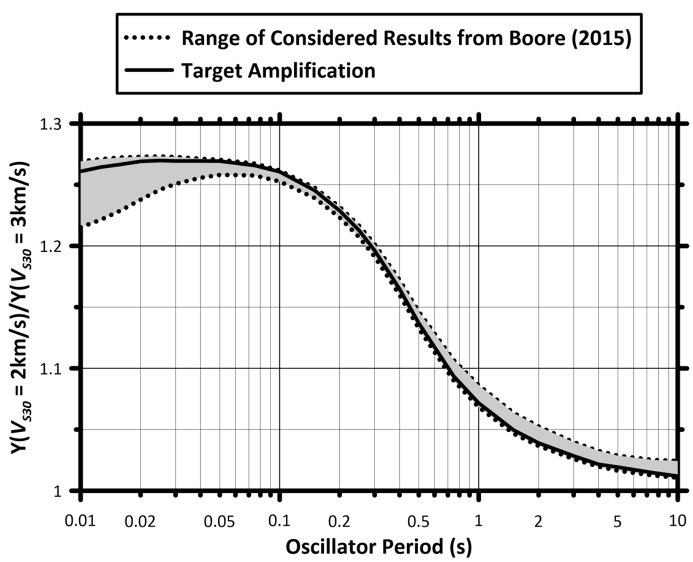

Figure 4.7. Simulation-based site amplification for V_{S30}=2000 m/s site relative to 3000 m/s reference condition, derived from Boore (2015b) using results for M=6-8 and R_{rup} = 10-100 km.

The selected values of site amplification at 2000 m/s, shown in Figure 4.7, were used to constrain the F_V model. The corresponding F_V values were computed from the results in Figure 4.7 as,

$$F_V(V_u) = ln(Y_{2000}) - F_{760}^{imp} \tag{4.8}$$

where V_u is 2000 m/s and Y_{2000} is the site amplification in Figure 4.7. We used the impedance model for F_{760} in this case, which we consider to be more appropriate for fast sites. The model in Eq. (4.3) is formulated to provide an amplification of $F_V(V_u)$ for $V_2 < V_{S30} < V_u$. This is obtained by adjusting V_2 from the original GWG-E values. The adjusted values of V_2 are shown in Figure 4.4 (labeled as computed). The last line of Eq. (4.3) provides for a linear decrease of amplification from $F_V(V_u)$ to $-F_{760}$ between V_u and 3000 m/s.

4.4.5 Model Uncertainty

We developed the model uncertainty shown in Figure 4.6 and the electronic supplement using engineering judgment, rather than through a formal calculation of standard deviations between models. This approach was applied for three principal reasons: (1) the variations among models is uneven across periods, being relatively low for $T > 1.0$s and large at smaller periods − in the judgment of the panel, these period-to-period features do not reflect true epistemic uncertainties in site amplification; (2) for many periods, the median model is not at the center of the range in log space (there are often more models above than below the median)−as a result, application of a formal standard deviation around the median model would not have encompassed the expected number of models; and (3) the panel judged that increases in the model uncertainty should be applied at upper and lower ends of the velocity range, where data are sparse −reliance on formal statistical methods would frequently not provide this.

We interpreted the distribution of results in the figures and proposed a range that can be interpreted as ± one standard deviation (σ_v). We sought to center the model on the median, to have the width of the range represent uncertainty in a smoothed manner across the velocity range (not fluctuating), and to increase the uncertainty at slow and fast velocities where data are relatively sparse. In Equation 4.4, term σ_{vc} represents the selected standard deviation in the central portion of the velocity range, and is plotted as a function of period in Figure 4.8. The relations in Equation (4.4) for $V_\ell < V_{S30} < V_1$ and $V_u > V_{S30} > V_2$ are polynomials constrained to have dispersion of σ_{vc} and zero slope at V_1 and V_2.

As shown in Figure 4.3, the uncertainty linearly decreases towards zero between V_u and 3000 m/s. This is applied because the model considered that the epistemic uncertainty for sites at or near 3000 m/s to be captured by the NGA-East GMMs (Goulet et al. 2017), with further uncertainty associated with site amplification being unnecessary.

We increased model uncertainty at short and long periods where coefficients were extrapolated. Figure 4.8 shows these increases to σ_{vc} beyond the observation range of 0.08-5.0s. Similar increases are provided for σ_ℓ and σ_u. These increases were largely based on expert judgement. Values of V_f were also increased in the extrapolation region, which has the effect of broadening the velocity range with increased uncertainty (i.e. lines 1 and 3 in Eq. 4.4).

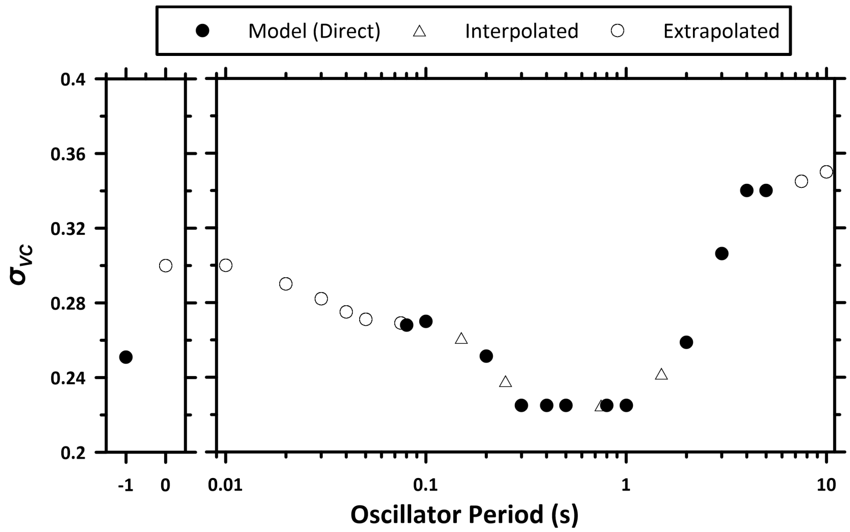

Figure 4.8. Trend with period of epistemic uncertainty parameter σ_{vc} as developed from observation and as extrapolated to short and long periods

4.5 F_{760} MODEL DEVELOPMENT

4.5.1 Models Considered

The proposed model for adjusting ground motion intensity measures from the V_S = 3000 m/s reference condition to V_{S30} = 760 m/s (F_{760}) is based on a number of alternative simulation results, all of which are based on one-dimensional ground response analyses of various types. This section presents the considered simulation results.

The panel considered results from four investigations – Boore and Campbell (2017), PEA, GWG-S, and Frankel et al. (1996) (later applied in Atkinson and Boore 2006). Boore and Campbell (2017) use both a square-root impedance approach and an approach that captures resonance effects. We consider the Boore and Campbell (2017) results to largely supersede results from previous related studies (Beresnev and Atkinson, 1997; Boore and Joyner, 1997; Boore, 2015; and Boore and Thompson, 2015). PEA and GWG-S used wave propagation analysis procedures (RVT-based equivalent linear and linear viscous-elastic, respectively) that capture resonance effects, and nonlinear effects in the case of PEA. Different material damping models were used in these studies, as discussed previously. The Frankel et al. (1996) study was re-done here for various magnitude and distance combinations using a square-root impedance approach.

Figure 4.9 shows the profiles used by PEA, GWG-S, and Frankel et al. (1996). The GWG-S profiles are based on measurements from CENA sites in which V_{S30} is between 700 and 800 m/s. The Boore and Campbell (2017) profiles (not shown in Figure 4.9) are similarly selected to be within 10% of 760 m/s), and as a group are qualitatively similar to those of GWG-S. The three PEA profiles are intended to be representative of three different CENA geologic conditions: glacial till, Piedmont saprolite, and a weathered rock gradient, all with V_{S30} = 760 m/s. They were constructed using suites of measured profiles reflecting these near surface geologies. The Frankel et al. (1996) profile represents a rather gradual increase of velocity with depth. A typical feature of the profiles considered by Boore and Campbell, PEA (till, saprolite), and GWG-S is the presence of impedance contrasts; these profiles were used to develop the impedance model (F_{760}^{imp}). The weathered rock (PEA) and gradient (Frankel et al.) profiles lack large impedance contrasts; these were used to develop the gradient model (F_{760}^{gr}).

Aside from V_S profiles, the other site parameter that strongly influences F_{760} is the site damping parameter κ_0. Based on an assessment by Boore and Campbell (2017), we use their simulation results for κ_0 = 0.01, 0.02, and 0.03s.PEA use κ_0 = 0.02s for 760 m/s profiles. The re-working of the Frankel et al. (1996) analyses that was performed here used site κ_0 = 0.01 and 0.02s. The GWG-S simulations employ a material damping model, which does not require specification of κ_0.

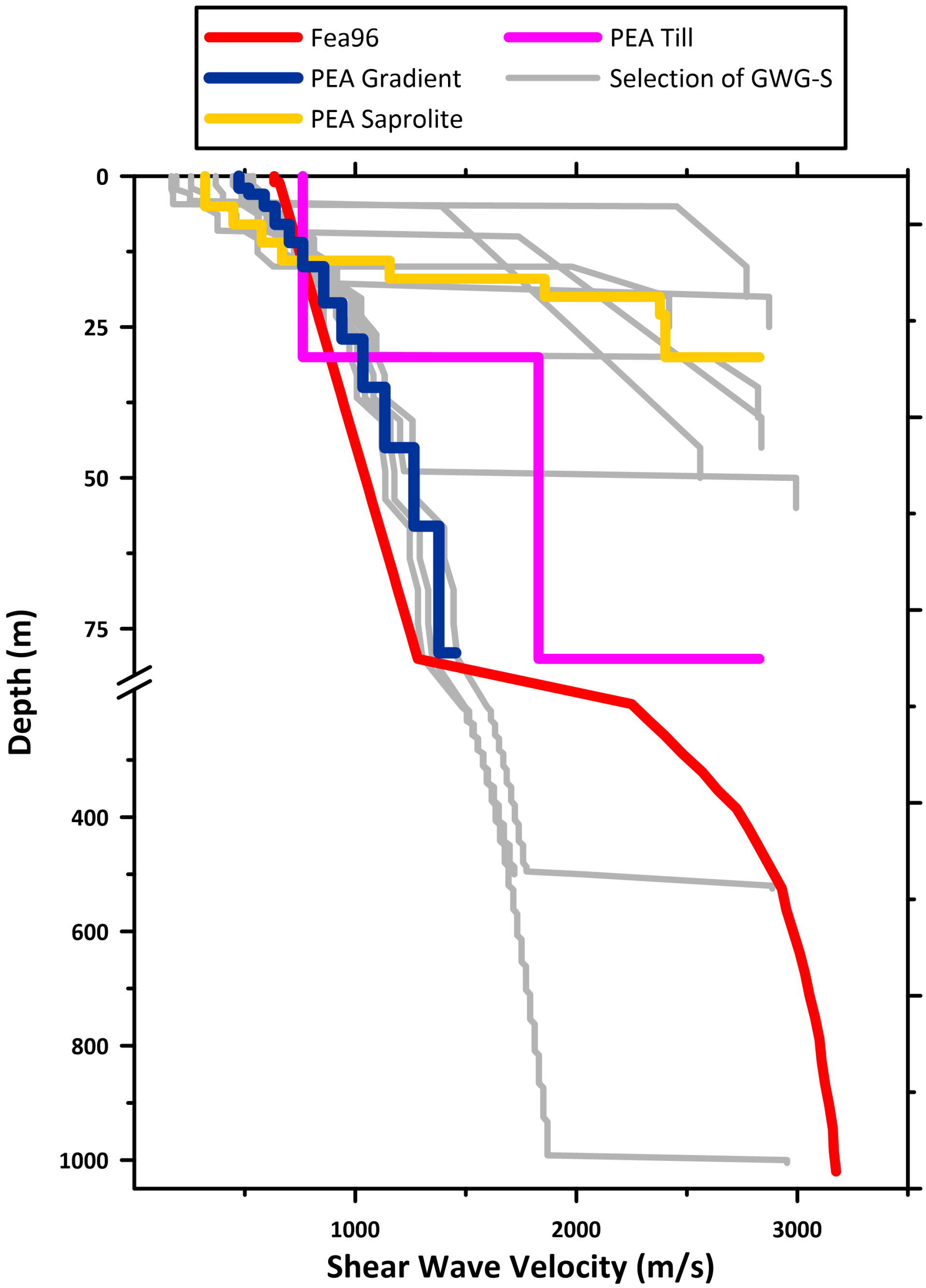

Figure 4.9. Shear-wave velocity vs depth profiles in CENA with V_{S30} between 700 and 800 m/s (marked as GWG-S in legend; Harmon et al. 2017a) or equivalent to 760 m/s as given by PEA (Piedmont saprolite, till, weathered firm rock) and Frankel et al. (1996) (gradient).

For the development of 5% damped pseudo spectral acceleration (PSA) ratios, it is necessary to compute ground response using acceleration time series, typically developed from point source simulations. To encompass a range of conditions, we took results from Boore and

Campbell for **M**5 at 10 km and **M**8 at 500 km. PEA results also apply for close distances. The GWG-S input motions cover a wide range of magnitudes and distances, but can generally be considered as having ample high-frequency energy as would be expected for ground motions reasonably near a seismic source for hard rock site conditions (V_S = 3000 m/s). Harmon et al. (2019b) have F_{760} models for a variety of depths to the 3000 m/s shear-wave horizon; the results presented here represent an average over the considered depth range. The gradient profile in Figure 4.9 was re-analyzed using input motions for **M** 4.5 and 6.5 and rupture distances of 10, 50, and 100 km.

4.5.2 Recommended Impedance and Gradient Models

Figure 4.10 shows the resulting 5% damped pseudo-spectral acceleration ratios from the three sets of simulations for impedance conditions. Most of the results have a similar shape, with a peak near 0.1-0.2s, decay towards no amplification (unity) at long periods, and highly variable behavior at periods below the peak as a result of model-to-model variability and variability between κ_0 values. We consider all of the results in Figure 4.10 to be credible representations of F_{760} behavior for impedance conditions. Accordingly, the recommended model is the median of the models shown in the figure. The uncertainty shown in the figure (σ_{lnF760}) represents a smoothed standard deviation between models, which decreases appreciably with period.

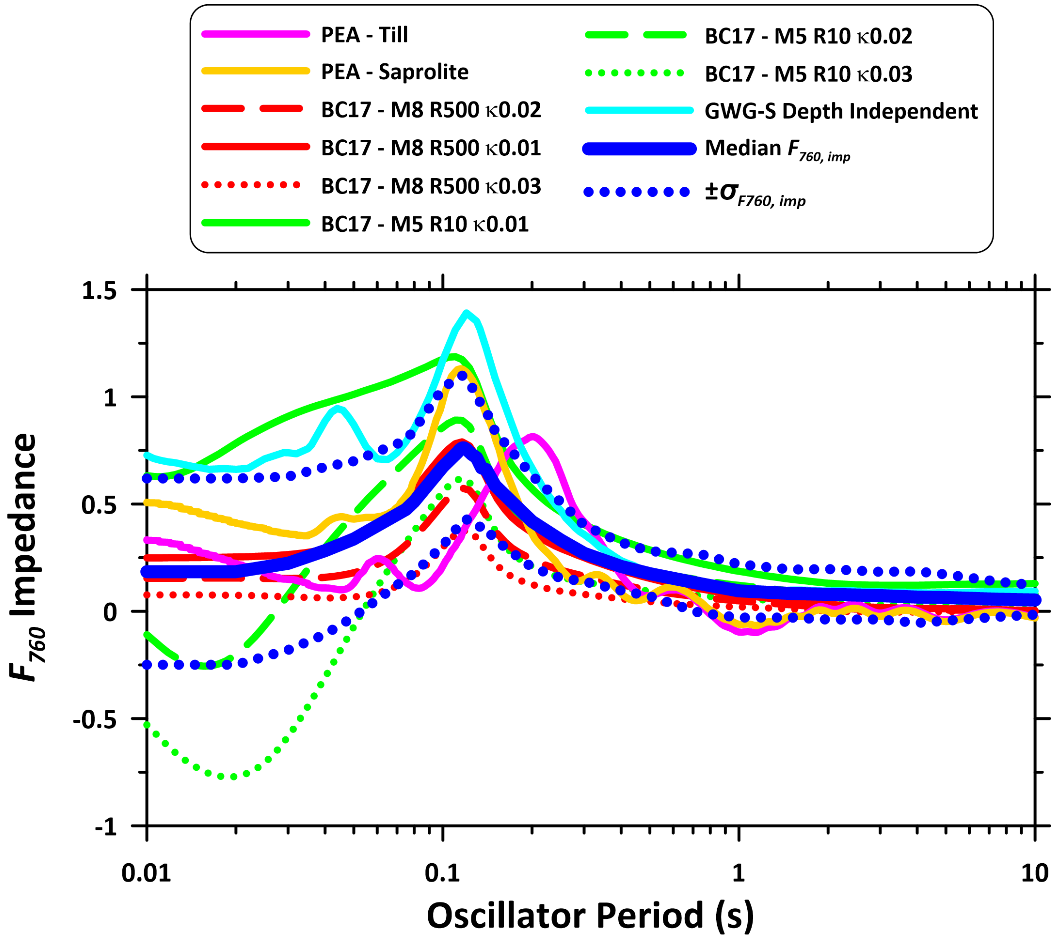

Figure 4.10. Reference site factor F_{760} for impedance profiles from Boore and Campbell (2017) (labelled BC17), PEA (Darragh et al. 2015), and GWG-S (Harmon et al. 2019b).

Figure 4.11 shows amplification results for gradient conditions. The gradient amplifications lack the peak near 0.1s and tend to have larger amplification at longer periods. The median model and uncertainty encompass the available models, with the exception of results for κ_0=0.01s for short periods.

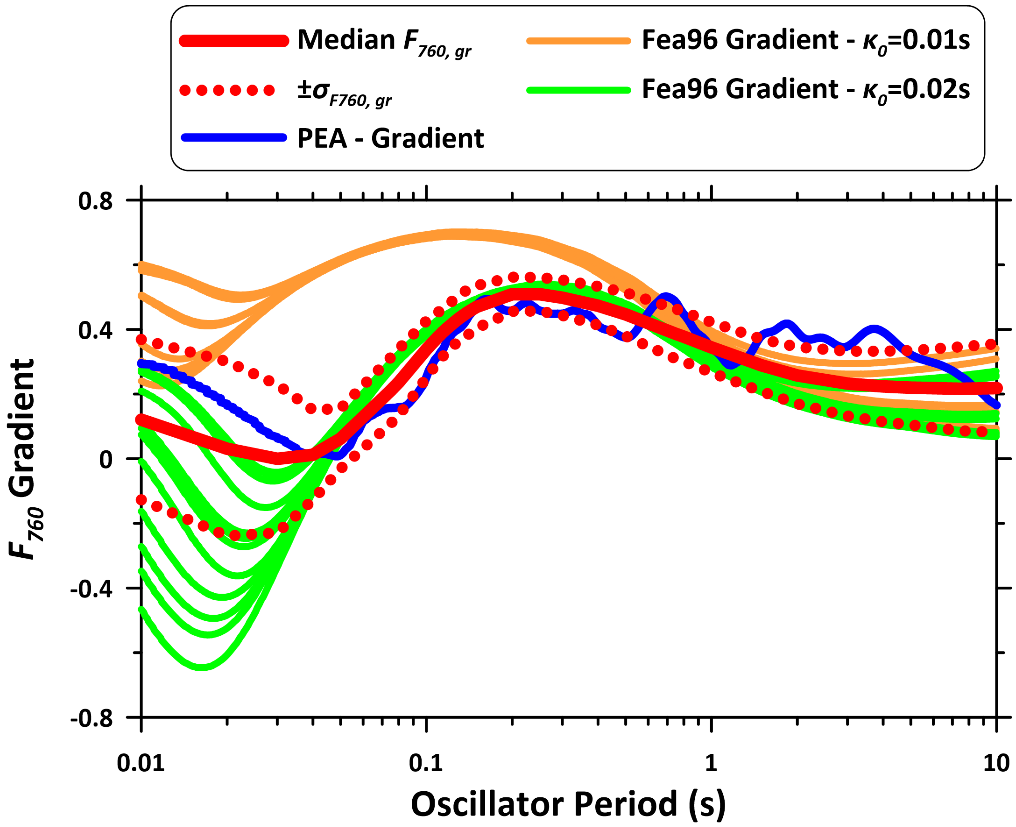

Figure 4.11. Reference site factor F_{760} for gradient profiles from PEA (Darragh et al. 2015) and Frankel et al. (1996), as re-analyzed in this study (labelled Fea96).

4.5.3 Model Weights

The impedance and gradient F_{760} models have distinct features, and for many applications, guidance is needed regarding the selection of the most appropriate model to pair with F_V. Particularly noteworthy is the peak at 0.1s in the impedance model, which can appreciably impact spectral shape.

To investigate the degree to which the 0.1s peak in the impedance model is realistic (or not), and to guide the selection of appropriate model weights, we examined spectral shapes from CENA ground motions for different V_{S30} ranges. After binning by earthquake magnitude (**M**), rupture distance (R_{RUP}), and V_{S30}, the available spectra were normalized by the average PSA between 0.08

and 1.5s oscillator periods. The spectral shapes for **M** = 4-5.5, R_{RUP} = 0-150 km and V_{S30} bins around 2000, 760, 500, and 260m/s are shown in Figure 4.12. The data show a strong peak near 0.1s in the mean spectral shape for V_{S30} exceeding 500 m/s, and a peak near 0.25s at 260 m/s. These trends match those observed by Hassani and Atkinson (2016b) in the NGA-East data, in which the peak of H/V (a proxy for site amplification) is near 0.1 s (10 Hz) for sites with V_{S30} values in the range from 500 m/s to 1000 m/s. Results qualitatively similar to Figure 4.12 are obtained from spectral shapes of simulated motions for different site conditions.

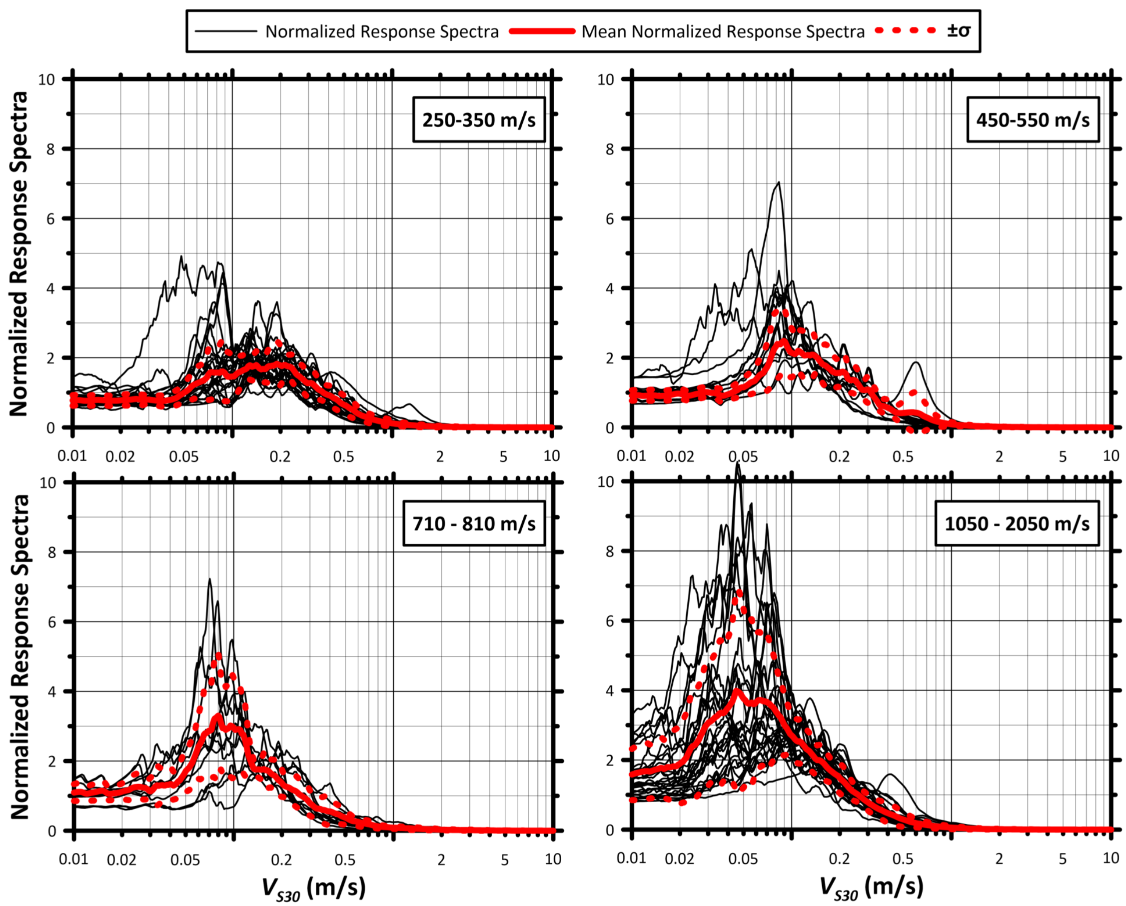

Figure 4.12. Spectral shapes of NGA-East data for M4-5.5 earthquakes recorded at R_{RUP} between 0-150 km at sites with the approximate V_{S30} values marked in the figures. Spectral shapes are normalized by the average response between 0.08-1.5s.

The weighting model in Eqs. 4.6-4.7 was formulated to enable the impedance and gradient models to be assigned different weights for different V_{S30} values. Alternate weight assignments have been discussed among the panel and between the second author and USGS technical staff. One approach, preferred by the panel, gives preference to the impedance model for fast sites, and to the gradient model for slow sites. Proponents suggested w_1 = 0.9, w_2= 0.1, V_{w1} = 600 m/s and V_{w2} = 400 m/s. Another approach, preferred by some USGS staff, gives equal weight to impedance and gradient models for fast V_{S30} sites and preference to the gradient model for soft sites. For use in the 2018 national maps, the decision was ultimately made to give 2/3 weight to approach 1 and 1/3 weight to approach 2 for firm sites, resulting in w_1 = 0.767. For soft sites, the gradient model was preferred by consensus, resulting in w_2 = 0.1. The transition velocities are V_{w1} = 600 and V_{w2} = 400 m/s.

4.6 SUMMARY AND DISCUSSION

4.6.1 Summary Recommendation

We recommend that ergodic (non-site specific) V_{S30}-based site amplification in central and eastern North America be computed using Eqs. 4.1-4.3 and 4.5-4.7, with the coefficients given in the electronic supplement for the F_{lin} model. The corresponding nonlinear model (F_{nl}) is given in Hashash et al. (2017, 201x). The model has three components in natural log units: F_V for V_{S30}-scaling referenced to V_{S30} = 760 m/s, F_{760} for amplification of the 760 m/s site condition relative to the CENA reference of V_S = 3000 m/s, and F_{nl} for nonlinear effects. These models are based on a combination of ground-motion data analysis and ground response simulations. The form of the F_V model is constrained by data, except for very stiff (fast) sites where it is constrained by simulations. The F_{760} models are simulation-based, with an impedance model representing conditions with large impedance contrast (applicable to stiff sites) and a gradient model representing conditions with a relatively deep weathering profile and no strong impedance contrasts (applicable to soft sites).

We recognize that our recommendations represent a substantial departure from past practice in CENA, which was based on site factors applicable to active tectonic regions. NGA-East data and simulations demonstrate that such models are biased for application to CENA sites.

Many CENA sites have strong resonance effects that can be better described by models that incorporate information on the site frequency. We encourage the use of such models as part of site-specific analyses. The use of such models was beyond the scope of the present study, but should be considered for future code developments.

4.6.2 Model Performance

The linear amplification resulting from the recommended model is given for various V_{S30} in Figure 4.13. The amplification is peaked near 0.1s for velocities up to about 500 m/s, as seen in data. The peak in the amplification then shifts to longer periods for softer sites. Including nonlinear effects (not shown in Figure 4.13) would further emphasize the shift to longer periods for strong shaking conditions.

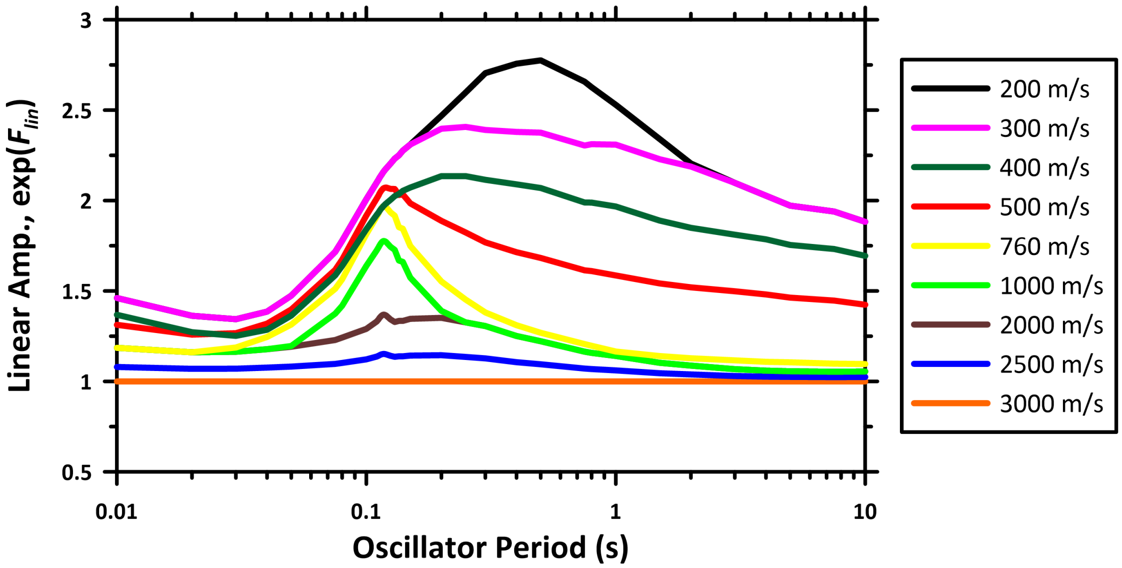

Figure 4.13. Linear amplification for oscillator periods from 0.01 to 10s for various V_{S30} using the proposed model with selected weights for USGS maps.

4.6.3 Model Rationale

Here we discuss several strategies that were employed in model development; they are presented as answers to frequently posed questions.

Why did we adopt a hybrid approach in which simulations are solely used for the nonlinear model while empirical data in conjunction with simulations were considered for the linear model? Our response is two-fold. First, there is precedent for combining information sources in a hybrid manner for application in active tectonic regions (e.g., Dobry et al., 2000; Seyhan and Stewart, 2014). Moreover, whereas the use of ground response simulations to predict absolute levels of site amplification have been shown to be potentially problematic (e.g., Baturay and Stewart 2003; Thompson et al. 2012; Kaklamanos and Bradley 2018), their application for prediction of nonlinear effects is often effective (e.g., Kwok and Stewart 2006).

Why do we split the linear amplification term into two components instead of using a single term referenced to V_S=3000 m/s? The empirical data are useful to constrain the changes in site amplification over the range of site conditions present in the dataset, which is approximately V_{S30} = 200 to 2000 m/s. There is no observational basis for extending this range to the 3000 m/s reference condition. As a result, the model discussed here uses data where it exists, and uses simulations for the step from 760 to 3000 m/s, which is considered preferable to the alternative of not using data and relying solely on simulations to evaluate site amplification for any V_{S30} relative to 3000 m/s (e.g., as in Darragh et al. 2015, Boore and Campbell 2017, and Harmon et al. 2019a,b).

4.6.4 Limitations

The models presented in this report are considered applicable for evaluation of ergodic site response effects for V_{S30} = 200 to 2000 m/s and intensity measures of PGA, PGV, and pseudo-spectral acceleration for oscillator periods between 0.01 and 10.0s.

Being ergodic, the models presented in this report do not provide site-specific estimates of site response effects, even if the V_{S30} value that is used is measured at the site of interest. Additional site-specific attributes could be introduced to the site response estimate by measuring site frequency, soil depth, and other dynamic material properties. Resonance effects are known to be strong at many CENA sites due to soil layers deposited over hard rock, so consideration of these effects can have a substantial impact on site response estimates and we recommend to do so. Such effects can be considered through the use of currently available empirical models (e.g., Hassani and Atkinson 2016a, 2017), simulation-based models (Harmon et al. 2019b), or site-specific analyses.

Finally, we have a recommendation associated with the application of the site response models in this report with NGA-East GMMs. Ideally, the development of GMMs and site terms should occur in a coordinated manner. For example, when performing regression of data for GMM development, site amplification models are often used to correct ground motion intensity measures to a reference site condition. Source and path attributes are then evaluated from regression on the site-corrected data. The coordination referred to above would require that the site models used to correct the data are the same as those used for the forward application. However, that was not the case for CENA with the NGA-East GMMs currently available (PEER 2015a, b; Goulet et al. 2017) and the site amplification model provided here. As a result, it is possible that bias will be found when CENA data are compared to NGA-East GMMs combined with our site amplification models. Accordingly, we recommend future work to re-evaluate the NGA-East GMMs using the available data and our site model, and that appropriate adjustments (likely to the constant term in the GMMs) be made to remove any bias that might be observed. An alternative approach that avoids these difficulties is to use the site variables directly within the GMM regression framework

5 NEXT GENERATION ATTENUATION GROUND MOTION MODEL FOR GLOBAL AND REGIONAL SUBDUCTION ZONE EARTHQUAKES

5.1 INTRODUCTION

The Next Generation Attenuation-Subduction (NGA-Sub) project is a large, multi-year, multi-disciplinary project with the goal of producing uniformly processed ground motion data, including time series and spectral data for earthquakes, and a suite of global and regional ground motion models (GMMs) for subduction zone earthquakes. This project is organized by the Pacific Earthquake Engineering Research Center (PEER), and encompasses subduction zones around the world, including those in Japan, Taiwan, British Columbia (Canada), Alaska and the Pacific Northwest (United States), New Zealand, Mexico, Chile, and Peru (shown in Figure 5.1).

The NGA-Sub project had two phases: (1) database development, including compilation of uniformly processed time series, computation of ground motion intensity measures, and development of metadata from global subduction zone earthquake events and recording sites (e.g. Kishida et al. 2018; Ahdi et al. 2017, Contreras 2017, PEER 2019), and (2) model development, in which a number of model developer teams worked to build models for predicting ground motion intensity measures (IMs) using the NGA-Sub database and auxiliary materials such as ground motion simulations. The work presented in this chapter is part of the second, model development phase of NGA-Sub and was performed in conjunction with collaborators Jonathan Stewart (UCLA), Gail Atkinson (Western University), Behzad Hassani (Western University), and David

Boore (USGS), with review from participants of the NGA-Sub project as a whole. The work presented in this chapter is closely related to the seismic site amplification model presented in Chapter 6.

5.2 LITERATURE REVIEW

Subduction zones are the descending limbs of mantle convection cells at convergent plate boundaries (Stern 2002), where one piece of lithosphere overrides a second, less buoyant section. Two types of earthquakes are generated in these regimes: (1) interface events, that occur due to the coupling of the subducting and overriding plate, and (2) intraslab earthquakes, that occur within the subducting plate. Interface earthquakes are controlled by the age of the down-going plate; fast subduction of young and buoyant lithosphere causes stronger coupling between the plates and thus larger events, and slow subduction of old, colder plates leads to weaker coupling. Intraslab events are controlled by the thermal state of the interior of the subducting slab (Stern 2002), and tend to be normal due to extension in the plate during subduction. This type of tectonic environment occurs in many regions globally as shown in Figure 5.1, making the resulting seismic hazard relevant for many populated areas including Japan and the Cascadia region of the U.S.

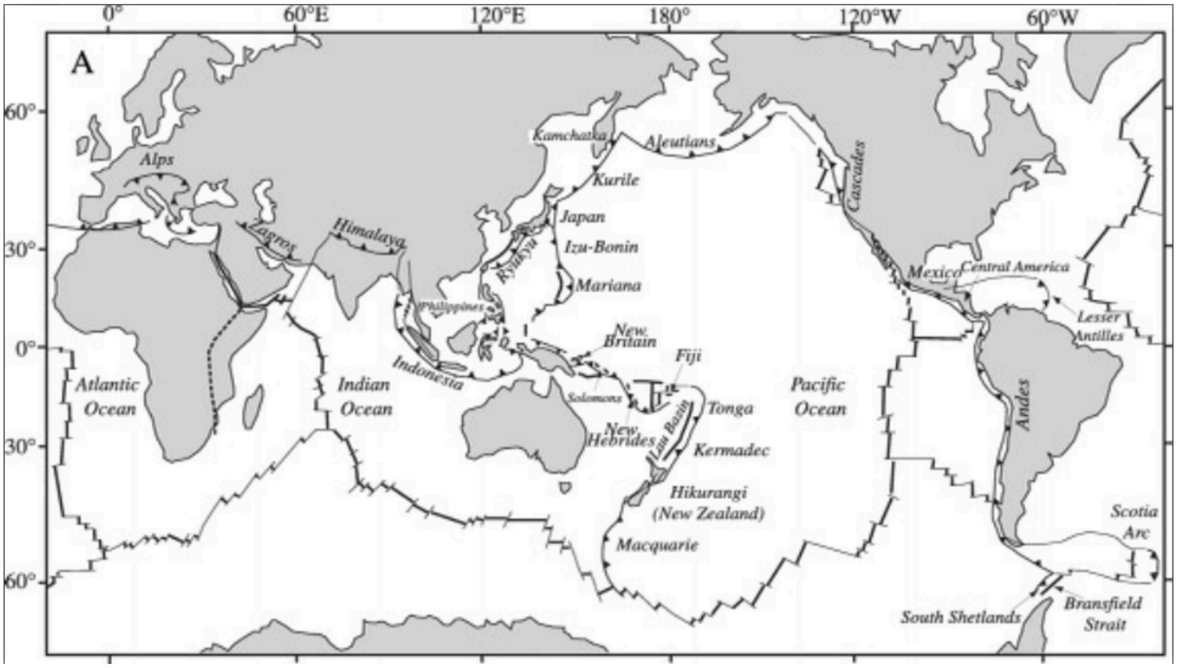

Figure 5.1. Global map of plate boundaries from Stern (2002). Convergent boundaries shown as solid lines with black teeth.

Because of the seismic hazard presented by subduction zones, the ground motions they produce have been the subject of much study, both empirical and simulation-based. Early studies of empirical ground motions from subduction zone regions did not investigate the difference in the ground motion amplitudes and behavior produced by different subduction zones globally (Atkinson 1997; Youngs et al. 1988; Crouse et al. 1988). Thus, the Frankel et al. (1996) U.S. Geological Survey national seismic hazard map was produced using the Youngs et al. (1997) subduction GMM for earthquakes occurring in the Cascadia region. Youngs et al. (1997) is an ergodic GMM developed using a mixed effects regression with 350 recordings from Alaska, Cascadia, Japan, Mexico, Peru, and the Solomon Islands, without consideration of potential regional effects. However, as the size and reliability of ground motion databases increased, this ergodic assumption was disproven. Atkinson and Boore (2003) used 1200 recordings from global events within the magnitude-distance range of engineering interest to develop a ground motion model for subduction zones. They found significant regional differences; the ground motion amplitude in Cascadia are reduced at long oscillator periods by up to two times from those in Japan for the same event type, magnitude, source-to-site distance, and NEHRP soil category. They also found that intraslab events produce larger ground motions than interface events within 100km of the fault, but decay faster with distance than interface events at larger distances.

Due to these global differences in ground motions, many regional ground motion models have been developed, in particular for data-rich regions such as Japan (Si and Midorikawa 1999; Zhao et al. 2006; Zhao et al. 2016a,b) and Taiwan (Lin and Lee 2008). In regions without much available data, such as Cascadia, simulations have been used to further inform our understanding of the earthquake ground motions and the hazard they pose. Gregor et al. (2002) performed a suite of stochastic finite fault simulations for **M**8.0, **M** 8.5, and **M** 9.0 interface earthquakes in Cascadia. The stochastic finite-fault model was validated against the 1985 **M**8.0 Michoacan, Mexico earthquake and the 1985 **M**8.0 Valpariso, Chile earthquake. A simple GMM was then fit to the computed intensity measures. The model predicts similar PGA at short distances (≤150km) as the Youngs et al. (1997) model, but larger PGA values at long distances due to slower distance attenuation. Atkinson and Macias (2009) also performed a suite of stochastic point source simulations for Cascadia interface events with **M**7.5-9.0. They validated their simulations using the **M**8.1 Tokachi-Oki earthquake sequence of Japan, and then adjusted the simulations accounting for the average source, attenuation, and site parameters of the Cascadia region. They find that uncertainties in input parameters due to regional differences, such as distance attenuation, produce large uncertainties in the resulting simulated ground motions, up to a factor of 2 at 100km.

More recently, motivated by high-impact infrastructure projects in the Cascadia region such as dams, a push has been made to accumulate a large empirical dataset for the development of semi-ergodic ground motions models for subduction zones. Abrahamson et al. (2016), a journal paper written to summarize the subduction zone GMM described in the BC Hydro (2012) report, used a mixed-effects regression approach following the Abrahamson and Youngs (1992) algorithm to develop an empirical GMM. This model was developed using an expanded dataset from Atkinson and Boore (2003), consisting of 9,946 horizontal time series pairs from 292 earthquakes. Their analysis finds that the same magnitude-scaling slope can be used for interface and intraslab events, and different distance-scaling slopes are needed in the forearc and back-arc regions of subduction zones. Comparisons to previous models shown that at short distances (≤100km) Abrahamson et al. (2016) predictions fall within the range of existing GMMs, but at longer distances the model predicts lower GMs due to faster distance attenuation. The Abrahamson et al. (2016) model is meant to be global, with a range of epistemic uncertainty in the constant term that can be used to represent regional variation in ground motion amplitudes. However, the model does not have a regionalized anelastic attenuation term or regionalized V_{S30}-scaling.

Lastly, Frankel et al. (2018) and Wirth et al. (2018) produced a set of broadband (0-10Hz) synthetic seismograms for **M**9 Cascadia interface events by combining synthetic seismograms derived from 3D finite-difference simulations (≤1Hz) with finite-source, stochastic synthetics (≥1Hz), informed by the **M**9.1 Tohoku earthquake and the **M**8.8 Maule earthquake. For sites not in sedimentary basins, the simulated ground motions match predictions from Abrahamson et al. (2016) for 0.1-6.0s, but are larger at longer periods. They also find that sites in Cascadia-area basins, such as the Tacoma and Seattle Basins, show site amplification factors of 2-5 for periods 1.0-10.0s, much larger than that predicted by the NGA-W2 GMMs (Bozorgnia et al. 2014).

The Next Generation Attenuation-Subduction project was started in 2014 with the goal of producing a uniformly processed ground motion database and a suite of improved GMMs to represent epistemic uncertainties in predicted ground motions. The rest of this chapter describes the development of a semi-empirical global ground motion model with regional adjustment factors for interface and intraslab subduction events using the NGA-Subduction ground motion database (Kishida et al. 2018; Ahdi et al. 2017; Contreras 2017; PEER 2019). The model presented herein uses a larger dataset than that of Abrahamson et al. (2016); considers regionalization in the ground motion amplitude, anelastic attenuation, source depth scaling, and V_{S30}-scaling terms; and treats the distance-, magnitude-, and depth-scaling terms differently between interface and intraslab event types.

5.3 DATABASE SELECTION AND SCREENING

5.3.1 NGA-Subduction Database Overview

The NGA-Sub database contains 71,343 three-component time series from 1,883 earthquakes that have been acquired from subduction zone regions around the world (Kishida et al. 2018; PEER 2019). Figure 5.2 shows the regional distribution of recordings, earthquakes, and recording stations. Table 5.1 gives the source organizations for the time series by region. The overall database is a combination of three individual databases: an earthquake source database, an earthquake recording database, and a recording station database. Data pulled from all three databases are combined into a single summary "flatfile" with one line per recording for use in the development of GMMs.

Table 5.1. Agencies that provided ground motion time series to the NGA-Subduction database (Table 1 in Kishida et al. 2018).

Region	Data Source Agencies
Japan	K-NET, KiK-net, PARI, JMA, NOAA, Hi-net
Taiwan	CWB, IES
Pacific Northwest	CESMD, COSMOS, IRIS, NSMP, NCEDC, GSC
Alaska	CESMD, COSMOS, IRIS, GSC
South/Central America	CESMD, COSMOS, NOAA, IRIS, GFZ, RENADIC, CSN, CISMID, NORSAR, Other regional networks

The ground motion recording database contains time series that have been processed using a set of instrument correction, filtering, and baseline correction algorithms developed by PEER (the procedures are similar to those described in Goulet et al. 2014). The resulting dataset includes acceleration time series, PGA, PGV, PGD, pseudo-spectral acceleration (PSA) for 110 oscillator periods between 0.01-20s, Fourier amplitude spectral ordinates (FAS), and significant durations based on Arias Intensity.

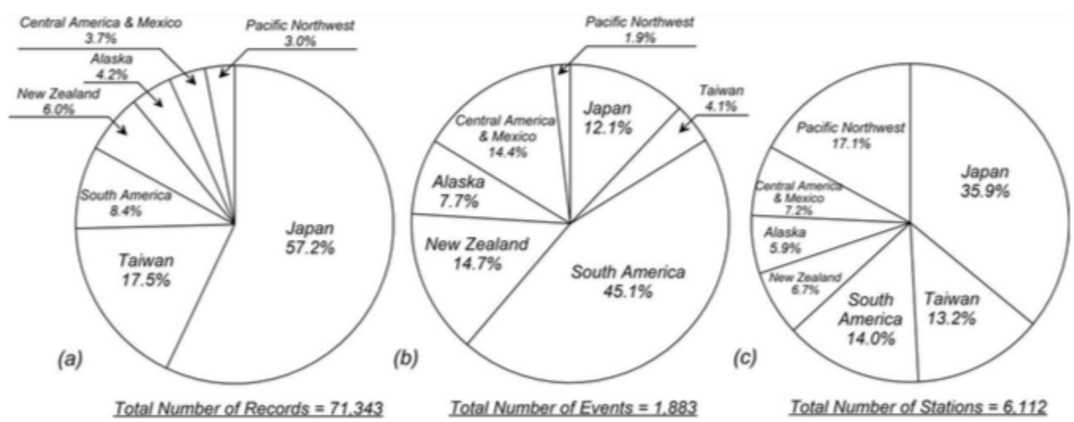

Figure 5.2. Regional distribution of (a) recordings, (b) events, and (c) stations in the NGA-Subduction Database (Figure 2 from Kishida et al. 2018).

The earthquake source database includes information on all earthquake events that produced recordings contained in the recording database, including location, magnitude, hypocentral depth, and finite-fault models (Contreras 2017; PEER 2019). The earthquakes have been classified into event-type categories: interface (43% of database), intraslab (42%), lower double-seismic zone (1.3%), outer rise (2.4%), or shallow crustal (11%) earthquakes.

The recording station database has information on site condition and instrument housing for the 6,112 strong motion stations that have recorded time series (Ahdi et al. 2017; PEER 2019;). This data includes station location, identification numbers, recommended V_{S30} and the corresponding source, and basin depths from measurement or models for stations in Japan and the Pacific Northwest. Seismic velocity measurements are used to calculated V_{S30} values for 39% of the stations in the database, whereas the rest are from proxy-based relations

5.3.2 Data Screening for Model Development

A subset of records from the NGA-Sub database was used for model development, which was selected based on the following criteria:

1. Flatfile entry for record was populated with the metadata necessary for model development (**M**, rupture distance (R_{rup}), hypocentral depth (d_{hyp}), V_{S30}, etc.)
2. Records from earthquakes classified as being interface, intraslab, or in the lower double seismic zone
3. Records from earthquakes classified as being class 1 (i.e. a mainshock) according to Wooddell (2018) method 2 using an 80km cutoff distance
4. $R_{rup} \leq \min(R_{max}, 1000\text{km})$
5. Sensor depth ≤ 2m

6. If interface event: $d_{hyp} \leq 40$km, if intraslab: $d_{hyp} \leq 200$km, where d_{hyp} is earthquake hypocentral depth

7. $T_{SU} \leq T_{osc} \leq T_{Oosc} \leq T_{LU}$, where T_{osc} refers to PSA oscillator period, and T_{SU} and T_{LU} refer to the shortest and longest useable periods based on the corner frequencies used to process the record, respectively

8. Epicenter and recording station both located in the forearc region

9. Multiple event flag ≠ 1, which excludes recordings that are from complex, multi-segment ruptures

10. Late P-wave trigger flag ≠ 1, which excludes recordings where the p-wave arrival was missed

11. Source review flag = 0, 1, 2 or 4, which excludes records with source properties that did not undergo quality control checks

12. At least 3 recordings/event after criteria 1-11 are applied

After the above screening criteria were applied, the number of events and recordings used for model development varied as a function of period, with a range of 1435-4618 records and 31-68 events (Figure 5.3). The magnitude-distance distribution of records for PGA from interface and intraslab events are shown in Figure 5.4. The data in Figure 5.4 is plotted with identification of the major regions that contribute data to NGA-Sub and for which regional effects were considered in model development.

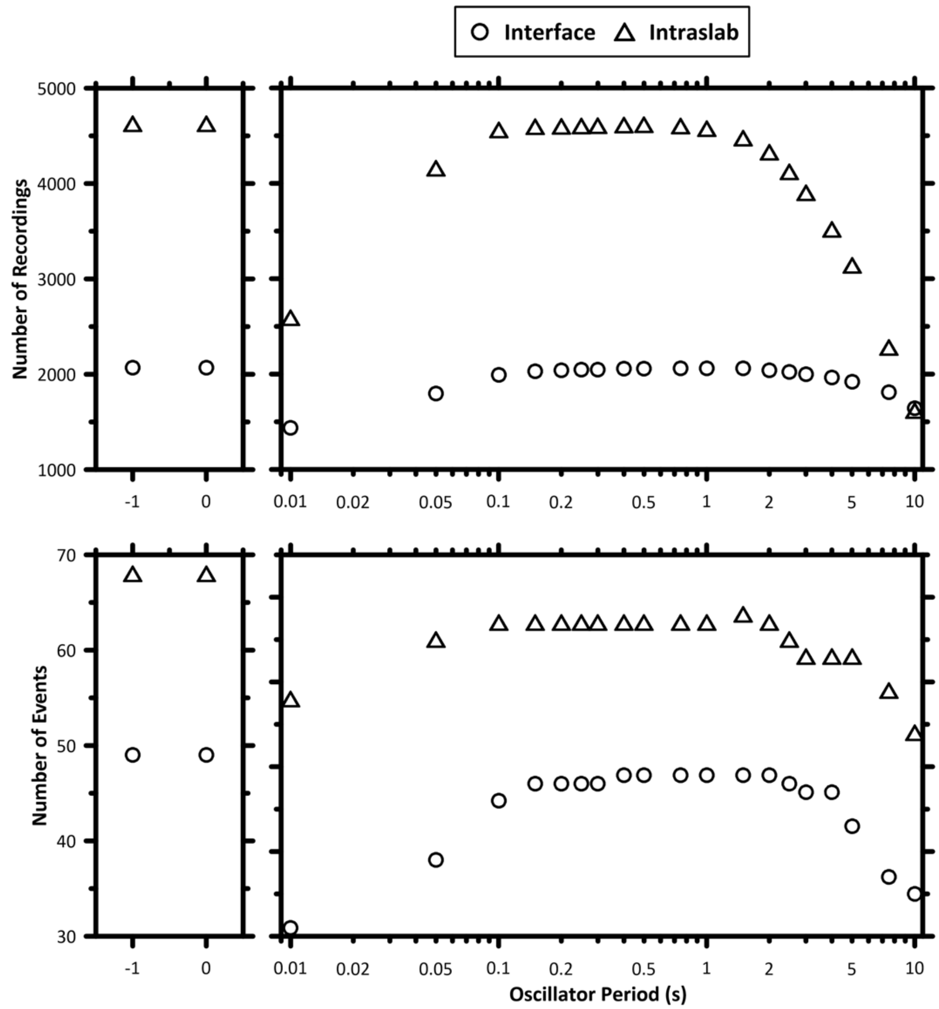

Figure 5.3. Number of events and number of recordings selected for model development according to the criteria in Chapter 5.2.2, for interface (circles) and intraslab (triangles) events.

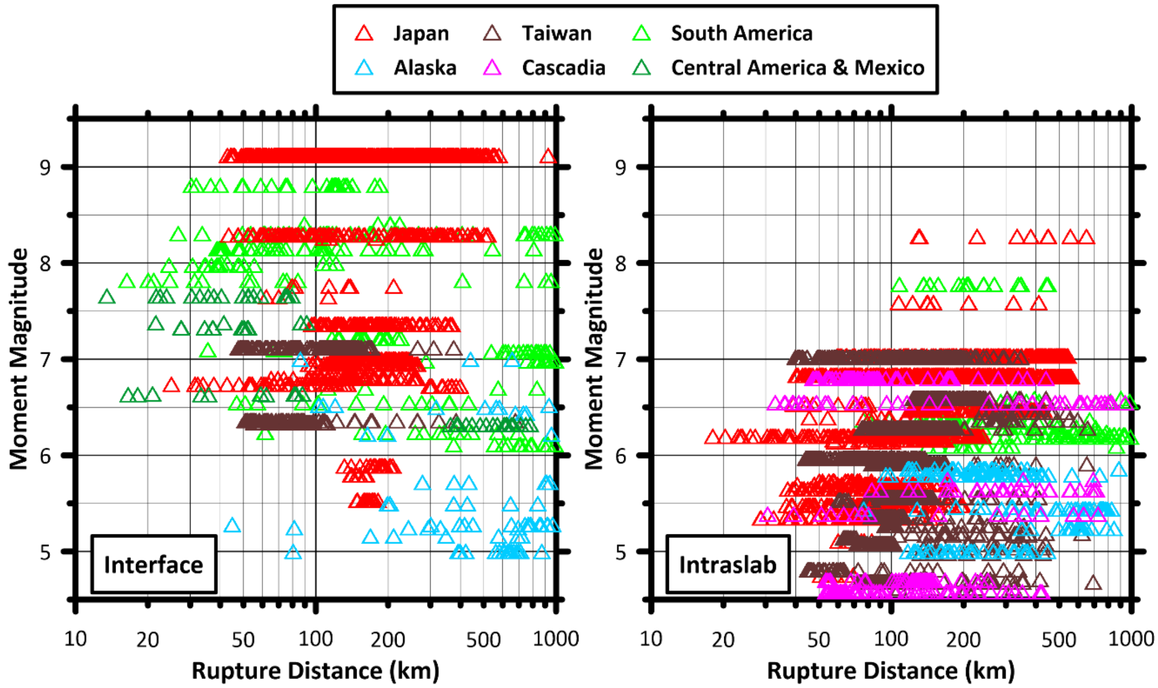

Figure 5.4. Magnitude-distance distribution of recordings from interface (left) and intraslab (right) events, color-coded by region.

5.4 GLOBAL AND REGIONAL MODELS

5.4.1 Global Models

Due to differences in path and source-scaling attributes, the global ground motion models are separate for interface and intraslab earthquakes. The two models have the same functional form, where some coefficients remain the same between the two, and other coefficients have two sets of values. Each median model has five terms: a constant (c_0) that controls the overall amplitude of the predicted ground motion; path term or distance-scaling term (F_P); the magnitude-scaling term (F_M); the source depth-scaling term (F_D); and the site amplification term (F_S). These terms are additive in natural log space:

$$\mu_{lnY} = c_0 + F_P + F_M + F_D + F_S \tag{5.1}$$

The path term consists of a magnitude dependent geometrical spreading (GS) term that represents the purely geometrical effect of the spreading of energy as seismic waves propagate from a point source along a spherical wave front, and an anelastic attenuation term that represents the per-cycle damping as seismic waves pass through the earth (Eq. 5.2). The F_P term accounts for near-source saturation using the variable h (Eq. 5.4), which is combined with site-to-source distance metric R_{rup} as a geometric mean (Eq. 5.3).

$$F_P = c_1 lnR + (b_3 + b_4\boldsymbol{M})lnR + \ a_0 R \tag{5.2}$$

$$R = \sqrt{R_{rup}^2 + h^2} \tag{5.3}$$

$$h = 10^{-0.82+0.252\boldsymbol{M}} \tag{5.4}$$

The magnitude-scaling term is a piecewise function with a parabolic and a linear segment, transitioning at a corner magnitude m_b:

$$F_M = \begin{cases} c_4(\boldsymbol{M} - m_b) + c_5(\boldsymbol{M} - \boldsymbol{m_b})^2 & for\ \boldsymbol{M} \leq m_b \\ c_6(\boldsymbol{M} - m_b) & for\ \boldsymbol{M} > m_b \end{cases} \tag{5.5}$$

The source-depth scaling term is a bi-linear function conditioned on hypocentral depth, with a two corner depths d_{b1} and d_{b2}:

$$F_D = \begin{cases} m(d_{b1} - d_{b2}) + d & for\ d_{hyp} < d_{b1} \\ m\left(d_{hyp} - d_{b2}\right) + d & for\ d_{b1} < d_{hyp} \leq d_{b2} \\ d & for\ d_{hyp} > d_{b2} \end{cases} \tag{5.6}$$

Where d_{b1} = 18km for interface events, and 20km for intraslab events.

Lastly, the site term, F_S, is comprised of two components, the linear term (F_{lin}), that represents the site amplification due to impedance contrasts in the near-surface, and the nonlinear term (F_{nl}) which accounts for soil damping and softening at high strain levels. The two terms are summed in natural logarithm space:

$$F_S = F_{lin} + F_{nl} \tag{5.7}$$

The functional forms for the linear and nonlinear terms are given in Eqs. 5.8 and 5.9-10. The linear term is tri-linear in V_{S30} space, but only data from Taiwan shows a break in slope at V_1, similar to that observed in previously in Japan (Campbell and Bozorgnia 2014) and CENA (Parker et al. 2019; Hassani and Atkinson 2017). In other words, for most regions s_1=s_2.

$$F_{lin} = \begin{cases} s_1 ln\left(\frac{V_{S30}}{V_1}\right) + s_2 ln\left(\frac{V_1}{V_{ref}}\right) & for\ V_{S30} \leq V_1 \\ s_2 ln\left(\frac{V_{S30}}{V_{ref}}\right) & for\ V_1 < V_{S30} \leq V_2 \\ s_2 ln\left(\frac{V_2}{V_{ref}}\right) & for\ V_{S30} > V_2 \end{cases} \tag{5.8}$$

The nonlinear term has the same functional form as the NGA-West2 Seyhan and Stewart (2014) model:

$$F_{nl} = f_1 + f_2 ln\left(\frac{PGA_r + f_3}{f_3}\right) \tag{5.9}$$

where f_1 is 0, meaning that the effect of nonlinearity disappears as PGA_r goes to 0, f_3 is taken as 0.01g across all periods, and f_2 is defined as:

$$f_2 = f_4[exp\{f_5(min(V_{S30}, 900) - 200)\} - exp\{f_5(900 - 200)\}] \tag{5.10}$$

Coefficients for the global interface model are given in Table E2(a) of the electronic supplement, and coefficients for the global intraslab are given in Table E3(a). The development of the constant, distance-scaling, magnitude-scaling, and source depth-scaling terms are discussed in Chapter 5.4. The development of the site term is treated separately in Chapter 6.

5.4.2 Regional Terms and Coefficients

Seven of the model coefficients in the global GMM are regionalized due to observed differences in the data: the constant c_0, the anelastic attenuation coefficient a_0, the magnitude breakpoint m_b, and the V_{S30}- scaling model coefficients s_1, s_2, V_1 and V_2.

Tables E2(b)-(g) in the electronic supplement give regional coefficients for the interface GMMs for use in Alaska and the Aleutians, Cascadia, Central America and Mexico, Japan, South America, and Taiwan, respectively. Tables E3(b)-(g) in the electronic supplement give regional coefficients for the intraslab GMMs for use in Alaska and the Aleutians, Cascadia, Central America and Mexico, Japan, South America, and Taiwan, respectively. For forward use in other regions, we recommend using the global model, with a range of epistemic uncertainty that represents, at a minimum, the effects of regional variation in the constant, anelastic attenuation, magnitude break-point, and V_{S30}-scaling coefficients.

5.5 MEDIAN MODEL DEVELOPMENT

The first step in model development was correcting the ground motion data to the same reference site condition. We chose to use an established site amplification model that is conditioned on V_{S30}, Seyhan and Stewart (2014) (SS14) as a first pass (denoted F_S^{atr}). Although SS14 was developed using data from active tectonic regions, a good substitute applicable to subduction zone regions did not exist at the time model development was occurring. The site-conditioned ground motions were used to develop all of the terms in Eq. (5.1) other than F_S. As described in Chapter 6, we then used the NGA-Subduction database to check and re-calibrate the linear and nonlinear site amplification terms from SS14. Finally, residuals analyses were performed using the revised site amplification model in combination with other model components. This led to minor adjustments to constant term c_0, and formed the basis for the development of aleatory variability models.

The following sections describe the development of the median models for source and path effects using the site-adjusted data. The near-source saturation model (Eq. 5.4), which is part of the path term, is discussed first. Then, other elements of the path term are then presented, followed by the magnitude and focal depth scaling.

5.5.1 Near-Source Saturation Model

The first model component our group worked on was the magnitude-dependent near source saturation model, h, given in Eq. 5.4. This model, also sometimes called the finite fault term, is necessary due to a geometric effect; at short source-to-site distances, the ground motion is controlled by the energy from the closest part of the finite fault (Yenier and Atkinson 2014; Rogers and Perkins 1996). Therefore, the ground motions appear to be from a smaller event or farther from the event. In the equivalent point source framework, this is corrected for, otherwise predictions of ground motions monotonically increase with decreasing distance, because the total energy is assumed to be released from a single point. This is done by incorporating a near source saturation term that is combined with the hypocentral distance as a geometric mean, treating the ground motion as coming from a virtual point off the fault plane to achieve the correct level of shaking (Yenier and Atkinson 2014). The same framework is commonly used in GMMs that employ rupture distance (R_{rup}), which is the closest distance from the site to any point on the fault plane. Here, this is done by combining R_{rup} with "h" (Eq 5.3-5.4).

We initially considered using the subduction data to constrain the near source saturation model, which controls the distance at which the slope of the path model changes as the distance decreases towards 0km. Figure 5.5 shows the data from NGA-Sub earthquake ID = 4000068 (**M**

8.29), where fitting of h was attempted (using data corrected to 760m/s using SS14). However, due to the typical offshore location of subduction earthquake epicenters and lack of nearby recording stations (Figure 5.4), there is not enough data close to the source to constrain this feature appropriately. Additionally, due to other magnitude-dependent terms in the model, including the magnitude-dependent geometrical spreading, and the magnitude-scaling, multiple trade-offs in the model exist. Therefore, the data from this and other subduction earthquakes cannot constrain a unique value of the h parameter; in other words, the near-source saturation model is under-determined. This is shown in Figure 5.5 by fitting two alternative path models to the data. The path models use different values of h, but provide equally good fits. This is not a new problem in GMM development for subduction regions; previous empirical determination of this portion of GMMs has been primarily for active tectonic regions (e.g. Boore et al. 2014; Yenier and Atkinson 2014; Abrahamson et al. 2014), and GMMs for subduction zones have mostly borrowed this portion of the model from other regions, or used simulations to constrain it.

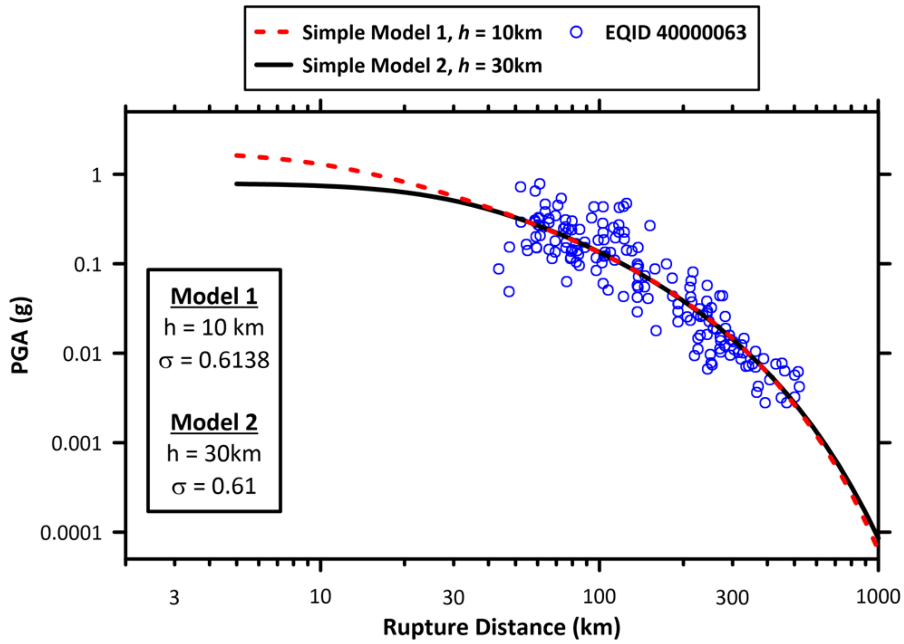

Figure 5.5. Example of the underdetermined problem of fitting *h* using NGA-Sub data. Two simple path models (Eq. 5.7) fit to an interface event from Japan with M = 8.29, one with *h* = 10, and *h* = 30km Both models produce a fit with negligible differences in the residual standard error (0.6138 versus 0.61, respectively).

Therefore, to constrain the *h* model given in Eq. 5.4, we used a combination of empirically based estimations of *h* from active tectonic regions at small magnitudes (Atkinson et al. 2016; Yenier and Atkinson 2014; YA14), and a suite of EXSIM simulations performed as part of the present work at large magnitudes (Figure 5.6). Atkinson et al. (2016) looked at a number of events that are well recorded at short source-to-site distances from the Geysers region of California to better constrain near source saturation effects for small magnitude earthquakes (**M**1.5-3.6). Their results support the modeling of these effects with the equivalent point-source approach, and validate the near source saturation model of YA14.

To constrain *h* at the large magnitudes necessary for subduction zone earthquakes, we ran a suite of ground motion simulations using EXSIM, an open-source stochastic finite-source simulation algorithm (Motazedian and Atkinson, 2005; Boore, 2009; Assatourians and Atkinson, 2012). Input parameters were chosen by looking at the properties of interface events in the NGA-Sub database; this work was started by Nicolas Kuehn and modified for the present application. Simulations were run for earthquakes with **M**= 3.75-9.5 in 0.25 magnitude unit intervals, with 5 runs per magnitude. For each run, the fault length and width were generated randomly using Strasser et al. (2010), and the hypocenter location on the fault plane was randomly sampled with a uniform distribution over the fault plane (N. Kuehn 2018, *pers. comm.*). Stress drop was taken as 150 bars. Depth to the top of rupture was limited at 5km to maintain a small source to site distance, and a fault dip between 15°-28° was assigned, in line with interface events in the database. Ground motions were generated at 36 sites located at 12 distances between 10 and 1000km along three azimuths (45°, 60° and 90°). The simulations use a simple attenuation model, with a geometrical spreading coefficient equivalent to c_1 in Eq. 5.2. Based on initial observations of empirical distance scaling, this input was set as -1.3.

Once the simulated ground motions were generated, the PSA values from the 5 runs per magnitude were combined, and a simple path model was fit to each magnitude bin,

$$F_{P,EXSIM} = c_0 + c_1 logR + c_2 R \tag{5.7}$$

where R is defined by Eq. (5.4). Coefficients c_1 and c_2, representing geometrical spreading and anelastic attenuation effects, respectively, were fit first using the simulated data at $R_{rup} \geq 40$km to avoid the influence near-source saturation effects at closer distances. Then, with c_1 and c_2 fixed, h and c_0 were fit using the simulated data over the entire distance domain. Figure 5.6 shows the resulting median h values along with their 95% confidence intervals for PGA, PGV, and 0.5 and 5.0s PSA as a function of magnitude, along with the empirical values and model from YA14. The estimates of h do not vary appreciably or in a systematic manner with period.

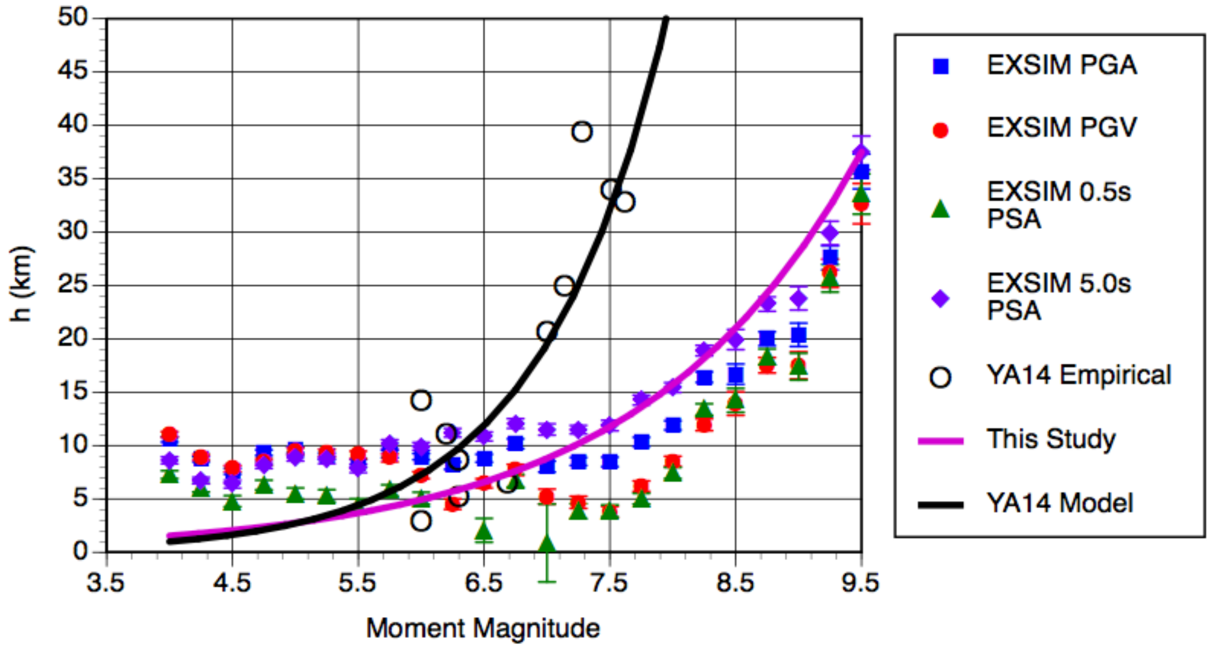

Figure 5.6. A comparison of the Yenier and Atkinson (2014) near-source saturation model (labeled YA14 above), *h* estimates from EXSIM simulations run to emulate subduction interface events, and the near source saturation model developed in this study, given in Eq. 5.4.

Figure 5.6 shows our proposed near source saturation model (Eq. 5.4), which is period-independent. We elected to maintain the functional form of the YA14 model, and constrained our model to be similar to the values of YA14 for magnitudes up to 5.5. However, we depart from YA14 and follow the trend of the EXSIM results for $5.5 < \mathbf{M} \leq 9.5$ (Figure 5.6).

5.5.2 Distance Scaling Model

With the near-source saturation term of the model set, we moved on to fitting the remaining elements of the path model, F_P (Eq. 5.2). The path model has two components, the geometrical spreading (GS) term, and the anelastic attenuation term. The geometrical spreading term represents the decay of energy from as it moves from a point source along a spherical wave front. In an idealized homogeneous elastic half-space, the energy at a point on the radius of the sphere will decay as R^{-1}. However, heterogeneities in the Earth cause a conversion of body waves to surface waves, and therefore the exponent is not unity. This empirical value is represented by coefficient c_1 in Eq. 5.2. The transition from Fourier amplitude space (FAS) to response spectra (RS) space introduces a magnitude-dependence in this term (Hassani and Atkinson 2018), which is represented by the $(b_3 + b_4\mathbf{M})lnR$ term in Eq. (5.2). In model space, these terms control the slope of decay with the natural logarithm of R, as shown in Figure 5.7.

The anelastic attenuation term represents the per-cycle energy dissipation, and is a property of the material the seismic wave is traveling through. In model space, this term controls the curvature of decay with the natural logarithm of R, which strongly influences the rate of distance attenuation at large distance (Figure 5.7).

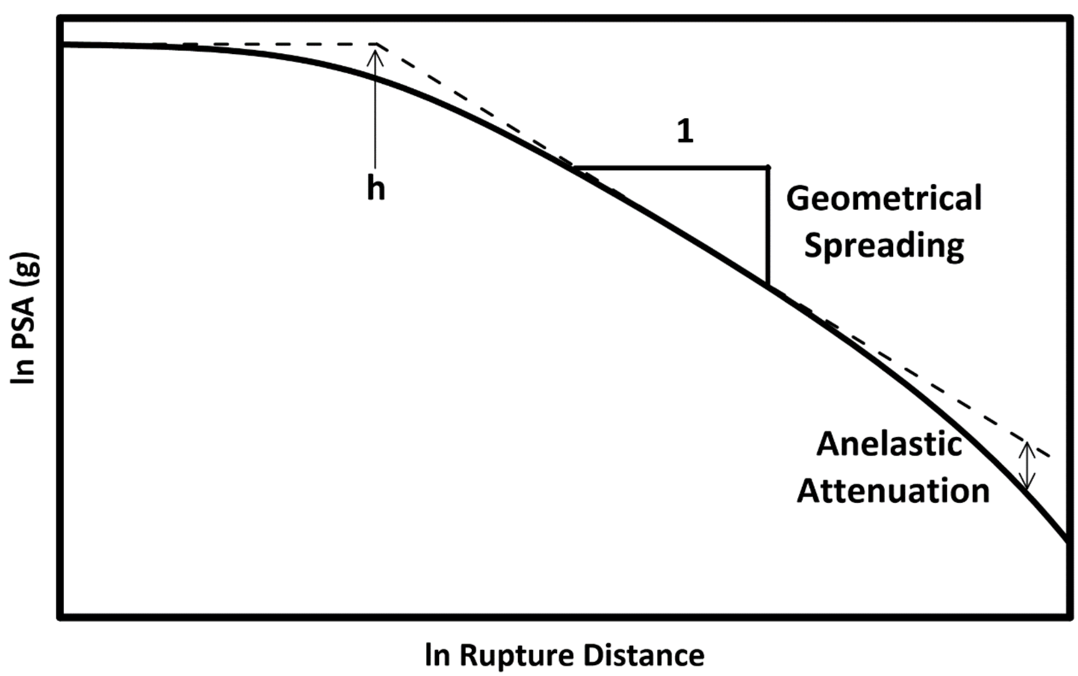

Figure 5.7. Schematic of path model, F_P (Eq. 5.2), showing the near-source saturation model, h, the geometrical spreading slope, and the curvature due to the anelastic attenuation term.

In order to fit the path model without a source term, we initially binned the data by magnitude and fit the data using the c_0 and F_P terms in Eq. (5.1). This allowed us to investigate the magnitude dependence of the geometrical spreading. The path coefficients derived from this process are the same for each magnitude bin, but the constant c_0 has a random effect on the magnitude bin; this constant was subsequently re-evaluated over the entire magnitude range of the dataset as part of the source term analyses. Our analyses indicated that the coefficients for the magnitude-dependent component of the geometrical spreading, specifically coefficient b_4, could be adopted from Hassani and Atkinson (2018; hereafter HA18). HA18 takes the generic point-source simulation-based GMM of Yenier and Atkinson (2015) and modified it to enable adjustments to the near surface attenuation parameter κ_0. These coefficients represent the transition of the geometrical spreading term between FAS and response spectral ordinates. The c_1 coefficient, which represents the geometrical spreading slope in FAS, is set empirically via regression.

Despite the large size of the NGA-Sub database, it is not possible to constrain both the slope and curvature of the path model simultaneously due to substantial trade-offs between these two model components. We address this by fitting c_1 to the subset of data with $R_{rup} \leq 125$km in order to avoid the portion of the data with the most curvature at large distances (Figure 5.8). These analyses were initially performed using b_3 and b_4 from HA18. Any period dependence that was observed in c_1 was transferred into b_3, smoothed across periods, and then c_1 was re-fit. The average of the second iteration of c_1 over the period domain was taken as the global value, with different values for interface and intraslab events. As shown in Figure 5.8, data from intraslab events show steeper geometrical spreading than data from interface events, which is similar to some prior results (Atkinson and Boore 2003; Abrahamson et al. 2016).

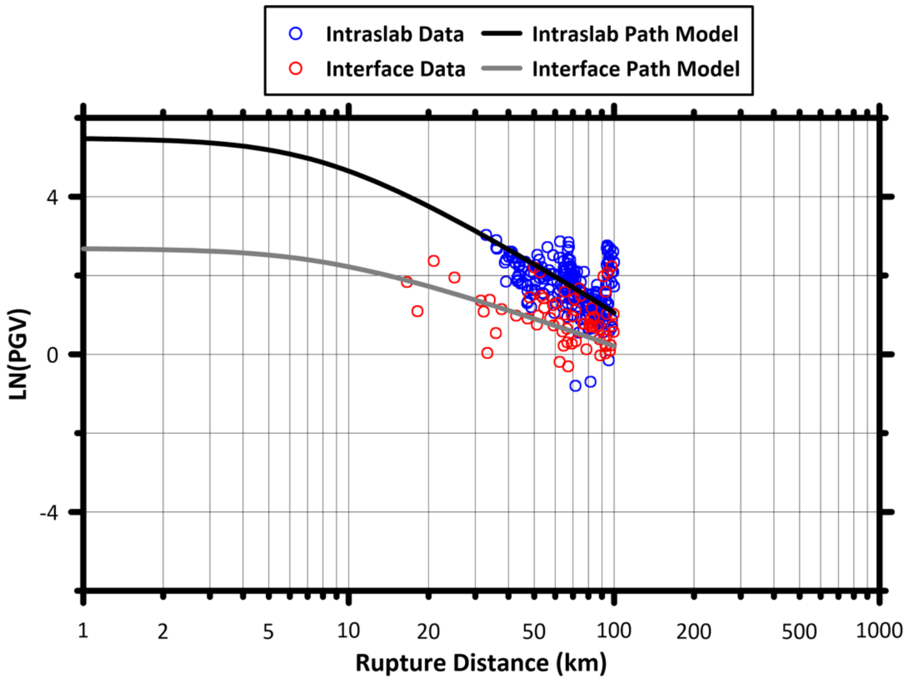

Figure 5.8. Example of peak ground velocity (PGV) data for magnitude bin 6.5-7 over the distance range ($R_{rup} \leq$ 100km) with decay controlled by geometrical spreading.

With the GS coefficients fixed, the anelastic attenuation coefficient, a_0, was fit with a random effect on region in order to produce both a global value and regional values. All values of a_0 were smoothed with respect to period, and constrained to go to zero at 10s, as the per-cycle damping at long oscillator periods is negligible. As shown in Figure 5.9, the anelastic attenuation is less for intraslab events than for interface events, indicating that although the interface data shows slower overall distance attenuation, there is more curvature in this data at large distances. The smoothed global and regional values of a_0 for interface and intraslab events are shown in Figure 5.10. In general, the anelastic attenuation in Central America and Mexico is less than the global value (absolute value of a_0 is smaller), the anelastic attenuation in South America and Alaska are closest to the global values, and the anelastic attenuation in Cascadia, Japan, and Taiwan is larger than the global value.

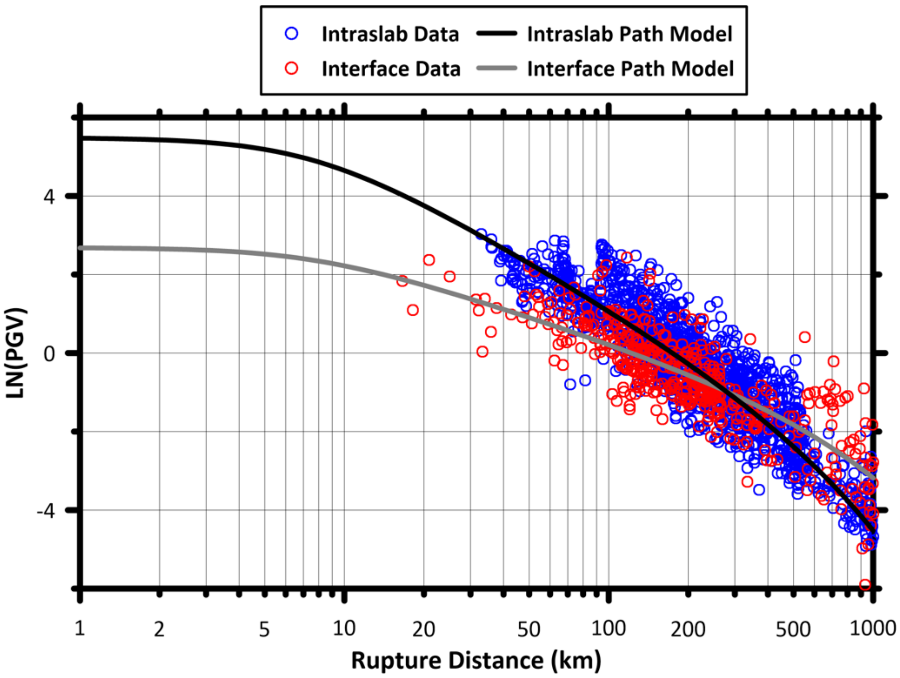

Figure 5.9. Example of peak ground velocity (PGV) data for magnitude bin 6.5-7 over the entire model distance domain ($R_{rup} \leq 1000$km) with the geometrical spreading terms from Figure 5.8, plus the best fit global anelastic attenuation term. Results shown for interface and intraslab events.

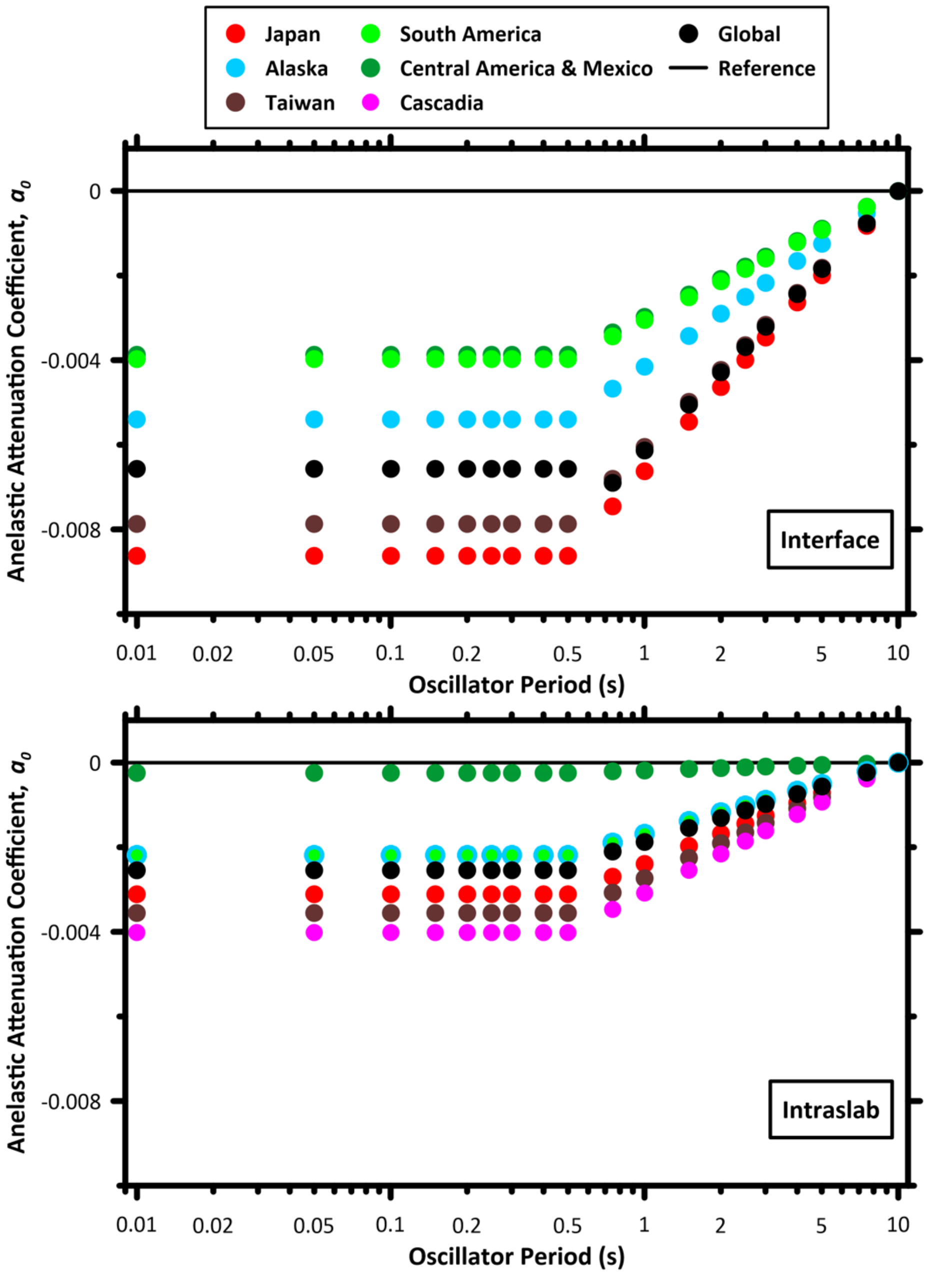

Figure 5.10. Anelastic attenuation coefficient, a_0, as a function of oscillator period for interface events (top) and intraslab events (bottom). Lack of data for interface events in Cascadia means there is no regional value of a_0; instead the global value is recommended.

5.5.3 Magnitude Scaling Model

Once the path model was set (Eqs. 5.2-5.4), it was subtracted out of the data and the resulting adjusted ground motions were used to fit the magnitude-scaling model. Equation 5.5 was first fit to the data with all coefficients set by the regression except for $\boldsymbol{m_b}$, which was constrained based on geometrical considerations specific to each subduction zone region. In the case of intraslab earthquakes, down-dip width of the event is limited by slab thickness. Events that rupture through the full slab thickness are expected to saturate when the rupture aspect ratio (i.e., ratio of along-strike length to down-dip width) exceeds about unity. This occurs because increasing magnitude produces increasing rupture far from the site, which would be expected to have little impact on high frequency ground motions (i.e., saturation). This concept was verified through the use of simulated data by Archuleta and Ji (2018), who also provided saturation magnitudes specific to each of the regions considered in NGA-Sub. We take these saturation magnitudes as $\boldsymbol{m_b}$ for use in Eq. (5.5) for intraslab events (given in Tables E2-E3 in the electronic supplement). In the case of interface events, Campbell (201x) has similarly derived saturation magnitudes, but instead of slab thickness, the seismogenic fault width is used to constrain the saturation magnitude. Table 5.2 gives these computed saturation magnitudes for each region considered in this study.

Table 5.2. Regional saturation magnitudes for interface events computed using seismogenic fault width (Campbell 201x) and for intraslab events computed using slab thickness (Archuleta and Ji 2018).

Region	Interface Saturation Magnitude (Campbell 201x)	Intraslab Saturation Magnitude (Archuleta and Ji 2018)
Alaska	8.0	7.20
Aleutian Islands	8.0	7.98
Cascadia	7.56	7.20
Northern CAM	7.45	7.40
Southern CAM	7.50	7.60
Japan – Kuril-Kamchatka Trench (Pacific Plate)	8.31	7.65
Japan – Nankai-Ryukyu Trench (Philippine Sea Plate)	7.28	7.55
Northern South America	8.49	7.30
Southern South America	8.49	7.25
Taiwan	8.0	7.70

With $\boldsymbol{m_b}$ fixed in this manner, the other magnitude-scaling coefficients were treated as fixed effects, and the constant c_0 was treated as a random effect conditioned on region and NGA-Sub earthquake ID. Using the results of this initial regression, c_5, the parameter that controls the parabolic behavior of the model below the break point, was smoothed and constrained.

Initially we expected to enforce $c_5 \leq 0$, meaning that the model would either be linear or concave downwards at small-to-moderate magnitudes. This expectation was met for interface events; the data displays curvature in this magnitude range that is concave downwards (Figure 5.11). However, the intraslab data at short periods exhibit different behavior. We attempted to fit the model with $c_5 = 0$, but this caused a positive bias in the resulting between event residuals at small magnitudes (**M** = 4-5). We then considered allowing a tri-linear model with an additional breakpoint at **M** ≅ 5, but did not want to introduce the complexity of a changing F_M functional form as a function of oscillator period. Ultimately, we decided to allow $c_5 > 0$ for short-period intraslab events as the data demands (shown for T=0.2s in Figure 5.11). After setting c_5, the mixed-effects regression was re-run to fit c_4 and c_6, the slope values before and after the breakpoint, respectively. These coefficients were constrained such that $c_6 \leq c_4$, in order to enforce saturation in the magnitude-scaling.

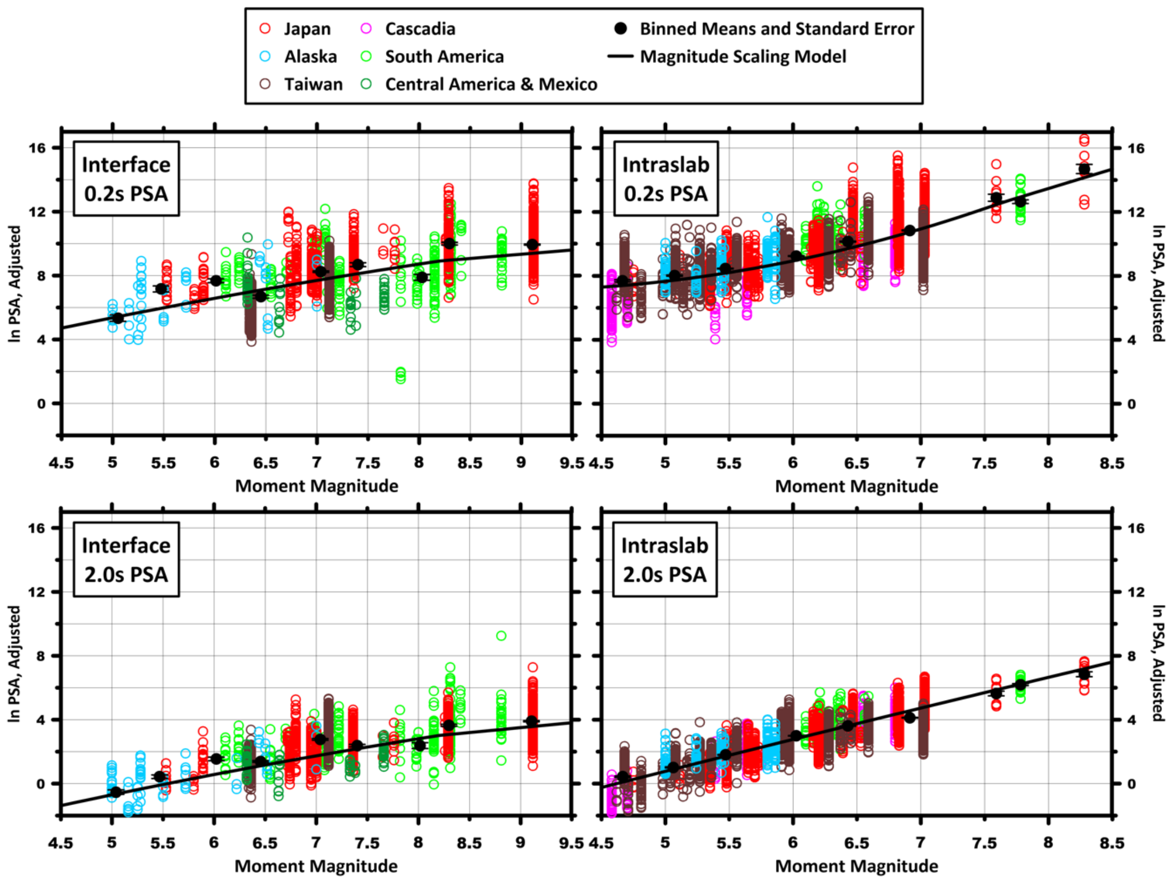

Figure 5.11. Global interface (left) and intraslab (right) magnitude-scaling models (F_M; Eq. 5.5) and site and path term-corrected data as a function of **M** for 0.2 and 2.0s PSA. For plotting purposes, the average of regional $\boldsymbol{m_b}$ values weighted by number of recordings was used for the intraslab and interface model.

5.5.4 Depth Scaling Model

At any stage of the model development process, it is possible to compute residuals, which are useful for examining model performance relative to predictor variables. Total residuals between an observed intensity measure for event i and site j (Y_{ij}) are computed as:

$$R_{ij} = ln(Y_{ij}) - \mu_{lnY}(\mathbf{M_i}, R_{rup,ij}, V_{S30,j}) \tag{5.8}$$

where μ_{lnY} is the natural log mean from the GMM at a particular step of model development. Total residuals R_{ij} can be partitioned into mean bias term (c_k), between-event residuals (η_{Ei}), and within-event residual (δW_{ij}) using mixed effects analysis (Bates et al., 2015; R Development Core Team, 2008).

$$R_{ij} = c_k + \eta_{E,i} + \delta W_{ij} \tag{5.9}$$

We developed the depth scaling model based on between-event residuals (also known as event terms) computed using site-adjusted data and the source and path models described in previous sections (i.e., $\mu_{lnY} = c_0 + F_P + F_M + F_S^{atr}$). Those event terms were examined for trends with earthquake source depth. Three measures of depth were considered: hypocentral depth (d_{hyp}), depth to top of rupture (Z_{tor}), and depth to the mid-point of the fault in the down-dip direction (Z_{mid}). There are two general considerations in selecting an appropriate depth metric – (1) predictive power and (2) convenience for forward application.

From a technical rigor point of view, we consider d_{hyp} to be preferred for two reasons: (i) it is a more fundamental parameter related to earthquake stress drop, especially for subduction zone earthquakes (Bilek and Lay 1998, 1999); and (ii) we believe there is less uncertainty in estimates of d_{hyp} than of Z_{tor} because the majority of events in the NGA-Subduction database do not have a published finite-fault model available, and thus have estimates of Z_{tor} from simulations (e.g. Contreras 2017). Unfortunately, d_{hyp} is not convenient for application, because including it in a hazard analysis would require adding randomization of source location on the fault, which would involve an additional loop in the hazard integral. Both Z_{tor} and Z_{mid} are equally convenient, because they are determined once a fault rupture plane is defined (which is already part of hazard analysis for distance calculation). Hence, no additional randomization (hazard integral loops) are required in forward analysis. For the present analysis, we use d_{hyp}, but as part of future work, we plan to investigate the application of empirical fault plane geometry – d_{hyp} relationships (e.g. Mai et al. 2005).

The bi-linear model given in Equation 5.6 was used for both the interface and intraslab events, with differing coefficients. We initially fit Eq. (5.6) to the event terms using a nonlinear least-squares regression with all parameters free. Based on these results, a single corner depth, d_b, was chosen for all periods: 27 km for interface events, and 67 km for intraslab. Then the regression was repeated with the corner depth constrained, and the slope m, and coefficient d were iteratively smoothed. The model slope m goes to 0 at the lower end of the depth range populated with data, which is18km for interface and 20km for slab. The model goes to zero at 0.75s for interface, and 2.0s for intraslab, as the increase in ground motion amplitudes due to increased stress is only observed at short periods. Figure 5.12 shows the model for PGA, 0.2s, and 1.0s for both event types.

This trend shown in Figure 5.12 can be interpreted as a consequence of the stress drop increasing with increasing depth in the equivalent point source, and has been observed previously for shallow events in active tectonic regions (Yenier and Atkinson 2015b; Hassani and Atkinson 2018), for events in stable continental regions such as CENA (Yenier and Atkinson 2015a), and for induced earthquakes in Oklahoma (Novakovic et al. 2018). Atkinson and Boore (2003) also has a linear source depth scaling term in their GMM for subduction zones.

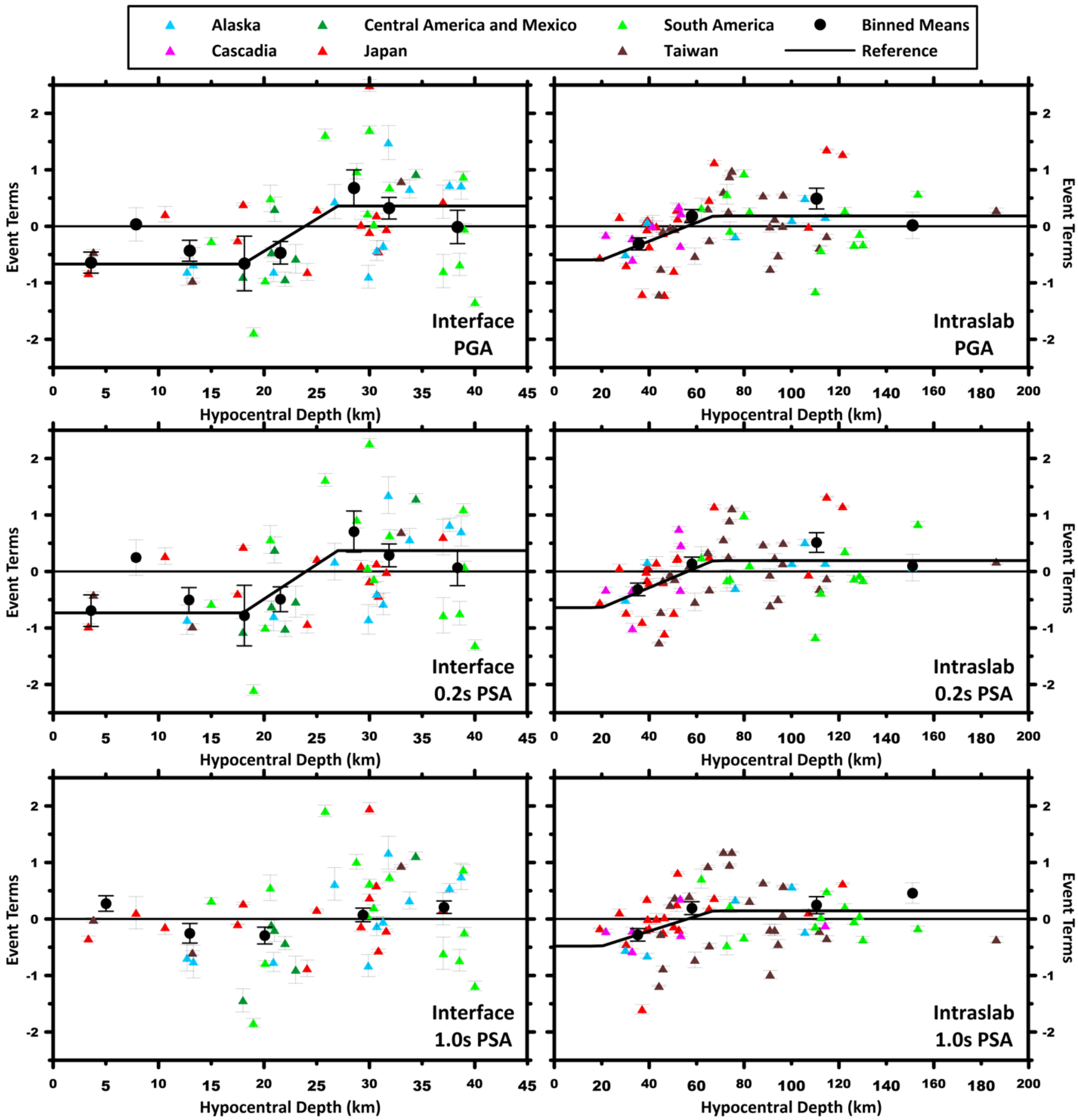

Figure 5.12. Variation of event terms as a function of hypocentral depth at PGA, 0.2s PSA and 1.0s PSA. Interface events are shown on the left and intraslab events are shown on the right. Binned means with standard errors and best-fit depth scaling model (Eq. 5.6) shown for both event types.

5.5.5 Regional and Global Constant Calibration

The last step in model development for the reference rock GMM was the determination of the global and regional model constants, c_0 (Eq. 5.1), and final event terms, η_E, through a mixed-effects residuals analysis. Total residuals are computed using Eq. (5.8) with the mean GMM taken as:

$$\mu_{lnY,ij} = F_P\left(R_{rup,ij}\right) + F_M(\boldsymbol{M_i}) + F_D\left(d_{hyp,i}\right) + F_S^{atr}(V_{S30}) \tag{5.10}$$

Then, the total residual for each recording was partitioned into the global constant, c_0, modifier on the constant for region k, $\Delta c_{0,k}$, and event terms $\eta_{E,i}$ using linear mixed effects in R (Bates et al. 2015):

$$R_{ijk} = c_0 + \Delta c_{0,k} + \eta_{E,i} + \delta W_{ij} \tag{5.11}$$

Eq. (5.11) is equivalent to Eq. (5.8), but with the general bias term c_k replaced with the sum of global constant c_0 and the regional variation $\Delta c_{0,k}$.

The global constant that is produced by the mixed effects analysis is an unweighted average of the regional constants. However, we think this gives too much weight to some regions without a large population of data. Instead, we compute the global constant by taking the average of the regional values, weighted by the inverse of their variances (i.e. the square of the standard deviation). The resulting global constant is shown as a function of period for interface and intraslab events in Figure 5.13, along with the regional constants for comparison.

Overall, the ground motion amplitude for interface events is lower than that of intraslab events (Figure 5.13). For intraslab events, the regional constants converge towards the global value at long periods (>0.75s). At short periods, Japan and South America have larger amplitudes than the average, and Cascadia and Central America and Mexico are slightly lower than the average.

For interface events, Japan has amplitudes larger than the global average. Alaska and Central America and Mexico have ground motion amplitudes lower than the average, and Taiwan and South America are near the average. Due to lack of data for interface events, we do not have an empirically-derived regional constant for Cascadia. Recommendations for forward use of both the interface and intraslab model in Cascadia are given in Chapter 5.6.2.

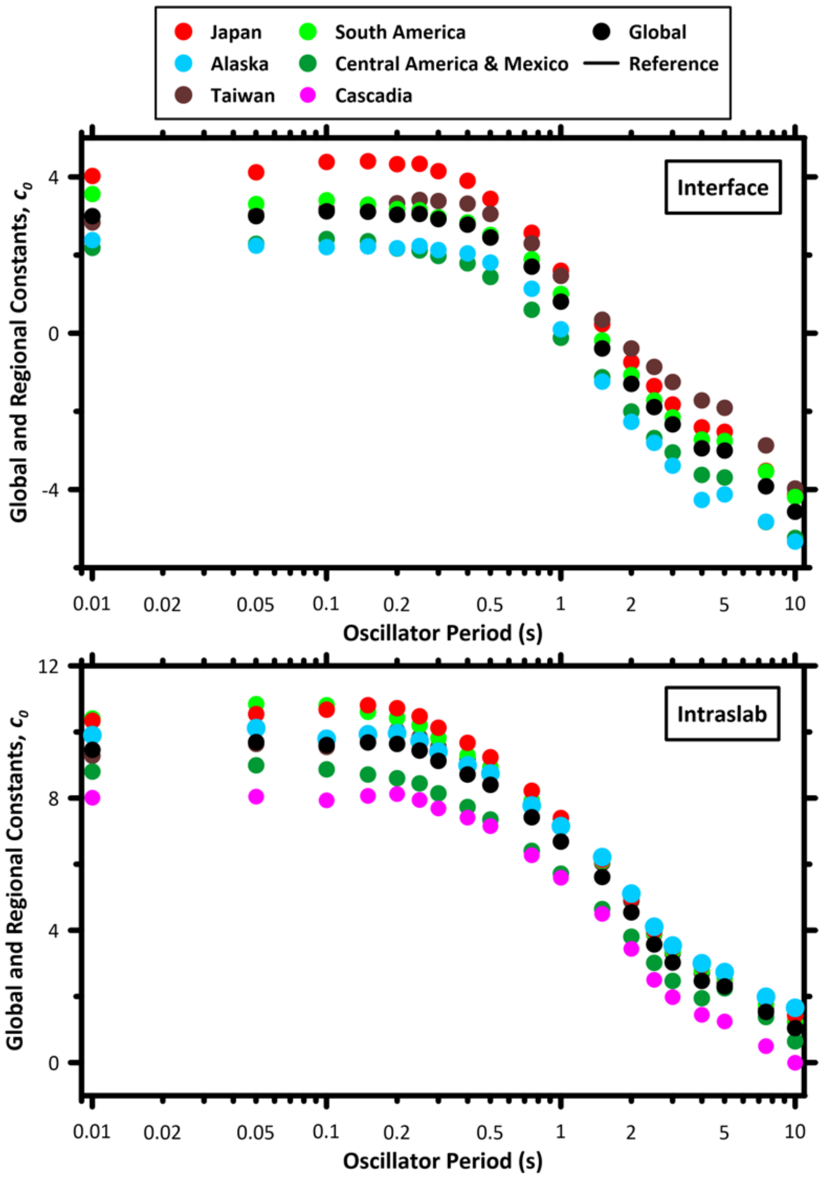

Figure 5.13. A comparison of the global model constant (c_0) with regional constants for interface (top) and intraslab (bottom) earthquakes.

5.5.6 Site Response

The development of the GMM, from path effects through to the constant term, was iterative with respect to the site response model. As described previously, we initially used the V_{S30}-based site response model for active tectonic regions (F_S^{atr}) from SS14 to develop all source- and path-related model coefficients. The resulting model is used in Chapter 6 to develop a site response model that is specific to subduction regions, including regional effects as appropriate. This site response model includes:

1. Global period-dependent V_{S30}-scaling specific to subduction zones. This scaling is the same for both source types (interface and intraslab).
2. Regional adjustment factors for the global V_{S30}-scaling coefficients (available for Alaska, Cascadia, Japan, South America, and Taiwan).
3. Nonlinear site response model that is adjusted from that in SS14.

All of these model elements were derived from the NGA-Sub data using the source and path models described here. With the site response model updated from F_S^{atr} (SS14) to F_S (Chapter 6), the global and regional constant terms were re-computed. The values shown in Figure 5.13 are based on the initial model; the trends following updating are qualitatively similar. The values tabulated in the electronic supplement reflect the updating to the subduction site amplification model.

5.6 MODEL RESIDUALS AND ALEATORY VARIABILITY

5.6.1 Model Residuals

Two types of model residuals were considered: within-event residuals (δW_{ij}) and between-event residuals ($\eta_{E,i}$), or event terms (Al Atik et al. 2010). The event terms and the within-event residuals are computed using Eqs. 5.8-5.9, with the final subduction-specific site amplification model F_S. Residuals analyses were performed to check model performance with respect to predictor variables. Between-event residuals are used to check the source model; within-event residuals are used to check the path and site models.

The event terms are shown as a function of moment magnitude for PGA, 0.2, 1.0, and 5.0s for the interface model in Figure 5.14, and the intraslab model in Figure 5.15. The event terms were computed using regional terms where applicable, and are color-coded by region in each plot. The overall regional mean (averaged across magnitudes) is zero for the data set as a whole, and for each region, due to the calibration of c_0 described in Chapter 5.4.5. Binned means are shown for the global data set in the figures. Not all regional data sets have a sufficient number of events over a wide enough **M** range (> ~ 2 **M** units) to judge model effectiveness (e.g., interface – Taiwan and Central America and Mexico; slab – Alaska and Central America and Mexico). For the other regions, the event terms do not appear to trend with magnitude. Similarly, the event terms are shown as a function of hypocentral depth for PGA, 0.2, 1.0, and 5.0s for the interface model in Figure 5.16, and the intraslab model in Figure 5.17.

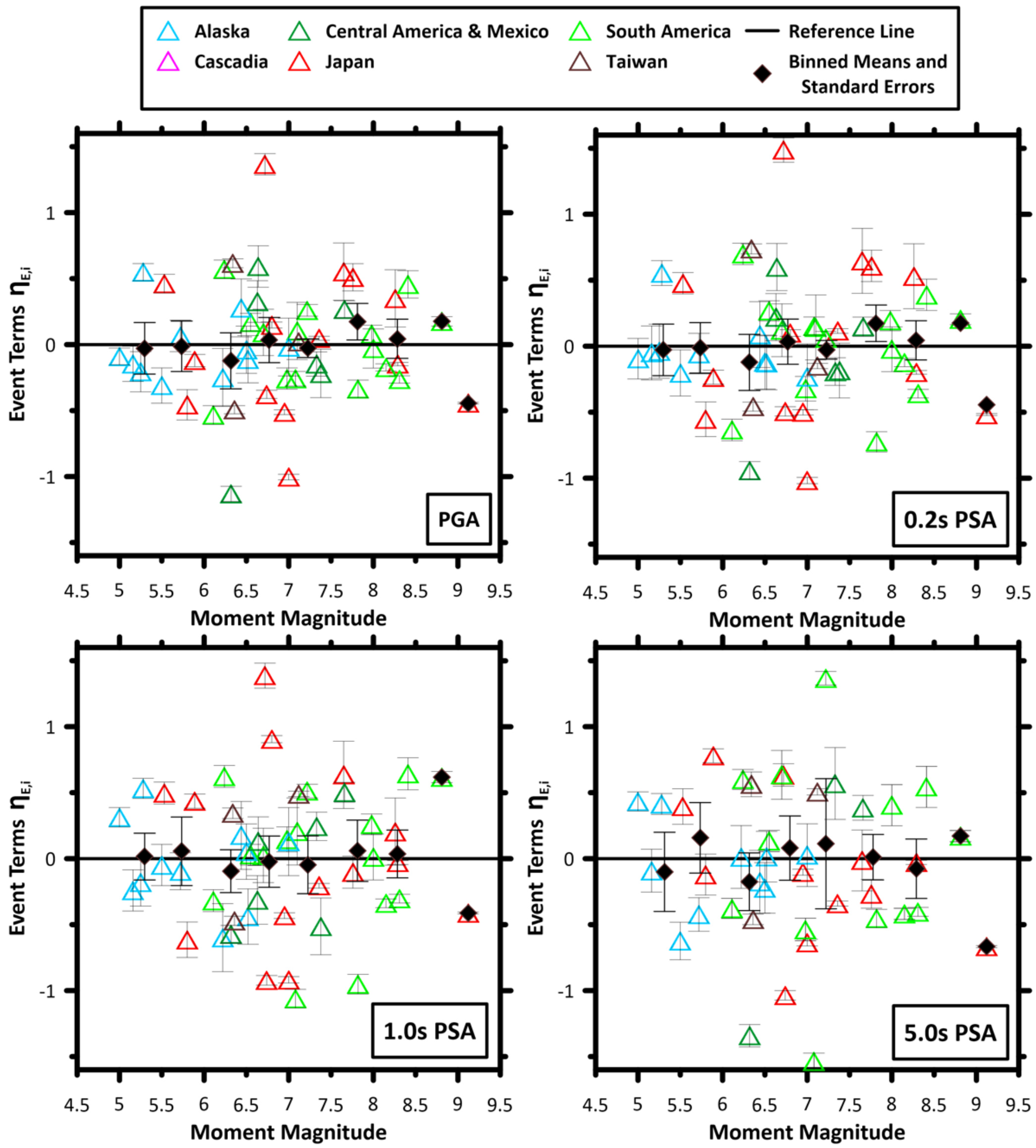

Figure 5.14. Event terms, $\eta_{E,i}$, from interface events, computed using mixed effects analysis (Eq. 5.9) as a function of moment magnitude for PGA and 0.2s, 1.0s, and 5.0s PSA. Event terms are color-coded by subduction zone region, and plotted with their standard errors (gray bars).

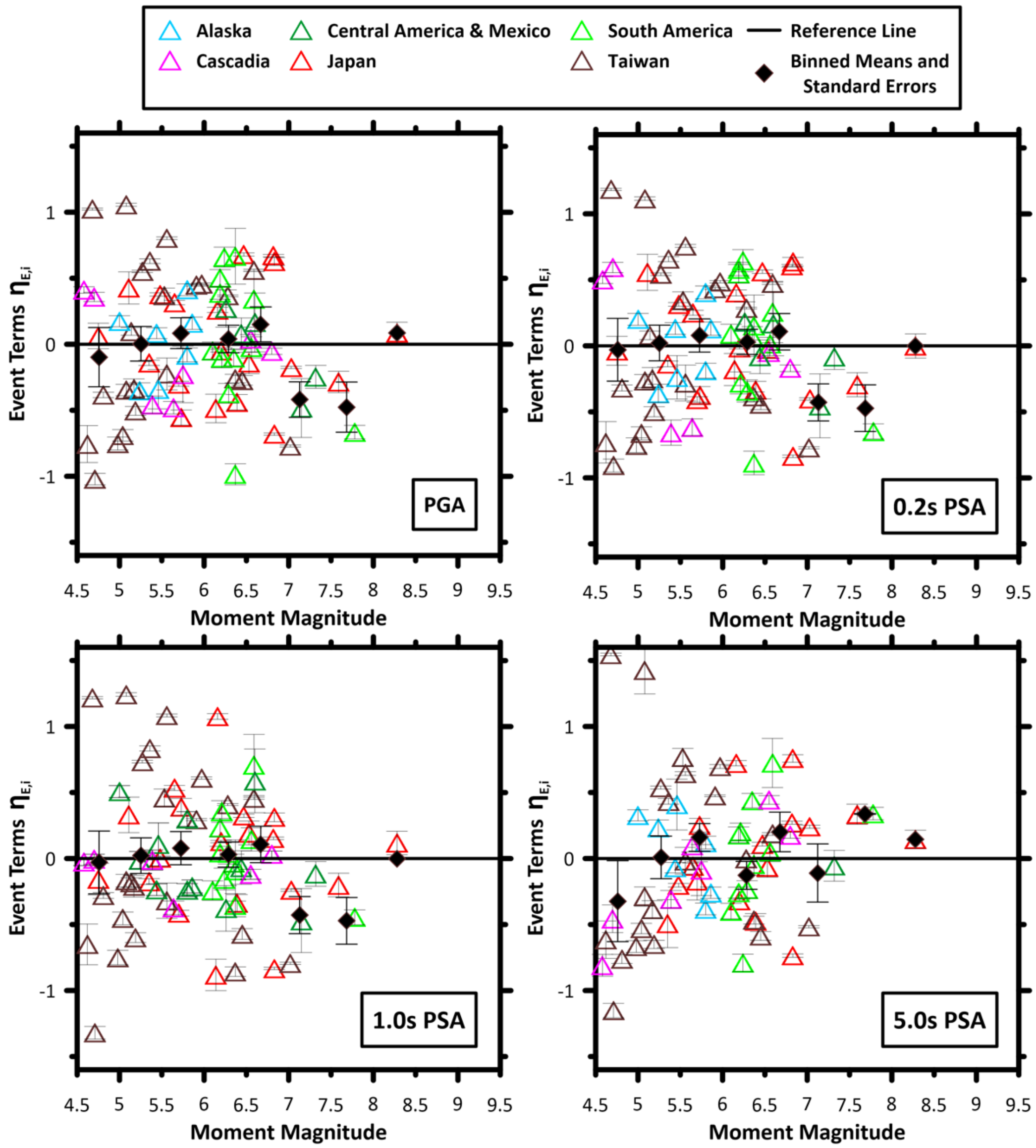

Figure 5.15. Event terms, $\eta_{E,i}$, from intraslab events, computed using mixed effects analysis (Eq. 5.9) as a function of moment magnitude for PGA and 0.2s, 1.0s, and 5.0s PSA. Event terms are color-coded by subduction zone region, and plotted with their standard errors (gray bars).

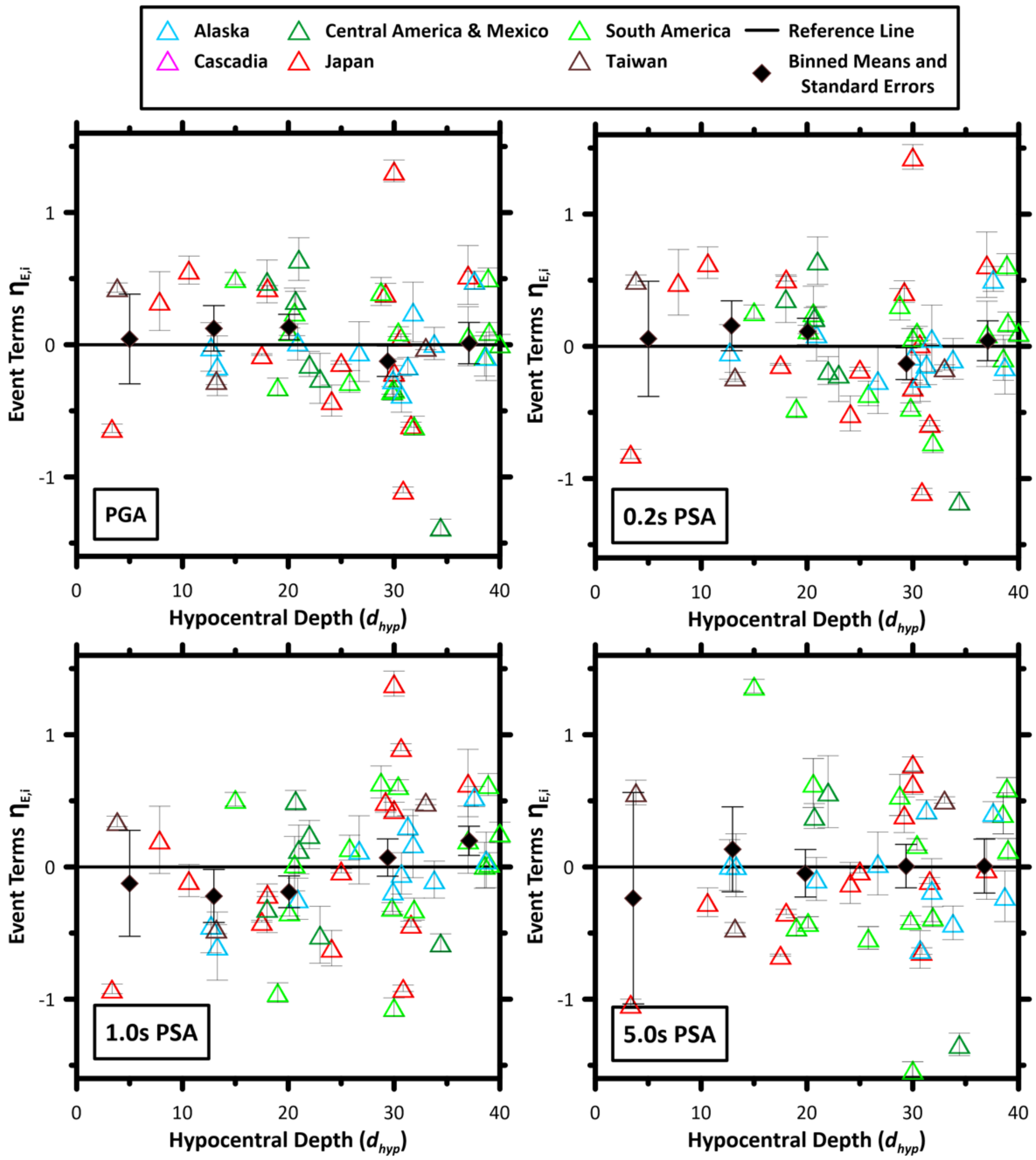

Figure 5.16. Event terms, $\eta_{E,i}$, from interface events, computed using mixed effects analysis (Eq. 5.9) as a function of hypocentral depth for PGA and 0.2s, 1.0s, and 5.0s PSA. Event terms are color-coded by subduction zone region, and plotted with their standard errors (gray bars).

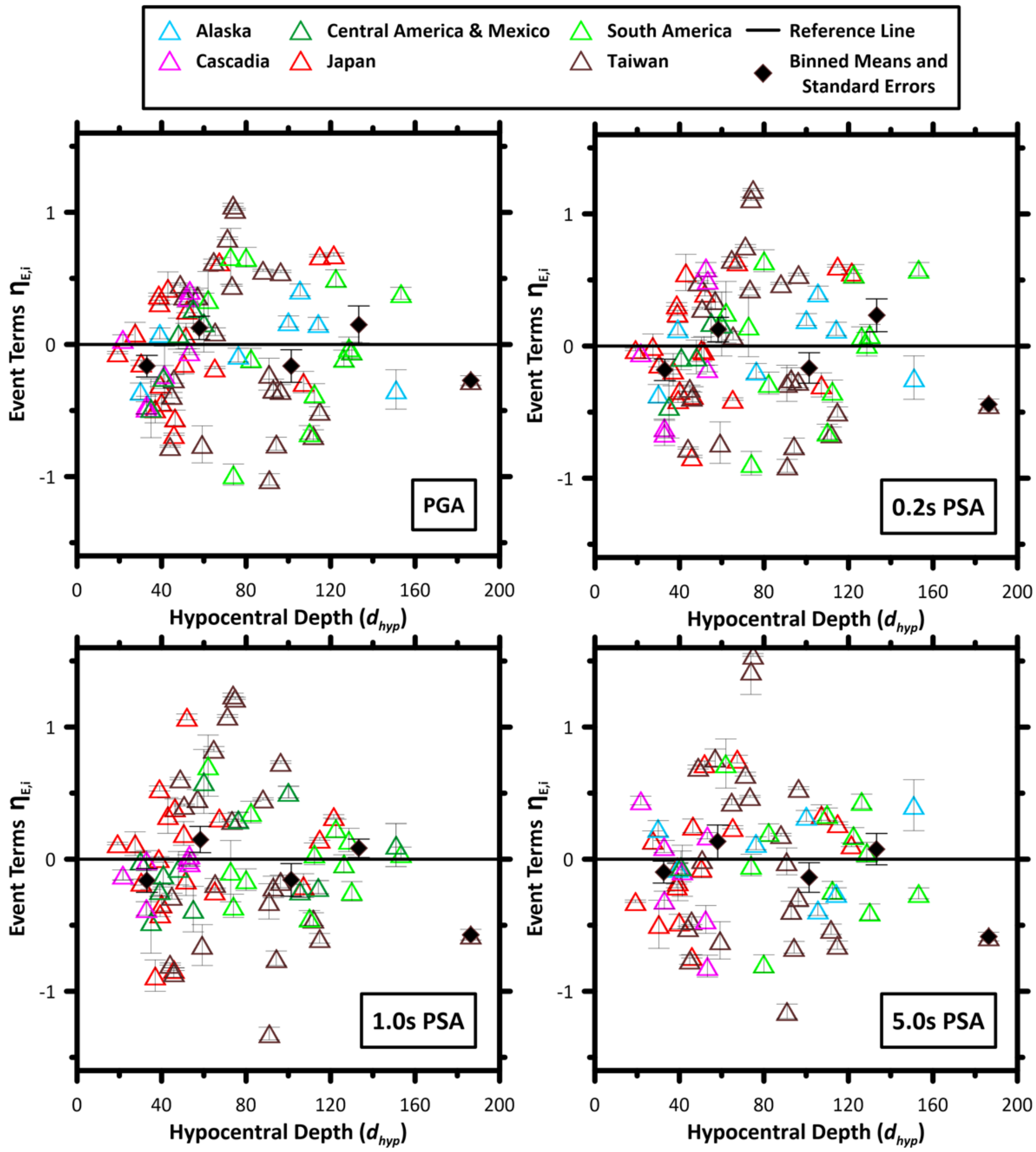

Figure 5.17. Event terms, $\eta_{E,i}$, from intraslab events, computed using mixed effects analysis (Eq. 5.9) shown as a function of hypocentral depth for PGA and 0.2s, 1.0s, and 5.0s PSA. Event terms are color-coded by subduction zone region, and plotted with their standard errors (gray bars).

The within-event residuals are shown as a function of distance for PGA, 0.2, 1.0, and 5.0s for the interface model in Figure 5.18, and for the intraslab model in Figure 5.19. Residuals were computed using regional terms where applicable, and are color-coded by region in each plot. Both for the overall data set and for the individual regional sets, the trend of residuals with distance are reasonably flat. For a clearer view of regional trends in residuals, the binned means of within-event residuals for each region are shown in Figures 5.20-5.21.

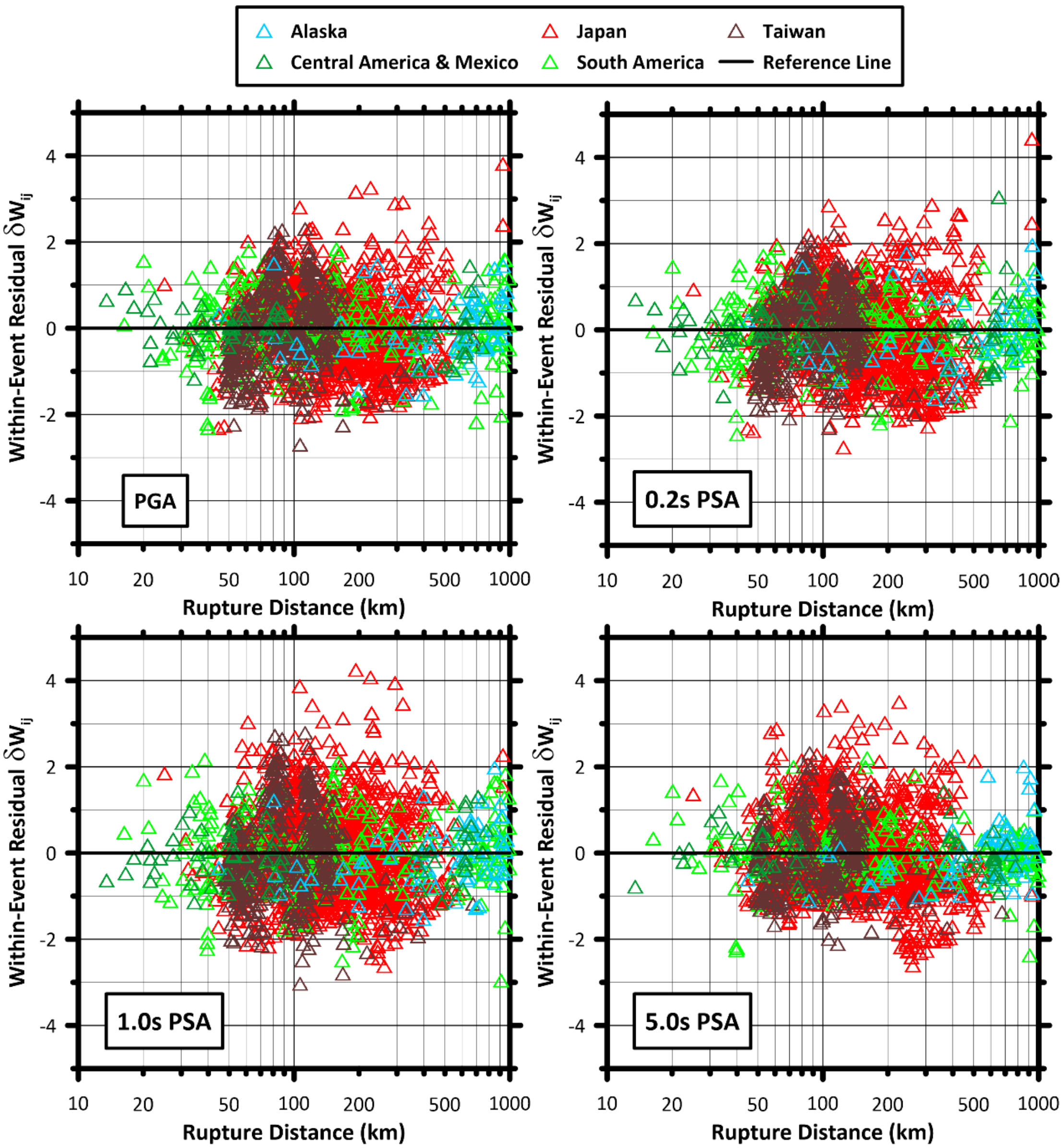

Figure 5.18. Within-event residuals, δW_{ij}, from interface events, computed using Eq. 5.10 for PGA and 0.2s, 1.0s, and 5.0s PSA. Residuals are color-coded by subduction zone region.

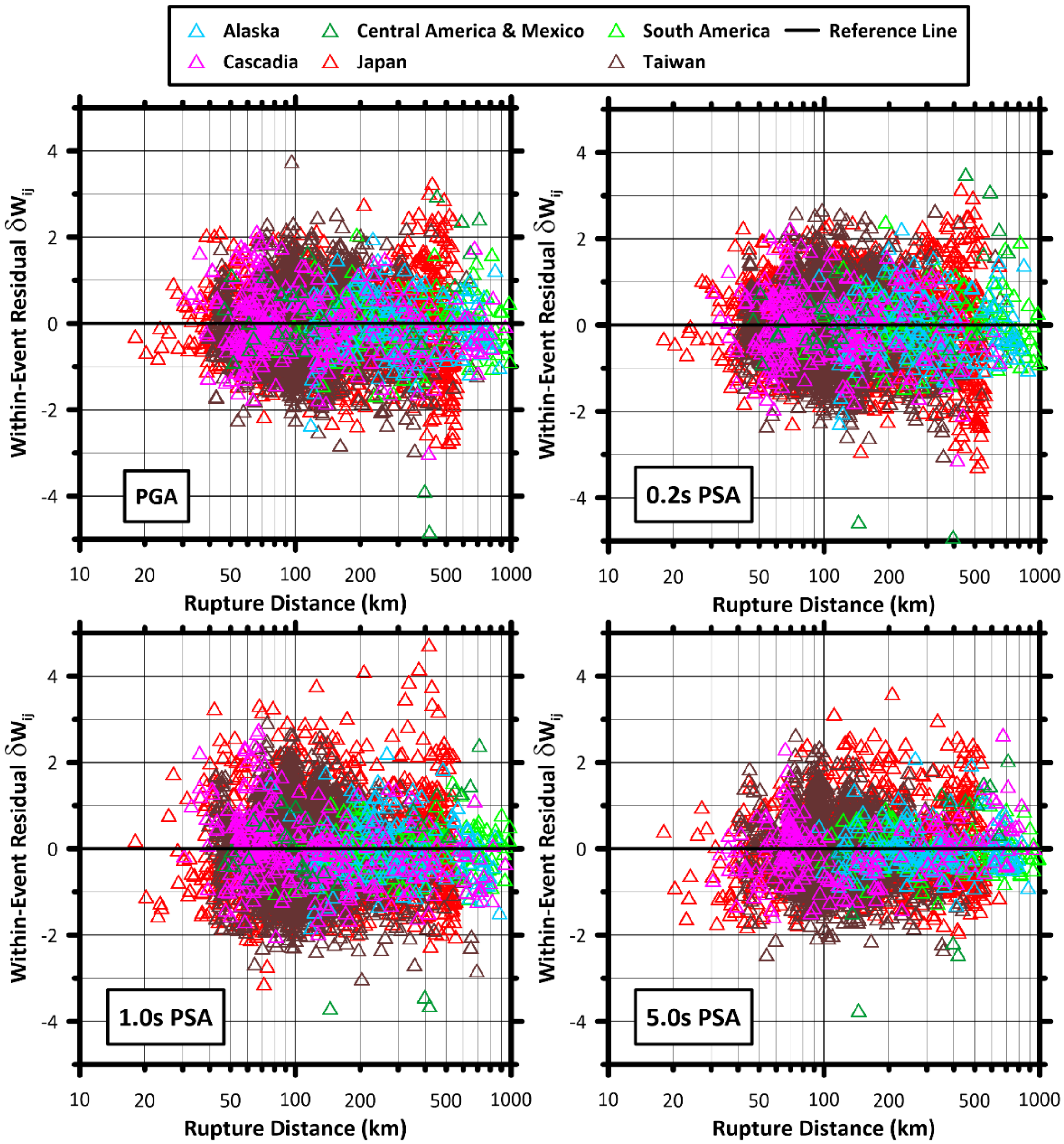

Figure 5.19. Within-event residuals, δW_{ij}, from intraslab events, computed using Eq. 5.10 for PGA and 0.2s, 1.0s, and 5.0s PSA. Residuals are color-coded by subduction zone region.

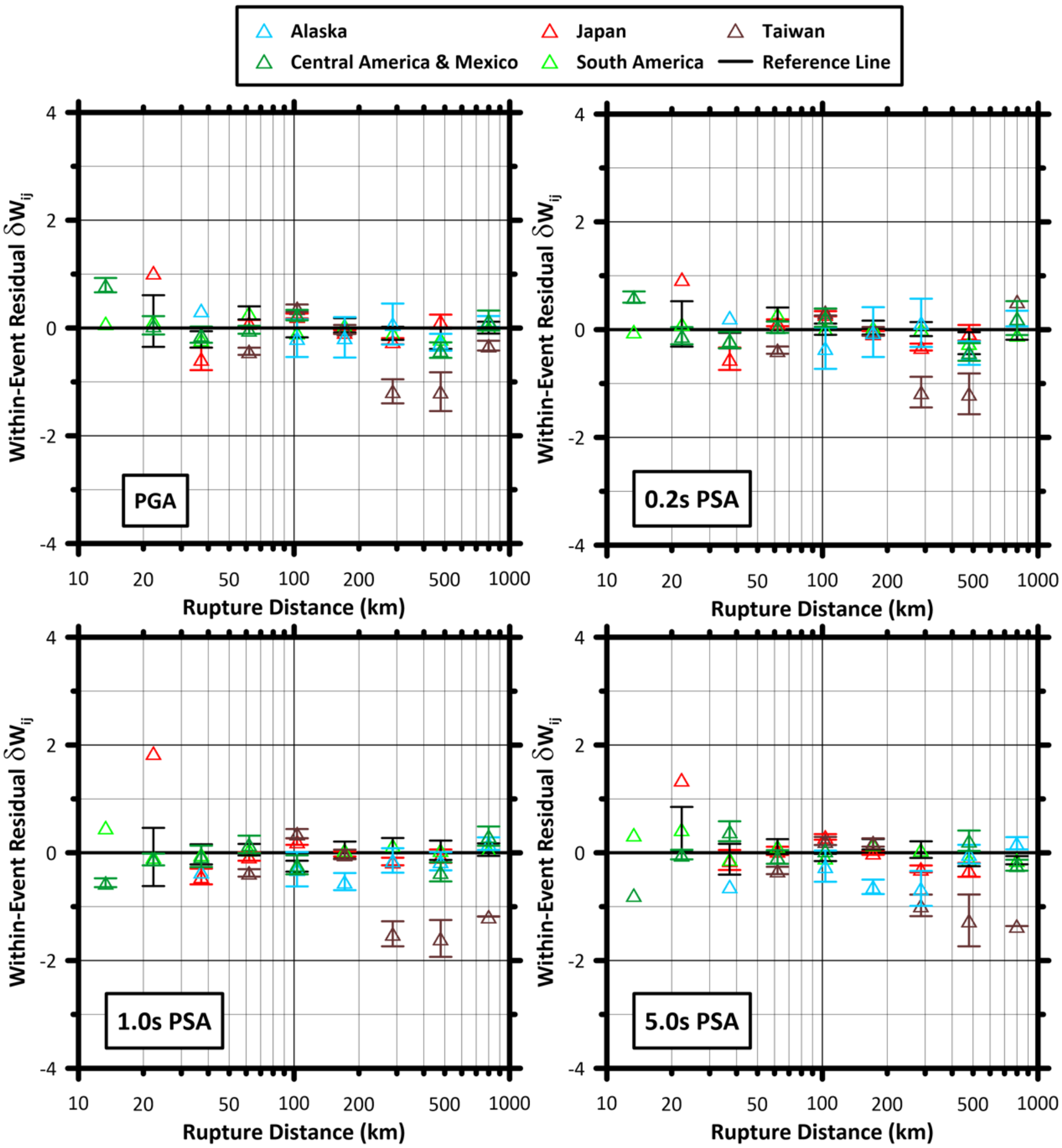

Figure 5.20. Regional binned means and standard errors of within-event residuals, δW_{ij}, from interface events for PGA and 0.2s, 1.0s, and 5.0s PSA. Residuals are color-coded by subduction zone region.

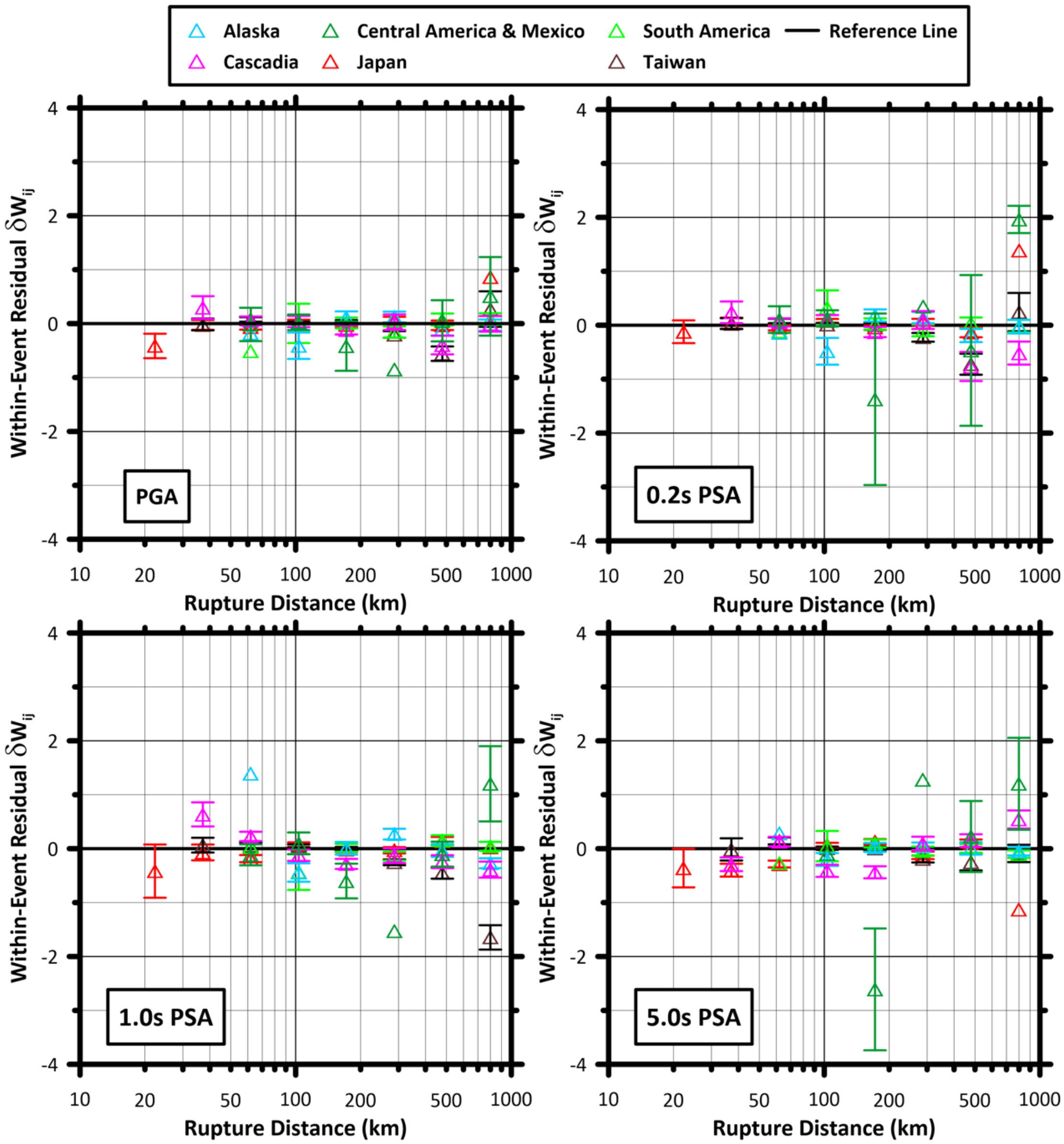

Figure 5.21. Regional binned means and standard errors of within-event residuals, δW_{ij}, from intraslab events for PGA and 0.2s, 1.0s, and 5.0s PSA. Residuals are color-coded by subduction zone region.

5.6.2 Aleatory Variability

The aleatory variability in the model represents the natural variation in earthquake ground motions relative to the median model predictions. For a given set of model input parameters, variations between realized ground motions and the model are possible due to differences in the earthquake source (represented by non-zero event term, η_E), site response (represented by non-zero site term, η_S) and additional variations in within-event residuals related both to path and site response (non-zero remaining residual, ε_{ij}). As shown in the previous figures (5.14-5.19), each of these terms has zero mean and no trend with predictor variables. Each also has an accompanying standard deviation (modified from Al Atik et al. 2010):

- Standard deviation of event terms η_E is denoted τ.
- Standard deviation of site terms η_S is referred to as site-to-site variability and is denoted ϕ_{S2S}.
- Standard deviation of remaining within-event variability ε_{ij} is due to path-to-path variations and randomness in site response for a given site, and is referred to as single-station within-event variability, ϕ_{SS}.

The total within-event variability combines the site-to-site and single-station terms:

$$\phi = \sqrt{\phi_{SS}^2 + \phi_{S2S}^2} \tag{5.12}$$

The total model uncertainty, σ, is given as the square root sum of squares of the between-event variability (τ) and the within-event variability (ϕ):

$$\sigma = \sqrt{\tau^2 + \phi^2} \tag{5.13}$$

Period-dependent values of τ, ϕ, and σ are given in Tables E2-E3 in the electronic supplement. Figures 5.22-5.23 show the global and regional values of τ as a function of period for interface and intraslab events, respectively. Figures 5.24-5.25 show the global and regional values of ϕ as a function of period for interface and intraslab events, respectively.

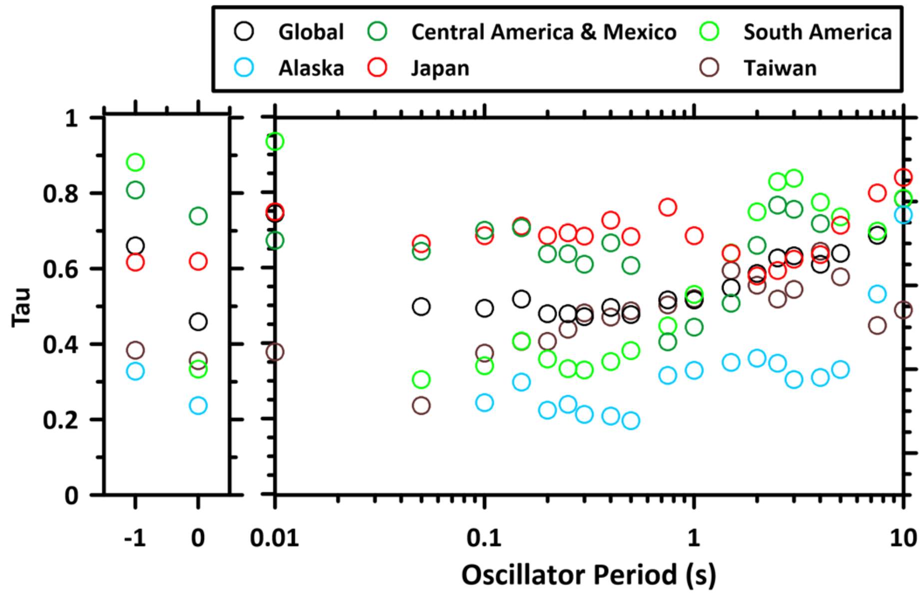

Figure 5.22. Global and regional values of tau as a function of oscillator period for the interface model.

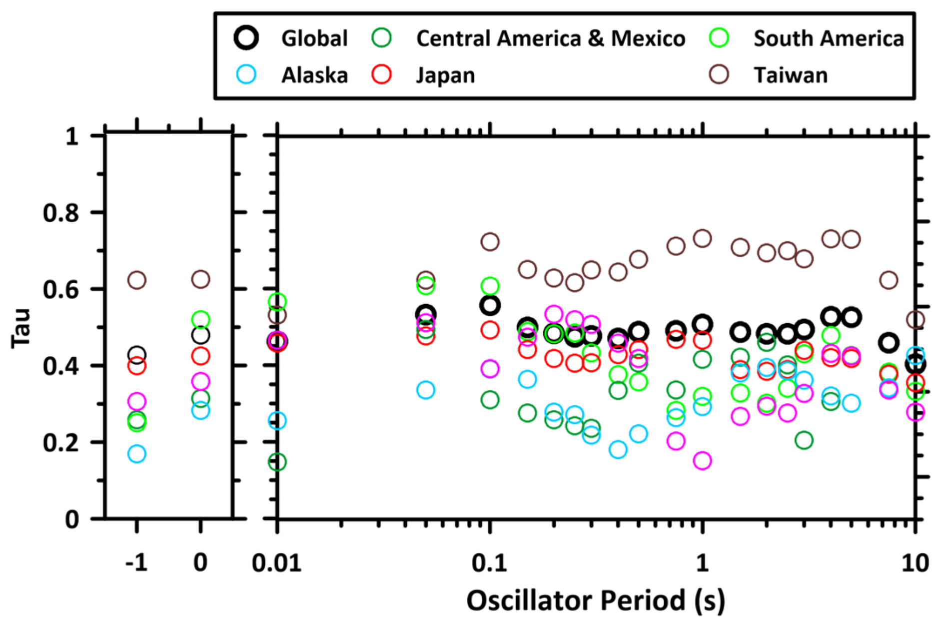

Figure 5.23. Global and regional values of tau as a function of oscillator period for the intraslab model.

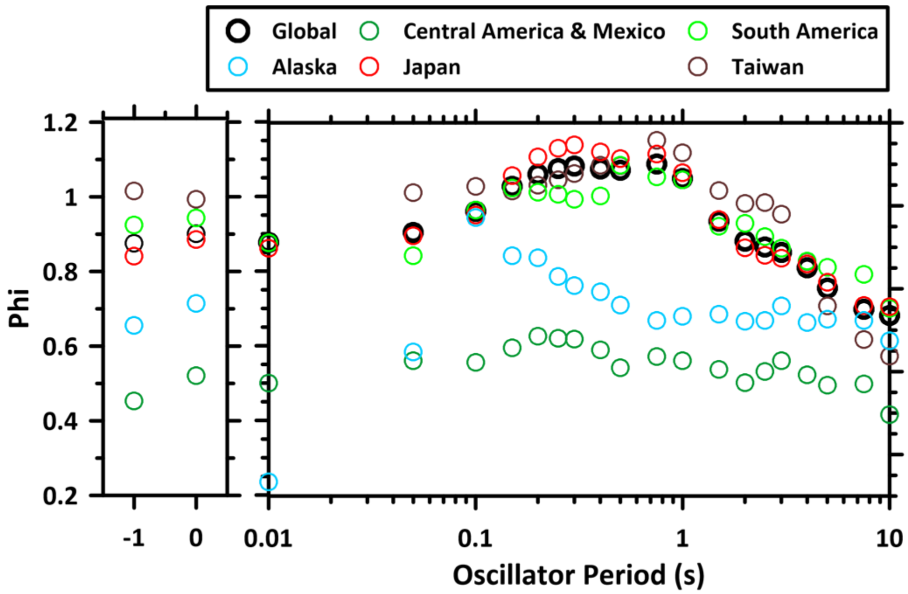

Figure 5.24. Global and regional values of phi as a function of oscillator period for the interface model.

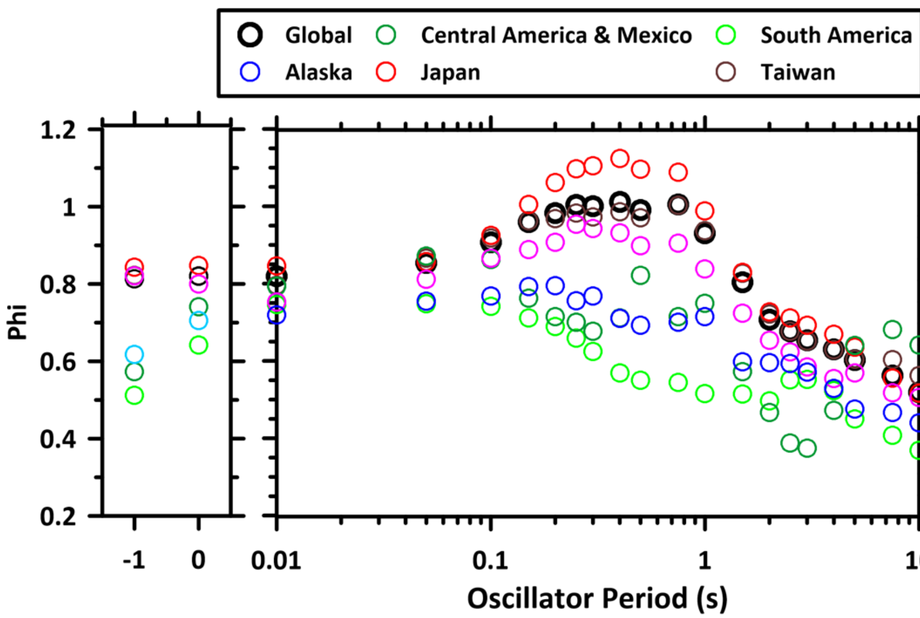

Figure 5.25. Global and regional values of phi as a function of oscillator period for the intraslab model.

5.7 MODEL PERFORMANCE AND VERIFICATION

5.7.1 General Model Performance

The main attributes of the global median model are illustrated in Figures 5.20-5.21, which show the variation of PGA, PGV, and 5.0s PSA as a function of magnitude (for R_{rup} = 100 km, d_{hyp} = 15km, and V_{S30} = 760 m/s) and as a function of distance (various magnitudes, d_{hyp} = 15km, V_{S30} = 760 m/s).

In general, intraslab ground motions are higher than those from interface events at 100km (Figure 5.20). The magnitude-scaling slope at small-to intermediate magnitudes (5-7) is similar, although the intraslab model has minor upward curvature where the interface model does not. The slope at large magnitudes (>$\boldsymbol{m_b}$) for interface is much shallower than for intraslab, in other words ground motion from interface events saturates more at large magnitudes than ground motion from intraslab events.

Figure 5.21 shows that although the near-source saturation model is the same for both event types, generally the intraslab model estimates higher ground motions than the interface model for short distances (40-200km) due to the steeper geometrical spreading coefficient. The interface model has more curvature of ground motion with distance across all periods due to the larger anelastic attenuation. Figure 5.21 shows the magnitude-dependent geometrical spreading, which is larger at short periods; the slope of the interface **M**8 model is at PGA is slightly larger than the slope of the interface M6 model. Overall, attenuation decreases as period increases, with ground motions attenuating slower at 5.0s PSA than at PGA.

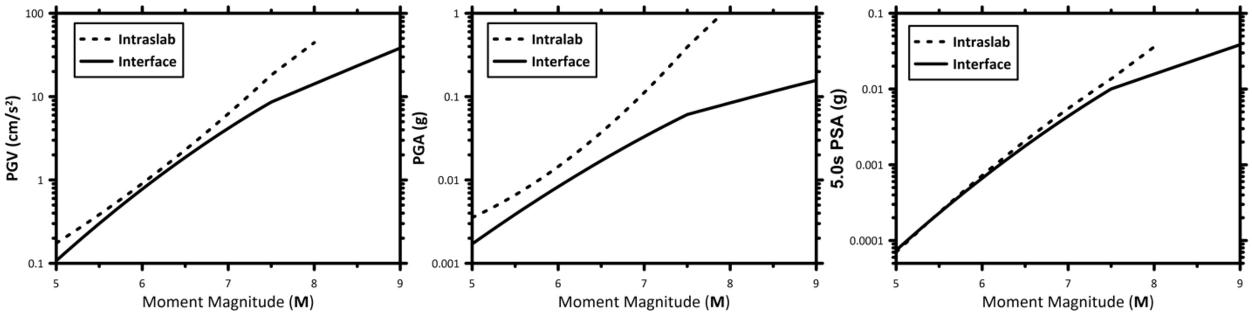

Figure 5.26. Variation of predicted ground motion with moment magnitude (**M**) for interface and intraslab events with hypocentral depth equal to 15km, at a rupture distance of 100km, and a site with the reference rock condition (V_{S30} = 760 m/s) for PGV, PGA and 5.0s PSA.

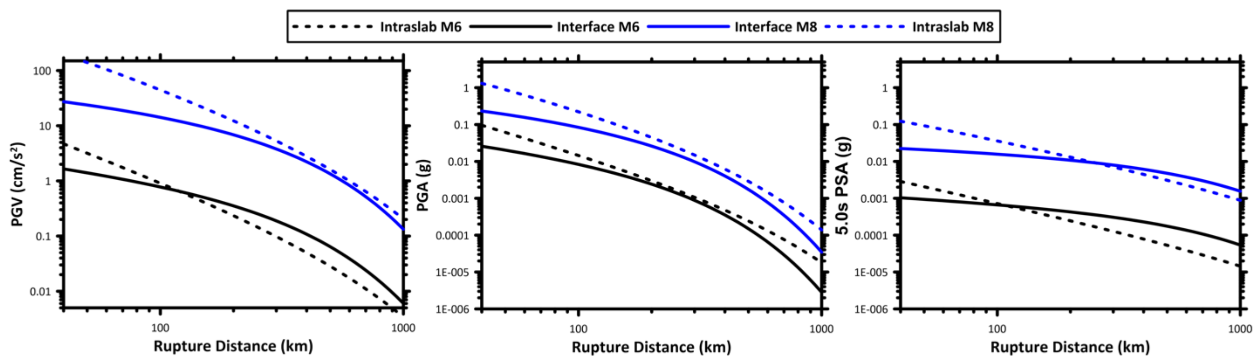

Figure 5.27. Variation of predicted ground motion with moment magnitude (**M**) for interface and intraslab events with hypocentral depth equal to 15km, at a rupture distance of 100km, and a site with the reference rock condition (V_{S30} = 760 m/s) for PGV, PGA and 5.0s PSA.

5.7.2 Application to Cascadia

In Cascadia, we lack data for interface events (Figure 5.4), and therefore recommend the use of the global model with additional uncertainty. We recommend to take epistemic uncertainty about two coefficients with regional values: the constant c_0, and the anelastic attenuation coefficient a_0, where the range of epistemic uncertainty covers the range of regional variation for each coefficient. We recommend using the Cascadia-specific V_{S30}-scaling slope s_1 for both event types. Because it is based on available geometry, we recommend taking $\boldsymbol{m_b}$ equal to the value recommended in Campbell (201x) for Cascadia interface events.

For intraslab events, while recorded ground motion data is available for Cascadia, it is mostly at small magnitudes (Figure 5.4) and exhibits weaker ground motions than the global average. An exception is the 2001 **M**6.8 Nisqually event that has event terms that are nearly zero. The main question in this case is whether to adjust the model to accommodate the average of the event terms, which is negative. Our recommendation, which was formulated in group discussion with other NGA-Subduction modelers, is to not allow this reduction. Rather we suggest to use the global constant c_0 and the regional Cascadia-specific anelastic attenuation constant a_0. We also recommend implementing epistemic uncertainty for Cascadia intraslab events about the constant, in the same way as recommended for interface events.

5.7.3 Model Verification

Model verification consists of comparisons between existing models. Part of the NGA-Subduction Project includes model-to-model comparisons of the four NGA-Subduction models (Parker et al. 201x; Chiou et al. 201x; Abrahamson and Gulerce 201x; Kuehn et al. 2018) with existing models for subduction zones: Atkinson and Boore (2003); Zhao et al. (2006); Gregor et al. (2006); Atkinson and Macias (2009); Zhao et al. (2016a); Zhao et al. (2016b); Abrahamson et al. (2016) and Abrahamson et al. (2018). Comparisons between the distance-scaling term of the global GMM presented in this chapter and those of existing subduction zone GMMs are given in Figures 5.28-5.31 for an **M**9 interface event with V_{S30} = 760 m/s at 1.0s PSA, and Figures 5.32-5.34 for an **M**8 intraslab event with V_{S30} = 760 m/s at 1.0s PSA. Additional comparisons between NGA-Subduction GMMs will be made when all four NGA-Subduction developer teams have complete models.

Figure 5.28 shows a comparison of the global model to the Atkinson and Boore (2003) GMM for NEHRP site classes B and C. The near-source saturation occurs at a much larger distance for Atkinson and Boore (2003), the geometrical spreading is slower, and the anelastic attenuation introduces a similar level of curvature to the path model at large distances (~500km). Figure 5.29 shows a comparison of the global model to the Zhao et al. (2006, 2016a) models for interface events. For both models the near-source saturation distance is similar, as is the geometrical spreading. The anelastic attenuation is less in the Zhao et al. (2006) model, but similar to the global model for Zhao et al. (2016a), perhaps due to similarities in the two datasets (large contributions of data for Japan). However, the overall predicted ground motion amplitude is less for Zhao et al. (2016a) than the global model. Figure 5.30 shows a comparison to the path term of Abrahamson et al. (2016) ("BC Hydro") and Abrahamson et al. (2018) ("Updated BC Hydro"). In both cases, the models are quite similar, except for an overall ground motion amplitude that is slightly less for the Abrahamson models. Lastly, Figure 5.31 shows a comparison to the simulation-based models of Atkinson and Macias (2009), and Gregor et al. (2006). The Gregor et al. (2006) path model does not agree with the distance-scaling implied by the NGA-Subduction dataset. The Atkinson and Macias (2009) model has a similar near-source saturation distance, flatter geometrical spreading, and a similar anelastic attenuation curvature.

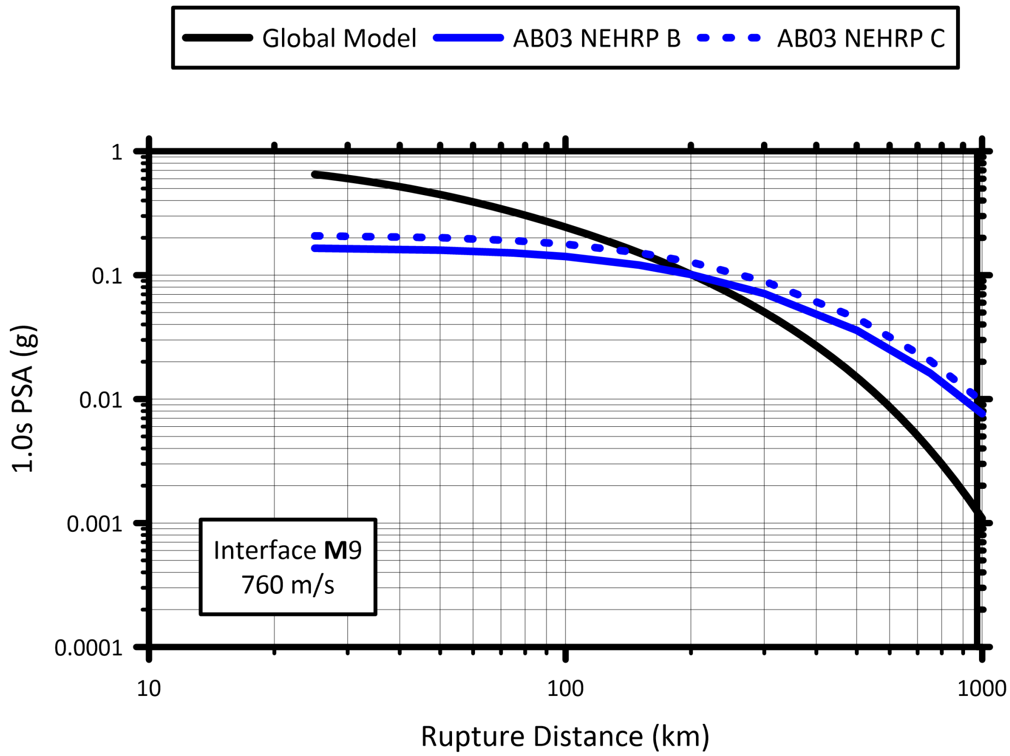

Figure 5.28. A comparison of the global path model for an interface **M**9 event at 1.0s PSA and a 760m/s site condition with the path model from Atkinson and Boore (2003) (AB03) for NEHRP site classes B and C.

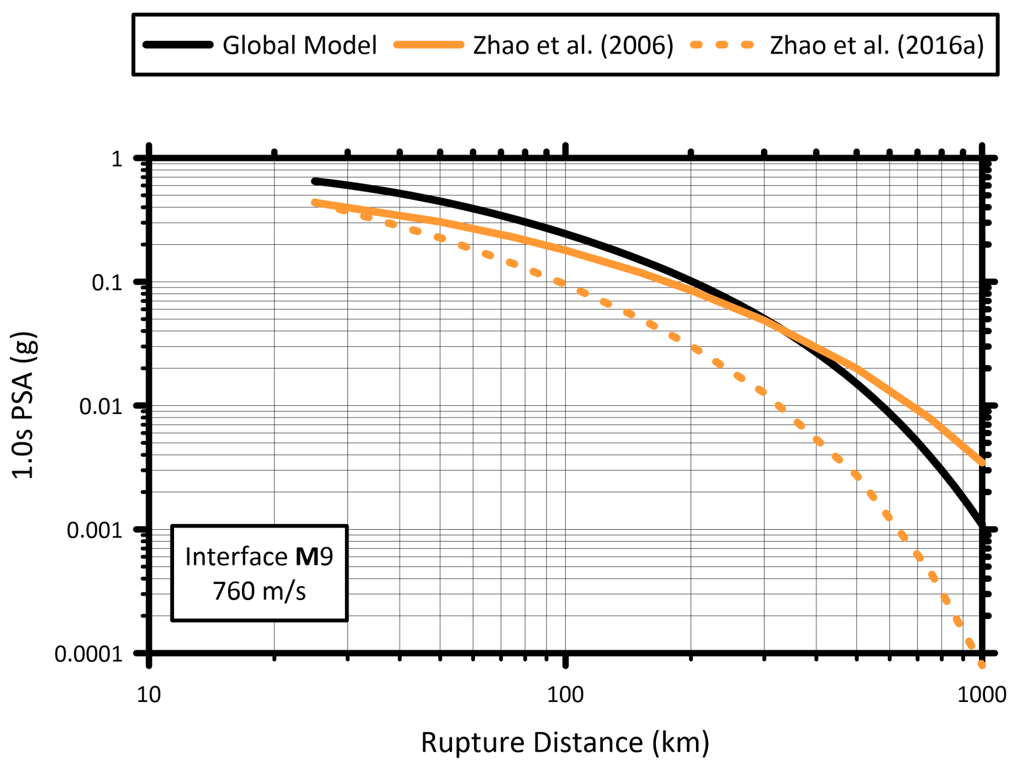

Figure 5.29. A comparison of the global path model for an interface **M**9 event at 1.0s PSA and a 760m/s site condition with the path model from Zhao et al. (2006) and Zhao et al. (2016a).

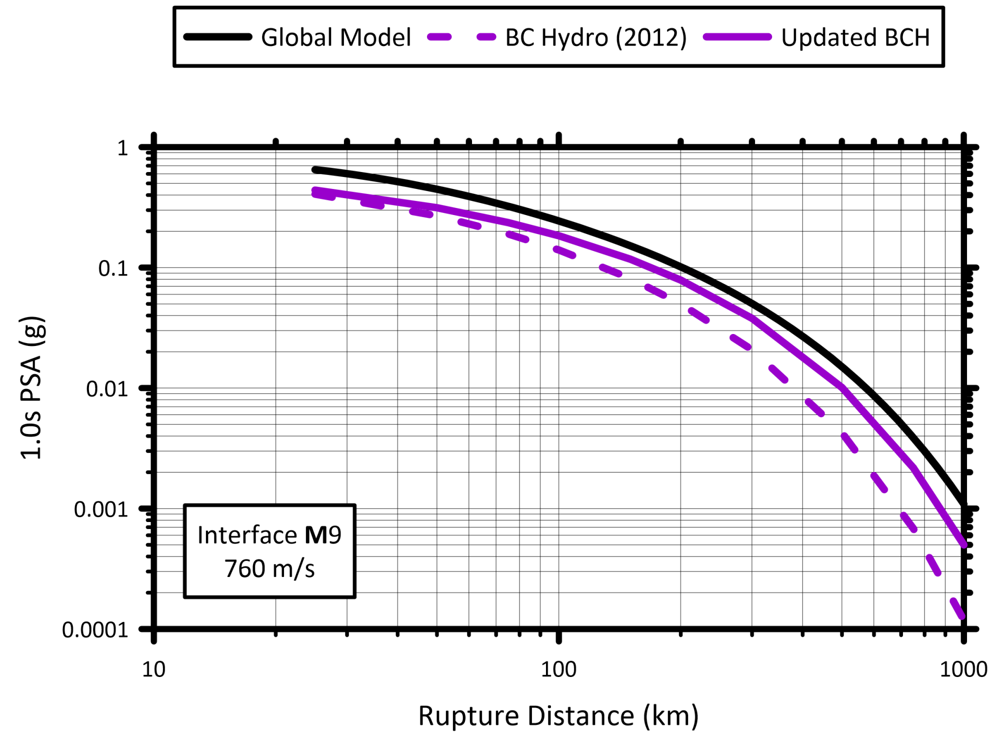

Figure 5.30. A comparison of the global path model for an interface **M**9 event at 1.0s PSA and a 760m/s site condition with the path model from Abrahamson et al. (2016)/BC Hydro (2012) and Abrahamson et al. (2018) (Updated BCH).

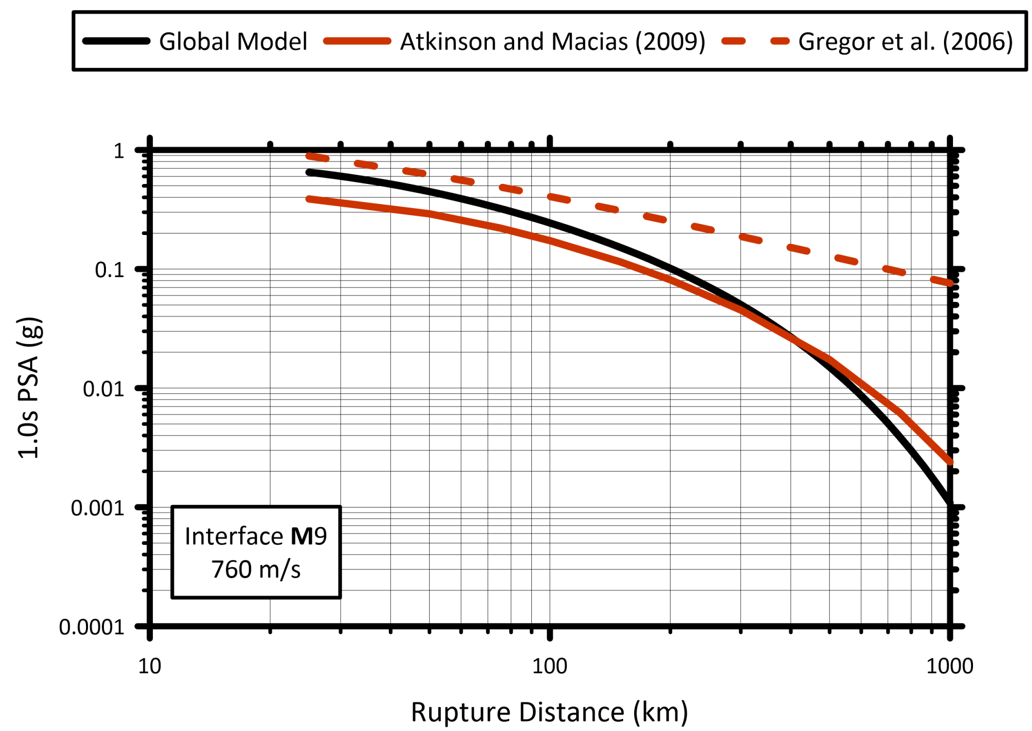

Figure 5.31. A comparison of the global path model for an interface **M**9 event at 1.0s PSA and a 760m/s site condition with the simulation-based path models from Atkinson and Macias (2009) and Gregor et al. (2006).

Figure 5.32 shows the same comparison as Figure 5.22, but for a **M**8 intraslab event. In this case, the Atkinson and Boore (2003) models also have a larger near-source saturation distance than the proposed global model, a similar distance-scaling slope in the intermediate distance range (100-200km), and more curvature due to stronger anelastic attenuation. Figure 5.33 shows a comparison of the global model to the Zhao (2006) and (2016b) models. The Zhao et al. (2016) model is very similar to the proposed global model for this case, with a slightly lower overall ground motion amplitude. The Zhao et al. (2016b) model has significantly slower distance-scaling, and almost no anelastic attenuation (i.e. curvature). Lastly, the proposed global intraslab model is compared to the Abrahamson et al. (2016) and Abrahamson et al. (2018) in Figure 5.34. Both of the Abrahamson models predict lower ground motion amplitudes than the proposed model, with more curvature at long distances (>300km), but comparable slope at intermediate distances (100-300km).

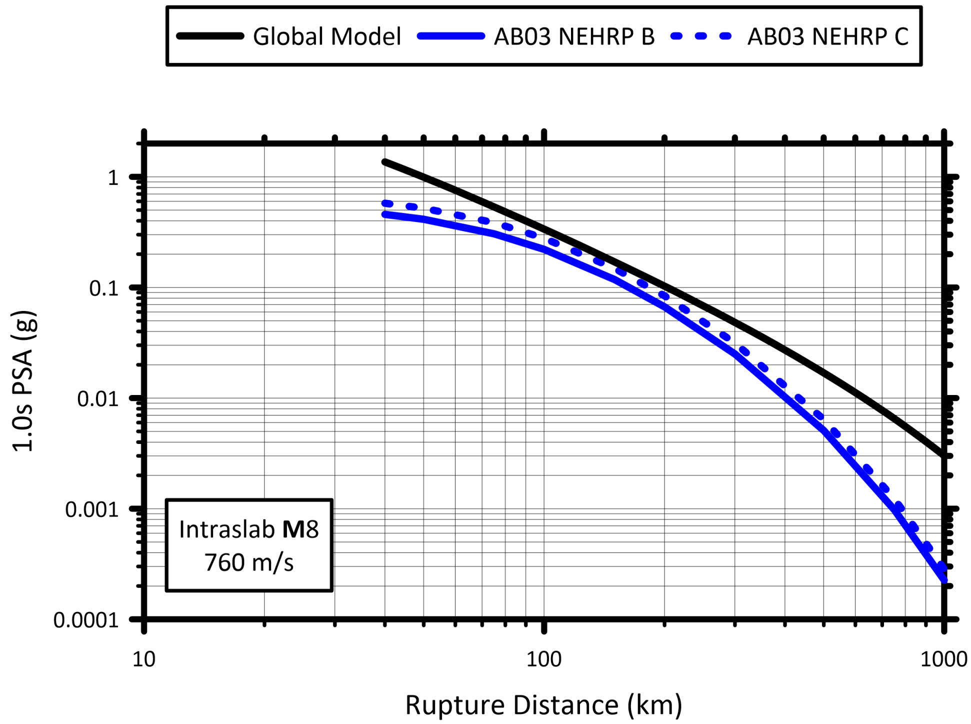

Figure 5.32. A comparison of the global path model for an interface **M**9 event at 1.0s PSA and a 760m/s site condition with the simulation-based path models from Atkinson and Macias (2009) and Gregor et al. (2006).

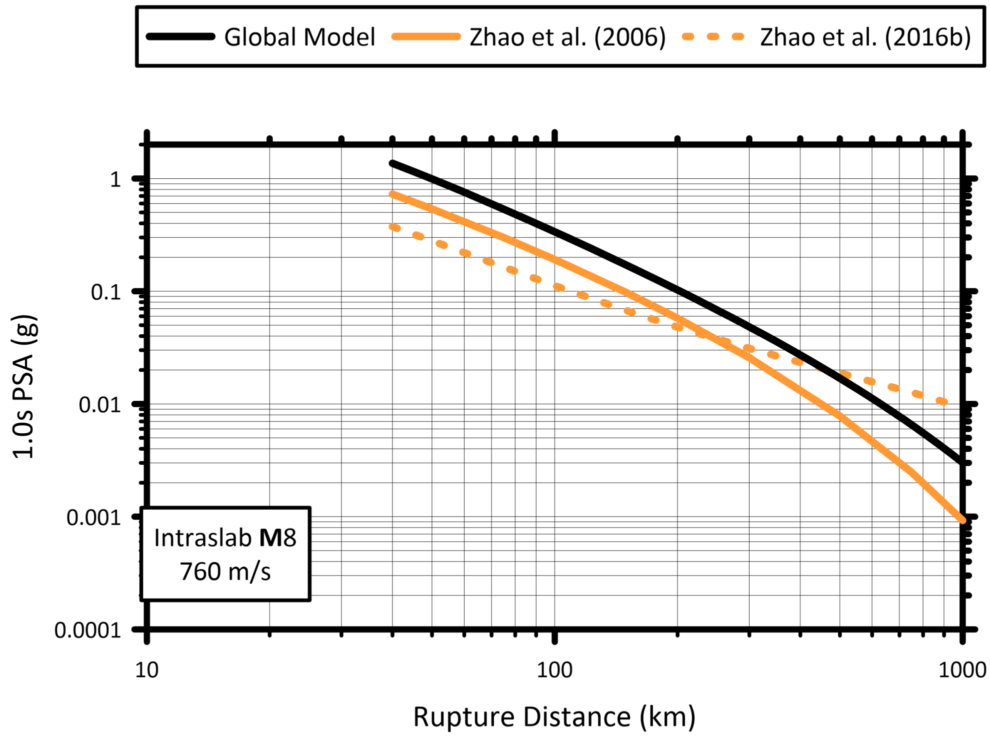

Figure 5.33. A comparison of the global path model for an interface **M**9 event at 1.0s PSA and a 760m/s site condition with the simulation-based path models from Atkinson and Macias (2009) and Gregor et al. (2006).

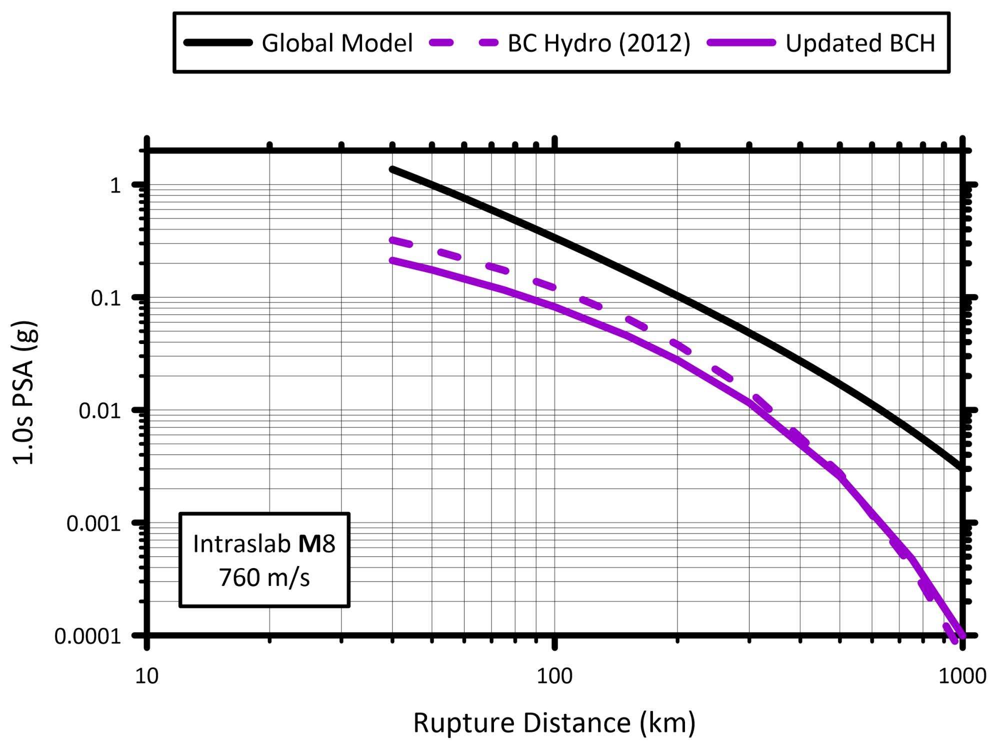

Figure 5.34. A comparison of the global path model for an interface **M**9 event at 1.0s PSA and a 760m/s site condition with the simulation-based path models from Atkinson and Macias (2009) and Gregor et al. (2006).

5.8 MODEL LIMITATIONS

The ground motion model presented in this chapter can be used to predict PGA, PGV and PSA at 19 oscillator periods between 0.01-10.0s for interface and intraslab subduction zone events. The interface model is valid over **M**4.5- **M**9.0, R_{rup} = 20-1000km, d_{hyp} = 0-40km, and V_{S30} = 150-2000m/s. The intraslab model is valid over **M**4.5- **M**8.5, R_{rup} = 35-1000km, d_{hyp} = 0-200km, and V_{S30} = 150-2000m/s. Both models are only applicable to sites in the forearc region of subduction zones. Future work is planned to evaluate model performance in regional back-arc regimes and create an additional anelastic attenuation term if necessary, and to evaluate the performance of the model in regional sedimentary basins (Chapter 7).

Regional modifications to the global models are available for Alaska, Cascadia, Central America and Mexico, Japan, South America, and Taiwan, which consist of regional coefficients for the constant term, anelastic attenuation term, magnitude break point, and site amplification model. For forward applications to regions not considered during model development, we advise that a range of epistemic uncertainty is taken about the global models that represents the regional variation in the constant, anelastic attenuation term, magnitude break-point, and V_{S30}-scaling.

6 NGA-SUBDUCTION GLOBAL AND REGIONAL SITE AMPLIFICATION

6.1 INTRODUCTION

The Next Generation Attenuation – Subduction project is a multi-year, multi-disciplinary project with the goal of producing uniformly processed ground motion data, including time series and spectral data, supporting metadata related to the earthquake events and recordings stations, and a suite of global and regional ground motion models (GMMs) for subduction zone earthquakes. This project considers subduction zones around the world for which ground motion data is available, including: Japan, Taiwan, British Columbia (Canada), Alaska and the Pacific Northwest of the United States, New Zealand, Mexico, Chile, and Peru. Phase 1 of the project, data acquisition and processing, is described in Kishida et al. (2018), Ahdi et al. (2017), Contreras (2017) and PEER (2019). Phase 2 of the project, model development, involves four teams developing independent subduction zone GMMs. The ground motion model described in Chapter 5 is one of these four, developed using a combination of empirical data analysis, finite-fault simulations, and geometrical constraints (e.g. Archuleta and Ji 2018, Campbell 201x) to predict PGA, PGV and PSA at oscillator periods between 0.01-10s. However, the model as given in Chapter 5 applies only to the reference rock condition of V_{S30} = 760 m/s (i.e. the NEHRP B/C boundary condition; Frankel et al. 1996). In order to use the model for other site conditions such as soil or weathered rock, an additional site amplification model is necessary, which is the subject of this chapter.

Site amplification models developed empirically for data-rich areas typically have three components:

(i) A linear site amplification term that expresses the effect of the shallow site condition on the ground motion intensity measure. Typically this is a V_{S30}-scaling term, although in some cases fundamental frequency of a site (f_0) is used;

(ii) A nonlinear term that decreases the amplitude of the ground motion intensity measure as the strength of shaking increases; and

(iii) Secondary terms beyond V_{S30}, which approximately account for 3-dimensional wave propagation effects due to basin geometry. These can include wave focusing (Baher and Davis, 2003; Stephenson et al., 2000) or body to surface wave conversion (Graves, 1993; Graves et al. 1998; Kawase, 1996; Pitarka et al., 1998).

When sufficient data is not available it is common for developers to use simulations to constrain all (e.g. Harmon et al. 2019b) or some components (e.g. Seyhan and Stewart 2014) of the site amplification model.

In the development of subduction zone GMMs, it is typical to borrow some or all of the above site amplification model components from active crustal regions or simulations when data is lacking. For example, Abrahamson et al. (2016) develop a global V_{S30}-scaling term, but use the simulation-based nonlinear site amplification term of Walling et al. (2008), constrained to be consistent with the Peninsular Range model from California, and do not consider basin depth or regionalization. Zhao et al. (2016a,b) use site class based on site period as their site term, binning sites into categories instead of having a continuous predictor variable, and adopt the 1-D simulation-based nonlinear model for site classes of Zhao et al. (2015). Moreover, in the Pacific northwest region of the US, site factors used in current building code applications are derived using data from active tectonic regions (i.e., the model of Seyhan and Stewart, 2014).

This chapter presents an empirical subduction-specific site amplification model to be paired with the reference-rock conditioned GMM of Chapter 5. Due to the large dataset compiled in the NGA-Subduction project (PEER 2019), a global model for V_{S30}-scaling is developed along with regional adjustments. The applicability of the nonlinear term in SS14 to the global subduction data is also investigated, and some modifications are proposed. Future work is planned to investigate basin effects in Taiwan, Japan and the Cascadia region.

6.2 GLOBAL AND REGIONAL MODELS

6.2.1 Global Site Amplification Model

In the absence of basin effects, the total site amplification model is given as the sum of two terms in natural logarithmic units:

$$F_S = F_{lin} + F_{nl} \tag{6.1}$$

where F_{lin} is the linear site amplification, or the V_{S30}-scaling term, and F_{nl} is the nonlinear site amplification term. The functional form for the linear term is given as:

$$F_{lin} = \begin{cases} s_1 ln\left(\frac{V_{S30}}{V_1}\right) + s_2 ln\left(\frac{V_1}{V_{ref}}\right) & for\ V_{S30} \le V_1 \\ s_2 ln\left(\frac{V_{S30}}{V_{ref}}\right) & for\ V_1 < V_{S30} \le V_2 \\ s_2 ln\left(\frac{V_2}{V_{ref}}\right) & for\ V_{S30} > V_2 \end{cases} \tag{6.2}$$

F_{lin} is tri-linear in V_{S30} space, but only data from Taiwan shows a break in slope at V_1, similar to that observed in previously in Japan (Campbell and Bozorgnia 2014) and CENA (Parker et al. 2019; Hassani and Atkinson 2017). In other words, for most regions $s_1=s_2$.

The nonlinear term has the same form as the NGA-West2 Seyhan and Stewart (2014) model:

$$F_{nl} = f_1 + f_2 ln\left(\frac{PGA_r+f_3}{f_3}\right) \tag{6.3}$$

where f_1 to f_3 are model parameters and PGA_r represents the peak acceleration expected for the reference site condition of 760 m/s). f_1 represents the level of amplification that is independent of PGA_r, which is accommodated by F_{lin}. As a result, f_1 is not needed in F_{nl} and is taken as zero in Eq. (6.3). f_3 represents a transition level of PGA_r, whereby for $PGA_r << f_3$, F_{nl} goes to zero and for $PGA_r >> f_3$, F_{nl} approaches a constant slope of f_2 with respect to the log of PGA_r/f_3. For modeling purposes, f_2 is related to V_{S30} as (modified from Chiou and Youngs 2008):

$$f_2 = f_4[exp\{f_5(min(V_{S30}, 900) - 200)\} - exp\{f_5(900 - 200)\}] \tag{6.4}$$

Coefficients for the global and regional site amplification models are independent of event-type, and given in Tables E2-E3 in the electronic supplement.

6.2.2 Regional V_{S30}-Scaling

Regional slopes s_1 and s_2 are given for Alaska, Cascadia, Japan, South America, and Taiwan in Tables E2-3 of the electronic supplement. Additionally, some regions require modification of V_2, the corner velocity at which the model goes flat at high V_{S30}. These regional V_2 values are also given in Tables E2-3 of the electronic supplement. Due to sparsity of data, the global site amplification model is recommended for Central America and Mexico. There is no regional variation in the nonlinear model, F_{nl}, due to lack of data to constrain coefficients for each independent region.

6.3 MODEL DEVELOPMENT AND RESULTS

6.3.1 Linear Site Amplification

The first step in computing the linear site amplification implied by the NGA-Subduction database is computing within-event rock residuals using the reference-rock conditioned GMM, μ_{ij}^{r}, given by Eq. 5.1, the event terms $\eta_{E,i}$ given by Eq. 5.9, and a nonlinear model, F_{nl}:

$$\delta W_{ij}^{r} = ln(Y_{ij}) - [\mu_{ij}^{r} + F_{nl,j} + \eta_{E,i}] \tag{6.5}$$

Subscripts i and j refer to event and station, respectively. Superscript r indicates the term is for the reference rock velocity condition of 760m/s. For this step of the model development process, data (Y_{ij}) from both interface and intraslab events are combined, as we do not expect differences in the source to affect amplification due to site properties. This expectation is tested subsequently using residuals analyses.

Within-event rock residuals δW^{r} are not expected to average to zero because they represent the difference between data for soil site conditions and model predictions for a reference rock condition. As such, when taken in aggregate, these residuals provide an estimate of site response per the non-reference site approach (Field and Jacob, 1995). Ideally, the differences between $ln(Y_{ij})$ and the quantity in brackets would be due to site response only, although in reality other factors contribute to non-zero realizations of δW^{r}. The event term is included in the sum within the brackets to remove bias in total residuals that is related to source, and hence unrelated to site. There can be biases associated with particular source-to-site paths, which are not accounted for in Eq. (6.5). An essential element of the non-reference site approach is that the path model should be unbiased in a broad sense, even if it may be biased for a particular realization. If this is the case, then many samples of path errors (over many observations) would average to zero, which in turn would leave site as the remaining source of non-zero mean δW^r values. The F_{nl} term is included within the brackets in Eq. (6.5) to remove nonlinear site effects because the initial focus is on the linear site response. This correction is small for most data points, only being appreciable for relatively near-fault (strong shaking) conditions, soft soils, and high-frequency IMs.

The within-event rock residuals are partitioned into reference-rock site terms (η_S^r), which represent the average site amplification observed over many events for each recording station, and the remaining residual (ε_{ij}),

$$\delta W_{ij}^r = \eta_{S,j}^r + \varepsilon_{ij} \tag{6.6}$$

The partitioning is done using mixed-effects analysis in R (Bates et al. 2015). The ε_{ij} term represents variation in ground motion due to event-to-event variations in site response and path errors.

The $\eta_{S,j}^r$ terms are examined for trends with V_{S30}, in order to develop the V_{S30}-scaling model (F_{lin}). The model development is iterative because of the use of a nonlinear model (F_{nl}) in Eq. (6.5). The first iteration uses an available F_{nl} term in the literature from SS14 (lines 2-3 of Eq. 6.2). Subsequent iterations used a modified F_{nl} term derived in the next section. The results shown here reflect the final outcome once the F_{nl} term was set.

Figure 6.1 shows the variation of η_S^r with V_{S30} using results from all regions together, along with the model from Eq. (6.2). The model fit was performed using nonlinear least-squares regression in R (R Core Team 2016). Each gray symbol in the figure represents a reference-rock site term for a single site. The scatter of these terms is appreciable. Data trends can be more readily appreciated by examining the variation with V_{S30} of binned means, which are shown along with their standard errors. The results indicate a steady increase in site amplification as V_{S30} decreases, and a flattening of the relationship for stiff sites, which is captured in the model by a flat trend for $V_{S30} > V_2$. As found previously in active regions, the strength of the trend is consistently found for all IMs considered, but is strongest for periods of 0.5 to 5.0s (along with PGV), and weaker at shorter and longer periods. The wavelengths associated with these intermediate periods are much longer than 30 m, so the strength of this trend is a result of correlation between V_{S30} and the average velocity structure at greater depths.

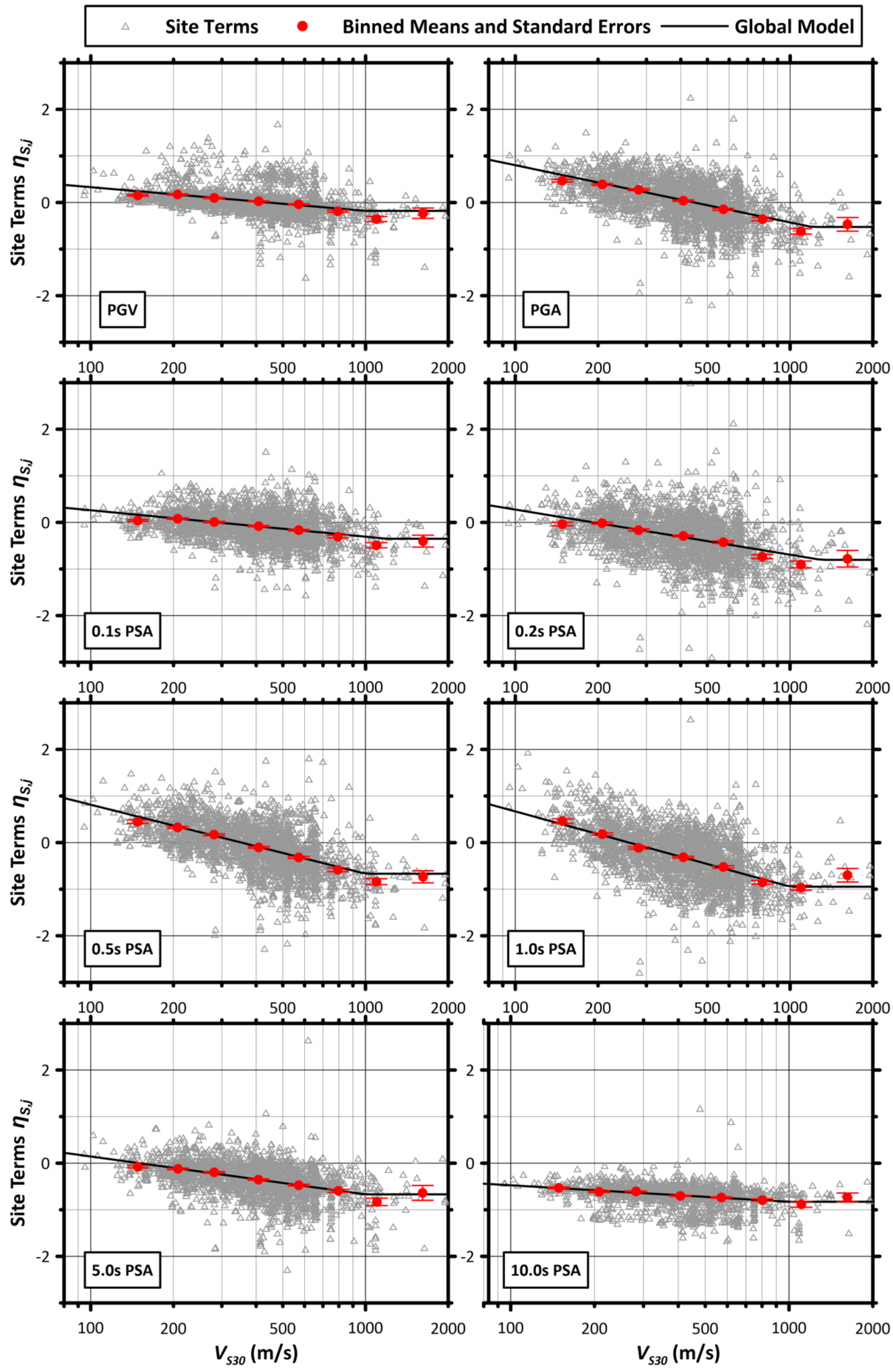

Figure 6.1. Global V_{S30}-scaling model, F_{lin}, for peak ground velocity (PGV), peak ground acceleration (PGA), and a range of PSA oscillator periods 0.1-10.0s.

Figure 6.2 compares slope parameter s_2 as derived for the global model in this study to comparable parameters in the SS14 model for active tectonic regions and the Parker et al. (2019) model for stable continental regions. The V_{S30}-scaling in the global subduction model is not as strong as in active regions, but is stronger than that in CENA for oscillator periods of 0.15-4.0s PSA. The shape of the c parameter in the global subduction model is very similar to that for ATRs, being more negative (stronger scaling) for intermediate periods (0.5-5.0 s) than shorter and longer periods. In contrast, the scaling for CENA is relatively constant with respect to period.

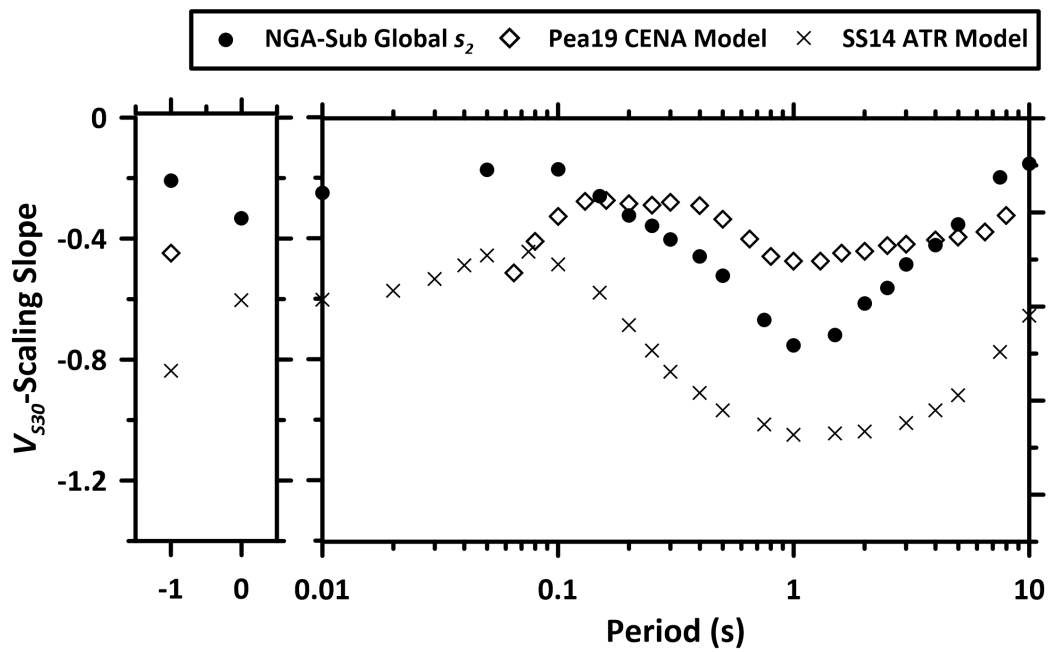

Figure 6.2. Comparison of V_{S30}-scaling slope between the global NGA-Subduction model, the Seyhan and Stewart (2014) (SS14) for active tectonic regions, and the Parker et al. (2019) (Pea19) model for central and eastern North America.

Once the global model was set, additional plots as in Figure 6.3 and 6.4 were prepared for each region individually. In Figure 6.3 the 0.2s PSA data for each region is compared to the global model (Eq. 6.2) and to a regional model reflecting regional coefficients. Figure 6.4 shows the same information for 1.0s PSA. The global value of V_2 is initially used for each period as the sparsity of data when split by region causes V_2 to be under-determined. Once the regional slopes were fit using nonlinear least-squares, the adequacy of the global V_2 was assessed, and regional adjustments were made by judgement as necessary. Lastly, V_1 and s_1 were fit to the Taiwan dataset, to allow a break in slope at slow V_{S30}.

Figure 6.5 shows the variation with period of the regional adjustment to the slope s_2. The regional variations from the global model are modest, but are large enough to be statistically significant. Alaska has the largest negative deviation in V_{S30}-scaling slope from the global model, exhibiting a stronger dependence on V_{S30} than other regions. Cascadia exhibits significantly weaker, and South America exhibits slightly weaker V_{S30}-scaling than the global model. Japan has similar V_{S30}-scaling as the global model; this is not surprising, as the majority of the data used to develop the global model comes from Japan. Due to lack of data, we recommend that the global model be used in Central America and Mexico.

The plots in Figures 6.1, 6.3, and 6.4 pass through zero at V_{S30} values lower that the desired reference condition of 760 m/s. This indicates that adjustment of the constant term is needed to shift the residuals up by a unit amount. This is applied during model development as described further in Chapter 5.4.5.

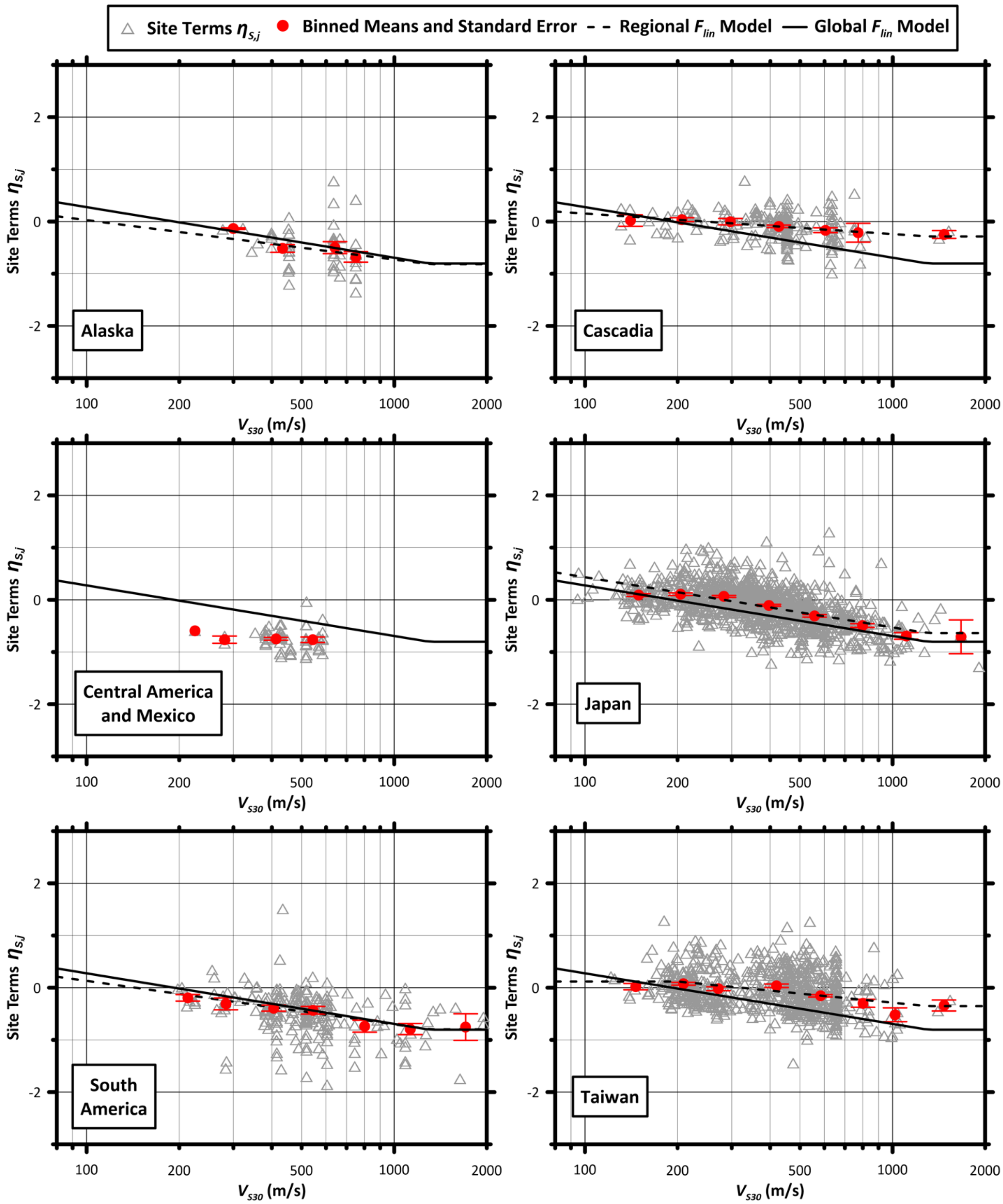

Figure 6.3. Regional V_{S30}-scaling model, F_{lin}, for 0.2s PSA shown in dashed line, compared to global model shown in solid line.

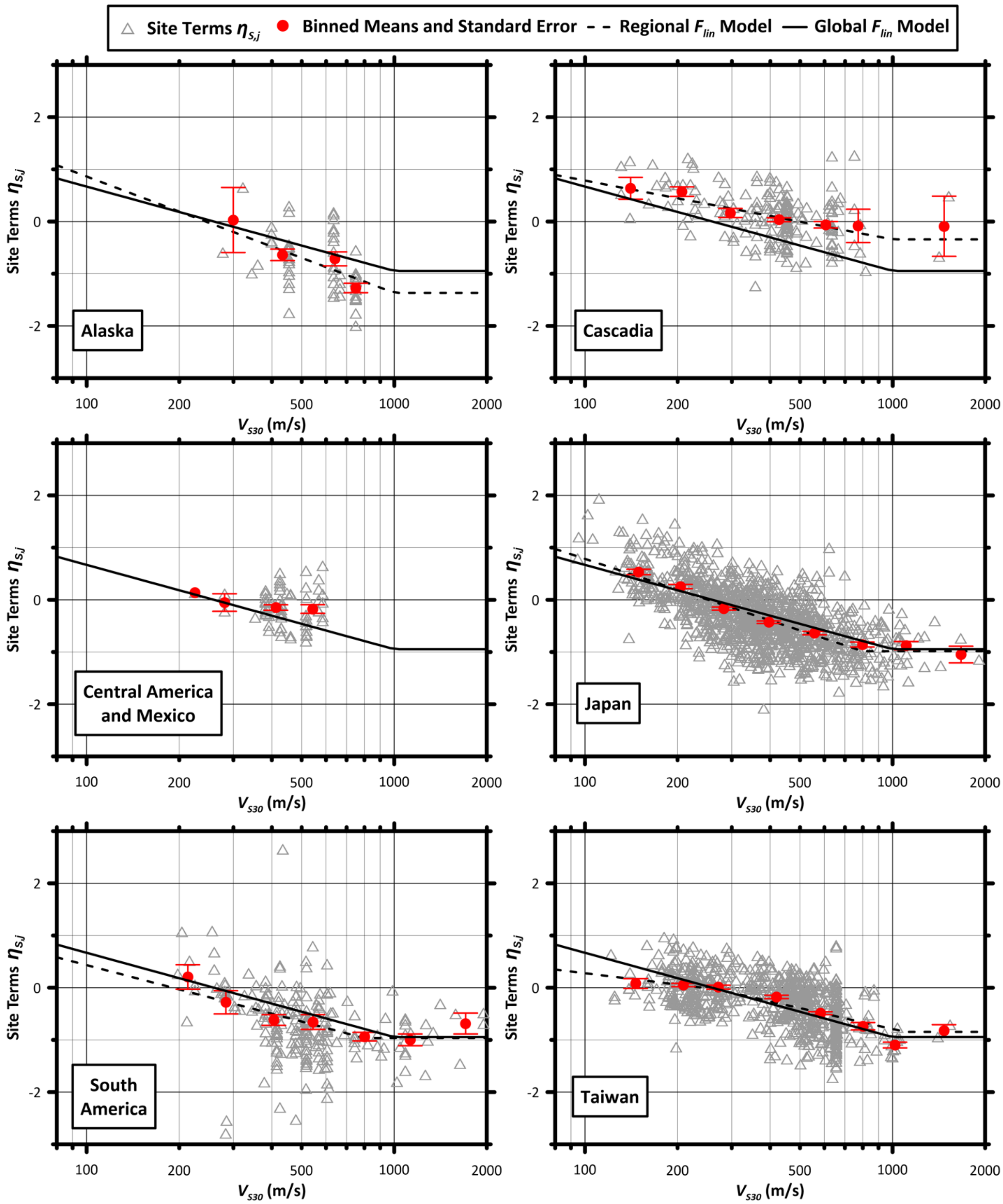

Figure 6.4. Regional V_{S30}-scaling model, F_{lin}, for 1.0s PSA shown in dashed line, compared to global model shown in solid line.

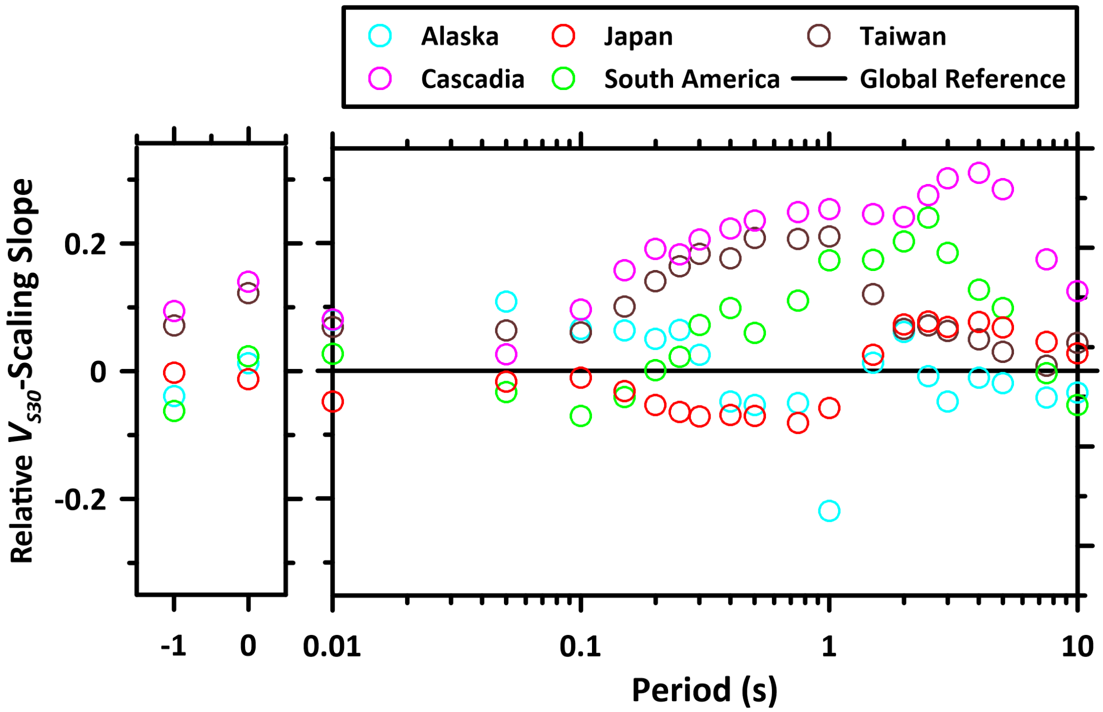

Figure 6.5. Regional variation in V_{S30}-scaling slope s_1 (Eq. 6.2), shown as a modifier relative to the global slope.

6.3.2 Nonlinear Site Amplification

To investigate the nonlinear component of site amplification implied by the NGA-Subduction data, the residuals between the data and a GMM with a linear site term are computed:

$$\delta W_{ij}^{lin} = \ln(Y_{ij}) - [\mu_{ij}^{r} + F_{S,j} + \eta_{E,i}] \tag{6.7}$$

The GMM is exercised for the reference rock condition (Eq. 5.1) and event terms are as given by Eq. (5.9). The linear site term is as given in Eq. (6.2). The mean of within-event residuals given by Eq. (6.7) should be zero if site response is linear. As a result, I look for conditions where the mean trend departs from zero to identify conditions giving rise to nonlinear site response.

The computed within-event residuals for linear site response (δW_{ij}^{lin}) are plotted in Figure 6.6 against the expected median PGA for the reference rock condition (760m/s). This median PGA is computed from the GMM reference rock mean and PGA event term,

$$PGA_{r,ij} = exp\left(\mu_{r,ij}^{PGA} + \eta_{E,i}^{PGA}\right) \tag{6.8}$$

where the PGA superscript indicates that the mean model and event term are taken for the IM of PGA. The PGA_r in Eq. (6.8) represents the expected shaking intensity that would have occurred at the site had the site condition been the reference condition. PGA_r affects the extent to which nonlinear soil behavior is expected (Eq. 6.3). As in SS14, these plots were made for V_{S30} bins ≤ 200, 200-310, 310-520, 520-760, and ≥ 760m/s.

As shown in Figure 6.6, the NGA-Subduction data shows a nonlinear trend with PGA_r that is most clearly evident for V_{S30} bins ≤ 200 and 200-310 m/s. The trend is demonstrated by a downward trend in δW_{ij}^{lin}, and its binned means, with respect to PGA_r. For each of the V_{S30} bins considered in Figure 6.6, the data trend is fit using Eq. (6.3), from which discrete values of f_2 are obtained for each period. The value of f_3 is period-independent and constrained to 0.05g based on visual inspection.

The f_2 results can fit as a function of V_{S30} in order to evaluate the applicable coefficients for the model in Eq. (6.4). Coefficients f_4 and f_5 are fit through this process, as are the velocities that appear in Eq. (6.4), which have been modified relative to those given in Chiou and Youngs (2008). The fit of the selected model to the f_2 values from individual bins is shown in Figure 6.7, which also shows f_2 values from NGA-West2 data and simulations, and the SS14 F_{nl} model. Overall, the nonlinear site amplification inferred from the NGA-Subduction data agrees with what was found in NGA-W2 for V_{S30} bins > 310m/s, but shows less nonlinearity for the 100-200 m/s and 200-310 m/s bins. There is no significant nonlinearity observed in the data for V_{S30} ≥ 760 m/s.

In SS14, f_3, the transition intensity, was constrained to 0.1g across all periods and V_{S30} bins. The nonlinear site amplification implied from the NGA-Subduction dataset suggests a smaller value of f_3, i.e. that nonlinearity starts to occur at a smaller shaking intensity. A value of 0.05g was chosen for all periods and V_{S30} bins based on judgement.

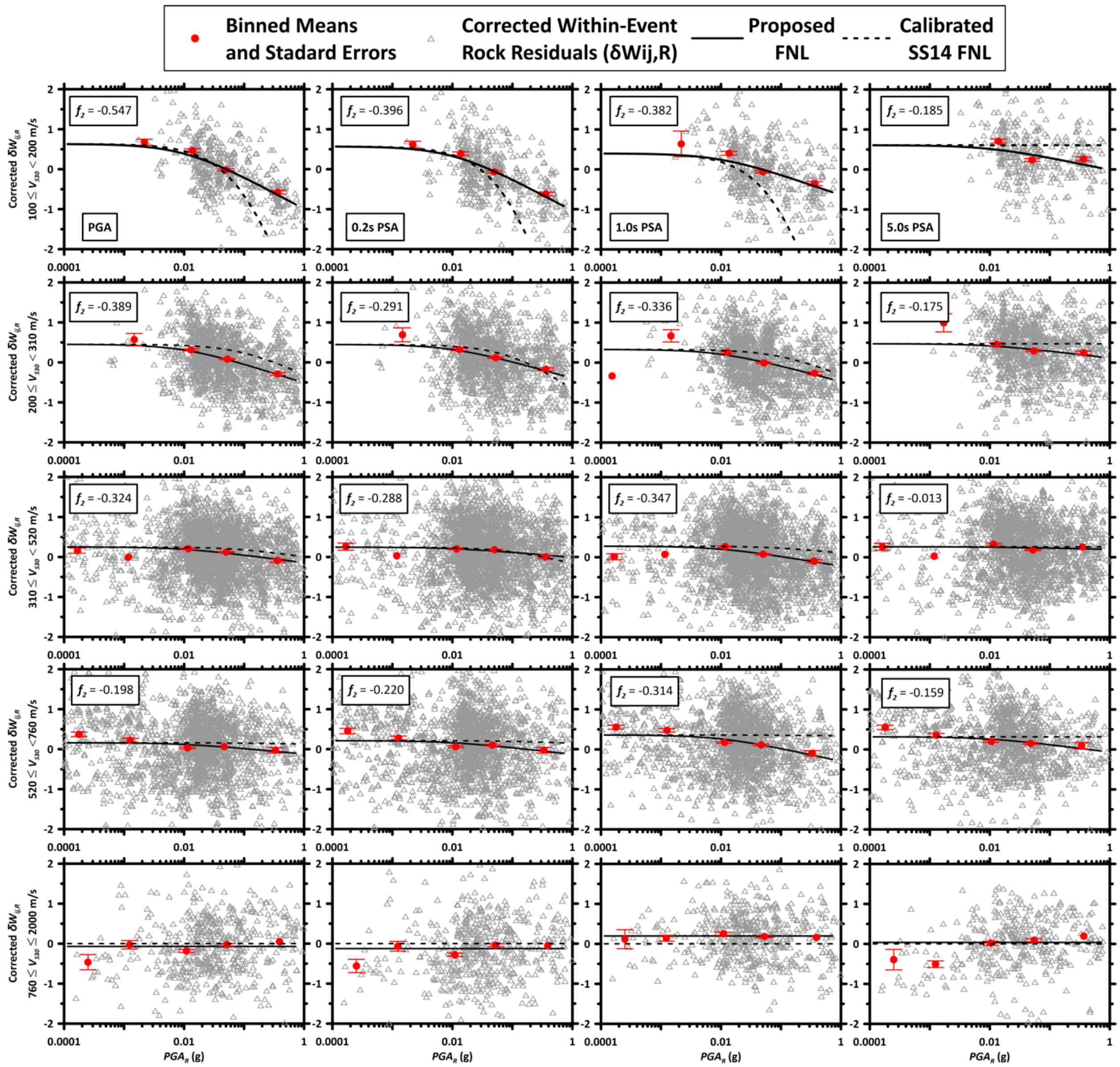

Figure 6.6. Nonlinear site model F_{nl} for PGA, 0.2s, 1.0s and 5.0s shown as a function of PGA_r for V_{S30} bins. The corresponding model from SS14 is shown for comparison.

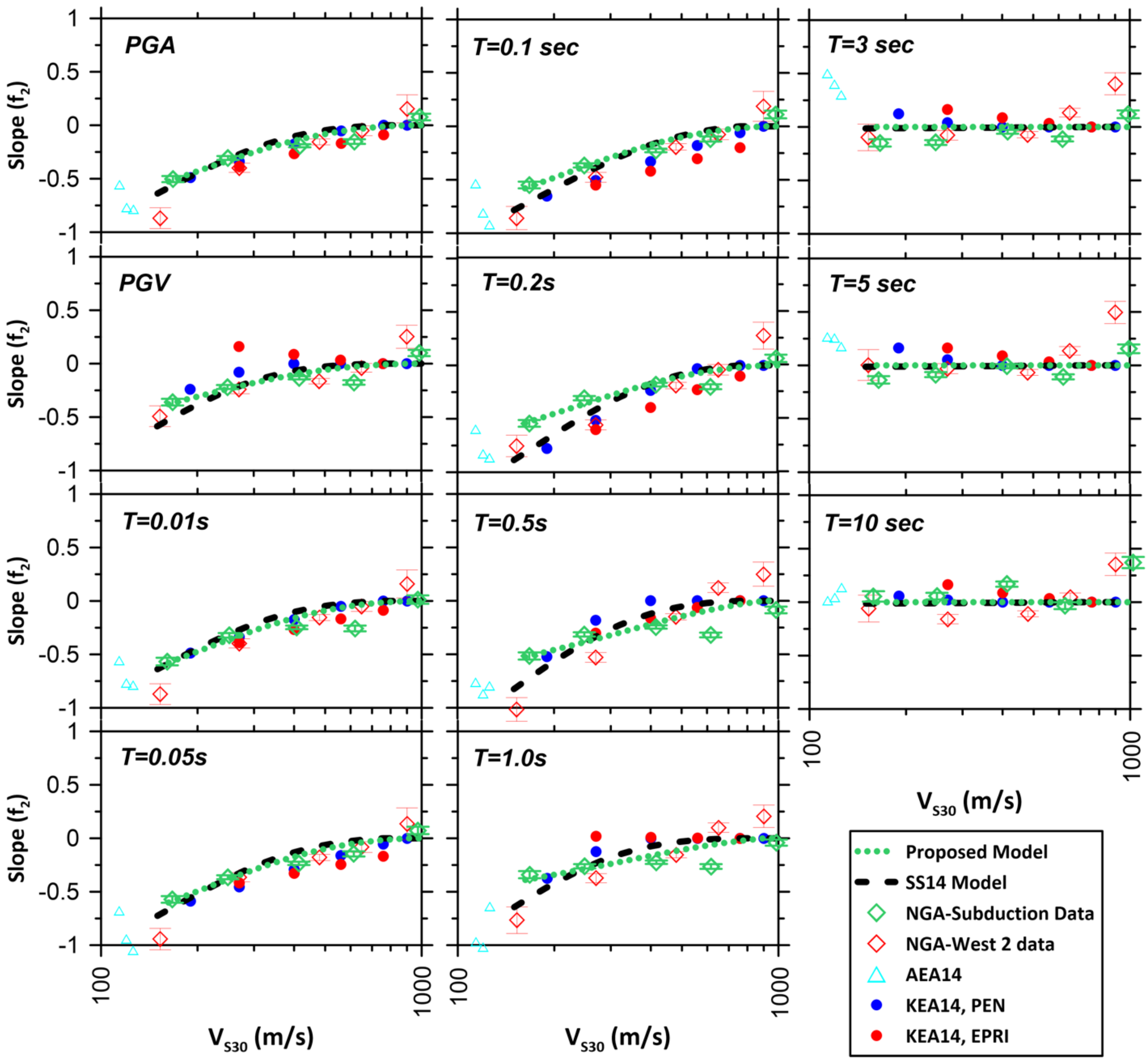

Figure 6.7. Values of parameter f_2 estimated using the NGA-Subduction dataset shown with the proposed form of Eq. 6.4, along with the model from SS14, empirical values of f_2 from NGA-West2, and simulation-based values of f_2 from NGA-West2.

6.4 MODEL PERFORMANCE

Figures 6.8-6.9 shows predictions of response spectra that are obtained by combining the reference rock GMM from Chapter 5 with the site amplification model described here. The global model parameters are used in the site amplification model. Figure 6.8 applies for a condition of an interface event of **M**8 and R_{rup} = 30 km, which produces strong shaking conditions. Figure 6.9 is similar, but now the event is an interface **M**7 earthquake at R_{rup} = 200 km, which produces much weaker shaking. In both cases, median spectra are shown for site conditions of V_{S30} = 200, 400, 700, 1000 m/s.

The relatively weak shaking condition (Figure 6.9) shows steady increases in spectral ordinates as site conditions become softer, with the strongest changes in the period range of 0.5-5.0 sec. The strong shaking condition (Figure 6.8) shows similarly steady increases in long period spectral ordinates, but a more complex pattern at short periods that is affected by differing amounts of nonlinearity.

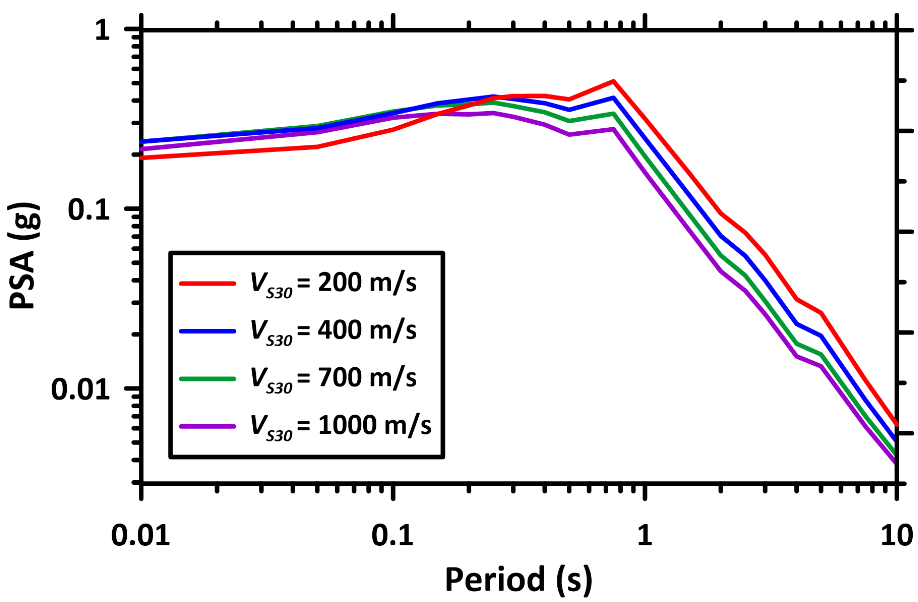

Figure 6.8. Predictions of response spectra computed using the global reference rock GMM from Chapter 5 with the global site amplification model described herein for an interface **M**8 event at R_{rup} = 30 km, for V_{S30} = 200, 400, 700, 1000m/s.

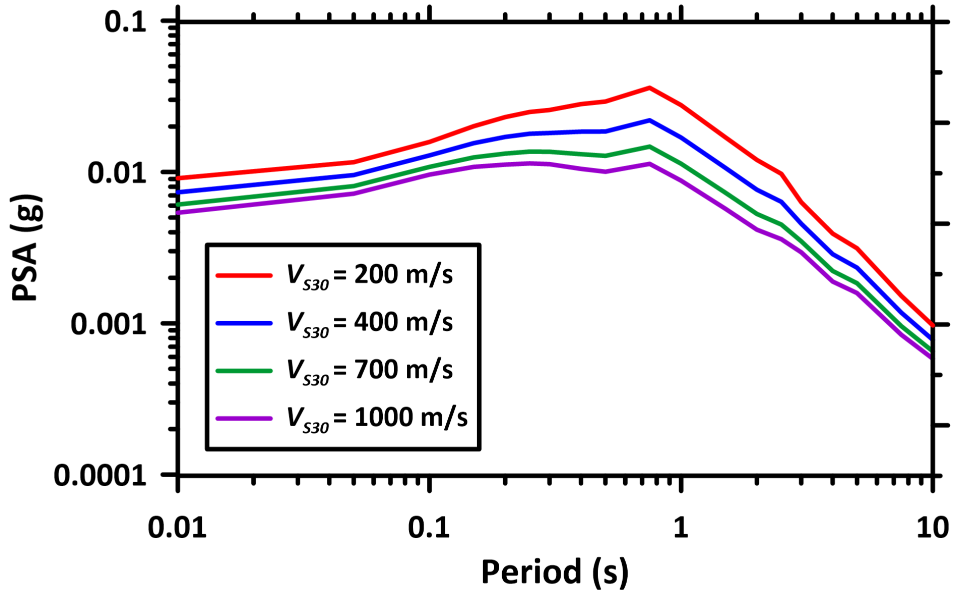

Figure 6.9. Predictions of response spectra computed using the global reference rock GMM from Chapter 5 with the global site amplification model described herein for an interface **M**7 event at R_{rup} = 200 km, for V_{S30} = 200, 400, 700, 1000m/s.

An assumption implicit to the model development is that site response is not affected by event type, meaning that the model applies equally to interface and slab events. This is checked by plotting within event residuals as a function of V_{S30} for both event types, as shown in Figure 6.10. The lack of bias and flatness of the trends demonstrates that the assumption is valid.

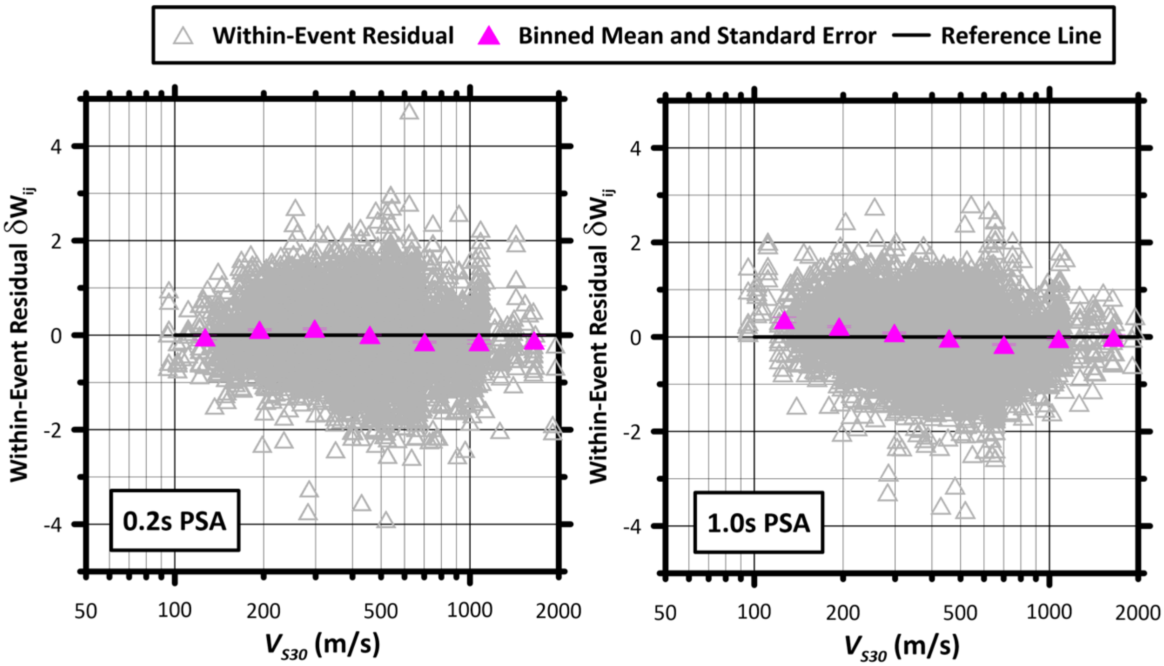

Figure 6.10. Within-event residuals for the global GMM presented in Chapter 5 in combination with the site amplification model presented herein, for 0.2 and 1.0s PSA. Data from interface and intraslab events are combined.

6.5 MODEL LIMITATIONS AND RECCOMENDED USE

The seismic site amplification model presented in this chapter is for use in conjunction with the NGA-Subduction GMM presented in Chapter 5. The model could also be used with other GMMs conditioned at 760 m/s, but a check should be performed for bias against ground motion data, which if present, would require adjustment of the GMM constant term.

The base seismic site amplification model, F_S is applicable to PGA, PGV, and PSA between 0.01-10s oscillator periods. It should not be used outside of the range of V_{S30} used in model building, 150-2000m/s. Regional coefficients are recommended for Alaska, Cascadia, Japan, South America, and Taiwan. The global model is recommended in Central America and Mexico. For forward use in regions not included in model development, we recommend using a range of s_2 values that captures the range of regional epistemic uncertainty.

The site response model presented here is ergodic. It will not produce a site-specific amplification factor, even with a measured V_{S30} profile from a site of interest. Site-specific site response can be evaluated separately using recordings at or near the site of interest or GRA simulations using a measured V_S profile (Stewart et al. 2017).

7 OVERVIEW OF FINDINGS AND FUTURE WORK

7.1 OVERVIEW OF DISSERTATION SCOPE AND FINDINGS

This dissertation focuses on model building for the prediction of earthquake ground motion intensity measures based on properties of the earthquake source, wave propagation path, and recording site. The first part of this thesis focuses on seismic site characterization and site amplification in CENA in the context of the Next Generation Attenuation-East project. Chapter 2 presents a hybrid geology-slope approach for V_{S30} estimation that utilized a new and expanded shear-wave velocity (V_S) measurement database for CENA. The proxy is conditioned on geologic category from newly considered large-scale geologic maps, the extent of Wisconsin glaciation, sedimentary basin structure, and 30 arc-sec topographic gradient. Nonglaciated sites were found to have a modest natural log dispersion of V_{S30} ($\sigma_{\ln V}$=0.36) relative to glaciated sites ($\sigma_{\ln V}$=0.66), indicating better predictability of V_{S30} for the former. These findings were used to estimate the mean and standard deviation of V_{S30} for NGA-East recording stations when measurements were not available. Chapter 3 presents empirical linear site amplification models conditioned on time-averaged shear wave velocity in the upper 30 m (V_{S30}) for CENA, developed using a combination of least-squares, mixed effects, and Bayesian techniques. Site amplification is found to scale with V_{S30} for intermediate to stiff site conditions (V_{S30} > 300 m/s) in a weaker manner than for active tectonic regions. For stiff sites (> 800 m/s), we find differences in amplification for previously glaciated and non-glaciated regions, with non-glaciated sites having lower amplification. The models account for predictor uncertainty, which does not affect the median model, but decreases model dispersion. Lastly, Chapter 4 presents recommendations for modeling of ergodic site amplification in CENA, based primarily on results from the literature (including Chapter 3), for application in the U.S. Geological Survey national seismic hazard maps. Previously, the maps have used site factors developed using data and simulations for active tectonic regions; however, results from NGA-East demonstrate different levels of site amplification in CENA as compared to active regions. The recommended model has three terms, two of which describe linear site amplification: an empirically constrained V_{S30}-scaling term relative to a 760 m/s reference, and a simulation-based term to adjust site amplification from the 760 m/s to the CENA reference of V_S=3000 m/s.

The second part of this thesis focuses on the development of a global GMM and site amplification model with regional adjustment factors for subduction zone regions as a part of the Next Generation Attenuation-Subduction (NGA-Sub) project. Chapter 5 presents global subduction zone GMMs for interface and intraslab events, with regionalized terms for Alaska, Cascadia, Central America. Mexico, Japan, South America, and Taiwan. The near-source saturation model, magnitude-dependent geometrical spreading, and magnitude-scaling break point are constrained using models derived by others from simulations and fault geometry, and the anelastic attenuation, magnitude scaling, and depth scaling terms are constrained empirically. The model is regionalized in the constant, anelastic attenuation, and depth-scaling terms, and in the magnitude break-point. When applying the model to a region not considered in the study, we recommend using an appropriate range of epistemic uncertainty that captures regional variation. Chapter 6 presents a subduction-specific site amplification model, meant to be paired with the reference-rock GMM of Chapter 5, that accounts for regional differences in V_{S30}-scaling, and re-calibrates the nonlinear term of SS14 using subduction data.

7.2 FUTURE WORK

7.2.1 Basin Depth Terms

For regions that contain basin structures, such as Tokyo in Japan, the Taipei Basin in Taiwan, and the Seattle and Tacoma Basins in Cascadia, we will define a period-dependent basin depth term, F_b, that is an adjustment to the base model, F_s, given in Eq. 6.1. The basin depth model will be conditioned on δz_x (Eq. 7.2), which represents the difference between the z_x at a site, and the z_x predicted by an empirical $V_{S30} - z_x$ relationship, $\mu_{zx}(V_{S30})$, where z_x is the depth to the x km/s crustal shear wave velocity horizon (x is taken as 2.5 km/s in Cascadia, and 1.0 km/s elsewhere).

$$F_b = f(T, \delta z_1) \tag{7.1}$$

$$\delta z_1 = z_1 - \mu_{z1}(V_{S30}) \tag{7.2}$$

This component of the model is still undergoing development, and a final functional form for F_b (Eq. 7.1) has not been chosen yet. It is likely we will use a version of the Nweke et al. (2018) μ_{z1} model (Eq. 6.7), adjusting coefficients using the NGA-Subduction database if necessary.

$$\mu_{z1} = \theta_1 \left[1 + erf\left(\frac{log(V_{S30}) - log(\upsilon_\mu)}{\upsilon_\sigma \sqrt{2}}\right)\right] + \theta_0 \tag{7.3}$$

We plan to fit F_b (Eq 7.1) to within-event residuals (δW_{ij}) at stations with an estimate of z_1 using a nonlinear least-squares approach. The within-event residuals will be computed using the GMM and event terms presented in Chapter 5, and the base site amplification model, F_S, given in Eqs. 6.1-6.4:

$$\delta W_{ij} = ln(Y_{ij}) - [\mu_{ij} + F_{S,j} + \eta_{E,i}] \tag{7.4}$$

Where Y_{ij} is a ground motion intensity measure from earthquake i recorded at station j. We will allow for regional F_B coefficients if we observe differences in the behavior of the data between Japan, Taiwan, and Cascadia.

7.2.2 NGA-Subduction Model Validation

After completion of the seismic site amplification model for subduction zones (Chapter 6) and the model verification process at multiple site conditions are complete, we plan to validate our NGA-Subduction ground motion model against data that was not included in the model development process (Chapter 5.2.2). There are three sources that we will consider:

1. Data from New Zealand (NZ) in the NGA-Subduction database (Kishida et al. 2018)
2. Data from subduction zone earthquakes that have occurred since NGA-Subduction database development, including the 2017 Chiapas and Puebla, Mexico earthquakes and the 2018 Gulf of Alaska earthquake, and
3. The Frankel et al. (2018) ground motions for Cascadia interface events produced using broadband simulations.

The NGA-Sub NZ dataset was excluded from model development due to a lack of Class 1/Class 2 classifications for the events (Wooddell 2018). We did not want to include ground motions from earthquakes that were potentially aftershock events in our regression analysis. However, we can use the 4,200 recordings from 263 events with **M**4-7.8 for validation purposes, and to see if any regional trends or differences exist in NZ.

The Frankel et al. (2018) simulations are only for **M**9 Cascadia interface events. Due to lack of data, we did not produce an interface model for the Cascadia region as part of the NGA-

Subduction project. However, we will compare our available interface models to the Frankel et al. (2018) simulations and seek insights from the process.

7.2.3 Residuals analysis in Regional Back-Arc Complexes

The NGA-Subduction ground motion model presented in Chapter 5 is only applicable to sites in the forearc regions of subduction zones. Data recorded at sites in the back-arc regions were excluded to simplify the path modelling process. Using the completed GMM and site amplification models we plan to undertake a residuals analysis to determine if there are any attenuation differences between the forearc and back-arc zones. Stronger attenuation in the back-arc has been observed in Japan, although we expect this to vary by region; some other NGA-Subduction model developers have gone through this process and have not found a difference between the forearc and back-arc ground motions for Alaska and Cascadia (N. Abrahamson, 2018, pers. comm.).

If a difference is observed, we plan to implement an additional anelastic attenuation term (Eq. 5.2) that modifies the distance-scaling for the fraction of the path inside the back-arc. In this case, the path model F_P may look like:

$$F_P = c_1 lnR + (b_3 + b_4\boldsymbol{M})lnR + \ a_0 R_{FA} + a_{BA} R_{BA} \tag{7.5}$$

Where the first two terms are the same as defined in Eqs. 5.2-5.3, a_0 is the same coefficient as given in Chapter 5.4.2, but R_{FA} and R_{BA} are defined as the fraction of the site-to-source distance R (Eq. 5.3) in the forearc and back-arc regions, respectively, and a_{BA} is an anelastic attenuation coefficient determined via regression on back-arc residuals.

7.2.4 Hypocenter Location Model Validation

Although our choice of hypocentral depth as a parameter during model development (Ch. 5.x) was based on a number of considerations, including predictive power and level of predictor variable uncertainty, it is not the ideal parameter for forward application in PSHA. This is because including it in a hazard analysis would require adding randomization of source location on the fault, which would involve an additional loop in the hazard integral. Both Z_{tor} and Z_{mid} are more convenient for PSHA application because they are determined once a fault rupture plane is defined, which is already part of hazard analysis for distance calculation.

Mai et al. (2005) examined the location of hypocenters within the finite fault rupture plane for 50 earthquakes in a number of tectonic settings, including subduction zones. Based on their empirical analysis, they define gamma probability distributions of down-dip hypocenter locations for strike-slip, dip-slip, crustal dip-slip, and subduction dip-slip events. We plan to use subduction dip-slip events from the NGA-Subduction database with finite fault models (PEER 2019, Contreras 2017) to validate the probability distribution of down-dip hypocenter location of Mai et al. (2005). Given a rupture plane defined in PHSA, this will allow for the estimation of d_{hyp} and the use of the GMM presented herein without an additional randomization over source properties. If desired, epistemic uncertainty in this model can be taken using the mean and standard deviation of the down-dip probability distribution.

7.2.5 Epistemic Uncertainty

The last component of planned future work is to develop quantitative epistemic uncertainty recommendations of two kinds:

1. Regional epistemic uncertainty based on the GMM presented herein, and

2. Model-to-model epistemic uncertainty for one region based on the suite of available subduction zone GMMs, including other NGA-Subduction GMMs (Abrahamson and Gulerce 201x; Kuehn et al. 201x; Chiou et al. 201x).

The first component can be carried out independently, and has been addressed qualitatively in this dissertation (e.g. Chapter 5.7.2). We will recommend specific epistemic ranges on regional coefficients c_0, a_0, and s_2. The second component will be addressed collectively by all of the NGA-Subduction GMM developers.

REFERENCES

Abrahamson, N.A., Gregor, N. and Addo, K. (2016). BC Hydro Ground Motion Prediction Equations for Subduction Earthquakes. *Earthq. Spectra*, **32**, 23-44.

Abrahamson, N.A., Kuehn, N., Gulerce, Z., Gregor, N., Bozorgnia, Y., Parker, G.A., Stewart, J.P., Chiou, B., Idriss, I.M., Campbell, K., and Youngs, R. (2018). Update of the BC Hydro Subduction Ground-Motion Model using the NGA-Subduction Dataset. *PEER Report 2018/02*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Abrahamson, N.A., Silva, W.J. and Kamai, R. (2014). Summary of the ASK14 ground motion relation for active crustal regions. *Earthq. Spectra*, **30**, 1025-1055.

Abrahamson, N.A., and Youngs, R.R. (1992). A stable algorithm for regression analyses using the random effects model, *Bull. Seismol. Soc. Am.*, **82**, 505–510.

Aboye S.A., Andrus R.D., Ravichandran N, Bhuiyan A.H., Harman N. (2015). Seismic site factors and design response spectra based on conditions in Charleston, South Carolina. *Earthq. Spectra*, **31**, 723–744.

Ahdi, S.K., Ancheta, T.D., Contreras, V., Kishida, T., Kwak, D.Y., Kwok A.O-L., Parker, G.A., Bozorgnia, Y., and Stewart, J.P. (2017). NGA-Subduction site database. *Proc. 16th World Conf. on Earthquake Eng.*, Santiago, Chile (Paper No. 4926).

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

Al Noman M.N., and Cramer C.H. (2015). Empirical ground-motion prediction equations for Eastern North America *PEER Report No. 2015/04*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, pp. 193–212.

Atkinson, G.M. (1997). Empirical ground motion relations for earthquakes in the Cascadia region. *Can. J. Civil Eng.*, **24**, 64-77.

Atkinson, G.M. (2008). Ground-motion prediction equations for eastern North America from a referenced empirical approach: Implications for epistemic uncertainty, *Bull. Seismol. Soc. Am.,* **98**, 1304–1318.

Atkinson G.M. (2012). Evaluation of attenuation models for the northeastern United States/southeastern Canada. *Seismol. Res. Letters* **83**, 166–178.

Atkinson, G.M. and Boore, D.M. (2003). Empirical ground-motion relations for subduction-zone earthquakes and their application to Cascadia and other regions. *Bull. Seismol. Soc. Am.*, **93**, 1703-1729.

Atkinson, G.M. and Boore D.M. (2006). Earthquake ground-motion prediction equations for eastern North America, *Bull. Seismol. Soc. Am.*, **96**, 2181–2205.

Atkinson, G.M. and Boore, D.M. (2011). Modifications to existing ground-motion prediction equations in light of new data, *Bull. Seismol. Soc. Am.*, **101**(3), 1121-1135.

Atkinson G.M., Hassani B., Singh A., Yenier E., and Assatourians K. (2015). Estimation of moment magnitude and stress parameter from ShakeMap ground-motions, *Bull. Seismol. Soc. Am.*, **105**, 2572–2588.

Atkinson, G.M. and Macias, M. (2009). Predicted ground motions for great interface earthquakes in the Cascadia subduction zone. *Bull. Seismol. Soc. Am.*, **99**, 1552-1578.

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

Atkinson, G.M., Yenier, E., Sharma, N. and Convertito, V. (2016). Constraints on the near-distance saturation of ground-motion amplitudes for small-to-moderate induced earthquakes. *Bull. Seismol. Soc. Am*, **106**, 2104-2111.

Archuleta, R. and Ji, C. (2018)

Assatourians, K. and Atkinson, G.M. (2012). EXSIM12: A Stochastic Finite-Fault Computer Program in FORTRAN, http://www.seismotoolbox.ca (last accessed November 2017).

Baher, S.A. and Davis, P.M. (2003). An application of seismic tomography to basin focusing of seismic waves and Northridge earthquake damage, *J. Geophy. Res.-Solid Earth*, **108**(B2), Art. No.2122.

Bates D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Statistical Software*, **67**, 1-48.

Baturay, M.B. and Stewart, J.P. (2003). Uncertainty and bias in ground-motion estimates from ground response analyses. *Bull. Seismol. Soc. Am.*, **93**, 2025-2042.

BC Hydro (2012). Probabilistic Seismic Hazard Analysis (PSHA) Model, vols. 1–4, *BC Hydro Engineering Report E658*, Vancouver, Canada.

Beresnev I.A., Atkinson G.M. (1997). Shear-wave velocity survey of seismographic sites in eastern Canada: Calibration of empirical regression method of estimating site response, *Seismol. Res. Lett.*, **68**, 981–987.

Bilek, S.L. and Lay, T. (1998). Variation of interplate fault zone properties with depth in the Japan subduction zone. *Science*, **281**, 1175-1178.

Bilek, S.L. and Lay, T. (1999). Rigidity variations with depth along interplate megathrust faults in subduction zones. *Nature*, **400**, 443.

Boore, D.M. (2009). Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM. *Bull. Seismol. Soc. Am.*, **99**, 3202-3216.

Boore D.M. (2013). The uses and limitations of the square-root impedance method for computing site amplification, *Bull. Seismol. Soc. Am.*, **103**, 2356–2368.

Boore D.M. (2015a). Point-source stochastic-method simulations of ground motions for the PEER NGA-East Project, *PEER Report No. 2015/04.* Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Boore, D.M. (2015b). Adjusting ground-motion intensity measures to a reference site for which V_{S30}=3000 m/s, *PEER Report 2015/06*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Boore, D. M. (2018). Ground-motion models for very-hard-rock sites in eastern North America: An update, *Seismol. Res. Lett.* **89**, 1172–1184.

Boore, D. M., and Campbell, K.W. (2017). Adjusting central and eastern North America ground-motion intensity measures between sites with different reference-rock site conditions. *Bull. Seismol. Soc. Am.*, **107**, 132-148.

Boore, D.M., and Goulet, C.A. (2014). The effect of sampling rate and anti-aliasing filters on high-frequency response spectra. *Bull. Earthq. Eng.*, **12**, 203-216.

Boore D.M. and Joyner W.B. (1997). Site amplifications for generic rock sites, *Bull. Seismol. Soc. Am.*, **87**, 327–341.

Boore, D.M. and Thompson, E.M. (2015). Revisions to some parameters used in stochastic-method simulations of ground motion, *Bull. Seismol. Soc. Am.*, **105**, 1029-1041.

Borcherdt, R.D., and Gibbs, J.F. (1976). Effects of local geological conditions in the San Francisco Bay region on ground motions and the intensities of the 1906 earthquake. *Bull. Seismol. Soc. Am.*, **66**, 467-500.

Borcherdt, R.D., and Glassmoyer, G. (1994). Influences of local geology on strong and weak ground motions recorded in the San Francisco Bay region and their implications for site-specific building-code provisions, *The Loma Prieta, California Earthquake of October 17, 198—-Strong Ground Motion,* U. S. Geological Survey Professional Paper 1551-A, A77-A108

Bozorgnia, Y., Abrahamson, N.A., Atik, L.A., Ancheta, T.D., Atkinson, G.M., Baker, J.W., Baltay, A., Boore, D.M., Campbell, K.W., Chiou, B.S.J.m and Darragh, R. (2014). NGA-West2 research project. *Earthq. Spectra*, **30**, 973-987.

Campbell, K.W. (2003). Prediction of strong ground motion using the hybrid empirical method and its use in the development of ground-motion (attenuation) relations in eastern North America, *Bull. Seismol. Soc. Am.*, **93**, 1012–1033.

Campbell, K.W. (2009). Estimates of shear-wave Q and κ_0 for unconsolidated and semiconsolidated sediments in Eastern North America, *Bull. Seismol. Soc. Am.*, **99**, 2365-2392.

Campbell, K.W. (201x). Reduced Magnitude Scaling of Earthquake Ground Motion and Source Dimensions in Subduction Mega-Earthquakes. *Earthq. Spectra*, *submitted.*

Campbell, K.W. and Bozorgnia, Y. (2014). NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthq. Spectra*, **30**, 1087-1115.

Chiou, B., and Youngs, R.R. (2008). An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthq. Spectra*, **24**, 173-215.

Coleman, J.L., Jr., and S.M. Cahan (2012). Preliminary catalog of the sedimentary basins of the United States. *Open-File Report 2012–1111*, U.S. Geological Survey, 27 p. (plus 4 figures and 1 table available as separate files) Available online at http://pubs.usgs.gov/of/2012/1111/.

Contreras, V. (2017). Development of a Chilean Ground Motion Database for the NGA-Subduction Project. *Master's Thesis*. University of California Los Angeles, Los Angeles, California.

Crouse, C.B., Vyas, Y.K. and Schell, B.A. (1988). Ground motions from subduction-zone earthquakes. *Bull. Seismol. Soc. Am.*, **78**,1-25.

Crow, H., Pyne, M. Hunter, J.A., Pullan, S.E., Motazedian, D., and Pugin, A. (2007). Shear wave measurements for earthquake response evaluation in Orleans, Ontario. *Geological Survey of Canada. Open-File*, *5579*, 192.

Darragh R.B., Abrahamson N.A., Silva W.J., and Gregor N. (2015). Development of hard rock ground-motion models for Region 2 of Central and Eastern North America, *PEER Report No. 2015/04*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, pp. 51–69.

Dawood, H.M., Rodriguez-Marek, A., Bayless, J., Goulet, C., and Thompson, E. (2016). A Flatfile for the KiK-net Database Processed Using an Automated Protocol. *Earthq. Spectra*, **32**, 1281-1302.

Day, S.M., Graves, R.W., Bielak, J., Dreger, D., Larsen, S., Olsen, K.B., Pitarka, A., and Ramirez-Guzman, L. (2008). Model for basin effects on long-period response spectra in southern California. *Earthq. Spectra*, **24**, 257-277.

Dobry, R., Borcherdt, R.D., Crouse, C.B., Idriss, I.M., Joyner, W.B., Martin, G.R., Power, M.S., Rinne, E.E. and Seed, R.B. (2000). New site coefficients and site classification system used in recent building seismic code provisions. *Earthq. Spectra*, **16**, 41-67.

EPRI (1993). Guidelines for determining design basis ground motions, *Report No. EPRI TR-102293*, Electrical Power Research Institute, Palo Alto, CA.

Field E.H., and Jacob K.H. (1995). A comparison and test of various site response estimation techniques, including three that are not reference site dependent, *Bull. Seismol. Soc. Am.*, **85**, 1127–1143.

Frankel A.D. (2015). Ground-motion predictions for Eastern North American earthquakes using hybrid broadband seismograms from finite-fault simulations with constant stress-drop scaling, *PEER Report No. 2015/04*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, pp. 149–163.

Frankel A.D., Mueller C.S., Barnhard T., Perkins D., Leyendecker E., Dickman N., Hanson S., and Hopper M. (1996). National seismic hazard maps: Documentation, *USGS Open-File Rept. 96-532*, 69 pgs.

Frankel, A., Wirth, E., Marafi, N., Vidale, J. and Stephenson, W. (2018). Broadband synthetic seismograms for magnitude 9 earthquakes on the cascadia megathrust based on 3D simulations and stochastic synthetics, Part 1: Methodology and overall results. *Bull. Seismol. Soc. Am.*, **108**, 2347-2369.

Fullerton, D.S., C.A. Bush and J.N. Pennell (2003). Map of surficial deposits and materials in the eastern and central United States (East of 102° West Longitude), *U.S. Geological Investigations Series* I-2789.

Fulton, R.J. (1996). Surficial Materials of Canada, Geological Survey of Canada, Natural Resources Canada. Ottawa. Map 1880A, Scale 1:5,000,000.

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

Goulet, C.A., Bozorgnia, Y., Kuehn, N., Al Atik, L., Youngs, R.R., Graves, R.W., and Atkinson, G.M. (2017). NGA-East ground-motion models for the U.S. Geological Survey national seismic hazard maps, *PEER Report 2017/03*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Graizer V. (2015). Ground-motion prediction equations for the central and eastern United States, *PEER Report No. 2015/04*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, pp. 51–69.

Graves, R.W. (1993). Modeling three-dimensional site response effects in the Marina District, San Francisco, California, *Bull. Seism. Soc. Am.*, **83**, 1042-1063.

Graves, R.W., Pitarka, A., and Somerville, P.G. (1998). Ground motion amplification in the Santa Monica area: effects of shallow basin edge structure, *Bull. Seism. Soc. Am.*, **88**, 1224-1242.

Gregor, N., Silva, W., Wong, I., and Youngs, R. (2002). Ground-motion attenuation relationships for Cascadia subduction zone megathrust earthquakes based on a stochastic finite-fault modeling, *Bull. Seismol. Soc. Am.*, **92**, 1923–1932.

Gregor, N., Silva, W., Wong, I. and Youngs, R. (2006). Updated response spectral attenuation relationship for Cascadia subduction zone megathrust earthquakes. *Seismol. Res. Lett.*, **77**, 325-326.

Harmon, J.A., Hashash, Y.M.A., Stewart, J.P., Rathje, E.M., Campbell, K.W., Silva, W.J., Xu, B., Musgrove, M., and Ilhan, O., (2019a). Site amplification functions for central and eastern North America – Part I: Simulation dataset development, *Earthq. Spectra, In Press.*

Harmon, J.A., Hashash, Y.M.A., Stewart, J.P., Rathje, E.M., Campbell, K.W., Silva, W.J., Xu, B., Musgrove, M., and Ilhan, O. (2019b). Site amplification functions for central and eastern North America – Part II: Modular simulation-based models, *Earthq. Spectra, In Press.*

Hashash, Y.M.A. and Park, D. (2001). Non-linear, one-dimensional seismic ground motion propagation in the Mississippi embayment, *Eng. Geo*., **62**, 185–206.

Hashash Y.M.A., Tsai C.C., Phillips C., Park D. (2008). Soil-column depth-dependent seismic site coefficients and hazard maps for the upper Mississippi Embayment, *Bull. Seismol. Soc. Am*., **98**, 2004–2021.

Hashash, Y.M.A., A.R. Kottke, J.P Stewart, K.W. Campbell, B. Kim, C. Moss, S. Nikolaou, E.M.Rathje, and W.J. Silva (2014). Reference rock site condition for central and eastern North America. *Bull. Seismol. Soc. Am.*, **104**, 684-701.

Hashash, Y.M.A., J.A. Harmon, O. Ilhan, G.A. Parker, J.P. Stewart (2017). Recommendations for Ergodic Nonlinear Site Amplification in Central and Eastern North America. *PEER Report 2017/05*, Pacific Earthquake Engineering Research Center, University of California Berkeley, Berkeley, CA.

Hashash, Y.M.A., J.A. Harmon, O. Ilhan, G.A. Parker, J.P. Stewart, E.M. Rathje, K.W. Campbell, and W.J. Silva (201x) – Nonlinear Site Amplification Model for Ergodic Seismic Hazard Analysis in Central and Eastern North America, *in preparation*.

Hassani, B., and Atkinson, G. M. (2015). Referenced empirical ground-motion model for eastern North America. *Seismol, Res. Lett.*, **86**, 477-491.

Hassani B, Atkinson G.M. (2016a). Site-effects model for central and eastern North America based on peak frequency, *Bull. Seismol. Soc. Am.*, **106**, 2197–2213.

Hassani, B. and G.M. Atkinson (2016b). Applicability of the site fundamental frequency as a V_{S30} proxy for central and eastern North America, *Bull. Seismol. Soc. Am.*, **106**, doi: 10.1785/0120150259.

Hassani, B. and Atkinson, G.M. (2017). Site-effects model for central and eastern North America based on peak frequency and average shear-wave velocity, *Bull. Seismol. Soc. Am.*, **108**, 338-350.

Hassani, B. and Atkinson, G.M. (2018). Adjustable Generic Ground-Motion Prediction Equation Based on Equivalent Point-Source Simulations: Accounting for Kappa Effects. *Bull. Seismol. Soc. Am.*, **108**, 913-928.

Hollenback J., Kuehn N., Goulet C.A., and Abrahamson N.A., (2015a). PEER NGA-East median ground-motion models, *PEER Report No. 2015/04*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, pp. 274–309.

Hollenback, J., Silva, W.J., Darragh, R.B., and Abrahamson, N.A. (2015b). Adjustment for gulf coast and Mississippi Embayment region. *In* NGA-East: adjustments to median ground-motion models for central and eastern North America. *PEER Report 2015/08,* Pacific Earthquake Engineering Research Center, Berkeley, CA.

Hwang H.H.M., Huo J.–R. (1994). Generation of hazard-consistent fragility curves for seismic loss estimation studies, *Technical Report NCEER-94-O015,* National Center for Earthquake Engineering Research, State University of New York at Buffalo, Buffalo, NY, 1994.

Hwang H.H.M., Lin H., Huo J.–R. (1997). Site coefficients for design of buildings in eastern United States. *Soil Dyn. Earthq. Eng.,* **16**, 29–40.

Kaklamanos, J., and Bradley, B. (2018). Challenges in predicting seismic site response with 1D analyses: Conclusions from 114 KiK-net vertical seismometer arrays, *Bull. Seismol. Soc. Am.*, **108**, 2816-2838.

Kishida T., Contreras V., Bozorgnia Y., Abrahamson N.A., Ahdi S.K., Ancheta T., Boore D.M., Campbell K., Chiou B., Darragh R., Gregor N., Kuehn N., Kwak D., Kwok A., Lin P., Magistrale H., Mazzoni S., Muin S., Midorikawa S., Si H., Silva W., Stewart J.P., Wooddell K., and Youngs R. (2018). NGA-Sub Ground Motion Database *Proceedings of the 11th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Los Angeles, CA.

Idriss, I.M. (1990). Response of soft soil sites during earthquakes*, Proc. H. Bolton Seed Memorial Symposium*, J.M. Duncan (ed.), Vol. 2, 273-290.

Kamai, R., Abrahamson, N. A., and Silva, W. J. (2014). Nonlinear horizontal site amplification for constraining the NGA-West2 GMPEs. *Earthq. Spectra*, **30**, 1223-1240.

Kawase, H. (1996). The cause of the damage belt in Kobe: 'The basin edge effect,' constructive interference of the direct S-wave with the basin induced diffracted/Rayleigh waves, *Seism. Res. Letters*, **67**, 25-34.

Kim, B., Y.M.A. Hashash, E.M. Rathje, J.P. Stewart, S. Ni, P.G. Somerville, A.R. Kottke, W.J. Silva, and K.W. Campbell (2016). Subsurface shear-wave velocity characterization using P-wave seismograms in Central and Eastern North America. *Earthq. Spectra*, **32**,143-169.

Khaheshi Banab, K., Kolaj, M., Motazedian, D., Sivathayalan, S., Hunter, J.A., Crow, H.L., Pugin, A.J.M., Brooks, G.R. and Pyne, M. (2012). Seismic site response analysis for Ottawa, Canada: A comprehensive study using measurements and numerical simulations. *Bull. Seismol. Soc. Am.*, **102**, 1976-1993.

Kottke, A.R., Y.M.A. Hashash, J.P. Stewart, C.J. Moss, S. Nikolaou, E.M. Rathje, W.J. Silva, and K.W. Campbell (2012). Development of geologic site classes for seismic site amplification for central and eastern North America. *Proc. 15th World Conf. on Earthquake Eng.*, Sept 24-28, 2012, Lisbon, Portugal. Paper No. 4557.

Kwak, D.Y., Stewart, J.P., Mandokhail, S.U.J., and Park, D. (2017). Supplementing V_{S30} with H/V Spectral Ratios for Predicting Site Effects. *Bull. Seismol. Soc. Am.*, **107**, 2028-2042.

Kwok, A.O. and Stewart, J.P. (2006). Evaluation of the effectiveness of theoretical 1D amplification factors for earthquake ground-motion prediction, *Bull. Seism. Soc. Am.*, **96**, 1422-1436.

Lin H., Hwang H., Huo J.-R. (1996). A study on site coefficients for new site categories specified in the NEHRP Provisions. *Technical Report*, Center for Earthquake Research and Information. The University of Memphis. Memphis, TN.

Lin, P.S., and Lee, C.T. (2008). Ground-motion attenuation relationships for subduction-zone earthquakes in Northeastern Taiwan, *Bull. Seismol. Soc. Am.*, **98**, 220–240.

Mai, P.M., Spudich, P., and Boatwright, J. (2005). Hypocenter locations in finite-source rupture models. *Bull. Seismol. Soc. Am.*, **95**, 965-980.

Moss, R.E.S. (2009). Reduced uncertainty of ground motion prediction equations through Bayesian variance analysis. *PEER Report 20019/105.* Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Moss, R. E. S. (2011). Reduced sigma of ground-motion prediction equations through uncertainty propagation. *Bull. Seismol. Soc. Am.*, **101**, 250-257.

Motazedian, D. and Atkinson, G.M. (2005). Stochastic finite-fault modeling based on a dynamic corner frequency. *Bull. Seismol. Soc. Am.*, **95**, 995-1010.

Novakovic, M., Atkinson, G.M. and Assatourians, K. (2018). Empirically calibrated ground-motion prediction equation for Oklahoma. *Bull. Seismol. Soc. Am.,* **108,** 2444-2461.

Nweke, C.C., Wang P., Brandenberg, S.J., and Stewart, J.P. (2018). Reconsidering Basin Effects in Ergodic Site Response Models. *Proceedings of the Strong Motion Instrumentation Program Seminar on Utilization of Strong Motion Data.* California Strong Motion Instrumentation Program, October 25, Sacramento, California.

Ontario Geological Survey (2000). Quaternary geology, seamless coverage of the Province of Ontario; Ontario Geological Survey, Data Set 14 - Revised.

Pacific Earthquake Engineering Research Center (2015a). NGA-East: Median ground-motion models for central and eastern North America. *PEER Report No. 2015/04.* Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Pacific Earthquake Engineering Research Center (2015b). NGA-East: Adjustments to median ground-motion models for central and eastern North America. *PEER Report 2015/08.* Pacific Earthquake Engineering Research Center, Berkeley, CA.

Pacific Earthquake Engineering Research Center (2019) – Next Generation Attenuation – Subduction database report, *in preparation*.

Palmer, J.R. (2006). Surficial material geologic map of the Wentzville 7 1/2' quadrangle, St. Charles County, Missouri: Missouri Department of Natural Resources, Division of Geology and Land Survey, Open-File Map OFM-06-518-GS, scale 1:24,000

Park D., Hashash Y.M.A. (2005a). Evaluation of seismic site factors in the Mississippi Embayment. I. Estimation of dynamic properties, *Soil Dyn. Earthq. Eng.*, **25**, 133–144.

Park D., Hashash Y.M.A. (2005b). Evaluation of seismic site factors in the Mississippi Embayment. II. Probabilistic seismic hazard analysis with nonlinear site effects, *Soil Dyn. Earthq. Eng.*, **25**, 145–156.

Parker, G.A., Harmon, J.A., Stewart, J.P., Hashash, Y.M.A., Kottke, A.R., Rathje, E.M., Silva, W.J. and Campbell, K.W. (2017). Proxy-based V_{S30} estimation in central and eastern North America. *Bull. Seismol. Soc. Am.*, **107**, 117-131.

Parker, G.A., Stewart, J.P., Hashash, Y.M.A., Rathje, E.M., Campbell, K.W., and Silva, W.J. (2019). Empirical linear seismic site amplification in central and eastern North America. *Earthq. Spectra, Preprint.* https://doi.org/10.1193/083117EQS170M.

Petersen, M.D., Moschetti, M.P., Powers, P.M., Mueller, C.S., Haller, K.M., Frankel, A.D., Zeng, Y., Rezaeian, S., Harmsen, S.C., Boyd, O.S., Field, N., Chen, R., Rukstales, K.S., Luco, N., Wheeler, R.L., Williams, R.A., and Olsen, A.H. (2015). The 2014 United States national seismic hazard model. *Earthq. Spectra*, **31**, S1-S30.

Petersen, M.D., Shumway, A.M., Powers, P.M., Mueller, C.S., Frankel, A.D., Rezaeian, S., Moschetti, M.P., McNamara, D.M., Thompson, E.M., Boyd, O.S., Luco, N., Hoover, S.M., and Rukstales, K.S. (2018). Preliminary 2018 National Seismic Hazard Model for the conterminous United States, *1th National Conf. on Earthquake Engineering,*, June 25-29, 2018, Los Angeles, CA.

Pezeshk S., Zandieh A., Campbell K.W., and Tavakoli B. (2015). Ground-motion prediction equations for CENA using the hybrid empirical method in conjunction with NGA-West2 empirical ground-motion models, *PEER Report No. 2015/04*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, pp. 119–147.

Pezeshk, S., Zandieh, A., and Tavakoli, B. (2011). Hybrid empirical ground-motion prediction equations for eastern North America using NGA models and updated seismological parameters, *Bull. Seismol. Soc. Am.*, **101**, 1859-1870.

Pitarka, A., Irikura, K., Iwata, T. and Sekiguchi, H. (1998). Three-dimensional simulation of the near-fault ground motion for the 1995 Hyogo-ken Nanbu (Kobe), Japan, earthquake, *Bull. Seism. Soc. Am.*, **88**, 428-440.

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

R Development Core Team (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0. Available online at: http://www.R-project.org.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reed, J.C., and C.A. Bush (2005). Generalized geologic map of the United States, Puerto Rico, and the U.S. Virgin Islands. *USGS National Atlas Small-Scale Collection*. Available online at: http://pubs.usgs.gov/atlas/geologic/.

Rogers, A. M., and Perkins, D.M. (1996). Monte Carlo simulation of peak acceleration attenuation using a finite-fault uniform-patch model including isochron and extremal characteristics, *Bull. Seismol. Soc. Am.* **86**, 79-92.

Romero S.M., and Rix G.J. (2001). Ground motion amplification of soils in the upper Mississippi Embayment, National Science Foundation Mid America Earthquake Center, *Report No. GIT-CEE/GEO-01-1*, Available at: http://hdl.handle.net/2142/8940

Sandıkkaya, M. A., Akkar, S., and Bard, P.Y. (2013). A nonlinear site-amplification model for the next pan-European ground-motion prediction equations. *Bull. Seismol. Soc. Am.*, **103**, 19-32.

Scherbaum, F., Kuehn, N.M., Ohrnberger, M. and Koehler, A. (2010). Exploring the proximity of ground-motion models using high-dimensional visualization techniques. *Earthq. Spectra*, **26**, 1117-1138.

Seed, H.B., Ugas, C. and Lysmer, J. (1976). Site-dependent spectra for earthquake-resistant design. *Bull. Seismol. Soc. Am.*, **66**, 221-243.

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

Seyhan, E., Stewart, J.P., Ancheta, T.D., Darragh, R.B., and Graves, R.W. (2014). NGA-West 2 site database*, Earthq. Spectra*, **30**, 1007-1024.

Shahjouei A., and Pezeshk, S. (2015). Hybrid empirical ground-motion model for Central and Eastern North America using hybrid broadband simulations and NGA-West2 GMPEs, *PEER Report No. 2015/04*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, pp. 165–192.

Si, H. and Midorikawa, S. (1999). New attenuation relationships for peak ground acceleration and velocity considering effect of fault type and site condition, *J. Struct. Constr. Eng.*, **523**, 63–70 in Japanese with English abstract and figure captions).

Silva W.J., Abrahamson N.A., Toro G., and Costantino C. (1997). Description and validation of the stochastic ground motion model*, Contract No. 770573*, *Report submitted to Brookhaven National Laboratory*, Associated Universities, Inc. Upton, NY.

Silva, W.J., Gregor, N., and Darragh, R. (2002). Development of regional hard rock attenuation relations for central and eastern North America, *Pacific Engineering and Analysis*, El Cerrito, California, 57 pp.

Skarlatoudis, A.A., Papazachos, C.B., Margaris, B.N., Ventouzi, C., Kalogeras, I. and EGELADOS Group (2013). Ground-motion prediction equations of intermediate-depth earthquakes in the Hellenic Arc, southern Aegean subduction area. *Bull.Seismol. Soc. Am.*, **103**, 1952-1968.

Snedecor, G.W. and Cochran, W.G. (1989). *Statistical Methods*, 8th edition, Ames, Iowa: Blackwell Publishing Professional. ISBN 978-0-8138-1561-9.

Soller, D.R., Reheis, M.C., Garrity, C.P., and Van Sistine, D.R. (2009). Map database for surficial materials in the conterminous United States, *U.S. Geological Survey Data Series 425,* scale 1:5,000,000. Available at: http://pubs.usgs.gov/ds/425/.

Stephenson, W.J., Williams, R.A., Odum, J.K., and Worley, D.M. (2000). High-resolution seismic reflection surveys and modeling across an area of high damage from the 1994 Northridge earthquake, Sherman Oaks, California, *Bull. Seism. Soc. Am.*, **90**, 643-654.

Stern, R.J. (2002). Subduction zones. *Rev. Geophys.*, **40**, 3-1.

Stewart, J. P., Liu, A.H., and Choi, Y. (2003). Amplification factors for spectral acceleration in tectonically active regions. *Bull. Seismol. Soc. Am.*, **93,** 332-352.

Stewart, J.P., Parker, G.A., Harmon, J.P., Atkinson, G.M., Boore, D.M., Darragh, R.B., Silva, W.J., and Hashash, Y.M.A. (2017a). Expert Panel Recommendations for Ergodic Site Amplification in Central and Eastern North America, *PEER Report 2017/04*, Pacific Engineering Research Institute, Berkeley, CA.

Stewart, J.P., Afshari, K., Goulet, C.A., (2017b). Non-ergodic site response in seismic hazard analysis. *Earthq. Spectra*, **33**, 1385–1414.

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

Thompson, E.M., Baise, L.G., Tanaka, Y., and Kayen, R.E. (2012). A taxonomy of site response complexity, *Soil Dyn. Earthq. Eng.*, **41**, 32–43.

Thompson, E.M. and Silva, W.J. (2013). Development of a hybrid slope-geology method of estimating VS30 for central and eastern North America. Report to USGS. Available at: http://earthquake.usgs.gov/research/external/reports/G12AP20003.pdf.

Thompson, E.M., Wald, D.J., Worden, C.B. (2014). A V_{S30} map for California with geologic and topographic constraints, *Bull. Seismol. Soc. Am*. **104**, 2313–2321.

Tukey, J. W. (1958). Bias and confidence in not quite large samples. *The Annals of Mathematical Statistics* **29**, 614–623.

Wald, D.J. and Allen, T.I. (2007). Topographic slope as a proxy for seismic site conditions and amplification. *Bull. Seisol. Soc. Am.*, **97**, 1379-1395.

Walling, M., Silva, W.J. and Abrahamson, N.A. (2008). Nonlinear site amplification factors for constraining the NGA models. *Earthq, Spectra*, **24**, 243-255.

Wills, C.J. and Clahan, K.B. (2006). Developing a map of geologically defined site-condition categories for California, *Bull. Seismol. Soc. Am*. **96**, 1483–1501.

Wirth, E.A., Frankel, A.D., Marafi, N., Vidale, J.E. and Stephenson, W.J. (2018). Broadband synthetic seismograms for magnitude 9 earthquakes on the Cascadia megathrust based on 3D simulations and stochastic synthetics, Part 2: Rupture parameters and variability. *Bull. Seismol. Soc. Am.*, **108**, 2370-2388.

Wooddell, K. (2018). New Earthquake Classification for the NGA-Subduction Project. *Proceedings of the 11th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Los Angeles, CA. 2018.

Yenier, E. and Atkinson, G.M. (2014). Equivalent point-source modeling of moderate-to-large magnitude earthquakes and associated ground-motion saturation effects. *Bull. Seismol. Soc. Am.*, **104**, 1458-1478.

Yenier, E. and Atkinson, G.M. (2015). Regionally adjustable generic ground-motion prediction equation based on equivalent point source simulations: Application to central and eastern North America. *Bull. Seismol. Soc. Am.*, **105**, 1989-2009.

Yenier, E., Atkinson, G.M. and Sumy, D.F. (2017). Ground motions for induced earthquakes in Oklahoma. *Bull. Seismol. Soc. Am.*, **107**, 198-215.

Yong, A., Hough, S.E., Iwahashi, J., and Braverman, A. (2012). A terrain based site conditions map of California with implications for the contiguous United States. *Bull. Seismol. Soc. Am.*, **102,** 114-1

Youngs, R.R., Chiou, S.J., Silva, W.J. and Humphrey, J.R. (1997). Strong ground motion attenuation relationships for subduction zone earthquakes. *Seismol. Res. Lett.*, ***68***, 58-73.

Youngs, R.R., Day, S.M. and Stevens, J.L. (1988). Near field ground motions on rock for large subduction earthquakes. In *Earthquake Engineering and Soil Dynamics II—Recent Advances in Ground-Motion Evaluation* (pp. 445-462). ASCE.

Zhao, J.X., Hu, J.S., Jiang, F., Zhou, J., and Rhoades, D.A. (2015). Nonlinear site models derived from 1-D analyses for ground-motion prediction equations using site class as the site parameter, *Bull. Seismol. Soc. Am.*, **105**, 2010–2022.

Zhao, J.X., Liang, X., Jiang, F., Xing, H., Zhu, M., Hou, R., Zhang, Y., Lan, X., Rhoades, D.A., Irikura, K. and Fukushima, Y. (2016a). Ground-motion prediction equations for subduction interface earthquakes in Japan using site class and simple geometric attenuation functions. *Bull. Seismol. Soc. Am.*, **106**, 1518-1534.

Zhao, J.X., Jiang, F., Shi, P., Xing, H., Huang, H., Hou, R., Zhang, Y., Yu, P., Lan, X., Rhoades, D.A. and Somerville, P.G. (2016b). Ground-motion prediction equations for subduction slab earthquakes in Japan using site class and simple geometric attenuation functions. *Bull. Seismol. Soc. Am.*, **106**, 1535-1551.

Zhao, J.X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H.K., Somerville, P.G. 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.