Title

Data-Informed Earthquake Ground Motion Characterization in Central and Eastern North America

Permalink

https://escholarship.org/uc/item/1g14983x

Author

Ramos-Sepulveda, Maria Elisa

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Data-Informed Earthquake Ground Motion Characterization in Central and Eastern North America

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

by

María Elisa Ramos-Sepúlveda

2024

ABSTRACT OF THE DISSERTATION

Data-Informed Earthquake Ground Motion Characterization in Central and Eastern North America

By

María Elisa Ramos-Sepúlveda

Doctor of Philosophy in Civil Engineering

University of California, Los Angeles, 2024

Professor Jonathan Paul Stewart, Chair

Earthquake ground motion intensity measures (*IM*s) are commonly estimated for engineering applications using ground motion models (GMMs) that estimate the median and lognormal standard deviation of *IM*s conditional on a series of source, path, and site parameters. Different GMMs are used in different tectonic regimes, and in the case of central and eastern North America (CENA), GMMs for stable continental regions are required. A major multi-agency, multi-investigator project (NGA-East) developed GMMs for CENA. That project, which was completed prior to the onset of the research presented in this thesis, provides the "initial conditions" from which the present work built upon.

Model development in the NGA-East project had two components: (1) GMMs that predict lognormal *IM* distributions conditional on magnitude and distance, derived for a reference condition with a shear wave velocity (V_S) = 3000 m/s; and (2) site amplification models that modify *IM*s for softer site conditions that are conditioned on time-averaged shear wave velocity in the upper 30 m of the site (V_{S30}). I investigate whether these models, when used in tandem, are compatible with an expanded inventory of ground motion recordings in CENA (i.e., the database was expanded relative to what had been developed during the NGA-East project). This effort was divided into six major tasks: (1) data processing; (2) consistent metadata compilation and organization of the data into a relational database; (3) residuals analyses; (4) analysis of mean misfits of the combined model (hard-rock GMM and site amplification model) relative to the data; (5) identification of potential path-model misfits for data in different geologic domains within CENA, with model adjustments as needed to remove those misfits; and (6) investigation of dependencies of residuals on site parameters, including V_{S30} and sediment depth in coastal plain regions using depth models published by the U.S. Geological Survey (USGS) in 2024. The site response work in Task 6 is intended to validate the original model and further investigate depth-dependence of site response, which had not been previously investigated.

To facilitate efficient data processing while maintaining protocols for human inspection of waveforms to ensure reliability, I worked with collaborators to expand the capabilities of the USGS's automated processing code, gmprocess, to include displacement drift checks in the selection of high-pass corner frequencies. These procedures check for, and as needed remove, low-frequency artifacts in the displacement record using polynomial fits, which can be used in combination with existing signal-to-noise ratio (SNR)-based corner frequency selection procedures. The automated selections are then efficiently verified and refined using a graphical user interface (GUI) that plots relevant ground-motion time series and spectra and tracks modifications to signal processing parameters. Ground motion recordings from a total of 186 events from November 2011 (end of NGA-East data curation) to April 2022 were processed using this method and then archived in a relational database along with metadata collected following the recommendations of the NGA project for subduction zones (NGA-Sub; Contreras et al., 2022). The expanded database includes 2096 sites and 16,272 three-component recordings from events having a magnitude range of ~4 to 5.8.

I computed residuals using 17 NGA-East GMMs and three data selection criteria that reflect within-CENA regional variations in certain ground motion attributes. Partitioning of residuals using mixed-effects analyses reveals a persistent pattern of misfits in which ground motions are overpredicted at short periods (0.01 to 0.6 sec, including PGA) and underpredicted at longer periods. These misfits are regionally variable, with the Texas-Oklahoma-Kansas (TOK) region having larger absolute misfits than other parts of CENA. Two factors potentially influencing these misfits are: (1) differences in the site amplification models used to adjust the data to the reference condition during GMM development relative to those developed in the later stages of NGA-East, and (2) potential bias in simulation-based factors used to adjust ground motions from the hard-rock reference condition to a V_{S30} = 760 m/s condition. Bias adjustment factors and their epistemic uncertainties are provided, which were considered in the development of the 2023 National Seismic Hazard Model (NSHM; Petersen et al., 2023).

I developed an equation-based version of the NGA-East tabular GMMs to provide physically meaningful parameters that could be adjusted, as needed, to remove scaling-trends in regions in CENA. After identifying five CENA regions (two coastal planes, two suture zones, and other

areas), I found no appreciable errors in the NGA-East path model for any of the regions other than the Gulf Coastal Plain (GCP). In the GCP, faster attenuation was encountered, which is similar to prior work, and period-dependent adjustments were provided to remove the trends. Using the adjusted path models, residuals were recomputed and used to investigate site response. I find the trends in the data with V_{S30} to be essentially unchanged relative to current models, whereas significant effects of depth in coastal plain sediments were encountered for GCP and the Atlantic Coastal Plain (ACP). The depth model is formulated using differential depths derived using a V_{S30}-conditioned depth model. Models for amplification conditioned on differential depth are different for GCP and ACP. ACP shows that shallower than average sites have higher short-period amplification and essentially average long-period amplification, whereas deeper-than-average ACP sites have reduced short-period amplification and increased long-period amplifications. For GCP, we see no bias of the V_{S30}-scaling model at short periods for shallow or average depth sites, but reduced amplifications are found for deep sites at short to intermediate periods.

The dissertation of María Elisa Ramos-Sepúlveda is approved.

Scott Joseph Brandenberg

Yousef Bozorgnia

Grace Alexandra Parker

Chukwuebuka C. Nweke

Jonathan Paul Stewart, Committee Chair

University of California, Los Angeles

2024

Pulgoso, esto es para ti y por ti.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Support for this research project was provided by a Eugene V. Cota-Robles Award provided by the UCLA Graduate Division, the USGS External Research Program under contract number G21AP10370, and gift funds provided by my advisor. This support is appreciated.

I would like to sincerely thank my committee for the time and effort to advise and mentor me during my graduate studies at UCLA. I would like to thank my colleagues as well, Grace Parker who helped me with coding and understanding gmprocess, Pengfei Wang who taught me how to process ground-motion data, and Tristan Buckreis for all the guidance regarding computing and analyzing residuals.

I thank Scott Brandenberg, Tristan Buckreis and Victor Contreras for their assistance in assembling the data into a database. I thank Scott Brandenberg and Eric Thompson for their collaboration in adapting and improving gmprocess for application to the data processing conducted in this study. I appreciate the helpful feedback on this work from many individuals, including Martin Chapman, Chris Cramer, Christine Goulet, and Oliver Boyd as well as current or former GWG student members Okan Ilhan, Meibai Li, and Dylan Centella Raef.

I thank the reviewers of the paper that is covered in Chapter 3 (Sanaz Rezaeian, Behzad Hassani, Norman Abrahamson, Annemarie Baltay and one anonymous reviewer), whose comments helped us improve the technical content of the paper and the clarity of presentation. Finally, we thank the members of a Tiger team appointed by USGS (Appendix F of Petersen et al. 2023) to review this model; that process resulted in modification of the weight factors applied to different datasets and

the introduction of the stiff site modification (Eq. 3.6). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Vita

2024	Hired, Senior Staff Engineer *AECOM*
2022,2024	Teaching Assistant *University of California, Los Angeles*
2020-2024	Graduate Research Assistant *University of California, Los Angeles*
2020	M.S. Civil Engineering *North Carolina State University*
2018-2019	Teacher Assistant *North Carolina State University*
2018	B.S. Civil Engineering *University of Puerto Rico, Mayagüez*

Publications

ME Ramos-Sepúlveda and A Cabas (2021) Site Effects on Ground Motion Directionality: Lessons from Case Studies in Japan. Soil Dynamics and Earthquake Engineering. Vol 147, 2021, 106755 doi: 10.1016/j.soildyn.2021.106755

ME Ramos-Sepulveda, GA Parker, M Li, O Ilhan, YMA Hashash, EM Rathje, JP Stewart (2022). Performance of NGA-East GMMs and site amplification models relative to CENA ground motions, *12th National Conference on Earthquake Engineering*, EERI, Salt Lake City, UT, Jun 27 - Jul 1, 2022 (electronic file)

ME Ramos-Sepulveda, TE Buckreis, M Li, GA Parker, SJ Brandenberg, EM Rathje, JP Stewart (2023). 2023 NGA-compatible CENA database for GMM validation and site response studies, DesignSafe-CI. https://doi.org/10.17603/ds2-ywqs-sp29

ME Ramos-Sepulveda, GA Parker, EM Thompson, SJ Brandenberg, M Li, O Ilhan, YMA Hashash, EM Rathje and JP Stewart (2023). High-pass corner frequency selection for implementation in the USGS automated ground-motion processing tool, Geo-Congress 2023: Geotechnical Data Analysis and Computation, Los Angeles, CA, March 2023, *Geotechnical Special Publication No. 342*, EM Rathje, B Montoya, and MH Wayne (Eds.), 327-335, ASCE Geo-Institute

ME Ramos-Sepúlveda, JP Stewart, GA Parker, MP Moschetti, EM Thompson, SJ Brandenberg, EM Rathje (2024) Data-Driven Adjustments for Combined Use of NGA-East Hard-Rock Ground Motion and Site Amplification Models. *Earthquake Spectra* doi: 10.1177/87552930241231825

1. Introduction

1.1 Background and Motivation

Two fundamental effects of earthquakes that produce hazards to civil infrastructure are dynamic ground shaking (oscillatory ground motion in which the position of a point moves during shaking but returns to its initial position at the end) and ground failure (permanent displacements). Ground shaking, which is the focus of the present work, is caused by waves that travel from the seismic source through the crust and that are modified near the site by the near-surface geological conditions. For engineering purposes, the main features of earthquake ground motions are typically described using intensity measures (*IM*s), different versions of which describe the amplitude, frequency content, and duration of ground shaking. Most engineering applications use the *IM*s of peak acceleration (PGA), peak velocity (PGV), and 5% damped spectral accelerations (Sa).

*IM*s are commonly estimated for engineering applications using ground motion models (GMMs) that predict median and lognormal standard deviation of *IM*s conditional on a series of source, path, and site parameters. Different GMMs are used in different tectonic regimes, and in the case of central and eastern North America (CENA), GMMs for stable continental regions are required. Section 1.3 presents a brief literature review of ground motion studies for CENA. The most important of those prior studies, which provided the baseline database and models that the present work built upon, was a major multi-agency, multi-investigator project (NGA-East) that developed GMMs for CENA (Goulet et al., 2021a; Youngs et al., 2021) that are used in both the 2018 and

2023 versions of the U.S. Geological Survey (USGS) National Seismic Hazard Model (NSHM) (Petersen et al., 2020, 2023). As such, the NGA-East models have had substantial impact.

The GMMs developed by the NGA-East project (Goulet et al., 2021a; Youngs et al., 2021) are based on the project database (Goulet et al., 2021b), which contains data from CENA events through November 2011, and a series of numerical simulations. Those GMMs apply for a hard-rock reference condition corresponding to a reference shear wave velocity (V_S) of 3000 m/s and site diminutive parameter (κ_0) of 0.006 sec (Hashash et al. 2014). A largely independent effort conducted within the NGA-East project was the development of site adjustment factors, which was done through ground motion data analyses (Parker et al., 2019) and ground response simulations (Harmon et al., 2019a,b). Subsequently, an expert panel reviewed those results and those of other studies from literature up to that time (Section 1.3) and recommended ergodic site amplification models for CENA that are conditioned on V_{S30} (Stewart et al., 2020; Hashash et al., 2020) and that were implemented in the 2018 and 2023 NSHMs.

The motivation for the present work drew from several considerations. First, an appreciable number of earthquakes have occurred since the “close date” for the NGA-East database in November 2011, and importantly, those earthquakes are recorded by more instruments, which are of higher quality, than what had been available previously. This additional data was considered to have the potential to expand the frequency and distance range for which ground motions could be investigated in CENA. Second, because the NGA-East hard rock and site amplification model development efforts were largely independent of each other (details in Chapter 3), there was an interest in examining whether the combined application of these models provided ground motion estimates that were consistent with the available data. Third, in consideration of the improved size and quality of the database developed in this work, as well as the publication in 2023 of maps showing the depth of coastal plain sediments (Boyd et al., 2024), we were interested to evaluate regional variations of path effects and potential depth-dependencies of site effects in CENA, which has not been possible previously.

1.2 Research Objectives and Scope

Based on the project motivation presented in Section 1.1, the first objective of the present work was to develop an updated version of the NGA-East database that includes newer events, many of which have large numbers of recordings, and improved site metadata. A second objective of the work was to investigate the potential for misfit between the predictions of NGA-East hard rock models combined with site amplification models relative to the expanded dataset. The third objective was to investigate regional variations of path and site effects in CENA, as well as dependencies of site amplification on depth in different regions.

In consideration of these objectives, the research scope included the following principle tasks, which are presented in Chapters 2 to 5 of this dissertation:

1. To facilitate efficient ground motion data processing while maintaining NGA data processing protocols, modify the USGS' open-source software, gmprocess (Hearne et al., 2019) to include routines to remove baseline drift, and provide tools to facilitate user review and adjustment of processing parameters. This work is presented in Chapter 2 of this thesis, a journal paper under review (Ramos-Sepulveda et al. 202x) and is reflected in a recent gmprocess update (Thompson et al., 2024a).

2. Expand the existing NGA-East ground-motion database (Goulet et al., 2021b) by processing ground motion data from 100 additional events. As a result of this effort, the database grew from 86 to 186 events and from 9376 to 16268 recordings.
3. Convert the database from excel files to a relational database format that is available online (DOI: 10.1061/9780784484692.047). This work involved extensive collaboration with others and is presented in Buckreis et al. (2023) and a journal paper under review (Buckreis et al., 202x).
4. Using the expanded dataset - relative to the NGA-East database - perform residuals analyses to compare the empirical data with *IM* predictions from the NGA-East GMMs and the CENA-specific site amplification model. This work is presented in a published journal paper (Ramos-Sepulveda et al., 2024) and Chapter 3 of this thesis.
5. To facilitate investigations of regional path and site effects using the expanded database, develop a version of the NGA-East GMMs, which are presented in the form of tables, that consists of an equation with source and path components with coefficients (such as anelastic attenuation) that have physical meaning. This exercise provides a stronger basis for evaluation of regional variations within the parameter space (i.e., magnitude and distance range) of the data, which can be quantified by regional adjustments to those coefficients. This work is presented in Chapter 4 of this thesis.
6. Using the equation-based GMM from Chapter 4, compute residuals and partition residuals into between-event components (event terms), within-event components, and site terms. Evaluate regional variations in path effects by investigating the distance dependence of within-event residuals. Develop models, as needed, to describe such trends. Using the updated models,

recompute residuals and investigate site effects with those results, including depth effects in different regions. This work is presented in Chapter 5 of this thesis.

This study is limited to the ground-motion data collected and processed for 187 earthquakes that originated in the CENA region with magnitudes from 4 to 6.7 and rupture distances up to 600km. This number of events and data range reflect the data selected for analysis, not the overall database size. The 187 selected events include some from the NGA-East database (35 earthquakes of tectonic origin) and 152 events that occurred after November 2011 (date of the latest event in NGA-East) until April 2022, most of which are potentially induced. Chapter 4 includes the **M**4.8 2024 Tewksbury, New Jersey Earthquake, which was not considered in the analyses presented in Chapter 3. The data is unevenly sampled, being particularly voluminous for **M** 4 to 5 events located in Texas, Oklahoma or Kansas, most of which are potentially induced (137 of the 152 that were added). This uneven data sampling was considered in the analyses presented in Chapters 3 and 5.

1.3 Literature Review on CENA ground-motion models

The NSHMs have been the basis for many public and private guidelines, including seismic-design regulations for buildings, bridges, highways, railroads, and other structures (Petersen et al., 2014). Hence, by examining the suite of GMMs implemented by previous NSHMs, one can track the evolution of ground motion modeling research for the CENA region.

Prior to the NGA-East project, the GMMs implemented in NSHMs since the mid-1990s were Frankel et al. (1996), Toro et al. (2014), Silva et al. (2002), Campbell (2003), Tavakoli and Pezeshk (2005), Atkinson and Boore (2006), Pezeshk et al. (2011), Atkinson (2008) and Somerville et al. (2001). These models are all based on numerical simulation of ground shaking, generally using semi-stochastic point source model with parameters such as stress parameter, quality factor, and site diminutive parameter (κ_0) estimated from limited available data that was assembled independently by different investigator teams.

These models were assigned different weights in the logic-trees based on model type, type of simulated databases, κ_0 models, and geometric spreading coefficients (Petersen et al., 2014). They all had different magnitude scaling, distance scaling and, if applicable, site response models based on the region and scope of the dataset. However, they all had to be applicable or reasonably extrapolated to magnitudes between 4.7 to 8.2 and up to distances of 1000 km. Numerous other models have been developed for CENA, but either they are predecessors (e.g Atkinson 2004) to the models incorporated by the NSHM, or they were incomplete models (e.g. Atkinson and Mereu, 1992, Boatwright and Seekins, 2011).

Currently, the 17 NGA-East GMMs are implemented in the NSHM logic-tree with high weights along with 14 seed models. Out of the 14 seed models, three of them (Graizer, 2016, 2017; and Shahjouei and Pezeshk, 2016) have been further improved since their publication as part of the NGA-East project in 2015. Recently, Pezeshk et al. (2024) presented a functional form for the NGA-East central branch GMM that is based on a point-source omega-squared model formulation (Aki, 1967; Boore, 1983, 2003). The general shape contains a geometric spreading, anelastic attenuation, stress drop, and a site attenuation term. Li et al. (under review) developed correction factors for source, distance and overall mean of the NGA-East central branch GMM for the Texas-Oklahoma-Kansas region. Alidadi et al. (2024) used machine learning techniques to generate a GMM for small-to-moderate potentially induced earthquakes.

In the 2018 and 2023 NSHMs, site response is derived using the amplification models presented by Stewart et al. (2020) and Hashash et al. (2020), with the main basis for these models being work that was carried out by the NGA-East Geotechnical Working Group (GWG). Prior NSHMs took site amplifications from Frankel et al. (1996) and Atkinson and Boore (2011) who developed factors using different κ_0 values depending on the site class (Petersen et al., 2014). Other earlier models based on simulated data include Hwang et al. (1997), Khaheshi Banab et al. (2012), Darragh et al. (2015), Atkinson et al. (2015), Hassani and Atkinson (2016), Harmon et al. (2019), and Aboye et al. (2015). Due to the lack of V_{S30} data at seismographic sites, there was a limited number of models based on empirical data (e.g., Hassani and Atkinson, 2017; and Yenier et al., 2017). These models were developed for different regions using various datasets and methods. These models were evaluated and compared in Stewart et al. (2020).

Chapman and Guo (2021) developed a model for site amplification in the coastal plain based on simulated data and they found a strong relationship between sediment depth and site amplification. Schleicher and Pratt (2021) used horizontal-to-vertical and surface-to-bedrock spectral ratios to measure site response in the Atlantic Coastal Plain (ACP). They found that linear site-response amplifications can be estimated for weak-motion in the ACP. Also using surface-to-bedrock spectral ratios from the ACP, Pratt and Schleicher (2021) developed equations to characterize site-response amplification as a function of sediment depth. With the recent sediment thickness model for the coastal plain constructed by Boyd et al. (2024), Boyd et al. (2024) and Akhani et al. (2024) created new depth adjustments to site amplification for sites in the coastal plains.

2. Processing Ground Motion Data

A version of this chapter is under review by the Seismological Research Letters. The manuscript title is "High-Pass Corner Frequency Selection and Review Tool for Use in Ground-Motion Processing" with the following authors María E. Ramos-Sepúlveda, Scott J. Brandenberg, Tristan E. Buckreis, Grace A. Parker and Jonathan P. Stewart.

2.1 Abstract

Raw seismological waveform data contains noise from the instrument's surroundings and the instrument itself that can dominate recordings at low and high frequencies. To use these data in ground-motion modeling, the effects of noise on the signals must be reduced and the signals' usable frequency range identified. We present automated procedures to efficiently reduce low-frequency noise that are implemented in the U.S. Geological Survey software package gmprocess. These procedures check for, and as needed remove, low-frequency artifacts in the displacement record using polynomial fits, which can be used in combination with existing signal-to-noise ratio (SNR)-based corner frequency selection procedures. The automated selections are then efficiently verified and refined using a graphical user interface (GUI) that plots relevant ground-motion time series and spectra and tracks modifications to signal processing parameters. We demonstrate these procedures using recordings from the 2020 **M**5.1 Sparta, North Carolina and the 2013 **M**4.7 Southern Ontario earthquakes. Data processed with the SNR-only and polynomial criteria for these events contain displacement artifacts in 37% and 23% of processed traces, respectively. Records with remaining artifacts are corrected manually using the GUI. These processing steps illustrate the workflow for efficient data processing with quality control.

2.2. Introduction

Seismometers and accelerometers record ground shaking before and during earthquakes. These recordings include the earthquake signal, non-seismic ground motions (ambient vibrations), and instrument noise. For ground-motion modeling purposes, processing is traditionally performed to remove instrument response effects and to reduce the effects of noise through the application of high- and low-pass filters. The selection of filter corner frequencies is a critical aspect of these analyses that was commonly performed manually in past work. These procedures, and considerations in corner frequency selection, are described by Boore (2005) and Boore and Bommer (2005).

Such manual procedures are increasingly falling out of favor due to the numbers of recordings produced by the large numbers of sensors in modern seismic networks. To facilitate rapid processing, the U.S. Geological Survey (USGS) developed open-source software, gmprocess (Hearne et al., 2019; Thompson et al., 2024a,b), that identifies corner frequencies based on ratios of Fourier amplitudes for the overall signal to the pre-event ambient vibrations (i.e., signal-to-noise ratio, SNR). These SNR-based criteria lack some of the checks in manual procedures, which creates the possibility of over- or under-filtering. The aims of the present work are to (1) augment the automated procedures to include checks for potential displacement artifacts that are functionally similar to those used in previous manual procedures and (2) allow for human review and adjustment of processing parameters for applications where quality control is prioritized.

Following this introduction, we describe and compare the current (SNR-only) and proposed (polynomial fit) methods in gmprocess v1.3.2 for selecting high-pass corner frequencies (f_{cHP}). We then present a tool for visual inspection of the processed records and, if needed, adjustment of filter corner frequencies to improve quality. An earlier conference paper (Ramos-Sepulveda et al., 2023) presented preliminary findings of related work, mainly comparing results of manual vs. automated ground-motion processing. The work presented here is expanded in scope from the earlier paper and entirely updated.

2.3 High-Pass Corner Frequency Selection

Two f_{cHP} selection algorithms are considered: (i) SNR-only (default in gmprocess), and (ii) SNR in combination with a displacement polynomial fit method (abbreviated as "polynomial" herein) that is intended to remove displacement artifacts.

The SNR-only method sets f_{cHP} to the Brune (1970) corner frequency (configurable) or to a smaller value if the SNR exceeds a threshold at the corner frequency, usually set to 2 (Thompson et al., 2024a,b). Figure 2.1 shows an example ground-motion where SNR criteria produce f_{cHP} = 0.014 Hz. Panel (a) depicts the location of f_{cHP} with respect to the normalized Fourier amplitude spectrum (FAS) of the unfiltered record (signal), pre-event signal (noise), and SNR. Panel (b) shows the displacement time series after filtering, which contains artifacts (i.e., drift before and following P-wave arrival, long-period drift beyond 70 s). Such results are not unusual, indicating that further adjustments are needed for applications requiring reliable low-frequency components of the ground motion.

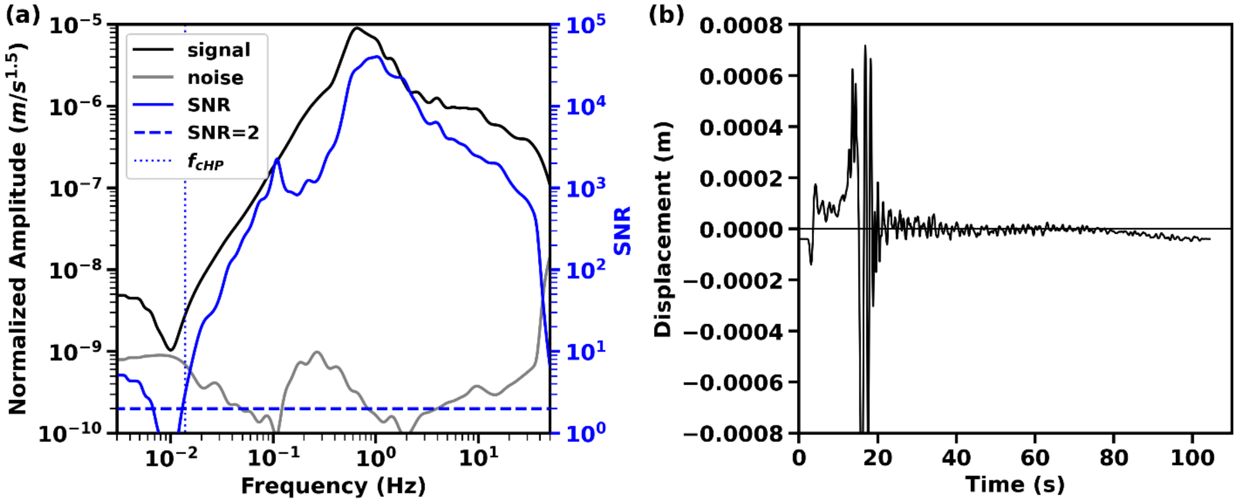


Figure 2.1. Example record (site ET GFM HHZ) in which SNR-only criteria were applied for selecting f_{cHP}: (a) Fourier Amplitude Spectra (FAS) for full signal and pre-event noise, along with their ratio (SNR). The selected f_{cHP} is based on a limiting value of SNR=2; and (b) the displacement time series after high-pass filtering at f_{cHP}, which demonstrates low-frequency artifacts.

The polynomial method initially selects a trial f_{cHP}, filters the record, computes the displacement time series, fits a polynomial (with configurable order; 6th order is used herein) to the filtered displacement record, computes the maximum amplitude of the polynomial fit within the time interval of the recording, and iterates on f_{cHP} until the difference between the *Ratio* (Eq. 2.1) and the target is within a specified tolerance.

$$Ratio = \frac{max(\,|polynomial\ fit|\,)}{max(\,|filtered\ displacement|\,)} \tag{2.1}$$

Iterations are performed using Ridders' (1979) method. Figure 2.2 illustrates the procedure using a target of 0.01 and tolerance of 0.001. These values were selected based on experience after applying this approach to recordings from multiple earthquakes with magnitudes ranging from 3.8 to 5.8 and epicentral distances from 10 to 700 km. In Figure 2.2a, the trial f_{cHP} is too small resulting in an under-filtered motion as depicted by the amplitude ratio being larger than 0.01 and a displacement with visible wobble (e.g., periodic long-period displacement features); in panel (b), the amplitude ratio is 0.01 and the record appears to be properly filtered; and in panel (c) the trial f_{cHP} is too high leading to an amplitude ratio that is smaller than 0.01 and a reduction of the displacement amplitude (compared to panel b) indicating over-filtering. If these criteria are not met for $f_{cHP} < 0.5$ Hz within 30 iterations, the record is rejected. This iterative method is coded by Brandenberg and Yang (2022).

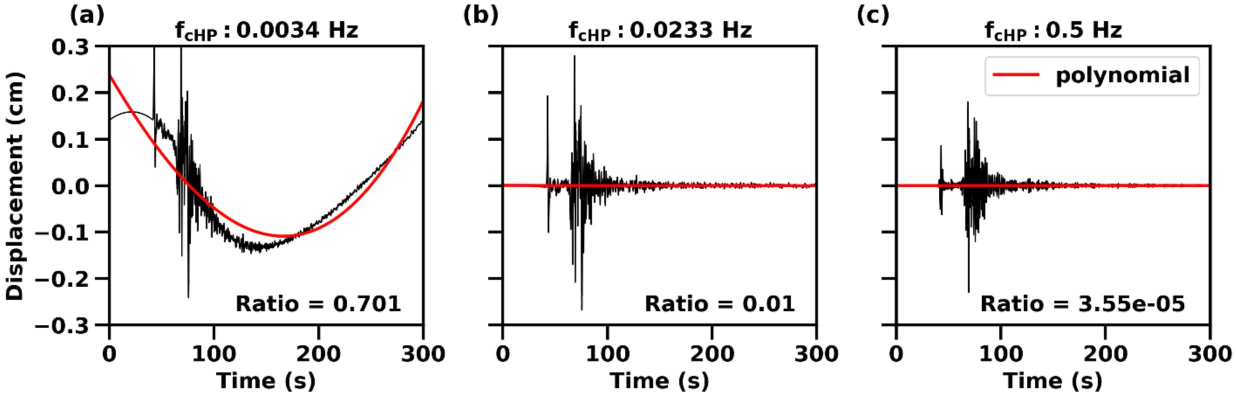


Figure 2.2. Displacement record processed with different f_{cHP} (increasing from left to right) along with the fitted polynomial and the corresponding Ratio for (a) under-filtered record, (b) properly processed record, and (c) over-filtered record.

One of the traditional considerations in f_{cHP} selection pertains to the FAS slope at low frequencies. When this slope flattens, which departs from classical expectation for earthquake signals (Brune, 1970), past work often considered the flat portion to be a manifestation of noise, thus requiring application of a filter to remove it (Boore and Bommer 2005). Such checks have been used in recent projects (e.g., Kishida et al. 2020). Where flat spectra at low frequencies are associated with low SNR or displacement artifacts, the procedures described here will remove the flat portion. However, we also encountered cases with modern sensors where data have flat FAS at low frequencies and meet SNR and displacement criteria; in such cases we recommend against applying filters to remove this portion of the spectrum. An example of this is shown in Figure 2.3

where the f_{cHP} for SNR is 0.008147Hz and the displacement trace shows no evidence of long-period noise.

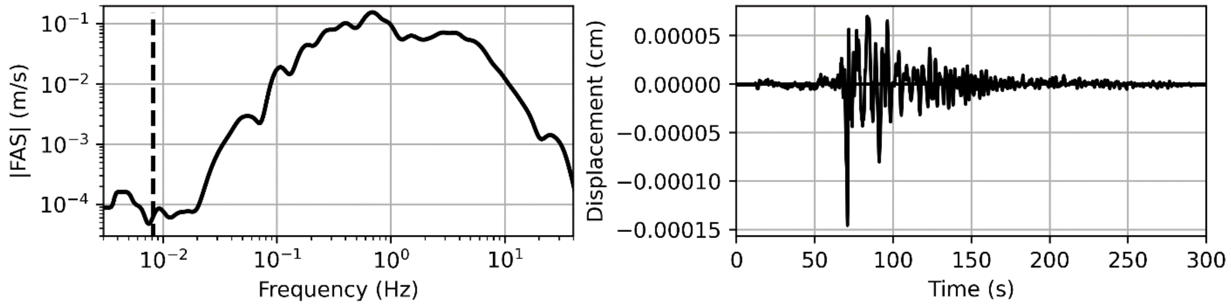


Figure 2.3. Example record (site PE PAKS HHN) in which f_{cHP} lies in the flat region of the FAS at short frequencies and no evidence of low-frequency artifacts is observed in the displacement trace.

Implementation of the polynomial criteria requires double integration of acceleration to compute displacement, which can be performed using time- or frequency-domain approaches. Time-domain integration assumes zero initial velocity and displacement. However, these initial conditions are incorrect due to the presence of ambient vibrations, and forcing the initial conditions can cause low-frequency drift in time-domain-integrated displacements that can either be removed by a baseline correction post-processing step, or by increasing f_{cHP} until the displacement trace flattens. This drift is not present when using frequency-domain integration, which forces the average velocity and displacement to be zero. Furthermore, it is not present when time-domain integration forces the average velocity and displacement to be zero instead of the initial conditions. We recommend utilizing either frequency-domain integration, or time-domain integration with average displacement equal to zero when selecting an appropriate value of f_{cHP}.

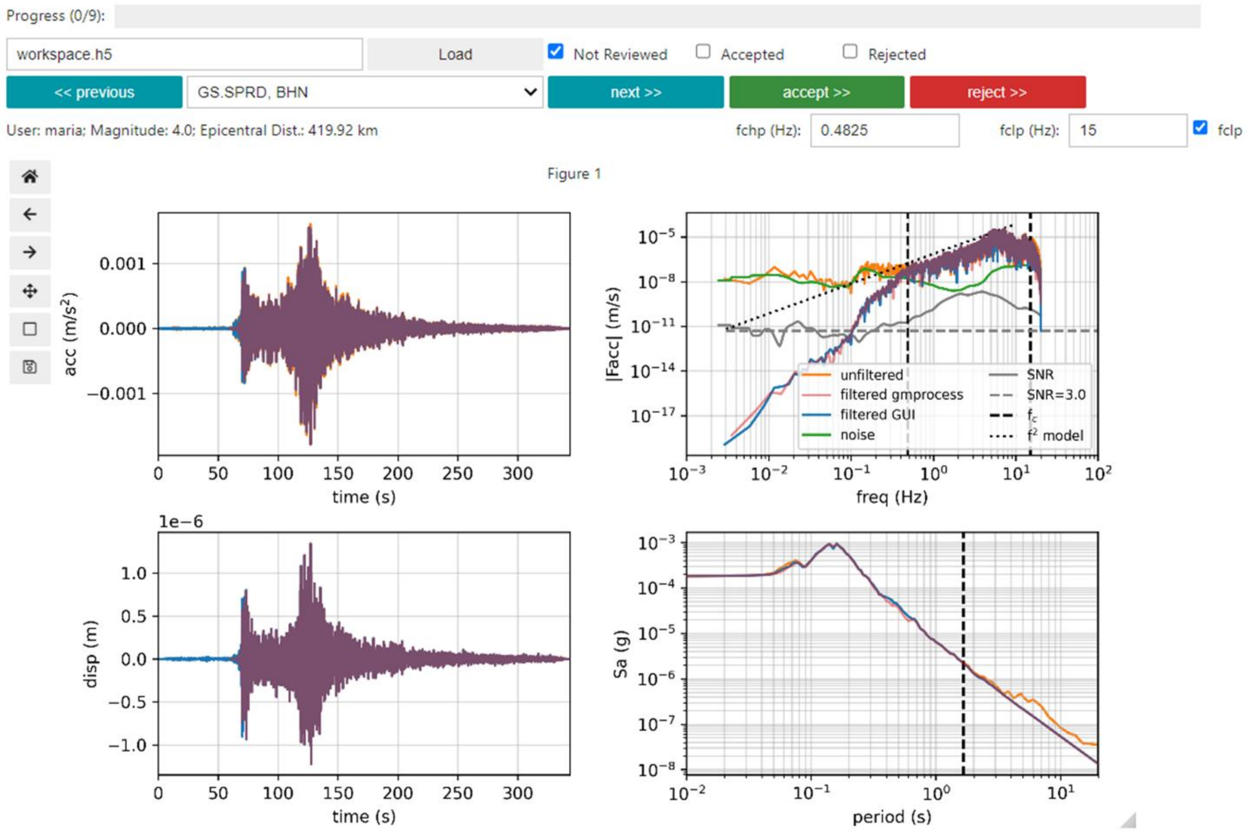
A 5th Butterworth acausal filter is utilized to reduce low-frequency components. There are two options for applying the filter in gmprocess. The first approach is to utilize the ObsPy (Krischer et al., 2015) filter implementation, which uses the SciPy (Virtanen et al., 2020) finite impulse response filter design. The result is a causal filter that approximately matches the specified target filter, and is applied in the time domain to the recorded signal. The causal filter induces phase lag that is removed by applying the filter twice; once on the as-recorded record, and again on a tail-reversed version of the filtered record. We found that this type of filter often does not reproduce the target Butterworth filter, particularly at low frequencies where we observed under-filtering. The second approach is to directly multiply the Fourier coefficients by the Butterworth filter ordinates. This filter is applied in the frequency domain, and does not have a time domain analog (i.e., it cannot be created from physical components). Its benefits are (1) it is acausal since the real and imaginary components of the Fourier coefficients are multiplied by the same constants, thereby preserving phase, and (2) it perfectly matches the target Butterworth filter. A downside is that the filter is applied in the frequency domain, and a Fourier transform must be applied to the signal prior to filtering. It therefore cannot be applied in real-time, and potentially cannot be applied to very long signals for which the Fourier transform would require too much memory. However, these limitations are not issues for the ground-motion records we have processed. We utilize the frequency-domain Butterworth filter approach in this paper.

2.4 Human Review

For research that produces ground-motion models used in engineering practice, it is important to incorporate human review into the data processing workflow for quality assurance. Improperly filtered records contain long-period noise, which would influence long-period components of the ground-motion model; the aim of the human review is to minimize such occurrences. To provide this capability for processing performed with gmprocess, we created a graphical user interface

(GUI) using a Jupyter Notebook (Kluyyer et al., 2016). The GUI operates on output files created by gmprocess in the Hierarchical Data Format 5 (HDF5; https://github.com/HDFGroup/hdf5). It allows users to inspect filtered time-domain traces, FAS and acceleration response spectra, and if modifications to corner frequencies are deemed necessary, it facilitates iterative modifications of corner frequencies until satisfactory results are obtained.

Figure 2.4 shows the GUI interface, which depicts acceleration and displacement time series on the left and FAS and 5% damped acceleration response spectra (Sa) on the right. These time series and spectra are plotted for three motions – the unfiltered record, filtered record as provided by gmprocess, and filtered record using modified corner frequencies from the GUI. Information such as user, magnitude, epicentral distance, f_{cHP}, and low-pass corner frequency (f_{cLP}) are shown in the header.



Progress (0/9):

workspace.h5

☑ Not Reviewed ☐ Accepted ☐ Rejected

GS.SPRD, BHN

User: maria; Magnitude: 4.0; Epicentral Dist.: 419.92 km

fchp (Hz): 0.4825

fclp (Hz): 15 ☑ fclp

Figure 2.4. Image showing the GUI used for human inspection of processed records. |Facc| indicates Fourier amplitudes and Sa indicates 5% damped spectral accelerations. The vertical dashed line in the Fourier amplitude plot shows f_{cHP}, while the same line in the Sa plot shows the maximum usable period. Note that when red traces overlie blue (i.e., "filtered GUI" overlaps "filtered gmprocess") they appear as purple.

Users can modify the f_{cHP} value by inspecting the displacement time series and FAS. The values of f_{cHP} can either be increased or decreased in the GUI. Furthermore, users can elect to apply and/or adjust the low-pass filter parameters as well, which is typically assessed by short-period anomalies in the acceleration response spectrum (shown in Figure 2.5d) or high-frequency anomalies in the FAS that are judged to be unrealistic (e.g., Boore and Goulet 2014). Depending on how the user configures gmprocess, low-pass filters may have been applied to every record or none. The default setting is to apply low-pass filters in the GUI.

The unfiltered record (plotted in orange) is obtained by trimming, demeaning, and tapering the record labeled as “unprocessed” in the HDF5. The "filtered GUI" record (blue) is a filtered version using the “apply_filter” function from gmprocess and changes if corner frequencies are modified in the GUI. The filtered GUI records include the pre-event noise and signal portions of the time series. The "filtered gmprocess" record (translucent red) is the record filtered by gmprocess; it does not change when corner frequencies are modified in the GUI. The purpose of including both filtered traces (blue and red) is so the user can visualize changes as corner frequencies are adjusted. The red trace is partially transparent to aid visualization. When red traces overlie blue (i.e., "filtered GUI" overlaps "filtered gmprocess") they appear as purple traces. The smoothed noise spectrum (FAS) is shown in green to provide a visual sense for SNR and the SNR is shown in grey. The SNR ordinates are normalized therefore the grey-dashed line is used as reference to identify if the SNR is below or above the acceptable threshold. The black dotted line is the theoretical acceleration decay at low frequency according to the f^2 model (Brune 1970; Boore and Bommer 2005), which has often been used as a reference slope against which to compare the data. The dashed-black lines depict corner frequencies in the FAS and the maximum usable period of response spectra, where maximum usable period is taken as $0.8/f_{cHP}$ as in standard practice (e.g., Ancheta et al., 2014).

2.5 Example Application

To illustrate the effects of different processing methods, we consider records from two well documented (i.e., more than 200 three-component raw-records) earthquakes with small to moderate magnitude: the 2020 **M**5.1 Sparta, North Carolina (244 three-component records) and the 2013 **M**4.7 Southern Ontario, Canada (251 three-component records). All records were obtained from the Incorporated Research Institutions for Seismology (IRIS) data center (see "Data Resources" section). After applying the gmprocess screening criteria (Thompson et al., 2024a), 1144 individual components remained and were processed. Following visual inspection in the GUI, 204 were rejected for ground-motion modeling applications. The reasons for rejection varied but were generally related to challenges associated with high noise levels such as inability to clearly identify the P-wave arrival time (Figure 2.5a), which in turn produces large uncertainty in identification of the pre-event noise window, or unrealistic behavior such as multiple peaks outside of the P-wave or S-wave windows (Figure 2.5b), which is suggestive of strong noise within the seismic signal.

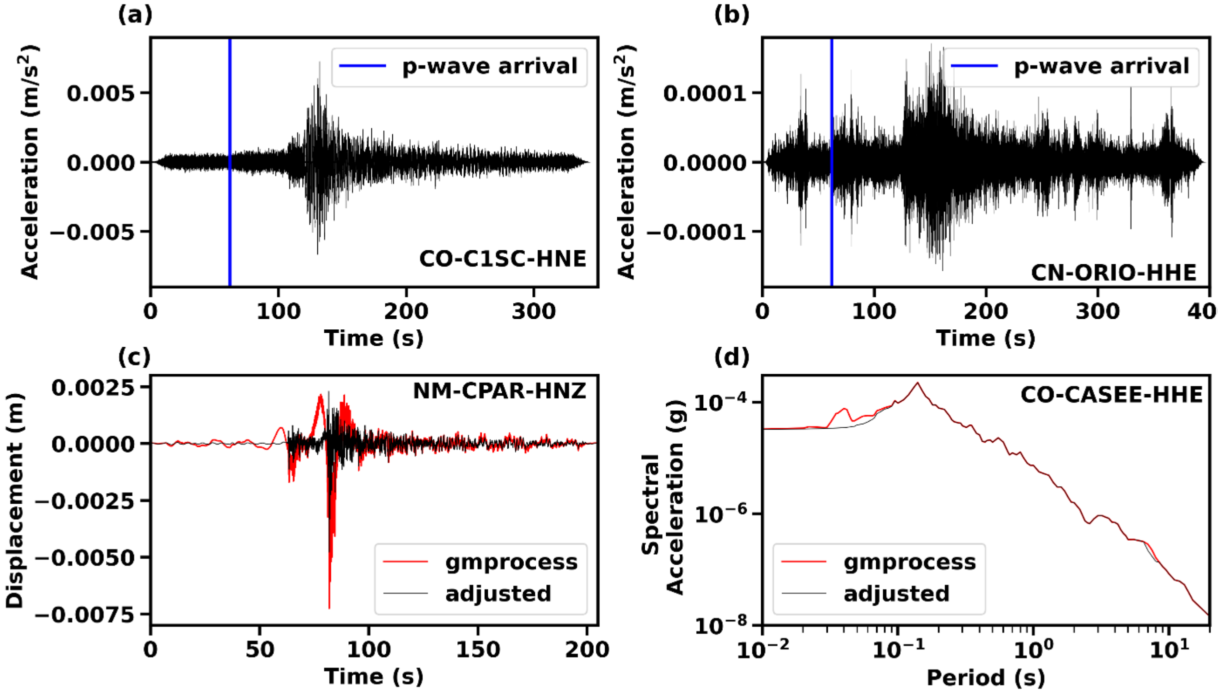


Figure 2.5. Example records subjected to automated processing in gmprocess and decisions made regarding record acceptance or filter adjustments: (a) rejected record due to indiscernible P-wave arrivals, (b) rejected record due to large noise amplitudes, (c) accepted displacement time series for which f_{cHP} was increased in the GUI to remove displacement artifacts, and (d) accepted

response spectra for which a low-pass filter was added at 14.6 Hz to remove high frequency artifacts evident in FAS and Sa plots.

Gmprocess was run on the remaining 940 raw records using automated procedures in a version of gmprocess that was updated to include the polynomial method described previously in this paper. The traces were evaluated in the GUI to adjust corner frequencies as needed to improve signal quality. Of the 940 records considered, 504 required no adjustments to corner frequencies (54%). The subset of records for which adjustments were made demonstrate features such as wobbly displacements both before and after P-wave arrival (Figure 2.5c), which was addressed by increasing f_{cHP} to produce an updated displacement time series; high-amplitude short-period noise indicated by an irregular peak in the response spectrum at short periods (Figure 2.5d); and irregular peaks in the FAS at high frequencies (not shown), which was addressed by low-pass filtering, in this case at f_{cLP} = 14.6 Hz.

The data from the two example events are used to illustrate the impact of the two principal contributions of this work on f_{cHP} selection: (1) the influence of the polynomial method relative to the SNR method, both consisting of automated selections in gmprocess (Figure 2.6a) and (2) influence of user adjustments in the GUI (Figure 2.6b) relative to the automated polynomial selection. Colors on the figure represents the binned difference between corner frequencies, which is defined as $\Delta ln f_{cHP} = ln\left(\frac{f_{cHP}^{abs}}{f_{cHP}^{ord}}\right)$ in Figure 2.6, where *abs* indicates abscissa and *ord* indicates ordinate. Natural logarithms are used to emphasize percent changes. Values along the diagonal indicate matching values of f_{cHP} whereas values below the diagonal indicate that f_{cHP} increases from the baseline (more aggressive filtering).

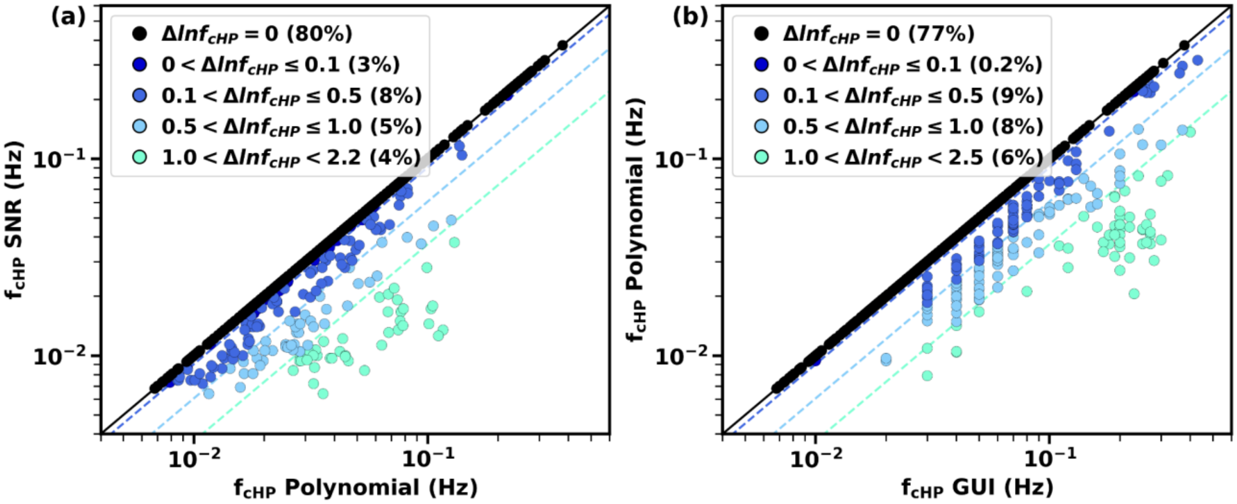


Figure 2.6. Changes in f_{cHP} for 940 records from two example events (a) use of polynomial vs SNR-only automated methods in gmprocess and (b) human inspection and adjustment (as needed) relative to polynomial automated criteria. The colors reflect the binned differences of Δlnf_{cHP}; e.g., dark colors on the diagonal indicate zero difference (most commonly encountered as depicted by the percentage shown in the legend), and light colors indicate the greatest distance below the diagonal (less than 4-6%).

The results in Figure 2.6a show that 20% of records have increased f_{cHP} from application of the polynomial method; among the subset for which increases were applied, about half were adjusted by $\Delta lnf_{cHP} < 0.5$ (65%) with the remainder of adjustments being larger. While the polynomial method takes additional computation time, these adjustments are low cost due to automation within gmprocess. Figure 2.6b shows that 23% of records have further adjustments (increased f_{cHP}) based on human inspection, which comes at higher cost due to the time spent by the analyst. In this case, the polynomial method saved analyst time because 37% of the records would have required modification through the GUI instead of 23%. Moreover, it likely that, on average, Δlnf_{cHP} would have been greater. Among the 23% of adjusted records, 6% correspond to records where the f_{cHP} was previously increased by applying the polynomial method (i.e. the polynomial method was not sufficient). Where adjustments were made, about 40% were adjusted by $\Delta lnf_{cHP} < 0.5$.

2.6 Discussion and Conclusion

Processing of earthquake ground motions is applied to reduce the effects of instrument noise and noise from ambient (non-seismic) ground vibrations, which can strongly influence recordings at low and high frequencies. In addition to reducing noise effects, this processing identifies the frequency range where signals are considered reliable for earthquake ground-motion studies. Procedures for record processing and identifying usable frequencies have developed over time to reflect the attributes of records from various instrument types that have been used at different times (e.g., Boore and Bommer 2005). These procedures have generally been applied manually, which has been practical because of the modest numbers of recordings per earthquake.

As the sizes of seismic networks have grown, the number of recordings per event has also grown to the point that automated processing is often considered more viable (e.g., Rekoske et al., 2020; Ji et al., 2023). The USGS developed gmprocess to meet this need, originally using relatively simple SNR-based criteria to guide the selection of filter corner frequencies. This can lead to records with undesirable noise impacts in some instances, which was the motivation for additional processing steps in traditional procedures (Boore and Bommer 2005).

This work seeks to reduce the number of records that require adjustment following automated processing by incorporating a key element of previous manual processing procedures (checks of displacement artifacts) into automated procedures, which is now included as an optional waveform processing step in gmprocess. We also provide a GUI that works with gmprocess (but is a separate tool) that allows for visual inspection and adjustment of corner frequencies for projects that prioritize quality control. Our experience, as illustrated with the two events considered here, and many additional events that have been evaluated in recent studies, is that the recommended procedures positively affect an appreciable percentage of recordings and should be included in ground-motion processing workflows.

2.7 Data Resources

Raw ground-motion recordings were retrieved from International Federation of Digital seismograph Networks (FDSN) data centers of IRIS Data Services using gmprocess (Hearne et al., 2019). IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR1851048. The code for the GUI is maintained on GitHub (https://github.com/meramossepu/HumanReviewGUI) and there is a versioned release available on Zenodo (v1.0, at the time of publication; doi: 10.5281/zenodo.10815137). The relevant documentation can be found on GitHub. This repository will be merged into gmprocess v2.1.0.

3. Data-Driven Adjustments for Combined Use of NGA-East Hard-Rock Ground Motion and Site Amplification Models

A version of this chapter has been published in Earthquake Spectra. The journal paper title is "Data-driven adjustments for combined use of NGA-East hard-rock ground motion and site amplification models" with the following authors María E. Ramos-Sepúlveda, Jonathan P. Stewart, Grace A. Parker, Morgan P. Moschetti, Eric M. Thompson, Scott J. Brandenberg, Youssef M.A. Hashash and Ellen M. Rathje.

3.1 Abstract

Model development in the Next Generation Attenuation-East (NGA-East) project included two components developed concurrently and independently: (1) earthquake ground-motion models (GMM) that predict the median and aleatory variability of various intensity measures conditioned on magnitude and distance, derived for a reference hard-rock site condition with an average shear-wave velocity in the upper 30 meters (V_{S30}) = 3000 m/s; and (2) a site amplification model that modifies intensity measures for softer site conditions. We investigate whether these models, when used in tandem, are compatible with ground-motion recordings in central and eastern North America (CENA) using an expanded version of the NGA-East database that includes new events from November 2011 (end date of NGA-East data curation) to April 2022. Following this expansion, the dataset has 187 events, 2096 sites and 16,272 three-component recordings, although the magnitude range remains limited (~4 to 5.8). We compute residuals using 17 NGA-East GMMs and three data selection criteria that reflect within-CENA regional variations in ground-motion attributes. Mixed-effects regression of the residuals reveals a persistent pattern in which ground

motions are over-predicted at short periods (0.01 to 0.6 sec, including PGA) and under-predicted at longer periods. These misfits are regionally variable, with the Texas-Oklahoma-Kansas region having larger absolute misfits than other parts of CENA. Two factors potentially influencing these misfits are: (1) differences in the site amplification models used to adjust the data to the reference condition during NGA-East GMM development relative to CENA amplification models applied since the 2018 National Seismic Hazard Model (NSHM), and (2) potential misfit in simulation-based factors used to adjust ground motions from the hard-rock reference condition to a V_{S30} = 760 m/s condition. We provide adjustment factors and their epistemic uncertainties and discuss implications for applications.

3.2 Introduction

In the 2018 and 2023 U.S. Geological Survey (USGS) National Seismic Hazard Model (NSHM; Petersen et al., 2020, 2023), ground-motion intensity measures for central and eastern North America (CENA) were evaluated using ground-motion models (GMMs) and site amplification models developed as a part of the Next Generation Attenuation-East (NGA-East) project (Goulet et al., 2021a; Youngs et al., 2021). These GMMs and site amplification models were developed by different teams of investigators and under different organizational frameworks. In the case of GMMs, 17 models and a weighted median (referred to as “central branch” below) were recommended by Goulet et al. (2021a) with the aim of capturing epistemic uncertainties related to the overall ground-motion space, including magnitude scaling, distance scaling, and other attributes. These recommended GMMs do not include individually developed “seed” GMMs by independent modelers (Pacific Earthquake Engineering Research Center, 2015; hereafter PEER 2015), although some of those seed models are considered in the NSHM (Rezaeian et al., 2021;

Moschetti et al., 2024), as those GMMs were argued to represent physical features that were not present in the 17 NGA-East models. The GMMs apply for a hard-rock reference site condition defined as having average shear-wave velocity in the upper 30 meters (V_{S30}) = 3000 m/s and site decay parameter (κ_0) = 0.006 sec (Hashash et al., 2014), which is often used as the reference site condition for applications in which site-specific site response is applied. The model development was conducted as a Senior Seismic Hazard Analysis Committee (SSHAC) Level 3 project (Budnitz et al., 1997; Nuclear Regulatory Commission, 2012), which is a formal process involving extensive review and documentation.

Because of the hard-rock reference site condition, development of the seed GMMs required adjustments to recorded ground motions in the NGA-East database (Goulet et al., 2021b), all of which were from softer-than-reference sites (average shear-wave velocities in the upper 30 meters, V_{S30} ~150 to 2000 m/s). The adjustments occurred relatively early in the project (PEER 2015), with each seed GMM developer team using their preferred site adjustment models (and initial site V_{S30} values from Goulet et al. 2014 that were later updated by Parker et al. 2017). The adjusted ground motions were used in GMM development as a constraint on scaling relations (with distance and magnitude), but also to set constant terms in the models that control the overall model amplitudes. Subsequently, the NGA-East Technical Integrators for GMM development applied four main amplification models to adjust data for residuals analyses. The purpose of those analyses was to screen GMMs and to thereby limit misfit in the recommended GMMs (Goulet et al. 2018). The amplification factors applied in those data adjustments, as with some of the seed models, employed models for active tectonic regions (e.g. Seyhan and Stewart 2014). It was later demonstrated by Hassani and Atkinson (2017), Parker et al. (2019), Zalachoris and Rathje (2019), and Boore (2020) that CENA has weaker V_{S30} scaling (i.e., smaller absolute value slopes), which means that V_{S30} has less predictive power in CENA than in active regions.

The site amplification models used in the 2018 and 2023 NSHMs for CENA were later developed by an expert panel based on a synthesis of available research (Stewart et al., 2020; Hashash et al., 2020). This synthesis drew heavily upon research products from the NGA-East Geotechnical Working Group (GWG; Parker et al., 2019; Harmon et al., 2019a,b). The GWG site amplification model development was reviewed extensively but this occurred outside of the NGA-East SSHAC process. The GWG site amplification (F_S) models are intended to represent amplification relative to V_{S30} = 3000 m/s and κ_0 = 0.006 sec. The reference condition was not defined relative to the NGA-East hard-rock GMMs; hence, if the GMMs are not centered with respect to the assumed reference condition, there is a potential for bias to propagate through the GWG site amplification model when used with the NGA-East models to predict ground motions at other V_{S30} values. Because the aim of the GWG was to derive amplification relative to the reference condition, the issue of potential bias was not addressed.

The site amplification model has linear (F_{lin}) and nonlinear (F_{nl}) components:

$$F_S = F_{lin} + F_{nl} \tag{3.1}$$

The linear component of the model has two components:

$$F_{lin} = F_V + F_{760} \tag{3.2}$$

in which F_V describes the amplification relative to a V_{S30}=760 m/s reference condition and F_{760} describes the amplification for 760 m/s sites relative to 3000 m/s sites. F_{760} carries significant parametric uncertainty in κ_0 due to the lack of empirical data at the reference site condition (e.g., Atkinson, 2012; Boore and Campbell, 2017) and the assumption that κ_0 reflects only material damping (Al Atik et al. 2022). Two terms are used in Eq. (3.2) because they were derived using different procedures. F_V is empirically constrained from NGA-East data (Parker et al., 2019), while F_{760} is derived from ground-response simulations (Frankel et al., 1996; Silva et al., 2015; Boore and Campbell, 2017; Harmon et al., 2019a,b). This two-tier approach was required because it was not possible to empirically derive site amplification relative to 3000 m/s conditions due to the lack of ground-motion recordings at this reference condition.

Because the site adjustments applied during GMM development used a modeling approach different from how the NGA-East GMMs are now applied, the current study was undertaken to assess whether the combined use of NGA-East GMMs and site amplification models is compatible with ground-motion data from CENA. We summarize an expanded (relative to NGA-East; Goulet et al., 2021b) CENA dataset that is presented in greater detail by Ramos-Sepulveda et al. (2023b). We then use residuals analyses to assess model performance with respect to independent variables (magnitude, distance, and V_{S30}) and overall mean model misfits. The residuals analyses consider different NGA-East GMMs (i.e., the 17 models presented by Goulet et al., 2021a) and alternative data selection criteria in consideration of observed regional features. Although the database has grown appreciably, the moment magnitude (**M**) range remains limited (**M** < ~5.8). For this reason, our aim is not to validate scaling relationships over the parameter ranges that control hazard (e.g., close distances and large magnitudes). Rather, we seek to evaluate the need for model adjustments to fit the available data, while leaving source and path scaling as-is. These analyses indicate a consistent pattern of mean misfit (misfit) with respect to period, which can be considered in forward applications through simple adjustments of GMM constant terms. We also investigate likely causes for the observed misfits.

3.3 CENA Ground-Motion Database

The database used in this research is an expanded version of the NGA-East ground-motion database (Goulet et al., 2021b). The NGA-East portion of the database consists of ground-motion intensity measures and metadata (event locations, magnitude, distance, V_{S30}) provided in the electronic supplements to Goulet et al. (2021b). The NGA-East data were merged into a relational database developed for ground-motion studies (see Section 3.7, and Ramos-Sepulveda et al. [2023b] for details) and expanded to include events in CENA since November 2011 (date of the latest event in NGA-East). Additional ground motions in the Texas-Oklahoma-Kansas (TOK) region were also considered in a parallel, collaborative component of this project (Zalachoris and Rathje, 2020; Li et al., 2023), but as of this writing those additional data have not been added to the relational database.

The database expansion considered all events with $\mathbf{M} \geq 4$ in CENA from November 2011 to April 2022, based on event hypocenter locations east of the boundary between the active tectonic and stable continental regions as provided by Dreiling et al. (2014). This boundary has been recently updated (Moschetti et al., 2024) as shown in Figure 3.1. All of the events we processed remain within the updated region. This comprised 187 earthquakes at the locations shown in Figure 3.1. Unprocessed ground motions from these events were downloaded as miniSEED files from the Incorporated Research Institutions for Seismology (IRIS) data center (see section 3.7, *Data Resources*). The number of records was over 73,111, which was considered too large for manual processing procedures as applied in past NGA projects (Goulet et al., 2014; Kishida et al., 2020). Accordingly, we sought an automated or semi-automated alternative, and ultimately decided to use the USGS open-source software gmprocess (Hearne et al., 2019). We introduced options in gmprocess to improve the high-pass corner frequency selection, to facilitate manual review of waveforms, and to resolve other differences between gmprocess and past NGA project processing protocols such as the choice to filter/differentiate/integrate in the time versus frequency domains (Ramos-Sepulveda et al., 2023a).

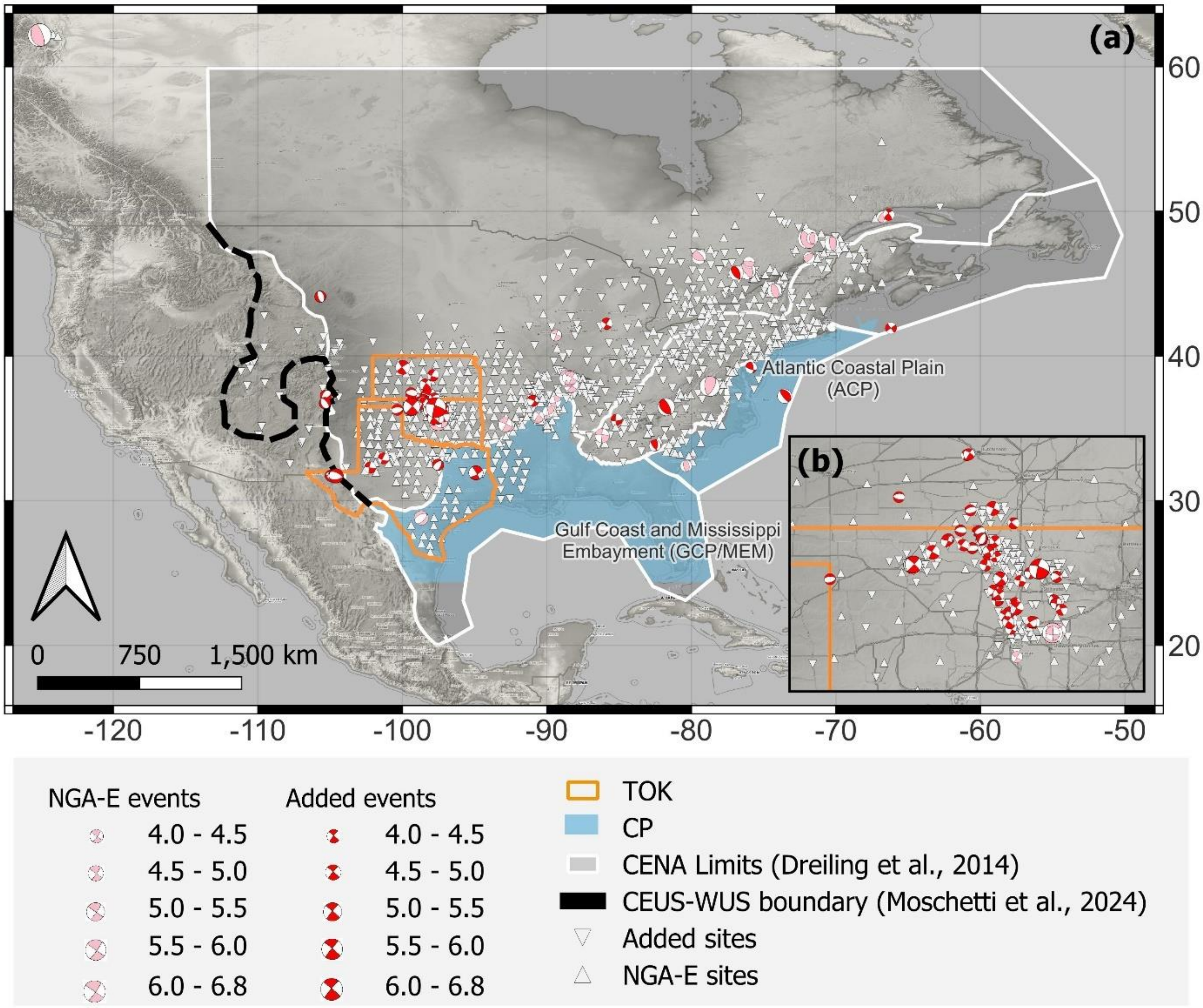


Figure 3.1. (a) Locations of CENA earthquakes and ground-motion recording stations considered in the present study (Ramos-Sepulveda et al., 2023b). Boundaries of Texas-Oklahoma-Kansas (TOK) and coastal plain (CP) regions (the CP boundary is defined using a minimum sediment depth, from Boyd et al. 2023, of 100 m). (b) Detailed view of Oklahoma where a high event density occurred.

The recent events typically have a greater density of recordings relative to the NGA-East events, which is a consequence of the growth of seismic instrumentation in CENA. The more recently installed instruments, including at prior locations of Temporary Array (TA) stations, have broader usable frequency ranges. Ground-motion intensity measures (peak ground acceleration, PGA; peak ground velocity, PGV; and spectral acceleration, Sa from 0.01 to 10 sec) from these events and information on their component-specific usable frequency ranges were uploaded to the relational database. For each of these intensity measures, as-recorded components are provided along with median-, minimum-, and maximum-component horizontal motions (RotD50, RotD00, and RotD100, respectively; Boore 2010).

Metadata describing the seismic sources, path, and site conditions were assembled. General information such as the name and location of the instrument, hypocenter location, and event origin date and time can be obtained from gmprocess' output. Newly added events include moment magnitudes from moment tensor solutions (Dziewonski et al., 1981; Ekström et al., 2012; and Guy et al., 2015) for 76 out of the 100 events, and estimates with the uncertainty of converted magnitudes (i.e., from events that did not have moment tensor solutions) were provided for the remaining 24 events (Nuclear Regulatory Commission, 2012; NUREG-2117). Procedures described in Contreras et al. (2022) were used to generate approximate rupture dimensions based on magnitude, event type, hypocenter, and orientation of one or two nodal planes. Contreras et al. (2022) modified a fault surface simulation routine previously presented by Chiou and Youngs (2008). The modifications incorporate a magnitude-area scaling relationship for stable continental regions (Leonard 2014). Site parameters were derived from shear-wave velocity (V_S) profiles where available and otherwise from the Li et al. (2022) geology-slope proxy for TOK sites and the Parker et al. (2017) geology-slope proxy for other CENA sites.

Figure 3.2 shows data distribution (stacked histograms) as functions of rupture distance (R_{rup}), V_{S30}, and **M**. The added data are mostly applicable for **M** < 5.8, R_{rup} > 10 km, and V_{S30} = 200 to 2000 m/s. The new data significantly increase the number of recordings per event and per station relative to what was available from the original NGA-East database, although it does not extend the magnitude range. Tables presenting source and site attributes, as well as a ground-motion flatfile, for the information considered in this project are provided by Ramos-Sepulveda et al. (2023b).

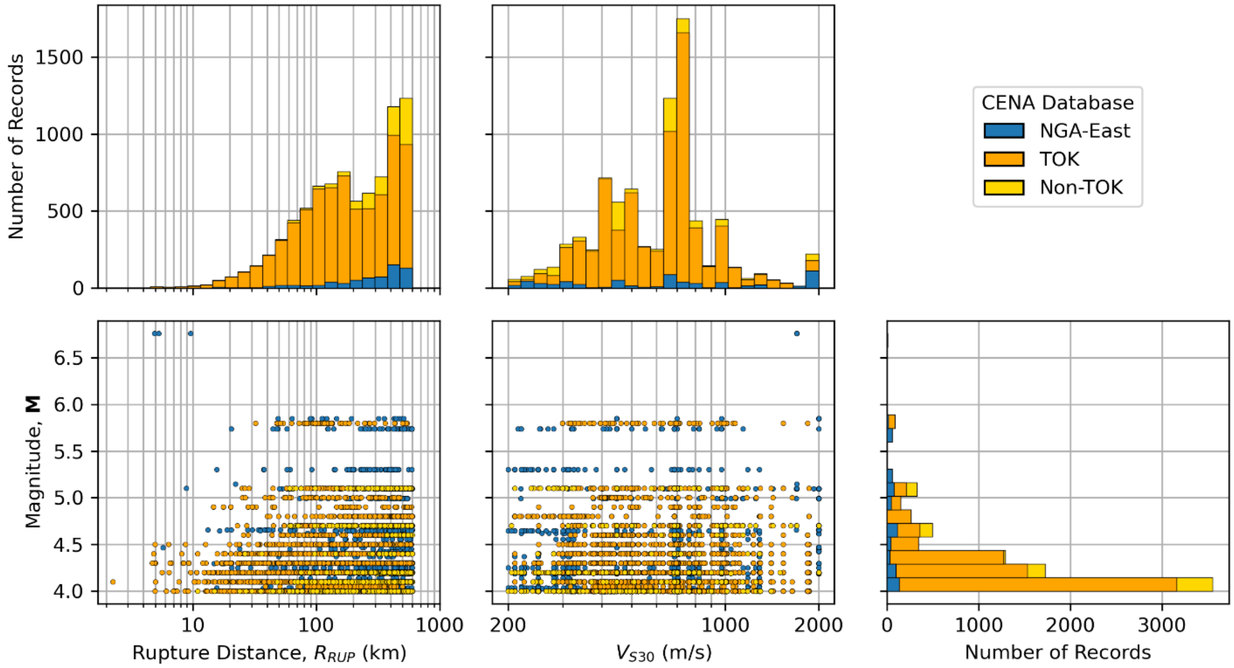


Figure 3.2. Distributions of CENA dataset with respect to rupture distance, V_{S30}, and magnitude, showing differences between NGA-East and added data (Ramos-Sepulveda et al., 2023b).

Figure 3.3 shows the numbers of available recordings and events before and after the expansion of the NGA-East database as a function of oscillator period. There are more records at all periods, but the increase is particularly pronounced at short periods because many ground-motion networks now have instrument sample rates that are adequate to capture these short-period responses, which had not been the case previously (Goulet et al. 2021b).

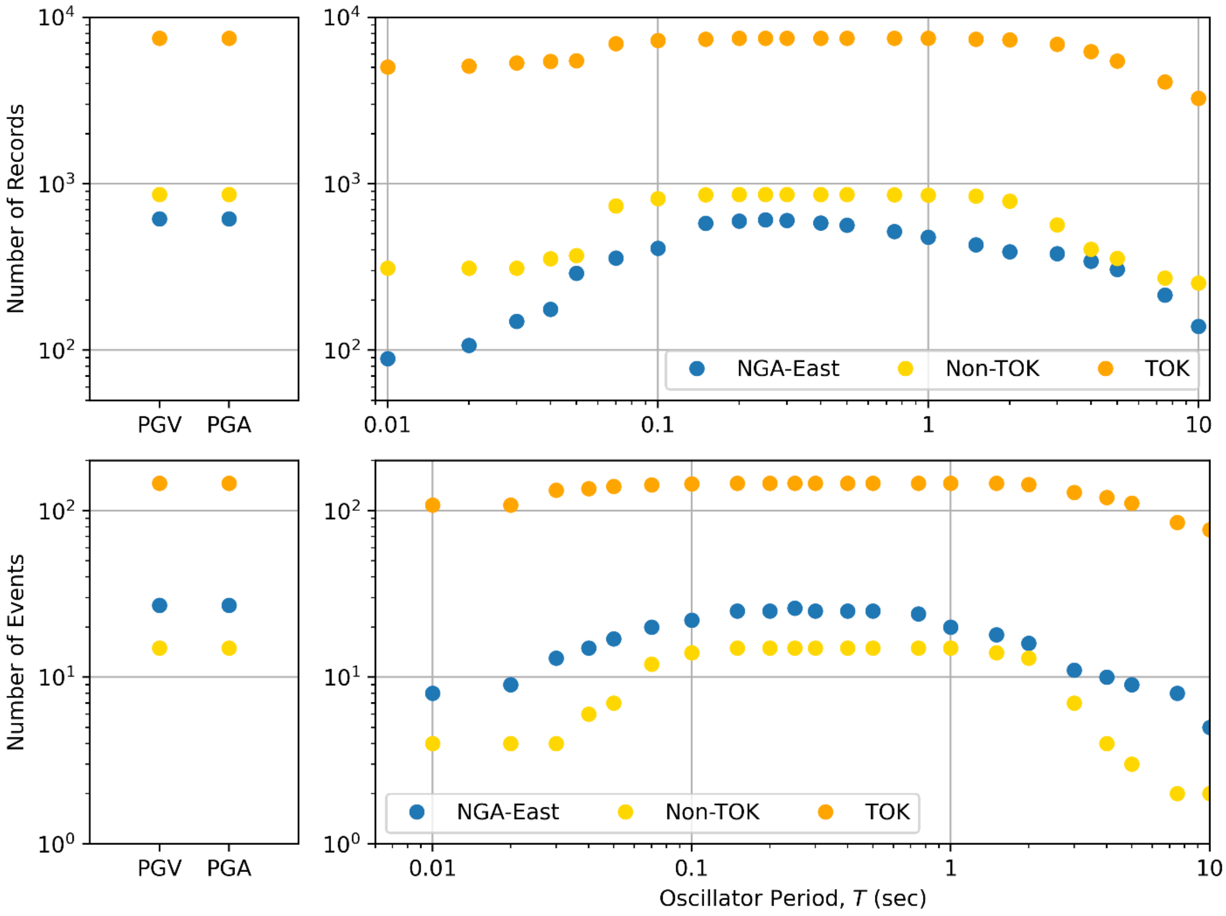


Figure 3.3. Numbers of available RotD50 spectral accelerations and events (Ramos-Sepulveda et al., 2023b) as a function of oscillator period. The NGA-East numbers are for M ≥ 4 events, to be compatible with the adopted data selection criteria.

The data was archived in a publicly web-serviced ground motion relational database that connects 25 tables of data, metadata and intensity measures by primary and secondary keys (Buckreis et al., 2023). This database also includes data from previous NGA projects (NGA-West1 and NGA-West2) and recent California studies (Buckreis et al., 2022, Nweke et al., 2022). The ground motion database uses MySQL as the management system, and an application programming interface (API) was written to facilitate queries using URLs. Ground motion records from NGA-East and newly added records are stored in the database under the “collection_id” of 3 and “user_id” of 2 respectively.

The database includes computed ground motion intensity measures (e.g., peak ground acceleration, peak ground velocity, response spectra from 0.01 to 10 sec, and effective amplitude spectra) as well as the acceleration time series. However, it does not include the time series processed as part of the NGA-East project, only the intensity measures and metadata as provided in Goulet et al. (2021b). Most of the NGA-East metadata was not modified or updated with the exception of sites with recently measured V_S profiles. Additional features related to CENA that are part of the database are the Wisconsin glaciation limits, coastal plains limits, basins used in the Parker et al. (2017) proxy, and coastal plain sediment depth as measured by Boyd et al. (2024). All of this information can be found in the geometry table.

3.4 Data Analysis

3.4.1 Residuals Calculations

GMM performance is assessed through residuals analyses. We define the residual as the difference between an observation (natural-log intensity measure from a recording) and a GMM estimate of the mean ground motion,

$$\delta_{ijk} = ln\left(Y_{ij}\right) - \left[\mu_{ln,k}\left(\mathbf{M}_i, \left(R_{rup}\right)_{ij}, V_S = 3000\right) + F_{lin}(V_{S30})\right] \qquad (3.3)$$

where $Y_{i,j}$ is the observed intensity measure for event i and station j, $\mu_{ln,k}(\mathbf{M}_i, R_{rup,ij})$ is the natural-log (ln) mean estimated intensity measure at the reference site with V_s = 3000 m/s for magnitude

Mi and rupture distance $R_{rup,ij}$ from GMM k, and F_{lin} is the CENA-specific linear site amplification model (Stewart et al., 2020; Eq. 3.2) in ln units. For Gulf Coast sites, the path component of the GMM was adjusted for additional anelastic attenuation as given in Eq. 10 of Goulet et al. (2021a). The residuals are partitioned into the following terms using mixed-effect regression analysis (Bates et al., 2015; R Core Team, 2019)

$$\delta_{ijk} = c_k + \eta_{E,i} + \delta W_{ij} \tag{3.4}$$

where c_k is the overall mean misfit for GMM k, $\eta_{E,i}$ is the event term for event i and δW_{ij} is the within-event residual. Note that subscript k is not used with the event term and within-event residual for brevity, although these terms are specific to a GMM.

Residuals analyses were performed using the NGA-East central branch GMM and the CENA dataset (as described in the previous section) with minimal screening in which **M** ≥ 4 events were considered with recordings at distances $R_{rup} \leq 600$ km (with some exceptions for the CP region, as described subsequently). The NGA-East GMMs can be used to estimate intensity measures for distances up to 1500 km, but the 600 km threshold was applied to avoid problems related to biased ground-motion sampling at larger distances. Ground-motion data were not considered beyond their maximum usable period (taken as 80% of the inverse of the high-pass corner frequency). No lowest-usable period was applied if the low-pass corner frequency (f_{cLP}) is 40 Hz or greater since Sa are usually controlled by lower frequencies (Douglas and Boore 2011; Boore and Goulet 2014); otherwise the lowest-usable period was taken as $1.25/f_{cLP}$. The purpose of these analyses was to examine trends in residuals with magnitude or distance, which if present would influence the need for (and magnitude of) adjustment factors for CENA. The regions considered are TOK, coastal plain (CP) (Gulf Coast, Mississippi Embayment, and Atlantic Coastal Plain), the remainder of CENA, and combinations thereof (Figure 3.1).

3.4.2 Regional Variations of Magnitude and Distance Scaling

As shown in Figure 3.1, the CENA database includes a large number of events (146) in the TOK region and a small number in the coastal plain (CP) region (8), with the balance being 33 events outside of those regions. The NGA-East project screened out potentially induced events (PIE) (Goulet et al. 2021a), which largely originated in the TOK region. This is consistent with other research suggesting that TOK may have distinct ground-motion magnitude- and distance-scaling characteristics (e.g., Zalachoris and Rathje, 2019; Moschetti et al., 2019). Regarding the CP region, the NGA-East GMM development team found that the Gulf Coast portion of CP had higher rates of anelastic attenuation (Goulet et al. 2021a; also later observed by Pezeshk et al., 2021), while the Atlantic Coastal Plain did not. Goulet et al. (2021a) provided adjustment factors for the Gulf Coast to account for this effect. Our CP region groups together both Atlantic and Gulf regions (Figure 3.1), for which we anticipate potentially distinct path characteristics and differences in site response relative to the rest of CENA due to the presence of relatively deep sediments (Guo and Chapman, 2019; Chapman and Guo, 2021; Schleicher and Pratt, 2021; Pratt and Schleicher, 2021; Boyd et al. 2023).

In consideration of these factors, we examined magnitude- and distance-scaling effects beginning with only the non-TOK / non-CP region to maximize consistency with data selection criterion used during NGA-East GMM development, then we examine differences for CP and TOK. While events in these regions are strictly incompatible with the originally applied criteria, they are nonetheless important to consider in this study because the combined NGA-East GMM and site factors are applied across CENA in the NSHM.

We performed residuals analysis followed by mixed-effects partitioning of the non-TOK / non-CP dataset (33 events, 1169 recordings) for the central branch GMM. Figure 3.4 shows the resulting trends of event terms ($\eta_{E,i}$) with magnitude (grey symbols; the blue symbols are for CP sites and are discussed subsequently). The binned means are positive for some bins and negative for others, but we observe no consistent trend with $\mathbf{M}$ for any of the intensity measures. This suggests that the magnitude scaling in the NGA-East GMM is consistent with the expanded CENA database. The short-period event terms are mostly negative for CP events but they fall within the range of the non-TOK / non-CP data. Differences are essentially imperceptible at long periods.

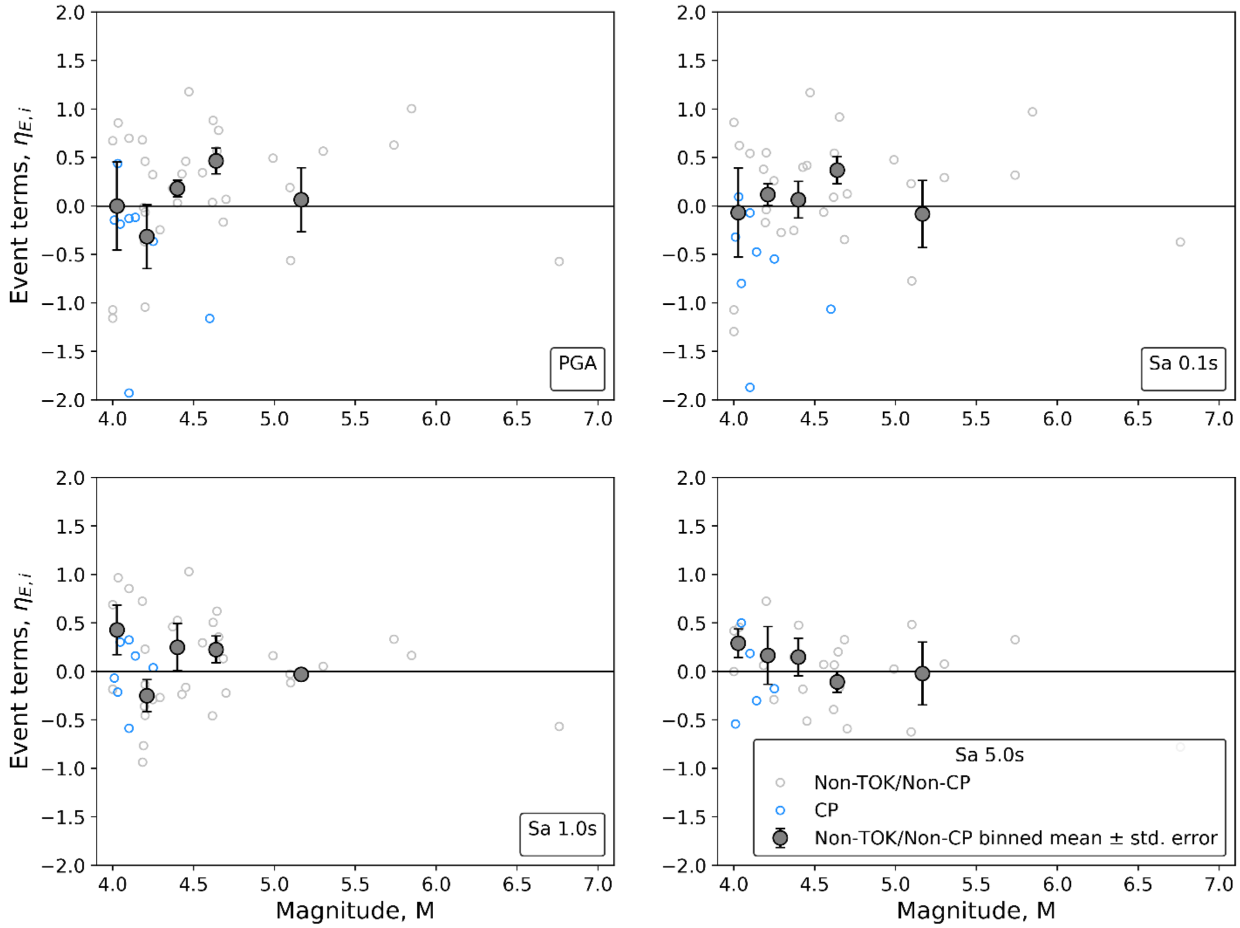


Figure 3.4. Trends of event terms with magnitude for non-TOK / non-CP and CP regions. Binned means are for non-TOK / non-CP data and vertical bars through binned means indicate ± one standard error of the mean. No binned means are shown for the CP region due to the limited number of events.

Figure 3.5 shows the trend of δW_{ij} with respect to R_{rup} (as before, gray symbols are for non-TOK / non-CP sites, blue are for CP sites). The non-TOK / non-CP trend is flat to 600 km, which suggests that the distance scaling of the NGA-East GMM is consistent with the expanded CENA database. For CP sites, at short periods we observe negative binned means at large distances (R_{rup} ≥ 300 km), which indicates a higher level of anelastic attenuation. For subsequent analyses, we only consider CP data to maximum distances of 300 km to avoid tradeoffs with distance-scaling misfits.

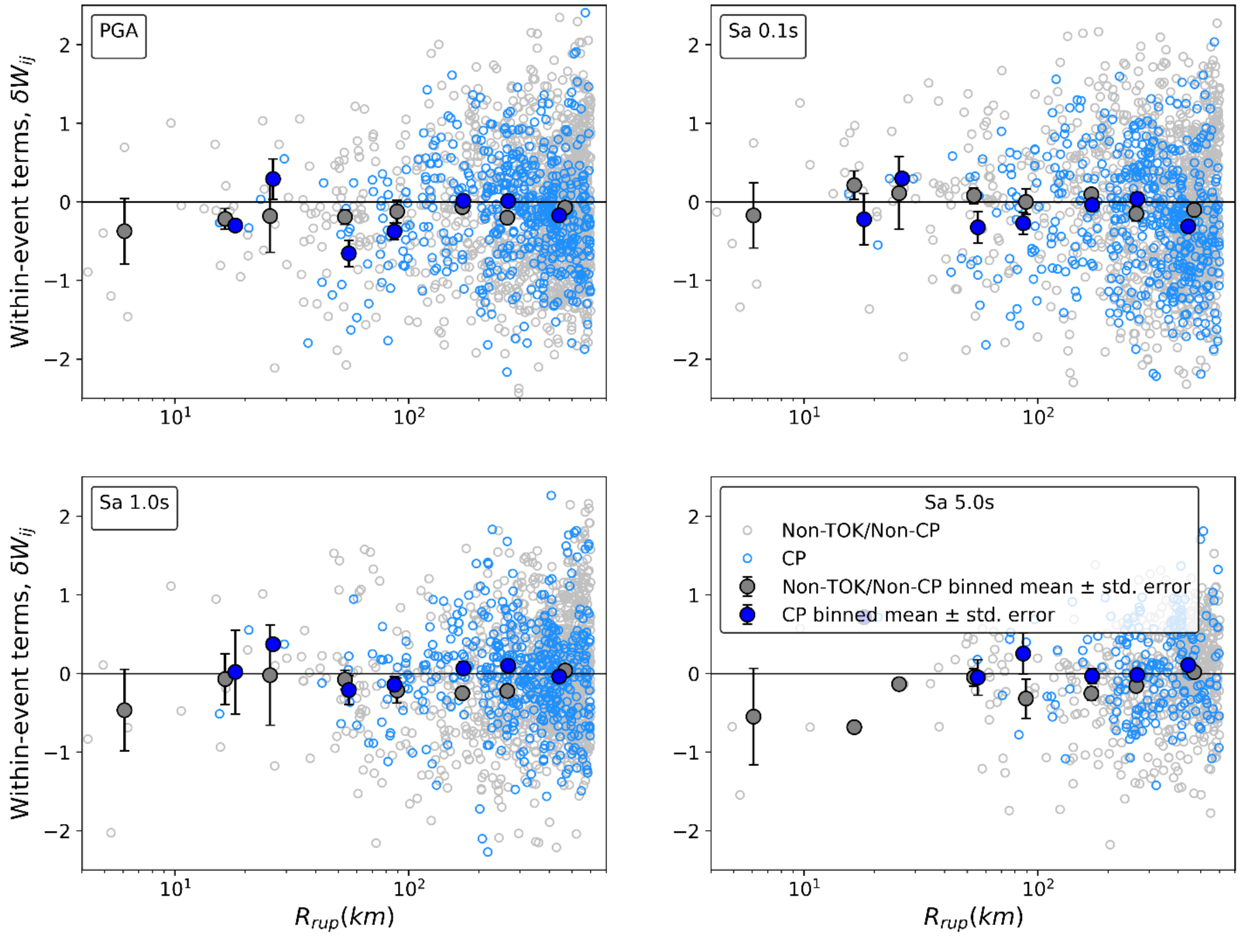


Figure 3.5. Trends of within-event residuals with distance (R_{rup}) for non-TOK / non-CP and CP regions. Vertical bars through binned means indicate ± one standard error of the mean.

We next examine regional variations between TOK (Figure 3.1) and the remaining CENA data (including CP) by computing residuals for the full dataset and then distinguishing the resulting outcomes by region. Figure 3.6 shows the resulting trends of event terms with magnitude. For TOK, the binned means of event terms are near zero for **M** = 4 – 5. For larger **M**, the event terms are negative, which has been observed previously (Zalachoris and Rathje 2019; Li 2022). Non-TOK event terms, which include both non-TOK / non-CP and CP, are generally positive at short periods, whereas TOK event terms mostly average nearly zero. This indicates that regional variations in ground-motion levels are present. This can be understood by recalling that a single misfit term is computed across all data; accordingly, deviations in mean event term for a particular region indicate different ground-motion levels in that region relative to the overall average. The near-zero mean of event terms for TOK is a consequence of that region dominating the dataset (146 of 187 events), whereas the positive mean of short-period event terms for non-TOK events indicates stronger average ground motions than the overall average for the CENA database.

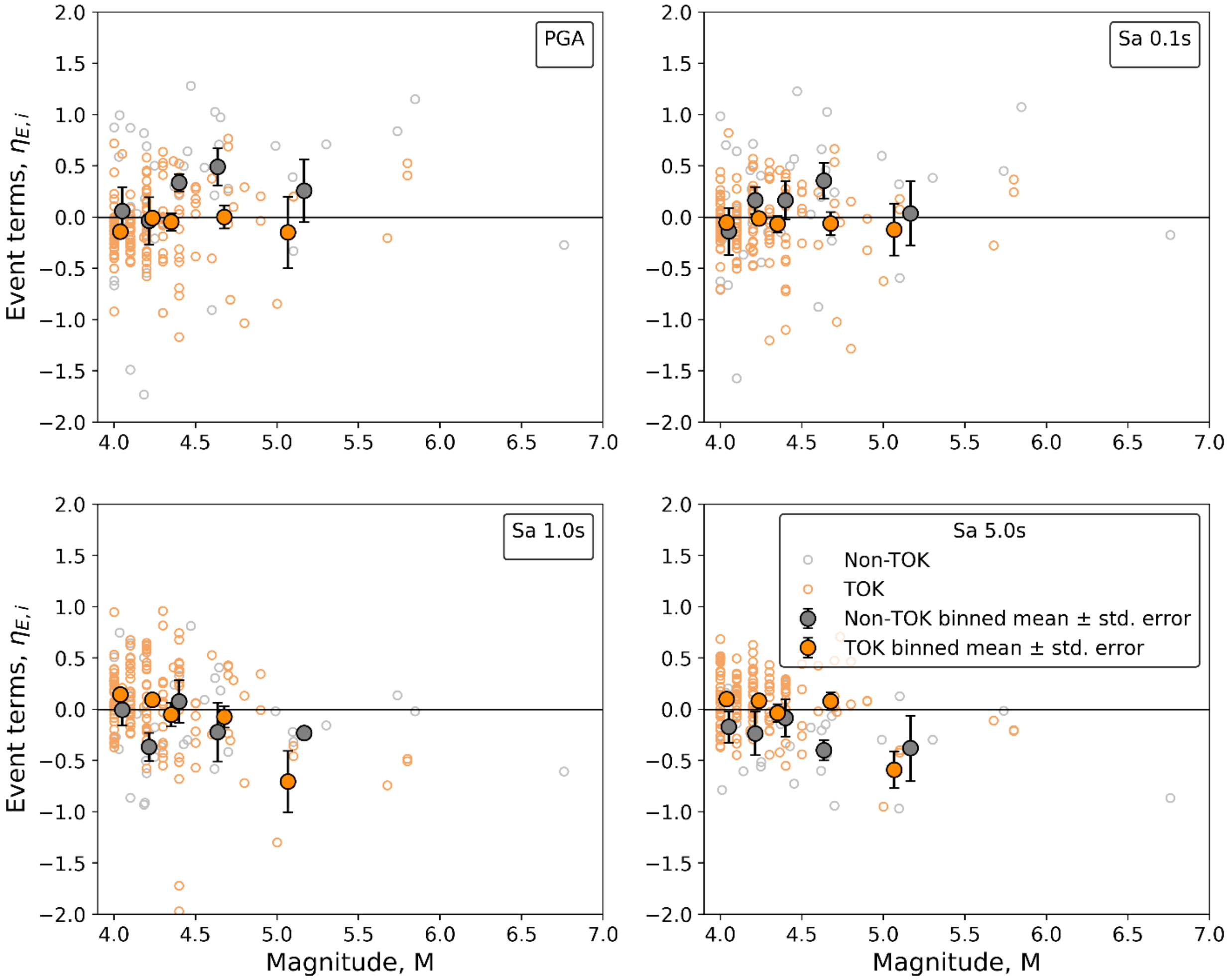


Figure 3.6. Magnitude-dependence of event terms for TOK and non-TOK regions. Vertical bars through binned means indicate ± one standard error of the mean.

Figure 3.7 shows the trend of δW_{ij} with respect to R_{rup}. At short periods, we observe a significant upward trend in binned means for TOK events as R_{rup} increases from approximately 10 to 150 km, whereas the non-TOK trends are flat, as before. Neither region has trends with distance for long-period intensity measures (Sa at 1.0 or 5.0 sec). These results indicate that the distance attenuation component of the NGA-East central branch GMM is biased in the TOK region, which is not surprising because it was not developed for application to induced earthquakes, which dominate the TOK dataset.

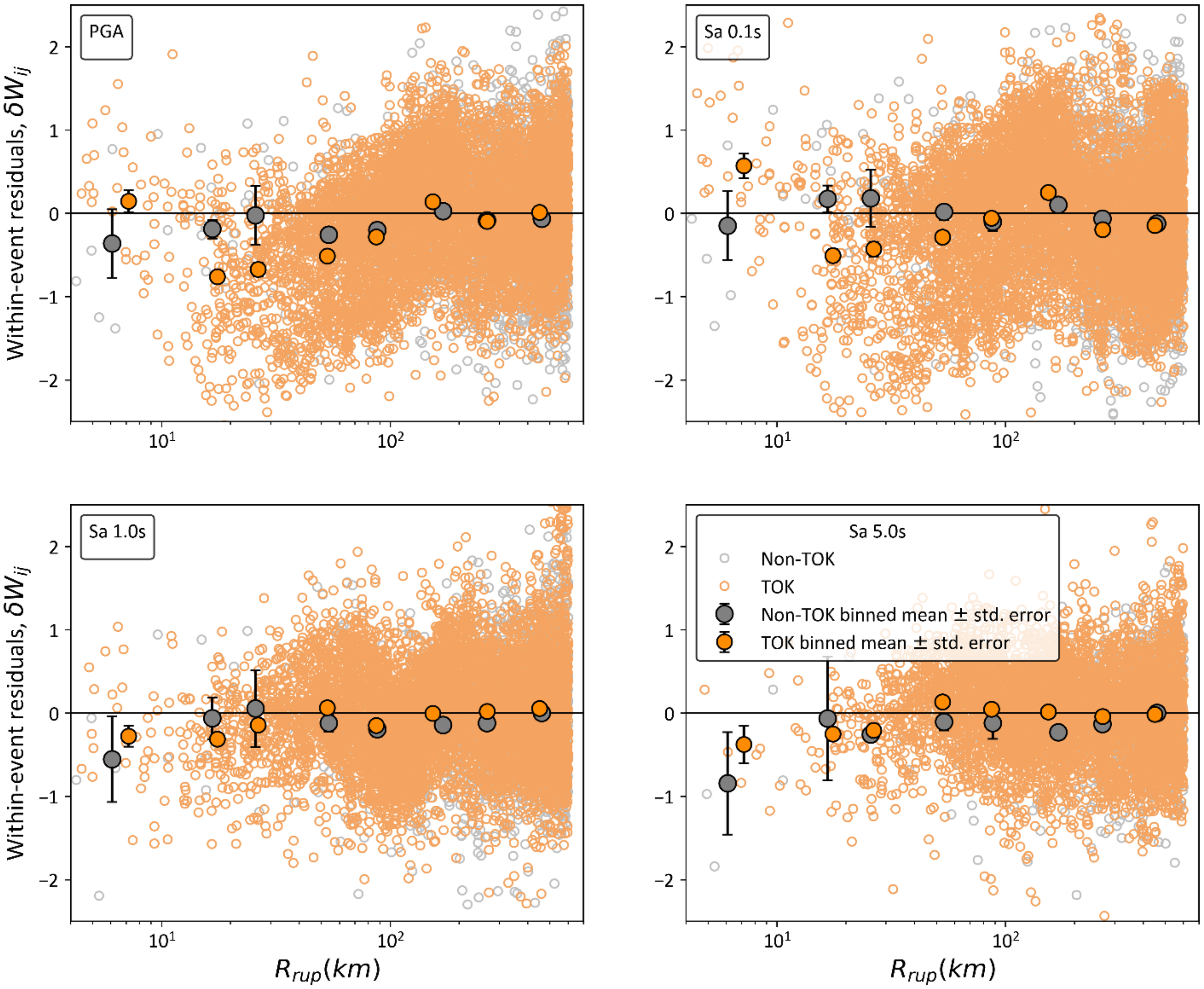


Figure 3.7. Distance-dependence of within-event residuals for full dataset. Vertical bars through binned means indicate ± one standard error of the mean.

In this paper, we do not attempt to model the scaling trends and mean model misfits for the TOK region, which is the subject of a parallel effort that has recommend a modified GMM for that region (Chapter 5 of Centella et al., 2023).

3.4.3 Validation of V_{S30}-Scaling Model

To evaluate the performance of the V_{S30}-dependent site amplification model (F_{lin} in Eq. 3.1), we partition the δW_{ij} using mixed effects analysis to evaluate site terms (η_S) and remaining residuals (ε_{ij}),

$$\delta W_{ij} = \eta_{S,j} + \varepsilon_{ij} \tag{3.5}$$

The site terms represent the approximate misfits of the model used in the original residuals calculation (Eq. 3.1) for sites in the full dataset, after model misfits and event-terms have been subtracted. Figure 3.8 shows the trend of site terms with V_{S30}. The results in Figure 3.8 show no appreciable trend, except for sites with V_{S30} > 1000 m/s, where downward trends are evident for some periods. The trends are similar for TOK sites as shown in Chapter 5 of Centella et al. (2023). These results indicate that the V_{S30}-scaling component of the site amplification model is consistent with the data to a maximum V_{S30} of 1000 m/s.

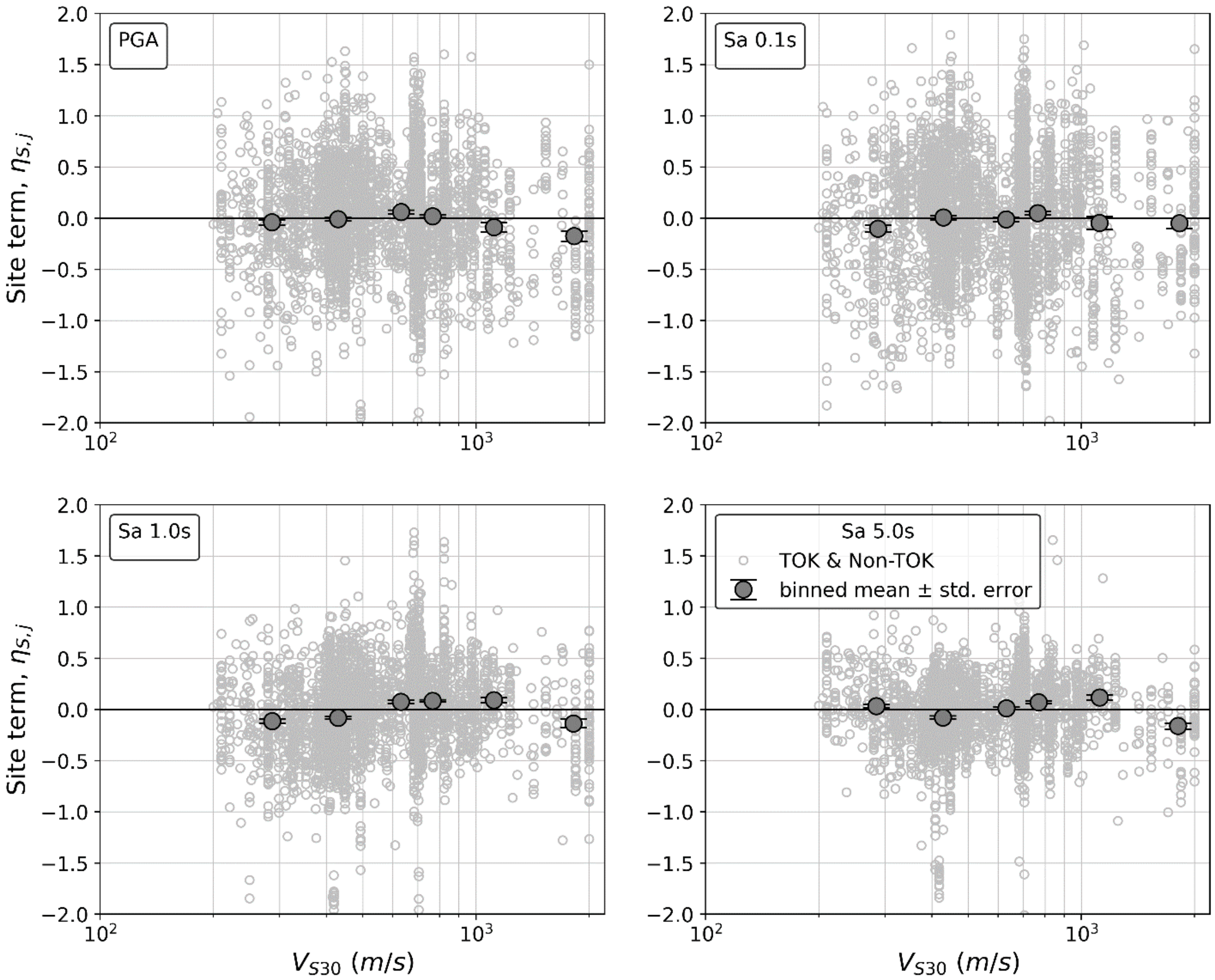


Figure 3.8. Site terms trends with V_{S30} for full dataset. Vertical bars through binned means indicate ± one standard error of the mean.

3.4.4 Analysis of Mean Misfits

The data analysis in the prior subsections suggests that the TOK region has distinct features that affect distance attenuation and overall ground-motion levels. Moreover, as shown in Figure 3.1, TOK has a substantial event concentration (events per area) compared to the rest of CENA, which to some extent produces results that largely reflect TOK attributes. Although less compelling, CP sites also have some different ground-motion features, mainly in relation to large-distance anelastic attenuation.

Accordingly, we developed different subsets of the CENA data for analysis of mean misfits:

1. Non-TOK / Non-CP: Events and sites within TOK and the CP are excluded.
2. Non-TOK: All non-TOK events are considered, including CP. No TOK events are considered.
3. Partial TOK: All non-TOK events are considered. Within TOK, 9 events from the 146 are selected (randomly) so that the event density (number of events per area) is consistent with other parts of CENA.

Mixed-effects analyses (Eq. 3.4) were repeated for each of these subsets of data using the central branch NGA-East GMM. This produces three sets of misfits (c_k), as shown in Figure 3.9, the mean of these sets ± 1 standard deviation are shown. For reference purposes, the c_k term for the complete CENA dataset is also shown. The results show a consistent trend of negative misfits at short periods (i.e., models are over-predicting these components of ground motion) and positive misfits at long periods (models are underpredicting). There are modest differences between the three data subsets, with partial TOK producing the largest misfits in terms of absolute value, non-TOK / non-CP the smallest, and non-TOK being an intermediate case. Nonetheless, all of the data subsets show less misfit than the complete CENA dataset, indicating that the concentration of data from TOK is contributing the large misfits at short and long periods.

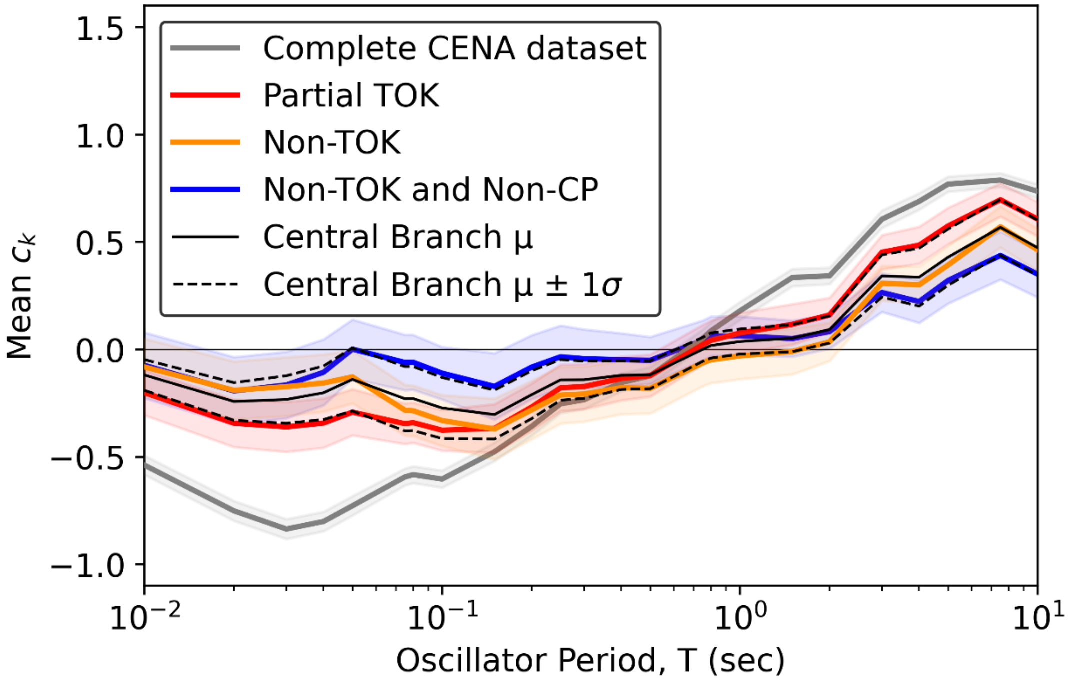


Figure 3.9. Period dependence of misfit term c_k for NGA-East central branch GMM for alternate datasets. The shaded regions enclose ± one standard error.

Figure 3.10 shows c_k for the non-TOK dataset using all 17 NGA-East GMMs, along with the population weighted mean and mean ± one weighted standard deviation. The weights applied in these calculations were taken from Goulet et al. (2021a) for the 17 GMMs and were equal for the three data selection criteria. We did not evaluate potential data misfits in relation to magnitude and distance scaling for each of these 17 GMMs as was done for the central branch. It is possible that individual GMMs with large misfits in Figure 3.10 are influenced by scaling problems. Comparing Figures 3.9 and 3.10, it is clear that the uncertainty introduced by alternate GMMs substantially exceeds that from alternative data selection protocols. As shown in Figure 3.11, the misfits found when only the NGA-East data are considered is similar to that shown for the non-TOK version of the expanded dataset in Figure 3.9. Moreover, Boore (2020) observed qualitatively similar misfit trends to those reported here, using the NGA-East dataset and the Boore (2018) GMM.

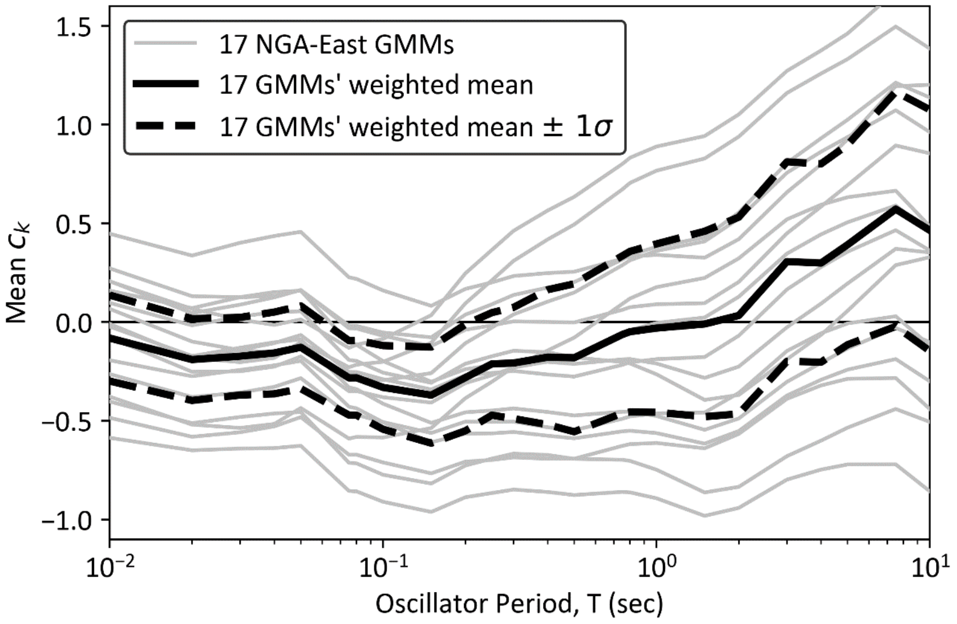


Figure 3.10. Period dependence of misfit term c_k for 17 NGA-East GMMs for Non-TOK data and weighted mean misfit.

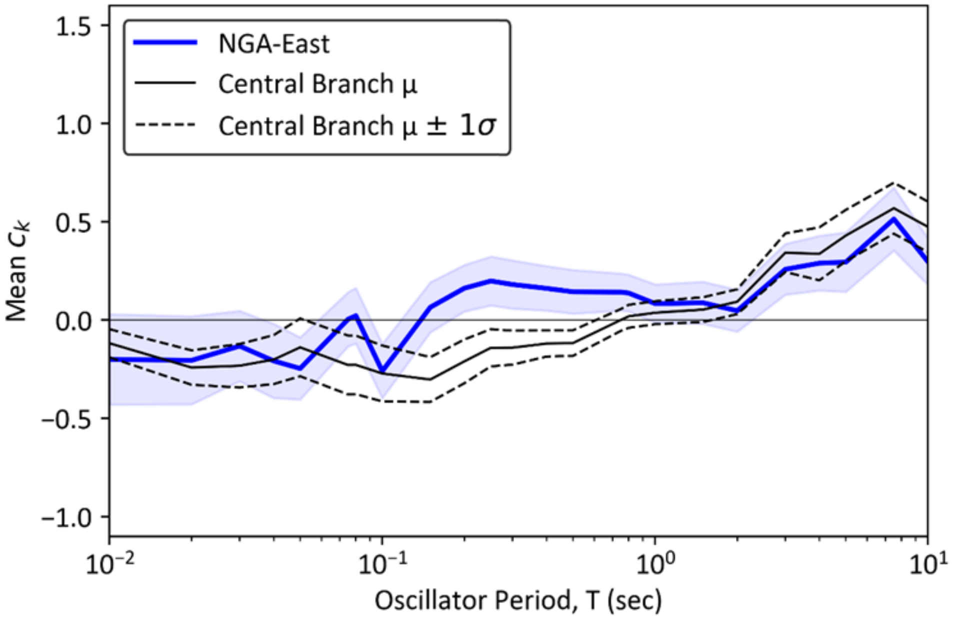


Figure 3.11. Period dependence of misfit term c_k for NGA-East central branch GMM for NGA-East data only compared to the central branch mean (same as Figure 3.9). The shaded regions enclose ± one standard error.

3.4.5 Model Adjustment Factors

We consider the 17 alternative NGA-East GMMs (weighted as given by Goulet et al. 2021a) and 3 data selection criteria (as given in *Analysis of Mean Misfits*) to compute 51 misfit terms. Various weighting combinations were considered for data selection, including equal weighting and weighting that gives preference to the non-TOK / non-CP dataset. Ultimately, the recommended weights were 0.2 for the non-TOK set, 0.2 for the partial TOK set, and 0.6 for the non-TOK / non-CP set. All sites are considered in the development of the model adjustment factors presented in this section, and modifications for stiff sites with V_{S30} > 1000 m/s are presented subsequently.

Figure 3.12a shows the resulting weighted mean of the 51 misfit terms (overall μ) ± one weighted standard deviation (σ_e). A smoothed version of the misfits is also shown for use in forward applications as a model adjustment factor. The weighted mean misfits from the 51 values was found to be equivalent to the weighted mean misfits obtained using only the single central GMM and the three alternative datasets, which confirms that the central branch model is the weighted mean of the 17 alternative NGA-East GMMs. Figure 3.12b shows the period dependence of the overall standard deviation (across the 51 misfit terms) and the standard deviation from alternative data selection criteria only ($\sigma_{e,data}$). The latter standard deviation ($\sigma_{e,data}$) is computed using mean misfits from the central branch GMM with the three datasets (between-GMM uncertainties are not included). This standard deviation was found to be nearly identical to those obtained with other single GMMs from the group of 17.

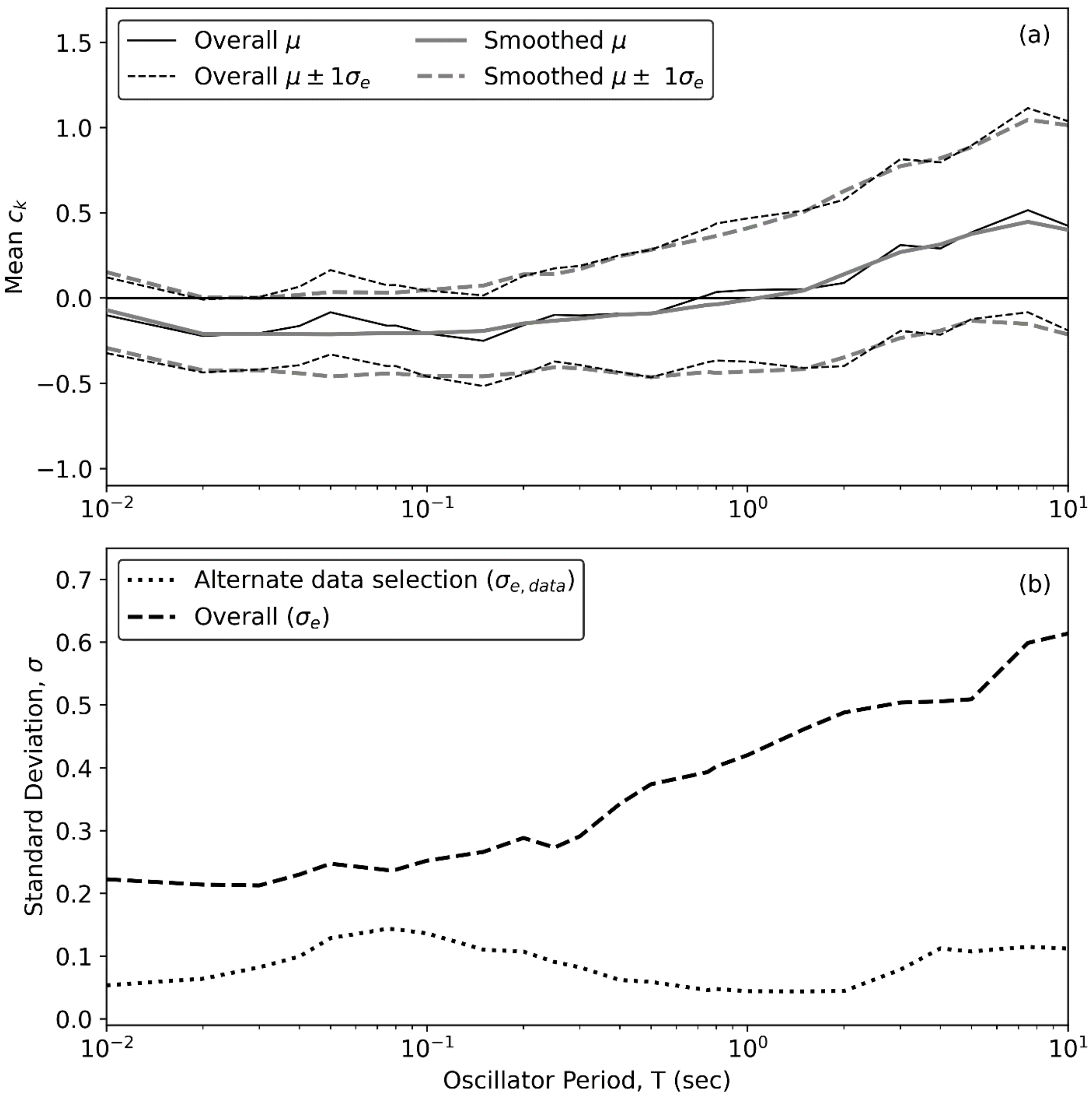


Figure 3.12. (a) Period dependence of model adjustment factor c_k and its uncertainty, as derived from 51 misfit terms and (b) overall standard deviations for all 51 misfit terms (3 datasets, 17 alternative NGA-East GMMs) and standard deviations solely related to alternate dataset selections.

For forward analysis in which the 17 NGA-East GMMs are considered in the logic tree, which is the preferred approach, there is no need to consider the between-GMM uncertainty in the logic tree (to do so would double-count this uncertainty). In this case, the applicable epistemic uncertainty is that labeled as “alternate data selection” in Figure 3.12b ($\sigma_{e,data}$). If only the central GMM is considered, the central branch model adjustment factors and the larger epistemic uncertainty should be considered. Table 3.1 provides values of the recommended model adjustment factors after smoothing (same values as in Figure 3.12a) and standard deviations representing epistemic uncertainties.

To apply the model adjustment factors in forward ground-motion analyses, the factors are simply added to the natural-log mean ground motions as computed from the combined models (hard-rock GMM and site response). Further recommendations on model application are provided in Section *3.6*.

Table 3.1. Recommended natural-log model adjustment factors (smoothed weighted mean of c_k values), epistemic uncertainties (expressed in the form of a natural-log standard deviation), and scaling coefficients for modifying model adjustment factors for high V_{S30} sites (Eq. 3.6).

Period, T (s)	**μ (smoothed weighted mean of c_k)**	**$\sigma_{e,\ data}$**	**σ_e**	**b**
PGA	-0.040	0.099	0.255	-0.346
PGV	-0.085	0.167	0.374	-0.250
0.010	-0.070	0.054	0.223	-0.352
0.020	-0.210	0.064	0.214	-0.305
0.030	-0.211	0.082	0.213	-0.261
0.040	-0.211	0.099	0.230	-0.230
0.050	-0.212	0.129	0.247	-0.220
0.075	-0.205	0.144	0.237	-0.220
0.080	-0.205	0.143	0.238	-0.220
0.100	-0.205	0.136	0.252	-0.220
0.150	-0.192	0.110	0.266	-0.220
0.200	-0.149	0.107	0.288	-0.220
0.250	-0.132	0.091	0.273	-0.220
0.300	-0.121	0.082	0.291	-0.220
0.400	-0.097	0.062	0.343	-0.220
0.500	-0.090	0.059	0.374	-0.220

0.750	-0.040	0.046	0.393	-0.205
0.800	-0.037	0.048	0.402	-0.193
1.00	-0.010	0.044	0.420	-0.175
1.50	0.045	0.044	0.461	-0.173
2.00	0.140	0.044	0.488	-0.173
3.00	0.270	0.079	0.504	-0.173
4.00	0.315	0.112	0.506	-0.173
5.00	0.377	0.108	0.509	-0.173
7.50	0.448	0.115	0.599	-0.173
10.0	0.401	0.112	0.614	-0.173

3.4.6 Stiff Site Modifications

The site term plots in Figure 3.8 indicate a downward trend for stiff sites with V_{S30} > 1000 m/s, whereas the mean of site terms is nearly zero with no trend for softer sites. To address this, we fit the V_{S30} trend of the site terms as follows:

$$\Delta c_k = \begin{array}{cc} 0 & V_{S30} \leq \ 1000\, m/s \\ b \times ln\left(\frac{V_{S30}}{1000}\right) & 1000 < V_{S30} \leq 2000\, m/s \\ b \times ln(2) & V_{S30} > 2000\, m/s \end{array} \tag{3.6}$$

where V_{S30} is in m/s units and b is a coefficient that represents the slope of the site terms between 1000 and 2000 m/s, with the constraint of zero ordinate at 1000 m/s. Figure 3.13 shows the resulting values of b and the smoothed version recommended for application. Table 3.1 provides the smoothed values. The effect of the modification is to reduce, but not eliminate, the recommended model adjustments for stiff sites.

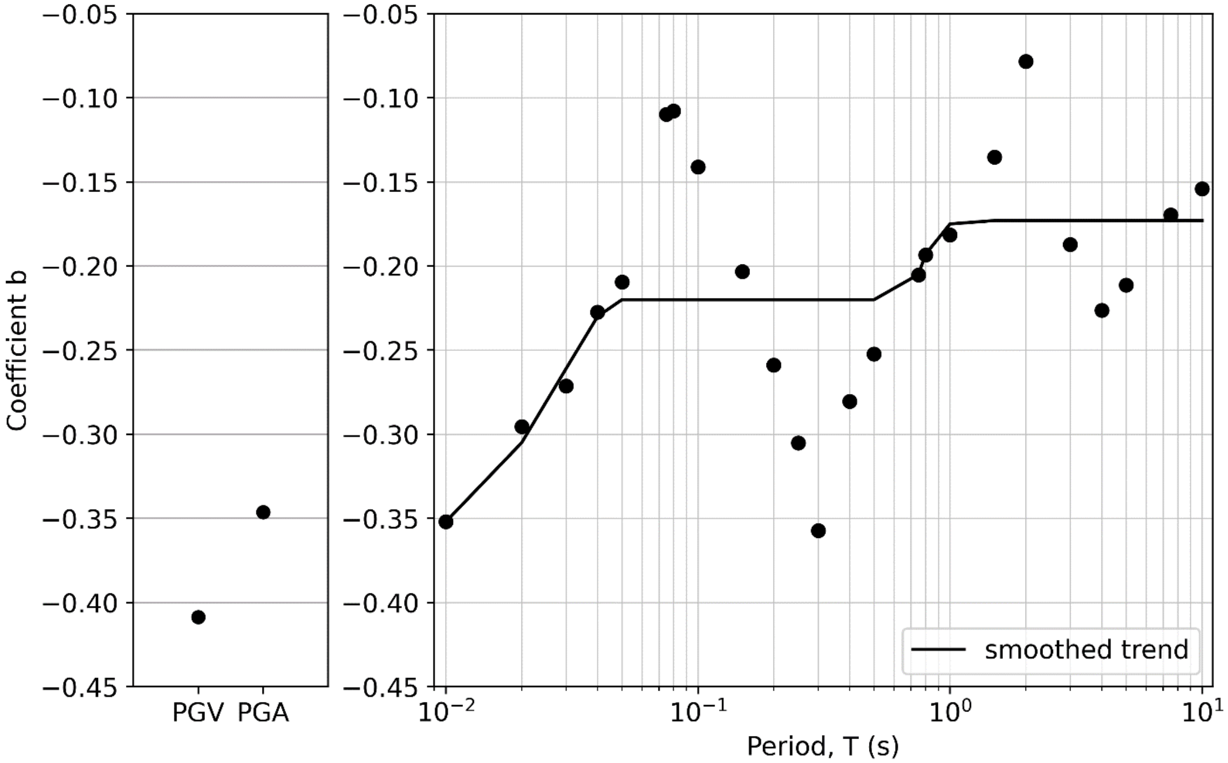


Figure 3.13. Slope of the site terms with V_{S30} between 1000 and 2000 m/s as a function of period.

3.5 Causes of Misfits

The GMM used for residuals analysis in the previous section has three components: a hard-rock GMM (i.e., $\mu_{ln,k}$ in Eq. 3.3), site amplification from hard-rock to V_{S30} = 760 m/s (F_{760}), and site amplification from 760 m/s to alternate V_{S30} values in the range of the data (F_V). Given the lack of trend of site terms with V_{S30} over the V_{S30} range containing most of the data (Figure 3.8), the F_V model is unlikely to be the cause of the misfits, which instead are likely associated with some combination of the hard-rock GMM and F_{760}. In this section, we assess the likely causes of the misfits.

3.5.1 Differences in NGA-West and NGA-East Site Amplification Models

In this sub-section we investigate the degree to which differences between the site amplification models employed during NGA-East GMM development under the SSHAC process, relative to those now used in forward application, may explain some of the observed misfits. The application of site factors in the development of NGA-East GMMs occurred in two phases: (1) development of seed models (PEER 2015), where in some cases data was adjusted to the 3000 m/s reference condition for model calibration; and (2) integration of seed models within the process of developing the final recommended GMMs (Goulet et al. 2018). As summarized by Parker et al. (2019) (their Table 1), the ten NGA-East seed models were developed using a variety of approaches, including simulations to directly estimate ground motions for the reference site condition, ground response simulations to compute site responses for different site conditions (which in turn were used for data adjustments), and two-step data adjustments as reflected by Eq. (3.2), in which the F_V term is based on a western US ergodic site amplification model (Seyhan and Stewart 2014; hereafter SS14) and F_{760} is based on CENA-specific simulations. Further information on the model development procedures is given in PEER (2015) and are summarized by Parker et al. (2019) and Section 4.4 of Centella et al. (2023).

For the present comparison, we focus on the second (integration) phase of this process, which used SS14 to adjust CENA ground motions to a site condition of V_{S30}=760 m/s, followed by adjustments from the 760 m/s condition to the reference (3000 m/s) condition using F_{760}. Four F_{760} models derived from one-dimensional ground-response simulations for CENA profiles were considered (Atkinson 2012, Boore 2015 [two models], and Grazier 2015). As noted by Goulet et al. (2018), these four F_{760} models are similar, particularly in the 1-10 Hz frequency range.

To investigate the potential impact of the different site amplification models, we consider differences between SS14 combined with Atkinson (2012) for site corrections applied during model development and the Stewart et al. (2020) CENA site response model used in forward applications (e.g., Petersen et al., 2020, 2023; Moschetti et al., 2024). The assumption made during NGA-East GMM development can be viewed as taking the natural-log mean motion for a given V_{S30} ($\mu_{ln,k}^{dev}(V_{S30})$) as,

$$\mu_{ln,k}^{dev}(V_{S30}) = \mu_{ln,k}(3000) + (F_V^{SS14} + F_{760}^{*}) \tag{3.7}$$

where $\mu_{ln,k}(3000)$ is the reference site GMM and F_V^{SS14} and F_{760}^{*} are the two components of the site amplification model (the * superscript for F_{760} indicates that multiple alternative models could be used, although Atkinson [2012] will be used here for demonstration purposes). In contrast, the mean model as applied in the 2018 and 2023 NSHMs is,

$$\mu_{ln,k}^{app}(V_{S30}) = \mu_{ln,k}(3000) + (F_V^{Sea20} + F_{760}^{Sea20}) \tag{3.8}$$

where F_V^{Sea20} and F_{760}^{Sea20} are the Stewart et al. (2020) (Sea20) site amplifications. The difference in predicted mean ground motions produced by the different site amplification models can be taken by subtracting Eq. 3.8 from Eq. 3.7 as shown in Eq. 3.9,

$$\delta\mu_{ln,k} = \mu_{ln,k}^{dev}(V_{S30}) - \mu_{ln,k}^{app}(V_{S30}) = (F_V^{SS14} + F_{760}^{*}) - (F_V^{Sea20} + F_{760}^{Sea20}) \tag{3.9}$$

To understand the linkage between the difference in Eq. 3.9 with the residuals in Eq. 3.3, it is useful to recognize that (*i*) the central tendency of the residuals, by definition, is c_k and (*ii*) the central tendency of the NGA-East data is $\mu_{ln,k}^{dev}(V_{S30})$, because in Eq. 3.7 $\mu_{ln,k}(3000)$ is fit to the data using the F_V^{SS14} and F_{760}^{*} models. Accordingly, Eq. 3.3 can be re-written as

$$c_k \cong \bar{\mu}^{dev}_{ln,k}(V_{S30}) - \left[\bar{\mu}_{ln,k}(3000) + \bar{F}_V^{Sea20} + \bar{F}_{760}^{Sea20}\right] \tag{3.10}$$

where the overbars represent means across the data population.

By substituting $\bar{\mu}_{ln,k}(3000) = \bar{\mu}^{dev}_{ln,k}(V_{S30}) - (\bar{F}_V^{SS14} + \bar{F}_{760}^{*})$ from re-arrangement of Eq 3.7 into Eq 3.10 we obtain:

$$c_k \cong \bar{\mu}^{dev}_{ln,k}(V_{S30}) - \left[\bar{\mu}^{dev}_{ln,k}(V_{S30}) - (\bar{F}_V^{SS14} + \bar{F}_{760}^{*}) + F_V^{Sea20} + F_{760}^{Sea20}\right] = \delta\mu_{ln,k} \tag{3.11}$$

Accordingly, a potentially reasonable hypothesis is that the mean misfits evaluated from residuals in this study may be influenced by the differences between the site amplification models.

Figures 3.14a and b show mean values of site amplification for the non-TOK dataset as derived from the two models (TOK is not included for this analysis due to the relatively strong biases in that region); the F_V and F_{760} values shown were obtained by exercising the models for each site and then averaging across sites. Considering first F_V, the amplification applied during model development (SS14) is stronger at all periods, but the differences are most pronounced at long period (about 0.15 natural log units). This difference should cause positive misfits (Eq. 3.11), as observed.

Multiple models for F_{760}^{*} are shown in Figure 3.14b, which are Atkinson (2012), Boore (2015) (two models), and Grazier (2015). The Boore and Grazier models are shown over the period range considered during the integration phase of NGA-East (Goulet et al. 2018), which are 0.1-1 s and 0.2-5 s, respectively, whereas the Atkinson model used for illustration purposes is shown over the full period range. The mean of F_{760}^{*} values over the period range 0.2-1 sec is closest to Atkinson among the four models (generally within 0.1 ln units).

The differences in the Sea20 and Atkinson (2012) F_{760} models in Figures 3.14a and 3.14b are small for $T > 0.4$ sec, but at short periods the Atkinson (2012) factors are much lower (about 0.4 natural log units between periods of 0.05 and 0.1 sec). These short-period differences are caused by distinctly different shapes of the Sea20 and Atkinson (2012) F_{760} models at short periods, likely indicating different implied κ_0 values. The relatively low F_{760} factors from Atkinson (also observed for Boore 2015) produce negative misfits, as observed.

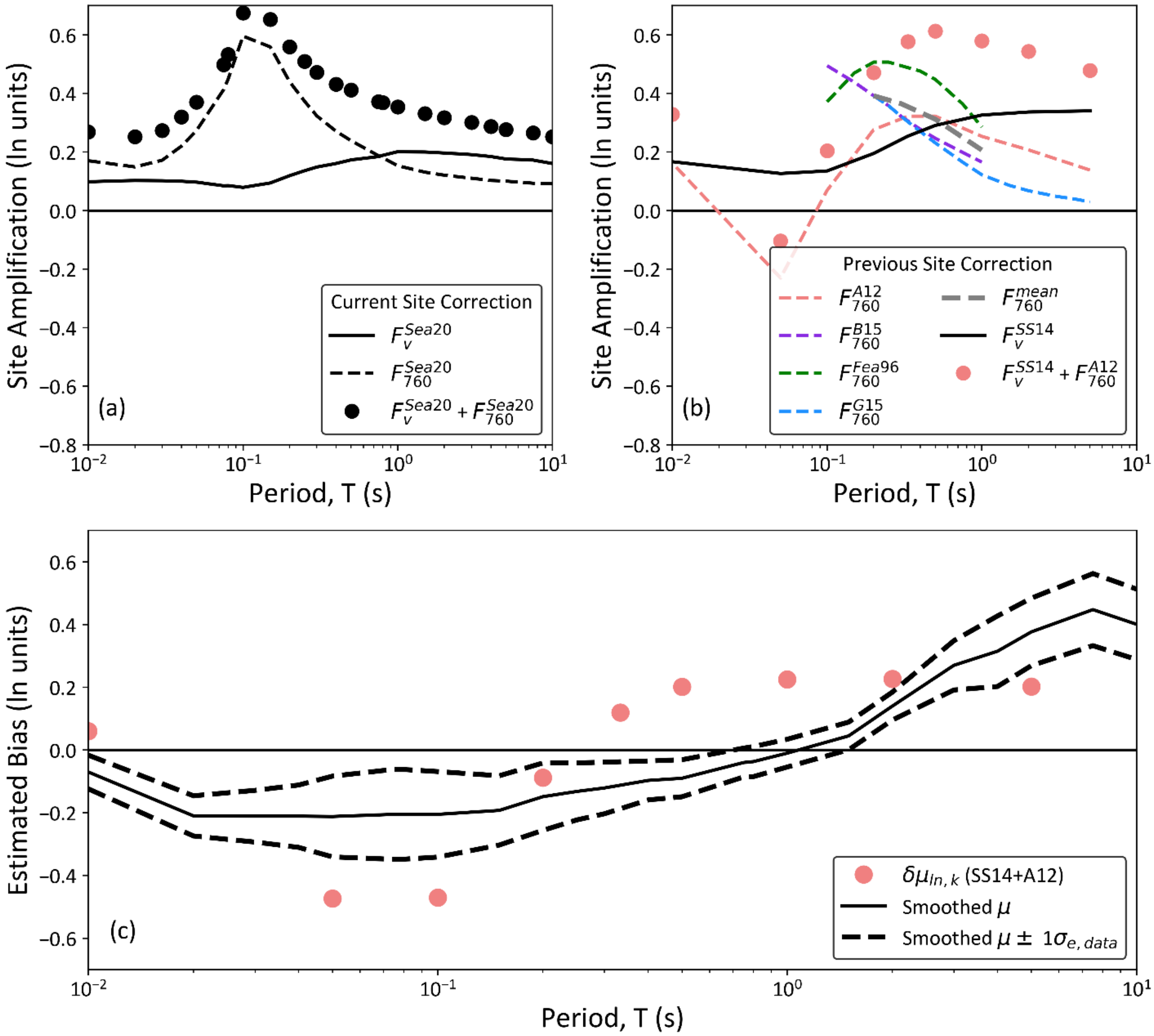


Figure 3.14. (a) Mean site amplification from F_V and F_{760} components of the Stewart et al. (2020) model across all CENA sites in the non-TOK dataset; (b) Mean site amplification from the FV

(SS14) and F_{760} (A12, B15, Fea96, G15) models across all CENA sites in the non-TOK dataset; and (c) comparison of site response differences (Eq. 3.9 and 3.11) to the recommended model adjustment factors. A12 = Atkinson 2012; B15 and Fea96 from Boore 2015; G15 = Grazier 2015; SS14 = Seyhan and Stewart (2014).

Figure 3.14c compares the differenced site corrections (Eq. 3.9) to the model adjustment factors from Figure 3.12 ± $\sigma_{e,data}$. These results demonstrate that the differences in site amplification and the adjustment factors have similar features; for example, the long-period underprediction (positive factors) appears to be influenced by the much stronger long-period amplification for active crustal regions than for stable continental regions (as contained in the F_V terms). Similarly, the short-period overprediction (negative factors) appears to align with differences in F_{760} models, in particular the strong peak in F_{760}^{Sea20} that is absent in the Atkinson (2012) model. However, the positive adjustment factors at long periods are larger than suggested by the model differences, while the negative factors at short periods are smaller than suggested by these model differences.

3.5.2 Modifications to F_{760} for CENA

Since the publication of the Sea20 F_{760} factors, additional simulations of site response for sites with V_{S30} = 760 m/s have been performed by Ilhan et al. (2024) and Chapter 7 of Centella et al. (2023). That work considered additional V_S profiles, additional material damping formulations, and explicit consideration of the range of κ_0 captured in the profiles. Most of the profiles apply for an impedance condition, as defined by Sea20. Figure 3.15 shows how F_{760} factors derived from that work compare to those in Sea20, separated according to whether the profiles represent impedance or gradient conditions. For impedance conditions, the newer results are larger at long periods and smaller at short periods ($T < \sim 0.015 – 0.03$ sec) than the mean factors in Sea20. For gradient conditions, the newer results are generally larger at short periods ($T < 0.25$ sec) and lower at long periods. Since the impedance condition is more typical for firm-ground sites (V_{S30}=760 m/s), the differences between impedance models are more important. If a new F_{760} model were to be developed that reflected these differences for impedance conditions, it would likely reduce the misfits at both long and short periods.

The F_{760} reduction for impedance conditions at short periods is qualitatively consistent with Ktenidou and Abrahamson (2016), who anticipated the potential for overprediction of short-period site response, which they attributed to the NGA-East hard-rock κ_0 = 0.006 sec being too small. However, the amount of short-period misfits are smaller than anticipated by Ktenidou and Abrahamson (2016).

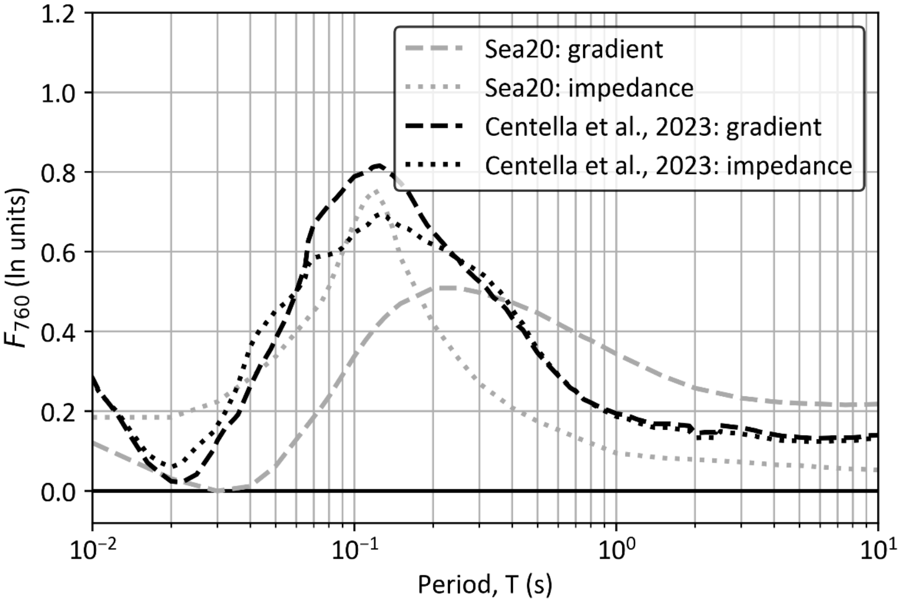


Figure 3.15. Average F_{760} values for CENA sites as derived from a weighted combination of the impedance and gradient models as given by Sea20 compared to F_{760} derived from recent simulations reported in Centella et al. (2023).

3.5.3 Misfit Attribution

The differences between site amplifications used in NGA-East GMM development relative to those used in application produce ground-motion changes that are generally consistent with the period-dependent pattern of the proposed model adjustments (Figure 3.14c). While it is difficult to know how much of the overall misfit can be attributed to this effect, it directly impacts multiple seed models and influenced the model integration process that led to the 17 NGA-East GMMs.

As noted previously, we do not anticipate F_V^{Sea20} as appreciably influencing the observed misfits. Some misfit may be from the F_{760}^{Sea20} model; at long periods the newer V_S profiles used for 760 m/s sites produce larger amplifications than were found previously for impedance conditions (Figure 3.15), which if adopted for applications would reduce but not eliminate the misfits. At short periods, the misfits are small and could easily be accounted for with adjustments to κ_0. The portions of the misfits that cannot be attributed to the F_{760}^{Sea20} model are likely associated with the hard-rock GMMs.

3.6 Conclusions

The USGS National Seismic Hazard Model uses hard-rock reference site GMMs from the NGA-East project (Goulet et al., 2021a) and site amplification models recommended by an expert panel (Stewart et al., 2020; Hashash et al., 2020) to estimate ground motions in CENA. Due to asynchronicity in the development of these models, they are not fully compatible with each other. The adjustment factors presented herein allow for compatibility in the joint application of these models. These factors, in effect, adjust the constant terms in the GMMs; the scaling relations (i.e., changes in ground motions with source, path, and site parameters) are unaffected.

Using an expanded CENA dataset (relative to that used in the NGA-East project), we examine residuals of the recommended GMMs. While expanded, the range of magnitudes in the database remains limited (**M** ~4.0-5.8), hence our focus was mainly to assess mean model misfit within the range of the data, rather than scaling relationships that are also important for hazard applications. Using the central branch GMM, these residuals analyses indicate that for data outside of the TOK region, there is no evidence for bias in the magnitude- and distance-scaling components of the GMM for $\mathbf{M} \geq 4$ events and $R_{rup} \leq 600$ km, with the exception of faster attenuation in coastal plain regions that manifests at distances > 300 km. However, persistent period-dependent mean misfits were observed from the data-to-model comparisons for a wide range of alternate NGA-East GMMs and alternate data selection criteria (i.e., excluding data from particular regions). These misfits are towards overprediction at short periods and underprediction at long periods. Different levels of mean misfits were found for the TOK region, the coastal plain region, and the remainder of CENA, with the TOK region having the most distinct characteristics. Model adjustment factors were derived in a manner that accounts for these regional differences by sampling data from the three regions in different ways, which contributes to epistemic uncertainty.

We anticipate that the misfits, which form the basis for proposed model adjustment factors, are associated with both the hard-rock GMMs and the F_{760}^{Sea20} factors, although the breakdown of relative contributions among these models is uncertain. For forward applications for commonly encountered site conditions in the range of the F_{V}^{Sea20} model (200 to 2000 m/s), whether the misfit arises from the hard-rock GMMs or from the F_{760}^{Sea20} factors is inconsequential. For these sites, we recommend applying the model adjustment factors and their uncertainties (Table 3.1), including the stiff site modification factors (Eq. 3.6) as applicable, to the sum of the hard-rock GMM and site response. The levels of uncertainty to be used depend on how the probabilistic seismic hazard analysis (PSHA) is conducted, as follows:

1. For PSHA in which all 17 NGA-East GMMs are used (such as in the 2018 and 2023 NSHM), the smoothed model adjustment factors should be used with the epistemic uncertainty given in Table 3.1 as $\sigma_{e,data}$. This is the preferred approach because the epistemic uncertainties contained within the 17 GMMs are preserved.
2. For PSHA in which only the central branch GMM is used, the smoothed model adjustment factors should be used with the uncertainty given in Table 3.1 as σ_e.

For applications where nonlinear site response is expected (i.e., the F_{nl} term in Eq. 3.1 is non-zero), the PGA term that drives the nonlinearity should be modified by the PGA adjustment factor.

For applications in which only the hard-rock GMMs are to be applied, contributions to the bias from the site factors (F_S) should be removed, which we anticipate to be solely related to F_{760}^{Sea20}. The amount of this adjustment is uncertain, and will depend on how knowledge of F_{760} evolves in future work. Increases in F_{760} at long periods, and decreases at short periods, which is possible based on recent results shown in Figure 3.15, would reduce misfits. For these hard-rock applications, we suggest the use of a logic tree in which different fractions of the smoothed model adjustment factors (μ in Table 3.1), including the stiff site modification, are attributed to the hard-rock GMM. Logic-tree branches in which the full adjustments, various percentages of the full adjustments, and no adjustments are applied are recommended. Weights given to these branches would be guided by the degree to which the misfits are believed to be attributed to the F_{760}^{Sea20} model, particularly for impedance conditions.

3.7 Data Resources

Raw ground-motion recordings were retrieved from International Federation of Digital Seismograph Networks (FDSN) data centers of Incorporated Research Institutions for Seismology (IRIS) Data Services using gmprocess (Hearne et al., 2019). IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR-1851048.

The time series and metadata used in this study are archived in a publicly web-serviced ground-motion relational database (Buckreis et al., 2023). The ground-motion database uses MySQL (by Oracle Corporation, http://www.mysql.com/) as the management system, and an application programming interface (API) was written to facilitate queries using URLs (https://uclageo.com/gm_database/api/index.php, last visited May 23, 2023). Ground-motion records from NGA-East and newly added records are stored in the database under the “collection_id” = 3 and “user_id” = 2 respectively.

The specific data resources applied in this project are provided as source, site, and ground-motion flatfile tables by Ramos-Sepulveda et al. (2023b).

4. Parametrization of the NGA-East GMMs

The NGA-East GMMs are a collection of tables where the user obtains Sa values based on period, rupture distance and magnitude. There is a total of 17 individual GMMs and a central branch (weighted mean). Since these models are tables and have no equation associated with them, when a trend in residuals is encountered, there are no scaling parameters that can be adjusted to remove the trend. For example, consider the distance attenuation trends shown in Section 3.4.2 – while the downward slope of within-event residuals is clearly a consequence of faster anelastic attenuation, there is no such parameter in the model to adjust to account for this effect.

Therefore, in order to facilitate investigations of regional path effects, which will be presented in Chapter 5, in this chapter I focus on finding a functional form that can fit the mean *IM* values from the NGA-East tables with appropriately selected scaling coefficients. Because the intended use of the equation-based models is parameter adjustments to remove residuals trends, my priority in selecting a functional form and deriving coefficients is to optimize the match to NGA-East median models over the parameter range of the data presented in Chapter 2. Since a similar fitting exercise was performed using a point source omega-square model by Pezeshk et al. (2024), I will compare my results to those presented in the Pezeshk et al. study.

4.1 Functional Form

The functional form I used for parametrizing the GMM was inspired by one of the seed models used in the development of the NGA-East GMMs, Yenier and Atkinson (2015, YA15 hereafter). YA15 is an equivalent point-source simulation model whose parameters for magnitude, distance, stress parameter, geometrical spreading rate, and anelastic attenuation have been calibrated to empirical data in California and adjusted to for use in CENA:

$$ln\left(\hat{Y}(\mathbf{M}, R_{rup})\right) = F_E(\mathbf{M}) + F_P(R_{rup}, \mathbf{M}) \tag{4.1}$$

where $\hat{Y}$ is the median value of an *IM* for an event with magnitude $\mathbf{M}$ and rupture distance R_{rup}, $F_E(\mathbf{M})$ is a model for source effects and $F_P(R_{rup}, \mathbf{M})$ is a model for path effects. Eq. (4.1) does not have a site term since the GMM is for a single, reference site condition. The source term from YA15 has the following functional form:

$$F_E = \begin{cases} e_0 + e_1(\mathbf{M} - \mathbf{M}_\mathrm{h}) + e_2(\mathbf{M} - \mathbf{M}_\mathrm{h})^2 & \mathbf{M} \leq \mathbf{M}_\mathrm{h} \\ e_0 + e_3(\mathbf{M} - \mathbf{M}_\mathrm{h}) & \mathbf{M} > \mathbf{M}_\mathrm{h} \end{cases} \tag{4.2}$$

where the first term represents a parabolic trend with magnitude below a hinge magnitude $\mathbf{M}_h$ and the second term represents a linear trend above the hinge magnitude. The hinge magnitudes used in this study vary from YA15. For each period, I examined the trends of Sa versus magnitude to find the change in slope (i.e., magnitude saturation). The path term has the following functional form:

$$F_P = F_Z + F_\gamma \tag{4.3}$$

where F_Z depicts the geometric spreading component and F_γ represents the anelastic attenuation term. Each component can be defined by:

$$F_\gamma = \gamma R_{rup} \tag{4.4}$$

$$F_Z = ln(Z) + (b_4 + b_5\mathbf{M})ln\left(\frac{R}{R_{ref}}\right) \tag{4.5}$$

where R and R_{ref} are defined in Eq. 4.6 and Eq. 4.7 respectively. The b_4 and b_5 coefficients are used to capture the magnitude-dependence of geometric spreading associated with oscillator response characteristics and γ captures anelastic attenuation,

The term Z in Eq. (4.5) is intended to model the observed geometric spreading effects for Fourier amplitude spectra. It is formulated as a trilinear function to capture the reflections and refractions off the Moho discontinuity which disrupts amplitude decay in the distance range from 50 to 200 km (e.g., Burger et al., 1987). This effect is also known as the Moho-bounce phenomenon and it is implemented in various CENA models (e.g., Boore, 2003; Campbell, 2003; Atkinson and Boore, 2011).

$$R = \sqrt{R_{rup}^2 + h^2} \tag{4.6}$$

$$R_{ref} = \sqrt{1 + h^2} \tag{4.7}$$

$$h(\mathbf{M}) = h_0 + 10^{h_1 + h_2 \mathbf{M}} \tag{4.8}$$

$$Z = \begin{cases} R^{b_1} & R < R_t \\ R_t^{b_1} \left(\frac{R}{R_t}\right)^{b_2} & R_t \le R < R_{tt} \\ R_t^{b_1} \left(\frac{R_{tt}}{R_t}\right)^{b_2} \left(\frac{R}{R_{tt}}\right)^{b_3} & R \ge R_{tt} \end{cases} \tag{4.9}$$

where Z has the shape of a continuous function that goes from direct-waves at near distances ($R<R_t$) to a transition zone between R_t and R_{tt} to surface-waves spreading at far distances ($R \geq R_{tt}$). The h term in Eq. (4.6) accounts for near-source saturation (e.g., Yenier and Atkinson, 2014) and the best fit was obtained with b_1 = -1.3, R_t = 60km and R_{tt} to 170km.

4.2 NGA-East Central Branch GMM

The fitting process using Eqs. (4.1-4.8) is performed individually for 23 different periods, PGA and PGV. The coefficients are not fit to empirical data, but to the tabulated NGA-East central branch GMM (herein referred to as tabulated GMM). This section describes this fitting process and level of misfit that was obtained. Similar fitting and parametrization of the 17 individual NGA-East GMMs is presented in Section 4.3.

I used the “nlsLM” function from the MINPACK library in R (Elzhov et al., 2015) to fit Eqs (4.1-4.9). The fitting process was undertaken in four steps, which allowed coefficients for different ranges of R_{rup} to be estimated in a way that reduces the trade-offs between coefficients, which is a very common problem in model fitting (e.g., Baltay et al., 2020). Table 4.1 summarizes the iterative fitting process that was used to estimate the coefficients. Figure 4.1 shows fits achieved at different steps for *Sa* at *T* = 0.1sec along with spectral ratios defined as,

$$\ln(SR) = \ln\left(\frac{Sa^{targ}}{\hat{Y}}\right) \tag{4.10}$$

where Sa^{targ} is the NGA-East tabulated *IM*, which serve as the target in the fitting process and $\hat{Y}$ is from Eq. (4.1).

The aim of Step 1, illustrated in Figure 4.1a—b, is to constrain parameters that control $\hat{Y}$ at close distances (R_{rup} < 60 km), including e_0, $h_0 - h_2$, and b_1. Additional parameters are regressed but later replaced in subsequent steps (marked as “temporarily regressed” in Table 4.1) and b_5 is fixed using values from YA15. Step 2 fixes the near-fault controlling parameters from Step 1 as well as b_5, and is used to constrain the intermediate-distance geometric spreading parameter b_2 using only Sa^{targ} values for R_{rup} > 60 km, while other parameters are temporarily regressed. Figures 4.1b-d shows these fits, which are quite accurate ($\ln(SR)$ values of nearly zero). Step 3 is used to constrain b_5, which had been fixed in Steps 1-2 at YA15 values, which required e_0 to be left free. Step 4 then operates similarly to Step 2 but targets large-distance geometric spreading parameter b_3 and produces final values of other parameters. Figure 4.1 e-f shows the resulting fits which have larger spectral ratios than in Figure 4.1b-d; this occurs because of changes in slope in the 60-120 km distance range that are difficult to match.

Table 4.1. Summary of steps used in the regression of the tabulated GMMs.

Parameter	**Step 1** $R_{rup} \leq 60$ **km** $\mathbf{M} \leq 6$	**Step 2** **60 km** $\leq R_{rup} \leq 600$ **km** $\mathbf{M} \leq 6$	**Step 3** $R_{rup} \leq 600$ **km** $\mathbf{M} \leq 6$	**Step 4** $R_{rup} \leq 600$ **km** $\mathbf{M} \leq 6$
e_0	Regressed	Fixed [step 1]	Temporarily regressed	Regressed
e_1, e_2, e_3, b_4, γ	Temporarily regressed	Temporarily regressed	Temporarily regressed	Regressed
b_3	--	Temporarily regressed	Temporarily regressed	Regressed
b_1, h_0, h_1	Regressed	Fixed [step 1]	Fixed [step 1]	Fixed [step 1]
b_2	Temporarily regressed	Regressed	Fixed [step 2]	Fixed [step 2]
b_5	Fixed [YA15]	Fixed [YA15]	Regressed	Fixed [step 3]

For the regressions performed in each step, there are multiple parameters being estimated, which can produce tradeoffs. These tradeoffs are addressed both through the sequenced four-step procedure and also by specifying bounds on the parameters in nlsLM. The fitting process for the parameters controlling h ($h_0 - h_2$) and b_1 were the most challenging. The largest misfits occur in the portion of the parameter space for which the fit is controlled by those parameters, which is R_{rup} < 60 km.

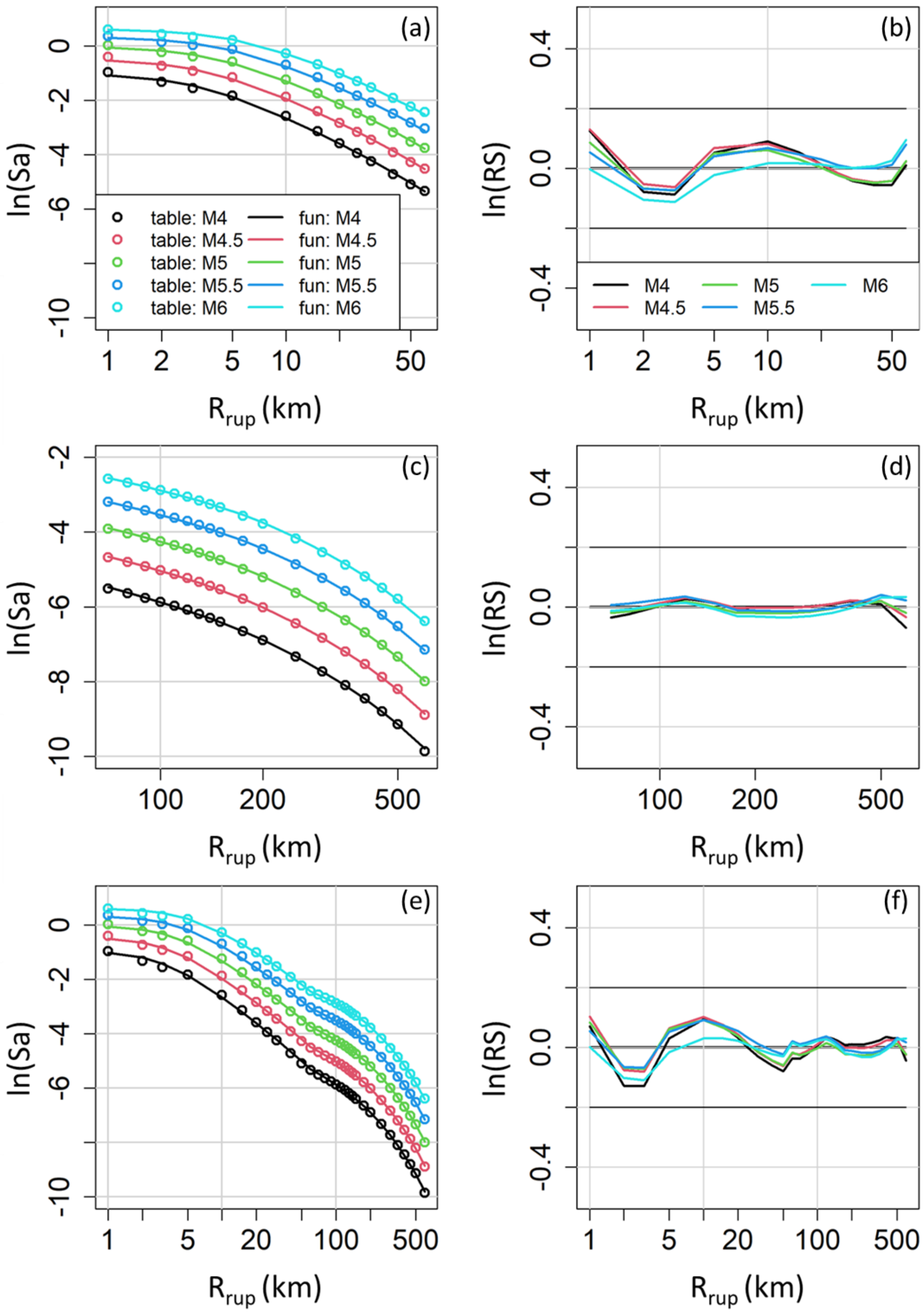


Figure 4.1. Comparisons of $ln(Sa)$ values obtained from the NGA-East tabulated GMM (S_a^{targ}) and the functional GMM ($\hat{Y}$) (labeled in the legend as *table* and *fun* respectively) obtained from Step 1 (a), Step 2 (c), and Step 4 (e). Step 3 results are not shown because they are indiscernible from those from Step 2. Corresponding misfits represented as spectral ratios [ln(RS), Eq. 4.10] are shown in (b), (d), and (f) for Steps 1, 2, and 4, respectively.

After the coefficients were parametrized, I smoothed them with respect to period. The smoothing of the coefficients was initially using moving averages. Coefficients were smoothed using a 3-point window. Since the moving average method is not applicable to the first and last numbers of the array, they were kept as-regressed. GMMs represent median Sa values, therefore, unlike Sa from individual records, the spectral shape should be smooth. Therefore, additional manual modifications were made if the Sa depicted a shape with non-physical bends and corners.

Figure 4.2 shows the coefficients as a function of period compared to the coefficients from YA15. The open symbols represent the coefficients as established by the fitting process, and the solid circles are the smoothed versions. The smoothed version of the coefficients had to be done carefully in order to obtain an acceptable response spectral shape while minimizing the degradation of the model fit (i.e., keeping the differences between Sa^{targ} and $\hat{Y}$ to a minimum). Table 4.1 presents the smoothed coefficients and Figure 4.3 shows the resulting spectral shapes, which indicate that the smoothing was successful because there are no sudden jumps with period or odd shapes.

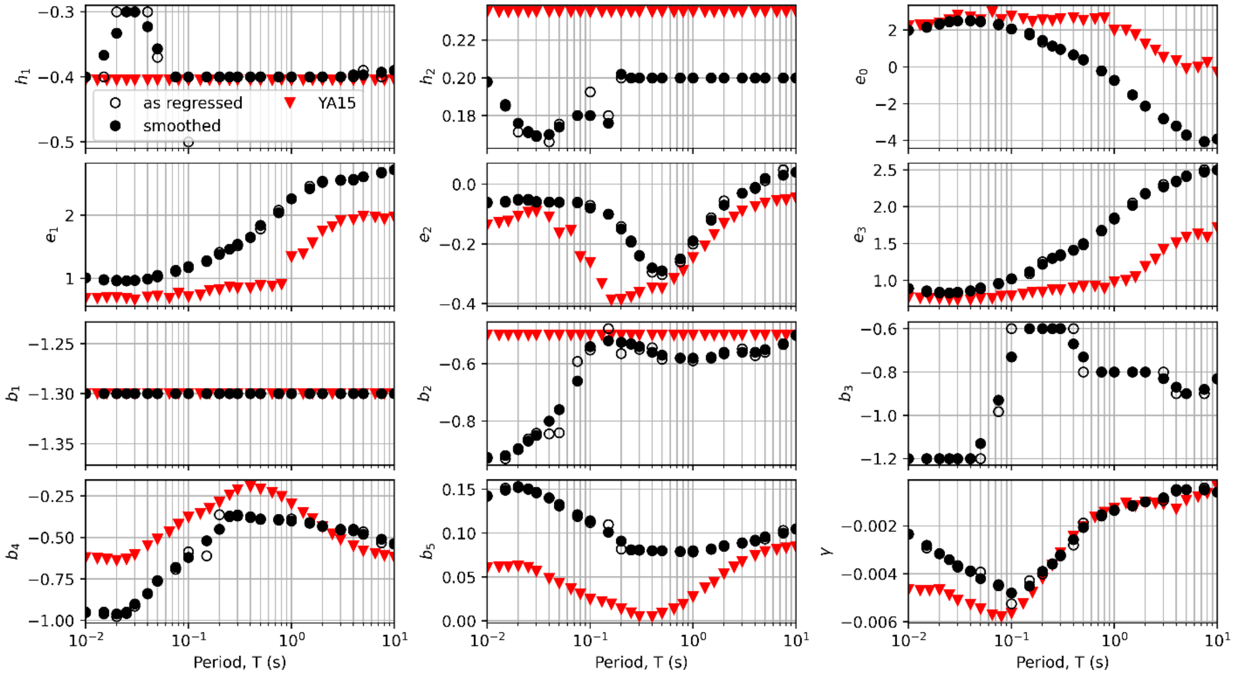


Figure 4.2. Parametrized fit coefficients for model presented in Eqs. (4.1-4.9), using target values from the NGA-East central branch GMM.

Table 4.2. Smoothed coefficients to be used with Eqs. (4.2), (4.4), (4.5), and (4.9).

period	**h_0**	**h_1**	**h_2**	**e_0**	**e_1**	**e_2**	**e_3**	**b_2**	**b_3**	**b_4**	**b_5**	**γ**	**M_h**
PGA	2	-0.400	0.190	1.756	1.04	-0.07	0.91	-0.76	-0.90	-0.91	0.142	-0.00327	5.1
PGV	2	-0.400	0.200	4.579	1.60	-0.06	1.43	-0.72	-0.90	-0.78	0.140	-0.00199	5.1
0.010	2	-0.400	0.198	1.990	1.01	-0.06	0.89	-0.93	-1.20	-0.95	0.142	-0.00235	5.1
0.015	2	-0.367	0.185	2.160	0.98	-0.06	0.86	-0.92	-1.20	-0.96	0.149	-0.00281	5.1
0.020	2	-0.333	0.176	2.320	0.97	-0.05	0.84	-0.90	-1.20	-0.96	0.152	-0.00317	5.1
0.025	2	-0.300	0.171	2.440	0.96	-0.05	0.83	-0.87	-1.20	-0.95	0.150	-0.00341	5.1
0.030	2	-0.300	0.169	2.500	0.97	-0.06	0.84	-0.85	-1.20	-0.90	0.146	-0.00372	5.1
0.040	2	-0.323	0.170	2.510	0.99	-0.06	0.86	-0.80	-1.20	-0.84	0.140	-0.00390	5.1
0.050	2	-0.357	0.174	2.480	1.05	-0.06	0.90	-0.76	-1.13	-0.76	0.131	-0.00420	5.1
0.075	2	-0.400	0.180	2.280	1.11	-0.06	0.96	-0.66	-0.93	-0.68	0.121	-0.00450	5.1
0.10	2	-0.400	0.180	2.070	1.17	-0.08	1.02	-0.54	-0.73	-0.62	0.114	-0.00480	5.1
0.15	2	-0.400	0.176	1.750	1.28	-0.10	1.12	-0.52	-0.60	-0.52	0.101	-0.00451	5.1
0.20	0	-0.400	0.202	1.440	1.38	-0.15	1.22	-0.53	-0.60	-0.45	0.091	-0.00390	5.1
0.25	0	-0.400	0.200	1.140	1.46	-0.19	1.30	-0.53	-0.60	-0.37	0.081	-0.00360	5.1
0.30	0	-0.400	0.200	0.950	1.54	-0.24	1.35	-0.54	-0.60	-0.37	0.081	-0.00321	5.1
0.40	0	-0.400	0.200	0.660	1.65	-0.28	1.41	-0.56	-0.67	-0.38	0.080	-0.00258	5.1
0.50	0	-0.400	0.200	0.400	1.84	-0.29	1.50	-0.57	-0.73	-0.39	0.080	-0.00207	5.1
0.75	0	-0.400	0.200	-0.220	2.04	-0.25	1.67	-0.58	-0.80	-0.39	0.079	-0.00159	5.1
1.0	0	-0.400	0.200	-0.730	2.26	-0.19	1.85	-0.58	-0.80	-0.40	0.080	-0.00135	5.1
1.5	0	-0.400	0.200	-1.520	2.42	-0.12	2.02	-0.58	-0.80	-0.41	0.082	-0.00117	5.1
2.0	0	-0.400	0.200	-2.120	2.52	-0.07	2.18	-0.56	-0.80	-0.43	0.085	-0.00100	5.1
3.0	0	-0.400	0.200	-2.820	2.55	-0.03	2.27	-0.56	-0.83	-0.45	0.089	-0.00083	5.2
4.0	0	-0.397	0.200	-3.200	2.57	-0.01	2.35	-0.56	-0.87	-0.46	0.091	-0.00049	5.3
5.0	0	-0.397	0.200	-3.700	2.61	0.02	2.42	-0.55	-0.90	-0.48	0.096	-0.00049	5.3
7.5	0	-0.393	0.200	-4.050	2.67	0.03	2.47	-0.53	-0.88	-0.51	0.100	-0.00050	5.5
10.0	0	-0.390	0.200	-3.910	2.72	0.04	2.50	-0.50	-0.83	-0.54	0.105	-0.00060	5.8

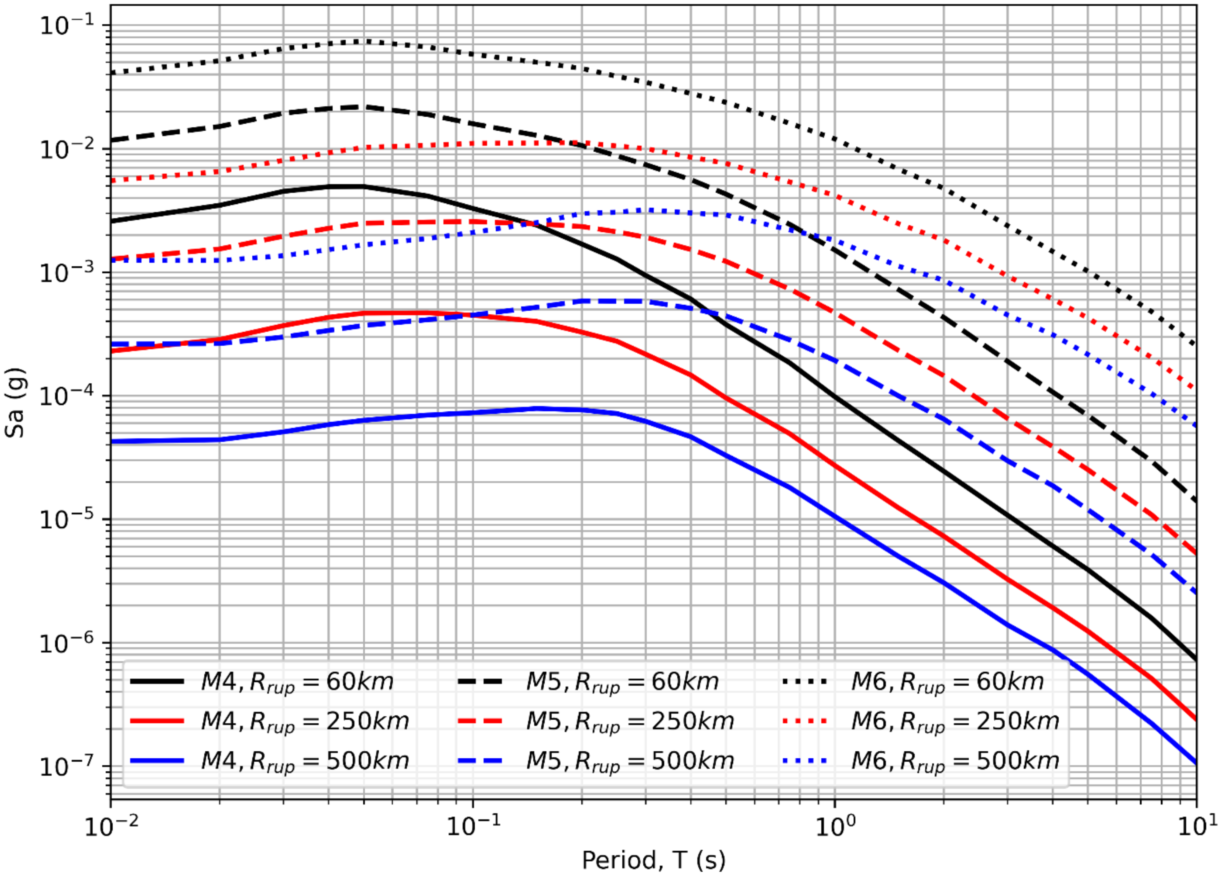


Figure 4.3. Response spectra for an array of magnitudes and rupture distances using Eqs. (4.1-4.9) with smoothed coefficients. No sudden jumps with period or odd shape confirms an adequate level of smoothing was applied.

Figure 4.4 shows spectral ratios versus distance for **M**4-6 earthquakes; the results are similar to Figure 4.1f but include additional *IM*s (the results do not exactly match those from Figure 4.1f at 0.1 sec due to smoothing of coefficients that affect the results in Figure 4.4). Figure 4.5 shows spectral ratios versus **M**. The misfits shown in Figures 4.4-4.5 are generally 0.1 (natural log units) or less and average approximately zero. Figure 4.6 shows similar misfit plots as a function of frequency from the Pezeshk et al. (2024) paper, for which misfits generally range from -0.15 to 0.15 (natural log units), which is comparable to the misfits developed in the present work (Figures 4.4-4.5).

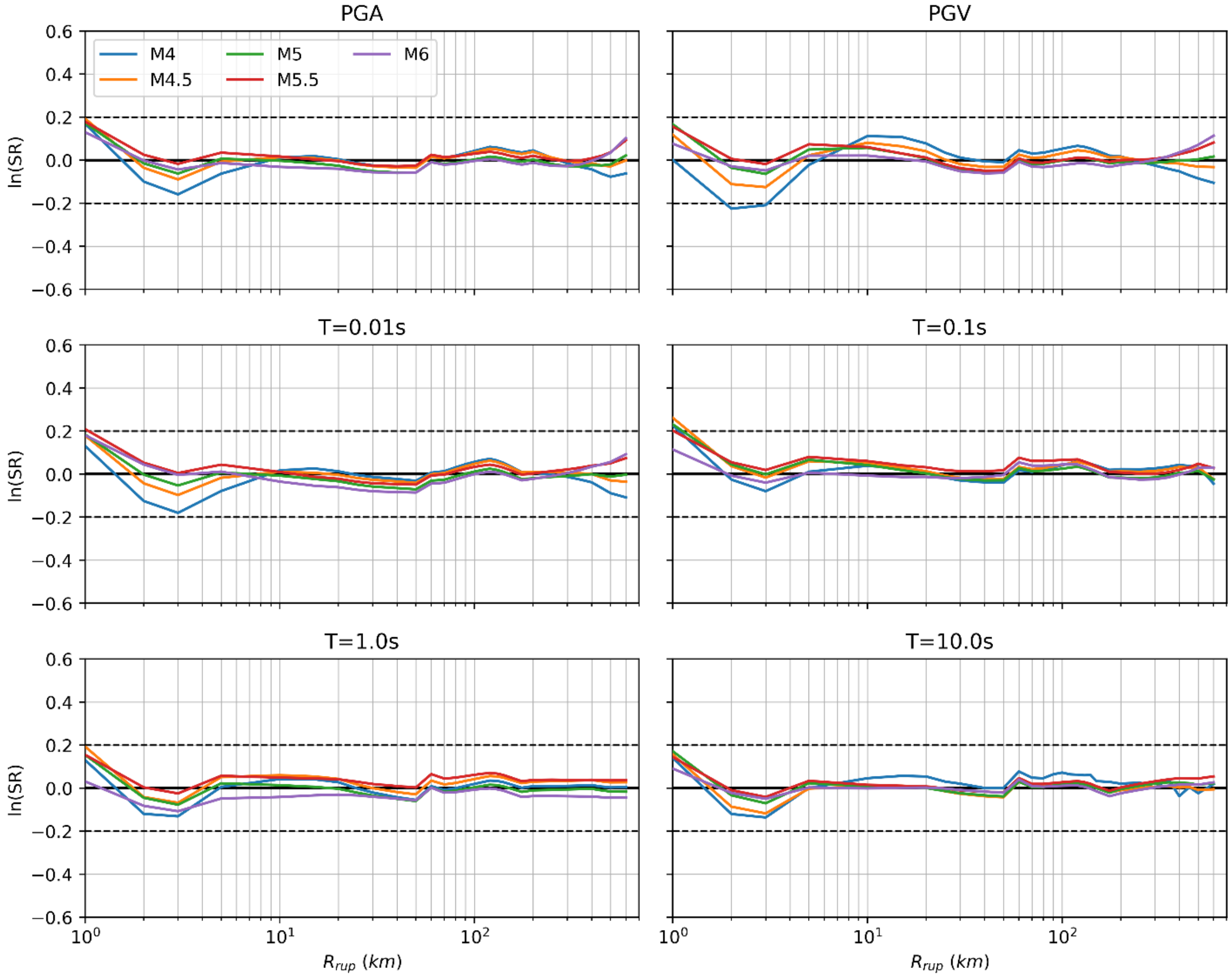


Figure 4.4. Distance-dependence of spectral ratios (Eq. 4.10) indicating misfit of equation-based model relative to tabulated NGA-East central branch GMM.

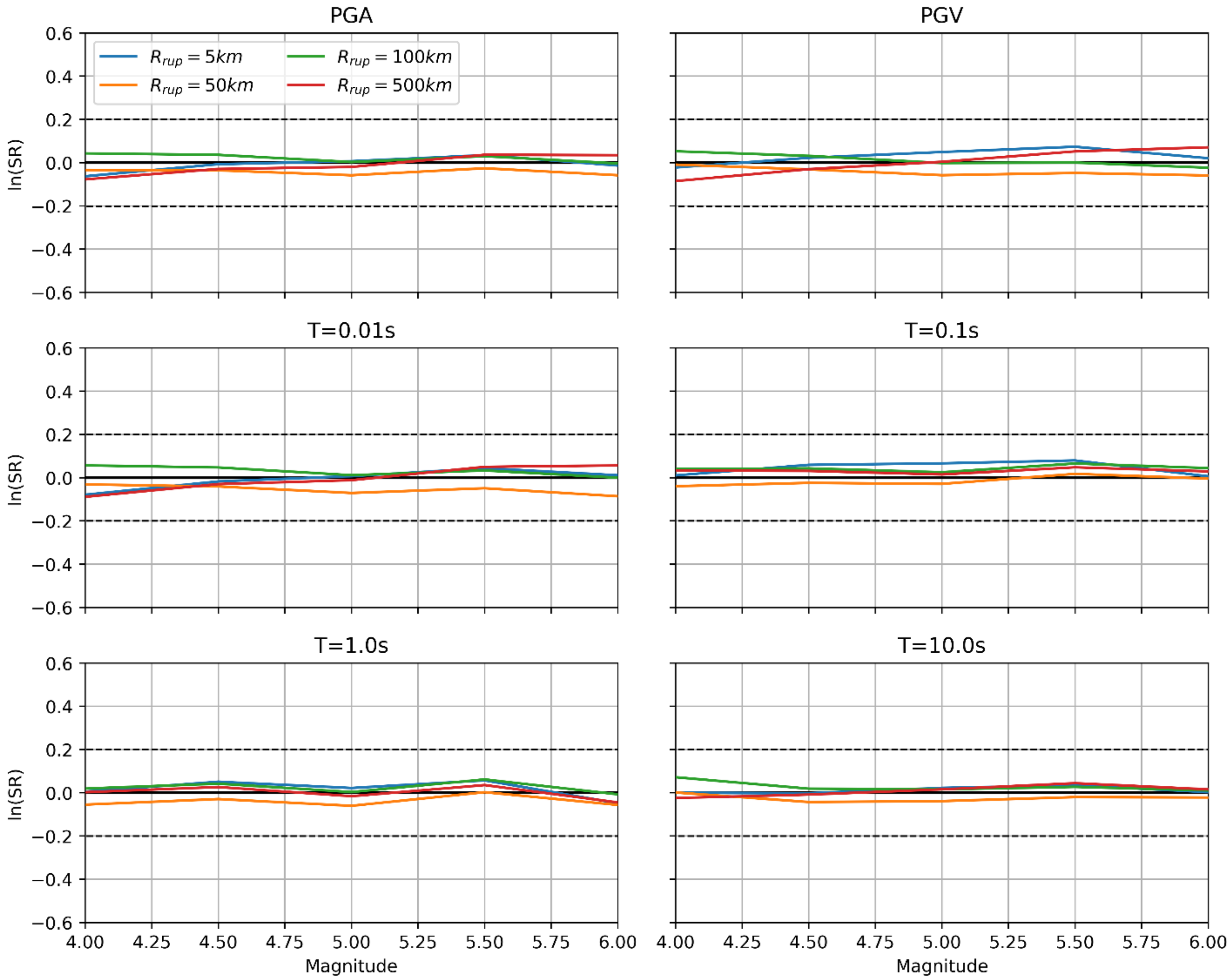


Figure 4.5. Magnitude-dependence of spectral ratios (Eq. 4.10) indicating misfit of equation-based model relative to tabulated NGA-East central branch GMM.

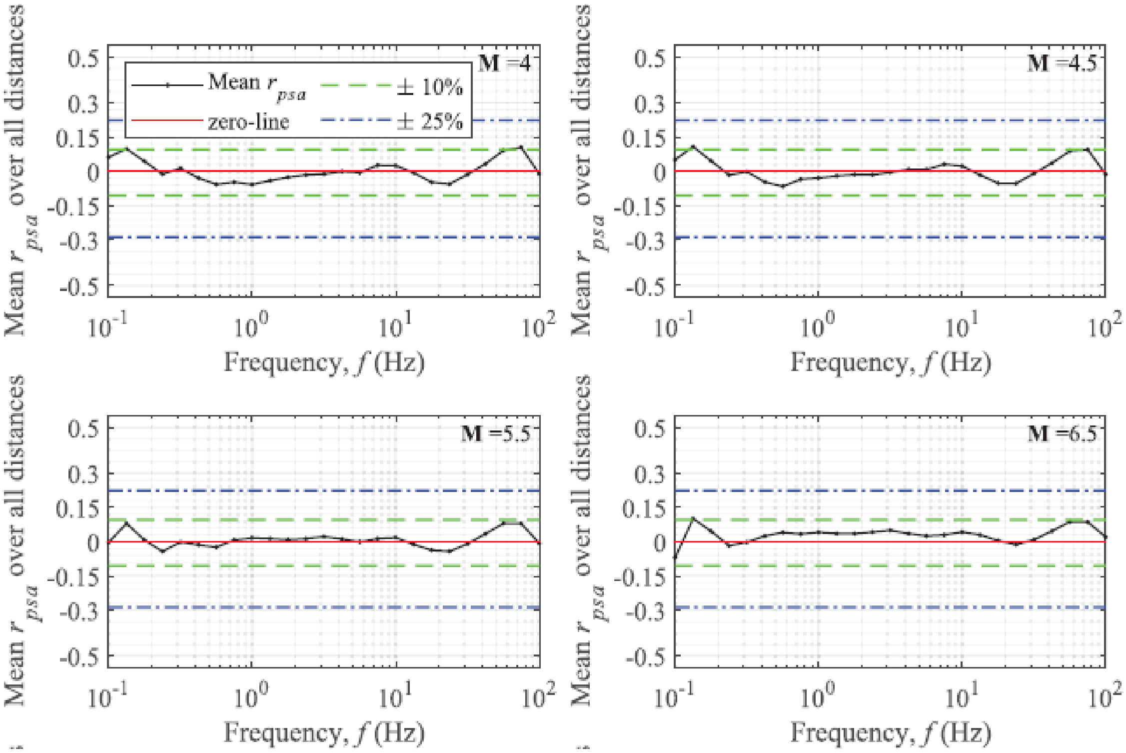


Figure 4.6 Distance-dependence of misfits of omega-squared model representation of tabulated NGA-East central branch GMM (Pezeshk et al., 2024)

4.3 NGA-East 17 Individual GMMs

The same procedure described in Section 4.2 for application to the central branch GMM was repeated for the 17 individual NGA-East GMMs. This was done to demonstrate that the functional form could be used to represent all 17 models and to gain insight into how the 17 models vary among each other in terms of parametric changes.

Figures 4.7 to 4.10 show spectral ratios versus distance for **M**4-6 earthquakes for PGA, PGV, and *Sa* at 0.1 sec and 1.0 sec, respectively. The levels of misfit vary across the GMMs, with the largest misfits being for $R_{rup} < \sim 8$ km and $R_{rup} > \sim 400$ km. For distances between those limits, the parametrizations of the GMMs are considered to be acceptable (i.e., mostly in the range of ±10% and none of the fits have systematic misfits across the full distance range).

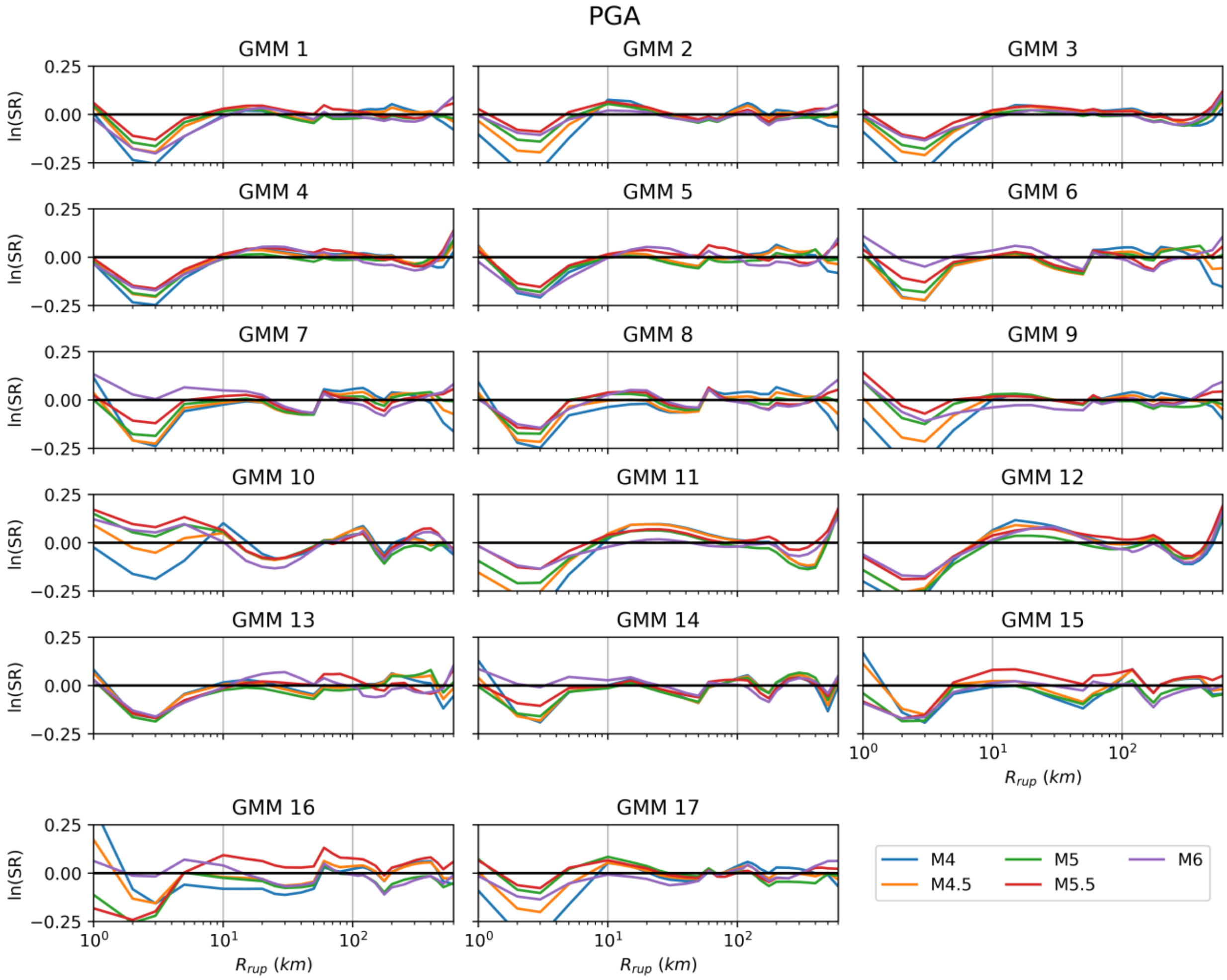


Figure 4.7. Distance-dependence of spectral ratios (Eq. 4.10) for PGA indicating misfit of equation-based model relative to tabulated NGA-East central branch GMM.

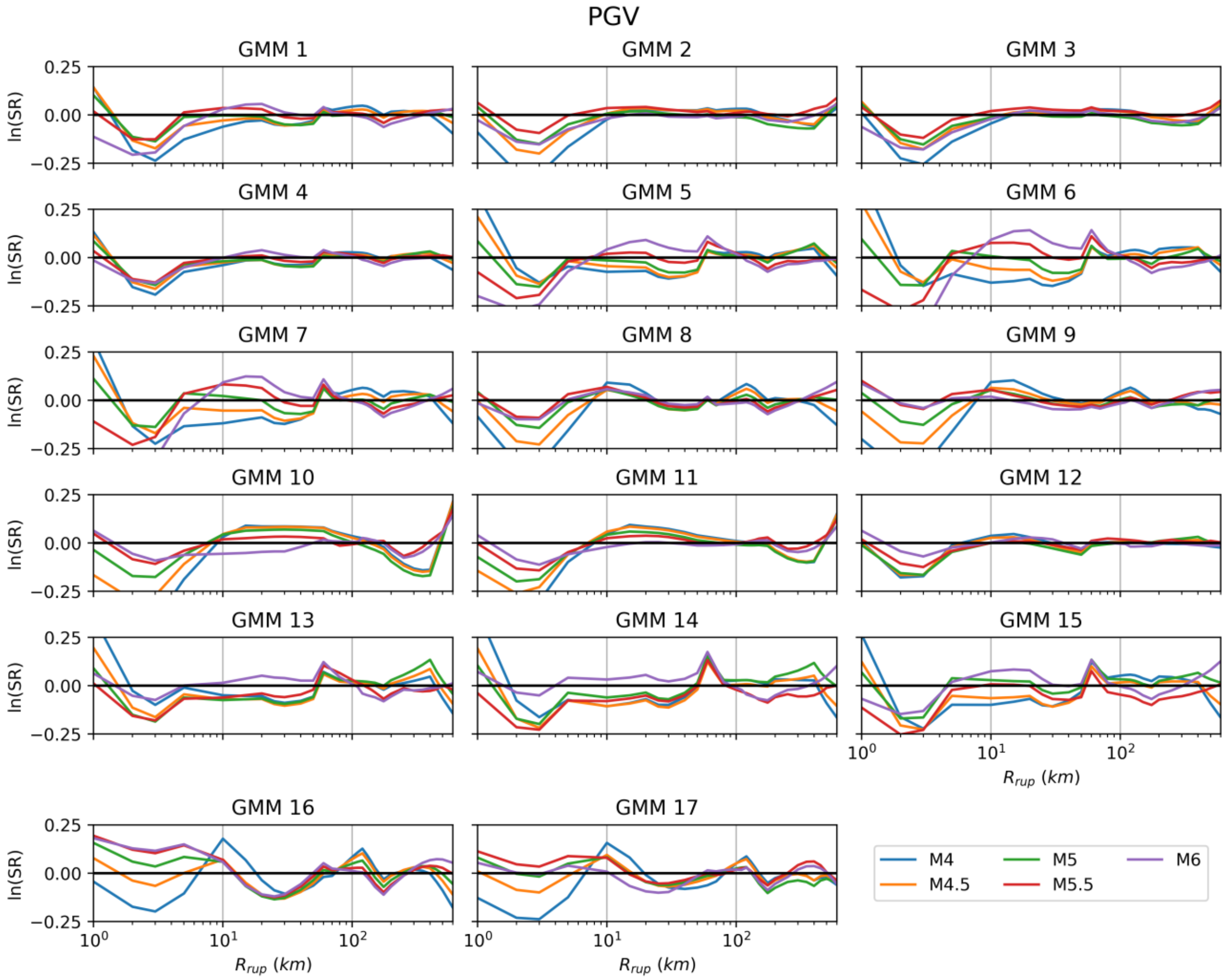


Figure 4.8. Distance-dependence of spectral ratios (Eq. 4.10) for PGV indicating misfit of equation-based model relative to tabulated NGA-East central branch GMM.

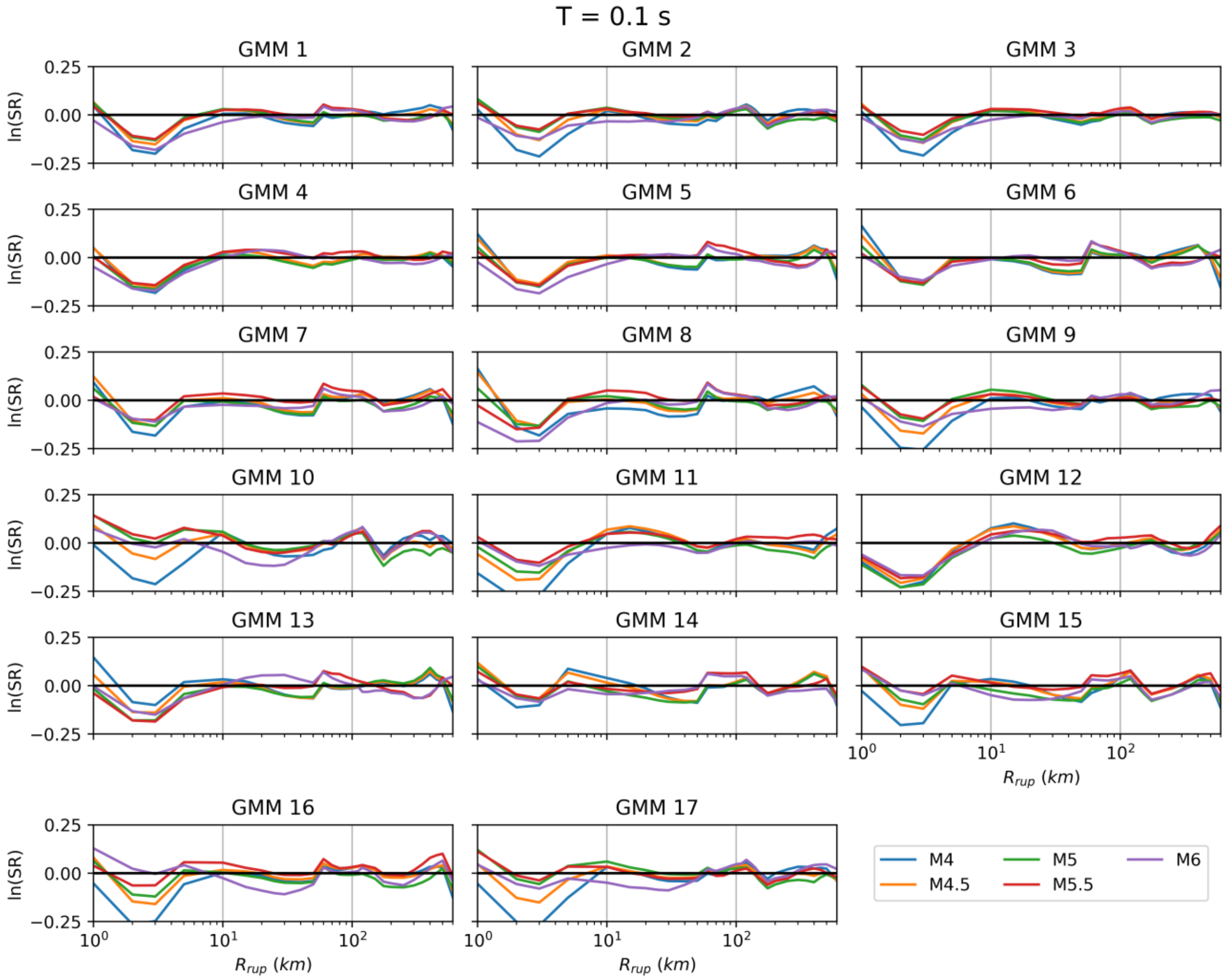


Figure 4.9. Distance-dependence of spectral ratios (Eq. 4.10) for *Sa* (0.1 sec) indicating misfit of equation-based model relative to tabulated NGA-East central branch GMM.

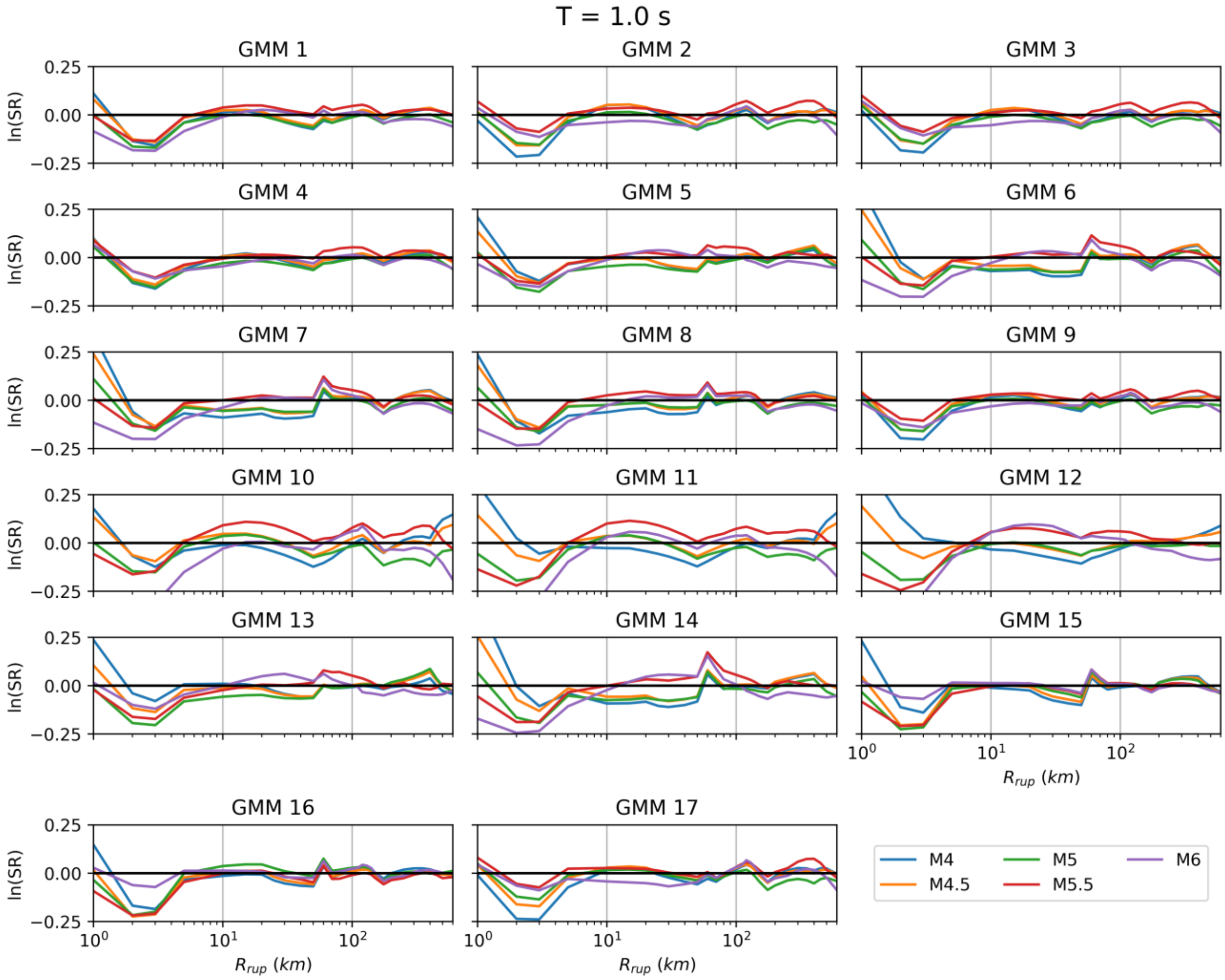


Figure 4.10. Distance-dependence of spectral ratios (Eq. 4.10) for *Sa* (0.1 sec) indicating misfit of equation-based model relative to tabulated NGA-East central branch GMM.

Smoothing of the coefficients was not applied for the 17 GMMs. The rationale for this is that I did not seek to customize these models for regional path effects in order to preserve the levels of epistemic uncertainties provided by the GMMs (customizing the path effect for the dataset in this project would have compromised that feature of the 17 models).

A Sammon's map (Sammon 1969) is used to visualize how the 17 models vary among each other in the parameterization provided by Eq. (4.1-4.10), which is also the method used in NGA-East (Goulet et al., 2021a). Figures 4.11 through 4.14 show the parametrized coefficients color-coded by location in the Sammons map developed for the NGA-East models by Goulet et al. (2021a). The map is centered at GMM 1 (closest model to the central branch) and different azimuths are intended to capture different types of ground motion model variations. Figure 4.11 demonstrates ground motion variation along a horizontal azimuth, which corresponds to changes in the constant term (e_0 in Eq. 4.2). Models with higher constants than the central branch GMM are located to the right of the map and those with smaller constants are located to the left.

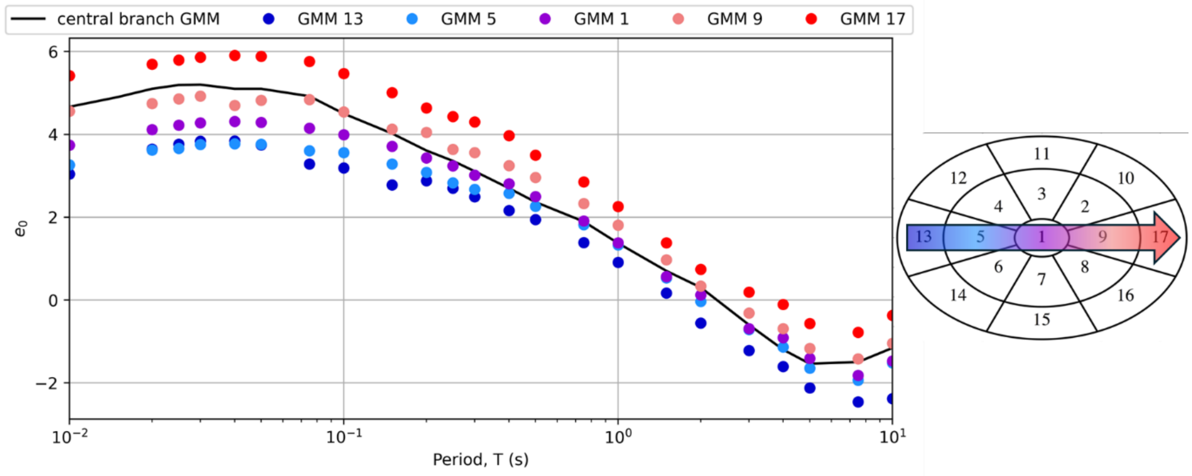


Figure 4.11. Variations of constant coefficient for selected NGA-East GMMs along a horizontal azimuth in the NGA-East Sammons map.

Figure 4.12 demonstrates differences in ground motion scaling for two azimuths across the Sammons map. In the upper row of plots, the azimuth is pointing to the “southeast” and the coefficients that are shown are those associated with magnitude-scaling (Eq. 4.2). The coefficients vary strongly along the azimuth in these plots. In the lower row of plots, the azimuth is pointing towards the “southwest” and the variations of the coefficients are relatively modest. This indicates that the southeast azimuth demonstrates magnitude-scaling features, whereas the southwest azimuth does not. Figure 4.13 shows the variations of path model parameters for the southwest azimuth (Eqs. 4.4-4.9). The large variations of the parameters along this azimuth indicate that it demonstrates ground motion variations with distance. Confirmation that the southeast azimuth does not represent distance-scaling is provided by Figure 4.14. In this figure the variations of distance-scaling parameters along the southeast azimuth are shown to be minimal.

In summary, the southeast azimuth represents magnitude scaling, with stronger scaling in the SE quadrant and weaker scaling to the NW. The southwest azimuth represents distance-scaling, with stronger scaling to the NE and weaker scaling to the SW.

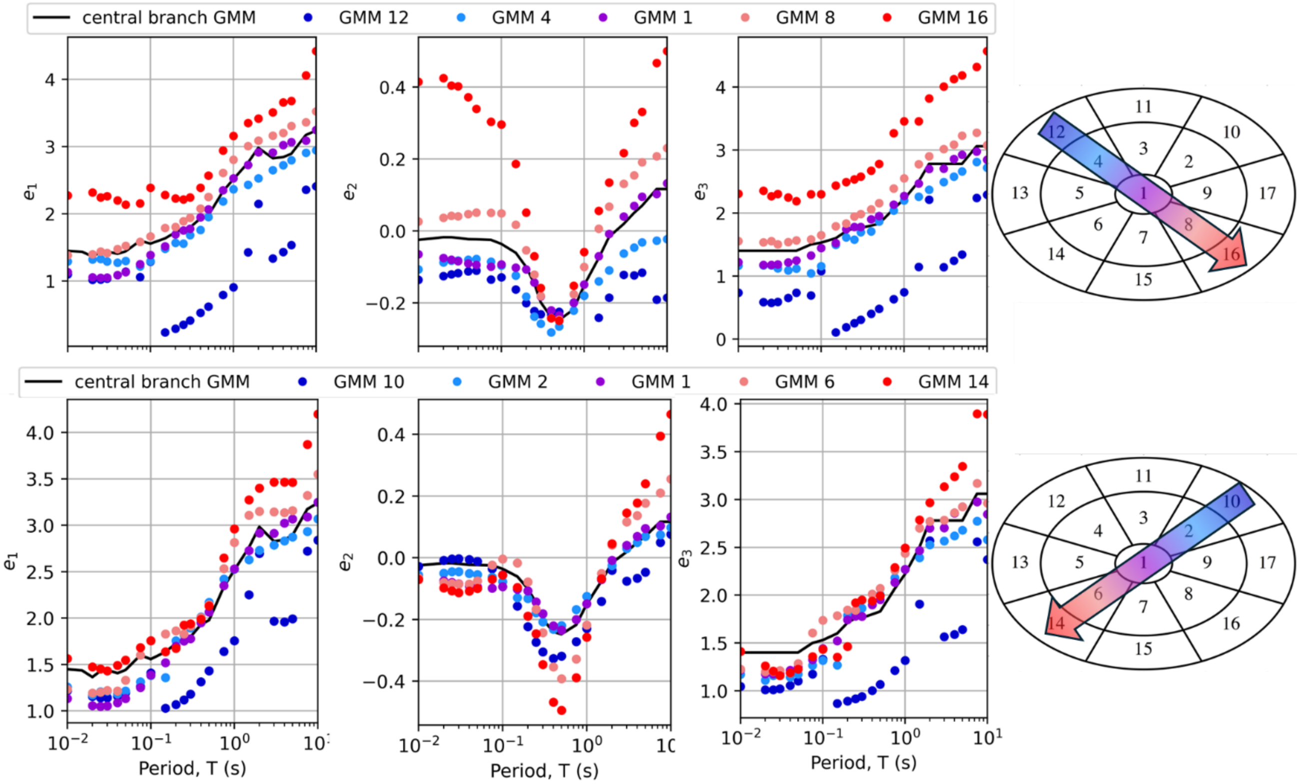


Figure 4.12. Variations of magnitude-scaling coefficients for selected NGA-East GMMs along a southeast-pointing azimuth (top) and southwest-pointing azimuth (bottom) in the NGA-East Sammons map. These results demonstrate the magnitude scaling is associated with the southwest azimuth.

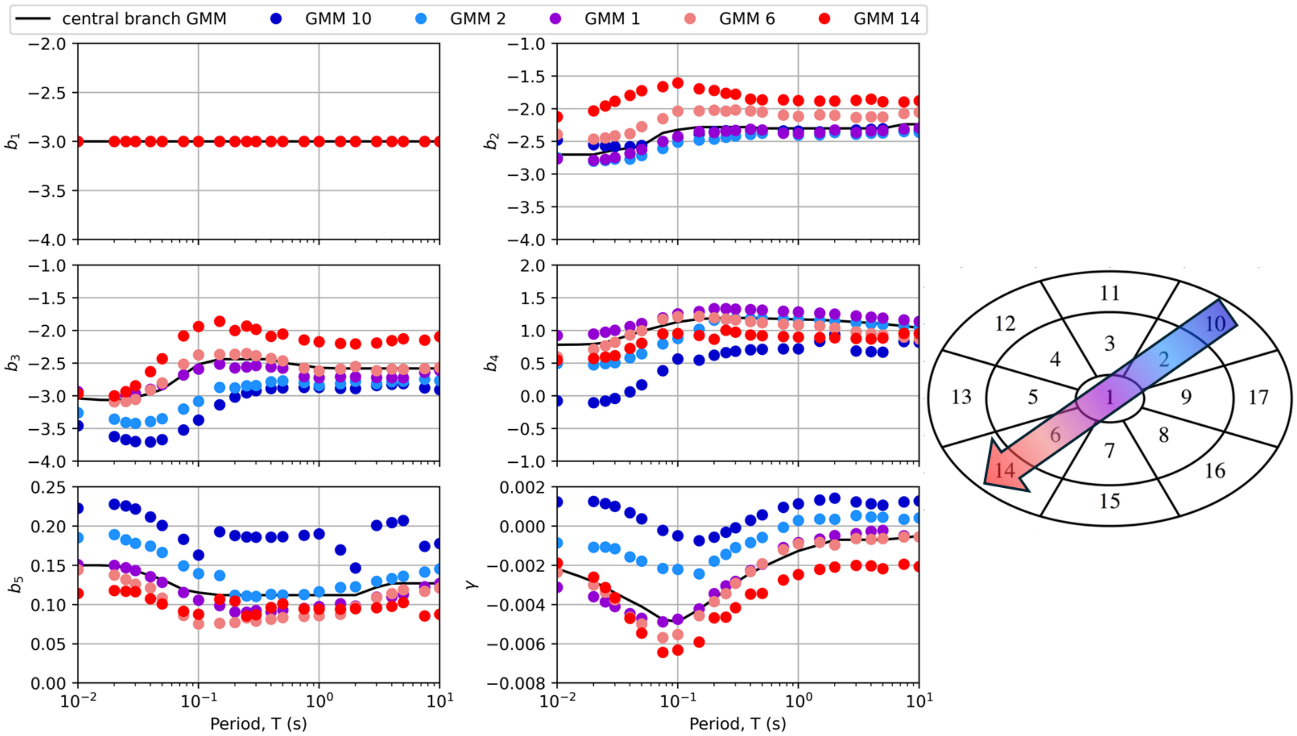


Figure 4.13. Variations of distance-scaling coefficients for selected NGA-East GMMs along a southwest-pointing azimuth in the NGA-East Sammons map. This azimuth reflects the widest difference in distance-scaling coefficients across GMMs.

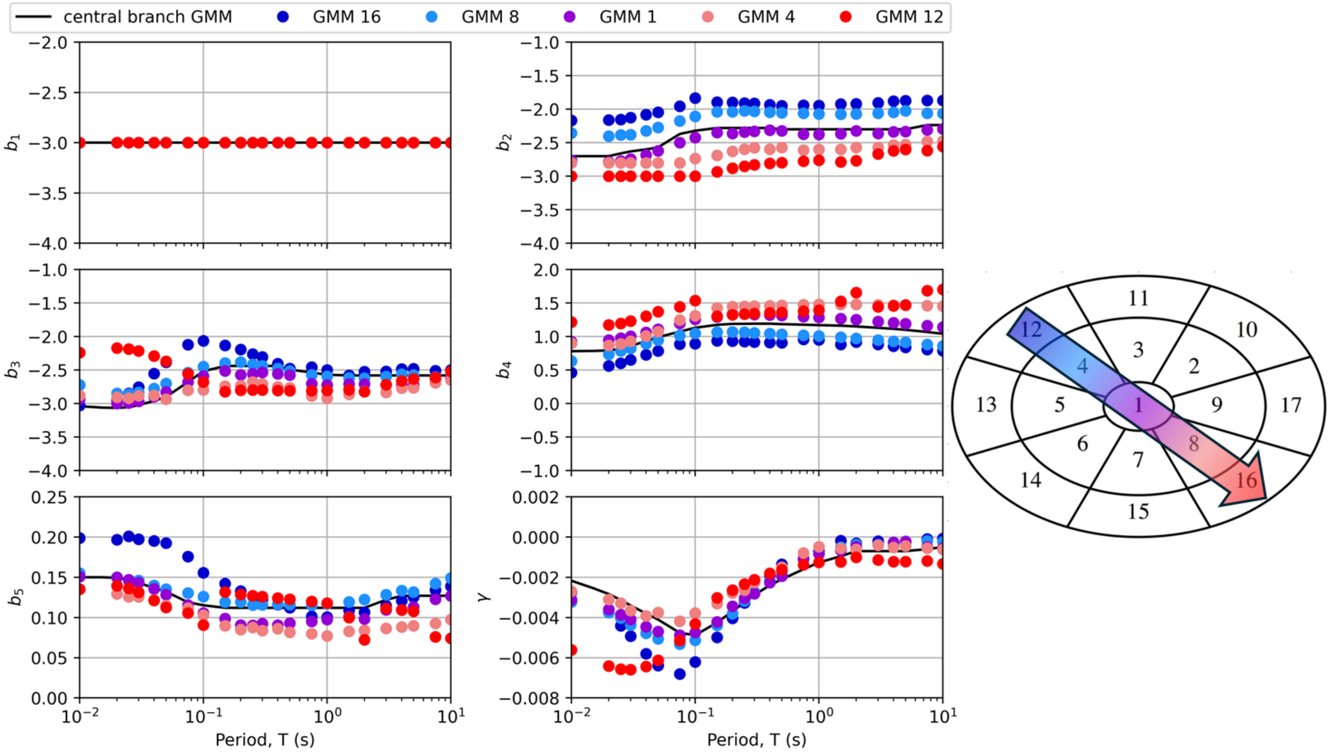


Figure 4.6. Variations of distance-scaling coefficients for selected NGA-East GMMs along a southeast-pointing azimuth in the Sammons map. This azimuth reflects the smallest differences in distance-scaling coefficients across GMMs.

4.4 Limitations of the Parametrized GMMs

The model fitting described in this chapter was done to facilitate adjustments to parameters so that the path component of the GMM could be adjusted to fit empirical data. Since there is no empirical data for **M** > 6, we focused on the parameter range of **M** < 6 and R_{rup} < 600km.

To investigate the performance of the model relative to the NGA-East central branch tabulated values for larger magnitudes (i.e., values not considered during fitting), Figure 4.15 shows the distance-dependence of spectral ratios (Eq. 4.10) for a wider range of magnitudes (up to 8.2) and distances out to 1000 km. Dashed lines are used for misfits that represent conditions beyond the fitted parameter range. At these larger magnitudes and for most distances, the misfits from Eq.

4.10) are negative, indicating that the fitted model over-predicts the tabulated NGA-East model. This suggests that the magnitude scaling at large magnitudes is too strong. As a consequence of these issues, I recommend against using the parametrized GMM for **M** > 6 earthquakes. Section 5.2.3 provides recommendations for estimating *IM*s for **M**>6 using the NGA-East GMMs (which extrapolate properly) and that incorporate findings from the present study.

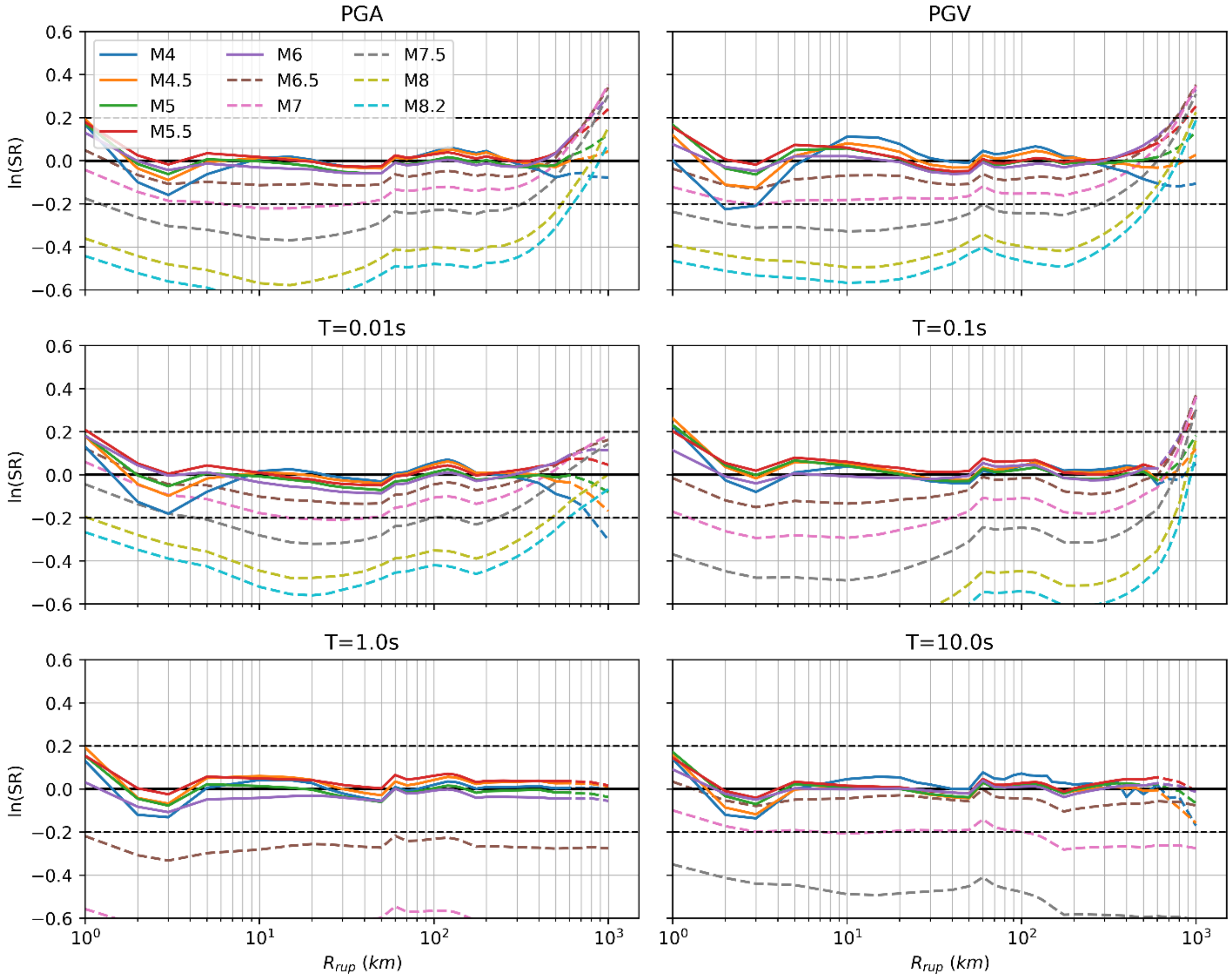


Figure 4.7. Distance-dependence of spectral ratios (Eq. 4.10) indicating misfit of equation-based model relative to tabulated NGA-East central branch GMM. Dashed lines are for extrapolated magnitudes and distances.

5. Adjustments to NGA-East Central Branch GMM

Using the equation-based GMM from Chapter 4, in this chapter I compute residuals and partition them into between-event components (event terms), within-event components, and site terms. I use the within-event residuals to investigate regional variations in path effects for four different geologic domains. The domains are described in Section 5.1 and the path analyses are described in Section 5.2. While there are similarities to the path investigations presented in Section 3.4.2, the distinctions of the present investigation are (1) the database is slightly expanded to include recordings of the **M**4.8 2024 Tewksbury, New Jersey earthquake and 9 events processed as part of the NGA-West3 project (Buckreis et al., 2024) that are located within the updated CENA boundary (see Figure 5.8); (2) different geologic domains are investigated, including separate consideration of the Gulf and Atlantic Coastal Plains; and (3) where data trends are identified, they are removed by parameterizing the fit, whereas in Section 3.4.2 the data was truncated at a certain distance as needed to avoid mapping path errors into model misfits.

Once the path models are updated, I re-compute residuals, which are used to investigate site response using both within-event residuals and site terms. I use the results to investigate whether the V_{S30}-scaling model from Stewart et al. (2020) requires adjustment and then investigate the potential scaling of site response with sediment depth in the coastal plain as provided by Boyd et al. (2024). This work, which is presented in Section 5.3, provides models for depth-related amplification that are fully compatible with the V_{S30}-scaling model currently used in practice.

5.1 CENA Regionalization

The area referred to as Central and Eastern North America (CENA) includes almost half of North America and it is known to have diverse and complex geologic formations. Because it is not on a plate boundary and seismicity rates are generally low, the locations of potentially active faults in the region are relatively poorly understood and zones have generally been used to model the locations and recurrence of future earthquakes (Figure 5.1). Although it is challenging to identify the individual structures responsible for earthquakes, the region has a history of damaging earthquakes and it is important to understand how broad variations in crustal structure affect wave propagation characteristics, including path and site response effects. In this section, I summarize information from literature pertaining to regional variations in crustal structure. These regions are considered in subsequent investigations of path effects (Section 5.2) and site effects (Section 5.3).

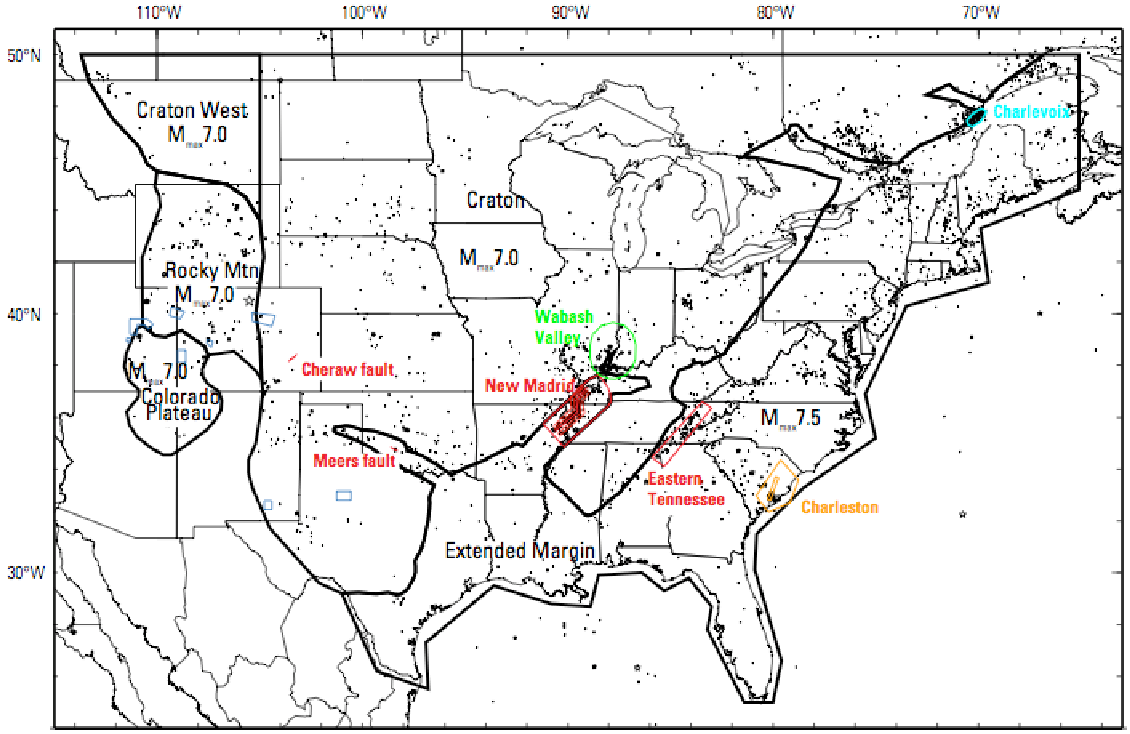


Figure 5.1. Map of CENA region showing source zones and mapped fault sourced considered in seismic source characterization model for national seismic hazard model (Petersen et al. 2008).

5.1.1 Coastal Plains

An important geologic feature of the CENA area is the presence of broad plains along the Atlantic and Gulf coasts, which are referred to in literature and here as the Atlantic Coastal Plain (ACP) and Gulf Coastal Plain (GCP). Due to the sparse data in CENA and similar seismological properties between the Gulf of Mexico and the Mississippi Embayment, these are often merged into one and referred to collectively as the GCP (e.g. Dreiling et al., 2014; Akhani et al., 2024).

The GCP region of CENA overlies a transition from a thick cratonic platform and Paleozoic orogenic belts to the north towards offshore oceanic crust that is much thinner in section (Salvador 1991; Chapmann and Conn 2016) to the south. As shown in Figure 5.2, between the thick craton and thin oceanic crust is a transitional continental crust that has undergone variable degrees of

Mesozoic extension and post-Jurassic sediment accumulation (Salvador 1991; Sawyer et al. 1991). Within this transitional zone, the basement rocks underlying the GCP in near-coastal areas are Jurassic to Upper Cretaceous carbonates and evaporites that include salt formations that have been observed to produce stronger than average attenuation (e.g. Gallegos et. al., 2014; Chapman and Conn, 2016; Pezeshk et al., 2018; Goulet et al., 2021; and Levandowski et. al., 2021). Such features form a strong argument for regionalization of the CENA path model.

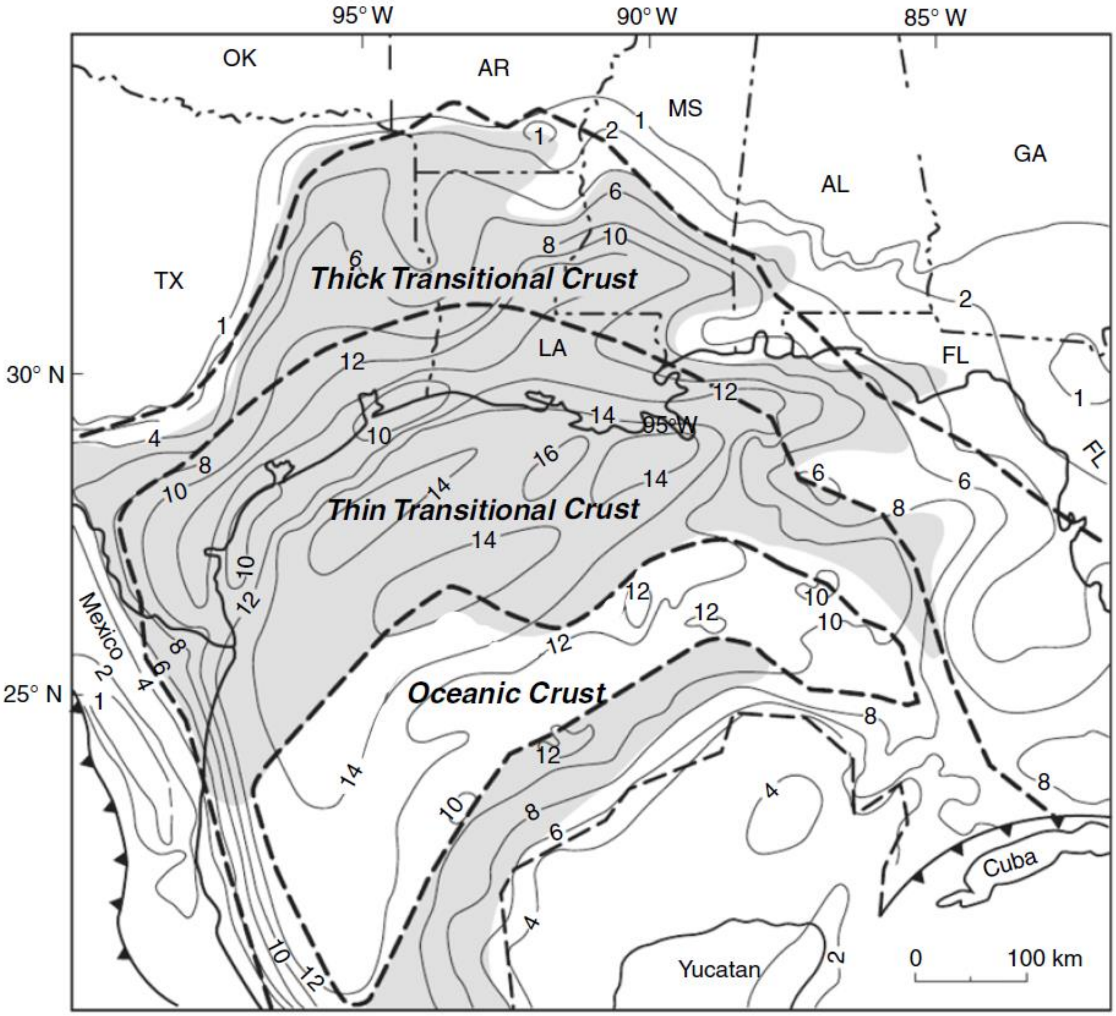


Figure 5.2. Map of Gulf of Mexico showing boundaries (dashed contours) between from north to south thick crustal craton, transition zone, and oceanic crust. (from Chapman and Conn 2016).

The ACP consists of predominantly marine sediments that gradually thicken towards the east and that overly Paleozoic and older crystalline bedrock (Maher 1965; Huddleston and Hetrick 1986; Huddleston 1988). The location of the ACP is the region to the south and east of the 0-depth contour in Figure 5.3, which also shows sediment thicknesses. The ACP sediments' rate of deposition is lower than the GCP, and as a consequence the young deposits are not as deep as in the GCP but are much deeper than the central parts of CENA north and west of the coastal plains (Dreiling et al., 2016). In contrast to the carbonates and evaporites underlying the GCP sediments, the basement beneath the ACP is predominantly Proterozoic to Paleozoic metamorphic or igneous rocks or well-consolidated sediments and volcanic rocks (Pratt et al. 2017; Pratt and Schleicher 2021). Given the different crustal structures beneath the ACP and GCP, and the differences in the sediments, it is reasonable to expect that both path and site effects for ACP may differ from those for GCP.

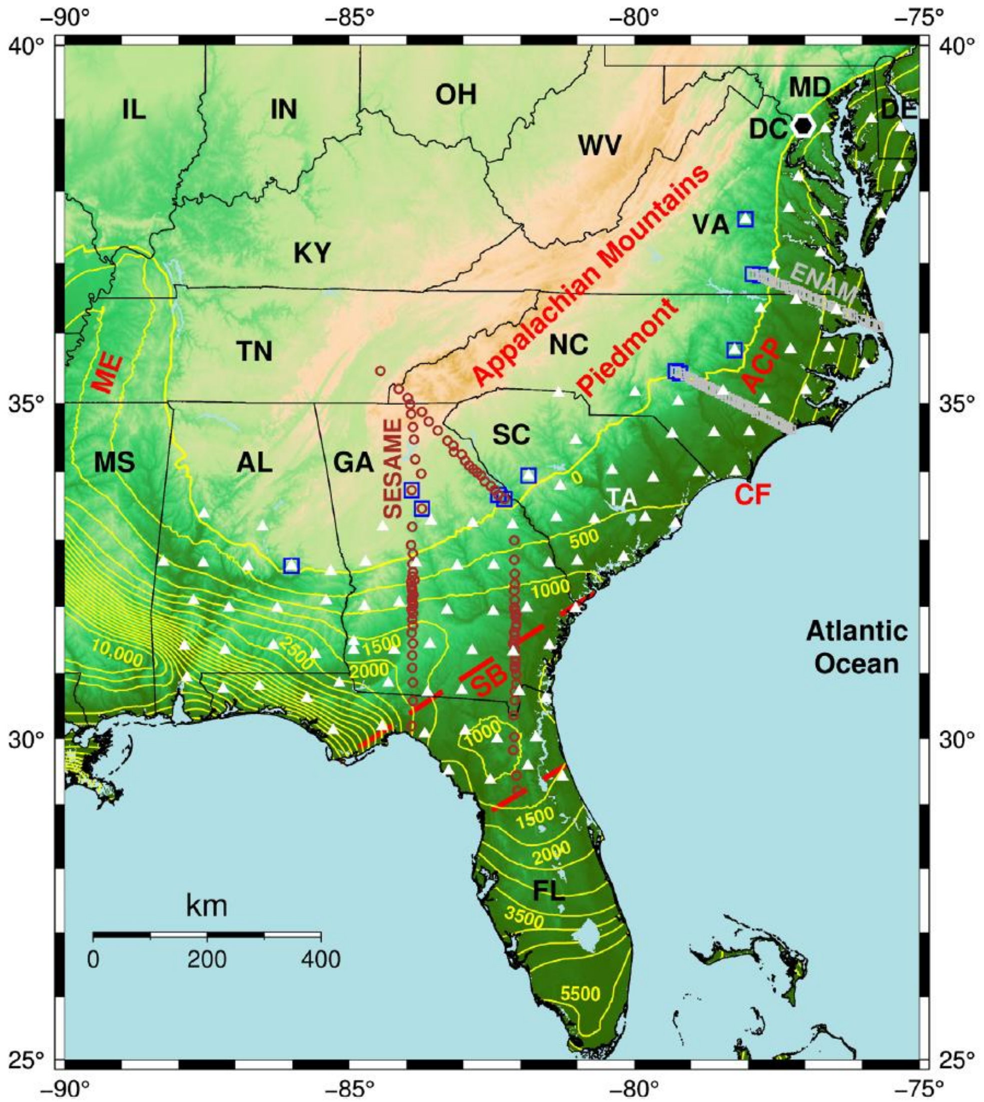


Figure 5.3. Map of the southeastern United States showing Atlantic Coastal Plain (ACP), which is located south and east of the 0-depth contour. The station locations and survey lines shown in the figure are not discussed in this document (from Pratt and Schleicher 2021).

Important distinctions between the GCP and ACP sedimentary structures and major basin structures in the western U.S. (e.g., Los Angeles, Seattle) are three fold: (1) the geologic composition of the underlying basement; (2) the unconfined geometry of the GCP (to the south) and the ACP (to the southeast), whereas the western basins consist of “bowl-like” structures bounded on all sides; and (3) faults immediately adjacent to or beneath western basins are typical, and hence the response of basins to motions from those faults are a critical consideration, whereas known sources beneath the coastal plains are generally not present, with the exception of Charleston South Carolina.

5.1.2 Central North America

The broad portion of CENA north of the GCP and west of the ACP is referred to here and elsewhere (Dreiling et al. 2016) as Central North America (CNA). The CNA region, shown in Figure 5.4, is tectonically stable and inactive since the middle Paleozoic. Paleozoic sediments are present in the shallow crust that are locally eroded to reveal Proterozoic igneous and metamorphic rocks (Bally, 1989; Mooney and Kaban, 2010). In contrast to the coastal plains, the late Cenozoic near-surface sediments are either absent or are a kilometer or less in thickness.

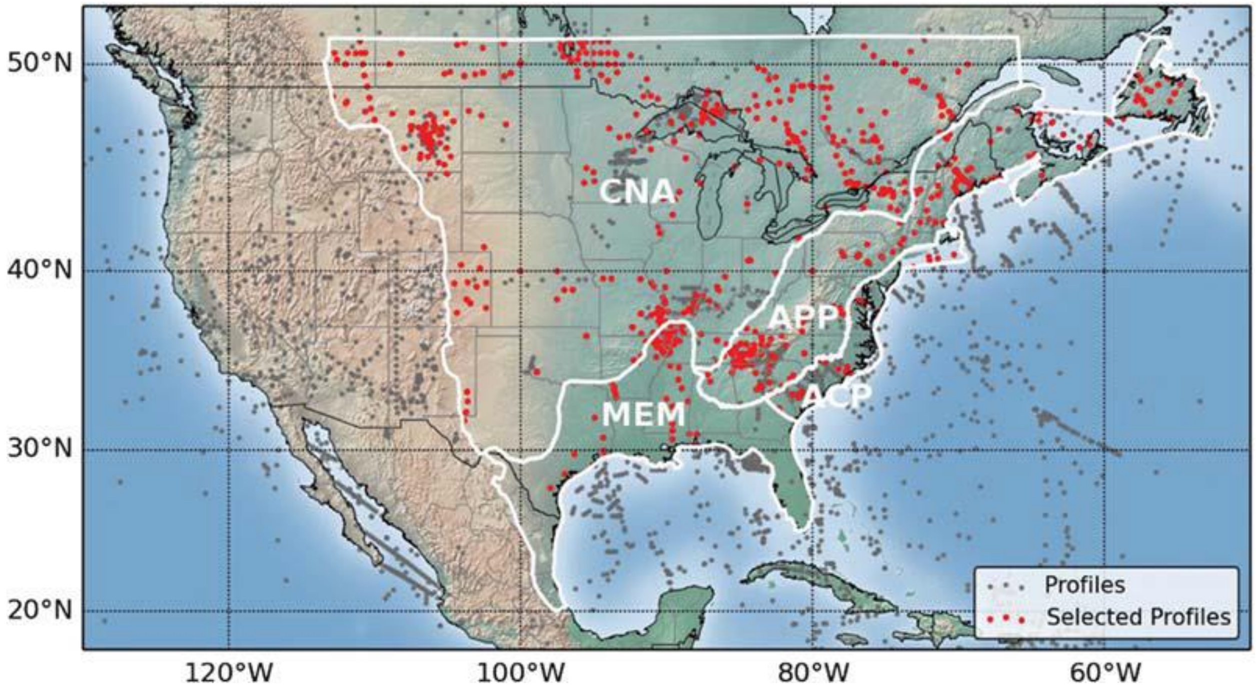


Figure 5.4. Regionalization of Central and Eastern North America (CENA) recommended by Dreiling et al. (2014, 2016). The regional labelled here as "MEM" matches "GCP" elsewhere in this document. (from Dreiling et al. 2016).

Within the broad CNA region, two subregions have been discussed in literature and were considered in this research: southern Granite-Rhyolite province (near Oklahoma) (Nelson and DePaolo 1985; Van Schmus et al. 1996) and the Yavapai-Mazatzal suture zone (Whitmeyer and Karlstrom 2007).

The southern Granite-Rhyolite province (Figure 5.5), which includes portions of Oklahoma and neighboring states, has some atypical features that include gravity anomalies and high velocity (>7 km/s) zones in the lower crust that are indicative of mafic composition which can be interpreted as evidence that this region has evolved from basaltic underplating and crustal melting (Ratre and Behm 2021). The zone of high velocity in Figure 5.5 also coincides with one of the deeper zones (~8km relative to present-day sea level) of the Great Unconformity formed by a slight rotation counterclockwise of the Tennessee block (Marshak et al., 2017). In Chapter 3, it was shown that the region in Figure 5.5 (referred to as the Texas-Oklahoma-Kansas – TOK – region) has distinct ground-motion magnitude- and distance-scaling characteristics, which has also been reported by Zalachoris and Rathje (2019), Moschetti et al. (2019), and Li et al., (under review). It is unknown whether the anomalies in crustal structure noted by Ratre and Behm (2021), which appear to be characteristic of the southern Granite-Rhyolite province, are responsible for these unusual features. Another hypothesis that has been put forward by Gallegos et al. (2014) argues that the lower-than-average crustal attenuations could be an effect of sediments 3 to 5 km deep with high concentrations of liquid and porosity.

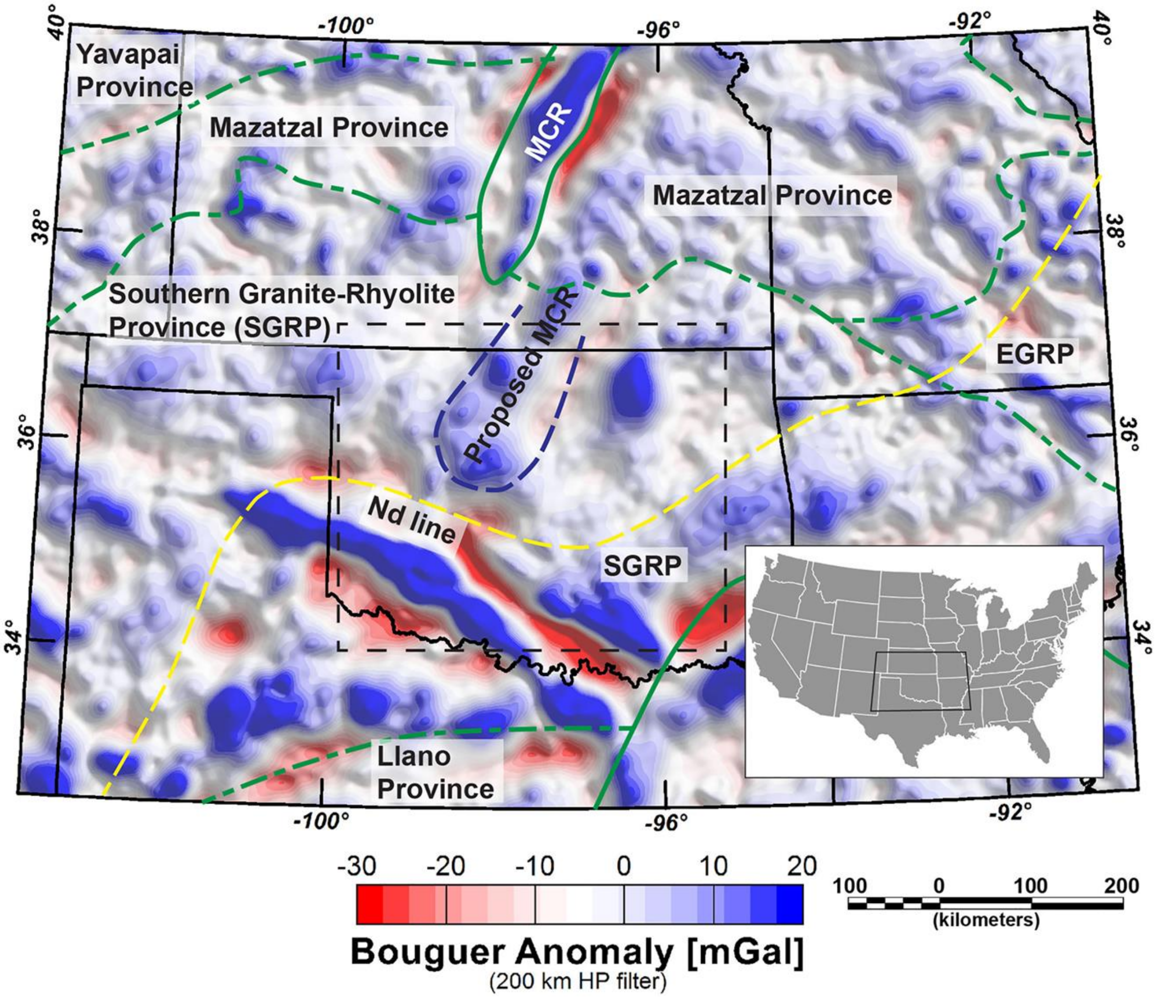


Figure 5.5. Map of Kansas, Oklahoma, and northern Texas region showing locations of Bouguer gravity anomalies. Figure from Ratre and Behm (2021).

The Yavapai-Mazatzal suture zones are associated with the lithosphere of CENA having been built by progressive addition of a series of geologically young volcanic arcs and accreted oceanic terranes. Two such accretionary processes are the Yavapai province (1.80–1.70 Ga), welded to North America during the 1.71–1.68 Ga Yavapai orogeny (Figure 5.6) and the Mazatzal province (1.70–1.65 Ga) added during the Mazatzal orogeny (Figure 5.7). These formations are associated with the Ancestral Rockies contraction (Zellman et al., 2021) and have experienced modern seismicity (i.e. in the last 350 years, e.g., Mueller, 2019). Low attenuations have been observed to align with the Yavapai-Mazatzal suture zone (Whitmeyer and Karlstrom, 2007). Gallegos et al.

(2014) estimated Lg Q_0 and found lower than average attenuations ranging from eastern New Mexico to Wisconsin which encompasses the transition zone between the Yavapai and Mazatzal provinces.

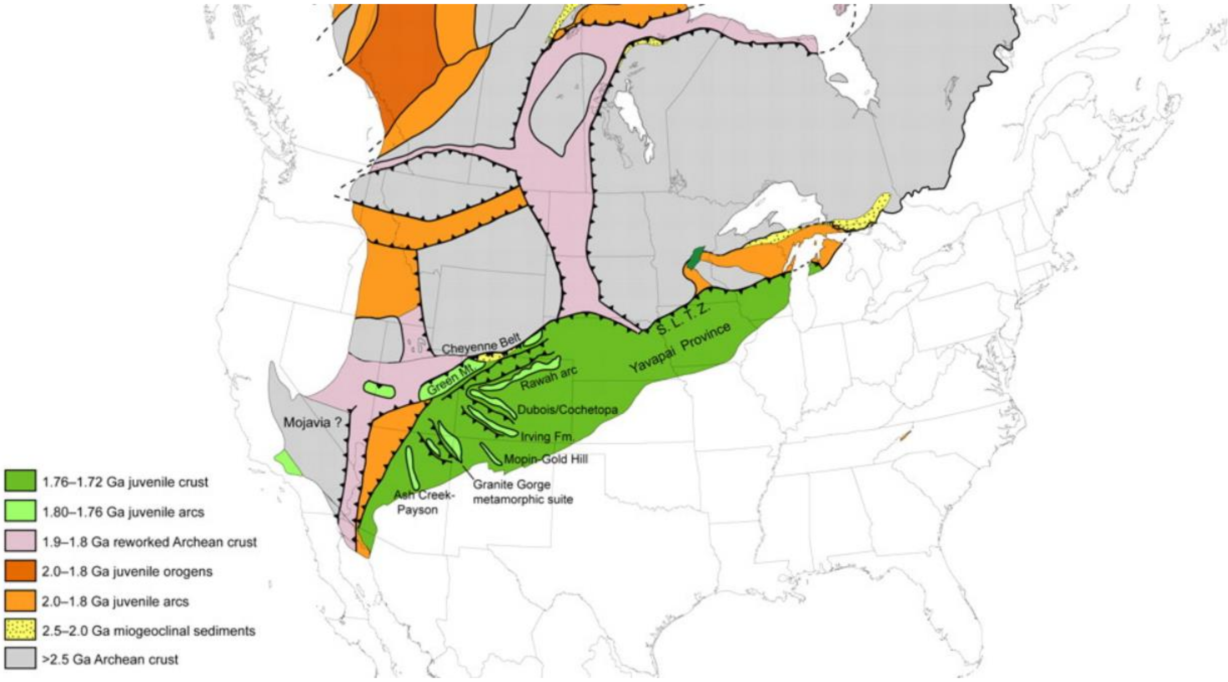


Figure 5.6. Map showing the location of the of the Yavapai suture zone. Figure from Whitmeyer et al. (2007).

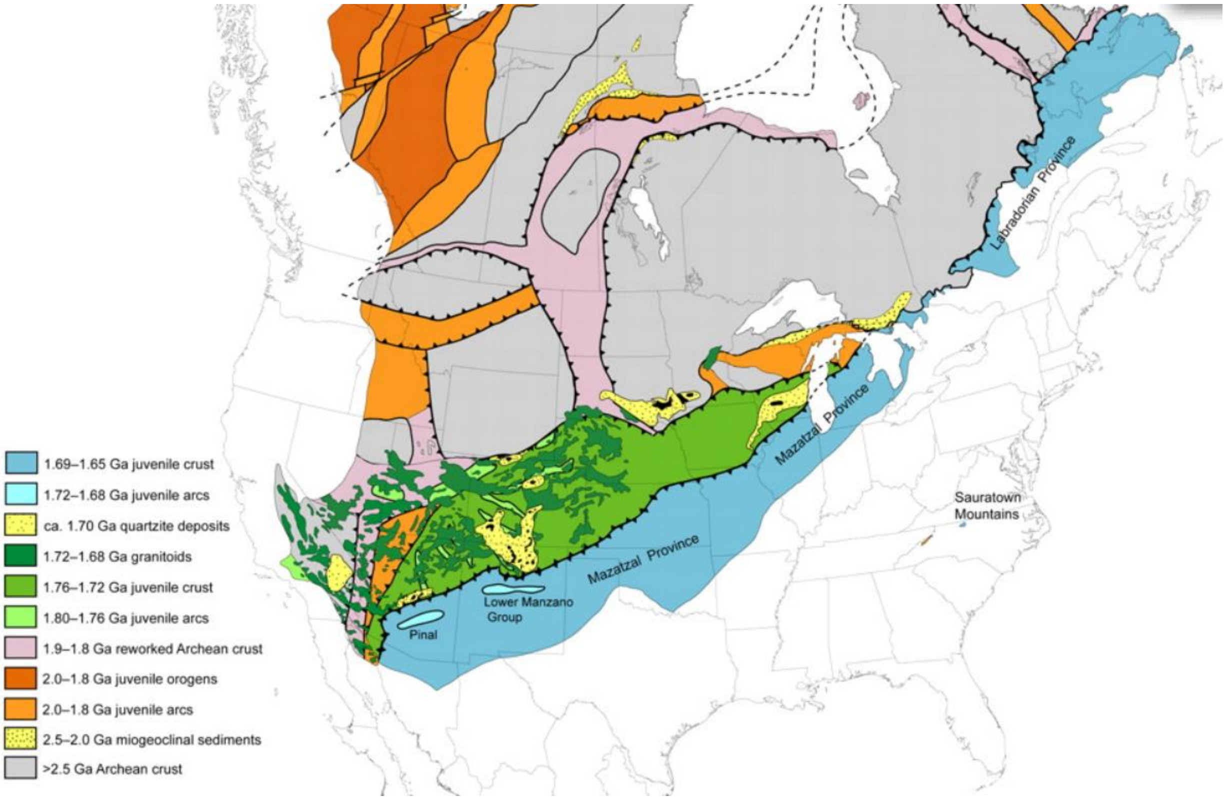


Figure 5.7. Map showing the location of the of the Mazatzal suture zone. Figure from Whitmeyer et al. (2007).

5.1.3 Regionalization in NGA-East and Present Work

The subject of regionalization was investigated in the NGA-East project using information on crustal quality factor (*Q*) and seismic velocities (Dreiling et al. 2014, 2016). This resulted in four broad regions that were investigated, as shown in Figure 5.4, which include (using the notation from this document) the CNA, ACP, and GCP regions and the APP region (Appalachia) not discussed above. Based on point source ground motion simulations using regional seismic velocities and *Q* values, and comparing the results to the baseline case (taken as CNA), it was found that CNA, ACP, and APP could be combined. Hence, two regions were ultimately suggested – a broad combined region (CNA/ACP/APP) and GCP.

For the regional path studies in this chapter, I selected four regions as shown in Figure 5.8. An additional “region” is added to encompass the totality of CENA (referred to as “other”) that includes CENA outside of the four colored polygons in Figure 5.8. Site response studies use data across CENA, with the exception of TOK events as noted below, for investigations of V_{S30}-scaling. Investigations of the effects of depth in the GCP and ACP on ground motion are necessarily regionalized.

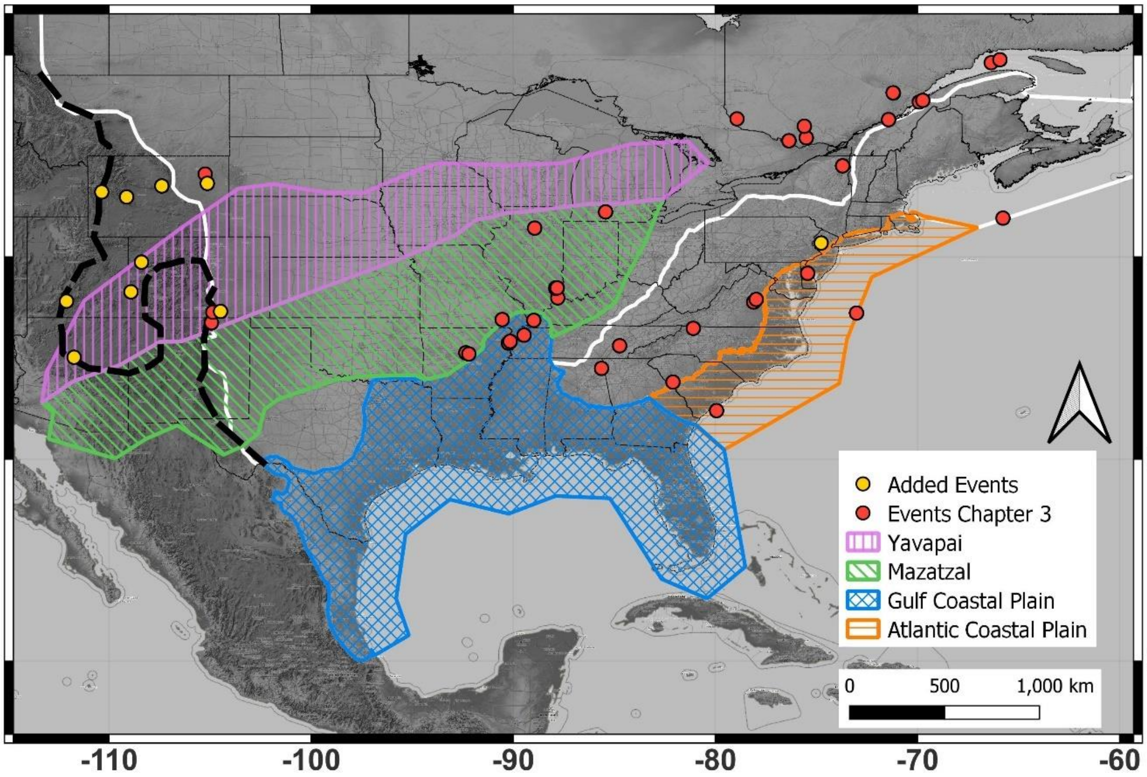


Figure 5.8. Map of CENA locating the four regions of interest and events used in this chapter.

For the work presented in Chapter 5, events located in TOK are excluded due to the unusual geologic features in the southern Granite-Rhyolite province (Section 5.1.2), which I anticipate could affect source and path effects in unusual ways, and the high concentration of potentially induced seismicity in the zone. Evidence for the unusual path effects in TOK was presented in Section 3.4.2 between rupture distances of 10 and 150km and by Li et al. (in review). Hence, the results presented here are exclusive of TOK events.

5.2 Regional Path Analyses

5.2.1 Data Resources Considered

The data that was considered in the regional path and site analyses is the same as that presented in Chapter 3 but with the addition of recordings of the **M**4.8 2024 Tewksbury, New Jersey earthquake and 9 events processed as part of the NGA-West3 project (Buckreis et al. 2024) that are located within the updated (i.e., Moschetti et al., 2024) CENA boundary. Figure 5.8 shows the locations of these events marked in yellow. Figure 5.9 shows the overall dataset (similar to Figure 3.2) with the new data added.

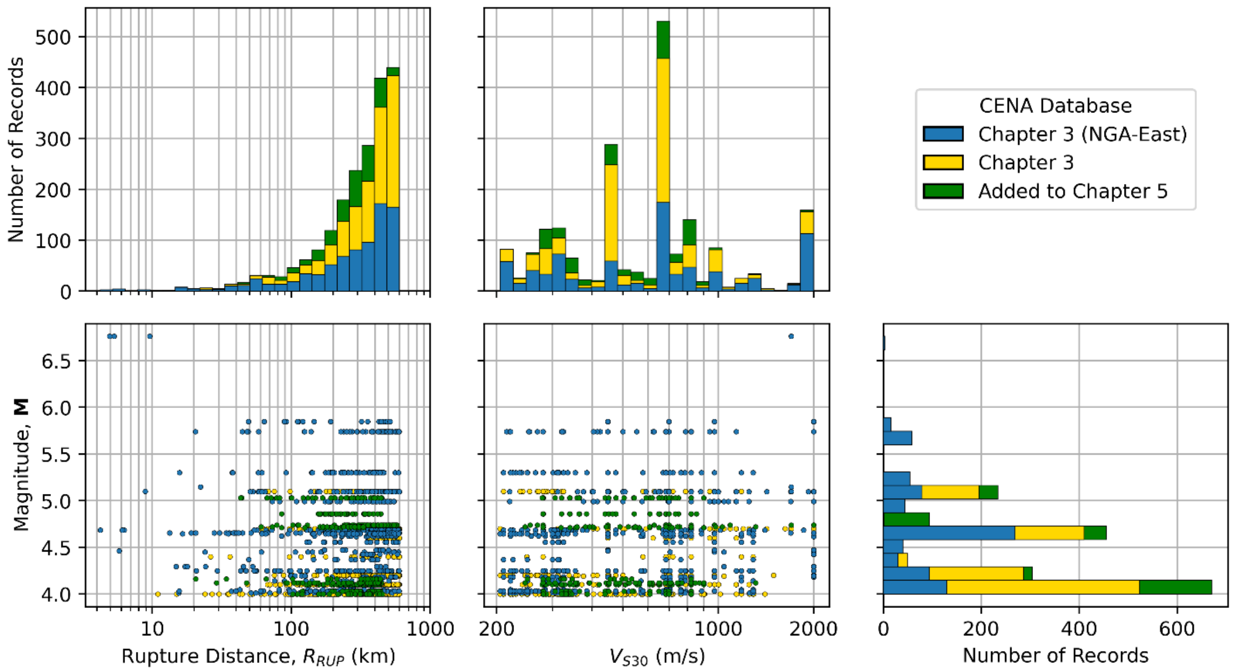


Figure 5.9. Distributions of CENA dataset with respect to rupture distance, V_{S30}, and magnitude, showing differences between data used for misfit analyses (Chapter 3) and the recently processed data.

The reason for the different datasets in different parts of the study is solely related to timing. The work presented in Chapter 3 was completed in the 2022-2023 time frame, which is prior to the New Jersey event and prior to the additional data processing from the NGA-West3 project. The work presented in this chapter was completed in mid-2024.

Figure 5.10 is a map of CENA with the regions from Section 5.1 that shows all considered events with source-site paths shown with lines. This shows the degree to which the different regions are sampled by the data. Mazatzal, ACP, and *other* are well-sampled. GCP is sampled near its northern margin including the Mississippi Embayment but sparsely sampled to the south. Yavapai is relatively sparsely sampled.

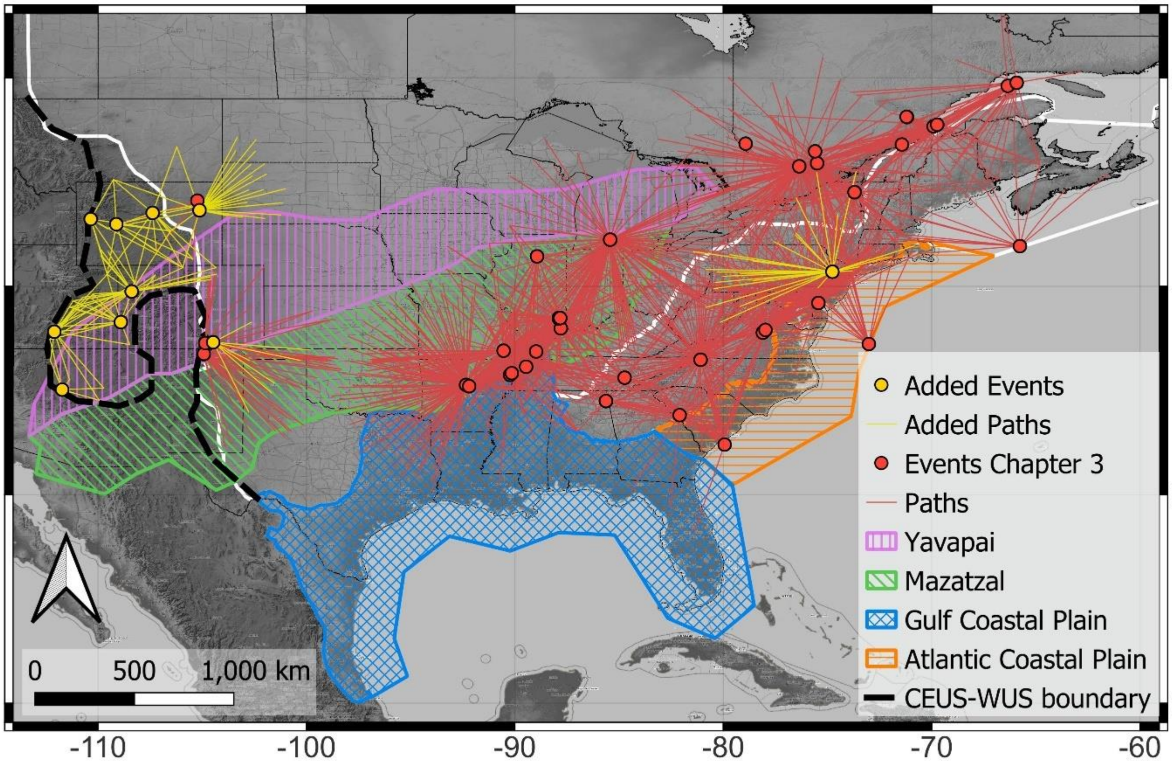


Figure 5.10. Map of CENA locating the four regions identified in Section 5.1, events and paths used in this chapter.

5.2.2 Residuals Analyses, Mean Misfits, and Event Terms

Total residuals were calculated as given by Eq. (3.3) and then partitioned using mixed-effects analyses as given by Eqs. (3.4-3.5). In the application of Eq. (3.3), we initially considered both the NGA-East tabulated values (Sa^{targ}) and the equation-based model (Eqs 4.1-4.9) for the hard-rock GMM (denoted $\mu_{ln,k}$ in Eq. 3.3). Figure 5.11 shows mean model misfit (c_k from Eq. 3.4) as-derived in Chapter 3 and as re-computed here using the function version of the GMM. In both cases $\mu_{ln,k}$ is taken from the dataset presented in Chapter 3. The ordinates are not significantly different, with the previous c_k term being within the standard error bounds of the current results. As expected, there are some differences because the models, although similar, are not identical. Figure 5.12

shows the mean misfit for the equation-based GMM and the expanded database. The newly added ground-motions do not change the trend of negative misfit at short periods and positive at long periods. Moreover, the c_k for the new dataset is relatively similar with the exception of *Sa* at T = 0.1 s where the misfit is more negative. Figure 5.13 shows the trend of event terms with magnitude for PGA and *Sa* for 0.1, 1, and 5 sec with the new events highlighted; the previous findings of no trends are encountered again.

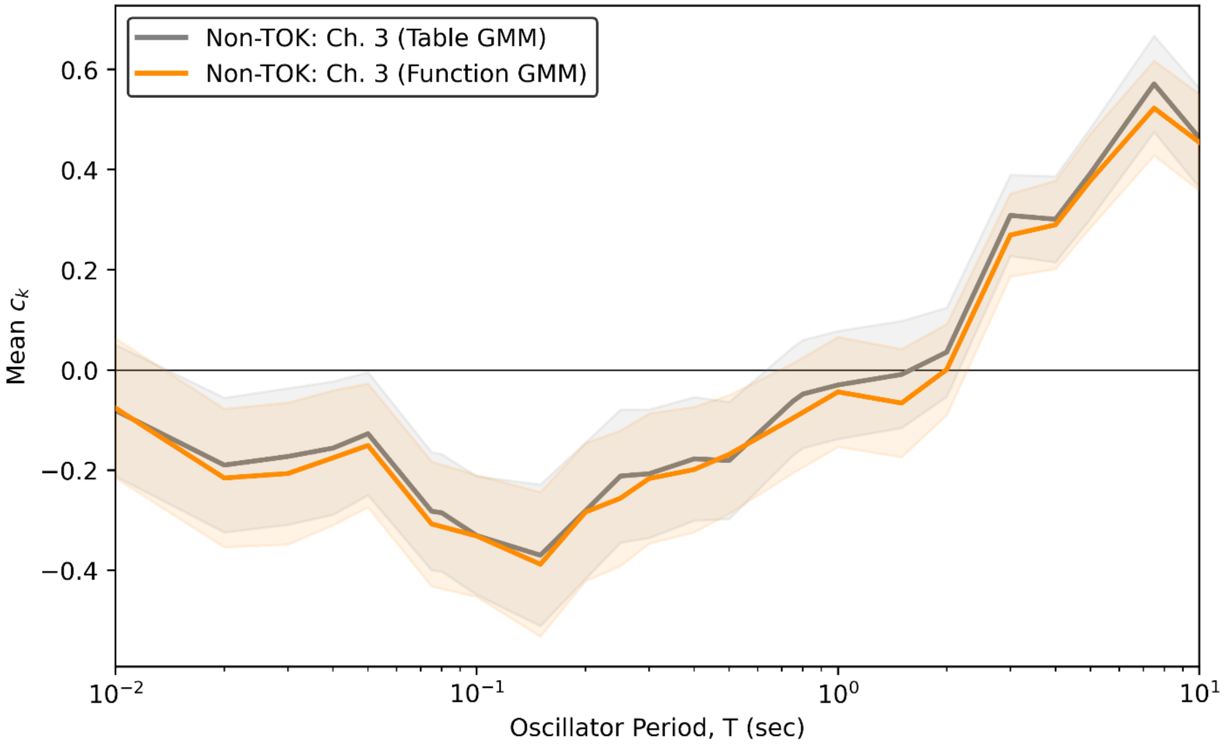


Figure 5.11. Period dependence of misfit term c_k for original, tabulated version of NGA-East central branch GMM and the equation-based version of the NGA-East central branch GMM. The shaded regions enclose ± one standard error.

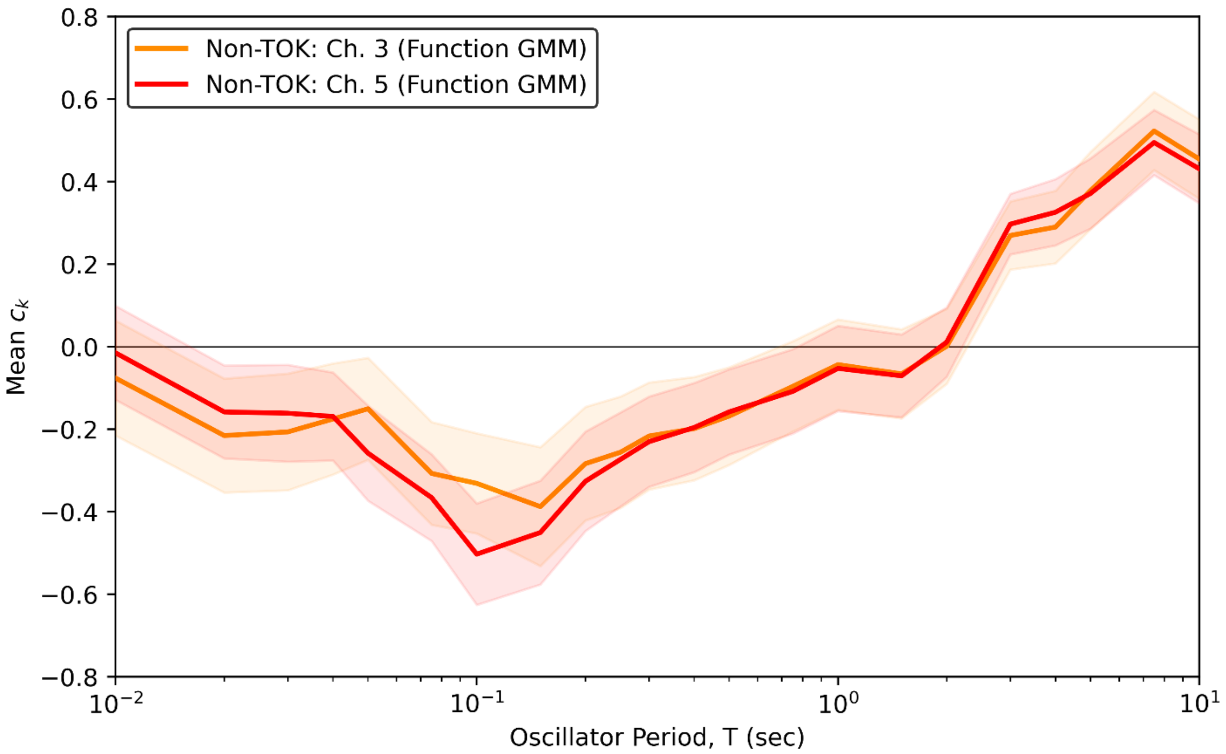


Figure 5.12. Period dependence of misfit term c_k for the equation-based versions of NGA-East central branch GMM with site adjustments for updated dataset from Figure 5.10. The shaded regions enclose ± one standard error.

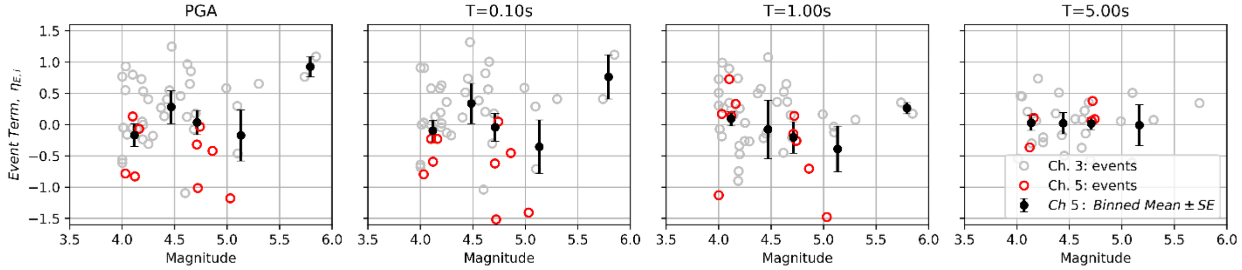


Figure 5.13. Trends of event terms with magnitude for non-TOK regions with updated database and newly added events highlighted.

5.2.3 Regional Path Regressions

The coefficients in the parametrized GMM presented in Chapter 4 (Eqs. 4.1-4.9), which match the NGA-East central branch model, represent a CENA-wide average. My intent is to regionalize the path model, which can be done through the use of an additive anelastic attenuation parameter, $\Delta\gamma$, as follows:

$$ln\left(\hat{Y}\left(\mathbf{M}, R_{rup}\right)\right) = F_E(\mathbf{M}) + F_Z + (\gamma + \Delta\gamma^*)R_{rup} \tag{5.1}$$

where $\Delta\gamma^*$ is computed from regional coefficients. To evaluate these coefficients, I compute total residuals (Eq. 3.3) and partition the residuals using Eqs. (3.4) and (3.5). The important residuals for the present analysis are the within-event residuals, δW_{ij}, where index i is for the event and index j is for the site. The estimation of $\Delta\gamma^*$ is conducted using "curve_fit" from the *Scipy* library in python using Eq. (5.2),

$$\delta W_{ij} = \Delta\gamma^*\left(R_{rup}\right)_{ij} + \Delta e_{0,r} \tag{5.2}$$

where $\Delta e_{0,r}$ is a constant needed for the fitting process. Parameter $\Delta e_{0,r}$ is an intercept analogous to e_0 in Eq. (4.2), because it shifts the fit to the data vertically. The subscript *r* refers to the region. Parameter $\Delta\gamma^*$ is a function of $\Delta\gamma_r$, the number of regions (*n*), and the percentage of the path traveled in region *r* (W_r),

$$\Delta\gamma^* = \sum_{r=1}^{n} \Delta\gamma_r W_r \tag{5.3}$$

To undertake the calculations, I first calculated W_r for each ground-motion / event pair using the polygons in Figure 5.10, I then computed $\Delta e_{0,r}$ to be the weighted mean of δW_{ij} up to 180km traveled within each region r to better constrain the regression given the sparse data, and lastly I ran a least-squares regression using Eq. (5.2)-(5.3) to estimate the $\Delta\gamma_r$ values.

The data within each region is plotted in Figure 5.14 as δW_{ij} versus the source-to-site distance as measured within the region. Following procedures introduced by Buckreis et al. (2023), data points are plotted with shading that reflects W_r as shown in the right margin of the plot. Trend lines are plotted using Eq. (5.2) with $W_r = 1$.

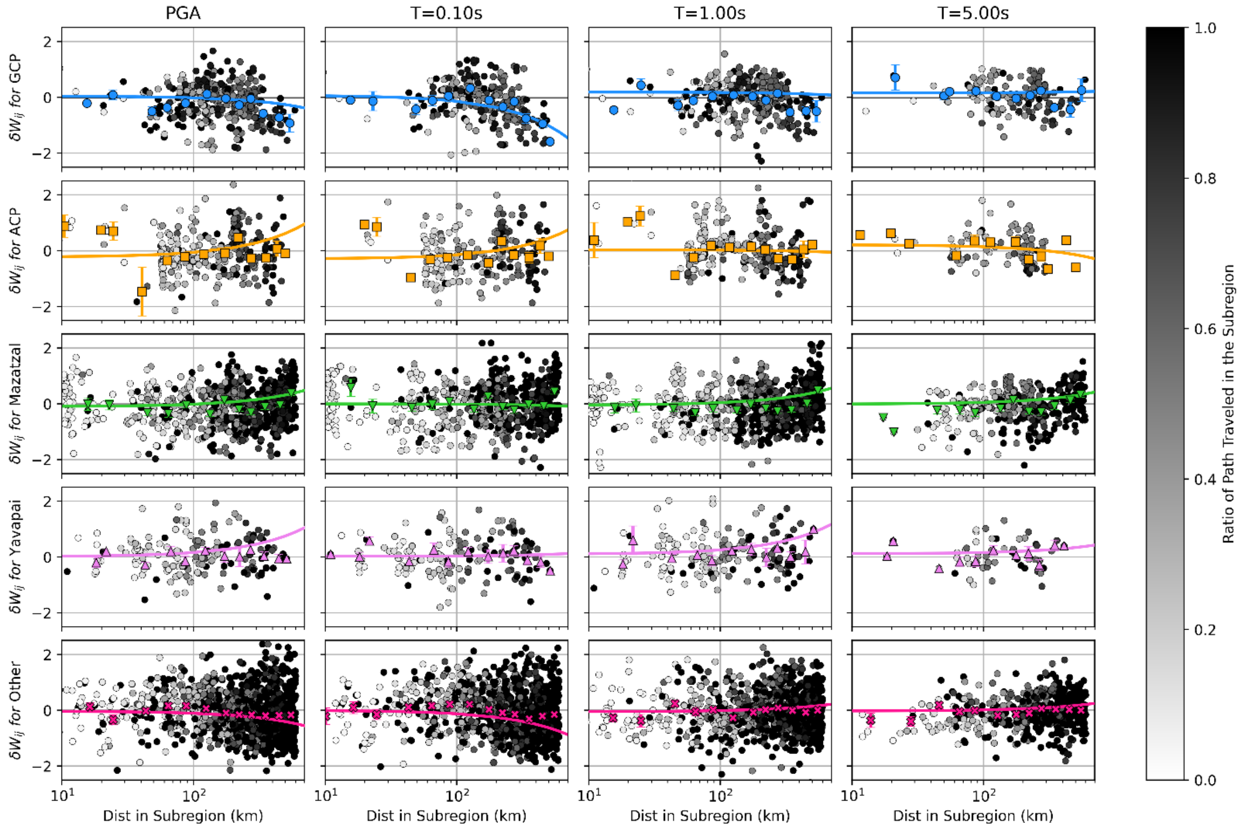


Figure 5.14. Within-event residuals (δW_{ij}) against the distance traveled in each considered subregion. The colored line is Eq. 5.2 with W_r=1. The colored symbols represent binned means of the shaded data points. The level of shade depends on the percentage of the distance from the epicenter to the site that falls within the subregion.

The $\Delta e_{0,r}$ values produced by the regression, and reflected in Figure 5.14 by offset in the trend line near 10 km, are presented in Figure 5.15 along with their standard errors. Since there are no trends in the event terms (Figure 5.13), I assume there is no magnitude-scaling bias, and because $\Delta\gamma$ should remove distance-scaling bias, the meaning of $\Delta e_{0,r}$ could be interpreted as a shift in mean site terms for region r. The results in Figures 5.15 show that $\Delta e_{0,r}$ for *Mazatzal* (abbreviated *Maza*) and *other* are unbiased, in the sense that zero lies within the bounds of $\Delta e_{0,r}$ ± its standard errors. On the other hand, *Yavapai* (abbreviated *Yava*) has negative bias at short periods, which could be a κ_0 effect (a higher value may be needed in that region). Both the GCP and ACP have positive bias at long periods. This could be a consequence of not considering sediment-depth effects in the model or bias in the V_{S30} scaling.

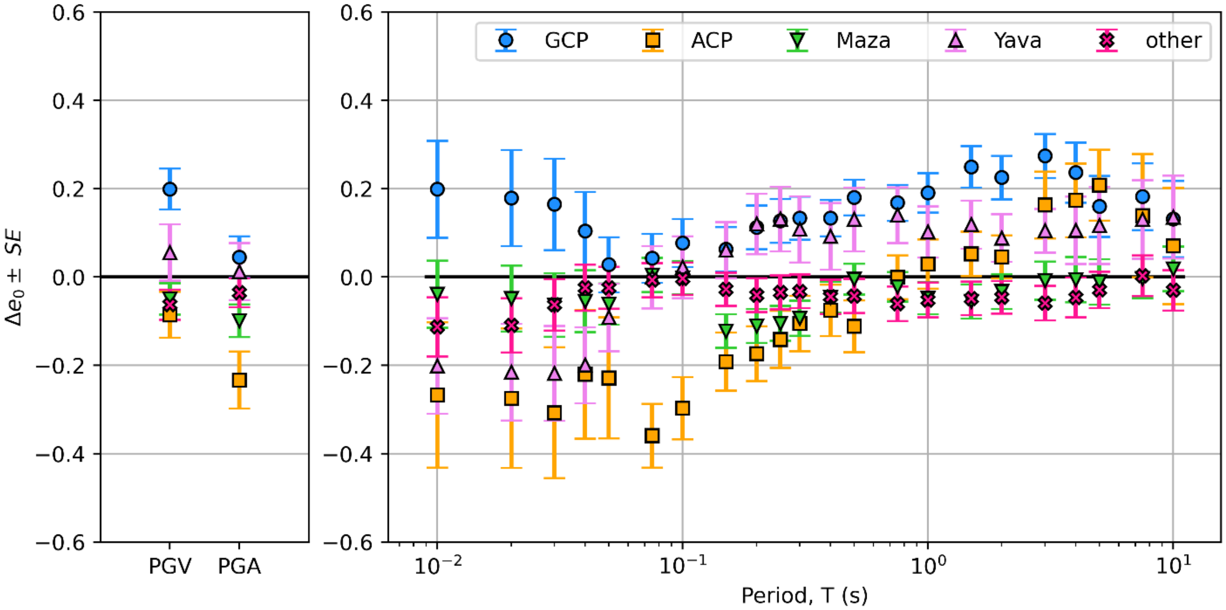


Figure 5.15. Trend of $\Delta e_{0,r}$ against period for five subregions: Gulf Coastal Plain (GCP), Atlantic Coastal Plain (ACP), Mazatzal providence (*Maza*), Yavapai province (*Yava*) and the rest of CENA that is not encompassed within the aforementioned regions (*other*).

Figure 5.16 shows the values of $\Delta\gamma_r$ as regressed and their standard errors. Where $\Delta\gamma_r$ is positive and negative, the trends lines in Figure 5.14 bend upward and downward at large distances, respectively. For GCP, the $\Delta\gamma_r$ values are consistently negative (i.e. stronger attenuation is observed) at short periods, the curves in Figure 5.14 bend downward, and the standard errors are small. This is an expected trend based on prior work (e.g. Gallegos et. al., 2014; Chapman and Conn, 2016; Pezeshk et al., 2018; Goulet et al., 2021; Levandowski et. al., 2021; and work presented in Section 3.4.2, Figure 3.7) and the characteristics of the basement rock (Section 5.1.1), reflecting faster than average attenuation rates. *Other* has $\Delta\gamma_r$ values that are relatively small and fluctuate around zero. This is also expected, as *other* should reflect the average attenuation of CENA as contained in the NGA-East central branch model. ACP has $\Delta\gamma_r$ values that fluctuate around zero with relatively high standard errors. I interpret the high standard errors and the lack of a trend as indicating no regional departure from the CENA average attenuation model. *Yava* has $\Delta\gamma_r$ values that fluctuate with periods and that have large standard errors, therefore, presenting no strong evidence for a regional attenuation rate distinct from the CENA-wide average. The $\Delta\gamma_r$ values for *Maza* are also small and fluctuate around zero without demonstrating a clear trend. To summarize, the results in Figures 5.14 and 5.15 do not support the presence of path misfits for ACP, *Yava*, *Maza*, and *other*. Only GCP exhibits persistent trends.

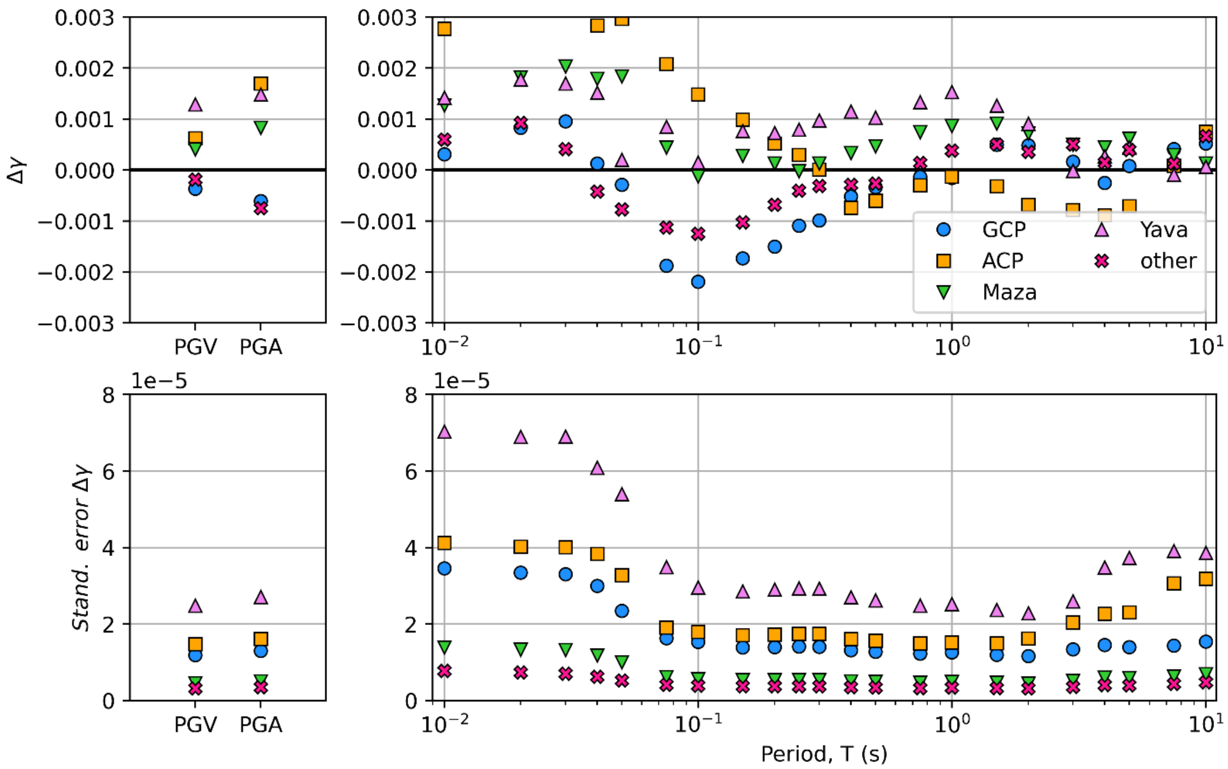


Figure 5.16. Trend of $\Delta\gamma_r$ and the standard error resulting from the regression against period for 5 subregions: Gulf Coastal Plain (GCP), Atlantic Coastal Plain (ACP), Mazatzal providence (Maza), Yavapai province (Yava) and the rest of CENA that is not encompassed within the aforementioned regions (*other*).

As a consequence of the lack of trend for all regions aside from GCP, all regions except for GCP are merged into *other*. With the new data grouping, I repeated the regression of Eq. (5.2) for two regions only with the results in Figures 5.17, 5.18, and 5.19 (equivalent to Figures 5.14, 5.15, and 5.16). The fits of Eq. (5.2) to the data, as shown in Figure 5.17, have a strong downward trend for GCP and more modest trends for *other*. The weighted binned means depict no significant trend for *other*, but this is not always captured by the regression. This could be explained by the constraints applied during regression (i.e., fitting of the constant term with a subset of the data before fitting the slope). Moreover, by observing Figure 5.19 for the behavior of $\Delta\gamma_r$ across periods, $\Delta\gamma_r$ values for GCP are negative beyond their standard errors at short periods whereas for *other* they fluctuate around zero within a range of ±0.001. Therefore, there is no clear evidence of a trend in *other*. As shown in Figure 5.18, GCP still presents positive $\Delta e_{0,r}$ at long periods, which I anticipate being a site effect related to the deep GCP sediments that will be addressed in Section 5.3. The $\Delta e_{0,r}$ values for the *other* group are generally small with zero falling within the standard error bounds.

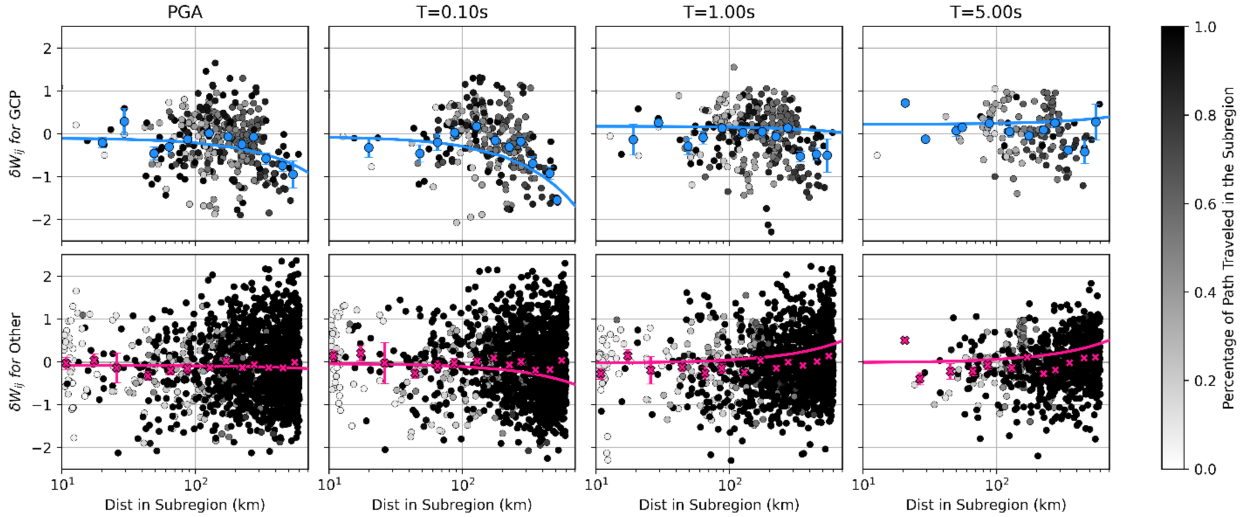


Figure 5.17. Within-event residuals (δW_{ij}) against the distance traveled in the two regions. The colored line is Eq. 5.2 with W_r=1. The colored symbols represent binned means of the shaded data points. The level of shade depends on the percentage of the distance from the epicenter to the site that falls within the subregion.

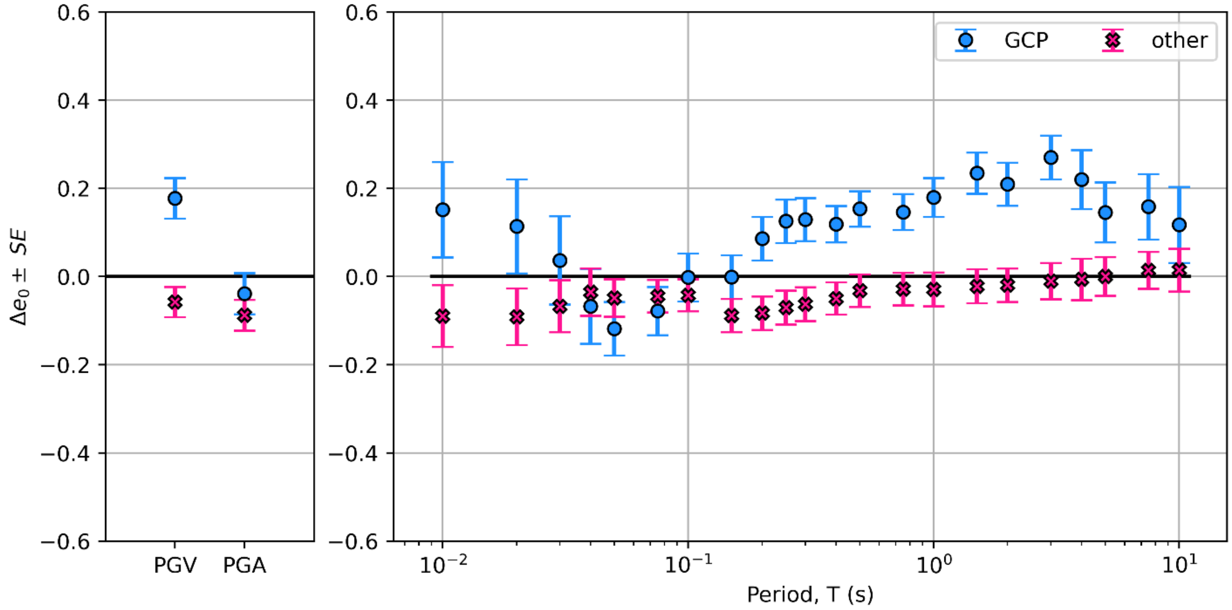


Figure 5.18. Trend of $\Delta e_{0,r}$ against period for two subregions: Gulf Coastal Plain (GCP) and everything else (*other*).

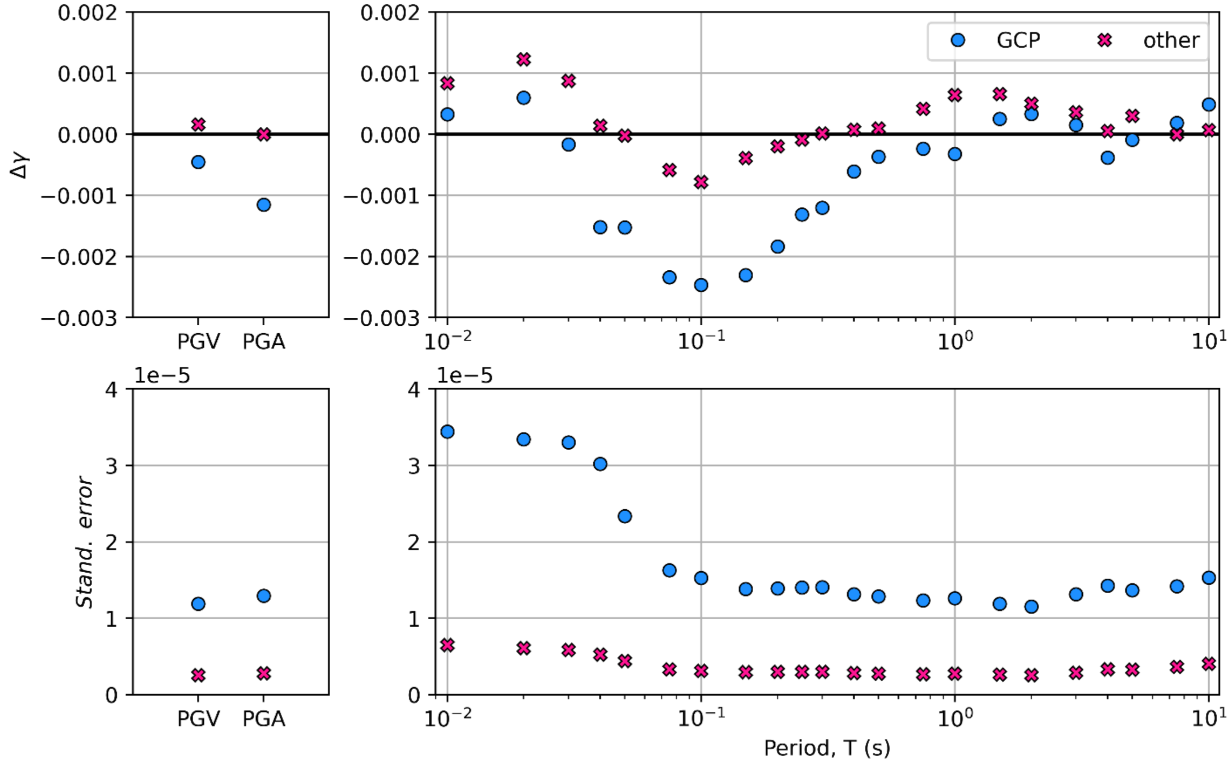


Figure 5.19. Trend of $\Delta\gamma_r$ and the standard error resulting from the regression against period for two subregions: Gulf Coastal Plain (GCP) and everything else (*other*).

In order for the coefficients of the subregional path model to be useful, I performed the following post-processing: (1) coefficients were smoothed with respect to period, (2) the sum of $\Delta\gamma_r$ and γ cannot be positive, which would lack physical meaning, and (3) response spectra must be smooth across periods without odd shapes.

If as-regressed $\Delta\gamma_r$ values were directly applied, the response spectral shapes would have artificial peaks and corners. Hence, a smoothing process was undertaken. Prior to smoothing, adjustments to $\Delta\gamma_r + \gamma$ were applied by simply shifting $\Delta\gamma_r + \gamma$ vertically for all periods to match the regressed value at 0.1sec. Next, the values at short periods were gradually transitioned based on visual inspection so that the value of $\Delta\gamma_r + \gamma$ at 0.01 sec matched that for PGA and a realistic spectral shape was obtained. At long periods $\Delta\gamma_r + \gamma \geq 0$ hence I did not allow $\Delta\gamma_r$ to exceed a minimum absolute value between 0 and γ. The resulting smoothed values of $\Delta\gamma_r$ and $\Delta\gamma_r + \gamma$ are plotted versus period in Figure 5.20 and listed in Table 5.1, which shows the adjustments to the "as-regressed" coefficients to meet the three aforementioned criteria.

The changes affect the path model performance. Performance evaluation involves comparing data to a revised model that incorporates the smoothed $\Delta\gamma_r$ values from Figure 5.20 into Eq. (5.3) with n = 2, and then applying the resulting $\Delta\gamma^*$ value into Eq. (5.1). Residuals computed with that model were then partitioned to evaluate within-event residuals, which are shown in Figure 5.21. The colored symbols in the figure represent binned means of the data points, which have no trend with distance.

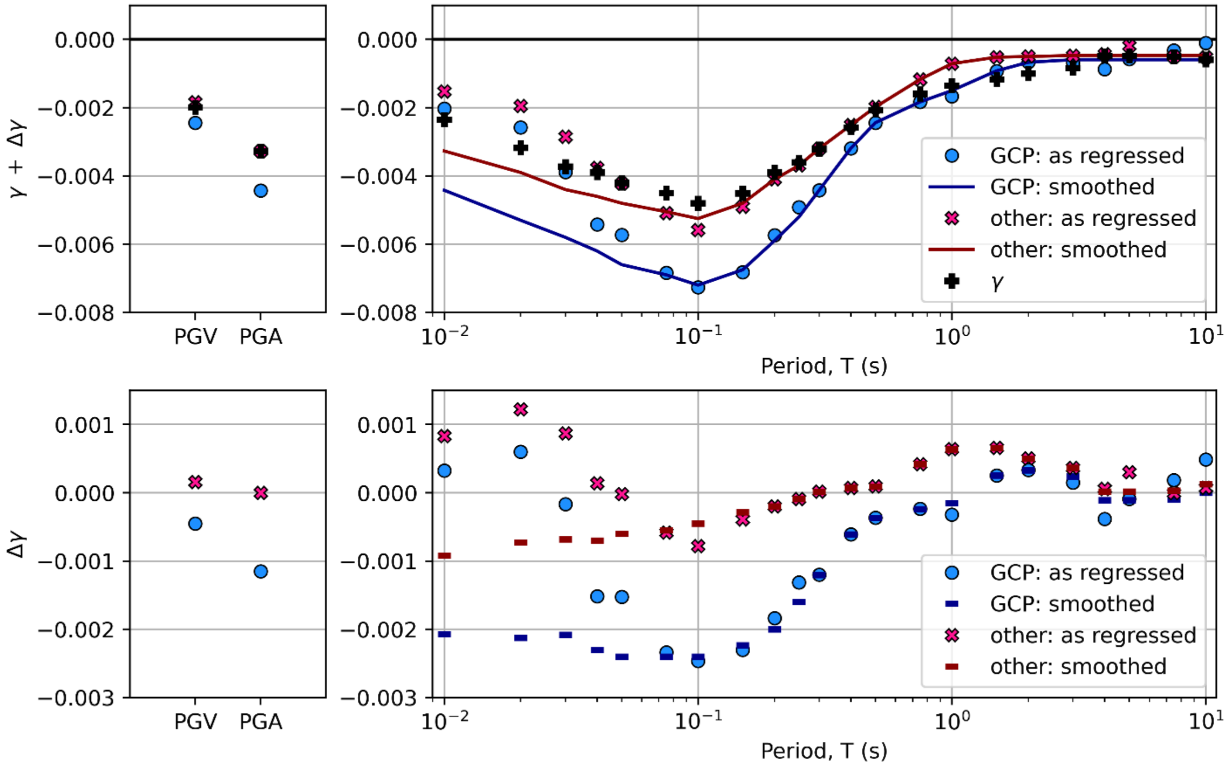


Figure 5.20. As regressed and smoothed subregional path attenuation coefficients compared to γ, the global anelastic attenuation coefficient.

Table 5.1. Smoothed coefficients of $\Delta\gamma_r$ for GCP and other

Period	$\Delta\gamma_r$: **GCP**	$\Delta\gamma_r$: **Other**
PGA	-0.00115	-0.000001
PGV	-0.00045	0.000159
0.010	-0.00207	-0.00092
0.020	-0.00213	-0.000728
0.030	-0.00208	-0.00068
0.040	-0.0023	-0.0007
0.050	-0.0024	-0.0006
0.075	-0.0024	-0.00055
0.10	-0.0024	-0.00045
0.15	-0.00224	-0.000286
0.20	-0.002	-0.0002
0.25	-0.0016	-0.00009
0.30	-0.00121	0.000013
0.40	-0.00061	0.00007
0.50	-0.00037	0.000092
0.75	-0.00024	0.000421
1.0	-0.00015	0.000638
1.5	0.00025	0.000653
2.0	0.00033	0.0005
3.0	0.00024	0.000364
4.0	-0.00011	0.000017
5.0	-0.00011	0.000017
7.5	-0.00009	0.000031
10.0	0.0	0.000125

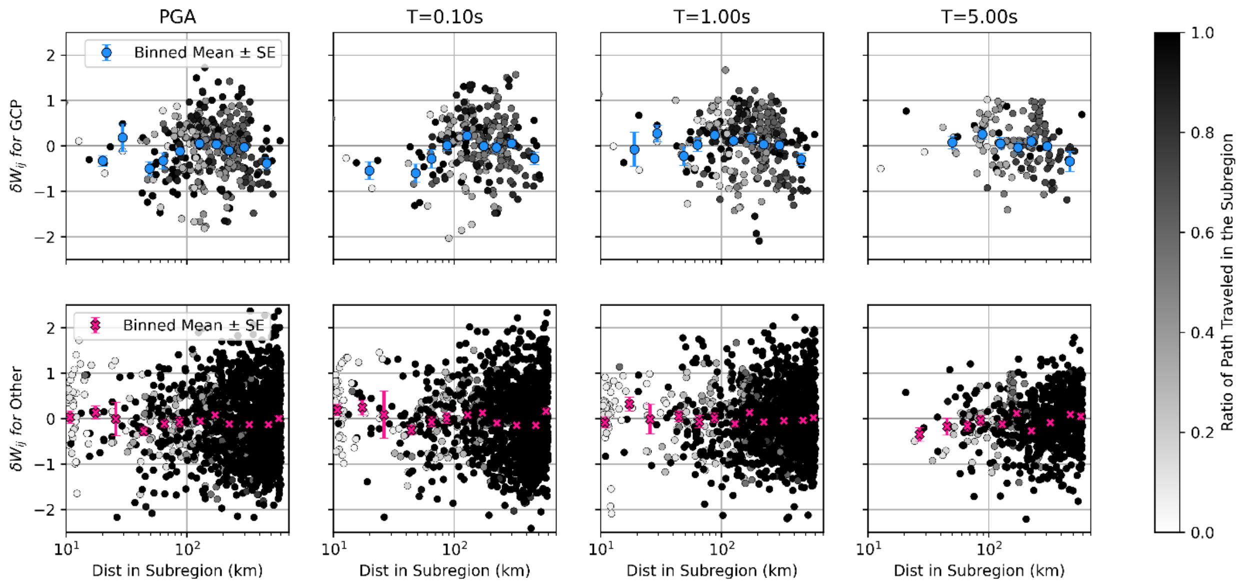


Figure 5.21. Within-event residuals (δW_{ij}) against the distance traveled in the subregion after the application of the smoothed coefficients of the subregional path model. The colored symbols represent binned means of the shaded data points. The level of shade depends on the percentage of the distance from the epicenter to the site that falls within the subregion.

For forward applications in hazard analyses where results are required for **M** > 6 earthquakes, it is necessary to merge the path adjustments developed in this chapter with the NGA-East reference rock model, which is denoted $Sa^{targ}(\boldsymbol{M}, R_{rup})$ (as presented in Chapter 4). Since the equation-based models used in the path adjustment calculations are practically equivalent to $Sa^{targ}(\boldsymbol{M}, R_{rup})$, and the anelastic attenuation effects are expected to be magnitude-independent, the implementation can be carried out using Eq. (5.4),

$$lnY(\mathbf{M}, R_{rup}) = \ln[Sa^{targ}(\boldsymbol{M}, R_{rup})] + (\textstyle\sum_{r=1}^{n} \Delta\gamma_r W_r)\, R_{rup} \tag{5.4}$$

where Sa^{targ} is the resulting *IM* from the tabulated GMM and $(\sum_{r=1}^{n} \Delta\gamma_r W_r)$ uses the coefficients in Table 5.1.

Within the NGA-East project, it was recognized that faster anelastic attenuation occurs in the GCP region (a different name was used) and two models for this effect were provided:

$$\ln\left(\frac{Sa_{GCP}}{Sa^{targ}}\right) = -0.00221 * max\left(0, R_{JB_{GCP}} - 100\right) \tag{5.5a}$$

$$\ln\left(\frac{Sa_{GCP}}{Sa^{targ}}\right) = \gamma(f) R_{JB_{GCP}} \tag{5.5b}$$

where Sa_{GCP} is the *IM* in the GCP and $R_{JB_{GCP}}$ is the Joyner–Boore distance traveled in the GCP.

Figure 5.22 shows the coefficients for the two NGA-East models next to the $\Delta\gamma_r$ parameter for GCP in Table 5.1. The present model has stronger period dependence than the NGA-East models, with stronger anelastic attenuation at short periods ($T <$ ~0.3 sec) and weaker at longer periods.

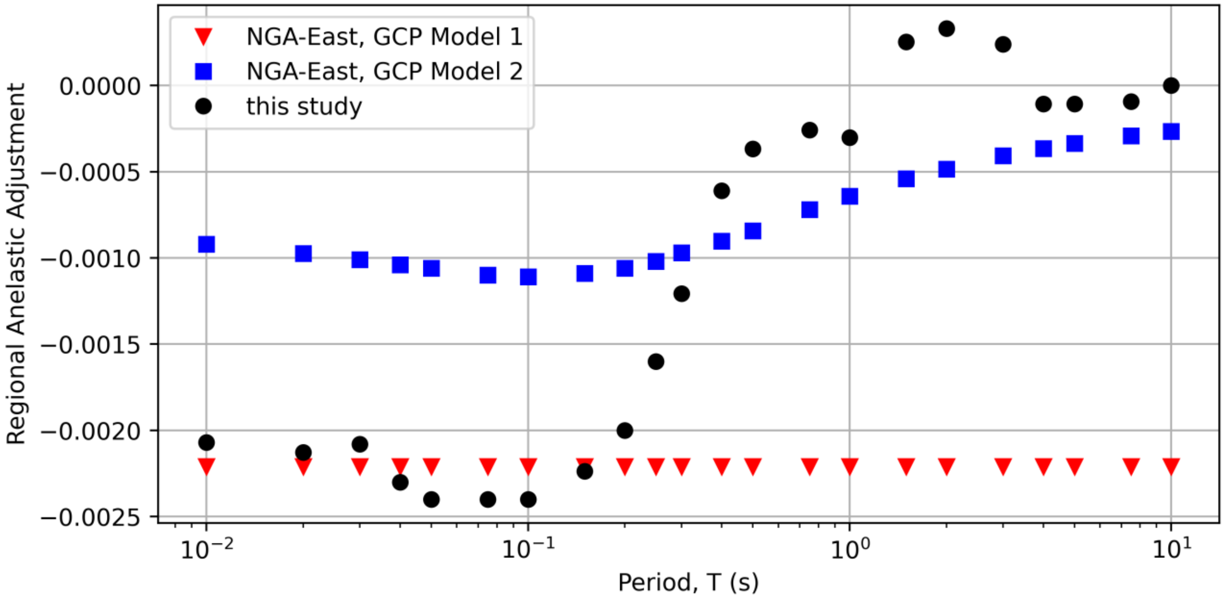


Figure 5.22. Comparison of period-dependence of anelastic attenuation coefficients from NGA-East project for GCP (Goulet et al. 2021a) to those developed in this study.

Figure 5.23 compares response spectra computed without a path correction (blue spectra), using the average of the two NGA-East path corrections (orange spectra), a record for which its full path is within *other* (W_{other} = 1, green spectra), and a record with its full path inside the GCP (W_{GCP}=1, red spectra). As expected, the present models reduce ground motions at short periods for large rupture distances, with the greatest reductions in GCP and generally small change outside of GCP.

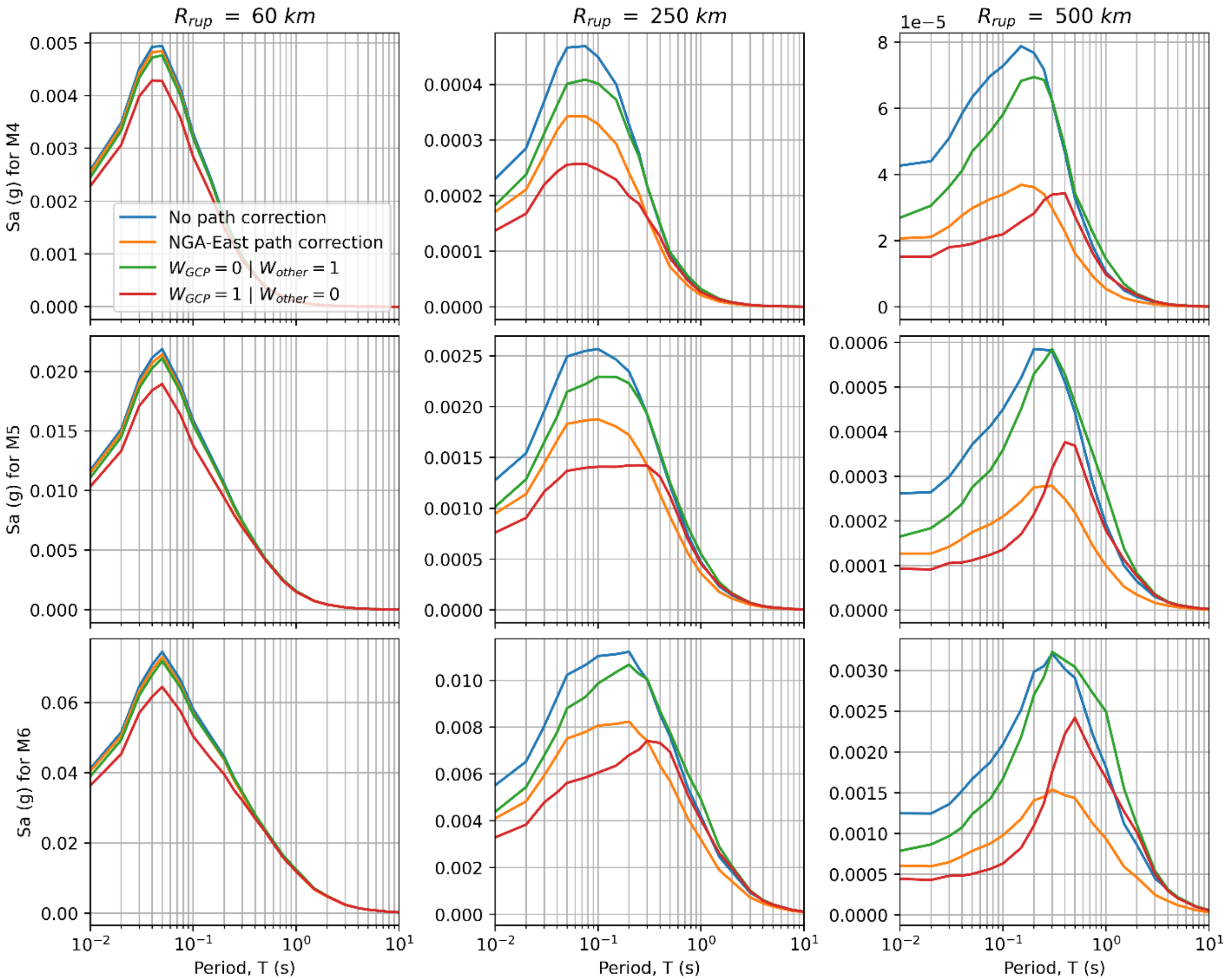


Figure 5.23. Spectral acceleration (Sa) computed by implementing different path corrections for selected magnitude and distances.

5.3 CENA-Wide and Regional Site Response

The remaining component of the model to be investigated is site response. To do so, I computed total residuals (Eq. 3.3) and performed mixed-effects regressions (Eqs. 3.4-3.5) using the function GMM (Eqs. 4.1-4.9) with the mean model misfit correction factors from Table 3.1 and path corrections from Eq. (5.1) and Table 5.1. Site terms (Eq. 3.5) represent the approximate misfits of the model for sites after misfits in the overall-, source- and path-model have been removed.

During the process of developing the path model (Section 5.2), I observed an upward trend of $\Delta e_{0,r}$ at long periods (Figure 5.18) in the coastal plains that could be a result of unmodeled depth effects. Hence, in this section I will also examine site terms relative to sediment-depths.

5.3.1 V_{S30}-scaling

Figure 5.24 shows binned means of site terms against V_{S30} values for four different *IM*s. These residuals include the CENA-wide average site response model (i.e., the F_V term in Eq. 3.2). Since the binned means are nearly zero and there is no visible trend, there is no need to modify the current V_{S30} scaling model for CENA.

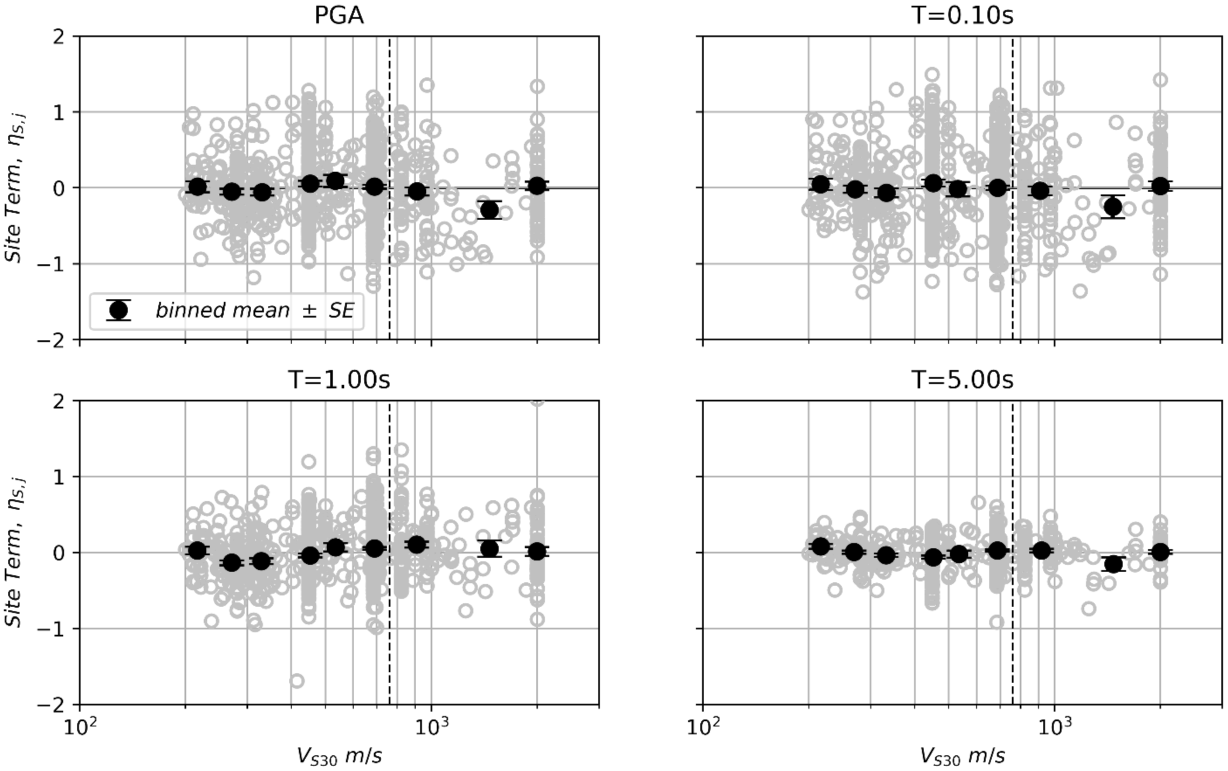


Figure 5.24. Site terms trends with V_{S30} for full dataset. Horizontal bars through binned means indicate ± one standard error of the mean.

In order to visualize the V_{S30}-scaling in my dataset, I recomputed residuals without applying the F_V term in Eq. (3.2) and plot the results with the model.

Figure 5.25 shows the resulting site terms compared to the F_V model for various *IM*s. Although the binned means do not match the mean model perfectly, they are within the epistemic uncertainty bounds provided by Stewart et al. (2020, herein as Sea20). My interpretation, as indicated above in the assessment of Figure 5.24, is that these results do not provide evidence for a need to adjust the existing V_{S30}-scaling model. The discrepancies between the mean model and the binned means are expected since the F_V model was not fitted only to ground motion data, instead they also considered the results of ground response simulations for CENA site conditions.

The uncertainty bounds shown in Figure 5.25 are intended to capture the range of models considered in the Sea20 study.

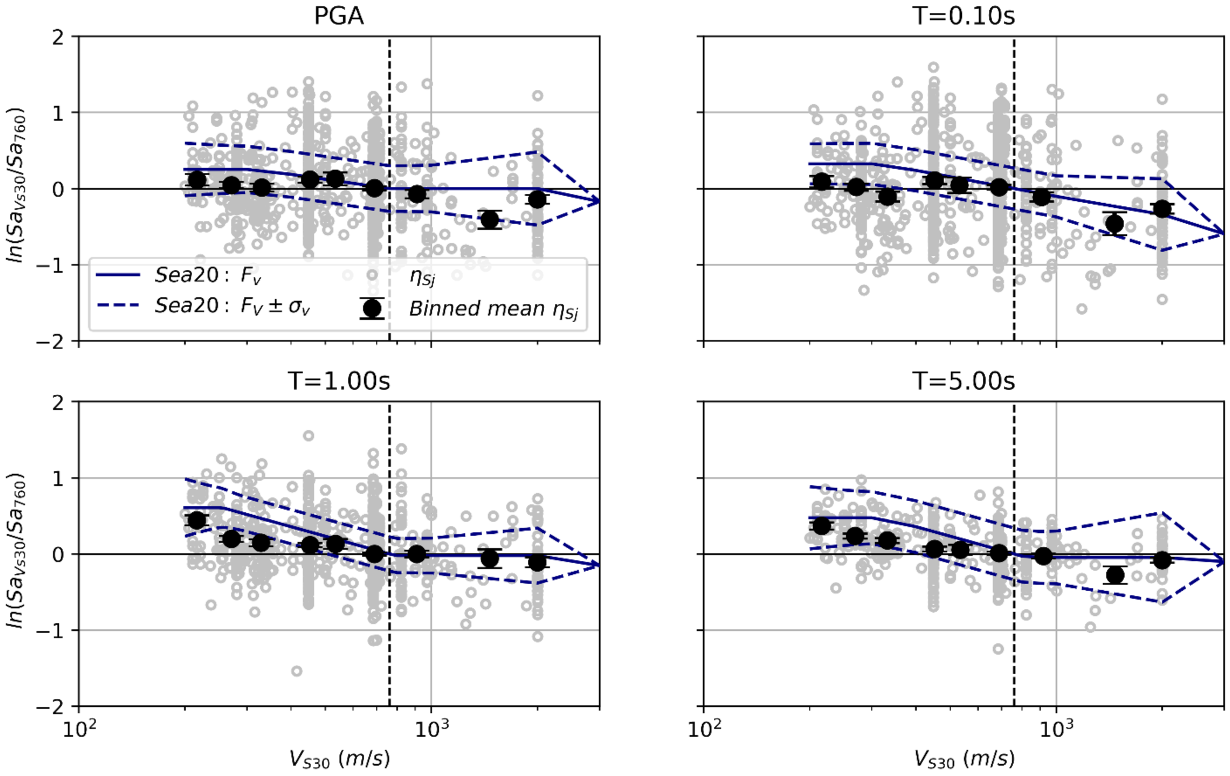


Figure 5.25. Within-event residuals and site terms against V_{S30} compared to Sea20 F_v model. The error bars depict the standard error.

5.3.2 Effects of Sediment Depth in the Coastal Plains

Current site amplification models for CENA do not include a depth term, mainly because depth parameters were not available for CENA sites at the time the site database for NGA-East was assembled (Parker et al. 2017, Goulet et al., 2021b). Recent studies (Chapman and Guo, 2021; Pratt and Schleicher, 2021; Boyd et al., 2024; and Akhani et al., 2024) have shown the importance of including depth as a site response parameter. Depth has also been widely considered in ergodic site response models for the NGA-West2 project (Abrahamson et al. 2014, BSSA14, Campbell and Bozorgnia 2014, Chiou and Youngs 2014), NGA-Sub project (Abrahamson and Gulerce 2022, Kuehn et al. 2020, Parker and Stewart 2022), for geomorphic provinces in southern California (Nweke et al. 2022) and for the Delta region in California (Buckreis et al. 2024).

Recently, Boyd et al. (2024) published depth maps of coastal plain sediments, shown in Figure 5.26, measured to a geologic horizon that depends on the type of bedrock existing in the region (Section 5.1.1). These maps provided sediment depths in the GCP and ACP regions but not elsewhere in CENA.

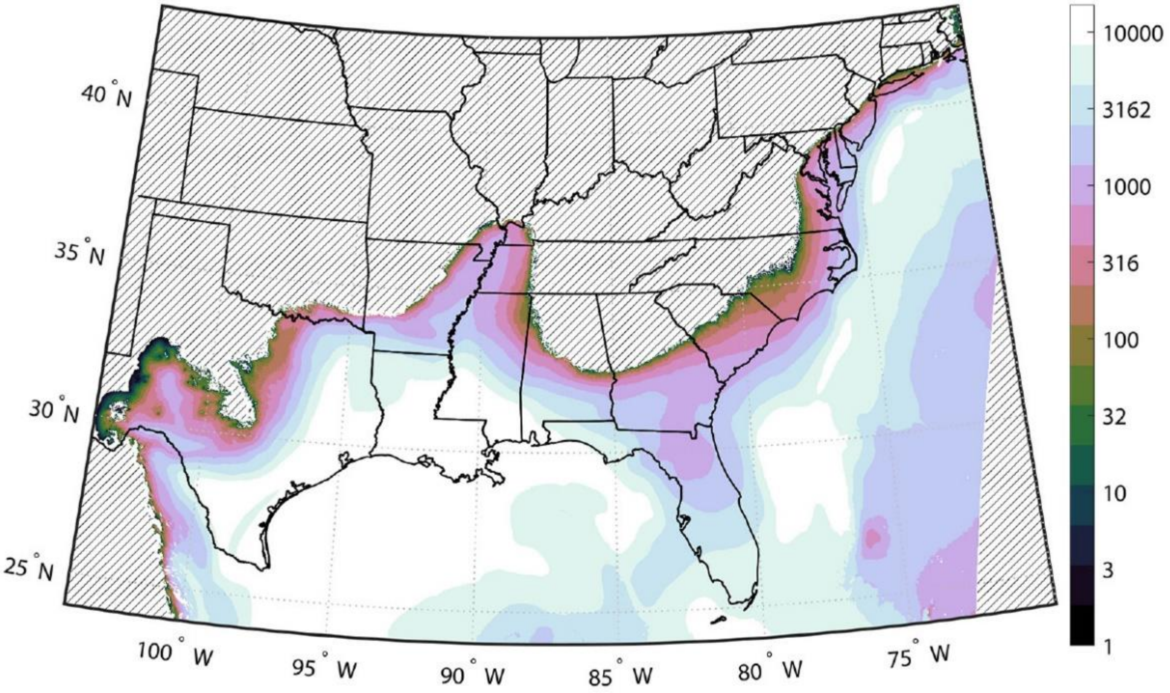


Figure 5.26. Map of coastal plain thicknesses as provided by Boyd et al. (2024).

The depths shown in Figure 5.26 have been used to develop depth-based site amplification models for CENA (e.g. Boyd et al., 2024; and Akhani et al., 2024). However, these recent models use depth directly, which is problematic because of the correlation of depth with V_{S30}. Lack of consideration of this correlation in model development makes the combined use of the F_V model

(Eq. 3.2) and a depth model likely to introduce bias. In the NGA-West2 models and Nweke et al. (2022), this problem was overcome using differential depth, which is the difference between a site-specific depth (e.g., as obtained for a given location using the Boyd et al. 2024 model) and the average depth conditioned on V_{S30}. I computed differential depth as

$$\delta \ln(z) = \ln(z) - \ln(\bar{z}) \tag{5.6}$$

where z is obtained from Boyd et al. (2024) and $\bar{z}$ is the estimated depth as a function of V_{S30} (Eq. 5.6). To develop a mean depth model, I plot in Figure 5.27 z-V_{S30} pairs, which are fitted using the functional form presented in Nweke et al. (2022),

$$\bar{z} = c_1 \left[1 + \operatorname{erf}\left(\frac{\log_{10} V_{S30} - \log_{10} \nu_\mu}{\nu_\sigma \sqrt{2}}\right)\right] + c_0 \tag{5.7}$$

where c_0 represents the deepest estimated depth (at the lowest V_{S30} values), c_1 is the slope of the scaling relationship, ν_μ defines the center where the slope is the steepest, and ν_σ represents the width of the ramp. The error function (erf) can be applied in most numerical software packages. The coefficients were fit to the CENA data and the resulting relationship is shown in Figure 5.27. The coefficients of the model are c_0 = 1.5352 km, c_1 = -0.6493 km, ν_μ = 440.4 m/s, ν_σ = 0.06.

I segregated the site terms according to the region where the ground-motion was recorded. Based on the discussion in Section 5.1.1, I am interested in examining how site response in the coastal plains deviate from the CENA-wide average and whether trends for the GCP deviate from trends in the ACP. It should be noted that the map provided by Boyd et al. (2024) has sediment-depth measurements for sites outside of the regions I define as GCP or ACP. These sites are labeled as "*Non-CP*". The distribution of site terms per region is shown in Figure 5.28.

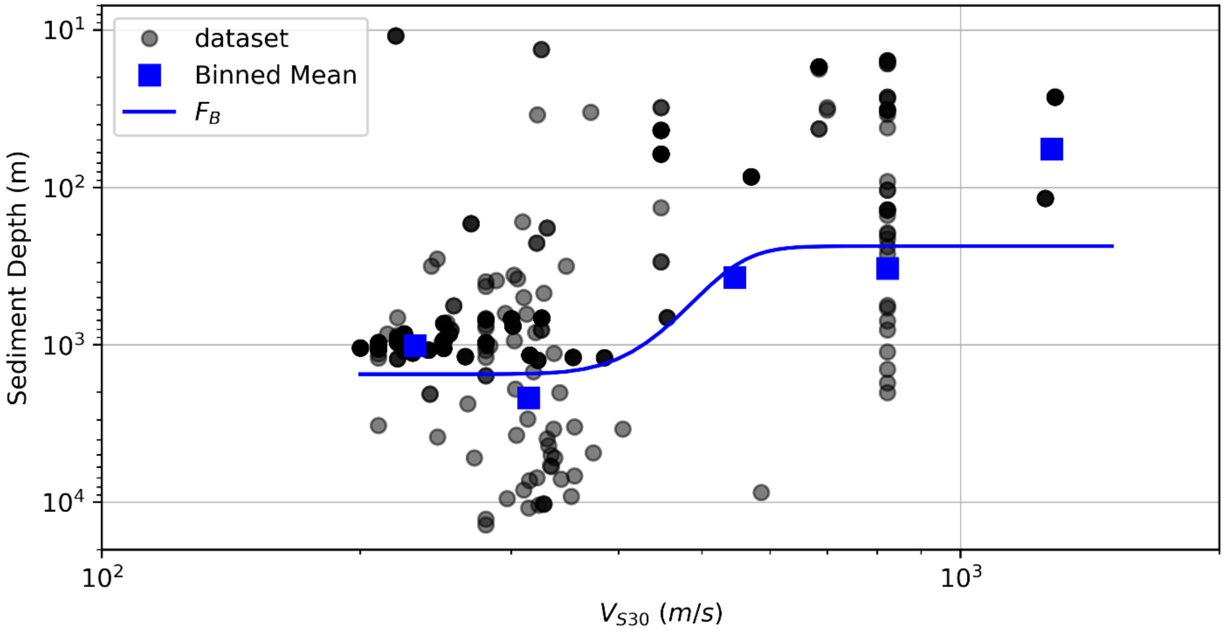


Figure 5.27. Correlation between V_{S30} and sediment depth as measured by Boyd et al. (2024).

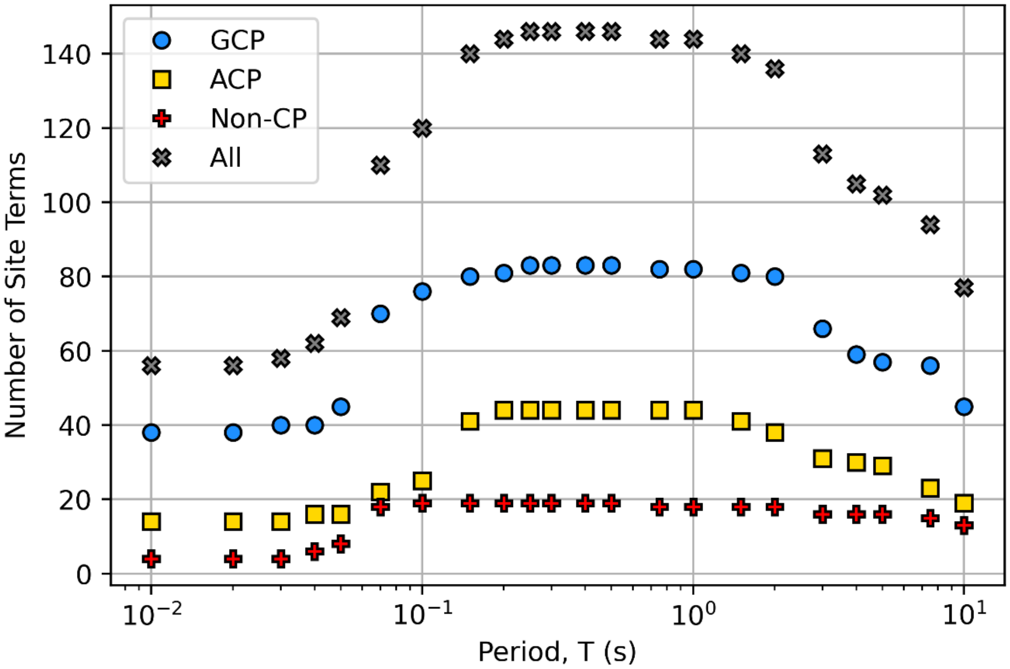


Figure 5.28. Site term distribution for three different regions and the combination of them all with depth measurements.

Figure 5.29 shows the variations of site terms with $\delta \ln(z)$, with one data point per site, for four *IM*s (PGA, *Sa* at 0.1, 1.0, and 5.0 sec). Results are shown for all sites and three groups: (1) non-CP (sites with depths in the Boyd et al. 2024 map but not in a coastal plain); (2) GCP; and (3) ACP. Considering first the GCP and ACP regions, relative to the V_{S30}-scaling model, different amplification trends are encountered for the two regions. ACP shows behavior that has been observed previously for western US basins (Nweke et al. 2022; Parker and Stewart 2022) in which shallower than average sites have higher short-period amplification (positive means of site terms) and essentially average long-period amplification (nearly zero means of site terms). For deeper-than-average ACP sites, smaller short period amplifications were found and larger long-period amplifications. For GCP, we see no bias of the V_{S30}-scaling model at short periods for negative and null $\delta \ln(z)$. For positive $\delta \ln(z)$, reduced amplifications are observed at short to intermediate periods and no bias relative to the depth model is found at long periods. Table 5.2 summarizes these observations. The average and higher amplifications for ACP and GCP observed at long periods coincide with the observed positive $\Delta e_{0,r}$ values in Figure 5.18 supporting the assumption that non-zero trends in $\Delta e_{0,r}$ could be interpreted as a shift in mean site terms.

Table 5.2. Summary of observed trends of site terms with respect to differential depth. *Deeper* stands for *deeper-than-average site* and *higher/smaller/average amp.* stands for *higher, smaller, or same as average amplifications as provided by the V_{S30}-scaling model*.

	Short periods	**Long periods**	**Region**
Shallower: $\delta \ln(z) < 0$	Higher amp.: $\eta_{S,j} > 0$	Average amp.: $\eta_{S,j} \approx 0$	**ACP**
Deeper: $\delta \ln(z) > 0$	Smaller amp.: $\eta_{S,j} < 0$	Higher amp.: $\eta_{S,j} > 0$	
Shallower: $\delta \ln(z) < 0$	Average amp.: $\eta_{S,j} \approx 0$	Higher amp.: $\eta_{S,j} > 0$	**GCP**
Deeper: $\delta \ln(z) > 0$	Smaller amp.: $\eta_{S,j} < 0$	Average amp. $\eta_{S,j} \approx 0$	

The trends for "all" sites are largely controlled by the combination of GCP and ACP sites, and are not discussed further. The data for the non-CP group are generally comparable to those for GCP. The non-CP sites are from stations in south Texas, which should have similar geologic features to the GCP. The presence of the non-CP group results from differences in the GCP boundaries between Boyd et al. (2024), which used a map by Garrity and Soller (2009), and the NGA-East project (and this study), which is based on a map presented by Pakiser and Mooney (1989).

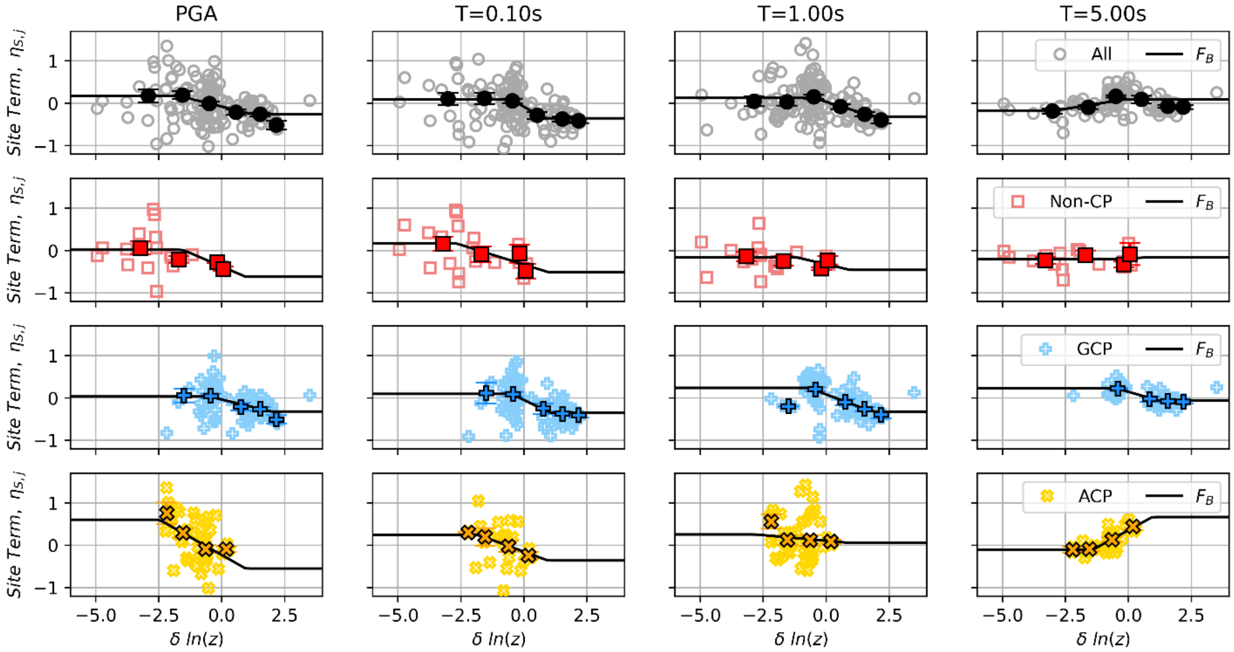


Figure 5.29. Site terms trends with differential depth for all the sites with depth measurement (All) and three subregions (GCP, ACP and Non-CP). Horizontal bars through binned means indicate ± one standard error of the mean.

I used the tri-linear function proposed by Nweke et al. (2022) to generate a site amplification model conditioned on differential depth,

$$F_B = \begin{cases} f_7 + f_6 f_8 & \delta\ln(z) < f_8 \\ f_7 + f_6\delta\ln(z) & f_8 \leq \delta\ln(z) \leq f_9 \\ f_7 + f_6 f_9 & \delta\ln(z) > f_9 \end{cases} \tag{5.8}$$

where f_7 is the ordinate at $\delta\ln(z) = 0$, f_6 is the slope, and f_8 and f_9 represent the lower and upper limits of the slope for shallower-than expected ($\delta\ln(z)<0$) and deeper-than expected ($\delta\ln(z)>0$) sites, respectively. The fits produced by Eq. (5.8) are shown in Figure 5.29, which appear to reasonable capture the trends in the data. Western models (e.g., Nweke et al., 2022; Parker and Stewart, 2022) show some similar trends to the ACP results in Figure 5.29, in the sense that at short periods, negative differential depths correlate with positive site terms, and at long periods, negative differential depths correlate with negative site terms, whereas the opposite occurs for positive differential depths (negative short-period site terms and positive long-period site terms). The trends observed for GCP have not been observed previously in other regions.

In Figure 5.30 I plot the coefficients from Eq. 5.8 with respect to period. Due to the lack of data for the *Non-CP* dataset, and the differences observed between the ACP and GCP model, I developed a depth model for the two aforementioned regions only. The smoothed coefficients for the model are also shown in Figure 5.30 and Table 5.3. Due to very limited data at short periods (which results from data screening related to the application of low-pass filters), the smoothed coefficients at short periods were selected based on the values for PGA.

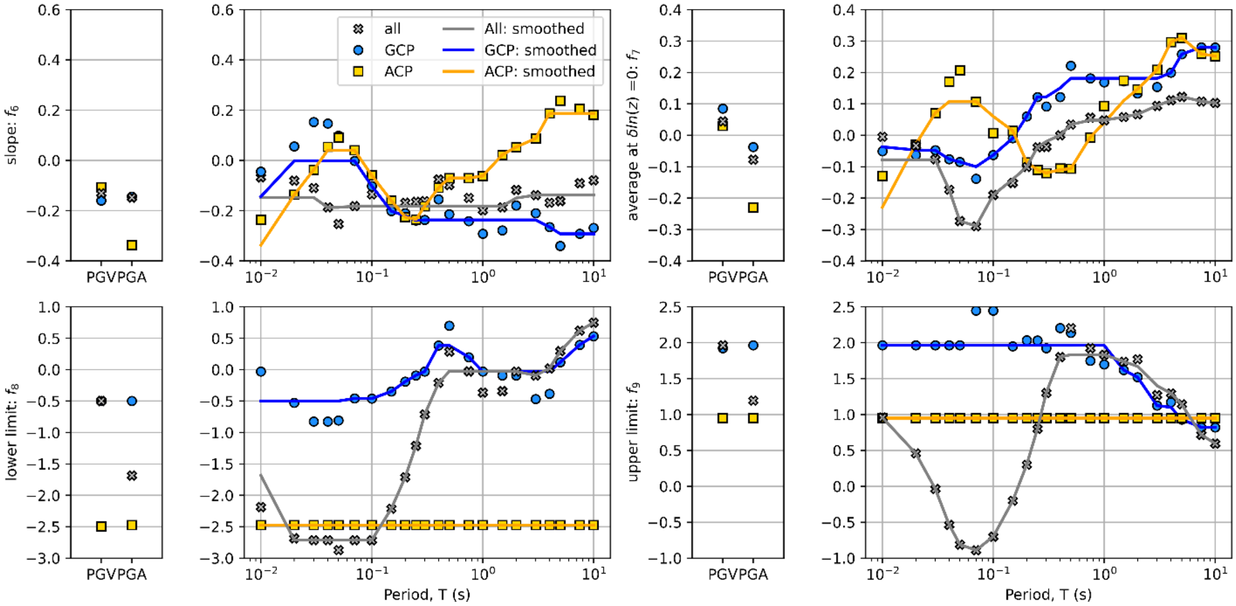


Figure 5.30. Depth model coefficients. Lines depict recommended period-dependent coefficients for the two coastal plain regions.

Table 5.3. Smoothed coefficients to be used with Eq. (5.8).

	GCP				**ACP**			
period	f_6	f_7	f_8	f_9	f_6	f_7	f_8	f_9
PGA	-0.144	-0.037	-0.5	1.963	-0.337	-0.229	-2.474	0.950
PGV	-0.159	0.085	-0.5	1.922	-0.106	0.03	-2.492	0.95
0.010	-0.144	-0.037	-0.5	1.963	-0.337	-0.229	-2.474	0.950
0.020	-0.002	-0.047	-0.5	1.963	-0.137	-0.029	-2.474	0.950
0.030	-0.002	-0.048	-0.5	1.963	-0.037	0.071	-2.474	0.950
0.040	-0.002	-0.076	-0.5	1.963	0.04	0.107	-2.474	0.950
0.050	-0.002	-0.085	-0.5	1.963	0.04	0.107	-2.474	0.950
0.075	-0.002	-0.1	-0.46	1.963	0.04	0.107	-2.474	0.950
0.10	-0.102	-0.063	-0.461	1.963	-0.06	0.06	-2.474	0.950
0.15	-0.202	-0.009	-0.349	1.963	-0.16	0.014	-2.474	0.950
0.20	-0.22	0.06	-0.192	1.963	-0.225	-0.086	-2.474	0.950
0.25	-0.239	0.122	-0.092	1.963	-0.235	-0.11	-2.474	0.950
0.30	-0.237	0.122	-0.03	1.963	-0.182	-0.12	-2.474	0.950
0.40	-0.237	0.15	0.383	1.963	-0.109	-0.11	-2.474	0.950
0.50	-0.237	0.181	0.383	1.963	-0.07	-0.106	-2.474	0.950
0.75	-0.237	0.181	0.198	1.747	-0.07	-0.006	-2.474	0.950
1.0	-0.237	0.181	-0.03	1.697	-0.062	0.04	-2.474	0.950
1.5	-0.237	0.181	-0.03	1.618	0.022	0.11	-2.474	0.950
2.0	-0.237	0.181	-0.03	1.519	0.053	0.147	-2.474	0.950
3.0	-0.237	0.181	-0.03	1.123	0.087	0.208	-2.474	0.950
4.0	-0.265	0.199	-0.03	1.1	0.187	0.297	-2.474	0.950
5.0	-0.292	0.258	0.115	0.93	0.187	0.31	-2.474	0.950
7.5	-0.292	0.28	0.395	0.822	0.187	0.259	-2.474	0.950
10.0	-0.292	0.279	0.532	0.822	0.187	0.252	-2.474	0.950

5.3.3. Residuals for the Final Model

In Chapter 3 I discussed the presence of overall misfits in the NGA-East central branch GMM and recommend correction factors to reduce the misfit. In Section 5.2 I use the function GMM to examine path-attenuation misfits for GCP and also developed a smoothed regional-path model to better capture faster than average anelastic attenuation in the GCP. In Section 5.3.2 I recommend a tri-linear function (Eq. 5.8) to model the dependence of site amplification on differential depths. By integrating all these elements into a combined model (Eq. 5.8), new residuals (total and partitioned) were computed.

$$lnY = F_E + F_Z + R_{rup}(\gamma + \sum_1^2 \Delta\gamma_r W_r) + F_S + \ln(c_k) + \ln(\Delta c_k) + \ln(F_B) \qquad (5.9)$$

where F_E, F_Z, F_S, Δc_k and F_B are Eqs. (4.2), (4.5), (3.1), (3.6) and (5.8), respectively. The smoothed coefficients are presented in Tables 3.1, 4.2, 5.1 and 5.3. The result of applying all these corrections to the NGA-East central branch GMM was successful in reducing misfit as shown in Figures 5.31 through 5.37. In Figure 5.31, the overall misfit resembles the inverse of Figure 3.9, nonetheless, if I do not apply the correction factors from Chapter 3, the overall misfit is even larger at long periods. This is demonstrated in Figure 5.32 where the overall bias was computed again without c_k correction factors. These results suggest that the bias factors presented in Chapter 3 were influenced by long period site response of CENA coastal plain sites and changing the site response model to address these effects would cause the bias factor to change. This is not surprising.

There are no visible trends of event terms with magnitude, within-event residuals with distance, and site terms with V_{S30}. I investigated the use of an iterative process whereby path coefficients are refined using the combined model (with path and site adjustments) followed by further adjustment of site models. Such processes failed to improve the path models due to weak data patterns that trend upward in some distance ranges and downward in others that cannot be captured with anelastic adjustment models (Eq. 5.2) and that produce no tangible impact on the site response results. For this reason, the model for which residuals are presented in this section (Eq. 5.9) is the final recommended model resulting from this research.

Finally, Figure 5.37 shows the within-event variability for residuals computed using the suggested model in this study and the NGA-East central branch GMM. The variability is reduced through the use of the model in Eq. (5.9).

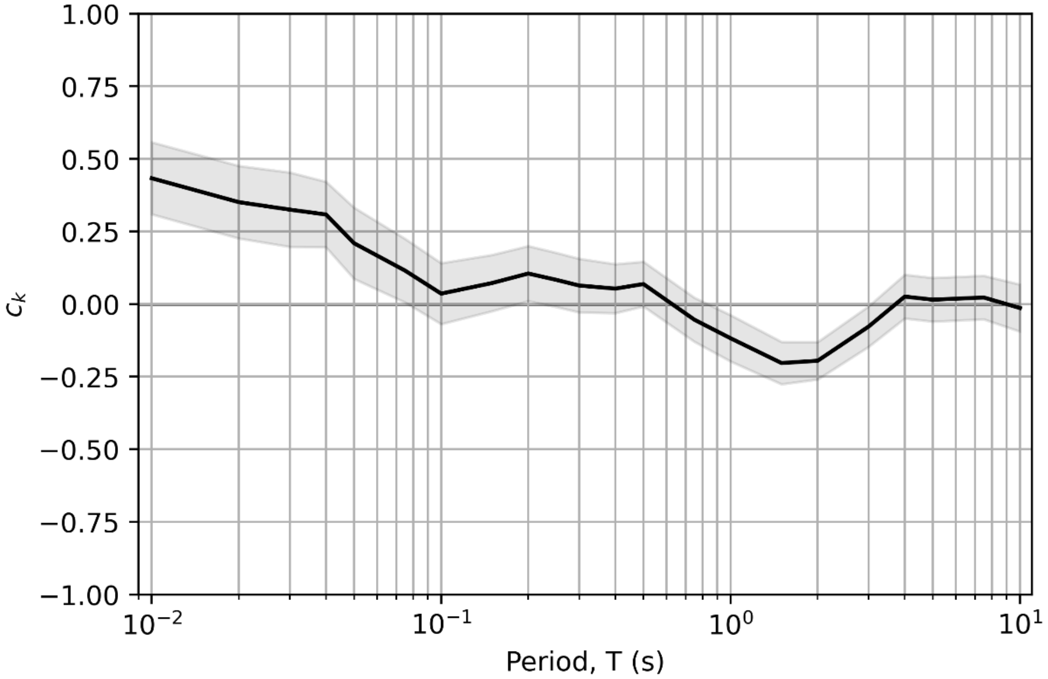


Figure 5.31. Period dependence of misfit term c_k for NGA-East central branch GMM in combination with the proposed model. The shaded regions enclose ± one standard error.

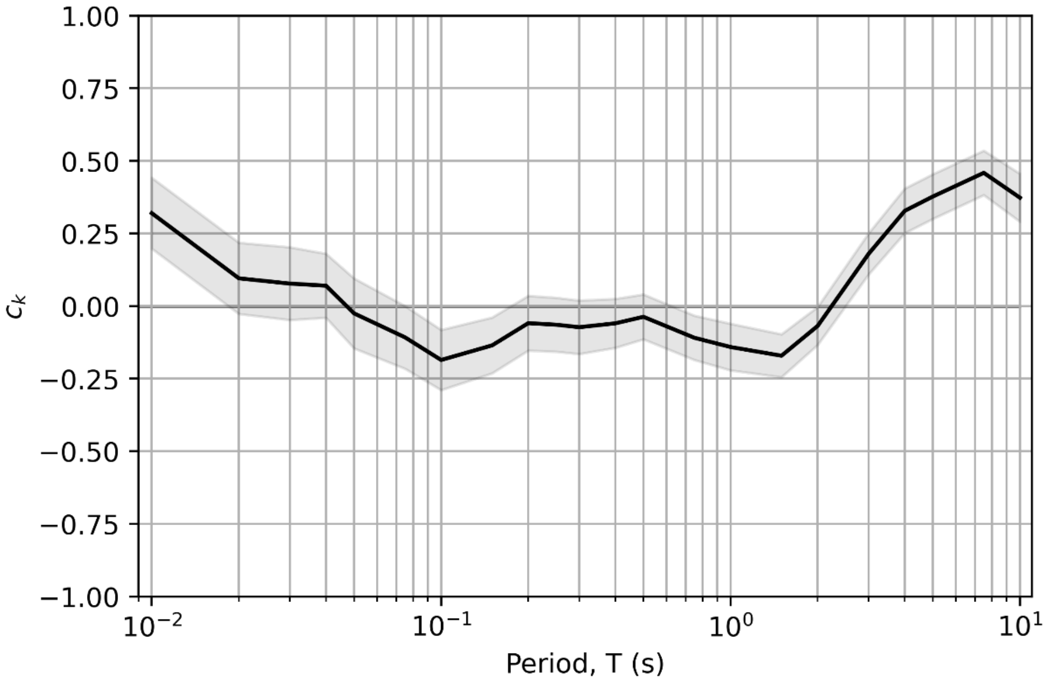


Figure 5.32. Period dependence of misfit term c_k for NGA-East central branch GMM in combination with the proposed model except for the c_k correction factors from Chapter 3. The shaded regions enclose ± one standard error.

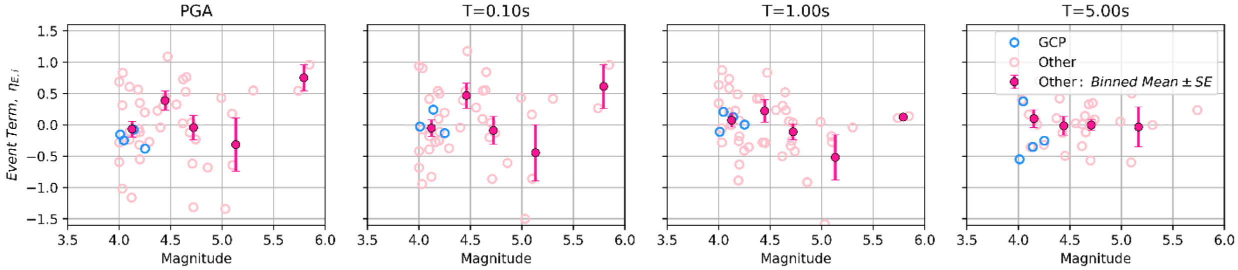


Figure 5.33. Magnitude-dependence of event terms for full dataset. Vertical bars through binned means indicate ± one standard error of the mean.

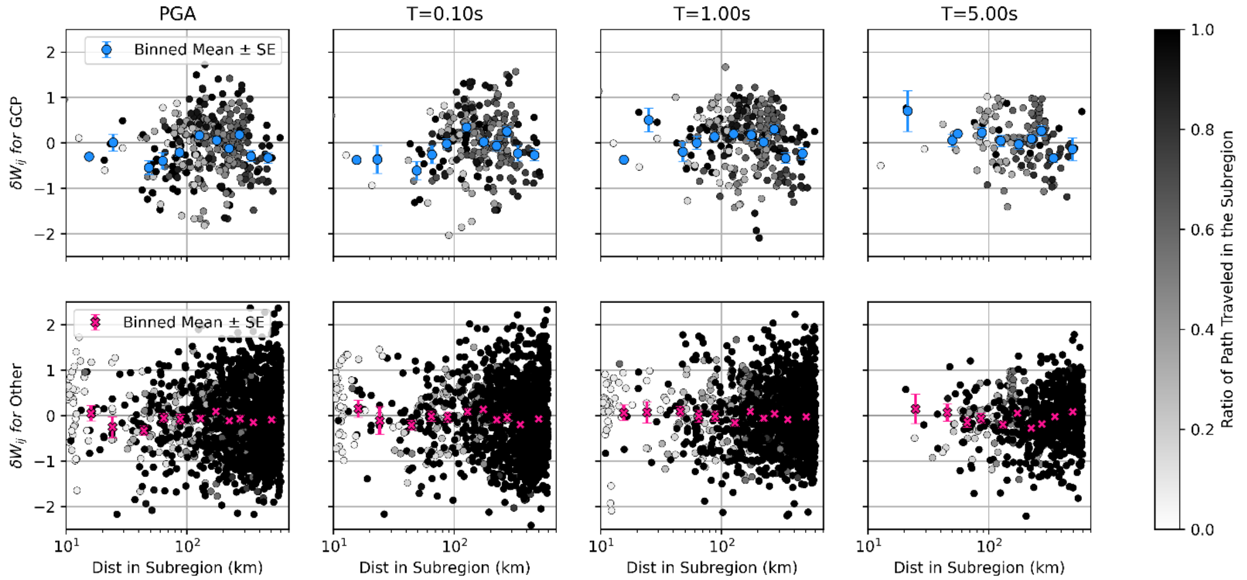


Figure 5.34. Distance-dependence of within-event residuals for full dataset. Vertical bars through binned means indicate ± one standard error of the mean.

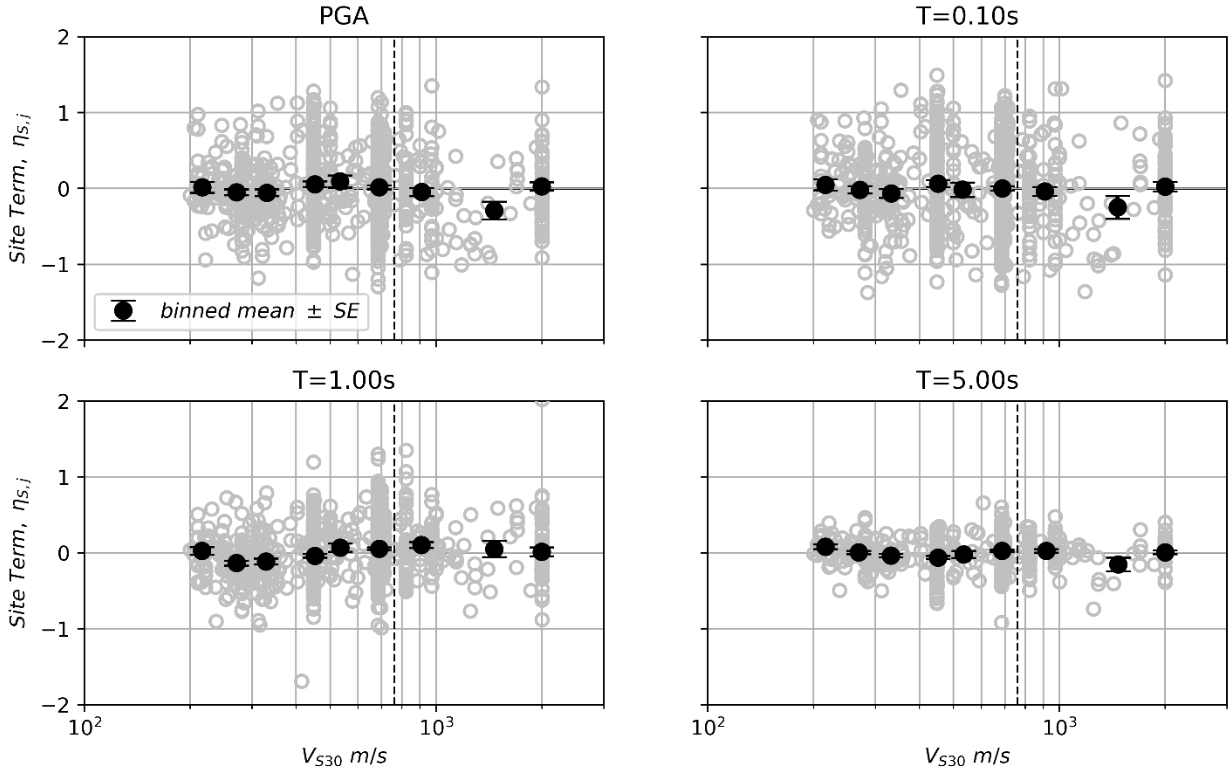


Figure 5.35. Site terms trends with V_{S30} for full dataset. Horizontal bars through binned means indicate ± one standard error of the mean.

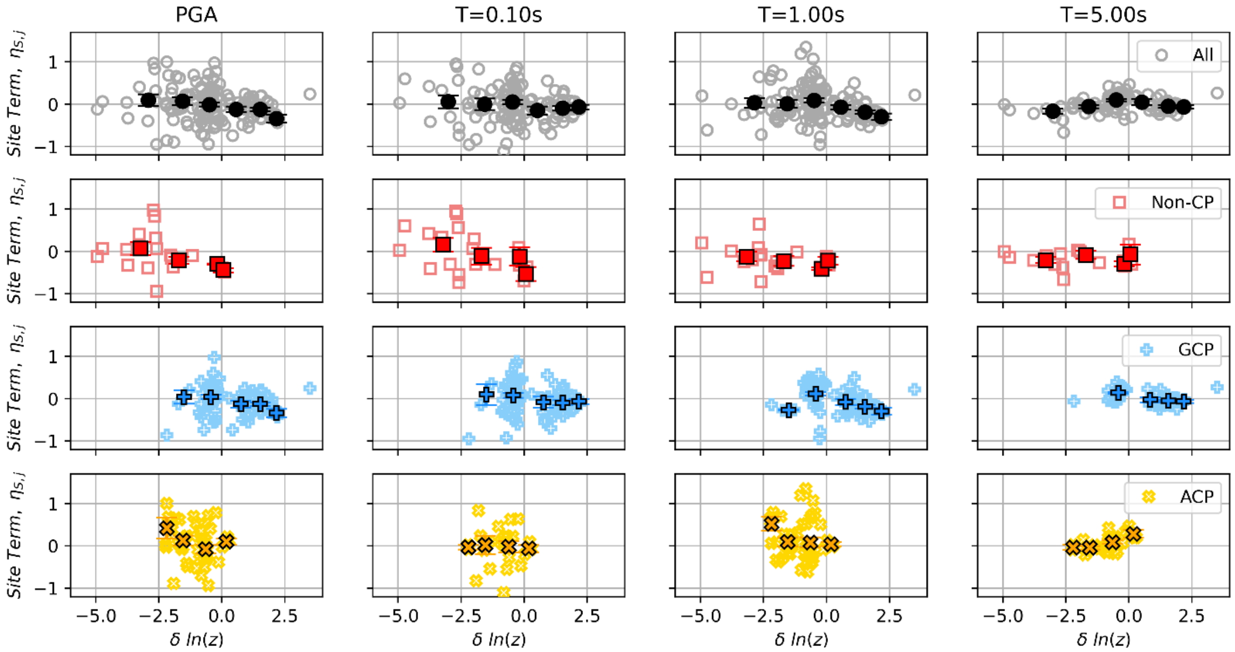


Figure 5.36. Site terms trends with differential depth for sites with sediment-depth measurement. Horizontal bars through binned means indicate ± one standard error of the mean.

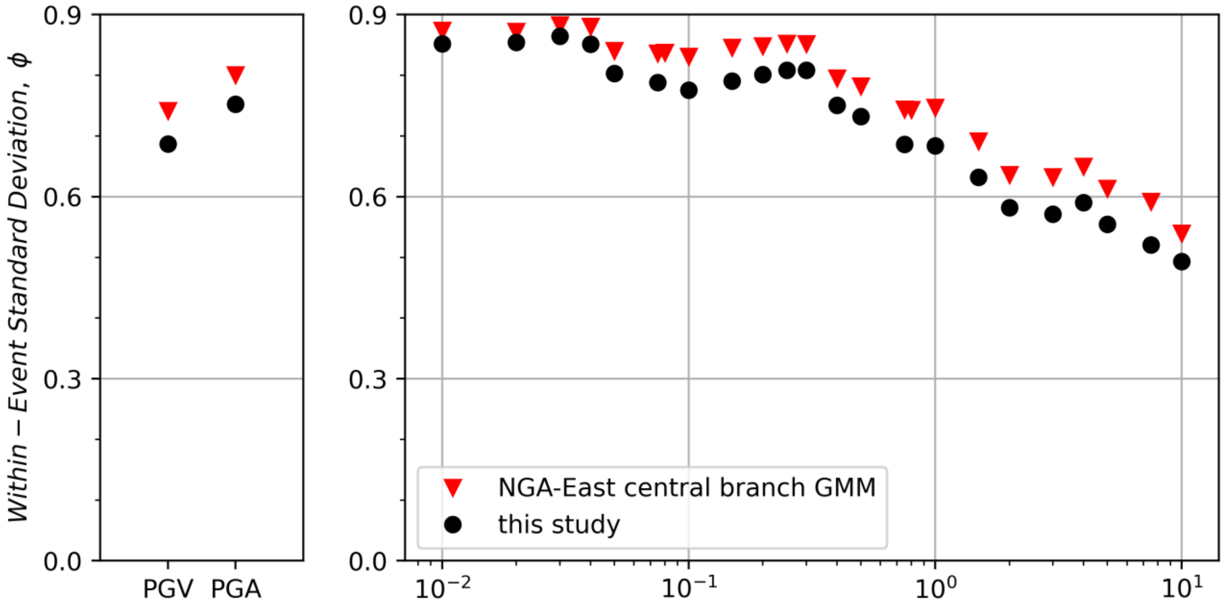


Figure 5.37. Within-event standard deviation (ϕ) for residuals computed using the suggested model in Eq. (5.9) and the tabulated NGA-East central branch GMM combined with the Sea20 site factors.

6. Summary and Conclusions

6.1 Scope of Research

A positive recent development in earthquake engineering and seismology is the increased density of high-quality seismometer and accelerometer arrays to record earthquake ground motions across the United States, including in Central and Eastern North America (CENA), which has traditionally been especially data-sparse. These arrays have dramatically increased the number of ground motions recorded in recent events.

In the first part of this work, I present efficient methods that can be used to facilitate record processing that produce comparable results to NGA processing procedures. At the time I started processing ground-motion data for CENA, manual procedures were the only accessible procedures that could be used that apply the processing standards developed during prior NGA projects. However, the challenge that I faced due to the large number of records that needed processing was the impracticability of manual procedures given the project timeline.

To address this problem, I worked with research partners and colleagues at the U.S. Geological Survey (USGS) to adapt their processing tool, gmprocess (Hearne et al., 2019; Thompson et al., 2024a,b), to conform with NGA protocols. Arguably the most critical step in manual or automated record processing is the selection of high-pass corner frequencies (f_{cHP}), which strongly impact displacements and may affect long-period response spectral ordinates. Prior to this effort, gmprocess applied criteria for selection of f_{cHP} based solely on signal-to-noise ratio (SNR). With collaborators, we developed procedures to find the smallest f_{cHP} for which the ratio between the displacement and a fitted third-order-polynomial is withing a certain threshold. The method results in an efficient, automated processing procedure that reduces long period noise. A limitation of the polynomial displacement fit method is that it is not suitable for cases with permanent displacement offset (from fling step or ground failure). This method assumes that the final displacement is zero. Gmprocess can be implemented by non-processing-experts for research purposes or for practical applications such as project-level non-ergodic site response analyses. A graphical user interface (GUI) that interfaces with gmprocess using Jupyter Notebooks is also provided to facilitate manual inspection of processed records for quality-control purposes.

I used the updated version of gmprocess with the GUI to process data from 100 events. Approximately 6892 additional three-component recordings were added to the dataset from NGA-East. All of the data was assembled into a relational database that can be accessed through Jupyter notebooks for efficient data analyses.

The USGS National Seismic Hazard Model (Petersen et al., 2020, 2023) uses reference site ground motion models (GMMs) from the NGA-East project (Goulet et al., 2021a; Youngs et al., 2021) and site amplification models recommended by an expert panel (Stewart et al., 2020; Hashash et al., 2020) to estimate ground motions in CENA. Due to asynchronicity in the development of these models, they are not fully compatible with each other. Using an expanded CENA dataset (relative to that used in the NGA-East project), I examine residuals of the recommended GMMs to evaluate whether the combination of the reference-rock model and site response model are compatible with the expanded NGA-East database.

To allow for examination of the possibility of regional path and site response features in CENA, I developed a function for the NGA-East central branch model with parametrized source and path components. This was needed because the NGA-East GMM is in the form of tables, hence, it does not provide the ability to adjust parameters. This version of the model is limited to $\mathbf{M} \leq 6$ and $R_{rup} \leq 600$km as it was intended to capture the range of empirical data. The fitting process was performed individually for 23 different periods, PGA and PGV and the fitting process was undertaken in four steps, which allowed coefficients for different ranges of R_{rup} to be estimated in a way that reduces the trade-offs between coefficients. First, I constrain parameters that control the intensity measures at close distances ($R_{rup} < 60$ km), then intermediate-distance geometric spreading parameter($R_{rup} > 60$ km), and then two more iterations for the complete range to refine the coefficients.

I used the equation-based reference rock model to perform residuals analyses, using a slightly enlarged version of the database. I reviewed a large body of literature on crustal structure in CENA and drew upon that body of knowledge to suggest five regions that could be considered for potentially distinct path and site response effects. I performed residuals analyses, to identify which of the regions had similar or distinct levels of anelastic attenuation. I performed some groupings and fit the trends for grouped regions. I also investigated site response effects CENA-wide in terms of V_{S30} and for the Gulf Coastal Plain (GCP) and Atlantic Coastal Plain (ACP) in terms of sediment depth.

6.2 Research Findings

The high-pass corner frequency selection tools developed in this research are integrated into a recently released version of gmprocess (Thompson et al. 2024a). The GUI is published in the USGS' GitLab repository for gmprocess (Thompson et al., 2024b) and is being actively used in ongoing research, including the NGA-West3 database (Buckreis et al. 2024).

Using the NGA-East central branch GMM, the residuals analyses presented in Chapter 3 indicate that for data outside of Texas, Oklahoma and Kansas, there is no evidence of misfit in the magnitude and distance scaling components of the GMM for $\mathbf{M} \geq 4$ events and $R_{rup} \leq 600$ km, with the exception of faster attenuation in the Gulf Coast region that manifests at distances > 300 km. The data demonstrates persistent period-dependent misfit for a wide range of alternate NGA-East GMMs and alternate data selection criteria (i.e., excluding data from particular regions). This misfit is towards overprediction at short periods and underprediction at long periods. Our findings suggest that the use of the site amplification model for active tectonic regions to adjust the NGA-East data is a strong contributor to the observed misfit. Considering various sources of uncertainties in the misfit calculation, I recommend smoothed correction factors for application in which all 17 NGA-East GMMs are used or where only the central branch GMM is used. I also provide epistemic uncertainties for these two cases. The resulting models were applied in the 2023 NSHM for CENA sites (Petersen et al. 2023).

The equation-based GMM provides a good match to the tabulated version. After the coefficients were smoothed, the mismatch between the function GMM and the tabulated version published by Goulet et al. (2021) is generally 0.1 or less with an average of approximately zero. This exercise was repeated for the individual 17 GMMs and the misfit was similar although higher for some GMMs.

Using the equation-based GMM, I investigated regional variations in anelastic attenuation and site response. I considered five regions, including the two coastal plains (GCP and ACP), the Yavapai and Mazatzal suture zones, and other areas. To reduce misfits, I developed a subregional path model that depends on the percentage of the path that is traveled within a certain region. I found a downward trend in within-event residuals at short periods for R_{rup}>200km for GCP but negligible trends elsewhere. The model coefficients required heavy smoothing since spectral shapes with multiple artificial peaks were observed. After the smoothing process, the subregional path model was found to capture the main features of the data. For subsequent analysis of site response, I truncated the dataset at R_{rup}=400km to avoid mapping path errors into site response.

I did not find evidence that the current CENA V_{S30}-scaling models need adjustments. Analyses of site terms with respect to differential depth revealed different trends for GCP and ACP. ACP shows that shallower-than-average sites have higher short-period amplification and essentially average long-period amplification. For deeper-than-average ACP sites, smaller short period amplifications were found and larger long-period amplifications. For GCP, we see no bias of the V_{S30}-scaling model at short periods for shallow or average depth sites. For deep sites, reduced amplifications are observed at short to intermediate periods and no bias relative to the V_{S30}-scaling model is found at long periods.

ACP shows behavior that has been observed previously for western US basins (Nweke et al. 2022; Parker and Stewart 2022) in which shallower than average sites have higher short-period amplification (positive means of site terms) and essentially average long-period amplification (nearly zero means of site terms). For deeper-than-average ACP sites, smaller short period amplifications were found and larger long-period amplifications. For GCP, we see no bias of the V_{S30}-scaling model at short periods for negative and null $\delta \ln(z)$ and also no bias at long periods for positive $\delta \ln(z)$. For positive $\delta \ln(z)$, reduced amplifications are observed at all periods.

6.3 Recommendations for Future Research

The relatively weak V_{S30}-scaling encountered in CENA is accompanied by relatively high site-to-site variability. This provides an opportunity for improvement in future research. What is needed is significantly improved information on site conditions, including more measured V_S profiles. A systematic program of measuring microtremor horizontal-to-vertical spectral ratios (HVSR) would also be extremely helpful, as it would allow characterization of resonance effects and their use in development of improved site response models. Models that consider both V_{S30} and HVSR in a combined format could significantly improve predictions.

If a future NGA-East project is undertaken, the problem of limited data at large magnitudes will remain, which can only be reasonably addressed through simulations. Consideration of epistemic uncertainties will also remain a challenge. If such a project were undertaken, an opportunity for improvement would be coordinated development of hard-rock models and site response models, so there would be no need for adjustment parameters.

7. Data Resources

Raw ground motion recordings from OK, TX, UU and ComCat networks were retrieved from FDSN data centers of IRIS Data Services using gmprocess (Hearne et al., 2019), which include the Oklahoma Geological Survey, 1978; Bureau of Economic Geology, The University of Texas at Austin, 2016; University of Utah, 1962; and U.S. Geological Survey, Earthquake Hazards Program, 2017. IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR-1851048.

The data used in this study are archived in a publicly web-serviced ground motion relational database (Buckreis et al., 2023). The ground motion database uses MySQL as the management system, and an application programming interface (API) was written to facilitate queries using URLs (https://uclageo.com/gm_database/api/index.php, last visited May 23, 2023). Ground motion records from NGA-East and newly added records are stored in the database under the "collection_id" = 3 and "user_id" = 2 respectively.

8. References

Aboye SA, Andrus RD, Ravichandran N, Bhuiyan AH, Harman N (2015) Seismic Site Factors and Design Response Spectra Based on Conditions in Charleston, South Carolina. *Earthquake Spectra*. 31(2):723-744 DOI:10.1193/041912EQS163M.

Abrahamson NA, Silva WJ, Kamai R (2014) Summary of the ASK14 ground motion relation for active crustal regions. *Earthquqake Spectra*. 30 (3): 1025-1055 DOI: 10.1193/070913EQS198M.

Abrahamson NA, Gulerce Z (2022) Summary of the Abrahamson and Gulerce NGA-SUB ground-motion model for subduction earthquakes. *Earthquake Spectra*. 38(4):2638-2681 DOI:10.1177/87552930221114374.

Akhani M, Davatgari-Tafreshi M, and Pezeshk S (2024) Adjusting Central and Eastern United States ground-motion models for use in the Coastal Plain considering the sediment thickness. *Earthquake Spectra*. 0(0) DOI:10.1177/87552930241258354.

Aki K (1967) Scaling law of seismic spectrum. *Journal of Geophysical Research*. 72: 1217–1231 DOI: 10.1029/jz072i004p01217.

Al Atik L, Gregor NJ, Abrahamson NA, Kottke AR (2022) GMPE-consistent hard-rock site adjustment factors for Western North America. *Earthquake Spectra*. 38(4): 2371-2397.

Alidadi N, Pezeshk S. (2024) Ground-Motion Model for Small-to-Moderate Potentially Induced Earthquakes Using an Ensemble Machine Learning Approach for CENA. *Bulletin of the Seismological Society of America.* 114 (4): 2202–2215DOI: 10.1785/0120230242.

Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BS-J, Wooddell KE, Graves RW, Kottke AR, Boore DM, Kishida T, Donahue JL (2014) NGA-West2 database. *Earthquake Spectra.* 30: 989-1005 DOI: 10.1193/070913EQS197M.

Atkinson GM, Mereu RF (1992) The shape of ground motion attenuation curves in southeastern Canada. *Bulletin of the Seismological Society of America.* 82 (5): 2014–2031 DOI: 10.1785/BSSA0820052014.

Atkinson GM (2004) Empirical attenuation of ground-motion spectral amplitudes in southeastern Canada and the northeastern United States: *Bulletin of the Seismological Society of America.* 94: 1079–1095.

Atkinson GM (2008) Ground-motion prediction equations for eastern North America from a referenced empirical approach—Implications for epistemic uncertainty. *Bulletin of the Seismological Society of America*. 98: 1304–1318 DOI:10.1785/0120070199.

Atkinson GM (2012) White paper on proposed ground-motion prediction equations (GMPEs) for 2015 national seismic hazard maps.

Atkinson GM, Boore DM (2006) Earthquake ground-motion prediction equations for eastern North America. *Bulletin of the Seismological Society of America*. 96: 2181–2205 DOI: 10.1785/0120050245.

Atkinson GM, Boore DM (2011) Modifications to Existing Ground-Motion Prediction Equations in Light of New Data. *Bulletin of the Seismological Society of America*. 101 (3): 1121–1135 DOI: 10.1785/0120100270.

Atkinson GM, Hassani B, Singh A, Yenier E, Assatourians K (2015) Estimation of moment magnitude and stress parameter from ShakeMap ground-motions. *Bulletin of the Seismological Society of America.* 105: 2572–2588

Boore DM (1983) Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. *Bulletin of the Seismological Society of America*. 73: 1865–1894 DOI:10.1785/BSSA07306A1865.

Bally AW (1989). Phanerozoic basins of North America, in The Geology of North America—An overview, A. W. Bally and A. R. Palmer (Editors), Vol. A, Geological Society of America, The Geology of North America, Boulder, Colorado, 397–446

Baltay AS, Abrahams LS, Hanks TC (2020) When source and path components trade-off in ground-motion prediction equations. *Seismological Research Letters*. *91*(4): 2259-2267.

Bates D, Maechler M, Bolker B, and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistic Software*. 67: 1–48.

Boatwright J, Seekins L (2011) Regional Spectral Analysis of Three Moderate Earthquakes in Northeastern North America. *Bulletin of the Seismological Society of America*. 101 (4): 1769–1782. DOI: 10.1785/0120100225.

Boore DM (2003) Simulation of ground motion using the stochastic method. *Pure and Applied Geophysics*. 160: 635–675 DOI:10.1007/978-3-0348-8010-7_10.

Boore DM (2005) On pads and filters: processing strong-motion data. *Bulletin of the Seismological Society of America*. 95:745-750. DOI: 10.1785/0120040160.

Boore DM, Bommer JJ (2005) Processing of strong-motion accelerograms: needs, options and consequences. *Soil Dynamics & Earthquake Engineering*. 25: 93–115 DOI: 10.1016/j.soildyn.2004.10.007.

Boore DM (2010) Orientation-independent, Non geometric-mean Measures of Seismic Intensity from Two Horizontal Components of Motion. *Bulletin Seismology Society America*. 100: 1830–1835.

Boore DM, Goulet CA (2014) The effect of sampling rate and anti-aliasing filters on high-frequency response spectra. *Bulletin of Earthquake Engineering.* 12: 203–216.

Boore DM (2015) 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, UC Berkeley (headquarters).

Boore DM, Campbell KW (2017) Adjusting central and eastern North America ground-motion intensity measures between sites with different reference-rock site conditions. *Bulletin of the Seismological Society of America.* 107: 132-148.

Boore DM (2018) Ground-motion models for very-hard-rock sites in eastern North America: An update. *Seismological Research Letter*. 89: 1172-1184.

Boore DM (2020) Revision of Boore (2018) Ground-motion predictions for central and eastern North America: Path and offset adjustments and extension to 200 m/s ≤ V_{S30} ≤ 3000 m/s. *Seismological Research Letters*. 91: 977-991.

Boore DM, Goulet CA (2014) The effect of sampling rate and anti-aliasing filters on high-frequency response spectra. *Bulletin of Earthquake Engineering.* 12:203–216.

Boyd OS, Churchwell D, Moschetti MP, Thompson EM, Chapman MC, Ilhan O, Pratt TL, Ahdi SK, Rezaeian S (2023) Sediment Thickness Map of Atlantic and Gulf Coastal Plain Strata, Central and Eastern U.S., and Their Influence on Earthquake Ground Motions. *Earthquake Spectra*. DOI: 10.1177/87552930231204880.

Brandenberg SJ, Yang Y (2022) "ucla_geotech_tools: A set of Python packages developed by the UCLA geotechnical group" (Version 1.1.0) [Computer software] DOI: 10.5281/zenodo.5621169.

Brune JN (1970) Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research*. 75:4997–5002. DOI: 10.1029/JB075i026p04997

Buckreis TE (2022) Customization of Path and Site Response Components of Global Ground Motion Models for Application in Sacramento-San Joaquin Delta Region of California [Dissertation]. University of California, Los Angeles

Buckreis TE, Nweke CC, Wang P, Brandenberg SJ, Mazzoni S and Stewart JP (2023) Relational database for California strong ground motions, Geo-Congress 2023: Geotechnical Data Analysis and Computation, Los Angeles, CA, March 2023. *Geotechnical Special Publication No. 342*. EM Rathje, B Montoya, and MH Wayne (Eds.), 461-470, ASCE Geo-Institute.

Buckreis TE, JP Stewart, R Shams, S Mohammed, M Li, T Kishida, G Lanzano, L Luzi, P Zimmaro, BA Bradley, J-Y Huang, CC Nweke, SJ Brandenberg, JL Donahue, Y Bozorgnia (2024) Overview of the NGA-West3 Ground Motion Database. 2024 SCEC Annual Meeting, Palm Springs, CA, abstract.

Budnitz RJ, Apostolakis G, Boore DM, Cluff LS, Coppersmith KJ, Cornell CA, Morris PA (1997) Recommendations for probabilistic seismic hazard analysis: Guidance on uncertainty and use of experts, US Nuclear Regulatory Commission. *Report NUREG/CR-6372*.

Bureau of Economic Geology, The University of Texas at Austin (2016) Texas Seismological Network [Data set]. International Federation of Digital Seismograph Networks. DOI: 10.7914/SN/TX

Burger RW, Somerville PG, Barker JS, Herrmann RB, Helmberger DV (1987) The effect of crustal structure on strong ground motion attenuation relations in eastern North America. *Bulletin of the Seismological Society of America.* 77 (2): 420–439 DOI: 10.1785/BSSA0770020420.

Campbell KW (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. *Bulleting of Seismologic Society of America.* 93: 1012–1033.

Campbell, KW, Bozorgnia Y (2014) NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthquake Spectra*. 30 (3): 1087-1115 DOI: 10.1193/062913EQS175M.

Centella D, Li M, Ramos-Sepulveda ME, Ilhan O, Hashash YMA, Rathje EM, Stewart JP (2023) Calibration of NGA-East GMMs and Site Amplification Models Against CENA Ground Motions. *Report to USGS External Research Program,* UIUC, UT, UCLA, Sept. 30. https://earthquake.usgs.gov/cfusion/external_grants/reports/G21AP10367.pdf

Chapman M, Conn A (2016) A model for Lg propagation in the Gulf Coastal Plain of the southern United States. *Bulletin of the Seismological Society of America.* 106(2): 349–363 DOI: 10.1785/0120150197.

Chapman MC, Guo Z (2021) A response spectral ratio model to account for amplification and attenuation effects in the Atlantic and Gulf Coastal Plain. *Bulletin of the Seismological Society of America*. 111: 1849–1867.

Chiou BS-J, Youngs RR (2008) NGA model for average horizontal component of peak ground motion and response spectra. Report no. 2008/09, *November. Berkeley, CA: Pacific Earthquake Engineering Research Center (PEER)*. November 2008. UC Berkeley.

Chiou B, Youngs R (2014) Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra.* 30 (3): 1117-1153 DOI: 10.1193/072813EQS219M.

Contreras V, Stewart JP, Kishida T, Darragh RB, Chiou BSJ, Mazzoni S, Youngs RR, Kuehn NM, Ahdi SK, Wooddell K, Boroschek R, Rojas F, Ordenes J (2022) NGA-Sub source and path database. *Earthquake Spectra*. 38: 799-840.

Darragh RB, Abrahamson NA, Silva WJ, and Gregor N (2015) Development of Hard Rock Ground-Motion Models for Region 2 of Central and Eastern North America. *PEERReport 2015/04*

Douglas J, Boore DM (2011) High-frequency filtering of strong-motion records. *Bulletin of Earthquake Engineering.* 9: 395–409 DOI: 10.1007/s10518-010-9208-4.

Dreiling J, Isken MP, Mooney WD, Chapman MC, Godbee RW (2014) NGA-East regionalization report: Comparison of four crustal regions within central and eastern North America using waveform modeling and 5%-damped pseudo-spectral acceleration response, *PEER Report 2014/15*. Pacific Earthquake Engineering Research Center, UC Berkeley (headquarters).

Dreiling J, Isken MP, Mooney WD (2016) Comparison of Synthetic Pseudoabsolute Response Spectral Acceleration (PSA) for Four Crustal Regions within Central and Eastern North America (CENA). *Bulletin of the Seismological Society of America*. 107 (1): 169–179 DOI: 10.1785/0120160121.

Dziewonski AM, Chou TA, and Woodhouse JH (1981) Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research.* 86: 2825-2852 DOI:10.1029/JB086iB04p02825.

Ekström G, Nettles M, and Dziewonski AM (2012) The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*. 200–201: 1-9 DOI: 10.1016/j.pepi.2012.04.002.Elzhov TV, Mullen KM, Spiess A, Bolker BM (2015) Title R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds.

Frankel AD, Mueller CS, Barnhard T, Perkins D, Leyendecker EV, Hanson S, Hopper M (1996) National seismic hazard maps: Documentation. *Open-file report no. 96-532*. U.S. Geological Survey, Reston, VA.

Gallegos A, Ranasinghe N, Ni J, Sandvol E (2014) Lg attenuation in the central and eastern United States as revealed by the EarthScope Transportable Array. *Earth and Planetary Science Letters.* 402:187-196.

Garrity CP, Soller DR (2009) Database of the Geologic Map of North America; adapted from the map by J.C. Reed, Jr. and others (2005): U.S. Geological Survey Data Series 424 [https://pubs.usgs.gov/ds/424/]

Goulet CA, Kishida T, Cramer CH, Darragh RB, Silva WJ, Hashash YMA, Harmon J, Stewart JP, Wooddell KE and Youngs RR (2014) PEER NGA-East Database. *Report 2014/17*. Pacific Earthquake Engineering Research Center, UC Berkeley (headquarters).

Goulet CA, Bozorgnia Y, Abrahamson NA, Kuehn N, Al Atik L, Youngs RR, Graves RW, Atkinson GM (2018) Central and Eastern North America ground motion characterization. *Report 2018/08*. Pacific Earthquake Engineering Research Center, UC Berkeley (headquarters).

Goulet CA, Bozorgnia Y, Kuehn N, Al Atik L, Youngs RR, Graves RW, Atkinson GM (2021a) NGA-East Ground-Motion Characterization model part I: Summary of products and model development. *Earthquake Spectra.* 37: 1231–1282.

Goulet CA, Kishida T, Ancheta TD, Cramer CH, Darragh RB, Silva WJ, Hashash YMA, Harmon J, Parker GA, Stewart JP, Youngs RR (2021b) PEER NGA-East database. *Earthquake Spectra*. 37: 1331-1353.

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. 51–69.

Graizer V (2016) Ground-motion prediction equations for central and eastern North America. *Bulletin of the Seismological Society of America.* 106: 1600–1612.

Graizer V (2017) Alternative (G-16v2) ground-motion prediction equations for central and eastern North America. *Bulletin of the Seismological Society of America.* 107(2): 869–886.

Guo Z, Chapman MC (2019) An examination of amplification and attenuation effects in the Atlantic and Gulf Coastal Plain using spectral ratios. *Bulletin of the Seismological Society of America* 109: 1855–1877.

Guy MR, Patton JM, Fee J, Hearne M, Martinez E, Ketchum D, Worden C, Quitoriano V, Hunter E, Smoczyk G, Schwarz S (2015) National Earthquake Information Center systems overview and integration. *U.S. Geological Survey Open-File Report 2015-1120.* DOI: 10.3133/ofr20151120.

Harmon J, Hashash YMA, Stewart JP, Rathje EM, Campbell KW, Silva WJ, Xi B, Ilhan O (2019a) Site amplification functions for Central and Eastern North America – Part I: Simulation data set development. *Earthquake Spectra*. 35: 787–814.

Harmon J, Hashash YMA, Stewart JP, Rathje EM, Campbell KW, Silva WJ, Ilhan O (2019b) Site amplification functions for Central and Eastern North America – Part II: Modular simulation-based models. *Earthquake Spectra*. 35: 815–847.

Hashash YMA, Kottke AR, Stewart JP, Campbell KW, Kim B, Moss C, Nikolaou S, Rathje EM, and Silva WJ (2014) Reference rock site condition for Central and Eastern North America. *Bulletin of the Seismological Society of America.* 104: 684–701.

Hashash YMA, Ilhan O, Harmon JA, Parker GA, Stewart JP, Rathje EM, Campbell KW, Silva WJ (2020) Nonlinear site amplification model for ergodic seismic hazard analysis in Central and Eastern North America. *Earthquake Spectra*. 36: 69–86.

Hassani B, Atkinson GM (2017) Site-effects model for central and eastern North America based on peak frequency and average shear-wave velocity. *Bulletin of the Seismological Society of America.* 108: 338–350.

Hearne M, Thompson EM, Schovanec H, Rekoske J, Aagaard BT, Worden CB (2019) USGS automated ground-motion processing software, USGS Software Release. DOI: 10.5066/P9ANQXN3.

Huddlestun PF (1988) A revision of the lithostratigraphic units of the Coastal Plain of Georgia—The Miocene through Holocene. *Georgia Geologic Survey, Bulletin.* 104: 162.

Huddlestun PF, Hetrick JH (1986) Upper Eocene stratigraphy of central and eastern Georgia. *Georgia Geologic Survey, Bulletin*. 95: 78.

Hwang HHM, Lin H, Huo J-R (1997) Site coefficients for design of buildings in eastern United States. *Soil Dynamics and Earthquake Engineering*. 16: 29–40.

Ilhan O, Hashash YMA, Stewart JP, Rathje EM, Nikolaou S, Campbell KW (2024) Simulated site amplification for Central and Eastern North America: Data set development and amplification models. *Earthquake Spectra*. DOI:10.1177/87552930231215723.

Ji C, Cabas A, Kottke A, Pilz M, Macedo J, Liu C (2023) A DesignSafe earthquake ground motion database for California and surrounding regions. *Earthquake Spectra*. 39(1):702-721. DOI:10.1177/87552930221141108.

Khaheshi Banab K, Kolaj M, Motazedian D, Sivathayalan S, Hunter JA, Crow HL, Pugin AJM, Brooks GR, Pyne M (2012) Seismic site response analysis for Ottawa, Canada: A comprehensive study using measurements and numerical simulations. *Bulletin of the Seismological Society of America.* 102: 1976–1993.

Kishida T, Darragh RB, Chiou BS, Bozorgnia Y, Mazzoni S, Contreras V, Boroschek R, Rojas F, Stewart JP (2020) Chapter 3: Ground Motions and Intensity Measures, in Data Resources for NGA-Subduction Project, *PEER Report 2020/02*, J.P. Stewart (editor), Pacific Earthquake Engineering Research Center, University of California, Berkeley, California (headquarters).

Kluyver T, Ragan-Kelly B, Pérez F, Granger B, Bussonnier M, Frederic J, Kelley K, Hamrick J, Grout J, Corlay S, Ivanov P, Avila D, Abdalla S, Willing C, Jupyter Development Team (2016) Jupyter Notebooks – a publishing format for reproducible computational workflows. Loizides, Fernando, and Scmidt, Birgit (eds.) In *Positioning and Power in Academic Publishing: Players, Agents, and Agendas*. IOS Press. 87-90 DOI: 10.3233/978-1-61499-649-1-87.

Krischer L, Megies T, Barsch R, Beyreuther M, Lecocq T, Caudron C, Wassermann J (2015) ObsPy: a bridge for seismology into the scientific Python ecosystem. *Computational Science & Discovery*. 8(1) DOI:10.1088/1749-4699/8/1/014003.

Ktenidou OJ, Abrahamson NA (2016) Empirical estimation of high-frequency ground motion on hard rock. *Seismological Research Letters.* 87(6): 1465–1478.

Kuehn NM, Kishida T, AlHamaydeh M, Lavrentiadis G, Bozorgnia Y (2020) A Bayseian model for truncated regression for the estimation of empirical ground-motion models. *Bulletin of the Seismological Society of America.* 18 (14), 6149-6179 DOI: 10.1007/s10518-020-00943-8.

Leonard M (2014) Self-consistent earthquake fault-scaling relations: Update and extension to stable continental strike-slip faults. *Bulletin of the Seismological Society of America.* 104: 2953–2965.

Li M (2022) Enhancements in earthquake ground motion assessment through regional V_{s30} characterization and V_{s30} scaling, Ph.D. Dissertation, University of Texas.

Li M, Rathje EM, Cox BR, Yust M (2022) A Texas-specific V_{s30} map incorporating geology and V_{S30} observations. *Earthquake Spectra*. 38(1): 521-542.

Li M, Rathje EM, Ramos-Sepúlveda M, Stewart J, Ilhan O and Hashash Y (2023) Evaluating bias of NGA-East GMMs and site factors for ground motions from natural and potentially induced earthquakes in Texas, Oklahoma, and Kansas. SSA 2023: USGS National Seismic Hazard Models: 2023 and Beyond. San Juan, PR April 2023.

Li M, Rathje EM, Stewart JS, Ramos-Sepulveda ME, Hashash YMA (submitted). Regional Adjustments to the NGA-East GMM for Texas, Oklahoma and Kansas. *Earthquake Spectra*.

Maher JC (1965) Correlations of subsurface Mesozoic and Cenozoic rocks along the Atlantic Coast. *American Association of Petroleum Geologists*. 3: 18.

Marshak S, Domrois S, Abert C, Larson T, Pavlis G, Hamburger M, Yang X, Gilbert H, Chen C (2017) The basement revealed: Tectonic insight from a digital elevation model of the Great Unconformity, USA cratonic platform. *Geology* 45(5): 391–394 DOI: 10.1130/G38875.1

Mooney WD, Kaban MK (2010) North American upper mantle: Density, composition and evolution. *Journal Geophysical Research*. 115: 24.

Moschetti MP, Thompson EM, Powers PM, Hoover SM, McNamara DE (2019) Ground motions from induced earthquakes in Oklahoma and Kansas. *Seismological Research Letters.* 90(1). 160-170.

Moschetti MP, Aagaard BT, Ahdi SK, Altekruse JM, Boyd OS, Frankel AD, Herrick JA, Petersen MD, Powers PM, Rezaeian S, Shumway AM, Smith J, Stephenson WJ, Thompson EM, Withers KB (2024). The 2023 U.S. National Seismic Hazard Model: Ground-motion characterization for conterminous U.S. *Earthquake Spectra*. 40(2):1158-1190 DOI:10.1177/87552930231223995.

Mueller CS (2019) Earthquake catalogs for the USGS National Seismic Hazard Maps. *Seismological Research Letters*. 90(1): 251–261.

Nelson BK, DePaolo DJ (1985) Rapid production of continental crust 1.7 to 1.9 b.y. ago: Nd isotopic evidence from the basement of the North American mid-continent. *Geological Society of America Bulletin.* 96(6): 746–754 DOI: 10.1130/0016-7606(1985)96⟨746:RPOCCT⟩2.0.CO;2.

Nuclear Regulatory Commission (2012) Practical implementation guidelines for SSHAC Level 3 and 4 Hazard Studies. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research. *Report NUREG/CR-2117*.

Nweke CC, Stewart JP, Wang P, Brandenberg SJ (2022) Site response of sedimentary basins and other geomorphic provinces in southern California. *Earthquake Spectra*. 38 (4):2341-2370 DOI: 10.1177/87552930221088609.

Oklahoma Geological Survey (1978) Oklahoma Seismic Network [Data set]. International Federation of Digital Seismograph Networks. DOI: 10.7914/SN/OK

Pacific Earthquake Engineering Research Center (2015) NGA-East: Median ground-motion models for the Central and Eastern North America region. *PEER Report No. 2015/04*, Pacific Earthquake Engineering Research Center, UC Berkeley (headquarters).

Pakiser LC, Mooney WD (1989) Geophysical framework of the Continental United States, Geological Society of America Memoir 172, Boulder, CO, 826

Parker GA, Harmon JA, Stewart JP, Hashash YMA, Kottke AR, Rathje EM, Silva WJ, Campbell KW (2017) Proxy-based V_{S30} estimation in central and eastern North America. *Bulletin of the Seismological Society of America*. 107: 117–131.

Parker GA, Stewart JP, Hashash YMA, Rathje EM, Campbell KW and Silva WJ (2019) Empirical linear seismic site amplification in Central and Eastern North America. *Earthquake Spectra*. 35: 849–881.

Parker GA, Stewart JP (2022) Ergodic site response model for subduction zone regions. *Earthquake Spectra*. 38 (2): 841-864 DOI: 10.1177/87552930211056963.

Petersen MD, Frankel AD, Harmsen SC, Mueller CS, Haller KM, Wheeler RL, Wesson RL, Zeng Y, Boyd OS, Perkins DM, Luco N, Field EH, Wills CJ, Ruksatles KS (2008) Documentation for the 2008 update of the United States national seismic hazard maps. *Open-File Report 2008-1128*, U.S. Geological Survey. DOI: 10.3133/ofr20081128.

Petersen MD, Moschetti MP, Powers PM, Mueller CS, Haller, KM, Frankel AD, Zeng Y, Rezaeian S, Harmsen SC, Boyd OS, Field N, Chen R, Rukstales KS, Luco Ni, Wheeler RL, Williams RA, and Olsen AH (2014) Documentation for the 2014 update of the United States national seismic hazard maps. U.S. *Geological Survey Open-File Report 2014–1091*. DOI: 10.3133/ofr20141091.

Petersen MD, Shumway AM, Powers PM, Muller CS, Moschetti MP, Rezaeian S, McNamara D, Luco N, Boyd OS, Rukstales KS, Jaiswal KS, Thompson EM, Hoover SM, Clayton BS, Field EH Zeng Y (2020) The 2018 update of the US National Seismic Hazard Model: Overview of model and implications. *Earthquake Spectra* 36: 5–41.

Petersen MD, Shumway AM, Powers PM, Field EH, Moschetti MP, Jaiswal KS, Milner KR, Rezaeian S, Frankel AD, Llenos AL, Michael AJ, Altekruse JM, Ahdi SK, Withers KB, Mueller CS, Zeng Y, Chase RE, Salditch LM, Luco N, Rukstales KS, Herrick JA, Girot DL, Taagaard B, Bender AM, Blanpied ML, Briggs RW, Boyd OS, Clayton BS, DuRoss CB, Evans EL, Haeussler PJ, Hatem AE, Haynie KL, Hearn EH, Johnson KM, Kortum ZA, Kwong NS, Makdisi AJ, Mason HB, McNamara DE, McPhillips DF, Okubo PG, Page MT, Pollitz FF, Rubinstein JL, Shaw BE, Shen ZK, Shiro BR, Smith JA, Stephenson WJ, Thompson EM, Thompson Jobe JA, Wirth EA, Witter RC (2023). The 2023 US 50-state national seismic hazard model: Overview and implications. *Earthquake Spectra*. DOI:10.1177/87552930231215428

Pezeshk S, Zandieh A, Tavakoli B (2011) Hybrid empirical ground-motion prediction equations for eastern North America using NGA models and updated seismological parameters: *Bulletin of the Seismological Society of America*. 101: 1859–1870 DOI:10.1785/0120100144.

Pezeshk S, Zandieh A, Campbell KW, Tavakoli B (2018) Ground-Motion Prediction Equations for Central and Eastern North America Using the Hybrid Empirical Method and NGA-West2

Empirical Ground-Motion Models. *Bulletin of the Seismological Society of America*. 108 (4): 2278–2304 DOI:10.1785/0120170179

Pezeshk S, Zandieh A, Haji-Soltani A (2021) A ground-motion model for the Gulf Coast Region of the United States. *Bulletin of the Seismological Society of America.* 111: 3261–3277.

Pezeshk S, Assadollahi C, Zandieh A (2024) An equivalent point-source stochastic model of the NGA-East ground-motion models. *Earthquake Spectra*. 40 (2): 1452–1478. DOI: 10.1177/87552930231225983.

Pratt TL and Schleicher LS (2021) Characterizing ground-motion amplification by extensive flat-lying sediments: The seismic response of the eastern U.S. Atlantic Coastal Plain strata. *Bulletin of the Seismological Society of America.* 111: 1795–1823.

RCoreTeam (2019). R: A Language and Environment for Statistical Computing. vol. ISBN 3-900.

Ramos-Sepulveda ME, Parker GA, Thompson EM, Brandenberg SJ, Li M, Ilhan O, Hashash YMA, Rathje EM, Stewart JP (2023a) High-pass corner frequency selection for implementation in the USGS automated ground-motion processing tool, Geo-Congress 2023: Geotechnical Data Analysis and Computation, Los Angeles, CA, March 2023, *Geotechnical Special Publication No. 342*, EM Rathje, B Montoya, and MH Wayne (Eds.), 327-335, ASCE Geo-Institute

Ramos-Sepulveda ME, Buckreis TE, Li M, Parker GA, Brandenberg SJ, Rathje EM, Stewart. JP (2023b). 2023 NGA-compatible CENA database for GMM validation and site response studies, DesignSafe-CI. DOI: 10.17603/ds2-ywqs-sp29.

Ramos-Sepúlveda ME, Stewart JP, Parker GA, Moschetti MP, Thompson EM, Brandenberg SJ, Hashash YMA, Rathje EM (2024) Data-driven adjustments for combined use of NGA-East hard-rock ground motion and site amplification models. *Earthquake Spectra*. 40(2):1132-1157. DOI:10.1177/87552930241231825

Ramos-Sepulveda ME, Brandenberg SJ, Buckreis TE, Parker GA, Stewart JP. High-Pass Corner Frequency Selection and Review Tool for Use in Ground-Motion Processing. *Seismological Research Letters* (Under review).

Ramos-Sepúlveda ME, Parker GA, Thompson EM, Brandenberg SJ, Li M, Ilhan O, Hashash YMA, Rathje EM and Stewart JP (2023a) High-pass corner frequency selection for implementation in the USGS automated ground-motion processing tool, Geo-Congress 2023: Geotechnical Data Analysis and Computation, Los Angeles, CA, March 2023, *Geotechnical Special Publication No. 342*, EM Rathje, B Montoya, and MH Wayne (Eds.), 327-335, ASCE Geo-Institute.

Ratre P, Behm M (2021) Imaging the Deep Crustal Structure of Central Oklahoma Using Stacking and Inversion of Local Earthquake Waveforms. *Journal of Geophysical Research: Solid Earth*. 126(5) DOI:10.1029/2020JB021368.

Rekoske J, Thompson EM, Moschetti MP, Hearne M, Aagaard BT, Parker GA (2020) The 2019 Ridgecrest, California, earthquake sequence ground motions: Processed records and derived intensity metrics. *Seismological Research Letters*. 91(4): 2010-2023 DOI: 10.1785/0220190292

Rezaeian S, Powers PM, Shumway AM, Petersen MD, Luco N, Frankel AD, Moschetti MP, Thompson EM, McNamara1 DE (2021) The 2018 update of the US National Seismic Hazard Model: Ground motion models in the central and eastern US. *Earthquake Spectra*. 37: 1354-1390.

Ridders C (1979) A new algorithm for computing a single root of a real continuous function, *IEEE Transactions on Circuits and Systems*, 26(11):979-980.

Salvador A (1991) Structure at the base and subcrop below Mesozoic marine section, Gulf of Mexico basin, in The Gulf of Mexico Basin, A. Salvador (Editor), Vol. J, Plate 3. *The Geology of North America, Geological Society of America.* Boulder, Colorado.

Sammon JW (1969) A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*. C-18: 401–409.

Sawyer DS, Buffler RT, Pilger AJr. (1991) The crust under the Gulf of Mexico basin, in The Gulf of Mexico Basin, R. H. Salvador (Editor), Vol. J, The Geology of North America, *Geological Society of America*, Boulder, Colorado, 53–72.

Schleicher LS and Pratt TL (2021) Characterizing fundamental resonance peaks on flat-lying sediments using multiple spectral ratio methods: An example from the Atlantic Coastal Plain, Eastern United States. *Bulletin of the Seismological Society of America.* 111:1824–1848.

Seyhan E, Stewart JP (2014) Semi-empirical nonlinear Site amplification from NGA-West2 data and simulations. *Earthquake Spectra*. 30: 1241–1256.

Shahjouei A, Pezeshk S (2016) Alternative Hybrid Empirical Ground-Motion Model for Central and Eastern North America Using Hybrid Simulations and NGA-West2 Models. *Bulletin of the Seismological Society of America*. 106(2): 734–754 DOI: 10.1785/0120140367.

Silva WJ, Gregor N, Darragh RB, Abrahamson NA (2015) Development of hard rock ground-motion models for region 2 of central and eastern North America. *Report PEER 2015/04201504*. Pacific Earthquake Engineering Research Center, UC Berkeley (headquarters).

Silva W, Gregor N, Darragh R (2002) Development of regional hard rock attenuation relations for central and eastern North America. *Pacific Engineering and Analysis Technical Report*.

Somerville P, Collins N, Abrahamson N, Graves R, Saikia C (2001) Ground motion attenuation relations for the Central and Eastern United States—Final report, June 30, 2001: Technical report to U.S. Geological Survey, Reston, Virginia, under Contract 99HQGR0098.

Stewart JP, Parker GA, Atkinson GM, Boore DM, Hashash YMA and Silva WJ (2020) Ergodic site amplification model for Central and Eastern North America. *Earthquake Spectra*. 36: 42–68.

Tavakoli B, Pezeshk S (2005) Empirical-stochastic ground-motion prediction for eastern North America. *Bulletin of the Seismological Society of America*. 95: 2283–2296.

Thompson E, Hearne M, Aagaard BT, Rekoske JM, Worden CB, Moschetti MP, Hunsinger HE, Ferragut GC, Parker GA, Smith JA, Smith KK, Kottke AR (2024a) Automated, near real-time ground-motion processing at the U.S. Geological Survey. *Seismological Research Letters* (Under review).

Thompson EM, Hearne M, Aagaard BT, Rekoske JM, Worden CB, Moschetti MP, Hunsinger HE, Ferragut GC, Parker GA, Smith JA, Smith KK, Kottke AR (2024b) USGS Automated Ground Motion Processing Software Version 2, U.S. Geological Survey software release. DOI: 10.5066/P13HMKFJ

Toro GR (2002) Modification of the Toro et al. (1997) attenuation equations for large magnitudes and short distances. *Risk Engineering Technical Report*.

University of Utah (1962) University of Utah Regional Seismic Network [Data set]. International Federation of Digital Seismograph Networks. DOI: 10.7914/SN/UU

U.S. Geological Survey, Earthquake Hazards Program (2017) Advanced National Seismic System (ANSS) Comprehensive Catalog of Earthquake Events and Products: Various, DOI: 10.5066/F7MS3QZH.

Van Schmus WR, Bickford ME, Turek A (1996) Proterozoic geology of the east-central Midcontinent basement. In Basement and basins of eastern North America. *Geological Society of America*. (308: 7–32) DOI: 10.1130/0-8137-2308-6.7.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat I, Feng Y, Moore EW, VanderPlas J (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 17, 261–272 DOI: 10.1038/s41592-019-0686-2.

Whitmeyer SJ, Karlstrom KE (2007) Tectonic model for the Proterozoic growth of North America. *Geosphere*. 3, 220–259. DOI:10.1130/GES00055.1.

Yenier E, Atkinson GM (2014) Equivalent point-source modeling of moderate-to-large magnitude earthquakes and associated ground-motion saturation effects. *Bulletin of the Seismological Society of America*. *104*(3): 1458-1478.

Yenier E, Atkinson GM (2015) Regionally adjustable generic ground-motion prediction equation based on equivalent point source simulations: Application to central and eastern North America. *Bulletin of the Seismological Society of America.* 105: 1989–2009.

Yenier E, Atkinson GM, and Sumy DF (2017) Ground motions for induced earthquakes in Oklahoma. *Bulletin of the Seismological Society of America*. 107: 198–215. Youngs RR, Goulet CA, Bozorgnia Y, Kuehn N, Al Atik L, Graves RW, Atkinson GM (2021) NGA-East ground-motion characterization model part II: Implementation and hazard implications. *Earthquake Spectra.* 37:1283-1330.

Zalachoris G, Rathje E. (2019) Ground Motion Model for Small-to-Moderate Earthquakes in Texas, Oklahoma, and Kansas. *Earthquake Spectra* 35(1):1-20.

Zalachoris G, He J, Rathje EM (2020) Earthquake time series from events in Texas, Oklahoma, and Kansas. DesignSafe-CI. DOI: 10.17603/ds2-v310-qc53.

Zellman MS, Ostenaa DA, Brown M (2021) Subsurface characterization of the quaternary active Cheraw Fault in southeastern Colorado based on seismic imaging. *The Seismic Record*, 1(1): 56–65 DOI: 10.1785/0320210016.