UNIVERSITY OF CALIFORNIA

Los Angeles

Customization of Path and Site Response Components of Global Ground Motion Models for Application in Sacramento-San Joaquin Delta Region of California

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Civil Engineering

by

Tristan Edward Buckreis

2022

ABSTRACT OF THE DISSERTATION

Customization of Path and Site Response Components of Global Ground Motion Models for Application in Sacramento-San Joaquin Delta Region of California

by

Tristan Edward Buckreis

Doctor of Philosophy in Civil Engineering

University of California, Los Angeles, 2022

Professor Scott J. Brandenberg, Co-Chair

Professor Jonathan P. Stewart, Co-Chair

Ground motion models (GMMs) are used to assess seismic demands given parameters descriptive of source, path, and site conditions. The current models used in California were developed from a global database, and incorporate path and/or site adjustments representing the statewide average. My research is concerned with improving ground motion predictions in the Sacramento-San Joaquin Delta region of northern California. The Delta is home to vital infrastructure which includes over 1,770 km of levees that serve as a conduit for approximately two-thirds of the state's drinking-water supply. As such, levee damage from earthquakes is potentially catastrophic. Seismic hazards are controlled mostly by nearby sources, however distant larger sources are also impactful. Seismic waves from these sources cross many physiographic provinces to reach the

Delta, hence complex path effects deviating from the statewide average are expected. Furthermore, much of the Delta is underlain by soft peaty-organic soils, which are absent from databases used to develop global GMMs. These factors can be accounted for by customizing GMMs for the Delta subregion, particularly the path and site response components.

The approach followed in this research was to first develop a California ground motion database with particular emphasis on northern California and the Delta subregion. Path effects were modeled to allow for different anelastic attenuation in each of nine distinct physiographic domains, which has a stronger physical basis than prior cell-specific methods. The regionally customized path model was used with the ground motion data to derive observation-based site amplification at 36 sites in and around the Delta. Based on those results, I then developed two subregional site response models. The first model relates small-strain amplification with the time-averaged shear wave velocity in the upper 30 m (V_{S30}), and can be applied for all sites in the Delta and immediately surrounding areas. The second model incorporates additive terms to the V_{S30}-based model to account for site resonance effects and additional levels of amplification; the independent variables are derived from microtremor horizontal-to-vertical spectral ratios (mHVSR). Collectively, these regionally-calibrated models significantly reduce bias and variability for ground motion predictions in the Delta region.

The dissertation of Tristan Edward Buckreis is approved.

Yousef Bozorgnia

Henry J. Burton

Scott J. Brandenberg, Committee Co-Chair

Jonathan P. Stewart, Committee Co-Chair

University of California, Los Angeles

2022

To my parents, without whom, none of this would have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEGEMENTS

The research presented in this dissertation was sponsored by the California Department of Water Resources (DWR) under Agreement No. 4600012415. Any opinions, findings, conclusions, and recommendations expressed here are those of the author and do not necessarily reflect the views of DWR. The Department of Civil and Environmental Engineering at UCLA also generously offered me teaching opportunities which I truly appreciate. Without these sources of funding I would not have been able to complete my Ph.D. research.

Chapter 4 is a version of Buckreis, TE, P Wang, SJ Brandenberg, and JP Stewart. (202x). Subregional anelastic path models for California, *Bull. Seismol. Soc. Am.* (submitted).

I would first like to personally thank my advisors past and present for their guidance and support throughout my academic career. Thank you Professor Lisa M. Star (California State University, Long Beach) for introducing me to geotechnical engineering and for providing me with my first research opportunity. Your support and guidance set me up for success and oriented me on a path which ultimately led me to UCLA. Thank you Professors Scott J. Brandenberg (UCLA) and Jonathan P. Stewart (UCLA) for guiding me throughout my M.S. and Ph.D. studies. I have learned a great deal from the classes you taught and our regular meetings. I truly appreciate the time you have invested in me to help conduct research and this dissertation. I want to thank all three of you for the insightful advice you have given me, and for the dedication to your work and research you have shown. I would also like to thank the other members of my committee, Professors Yousef Bozorgnia (UCLA) and Henry V. Burton (UCLA) for your valuable suggestions.

Next I would like to thank those individuals which have contributed directly to the research I was lucky enough to be involved in:

Thank you to Mike Driller, Nicholas Novoa, Tim Wehling, Teresa Butler, Stephan Neumayr, Mark Pagenkopp, Brad Von Dessonneck, Tyler Mitchell, Steven Freisen, and countless others at the DWR for your collaboration and support throughout the *Delta Response Project*. Similarly, thank you Ariya Balakrishnan (DSOD), Albert Kottke (PG&E), Jamie Steidl (USGS and UCSB), and Ivan Wong (LCI) for your contributions and guidance while serving as an advisory panel on the *Delta Response Project*.

Thank you Alan Yong (USGS) for training me on how to perform microtremor H/V tests, which made much of the site-characterization work and subsequent analysis possible. Thank you Jamile Erdem (USGS) for letting me borrow the equipment on multiple occasions. Lastly, thank you Andrew Cowell (DWR) and Michael Winders (UCLA) for helping me in the field to collect H/V data during the intense August heat.

I would like to thank the *GMDB Project Team* which has contributed to the relational ground motion database. Thank you Pegnfei Wang (Old Dominion University, formally UCLA), Chukwuebuka C. Nweke (USC, formally UCLA), Scott J. Brandenberg, and Jonathan P. Stewart for helping to build the database up from scratch. Thank you Silvia Mazzoni (UCLA Garrick Institute for the Risk Sciences) for joining the team and for providing your many insights, your experience has proven to be invaluable. Thank you Maria Ramos-Sepulveda (UCLA) for your contributions towards incorporating NGA-East into the database. Lastly, thank you Jeffery Bayless (AECOM) for helping to identify shortcomings and for providing much needed user feedback.

I would also like to thank the many professionals and researchers who helped to contribute data for the various projects I was a part of. The list of people to thank is too grand, so as to not omit anyone I will give a collective “thank you”.

In a similar manner, I would like to thank the past and present collaborators of the *Geotechnical Research Group* at UCLA. Not only have you all provided a community which I could rely on for help with technical issues, but also a network of friendships. I would like to particularly thank Pengfei Wang for your collaboration on the *Delta Response* and *GMDB* projects, as well as your support with many technically related issues.

Finally, I would like to thank my family, friends, and pets for their support throughout this endeavor. Special thanks to my parents, Heather and Edward, without whom none of this would have been possible.

CURRICULUM VITA

Education

M.S., 2018, Civil (Geotechnical) Engineering, University of California, Los Angeles
B.S., 2017, Civil Engineering, California State University, Long Beach

Experience

Mentor, Southern California Earthquake Center (SCEC) Supported Opportunities for Undergraduate and Research to Collaborate on Earthquake Science (SOURCES) (2022)
Teaching Assistant, University of California, Los Angeles, Department of Civil and Environmental Engineering, Design of Foundations and Earth Structures (2022)
Mentor, Natural Hazards Engineering Research Infrastructure (NHERI) Research Experience for Undergraduates (REU) (2021)
Teaching Assistant, University of California, Los Angeles, Department of Civil and Environmental Engineering, Geotechnical Earthquake Engineering (2020)

Publications

Journal Articles

Buckreis, TE, P Wang, SJ Brandenberg, and JP Stewart. (202x) Subregional anelastic attenuation model for California, *Bul. Seimol. Soc. Am.*, (in review).

Wang, P, P Zimmaro, **TE Buckreis**, T Gospe, SJ Brandenberg, SK Ahdi, A Yong, JP Stewart. (2022). Relational database for Horizontal-to-Vertical Spectral Ratios, *Seism. Research Letters*, 93(2A), 1075-1088.

Brandenberg, SJ, JP Stewart, P Wang, CC Nweke, K Hudson, CA Goulet, X Meng, CA Davis, SK Ahdi, MB Hudson, A Donnellan, G Lyzenga, M Pierce, J Wang, MA Winters, MP Delisle, J Lucey, Y Kim, TW Gallien, A Lyda, JS Yeung, O Issa, **T Buckreis**, Z Yi. (2020). Ground Deformation Data from GEER Investigations of Ridgecrest Earthquake Sequence. *Seism. Research Letters*. 91 (4): 2024–2034.

Conference Papers

Buckreis, TE, P Wang, SJ Brandenberg, S Friesen, JP Stewart (2023). Site characterization data for site response modeling in Sacramento-San Joaquin Delta Region of California. *Proceeding of the Geo-Congress 2023 Conference, ASCE*. (Accepted).

Buckreis, TE, CC Nweke, P Wang, SJ Brandenberg, S Mazzoni, and JP Stewart. (2023). Relational database for California strong ground motions. *Proceeding of the Geo-Congress 2023 Conference, ASCE*. (Accepted).

Wang, P, **TE Buckreis**, SJ Brandenberg, JP Stewart. (2023). Modified hyperbolic model for dynamic properties of peaty organic soils. *Proceeding of the Geo-Congress 2023 Conference, ASCE*. (Accepted).

Buckreis, TE, P Wang, SJ Brandenberg, and JP Stewart. (2022). Seismic ground motion prediction for Delta Region of California including regional path effects and local site response, *12th National Conference on Earthquake Engineering*, EERI, Salt Lake City, UT, Jun 27 - Jul 1, 2022.

Gospe, T, P Zimmaro, P Wang, **TE Buckreis**, SK Ahdi, AK Yong, SJ Brandenberg and JP Stewart. (2020) "Supplementing shear wave velocity profile database with microtremor-based H/V spectral ratios." *17th World Conference on Earthquake Engineering*, (Paper No. C003010)

Buckreis, TE, SJ Brandenberg, and JP Stewart. (2019). Regional semi-empirical site response model for Sacramento-San Juaquin Delta levees. *ASDSO Dam Safety*, Orlando, FL, Sep 8-12, 2019.

Thesis

Buckreis, TE. (2017). Numerical evaluation of laterally loaded rock-socketed deep foundations. Honors thesis. California State University, Long Beach.

Curated Datasets

Buckreis, TE, JP Stewart, SJ Brandenberg, P Wang, and CC Nweke. (2022) “Data Files for California Subregional Path Effect Study.” DesignSafe-CI. 10.17603/ds2-mbee-8941.

Nweke, CC, JP Stewart, P Wang, SJ Brandenberg, and **T Buckreis.** (2022) "Data Files for Ground Motion Studies Pertaining to Southern California Basins and Other Geomorphic Provinces." 10.17603/ds2-93rk-hz83.

Buckreis, T, A Winders, P Wang, SJ Brandenberg and JP Stewart. (2021) "Microtremor Data Collected in Sacramento-San Joaquin Delta Region of California." 10.17603/ds2-dk6t-8610.

Rathje, E, A Mathews, SJ Branenberg, **T Buckreis**, and K Hudson. (2021). “Interpolating ground motion intensity at sites in the Next Generation Liquefaction database.” 10.17603/ds2-ab8v-z026.

Buenker, JM, SJ Brandenberg, MM Eslami, D Abundis, **T Buckreis**, and JP Stewart. (2020) "Centrifuge Test on Bentonite Clay - Test UCLA JZB01." 10.17603/ds2-e7s5-b025.

Lyda, A, J Yeung, J, **T Buckreis**, O Issa, SJ Brandenberg, and Z Yi (2019) "'UW RAPID UAV Imaging', in Ridgecrest, CA earthquake sequence, July 4 and 5, 2019." 10.17603/ds2-tyca-se83.

CHAPTER 1

Introduction

1.1. Motivation of Research

The Sacramento-San Joaquin Delta region of California – hereafter *Delta* – is the largest freshwater tidal estuary on the west coast of the United States, which provides hydrological conditions well suited for the deposition of peaty-organic soils (Drexler 2011). Peat is a unique material in the context of geotechnical engineering not only for its problematic characteristics (e.g., low shear strength, high compressibility, high water content, etc.) but also for its ability to vary widely spatially and temporally (Kazemian et al. 2011). Nonetheless, the Delta region is home to more than 1,700 km of levees that serve as a conduit for approximately two-thirds of California's drinking water and protect roughly 700,000 acres of land, much of it below sea level due to peat subsidence.

Seismic hazards in the Delta region are appreciable, but ground motion prediction is complicated by large epistemic uncertainty associated with site response through the deep soft soils in the region, including the surficial peats. Such site conditions are well outside of the range considered in databases used for ground motion model (GMM) development, including the site database developed in the NGA-West2 project (Seyhan et al. 2014). For example, the softest sites in that database have 30-m time-averaged shear wave velocities (V_{S30}) of about 150 m/s, whereas

many Delta sites have much softer conditions. As a result, NGA-West2 GMMs, and in particular their site terms, are not calibrated for the geologic conditions encountered in the Delta.

A major, but heretofore, untapped resource was available to address the site response problem in the Delta – a network of modern ground motion sensors across the Delta and elsewhere in California (e.g., Clayton et al. 2013; Kuyuk and Allen 2013; Krischer et al. 2015; Li 2021). Moreover, processing tools are coming online that streamline the task of signal processing (e.g., Hearne et al. 2019, Ramos-Sepulveda et al. 2023). As a result, substantial amounts of ground motion data are now available, often > 100 recordings per event in northern California, albeit generally for relatively low shaking amplitudes. To apply this information requires careful attention to quality control, storage, and management of data. It has been over a decade since the conclusion of the data synthesis component of the NGA-West2 project in 2011, and since that time no large-scale efforts were made to assemble and disseminate more recent high-quality ground motion data in California.

While some researchers have taken on the burden of processing and assembling more recent data, the majority have not. As a result, most of the source, path, and site models for California applications published in recent years (since NGA-West2) are either developed using the NGA-West2 dataset, which encompasses a comparatively small parametric range (i.e., fewer exceptionally soft/hard sites, fewer near-source observations, etc.), or are potentially developed using inconsistent databases (i.e., NGA-West2 in combination with more recent ground motion data with less carefully, fully automated processing, and inconsistent metadata assignment protocols). As described below, my thesis work and that of several collaborators at UCLA addresses this problem, and forms the starting point for the systematic data compilation in the pending NGA-West3 project.

1.2. Research Objectives and Scope

The overall objective of the research presented in this dissertation is to develop regionally-calibrated source, path, and site response models applicable for the unique conditions encountered in the Delta region. This involves, assignment of accurate site parameters from geotechnical site characterization data, seismic path effects, site response in very soft soils, and a multitude of related problems in geotechnical engineering and engineering seismology. The scope of my work may be summarized by six items, as follows:

1. Development of a publically available ground motion database with emphasis on events that produced recordings at ground motion recording sites at locations within the Delta and the surrounding areas (i.e., northern California). This work benefits the greater scientific community by adding to the global database of ground motions for active tectonic regions.
2. Collection and organization of geotechnical site characterization data, such as boring logs, cone penetration testing, seismic velocity measurements, and microtremor testing for horizontal-to-vertical spectral ratios (mHVSR). All collected data have been made available through the velocity profile database (VSPDB; Ahdi et al. 2018).
3. Based on the data in the VSPDB, development of regionally calibrated proxy-based models for assigning site parameters for soft soil sites in the Delta without a seismic velocity profile.
4. Analysis of ground motions to investigate regional path effects and source biases. Development of adjusted path terms in GMMs to account for regional variations.
5. Nonergodic analysis of ground motions to estimate site terms, from which site response for linear conditions at locations across the Delta and surrounding areas was derived.

Development of regionally-calibrated site response model(s) to capture the main attributes of the observed site responses conditional on various site parameters.

6. Assessment of all produced models with respect to performance and variability, and development of regionally-calibrated aleatory variability model(s). Models should produce unbiased predictions and reduce uncertainty when compared to ergodic models.

1.3. Organization

The organization of this dissertation is described as follows:

Chapter 2 describes the development of a database for ground motion studies. In this chapter, I present my data collection effort which was targeted at ground motions in northern California. Next, I discuss the assignment of source, site, and distance metadata and the integration of several data collection efforts into a centralized Californian dataset. Lastly, I present the publically available ground motion relational database (GMDB), tools to transfer data into the database, and an online interface to help users access the data. The work related to developing the GMDB is described in a paper to be presented at *Geo-Congress 2023*.

Chapter 3 focuses on site characterization data to facilitate the assignment of site parameters in the Delta. I first describe the collection and synthesis of geotechnical site characterization data from publically available sources and field explorations. Then I describe the analysis of these data to assign site parameters related to peat thickness, V_{S30}, and mHVSR with special attention given to peak features.

Chapter 4 focuses on subregional path effects in California. I will first introduce relevant literature regarding ergodic and nonergodic path models, as well as prior regionalization of

California. Next, I describe the subset of ground motion data from the GMDB that was used in my study. I outline the residuals analyses and novel modeling approach which was used to investigate spatially variable path effects. I also investigated unique source effects attributed to induced earthquakes originating in The Geysers region. Lastly, I developed a subregional anelastic path model and assessed its performance against an ergodic GMM. Much of the contents of Chapter 4 are contained in a paper submitted to the *Bulletin of the Seismological Society of America* (in review).

Chapter 5 presents the empirical study of site response for the Delta subregion which is the culmination of my work, and was made possible by efforts presented in Chapters 2 through 4. In this chapter, I discuss the subset of ground motion data recorded by seismic instruments in the Delta; outline residuals analysis to extract nonergodic site response at each site; and develop subregion-specific site amplification and aleatory variability models conditioned on the site parameters discussed in Chapter 2.

Finally, Chapter 6 summarizes the scope and major findings from this study, and provides recommendations for future work based on what I have learned throughout my research.

CHAPTER 2

Development of a Ground Motion Relational Database

2.1. Introduction and Project Motivation

Several Next-Generation Attenuation (NGA) research programs have developed collections of ground motion data and metadata. These projects have considered shallow crustal events in active tectonic regions (NGA-West2; Bozorgnia et al. 2014), crustal events in stable continental regions (NGA-East; Goulet et al. 2021a), and subduction earthquakes (NGA-Sub; Bozorgnia et al. 2022). Each of these programs has developed a uniformly processed dataset of recorded earthquake ground motions and associated source, path, and site metadata for use in their respective tectonic regimes. The metadata compiled in these projects, while broadly similar, have differences that reflect considerations that may be unique for a given tectonic regime; e.g. event-type assignments (intraslab, interface, etc.) for subduction zones. The ground motion data are provided as intensity measures (e.g., peak metrics and pseudo-spectral acceleration at a specified damping level: *PSA*) in all three datasets, and in most cases, ground motion time series are available for download from publicly accessible web sites (https://ngawest2.berkeley.edu/ and https://www.risksciences.ucla.edu/nhr3/gmdata). The NGA datasets are used by researchers to develop GMMs and by practicing engineers as part of site-specific hazard characterization studies.

The NGA-West2 dataset includes a large volume of data from California. The overall dataset contains 600 events, 4,151 stations, and 21,539 multi-component recordings (generally 3-component, but some recordings are missing an individual direction). The California component

of this database consists of 376 events, 1,647 stations, and 15,397 recordings. A majority of the California data are from small magnitude (**M** < 4.5) events in southern California or near the San Francisco Bay Area, with less populated areas having fewer – if any – recordings. Since completion of the NGA-West2 data collection efforts in 2011, the amount of seismic instrumentation across the state has increased (Kuyuk and Allen 2013), which means that the quantity and spatial distribution of data produced from a single event is much greater than a similar event included in NGA-West2. The subject of this chapter is a ground motion database (GMDB) that has been developed with emphasis on California data that is inclusive of NGA-West2 and more recent events. The GMDB has been assembled as a relational database that is publicly accessible through an application programming interface (API). The organization of the GMDB is similar to that used in NGA-Sub (Mazzoni et al. 2022).

2.2. Ground Motion Data

A number of recent studies have developed California ground motion data for events since 2011 in a consistent manner. These studies include ground motion analysis for the 2019 Ridgecrest earthquake sequence (Ahdi et al. 2019) and non-ergodic site response studies to assess site response predictability (Wang 2020). These efforts increased the quantity of available ground motions in southern California. A key element in my doctoral work is an empirically based site response model for the Sacramento-San Joaquin Delta (hereafter *Delta*) region (discussed in Chapter 5), therefore the need existed to assemble additional ground motion data in northern California. The following sections describe my data collection effort (which is published as a curated dataset on DesignSafe; Buckreis et al. 2022), the integration of available California ground motion data into a unified dataset, and development of the relational database itself.

2.2.1. Data Selection, Processing, and Distributions

Ground motion data within the Delta is necessary to facilitate development of the regional site response model discussed in Chapter 5. I identified events that occurred between 2011 and 2021 and that were recorded by one or more stations within the Delta. Additional screening was performed to consider only events with **M** ≥ 4.0 since spectral shape effects at low **M** can produce apparent site nonlinearities (Stafford et al. 2017), which can complicate the analysis of site terms. Moreover, I only considered events within 400 km of the Delta since weak motions beyond this distance may be screened out, thereby favoring unusually strong records (introducing bias). The search produced 71 events with **M** between 4.0 and 6.5 which were recorded by over 1,309 seismic stations. Figure 2.1(a) and (b) present maps showing the event and station locations, respectively.

Raw time-series records for all available stations for each event were obtained from the Incorporated Research Institutions for Seismology (IRIS) using the "obspy" package in Python (Krischer et al. 2015), and cross-checked against those available from the Center for Engineering Strong Motion Data (CESMD) data repository maintained by the California Strong Motion Instrumentation Program (CSMIP) (https://www.strongmotioncenter.org/). In the case of some Delta stations, records were obtained directly from the seismic network maintained by the California Department of Water Resources (DWR) with the assistance of Mike Driller and Nick Novoa. Additional screening was performed to remove apparent unreliable and duplicate records which may exist due to multiple co-located instruments at a site. In the event of co-located accelerometer and seismometer, both recording meaningful signals, I prefer the motion recorded by the seismometer unless there is evidence of amplitude clipping, in which the time-series recorded by the accelerometer is used.

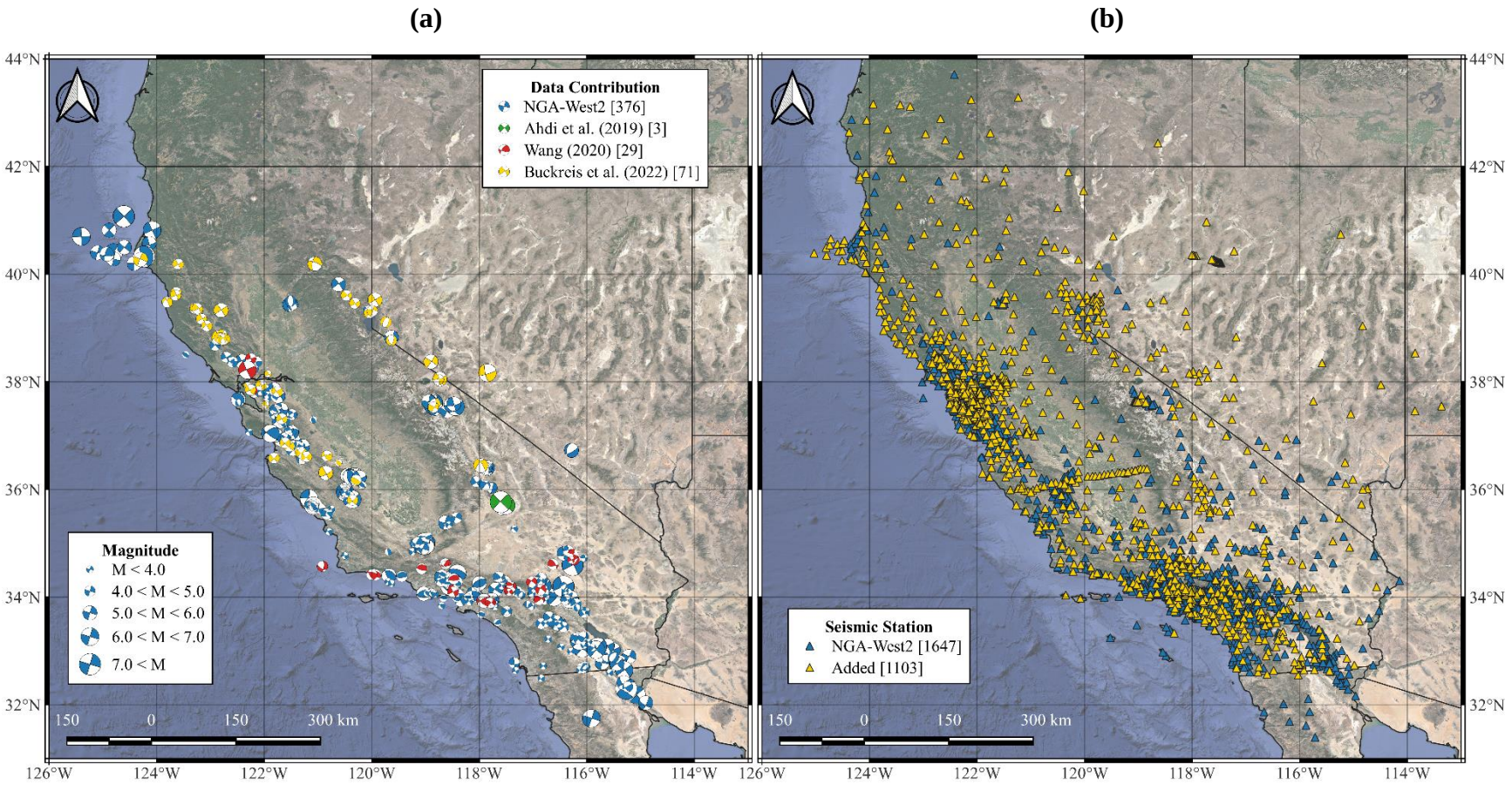


Figure 2.1: (a) Map of earthquake locations and mechanisms – symbol color represents corresponding data collection effort [blue – NGA-West2; green Ahdi et al. (2019); red – Wang (2020); and yellow – Buckreis et al. (2022)]; (b) Map of seismic station location – symbol color represents if the station was included in NGA-West2 database (blue) or new from one of the three recent collection efforts (yellow).

Each of the three-component records are processed individually according to standard protocols developed during the Pacific Earthquake Engineering Research center (PEER)-NGA projects (e.g., Goulet et al 2021b). Signal processing was performed in R (R Core Team 2022), where for each component I manually:

1. Accepted or rejected the raw time-series,
2. Selected P-wave and S-wave arrival times,
3. Selected high- and low-pass corner frequencies for filtering, and
4. Inspected the quality and acceptability of the processed displacement trace, response spectrum, and Fourier amplitude spectrum (*FAS*) results after filtering.

Task (1) requires that I examine the raw time-series to determine if the record contains a meaningful signal. The two extreme cases are: (1) a clean record which has a pronounced (earthquake) signal which will be accepted; and (2) a record which appears to be entirely noise (the amplitude of the signal does not exceed the noise threshold of the instrument) which will be rejected. Records which are not easily identified as clean or noise require windowing via selection of P-wave and S-wave arrival times to identify a "noise" and "signal" window, although all records which are initially accepted go through this process. The signal-to noise ratio (SNR) of the *FAS* is utilized to objectively assess whether a particular record contains a meaningful signal or not over some frequency bandwidth. Frequency bands with corresponding SNR ≥ 3 are deemed to contain meaningful signals. The high-pass corner frequency is selected as the lowest frequency with SNR>3 that adequately reduces low-frequency artifacts from the displacement trace. I did not consider shape of the low frequency portion of the *FAS* when assigning the high-pass corner frequency (e.g., if the shape of the *FAS* was relatively flat for a low-frequency interval above the high-pass frequency, in

apparent violation of seismological theory, but the SNR and displacement checks are satisfied, the high-pass corner frequency is nonetheless accepted). A low-pass filter was also sometimes applied to reduce the effects of noise on short-period spectral accelerations. Ground motion processing is highly sensitive to the selection of corner frequencies, therefore the last two tasks which require judgement are critical to ensure that the processed ground motions are not under- or over-processed.

The 71 events produced 9,875 usable three-component records. The distribution of these data with respect to **M**, Joyner-Boore distance [R_{JB}; discussed in Section 2.3.1(c)], and site condition [30 m time-averaged shear wave velocity - V_{S30}; discussed in Section 2.3.1(b)] is shown in Figure 2.2. The dataset, in combination with Ahdi et al. (2019) and Wang (2020), significantly expands the number of records for events with $4 < \mathbf{M} < 5$ and for soft soil sites ($V_{S30} < 300$ m/s). Figure 2.3 presents the number of usable horizontal ground motions per period defined as $T < 1/(1.25 f_{cHP})$, where f_{cHP} is the greater of the two horizontal high-pass corner frequencies selected during signal processing. The recently added data has a more rapid rate of decay of number of usable records with period for $T > 10$ sec than the NGA-West2 data. This likely results from the relatively small magnitudes of the recent data, which produce much less long-period energy. The reduced long-period energy increases the likelihood that records will be noise-dominated at these long periods.

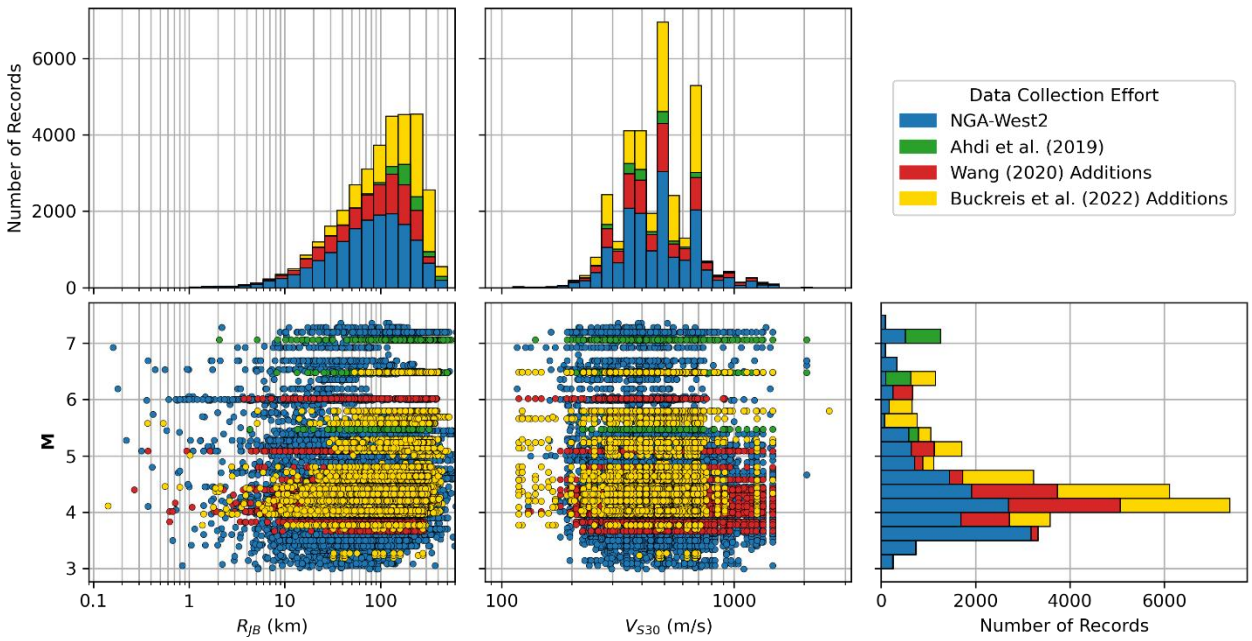


Figure 2.2: Data distributions for NGA-West2 (blue), Ahdi et al. (2019) (green), Wang (2020) additions (red), and Buckreis et al. (2022) additions (yellow) datasets.

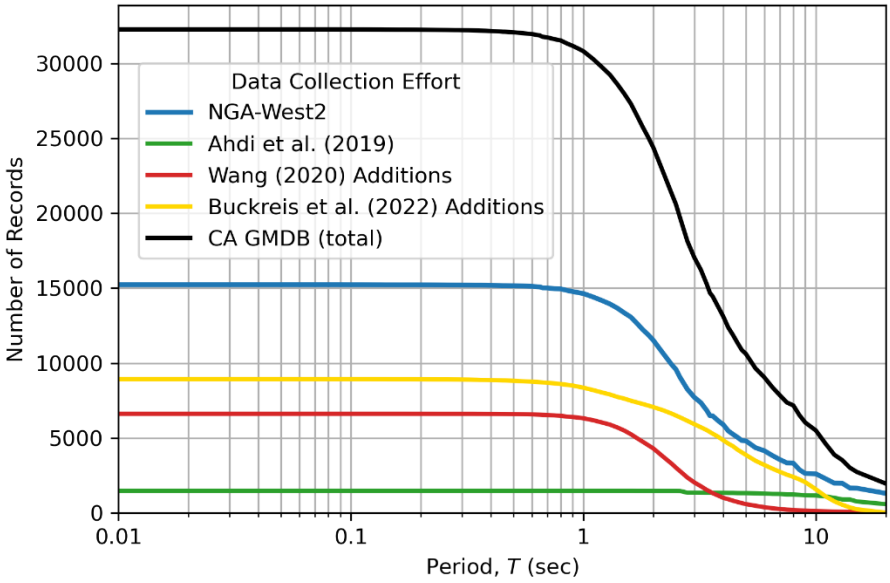


Figure 2.3: Number of usable ground motions per period.

2.2.2. Intensity Measure Calculations

Intensity measures (IMs) are parameters computed from ground motions and are used quantify certain attributes. NGA-West2 included peak IMs such as peak ground acceleration (*PGA*) and peak ground velocity (*PGV*), as well as 5% damped *PSA* for 111 oscillator periods between 0.01 s and 20 s. NGA-Sub and NGA-East expanded the selection of IMs to include Arias intensity (I_A; Arias 1970), times corresponding to select percentiles of I_A used to facilitate calculation of significant durations (I_A-times), and *FAS*. Other commonly used IMs include the cumulative absolute velocity (*CAV*; EPRI 1988) and standardized *CAV* (CAV_5).

All IMs excluding *FAS* are calculated for individual components (two orthogonal-horizontal and one vertical) as well as the combined horizontal minimum-, median-, and maximum-component (RotD00, RotD50, and RotD100, respectively) defined by Boore (2010). *PGA*, *PGV*, *PSA*, I_A, and I_A-times are obtained using the "RCTC" package in R (Wang et al. 2017). The orientation-independent horizontal component *FAS* (effective amplitude spectra; *EAS*) as defined by Kottke et al. (2021) is calculated using Python routines consistent with Kottke (2020).

2.3. Ground Motion Metadata

The ground motion data discussed in the previous section do not provide significant utility on their own. Parameters which describe the source, site, and path attributes associated with each individual ground motion are therefore required for model development and for use as additional screening criteria in forward applications. This section presents the metadata sources, methods, and routines used to assign ground motion metadata, which includes resolving differences between data from distinct datasets.

2.3.1. Assignment of Parameters

(a) Source Parameters

Earthquake-source parameters include origin date and time, hypocenter coordinates (longitude, latitude, and depth), seismic moment and moment magnitude (**M**), and nodal plane solution (strike, dip, and rake). These data are taken from the Global Centroid Moment Tensor (CMT) project (Ekström et al. 2012), the National Earthquake Information Center (NEIC) at the United States Geological Survey (USGS) or the Northern California Earthquake Data Center (NCEDC) housed at UC Berkeley. CMT origin times, **M**, and nodal plane solutions are preferred because they are derived using global recordings that average out radiation pattern variability and provide between-region consistency. CMT hypocenter coordinates are not preferred because the solution corresponds to the center of the earthquake moment distribution in time and space (Ekström et al. 2012), which may not align well with the location of the initial slip. Therefore, the preferred location as reported by the Advanced National Seismic System (ANSS) Comprehensive Earthquake Catalog (ComCat) is used (generally the NCEDC solution). The CMT did not contain data for 50 out of the 71 new events, in which case the preferred origin solution (i.e., time, **M**, and nodal plane solutions) as reported by the ANNS ComCat were adopted. Faulting mechanisms are assigned based on rake angle relationships given in Ancheta et al. (2013).

Parameters that describe the fault rupture surface as one or more rectangles (upper-left corner coordinates and dimensional length and width) are necessary to calculate source-to-site distances described in Section 2.3.1(c). All 71 events I collected data for do not have available finite-fault models in literature, therefore the simulation procedure described in Contreras et al. (2022) was performed for all 71 events using the CCLD5 program to obtain rupture surface parameters. When

simulating ruptures for events near the plate boundary region I used only the nodal plane orientation parallel to the San Andreas Fault, otherwise simulations were ran using both nodal plane solutions. Appropriate models for rupture area and aspect ratio given **M** are selected by CCLD5 by specifying the tectonic regime (i.e., shallow crustal).

(b) Site Parameters

Site parameters include location (latitude, longitude, elevation, and depth), topographic slope, terrain class, surficial geological unit, V_{S30}, and depth parameters $z_{1.0}$ and $z_{2.5}$ which represent depths to a shear wave velocity (V_S) of 1.0 and 2.5 km/s, respectively. Locations of each station were compared between those obtained from IRIS, NCEDC, CESMD, and the Southern California Earthquake Data Center (SCEDC) housed at Caltech, as available. For nearly all stations, locations agreed well and data from the catalog with the most significant digits was adopted. Stations with contradicting locations were handled on a case-by-case basis.

Coordinates were used in GIS software to assign terrain class (Iwahashi and Pike 2007) and calculate topographic slope using 3 arc-sec resolution digital elevation maps. Geological units for sites within California were assigned from Wills et al. (2015), while sites in Oregon, Nevada, and Utah were assigned units from local geologic maps (Ludington et al. 2005; Rowley et al. 2006; Biek et al. 2010; and Smith and Roe 2015).

V_{S30} values were assigned according to the following hierarchy:

1. Computed from a nearby measured V_S profile to a depth (z_p) of at least 30 m.

2. Estimated using the extrapolation relationship of Boore (2004) with California-specific regression coefficients from Kwak et al. (2017a) when a nearby shallow measured V_S profile exists (z_p < 30 m).
3. Estimated using regional V_{S30} proxy-based relationships (e.g., peat-thickness proxy model for sites located in the Delta, described in Section 3.3.2a).
4. Estimated using a weighted combination of kriging-interpolated (Thompson 2018), slope (Wald and Allen 2007), terrain (Yong et al. 2012), and/or geology based proxy models as described in Wang (2020).

Geology-based V_{S30} proxy models exist for California (Wills et al. 2015), Oregon (Ahdi et al. 2017), and Utah (McDonald and Ashland 2008), however no models are published for Nevada, where a significant number of new sites are located. Given the relative proximity of most Nevada sites to California, I felt it would be acceptable to relate geologic units given by Ludington et al. (2005) to counterparts in Wills et al. (2015) so that geology-based V_{S30} estimates could be considered. Table 2.1 summarizes these mappings, and Figures 2.4 through 2.6 present maps illustrating their spatial distributions. As observed in Figures 2.4 through 2.6, there is good agreement at the border between California and Nevada for respective units, which suggests that the mappings presented in Table 2.1are appropriate since geologic properties are not expected to change based on governmental borders.

Table 2.1: Summary of mapping Ludington et al. (2005) geologic units to Wills et al. (2015) geologic units.

Wills et al. (2015)			Ludington et al. (2005)	
Geologic Unit	**V_{S30} (m/s)**	**Description**	**Geologic Unit**	**Description**
Xtaline	710.1 ± 393.8	Crystalline rocks, including Cretaceous granitic rocks, Jurassic metamorphic rocks, schist, and Precambrian gneiss.	CZq	Quartzite and minor amounts of conglomerate, phyllitic siltstone, limestone, and dolomite. Includes Prospect Mountain Quartzite, Osgood Mountain Quartzite, and Gold Hill Formation in northern Nevada, and Stirling Quartzite, Wood Canyon Formation, and Zabriskie Quartzite in southern Nevada.
			CZs	Phyllitic siltstone, quartzite, and lesser amounts of limestone and dolomite. Includes Reed Dolomite; Deep Spring, Campito, Poleta, Harkless, and Saline Valley Formations; and Mule Spring Limestone
			Kgr	Granitic rocks. Mostly quartz monzonite and granodiorite
			KJim	Igneous and metamorphic complex. Pegmatitic granite and other granitic rocks complexly intermixed with metasedimentary rocks. Considered to be Mesozoic igneous complex intruding lower Paleozoic and possibly Precambrian Z sedimentary rocks. Grades into units shown on map as lower Paleozoic. Ruby Mountains and East Humboldt Range, Elko County
			MZgr	Granitic rocks, western Nevada (Mesozoic). Mostly quartz monzonite and granodiorite. Inconclusively dated or not dated radiometrically
			PMh	Havallah sequence of Silberling and Roberts (1962). Chert, argillite, shale, greenstone, and minor amounts of siltstone, sandstone, conglomerate, and limestone. Includes Schoonover Formation of Fagan (1962) and Reservation Hill Formation in Elko County, Farrel Canyon Formation in southwestern Humboldt County, Havallah and Pumpernickel Formations in Pershing, Lander, and parts of Humboldt Counties, and rocks originally considered a part of the Pablo and Excelsior Formations in northern Nye, northern Esmeralda, and southern Mineral Counties. Assignment of some rocks to the Havallah sequence in the East Range, Pershing County, is highly uncertain. Includes rocks ranging in age from Late Mississippian to Early Permian
			Xm	Metamorphic rocks. Gneiss and schist and lesser amounts of gneissic granite, pyroxenite, hornblendite, migmatite, pegmatite, and marble.
			Zw	Wyman formation. Phyllite and phyllitic siltstone and minor amounts of limestone, dolomite, and sandstone

Wills et al. (2015)			Ludington et al. (2005)	
Geologic Unit	**V_{S30} (m/s)**	**Description**	**Geologic Unit**	**Description**
Qal1	228.2 ± 48.0	Quaternary (Holocene) alluvium in areas of very low slopes (less than 0.5%)	Qa†	Alluvial deposits. Locally includes beach and sand dune deposits
Qal2	293.5 ± 73.5	Quaternary (Holocene) alluvium in areas of moderate slopes (0.5 - 2.0%)		
Qal3	351.9 ± 112.2	Quaternary (Holocene) alluvium in areas of steep slopes (greater than 2%)	Qp†	Playa, marsh, and alluvial-flat deposits, locally eroded.
Qoa	386.6 ± 145.1	Quaternary (Pleistocene) alluvium	QToa	Older alluvial deposits.
Tv	518.9 ± 172.0	Tertiary volcanic units including the Conejo Volcanics in the Santa Monica Mountains and the Leona Rhyolite in the East Bay Hills.	Ta2	Andesite and related rocks of intermediate composition. Flows and breccias
			Ta3	Andesite and related rocks of intermediate composition. Flows and breccias
			Tb	Basalt flows.
			Tba	Andesite and basalt flows. Mostly in about 17 to about 6 m.y. age range. In Humboldt County, locally includes rocks as old as 21 m.y. May include rocks younger than 6 m.y. in places
			Tgr	Granitic rocks. Mostly quartz monzonite and granodiorite
			Tt2	Welded and nonwelded silicic ash-flow tuffs. Locally includes thin units of air-fall tuff and sedimentary rock
			Tt3	Welded and nonwelded silicic ash-flow tuffs. Locally includes thin units of air-fall tuff and sedimentary rock

† Calculate topographic slope to correctly map to Wills et al. (2015) "Qal" unit.

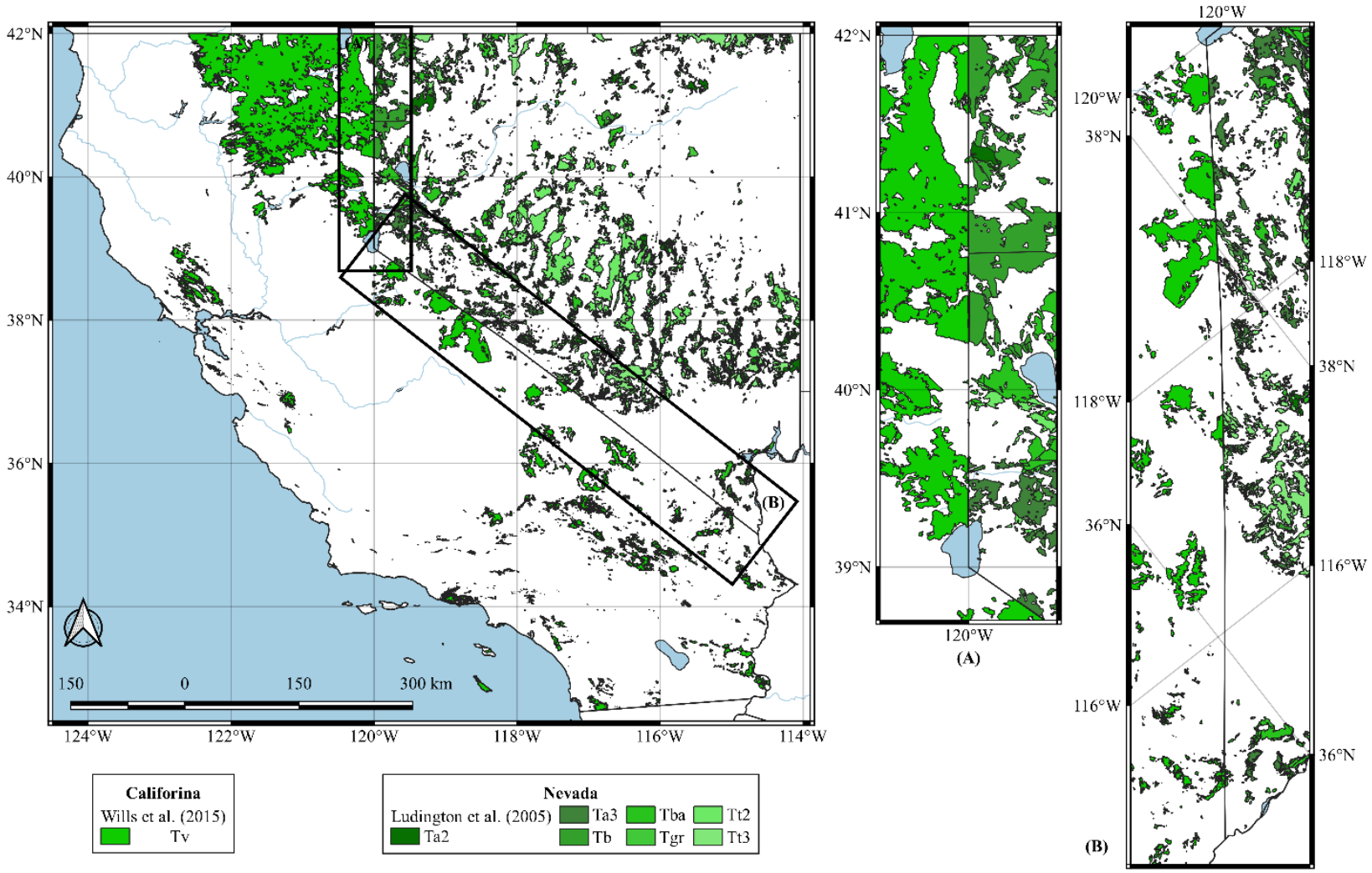


Figure 2.4: Spatial distribution of Nevada geologic units mapped as "Tv" in Wills et al. (2015). Inset maps illustrate good agreement of geologic unit across the state border.

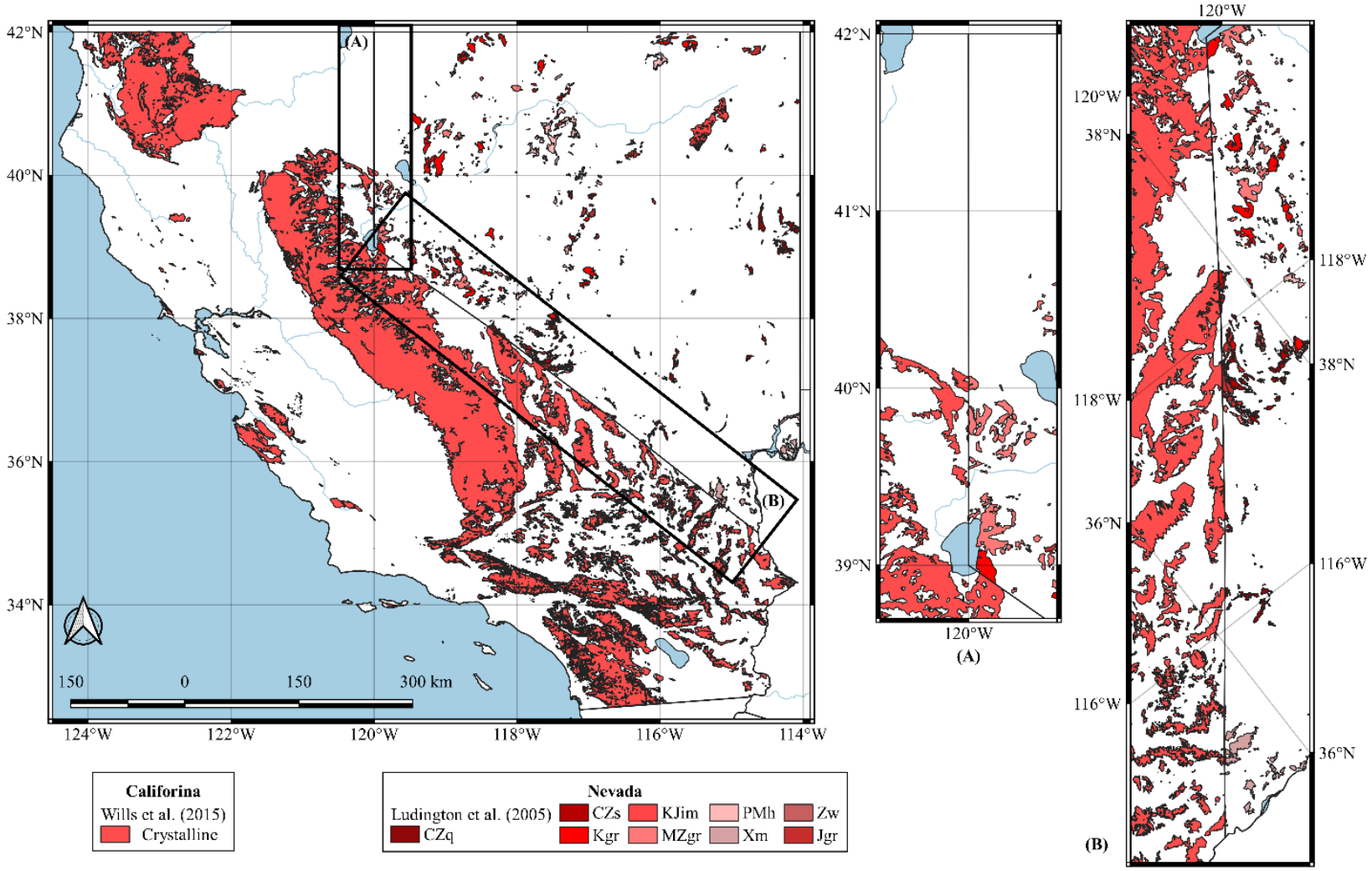


Figure 2.5: Spatial distribution of Nevada geologic units mapped as "Xtaline" in Wills et al. (2015). Inset maps illustrate good agreement of geologic unit across the state border.

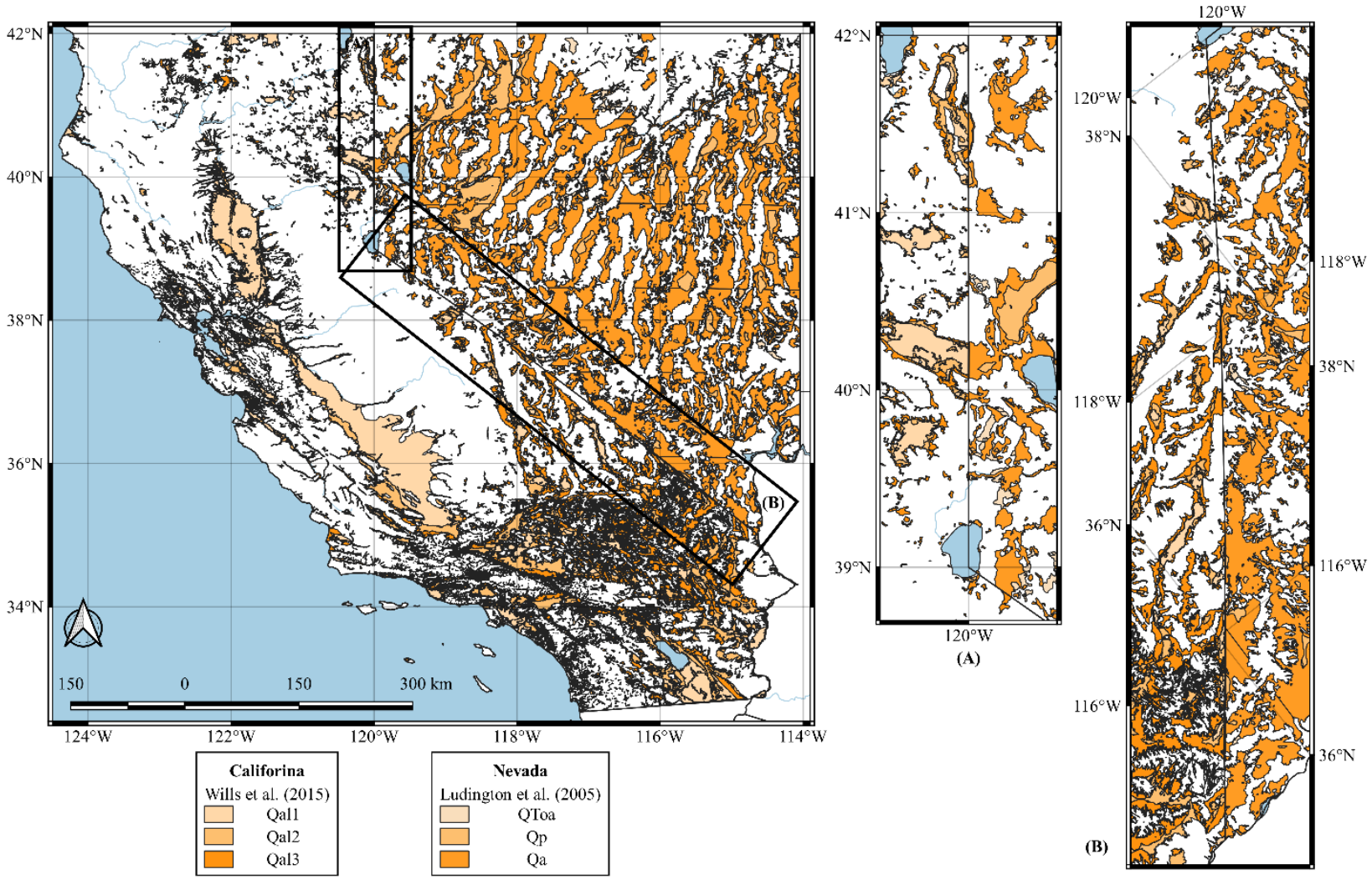


Figure 2.6: Spatial distribution of Nevada geologic units mapped as "Qal1", "Qal2", and "Qal3" in Wills et al. (2015). Inset maps illustrate good agreement of geologic unit across the state border.

Basin depth parameters ($z_{1.0}$ and $z_{2.5}$) were obtained from the most recently published California seismic velocity models (Lee et al. 2014; Shaw et al. 2015; Chen and Lee 2017; and Aagaard and Hirakawa 2021). Sites located within the modeled domain of more than one model are assigned multiple depth values for each depth term.

(c) Distance Parameters

Site-to-source distances are calculated using the P4CF routine which considers the rupture representations described in Section 2.3.1(a) and station coordinates. Details of the routine are described in Contreras et al. (2020). The recorded distance metrics include the closest distance from the site to any point on the rupture surface (R_{RUP}), closest distance from the site to any point on the surface projection of the rupture surface (R_{JB}), distance measured perpendicular to the fault strike from the surface projection of the top edge of the rupture surface (R_x), distance measured parallel to the fault strike from the midpoint of the surface projection of the rupture surface (R_y), the root-mean-square distance (R_{rms}), the epicentral distance (R_{epi}), and the hypocentral distance (R_{hyp}).

2.3.2. Integration of Ground Motion Data from Different Collection Efforts

The ground motion datasets of NGA-West2, Ahdi et al. (2019), Wang (2020), and Buckreis et al. (2022) contain similar fields, however significant effort was needed to integrate them into a unified dataset. The number of usable records available at a site is a principal consideration in determining whether or not a site-specific ground motion study is feasible. Therefore, mapping records to their appropriate seismic station across datasets is vital to integrating the separate collection efforts.

The recent datasets [Ahdi et al. (2019), Wang (2020), and Buckreis et al. (2022)] were coordinated in such a manner that stations could easily be matched by their unique network and

station code combinations, however no network information was recorded in NGA-West2 and many station codes were not populated. Furthermore, station coordinates may be rounded and/or their precision truncated. Manual screening of the more than 2,754 seismic stations is not practical, therefore a semi-automated procedure was developed to reduce the number of stations requiring manual inspection.

Station codes, names, and coordinates represent a combination of metadata that should be unique to a given station. Under this premise, we can first identify potential matches by computing distances between pairs of stations using coordinate information from two separate datasets. Pairs which are relatively close (within 1 km or 0.01^{o} in coordinate precision) may be a match and require further screening, however stations with no other nearby stations in the separate dataset are deemed to be distinct. During the second screening, pairs of station codes and names are compared using *fuzzy logic*. Fuzzy logic deals with approximate reasoning by assigning logic values ranging from zero to one, which represent *completely false* or *completely true*, respectively (Zadeh 1988). "Fuzzy string matching" is an application of fuzzy logic with the goal of identifying strings that approximately match a pattern. Using the "fuzzywuzzy" package in Python (Inc 2014), this approach assigns similarity scores to pairs of station codes and names for the potential matches. A normalized aggregate score is computed as the sum of individual scores divided by two, and represents the degree to which two stations can be considered a match.

Stations with normalized aggregate scores equal to one are interpreted to be perfect matches, however these are rare. Relatively large aggregate scores (> 0.85) generally correspond to matching stations, and are flagged as such after a quick review. Pairs with low aggregate scores (< 0.15) and large in-between distance (> 0.5 km or 0.005^{o} in coordinate precision) are separated, and any stations

which do not have another potential match with corresponding aggregate score greater than 0.15 are considered distinct. The number of remaining pairs of potential matches is manageable, and as such are sorted by aggregate scores in descending order and screened manually. Using this approach, 2,754 seismic stations were deemed to be unique after unifying the four datasets, the locations of which are presented in Figure 2.1(b).

In addition to unifying the station metadata, I update site parameters for all stations in a consistent manner as that described in Section 2.3.1(b). Given the recent efforts to make measured V_S profiles available (e.g., Ahdi et al. 2018), V_{S30} values for many NGA-West2 stations have been updated. NGA-West2 formally had 391 sites in CA with V_{S30} computed from a measured V_S profile, however the updated CA GMDB contains 651 sites with computed V_{S30} values (i.e., "measured"). Distributions of measured and inferred V_{S30} values are shown in Figure 2.7 for NGA-West2 and the unified California dataset (referred to as "CA GMDB"). The extremes of the V_{S30} distributions (<200 m/s and >700 m/s) are dominated by sites with measured data.

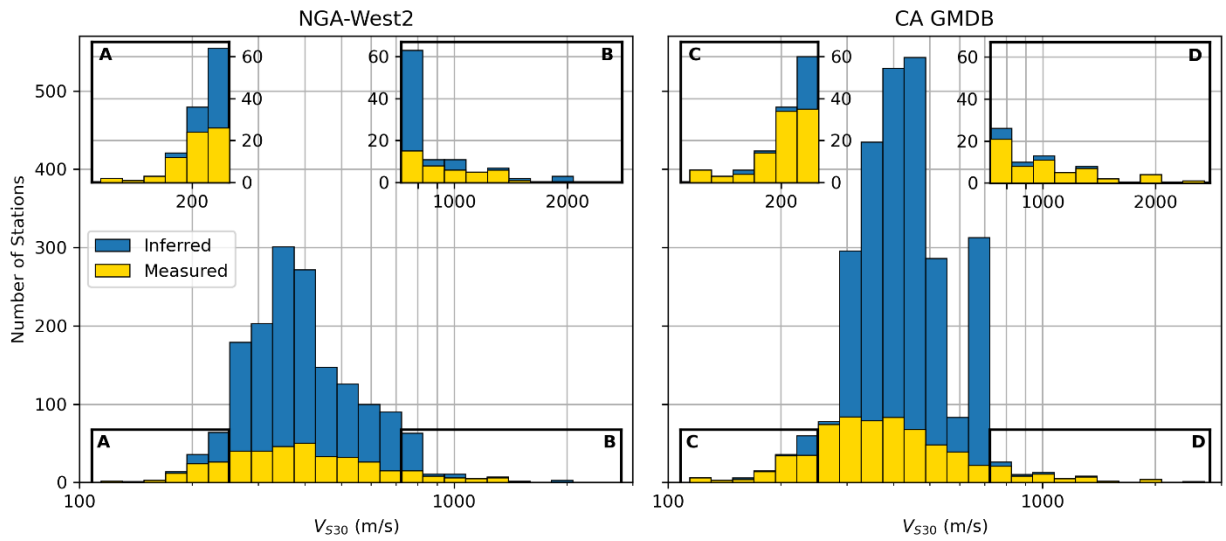


Figure 2.7: Distribution of measured (yellow) and inferred (blue) V_{S30} values for NGA-West2 (left) and unified dataset (right). Insets show expanded views of distribution tails.

In addition to screening for duplicate stations, the aforementioned approach was adapted to help assign networks (and their associated network codes) to the majority of NGA-West2 California stations. The motivation for this effort arose from the fact that it is easier to screen for duplicate stations based on their unique network and station code combinations, and that this process will need to be undertaken each time new data is added. I set out to assign networks to the remaining NGA-West2 California stations by comparing NGA-West2 station metadata to all available California stations downloaded from IRIS, NCEDC, SCEDC, and CESMD. I was able to identify the corresponding networks from the available metadata for all but 94 of the 1,647 stations, of which I suspect most to be decommissioned.

Unifying the event-related fields is trivial in comparison to that required for the stations. The number of events is of manageable size, and the unique origin information (date, time, and location) is sufficient enough to manually screen for event matches. Unsurprisingly, no event conflicts were

identified between the four datasets, and the locations of 479 unique events are presented in Figure 2.1(a). Event metadata were adopted as those recorded in each dataset.

Now that the datasets have been methodically unified, I assign new unique identifiers to each earthquake, station, and ground motion record which I refer to as *event_id*'s, *station_id*'s, and *motion_id*'s, respectively. The unified California dataset is made up of 33,370 ground motions, which is more than double the size of the California NGA-West2 subset.

2.4. Relational Database

NGA-West2 and NGA-East were managed as large spreadsheet files, while NGA-Sub was organized as a relational database hosted and managed on a local PC (e.g., Mazzoni et al. 2022). Products from all three NGA databases were published as a series of spreadsheet files referred to as flatfiles. The objective of the work presented here is to develop a web-served publicly accessible relational database that users can query. Motivation for a web-served relational database is driven by four factors:

1. The spreadsheet flatfiles have become rather large and unwieldy as the amount of data has grown;
2. Users often desire specific fields of data that can be accessed by targeted queries rather than having to download undesired data with the desired data;
3. Database queries are generally significantly faster than file input/output operations since relational databases are stored in a combination of random access memory and disk space; and

4. The relational structure reduces the potential for inconsistent metadata (e.g., different site parameters for different recordings made at the same site).

The relational database is created using the Structured Query Language (SQL) database format utilizing the MySQL InnoDB storage engine, which balances reliability and performance. This management system allows for fields to have predefined data types which include numeric (e.g., “INT”, “FLOAT”, etc.), temporal (e.g., “DATE” and “DATETIME”), or string types (e.g., “VARCHAR”, “MEDIUMTEXT”, etc.), which helps to ensure data integrity by only allowing entries conforming to valid data types. Furthermore, tables are arranged on the server to optimize queries based on table relationships defined by shared fields called “keys”. A *primary key* denotes a unique identifier for each table entry, and a *foreign key* is a field in a separate table that identifies the corresponding entry from the table where it is the primary key. Since it is most efficient to query integer values in SQL, keys are defined as unique integer numbers. This section describes the database structure, reasoning for organization methods, the methods used to prepare data for integration into the database, and other elements developed to help users access the data.

2.4.1. Database Structure

Querying the database requires knowledge of the schema, which describes the tables, fields, and relationships among the many tables in the database. One advantage of a relational schema is that it can easily accommodate changes and additions over time as warranted. This is beneficial because future research will likely include additional information which we do not currently store, but simple alterations could be made to accommodate the data. The current GMDB schema is a result of extensive discussion among the project team members, and is comprised of 25 distinct tables. A list of table names is provided in Table 2.2, which are grouped into five

categories: event tables, station-site tables, ground motion metadata, ground motion data, and auxiliary information. A simplified diagram describing the full schema is presented in Figure 2.8. The following subsections discuss each category in detail.

Table 2.2: List of table names in GMDB schema.

Group Type	Table Name	Number of Fields
Event Tables	event_type	2
	event	28
	finite_fault	8
	finite_fault_seg	11
	finite_fault_kinematic_parameter	10
	aftershock_mainshock	4
	event_eqid	4
Station-Site Tables	network	8
	station	10
	site	20
	vs30_code	3
	basin_model	4
	basin_site	5
	station_ssn	4
Ground Motion Metadata	motion	30
	collection_motion	4
Ground Motion Data	component	3
	response_spectra	5
	fourier_spectra	7
	intensity_measure	9
	time_series_metadata	16
	time_series_data	3
Auxiliary Information	collection	4
	geometry	7
	user	6

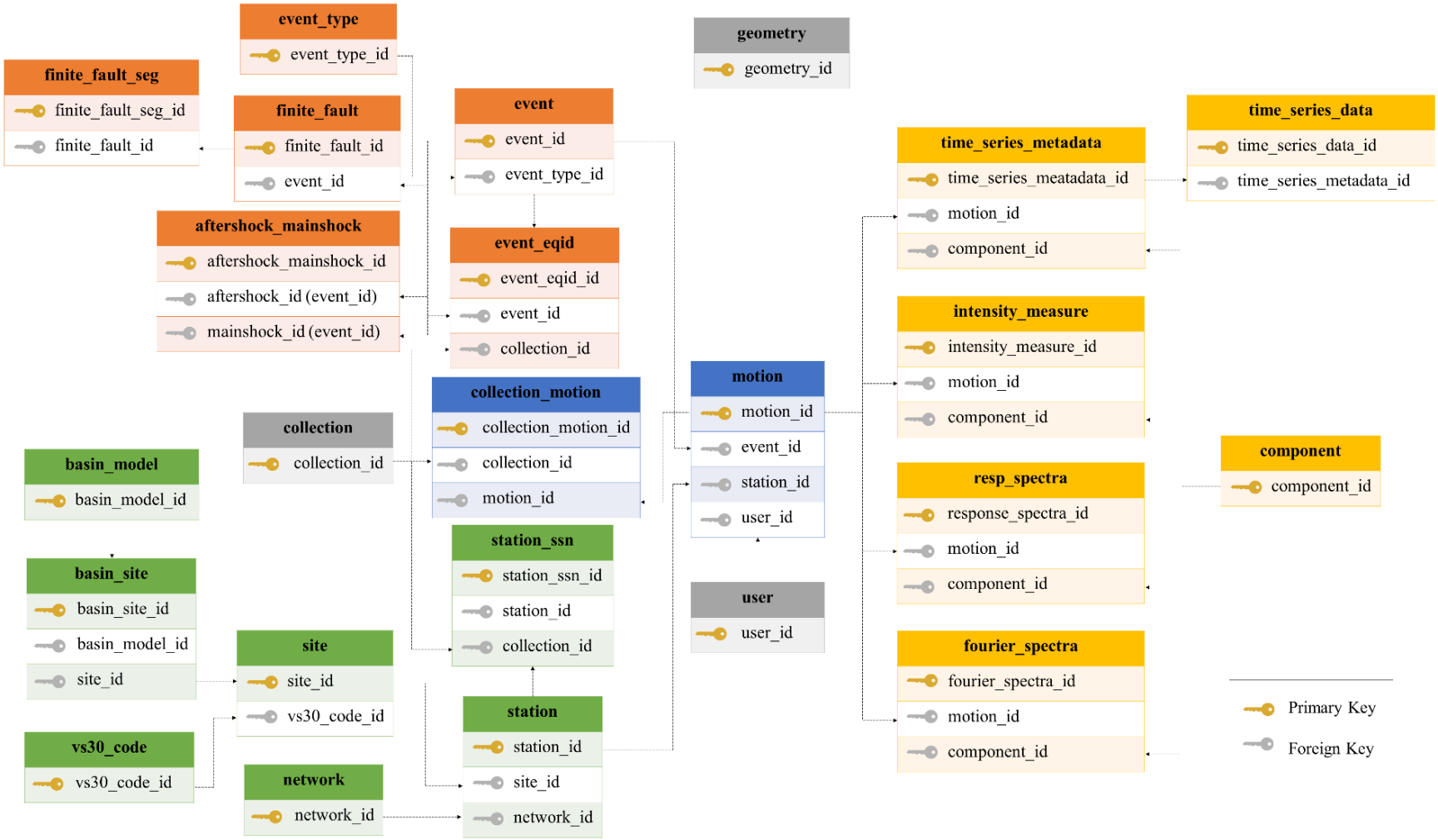


Figure 2.8: Diagram of simplified schema including primary and foreign key relationships; event tables colored red; station-site tables colored green; ground motion metadata tables colored blue; ground motion data tables colored yellow; and auxiliary information tables colored gray.

(a) Event Tables

Event metadata are any information related to characterizing an earthquake, such that that discussed in Section 2.3.1(a), and are organized into seven tables: *event*, *finite_fault*, *finite_fault_segment*, *finite_fault_kinematic_parameter*, *aftershock_mainshock*, *event_type*, and *event_eqid*. Figure 2.9 presents the detailed fields, their data types, and key relationships between event tables in the GMDB.

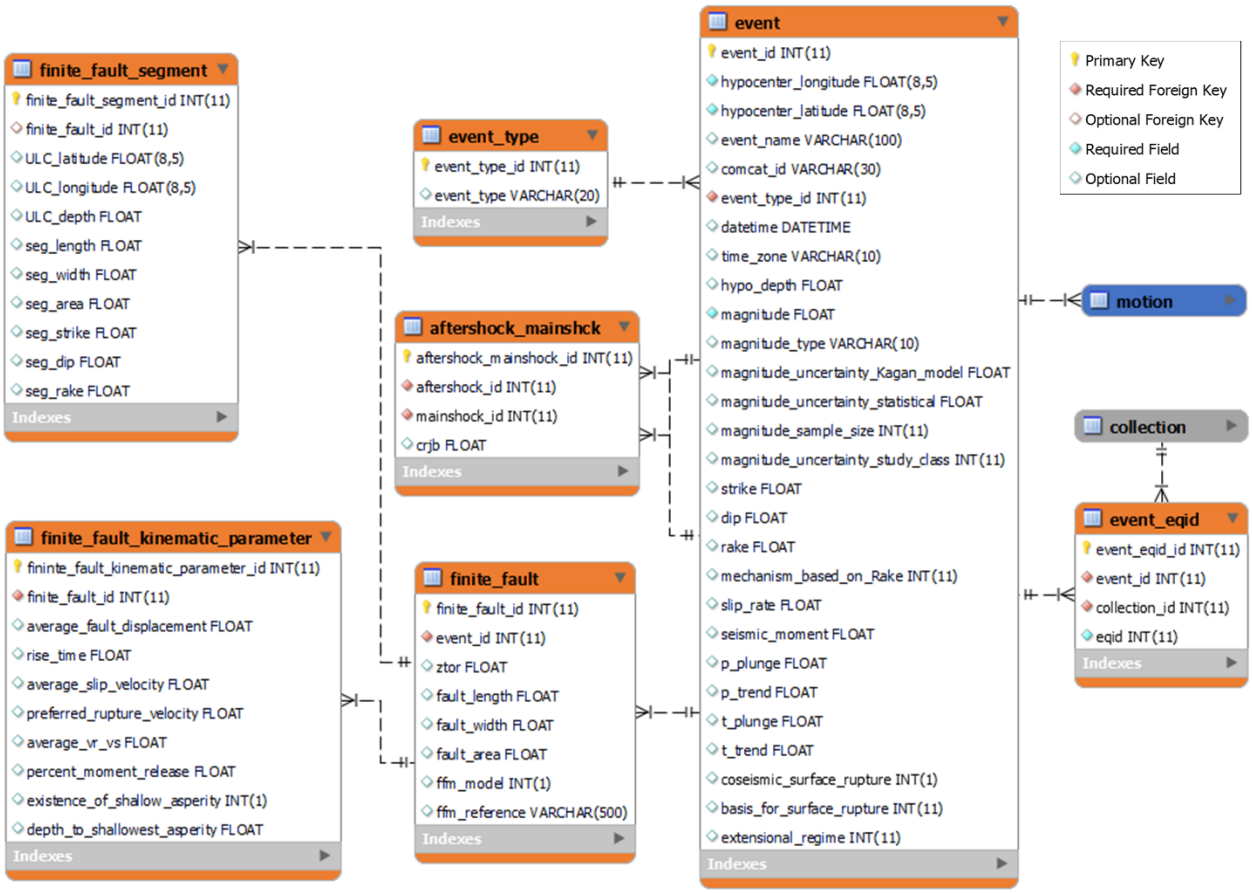


Figure 2.9: Detailed fields and datatypes for event tables.

The *event* table stores the most common earthquake related metadata, with *event_id* serving as the primary key. In addition to source parameters, the *event* table also records the unique ID assigned by the ANSS ComCat in the *comcat_id* field. All numeric fields are stored as "FLOAT" data types, and descriptive fields including *event_name, magnitude_type*, and *comcat_id* are stored as "VARCHAR" string objects (this extends to all other tables). Longitude and latitude fields are stored as "FLOAT(8,5)" datatypes, which preserves up to five significant digits after the decimal point (approximately 1.11 m resolution, although event locations are typically reported with three significant digits after the decimal point). Origin times are stored as "DATETIME" objects in the *datetime* field which allows the use of time operations to easily and efficiently construct queries with temporal constraints. Faulting mechanisms are assigned based on rake angle relationships defined in Ancheta et al. (2013) in the *mechanism_based_on_rake* field, and the presence of surface rupture is stored as a binary indicator (0 = "no" or 1 = "yes") in the *coseismic_surface_rupture* field. The *mechanism_based_on_rake*, *magnitude_uncertainty_study_class*, *basis_for_surface_rupture*, and *extensional_regime* fields are stored as integer-valued codes to facilitate efficient querying, the definitions of which are provided in Tables 2.3 – 2.6.

Table 2.3: Integer definitions for *mechanism_based_on_rake*, adopted from Ancheta et al. (2013).

Code	Mechanism	Criteria Based on Rake (λ)
0	Strike-Slip	$-180^o < \lambda < -150^o$
		$-30^o < \lambda < 30^o$
		$150^o < \lambda < 180^o$
1	Normal	$-120^o < \lambda < -60^o$
2	Reverse	$60^o < \lambda < 120^o$
3	Reverse – Oblique	$30^o < \lambda < 60^o$
		$120^o < \lambda < 150^o$
4	Normal – Oblique	$-150^o < \lambda < -120^o$
		$-60^o < \lambda < -30^o$

Table 2.4: Integer definitions for *magnitude_uncertainty_study_class*, adopted from Ancheta et al. (2013).

Code	Magnitude Uncertainty Class Description
1	Recent events well studied
2	Older events well studied, or recent events not well studied
3	Older events not well studied

Table 2.5: Integer definitions for *basis_for_surface_rupture*, adopted from Ancheta et al. (2013).

Code	Ancheta et al. (2013)	Magnitude Uncertainty Class Description
1	sfdoc	Reference documents the presence or absence or surface faulting
2	sfdis	Surface faulting discussed in references
3	locdis	Location of earthquake discussed in references
4	M<6	**M** < 6, likelihood of existence of surface rupture is small
5	M~6	**M** ≈ 6, likelihood of existence of surface rupture is small
6	M>7	**M** ≥ 7, likelihood of existence of surface rupture is large

Table 2.6: Integer definitions for *extensional_regime*, adopted from Ancheta et al. (2013).

Code	Extensional Regime Description
0	No
1	Yes
2	Convergent
3	Transform

Rupture surface representations and associated data for each event are stored in the *finite_fault*, *finite_fault_segment*, and *finite_fault_kinematic_parameter* tables. The *finite_fault* table stores general rupture surface information including the total length (*fault_length*), width (*fault_width*), area (*fault_area*), and depth to top of the rupture model (*ztor*), as well as an integer field (*ffm_model*) which takes the value of "1" if the representation is derived from a published

finite-fault model, or "0" if the rupture surface is simulated using the procedure described in Contreras et al. (2022). The *finite_fault* table is connected to the *event* table via the *event_id* foreign key. Although it is possible for several finite-fault model representations to exist for a single event, the GMDB only records information derived from the preferred model as selected using the approach described in Contreras et al. (2022). The location and geometric properties of each rupture surface are stored in the *finite_fault_segment* table, and are linked to the *finite_fault* table via the *finite_fault_id* foreign key. Fault ruptures may be represented by multiple segments, therefore the *finite_fault_id* foreign key represents a "one-to-many" relationship, meaning that one *finite_fault* entry can be related to multiple *finite_fault_segment* entries. Kinematic parameters including average fault displacement, average slip velocity, percent moment release in the top 5 km of crust, and more are stored in the *finite_fault_kinematic_parameter* table for events with published finite-fault models.

The *aftershock_mainshock* table stores data which can be used to identify aftershocks that re-rupture the mainshock fault plane or occur within the damaged zone within a time window for aftershocks (i.e., "Class 2" events). Class 2 events have been postulated to produce systematically lower event terms when compared to "Class 1" events (mainshocks, foreshocks, triggered events, and off-plane aftershocks) (Ancheta et al. 2013; Boore et al. 2014). The table functions as a junction table between two events in the *event* table, where *aftershock_id* corresponds to the *event_id* of the aftershock and *mainshock_id* is the *event_id* of the corresponding mainshock. The shortest distance between the surface projection of the centroid of the rupture surface of the trial event and the closest point on the edge of the surface projection of the mainshock rupture surface (CR_{JB}) is stored in the *crjb* field. Only pairs of events which occur within the Gardner and Knopoff (1974) time window are considered, all others are assumed to be Class 1. Users distinguish Class

1 and Class 2 events in the *aftershock_mainshock* table by comparing CR_{JB} to a specified threshold distance of their choosing (e.g., CR_{JB} < 15 km suggests Class 2; Ancheta et al. 2013).

The broader motivation behind the GMDB goes beyond developing a unified relational database for California ground motions – indeed, the aim is to include ground motions from all tectonic regimes within a uniform framework. This is different from past NGA projects, which have separated ground motion data based on tectonic regime, which results in three separate and distinct databases:

1. NGA-West 2 for shallow crustal earthquakes (Bozorgnia et al. 2014).
2. NGA-Sub for subduction-type earthquakes (Bozorgnia et al. 2022).
3. NGA-East for stable continental earthquakes (Goulet et al. 2021a).

As a result of these databases having been developed separately, there is no centralized resource that users can query for uniformly processed ground motion data with consistently developed metadata fields. The reasoning which led to this decision was likely based on the truth that different phenomena are responsible for causing each of the three types of events, therefore they should be modeled separately. However, there are regions of the world that experience earthquakes of more than one type (e.g., Alaska, Japan, New Zealand, and California), and querying from multiple non-unified databases may result in data inconsistencies during model development or forward applications.

Modeling the effects of different earthquake types requires that the data be partitioned based on type, however there is no requirement that the data be organized and stored separately. We introduce an *event_type* table linked to the *event* table through the *event_type_id* foreign key as a means to identify the type of earthquake, which represents a “one-to-one” relationship

meaning that a given event can only have one event type. This way, all ground motion data from around the world can be integrated into the GMDB, and users can query from a centralized databased to obtain consistent metadata. This is a long term goal, however as of writing the GMDB only contains ground motion data from around CA and central and eastern North America (CENA). Table 2.7 presents the current *event_type* table.

Table 2.7: Current GMDB *event_type* table.

event_type_id	***event_type***
1	Shallow Crustal
2	Interface
3	Intraslab
4	Outer-rise
5	Induced
6	Undetermined
7	Stable Continental

The "earthquake ID" (*EQID*) assigned to events in NGA products are not adopted as *event_id*'s in the GMDB because NGA-West2 and NGA-East each started their assignments at "1". We assign unique *event_id*'s in the *event* table to ensure that the primary keys are unique, however users may wish to search the GMDB using familiar *EQID*'s assigned by individual collection efforts unified in the GMDB (e.g., NGA-West2, NGA-East, etc.). The *event_eqid* table provides this mapping through four fields: (1) *event_eqid_id* which acts as the primary key, (2) *event_id* foreign key linked to the *event* table, (3) *collection_id* foreign key linked to the *collection* table [discussed in Section 2.4.1(e)], and (4) *eqid* which records the *EQID* assigned by the individual collection (e.g., NGA-West2). The relationships of *event_eqid_id* to *collection_id* and

event_id are both “one-to-many” relationships meaning that a given *event_id* can only have one *EQID* for a given *collection_id,* however individual *event_id*’s may appear numerous times in *event_eqid* if they exists in multiple collections.

(b) Station-Site Tables

Station-site metadata are any information related to the seismic station, network or site characterization, and are organized into seven tables: *network*, *station*, *site*, *vs30_code*, *basin_model*, *basin_site*, and *station_ssn*. Figure 2.10 presents the detailed fields, their data types, and key relationships between station-site tables in the GMDB. Within the context of the database, “station” refers to a particular instrument which provides the ground motion record and “site” refers to the physical location of the recording.

The *network* and *station* tables together contain information about the recording instrument. The *network* table stores metadata about seismic networks including their name, network code, operator (*operation_org*), and operating dates (*start_date* and *end_date*) assigned by the International Federation of Digital Seismograph Networks (FDSN). The *network_id* is used as a foreign key in the *station* table, which contains information about individual stations including the name, station code, housing, depth, and location (*station_latitude*, *station_longitude*, and *station_elevation*). The *housing* field describes the instrument structure type, for which we adopt the GMX first letter codes summarized in Table 2.8. The primary key in the *station* table is *station_id*.

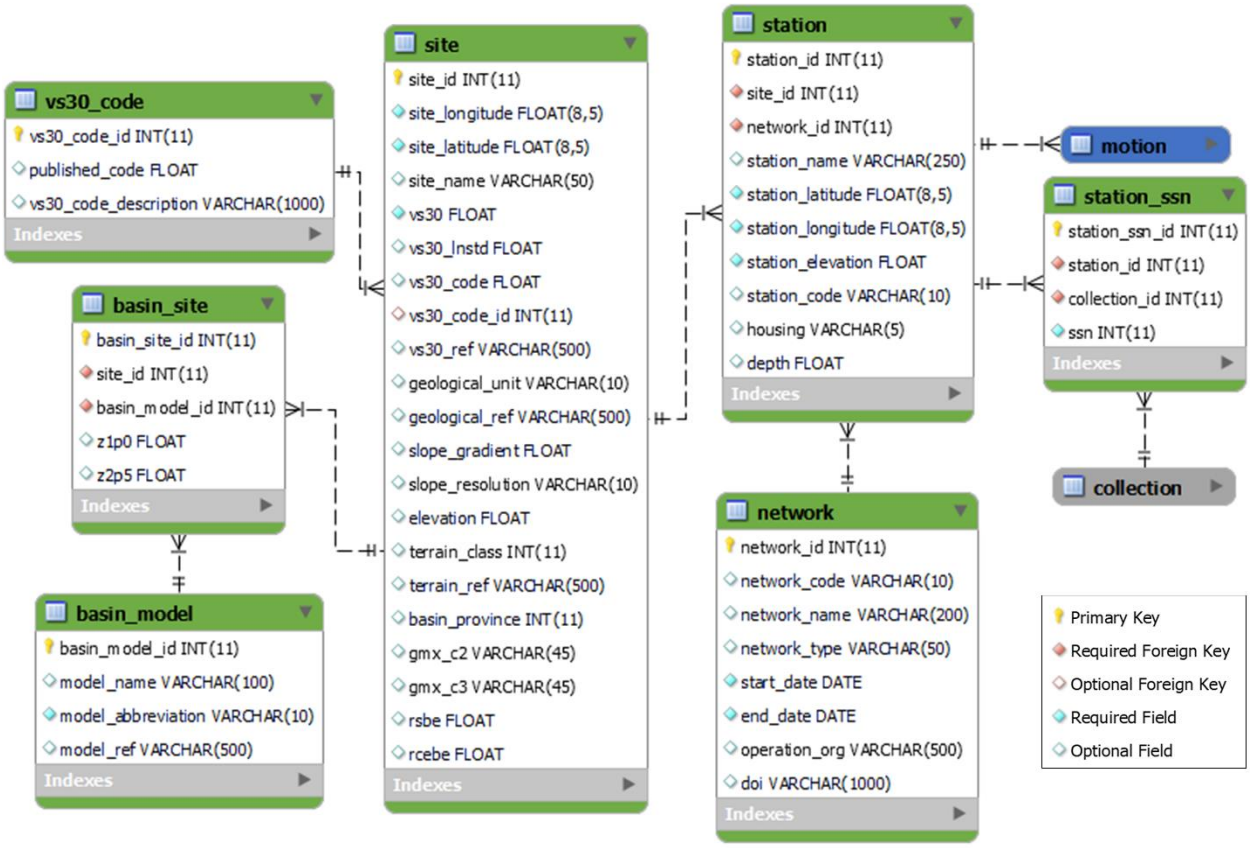


Figure 2.10: Detailed fields and datatypes for station-site tables.

Table 2.8: GMX first letter code descriptions, adopted from Ancheta et al. (2013).

GMX First Letter	Instrument Structure Type
A	One-story structure of lightweight construction. Instrument is located at the lowest level and within several feet of the ground surface.
A,B	Small generally lightweight structure for which we cannot determine the number of stories from the available information. These sites generally have COSMOS site code 4 which defines a reference station described as either a 1- or 2-story, small, light building.
B	Two- to four-story structure of lightweight construction, or very large tall one-story warehouse-type building. Instrument is located at the lowest level and within several feet of the ground surface.
C	One- to four-story structure of lightweight construction. Instrument is located at the lowest level in a basement and below the ground surface.
D	Five or more story structure of heavy construction. Instrument is located at the lowest level and within several feet of the ground surface.
E	Five or more story structure of heavy construction. Instrument is located at the lowest level in a basement and below the ground surface.
F	Structure housing instrument is buried below the ground surface about 1-2m, at a shallow depth. e.g., tunnel or seismic value but shallow embedment. (use 'T' for deeper embedment or 'V' for deeply embedded vaults, both not considered 'free-field')
G	Structure of light or heavyweight construction, instrument not a lowest level.
H	Dam either earth or concrete (station at toe of embankment or on abutment)
I	Free-field instrument or instrument shelter. Instrument is located at or within several feet of the ground surface, and not adjacent to any structure.
I,F	These sites generally have COSMOS site code 3 for which the sensors have been buried/set in ground at shallow or near surface depths.
J	Concrete Dam structural instrumentation.
K	Near a one-story structure of lightweight construction. Instrument is located outside on the ground surface, within approximately 3 m from the structure.
L	Near a two- to four-story structure. Instrument is located outside on the ground surface, within approximately 6 m of the structure.
M	Near a two- to four-story structure with basement. Instrument is located outside on the ground surface, within approximately 6 m of the structure.

GMX First Letter	Instrument Structure Type
N	Near a five- to eight-story structure. Instrument is located outside on the ground surface, within approximately 10 m of the structure.
O	Near a five- to eight-story structure with a basement. Instrument is located outside on the ground surface, within approximately 10 m of the structure.
P	Castle of masonry construction, massive 1-3 stories.
Q	Associated with a structure, size of structure is not known.
S	Associated with a structure and in the basement, size of structure is not known.
T	Associated with a deep tunnel.
U	On an embankment between two roads and retaining walls.
V	Deeply embedded seismic vault.
W	Structural response.
Z	At depth in borehole or missile silo.

The *site* table contains metadata related to the physical attributes of the recording location such as geologic and topographic information. Latitude, longitude, and elevation are also assigned to the site (i.e., *site_latitude, site_longitude,* and *site_elevation*), which are used to populate other metadata fields from published maps. Sites are assigned to stations through the *site_id* foreign key in the *station* table, which is a one-to-many relationship. Most sites have only been inhabited by one station, so site location is usually the same as station location, however this is not always the case. For example, a vertical array consists of two or more instruments installed at different depths, and the latitude and longitude may be identical or very similar, while elevations will differ. For all intents and purposes, stations within a vertical array are located at the same site. Moreover, a temporarily deployed station may be decommissioned and replaced by one maintained by a permanent network, in which case the new station will be assign to the existing site. Instruments in dense arrays, where intra-instrument spacing may be as little as 10 m (e.g., Borrego Valley Differential Array), are assigned unique *station_id*'s and *site_id*'s because geologic conditions may vary considerably over relatively short distances.

The *site* table also records the preferred V_{S30} value recommended for analysis. The preferred V_{S30} value is that which best reflects the median estimate as obtained through agreed upon protocols that have evolved over time (e.g., Chiou et al. 2008; Ancheta al. 2014; Wang 2020; Ahdi et al. 2022). As discussed in Section 2.3.1(b), V_{S30} may be computed from a measured V_S profile or estimated from proxy-based models. The *site* table includes *vs30_code_id* and *vs30_ref* fields to record this type of information. An integer valued primary key is assigned to each method discussed in Section 2.3.1(b) in the *vs30_code* table, which is used to populate the *vs30_code_id* field. Current methods and their corresponding integer values are summarized in Table 2.9. The

vs30_lnstd field records the uncertainty associated with the preferred V_{S30} value. The *vs30_ref* entry provides additional information related to how the preferred V_{S30} was obtained, such as the reference to the V_S profile used to calculate V_{S30} through an abbreviation to the data provider and an associated identifier. Current providers of measured-V_S information include:

- CESMD (stations with calculated V_{S30}): "CESMD – network-station code"
- NGA-West 2 flatfile (stations with V_{S30} code = 0): "NGAW2 – ssn"
- NGA-East flatfile (stations with V_{S30} code = 0): "NGAE – ssn"
- USGS V_{S30} Compilation (Yong et al. 2015): "USGS – Id"
- VSPDB: "VSPDB – profile id"

For sites where V_{S30} is estimated from proxy-based methods, *vs30_ref* describes the type of proxy model(s) used. Abbreviations for proxy-models types include:

- "Geo": geology- or hybrid geology-slope proxies (e.g., Wills et al. 2015).
- "Kri": Kriging interpolation if site is located near measured V_{S30} (e.g., Thompson 2018).
- "Terr": terrain-based proxies (e.g., Yong et al. 2016).
- "Pea17 – *group number*": proxy-based V_{S30} estimation in Central and Eastern North America (CENA) proposed by Parker et al. (2017).

Note that the expanded list of proxy models implemented in the NGA-Sub project (Ahdi et al. 2022) have not yet been implemented in the GMDB.

Table 2.9: Summary of V_{S30} codes and their definitions.

ID	Description	Standard Deviation ($\sigma \ln V$)
0	V_{S30} computed using profile with $z_p \geq 30$ m.	0.1
1	Profile is available but maximum V_S profile depth $z_p < 30$ m. V_{S30} is estimated using the extrapolation relationship of Boore (2004), with California-specific regression coefficients from Kwak et al. (2017a).	$(0.12 + \sigma_e)^{0.5}$ Values of σ_e given in Kwak et al. (2017a)
2	Estimate V_S profile from correlations between V_S and penetration resistance/effective stress. V_{S30} computed from estimated profile.	0.25
3	No Profile available. V_{S30} and its variability are assigned based on the weighted combination of (1) geology (pre-Quaternary) or hybrid geology-slope proxies (Wills et al. 2015 or local correlation for particular geologic units) and (2) CA-specific models based on geomorphic terrain categories (Yong 2016; Iwahashi and Pike 2007). For the geology-based model, Kriging is used if location is near measured V_{S30} (Thompson et al. 2014; Thompson 2018).	based on weighted combination
4	No profile available. Surface geology not available. V_{S30} and its variability are estimated using CA-specific models based on geomorphic terrain categories (Yong 2016; Iwahashi and Pike 2007).	based on category statistics
5	No profile available. Terrain-based V_{S30} estimate not available (typically because terrain class has null or sparse data). V_{S30} estimated using CA-specific models based on geology (pre-Quaternary) or hybrid geology-slope proxies (Wills et al. 2015). Kriging used if near measured V_{S30} (Thompson et al. 2014; Thompson 2018).	based on category statistics
6	No profile available. Site in Central and Eastern North America (CENA), V_{S30} and its variability are estimated using CENA-specific model based on geology, topography, glaciation, and location in or outside of a basin (Parker et al. 2017).	based on category statistics
7	Known site conditions and geology based on measurements of V_S profiles at different location but the same geological condition. This assignment is only used based on a recommendation or site visit from a geologist.	0.3 (Goulet et al. 2014)
8	Estimate mean V_{S30} by P-wave seismogram method (Kim et al., 2016) for sites having multiple ground-motion recordings and corresponding V_{S30} values from measurements.	0.456

The *site* table includes a *basin_province* field which corresponds to the site geomorphic province classification as defined by Nweke et al. (2022). Integer-valued codes are used to facilitate efficient queries. Table 2.10 provides the definitions for *basin_province* codes, adopted from Nweke et al. (2022).

Table 2.10: Definition of *basin_province* codes, adopted from Nweke et al. (2022).

Code	Definition	Description
0	Mountain-Hill	Sites without significant sediments, generally having topographic relief
1	Valley	"Small" sedimentary structure
2	Basin Edge	Along the basin margin
3	Basin	Site located within the basin interior

The *basin_model* and *basin_site* tables are used to assign seismic velocity isosurface depths ($z_{1.0}$ and $z_{2.5}$) to sites within the geographic domain of basin models. Given that the boundaries described by several velocity – or basin – models may overlap (e.g., CVM-S and CVM-H in southern California), the GMDB needed a way to allow for multiple estimates to be assigned. Furthermore, we store multiple depth estimates because the seismological and earthquake engineering communities have yet to agree upon protocols to assign preferred depths. To meet this need, the *basin_site* table acts as a "many-to-many" junction between the *site* and *basin_model* tables via relationships of *site_id* and *basin_model_id* foreign keys. The *basin_model* table contains identifying information for each model (i.e., full name, abbreviation, and reference citation), and each model is assigned a unique *basin_model_id*. Within the *basin_site* table pairs of *basin_model_id* and *site_id* are assigned appropriate $z_{1.0}$ (*z1p0*) and $z_{2.5}$ (*z2p5*) values. An

advantage of this structure is that it is flexible enough to accommodate additional estimates as newer models are published.

Just as NGA *EQID*'s are not used as *event_id*'s, station sequence numbers (*SSN*'s) assigned by NGA products are not adopted for *station_id*'s. Stations found in multiple NGA databases (e.g., NGA-West2 and NGA-Sub or NGA-West2 and NGA-East) sometimes have different *SSN*'s, so we assign unique *station_id*'s in the *station* table to ensure that the primary keys are unique. Users may wish to search the GMDB using familiar *SSN*'s assigned by individual collection efforts unified in the GMDB (e.g., NGA-West2, NGA-East, etc.), therefore the *station_ssn* table provides this mapping. There are four fields in the *station_ssn* table: (1) *station_ssn_id* which acts as the primary key, (2) *station_id* foreign key linked to the *station* table, (3) *collection_id* foreign key linked to the *collection* table [discussed in Section 2.4.1(e)], and (4) *ssn* which records the *SSN* assigned by the individual collection (e.g., NGA-West2). The relationships of *station_ssn_id* to *collection_id* and *station_id* are both "one-to-many" relationships meaning that a given *station_id* can only have one *SSN* for a given *collection_id*, however individual *station_id*'s may appear numerous times in *station_ssn* if they exists in multiple collections.

(c) Ground Motion Metadata Tables

Ground motion metadata are any remaining metadata not directly associated with the time-series themselves, and are organized into two tables: *motion* and *collection_motion*. Figure 2.11 presents the detailed fields, their data types, and key relationships between ground motion metadata tables in the GMDB. The *motion* table acts as a junction table between the *event* and *station* tables by assigning unique *motion_id*'s to each source-to-site record. All distance metrics calculated in Section 2.3.1(c) are stored within the *motion* table, in addition to a *hanging_wall* indicator.

Hanging wall indicators are defined in Table 2.11. The GMDB also assigns a *user_id* [discussed in Section 2.4.1(e)] to each record to document who is responsible for uploading the ground motion data.

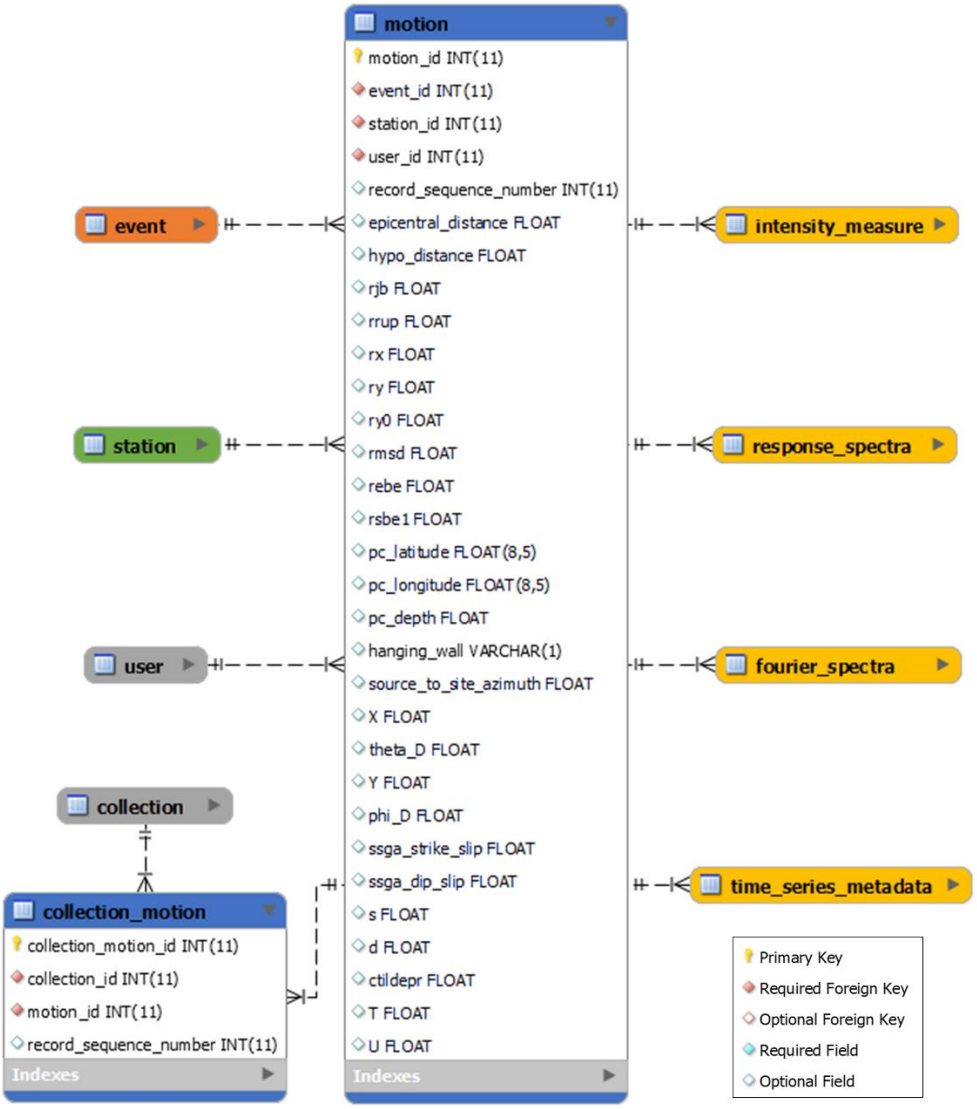


Figure 2.11: Detailed fields and datatypes for ground motion metadata tables.

Table 2.11: Summary of hanging wall indicators, adopted from Abrahamson and Somerville (1996) and used by Ancheta et al. (2013).

Code	Abrahamson and Somerville (1996)	Definition
1	hw	Site is within the hanging wall region
2	fw	Site is within the footwall region
3	nu	Site is in the neutral region
4	na	Not applicable as fault dip is greater than 70^{o}

The *collection_motion* table is a junction table relating the *collection* and *motion* tables that contains a primary key (*collection_motion_id*), two foreign keys (*collection_id* and *motion_id*), and a record sequence number (*RSN*; *record_sequence_number*) assigned by an individual collection (e.g., NGA-West2). The purpose of the table is to group sub-datasets of records together within the database structure to facilitate efficient group requests (i.e. particular subsets of ground motions), and to map *RSN*'s from external databases to GMDB *motion_id*'s (only for motions from unified datasets). The many-to-many relationships defined in the table mean that a *collection_id* may have many *motion_id*'s, and an individual *motion_id* may belong to several *collection_id*'s. Definition of what a "collection" represents is discussed in Section 2.4.1(e).

(d) Ground Motion Data Tables

Ground motion data include acceleration time-series files, associated processing metadata, and any calculated metrics, and are organized into six tables: *component*, *time_series_metadata*, *time_series_data*, *intensity_measure*, *response_spectra*, and *fourier_spectra*. Figure 2.12 presents the detailed fields, their data types, and key relationships between ground motion data tables in the GMDB.

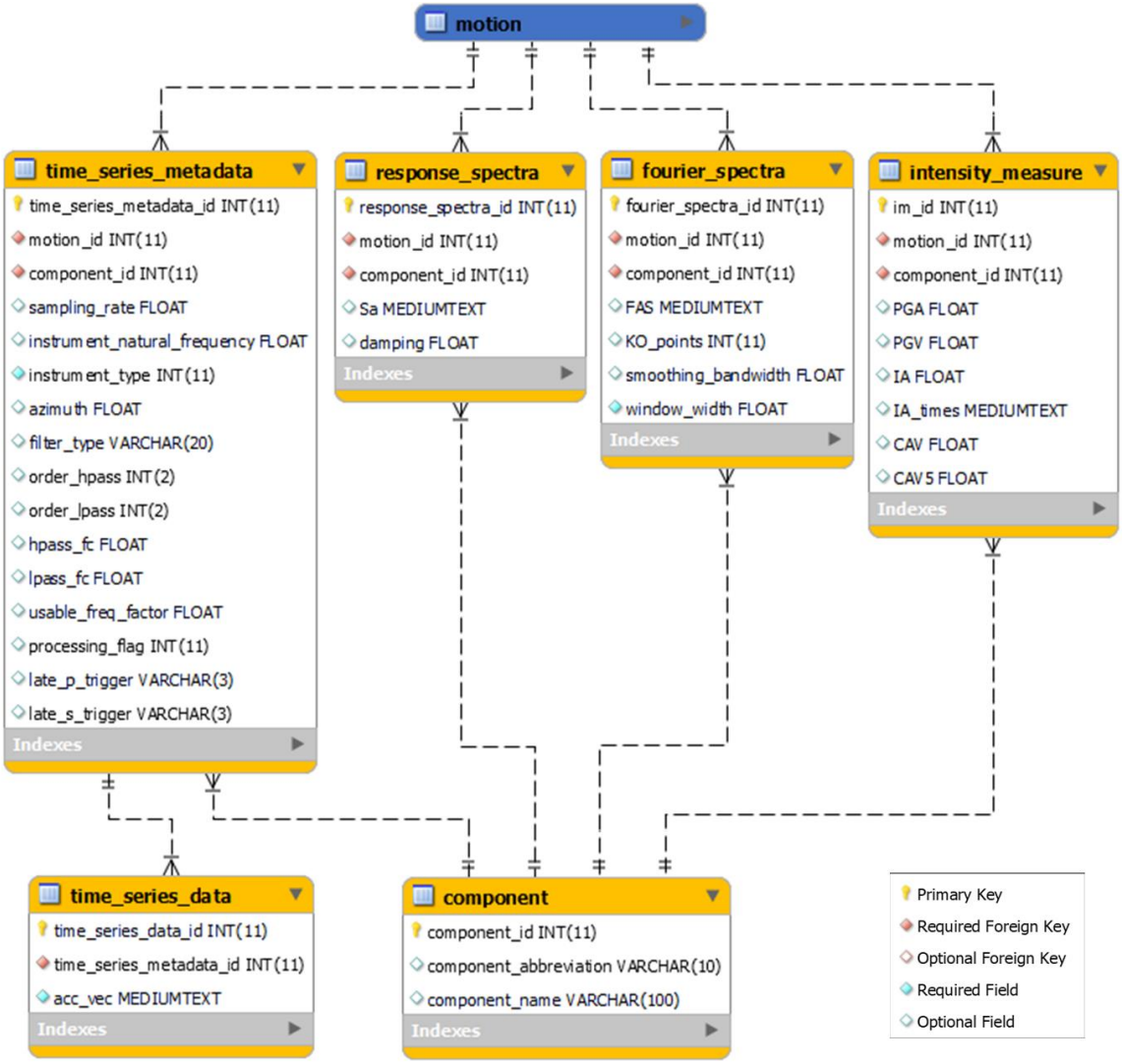


Figure 2.12: Detailed fields and datatypes for ground motion data tables.

Ground motions are comprised of three as-recorded components: two orthogonal horizontal and one vertical. Engineering applications simplify analyses by combining the two as-recorded horizontal time-series into a single orientation-independent representation, through the use of different algorithms (Boore et al. 2006; Boore 2010). By convention, the resulting combination is identified as a distinct component. The *component* table defines and assigns a unique *component_id* to each component commonly used in analyses, as shown in Table 2.12.

Table 2.12: Current GMDB *component* table.

component_id	***component_abbreviation***	***component_name***
1	H1	As-recorded horizontal 1
2	H2	As-recorded horizontal 2
3	V	As-recorded vertical
4	RotD50	Median single-component horizontal ground motion across all non-redundant azimuths
5	RotD100	Maximum single-component horizontal ground motion across all non-redundant azimuths
6	EAS	Effective amplitude spectrum; orientation-independent Fourier amplitude spectrum (FAS)
7	RotD00	Minimum single-component horizontal ground motion across all non-redundant azimuths

The *time_series_metadata* table stores information related to the time-series data and associated metadata. Each *motion_id* in the *motion* table relates to at-most three entries in the *time_series_metadata* table corresponding to the three as-recorded components (*component_id* = 1, 2, and 3). The azimuths of the two horizontal components are stored in the *azimuth* field. The type of instrument, which is an important factor when considering the reliability of the recorded data, is indicated by an integer code in the *instrument_type* field and options are defined in Table 2.13. The sampling rate at which the instrument makes measurements is recorded in the *sampling_rate* field. Processing metadata which includes the type of filter used (*filter_type*), order (*order_hpass* and *order_lpass*), and selected high- and low-pass corner frequencies (*hpass_fc* and *lpass_fc*, respectively) are also stored. NGA-West2 filter-type codes (defined in Table 2.14) are adopted and used to populate the *filter_type* field.

Table 2.13: Definition of *instrument_type* codes.

Code	Instrument Type	Type of Recording
1	Accelerometer	Acceleration
2	Seismometer	Velocity
3	MEMS Accelerometer (CSN seismic network)	Acceleration

Table 2.14: Definition of *filter_type* codes, adopted from Ancheta et al. (2013).

Code	Ancheta et al. (2013)	Filter Type
1	O	Ormsby
2	A	Acausal Butterworth
3	C	Causal Butterworth

Table 2.15: Definition of late trigger codes, adopted from Ancheta et al. (2013).

Code	Ancheta et al. (2013)	Description
1	d	Digital with pre-event memory, did not record the first P-wave onset. In the digital case we are confident that we recorded the largest amplitude (provided, later peaks are > the nominal trigger level of 0.005 g, which depends on instrument, array, etc.).
2	?	Analog recording that probably recorded largest amplitudes on the vertical component but the P-wave onset not recorded, OR probably recorded largest amplitudes on the horizontal component(s) for the S-wave.
3	Y	Late trigger, probably did not record the largest vertical amplitude, The characteristic diagnostic is that the largest amplitude is at start of recording on the vertical component.
4	DNP	Did not process the vertical component (e.g., late trigger)
5	DNR	Did not record the vertical component.

Acceleration time-series data are stored separately in the *time_series_data* table, where a *time_series_metadata_id* foreign key is used to relate each *time_series_data_id* to the *time_series_metadata* table. A single-component time-series may have upwards of 40,000 data

points, meaning that the time-series data would require around 4 billion rows for the 33,370 motions currently in the GMDB if a traditional table structure was used [i.e., primary key, foreign key, time (FLOAT), and acceleration (FLOAT) fields]. The scalability of this approach would introduce performance issues as additional ground motion data is added. Our solution is to store these data as JavaScript Object Notation (JSON) string objects in a "MEDIUMTEXT" field (*acc_vec*), which can be read and parsed by most programming languages. The JSON-strings have a dictionary-like structure where acceleration (in units of g) is represented as a list of values (e.g., {"acceleration (g)":[a_1, a_2, … , a_N]}, where a_i represents the acceleration value and N is the number of points). Times are not stored in *acc_vec* because a time-array can easily be constructed from the corresponding *sampling_rate* in the *time_series_metadata* table. A secondary benefit to this structure is that the time-series data for a single component are stored collectively, which reduces the potential of introducing errors when queried (i.e., everything is in one row instead of many).

The *intensity_measure* table stores all IMs discussed in Section 2.2.2 except *PSA* and *EAS* for each individual component. *PGA*, *PGV*, *CAV*, CAV_5, and I_A are stored as "FLOAT" objects, however JSON-strings are used to store the I_A-*times* corresponding to I_A-percentages of 5 to 95 in increments of 5 % as "MEDIUMTEXT" objects (*IA_times*). The structure of the JSON-string includes a list of percentages and a list of corresponding times (i.e., {"percentage":[5,10,…,95], "time (s)":[…]}).

PSA are stored in the *response_spectra* table as JSON-strings in the *Sa* field. The structure of these JSON-strings includes a list of non-uniformly spaced oscillator periods and a list of spectral accelerations (i.e., {"period (s)":[0.01,0.02,…,20], "Sa":[…]}). The associated damping value is stored in the *damping* field as a percent.

Similarly, *EAS* are stored in the *fourier_spectra* table as JSON-strings which includes an initial frequency (f_0), a log-normally spaced frequency step ($\Delta \log f$), and a list of *FAS* values (i.e., {"f0 (Hz)": f_0, "dlogf": $\Delta \log f$, "FAS (g-s)":[…]}). The *fourier_spectra* table also stores the number of frequencies per log-cycle (*KO_points*) and the frequency multiplier defining width (*window_width*) used for Konno-Omachi smoothing as well as the bandwidth window defined by Kottke et al. (2021) (*smoothing_bandwidth*).

All three IM data tables include *motion_id* and *component_id* as foreign keys.

(e) Auxiliary Information Tables

Auxiliary information describes anything that is not inherently related to ground motion data, and are organized into three tables: *collection*, *user*, and *geometry*. Figure 2.13 presents the detailed fields, their data types, and key relationships between auxiliary information tables in the GMDB. The *collection* table stores metadata related to a "collection of motions", such as a name, description, and reference-citation (if related to a publication). Types of collections include published datasets (e.g., NGA-West2), sub-datasets describing regional data (e.g., California ground motions), and specific datasets used by researchers. Collections referenced in the *event_eqid* and *station_ssn* tables will only be of the first type, however those referenced in the *collection_motion* table can be anything. The primary function of "collections" are to provide a means to quickly and efficiently query sub-datasets associated with particular research projects. Only authorized and authenticated users will have permission to upload and edit ground motion data and metadata, but all users will be able to define a collection (not yet implemented).

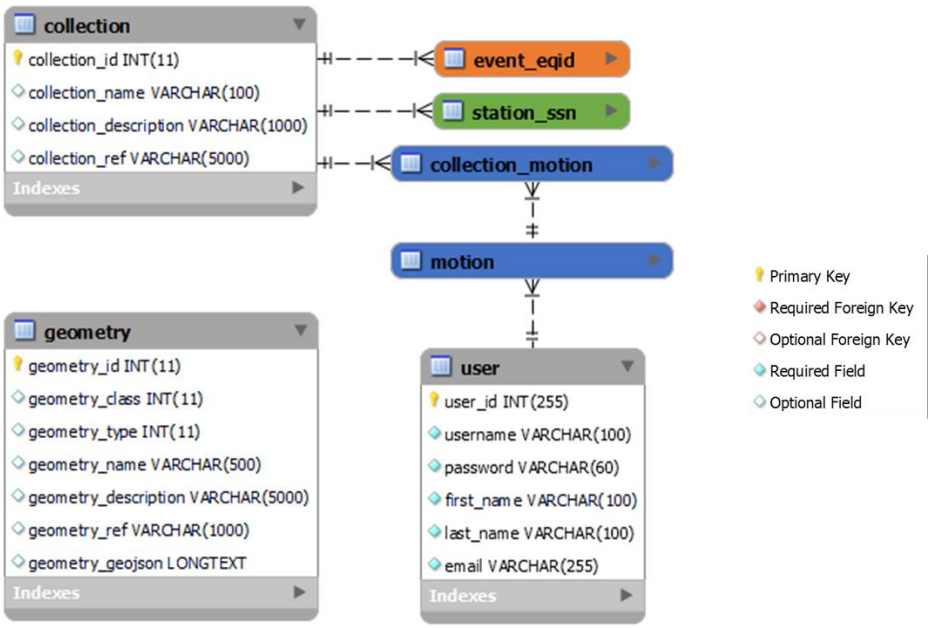


Figure 2.13: Detailed fields and datatypes for auxiliary information tables.

The *user* table contains information about each user including a username, email, and full name. Passwords are also stored in an encrypted *password* field, and are used to authenticate when users connect to the GMDB server. As a means of quality assurance, *user_id*'s are assigned to each ground motion in the *motion* table and correspond to the user who uploaded the data. Only these users will have permission to edit the ground motion data/metadata through the API [discussed in Section 2.4.3(a)]. Authorized users maintain the ability to edit all data through the back-end.

Lastly, the *geometry* table stores geographic and associated metadata for defining geographic features. The purpose of this table is to provide a container for supplementary geospatial data commonly used in ground motion analyses, and as such it is the only table in the GMDB to not have any foreign key relationships. Data are stored as geospatial JSON-strings (geoJSON) in the *geometry_geojson* field ("LONGTEXT"), which can be read and parsed by most

programming languages or GIS software. The *geometry_class* field is used to identify the nature of the type of data contained within the geoJSON, and is populated using codes defined in Table 2.16. The geoJSON geometry type is stored in the *geometry*_type field using codes defined in Table 2.17. Name, description, and reference (if associated with a publication) are also stored for each entry.

Table 2.16: Definition of *geometry_class* codes.

Code	Description	Examples
1	Global region	NGA-West2 regions (California, Japan, New Zealand, etc.)
2	Local region	Sacramento-San Joaquin Delta; *The Geysers* Boundary
3	Geomorphic feature	Basin Outline; Volcanic Arc; Tectonic Regime; Glacial Extents
4	Modeling related	Described boundaries; Sub-regions for anelastic path model

Table 2.17: Definition of *geometry_type* codes.

Code	Description
0	Point
1	Multipoint
2	LineString
3	MultiLineString
4	Polygon
5	MultiPolygon
6	GeometryCollection

2.4.2. Data Organization and Transfer to Relational Database

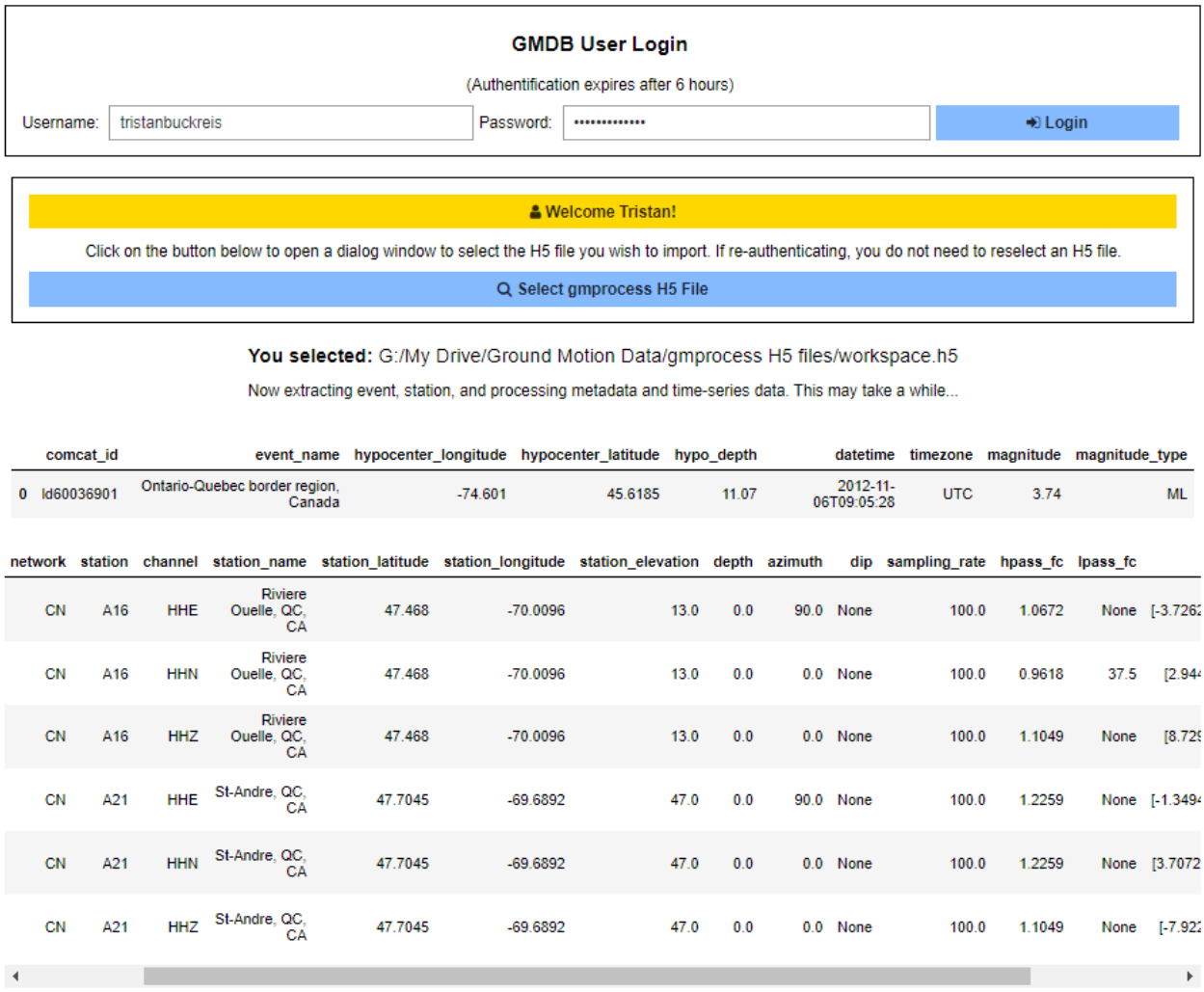
An advantage of the GMDB having been developed as a web-serviced and publicly accessible resource is that users can access the data as soon as they are uploaded. The main drawback to this

approach, however, is the added responsibility to ensure that the data are high-quality and the associated metadata are reliable at the time of upload (i.e., adequately checked to minimize errors). Ground motion processing has evolved to point where semi-automated procedures can be used to select corner frequencies. One such approach is implemented in the "gmprocess" package in Python (Hearne et al. 2019) with the addition of a manual review step where an analyst has the option to accept, reject, or modify the corner frequencies selected by the automated processing procedure based on visual inspection of the acceleration and displacement traces, *FAS*, and response spectrum (Ramos-Sepulveda et al. 2023). The output of this procedure is a hierarchical data format file (HDF5) that contains information about the event, stations with recordings, processing parameters, and the processed time series data. The processing protocols used are consistent with those recommended by PEER (e.g., Goulet et al 2021b). With the processing tools having been developed by others, my work was focused on metadata, specifically the development of tools to ensure that the associated metadata are correct.

I developed a Python based Jupyter Notebook (Kluyver et al. 2016) that takes as an input the HDF5 file produced by gmprocess. The tool facilitates the assignment of associated metadata and uploads that information to the GMDB using interactive widgets from the "ipywidgets" package (Jupyter Development Team 2022). Given the fluidity of the GMDB schema to change over time, it is necessary to update this Jupyter Notebook in parallel as changes occur. Therefore, the version discussed herein (version 1.2) reflects the current schema at the time of writing. The application programming interface (API) discussed in the following section [Section 2.4.3(a)] allows this Notebook to interact directly with the GMDB through a secure connection using the "requests" package (Reitz 2022). Other packages which are fundamental to the functionality of the Notebook include "folium" (Python-Visualization 2020), "json" (Davis-Foster 2022), "matplotlib"

(Hunter 2007), “numpy” (Harris et al. 2020), “pandas” (Pandas Development Team 2020), “pyproj” (Snow et al. 2021), and “scipy” (Virtanen et al. 2020).

Only authenticated users have permission to upload data to the GMDB, therefore the first step in the Notebook is for users to log into their GMDB account. If unsuccessful, a warning will be displayed and all subsequent functionality of the Notebook will be disabled. If successful, the Notebook will prompt users to select an HDF5 file containing the ground motion data they wish to upload. The Notebook will then automatically extract the available event, station, and processing metadata as well as the processed time-series and add them to pandas DataFrames for review and manipulation within the code. Figure 2.14 shows a screenshot of the interface after the Notebook has successfully extracted all available information, providing sample displays for the user to confirm that the data have been correctly extracted.



	comcat_id	event_name	hypocenter_longitude	hypocenter_latitude	hypo_depth	datetime	timezone	magnitude	magnitude_type
0	id60036901	Ontario-Quebec border region, Canada	-74.601	45.6185	11.07	2012-11-06T09:05:28	UTC	3.74	ML

network	station	channel	station_name	station_latitude	station_longitude	station_elevation	depth	azimuth	dip	sampling_rate	hpass_fc	lpass_fc	
CN	A16	HHE	Riviere Ouelle, QC, CA	47.468	-70.0096	13.0	0.0	90.0	None	100.0	1.0672	None	[-3.726
CN	A16	HHN	Riviere Ouelle, QC, CA	47.468	-70.0096	13.0	0.0	0.0	None	100.0	0.9618	37.5	[2.94
CN	A16	HHZ	Riviere Ouelle, QC, CA	47.468	-70.0096	13.0	0.0	0.0	None	100.0	1.1049	None	[8.72
CN	A21	HHE	St-Andre, QC, CA	47.7045	-69.6892	47.0	0.0	90.0	None	100.0	1.2259	None	[-1.349
CN	A21	HHN	St-Andre, QC, CA	47.7045	-69.6892	47.0	0.0	0.0	None	100.0	1.2259	None	[3.7072
CN	A21	HHZ	St-Andre, QC, CA	47.7045	-69.6892	47.0	0.0	0.0	None	100.0	1.1049	None	[-7.92

Figure 2.14: Screenshot of Jupyter Notebook authentication step, including HDF5 selection and display of extracted data.

The next step is to search the GMDB event metadata tables (i.e., GMDB earthquake catalog) for potential matches to the event under consideration. Origin date, time, **M**, and hypocenter latitude, longitude, and depth represent unique identifying attributes, and are auto-populated into a form the user can edit. These parameters are used by the Notebook to constrain the query. The objective is to identify existing events within the GMDB with similar identifying attributes (i.e., events occurring on the same day in the general vicinity with **M** ± 0.75). If any

potential matches exist, the Notebook prompts the user to identify the correct entry as shown in Figure 2.15. The user then has the option to select the *event_id* of the existing earthquake from a drop-down list of potential matches, or to specify that the earthquake does not exist in the GMDB. If they select "Earthquake Already in GMDB", the Notebook takes note of the selected *event_id* and proceeds to the next step. Otherwise, a new form appears requesting the user to input the preferred earthquake metadata. This prompt also appears if the initial event search yields no potential matches, as shown in Figure 2.16.

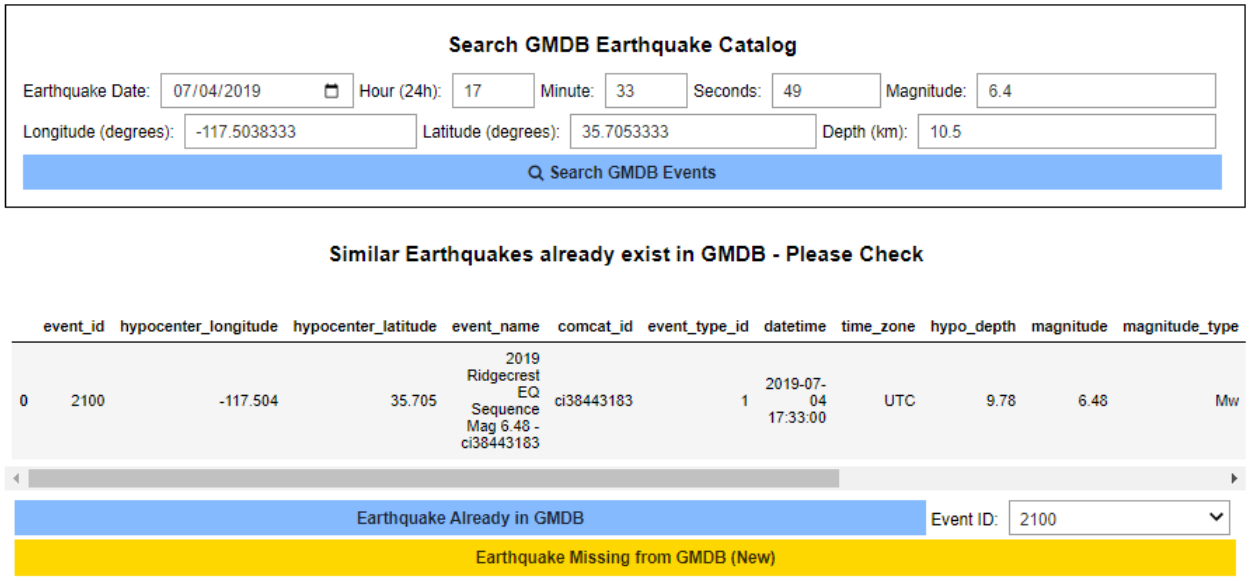


Figure 2.15: Screenshot of Jupyter Notebook event screening with potential existing event.

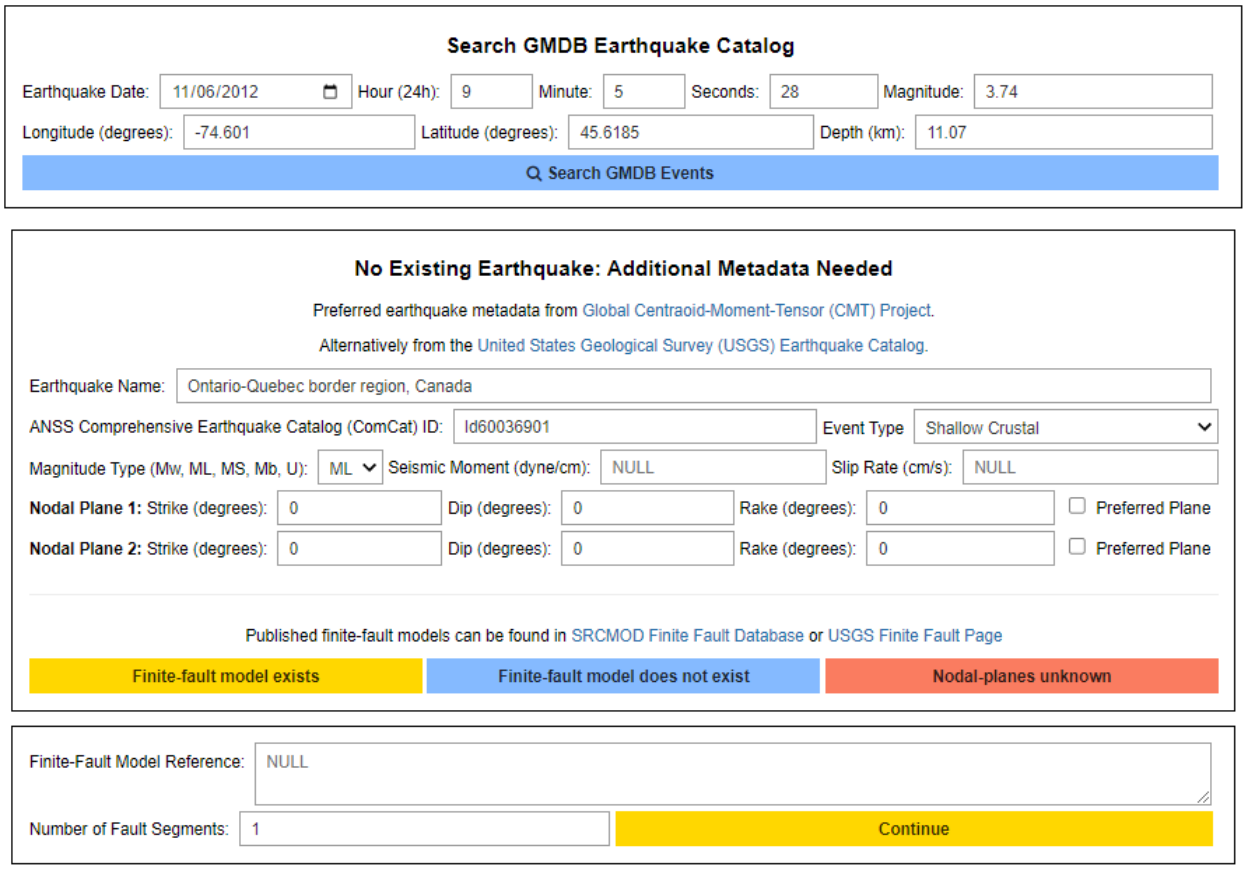


Figure 2.16: Screenshot of Jupyter Notebook event screening with no potential existing events assuming the event has a published finite-fault model.

The event metadata provided in the HDF5 file produced by gmprocess does not generally match metadata from preferred sources [Section 2.3.1(a)], therefore the user is responsible for ensuring that preferred values are uploaded. The Notebook provides links to the CMT search portal and USGS event page (if it exists for the event in question). Authenticated users should be trained on how to identify this metadata and navigate those resources to find the necessary information and manually populate the corresponding fields in these forms.

A potentially important aspect of event metadata pertains to the geometry of the finite fault. Users are prompted to select whether to upload a rupture surface representation derived from a

published finite-fault model (if one exists) or to simulate one. The Notebook provides links to the Earthquake-RC Source Model (SRCMOD) Database (Mai and Thingbaijam 2014) and the USGS finite-fault page where published finite-fault representations can be found. If multiple finite-fault models exist, the user is responsible for selecting and trimming the preferred model using the criteria and procedures outlined in Contreras et al. (2022). To upload a representation from literature, users first select "Finite-fault model exists" where they are then prompted to enter the reference citation and the number of rupture segments, as shown in Figure 2.16. After clicking "Continue", a new form will be displayed which allows the user to enter information for each segment individually, which is shown in Figure 2.17. Alternatively, if no finite-fault models exist the user can select one of two options to simulate the rupture surface: (1) "Finite-fault model does not exist" or (2) "Nodal-planes unknown". If nodal planes are known users are advised to select the first option in which the orientation of the two nodal planes are used during the simulation procedure, otherwise the second option should be selected. The Notebook will automatically run CCLD5 (Contreras et al. 2022) implementing the inputs specified by the user to simulate the rupture surface.

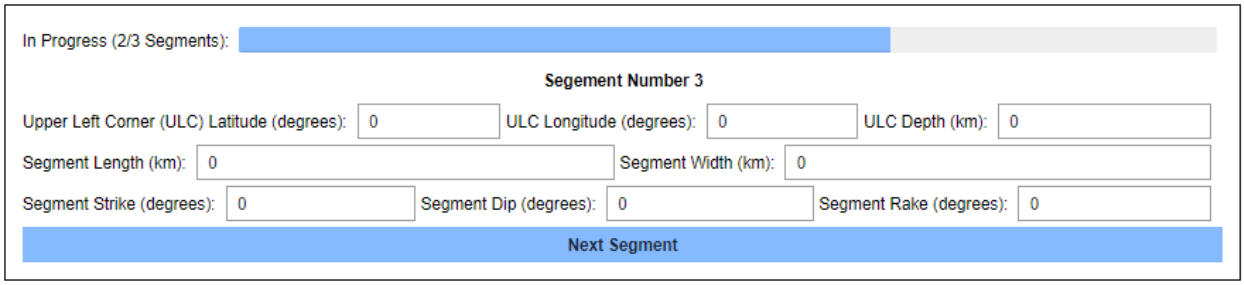


Figure 2.17: Screenshot of Jupyter Notebook finite-fault segment entry form.

After the user is finished entering the finite-fault segment information or the simulation procedure is run, the Notebook will display the event metadata for review, as shown in Figure 2.18. If the metadata are satisfactory the user can select "Push Earthquake Metadata to GMDB Server", after which the Notebook will construct SQL commands to insert new rows into the *event*, *finite_fault*, and *finite_fault_segment* tables. These commands are sent to the GMDB via a "post" request through the API which requires authentication. Primary keys (e.g., *event_id*, *finite_fault_id*, and *finite_fault_segment_id*) are automatically assigned by the database, therefore SQL commands are executed in a systematic order to ensure that primary and foreign key relationships remain intact. The Notebook will inform the user of the *event_id* assigned to the event, and prompt them to move forward to the next step.

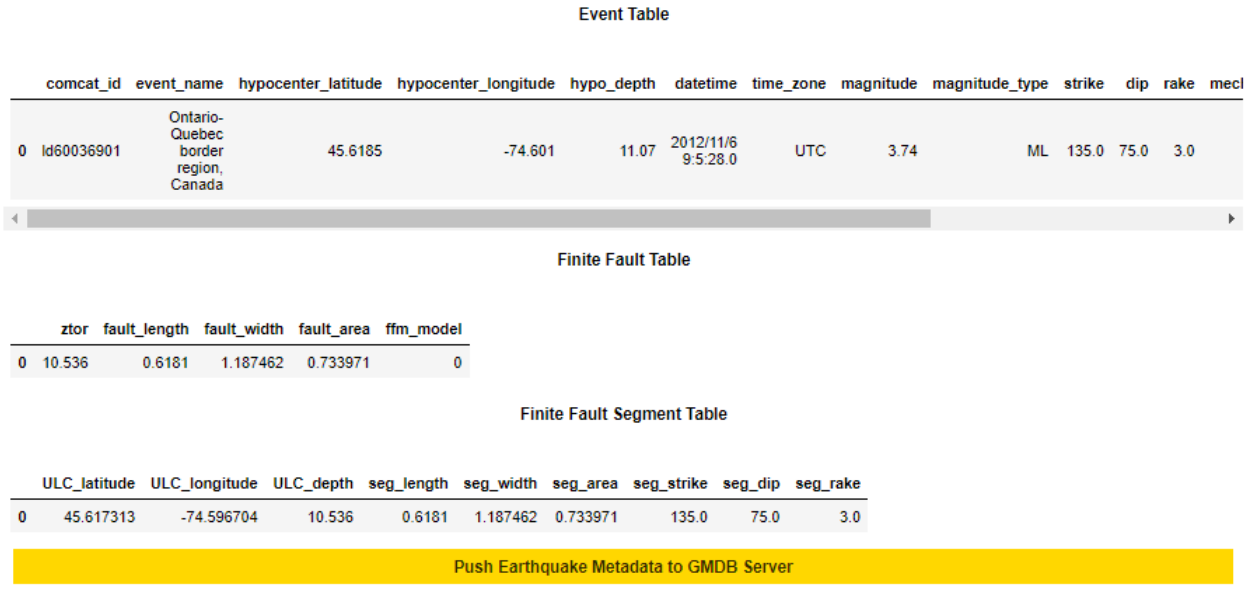


Figure 2.18: Screenshot of Jupyter Notebook display of event metadata tables ready to upload to GMDB after user review.

New ground motion data may become available after an initial batch of records have been uploaded for an event. To accommodate these scenarios, the Notebook next searches the existing GMDB for matches with the new records to ensure that no duplicate records are uploaded. This check is performed by first identifying GMDB *station_id*'s which correspond to the new dataset based on the unique network code and station code assignments. If a particular station is not able to be mapped to a *station_id* in the GMDB, it is flagged as being a "new station" and the site assignment is resolved in a future step. Otherwise, the Notebook will query the GMDB through the API for all *station_id*'s with records of the event, and compare those to the *station_id*'s in the new dataset. If any potential matches exist, the Notebook alerts the user, otherwise the Notebook proceeds to the next step.

Seismic networks are assigned a two-digit network code by the FDSN, which is unique for any permanent network. However, temporary networks operating over non-overlapping time spans may be assigned the same network code (e.g., "YU" was assigned to "Northern California Delta" which operated from 2006 to 2014, and later assigned to the "Caribbean-Merida Andes Experiment" which operated from 2016 to 2018). The Notebook will compare all network codes within the new dataset to those in the *network* table conditioned on the origin time of the earthquake. To ensure that the Notebook correctly identifies network assignments, it will display the results for the user to review. If a network does not currently exist in the GMDB *network* table, the Notebook will prompt the user to upload the network metadata during this step, as illustrated in Figure 2.19. Most of the information in the network form is automatically populated using a web-services "get" request through the FDSN's API. A link to the network's FDSN page is provided in the Notebook for users to gather the missing information needed to populate the operation organization and DOI fields. When the "Push Network Metadata to GMDB Server"

button is pressed, the Notebook constructs an SQL command to insert the new network into the *network* table and posts it through the GMDB API. A *network_id* is automatically assigned to any new network. Once all networks in the new dataset are identified in the *network* table, the Notebook will assign appropriate *network_id*'s to all records.

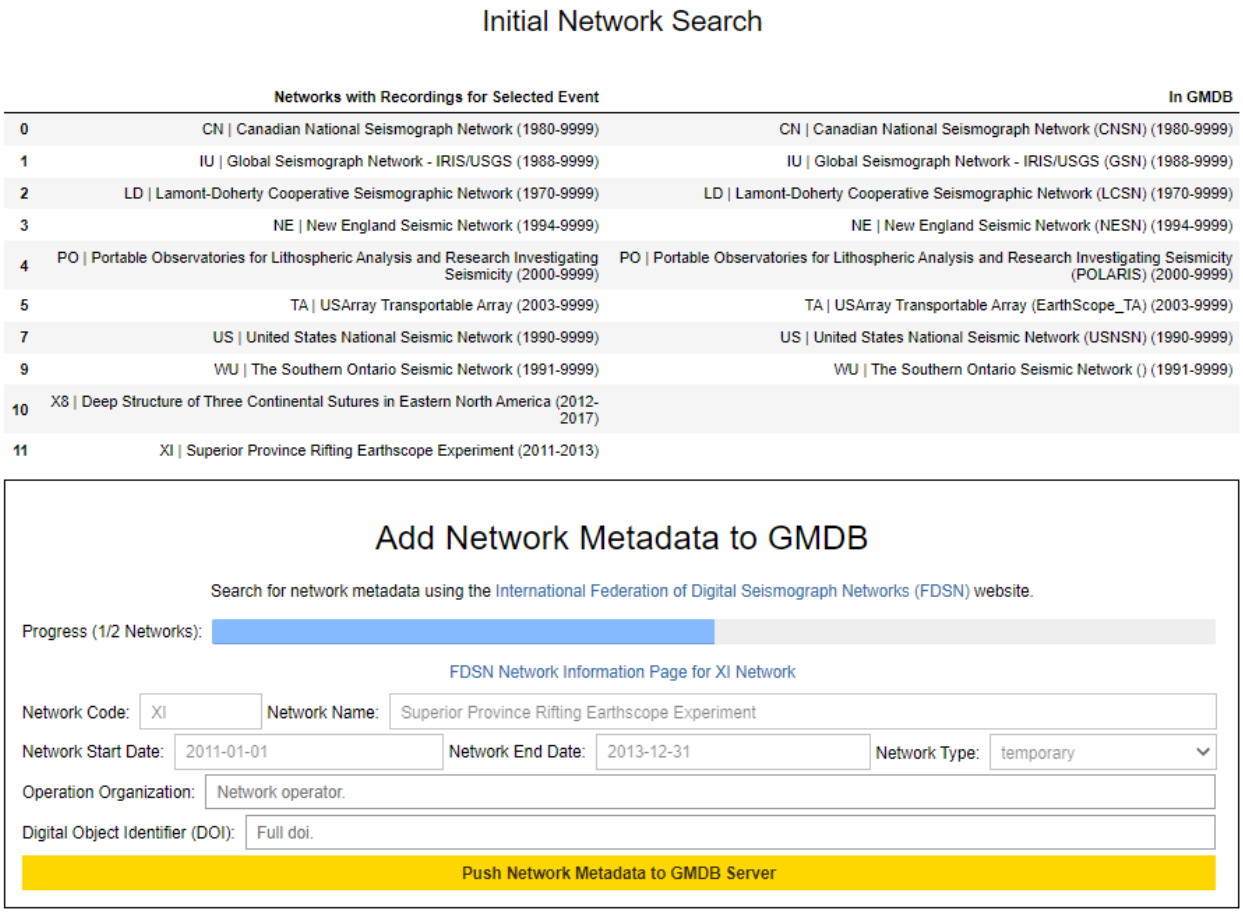


Figure 2.19: Screenshot of Jupyter Notebook screening for seismic networks and form to upload new network metadata.

Following completion of the above steps, the Notebook will have assisted the user in uploading event and network related metadata information, as well as assigning *event_id* and *network_id* keys. If there are no new stations, the Notebook will assign *station_id*'s to each record and move on to the next step. However, if there are any new stations, the next step is to assign *station_id*'s and their corresponding *site_id*'s, which is not trivial as demonstrated in Section 2.3.2. To reduce the possibility of entering duplicate information, the Notebook requires the user to manually screen all potentially new sites (i.e., new stations) against those already in the GMDB on a one-by-one basis. This step in the process was implemented because (1) the number of stations which record a particular event is seldom greater than about a hundred, (2) in regions with existing ground motion data (e.g., California) the number of new stations will be a fraction of the total number, and (3) individual inspection helps to ensure assignment of correct station and site metadata.

The Notebook will create an interactive form which includes a dropdown list of all the new stations. For each new station the Notebook will:

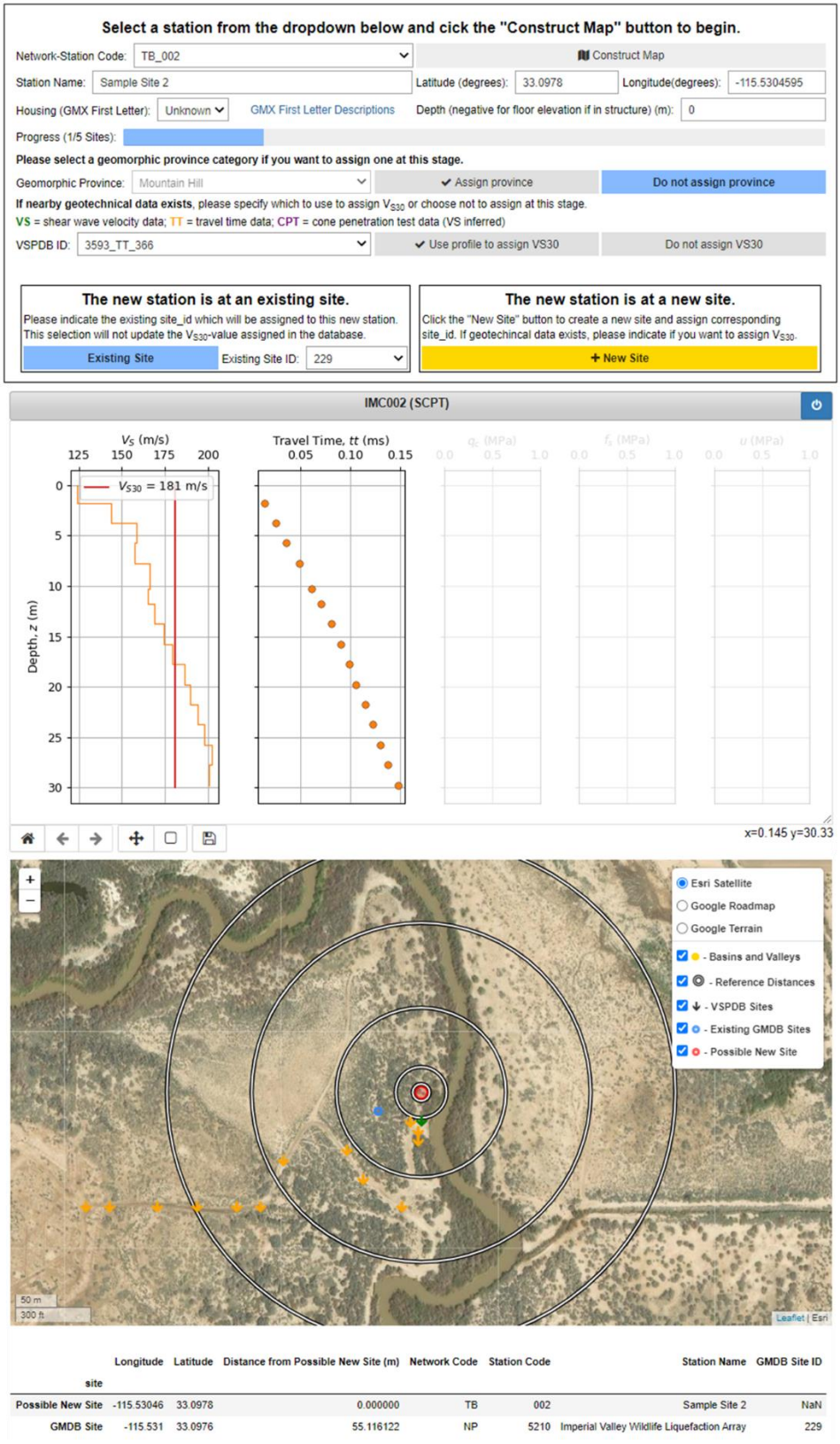
1. Automatically populate editable fields with the station name, latitude, longitude, and depth provided by gmprocess;
2. Display fields which may be used to assign station housing, site geomorphic province, and V_{S30} from measured data;
3. Query the GMDB for any nearby existing sites;
4. Query the VSPDB (Ahdi et al. 2018) for nearby geotechnical data including V_S profiles, travel time data, and cone penetration test (CPT) data;

5. Construct an interactive map that includes the location of the new station, locations of any nearby existing GMDB sites, locations and type of any nearby VSPDB geotechnical data, and concentric circles representing reference distances of 10, 30, 100, 200, and 300 m centered about the new station; and
6. Allow the user to specify if the new station is at an existing site, or if a new site should be created and *site_id* assigned.

A screenshot of the display is presented in Figure 2.20.

The decision to allow users the ability to manually override the station name, latitude, longitude, and depth provided by gmprocess was made because the information provided by gmprocess may not be fully accurate. Section 2.3.2 discussed the criteria and methodology used to assign these metadata. The Notebook does not currently provide links to appropriate resources because it is impractical to provide links for all regional data providers across the globe. Therefore, the burden falls on the user to gather these data in the current version of the Notebook. However, the map constructed by the Notebook includes satellite imagery which may be useful to improve coordinate locations of free-field stations.

The station housing field adopts the GMX first letter codes defined in Table 2.8, and can sometimes be assigned based on satellite imagery. The geomorphic province classes are defined in Table 2.10 and are selected using judgement based on knowledge of the local geologic structures and inspection of terrain maps, as discussed in Nweke et al. (2022). If the user is not confident assigning these parameters at this stage, they have the option to leave them as "NULL" (i.e., select "Unknown" and/or "Do not assign province"). Authenticated users have the ability to update entries through the API outside of the Notebook, so NULL entries are expected to be updated in the future.



site	Longitude	Latitude	Distance from Possible New Site (m)	Network Code	Station Code	Station Name	GMDB Site ID
Possible New Site	-115.53046	33.0978	0.000000	TB	002	Sample Site 2	NaN
GMDB Site	-115.531	33.0976	55.116122	NP	5210	Imperial Valley Wildlife Liquefaction Array	229

Figure 2.20: Screenshot of Jupyter Notebook available options when screening new stations.

The preferred methods to assign V_{S30} are from co-located or nearby geotechnical data, as summarized in Table 2.9. If nearby geotechnical data exists, the Notebook will populate the "VSPDB ID" dropdown list and construct data plots versus depth. V_{S30} is calculated (or estimated) from data at the selected profile and plotted as a vertical red band through the top 30 m of the V_S plot. Users have the option to "Use profile to assign V_{S30}" or "Do not assign V_{S30}", one of which must be selected. The option to not assign exists because the available data may be relatively far away in what the analyst suspects to be incompatible geology, or the data may have apparent issues. If the user selects to assign V_{S30}, then the Notebook also estimates the uncertainty (*vs30_lnstd*) and assigns the appropriate *vs30_code_id* as defined in Table 2.9 and *vs30_ref*.

If the new station is co-located at an existing GMDB site, the user does not need to assign geomorphic province or V_{S30}. This is because those parameters have previously been assigned to the existing site, and the user should indicate this by selecting the appropriate *site_id* from the "Existing Site ID" dropdown list and then click "Existing Site". Otherwise, the user is urged to assign geomorphic province and V_{S30} (if available) before selecting "+ New Site". Station metadata (name, latitude, longitude, depth, and housing) will be updated for the station regardless of the site selection. Once either button is clicked the Notebook will construct and post SQL commands to insert station (and site, if new) metadata into the GMDB, assign *station_id*, *site_id*, and a "reviewed" flag to the new station, and automatically move forward to the next un-reviewed station.

The Notebook interface also allows users to go back to any reviewed station to update the station metadata or change their decision of "existing" or "new" site. However, site metadata at only "new" sites added by the user can be changed after they are initially created through the

Notebook. Furthermore, if the user updates a station from being a “new” site to an “existing” site, the Notebook will delete the “new” site if no other stations exist at that site.

This completes my presentation of the interactive portions of the Notebook. The remaining automated features perform the following tasks:

1. Calculate source-to-site distances as discussed in Section2.3.1(c);
2. Assign *user_id* to each record;
3. Upload the *motion* table and assign *motion_id*’s to each record;
4. Upload the *time_series_metadata* table and assign *time_series_metadata_id*’s;
5. Upload the processed time-series to the *time_series* table;
6. Calculate response spectra and intensity measures for the as-recorded components as discussed in Section 2.2.2; and
7. Upload the *response_spectra* and *intensity_measure* tables.

It is important to note that the current version of the Notebook (version 1.2) does not populate all fields in the GMDB schema. All essential information has been considered in the development of Notebook, but external routines are performed on the back-end to assign V_{S30} from proxy-based methods, basin depth terms $z_{1.0}$ and $z_{2.5}$, surficial geologic unit, topographic slope, and terrain class as well as to calculate response spectra and intensity measures for rotated horizontal components (i.e., RotDXX) and *EAS*. Additional work is required to manually populate other fields not currently handled within the Notebook.

2.4.3. Online Interface

The primary mission of the GMDB is to make desired ground motion data and associated metadata readily and easily accessible to the public. The SQL based structure has many advantages with

respect to storage efficiency and query performance, however not all users will be proficient enough in SQL to construct the queries needed to extract data across several tables. As a means to circumnavigate user challenges, an application programing interface (API) and online tool have been developed to assist users with accessing the data.

(a) Application Programing Interface (API)

An API is an intermediary which interacts directly with the database requiring only basic knowledge of its implementation. Direct interaction with the GMDB requires knowledge of structuring SQL commands and meticulous understanding of the schema, both of which represent potential barriers to users. We develop an API which allows users to easily query the GMDB by constructing simple query strings with only a basic understanding of the database structure and no SQL knowledge required. Complete documentation of the API can be found at https://uclageo.com/gm_database/api.

Queries can be submitted through web services in the form of a uniform resource locator (URL) in most web browsers or programming languages. The fundamental components of the URL are:

1. A “base url” utilized in the API: https://uclageo.com/gm_database/api/index.php;
2. A “resource” which describes the nature of the query (i.e., functionality); and
3. A list of parameters used to specify the query constraints and return structure (i.e., “query string”).

The GMDB API currently supports four publicized resources:

1. "schema": Provides access to the organizational structure of the database, which users must be familiar with to utilize the following resources.
2. "flatfile": Provides a single table synthesized from many tables (i.e., NGA flatfile).
3. "response_spectra": Provides a flatfile-style table of response spectra data and associated metadata along with an interactive plotting tool.
4. "fourier_spectra": Provides a flatfile-style table of *FAS* and associated metadata along with an interactive plotting tool.

A valid URL takes the form of "*base_url*/*resource*?*query_string*", where *query_string* parameters are separated by an "&" and cannot contain any spaces. The *query_string* parameters currently supported by the API are summarized in Table 2.18. Each *resource* behaves a little differently, therefore Table 2.18 also includes the associated *resources* which support each parameter. For example, if a user wishes to find out which fields are in the *event* table, their URL would be https://uclageo.com/gm_database/api/index.php/schema?tables=event&table_list=false. Additional example URLs are shown in Section 2.4.3(b).

Table 2.18: Summary of GMDB API query string parameters.

Parameter	Description	Options	Supported Resources
format	Controls the output format.	html, json, csv	schema, flatfile
limit	Controls the number of rows returned; default is 50	non-zero integer values	schema, flatfile, response_spectra, fourier_spectra
offset	Controls the starting index of the returned data; default is 0	non-zero integer values	schema, flatfile, response_spectra, fourier_spectra
table_list	Control whether a list of tables in the database is returned	true, false	schema

Parameter	Description	Options	Supported Resources
tables	Controls which tables to be included in the query; comma separated list of table names (no spaces)	"Table Name" listed in Table 2.2	schema, flatfile
fields	Controls which fields are returned by the query;	All non JSON-string fields shown in Figures 2.9-13	flatfile
<fields>	Controls the range of values for a specific field; bounds separated by a "-"; parenthesis can be used around negative bounds; example: PGA=0.1-0.2 will return entries with PGA values between 0.1 and 0.2g.	varies	flatfile, response_spectra, fourier_spectra
sortby	Controls the field by which to sort the data	All non JSON-string fields shown in Figures 2.9-13	flatfile, response_spectra, fourier_spectra
order	Controls whether to sortby field should be ascending or decending; default is ascending	ASC, DESC	flatfile, response_spectra, fourier_spectra
parse_json	Controls whether to parse data from native JSON-string formats into a column format; default is true	true, false	flatfile
fill_null	Controls the number to be inserted into NULL fields when parsing JSON-strings; only applicable if parse_json=true; example: NGA products use "-999"	Any number	flatfile

URL's passed through the API are parsed on the server's end to construct the necessary SQL command before returning data to the user. In the event that users specify more than one table for the *tables* parameter, the API will check that all tables are joined based on the expected primary-to-foreign "key chain". Best practice is to format *query strings* with table lists that form continuous key chains. For example, "*tables=network,station,site*" is a continuous key chain from the *network* to the *site* table through mutual relationships with the *station* table. A broken key chain is one which contains one or more gaps (e.g. "*tables=network,site*"). The API will fill in any gaps as a means to construct a valid query, however the data which is returned will contain entries from

tables not specified in the query string. It follows that the API is not a fool-proof method to retrieving only desired data, but does simplify the overall process.

(b) URL Builder Tool

To help users construct valid URL's, I developed a "URL Builder Tool" in hypertext markup language (HTML) which uses JavaScript to facilitate dynamic interaction. The tool can be accessed at https://uclageo.com/gm_database/api/url_builder.php, and currently only supports the *schema* and *flatfile* resources. The advantage of this format is that users do not need any pre-installed software to use the tool, so long as they have a functioning web browser. The webpage provides an interactive user interface to construct URL's conditioned on the *query_string* parameters described in Table 2.18.

Users first select the API *resource* they wish to utilize, after which the webpage will display the appropriate forms the user can populate to construct the *query_string*. Figure 2.21 presents a screenshot of the options available under the *schema* resource. The URL constructed using the available options is displayed in a box labeled "Your URL", and is visible throughout the process even when the page is scrolled down. Users have the option to click the URL, which will open the request in a new tab, or to copy the text to their clipboard for easy integration into programming language-based web services requests (e.g., Python, R, Matlab, etc.).

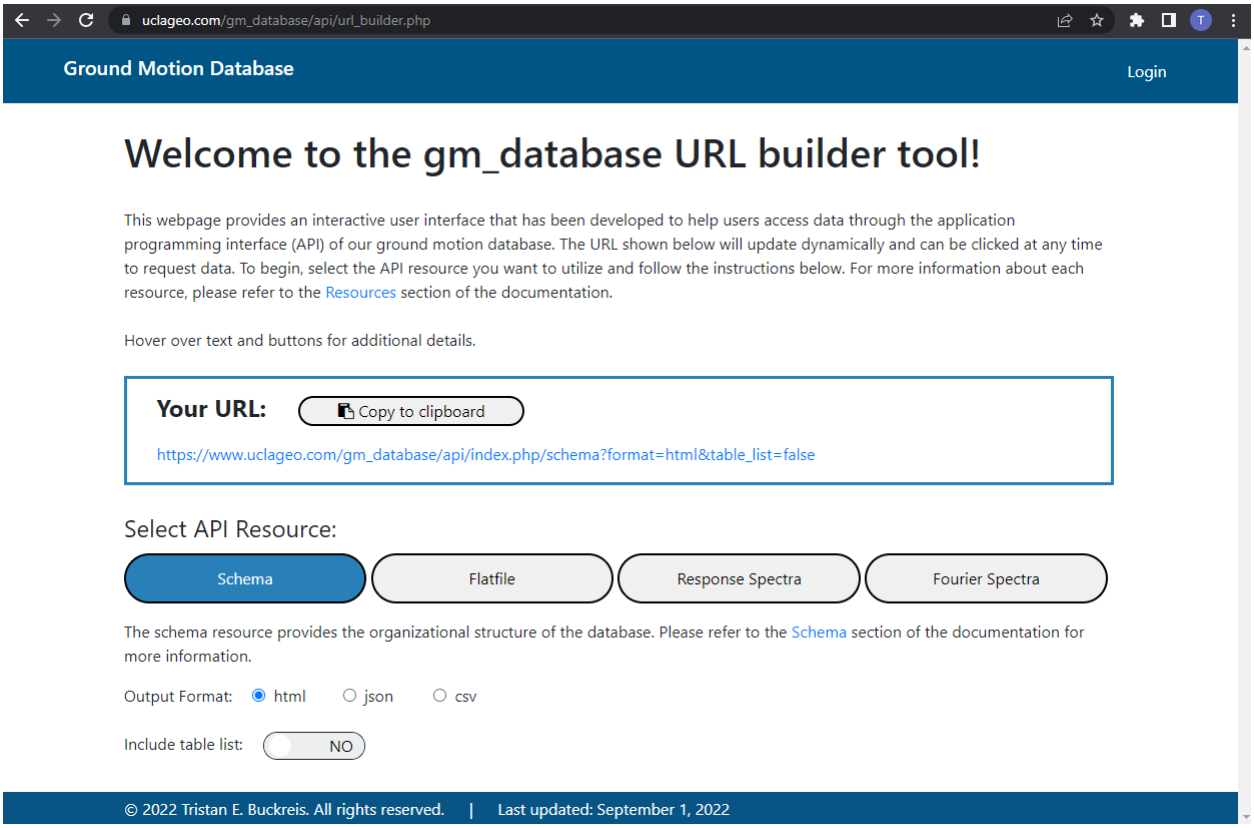


Figure 2.21: Screenshot of API URL Builder Tool schema resource options.

The *flatfile* resource is the primary mechanism to access data from the GMDB, which accordingly requires the most flexibility to accommodate tailored requests. The online tool utilizes a five step process to simplify the *query_string* construction as shown in Figure 2.22 and outlined below:

1. Users first select which tables they want to query, as shown in Figure 2.23. The tool does not currently allow users to specify desired fields, however those may be added externally (e.g., if users are only interested in magnitude and hypocentral depth data in the *event* table, they can add "&fields=magnitude,hypo_depth" to the end of the *query_string*). With respect to ground motion attributes, the tool is set up to return

intensity measures but does not allow queries from the *time_series_data* table, which could produce unmanageably large data files returned by API. If the user does not specify any tables, the API will return a default set of tables and fields.

2. If the request requires screening based on earthquake related metadata (origin time, location, magnitude, type, and/or mechanism), the user can specify these constraints, as shown in Figure 2.24. This form includes an interactive map where the user can preview event-specific metadata and dynamically draw a coordinate bounding box.
3. If the request requires screening based on station-site related metadata (location, V_{S30}, $z_{1.0}$, $z_{2.5}$, or geomorphic province classification), the user can specify these constraints. An interactive map of station locations is included within this form, with functionality similar to that described in Step (2).
4. If the request requires screening based on information about the ground motion records themselves (distance, *PGA*, spectral damping, component, and/or collection information), the user can specify these attributes during this step.
5. Lastly, the user can customize the returned data structure (format, JSON-string parsing, sorting, limit, offset, and/or filling NULL entries), as shown in Figure 2.25.

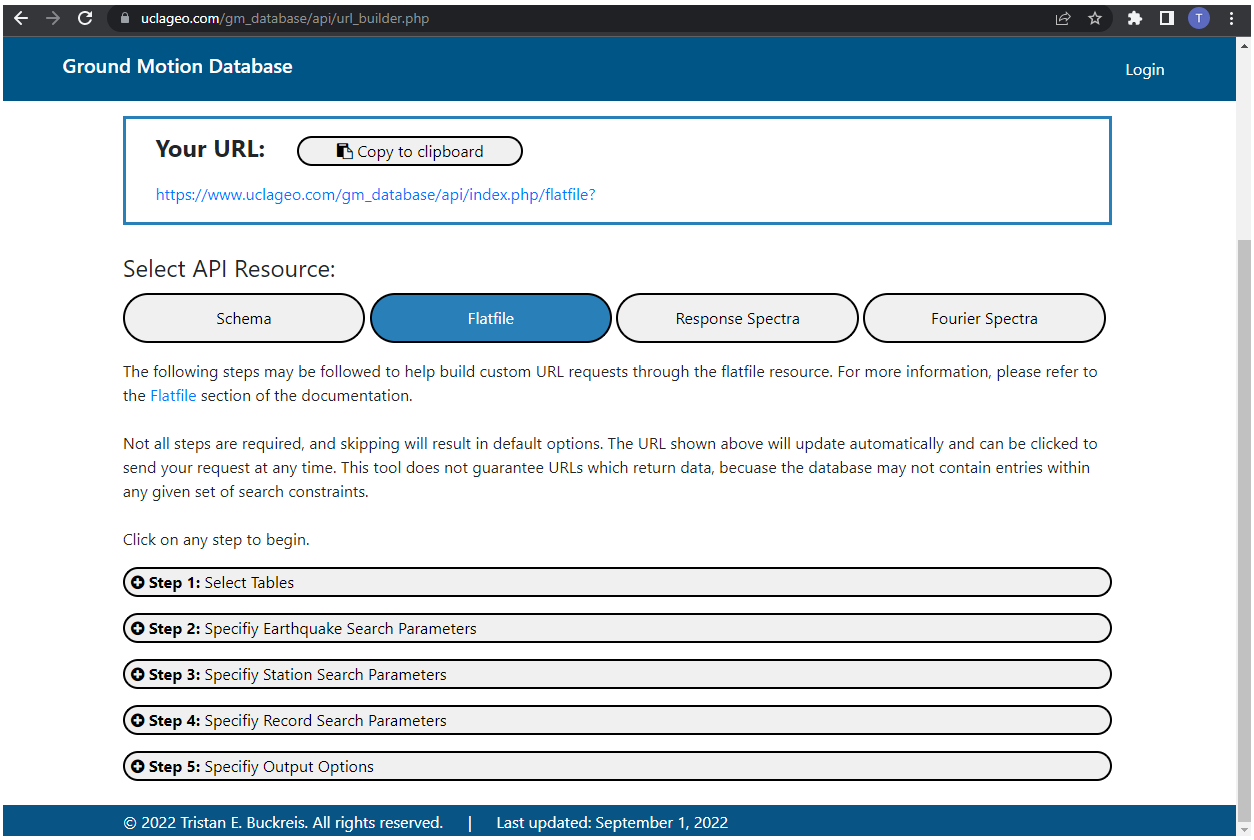


Figure 2.22: Screenshot of API URL Builder Tool flatfile resource options.

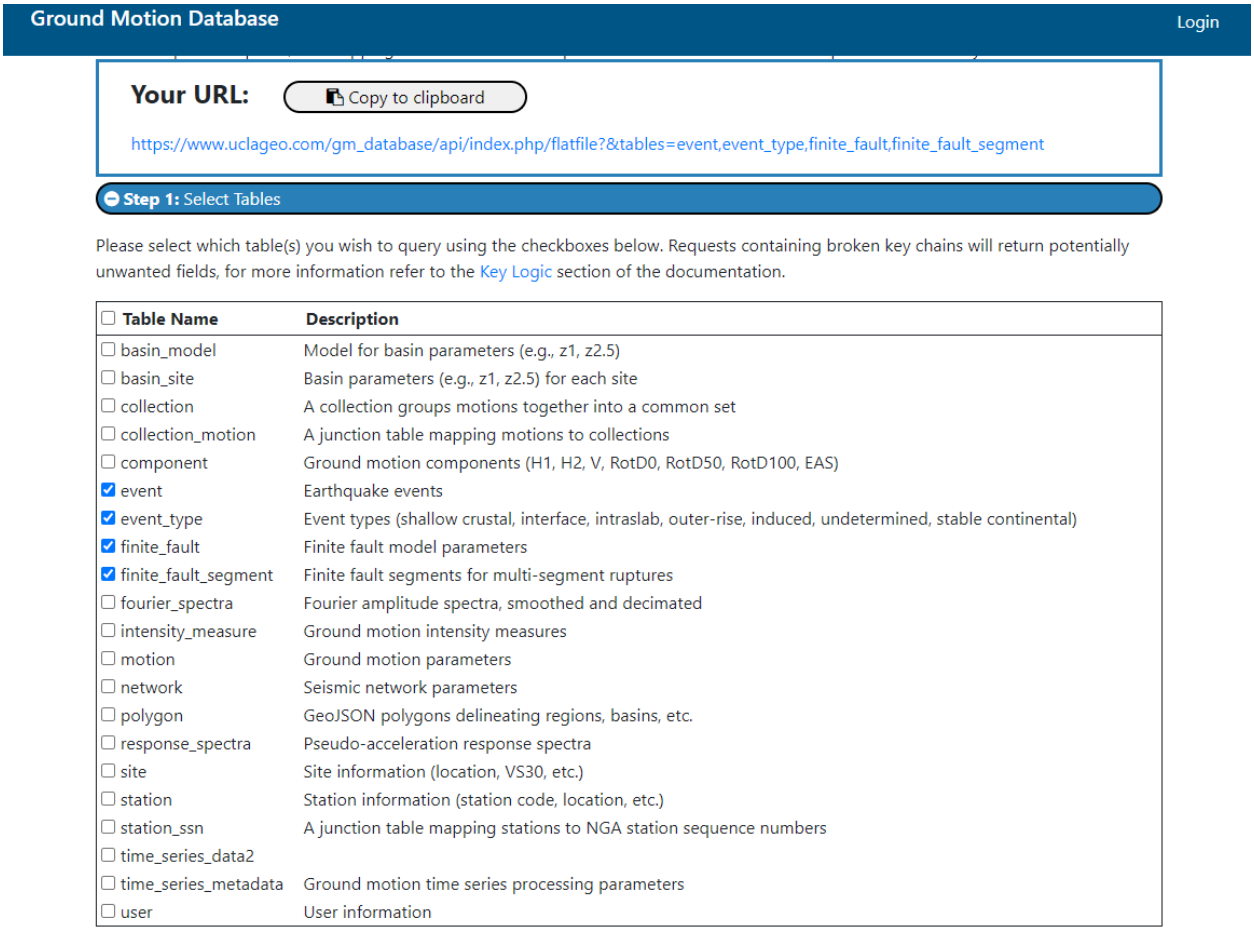


Figure 2.23: Screenshot of API URL Builder Tool flatfile resource step 1 options.

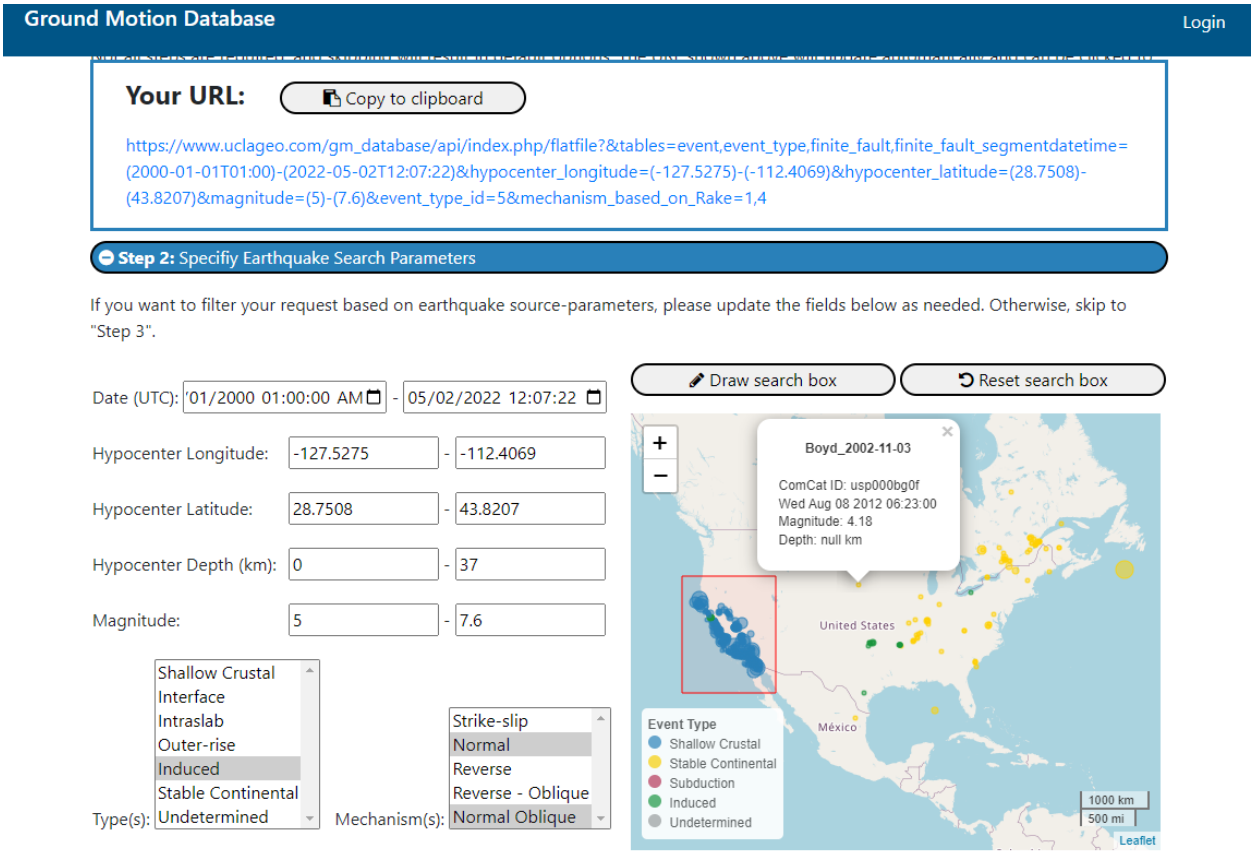


Figure 2.24: Screenshot of API URL Builder Tool flatfile resource step 2 options.

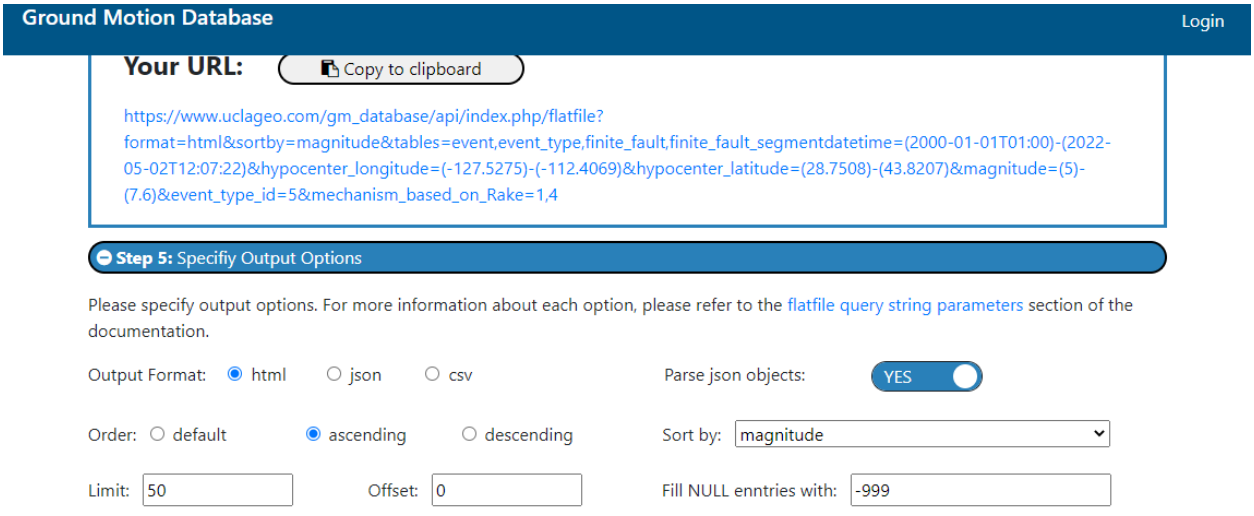


Figure 2.25: Screenshot of API URL Builder Tool flatfile resource step 5 options.

As illustrated in Figures 2.22 through 2.25, the URL is updated dynamically as the forms are populated. Individual fields within each form are automatically populated with the parametric ranges of data currently stored in the GMDB, and include sanity checks to ensure that the constraints are valid (e.g., longitude values must be between -180^{o} and 180^{o}; minimum bounds are less than maximum bounds; etc.). Users are not required to complete each step; if any steps are skipped the API will use default parameters.

2.5. Summary and Conclusions

The GMDB is a resource to provide public access to query up-to-date available ground motion data and related metadata from earthquakes around the world. Currently the GMDB only stores the unified California ground motion data assembled by NGA-West2, Ahdi et al. (2019), Wang (2020), and Buckreis et al. (2022). As of this writing, multiple efforts are ongoing to extend the reach of the GMDB, including (1) integration of NGA-East ground motion data (Goulet et al. 2021); (2) adding recently processed data from the Community Seismic Network in southern California (Clayton et al. 2011) and (3) updating of the NGA-East data for events since 2012 (Ramos-Sepulveda et al. 2022). The schema was developed with these expansions in mind, however organizational additions and/or alterations are anticipated as the database grows.

A workflow to aid authenticated users in uploading ground motion data and assigning metadata has been implemented in a Jupyter Notebook. Back-end routines are required to assign site related metadata from maps and proxy-based models and to calculate IMs for combined horizontal components, however the intention is to incorporate these routines into the Notebook in the future. Tools in the form of an API and online URL Builder have been developed to help users access the data.

CHAPTER 3

Site Characterization in the Sacramento-San Joaquin Delta

3.1. Introduction

This chapter presents the collection, synthesis, and analysis of geotechnical site characterization data in the Sacramento-San Joaquin Delta region (hereafter *Delta*). The motivation for undertaking this work was to provide the necessary data to reasonably estimate site parameters at seismic stations in the Delta which have contributed data to the GMDB (discussed in Chapter 2) and are used to investigate regional site response (presented in Chapter 5). An additional objective of the work presented here is to provide a model for estimation of site parameters in the Delta when site-specific velocity profiles are not available, which is useful for forward application of ground motion models and probabilistic seismic hazard analysis.

Following this introduction, different types of geotechnical data are reviewed which are used to characterize site conditions. Afterwards, I present the data assembled from publically available resources. Given the quantity of critical infrastructure distributed across the Delta, it is no surprise that a significant amount of data have been collected. Nevertheless, there existed a dearth in information at locations of seismic stations. To address this deficiency, I was actively involved with several field explorations to collect vital geotechnical data. A summary of these explorations is presented within this chapter. Lastly, I discuss in detail the analysis of peat

thickness (t_p), shear wave velocity (V_S), and horizontal-to-vertical (H/V) data to facilitate the assignment of site parameters.

3.2. Site Characterization Data

Site characterization is defined as "the process of developing an understanding of the geologic, hydrologic and engineering properties at the site including the soil, rock, along with groundwater and in many cases, man-modified conditions" (Benson and Yuhr 2016). Accordingly, site characterization data is anything that increases the subsurface or surficial understanding of the physical site. These data can be measured or observed *in situ* (at the site), and include information about stratigraphy, penetration resistance, and shallow-subsoil seismic properties (i.e., V_S and H/V).

Published maps are used to assign information related to surficial geology, elevation, topographic slope, and terrain class. Surficial geology refers to the geologic materials at the ground surface, usually unconsolidated Quaternary soils, and is represented by qualitative symbols (i.e., geological units). Elevation, topographic slope, and terrain class are metrics which describe the ground surface, and are obtained from digital elevation models (DEMs); for the present applications I use the NASA Shuttle Radar Topographic Mission (SRTM) DEM, which is at a horizontal resolution of 3 arc-sec (CGIAR-CSI 2006). Slope and terrain class are computed from DEMs through the use of different algorithms that describe the percent change in elevation over a certain distance and the general physical attributes, respectively. I perform these calculations using QGIS. The assigned terrain classes are based on Iwahashi and Pyke (2007) and are evaluated from the DEM using the algorithm presented in Riley et al. (1999).

The most direct way to measure stratigraphic information is by drilling a bore hole and recovering *in situ* samples. Cuttings from the boring process and recovered samples are inspected by the field engineer to assign soil-type classifications (e.g., Unified Soil Classification System; Howard 1986). The depth to the ground water table (GWT) is also recorded. Stratigraphic information can also be inferred from cone penetrometer testing (CPT) data. During a CPT, a "cone" is advanced into the ground and the tip resistance (q_c), sleeve friction (f_s), and excess pore water pressure (u_2; if equipped) are recorded (ASTM D3441). Tip resistance refers to the force per unit area (i.e., stress) required to push the tip of the cone through the soil, and f_s measures the stress required to overcome the friction between the soil and the sides of the instrument (i.e., the *sleeve*). Relationships are used to compute soil behavior type (SBT) from CPT data (e.g., Robertson 2016), from which stratigraphic information may be inferred. Location of the GWT is interpreted from u_2 measurements.

Penetration resistance is measured directly during a CPT (i.e., q_c), however it can also be measured in a boring using the standard penetration test (SPT). During a SPT, the field engineer records the number of hammer blows (*N*) required to drive a standardized sampler over a depth interval of 0.45 m (ASTM D1586). Given the circumstances, *N* represents a discrete measurement of penetration resistance, when compared to the relatively continuous q_c measurements achieved through CPT. Many correlations exist which relate measured penetration resistance to soil mechanical properties used by engineers (i.e., strength, stiffness, compressibility, etc.).

Shear wave velocity (V_S), which relates to the shear modulus of the soil, is a common property used in soil dynamics, and can be measured using either invasive or non-invasive geophysical methods. Invasive methods include downhole (ASTM D7400) and crosshole testing

(ASTM D4428), P-S suspension logging (ASTM D5753), and seismic CPT (sCPT; ASTM D7400). Non-invasive methods measure frequency-dependent surface waves (Rayleigh or Love) to develop a dispersion curve (plot of surface wave velocity vs. frequency), which can then be inverted to estimate a representative V_S profile (Menke 1989). These methods include the spectral analysis of surface waves (SASW; Stoke et al. 1994), multi-channel analysis of surface waves (MASW; Park et al. 1999), and passive-source microtremor array measurements (MAM; Kanai et al. 1954; Horike 1985; Okada et al. 1990). S-wave seismic refraction is a unique type of non-invasive method which measures body waves that are reflected (or refracted) off layer interfaces with impedance contrasts in the subsurface (ASTM D5777).

H/V tests measure vibrations at a site using a three-component seismometer or geophones (two-orthogonal horizontal and one vertical). H/V data are used to construct mean horizontal-over-vertical spectral ratio curves (HVSR; discussed in Section 3.3.3), which relate information about resonant frequencies and site amplification. HVSR can be developed from earthquake recordings (eHVSR) or from microtremor ambient vibrations recorded using permanent (mHVSR-P) or temporarily deployed sensors (mHVSR-T).

3.2.1. Synthetization of Available Geotechnical Data

The primary motivation of the work presented in this chapter is to collect site-specific geotechnical site characterization data which can aid in the assignment (or estimation) of the site parameters discussed in Section 2.3.1(b) of this dissertation. Accordingly, I performed an exhaustive literature search for the types of data described in the previous section with emphasis on the Delta region. The findings from this effort are presented below as subsections related to geologic and topographic data, stratigraphic profiles, V_S profiles, and H/V data. The V_S and H/V data, including

co-located supporting data such as boring or CPT logs, have been entered in into the shear wave velocity profile database (www.vspdb.org; Kwak et al. 2021) and as such are readily accessible for the broader community.

(a) Geologic and Topographic Data

Published resources are commonly used to assign geologic and topographic site parameters, as discussed in Section 2.3.1(b). Elevation, topographic slope, and terrain class for Delta sites are assigned using the methods described in Chapter 2. Previously I used Wills et al. (2015) to assign surficial geologic units to all sites in California, including those in the Delta. Wills et al. (2015) provides a simplified representation of regional geologic maps and combines geologic units into 15 distinct groups that are expected to have similar V_{S30} distributions (i.e., mean and standard deviation). The advantage of this approach is that geology-based V_{S30}-proxy models can be used to estimate V_{S30} at sites with no V_S measurements at a state-wide scale. The fundamental drawback of Wills et al. (2015) is that it does not distinguish between nuanced geologic features, which may prove to be significant in a smaller regional study such as the Delta.

The young (Holocene) alluvial deposits which occupy much of the Delta are classified within the "Qal1" geological unit of Wills et al. (2015) – Quaternary-Holocene alluvium in areas of very low slopes. Furthermore, much of the Delta is underlain by peaty-organic deposits which are not aptly represented in Wills et al. (2015). Peat is a unique material because it possesses problematic characteristics (e.g., low shear strength, high compressibility, high water content, low V_S, etc.), but also because it has the ability to vary widely spatially and temporally (Kazemian et al. 2011). Given the significant potential peat has to influence site response, it is important to

identify zones within the Delta where it occurs, therefore a detailed geologic map of the Delta is needed.

Bay Delta Live (2021) provides a detailed geologic map applicable to much of the Delta region, and is presented in Figure 3.1. Much of the Delta can be described by four different types of alluvium or “peat and muck”, however there is also a large older eolian deposit along the southern bank of the San Joaquin River in the western portion of the Delta. Dredge soils, defined as an unconsolidated mixture of sediments produced from dredging and dumping activities, are encountered in the Sacramento River and near the Port of Stockton. Stiffer more-competent units (e.g., Montezuma Formation and Tertiary and Cretaceous Bedrock) are not encountered near the ground surface within the Delta legal boundaries (DWR 2018).

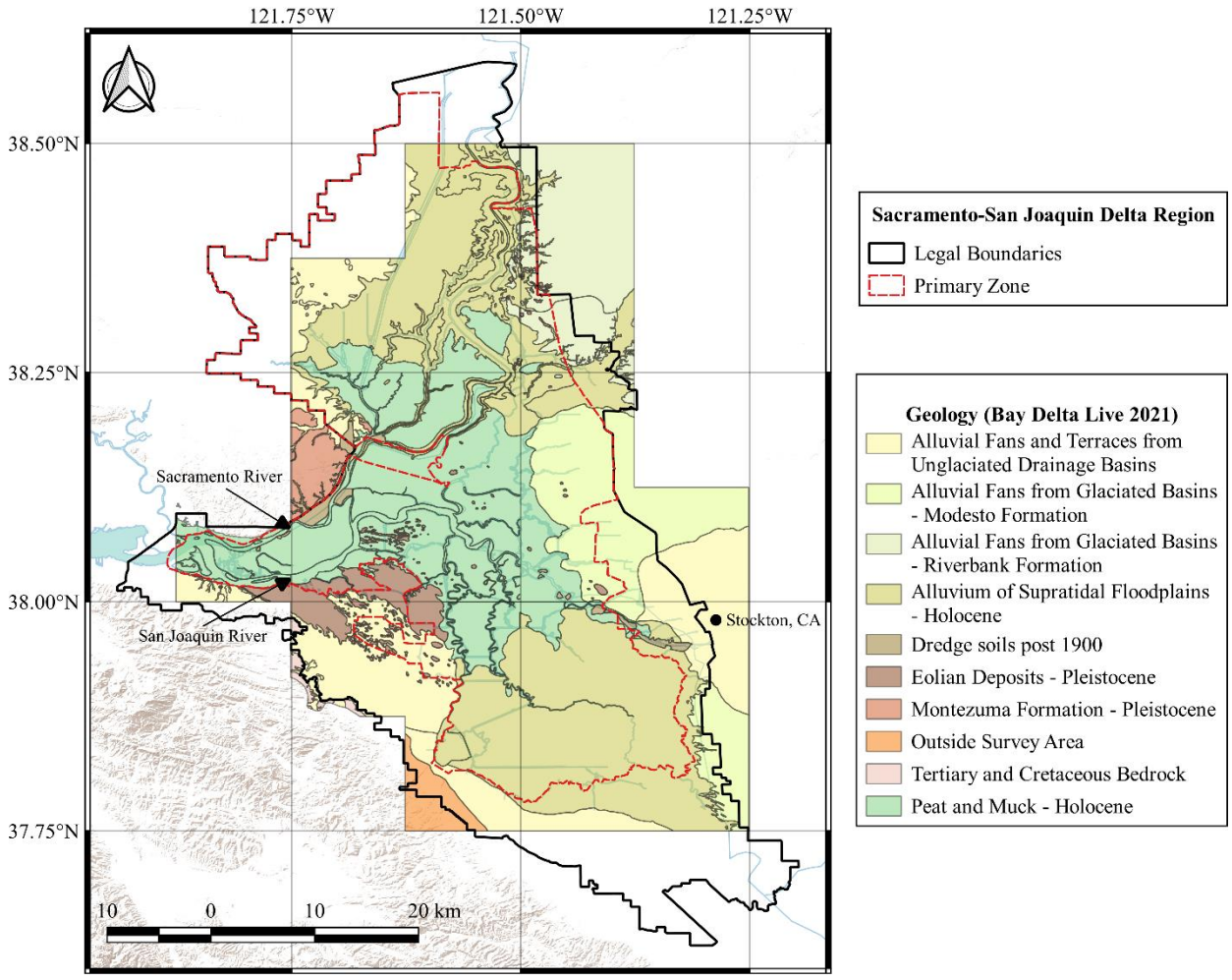


Figure 3.1: Detailed geologic map of the Delta region (source: Bay Delta Live 2021).

(b) Stratigraphic Profiles

The Department of Water Resources (DWR) has collected a considerable amount of data within the Delta region through field investigations for several projects, including the Peripheral Canal (Water Education Foundation 2010) and the Delta Habitat Conservation and Conveyance Program (DWR 2022). The alignments for each of these dictate where most of the data has been collected; Figure 3.2 shows the locations of 10,137 borings and 4,389 CPTs along the project alignments.

Additional data is available through the community shear wave velocity profile database (VSPDB; Ahdi et al. 2018) for select locations from Caltrans for sites near the Antioch Bridge and from various research studies. It would take a monumental effort to digitize the large number of DWR CPTs and boring logs, therefore these data have not been organized and uploaded to the VSPDB except in cases where they are co-located with V_S profiles. The borehole and CPT data can be accessed through the California Natural Resources Open Data web portal (https://data.cnra.ca.gov/).

The available stratigraphic data indicate that peat is present in 89% of borings and CPTs performed for several specific DWR projects in much of the central Delta (i.e., Primary Zone), but its thickness is typically modest. Peat makes up about 18% of soils encountered within depths of 0-10 m across much of the Delta with notable exceptions on Sherman, Twitchell, and several other neighboring islands where deposits can be up to 15 m thick. Underlying the peat are interbedded layers of clays, silts, and sands. In the interiors of Delta islands, ground water is approximately 1.5 m below the ground surface, and is controlled by pumping related to agricultural activities, otherwise ground water would seep out the surface.

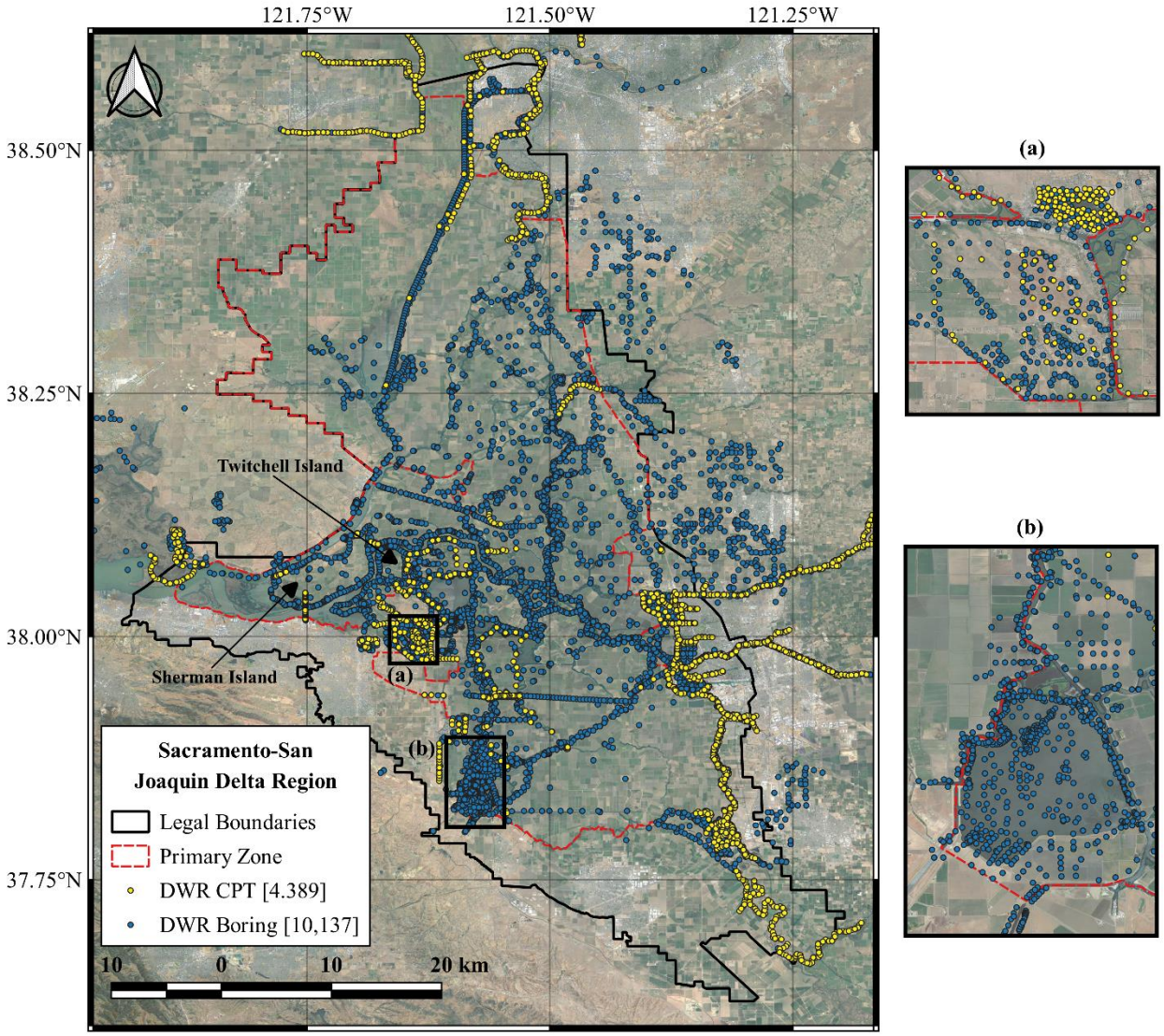


Figure 3.2: Locations of measured stratigraphy via CPT or boring. Inset maps present areas with a high density of measurements: (a) near Bethel Island and (b) Clifton Court. Delta legal boundary and primary zone delineation is from DWR (2018).

(c) Shear Wave Velocity Profiles

At the onset of my work, the VSPDB included only 27 V_S profiles in and around the Delta, most of which located near the edge of the legal boundary (i.e., in environments where peat is not

typically encountered). I performed an exhaustive search for data from V_S profiles published in literature or made available through publically available means (e.g., Caltran's digital archive of geotechnical data – GeoDOG; https://geodog.dot.ca.gov/). This effort produced 17 V_S profiles shared by Mitchell Craig of California State University, East Bay, one profile from Caltrans (Bridge 28-0009), and two profiles shared by Albert Kottke of Pacific Gas and Electric (PG&E). DWR, who sponsored much of the work presented throughout this dissertation, shared data from a total of 122 new V_S profiles (DWR 2012, 2013a-b, 2014a-e, 2015a-b, 2016, 2017; Mike Driller *personal communication* 2018; Mark Pagenkopp *personal communication* 2019). Lastly, I was actively involved with measuring V_S data at six additional sites via the sCPT method, as discussed in Section 3.2.2(a). All newly assembled data from 155 V_S profiles were uploaded to the VSPDB, and are accessible at https://vspdb.org/ (Kwak et al. 2021). The locations of these profiles are presented in Figure 3.3.

The V_S data were generally obtained from downhole testing during sCPT soundings, suspension logging in boreholes, or surface wave testing (SASW or MASW). The majority of profiles do not extend deeper than approximately 30 m with velocities at this horizon ranging from 300 to 400 m/s. Measurements from four suspension logs performed at depths around 90 m encountered deep layers with V_S between 400 to 600 m/s, these results indicate that much of the Delta is composed of a relatively thick layering of soils overlying more competent materials at great depths. The typical V_S for consolidated peat below Delta levees ranges between 40 to 110 m/s, and 25 to 80 m/s for unconsolidated free-field deposits. Even underlying nonorganic deposits (e.g., soft silty clay to stiff clay) can be exceptionally low ranging between 70 to 190 m/s.

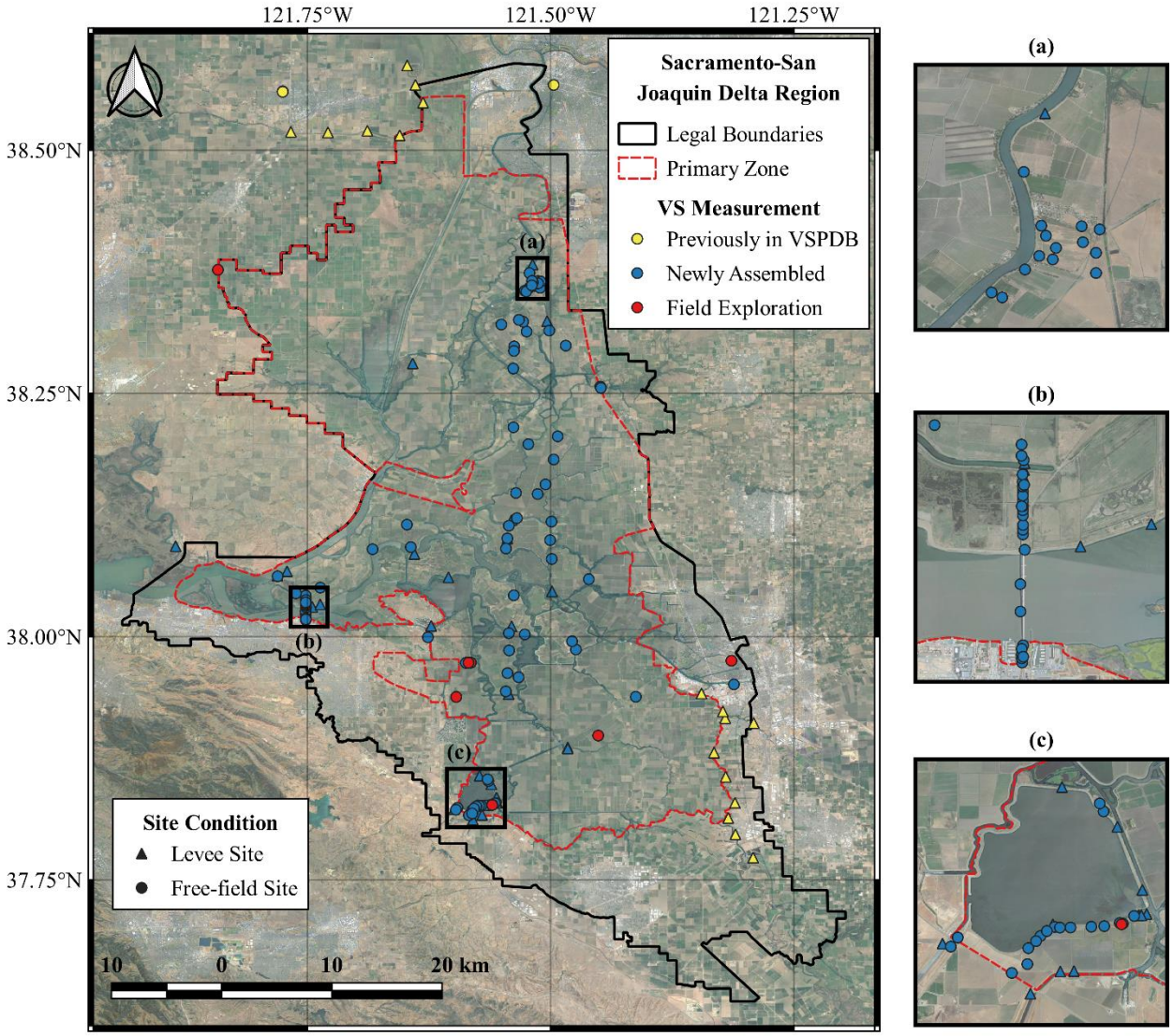


Figure 3.3: Locations of measured shear wave velocity (V_S) profiles, the corresponding source of data (VSPDB, newly assembled, or measured as part of field explorations related to the work presented in this dissertation), and type of site condition (levee or free-field). Inset maps present areas with a high density of measurements: (a) Hood, CA, (b) Antioch Bridge, and (c) Clifton Court.

(d) H/V Measurements

H/V data has not historically been widely used by practitioners or government agencies in the United States, therefore there are very few publically-available measurements in the Delta. One firm at the forefront of advocating the collection and use of H/V data is Geometrics, who have gathered data worldwide (https://seisimager.com/). Their collection efforts include measurements at four sites within the Delta, which are stored in the VSPDB as processed mHVSR curves with associated metadata. Given the relevance of information provided from HVSR to site response studies, I deployed temporary seismometers at 41 locations to measure H/V which facilitates the development of mHVSR-T curves (further discussion provided in Sections 3.2.2(b) and 3.3.3). Ambient ground vibrations (i.e., noise) recorded by six permanent seismometers were used where field testing was infeasible (i.e., mHVSR-P). Locations of 52 H/V measurements with corresponding mHVSR curves are shown in Figure 3.4. Processed mHVSR curves for all sites are accessible through the VSPDB at https://vspdb.org/ (Kwak et al. 2021).

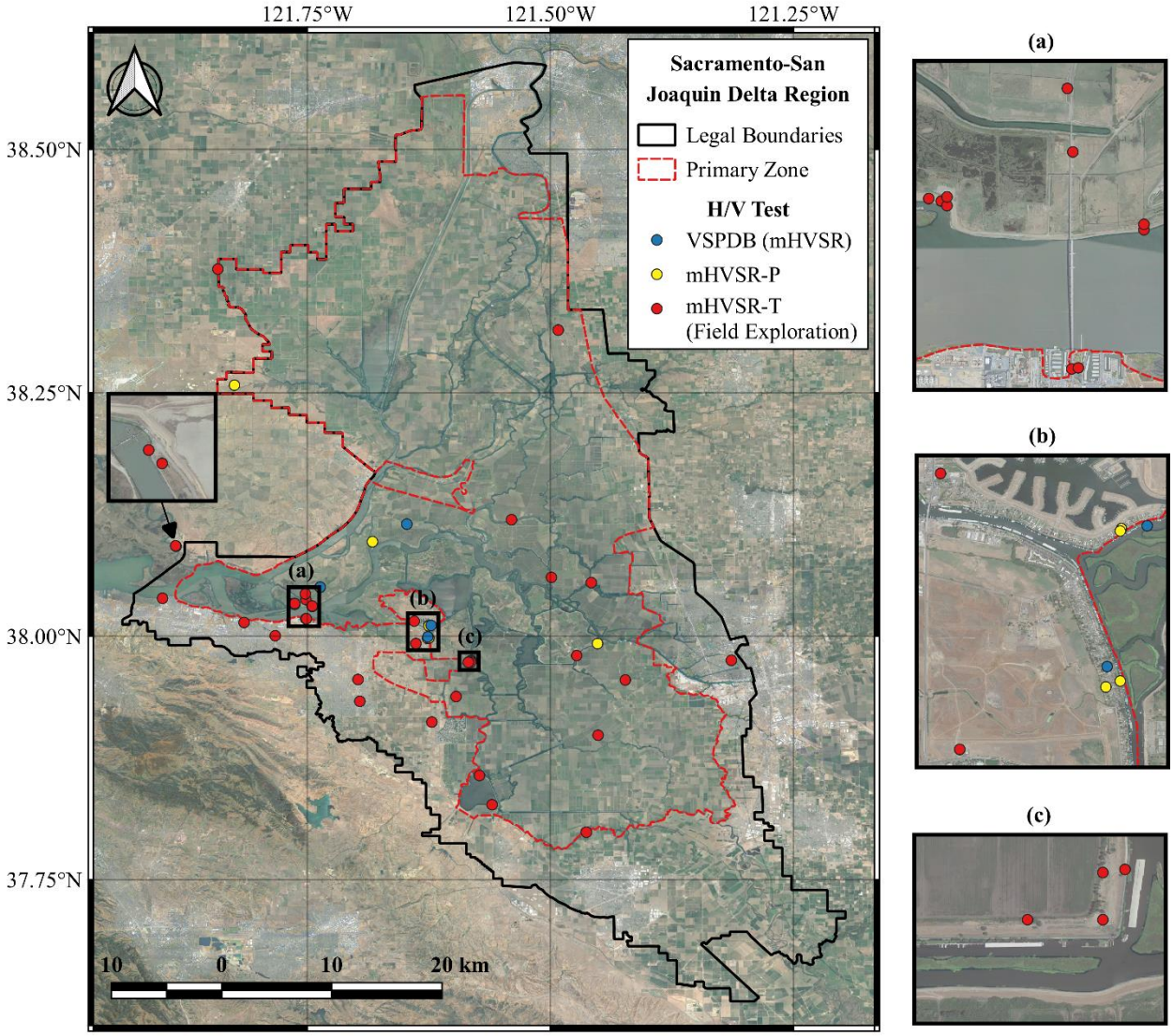


Figure 3.4: Locations of H/V data and the corresponding source of data (VSPDB, mHVSR-P, or measured as part of field explorations related to the work presented in this dissertation – mHVSR-T). Inset maps present areas with a high density of measurements: (a) Antioch Bridge, (b) near Bethel Island, and (c) Holland Tract.

3.2.2. Field Explorations

The motivation for assembling geotechnical site characterization data in the Delta stems from the need to develop site parameters at seismic stations for the purpose of modeling site response. A total of 54 stations have recorded earthquake shaking in the Delta since 1969. Figure 3.5 presents the locations of these "Delta stations" and whether or not V_S and/or H/V data were available following the data synthetization effort discussed in Section 3.2.1. A total of 8 stations (or 15%) had nearby V_S profiles and only 2 (or 4%) had an available mHVSR. Many of the Delta stations have significant numbers of earthquake recordings, however their value to the seismological and earthquake engineering communities would be significantly enhanced with additional geotechnical work. As a result, three field-testing programs were organized and executed in coordination with DWR during August-September 2019, December 2020, and March 2022. The aims of these investigations were to measure V_S profiles (from which V_{S30} could be computed), stratigraphic logs (from which t_p can be obtained), and H/V data (from which HVSR curves can be derived).

Delta stations were assigned one of three priority levels conditioned on the number of available earthquake recordings and available site characterization data. Sites with ≥ 5 recordings that were missing t_p, V_S, and/or H/V data were assigned the highest priority level. Sites with no missing data or decommissioned stations with few recordings were assigned the lowest priority. Active stations with few recordings and missing geotechnical data were given an intermediate level of priority, because they have the potential to record additional earthquakes in the future. Table 3.1 summarizes the type of geotechnical data at each site prior to the field explorations and

associated priority level for Delta stations. The following subsections describe the aspects of the field explorations related to collecting stratigraphic, V_S, and H/V data.

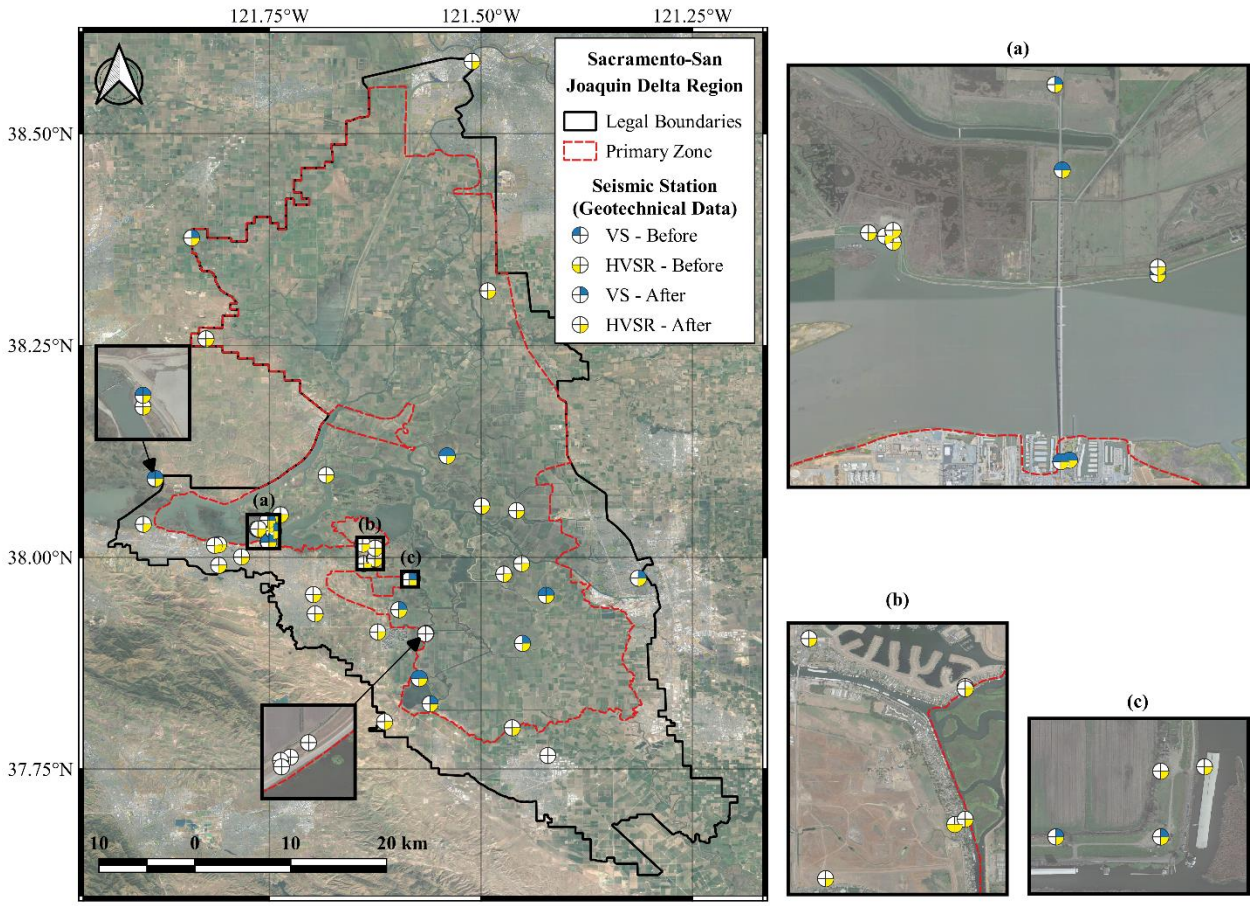


Figure 3.5: Locations of 54 "Delta seismic stations" and available site characterization data "before" and "after" field explorations. Inset maps present areas with many instruments: (a) Antioch Bridge, (b) near Bethel Island, and (c) Holland Tract.

Table 3.1: Summary of missing geotechnical site characterization data and priority level.

Delta Station	Latitude (°)	Longitude (°)	mHVSR	Vs	Stratigraphy	Priority Level
BK_JEPS	38.2579	-121.8252	No	No	No	1
BK_TWIT	38.0971	-121.6832	No	No	No	1
CE_57195	37.9753	-121.3140	No	No	No	1
CE_57531	37.9332	-121.6956	No	No	No	1
CE_57534	37.9119	-121.6219	No	No	No	1
CE_67070	38.0150	-121.8130	No	No	No	3
CE_67215	37.9925	-121.6384	No	No	No	1
CE_67265	38.0377	-121.7515	No	Yes	Yes	2
CE_67266	38.0179	-121.7516	No	Yes	Yes	1
CE_67523	38.0183	-121.7509	No	Yes	Yes	1
CE_67533	38.0155	-121.6396	No	No	No	1
CE_67557	38.0141	-121.8148	No	No	No	2
CE_67587	38.0390	-121.8986	No	No	No	1
CE_67615	38.0005	-121.7830	No	No	No	1
CE_67910	38.0435	-121.7520	No	No	No	2
NC_C057	37.9555	-121.6979	No	No	Yes	1
NP_DIX	38.3771	-121.8424	No	No	No	1
NP_EMR	38.0605	-121.4993	No	No	Yes	1
NP_KIR	38.0551	-121.4582	No	No	Yes	1
NP_LVA1	38.0334	-121.7647	No	No	No	2
NP_LVA2	38.0332	-121.7635	No	No	No	2
NP_LVA3	38.0328	-121.7631	No	No	No	1
NP_LVA4	38.0335	-121.7634	No	No	No	1
NP_LVB1	37.9099	-121.5654	No	No	No	2
NP_LVB2	37.9098	-121.5656	No	No	No	2
NP_LVB3	37.9102	-121.5649	No	No	No	2
NP_LVB4	37.9097	-121.5656	No	No	No	2
NP_MCD	37.9802	-121.4735	No	No	No	1
NP_PLA	37.7987	-121.4632	No	No	No	1
NP_SIA	38.0503	-121.7367	Yes	No	No	1
WR_AWTP	37.9906	-121.8100	No	No	No	2
WR_CKR	38.3145	-121.4920	No	No	Yes	1
WR_CLFN	37.8570	-121.5730	No	Yes	Yes	1
WR_CLFS	37.8273	-121.5604	No	No	No	1
WR_HOLT	37.9551	-121.4227	No	No	No	1
WR_MOFF	38.0925	-121.8849	No	No	Yes	2
WR_MONN	38.0929	-121.8855	No	Yes	Yes	1
WR_SHER	38.0306	-121.7450	No	Yes	Yes	1

Delta Station	Latitude (º)	Longitude (º)	mHVSR	Vs	Stratigraphy	Priority Level
WR_SIFF	38.0311	-121.7450	No	No	Yes	1
WR_STNI	38.1197	-121.5401	No	Yes	Yes	1
YU_CEC	37.9381	-121.5968	No	No	No	1
YU_HMT	37.9749	-121.5813	No	No	No	1
YU_HOL1	37.9748	-121.5818	No	No	No	1
YU_HOL2	37.9733	-121.5823	No	No	Yes	1
YU_HOL3	37.9733	-121.5844	No	No	No	1
YU_SMB	37.9977	-121.6262	Yes	No	No	3
YU_SMT	37.9982	-121.6252	No	No	No	1
YU_SRB	38.0109	-121.6249	No	No	No	3
YU_SRT	38.0107	-121.6246	No	No	No	1
YU_STF	37.9925	-121.4516	No	No	No	3
YU_WHR	37.8984	-121.4508	No	No	No	1

(a) Seismic Cone Penetrometer (sCPT) Testing

sCPT is a quick, reliable, and relatively cheap method used to measure V_S *in situ* using a surface source and downhole (within-CPT) geophone, and with the added benefit of also collecting information related to penetration resistance and stratigraphic layering (i.e., t_p). Original plans included performing sCPTs at all intermediate and high priority level sites, however obtaining permission from land owners proved to be difficult. Ultimately, we received permission to perform the tests at a total of seven sites, the locations of which are represented by red markers in Figure 3.3. Gregg Drilling, LLC was contracted to perform the measurements within 5-50 m of each seismic station to a terminating depth of 35 m or until refusal is encountered (whichever came first). Shear wave travel time data was collected every 1.5 m using an electronic wave generator attached to the CPT rig coupled to the ground surface, as shown in Figure 3.6(a).

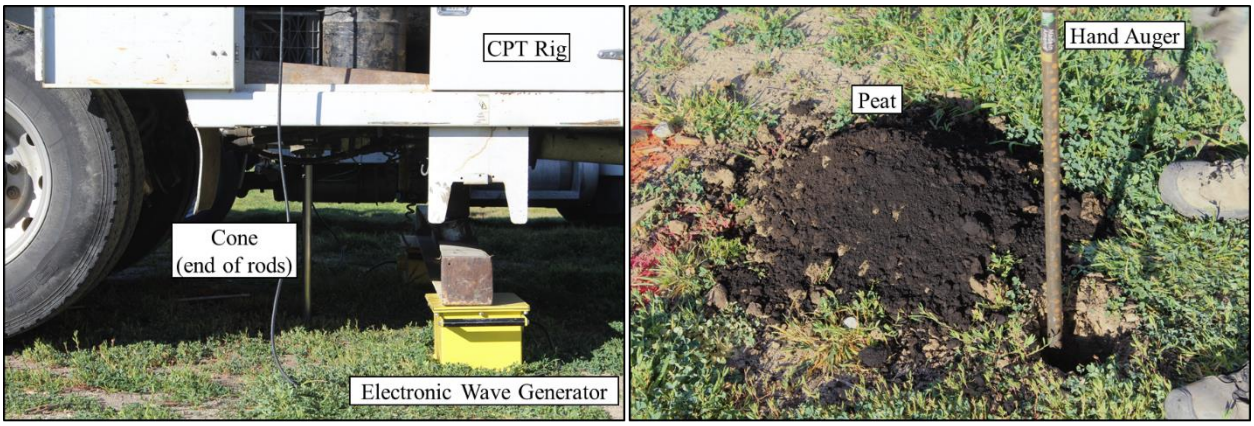


Figure 3.6: (a) Photo of CPT rig and electronic wave generator; (b) Peat cuttings excavated when drilling the pilot hole with a hand auger.

The objectives of these tests included obtaining a representative V_S profile which could be used to calculate V_{S30} and to identify the presence and thickness of peaty-organic layers. Peat is usually encountered near the ground surface where it is difficult to accurately measure penetration resistance (from which SBT is calculated to infer soil-type). A pilot hole was usually drilled to a depth of 1-2 m to ensure that no utilities are severed during the test and to assist in aligning the cone vertically. An added benefit of this step was that we could inspect the cuttings directly to identify the soil-type and relative thicknesses of shallow layers, as shown in Figure 3.6(b). Methods presented in Robertson (2016) were used to correct for overburden effects and to normalize the as-recorded q_c and f_s, from which SBT are calculated. Layers with SBT values of 2 are interpreted to be organic soil (i.e., peat). Peat was encountered at three of the seven sites investigated with low-to-moderate thicknesses (<1 m – 2 m).

All collected data (q_c, f_s, u_2, and V_S) are available through the VSPDB at https://vspdb.org/ (Kwak et al. 2021). Interpretation and analysis of the V_S results is presented in Section 3.3.2.

(b) Microtremor Horizontal-Over-Vertical Measurements

H/V tests are inexpensive non-invasive geophysical measurements that do not require the level of permitting or time that invasive site characterization methods need (e.g., CPT and borings). HVSR curves can be derived from permanent broadband seismometers (i.e., mHVSR-P) if they continuously stream data that is accessible through web services. In several instances, we have also deployed temporary three-component broadband seismometers where possible to develop usable HVSR curves (i.e., mHVSR-T). The purpose of these mHVSR-T campaigns was to (1) test the similarity of mHVSR from permanent and temporary sources and (2) to obtain reliable mHVSR data where broadband seismometers are not present, which is the case at 30 of the 54 stations. Jemile Erdem of the USGS loaned the necessary equipment which facilitated the measurement of H/V data at 41 locations, which are shown by red markers in Figure 3.4. Special access onto private property was granted through correspondence with DWR, Pittsburg Power Plant, Reclamation District No. 341, and four private land owners. Several tests were performed on public land near the seismic station (< 100 m).

The H/V method involves placing a three-component seismometer on the ground surface (or buried at a shallow depth – approximately 7.5 to 15 cm) over a length of time to record the natural vibrations of the earth, as shown in Figure 3.7. Each measurement was performed within practical distances to the seismic station or historic coordinates at decommissioned stations (<10 m for private sites or <100 m for tests on public lands). The total duration of time is proportional

to the lowest usable frequency of the HVSR curve [discussed in Section 3.3.3(a)]. Since lower frequencies potentially provide information about larger geologic structures (i.e., basin resonance), I recorded ambient vibrations at each site between 2.5 to 4 hours, which theoretically is sufficiently long to resolve resonant frequencies as low as 0.1 Hz. Two seismometers (Güralp 3ESP and Trillium Compact) are used as a standard for redundancy.

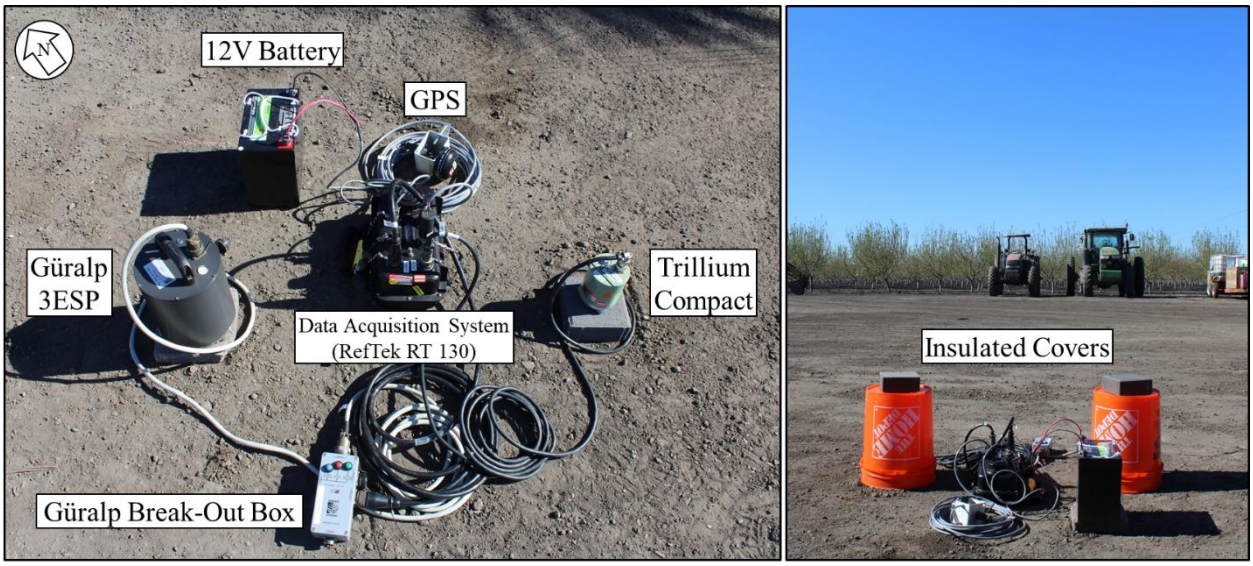


Figure 3.7: (a) Photo of typical H/V equipment setup using two different seismometers (Güralp 3ESP and Trillium Compact); (b) H/V equipment setup when recording with insulated covers to protect from the elements (primarily wind and heat).

The quality of ground coupling can significantly impact the results of the processed HVSR curve (SESAME 2004). Burying the sensor into dry somewhat competent soil provides the best results and was exercised whenever possible. In urban settings or on surfaces which were too stiff to excavate using hand tools, the sensor was placed on the ground surface. Insulated covers

(buckets with Mylar foil bubble insulation on their interior) are used to protect the sensor from wind and extreme heat, which may introduce artificial noise into the recording, thus contaminating the HVSR.

The raw data and associated metadata collected through my efforts are published as a curated dataset on DesignSafe (Buckreis et al. 2021). Processed mHVSR-T are available through the VSPDB at https://vspdb.org/ (Kwak et al. 2021). Processing procedures, interpretation, and analysis of H/V data are presented in Section 3.3.3.

3.3. Analysis of Site Characterization Data

Traditional geotechnical engineering is generally concerned with the mechanical properties of different soils and rock (i.e., strength, compressibility, plasticity, etc.). While these metrics are important, ground motion modeling utilizes representative *site parameters* to describe the average site condition for use in site response models. Early GMMs explored qualitative parameters which describe the general site condition as either "soil" or "rock". However, modern modelers have adopted the use of quantitative parameters such as V_{S30}. The most accurate site parameters are those which are derived from site-specific data, although proxy-based methods can be used to estimate them in the absence of site characterization data. This section presents analyses of the site characterization data discussed in Section 3.2 with special emphasis on assigning site parameters to the Delta stations shown in Figure 3.5.

3.3.1. Peat Thickness

Peat is an unusually soft and weak material with high small-strain damping (Wehling et al. 2003; Kishida et al. 2009). Numerical studies have shown that the presence of peat within the soil column

can have an impact on the overall site response (Kishida et al. 2008; Wang et al. 2022b). Accordingly, the thickness of peat (t_p) encountered at a site is considered a potentially useful site parameter. The utility of t_p goes beyond site response modeling to include application in consolidation, settlement, and other traditional geotechnical analyses.

Site-specific values for t_p are easily measured or interpreted from the stratigraphic data presented in Section 3.2.1(b). However, t_p may need to be estimated for locations far from any *in situ* measurement. Deverel and Leighton (2010) present a peat thickness map developed using kriging interpolation of data from Atwater (1982), which is shown in Figure 3.8. Estimates for t_p can be easily queried from the raster using geodetic coordinates.

A caveat to this map is that is represents peak thickness circa 1982. Due to subsidence, lower peat thicknesses can be anticipated at the present time where the peat is not replenished via an active depositional process. To examine the possibility of over-prediction of peat thickness using the map in Figure 3.8, I assembled an empirical t_p dataset from 401 boring logs and CPTs performed since 1990, the locations of which are shown in Figure 3.8. Evidence of bias is observed when plotting t_p against the predicted peat thickness ($\hat{t}_p$), as demonstrated in Figure 3.9(a).

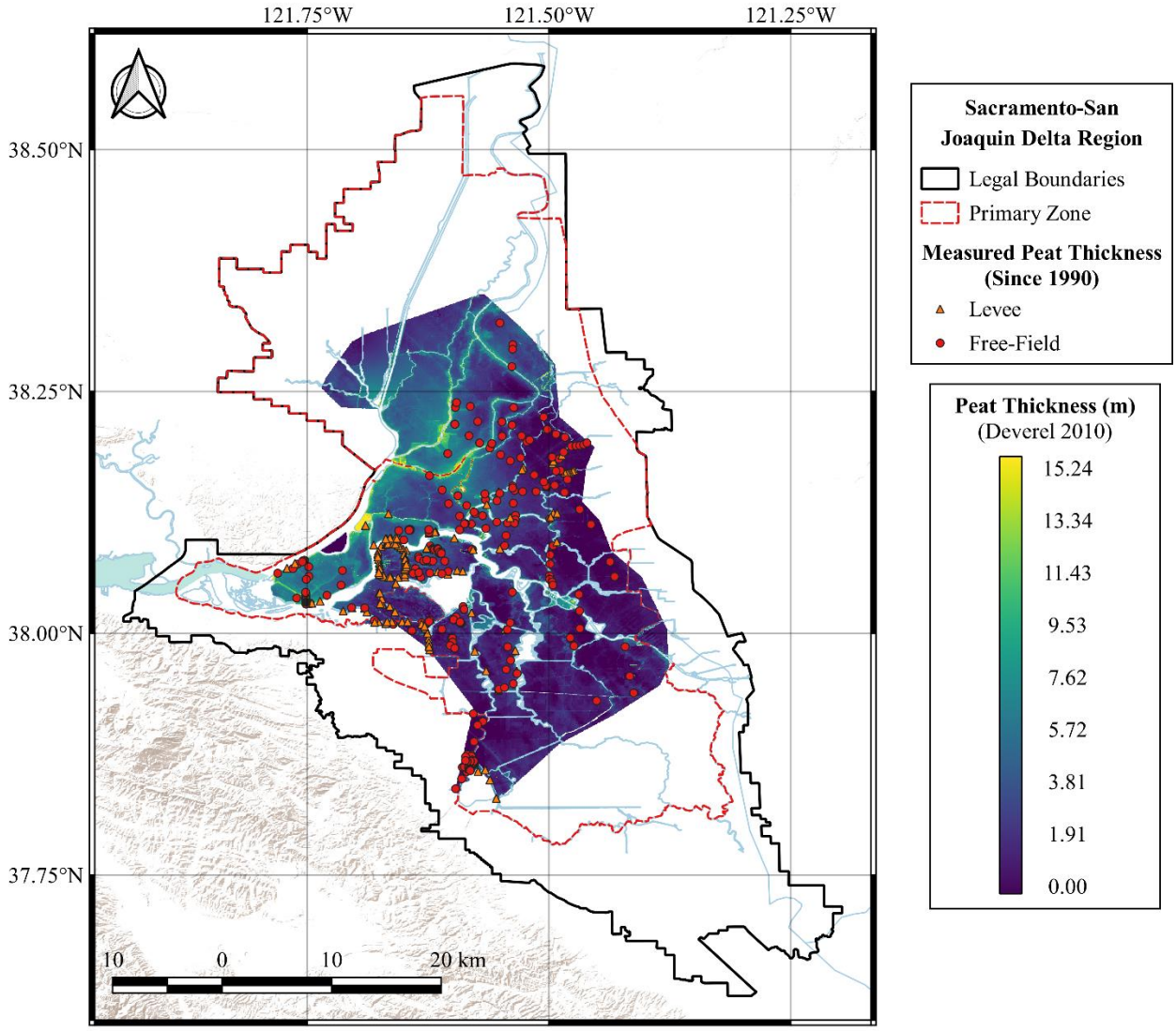


Figure 3.8: Locations of CPT and borings performed since 1990 within the area described by the Deverel and Leighton (2010) peat-thickness map; Triangle symbols represent measurements through a levee and circle symbols represent free-field measurements.

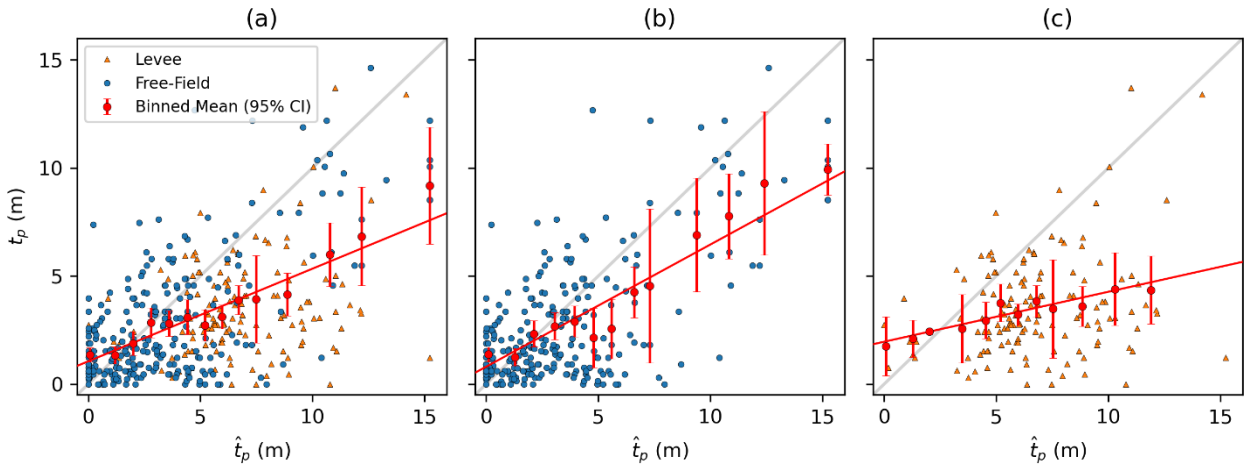


Figure 3.9: Plot comparing post-1990 measured peat thicknesses (t_p) versus estimated thickness ($\hat{t}_p$) from Deverel and Leighton (2010); (a) all sites together; (b) free-field only sites; (c) levee only sites.

The empirical t_p dataset includes free-field and through-levee measurements. The degree of over-prediction observed for the two site conditions is different, as shown in Figure 3.9(b-c). Several factors contribute to the general subsidence observed including oxidation and wind erosion of exposed desiccated peat, consolidation, and ground water extraction (Deverel and Leighton 2010). Of these factors, consolidation is likely the mechanism driving the over-prediction at levee sites, with oxidation and wind erosions likely negligible. Peat possesses relatively large coefficients of primary, secondary, and tertiary compression (Wong et al. 2009). Soils beneath levees are exposed to greater vertical consolidation stresses imposed by the self-weight of the levee fill. The ultimate amount of consolidation is directly proportional to the coefficients of consolidation, the consolidation stress, and stress-history of the soil (i.e., over-consolidation ratio, OCR). Soils under larger loading will experience greater amounts of consolidation, all other factors held constant. In the free-field sites, consolidation would be expected to be less significant, but

oxidation and wind erosion could be appreciable because these peats extend to the surface where they are exposed to the atmosphere. The results in Figure 3.9 suggest that of these competing factors, the high degrees of consolidation at levee sites are more significant and lead to greater reductions of peat thickness since 1982.

To evaluate trends in the bias, I compute residuals ($\ln t_p - \ln \hat{t}_p$) and plot them against $\hat{t}_p$, as shown in Figure 3.10. A rather constant trend of over-prediction is observed at both free-field and levee sites when $\hat{t}_p$ > 1-3 m, which I interpret to be a result of subsidence effects. Conversely, t_p appears to be under-predicted at sites predicted to have minor deposits ($\hat{t}_p$< 1-3 m). This observation is likely an artifact of available data and kriging interpolation, in combination with the fact that subsurface layers of peat will never fully disappear. In other words, thin layers of peat will experience some degree of subsidence, however the thicknesses are unlikely to go to zero unless removed through natural processes (i.e., weathering and erosion) or human excavation.

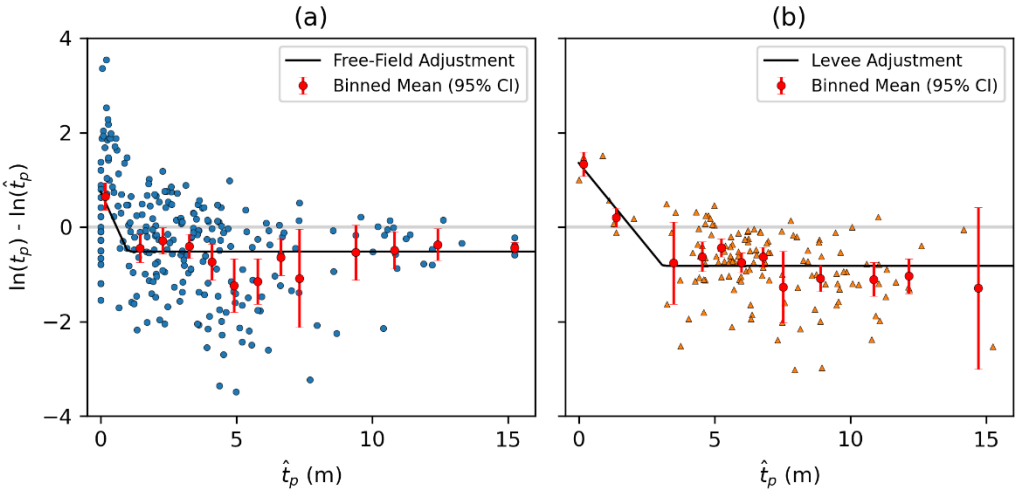


Figure 3.10: Residuals versus estimated peat thickness ($\hat{t}_p$) from Deverel and Leighton (2010) and proposed adjustment; (a) free-field only sites; (b) levee only sites.

To correct for subsidence effects, I apply a bilinear adjustment to re-center the residuals about zero:

$$t_p^* = \begin{cases} \hat{t}_p \times \exp(a\hat{t}_p + b) & , \hat{t}_p < t \\ \hat{t}_p \times \exp(at + b) & , \hat{t}_p \geq t \end{cases} \quad (3.1)$$

where t represents the predicted thickness beyond which constant over-prediction is observed (in meters), a is the slope of the adjustment ramp (i.e., trend for minor deposits), and b is a constant adjustment when $\hat{t}_p$ is zero. Coefficients are regressed separately for free-field and levee sites, and are reported in Table 3.2. The proposed adjustment models for each site condition are shown in Figure 3.10.

Table 3.2: Coefficients of peat thickness adjustment model (for Deverel and Leighton 2010).

Site Condition	t (m)	a	b (m)	σ (m)
Free-Field	1.00	-1.2665	0.75	0.99
Levee	3.10	-0.7031	1.36	0.70

3.3.2. V_{S30}

(a) Data Distributions

The most commonly used site parameter in site response models is V_{S30}, which represents the time-average V_S in the upper 30 m of the soil column (e.g., Borcherdt 1994):

$$V_{S30} = \frac{30\text{ m}}{tt_{30}} \quad (3.2)$$

where tt_{30} is the duration of time needed for shear waves to travel from a depth of 30 m to the ground surface,

$$tt_{30} = \int_{z=0\text{ m}}^{z=30\text{ m}} \frac{dz}{V_S(z)} \tag{3.3}$$

In practice V_{S30} is computed as a summation across depth intervals of constant velocities, and provides a quantitative basis for assigning site class (i.e., NEHRP site class; Dobry et al. 2000). The formulation of Equation (3.2) is based on the theoretical definition for *velocity*, which leads to softer layers (i.e., lower V_S) having a greater influence on V_{S30}. As discussed in Section 3.2.1(c), peats have characteristically low V_S values, which lead to relatively low V_{S30}. Values of V_{S30} computed from all profiles measured to depths greater than 30 m in the Delta range from 59 to 392 m/s, with an average value of approximately 206 m/s as shown in Figure 3.11.

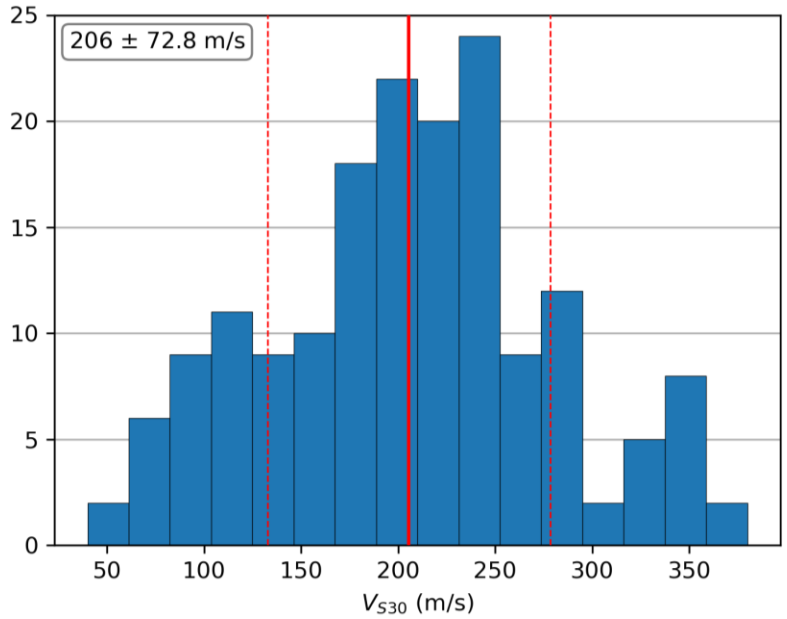


Figure 3.11: Histogram of V_{S30} computed from measured V_S profiles with $z_p \geq 30$ m in the Delta.

The V_{S30} distributions shown in Figure 3.11 are computed from profiles with and without peaty layers, so the average value shown does not provide a good representation for sites with peat. These sites can be partitioned based on expected surficial geology (Bay Delta Live 2021) into six sub-groups:

1. Peat and Muck: peat and mud of tidal wetlands and waterways (Holocene).
2. Alluvium (Supratidal Floodplain): flood basin deposits (Holocene).
3. Eoloian Deposits: eolian deposits of upper Modesto formation (upper Pleistocene).
4. Alluvial Fan (Modesto Formation): alluvial fans from glaciated basins – Modesto formation (Pleistocene).
5. Alluvial Fan (Riverbank Formation): alluvial fans from glaciated basins – Riverbank formation (upper Pleistocene).
6. Alluvial Fan (Unglaciated): alluvial fans and terraces from unglaciated drainage basins (Holocene – upper Pleistocene).

Figure 3.12 presents individual histograms of V_{S30} for each of the six geologic groups. Peat sites (Group 1) have an average V_{S30} of 157 m/s, while most non-peat sites (Groups 2-6) have average V_{S30}'s ranging from 218 to 261 m/s. The *Alluvial Fan (Riverbank Formation)* group is atypically stiff with an average V_{S30} value of 340 m/s, however this unit is not typically encountered within the Delta legal boundaries (illustrated in Figure 3.1). As discussed in Section 3.2.1(a), all of these sites are mapped as "Qal1" in Wills et al. (2015), which has a predicted V_{S30} value of 228 m/s. The Wills et al. (2015) V_{S30} prediction reasonably captures the behavior observed for non-peat sites with the exception of *Alluvial Fan (Riverbank Formation)*, as shown in Figure 3.12.

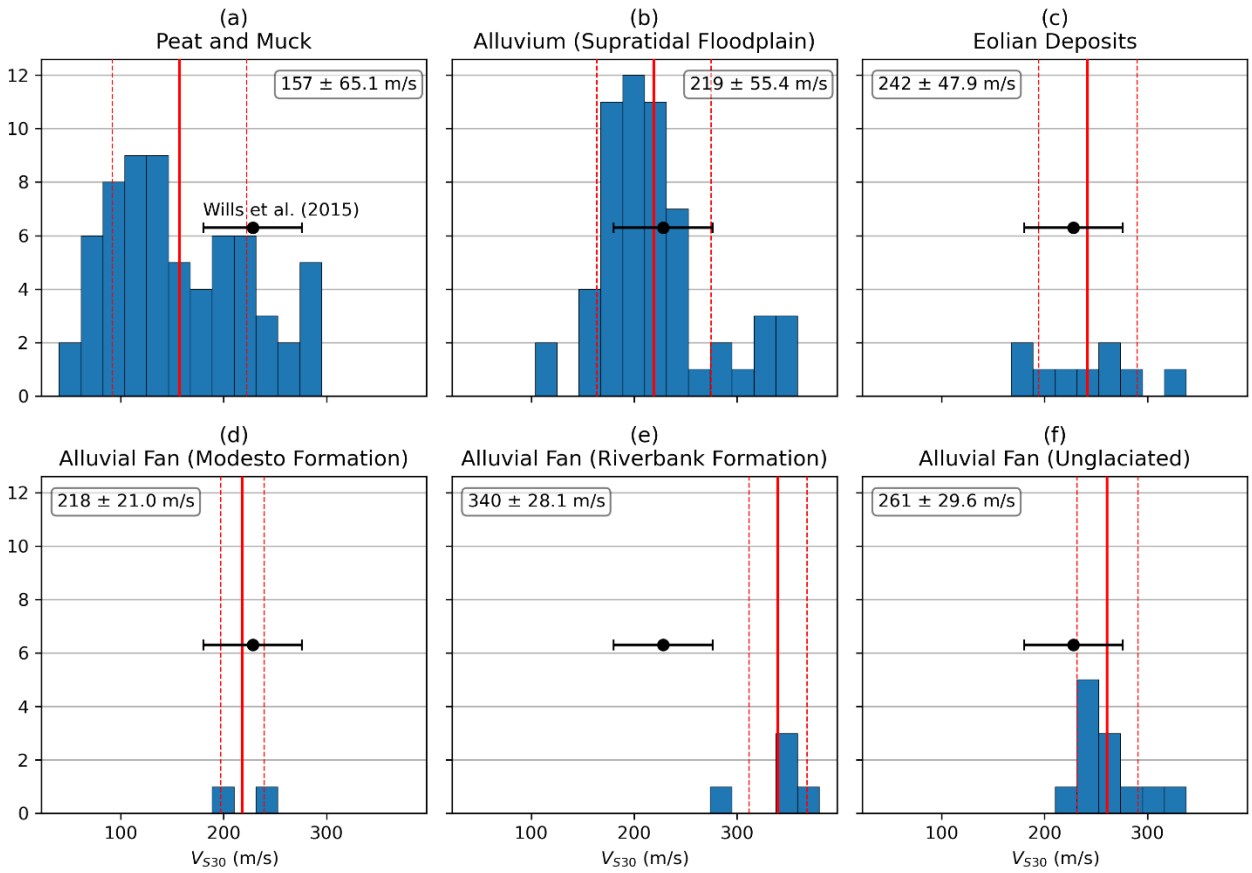


Figure 3.12: Histograms of V_{S30} grouped by Bay Delta Live (2021) geologic unit; mean and standard deviation are shown by solid and dashed red lines, respectively; geology-based V_{S30} prediction from Wills et al. (2015) shown by black marker with error bar; (a) Peat and Muck; (b) Alluvium (Supratidal Floodplain); (c) Eolian Deposits; (d) Alluvial Fan (Modesto Formation); (e) Alluvial Fan (Riverbank Formation); (f) Alluvial Fan (Unglaciated).

(b) Peat Thickness V_{S30} Proxy Model

V_{S30} values computed from measured V_S profiles are used as site parameters for 13 Delta stations, as shown in Figure 3.5. Proxy-based methods are required to estimate V_{S30} at the remaining 41 stations from other site parameters including geology, terrain, and topographic slope, however, existing models are not calibrated for sites with peat. One option is to develop a new geology category calibrated for *Peat and Muck* sites as assigned from the Bay Delta Live (2021) map, for

which a constant value would be assigned. The constant value is estimated by computing the average V_{S30} in natural log units for the combination of all *Peat and Muck* sites and profiles which include peat. I refer to this option as proxy model "0":

$$\ln(\bar{V}_{S30})_0 = \frac{1}{N}\sum_{i=1}^{N} \ln(V_{S30,i}) \tag{3.4}$$

where $\ln(\bar{V}_{S30})$ represents the V_{S30} estimate in natural log units, $V_{S30,i}$ is the V_{S30} value for profile *i*, and *N* represents the total number of profiles. $\ln(\bar{V}_{S30})_0$ is estimated to be equal to 5.0516, which corresponds to a V_{S30} value of about 156 m/s.

Since V_{S30} is strongly influenced by soft layers, t_p holds significant potential to predict V_{S30}. Figure 3.13 plots V_{S30} against t_p, and in general as t_p increases V_{S30} decreases. A V_{S30} proxy model for peat sites which uses t_p as the sole independent variable may be expressed as:

$$\ln(\bar{V}_{S30}) = \mu_{\ln V_{S30}} + C(t_p - \mu_{t_p}) \tag{3.5}$$

where $\mu_{\ln V_{S30}}$ is the natural log mean of all V_{S30} computed from sites with peat, μ_{t_p} is the mean peat thickness in units of meters, and t_p is the peat thickness at the site of interest. The purpose of $\mu_{\ln V_{S30}}$ and μ_{t_p} are to center the model with respect to $\ln(V_{S30})$ and t_p so that only a slope coefficient (C) need be regressed. Slight differences are observed in the trends between free-field and levee sites, therefore two alternate proxy models are considered: $\ln(\bar{V}_{S30})_1$ which uses a general set of coefficients and $\ln(\bar{V}_{S30})_2$ which uses separate sets of coefficients conditioned on site condition (free-field or levee). Table 3.3 summarizes the coefficients of each model, which are shown as solid curves in Figure 3.13.

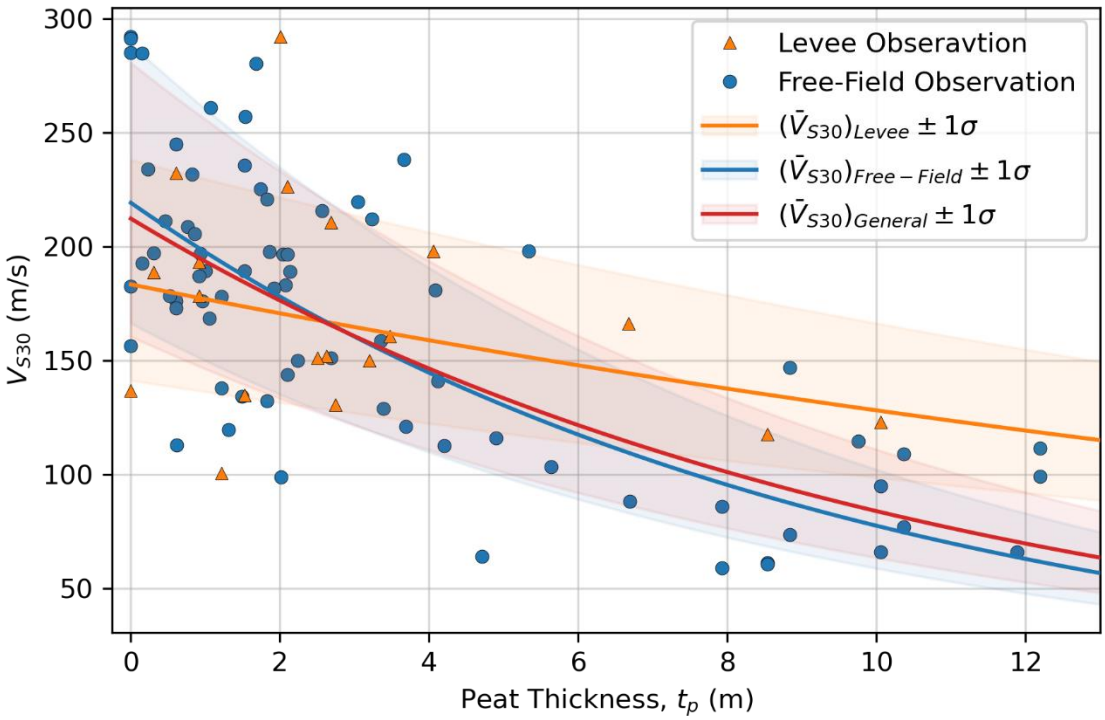


Figure 3.13: Plot of peat thickness (t_p) versus V_{S30}; regression models for levee, free-field, and general site conditions included.

Table 3.3: Summary of peat-thickness V_{S30} proxy model coefficients.

Proxy Model	Site Condition	$\mu_{\ln V_{S30}}$	μ_{t_p} (m)	C	$\ln \sigma$
$\ln(\bar{V}_{S30})_0$	Any	5.0516	-	-	0.3953
$\ln(\bar{V}_{S30})_1$	Any	5.0516	3.3000	-0.0928	0.2797
$\ln(\bar{V}_{S30})_2$	Free-Field	5.0378	3.3885	-0.1039	0.2760
	Levee	5.1056	2.9547	-0.0358	0.2608

To evaluate the performance of each model, I compute residuals, $\ln(V_{S30}) - \ln(\bar{V}_{S30})$, from which model bias and variability can be assessed. Additionally, I compare the performance of established models which are not calibrated to the soft-peaty conditions found in the Delta:

- Wald and Allen (2007) – topographic slope based model.

- Yong et al. (2012) – terrain based model.
- Wills et al. (2015) – geology and gradient based model.

Histograms of residuals computed from each proxy model are shown in Figure 3.14. Model bias is represented by the mean value, where positive and negative values suggest under- and over-prediction, respectively. Variability is quantified by the standard deviation. Table 3.4 provides a summary of each proxy model's performance.

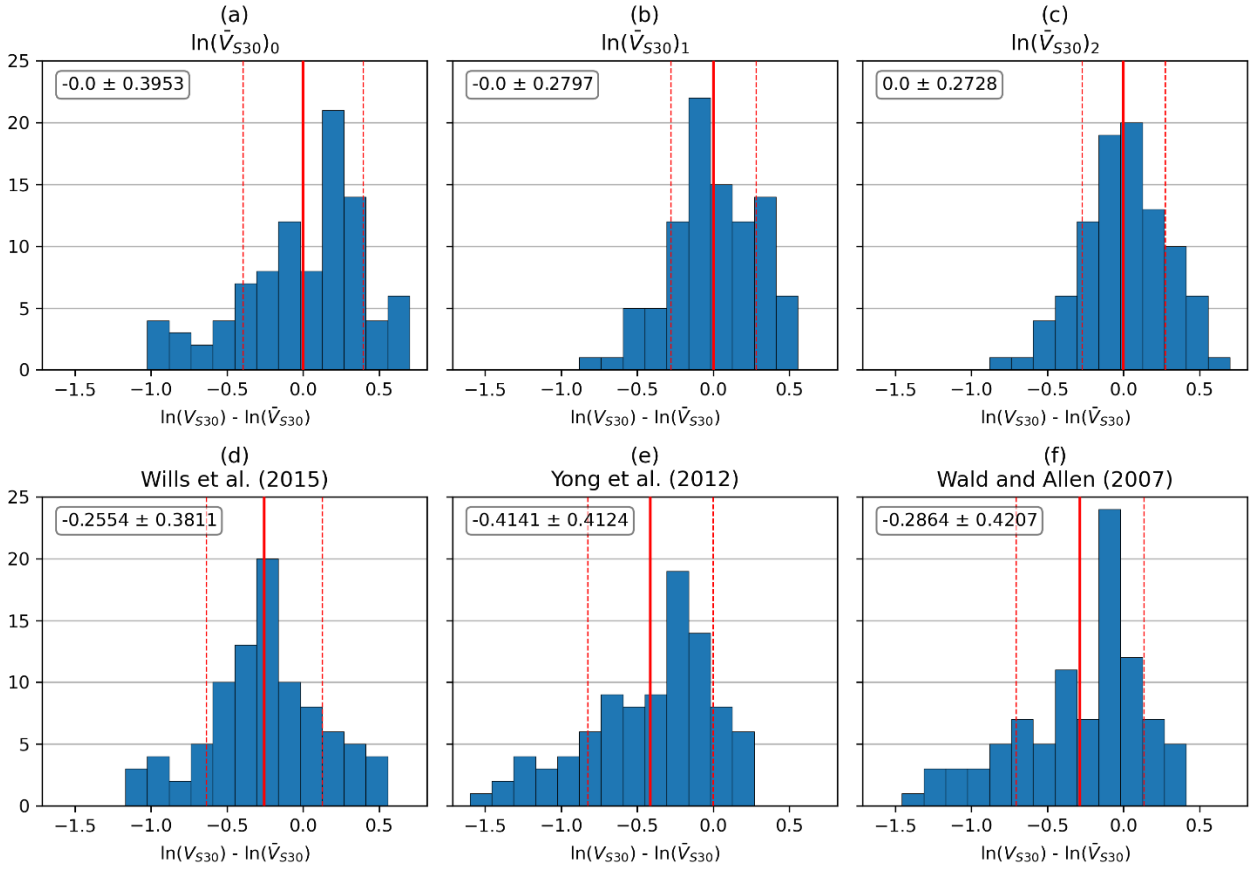


Figure 3.14: Histograms of V_{S30} residuals computed from different proxy-based models; (a) peat-geology model [$\ln(\bar{V}_{S30})_0$] (b) peat-thickness general model [$\ln(\bar{V}_{S30})_1$]; (c) peat-thickness levee and free-field specific coefficients [$\ln(\bar{V}_{S30})_2$]; (d) geology based (Wills et al. 2015); (e) terrain based (Yong et al. 2012); and (f) slope based (Wald and Allen 2007).

Table 3.4: Summary of V_{S30}-proxy model performance in the Delta.

Proxy Model	Bias (mean of residuals)	Variability (standard deviation of residuals)
$\ln(\bar{V}_{S30})_0$	0.0000	0.3953
$\ln(\bar{V}_{S30})_1$	0.0000	0.2797
$\ln(\bar{V}_{S30})_2$	0.0000	0.2728
Wald and Allen (2007)	-0.2864	0.4207
Yong et al. (2012)	-0.4141	0.4124
Wills et al. (2015)	-0.2553	0.3811

All three established models over-predict V_{S30} at peaty sites and have relatively large variabilities (0.38 – 0.42). This is to be expected as these models are not well calibrated for soft peaty soil conditions. Therefore, estimated values using these proxy models will result in inaccurate results. The geology-based constant adjustment presented at the start of this section ($\ln(\bar{V}_{S30})_0$) significantly reduces the bias, however the variability (0.40) is comparable to the established models. It is for this reason that I considered a proxy model conditioned on t_p, aiming to not only reduce the bias, but also to reduce variability. As expected, bias from the t_p based models are practically zero (expected because dispersion is computed from the same data used during model development), and uncertainty is the lowest (0.27-0.28).

The performance of $\ln(\bar{V}_{S30})_2$ is slightly better than $\ln(\bar{V}_{S30})_1$, which is to be expected given the use of additional independent variables. V_{S30} at levee sites with negligible t_p is predicted to be about 180 m/s as compared to 220 m/s in the free-field. There is a reversal of this trend at a t_p of about 2.5 m, as observed in Figure 3.14. One justification may be that the levee-fill materials (usually native clays and silts mixed with organics) have unusually low V_S, which results in a low V_{S30}. Furthermore, the stronger scaling of V_{S30} with t_p for free-field conditions than for levee sites

may be a result of smaller confining stresses. Confining stresses imposed on peat underneath levees will increase their stiffness and reduce their thickness; both of these factors would be expected to reduce the overall influence of the peat layer in the V_{S30} calculation. However, this reasoning is speculative, and given the limited available data for levee sites with $t_p \geq 4$ m, there is insufficient evidence to conclude that the differences observed between free-field and levee sites is meaningful. Therefore, the generalized t_p based proxy model [$\ln(\bar{V}_{S30})_1$] is preferred and applied to assign site parameters in this project.

3.3.3. mHVSR

H/V data is used to develop HVSR curves from which site period and possible resonance effects in the site amplification can be identified (Nakamura 1989). The data described in Section 3.2.2(b) are time-domain signals of ground vibrations, which require processing and post-processing before site parameters can be estimated. This section presents the processing protocols and post-processing routine which was implemented to assign site parameters at Delta stations.

(a) Processing

I implement the procedures proposed by Wang et al. (2022a) in R (R Core Team 2022), which are an update to traditional methods (SESAME, 2004; Molnar et al. 2022) to convert time-domain signals to mHVSR. These procedures differ slightly from those built into the widely used *Geopsy* software (Wathelet et al. 2020), and can be summarized into six steps:

1. Windowing
2. Antitriggering
3. Signal processing

4. Combining horizontal components
5. HVSR calculation
6. Resampling and decimation

The first step requires subdividing the recorded ambient vibrations into windows of shorter durations. The lowest prominent peak (f_0) in the mHVSR spectrum should be greater than $10/T_{win}$, where T_{win} is the window length used in seconds (SESAME 2004). However, f_0 can only be reliably used if there are a sufficient number of significant cycles (N_{cyc}) within each window, defined as:

$$N_{cyc} = T_{win} f_0 N_{win} \tag{3.6}$$

where N_{win} represents the number of windows used in the mHVSR spectrum computation. To reduce dispersion, especially at low-frequencies, more windows are recommended. Recall, ambient vibrations were recorded at each site spanning durations between 2.5 to 4 hours. T_{win} of 300 to 450 sec where used during processing (satisfying $N_{win} \geq 20$) to ensure that each mHVSR could be reliably used at frequencies above 0.1 Hz (within the frequency domain described by GMMs). Since each site possessed ample data, non-overlapping windows are defined. Figure 3.15 illustrates the windowing procedure for site CE 67265.

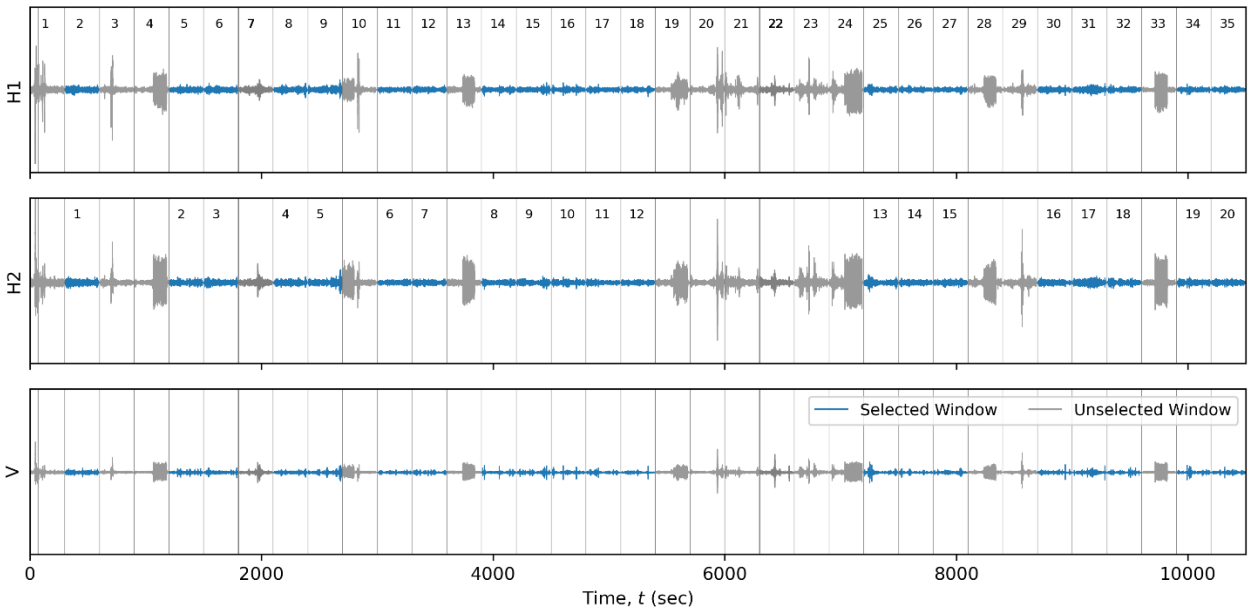


Figure 3.15: Ambient vibrations recorded at CE_67265. Windowing is performed using T_{win} = 300 sec, and individual windows are selected to exclude transient excitations (i.e., antitriggering).

The second step is to perform antitriggering. "Triggering" refers the recording of temporary vibrations (or excitations) within a window. Ambient vibrations have relatively uniform amplitudes over most time intervals (Molnar et al. 2018), while vibrations caused by traffic or other industrial sources will produce transient signals. These transient signals can introduce undesirable variability into mHVSR, which is the motivation for antitriggering, or rejection of windows with contaminated data. Antitriggering is performed in the time-domain, as illustrated in Figure 3.15, and subsequently in the frequency-domain, if necessary. If any component for a particular window is found to be problematic, the entire window (three components) is removed.

The vast majority of Delta stations are in rural locations away from potential sources of transient signals, however seven sites are located in busy urban environments (CE_57195, CE_57531, CE_67215, CE_67577, CE_67615, CE_67642, and NC_C057), and four others are

located within close proximity to the Antioch Bridge (CE_67265, CE_67266, CE_67523, and CE_67910). Careful considerations for antitriggering were made for these sites, however the accepted windows likely include a mix of true-ambient (low level noise) and low-amplitude triggered vibrations (high-level noise). HVSR methods utilize near-vertical incident waves, but also involve a significant contribution of surface waves. Triggered vibrations from industrial and traffic sources in close proximity to the sensor contain a mixture of surface and reflected, refracted, and direct body waves (Mihaylov et al. 2016). The mixture of wave types complicates the interpretation of HVSR, because transients may enhance peaks if they have the same frequency content as the site, or produce high-frequency peaks not representative of the site condition (Parolai and Galiana-Merino 2006). In such cases, it is difficult to determine if the fundamental resonance is suppressed by or enhanced by transient signals without a rigorous study. Furthermore, some studies suggest that the low-frequency portion of the mHVSR will have greater uncertainty when high-level noise is present (Castellaro and Mulargia 2009). These concerns are taken into consideration when deriving site parameters from mHVSR curves (discussed in the following section).

Signal processing includes tapering, filtering, and smoothing which is applied to each window. A cosine taper with a length of 5% of T_{win} (Chatelain et al. 2008) and a high-pass corner frequency of 0.1 Hz are applied to each window to reduce low-frequency drift in the waveforms. Filtering is required because it can influence the outcome of the smoothing operator. In principle, mHVSR should theoretically not be affected by filtering because the same filter is applied to all three components – this hypothesis was tested and some sensitivity to filtering was encountered for frequencies below the high pass filter corner frequency, but negligible differences were observed for frequencies above the high pass filter corner frequency, as illustrated in Figure 3.16.

Konno and Ohmachi (1998) smoothing is applied to reduce high-frequency noise and to facilitate identification of peak features. Spectral smoothing is applied to each component for each window individually using a bandwidth parameter equal to 30.

6.

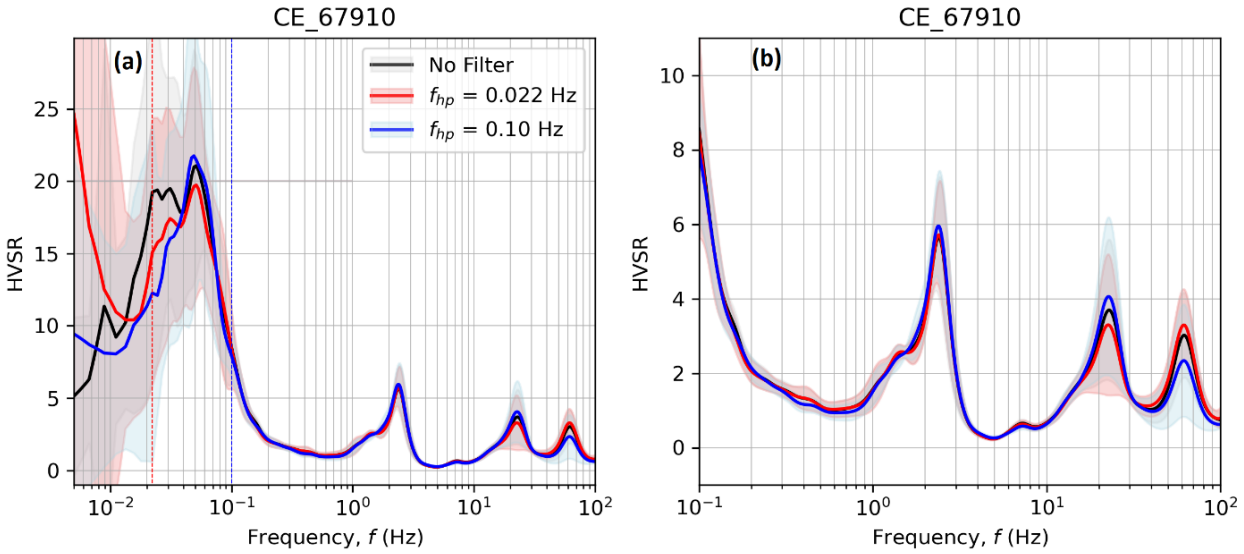


Figure 3.16: Comparison of mHVSR when different high-pass corner frequencies are used in record processing. (a) HVSR for wide frequency range (0.0005 to 100 Hz) showing differences below 0.1 Hz; (b) HVSR for more typical frequency range (0.1 to 100 Hz) showing negligible differences.

It is necessary to combine the two as-recorded orthogonal horizontal components into a single representative horizontal component to facilitate the HVSR computation. The median component (RotD50; Boore 2010) is preferred given its general acceptance and use in ground-motion studies. However, variable-azimuth components are also computed for distinct azimuths between 0 to 180 degrees by geometrically combining the two as-recorded horizontal components and both sets of results are provided in the HVSR component of the database (Kwak et al. 2021).

HVSR is computed by dividing the smoothed horizontal-component Fourier amplitudes by the smoothed vertical-component Fourier amplitude. This is performed for each window, and a mean HVSR is obtained by taking the average across windows at each frequency. This averaging provides additional smoothing beyond the Konno Ohmachi (1998) smoothing applied to individual components. The dispersion of HVSR is computed as the standard deviation of all windows at each frequency.

Lastly, resampling and decimation is required to provide a uniform level of resolution on a log frequency scale. The effect of this step is an increase of resolution at low frequencies and decrease at high frequencies (the number of samples does not change).

Figure 3.17(a) shows an example mHVSR curve for the CE 67265 site. Variable components are used to construct a polar curve (as shown in Figure 3.17b), from which azimuthal variation of mHVSR can be inferred. Polar curves are used to detect complexity that does not conform with the assumptions of one-dimensional wave propagation (Cheng et al. 2020). The 0.9 Hz resonance clearly visible in Figure 3.17 spans across all azimuths, but is strongest in the north-south direction (0^o – 20^o and 160^o – 180^o). Site CE 67265 is positioned between two piers of the Antioch Bridge (~60 m span), which is oriented in the north-south direction. The higher energy associated with the north-south direction is likely a result of vibrations transferring from the structure into the ground, which is generally observed for frequencies in the range of 0.3 – 20 Hz.

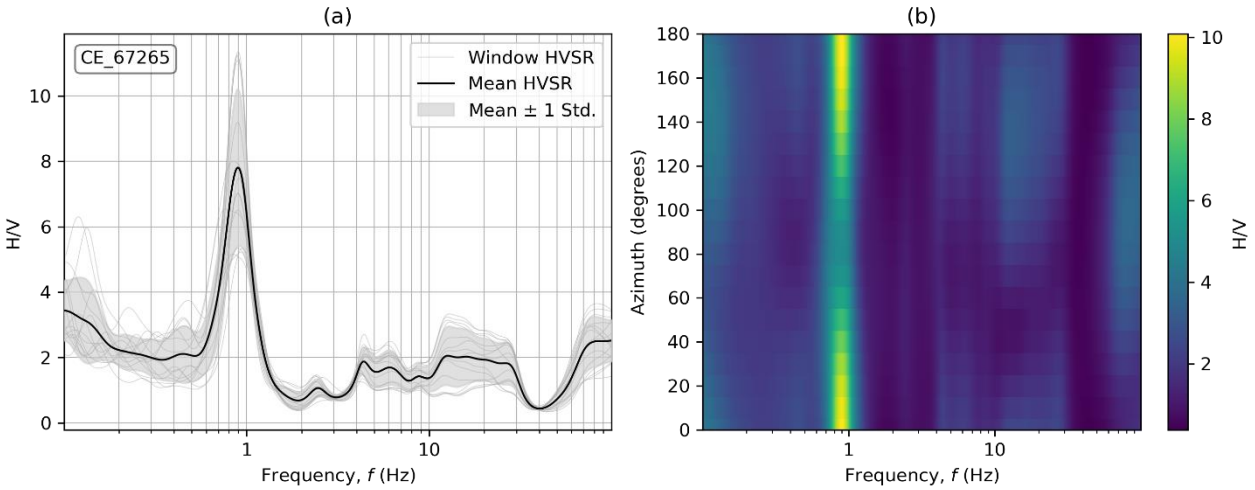


Figure 3.17: (a) Mean mHVSR curve for CE_67265, HVSR for individual windows shown by light gray lines; (b) Azimuthal variation of mHVSR at CE_67265.

(b) Peak Identification

I examined each mean mHVSR curve individually to identify "peak parameters" which may be useful in site response modeling. *Peak parameters* are those which describe peak features in the mean HVSR curve and include the frequency (f_p), absolute amplitude (a_p), relative amplitude to adjacent plateaus (c_1), and the half-width (w_p). In previous sections f_0 was used as a general descriptor of site fundamental frequency, however f_p is now used to describe a fit parameter which can be used as a site parameter, and is expected to relate to f_0. The peak detection algorithm proposed by Wang et al. (2021) is used to objectively identify peak features and assign corresponding peak parameters in R (R Core Team 2022).

The Wang et al. (2021) algorithm implements a regression tree (Breiman et al. 1984) to simplify the mean HVSR curve into a step-wise function. The relative amplitudes and lengths of adjacent steps are used to identify peak features – if several peaks are identified the one at the

lowest frequency is preferred. If a peak feature is identified, the simplified regression tree data within the frequency range between the left-lower step to the right-lower step (relative to the peak step) is used to fit a Gaussian pulse function adapted from Ghofrani and Atkinson (2014):

$$F_{H/V} = c_0 + c_1 \exp\left[-\frac{1}{2}\left(\frac{\ln(f/f_p)}{2w_p}\right)^2\right] \tag{3.7}$$

where c_0 is a frequency-independent constant representing the amplitude of the flat tails. The absolute amplitude of the peak is calculated as:

$$a_p = c_0 + c_1 \tag{3.8}$$

I added the ability to manually constrain the c_0 parameter during the regression. This modification is necessary to adjust the fit at sites where the optimal fit is deemed to be unsatisfactory based on visual inspection. Figure 3.18 presents examples illustrating results of the peak identification algorithm and fitting routines for sites with no peak, one peak, and multiple peaks. Results for 45 Delta stations are summarized in Table 3.5 and shown in Figure 3.19.

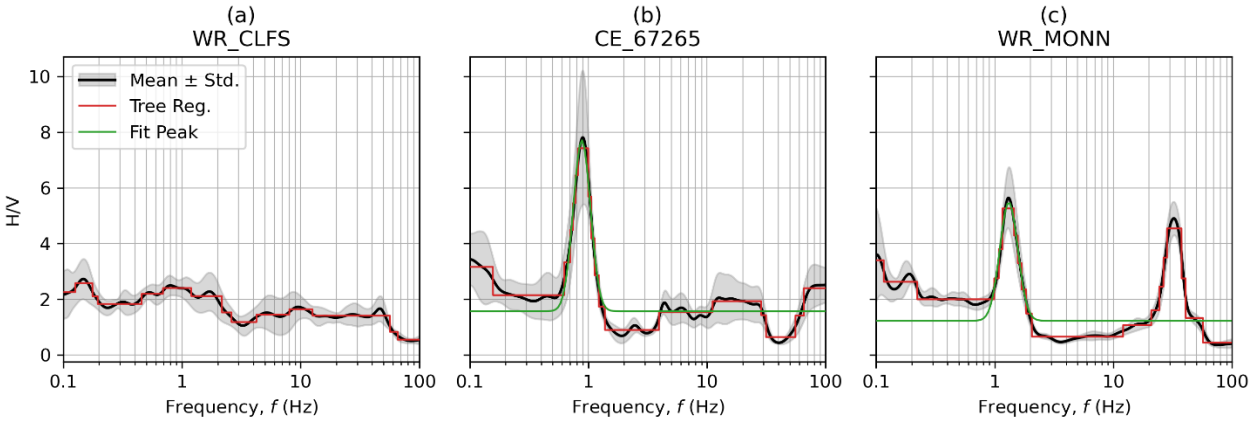


Figure 3.18: Examples results for peak identification algorithm and peak fitting routine of Wang et al. (2021) for Delta sites with (a) no peak, (b) one peak, and (c) multiple peaks.

Table 3.5: Summary HVSR peak fitting.

Delta Station	Peak Presence	c_0	c_1	a_p	f_p (Hz)	w_p
BK_JEPS	Y	0.8334	1.4573	2.2907	0.6613	0.4950
BK_TWIT	Y	0.9442	2.9397	3.8839	0.6913	0.3133
CE_57195	N	-	-	-	-	-
CE_57531	Y	0.8142	1.5074	2.3216	0.7505	0.5000
CE_57534	N	-	-	-	-	-
CE_67215	Y	1.1000	0.9144	2.0144	0.6945	0.3826
CE_67265	Y	1.5713	6.1371	7.7084	0.8809	0.1677
CE_67266	Y	1.5890	2.2760	3.8650	1.7009	0.1206
CE_67523	Y	1.1000	0.9444	2.0444	2.8892	0.2637
CE_67533	N	-	-	-	-	-
CE_67557	N	-	-	-	-	-
CE_67587	Y	1.1000	4.1500	5.2500	1.1052	0.2888
CE_67615	N	-	-	-	-	-
CE_67910	Y	0.8000	4.5000	5.3000	2.3514	0.1821
NC_C057	N	-	-	-	-	-
NP_DIX	Y	1.2000	1.0257	2.2257	0.6891	0.3130
NP_EMR	Y	1.0000	3.8300	4.8300	1.0305	0.2848
NP_KIR	Y	1.4560	7.9000	9.3560	2.0822	0.3471
NP_LVA1	Y	1.1533	3.0546	4.2079	0.7870	0.3674
NP_LVA2	Y	1.1591	3.5715	4.7306	0.7405	0.3889
NP_LVA3	Y	1.7500	3.4300	5.1800	1.0001	0.2506
NP_LVA4	Y	1.1064	5.0877	6.1941	0.8552	0.3434
NP_MCD	N	-	-	-	-	-
NP_PLA	N	-	-	-	-	-
NP_SIA	N	-	-	-	-	-
WR_CKR	N	-	-	-	-	-
WR_CLFN	N	-	-	-	-	-
WR_CLFS	N	-	-	-	-	-
WR_HOLT	Y	1.2300	1.4383	2.6683	0.6471	0.3199
WR_MOFF	Y	0.9700	3.1800	4.1500	1.6487	0.3195
WR_MONN	Y	1.2299	4.2482	5.4781	1.3139	0.1755
WR_SHER	Y	1.1524	4.4550	5.6074	0.7317	0.4024
WR_SIFF	Y	1.0195	2.6451	3.6646	0.7347	0.2452
WR_STNI	Y	1.7900	1.3200	3.1100	1.8221	0.1000
YU_CEC	N	-	-	-	-	-
YU_HMT	Y	1.2900	2.8563	4.1463	1.5527	0.2391
YU_HOL1	Y	1.1500	2.6163	3.7663	2.1789	0.3855
YU_HOL2	Y	1.1900	4.1514	5.3414	2.1240	0.2665

Delta Station	Peak Presence	c_0	c_1	a_p	f_p (Hz)	w_p
YU_HOL3	Y	1.0500	4.1947	5.2447	4.6526	0.2718
YU_SMB	Y	1.0000	4.2000	5.2000	8.4149	0.2000
YU_SMT	Y	1.5200	1.1500	2.6700	3.1582	0.2000
YU_SRB	Y	1.1020	3.4153	4.5172	5.9005	0.3105
YU_SRT	Y	1.5918	2.8039	4.3957	2.8786	0.2230
YU_STF	Y	1.2300	2.5040	3.7340	2.7315	0.1935
YU_WHR	Y	0.8694	1.0757	1.9451	3.0714	0.3752

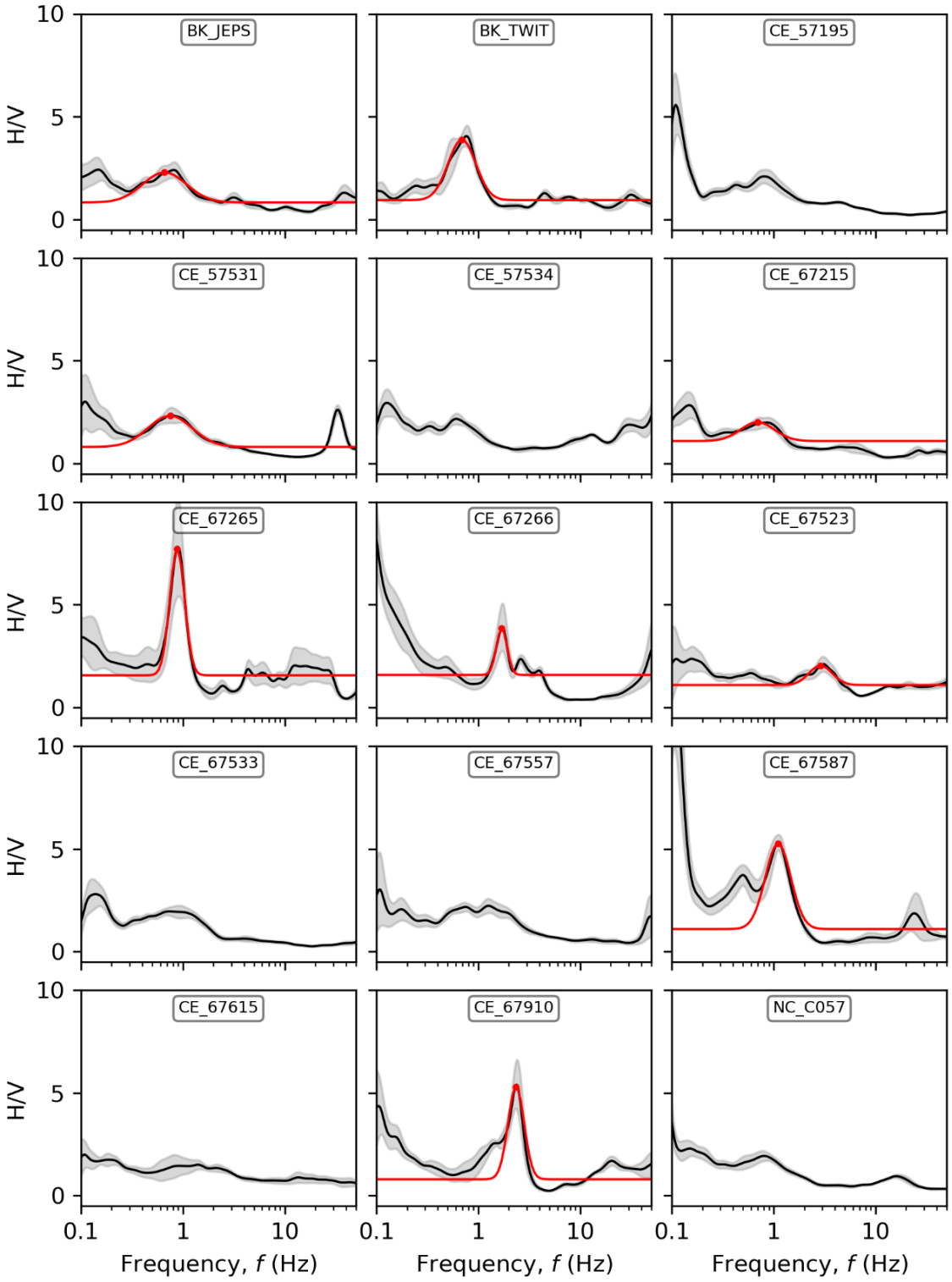


Figure 3.19: Processed mHVSR curves for 45 Delta stations. Fit Gaussian peak (Equation 3.7) is shown in red for sites with identified peaks.

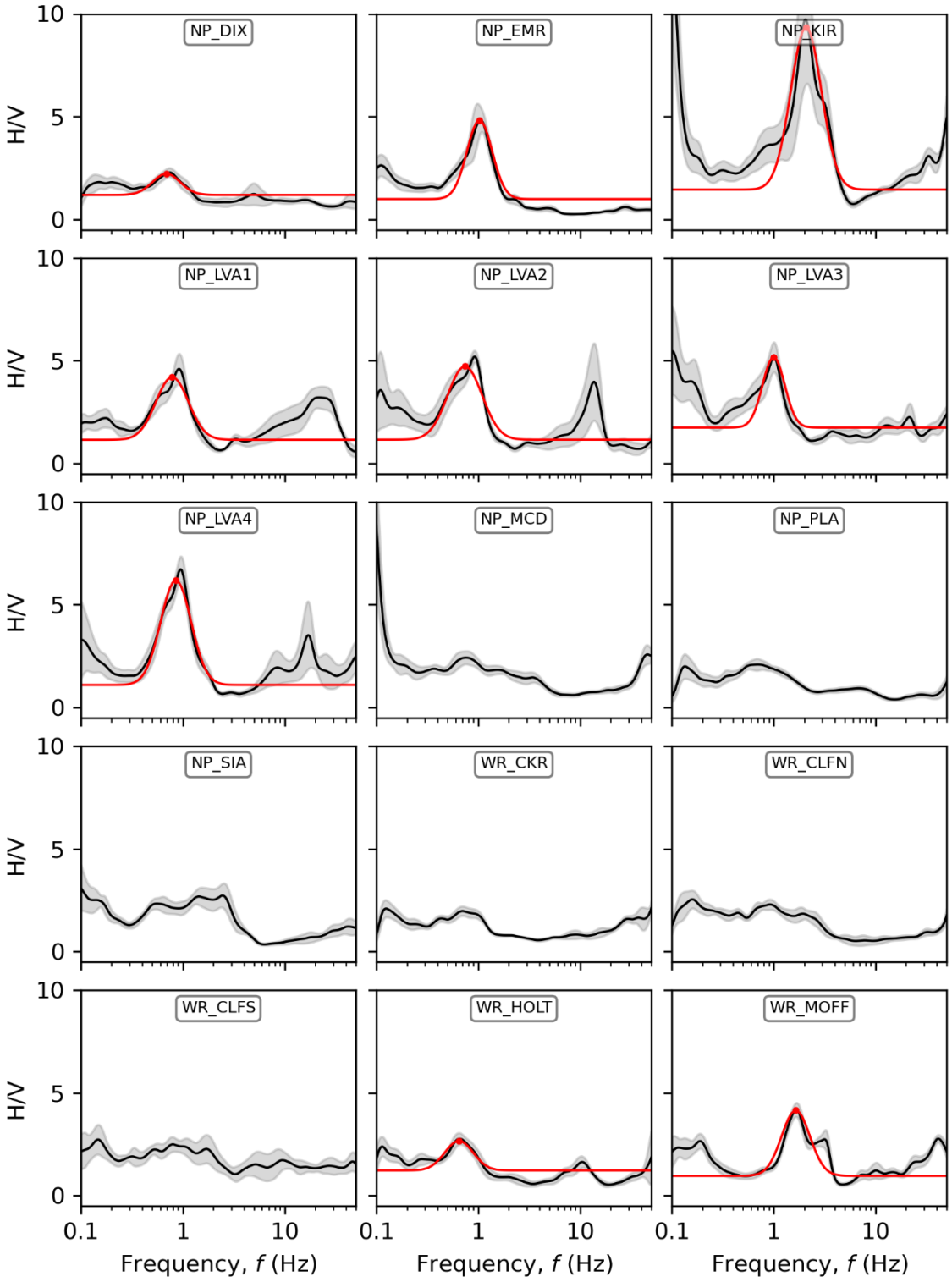


Figure 3.19: *Continued.*

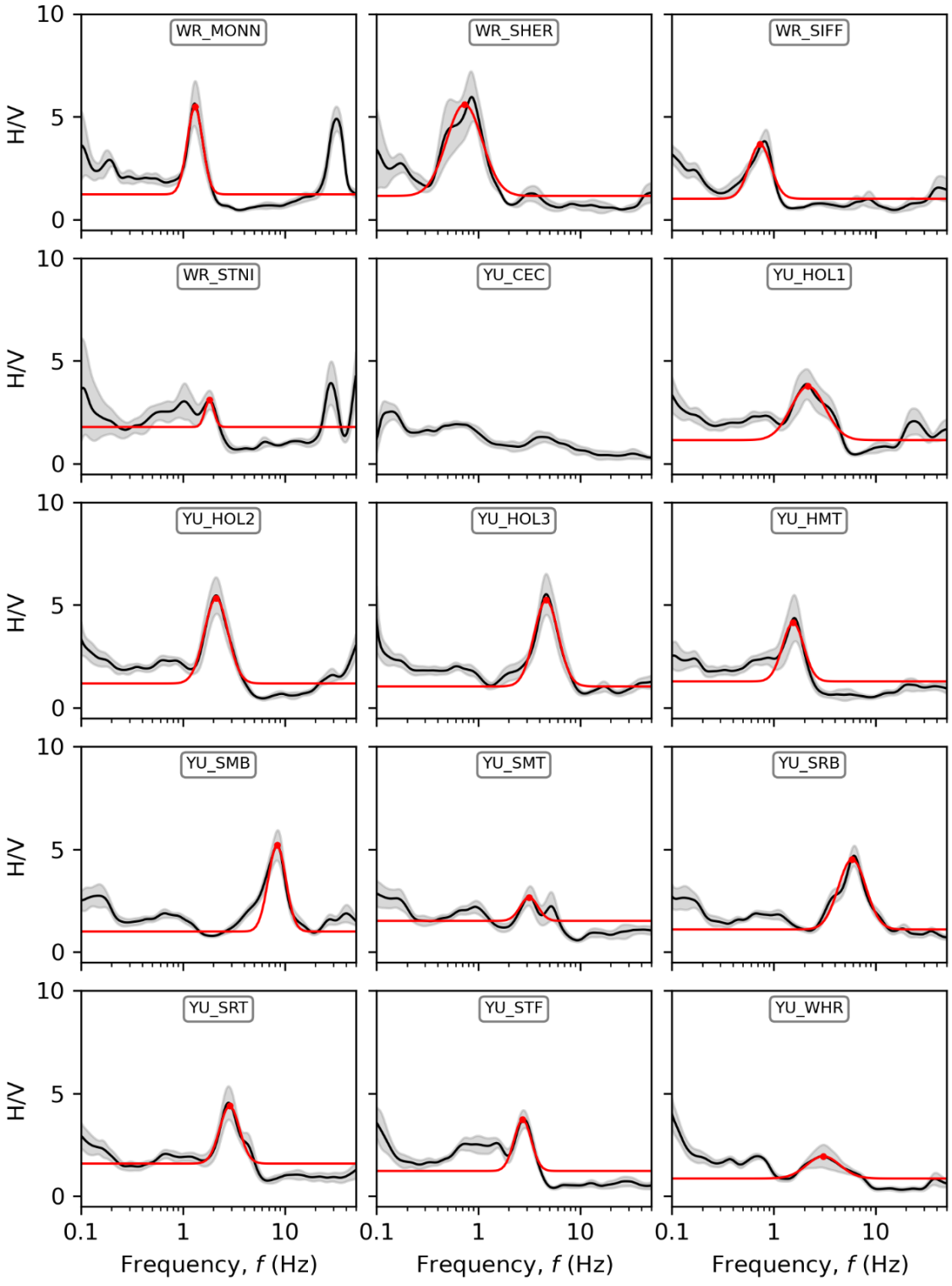


Figure 3.19: *Continued.*

Figure 3.20 presents the summary statistics for 45 Delta stations with mean mHVSR curves derived from ambient vibrations. Approximately 50-70% of California sites do not have HVSR peaks depending on whether relatively liberal or conservative peak identification thresholds are adopted (Wang et al. 2021), however 32 (71%) of Delta sites are identified as having peaks, even using conservative criteria. One factor which produces peak features is the presence of a strong impedance contrast in the soil column, such as that observed between soft-peaty layers and stiffer underlying non-organic soils typically encountered at Delta sites. I interpret these peaks as being associated with resonance of the portion of the soil column above an impedance contrast, which includes the peat layers. The peak features are described by a relatively wide parametric range (i.e, 0.65 Hz $\leq f_p \leq$ 8.41 Hz, 1.95 $\leq a_p \leq$ 9.36, etc.).

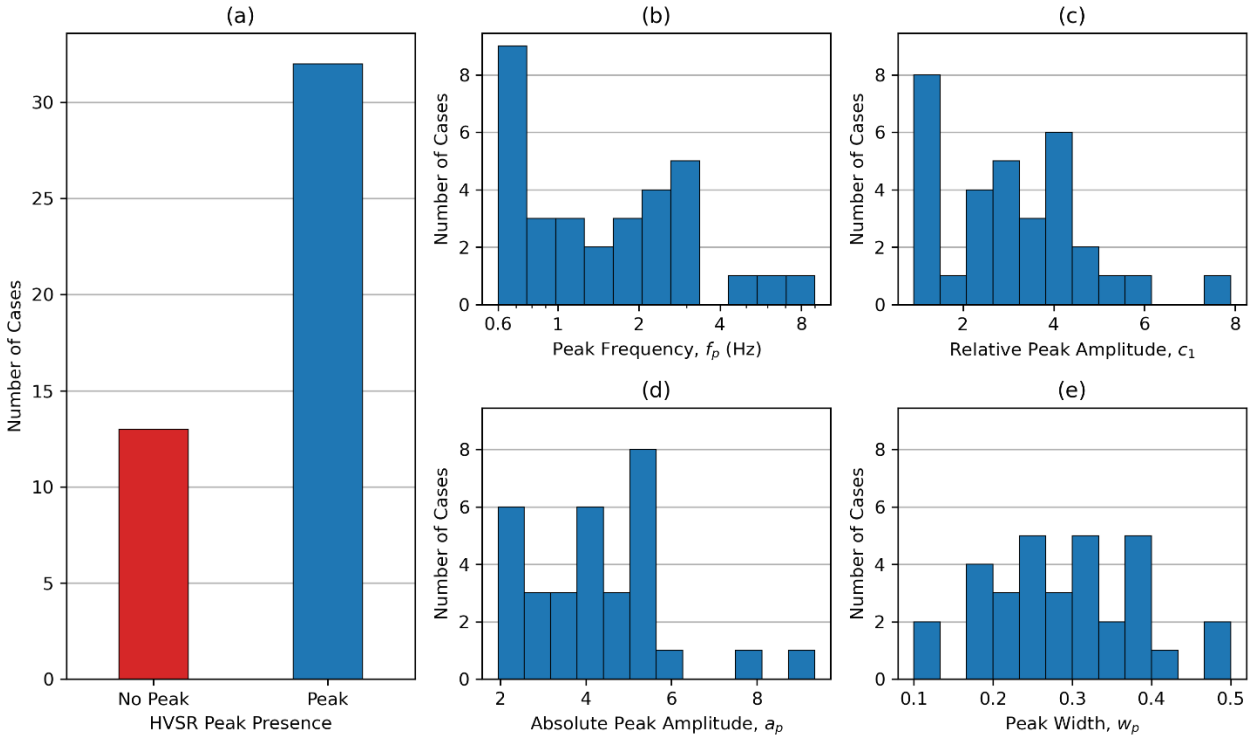


Figure 3.20: Mean mHVSR peak summary statistics; (a) peak presence; (b) peak frequency; (c) relative peak amplitude; (d) absolute peak amplitude; and (e) peak width.

In most cases, I associate the peaks with soft peaty soils and soft non-organic soils, with the impedance contrast occurring at the Holocene-Pleistocene interface. For example, the WR_MONN site possesses a prominent peak (i.e., f_p) in the mHVSR at about 1.3 Hz, which is most likely controlled by the peat layer at a depth of six to nine meters, as shown in Figure 3.21. This hypothesis is tested by calculating the fundamental site period ($T = 1/f_0$) from the measured V_S profile as a function of depth to see which depth range controls the resonant peak:

$$T = \frac{4h}{V_{eq}} \tag{3.9}$$

where h is the total thickness and V_{eq} represents an "effective velocity" for the total soil profile. *Harmonic averaging* (Urzúa 1974) is a common method used to estimate V_{eq}:

$$V_{eq} = \overline{V_S} = \frac{\sum_{i=1}^{n} d_i}{\sum_{i=1}^{n} d_i / V_{Si}} \tag{3.10}$$

where d_i is the thickness of any layer, V_{Si} is the corresponding shear wave velocity, and n is the number of layers in the soil profile. A drawback of this method is that it tends to overestimate T for profiles where V_S increases with depth, such as observed in the Delta.

The *simplified Rayleigh method* (Dobry et al. 1976) provides more accurate estimates of T by accounting for variations of soil properties with depth:

$$T \approx T_{SR} = \frac{2\pi}{\omega_{SR}} \tag{3.11}$$

and

$$\omega_{SR}^2 = \frac{4 \sum_{i=1}^{n} \frac{(h - z_{mi})^2}{V_{Si}^2} d_i}{\sum_{i=1}^{n} (X_i + X_{i+1})^2 d_i} \tag{3.12}$$

where the layers, i = 1, 2, 3, …, n are counted from the bottom up, ($h - z_{mi}$) is the depth of the midpoint of layer i, and X_i and X_{i+1} represent the approximate modal shapes at the bottom and top of layer i, respectively. Modal shapes are computed from the bottom up with $X_0 = 0$ at the bottom of the profile:

$$X_{i+1} = X_i + \frac{h - z_{mi}}{V_{Si}^2} d_i \tag{3.13}$$

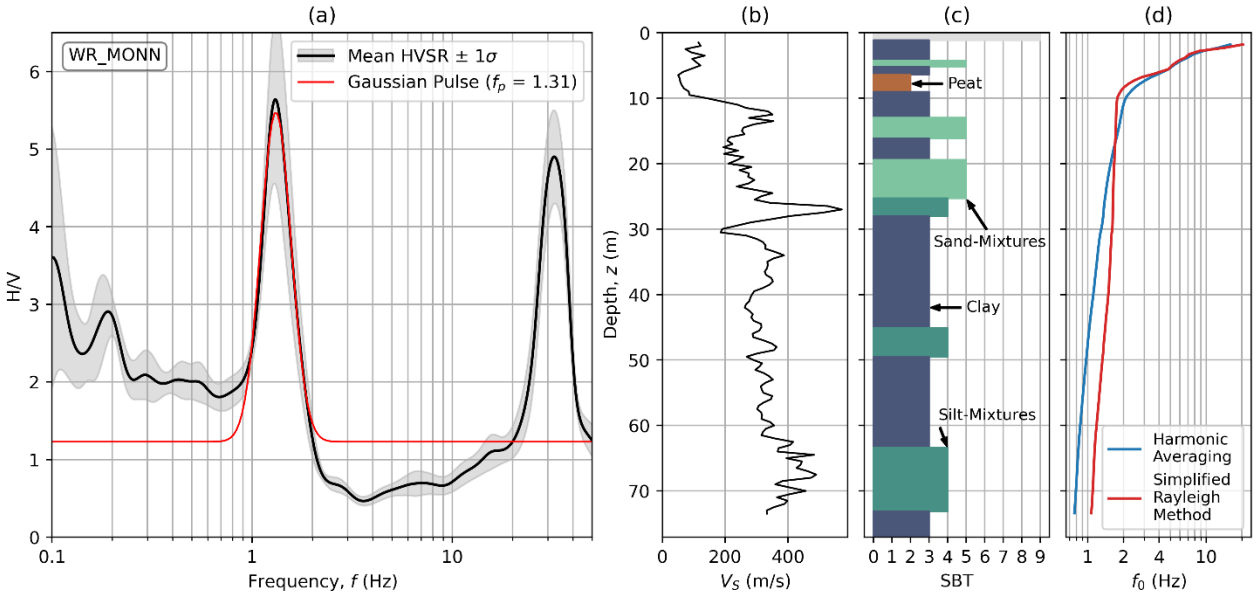


Figure 3.21: Summary of geotechnical data at site WR_MONN: (a) mHVSR and fit peak frequency; (b) V_S profile; (c) soil behavior type index (SBT) interpreted from boring log; and (d) theoretical resonant frequency calculated by harmonic averaging or the simplified Rayleigh method (Dobry et al. 1976) as a function of depth.

Figure 3.21(d) presents the variation of f_0 with depth computed from harmonic averaging and the simplified Rayleigh methods. The trends of both methods are consistent with high-variability observed at shallow layers, reaching a stable f_0 value of about 1 - 2 Hz for soil profiles

with depths greater than about 10 m and a minor trend towards higher frequency as deeper layers are included (i.e., layers deeper than 10 m do not appreciably change the computed site frequency). Values obtained from the harmonic averaging method are consistently lower than those produced from the simplified Rayleigh method at depths beyond about 17 m, which is expected since there is a modest positive gradient in the V_S profile. Furthermore, the f_0 value computed using the total soil profile (73 m) agrees well with the f_p interpreted from mHVSR (i.e., 1.3 Hz ≈ 1.1 Hz). These results suggest that the soil profile above 10 m controls the site resonance, meaning that the peak observed at about 1.3 Hz in the mHVSR for site WR_MONN relates to the soil column including and above the peat layer. These results are observed consistently at other sites with co-located V_S and stratigraphic data, providing evidence that mHVSR peaks at Delta sites can be associated with soft peaty soils and soft non-organic soils.

CHAPTER 4

Subregional Anelastic Path Effects in California

The work presented throughout this chapter has been submitted for publication to the *Bulletin of the Seismological Society of America* (Buckreis et al. 202x).

4.1. Introduction

Ground motion models (GMMs) predict the natural log mean and standard deviation of intensity measures as a function of predictor variables such as magnitude, site-to-source distance, and site parameters. The NGA-West1 project (Power et al. 2008) developed GMMs from global databases such that the scaling of ground motions with predictor variables represented globally-averaged values given the data available at that time. In the NGA-West2 and NGA-Sub projects (Bozorgnia et al. 2014, 2022), GMMs were developed for global application, but with regionalization of certain parameters, including the GMM constant term (shifting the overall ground motion levels up or down relative to the global average); anelastic attenuation coefficients, which control ground motion scaling with distance for distances greater than about 80 km, and coefficients describing the scaling of ground motions with site parameters.

An interesting question for ground motion analysts is what degree of regionalization is justified by data trends and by regional geology. In California, the regional NGA-West2 relations apply state-wide. However, there is past evidence to suggest that refinement within the state is needed to account for significant differences in geologic and tectonic conditions. Chiou et al. (2010) found different rates of magnitude-scaling in central and southern California while

Atkinson and Morrison (2009) found different rates of distance attenuation for these regions (faster in the north), with the division located approximately at San Luis Obispo. Kuehn et al. (2019) extend regionalization to much finer resolution of 30 × 30 km cells. Such fine discretization facilitates evaluation of spatial variations in GMM coefficients, but there are drawbacks, including: (1) there can be few paths passing through cells resulting in unstable coefficients; (2) gridded cell boundaries are not optimized to capture the limits of geologic domains; and (3) the model could be overfitting the limited data.

In this chapter, I utilize the expanded ground motion database (relative to NGA-West2) described in Section 2.4 to evaluate regional variations in anelastic attenuation coefficients and constant terms. Based on geologic maps and prior models for rock-quality factor, I discretize the state into nine domains, which is conceptually similar to a discretization of Europe by Kotha et al. (2020). These domains represent a substantial refinement of path models relative to those used in NGA-West2 (9 vs 1) but are much coarser than those used in Kuehn et al. (2019) (513 cells in CA). The models presented here represent a refinement to the ergodic, region-specific path model for California in Boore et al. (2014; hereafter BSSA14). The results are expected to be useful for regional hazard analyses and non-ergodic site response applications.

Following this introduction, I present background information on the functional form of ergodic path models and prior studies on path regionalization and non-ergodic path modeling. I then present a subregionalization model for California modified from Chiou et al. (2010). I describe data selection protocols used to obtain a California dataset that is a subset of the GMDB described in Section 2.4, and then present data analyses that include residuals calculations. The interpretations of residuals indicate some unique conditions in the Geysers region of California and path biases for different subregions. The results are used to develop subregional path models.

4.2. Background

4.2.1. Ergodic Path Models

Ground motion models (GMMs) such as the NGA-West2 models by Abrahamson et al. (2014; ASK14), Boore et al. (2014; BSSA14), Campbell and Bozorgnia (2014), and Chiou and Youngs (2014) utilize equations to characterize attributes of ground motions. A typical GMM takes the following form:

$$\ln Y = c_0 + F_E + F_P + F_S + \epsilon_n \sigma \quad (4.1)$$

where ln*Y* represents the natural logarithm of a ground motion intensity measure (IM; *PGA* – peak ground acceleration, *PGV* – peak ground velocity, or *PSA* – pseudospectral acceleration at various oscillator periods); c_0 is a constant term that adjusts ground motions up and down irrespective of the effects of independent variables, and F_E, F_P, and F_S represent scaling functions for how ground motion changes with source, path, and site parameters. Error term ϵ_n is a fractional number of standard deviations of ln*Y* away from the mean and σ represents the total standard deviation of the GMM. Typical predictor variables for source, path, and site are moment magnitude (**M**), rupture distance or closest distance to the surface projection of the fault (R_{rup} or R_{JB}, respectively), and 30 m time averaged shear wave velocity (V_{S30}). The focus in this Chapter is on the path model (F_P) and how changes to that model influence the constant term, c_0, both of which can be regionalized.

The path term, F_P, models the decrease of ground motion with distance from the source. The BSSA14 path model is formulated as,

$$F_P(R_{JB}, \mathbf{M}, region) = \left[c_1 + c_2(\mathbf{M} - \mathbf{M}_{ref})\right] ln(R/R_{ref}) + (c_3 + \Delta c_3)(R - R_{ref}) \quad (4.2)$$

where $c_1 - c_3$ are fixed (global) model coefficients, R_{ref}=1 km, $\mathbf{M}_{ref}$ = 4.5, and Δc_3 is a coefficient for regional differences in anelastic attenuation from the global mean. The first term in Equation

(4.2) models geometric spreading and the second models anelastic attenuation, which scales ground motion with arithmetic distance. Anelastic attenuation is related to the energy loss associated with internal friction during wave propagation (e.g., Boore 2003) and introduces downward curvature in log-ground motion vs log (distance) plots for distances greater than about 80 km. In Equation (4.2), the distance *R* term is given as,

$$R = \sqrt{R_{JB}^2 + h^2} \tag{4.3}$$

where *h* is a model coefficient that is introduced to provide near-source saturation.

4.2.2. Prior Regionalization and Non-Ergodic Path Models

Whereas Chiou et al. (2010) and Atkinson and Morrison (2009) demonstrated sub-regional features in California data at a broad scale (i.e., the northern and southern halves of the state), more recent studies have examined certain ground motion features at finer levels of resolution (Landwehr et al. 2016, Kuehn et al. 2019, and Erdem et al. 2019).

Landwehr et al. (2016) produced a GMM for California with spatially varying coefficients using the NGA-West2 dataset. The model depends on typical input parameters in addition to considering site and event coordinates. Spatially dependent and correlated coefficients were derived to develop a varying coefficient model whose values change smoothly in space. Among the spatially variable coefficients are those for geometric spreading and anelastic attenuation (similar to c_1 and c_3 in Equation 4.2). A limitation of this approach is that source-to-site specific path effects are determined solely by the earthquake location.

Kuehn et al. (2019) applied a cell-specific attenuation approach originally proposed by Dawood and Rodriquez-Marek (2013) to California. The state was divided into non-overlapping rectangular cells 30×30 km in size, and for each cell, the anelastic attenuation portion of the path

model (similar to the $(c_3 + \Delta c_3)(R - R_{ref})$ term in Equation 4.2) was replaced with a term of the form (using similar notation to Equation 4.2):

$$f_{attn}(R) = \sum_i^{N_c} c_{3,i} \Delta R_i \quad (4.4)$$

where N_c is the number of cells the source-to-site path crosses, $c_{3,i}$ are gridded anelastic attenuation coefficients for each of the N_c cells, and ΔR_i is the path length within the i^{th} cell using the appropriate distance parameter for the GMM. Kuehn et al. (2019) performed the modification on the ASK14 GMM, which uses the R_{rup} distance parameter, so in their application R in Equation (4.4)was R_{rup} and the gridded attenuation coefficients were $\theta_{17,i}$. The cell-specific coefficients ($\theta_{17,i}$) were derived using a Bayesian hierarchical approach using only data from paths that traversed the cell; if no paths crossed the cell then the regional mean value (for all of California) was assigned. The NGA-West2 database was used to estimate coefficients. This approach separates systematic path effects from event and site effects in areas having sufficient data. Challenges associated with the application of this approach are data limitations that cause cell-specific coefficients to be poorly constrained and substantial increase in the complexity of ground motion calculations due to the large number of cells that need to be considered. Moreover, it could be argued that the use of 513 cell-specific coefficients (or the subset with sufficient data to have regressed coefficients) could be an overfit of the training dataset (Babyak 2004).

Erdem et al. (2019) presented a regional path adjustment to the BSSA14 model that is intended for application in the Sacramento-San Joaquin River Delta region of northern California. Data from 14 Bay Area events were considered, and instead of using the built-in mechanism for accommodating regional variations in attenuation (i.e., the Δc_3 term in Equation 4.2), a third path term was added to Equation (4.2) that is dependent on distance and hypocentral depth. Shallow

earthquakes (hypocentral depths < 9 km) were found to attenuate faster while deeper events exhibited average attenuation when compared to BSSA14.

The effect of each of these models is to reduce the spatial variability of residuals by accounting for regional path effects. These effects indicate that the ground motion residuals are spatially non-stationary, which is also known as a non-ergodic process (e.g., Anderson and Brune 1999). The phrase "non-ergodic" can also be applied to GMMs, which I posit requires that the models transform the underlying non-ergodic process into an ergodic one, by removing spatial trends in the residuals. Whereas a fully non-ergodic site response model can be achieved when sufficient ground motion and geotechnical data are available from the site (Stewart et al., 2017), this non-ergodic standard is more difficult to achieve for path modelling. The subregional and cell-specific models of Kuehn et al. (2019) and Erdem et al. (2019) certainly account for regional variations that are missed by the California-wide model of BSSA14, but it is impossible to know with currently available information if the level of localization in these models fully captures regional variability, thereby rendering residuals as spatially stationary. For this reason, I do not refer to the aforementioned models from literature, nor the models developed subsequently in this chapter, as non-ergodic.

4.3. Subregionalization of California

To construct the initial subregional boundaries for California, I considered a physiographical provinces map for the state by Chiou et al. (2010) and spatial trends found in prior studies of rock quality factor (Q_S). Figure 4.1 shows the physiographical provinces map from Chiou et al. (2010). In coastal areas, this map distinguishes hilly areas co-located with relatively straight portions of the San Andreas plate boundary (coast range), the roughly east-west striking transverse range in

southern California south of the big bend in the San Andreas fault, and the peninsular range in coastal southern California. Major inland mountainous areas include the Klamath mountains, Cascades, and Sierra Nevada. Major basin structures include the central valley, Colorado Desert (which includes Imperial Valley), and the peninsular range (which includes large basins in southern California). It is noteworthy that the provinces in Figure 4.1 are orders of magnitude greater in size than the cells used in Kuehn et al. (2019).

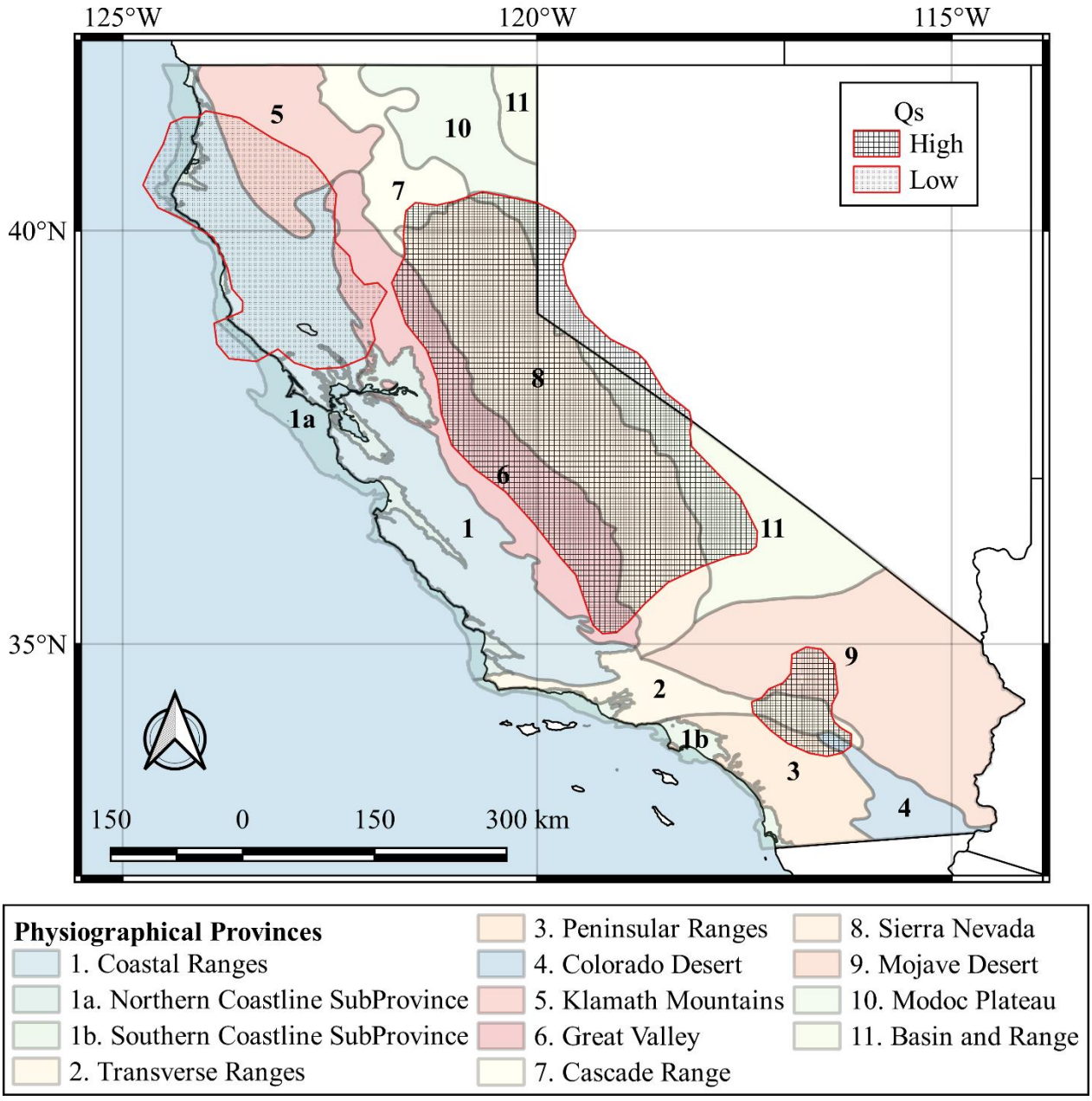


Figure 4.1: California physiographical provinces from Chiou et al. (2010) with zones of relatively low and high Q_S from Hauksson and Shearer (2006) and Eberhart-Phillips (2016).

To check the significance of the physiographical province boundaries, I examined the variability of Q_S across California (Hauksson and Shearer 2006; Eberhart-Phillips 2016). The studies by Hauksson and Shearer (for southern California) and Eberhart-Phillips (for northern

California) used different methods to estimate Q_S, and their estimates do not align perfectly in overlap zones in the central part of the state. Accordingly, I interpret their results to identify areas of relatively high, medium, and low Q_S to a depth of 20 km, which are shown in Figure 4.1. Based on the two studies and as shown in Figure 4.1, there is evidence of low Q_S in the coast range north of the Bay Area, high Q_S in the Sierra Nevada and San Bernardino Mountains, and varying but intermediate levels of Q_S elsewhere. With the exception of the San Bernardino Mountain local zone of high Q_S, these zones corroborate some of the physiographical province boundaries. I conclude that these boundaries provide a reasonable basis for regionalization of anelastic attenuation, but anticipate that further discretization of the coast range may be needed (given the apparently low Q_S in the north coast area) and that some provinces outside of the coast range and Sierra Nevada could potentially be combined. This will be explored subsequently.

The low-Q_S region in the north coast range includes a geothermal field colloquially known as *The Geysers*, which is an area characterized by relatively active seismicity. The Geysers geothermal field is located about 115 km north of San Francisco as shown in Figure 4.2. Geyser events differ from tectonic events in that they are likely induced by geothermal drilling and exhibit low stress drops and shallow focal depths (Johnson and Majer, 2017). Previous studies (e.g., Chiou et al., 2010) excluded data from these events from analysis citing observed lower than average amplitudes. I consider the Geyser events to maximize database breadth and size, but consider potential biases in the ground motions.

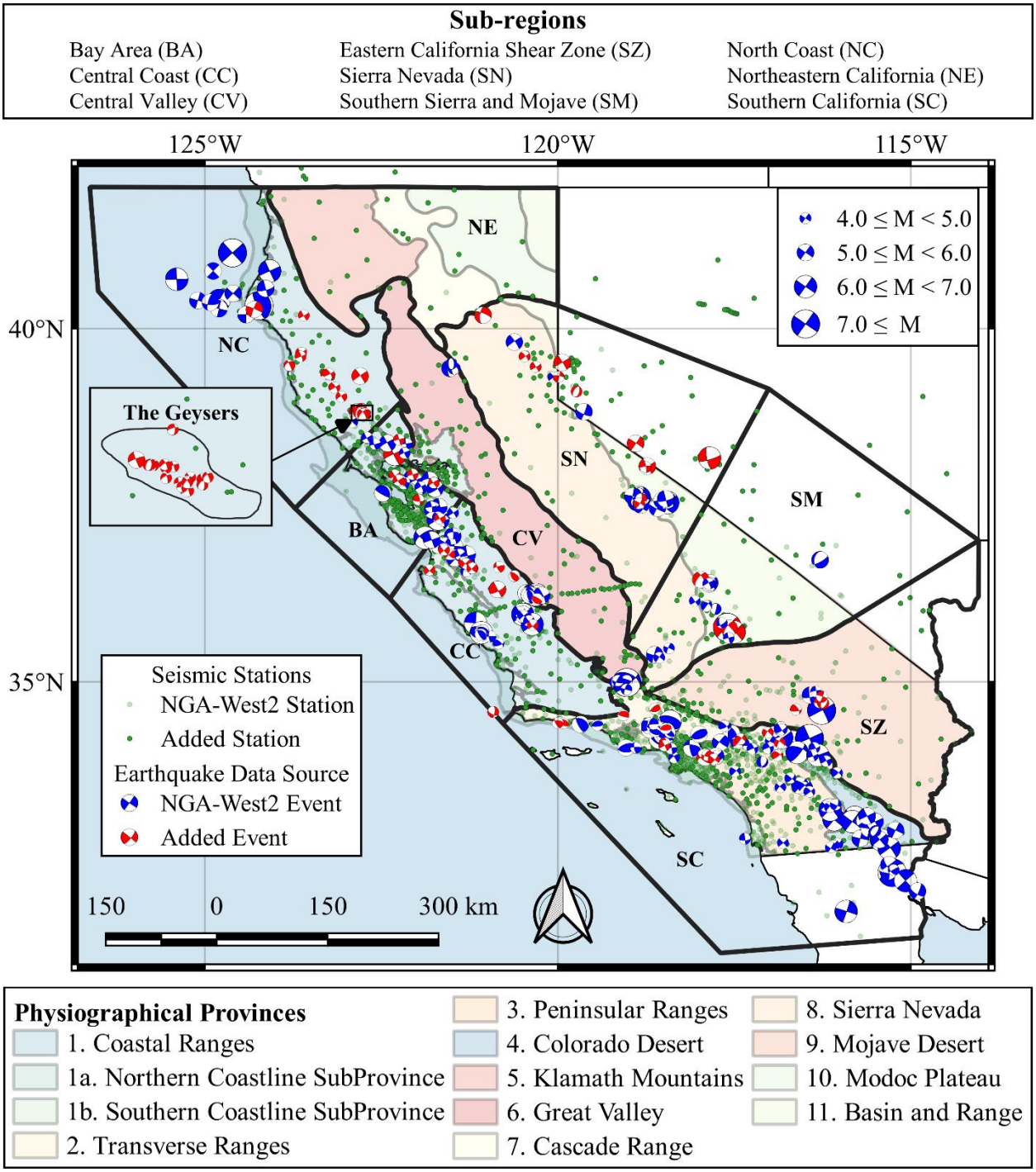


Figure 4.2: Locations of earthquakes used in study; California physiographical provinces and subsequent nine study regions outlined by solid black lines. Inset shows outline of Geysers region from Viegas and Hutchings (2010) and locations of Geysers events.

4.4. Data Resources

The database applied for the present analyses is a subset of the complete GMDB described in Section 2.4. Since spectral shapes have been observed to be sensitive to event magnitude (Stafford et al. 2017), I considered all readily available ground motion data from events with **M** > 4.0 originating in California, western Nevada, and northern Baja, Mexico. Figure 4.2 shows the epicenters of 313 events and locations of 2,519 distinct seismic recording instruments that have produced over 26,209 three-component ground motion records used during analysis.

Data from sites within the Sacramento-San Joaquin Delta, which have unusually strong site effects, are removed to avoid introducing additional bias. Magnitude-distance cutoffs suggested by BSSA14 were enforced for all ground motions resulting in 26,209 usable ground motions within the study region. Horizontal components are combined to median-component (RotD50) intensity measures as defined by Boore (2010), and the lowest usable period of the two individual components defined as $1/1.25f_{hp}$, where f_{hp} represents the high-pass corner frequency used during signal processing, is used. Figure 4.3 compares the parameter ranges covered by the NGA-West2 and newly added data. The added events have nearly doubled the database size, mostly from events with **M** < 5 and relatively distant ($R_{JB} \geq 100$ km) recordings. This expansion of the database is useful for analysis of anelastic path effects, which is the primary objective of this Chapter.

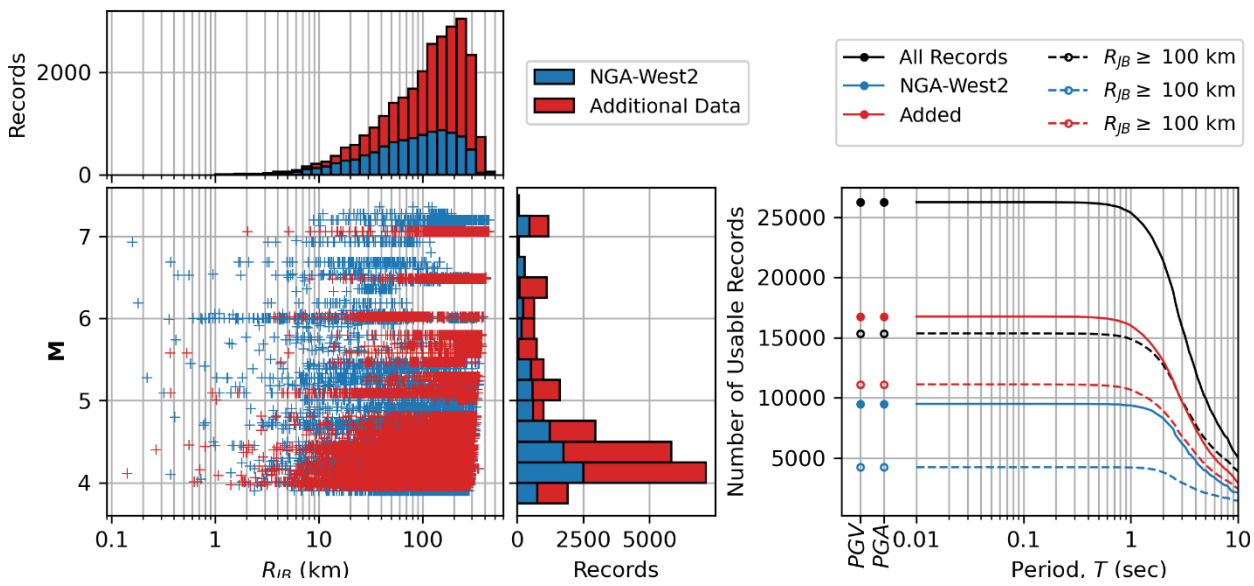


Figure 4.3: Distribution of NGA-West2-subset (blue) and additional (red) datasets used.

4.5. Data Analysis and Model Development

Ground motion analyses were performed to characterize spatial variations of path effects in California and to develop regional models that capture such effects for ground motion and seismic hazards analyses. The approach that was applied consists of the following steps:

1. Screen the database to remove long-period components of records that are beyond their usable oscillator period range.
2. Perform residuals analyses using the original GMM (denoted here GMM0) to evaluate total residuals (R_{ij}) and then partition R_{ij} using a mixed effects approach into model bias terms (c_k), event terms ($\eta_{E,i}$), within-event residuals (δW_{ij}), site terms ($\eta_{S,j}$), and remaining residuals (ϵ_{ij}). These analyses are described, and the meaning of the various terms are provided, in Section 4.5.1 below.

3. Investigate bias of $\eta_{E,i}$ for Geyser events and modify the GMM0 constant term and (if necessary) source model (i.e. $c_0 + F_E$ in Equation 4.1) to capture these effects (revised model denoted GMM1).
4. Re-compute residuals using GMM1. If trends of δW_{ij} with R_{JB} are found, identify a path function to capture those trends. Perform regression to estimate path model coefficients (revised model denoted GMM2). As appropriate, combine regions in Figure 4.1 for which distinct attenuation features are not identified.
5. Re-compute R_{ij} using the GMM2 models and obtain new estimates for $\eta_{E,i}$, δW_{ij}, and $\eta_{S,j}$.
6. Repeat steps (3) and (5) until an acceptable level of convergence is observed on the model coefficients and in random effects $\eta_{E,i}$ and $\eta_{S,j}$.
7. Evaluate model performance through residuals analyses and dispersion calculations. Compare the dispersion of within-event residuals to that achieved with more localized cell-specific models (Kuehn et al. 2019).

4.5.1. Residuals Calculations

Total residuals (R_{ij}) represent differences between data and the median prediction of a GMM,

$$R_{ij} = \ln(Y_{ij}) - \mu_{ij}(\mathbf{M}_i, R_{JB,ij}, V_{S30,j}, z_{1,j}) \quad (4.5)$$

where $\ln(Y_{ij})$ is the natural logarithm of the observed IM at site j from event i and μ_{ij} is the natural log mean from a GMM conditioned on the indicated parameters (BSSA14 is used here). Total residual R_{ij} can be partitioned using mixed-effects analyses (Abrahamson and Youngs 1992, Gelman et al. 2014):

$$R_{ij} = c_k + \eta_{E,i} + \delta W_{ij} \quad (4.6)$$

The within-event residual is further partitioned as:

$$\delta W_{ij} = \eta_{S,j} + \epsilon_{ij} \quad (4.7)$$

Term ϵ_{ij} is the remaining residual after systematic effects from source and site have been removed. Path effects are investigated in this paper using δW_{ij}; the partitioning of this residual in Equation (4.7)is shown because of its subsequent use in dispersion calculations. The mixed-effects analyses are performed in R (Bates et al., 2015) as a subroutine in Python (Gautier, 2009). For brevity, the subscript *k* is omitted from subsequent equations.

As shown subsequently, residuals (both R_{ij} and δW_{ij}) exhibit trends with R_{JB} for multiple regions. When such path model biases are present, they contaminate $\eta_{E,i}$ in the sense that these event terms do not solely reflect source attributes. In order to obtain accurate estimates for $\eta_{E,i}$, which is particularly important for the Geysers analyses, I only consider observations up to a limiting distance within which the trend of residuals with distance are relatively flat. From visual inspection, that limiting distance was found to be approximately 100 km. These distance limits are relaxed after the first several iterations because at that point path biases have been corrected.

4.5.2. Analysis of Geyser Event Terms

The database contains 19 Geyser events that were recorded by many stations over a wide range of distances. These events are known to have low stress parameters (Chiou et al., 2010; Johnson and Majer, 2017), which in the present context translate into strongly negative $\eta_{E,i}$ (average of ~ -1.5), as shown in Figure 4.4. The Geysers event terms have no clear trends with the source parameters of magnitude, hypocentral depth, or rake angle.

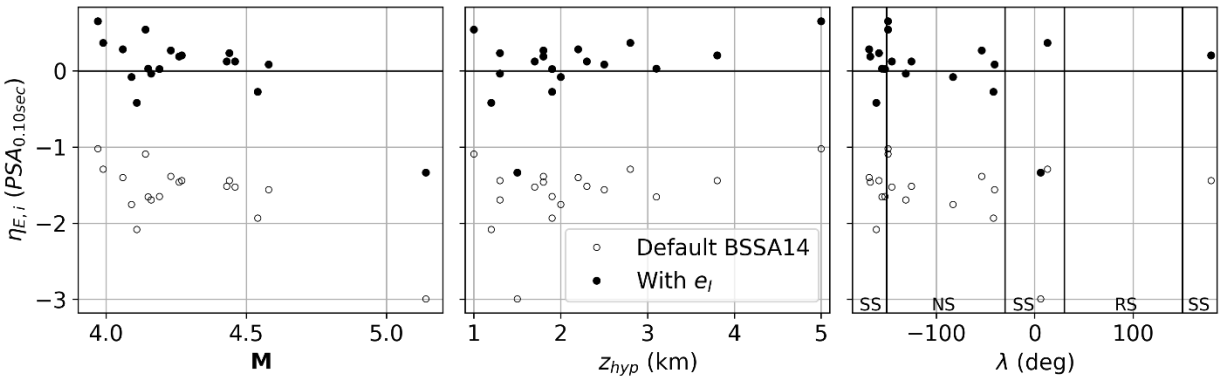


Figure 4.4: Geyser event terms trends against M (left), hypocenter depth (center), and rake angle (right) for the IM of *PSA* at 0.10 sec.

In principle, source-related biases should not affect path-related analyses because event terms are removed in the calculation of δW_{ij} (from which path models are regressed). Nonetheless, I seek to remove these biases so that they do not inflate between-event variability (τ). Due to the lack of trends of Geyser event terms with source attributes (Figure 4.4), the only modification to the original GMM (Equation 4.1) is in the constant term,

$$\ln Y = (c_0 + Ie_I) + F_E + F_P + F_S + \epsilon_n \sigma \tag{4.8}$$

where I represents an indicator variable which takes the value of 1 if the event is induced (i.e., Geyser) and 0 otherwise; and e_I is a constant induced adjustment term.

Using the subset of R_{ij} for Geyser events with R_{JB} < 100 km (to minimize the influence of relatively complex anelastic path effects in this region), I perform mixed-effects regression to estimate e_I and updated event terms ($\eta^*_{E,i}$). Figure 4.5 shows that parameter e_I is negative for most periods thus reducing median ground motions for Geyser events, but for periods T > ~2.5 sec e_I becomes slightly positive. Due to the relatively small sample size for this period range, I constrain $e_I \leq 0$ and apply locally weighted scatterplot smoothing (Cleveland 1979) with the weight

parameter selected by eye. The resulting event terms average to zero (Figure 4.4). Coefficient values for e_I are tabulated in Table 4.1.

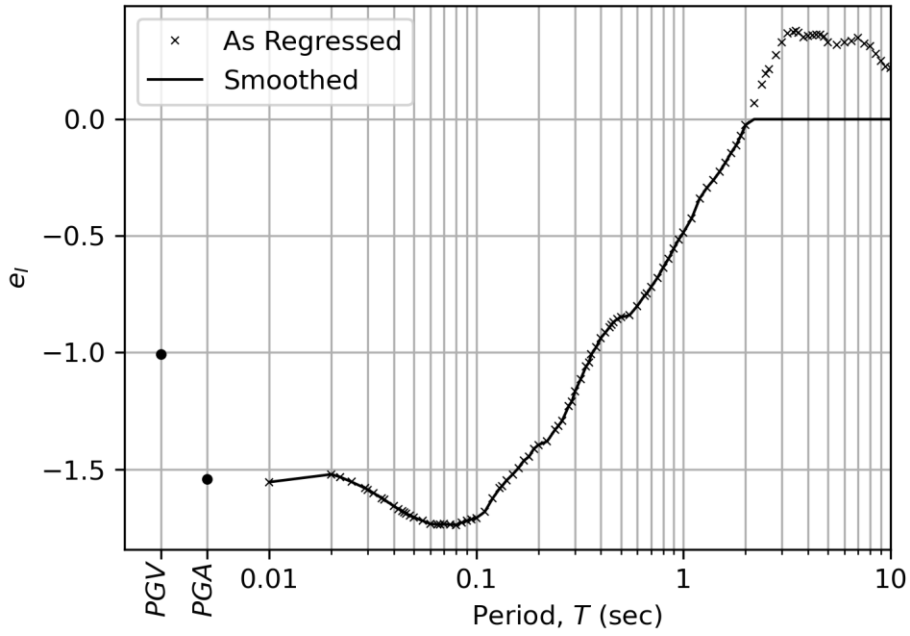


Figure 4.5: Induced source correction term (e_I) variation with oscillator period.

Table 4.1: Coefficient values for e_1.

IM	e_1	IM	e_1	IM	e_1
PGV	-1.00587	T = **0.170 s**	-1.46288	T = **1.300 s**	-0.29434
PGA	-1.54118	T = **0.180 s**	-1.44629	T = **1.400 s**	-0.26058
T = **0.010 s**	-1.55497	T = **0.190 s**	-1.41192	T = **1.500 s**	-0.22454
T = **0.020 s**	-1.52176	T = **0.200 s**	-1.39553	T = **1.600 s**	-0.18558
T = **0.022 s**	-1.53302	T = **0.220 s**	-1.38070	T = **1.700 s**	-0.14523
T = **0.025 s**	-1.55418	T = **0.240 s**	-1.33160	T = **1.800 s**	-0.11247
T = **0.029 s**	-1.58064	T = **0.250 s**	-1.31319	T = **1.900 s**	-0.07033
T = **0.030 s**	-1.58797	T = **0.260 s**	-1.29240	T = **2.000 s**	-0.02449
T = **0.032 s**	-1.60231	T = **0.280 s**	-1.22925	T = **2.200 s**	0.00000
T = **0.035 s**	-1.62408	T = **0.290 s**	-1.20887	T = **2.400 s**	0.00000
T = **0.036 s**	-1.63029	T = **0.300 s**	-1.16721	T = **2.500 s**	0.00000
T = **0.040 s**	-1.65806	T = **0.320 s**	-1.11291	T = **2.600 s**	0.00000
T = **0.042 s**	-1.66991	T = **0.340 s**	-1.05849	T = **2.800 s**	0.00000
T = **0.044 s**	-1.67940	T = **0.350 s**	-1.04339	T = **3.000 s**	0.00000
T = **0.045 s**	-1.68509	T = **0.360 s**	-1.00809	T = **3.200 s**	0.00000
T = **0.046 s**	-1.69156	T = **0.380 s**	-0.97669	T = **3.400 s**	0.00000
T = **0.048 s**	-1.69806	T = **0.400 s**	-0.93761	T = **3.500 s**	0.00000
T = **0.050 s**	-1.70557	T = **0.420 s**	-0.91489	T = **3.600 s**	0.00000
T = **0.055 s**	-1.72066	T = **0.440 s**	-0.89278	T = **3.800 s**	0.00000
T = **0.060 s**	-1.73401	T = **0.450 s**	-0.88045	T = **4.000 s**	0.00000
T = **0.065 s**	-1.73638	T = **0.460 s**	-0.86938	T = **4.200 s**	0.00000
T = **0.067 s**	-1.73708	T = **0.480 s**	-0.85690	T = **4.400 s**	0.00000
T = **0.070 s**	-1.73502	T = **0.500 s**	-0.84871	T = **4.600 s**	0.00000
T = **0.075 s**	-1.73719	T = **0.550 s**	-0.84023	T = **4.800 s**	0.00000
T = **0.080 s**	-1.73847	T = **0.600 s**	-0.80034	T = **5.000 s**	0.00000
T = **0.085 s**	-1.72901	T = **0.650 s**	-0.75637	T = **5.500 s**	0.00000
T = **0.090 s**	-1.72181	T = **0.667 s**	-0.74704	T = **6.000 s**	0.00000
T = **0.095 s**	-1.71417	T = **0.700 s**	-0.71868	T = **6.500 s**	0.00000
T = **0.100 s**	-1.70934	T = **0.750 s**	-0.68039	T = **7.000 s**	0.00000
T = **0.110 s**	-1.68244	T = **0.800 s**	-0.63555	T = **7.500 s**	0.00000
T = **0.120 s**	-1.62517	T = **0.850 s**	-0.59730	T = **8.000 s**	0.00000
T = **0.130 s**	-1.58168	T = **0.900 s**	-0.55478	T = **8.500 s**	0.00000
T = **0.133 s**	-1.57314	T = **0.950 s**	-0.51648	T = **9.000 s**	0.00000
T = **0.140 s**	-1.54847	T = **1.000 s**	-0.48678	T = **9.500 s**	0.00000
T = **0.150 s**	-1.52205	T = **1.100 s**	-0.42390	T = **10.000 s**	0.00000
T = **0.160 s**	-1.49473	T = **1.200 s**	-0.33964		

4.5.3. Path Biases

I begin by examining in Figure 4.6 the performance of the BSSA14 regional California model (with $\Delta c_3 = 0$) against the full dataset, which is much larger than the California dataset used in the derivation of BSSA14. Figure 4.6 shows flat trends of residual δW_{ij} with distance for **M** > 5 events. A modest downward trend is present for **M** < 5 events for large distances (R_{JB}>80 km) where anelastic effects are strong; that trend is also visible for all events.

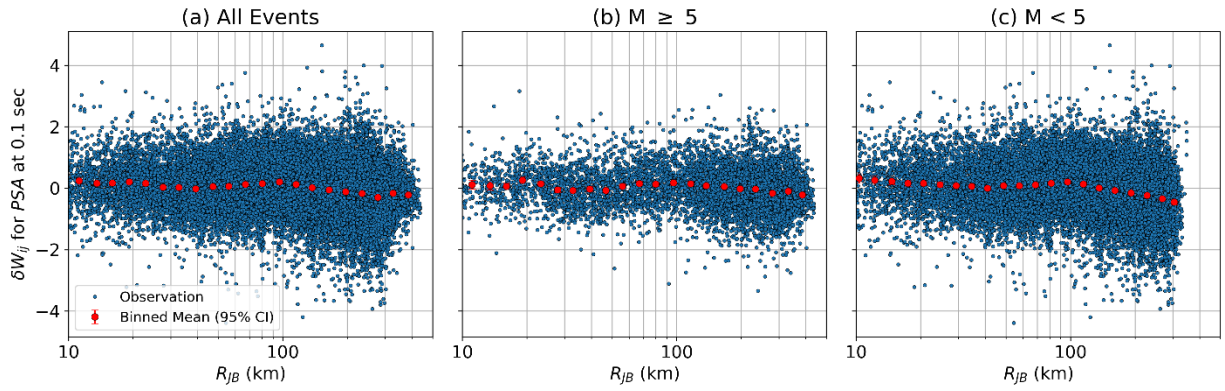


Figure 4.6: Trends of within-event residual (δW_{ij}) for *PSA* at 0.10 sec with distance using (a) full California dataset and subsets of earthquakes with (b) **M** > 5 and (c) **M** < 5.

Figure 4.7 shows subsets of the data for six sub-regions, which are assigned based on epicenter location: North Coast, Bay Area, Central Coast, Sierra-Nevada, Southern Sierra and Mojave, and Southern California (sub-region boundaries are shown in Figure 4.2). The sub-region plots shade individual residuals according to the fraction of the path within the epicenter sub-region (darker indicating higher fractions). Binned means are shown as is a fit line discussed further below. These plots demonstrate several notable features:

1. The North Coast region has high residuals at close distance (R_{JB}<10 km). These biases are caused by Geyser events and are characteristic of bias in the finite fault term *h* (Equation 4.3), which in this case is too large. The concern in this Chapter is anelastic effects at larger distances, hence I do not modify GMM *h* terms.
2. For R_{JB}>60 km, trends in residuals are downward for the North Coast subregion, upward for Sierra Nevada and Southern Sierra and Mojave subregions, and horizontal for the Bay Area, Central Coast, and Southern California subregions. Downward and upward curvature requires negative and positive Δc_3, respectively.
3. The different attenuation trends for North Coast and other coast range areas justify subdivision of the coast range physiographical province (Figure 4.1). Bay Area and central coast groups are also considered due to different geological conditions (Wills et al. 2015).
4. Variable levels of shading of the data symbols in Figure 4.7 show that attenuation trends for events in one subregion are often controlled by neighboring subregions. This feature of the data creates a need to consider multiple subregions in the path model, which is addressed in the next section.

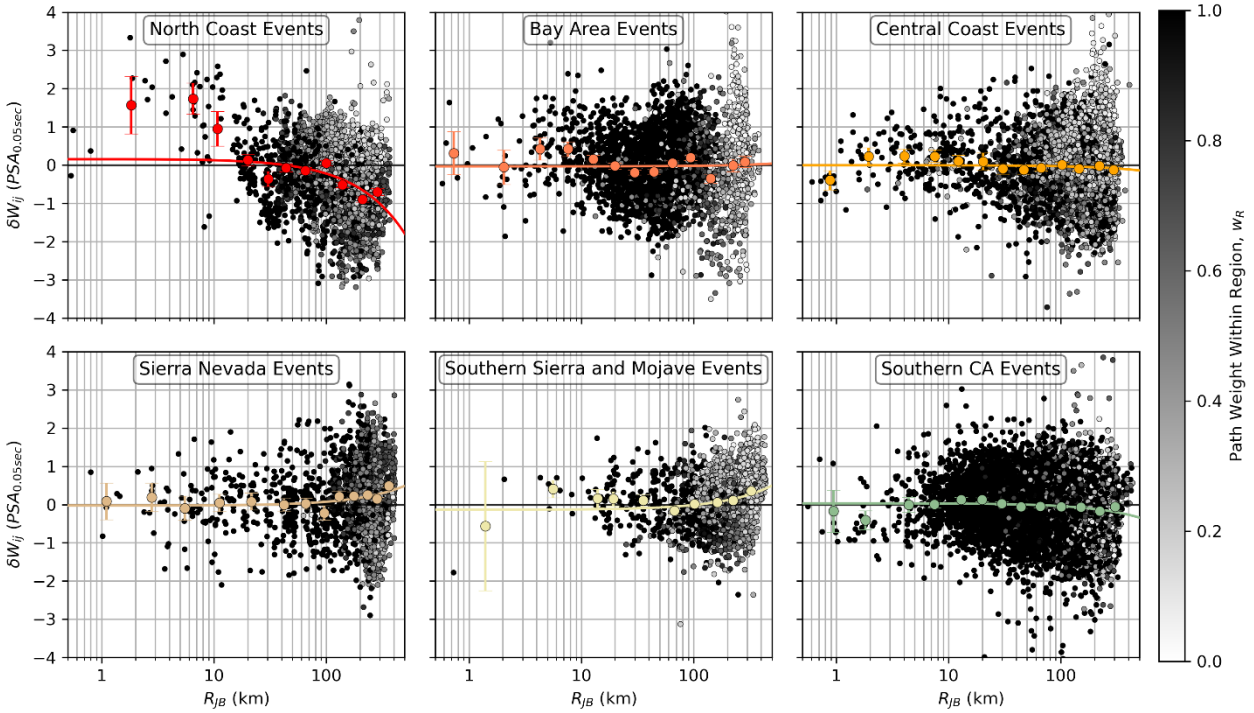


Figure 4.7: Trend of within-event residuals (δW_{ij}) against Joyner-Boore distance (R_{JB}) for *PSA* at 0.05 sec grouped by region of event origination; data points shaded relative to the path weight within the region of event origination; results shown for the first iteration.

4.5.4. Path Model Development

The development of path models has three major steps: (*i*) identify suitable subregions for modeling purposes; (*ii*) select functions to capture path effects not accounted for with the BSSA14 path model(Equation 4.2); and (*iii*) regress model coefficients using the iterative procedure outlined at the beginning of this section. As part of these analyses, I also consider differences between Geyser and non-Geyser events in the North Coast subregion.

(a) Subregion Identification

Plots of δW_{ij} with distance as in Figure 4.7 are helpful to visualize differences in attenuation trends between subregions. However, I sought a more quantitative basis for evaluating attenuation

differences between subregions. This was accomplished by performing event-specific regressions as follows:

$$\delta W_{ij} = \Delta c_{3,i}\left(R - R_{ref}\right) + \Delta c_{0,i} \tag{4.9}$$

where $\Delta c_{3,i}$ is an adjustment to Δc_3 for event i and $\Delta c_{0,i}$ represents a constant term which allows the function to shift vertically to fit the data. Results are plotted in Figure 4.8 for *PSA* at 0.1 s by coloring the focal mechanism diagram according to the $\Delta c_{3,i}$ scale bar. Major findings of this exercise are can be summarized as follows:

- The northern portion of the coast range has more negative $\Delta c_{3,i}$ values for short-period IMs than any other subregion, including other portions of the coast range. The attenuation bias decreases towards the south, being small for the central coast and having mixed but intermediate results in the Bay Area. These distinctions motivated the subdivision of the coast range shown in Figures 4.2 and 4.8. The straight-line boundaries, drawn approximately perpendicular to the San Andreas Fault, are based on interpretation of attenuation results for many IMs and was aided by surficial geologic maps (Wills et al. 2015).
- While events in the Sierra Nevada and Basin and Range physiographical provinces both produced positive $\Delta c_{3,i}$ values, which can be attributed to the relative integrity of rock in the Sierra batholith, the attenuation rates were slightly slower for northern events than those in the southern Sierra and adjoining basin and range. This led to the use of two sub-regions: *Sierra Nevada* (SN) and *southern Sierra and Mojave* (SM).
- Several physiographical provinces were grouped for practical reasons. The transverse ranges, southern coastline sub-province, peninsular ranges, and Colorado Desert were combined into the *Southern California* (SC) subregion due in part to their relatively small

size (providing few within-region paths of sufficient length). The Klamath Mountains, Cascade Range, Modoc plateau, and northern-most portion of the Basin and Range provinces were grouped into the *northeastern California* (NE) subregion, mainly because of sparse data in these areas.

- Physiographical provinces retained as subregions are the Mojave Desert province, denoted here the *eastern California Shear Zone* (SZ), and the *Central Valley* (CV).

It follows that this discretization of California results in nine subregions (Figures 4.2 and 4.8).

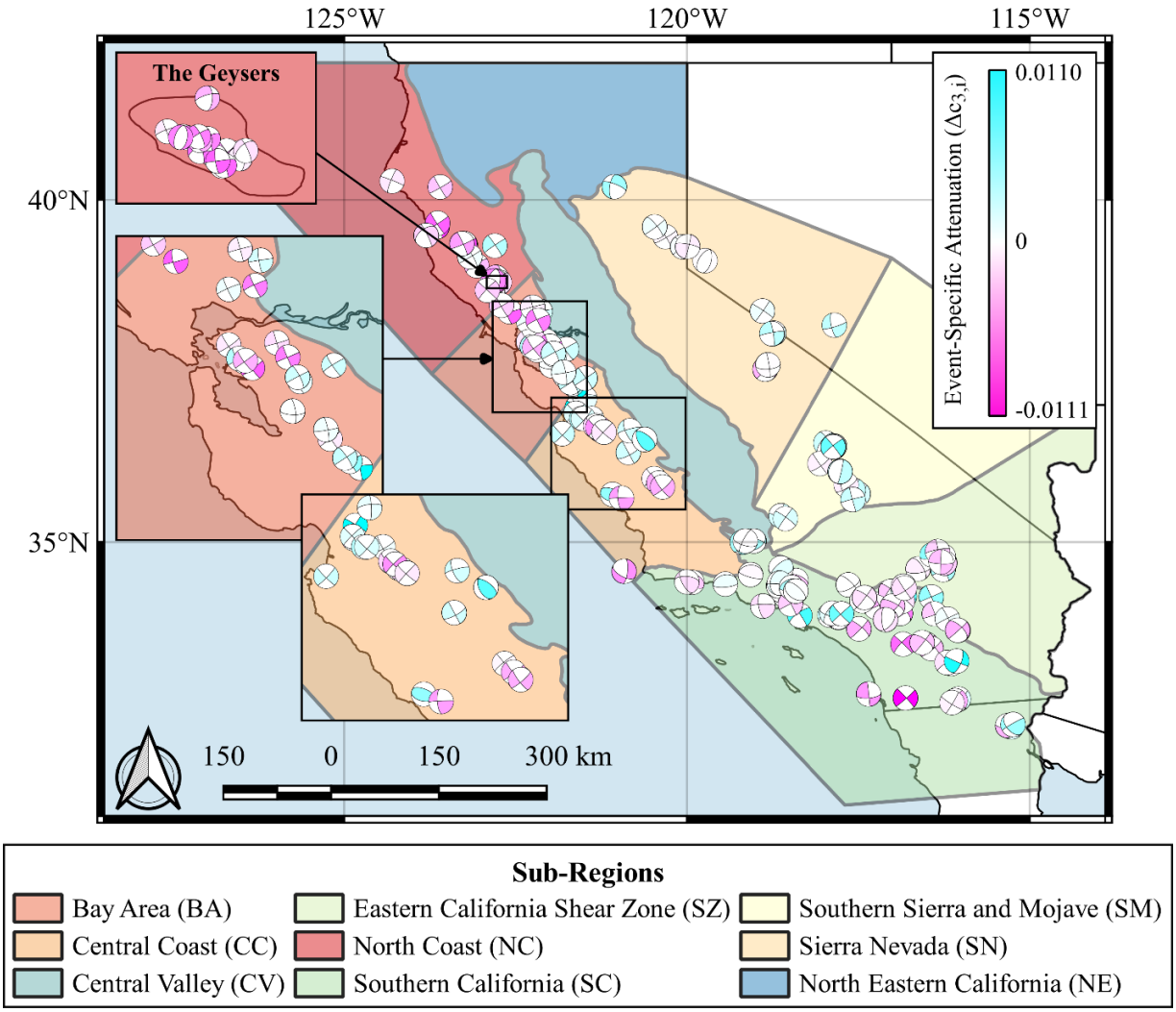


Figure 4.8: Spatial distribution of event-specific attenuation coefficients ($\Delta c_{3,i}$) for *PSA* at 0.10 s.

(b) Anelastic Attenuation Models

The general functional form for modeling departures of subregional anelastic attenuation from the regional mean is similar to Equation (4.9), but without event indices on coefficients,

$$\delta W_{ij} = \Delta c_3^* \left(R_{ij} - R_{ref}\right) + \Delta c_{0,r} \tag{4.10}$$

For application purposes, the Δc_3^* coefficients in Equation (4.10)are applied in the path model as,

$$F_P\left(R_{JB}, \mathbf{M}, region\right) = \left[c_1 + c_2\left(\mathbf{M} - \mathbf{M}_{ref}\right)\right] ln\left(R/R_{ref}\right) + (c_3 + \Delta c_3^*)(R - R_{ref}) \tag{4.11}$$

The $\Delta c_{0,r}$ coefficients are applied in the overall GMM expression in Equation (4.8) as,

$$\ln Y = \left(c_0 + \Delta c_{0,r} + Ie_I\right) + F_E + F_P + F_S + \epsilon_n \sigma \tag{4.12}$$

In the *subregional path model*, parameter Δc_3^* is computed for a particular source-site path as,

$$\Delta c_3^* = \sum_{r=1}^{9} \Delta c_{3r} W_r \tag{4.13}$$

where Δc_{3r} comprise subregional anelastic coefficients (for r=1:9), W_r indicates the proportion of the total path which traverses subregion r, and $\sum_{r=1}^{9} W_r = 1$. Weights W_r are computed from a linear source-to-site path drawn from the closest point on the surface projection of the fault to the site; using that line the proportion of the path in each subregion is computed, as illustrated in Figure 4.9. Parameter $\Delta c_{0,r}$ is taken as the value for the subregion where the event originates.

An alternate form of the path model was also considered in which Equations (4.10)-(4.12) are retained but Δc_3^* is based only on source location. This *source location* approach is simpler to apply than the subregional path model because fractions of source-site paths need not be computed, but has a weaker physical basis because path effects should depend upon crustal attributes along the path, not just those at the hypocenter. Moreover, this alternate form cannot be developed for subregions with few if any events in the database (central valley and northeastern California).

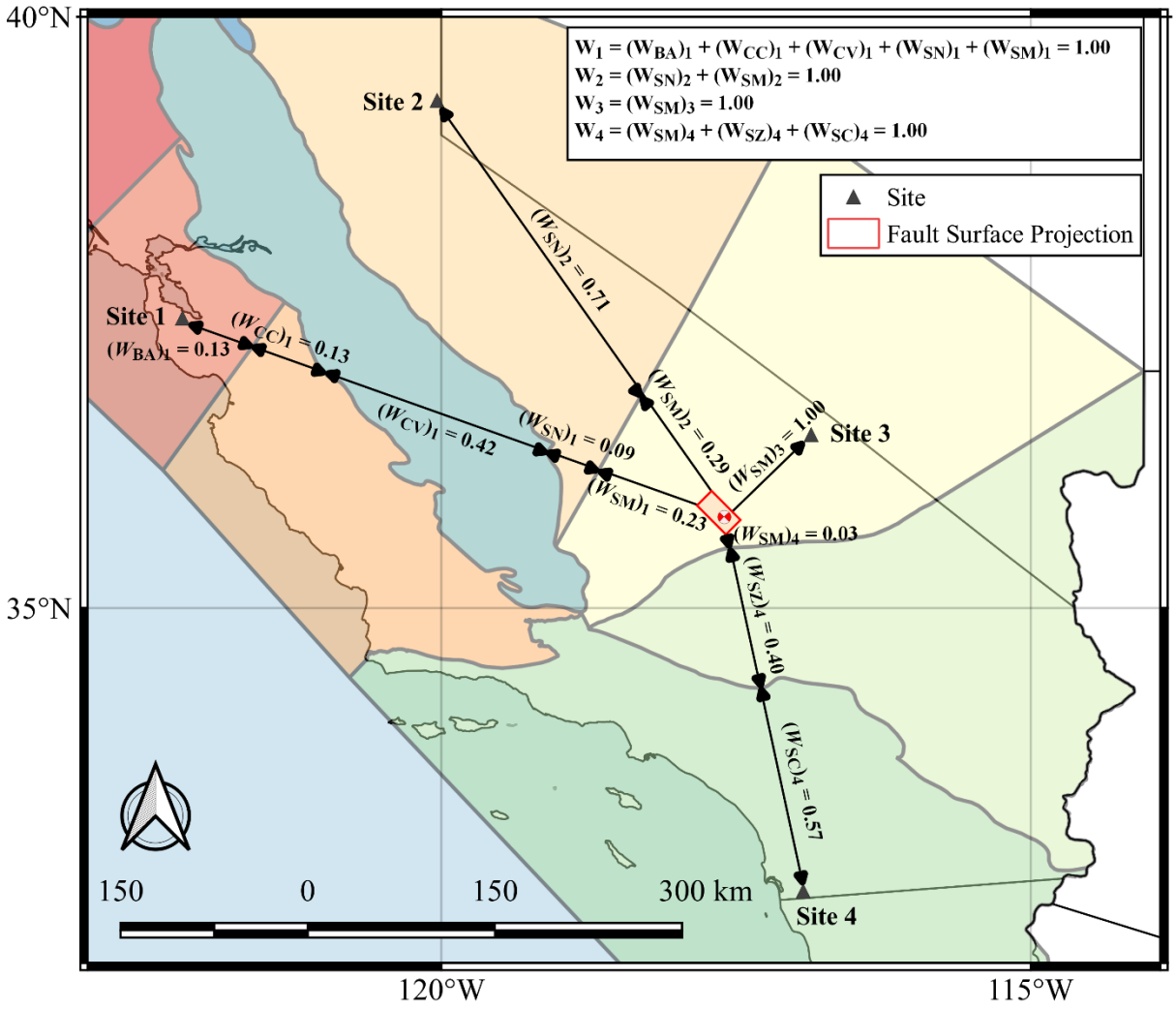


Figure 4.9: Illustration of how to compute subregion weights (W_r) for paths between four sites and a single event.

(c) Regression of Coefficients

I develop subregional constant terms ($\Delta c_{0,r}$) and attenuation coefficients (Δc_{3r}) for *PGA*, *PGV*, and *PSA*s for 105 oscillator periods between 0.01 and 10 sec for both path models using damped least-squares (Levenberg-Marquardt algorithm; More 1977) regression. To estimate W_r, for the subregional path model regressions, I approximated the source-to-site path as a ray between the

earthquake epicenter and station coordinates, which is nearly equivalent to the ray between the station coordinates and closest surface projection of the fault (used in calculating R_{JB}) for the small **M** events that dominate the dataset.

The iterative procedure described at the beginning of this section updates coefficients after each iteration:

$$P = P_1 + \sum_{n=2}^{N} (dP)_n \tag{4.14}$$

where P is the coefficient of interest (e.g., $\Delta c_{3,r}$ or $\Delta c_{0,r}$), $(dP)_n$ is the difference between the n^{th}+1 and n^{th} iterations, P_1 is the initial estimate from the first iteration, and N is the number of iterations required for coefficient estimates to become stable out to the seventh decimal point (typically N=5 provides good results). Example convergence plots for $\Delta c_{3,r}$ and $\Delta c_{0,r}$ are presented in Figure 4.10.

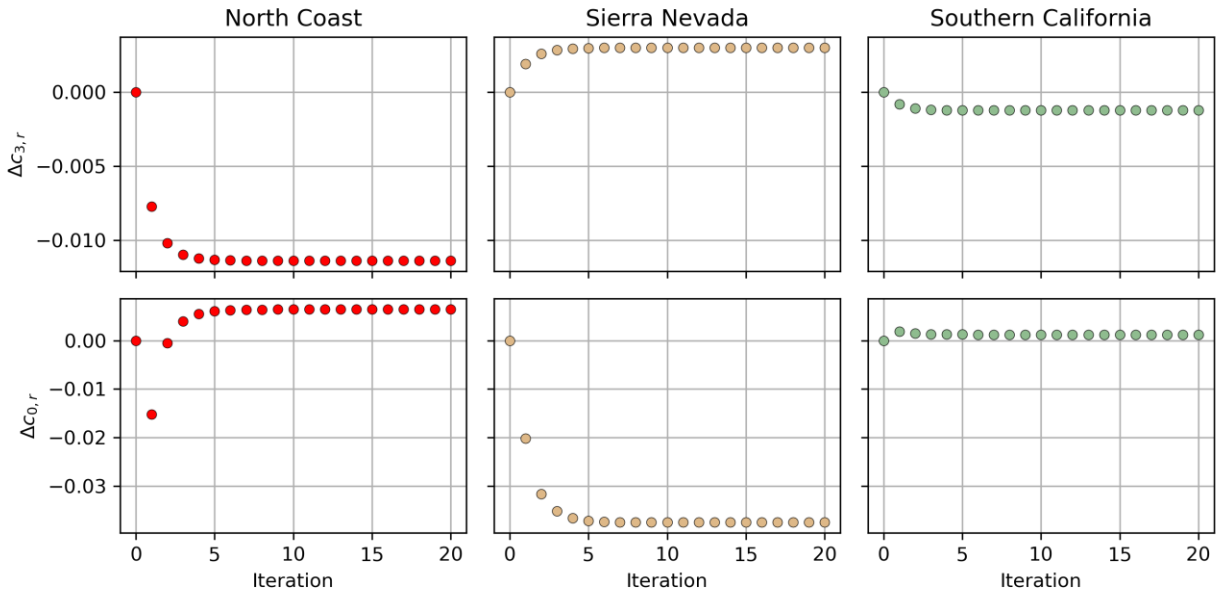


Figure 4.10: Convergence of $\Delta c_{3,r}$ (top) and $\Delta c_{0,r}$ (bottom) coefficients for north coast (left), Sierra Nevada (center), and southern California (right) subregions for *PSA* at 0.10 sec.

Figure 4.11(c) shows that the constant term coefficients ($\Delta c_{0,r}$) produced by the regressions are close to zero for all regions. These coefficients, which are needed to fit the data (per Equation 4.10), do not indicate the relative strengths of ground motions in sub-regions (e.g., from high or low stress parameters), which would be better assessed from event term statistics and are not the subject of this chapter. The small values of $\Delta c_{0,r}$ are convenient for forward applications, because they imply that the anelastic adjustments can be made without other changes to the GMM.

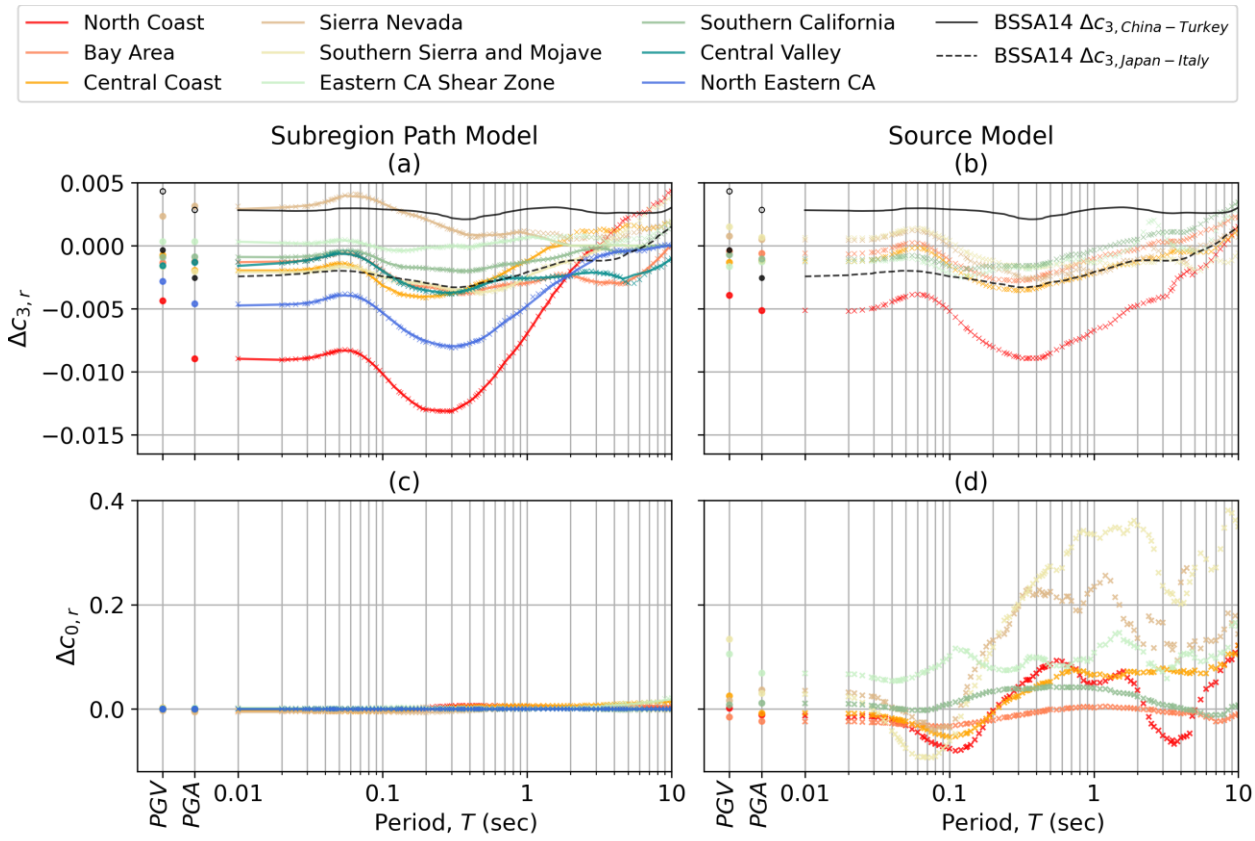


Figure 4.11: Regressed model coefficients for preferred subregion path model (left) and source model (right). Parts (a) and (b) are regional attenuation coefficient, $\Delta c_{3,r}$; Parts (c) and (d) are for constant term adjustments ($\Delta c_{0,r}$).

Figure 4.11(a) shows that the attenuation coefficients ($\Delta c_{3,r}$) are strongly period-dependent, being close to zero at long periods but exhibiting a wide range of variations from zero at short periods. As shown previously in Figures 4.7 and 4.8, the North Coast subregion has the strongest fast attenuation effect while the Sierra Nevada subregion has the strongest slow attenuation effect. Other regions with faster-than-average attenuation at short periods are Northeastern California, Bay Area, and Central Valley. Southern Sierra and Mohave is the other region with slower attenuation, whereas Southern California, Central Coast, and Eastern California shear zone have nearly zero values of $\Delta c_{3,r}$. For comparison, values of Δc_3 for fast- and slow-attenuation global regions are shown (Japan and Turkey, respectively), which are similar in magnitude to some subregional California corrections. In both Figure 4.11(a,c), regressed values are shown with discrete symbols, while the lines indicate smoothed representations intended for modeling purposes. The smoothed lines trend towards zero at long periods and in no case is the sum $c_3 + \Delta c_3$ positive. Subregion coefficients for $\Delta c_{0,r}$ and $\Delta c_{3,r}$ are tabulated in Tables 4.2 and 4.3.

Figure 4.11 (b,d) show regressed coefficients for the source-based model. This model is not recommended for application, but is shown mainly as a point of comparison to illustrate the robustness of regional variations, even when a different modeling framework is used. Constant term adjustments (Figure 4.11d) are much larger in this case, particularly at mid-periods. This is a result of tradeoffs between constant and slope terms in the regressions that produces non-physical results when a single attenuation coefficient is used for all recordings produced by events in a single subregion. The attenuation coefficients (Figure 4.11b) have qualitatively similar trends but are generally closer to zero than for the recommended model (Figure 4.11a).

Table 4.2: Coefficient values for $\Delta c_{3,r}$.

$\Delta c_{3,r}$									
Intensity Measure	**North Coast**	**Bay Area**	**Central Coast**	**Sierra Nevada**	**Southern Sierra and Mojave**	**Eastern CA Shear Zone**	**Southern California**	**Central Valley**	**Northeastern California**
PGV	-0.0043520	-0.0013296	-0.0007512	0.0023554	-0.0016353	0.0003446	-0.0008945	-0.0015673	-0.0028134
PGA	-0.0089439	-0.0012440	-0.0019318	0.0031416	-0.0020302	0.0003259	-0.0008783	-0.0012976	-0.0046088
***T* = 0.010 s**	-0.0089725	-0.0013082	-0.0019623	0.0029033	-0.0021027	0.0003136	-0.0008951	-0.0016015	-0.0047444
***T* = 0.020 s**	-0.0090577	-0.0012617	-0.0019683	0.0030308	-0.0020994	0.0002204	-0.0009151	-0.0013839	-0.0046733
***T* = 0.022 s**	-0.0090371	-0.0012554	-0.0019754	0.0030590	-0.0020981	0.0002038	-0.0009111	-0.0013382	-0.0046587
***T* = 0.025 s**	-0.0090039	-0.0012366	-0.0019521	0.0031037	-0.0020961	0.0001864	-0.0008910	-0.0012688	-0.0046375
***T* = 0.029 s**	-0.0089442	-0.0011654	-0.0019092	0.0031698	-0.0020700	0.0001669	-0.0008622	-0.0011729	-0.0045659
***T* = 0.030 s**	-0.0089314	-0.0011480	-0.0018900	0.0031865	-0.0020613	0.0001718	-0.0008482	-0.0011476	-0.0045436
***T* = 0.032 s**	-0.0088767	-0.0011010	-0.0018539	0.0032190	-0.0020445	0.0001798	-0.0008205	-0.0010942	-0.0044990
***T* = 0.035 s**	-0.0087800	-0.0010154	-0.0017810	0.0032956	-0.0019978	0.0002102	-0.0007631	-0.0010048	-0.0044058
***T* = 0.036 s**	-0.0087454	-0.0009831	-0.0017544	0.0033305	-0.0019750	0.0002230	-0.0007408	-0.0009763	-0.0043678
***T* = 0.040 s**	-0.0086056	-0.0008551	-0.0016424	0.0034693	-0.0018849	0.0002808	-0.0006455	-0.0008575	-0.0042201
***T* = 0.042 s**	-0.0085399	-0.0007918	-0.0015855	0.0035400	-0.0018386	0.0003122	-0.0005960	-0.0007966	-0.0041444
***T* = 0.044 s**	-0.0084763	-0.0007323	-0.0015351	0.0036165	-0.0017881	0.0003472	-0.0005486	-0.0007383	-0.0040730
***T* = 0.045 s**	-0.0084463	-0.0007028	-0.0015129	0.0036542	-0.0017629	0.0003647	-0.0005254	-0.0007116	-0.0040384
***T* = 0.046 s**	-0.0084166	-0.0006781	-0.0014906	0.0036911	-0.0017382	0.0003821	-0.0005046	-0.0006829	-0.0040089
***T* = 0.048 s**	-0.0083654	-0.0006373	-0.0014524	0.0037619	-0.0016953	0.0004167	-0.0004631	-0.0006525	-0.0039624
***T* = 0.050 s**	-0.0083158	-0.0006179	-0.0014416	0.0038292	-0.0016694	0.0004513	-0.0004289	-0.0006256	-0.0039296
***T* = 0.055 s**	-0.0083265	-0.0005979	-0.0014140	0.0039738	-0.0017141	0.0004099	-0.0003968	-0.0006601	-0.0039489
***T* = 0.060 s**	-0.0083236	-0.0006509	-0.0015522	0.0039259	-0.0017295	0.0003762	-0.0004002	-0.0006969	-0.0039713
***T* = 0.065 s**	-0.0084686	-0.0007483	-0.0016722	0.0039044	-0.0017492	0.0003124	-0.0004398	-0.0008095	-0.0040220
***T* = 0.067 s**	-0.0085506	-0.0007939	-0.0017399	0.0038867	-0.0017784	0.0002808	-0.0004639	-0.0008344	-0.0040718
***T* = 0.070 s**	-0.0086136	-0.0008815	-0.0018362	0.0038750	-0.0018241	0.0002284	-0.0005152	-0.0009177	-0.0041528
***T* = 0.075 s**	-0.0088506	-0.0010441	-0.0020546	0.0038536	-0.0019240	0.0001408	-0.0006100	-0.0010735	-0.0043161

$\Delta c_{3,r}$									
Intensity Measure	**North Coast**	**Bay Area**	**Central Coast**	**Sierra Nevada**	**Southern Sierra and Mojave**	**Eastern CA Shear Zone**	**Southern California**	**Central Valley**	**Northeastern California**
***T* = 0.080 s**	-0.0091017	-0.0012223	-0.0022774	0.0037816	-0.0020366	0.0000551	-0.0007147	-0.0012537	-0.0045074
***T* = 0.085 s**	-0.0093652	-0.0014126	-0.0025035	0.0036829	-0.0021494	-0.0000277	-0.0008325	-0.0014501	-0.0047091
***T* = 0.090 s**	-0.0096319	-0.0015992	-0.0027273	0.0035783	-0.0022623	-0.0001078	-0.0009577	-0.0016441	-0.0049098
***T* = 0.095 s**	-0.0099056	-0.0017816	-0.0029498	0.0034674	-0.0023746	-0.0001625	-0.0010775	-0.0018254	-0.0051134
***T* = 0.100 s**	-0.0101687	-0.0019562	-0.0031709	0.0033570	-0.0024833	-0.0002203	-0.0011909	-0.0019792	-0.0053163
***T* = 0.110 s**	-0.0106865	-0.0022661	-0.0033701	0.0031543	-0.0025785	-0.0003256	-0.0013997	-0.0022965	-0.0056852
***T* = 0.120 s**	-0.0111772	-0.0025268	-0.0035739	0.0029602	-0.0026685	-0.0003548	-0.0015106	-0.0025598	-0.0060358
***T* = 0.130 s**	-0.0115260	-0.0027519	-0.0038198	0.0028032	-0.0027543	-0.0003754	-0.0016150	-0.0027801	-0.0063469
***T* = 0.133 s**	-0.0116228	-0.0028188	-0.0038404	0.0027527	-0.0027951	-0.0003816	-0.0016546	-0.0028316	-0.0064436
***T* = 0.140 s**	-0.0118747	-0.0029386	-0.0038835	0.0026644	-0.0028707	-0.0003970	-0.0016905	-0.0029646	-0.0066310
***T* = 0.150 s**	-0.0121446	-0.0030600	-0.0039349	0.0025580	-0.0029209	-0.0003486	-0.0017110	-0.0031192	-0.0068563
***T* = 0.160 s**	-0.0124162	-0.0031336	-0.0039815	0.0024587	-0.0029358	-0.0003234	-0.0017246	-0.0032340	-0.0070485
***T* = 0.170 s**	-0.0125828	-0.0031696	-0.0040234	0.0023649	-0.0029577	-0.0002892	-0.0017341	-0.0033492	-0.0072026
***T* = 0.180 s**	-0.0127737	-0.0032184	-0.0040640	0.0022703	-0.0029842	-0.0002378	-0.0017508	-0.0034246	-0.0073508
***T* = 0.190 s**	-0.0129521	-0.0032416	-0.0040343	0.0021780	-0.0030184	-0.0001878	-0.0017684	-0.0034820	-0.0074708
***T* = 0.200 s**	-0.0129841	-0.0032656	-0.0040335	0.0020855	-0.0030584	-0.0001390	-0.0017893	-0.0035406	-0.0075705
***T* = 0.220 s**	-0.0130433	-0.0033197	-0.0040207	0.0018986	-0.0031527	-0.0001408	-0.0018276	-0.0036217	-0.0077174
***T* = 0.240 s**	-0.0130815	-0.0033889	-0.0039762	0.0017171	-0.0032509	-0.0001300	-0.0018605	-0.0036865	-0.0078351
***T* = 0.250 s**	-0.0130964	-0.0034380	-0.0039428	0.0016372	-0.0032980	-0.0001179	-0.0018753	-0.0037106	-0.0078754
***T* = 0.260 s**	-0.0131082	-0.0034854	-0.0039055	0.0015563	-0.0033444	-0.0000989	-0.0018894	-0.0037292	-0.0079173
***T* = 0.280 s**	-0.0131177	-0.0035852	-0.0038353	0.0014137	-0.0034177	-0.0000605	-0.0019188	-0.0037356	-0.0079674
***T* = 0.290 s**	-0.0131232	-0.0036291	-0.0037989	0.0013502	-0.0034447	-0.0000431	-0.0019366	-0.0037366	-0.0079769
***T* = 0.300 s**	-0.0131291	-0.0036757	-0.0037694	0.0012937	-0.0034624	-0.0000356	-0.0019540	-0.0037319	-0.0079728
***T* = 0.320 s**	-0.0129679	-0.0037321	-0.0037010	0.0011760	-0.0035033	-0.0000384	-0.0019868	-0.0037011	-0.0079689
***T* = 0.340 s**	-0.0128240	-0.0037592	-0.0036308	0.0010688	-0.0035300	-0.0000506	-0.0020093	-0.0036688	-0.0079434

$\Delta c_{3,r}$									
Intensity Measure	**North Coast**	**Bay Area**	**Central Coast**	**Sierra Nevada**	**Southern Sierra and Mojave**	**Eastern CA Shear Zone**	**Southern California**	**Central Valley**	**Northeastern California**
***T* = 0.350 s**	-0.0127403	-0.0037756	-0.0035895	0.0010193	-0.0035387	-0.0000585	-0.0020154	-0.0036369	-0.0079215
***T* = 0.360 s**	-0.0126614	-0.0037692	-0.0035384	0.0009700	-0.0035484	-0.0000734	-0.0020156	-0.0036154	-0.0078941
***T* = 0.380 s**	-0.0125069	-0.0037550	-0.0034301	0.0008840	-0.0035746	-0.0001007	-0.0020116	-0.0035479	-0.0078281
***T* = 0.400 s**	-0.0123510	-0.0037226	-0.0033172	0.0008725	-0.0036041	-0.0000865	-0.0020018	-0.0034746	-0.0077519
***T* = 0.420 s**	-0.0121609	-0.0036810	-0.0032077	0.0008783	-0.0036398	-0.0000612	-0.0019755	-0.0034002	-0.0076718
***T* = 0.440 s**	-0.0119627	-0.0036286	-0.0031005	0.0008495	-0.0036730	-0.0000551	-0.0019460	-0.0033222	-0.0075896
***T* = 0.450 s**	-0.0118518	-0.0035994	-0.0030543	0.0008453	-0.0036783	-0.0000193	-0.0019293	-0.0032808	-0.0075435
***T* = 0.460 s**	-0.0117399	-0.0035706	-0.0030106	0.0008428	-0.0036782	0.0000119	-0.0019112	-0.0032394	-0.0074949
***T* = 0.480 s**	-0.0115206	-0.0035116	-0.0029279	0.0008266	-0.0036841	0.0000682	-0.0018738	-0.0031580	-0.0073964
***T* = 0.500 s**	-0.0113029	-0.0034534	-0.0028647	0.0008319	-0.0036769	0.0000964	-0.0018319	-0.0030845	-0.0072856
***T* = 0.550 s**	-0.0108099	-0.0033347	-0.0026981	0.0008474	-0.0036239	0.0001679	-0.0017270	-0.0029233	-0.0069697
***T* = 0.600 s**	-0.0103136	-0.0032530	-0.0025172	0.0008526	-0.0035546	0.0002326	-0.0016336	-0.0028000	-0.0066413
***T* = 0.650 s**	-0.0097797	-0.0031847	-0.0023300	0.0008841	-0.0034767	0.0002911	-0.0015514	-0.0027008	-0.0063330
***T* = 0.667 s**	-0.0095988	-0.0031627	-0.0022636	0.0008988	-0.0034418	0.0003153	-0.0015254	-0.0026798	-0.0062435
***T* = 0.700 s**	-0.0092486	-0.0031296	-0.0021341	0.0009232	-0.0033621	0.0003450	-0.0014886	-0.0026405	-0.0060774
***T* = 0.750 s**	-0.0088653	-0.0030907	-0.0019351	0.0009579	-0.0032265	0.0003894	-0.0014646	-0.0026036	-0.0058375
***T* = 0.800 s**	-0.0084834	-0.0030619	-0.0017367	0.0010055	-0.0030790	0.0004419	-0.0014309	-0.0025841	-0.0056128
***T* = 0.850 s**	-0.0081050	-0.0030414	-0.0015982	0.0010607	-0.0029305	0.0004972	-0.0014066	-0.0025801	-0.0053952
***T* = 0.900 s**	-0.0077284	-0.0030166	-0.0014760	0.0010870	-0.0027861	0.0005507	-0.0013655	-0.0025809	-0.0051812
***T* = 0.950 s**	-0.0073511	-0.0029877	-0.0013424	0.0010215	-0.0026391	0.0005999	-0.0013222	-0.0025835	-0.0049749
***T* = 1.000 s**	-0.0069728	-0.0029572	-0.0012053	0.0009335	-0.0024898	0.0006380	-0.0012796	-0.0025853	-0.0047667
***T* = 1.100 s**	-0.0062904	-0.0028740	-0.0009333	0.0008366	-0.0021999	0.0007037	-0.0012093	-0.0025780	-0.0043828
***T* = 1.200 s**	-0.0056003	-0.0027712	-0.0006778	0.0007322	-0.0020833	0.0006126	-0.0011520	-0.0025574	-0.0040467
***T* = 1.300 s**	-0.0049973	-0.0026647	-0.0004244	0.0006354	-0.0019087	0.0005543	-0.0011019	-0.0025376	-0.0037110
***T* = 1.400 s**	-0.0045639	-0.0025499	-0.0001770	0.0005121	-0.0017622	0.0004882	-0.0010557	-0.0025193	-0.0034064

$\Delta c_{3,r}$									
Intensity Measure	**North Coast**	**Bay Area**	**Central Coast**	**Sierra Nevada**	**Southern Sierra and Mojave**	**Eastern CA Shear Zone**	**Southern California**	**Central Valley**	**Northeastern California**
***T* = 1.500 s**	-0.0038953	-0.0024802	-0.0001259	0.0003985	-0.0016739	0.0003738	-0.0010003	-0.0025013	-0.0031367
***T* = 1.600 s**	-0.0035275	-0.0024287	-0.0000584	0.0002954	-0.0015487	0.0002722	-0.0009313	-0.0024772	-0.0029312
***T* = 1.700 s**	-0.0031504	-0.0023784	0.0000069	0.0002211	-0.0014325	0.0001787	-0.0008555	-0.0024451	-0.0027570
***T* = 1.800 s**	-0.0027644	-0.0023632	0.0000981	0.0001615	-0.0013422	0.0000945	-0.0007774	-0.0024032	-0.0025872
***T* = 1.900 s**	-0.0023743	-0.0023469	0.0000602	0.0001208	-0.0012731	0.0000830	-0.0006996	-0.0023573	-0.0024227
***T* = 2.000 s**	-0.0019816	-0.0023671	0.0000000	0.0000000	-0.0012039	0.0000000	-0.0006203	-0.0023130	-0.0022614
***T* = 2.200 s**	-0.0015498	-0.0024838	0.0000000	0.0000000	-0.0010575	0.0000000	-0.0004594	-0.0022395	-0.0019425
***T* = 2.400 s**	-0.0010001	-0.0026066	0.0000000	0.0000000	-0.0009078	0.0000000	-0.0003705	-0.0021853	-0.0016800
***T* = 2.500 s**	-0.0007738	-0.0026665	0.0000000	0.0000000	-0.0008337	0.0000000	-0.0003280	-0.0021652	-0.0015652
***T* = 2.600 s**	-0.0005474	-0.0027221	0.0000000	0.0000000	-0.0007582	0.0000000	-0.0003014	-0.0021656	-0.0014477
***T* = 2.800 s**	-0.0003704	-0.0028153	0.0000000	0.0000000	-0.0005789	0.0000000	-0.0002032	-0.0021380	-0.0012488
***T* = 3.000 s**	-0.0002187	-0.0028632	0.0000000	0.0000000	-0.0004601	0.0000000	-0.0002023	-0.0021396	-0.0009750
***T* = 3.200 s**	-0.0000788	-0.0029038	0.0000219	0.0000245	-0.0003537	0.0000221	-0.0002036	-0.0021627	-0.0008049
***T* = 3.400 s**	0.0000401	-0.0029263	0.0000360	0.0000331	-0.0002432	0.0000360	-0.0002368	-0.0021985	-0.0007390
***T* = 3.500 s**	0.0000444	-0.0029273	0.0000408	0.0000363	-0.0002019	0.0000430	-0.0002479	-0.0022277	-0.0007007
***T* = 3.600 s**	0.0000477	-0.0029307	0.0000461	0.0000395	-0.0001349	0.0000461	-0.0002871	-0.0022498	-0.0006627
***T* = 3.800 s**	0.0000513	-0.0029366	0.0000495	0.0000416	-0.0000101	0.0000495	-0.0003658	-0.0023163	-0.0005990
***T* = 4.000 s**	0.0000502	-0.0029508	0.0000492	0.0000412	0.0000510	0.0000492	-0.0004653	-0.0023829	-0.0005490
***T* = 4.200 s**	0.0000448	-0.0029699	0.0000441	0.0000368	0.0000430	0.0000448	-0.0004562	-0.0024561	-0.0005101
***T* = 4.400 s**	0.0000368	-0.0029954	0.0000354	0.0000312	0.0000341	0.0000368	-0.0004544	-0.0025330	-0.0004741
***T* = 4.600 s**	0.0000261	-0.0029701	0.0000251	0.0000252	0.0000260	0.0000254	-0.0004354	-0.0026125	-0.0004411
***T* = 4.800 s**	0.0000137	-0.0028732	0.0000167	0.0000198	0.0000200	0.0000143	-0.0004150	-0.0025620	-0.0004253
***T* = 5.000 s**	0.0000000	-0.0027701	0.0000000	0.0000000	0.0000145	0.0000000	-0.0003774	-0.0025057	-0.0004108
***T* = 5.500 s**	0.0000000	-0.0025466	0.0000000	0.0000000	0.0000082	0.0000000	-0.0002752	-0.0023188	-0.0003932
***T* = 6.000 s**	0.0000000	-0.0022925	0.0000000	0.0000000	0.0000037	0.0000000	-0.0001716	-0.0022293	-0.0003540

$\Delta c_{3,r}$									
Intensity Measure	**North Coast**	**Bay Area**	**Central Coast**	**Sierra Nevada**	**Southern Sierra and Mojave**	**Eastern CA Shear Zone**	**Southern California**	**Central Valley**	**Northeastern California**
***T* = 6.500 s**	0.0000000	-0.0019967	0.0000000	0.0000000	0.0000016	0.0000000	-0.0001191	-0.0021174	-0.0003005
***T* = 7.000 s**	0.0000000	-0.0016923	0.0000000	0.0000000	0.0000002	0.0000000	-0.0000627	-0.0020158	-0.0002444
***T* = 7.500 s**	0.0000000	-0.0013939	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	-0.0018697	-0.0001863
***T* = 8.000 s**	0.0000000	-0.0010915	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	-0.0017113	-0.0001352
***T* = 8.500 s**	0.0000000	-0.0007839	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	-0.0015486	-0.0000843
***T* = 9.000 s**	0.0000000	-0.0004767	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	-0.0013841	-0.0000333
***T* = 9.500 s**	0.0000000	-0.0001702	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	-0.0012181	0.0000000
***T* = 10.000 s**	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	-0.0010506	0.0000000

Table 4.3: Coefficient values for $\Delta c_{0,r}$.

$\Delta c_{0,r}$									
Intensity Measure	**North Coast**	**Bay Area**	**Central Coast**	**Sierra Nevada**	**Southern Sierra and Mojave**	**Eastern CA Shear Zone**	**Southern California**	**Central Valley**	**Northeastern California**
PGV	0.0007371	0.0005161	0.0023289	-0.0033628	0.0008042	0.0021270	0.0001762	0.0000000	0.0000000
PGA	0.0003538	-0.0035011	-0.0030760	-0.0059568	0.0005254	0.0023358	0.0001337	0.0000000	0.0000000
***T* = 0.010 s**	0.0003879	-0.0033996	-0.0029975	-0.0060205	0.0004986	0.0022212	0.0001552	0.0000000	0.0000000
***T* = 0.020 s**	0.0000007	-0.0037354	-0.0033778	-0.0058507	0.0005243	0.0021375	0.0001556	0.0000000	0.0000000
***T* = 0.022 s**	-0.0000776	-0.0038030	-0.0034509	-0.0058170	0.0005296	0.0021214	0.0001556	0.0000000	0.0000000
***T* = 0.025 s**	-0.0001947	-0.0039029	-0.0035586	-0.0057686	0.0005374	0.0020976	0.0001555	0.0000000	0.0000000
***T* = 0.029 s**	-0.0003578	-0.0040548	-0.0037868	-0.0056780	0.0005480	0.0020668	0.0001553	0.0000000	0.0000000

$\Delta c_{0,r}$									
Intensity Measure	**North Coast**	**Bay Area**	**Central Coast**	**Sierra Nevada**	**Southern Sierra and Mojave**	**Eastern CA Shear Zone**	**Southern California**	**Central Valley**	**Northeastern California**
***T* = 0.030 s**	-0.0004047	-0.0040999	-0.0038452	-0.0056532	0.0005506	0.0020593	0.0001552	0.0000000	0.0000000
***T* = 0.032 s**	-0.0004939	-0.0041897	-0.0039629	-0.0056019	0.0005559	0.0020444	0.0001551	0.0000000	0.0000000
***T* = 0.035 s**	-0.0006282	-0.0043249	-0.0041570	-0.0055114	0.0005639	0.0020230	0.0001548	0.0000000	0.0000000
***T* = 0.036 s**	-0.0006793	-0.0043678	-0.0042220	-0.0054820	0.0005666	0.0020163	0.0001547	0.0000000	0.0000000
***T* = 0.040 s**	-0.0008856	-0.0045365	-0.0044819	-0.0053681	0.0005774	0.0019934	0.0001544	0.0000000	0.0000000
***T* = 0.042 s**	-0.0009915	-0.0046184	-0.0046002	-0.0053136	0.0005828	0.0019935	0.0001543	0.0000000	0.0000000
***T* = 0.044 s**	-0.0010998	-0.0046938	-0.0047180	-0.0052649	0.0005882	0.0019898	0.0001542	0.0000000	0.0000000
***T* = 0.045 s**	-0.0011536	-0.0047274	-0.0047713	-0.0052469	0.0005910	0.0019875	0.0001542	0.0000000	0.0000000
***T* = 0.046 s**	-0.0012073	-0.0047647	-0.0048220	-0.0052243	0.0005937	0.0019859	0.0001542	0.0000000	0.0000000
***T* = 0.048 s**	-0.0013135	-0.0048272	-0.0049227	-0.0052029	0.0005992	0.0019868	0.0001541	0.0000000	0.0000000
***T* = 0.050 s**	-0.0014112	-0.0048706	-0.0050029	-0.0051829	0.0006047	0.0019952	0.0001542	0.0000000	0.0000000
***T* = 0.055 s**	-0.0016349	-0.0049518	-0.0051813	-0.0051379	0.0006185	0.0020140	0.0001545	0.0000000	0.0000000
***T* = 0.060 s**	-0.0017995	-0.0049903	-0.0053372	-0.0050880	0.0006323	0.0020488	0.0001551	0.0000000	0.0000000
***T* = 0.065 s**	-0.0019317	-0.0050229	-0.0054512	-0.0050531	0.0006463	0.0021139	0.0001559	0.0000000	0.0000000
***T* = 0.067 s**	-0.0019774	-0.0050328	-0.0054787	-0.0050509	0.0006519	0.0021402	0.0001562	0.0000000	0.0000000
***T* = 0.070 s**	-0.0020146	-0.0050325	-0.0054979	-0.0050709	0.0006603	0.0021943	0.0001564	0.0000000	0.0000000
***T* = 0.075 s**	-0.0020073	-0.0050063	-0.0054667	-0.0051419	0.0006744	0.0022767	0.0001671	0.0000000	0.0000000
***T* = 0.080 s**	-0.0019666	-0.0049716	-0.0054158	-0.0052731	0.0006884	0.0023733	0.0001697	0.0000000	0.0000000
***T* = 0.085 s**	-0.0019057	-0.0049285	-0.0053404	-0.0053881	0.0007025	0.0024718	0.0001981	0.0000000	0.0000000
***T* = 0.090 s**	-0.0018204	-0.0048588	-0.0052219	-0.0055197	0.0007165	0.0025833	0.0002254	0.0000000	0.0000000
***T* = 0.095 s**	-0.0017221	-0.0047729	-0.0050974	-0.0056645	0.0007306	0.0026957	0.0002518	0.0000000	0.0000000
***T* = 0.100 s**	-0.0015895	-0.0046561	-0.0049504	-0.0058110	0.0007446	0.0028060	0.0002775	0.0000000	0.0000000
***T* = 0.110 s**	-0.0012471	-0.0043758	-0.0046176	-0.0060780	0.0007730	0.0030192	0.0003276	0.0000000	0.0000000
***T* = 0.120 s**	-0.0008595	-0.0040884	-0.0042346	-0.0063429	0.0008021	0.0031056	0.0003539	0.0000000	0.0000000
***T* = 0.130 s**	-0.0003986	-0.0037811	-0.0038078	-0.0065771	0.0008330	0.0031284	0.0003598	0.0000000	0.0000000

$\Delta c_{0,r}$									
Intensity Measure	**North Coast**	**Bay Area**	**Central Coast**	**Sierra Nevada**	**Southern Sierra and Mojave**	**Eastern CA Shear Zone**	**Southern California**	**Central Valley**	**Northeastern California**
***T* = 0.133 s**	-0.0002464	-0.0036854	-0.0036446	-0.0066433	0.0008429	0.0031073	0.0003609	0.0000000	0.0000000
***T* = 0.140 s**	0.0001137	-0.0034617	-0.0032592	-0.0067936	0.0008711	0.0030817	0.0003688	0.0000000	0.0000000
***T* = 0.150 s**	0.0006632	-0.0031385	-0.0026479	-0.0069894	0.0009136	0.0030986	0.0003721	0.0000000	0.0000000
***T* = 0.160 s**	0.0012322	-0.0028157	-0.0020290	-0.0071411	0.0009612	0.0031048	0.0003597	0.0000000	0.0000000
***T* = 0.170 s**	0.0017822	-0.0024946	-0.0014021	-0.0072358	0.0010063	0.0030868	0.0003325	0.0000000	0.0000000
***T* = 0.180 s**	0.0022500	-0.0021764	-0.0008503	-0.0071902	0.0010477	0.0030449	0.0002909	0.0000000	0.0000000
***T* = 0.190 s**	0.0026773	-0.0018634	-0.0003317	-0.0070653	0.0010851	0.0029862	0.0002393	0.0000000	0.0000000
***T* = 0.200 s**	0.0030947	-0.0015565	0.0001122	-0.0068385	0.0011206	0.0029069	0.0001909	0.0000000	0.0000000
***T* = 0.220 s**	0.0038593	-0.0009691	0.0008664	-0.0065409	0.0011781	0.0027005	0.0000806	0.0000000	0.0000000
***T* = 0.240 s**	0.0045241	-0.0004272	0.0015825	-0.0061199	0.0012239	0.0024697	-0.0000476	0.0000000	0.0000000
***T* = 0.250 s**	0.0048484	-0.0001592	0.0019070	-0.0058923	0.0012466	0.0023581	-0.0001086	0.0000000	0.0000000
***T* = 0.260 s**	0.0051260	0.0000979	0.0022143	-0.0056651	0.0012684	0.0022496	-0.0001680	0.0000000	0.0000000
***T* = 0.280 s**	0.0055287	0.0005679	0.0028106	-0.0052466	0.0013165	0.0020513	-0.0002761	0.0000000	0.0000000
***T* = 0.290 s**	0.0056606	0.0007636	0.0031036	-0.0050800	0.0013416	0.0019633	-0.0003209	0.0000000	0.0000000
***T* = 0.300 s**	0.0058114	0.0009659	0.0033173	-0.0048924	0.0013674	0.0018700	-0.0003607	0.0000000	0.0000000
***T* = 0.320 s**	0.0060494	0.0013150	0.0037700	-0.0045677	0.0014197	0.0016981	-0.0004189	0.0000000	0.0000000
***T* = 0.340 s**	0.0063089	0.0016557	0.0041447	-0.0042673	0.0014698	0.0015652	-0.0004864	0.0000000	0.0000000
***T* = 0.350 s**	0.0064042	0.0018031	0.0042744	-0.0041118	0.0014919	0.0014967	-0.0005126	0.0000000	0.0000000
***T* = 0.360 s**	0.0064782	0.0019405	0.0044582	-0.0039529	0.0015125	0.0014452	-0.0005481	0.0000000	0.0000000
***T* = 0.380 s**	0.0065590	0.0021829	0.0046484	-0.0036166	0.0015404	0.0013344	-0.0005913	0.0000000	0.0000000
***T* = 0.400 s**	0.0066122	0.0024113	0.0047819	-0.0032605	0.0015459	0.0012455	-0.0006371	0.0000000	0.0000000
***T* = 0.420 s**	0.0066330	0.0026010	0.0049113	-0.0029167	0.0015386	0.0011553	-0.0006638	0.0000000	0.0000000
***T* = 0.440 s**	0.0066449	0.0027515	0.0050390	-0.0025897	0.0015259	0.0010903	-0.0006713	0.0000000	0.0000000
***T* = 0.450 s**	0.0066522	0.0028248	0.0051027	-0.0024314	0.0015169	0.0010607	-0.0006818	0.0000000	0.0000000
***T* = 0.460 s**	0.0066559	0.0028777	0.0051665	-0.0023246	0.0015073	0.0010351	-0.0006824	0.0000000	0.0000000

$\Delta c_{0,r}$									
Intensity Measure	**North Coast**	**Bay Area**	**Central Coast**	**Sierra Nevada**	**Southern Sierra and Mojave**	**Eastern CA Shear Zone**	**Southern California**	**Central Valley**	**Northeastern California**
***T* = 0.480 s**	0.0066584	0.0029640	0.0053001	-0.0021319	0.0014911	0.0010045	-0.0006888	0.0000000	0.0000000
***T* = 0.500 s**	0.0066490	0.0030429	0.0054397	-0.0019850	0.0014732	0.0009986	-0.0006857	0.0000000	0.0000000
***T* = 0.550 s**	0.0064680	0.0031364	0.0057203	-0.0016661	0.0014224	0.0009795	-0.0006751	0.0000000	0.0000000
***T* = 0.600 s**	0.0062704	0.0032372	0.0059508	-0.0013325	0.0013642	0.0009958	-0.0006460	0.0000000	0.0000000
***T* = 0.650 s**	0.0060451	0.0033177	0.0061445	-0.0010222	0.0013122	0.0010111	-0.0006071	0.0000000	0.0000000
***T* = 0.667 s**	0.0059793	0.0033439	0.0062035	-0.0009205	0.0012963	0.0010233	-0.0005972	0.0000000	0.0000000
***T* = 0.700 s**	0.0058457	0.0033928	0.0063048	-0.0007217	0.0012585	0.0010428	-0.0005735	0.0000000	0.0000000
***T* = 0.750 s**	0.0056574	0.0034604	0.0064312	-0.0004223	0.0012020	0.0010720	-0.0005398	0.0000000	0.0000000
***T* = 0.800 s**	0.0054748	0.0035149	0.0065382	-0.0001585	0.0011421	0.0010979	-0.0005056	0.0000000	0.0000000
***T* = 0.850 s**	0.0052955	0.0035556	0.0066344	0.0001054	0.0010831	0.0011201	-0.0004681	0.0000000	0.0000000
***T* = 0.900 s**	0.0051218	0.0035784	0.0066991	0.0003544	0.0010276	0.0011363	-0.0004269	0.0000000	0.0000000
***T* = 0.950 s**	0.0049555	0.0035787	0.0067557	0.0005539	0.0009771	0.0011535	-0.0003824	0.0000000	0.0000000
***T* = 1.000 s**	0.0047930	0.0035837	0.0067768	0.0007257	0.0009316	0.0011693	-0.0003379	0.0000000	0.0000000
***T* = 1.100 s**	0.0045063	0.0035558	0.0067462	0.0009381	0.0008521	0.0011972	-0.0002569	0.0000000	0.0000000
***T* = 1.200 s**	0.0042576	0.0035171	0.0066872	0.0011090	0.0007916	0.0012227	-0.0001874	0.0000000	0.0000000
***T* = 1.300 s**	0.0040360	0.0034764	0.0065882	0.0012550	0.0007449	0.0012466	-0.0001265	0.0000000	0.0000000
***T* = 1.400 s**	0.0038001	0.0034412	0.0065077	0.0013922	0.0007020	0.0012751	-0.0000710	0.0000000	0.0000000
***T* = 1.500 s**	0.0035939	0.0034316	0.0064361	0.0014893	0.0006849	0.0013259	-0.0000274	0.0000000	0.0000000
***T* = 1.600 s**	0.0033990	0.0034300	0.0064049	0.0015918	0.0006786	0.0013897	-0.0000200	0.0000000	0.0000000
***T* = 1.700 s**	0.0032396	0.0034545	0.0063673	0.0017083	0.0006751	0.0014929	-0.0000432	0.0000000	0.0000000
***T* = 1.800 s**	0.0031052	0.0034996	0.0063625	0.0018300	0.0006723	0.0016211	-0.0000868	0.0000000	0.0000000
***T* = 1.900 s**	0.0030108	0.0035970	0.0063882	0.0019653	0.0006649	0.0017606	-0.0001436	0.0000000	0.0000000
***T* = 2.000 s**	0.0028945	0.0036718	0.0064470	0.0021014	0.0006521	0.0019154	-0.0001992	0.0000000	0.0000000
***T* = 2.200 s**	0.0026935	0.0038541	0.0066991	0.0023031	0.0006257	0.0022037	-0.0003276	0.0000000	0.0000000
***T* = 2.400 s**	0.0025437	0.0040554	0.0069413	0.0024840	0.0006064	0.0025146	-0.0004730	0.0000000	0.0000000

$\Delta c_{0,r}$									
Intensity Measure	**North Coast**	**Bay Area**	**Central Coast**	**Sierra Nevada**	**Southern Sierra and Mojave**	**Eastern CA Shear Zone**	**Southern California**	**Central Valley**	**Northeastern California**
***T* = 2.500 s**	0.0024944	0.0041648	0.0070536	0.0025438	0.0006020	0.0026855	-0.0005406	0.0000000	0.0000000
***T* = 2.600 s**	0.0024601	0.0042761	0.0071720	0.0026223	0.0005988	0.0028719	-0.0006042	0.0000000	0.0000000
***T* = 2.800 s**	0.0024101	0.0044949	0.0074129	0.0027284	0.0006016	0.0034452	-0.0007334	0.0000000	0.0000000
***T* = 3.000 s**	0.0023328	0.0046868	0.0076676	0.0027756	0.0006138	0.0041175	-0.0008778	0.0000000	0.0000000
***T* = 3.200 s**	0.0022556	0.0048614	0.0079211	0.0028270	0.0006282	0.0047139	-0.0010289	0.0000000	0.0000000
***T* = 3.400 s**	0.0021857	0.0050246	0.0081951	0.0028909	0.0006413	0.0053066	-0.0011839	0.0000000	0.0000000
***T* = 3.500 s**	0.0021423	0.0050881	0.0083394	0.0029260	0.0006463	0.0056036	-0.0012623	0.0000000	0.0000000
***T* = 3.600 s**	0.0021152	0.0051620	0.0084949	0.0029608	0.0006506	0.0059004	-0.0013398	0.0000000	0.0000000
***T* = 3.800 s**	0.0020242	0.0052419	0.0087959	0.0030300	0.0006546	0.0065005	-0.0014951	0.0000000	0.0000000
***T* = 4.000 s**	0.0019610	0.0053142	0.0090483	0.0030766	0.0006533	0.0070943	-0.0016024	0.0000000	0.0000000
***T* = 4.200 s**	0.0019243	0.0053873	0.0092684	0.0030191	0.0006490	0.0072719	-0.0016619	0.0000000	0.0000000
***T* = 4.400 s**	0.0019127	0.0054507	0.0094703	0.0030126	0.0006427	0.0075982	-0.0016885	0.0000000	0.0000000
***T* = 4.600 s**	0.0019819	0.0055181	0.0096650	0.0029828	0.0006345	0.0080540	-0.0016748	0.0000000	0.0000000
***T* = 4.800 s**	0.0020433	0.0055794	0.0098488	0.0029379	0.0006260	0.0084462	-0.0016414	0.0000000	0.0000000
***T* = 5.000 s**	0.0021676	0.0056476	0.0102346	0.0028622	0.0006184	0.0089168	-0.0015837	0.0000000	0.0000000
***T* = 5.500 s**	0.0024661	0.0056850	0.0109673	0.0025728	0.0006019	0.0099181	-0.0014884	0.0000000	0.0000000
***T* = 6.000 s**	0.0027645	0.0057639	0.0115453	0.0019241	0.0005876	0.0108898	-0.0014697	0.0000000	0.0000000
***T* = 6.500 s**	0.0029583	0.0058320	0.0121279	0.0006702	0.0005749	0.0118528	-0.0014517	0.0000000	0.0000000
***T* = 7.000 s**	0.0029910	0.0058669	0.0125753	-0.0000389	0.0005621	0.0128359	-0.0014447	0.0000000	0.0000000
***T* = 7.500 s**	0.0030345	0.0058622	0.0128972	-0.0007060	0.0005487	0.0138337	-0.0014412	0.0000000	0.0000000
***T* = 8.000 s**	0.0030442	0.0057885	0.0130802	-0.0011496	0.0005344	0.0148413	-0.0014367	0.0000000	0.0000000
***T* = 8.500 s**	0.0030611	0.0056969	0.0132674	-0.0015064	0.0005196	0.0158558	-0.0014301	0.0000000	0.0000000
***T* = 9.000 s**	0.0030726	0.0055922	0.0134478	-0.0018296	0.0005042	0.0168750	-0.0014216	0.0000000	0.0000000
***T* = 9.500 s**	0.0030750	0.0054790	0.0136200	-0.0021344	0.0004886	0.0178976	-0.0014113	0.0000000	0.0000000
***T* = 10.000 s**	0.0030684	0.0053607	0.0137853	-0.0024295	0.0004729	0.0189226	-0.0013999	0.0000000	0.0000000

(d) Geysers and Tectonic North Coast Events

In this section, I investigate whether Geyser and tectonic events in the North Coast subregion exhibit different attenuation characteristics. To evaluate these effects, event-specific within-event residuals are regressed using Equation (4.9)for both event types, with the results shown in Figure 4.12 for the IM of *PSA* at 1.0 sec. I find little difference between data trends for the two event types. Statistical F-tests were used to evaluate the degree to which data from the two event types are distinct (Mishra et al. 2019). These F-tests are performed on residuals from two path models. One model is fit using the combined dataset of Geyser and non-Geyser (tectonic) North Coast data, while the second is comprised of two separate sets of coefficients fit from the individual datasets. By comparing the statistical properties of the two residual datasets I can infer if the path effects are different, which is quantified using an F-statistic. P-values computed from the F-statistic range from 0.18 to 0.99 for all IMs, which exceed a commonly applied significance level of 0.05. This indicates that submodels vs combined model cannot be considered distinct. As a result, I conclude that it is unnecessary to discriminate between these event types in model development, which was applied in the regressions for the North Coast subregion described above.

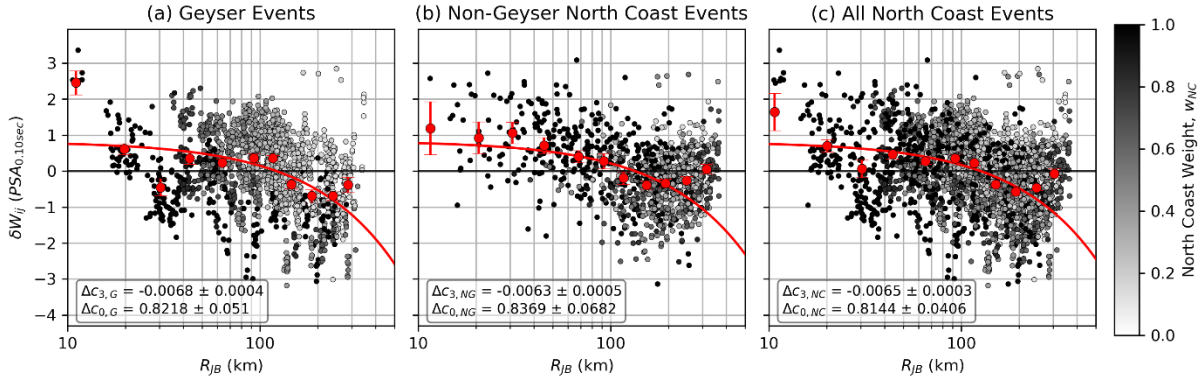


Figure 4.12: Within-event residuals (δW_{ij}) against R_{JB} for *PSA* at 0.10 sec for (a) Geyser-only events, (b) tectonic north coast events, and (c) all north coast (NC) events. Data points shaded relative to the path weight within the north coast region.

4.6. Model Performance

Model performance is evaluated based on attributes of residuals computed using Equation (4.5)but with the BSSA14 model updated to include the path adjustments from the subregional path model (Equation 4.11) and constant term adjustment from the subregional path model and coefficients for Geyser events (Equation 4.12). All model coefficients were taken from Tables 4.1 – 4.3. Total residuals were partitioned into event terms and within-event residuals (δW_{ij}) usingEquation (4.6). Figure 4.13(a) shows constant terms from the original BSSA14 ergodic model and the updated model. I observe a reduction of model bias for *PGV*, *PGA* and for periods shorter than about 1.0 sec with longer periods having similar bias when subregional path effects are considered. This is expected since anelastic effects are more pronounced for short period intensity measures, and the proposed subregional path model does not modify the ergodic anelastic attenuation slope for longer periods. Figure 4.13(b-d) show δW_{ij} and binned means for all magnitudes and events with **M** > 5 and **M** < 5 (as in Figure 4.6). The downward trend for **M** < 5 events has largely been removed.

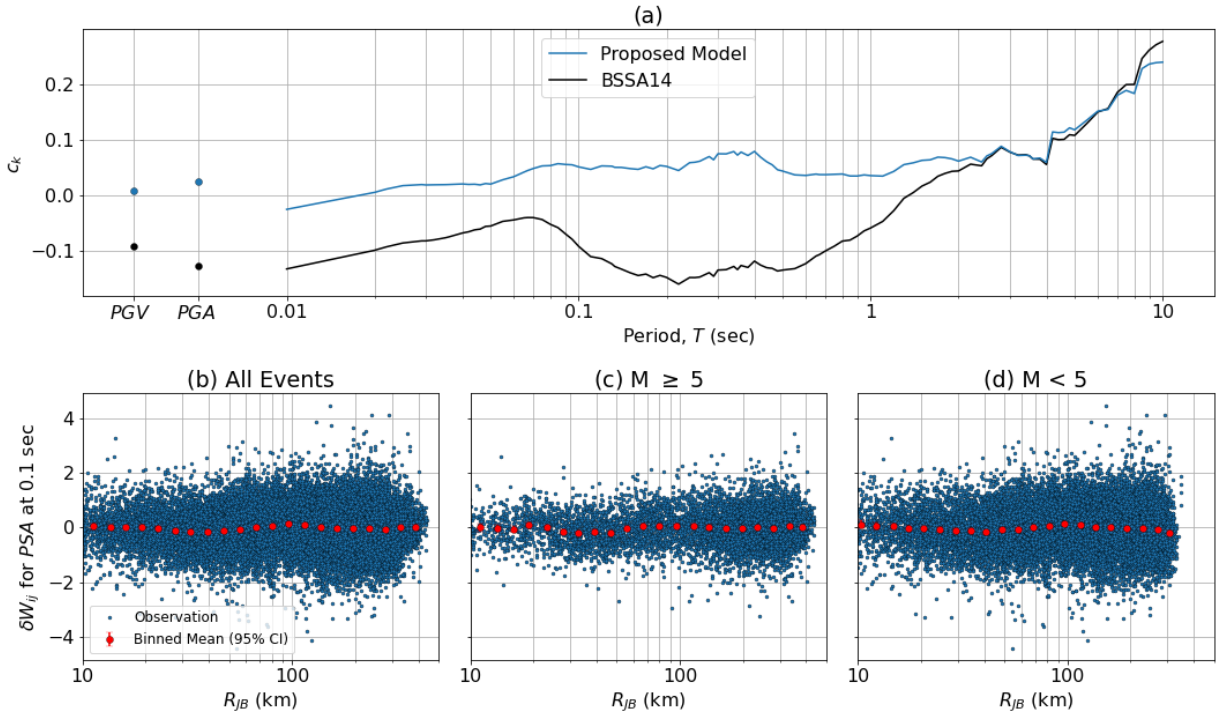


Figure 4.13: (a) Model bias from residuals analysis for the original regional BSSA14 model and adjusted subregional model for *PGV*, *PGA* and 105 oscillator periods; trends of within-event residual (δW_{ij}) for *PSA* at 0.10 sec with distance using (b) full California dataset, (c) subset of earthquakes with **M** > 5 and (d) subsets of earthquakes with **M** < 5.

Figure 4.14 shows δW_{ij} vs. distance computed for the updated model for the six subregions shown previously in Figure 4.7. For each subregion, the trend of binned means with distance is flat, indicating that the variable path effects have been accounted for in the revised model. In the case of the North Coast subregion, the misfit at close distances (R_{JB} < 20 km) remains due to the lack of adjustment of the *h* term, but binned means at larger distances retain a flat trend.

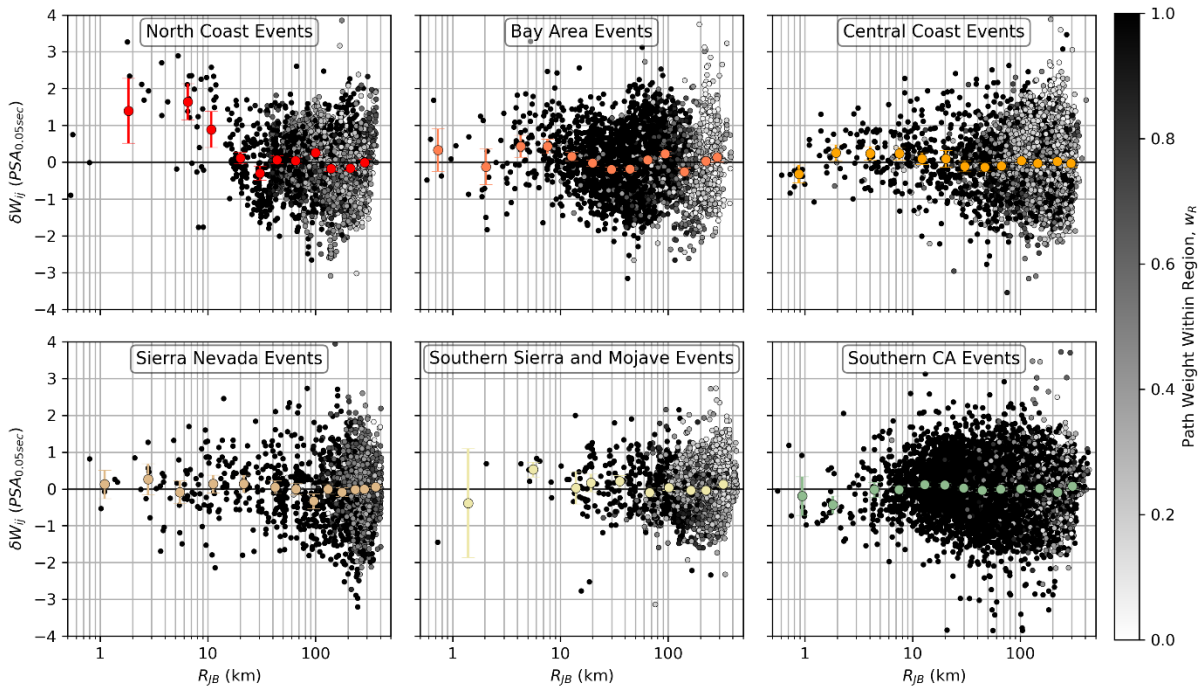


Figure 4.14: Trend of within-event residuals (δW_{ij}) against Joyner-Boore distance (R_{JB}) for *PSA* at 0.05 sec grouped by region of event origination with updated model; data points shaded relative to the path weight within the region of event origination; results shown for the first iteration.

Figure 4.15 shows total, between-event, and within-event standard-deviations (σ, τ, and ϕ, respectively) obtained from residuals computed using the BSSA14 regional GMM for California and two datasets: (1) California data in NGA-West2 data and (2) the full dataset utilized in this study. These standard-deviations are related as,

$$\sigma = \sqrt{\tau^2 + \phi^2} \tag{4.15}$$

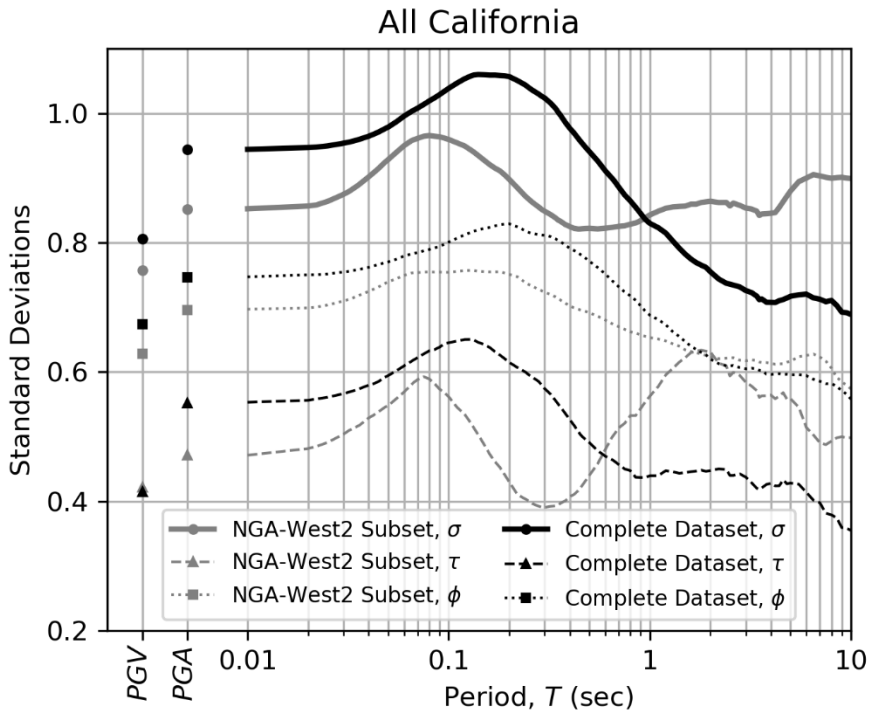


Figure 4.15: Standard-deviations calculated using the default BSSA14 GMM for the subset of NGA-West2 data and the complete dataset utilized in this study.

All three standard deviations are larger at short periods and smaller at long periods from the expanded dataset than from the NGA-West2 data. These increases are likely caused by a relatively large amount of small magnitude data, which has large short-period dispersion and smaller long-period dispersion. Also potentially contributing to these trends are the increased diversity of earthquakes and recording stations in the expanded data, particularly in the northern and eastern portions of California.

Figure 4.16(a) shows the same dispersion terms along with within-event, single station variability (ϕ_{SS}) (standard deviation of ϵ) when residuals are computed using the BSSA14 GMMs and using the updated GMMs developed here. The proposed path and constant term models reduce variability for all IMs, but particularly at short periods. The bump in dispersion near 0.1 sec in the

NGA-West2 model has been repeatedly observed in different studies, and has been attributed by some to be a consequence of site response (Al Atik and Abrahamson, 2010). The proposed model eliminates the bump in ϕ_{SS} and substantially reduces it in τ. The effect on τ likely stems from path corrections for distant recordings reducing path dispersion that had influenced $\eta_{E,j}$ estimates.

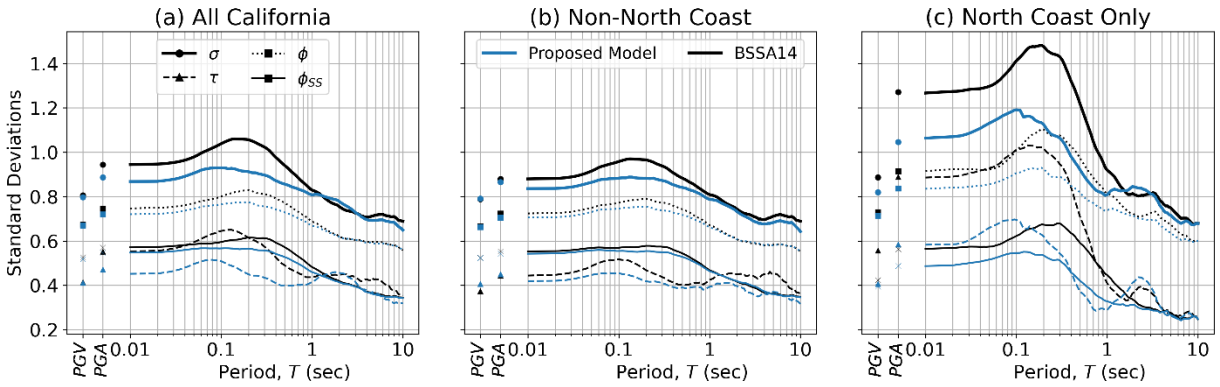


Figure 4.16: Standard-deviations calculated using the BSSA14 GMM and updated model and (a) the full California dataset, (b) the data set exclusive of north coast data, and (c) only north coast data.

I investigated causes of the elevated dispersion from the expanded data set (Figure 4.13), and found that a contributing factor is the added data for the North Coast subregion. Figure 4.16(b,c) show the same dispersion terms as used for the full dataset in Figure 4.16(a), but now computed from datasets exclusive of North Coast events (Figure 4.16b) and for North Coast events only (Figure 4.16c). The results demonstrate variability reduction for non-North Coast data and variability increase for North Coast data. The large variability coming from the North Coast subregion may be influenced by unmodelled source effects for Geysers events or from especially complex wave propagation effects.

After excluding the North Coast data, Figure 4.17(a) shows the magnitude-dependence of between-event variability terms (τ), which were computed by binning the data into **M** groups in a similar manner to that of BSSA14. Results for both the original BSSA14 model and the updated model are shown. The change in binned τ values is small relative to the standard errors. In both cases, the data show a decrease of dispersion with increasing **M**, whereas BSSA14 show a small increase (this increase is only for oscillator periods between 0.08-0.17 sec, elsewhere τ decreases with **M)**. Given the similar levels of τ in this study with BSSA14, no revisions to the τ model are proposed. Figure 4.17(b,c) show ϕ and ϕ_{SS} binned by **M** and R_{JB} for the original and updated GMMs. The model drawn with the ϕ_{SS} results is from Goulet et al. (2018). Comparing binned ϕ and ϕ_{SS} values derived using BSSA14 to those from the proposed model, I observe no significant within-event variability reductions at close distances for the proposed model, as expected. On the other hand, dispersions are reduced at larger distance for **M** $\leq$ 4.5 events, which is the source of most of the data considered in this study. The lack of dispersion reduction for larger magnitudes is based on a small number of well-sampled, large **M** events, and hence it is unclear to what extent regionalized path models are helpful for such earthquakes. Overall, the small difference between ergodic aleatory model and the data derived aleatory variations in Figure 4.17 indicate that updates to aleatory variability models are not needed when the proposed model is applied. This occurs because of two offsetting effects – the enlarged database increases dispersion in the BSSA14 regional model (Figure 4.15), but the subregional model reduces uncertainty at large distances to levels consistent with BSSA14 (Figure 4.17).

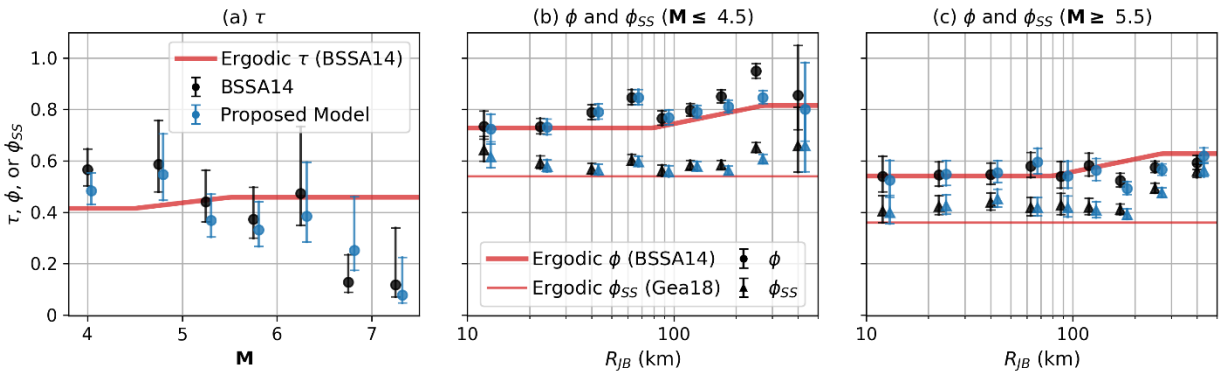


Figure 4.17: Binned standard deviation terms with their standard errors for *PSA* at 0.10 sec showing (a) magnitude effect on between-event variability (τ), (b) distance effect on within-event variability (ϕ) and single-station within-event variability (ϕ_{SS}) for small **M**, and (c) distance effect on ϕ and ϕ_{SS} for large **M**. Results are for California dataset exclusive of North Coast events. BSSA14 = Boore et al. (2014); Gea18 = Goulet et al. (2018).

To investigate effects of the subregional model on dispersion, I plot in Figure 4.18 the changes in overall within-event variability ($\Delta\phi$) and single station variability ($\Delta\phi_{SS}$), where negative changes indicate reductions. The results indicate standard deviations reductions of up to 0.07 at short periods and large distances (R_{JB} > 100 km). These changes are comparable to changes in total dispersion (σ) from non-ergodic path models as reported in Figure 6 of Kuehn et al. (2019), which suggests that the present work, despite the coarser spatial resolution, achieves similar levels of within-event variability reduction.

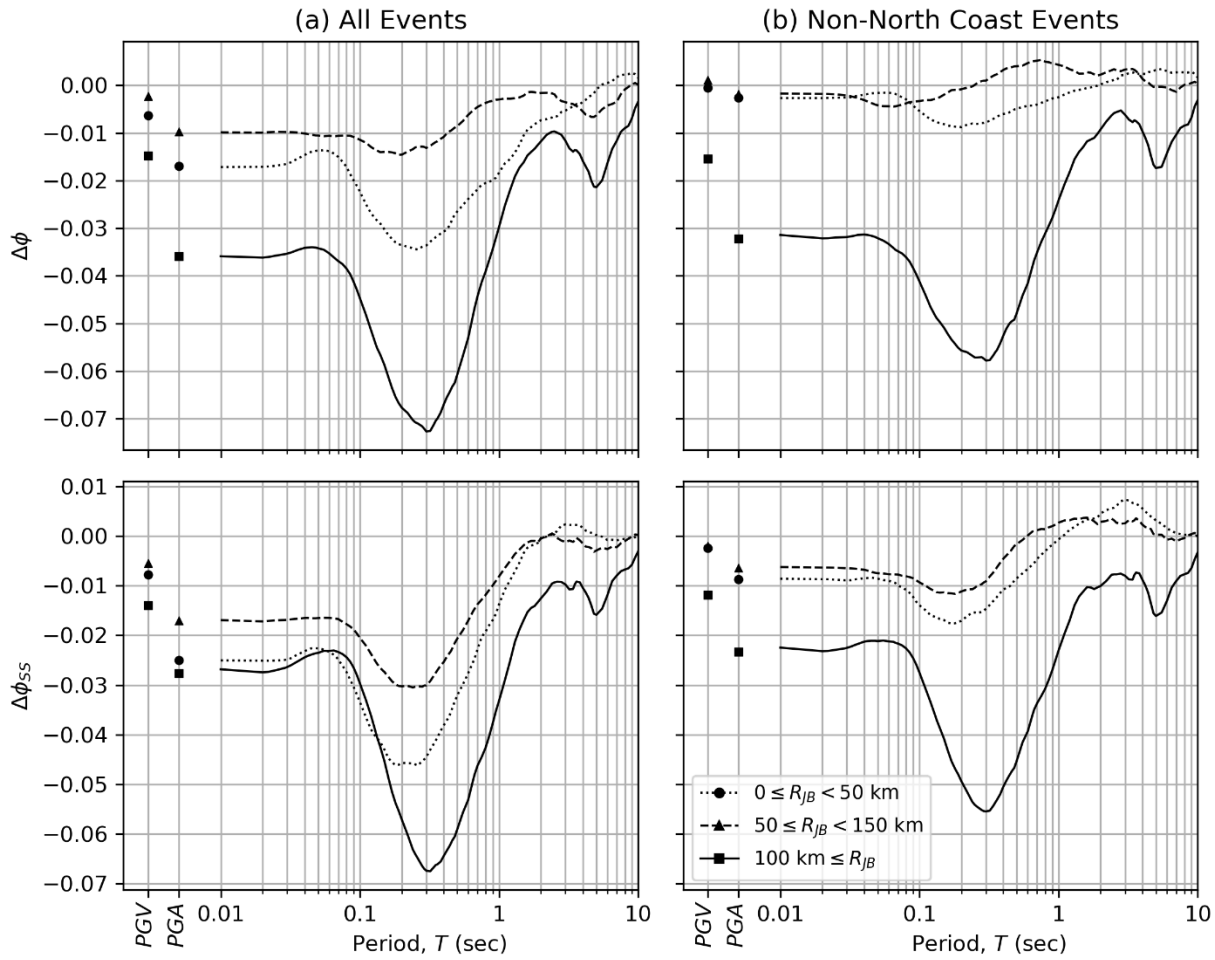


Figure 4.18: Changes in within-event variability ($\Delta\phi$) and single-station within-event variability ($\Delta\phi_{SS}$) from BSSA14 to proposed subregional model. Results are binned by distance and computed for (a) all California events (b) California dataset exclusive of north coast. Negative values indicate variability reduction relative to BSSA14.

4.7. Conclusions

The anelastic path models in the NGA-West2 GMMs consider broad geopolitical regionalization which lacks physical basis and represent average attenuation for large areas. Previous studies have shown evidence of variable ground motion characteristics across California, some of which proposed regional or cell-based path models to capture variable rates of attenuation. I propose nine

physiographical subregions for California derived from prior work (Hauksson and Shearer 2006; Chiou et al. 2010; Eberhart-Phillips 2016). A substantially augmented version of the NGA-West2 California ground motion database was used to facilitate a residuals-based analyses of subregional path-effects.

I demonstrated that Geyser (induced) events possess different source effects when compared to other north coast tectonic events. I find path effects between induced and tectonic not to be statistically distinct, therefore my analyses of path effects combine them. I introduce an induced constant adjustment term (e_I) to remove biases related to induced source effects so that they do not inflate between-event variability (τ).

I present an iterative approach that identifies subregions with path bias (relative to the California-wide model) and adjusts subregional anelastic path coefficients (Δc_3^*; Equation 4.13) and constant term adjustment ($\Delta c_{0,r}$). The resulting model, which adjusts the BSSA14 GMM, is given by Equation (4.10), in which the F_P term that accounts for subregional path effects is taken from Equation (4.11). Other than the adjusted constant and anelastic attenuation terms, other terms are not modified from BSSA14. The aleatory variability is generally unchanged from prior models (BSSA14 for τ and ϕ; Goulet et al. 2018 for ϕ_{SS}), although higher variability is found for North Coast events. Subregional model coefficients $\Delta c_{3,r}$, $\Delta c_{0,r}$, and e_I are provided in Tables 4.1 – 4.3.

Given that the database is dominated by **M** 4-5 events, an implicit assumption in the proposed model that is important for hazard applications is that the frequency-dependent spatial variations of anelastic attenuation are equally applicable for the data range and the larger, hazard-critical magnitude range. Further subregionalization may be needed in some cases such as southern California, which was not the main focus of this work.

CHAPTER 5

Region-Specific Site Response in the Sacramento-San Joaquin Delta

5.1. Introduction

Site response, in the context of geotechnical earthquake engineering, represents the effects of local geological and geotechnical site conditions on earthquake ground motions. Site response can be evaluated from one-dimensional or multi-dimensional wave propagation analyses that account for the seismic velocity structure or can be inferred from local ground motion recordings. Site response models utilize site parameters, such as those discussed in Chapter 2, to estimate the changes in ground motions at the surface of a site relative to ground motions for a reference site condition. Many factors influence site response, including:

1. Impedance related to the vertical propagation of seismic waves, generally from stiffer (high V_S) to softer layers (low V_S), which results in amplification to maintain the conservation of energy (i.e., as waves slowdown in softer layers their amplitudes must increase);
2. Basin effects associated with large sedimentary structures in the form of basin edge effects, focusing, and the generation of surface waves (Graves 1993);
3. Resonance of particular ground motion frequencies with one or more natural frequencies of the site, which results in a relatively narrow band of higher amplification (e.g., Idriss and Seed 1968; Seed et al. 1976); and

4. Nonlinearity associated with soil modulus reduction and increased damping under strong shaking conditions that produce large soil strains, which attenuates high-frequency ground motions (e.g., Seyhan and Stewart 2014 – SS14; Hashash et al. 2020).

The ground motion models (GMMs) used in California are from the NGA-West2 project (Bozorgnia et al. 2014). These models have site response components that apply either globally for active tectonic regions (e.g., Boore et al. 2014) or apply for broad regions (e.g., California vs. Japan; Campbell and Bozorgnia 2014). However, differences have been observed when investigating regional site effects at a more local scale (e.g., Landwehr et al. 2016; Nweke et al. 2022; Wang et al. 2022a). The peaty-organic soils in the Delta have characteristic V_{S30} values typically in the range of 100 to 200 m/s, which is softer than the lower limit for NGA-West2 site response models, as shown in Figure 5.1. Furthermore, the soft soils in the Delta typically overlie relatively firm, non-organic soils, which can give rise to more pronounced impedance and resonance effects than would be typical at non-Delta sites. Accordingly, the peaty soils in the Delta produce site effects that are not well represented by site terms in current GMMs.

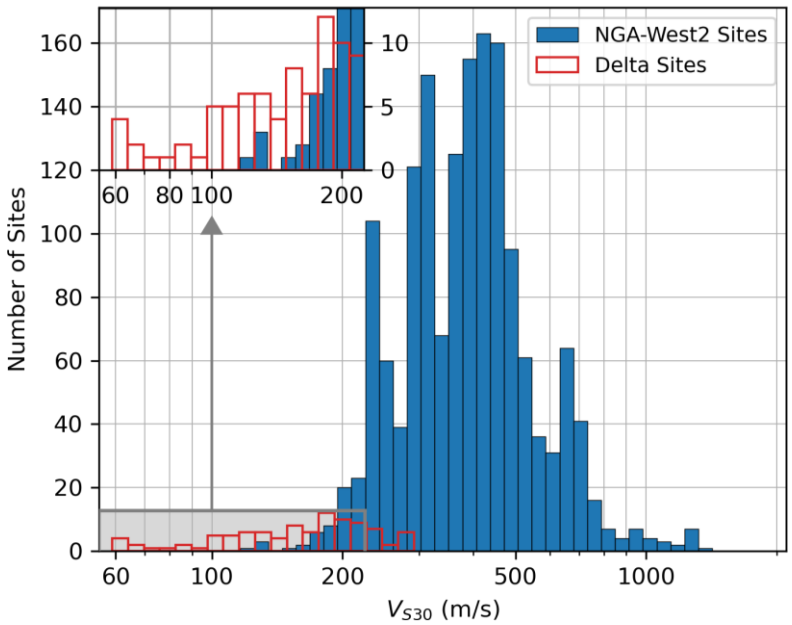


Figure 5.1: V_{S30} distributions of sites used in NGA-West2 models and at Delta sites (red).

The peaty organic soils in the Delta affect seismic risk for the regional infrastructure in two ways: (1) seismic ground motions are affected by site amplification and (2) ground deformation potential associated with permanent shear and/or volumetric deformations in peat (Shafiee et al. 2015). This chapter is concerned with the first of these issues (site response). In this chapter I utilize the expanded ground motion database described in Section 2.4 to evaluate empirical site response in the Delta. These data have relatively low amplitudes (i.e., weak motions), for which significant nonlinear effects are not expected, thus allowing me to investigate the features of local site response from non-ergodic analyses. Following this introduction, I describe the selected data and then present data analyses that include residuals calculations. The results are used to develop site response models conditioned on the site parameters discussed in Section 3.3.

5.2. Non-Ergodic Site Responses for Delta Region Stations

In this section I use ground motion recordings from sites in the Delta subregion as defined by DWR (2018) to derive site-specific site responses for use in subsequent modeling. The non-ergodic method of site response analysis that is applied here is conceptually similar to Stewart et al. (2017).

The Delta subregion includes sites with and without peaty-organic soils. This region contains 54 seismic instruments. At the time of writing, 43 of these stations have recorded ground motions from 69 events (**M** ≥ 4) in the GMDB discussed in Chapter 2. The vast majority of these ground motions were processed as part of this study.

5.2.1. Data Selection and Attributes

Table 5.1 provides the metadata, including V_{S30}, peak parameters derived from mHVSR [f_p, c_0, c_1, a_p, and w which are defined in Section 3.3.3(b)], and the number of usable recordings, for these 43 stations. Magnitude-distance criteria defined by Boore et al. (2014; BSSA14) are enforced to avoid potential bias with instrument trigger levels, and operate by eliminating data beyond a maximum distance where an appreciable fraction of recordings may not appear in the data as a result of having amplitudes below trigger or noise thresholds. Event metadata for the 69 events are provided in Table 5.2. Figure 5.2 presents a map of the locations of the events and stations used during model development.

Table 5.1: Metadata for Delta stations with ground motion records.

Station Code	Latitude (°)	Longitude (°)	V_{S30} (m/s)	t_p (m)	HVSR Peak	c_0	c_1	a_p	f_p (Hz)	w	No. of Records
BK_JEPS	38.2579	-121.8252	389.8	0.00	Y	0.833	1.457	2.291	0.661	0.495	20
BK_TWIT	38.0971	-121.6832	157.6	3.10	Y	0.944	2.940	3.884	0.691	0.313	21
CE_57195	37.9753	-121.3140	306.2	0.00	N	-	-	-	-	-	19
CE_57531	37.9332	-121.6956	250.6	0.00	Y	0.814	1.507	2.322	0.750	0.500	18
CE_57534	37.9119	-121.6219	238.3	0.00	N	-	-	-	-	-	28
CE_67215	37.9925	-121.6384	227.5	0.00	Y	1.100	0.914	2.014	0.695	0.383	9
CE_67266	38.0179	-121.7516	271.6	0.00	Y	1.589	2.276	3.865	1.701	0.121	2
CE_67523	38.0183	-121.7509	272.0	0.00	Y	1.100	0.944	2.044	2.889	0.264	1
CE_67533	38.0155	-121.6396	188.4	0.91	N	-	-	-	-	-	16
CE_67557	38.0141	-121.8148	324.6	0.00	N	-	-	-	-	-	2
CE_67587	38.0390	-121.8986	226.8	0.00	Y	1.100	4.150	5.250	1.105	0.289	20
CE_67615	38.0005	-121.7830	287.1	0.00	N	-	-	-	-	-	13
NC_C057	37.9555	-121.6979	344.9	0.00	N	-	-	-	-	-	13
NP_DIX	38.3771	-121.8424	284.8	0.00	Y	1.200	1.026	2.226	0.689	0.313	39
NP_EMR	38.0605	-121.4993	134.1	3.75	Y	1.000	3.830	4.830	1.030	0.285	17
NP_KIR	38.0551	-121.4582	158.5	3.03	Y	1.456	7.900	9.356	2.082	0.347	29
NP_LVA3	38.0328	-121.7631	154.6	4.36	Y	1.750	3.430	5.180	1.000	0.251	6
NP_LVA4	38.0335	-121.7634	139.8	4.57	Y	1.106	5.088	6.194	0.855	0.343	11
NP_LVB3	37.9102	-121.5649	195.0	0.54	NA	NA	NA	NA	NA	NA	4
NP_LVB4	37.9097	-121.5656	190.1	2.03	NA	NA	NA	NA	NA	NA	6
NP_MCD	37.9802	-121.4735	105.8	6.32	N	-	-	-	-	-	11
NP_PLA	37.7987	-121.4632	184.7	0.00	N	-	-	-	-	-	29
NP_SIA	38.0503	-121.7367	135.8	4.92	N	-	-	-	-	-	38
WR_CKR	38.3145	-121.4920	319.9	0.00	N	-	-	-	-	-	26
WR_CLFN	37.8570	-121.5730	160.6	3.05	N	-	-	-	-	-	9
WR_CLFS	37.8273	-121.5604	185.2	0.00	N	-	-	-	-	-	14

Station Code	Latitude (º)	Longitude (º)	V_{S30} (m/s)	t_p (m)	HVSR Peak	c_0	c_1	a_p	f_p (Hz)	w	No. of Records
WR_HOLT	37.9551	-121.4227	257.0	0.40	Y	1.230	1.438	2.668	0.647	0.320	7
WR_MOFF	38.0925	-121.8849	169.4	0.00	Y	0.970	3.180	4.150	1.649	0.319	1
WR_MONN	38.0929	-121.8855	152.0	2.62	Y	1.230	4.248	5.478	1.314	0.176	7
WR_SHER	38.0306	-121.7450	122.4	10.06	Y	1.152	4.455	5.607	0.732	0.402	7
WR_SIFF	38.0311	-121.7450	103.4	6.55	Y	1.019	2.645	3.665	0.735	0.245	4
WR_STNI	38.1197	-121.5401	165.8	6.68	Y	1.790	1.320	3.110	1.822	0.100	4
YU_CEC	37.9381	-121.5968	228.3	0.00	N	-	-	-	-	-	9
YU_HMT	37.9749	-121.5813	191.5	3.21	Y	1.290	2.856	4.146	1.553	0.239	4
YU_HOL1	37.9748	-121.5818	134.8	3.87	Y	1.150	2.616	3.766	2.179	0.385	4
YU_HOL2	37.9733	-121.5823	220.7	1.83	Y	1.190	4.151	5.341	2.124	0.266	9
YU_HOL3	37.9733	-121.5844	235.5	1.52	Y	1.050	4.195	5.245	4.653	0.272	4
YU_SMB	37.9977	-121.6262	193.2	0.60	Y	1.000	4.200	5.200	8.415	0.200	2
YU_SMT	37.9982	-121.6252	136.8	3.57	Y	1.520	1.150	2.670	3.158	0.200	5
YU_SRB	38.0109	-121.6249	180.7	1.41	Y	1.102	3.415	4.517	5.901	0.310	1
YU_SRT	38.0107	-121.6246	166.5	1.91	Y	1.592	2.804	4.396	2.879	0.223	13
YU_STF	37.9925	-121.4516	124.8	4.41	Y	1.230	2.504	3.734	2.731	0.193	2
YU_WHR	37.8984	-121.4508	197.0	0.00	Y	0.869	1.076	1.945	3.071	0.375	13

NA = No HVSR available.

Table 5.2: Metadata for events recorded by Delta stations.

event_id	Name	Datetime (UTC)	Latitude (º)	Longitude (º)	Depth (m)	M_w	Mechanism	Geyser Event	Usable Records	Delta Records
53	Livermore-01	1980-01-24 19:00	37.855	-121.816	12.0	5.80	SS	N	7	1
54	Livermore-02	1980-01-27 02:33	37.737	-121.740	14.5	5.42	SS	N	8	1
1910	N of Piedmont, CA	2015-08-17 13:49	37.840	-122.230	5.0	4.01	SS	N	208	1

event_id	Name	Datetime (UTC)	Latitude (º)	Longitude (º)	Depth (m)	M_w	Mechanism	Geyser Event	Usable Records	Delta Records
1913	ENE of Yountville, CA	2015-05-22 02:53	38.430	-122.250	12.8	4.09	SS	N	99	1
1915	South Napa, CA	2014-08-24 10:20	38.220	-122.310	11.1	6.02	SS	N	412	6
2012	Central California	2006-12-16 06:14	36.174	-120.294	9.6	4.15	SS	N	30	3
2013	NW of Mogul, NV	2008-04-26 06:40	39.523	-119.939	2.8	5.01	SS	N	44	2
2015	Central California	2011-08-27 07:18	36.584	-121.818	7.5	4.64	SS	N	218	6
2016	Northern California	2011-10-27 06:37	39.606	-120.472	12.0	4.73	SS	N	132	3
2017	Northern California	2012-02-13 04:47	38.793	-122.743	1.5	4.17	SS	Y	89	7
2018	El Cerrito, CA	2012-03-05 13:33	37.927	-122.310	8.1	3.98	SS	N	78	10
2019	Northern California	2012-05-05 09:23	38.796	-122.761	1.8	4.26	SS	Y	86	8
2020	Nevada	2012-06-23 03:51	39.318	-119.990	6.4	4.02	SS	N	62	8
2021	offshore Northern California	2012-07-08 12:05	39.480	-123.804	7.5	4.42	SS	N	50	5
2022	Northern California	2012-09-25 15:15	39.168	-123.166	10.8	4.51	SS	N	59	2
2023	Central California	2012-10-21 06:55	36.310	-120.856	8.6	5.29	SS	N	259	14
2025	WNW of Greenville, CA	2013-05-24 03:47	40.192	-121.060	8.0	5.69	SS	N	183	12
2026	NW of The Geysers, CA	2014-01-12 20:24	38.814	-122.816	1.9	4.54	NS	Y	91	5
2028	SSW of South Dos Palos, CA	2014-09-28 20:45	36.626	-120.834	7.0	4.43	NS	N	127	4
2029	S of San Juan Bautista, CA	2014-11-20 06:26	36.806	-121.536	6.2	4.23	SS	N	139	8
2033	SSE of Ridgemark, CA	2016-07-19 21:38	36.693	-121.330	5.9	4.21	SS	N	138	7
2034	NNE of Upper Lake, CA	2016-08-10 02:57	39.329	-122.802	14.5	5.09	SS	N	86	10
2035	NW of The Geysers, CA	2016-12-14 16:41	38.822	-122.841	1.5	5.14	SS	Y	101	9
2036	SW of Hawthorn, NV	2016-12-28 08:18	38.376	-118.899	11.3	5.66	SS	N	334	16
2040	NW of The Geysers, CA	2017-04-30 10:12	38.788	-122.777	2.8	3.99	SS	Y	99	6
2041	ESE of Alum Rock, CA	2017-10-10 00:53	37.314	-121.672	9.7	4.09	SS	N	201	11
2042	NE of Gonzales, CA	2017-11-13 19:31	36.631	-121.244	6.3	4.58	SS	N	198	12
2043	SE of Berkeley, CA	2018-01-04 10:39	37.855	-122.257	12.3	4.37	SS	N	136	13
2044	NE of The Geysers, CA	2018-01-18 05:55	38.785	-122.743	1.8	4.23	NS	Y	93	6

event_id	Name	Datetime (UTC)	Latitude (º)	Longitude (º)	Depth (m)	M_w	Mechanism	Geyser Event	Usable Records	Delta Records
2045	WSW of Byron, CA	2019-07-16 20:11	37.819	-121.757	12.4	4.30	SS	N	201	14
2046	SSE of Pleasant Hill, CA	2019-10-15 05:33	37.938	-122.057	14.0	4.46	SS	N	297	18
2047	SSE of Tres Pinos, CA	2019-10-15 19:42	36.646	-121.274	10.1	4.71	SS	N	272	17
2048	WSW of The Geysers, CA	2019-11-03 20:34	38.775	-122.767	3.1	4.15	SS	Y	132	4
2049	NNW of Cholame, CA	2019-12-17 18:29	35.806	-120.356	6.3	4.28	SS	N	162	3
2050	WNW of Toms Place, CA	2020-02-01 18:36	37.589	-118.822	9.8	4.43	SS	N	150	8
2051	NNE of New Idria, CA	2020-03-03 15:01	36.500	-120.617	10.9	3.99	RS	N	123	6
2052	Northern California	2006-10-20 17:00	38.867	-122.787	2.5	4.58	NS	Y	62	1
2053	Northern California	2006-11-09 08:38	39.359	-123.282	4.1	4.05	SS	N	31	2
2054	Northern California	2007-04-24 21:08	38.795	-122.797	1.7	4.46	NS	Y	72	3
2055	Northern California	2008-05-30 04:48	38.776	-122.764	1.0	4.14	NS	Y	51	5
2056	Northern California	2009-01-04 17:27	38.782	-122.773	3.8	4.27	SS	Y	63	8
2057	Northern California	2011-03-01 02:19	38.815	-122.820	2.3	4.43	NS	Y	63	6
2058	WSW of Cobb, California	2013-03-14 09:09	38.812	-122.786	1.3	4.44	SS	Y	171	9
2059	NE of The Geysers, CA	2017-02-21 00:57	38.796	-122.735	1.3	4.16	NS	Y	63	4
2060	NW of The Geysers	2018-05-10 02:58	38.810	-122.797	1.9	4.19	SS	Y	116	7
2061	NW of The Geysers	2019-04-15 11:54	38.815	-122.799	1.2	4.11	SS	Y	88	6
2062	NW of The Geysers	2019-06-11 07:46	38.814	-122.823	2.0	4.09	NS	Y	105	8
2063	N of Johnson Lane, Nevada	2020-03-21 01:33	39.111	-119.736	17.5	4.81	NS	N	223	16
2064	WSW of Laytonville, CA	2016-11-06 13:00	39.660	-123.637	3.5	4.06	SS	N	34	1
2065	WSW of Laytonville, CA	2017-12-14 04:57	39.608	-123.646	11.0	4.25	NS	N	61	1
2066	NNW of Redwood Valley, CA	2019-04-29 07:16	39.348	-123.235	18.0	4.33	NS	N	96	5
2067	NNE of Fremont, CA	2015-07-21 09:41	37.578	-121.974	8.4	3.97	SS	N	286	7
2070	NNW of The Geysers, CA	2020-03-25 11:57	38.790	-122.765	2.2	4.06	NS	Y	159	10
2071	WSW of Toms Place, CA	2020-04-05 19:20	37.527	-118.853	14.0	4.05	NS	N	150	9
2072	SE of Bodie, CA	2020-04-11 14:36	38.053	-118.733	11.0	5.24	NS	N	317	16

event_id	Name	Datetime (UTC)	Latitude (º)	Longitude (º)	Depth (m)	M_w	Mechanism	Geyser Event	Usable Records	Delta Records
2073	SE of Bodie, CA	2020-04-11 16:22	38.042	-118.758	11.0	4.57	NS	N	187	16
2074	Monte Cristo Range, NV Earthquake	2020-05-15 11:03	38.169	-117.850	11.5	6.49	NS	N	522	19
2075	SSE of Lone Pine, CA	2020-06-24 17:40	36.447	-117.975	8.0	5.80	RS	N	481	14
2076	NW of The Geysers, CA	2020-07-05 16:14	38.788	-122.766	5.0	3.97	RS	Y	121	5
2077	Central California	2008-12-21 17:35	36.675	-121.300	6.8	4.02	NS	N	100	9
2078	SSW of Petrolia, CA	2019-06-23 03:53	40.274	-124.300	18.0	5.58	NS	N	166	11
2079	SE of Bodie, CA	2020-05-31 01:07	38.045	-118.762	8.1	4.34	NS	N	168	14
2080	ESE of Willits, CA	2020-08-19 00:55	39.361	-123.246	11.0	4.22	NS	N	81	2
2081	W of Lakeport, CA	2020-12-06 15:03	39.038	-123.067	5.0	4.41	NS	N	158	4
2082	NW of Pinnacles, CA	2021-01-02 14:42	36.609	-121.213	5.0	4.32	NS	N	274	10
2083	SE of Aromas, CA	2021-01-17 04:01	36.867	-121.611	11.0	4.20	NS	N	333	10
2085	NW or Truckee, CA	2021-05-07 04:35	39.461	-120.315	14.0	4.64	NS	N	218	15
2102	2019 Ridgecrest EQ Sequence Mag 7.06	2019-07-06 03:19	35.770	-117.599	8.0	7.06	SS	N	719	2

Mechanism abbreviations: SS = Strike-Slip; NS = Normal-Slip; RS = Reverse-Slip

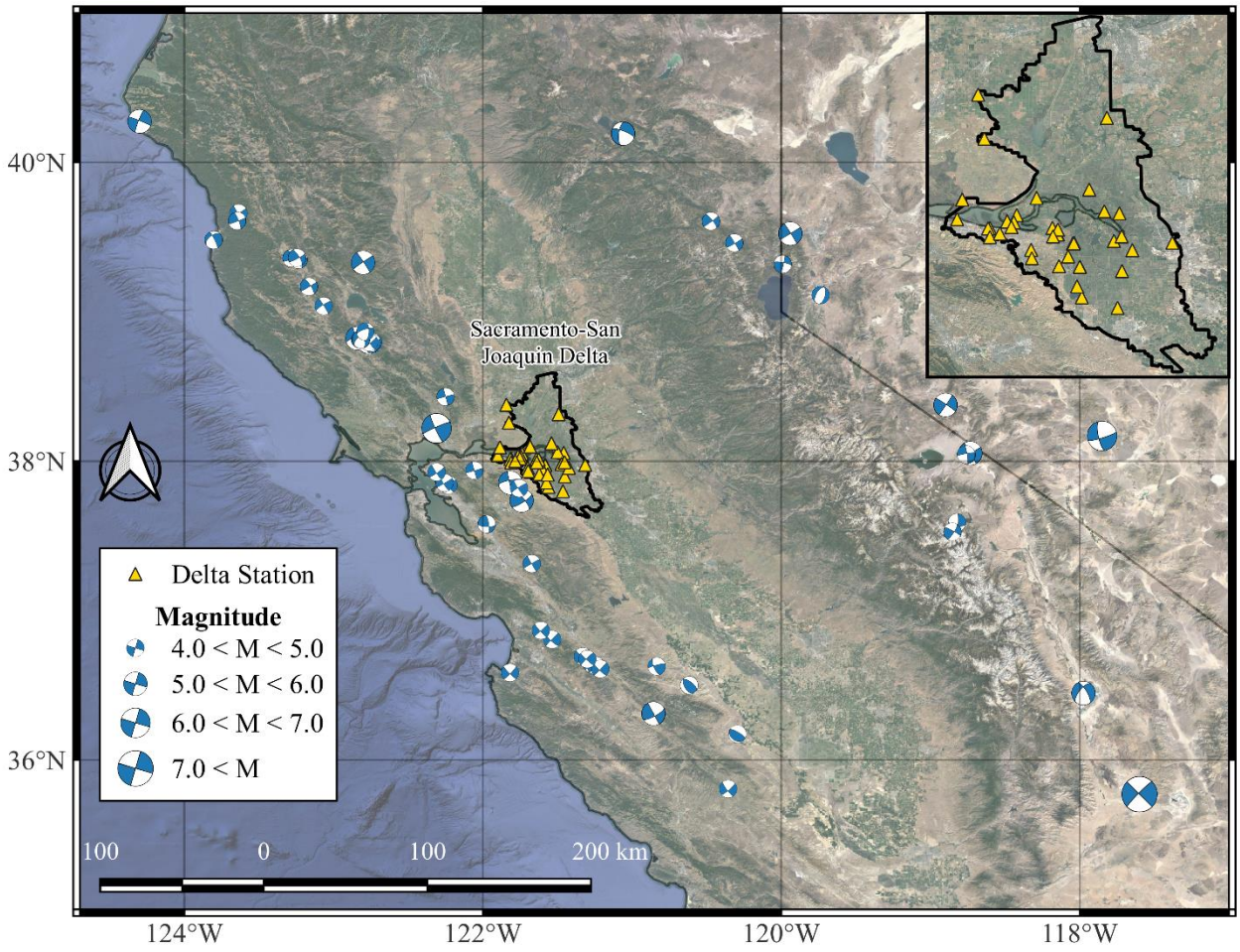


Figure 5.2: Map of northern California showing locations of Delta stations and event locations considered in this study.

Figure 5.3 presents the distribution of the subregional dataset with respect to **M**, R_{JB}, and V_{S30}. Delta stations have V_{S30} values ranging from 100 to 400 m/s, with a significant fraction of recordings coming from stations with V_{S30} < 200 m/s. Most data are associated with small magnitude events (**M** < 4.75) originating relatively far from the Delta (R_{JB} > 100 km). Accordingly, these motions are relatively weak as shown by the distribution of median component peak ground accelerations (*PGA*) in Figure 5.4. These observations provide evidence that nonlinearity is not expected to be significant in the dataset, therefore linear site response is anticipated.

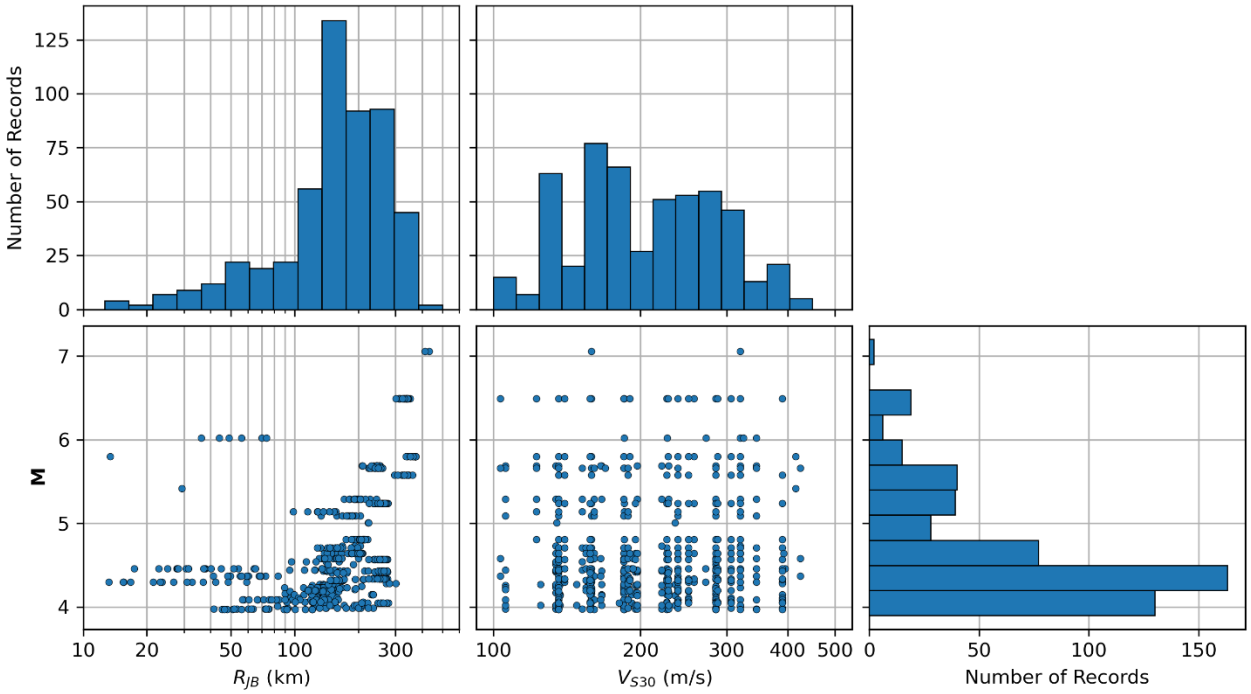


Figure 5.3: Data distribution for ground motions recorded at Delta stations.

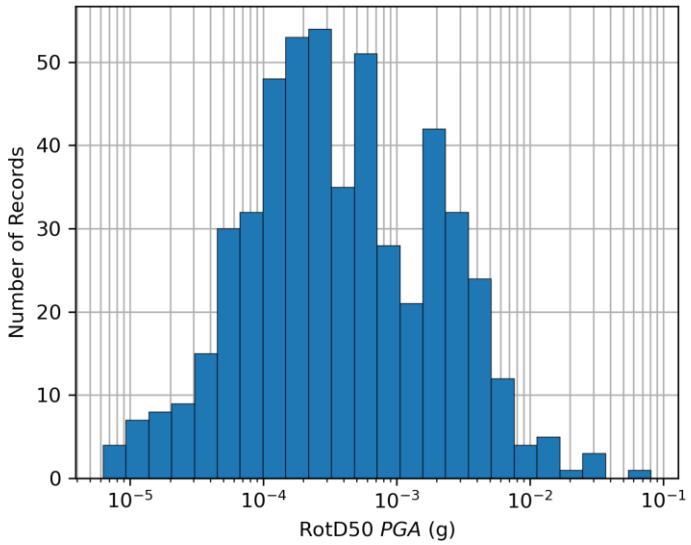


Figure 5.4: Distribution of RotD50 peak ground acceleration (*PGA*) at Delta stations.

5.2.2. Residuals Calculations

Total residuals (R_{ij}) are calculated from RotD50 intensity measures computed for the selected data from Section 5.2.1:

$$R_{ij} = \ln(Y_{ij}) - (\mu_{\ln Y})_{ij} \tag{5.1}$$

where $\ln(Y_{ij})$ is the intensity measure for event i and recording j and $(\mu_{\ln Y})_{ij}$ is the natural log mean prediction of that intensity measure using a GMM. I implement the GMM described in Chapter 4, which is modified from BSSA14 to include induced source corrections for the Geysers region and the subregional anelastic path model. This model takes the form,

$$(\mu_{\ln Y})_{ij} = (c_0 + \Delta c_{0,r} + Ie_I) + F_E + F_P + F_S \tag{5.2}$$

where F_E and F_S represent the source and site terms of BSSA14, respectively, and F_P is described by Equation (4.11). The modifications made to BSSA14 are necessary to remove systematic induced-source and path effects, which otherwise would bias the non-ergodic site responses.

Total residuals (Equation 5.1) are partitioned into event terms ($\eta_{E,i}$) and within-event residuals (δW_{ij}) using data for all **M** > 4.0 events in the GMDB (for shallow crustal type events) by applying mixed-effects analyses (Abrahamson and Youngs 1992; Gelman et al. 2014),

$$R_{ij} = c_k + \eta_{E,i} + \delta W_{ij} \tag{4.6, 5.3}$$

These analyses exclude observations from Delta stations. I assume that event biases are accurately estimated using only non-Delta observations (> 27 observations for each event, with the exception of events 53 and 54). In doing so, $\eta_{E,i}$ values are independent of Delta observations, thus eliminating the need to iterate to account for the influence of Delta observations on $\eta_{E,i}$. Within-event residuals (δW_{ij}) for Delta stations are obtained using Equation (5.3) and represent errors in GMM predictions that can be attributed to a combination of systematic site effects at each station

and random, event-to-event path errors. If the path errors are indeed random, which they should be following the work presented in Chapter 4, they will average to zero when summed over many observations. Within-event residuals are further partitioned into site terms ($\eta_{S,j}$) and remaining residuals (ε_{ij}) using additional mixed effects analyses,

$$\delta W_{ij} = \eta_{S,j} + \epsilon_{ij} \quad (4.7, 5.4)$$

The site terms ($\eta_{S,j}$) represent approximately the mean misfit of the site ground motions from the event-adjusted GMM, which can be used to estimate site response (Section 5.2.3). The standard errors ($SE_{\eta_{S,j}}$) of site terms are also are estimated.

5.2.3. Observed Linear Site Response

Site response models (F_S), whether site-specific or ergodic, usually can be expressed in the general form of:

$$F_S = F_{lin} + F_{nl} \quad (5.5)$$

where F_{lin} quantifies the linear amplification and F_{nl} quantifies nonlinear effects. For a particular site, Stewart et al. (2017) suggested the following form:

$$F_S = f_1 + f_2 \ln\left(\frac{x_{IMr}+f_3}{f_3}\right) \quad (5.6)$$

where the first and second terms represent F_{lin} and F_{nl}, respectively. x_{IMr} is the amplitude of shaking for a reference site condition (generally rock) for a particular event at a particular site (expressed as an intensity measure, which is often *PGA*), f_2 represents the slope (generally negative) in ln(amplification)-ln(x_{IMr}) space for $x_{IMr} \gg f_3$, and f_3 represents a transitional value of the reference site intensity measure below which the site response is nearly linear, and above which the trend of amplification with x_{IMr} is nearly linear in log-log space (e.g., 0.1 g in SS14).

With the specific form of the site response model having been established, I next turn to the estimation of site response using non-ergodic methods. Those methods take the site response as the sum of the site response model (used in the GMM applied for residuals analyses) and the site term. For this application, we estimate the linear site response (f_1) as the sum of the site term (established from weak motion data) and the linear part of the site response GMM,

$$(f_1)_j^o = \eta_{S,j} + [F_{S,j} - F_{nl,j}] \tag{5.7}$$

where $(f_1)_j^o$ represents the data-derived estimate of the linear site response (the 'o' superscript is for "observed" and should not be confused with the model estimate of f_1 from Equation 5.6). The subtraction of F_{nl} in Equation (5.7) is to remove first-order nonlinear effects in the site response estimate, although this term is nearly zero in the present case as demonstrated subsequently (Section 5.3.5).

5.3. Site Amplification Model Development

The modeling approach outlined in this section aims to develop a linear site amplification model for the Delta region using weak ground motions (relatively low amplitudes), where significant nonlinear effects are not expected. The computed $(f_1)_j^o$ values represent the average site amplification observed over multiple events for station *j*, and are used as the observations from which site amplification models can be developed.

The advantage of using $(f_1)_j^o$ over δW_{ij} during model development is that $(f_1)_j^o$ reduces the influence of the number of observations at each site during regression methods. This is important because sites with the most recordings would otherwise control the regional site response, and sites with few recordings would have limited impact. However, the disadvantage is that it can be difficult to accurately estimate $(f_1)_j^o$ at sites with relatively few recordings – even

through mixed-effects methods. To address this issue, I use only data from stations which have recorded at least four events (per intensity measure) and consider their associated $SE_{\eta_{S,j}}$ during model development through weighted least squares (WLS) regression (Strutz 2016). Each observation is assigned a weight (w_j) inversely proportional to its $SE_{\eta_{S,j}}$:

$$w_j = \frac{n\left(SE_{\eta_{S,j}}\right)^{-2}}{\sum_{j=1}^{n}\left(SE_{\eta_{S,j}}\right)^{-2}} \tag{5.8}$$

where n represents the number of sites used during regression for each intensity measure ($n \leq 36$). It follows that stations with smaller $SE_{\eta_{S,j}}$ values will have greater influence during the regression, however the influence is inversely proportional to the uncertainty of the $(f_1)_j^o$ estimate instead of the number of recordings.

5.3.1. Linear V_{S30} – Scaling

Ergodic site response models are typically conditioned on V_{S30} to capture first-order site effects (e.g., SS14; Parker and Stewart 2022). Figure 5.5 presents the distribution of V_{S30} for Delta stations with at least four ground motion records, which ranges from 103 to 390 m/s. Stations with V_{S30} > 275 m/s are unlikely to have peat (illustrated in Figure 3.13), hence the Delta subregion includes both peat and non-peat sites. To expand the size of the database for non-peat sites, I supplemented the dataset with data from 45 "non-Delta" stations in the surrounding area, as shown in Figure 5.6. The motivation for including these additional data at this stage of model development is as follows:

1. The Delta is located within the Central Valley – a large sedimentary basin structure. I anticipate basin effects to affect Delta stations, which would also influence site response in the greater Delta area.

2. To investigate basin effects (presented in Section 5.3.2), residuals must be unbiased with respect to first-order site effects modeled by V_{S30}. Therefore, the V_{S30}-scaling model should be suited for all site conditions in the Delta region, which includes sites without peaty-organic soils.
3. A V_{S30}-scaling model developed using solely data from Delta stations would only be applicable to the parametric range of the data used during development (i.e., $V_{S30} < 300$ m/s). The resulting model would likely be biased for stiffer soil sites, so data in the general vicinity of the Delta with similar geologic environments (i.e., within the Central Valley) are used to constrain the model for moderately-soft-to-stiff site conditions.

Figure 5.5 includes the distribution of V_{S30} for non-Delta stations in the region surrounding the Delta. There are five non-Delta stations with 230 m/s $\leq V_{S30} \leq$ 280 m/s, which span a V_{S30} range which includes Delta stations. However, the majority of non-Delta stations are considerably stiffer. I assume that the inclusion of non-Delta stations does not have a significant impact on the resulting model for the soft-soil conditions ($V_{S30} <$ 200 m/s), which are the primary concern of this study (this is confirmed during a subsequent stage of model development).

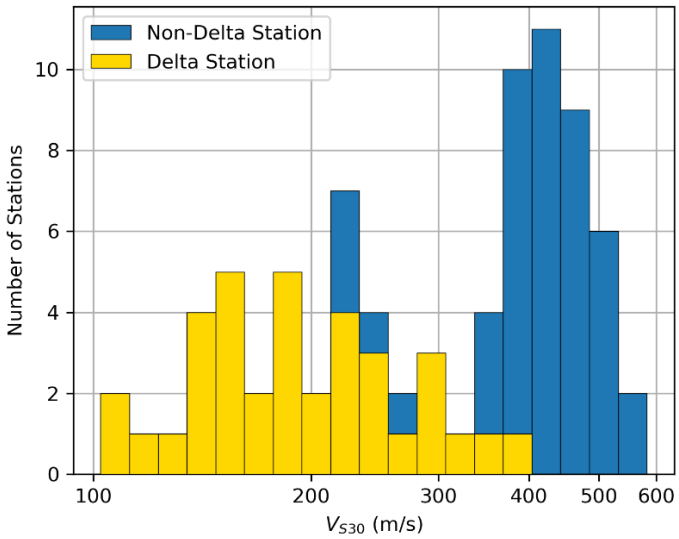


Figure 5.5: Distribution of V_{S30} at sites with at least four ground motion records for Delta stations (yellow) and non-Delta stations in the surrounding region (blue).

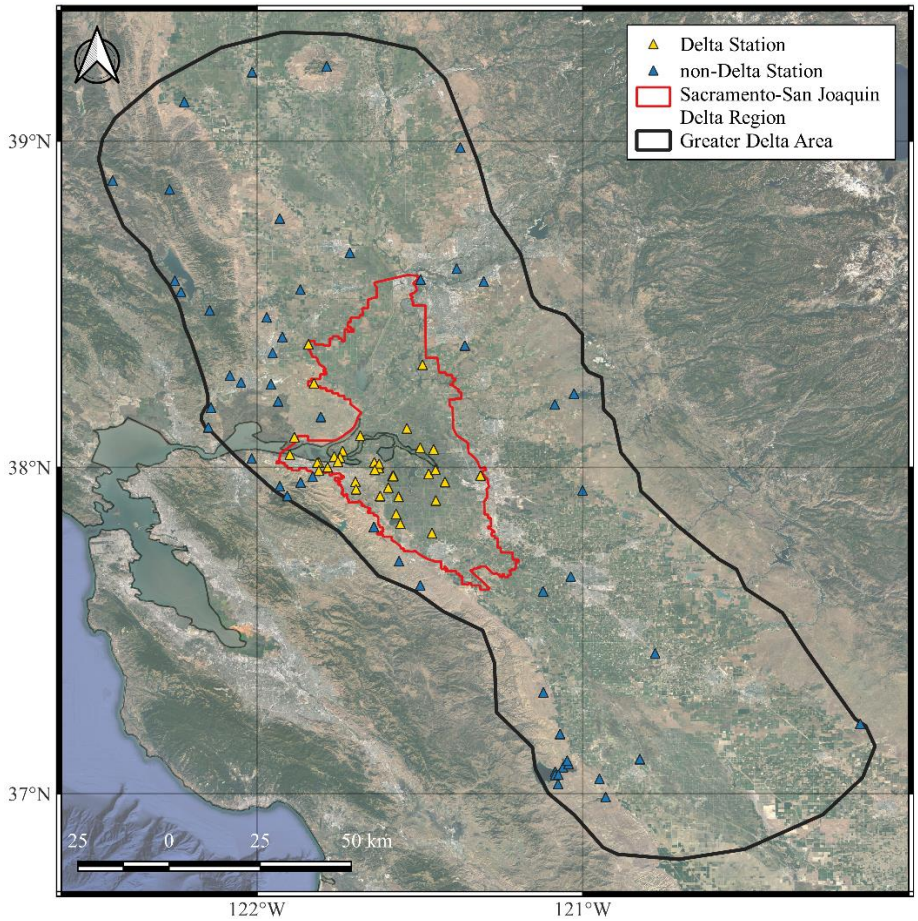


Figure 5.6: Map of the Greater Delta Area, showing the locations of Delta and non-Delta stations.

Figure 5.7 presents plots of $(f_1)_j^o$ against V_{S30} for *PGV*, *PGA*, and *PSA* at 0.1, 0.3, 1.0, 3.0, and 10.0 s, compared to the linear component of the ergodic SS14 model. Binned means of the data are plotted with their 95% confidence intervals (the bins are equally spaced on the log scale and the binned values are plotted at the median V_{S30} within each bin). Several noteworthy observations can be made:

1. The ergodic model reasonably captures the observed amplification and V_{S30}-scaling [slope of $(f_1)_j^o$ with V_{S30}] of non-Delta stations for *PGV*, *PGA*, and *PSA* at periods shorter than approximately 3.0 s. However, SS14 under-predicts the amplification and V_{S30}-scaling for non-Delta stations for longer periods. This observation provides evidence that the ergodic model is generally appropriate for typical non-peaty sites in the greater Delta area. The misfits at long periods may result from sedimentary basin effects in the Central Valley.
2. The ergodic model over-predicts site response of Delta stations for *PGA* and periods shorter than about 1.0 s.
3. The ergodic model seems to reasonably capture the V_{S30}-scaling (i.e., slope with respect to V_{S30}) of Delta stations with V_{S30} > 200 m/s for periods shorter than about 1.0 s.
4. Across all observed intensity measures, the V_{S30}-scaling appears to saturate (i.e., slope goes to zero) for exceptionally soft soil sites. This observation is consistent with the findings of recent studies for other regions (Parker and Stewart 2022 for global subduction zones; Nweke et al. 2022 for southern California basins).
5. The ergodic model appreciably under-predicts the site response and V_{S30}-scaling for Delta stations at periods longer than about 3.0 s. This observation is similar to (1) and may result from Central Valley sedimentary basin response long-period site response.

6. Lastly, the variability of the $(f_1)_j^o$ data with respect to V_{S30} appears from visual inspection to be quite large (especially for short-to-intermediate periods). This suggests that site response in the Delta may be influenced by phenomena that are poorly represented by V_{S30}; subsequently, I consider additional parameters for their ability to capture additional site response effects.

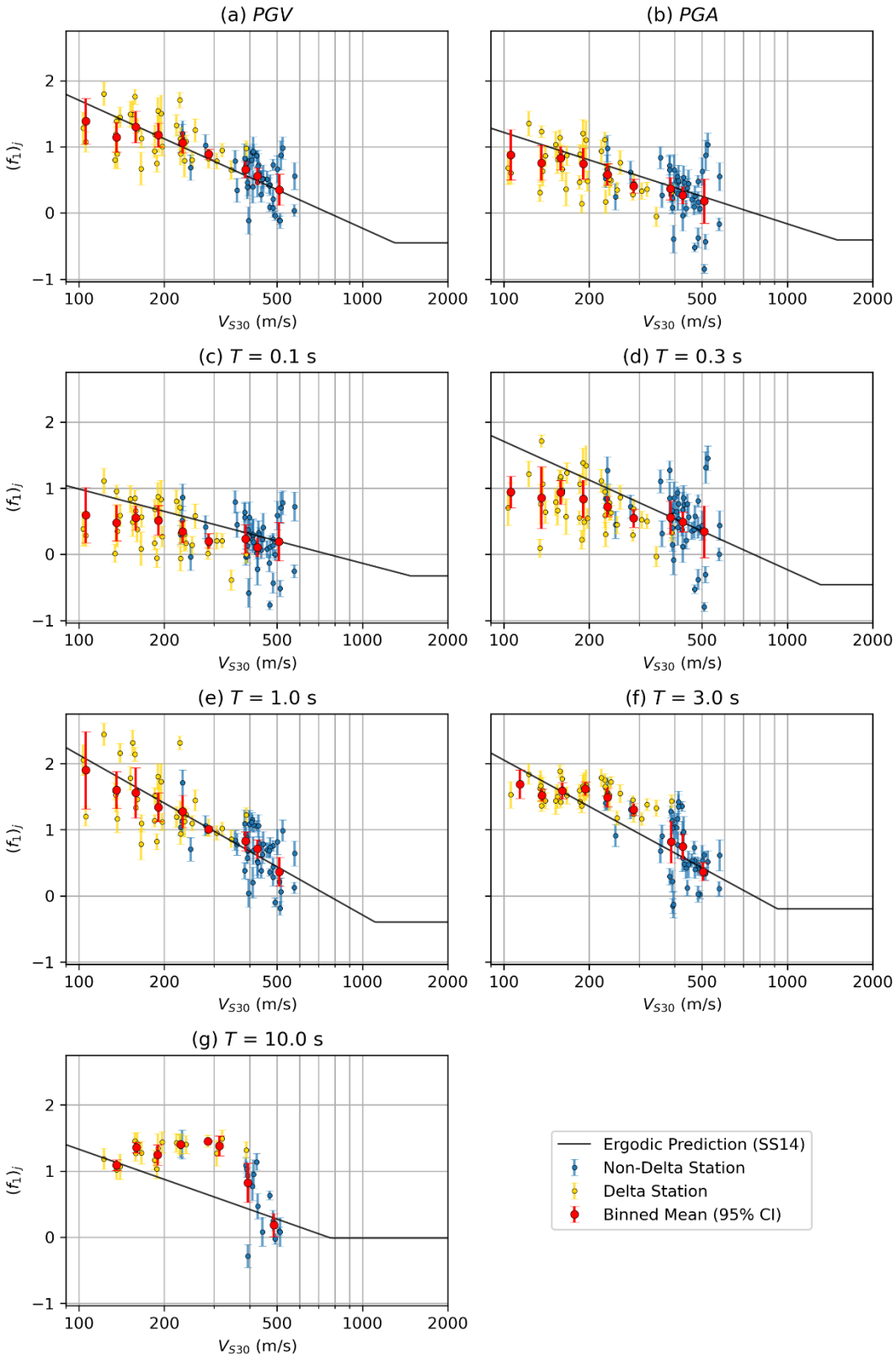


Figure 5.7: Plots of observed linear amplification $[(f_1)_j^o]$ versus V_{S30} for (a) *PGV*, (b) *PGA*, and *PSA* at (c) T = 0.1s, (d) T = 0.3s, (e) T = 1.0s, (f) T = 3.0s, and (g) T = 10.0s.

Based on visual inspection, I developed a functional form to represent the observed data trends,

$$F_{lin}(V_{S30}) = \begin{cases} c_1 \ln\left(\frac{V_{S30}}{V_1}\right) + c_2 \ln\left(\frac{V_1}{V_2}\right) + c \ln\left(\frac{V_2}{V_{ref}}\right) & V_{S30} < V_1 \\ c_2 \ln\left(\frac{V_{S30}}{V_2}\right) + c \ln\left(\frac{V_2}{V_{ref}}\right) & V_1 \leq V_{S30} < V_2 \\ c \ln\left(\frac{V_{S30}}{V_{ref}}\right) & V_2 \leq V_{S30} < V_c \\ c \ln\left(\frac{V_c}{V_{ref}}\right) & V_c \leq V_{S30} \end{cases} \quad (5.9)$$

where c_1, c_2, and c are V_{S30}-scaling coefficients, V_1, V_2, and V_c are limiting velocities, and V_{ref} is the reference site condition where F_{lin} = 0 (V_{S30} = 760 m/s). The first two model components (1st and 2nd rows) effectively represent the V_{S30}-scaling in the Delta for exceptionally soft and soft-to-moderate site conditions, respectively. The third and fourth components match the SS14 model; the only modification is an added lower limiting velocity on the third component (i.e., V_2). The coefficients c and V_c are adopted from SS14, and c_1, c_2, V_1, and V_2 are estimated for the Delta subregion.

The model is first fit to the data with all four free parameters using WLS regression. V_{S30}-scaling coefficients c_1 and c_2 are constrained to be zero or negative, since site response has been observed to increase as sediment stiffness decreases across many investigations spanning over 50 years (e.g., Borcherdt and Gibbs 1976, Seed et al. 1976, Idriss 1990, Borcherdt and Glassmoyer 1994). V_1 is constrained to be less than V_2, which is constrained to be less than or equal to V_{ref} to ensure a continuous function and enforce a reference velocity V_{ref} = 760 m/s. From these results, c_1 was identified to be the most stable parameter in terms of lacking sudden between-period fluctuations and was set to a constant value of zero, effectively removing the first term in the first model component of Equation (5.9).

A second WLS regression was performed with c_2, V_1, and V_2 estimated as free parameters. These results indicated that V_1 was the most stable parameter. Regressed values of V_1 were smoothed using the locally weighted scatterplot smoothing algorithm (Cleveland 1979). A third and fourth round of WLS regressions and smoothing were conducted to set V_2 and c_2, respectively. Figure 5.8 presents the as-regressed and final smoothed-coefficients of the proposed V_{S30}-scaling model. Coefficients are listed in Table 5.3, and Figure 5.9 compares the fit of SS14 and the proposed model to the data for the same intensity measures shown in Figure 5.7.

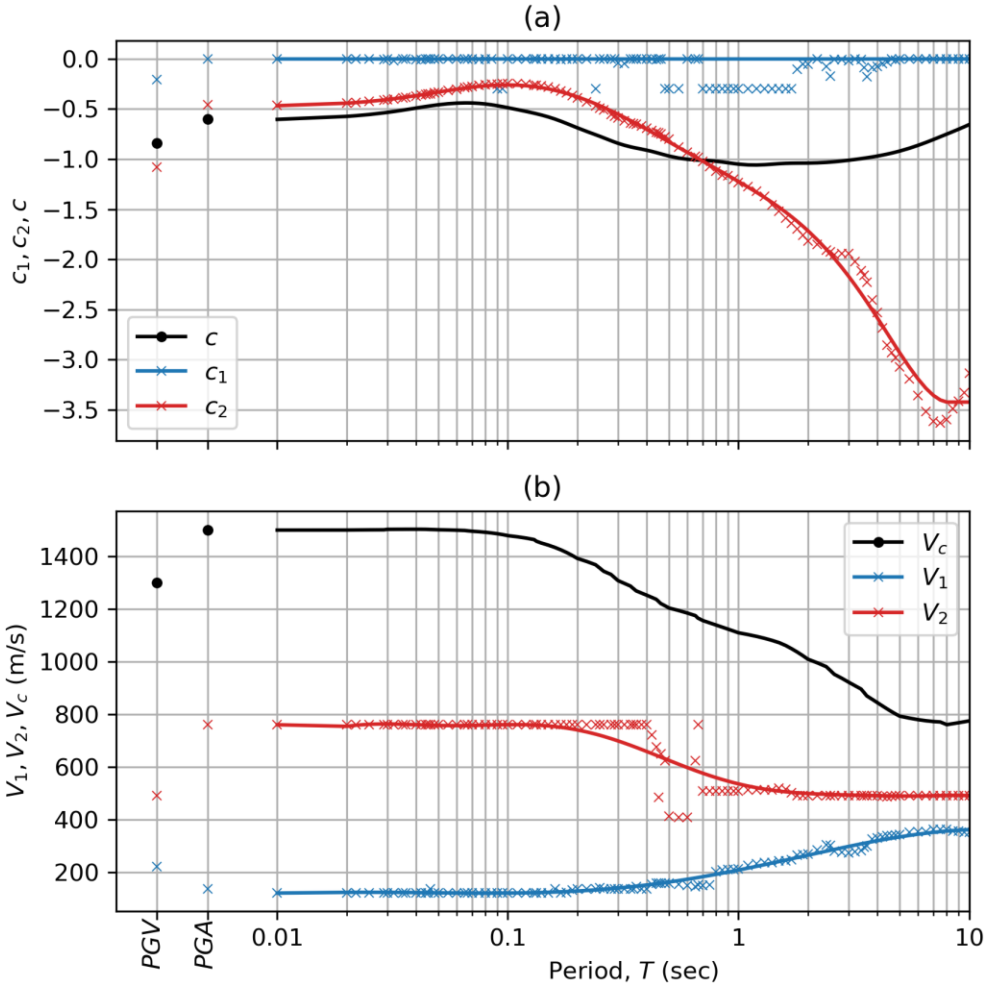


Figure 5.8: As-regressed and final-smoothed coefficients for V_{S30}-scaling model; (a) V_{S30}-scaling coefficients c, c_1, and c_2; (b) limiting velocities V_c, V_1, and V_2. In some cases, the result of multiple rounds of smoothing are shown as crosses.

Table 5.3: Coefficient values for c, c_1, c_2, V_c, V_1, and V_2.

Intensity Measure	c_1	V_1 (m/s)	c_2	V_2 (m/s)	c	V_c (m/s)	V_{ref} (m/s)
PGV	-0.20509	221.47	-1.07806	491.32	-0.84000	1300.00	760.00
PGA	0.00000	135.82	-0.45979	760.00	-0.60000	1500.00	760.00
T = 0.010 s	0.00000	122.35	-0.46808	760.00	-0.60372	1500.20	760.00
T = 0.020 s	0.00000	122.35	-0.43790	760.00	-0.57388	1500.36	760.00
T = 0.022 s	0.00000	122.35	-0.43182	760.00	-0.56675	1500.68	760.00
T = 0.025 s	0.00000	122.35	-0.42275	760.00	-0.55520	1501.04	760.00
T = 0.029 s	0.00000	122.35	-0.41004	760.00	-0.53850	1501.26	760.00
T = 0.030 s	0.00000	122.35	-0.40643	760.00	-0.53414	1502.95	760.00
T = 0.032 s	0.00000	122.35	-0.39943	760.00	-0.52529	1503.12	760.00
T = 0.035 s	0.00000	122.35	-0.38881	760.00	-0.51192	1503.24	760.00
T = 0.036 s	0.00000	122.35	-0.38503	760.00	-0.50752	1503.32	760.00
T = 0.040 s	0.00000	122.35	-0.36995	760.00	-0.49065	1503.35	760.00
T = 0.042 s	0.00000	122.35	-0.36244	760.00	-0.48290	1503.34	760.00
T = 0.044 s	0.00000	122.35	-0.35509	760.00	-0.47572	1503.13	760.00
T = 0.045 s	0.00000	122.35	-0.35155	760.00	-0.47236	1502.84	760.00
T = 0.046 s	0.00000	122.35	-0.34788	760.00	-0.46915	1502.47	760.00
T = 0.048 s	0.00000	122.35	-0.34089	760.00	-0.46321	1502.01	760.00
T = 0.050 s	0.00000	122.35	-0.33437	760.00	-0.45795	1501.42	760.00
T = 0.055 s	0.00000	122.35	-0.31860	760.00	-0.44787	1500.71	760.00
T = 0.060 s	0.00000	122.35	-0.30423	760.00	-0.44186	1499.83	760.00
T = 0.065 s	0.00000	122.35	-0.29096	760.00	-0.43951	1498.74	760.00
T = 0.067 s	0.00000	122.35	-0.28600	760.00	-0.43950	1497.42	760.00
T = 0.070 s	0.00000	122.35	-0.27961	760.00	-0.44040	1495.85	760.00
T = 0.075 s	0.00000	122.35	-0.27173	760.00	-0.44411	1494.00	760.00
T = 0.080 s	0.00000	122.35	-0.26531	760.00	-0.45020	1491.82	760.00
T = 0.085 s	0.00000	122.35	-0.26017	760.00	-0.45813	1489.29	760.00
T = 0.090 s	0.00000	122.35	-0.25682	760.00	-0.46732	1486.36	760.00
T = 0.095 s	0.00000	122.35	-0.25414	760.00	-0.47721	1482.98	760.00
T = 0.100 s	0.00000	122.35	-0.25327	760.00	-0.48724	1479.12	760.00
T = 0.110 s	0.00000	122.35	-0.25546	760.00	-0.50632	1474.74	760.00
T = 0.120 s	0.00000	122.35	-0.25972	760.00	-0.52438	1469.75	760.00
T = 0.130 s	0.00000	122.35	-0.26737	759.68	-0.54214	1464.09	760.00
T = 0.133 s	0.00000	122.35	-0.27054	759.17	-0.54752	1457.76	760.00
T = 0.140 s	0.00000	122.35	-0.27864	756.77	-0.56032	1450.71	760.00
T = 0.150 s	0.00000	122.42	-0.29305	750.93	-0.57962	1442.85	760.00
T = 0.160 s	0.00000	123.34	-0.30931	746.44	-0.60052	1434.22	760.00
T = 0.170 s	0.00000	125.02	-0.32719	743.63	-0.62252	1424.85	760.00

Intensity Measure	c_1	V_1 (m/s)	c_2	V_2 (m/s)	c	V_c (m/s)	V_{ref} (m/s)
T = 0.180 s	0.00000	127.46	-0.34627	741.27	-0.64486	1414.77	760.00
T = 0.190 s	0.00000	129.65	-0.36569	738.83	-0.66681	1403.99	760.00
T = 0.200 s	0.00000	131.86	-0.38570	736.22	-0.68762	1392.61	760.00
T = 0.220 s	0.00000	134.52	-0.42744	730.49	-0.72431	1380.72	760.00
T = 0.240 s	0.00000	136.07	-0.46906	724.54	-0.75646	1368.51	760.00
T = 0.250 s	0.00000	136.18	-0.48940	721.76	-0.77177	1356.21	760.00
T = 0.260 s	0.00000	136.17	-0.50886	718.94	-0.78697	1343.89	760.00
T = 0.280 s	0.00000	136.03	-0.54391	713.80	-0.81613	1331.67	760.00
T = 0.290 s	0.00000	135.92	-0.55842	711.27	-0.82950	1319.83	760.00
T = 0.300 s	0.00000	135.84	-0.57401	708.68	-0.84165	1308.47	760.00
T = 0.320 s	0.00000	135.82	-0.60141	702.98	-0.86175	1297.65	760.00
T = 0.340 s	0.00000	135.82	-0.62957	690.97	-0.87726	1287.50	760.00
T = 0.350 s	0.00000	135.82	-0.64156	684.88	-0.88375	1278.06	760.00
T = 0.360 s	0.00000	137.96	-0.65254	680.16	-0.88965	1269.19	760.00
T = 0.380 s	0.00000	143.87	-0.67259	670.54	-0.90038	1260.74	760.00
T = 0.400 s	0.00000	148.29	-0.69236	661.56	-0.91092	1252.66	760.00
T = 0.420 s	0.00000	154.05	-0.71227	652.47	-0.92241	1244.80	760.00
T = 0.440 s	0.00000	156.64	-0.73321	643.10	-0.93459	1237.03	760.00
T = 0.450 s	0.00000	157.29	-0.74383	638.34	-0.94075	1229.23	760.00
T = 0.460 s	0.00000	157.86	-0.75511	633.28	-0.94686	1221.16	760.00
T = 0.480 s	0.00000	158.69	-0.77858	623.37	-0.95863	1212.74	760.00
T = 0.500 s	0.00000	158.86	-0.80306	612.95	-0.96930	1203.91	760.00
T = 0.550 s	0.00000	167.18	-0.85829	607.39	-0.98925	1194.59	760.00
T = 0.600 s	0.00000	170.89	-0.91435	597.11	-1.00120	1184.93	760.00
T = 0.650 s	0.00000	175.24	-0.97039	585.04	-1.00780	1175.19	760.00
T = 0.667 s	0.00000	177.48	-0.98894	581.36	-1.00930	1165.69	760.00
T = 0.700 s	0.00000	183.28	-1.01742	574.21	-1.01170	1156.46	760.00
T = 0.750 s	0.00000	190.89	-1.05163	563.25	-1.01540	1147.59	760.00
T = 0.800 s	0.00000	197.85	-1.08814	552.73	-1.02100	1139.21	760.00
T = 0.850 s	0.00000	206.50	-1.12254	544.38	-1.02820	1131.34	760.00
T = 0.900 s	0.00000	209.46	-1.15520	543.47	-1.03600	1123.91	760.00
T = 0.950 s	0.00000	213.06	-1.18624	541.29	-1.04360	1116.83	760.00
T = 1.000 s	0.00000	217.16	-1.21816	537.68	-1.05000	1109.95	760.00
T = 1.100 s	0.00000	224.23	-1.27804	531.07	-1.05730	1103.07	760.00
T = 1.200 s	0.00000	229.98	-1.33802	525.56	-1.05840	1096.04	760.00
T = 1.300 s	0.00000	235.08	-1.39817	520.37	-1.05540	1088.67	760.00
T = 1.400 s	0.00000	238.99	-1.45739	515.41	-1.05040	1080.77	760.00
T = 1.500 s	0.00000	242.25	-1.51790	511.41	-1.04540	1072.39	760.00
T = 1.600 s	0.00000	248.08	-1.57875	507.67	-1.04210	1061.77	760.00

Intensity Measure	c_1	V_1 (m/s)	c_2	V_2 (m/s)	c	V_c (m/s)	V_{ref} (m/s)
T = 1.700 s	0.00000	254.18	-1.63804	505.87	-1.04040	1049.29	760.00
T = 1.800 s	0.00000	259.65	-1.67895	503.48	-1.03970	1036.42	760.00
T = 1.900 s	0.00000	265.49	-1.70020	501.64	-1.03950	1023.14	760.00
T = 2.000 s	0.00000	271.18	-1.73884	499.91	-1.03920	1009.49	760.00
T = 2.200 s	0.00000	273.48	-1.81489	497.09	-1.03680	995.52	760.00
T = 2.400 s	0.00000	275.99	-1.88944	495.09	-1.03230	981.33	760.00
T = 2.500 s	0.00000	276.46	-1.92405	494.26	-1.02940	966.94	760.00
T = 2.600 s	0.00000	278.68	-1.95588	493.52	-1.02620	952.34	760.00
T = 2.800 s	0.00000	277.29	-2.02305	492.39	-1.01900	937.52	760.00
T = 3.000 s	0.00000	278.12	-2.11165	491.68	-1.01120	922.43	760.00
T = 3.200 s	0.00000	285.99	-2.19226	491.18	-1.00320	908.79	760.00
T = 3.400 s	0.00000	292.42	-2.27232	490.93	-0.99512	896.15	760.00
T = 3.500 s	0.00000	296.02	-2.31592	490.82	-0.99100	883.16	760.00
T = 3.600 s	0.00000	300.79	-2.33828	490.74	-0.98682	870.05	760.00
T = 3.800 s	0.00000	310.68	-2.41052	490.63	-0.97826	857.07	760.00
T = 4.000 s	0.00000	318.94	-2.53754	490.58	-0.96938	844.48	760.00
T = 4.200 s	0.00000	327.33	-2.66729	490.56	-0.96012	832.45	760.00
T = 4.400 s	0.00000	336.53	-2.79888	490.56	-0.95049	821.18	760.00
T = 4.600 s	0.00000	339.52	-2.93050	490.61	-0.94050	810.79	760.00
T = 4.800 s	0.00000	341.65	-2.98538	490.64	-0.93018	801.41	760.00
T = 5.000 s	0.00000	343.95	-3.04025	490.68	-0.91954	793.13	760.00
T = 5.500 s	0.00000	348.59	-3.11297	490.80	-0.89176	785.73	760.00
T = 6.000 s	0.00000	352.61	-3.22161	490.90	-0.86286	779.91	760.00
T = 6.500 s	0.00000	356.10	-3.30329	490.99	-0.83355	775.60	760.00
T = 7.000 s	0.00000	358.06	-3.34135	491.07	-0.80457	772.68	760.00
T = 7.500 s	0.00000	358.96	-3.34668	491.15	-0.77665	771.01	760.00
T = 8.000 s	0.00000	359.90	-3.34912	491.24	-0.75033	760.81	760.00
T = 8.500 s	0.00000	358.21	-3.35052	491.33	-0.72544	764.50	760.00
T = 9.000 s	0.00000	356.26	-3.35141	491.41	-0.70161	768.07	760.00
T = 9.500 s	0.00000	354.19	-3.35200	491.49	-0.67850	771.55	760.00
T = 10.000 s	0.00000	352.06	-3.35200	491.56	-0.65575	775.00	760.00

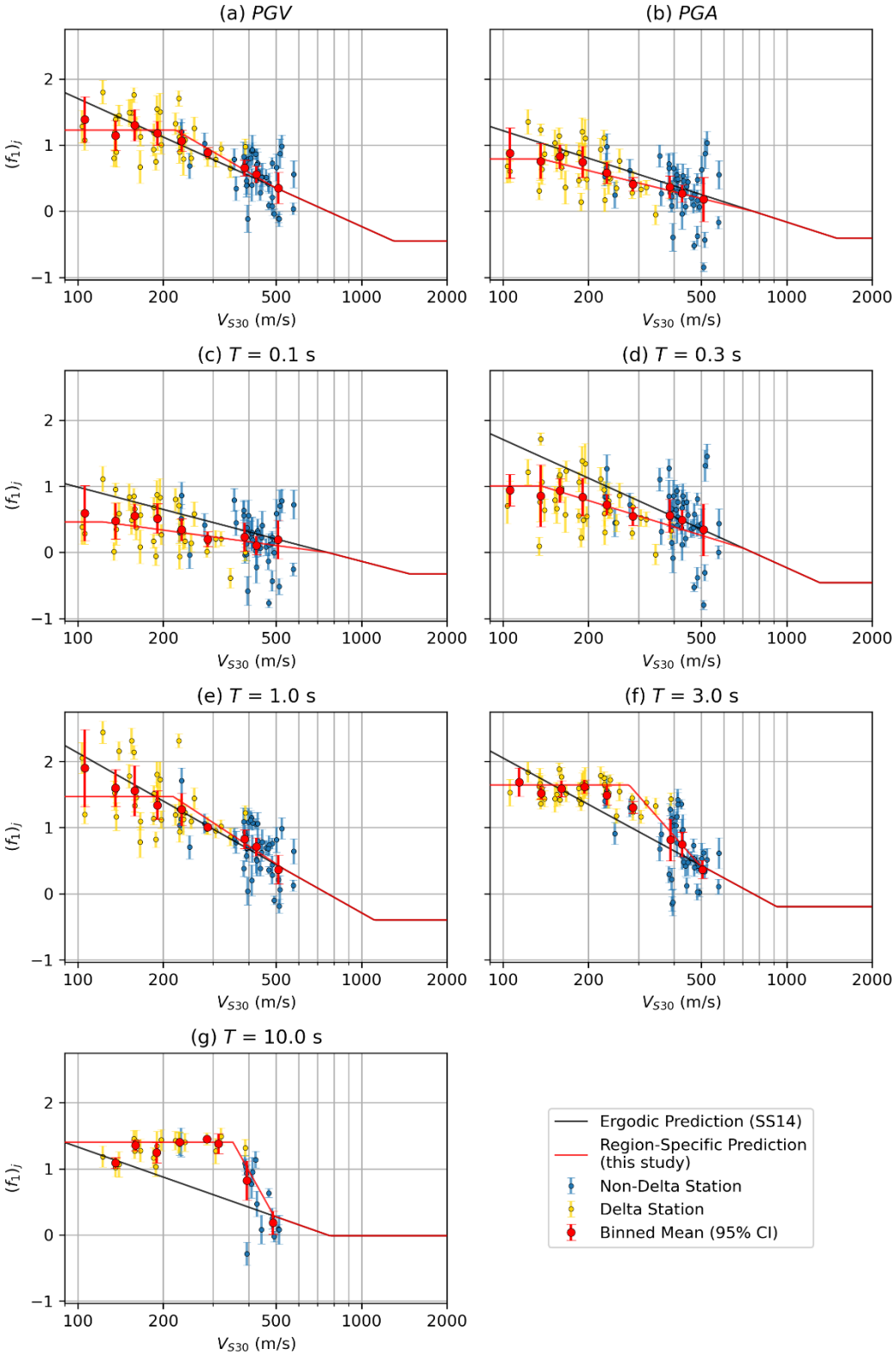


Figure 5.9: Plots comparing the fit of ergodic model (SS14) and region-specific model (this study) to observed linear amplification $[(f_1)_j^o]$ versus V_{S30} for (a) *PGV*, (b) *PGA*, and *PSA* at (c) T = 0.1s, (d) T = 0.3s, (e) T = 1.0s, (f) T = 3.0s, and (g) T = 10.0s.

Figure 5.10 compares median amplification as predicted by the ergodic model (SS14) and the Delta-specific V_{S30}-scaling model for V_{S30} = 100, 200, 300, 400, and 500 m/s. In general, the differences in amplification increase as V_{S30} decreases. This is to be expected since SS14 extrapolates V_{S30}-scaling for soft-soil conditions. The SS14 model does not have V_{S30}-saturation for soft sites, whereas Delta subregional data saturate at V_{S30} < 200 ~ 300 m/s. For all site conditions the region-specific amplification is smaller than SS14 at periods shorter than about 0.7 s, whereas long-period amplification is larger for V_{S30} = 200 – 400 m/s. Amplification predicted by the two models for non-Delta site conditions (i.e., V_{S30} > 400 m/s) have better agreement, which indicates that the transition from Delta-specific to SS14 for stiffer sites operates as expected.

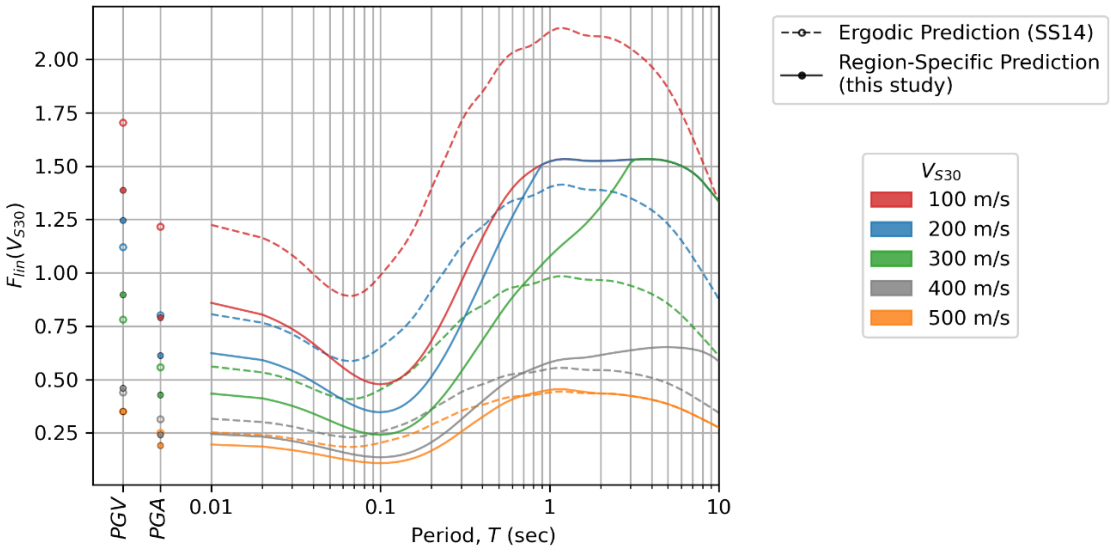


Figure 5.10: Comparisons of the predicted amplification from the proposed model for the Delta sub-region to the ergodic prediction from SS14. Amplification shown for V_{S30} = 100, 200, 300, 400, and 500 m/s.

As shown in Figures 5.8(b) and 5.10, an unexpected feature of the sub-regional model is that V_{S30}-scaling saturation affects a wider range of velocities as the period increases (i.e., V_1

increases with period). This observation differs from other recent models, for which saturation is relatively stable with period or occurs at lower V_{S30}'s as periods increase (e.g., Nweke et al. 2022; Parker et al. 2022). When saturation occurs, it indicates that V_{S30} has lost its predictive power for site response. This does not necessarily mean that site response is invariant across the saturation range, but that underlying physics are associated with phenomena that are not captured by the V_{S30} parameter (e.g., resonance effects, multi-dimensional wave propagation). The predictive power of additional parameters is investigated in subsequent subsections.

Within-event residuals (δW_{ij}^{v}) are computed from the BSSA14 GMM modified to include the site response model in Equation (5.9) and the regional path model from Chapter 4,

$$\delta W_{ij}^{v} = \ln(Y_{ij}) - [(c_0 + \Delta c_{0,r} + Ie_I) + F_E + F_P + F_{lin,ij}(V_{S30}) + F_{nl,ij} + \eta_{E,i}] \quad (5.10)$$

where the site response is solely a function of V_{S30} and PGA_r (i.e., F_{nl}). The 'v' superscript indicates that the residuals are computed using a model that includes V_{S30}-scaling. For the present analysis, the nonlinear model is taken from SS14. These residuals are partitioned using mixed effects analyses to estimate site terms ($\eta_{S,j}^{v}$), which represent the difference between site-specific observed amplification and the GMM described by Equation (5.10), and the remaining residual (ϵ_{ij}).

$$\delta W_{ij}^{v} = \eta_{S,j}^{v} + \epsilon_{ij} \quad (5.11)$$

Figure 5.11 shows the variation of $\eta_{S,j}^{v}$ with V_{S30} for the sites used during model development for *PGV*, *PGA*, and *PSA* at 0.1, 0.3, 1.0, 3.0, and 10.0 s. As expected, trends for the data are relatively flat and binned means are near zero, which indicates that the V_{S30}-scaling model accurately captures first-order site effects (i.e., those described by V_{S30}). As noted previously, there is relatively large variability for sites with V_{S30} < 250 m/s.

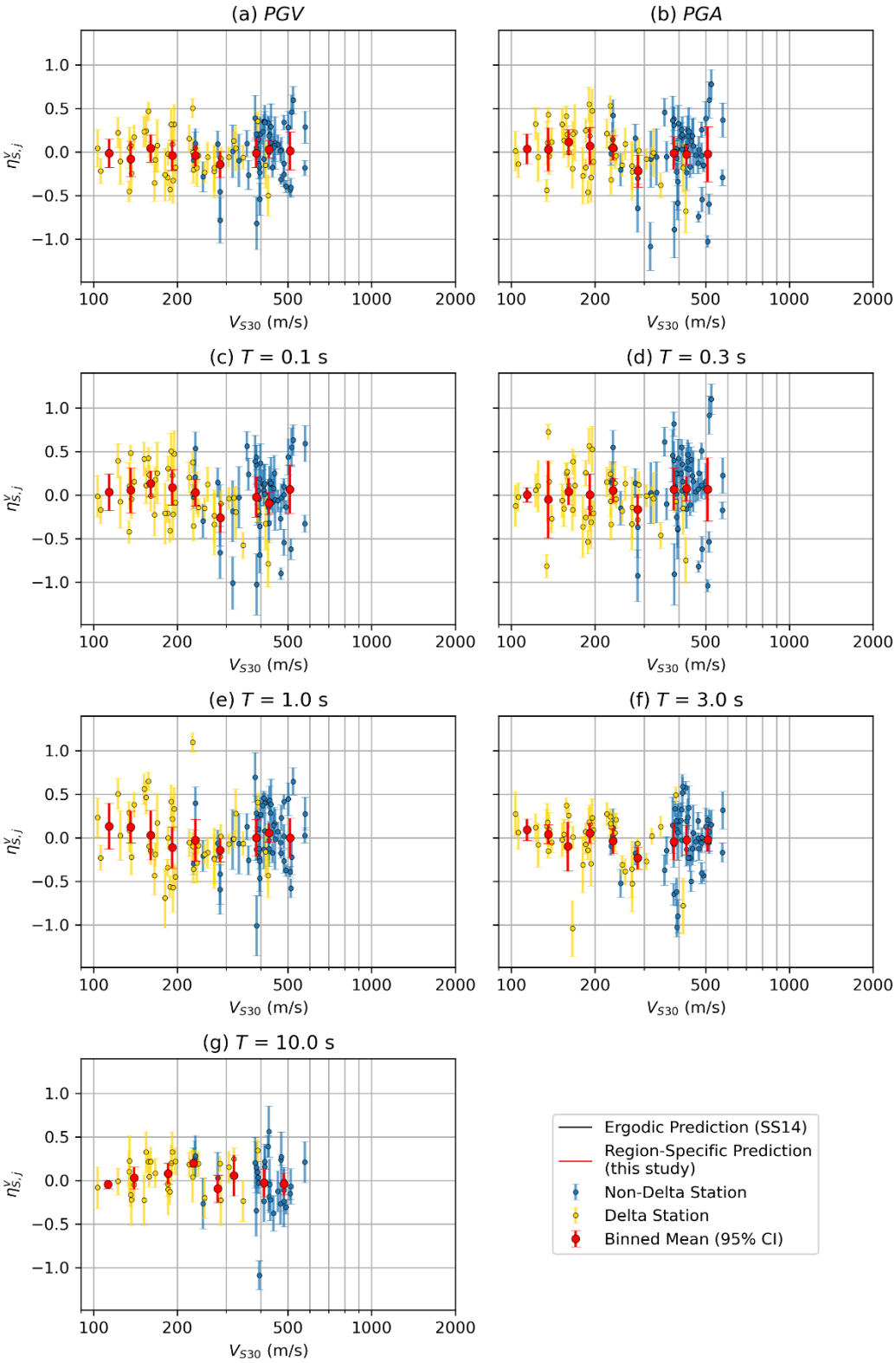


Figure 5.11: Trends of V_{S30}-dependent site terms ($\eta^{v}_{S,j}$) against V_{S30} for Delta and non-Delta stations for (a) *PGV*, (b) *PGA*, and *PSA* at (c) T = 0.1s, (d) T = 0.3s, (e) T = 1.0s, (f) T = 3.0s, and (g) T = 10.0s.

5.3.2. Dependence of Site Response on Isosurface Depth

The principal site parameter beyond V_{S30} that has been found to be correlated to site response is the depth to a shear wave velocity isosurface, generally 1.0 or 2.5 km/s ($z_{1.0}$ or $z_{2.5}$, respectively). Relationships between site response and $z_{1.0}$ or $z_{2.5}$ have been established for southern California basins (Day et al. 2008), for broad regions (California and Japan) in the NGA-West2 project (Abrahamson et al. 2014, BSSA14, Campbell and Bozorgnia 2014, Chiou and Youngs 2014), for basins in the Pacific Northwest and Japan in the NGA-Sub project (Abrahamson and Gulerce 2022, Kuehn et al. 2020, Parker and Stewart 2022), and for geomorphic provinces in southern California (Nweke et al. 2022). Here I examine potential variations of site response with $z_{1.0}$ for the Delta subregion.

To estimate depth parameters for Delta sites, I use the USGS San Francisco Community Velocity Model (Hirakawa and Aagaard 2021; SFCVM v21.1), which is currently the only such model for the Delta region and surrounding portions of the Central Valley. An important difference between USGS SFCVM and southern California models (CVM-S4, Lee et al. 2014; CVM-H, Shaw et al. 2015) is that seismic tomography is not used in the development of SFCVM. Tomography provides a means by which spatial variations in isosurface depths can be established. Because that information is not used in SFCVM, velocity profiles are spatially invariant within geologic units. The procedures used to establish unit-specific velocity profiles are described by Brocher (2005a-b; 2006), Hardebeck et al. (2007), Thurber et al. (2007), Aagaard et al. (2020) and Hirakawa and Aagaard (2021).

Previous research has shown that $z_{1.0}$ and V_{S30} are correlated in California (e.g., Chiou and Youngs 2014; Nweke et al. 2022). It is important to account for this correlation when developing models jointly conditioned on this combination of site parameters. To reduce the multicollinearity

of $z_{1.0}$ and V_{S30}, I evaluate mean basin depths ($\bar{z}_{1.0}$) conditioned on V_{S30}, compute the differential depth ($\delta z_{1.0}$) as defined in Equation (5.12), and investigate dependencies of site response on differential depth.

$$\delta z_{1.0} = z_{1.0} - \bar{z}_{1.0} \tag{5.12}$$

Separate $\bar{z}_{1.0}$ relationships have been developed for California and Japan (Chiou and Youngs 2013) and Cascadia (Parker and Stewart 2022). Moreover, Nweke et al. (2022) demonstrated that $\bar{z}_{1.0}$ relationships for a given region can vary based on seismic velocity model by comparing results between the CVM-S4 and CVM-H models. Given these between-region and within-region variations, it is possible that the relationship for the Delta may differ from the state-wide model from Chiou and Youngs (2014). Therefore, I sought to develop a centering model for the Central Valley (which encompasses the Delta region).

The necessary data required to develop a $\bar{z}_{1.0}$-V_{S30} model consists of V_{S30} and $z_{1.0}$ values (measured or estimated) for sites within the Central Valley. I adopt the Central Valley boundary as defined by Chiou et al. (2010) (identical to the Central Valley subregion utilized in the subregional anelastic path model presented in Section 4.3). The ground motion database (GMDB) assembled as part of Chapter includes 136 sites within this domain. Querying the community shear wave velocity profile database (Ahdi et al. 2018; VSPDB) produced 104 sites which are used to supplement the GMDB data. These are screened to remove duplicates yielding 225 distinct sites, the locations of which are shown in Figure 5.12.

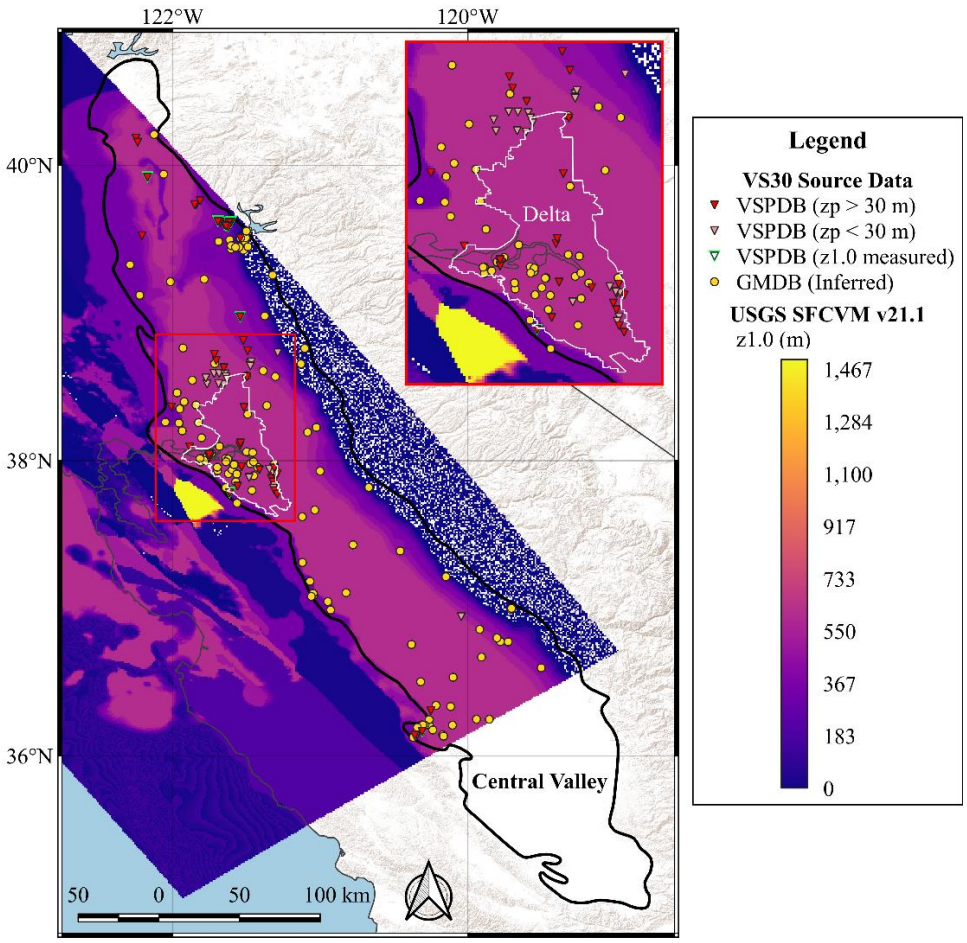


Figure 5.12: Map of the Central Valley and Delta showing locations of sites with V_{S30} from the Velocity Profile DataBase (VSPDB) and the Ground Motion DataBase (GMDB) overlaid with $z_{1.0}$ depth raster from the USGS San Francisco Seismic Velocity Model (SFCVM v21.1).

V_{S30} values are computed from deep V_S profiles ($z_p \geq 30$ m) obtained from the VSPDB, or extrapolated from shallow profiles ($z_p < 30$ m) using the approach presented by Boore (2004) with the California-specific coefficients reported by Kwak et al. (2017). Many of the sites incorporated into this compilation were prepared for use and uploaded to the VSPDB as part of this project [Section 3.2.1(c)]. Measured V_S profiles are also examined individually for $z_{1.0}$. Twelve sites encounter $V_S \geq 1{,}000$ m/s, for which the measured $z_{1.0}$ is used in lieu of the SFCVM model estimate. These sites correspond to locations along the edge of the basin structure. The distribution

of Central Valley data in V_{S30}-$z_{1.0}$ space is shown in Figure 5.13. Individual data points are shown along with binned means and their 95% confidence intervals (bins are evenly spaced in log units of V_{S30}). The binned means indicate that the relationship proposed by Chiou and Youngs (2014) does not properly capture the observed trend, therefore a regionally-specific relationship is warranted.

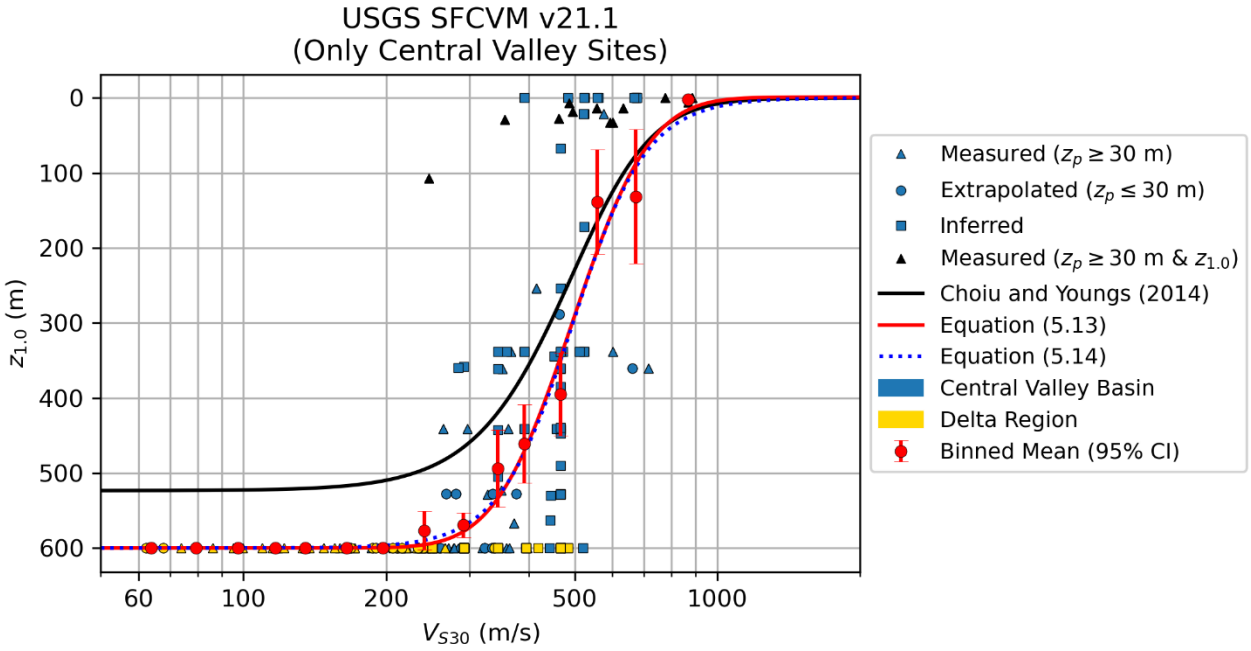


Figure 5.13: Distribution of Central Valley data in V_{S30}-$z_{1.0}$ space, including mean depth ($\bar{z}_{1.0}$) models from Chiou and Youngs (2014) for California and as regressed from the data using the functional forms presented in Equations (5.13) and (5.14).

I sought to develop an improved fit to the data while also enforcing physical bounds at the limits, where depth scaling saturates, by adopting the functional form presented in Nweke et al. (2022) which returns $\bar{z}_{1.0}$ in units of km:

$$\bar{z}_{1.0} = c_1 \left[1 + \text{erf}\left(\frac{\log(V_{S30}) - \log(\nu_\mu)}{\nu_\sigma \sqrt{2}}\right)\right] + c_0 \quad (5.13)$$

where c_0 represents the deepest depth within the basin (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 *erf* function can be solved in most numerical software packages. A simplified functional form which eliminates the c_1 coefficient and replaces the *erf* function is also investigated:

$$\bar{z}_{1.0} = \frac{c_0}{1+\exp\left[1.702\left(\frac{\log(V_{S30})-\log(\nu_\mu)}{\nu_\sigma}\right)\right]} \tag{5.14}$$

where 1.702 renders a logistic function that very closely approximates a cumulative distribution function.

The dataset shown in Figure 5.13 includes V_{S30}-values calculated from deep profiles, extrapolated from shallow profiles, and inferred from proxy-based model(s) – each of which possess different degrees of uncertainty ($\sigma_{lnV_{S30}}$) as summarized in Table 2.9. To account for this during the regression, I apply weights (w_i) inversely proportional to $\sigma_{lnV_{S30}}$:

$$w_i = \frac{n\left(\sigma_{lnV_{S30}}\right)^{-2}}{\sum_{i=1}^{n}\left(\sigma_{lnV_{S30}}\right)^{-2}} \tag{5.15}$$

The binned means shown in Figure 5.13 were calculated considering these weights.

The data in Figure 5.13 show that SFCVM predicts a maximum $z_{1.0}$ of 600 m in the Central Valley, therefore I constrain c_0 to be 0.6 km for both functional forms. To enforce the $z_{1.0}$ = 0 m boundary condition for large values of V_{S30} in Equation (5.13), I pad the dataset with 80 fictitious observations (i.e., 80 points between 850 and 3000 m/s with $z_{1.0}$ = 0 m). This padding is not needed when regressing Equation (5.14). WLS regression is applied for both functionals to obtain estimates for c_1 (Equation 5.13) only], ν_μ, and ν_σ – the values of which are tabulated in Table 5.4.

Table 5.4: Basin depth predictive model coefficients for Equations (5.13) and (5.14).

Equation	c_0 (km)	c_1	ν_μ (m/s)	ν_σ
(5.13)	0.6	-0.3006 ± 0.0025	494.20 ± 4.04	0.1282 ± 0.0053
(5.14)	0.6	-	496.00 ± 8.73	0.1324 ± 0.0140

The results for both functions, shown by the solid red and dotted blue lines in Figure 5.13, provide reasonable fits to the binned means and are nearly identical. Central Valley-specific models suggest a steeper depth scaling relationship and deeper maximum depths when compared to the California-specific model proposed by Chiou and Youngs (2014). Another notable observation is a perfect fit at V_{S30} less than about 250 m/s, which is exclusively sites within the Delta. This results from constant values of depth (600 m) within this domain, because of the consistent geology. In reality, some variations are likely, due in part to known faulting in the region (e.g., Midland fault, Deverel et al. 2016). A consequence of these uniform depths is that $\delta z_{1.0}$ values, whether computed using $\bar{z}_{1.0}$ from Equations (5.13) or (5.14), have practically no variation for Delta-only stations, as shown in Figure 5.14.

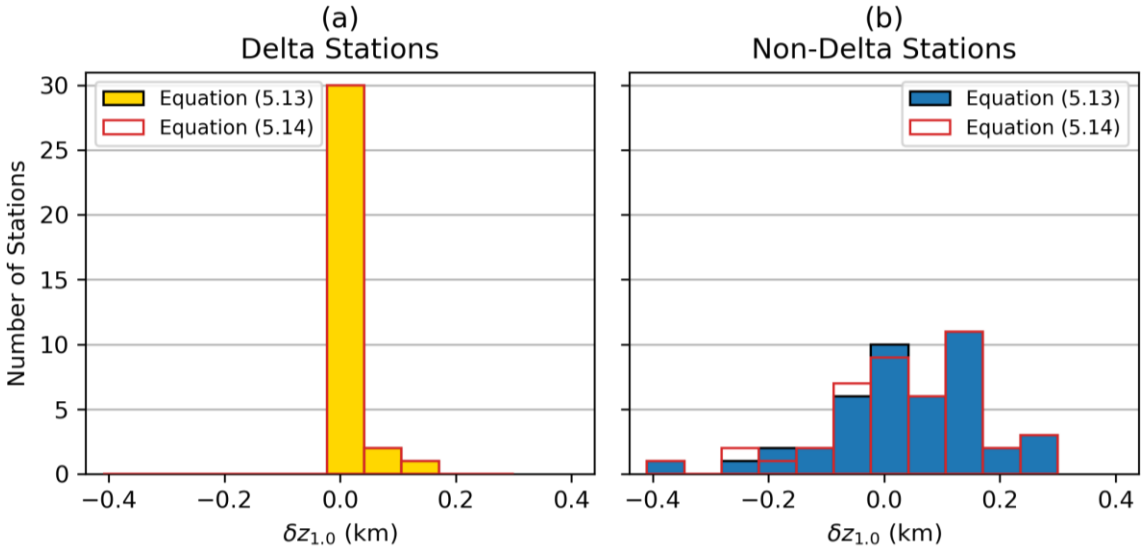


Figure 5.14: Histograms of differential basin depth ($\delta z_{1.0}$) for (a) Delta stations and (b) non-Delta stations using Equations (5.13) and (5.14).

Equation (5.14) was used to develop $\delta z_{1.0}$ for each of the Delta and non-Delta Central Valley sites. The depth-dependence of site response is then investigated by plotting in Figure 5.15 $\eta_{S,j}^{v}$ against $\delta z_{1.0}$ for *PGV*, *PGA*, and *PSA* at 0.1, 0.3, 1.0, 3.0, and 6.0 s. A binned mean for all Delta stations is positioned at $\delta z_{1.0}$ = 0 km, other binned means represent the mean trend of non-Delta stations. In general, the data indicate systematic trends between $\eta_{S,j}^{v}$ and $\delta z_{1.0}$ for long periods (T > 2.4 s), but no appreciable depth dependence for short periods (T < 0.45 s). To better visualize these trends, I use WLS regression to fit a preliminary basin model (F_B) of the tri-linear functional form proposed by Nweke et al. (2022):

$$F_B = \begin{cases} f_7 + f_6 f_8 & \delta z_{1.0} < f_8 \\ f_7 + f_6 \delta z_{1.0} & \text{otherwise} \\ f_7 + f_6 f_9 & \delta z_{1.0} > f_9 \end{cases} \tag{5.16}$$

where f_6 is the slope of the differential depth scaling (natural log amplification per unit depth), f_7 represents $F_B(\delta z_{1.0} = 0)$ in natural log units, and f_8 and f_9 represent the lower and upper differential depth limits between which site response scales with $\delta z_{1.0}$. Parameters f_8 and f_9 were set to -0.25 km and 0.2 km by visual inspection.

The results in Figure 5.15 conform well to the current understanding of basin effects (i.e., larger amplification for sites on thicker sediment deposits especially at long periods), however the model requires further refinement before it can be recommended for use. In this project, our aim is to predict site response within the Delta, with special consideration of sites with peaty-organic soils. Delta sites generally have V_{S30} < 250 m/s, as shown in Figure 5.1. Because SFCVM model depth estimates are uniformly 600 m for all Delta sites, a depth-based model has no ability to resolve differences in site response between Delta sites. Accordingly, I recommend against using an F_B model in engineering application. However, the work presented herein may support future work to better model basin effects in the greater Central Valley.

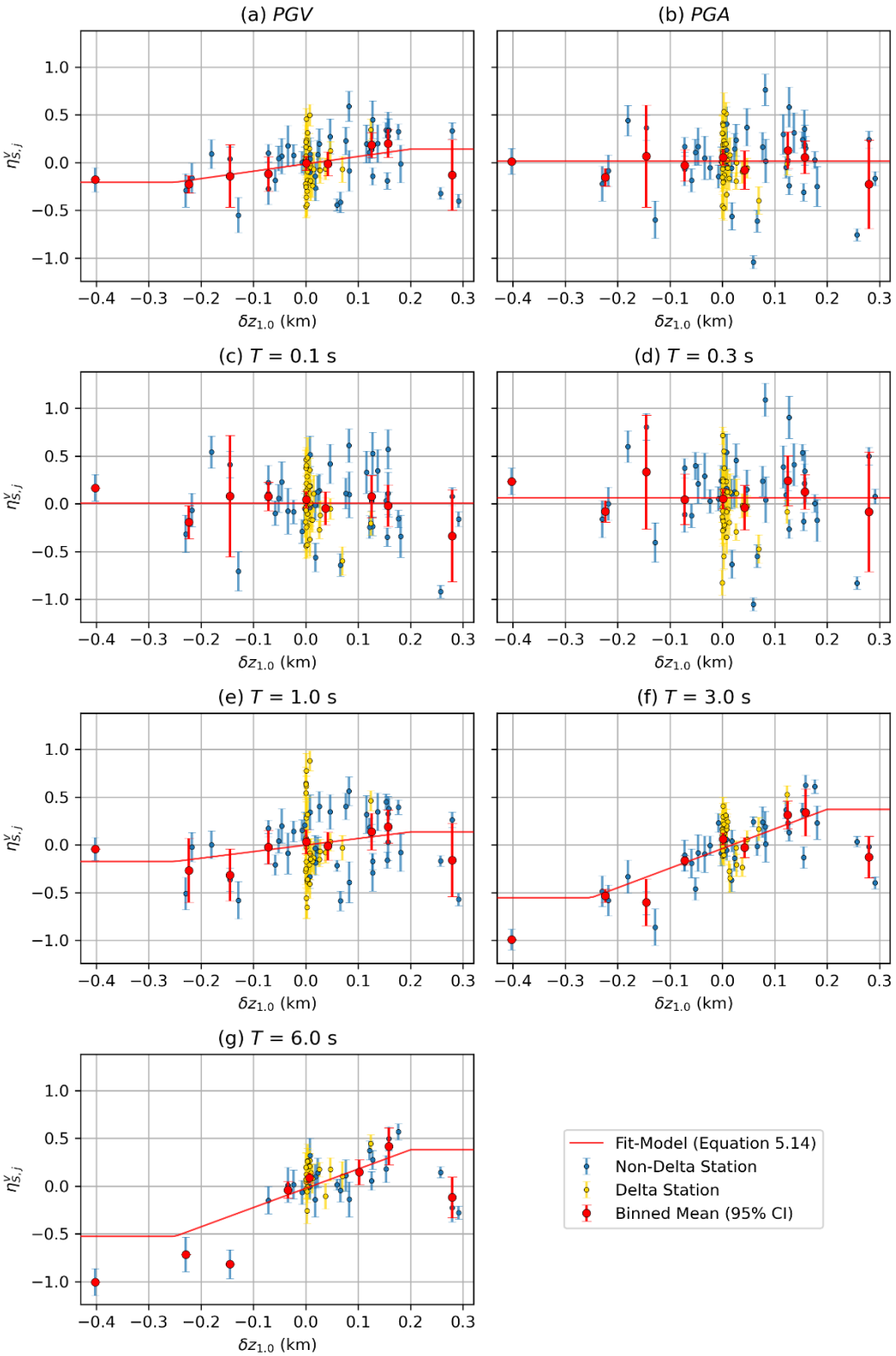


Figure 5.15: Trends of V_{S30}-dependent site terms ($\eta^{v}_{S,j}$) against $\delta z_{1.0}$ for Delta and non-Delta Central Valley stations for (a) *PGV*, (b) *PGA*, and *PSA* at (c) T = 0.1s, (d) T = 0.3s, (e) T = 1.0s, (f) T = 3.0s, and (g) T = 6.0s.

5.3.3. Site Resonance Effects

The average $\eta_{S,j}^{v}$ for Delta sites suggests that the current F_{lin} model, which captures only V_{S30}-scaling, provides unbiased predictions for the Delta sites as a whole, as illustrated in Figure 5.16. However, such models cannot capture site-specific resonance effects, because the model prediction is smoothed over a large number of sites with varying characteristics. For example, when I examine $\eta_{S,j}^{v}$ for individual sites, two trends with respect to period emerge:

1. A flat trend for periods less than about 0.1 s, which sometimes extends to longer periods as well.
2. Significant peak features for many sites at intermediate-to-long periods and generally larger dispersion.

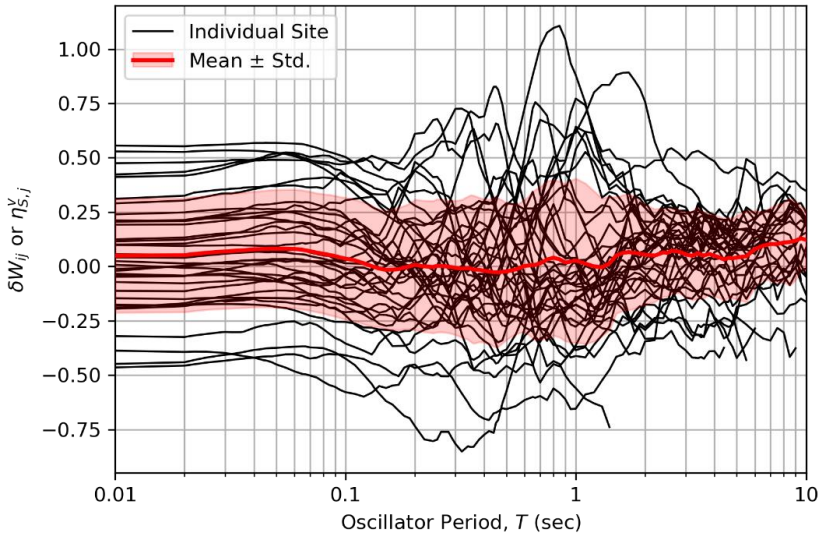


Figure 5.16: Average un-modelled site response ($\eta_{S,j}^{v}$) for Delta sites with reliable estimates.

Figure 5.17 shows V_{S30}-dependent within-event residuals using Equation (5.10) and the site terms derived from them ($\eta_{S,j}^{v}$) for two sites. Site WR_CKR has a flat trend with period, indicating

a lack of peaks, whereas CE_67587 has a strong peak near 0.8 s. The behavior of sites with no systematic features (such as WR_CKR) suggests that the period-dependence of site response is adequately captured by the V_{S30}-scaling model (although a general bias may still be present), while the behavior of other sites (such as CE_67587) indicates that site resonances are influencing site response in a manner that is not captured by the V_{S30}-scaling model presented in Section 5.3.1.

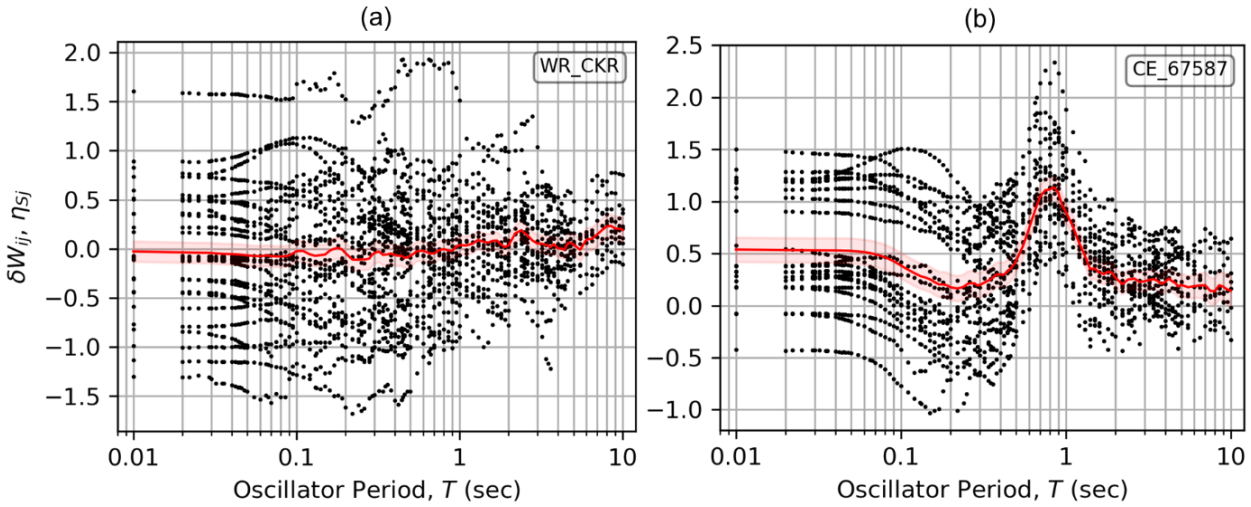


Figure 5.17: Examples of individual un-modelled site response ($\eta_{S,j}^{v}$) for sites with (a) no systematic features (WR_CKR) and (b) significant peak features (CE_67587).

Given the complex processes that produce ground motions relative to the simplicity of available predictive tools, I expect some randomness in the observed site response. This randomness often results in many local maxima and minima, which are not always repeatable across multiple events. However, peaks are consistently observed at many Delta stations, and because such peaks have a strong physical basis (i.e., increased amplification associated with site resonances), I seek to incorporate these features into the regional site response model. In the

remainder of this section, I identify peaks in the site response. The following section (5.3.4) links the site response peaks to peaks in mHVSR measurements from Delta sites.

(a) Peak-Resonance Identification Algorithm

When investigating site resonance using HVSR methods, analysists rely on criteria for the objective identification of peak features (e.g., SESAME 2004; Wang et al. 2021; Molnar et al. 2022). I posit that similar criteria should be applied to assess site resonance from non-ergodic site responses, although adjustments to the criteria are required given the differences in the nature of the data. In general, "peaks" are defined as features possessing the following key attributes:

1. Relatively localized (i.e., the width should not span too large of a period/frequency range),
2. Have sufficiently large mean amplitude relative to adjacent periods/frequencies, and
3. Have sufficient confidence that the feature is meaningful (i.e., uncertainty in amplitudes or frequencies should not be too large).

I propose a peak detection algorithm adapted from that recommended by Wang et al. (2021) for application to site response, which considers the three attributes listed above. For the purposes of this section, site response refers to the unmodeled site response after V_{S30}-scaling effects have been accounted for (i.e., $\eta_{S,j}^{v}$ as computed in Equation 5.11).

The principal challenge in assessing the first two attributes is defining the amplitudes and locations (periods/frequencies) of site response peaks; what is required is a parameterization of site response amplitudes adjacent to peak features and within peak features. I implement a regression tree (Breiman et al. 1984), which is a predictive modeling approach in machine learning, to effectively smooth and simplify the empirical site response as a piecewise function of non-

overlapping linear segments (i.e., steps). A complexity parameter (c_p) is used to specify the penalty in tree regression. Large values of c_p produce relatively crude fits with wide steps, whereas smaller values produce better fits with narrow steps. Figure 5.18 illustrates the influence of the c_p parameter on the tree regression for an individual site. If c_p is too large, the fit is poor, whereas if c_p is too small, there is the potential that the tree regression captures too many small peaks, which is not amendable to defining a stable peak-adjacent plateau.

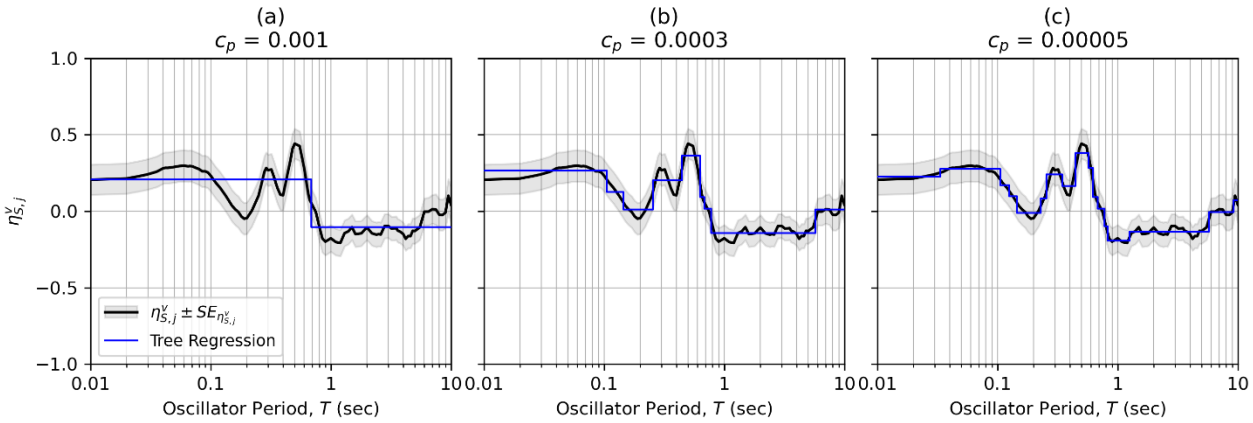


Figure 5.18: Schematic illustration of regression tree fitting of site response data with different values of the complexity parameter (c_p) for site NP_KIR; (a) c_p = 0.001; (b) c_p = 0.0003; (c) c_p = 0.00005.

An optimal (or preferred) value of c_p captures the main features of site response peaks without over-fitting. Selection of the preferred value of c_p is subjective. Using trial and error with visual inspection, I find that c_p = 0.0003 provides a reasonable balance between accuracy and reliability (in Python; Pedregosa et al. 2011). This value is lower than the recommended values suggested by Wang et al. (2021) for peak detection using HVSR data (i.e., c_p = 0.002-0.003), which is to be somewhat expected since site response features are generally more subtle than those

observed in typical HVSR. Specifically, when comparing the amplitudes of peak features to those of stable plateaus in mHVSR and $\eta_{S,j}^{v}$, the differences between the two for mHVSR are usually greater than $\eta_{S,j}^{v}$. The relative difference in amplitudes, which is usually related to the overall range of amplitudes, has significant impact on the regression tree results. In other words, a relatively simple model can capture prominent peaks and adjacent plateaus (such as in mHVSR), while higher complexity is required to be able to distinguish less pronounced peak features from adjacent plateaus (as with $\eta_{S,j}^{v}$).

The peak detection algorithm operates on the stepped results of tree regression and is shown as a flow chart in Figure 5.19. The algorithm operates as follows:

1. Identify potential peak steps: Step i is a peak if its amplitude (amp_p) is larger than those of steps $i - 1$ and $i + 1$.
2. For each potential peak step in Task (1), identify the left-peak-adjacent step-plateau ($step_l$). This identification has the following sub-steps:
    i. Let j denote the number of steps to the left of the peak step (i.e., $j = 0$ is the peak step i; $j = 1$ is the step immediately to the left of step i; $j = 2$ is the second step to the left of step i; and so on).
    ii. Starting at $j = 1$, obtain the left step amplitude (amp_{i-j}) and calculate the width (natural log difference of maximum and minimum period of step $i - j$; wid_{i-j}).
    iii. If $amp_{i-j} > amp_{i-j+1}$, then step $i - j + 1$ is the left-peak-adjacent step plateau, and proceed to Task (3). Otherwise, continue to Sub-task (iv).

iv. If $id_{i-j} > step_{thres}$, the step is sufficiently wide to be considered a plateau, and step $i - j$ is the left-peak-adjacent step plateau, and proceed to Task (3). Otherwise, continue to Sub-task (v).

v. If $i - j = 1$ (i.e., the first step), then go back through all left-steps and select the step with the largest width, and proceed to Task (3).

3. For each potential peak step in Task (1), identify the right-peak-adjacent step-plateau ($step_r$).

   i. Let j denote the number of steps to the right of the peak step (i.e., $j = 0$ is the peak step i; $j = 1$ is the step immediately to the right of step i; $j = 2$ is the second step to the right of step i; and so on).

   ii. Starting at $j = 1$, obtain the right step amplitude (amp_{i+j}) and calculate the width (log difference of maximum and minimum period of step $i + j$; wid_{i+j}).

   iii. If $amp_{i+j} > amp_{i+j-1}$, then step $i + j - 1$ is the right-peak-adjacent step plateau, and proceed to Task (4). Otherwise, continue to Sub-task (iv).

   iv. If $id_{i+j} > step_{thres}$, the step is sufficiently wide enough to be considered a plateau, and step $i + j$ is the right-peak-adjacent step plateau, and proceed to Task (4). Otherwise, continue to Sub-task (v).

   v. If $i + j = n$, where n represents the number of steps in the piecewise function, then go back through all right-steps and select the step with the largest width, and proceed to Task (4).

4. For each potential peak step in Task (1), compute the peak width (wid_p):

   i. Identify the maximum period of the left-peak-adjacent step plateau (T_l).

   ii. Identify the minimum period of the right-peak-adjacent step plateau (T_r).

iii. Compute wid_p = ln(T_r) – ln(T_l).

5. For each potential peak step in Task (1), compute $k = (amp_p - \bar{\eta}^v_{S,j})/\ \overline{SE}_{\eta^v_{S,j}}$ for the left- and right-peak-adjacent step plateaus, where $\bar{\eta}^v_{S,j}$ and $\overline{SE}_{\eta^v_{S,j}}$ represent the average amplitude and standard error over the step width, and k represents a non-zero multiplier.
6. Identify clear peak features from among potential peaks: Clear peaks are those which satisfy the following criteria:
   (a) The difference between amp_p and the maximum of amp_l and amp_r should exceed a threshold: $amp_p - \max(amp_l, amp_r) \geq amp_{thres}$.
   (b) The peak should not be too wide: $wid_p \leq wid_{thres}$.
   (c) There should be sufficient confidence that the mean peak amplitude (amp_p) is greater than the right- and left-peak-adjacent step plateau amplitudes: $\min(k_l, k_r) \geq k_{thres}$.

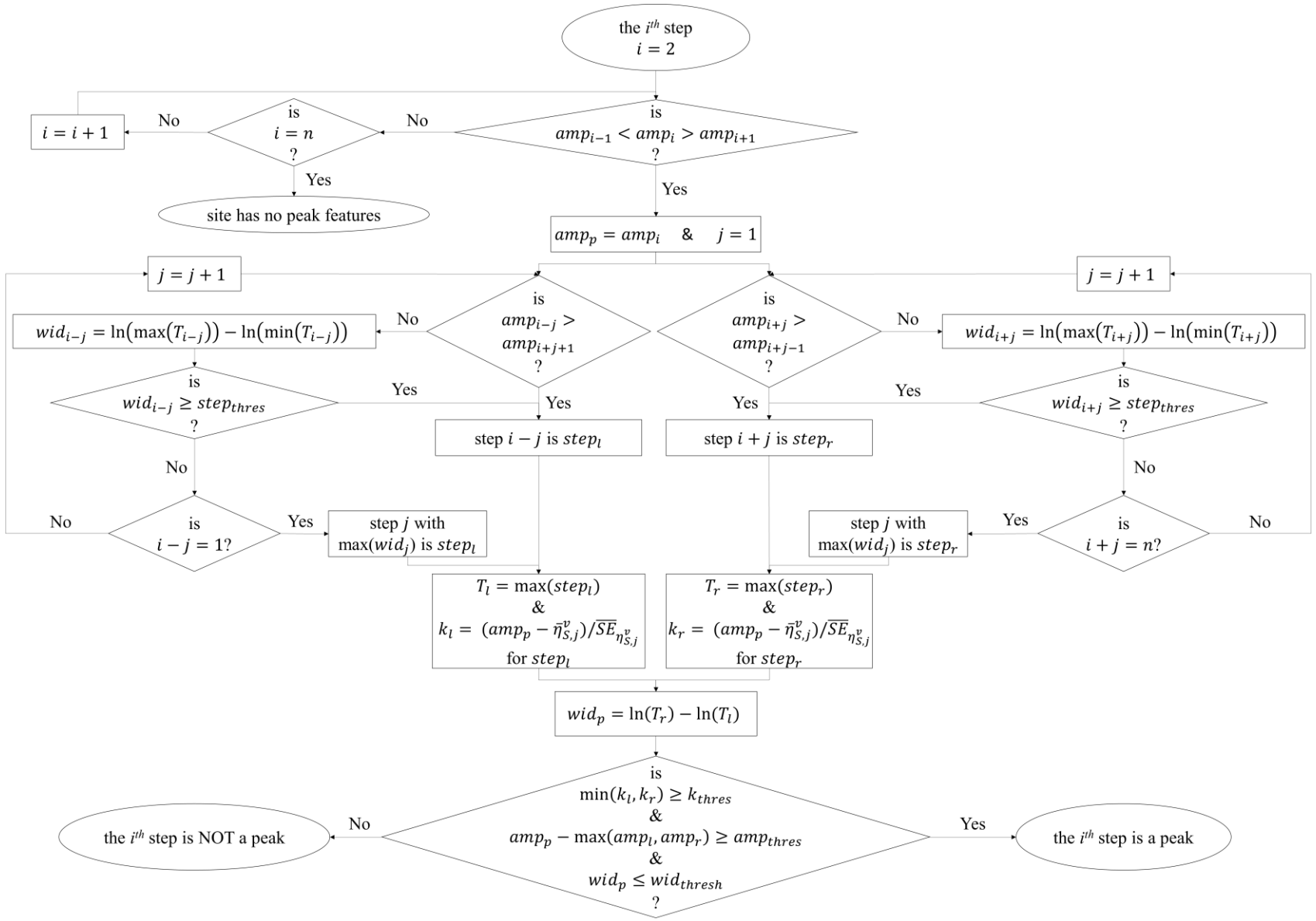


Figure 5.19: Flowchart illustrating site response peak detection algorithm utilizing tree regression of $\eta^v_{S,j}$.

The algorithm parameters ($step_{thres}$, amp_{thres}, wid_{thres}, and k_{thres}) can be applied with different levels of conservatism, depending on user preference. Smaller values of $step_{thres}$ and wid_{thres}, and larger values of amp_{thres} and k_{thres} result in a more conservative algorithm (i.e., fewer peaks identified). Conversely, larger values of $step_{thres}$ and wid_{thres}, and smaller values of amp_{thres} and k_{thres} result in a more liberal algorithm (i.e., more peaks identified). Sites which clearly possess or do not possess peak features (Figure 5.20a,b) are less sensitive to selection of these parameters. However, the algorithm results for sites with ambiguous peak features (Figure 5.20c,d) can vary depending on the selections. The $step_{thres}$ parameter carries significant influence because it is used to define the left- and right-peak adjacent step-plateaus, which are used to derive the quantities that the other parameters are checked against (in cases where a local trough is not present). The tree regression utilizes many steps with relatively short widths to capture steep slopes (such as those coming off of peak amplitudes). It follows that once the step width becomes subjectively large enough, then the peak feature has transitioned to a plateau. Similarly, amp_{thres} and wid_{thres} are used to check if the potential peak has sufficiently large amplitude relative to adjacent amplitudes and that the feature is localized, respectively.

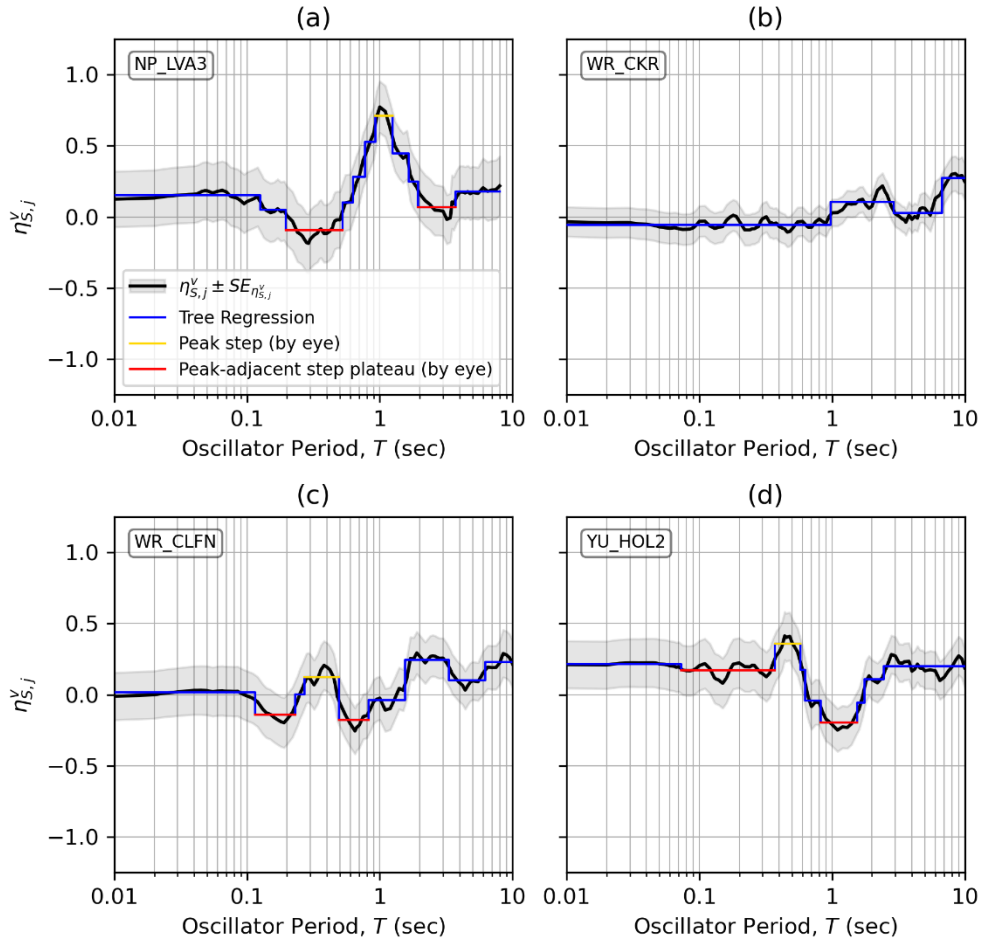


Figure 5.20: Examples of features observed in site response: (a) site with a clear peak (NP_LVA3); (b) site with no peak (WR_CKR); and (c) site with an ambiguous peak feature (WR_CLFN); and (d) site with an ambiguous peak feature (YU_HOL2).

The values of $step_{thres}$, amp_{thres}, and wid_{thres} were initially selected by calibrating the algorithm to correctly predict peaks using only sites which clearly possessed or did not possess peak features. I later reduced amp_{thres} to account for the smaller relative amplitudes which were observed in the sites with ambiguous peaks, and relied on the k_{thres} parameter to distinguish if the computed relative amplitudes were reliable. It follows that the recommended values for $step_{thres}$, amp_{thres}, wid_{thres}, and k_{thres} are 0.65, 0.27, 2.3, and 1.0, respectively. Figure 5.21 presents the algorithm results for the Delta stations used in this study. Of the 36 sites with reliable $\eta^{\nu}_{S,j}$

estimates, 17 (47%) are identified as possessing a peak. The following sub-section aims to parameterize these peak features for the purpose of model development (discussed in Section 5.3.4).

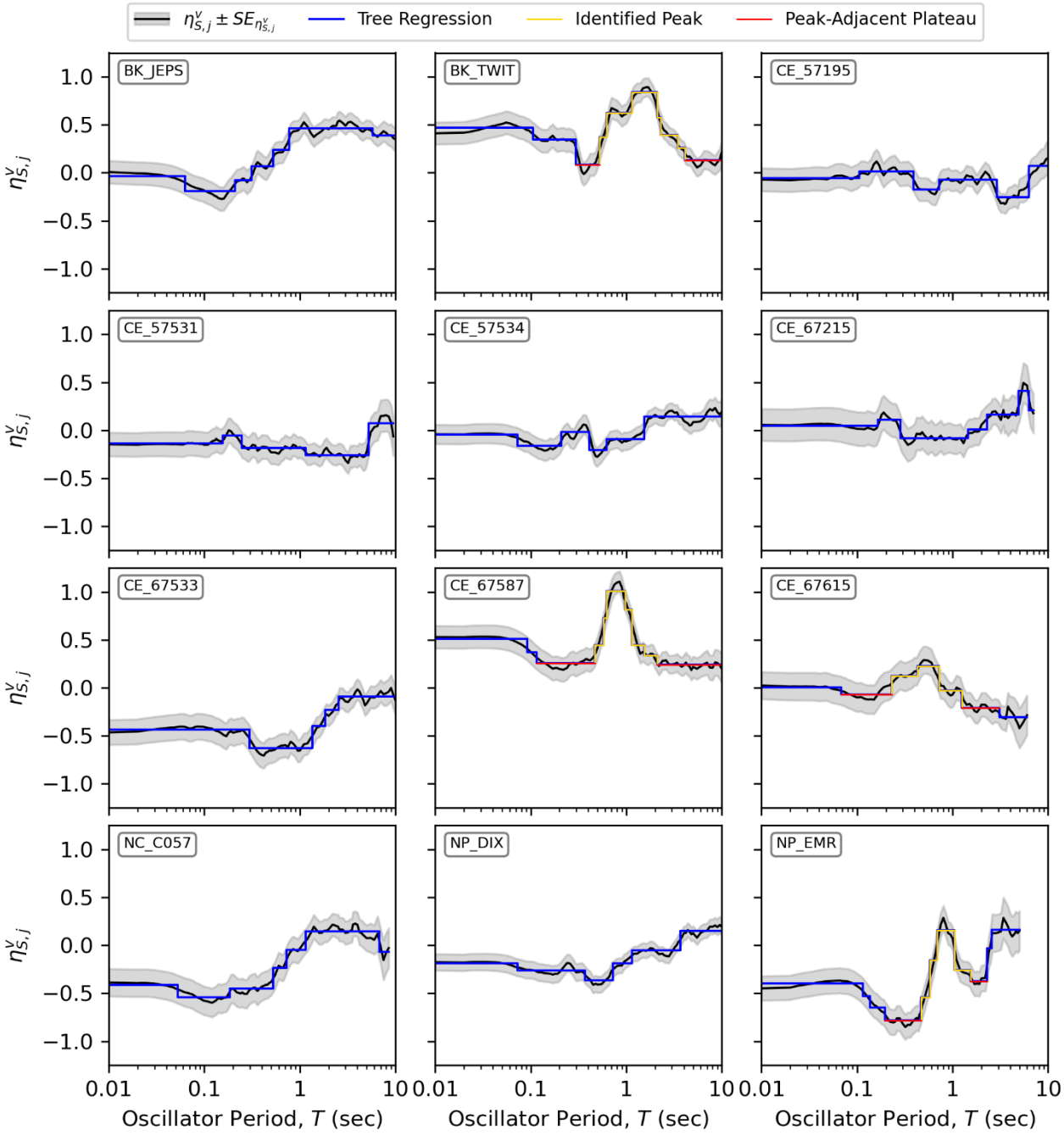


Figure 5.21: Results of peak-resonance identification algorithm for 36 Delta sites. Fit tree-regression of $\eta^{v}_{S,j}$ shown by blue step-function, and identified peaks in the site response are indicated by yellow (sites with no peak in the site response do not have any yellow in their plots).

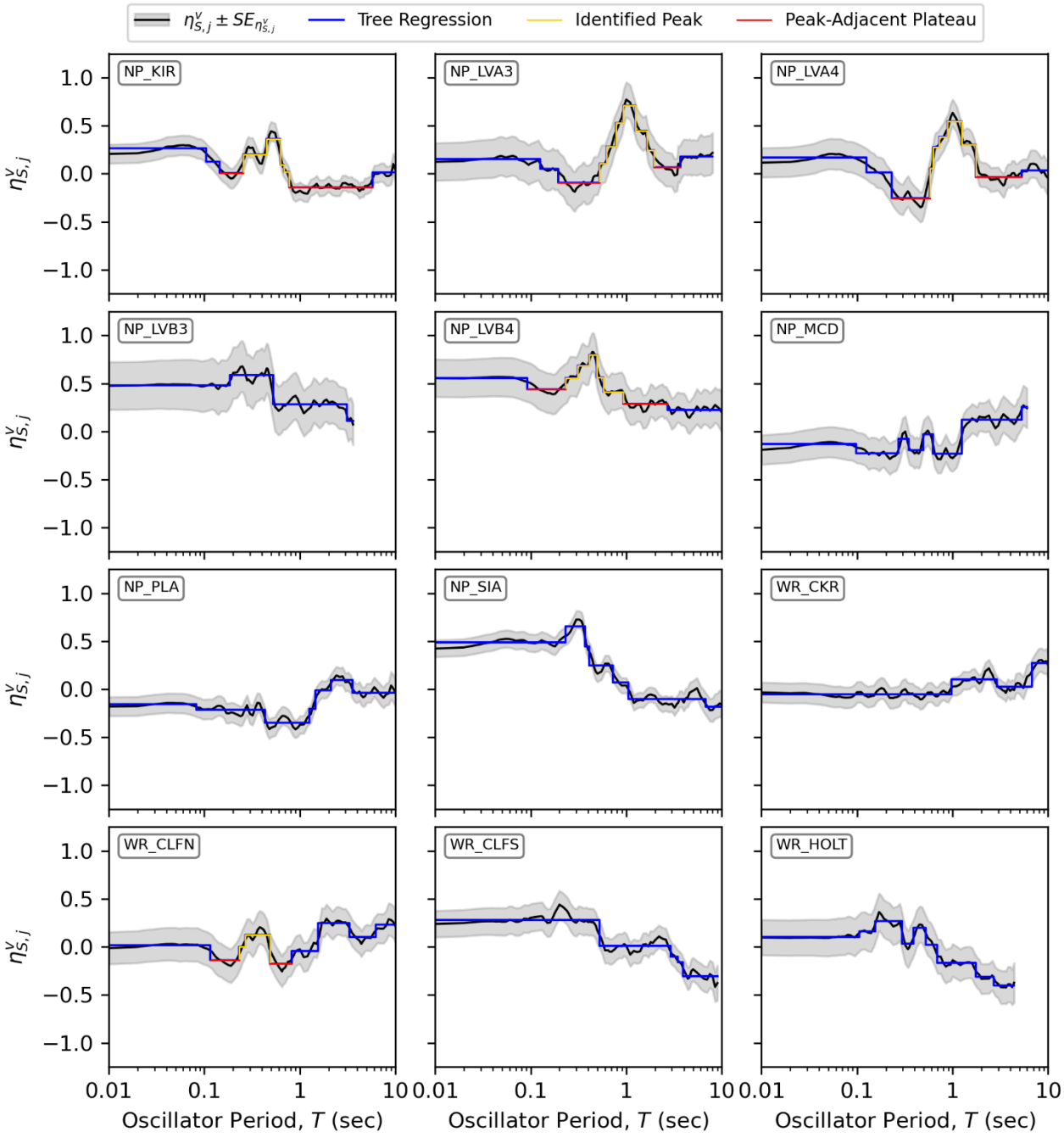


Figure 5.21: *continued.*

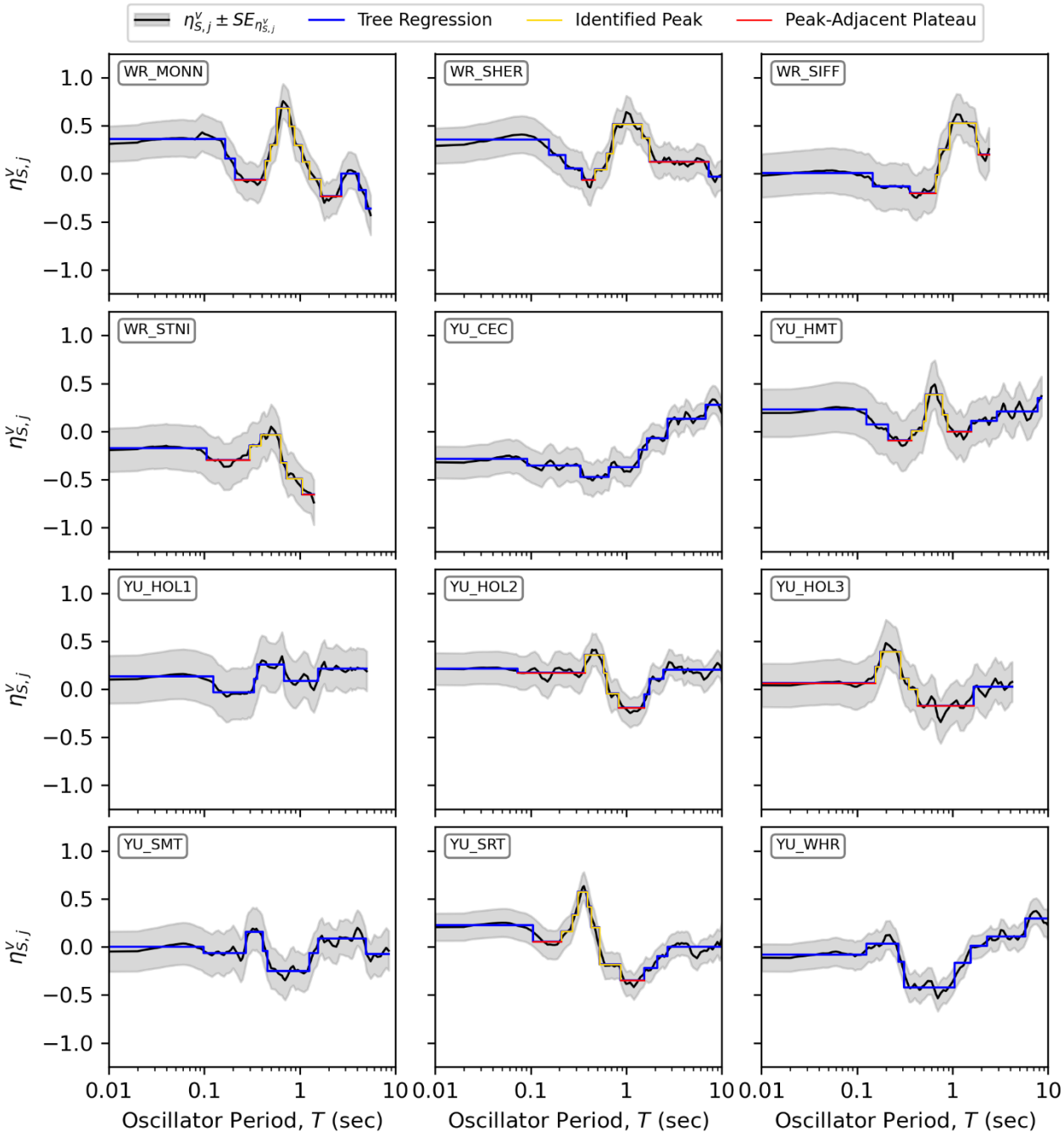


Figure 5.21: *continued.*

(b) Parameterization of Site Resonance Peak Features

In order for peak features to be modelled in a practical context, it is essential to parameterize site response peaks like in Figure 5.17(b). Based on observed features of peaks (i.e., the presence of a dip in amplitude immediately before the peak feature for all sites, which is not observed immediately following the peak feature for most sites), I select a hybrid Ricker Wavelet (Ryan 1994) and Gaussian Pulse function (Squires 2001) to capture the peaked shape of $\eta_{S,j}^{v}$ at Delta sites. The Ricker Wavelet function is used to capture unmodeled site resonance effects that dominate amplification shapes at short to intermediate periods ($T < \hat{f}_p^{-1}$), while the Gaussian Pulse function is used at longer periods ($T > \hat{f}_p^{-1}$):

$$f_1\left(T, \hat{f}_p\right) = \begin{cases} a_0 + \frac{2a_1}{\sqrt{3a_2}\pi^{\frac{1}{4}}}\left[1 - \left(\frac{\ln\left(T\hat{f}_p\right)}{a_2}\right)^2\right]\exp\left[-\frac{1}{2}\left(\frac{\ln\left(\hat{f}_p\right)}{a_2}\right)\right]^2 & T \le \hat{f}_p^{-1} \\ \left[f_1\left(\hat{f}_p^{-1}, \hat{f}_p\right) - a_3\right]\exp\left[-\frac{1}{2}\left(\frac{\ln\left(\left(T\,\hat{f}_p\right)^{-1}\right)}{2a_2/3}\right)^2 + a_3\right] & T > \hat{f}_p^{-1} \end{cases} \quad (5.17)$$

where a_0 controls the overall level of site amplification (i.e., constant adjustment up or down), a_1 scales the amplitude of the Ricker Wavelet function, a_2 describes the width of the Ricker Wavelet peak in natural log period units (a_2/3 approximates the width of the Gaussian Pulse peak), a_3 describes the difference in amplification between short and long periods, and $\hat{f}_p$ represents the frequency of the peak.

I perform least squares regression of Equation (5.17) on $\eta_{S,j}^{v}$ for the 17 Delta sites which were identified to have peaks by the algorithm presented in Section 5.3.3(a), which are identified in Figure 5.21. The results of these regressions are shown in Figure 5.22 and summarized in Table 5.5.

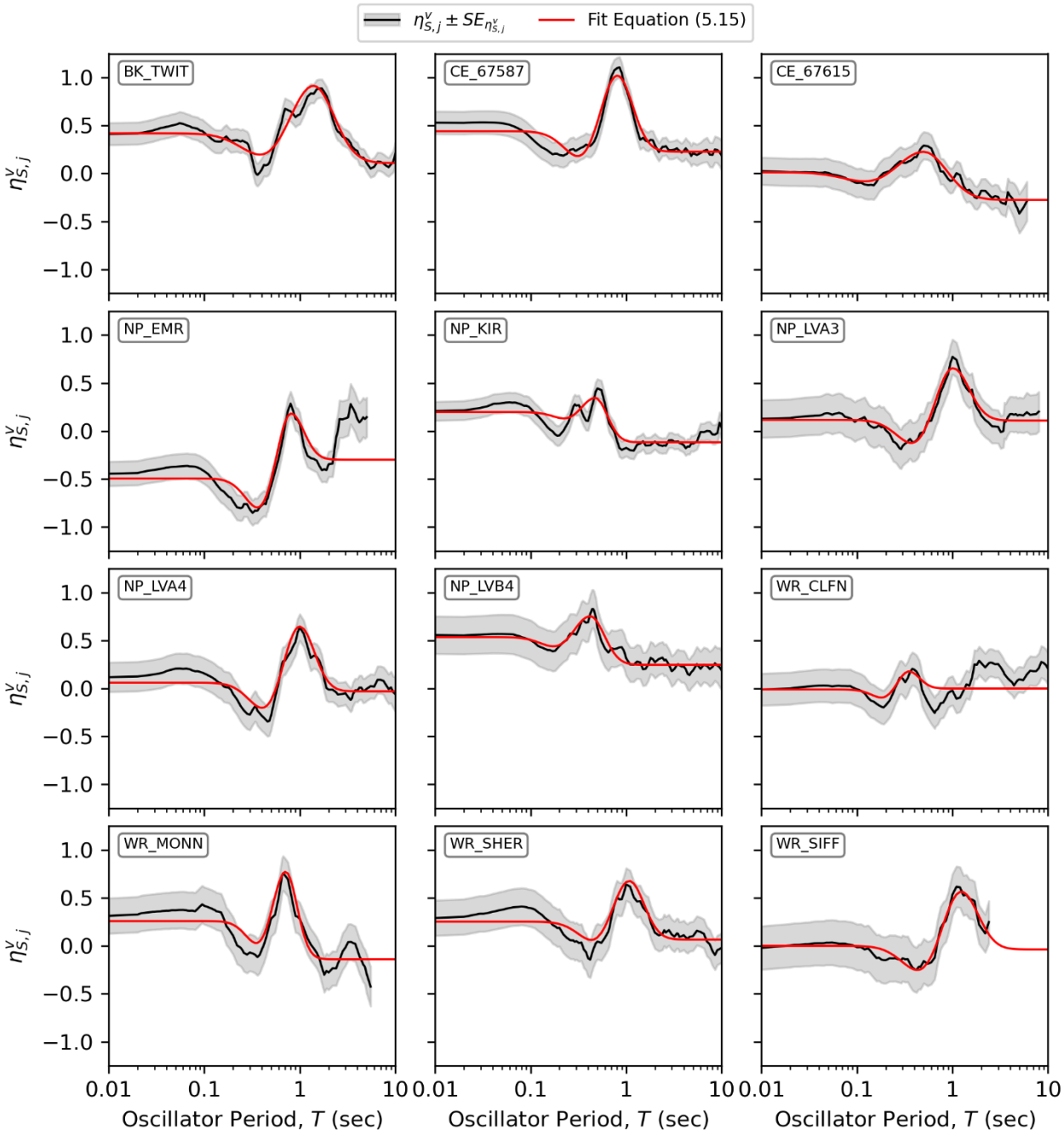


Figure 5.22: Results of fitting Equation (5.17) on $\eta^{v}_{s,j}$ for Delta sites with identified peak-resonance; fit shown by solid red line.

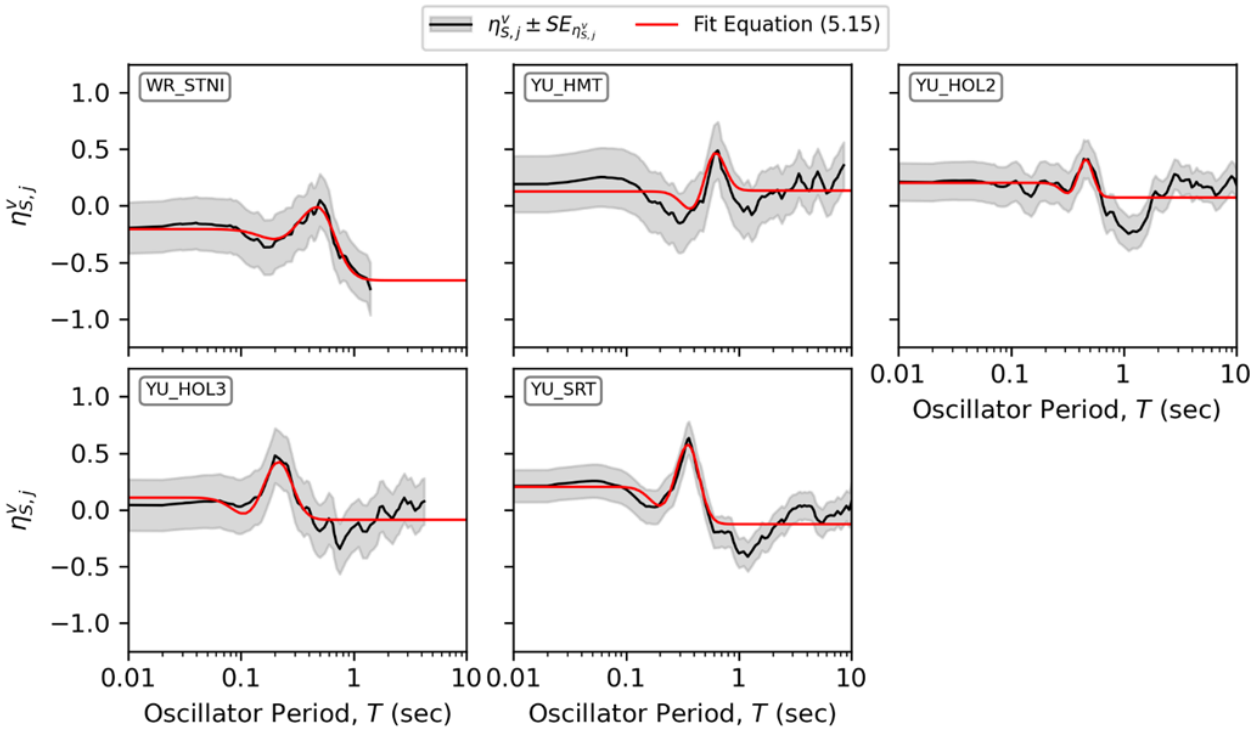


Figure 5.22: *continued.*

Table 5.5: Summary of $\eta^{v}_{S,j}$ peak fitting using Equation (5.17).

Delta Station	$\hat{f}_p$ (Hz)	a_0	a_1	a_2	a_3
BK_TWIT	0.7207	0.4191	0.2786	0.7472	0.1095
CE_67587	1.2382	0.4411	0.2804	0.5526	0.2274
CE_67615	1.9994	0.0138	0.1260	0.8292	-0.2751
NP_EMR	1.2380	-0.4949	0.3023	0.4715	-0.2990
NP_KIR	2.1100	0.1972	0.0635	0.4366	-0.1176
NP_LVA3	0.9882	0.1162	0.2672	0.5812	0.1087
NP_LVA4	1.0019	0.0596	0.2768	0.5275	-0.0300
NP_LVB4	2.3831	0.5355	0.1011	0.5130	0.2457
WR_CLFN	2.8571	-0.0110	0.0760	0.3851	0.0000
WR_MONN	1.4084	0.2590	0.2144	0.4117	-0.1388
WR_SHER	0.9276	0.2526	0.2015	0.5302	0.0660
WR_SIFF	0.8120	0.0010	0.2869	0.6113	-0.0363
WR_STNI	2.0911	-0.2055	0.0907	0.5151	-0.6571
YU_HMT	1.6127	0.1276	0.1216	0.3022	0.1348
YU_HOL2	2.1757	0.2032	0.0621	0.2201	0.0739
YU_HOL3	4.6360	0.1069	0.1326	0.4154	-0.0878
YU_SRT	2.8219	0.2025	0.1398	0.3354	-0.1263

5.3.4. Linking Site Response to mHVSR Site Parameters

The resonance effects described and quantified in the previous section are in site response. Such effects cannot be predicted by V_{S30}-scaling models such as that developed in Section 5.3.1. Here I investigate whether these effects are related to peaks in microtremor-based horizontal-to-vertical spectral ratios (mHVSR), as described for Delta sites in Section 5.3.3(b).

There are two general types of HVSR measurements – from microtremors (mHVSR) and earthquake recordings (eHVSR). Most past research relating HVSR to site response has used eHVSR, including Héloïse et al. (2012), Zhao and Xu (2013), Ghofrani et al. (2013) and Kwak et al. (2017b) for Japan, Hassani and Atkinson (2016, 2018b) for central and eastern North America, Hassani and Atkinson (2018a) for California, and Panzera et al. (2021) for Switzerland. The use of eHVSR is convenient because the data required to measure HVSR is readily available. The problem with using eHVSR is that the recordings used to develop the HVSR site parameters are not independent of those used to compute the site response. As a result, the correlation between eHVSR site parameters and site response is likely to be stronger than that provided by independent data such as mHVSR. For this reason, we only use mHVSR as the basis for site parameters used to develop the model components in this section. A similar approach was used previously for another subregion having peaty organic soil – Obihiro in Hokkaido, Japan (Wang et al. 2022a).

An effort has been underway to develop mHVSR data for sites in California and to evaluate the characteristics of that data. A database and data processing protocols are described in Wang et al. (2022b). Analyses of the data have shown that a modest percentage of sites (25-35%) have significant HVSR peaks, using peak identification protocols similar to those described in Section 5.3.3 (Wang et al. 2021). As part of the present study, the database was augmented for Delta sites as described in Section 3.3.3.

The Delta is rather unique in California because the geologic environment and depositional history results in a high percentage of sites (~71%) having strong impedance contrasts, which produce significant mHVSR peaks. It follows that site parameters, such as those summarized in Table 5.1 (c_0, c_1, a_p, f_p and w_p), may have significant potential to improve site response in this subregion. The questions addressed in this sub-section include the following:

1. Can mHVSR be used to predict the presence of peaks in the site response?
2. When mHVSR contain peak features, can the associated site parameters be used to improve predictions of site response?
3. Can other site parameters (unrelated to peaks) derived from mHVSR be used to improve predictions of site response, at sites with and without mHVSR peak features?

(a) Predicting Resonance in Site Response from mHVSR

Of the 54 Delta seismic stations, 36 have sufficient numbers of recordings to provide reliable η_S estimates (Section 5.2.1). Of those 36 sites, 34 have on-site measurements of mHVSR where 16 (47%) possess $\eta_{S,j}^{v}$ peak features. In this section, I compare the presence of peak features in mHVSR and $\eta_{S,j}^{v}$ for each site and classify them into one of four categories defined in the confusion matrix in Table 5.6. Table 5.7 provides a detailed summary of the presence of peak features in both mHVSR and $\eta_{S,j}^{v}$ for the 34 sites considered during modeling.

Table 5.6: Qualitative class definition for observed peak features.

		mHVSR	
		No Peak	**Peak**
$\boldsymbol{\eta_{S,j}^{v}}$	**No Peak**	Class I (10 sites – 29%)	Class III (8 sites – 24 %)
	Peak	Class IV (2 sites – 6%)	Class II (14 sites – 41%)

Table 5.7: Summary of presence of peak features in $\eta_{S,j}^{v}$ mHVSR.

Delta Station	Peak in $\eta_{S,j}^{v}$	Peak in mHVSR	Class
BK_JEPS	No Peak	Peak	III
BK_TWIT	Peak	Peak	II
CE_57195	No Peak	No Peak	I
CE_57531	No Peak	Peak	III
CE_57534	No Peak	No Peak	I
CE_67215	No Peak	Peak	III
CE_67533	No Peak	No Peak	I
CE_67587	Peak	Peak	II
CE_67615	Peak	No Peak	IV
NC_C057	No Peak	No Peak	I
NP_DIX	No Peak	Peak	III
NP_EMR	Peak	Peak	II
NP_KIR	Peak	Peak	II
NP_LVA3	Peak	Peak	II
NP_LVA4	Peak	Peak	II
NP_MCD	No Peak	No Peak	I
NP_PLA	No Peak	No Peak	I
NP_SIA	No Peak	No Peak	I
WR_CKR	No Peak	No Peak	I
WR_CLFN	Peak	No Peak	IV
WR_CLFS	No Peak	No Peak	I
WR_HOLT	No Peak	Peak	III
WR_MONN	Peak	Peak	II
WR_SHER	Peak	Peak	II
WR_SIFF	Peak	Peak	II
WR_STNI	Peak	Peak	II
YU_CEC	No Peak	No Peak	I
YU_HMT	Peak	Peak	II
YU_HOL1	No Peak	Peak	III
YU_HOL2	Peak	Peak	II
YU_HOL3	Peak	Peak	II
YU_SMT	No Peak	Peak	III
YU_SRT	Peak	Peak	II
YU_WHR	No Peak	Peak	III

The categories in the confusion matrix (Table 5.6) indicate the consistency of peak identification: Cases I and II indicate consistent results (both without peaks, both with peaks, respectively, which can also be referred to as true negatives and true positives). Cases III and IV indicate inconsistent results, with Case III being false positives (mHVSR peak, site response no-peak) and Case IV being false negatives (mHVSR no peak, site response peak). The *accuracy* of mHVSR is the percentage of peak or no-peak identifications that match the site response result, and is computed as the ratio of Case I and II sites divided by the total (71%). Considering sites with mHVSR peaks, the *precision* (percentage of sites with mHVSR peaks correctly predicted) is about 64%. The *recall* of the dataset, or the percentage of total results correctly predicted when there is a $\eta_{S,j}^{v}$ peak, is about 87.5%, which does not consider the results of non-peak sites. The F_1-score computed from the *precision* and *recall* is 73.7%. These metrics suggest that the existence of a mHVSR peak is a good indicator of when we can expect site-resonant effects in the site response, however there are other contributing factors which sometimes result in inaccurate predictions. Therefore, I plotted together in Figure 5.23 mHVSR and $\eta_{S,j}^{v}$ trends for each site to facilitate visual inspections that may help to improve predictions. Both plots use period axes – for $\eta_{S,j}^{v}$ period indicates oscillator periods, whereas for mHVSR period is the inverse of frequency.

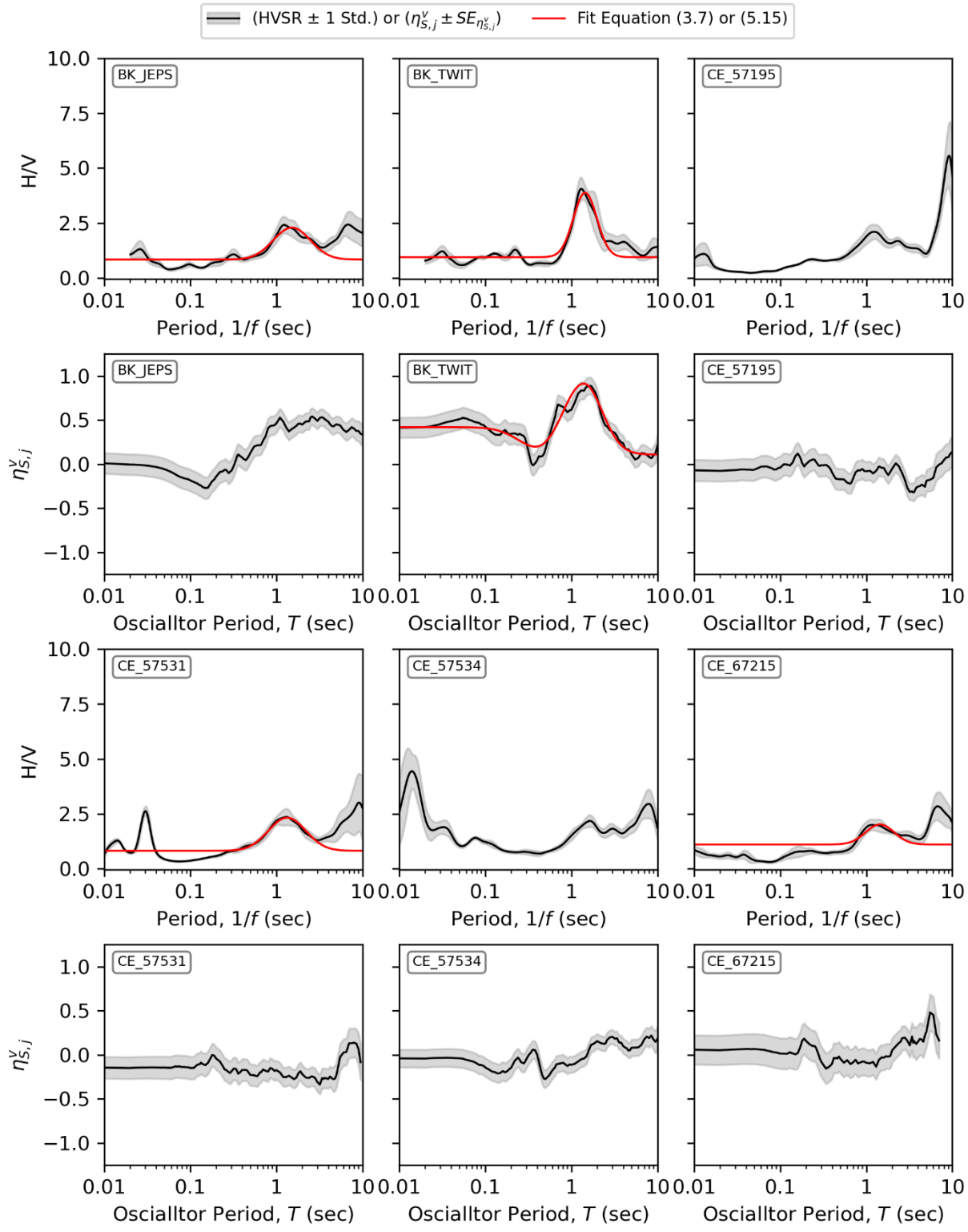


Figure 5.23: Comparisons of mHVSR and $\eta^{v}_{S,j}$ features for Delta stations. Fit of mHVSR peak shown by red Gaussian pulse (Equation 3.7). Fit of $\eta^{v}_{S,j}$-parameterization per Equation (5.17) shown in red. Sites with no peaks (mHVSR or $\eta^{v}_{S,j}$) do not have red lines shown.

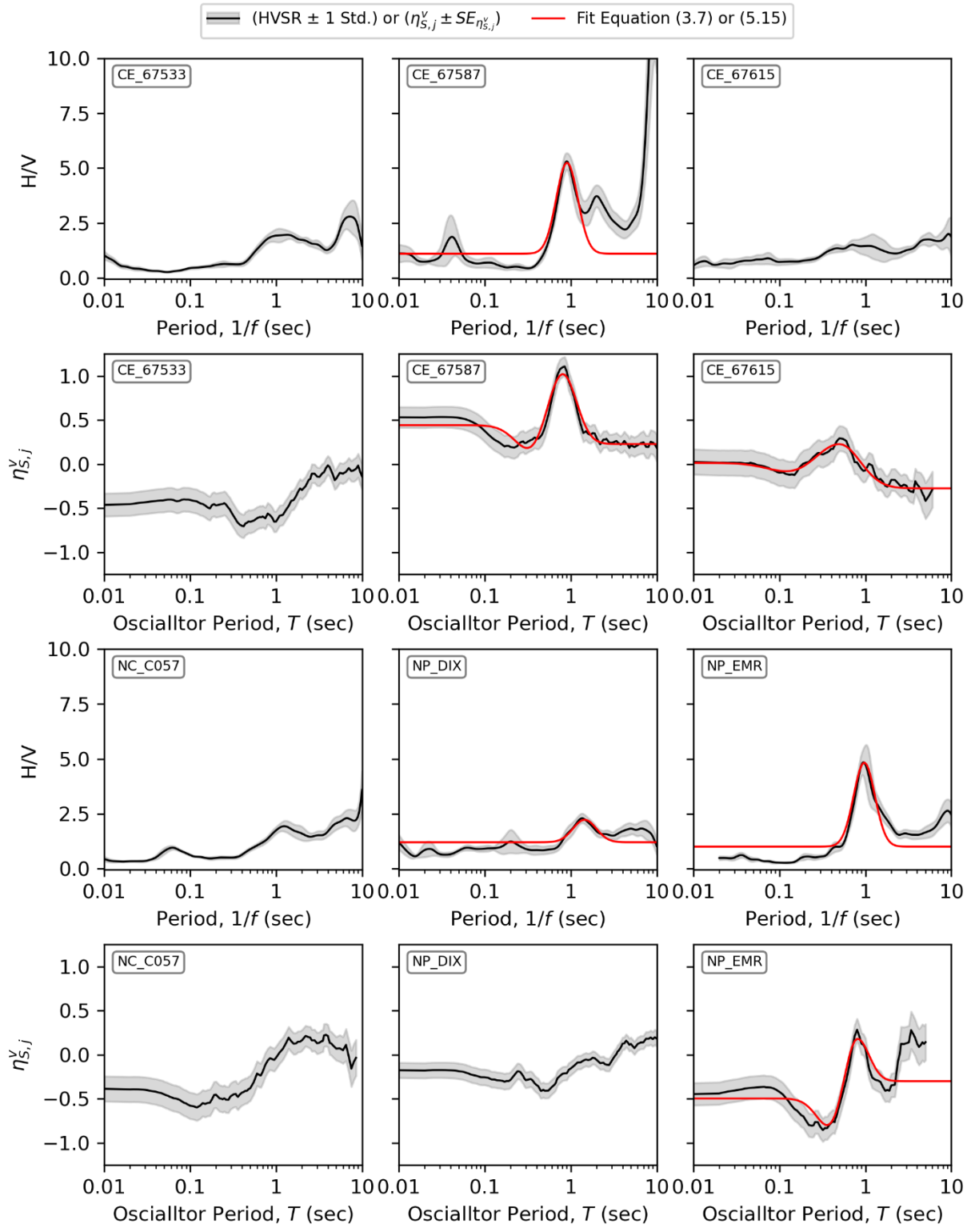


Figure 5.23: *continued.*

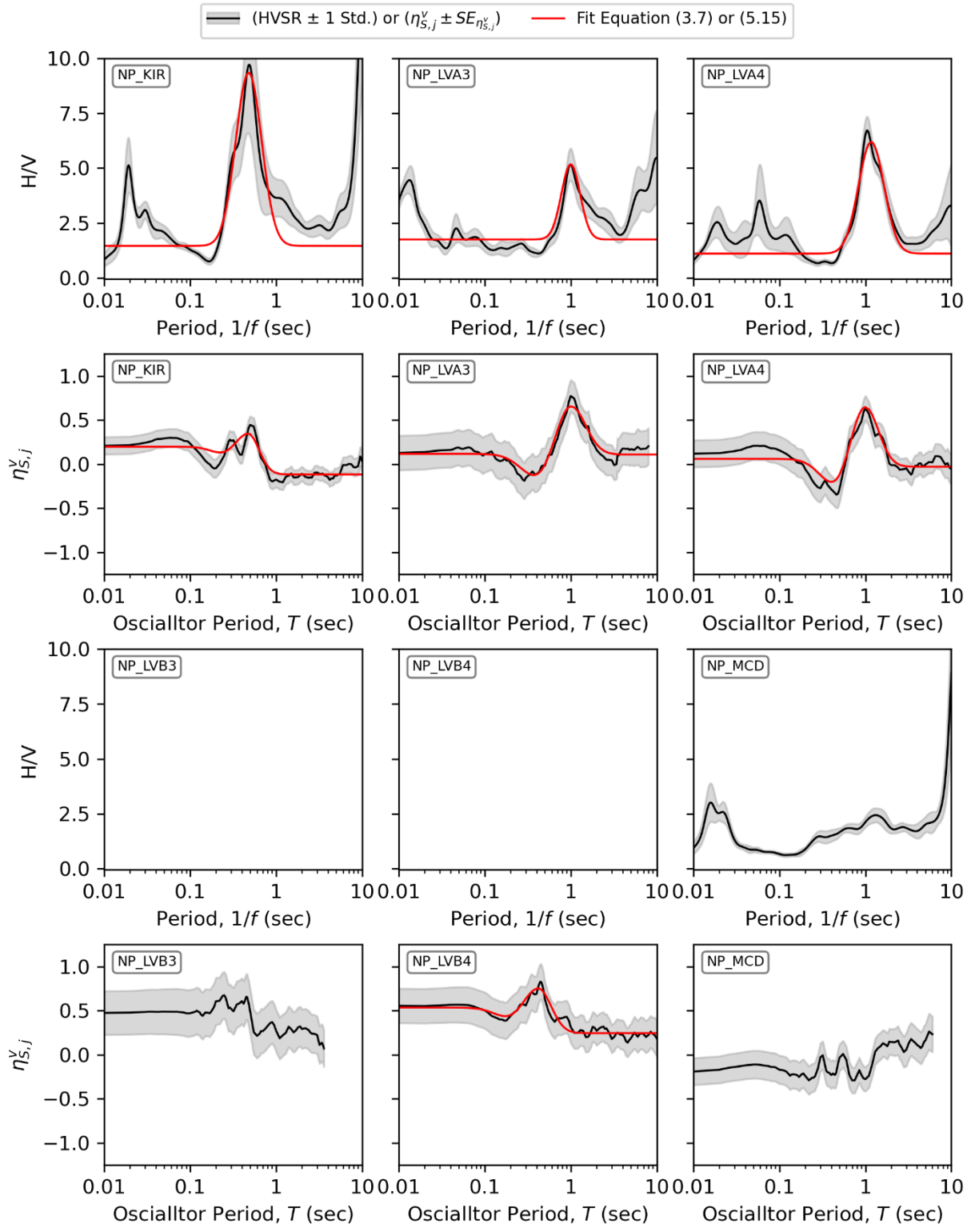


Figure 5.23: *continued.*

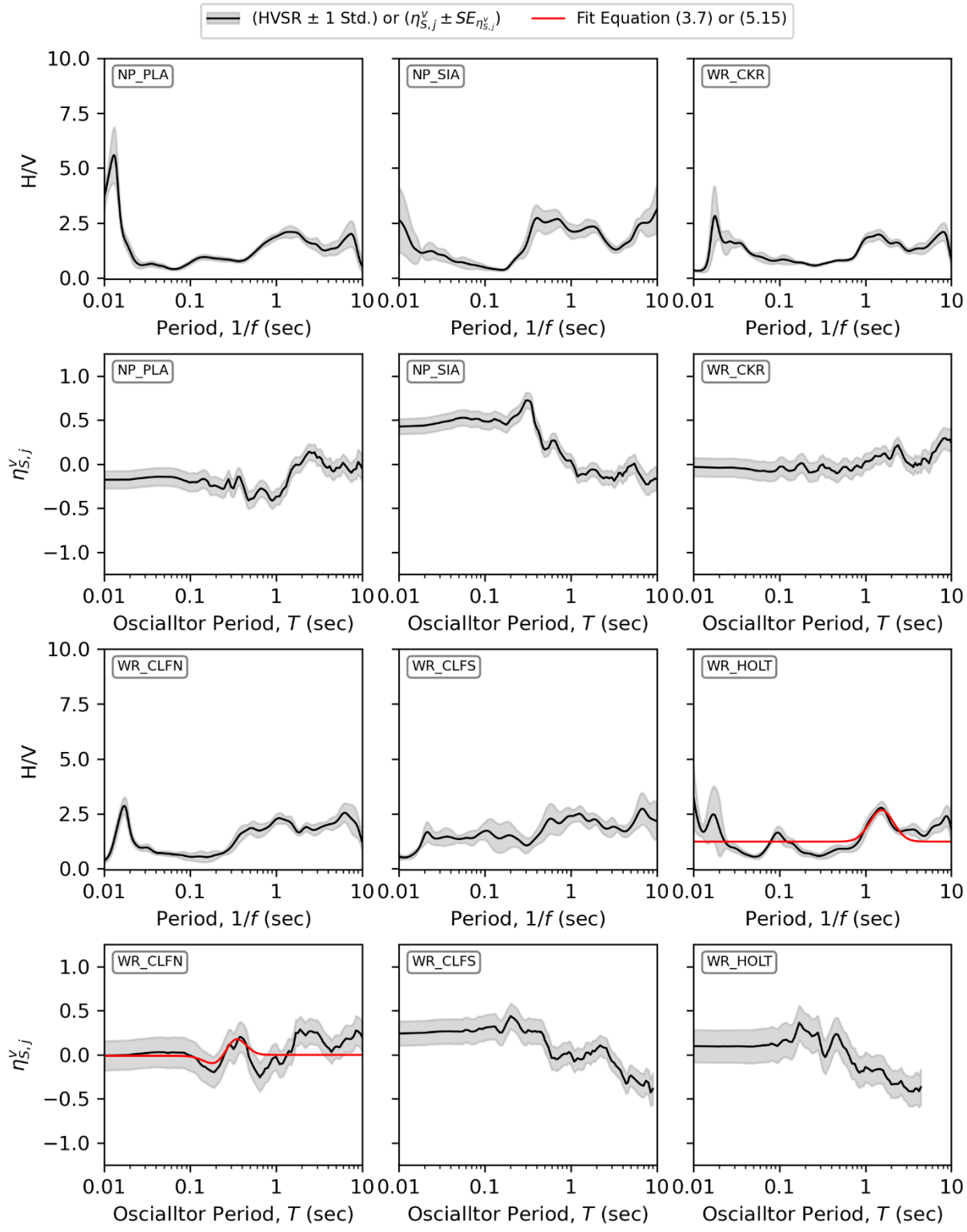


Figure 5.23: *continued.*

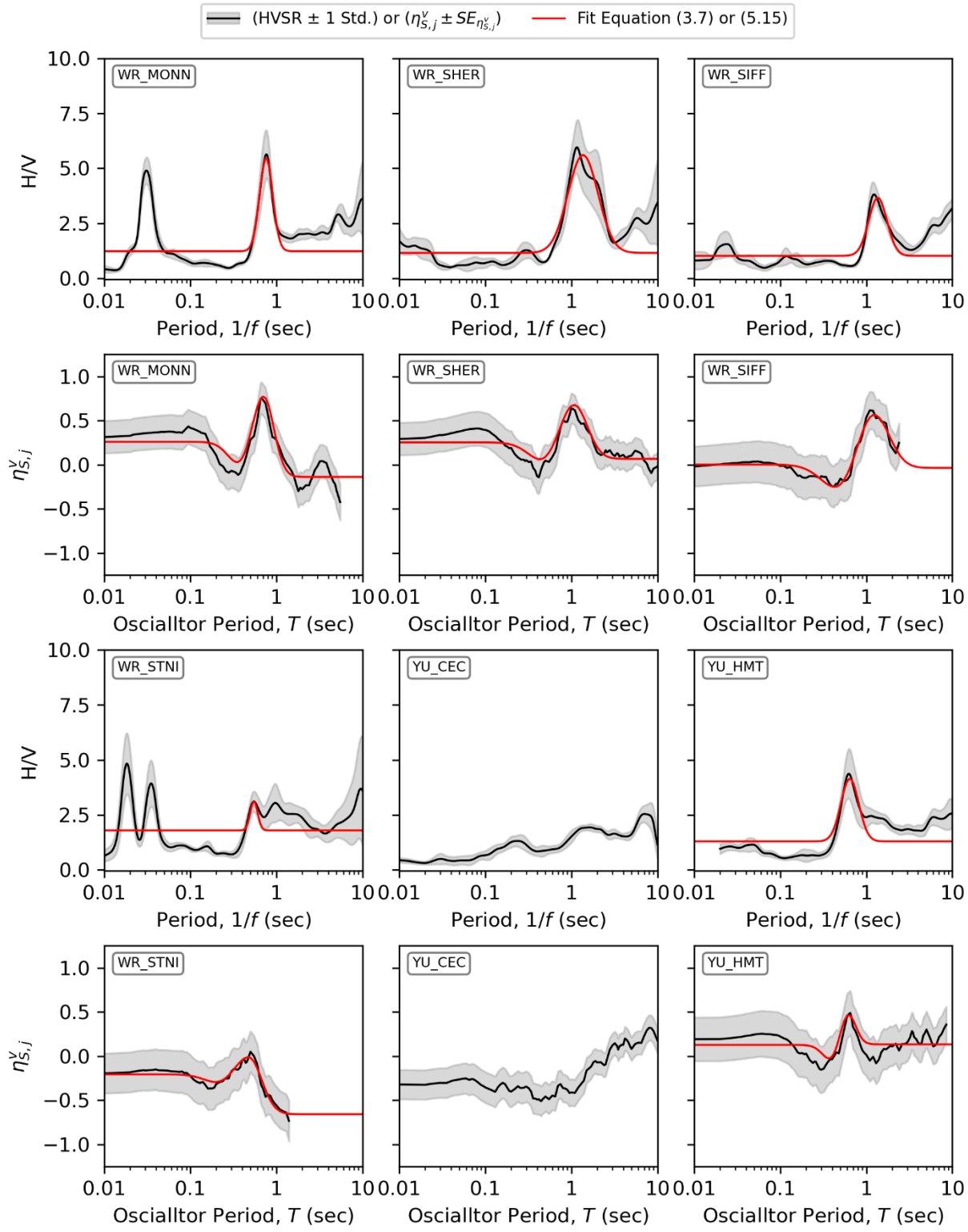


Figure 5.23: *continued.*

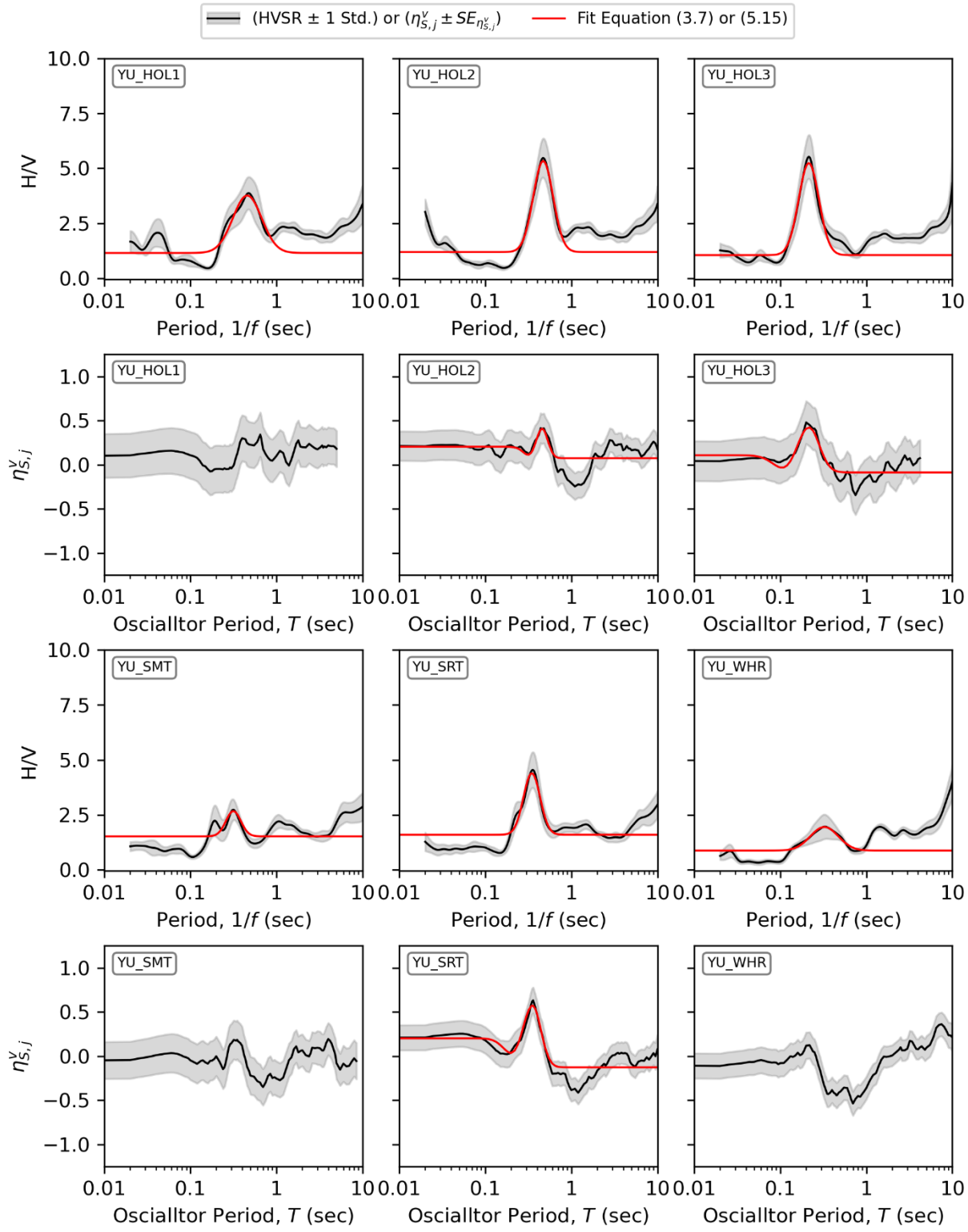


Figure 5.23: *continued.*

A recurring feature of the plots shown in Figure 5.23 is that site-resonance effects are generally only observed at sites with prominent mHVSR peaks (i.e., relatively large a_p). To test this observation, I construct correlation plots between pairs of peak site parameters derived from mHVSR distinguishing sites with and without $\eta_{S,j}^{v}$ peak features, as shown in Figure 5.24.

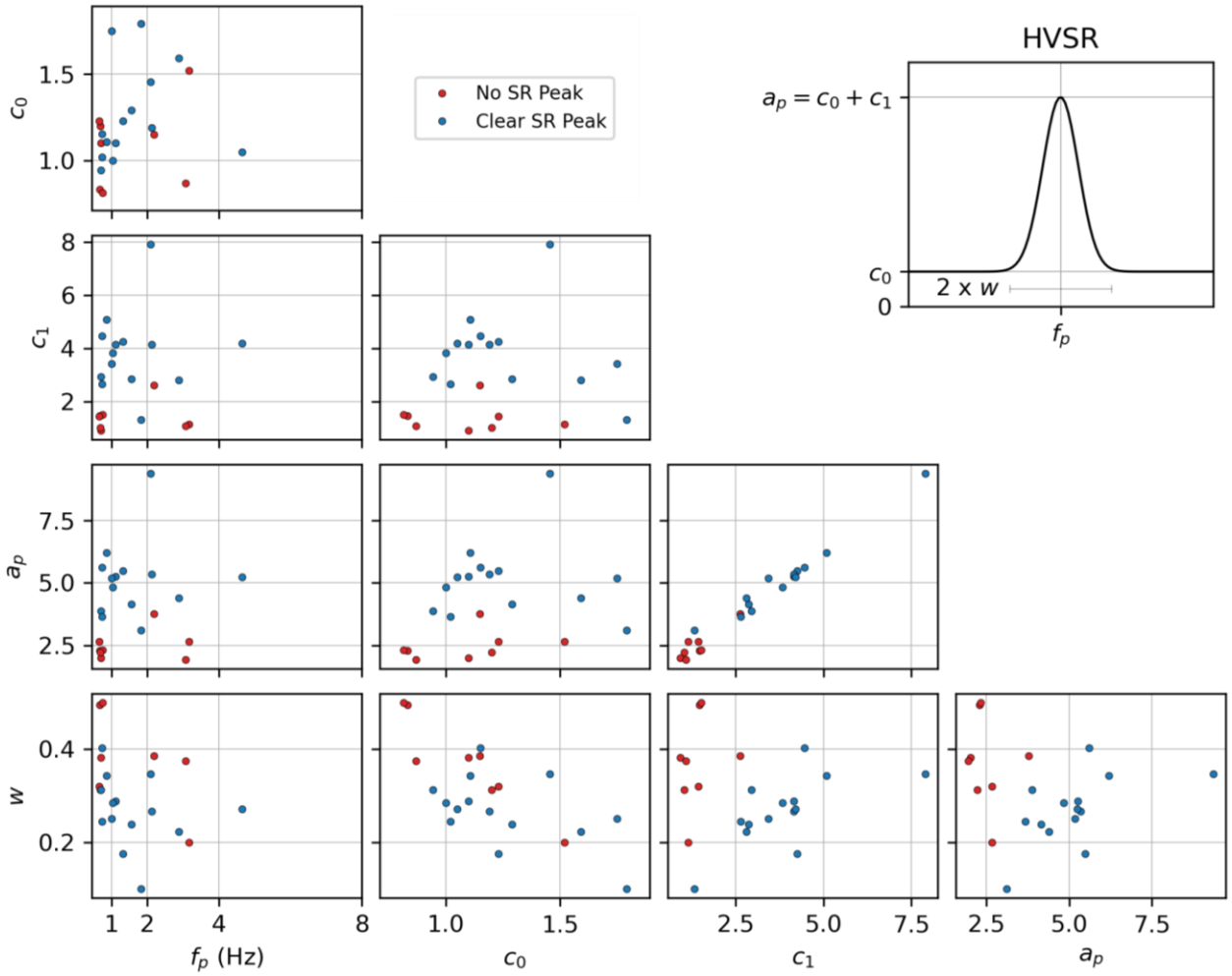


Figure 5.24: Correlation plots between pairs of peak site parameters derived from mHVSR.

The plots in Figure 5.24 show that sites with clear site response peaks generally have mHVSR peaks with $a_p \geq 3$, and small tail amplitudes (c_0). Using logistic regression (Lottes et al.

1996), I develop a probabilistic model conditioned on c_0 and a_p to predict the presence of site response peaks:

$$P_{peak}(c_0, a_p) = \frac{1}{1+\exp[-Q_{peak}(c_0,a_p)]} \tag{5.18}$$

where

$$Q_{peak}(c_0, a_p) = \beta_0 + c_0\beta_1 + a_p\beta_2 \tag{5.19}$$

where P_{peak} represents the probability of a site response peak ($0 \leq P_{peak} \leq 1$), and estimates for β_0, β_1, and β_2 are tabulated in Table 5.8.

Table 5.8: Estimates and standard errors for coefficients in Equation (5.19).

Coefficient	Value	Standard Error
$\boldsymbol{\beta_0}$	-19.2471	11.2033
$\boldsymbol{\beta_1}$	3.8467	3.0136
$\boldsymbol{\beta_2}$	4.3943	2.3088

Figure 5.25 illustrates the variation of peak probabilities in c_0 versus a_p space. A relatively narrow band separates sites predicted to have or not have site response peaks. The main outlier in Figure 5.25 is the YU_HOL1 site, which was not identified as having a site response peak but has a predicted peak probability of P_{peak} = 0.85. YU_HOL1 is characterized with a_p and c_0 values similar to other sites exhibiting site-resonance effects in $\eta^v_{S,j}$. However, as shown by the plots in Figure 5.23, a local high in the mean $\eta^v_{S,j}$ is present at around 0.4-0.6 s, but a peak is not identified because the uncertainty is relatively large.

To evaluate the model performance, we compare its predictions to a reference model, referred to here as a binomial model, in which the presence of a mHVSR peak is assumed to correspond to a site response peak. This comparison, shown in Table 5.9, shows that the logistic

model is an improvement to the binomial model as reflected by increased accuracy, precision, recall, and F_1 scores. It is important to note that the predictive model for peak probabilities in Equations (5.18) and (5.19) was developed for Delta sites and has unknown applicability for sites in other domains.

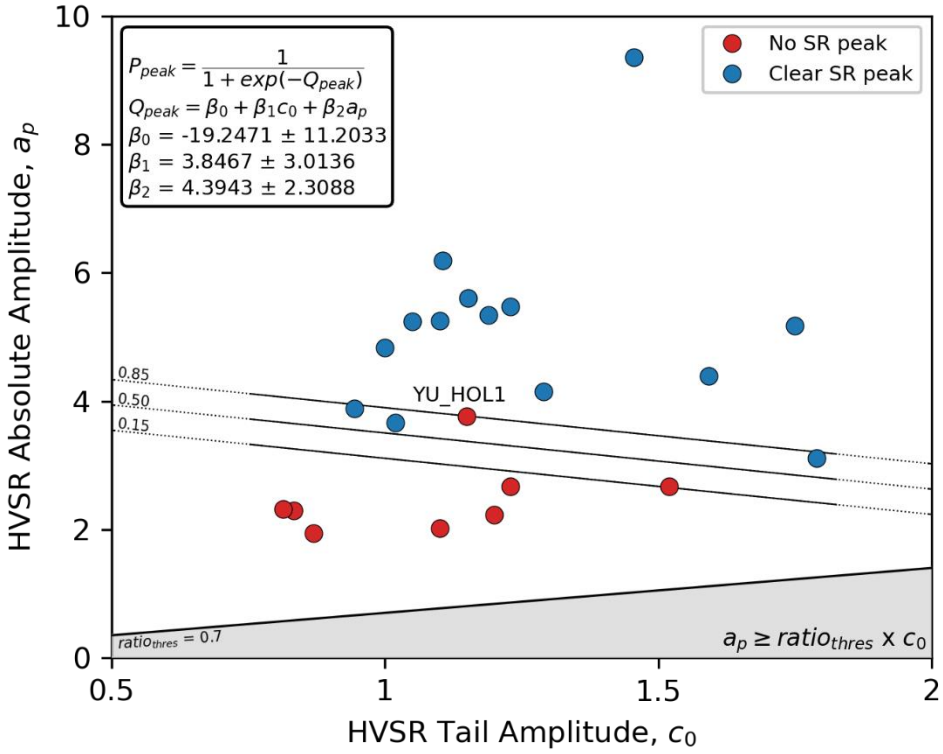


Figure 5.25: Probability distribution of logistic model for expecting site-resonance effects in c_0 versus a_p space.

Table 5.9: Qualitative comparison of binomial and logistic site-resonance effects predictive models based on mHVSR peak features.

Binomial Model (peak in mHVSR = peak in $\eta_{S,j}^{v}$ and no peak in mHVSR = no peak in $\eta_{S,j}^{v}$)						
	Correct Prediction	**Incorrect Prediction**	**Accuracy**	**Precision**	**Recall**	**F_1-Score**
mHVSR Peak	14	8	70.59%	63.64%	87.50%	73.68%
mHVSR No Peak	10	2				
Logistic Model [Equation (5.18), assuming P_{peak} threshold of 50%]						
	Correct Prediction	**Incorrect Prediction**	**Accuracy**	**Precision**	**Recall**	**F_1-Score**
mHVSR Peak	21	1	91.18%	95.45%	91.30%	93.33%
mHVSR No Peak	10	2				

(b) Using mHVSR Peak Parameters to Predict Amplification from Site Resonance

As discussed in Section 5.3.3, about 70% of Delta sites were classified as having site response peaks. In this sub-section, for those sites I relate site response peaks to site parameters derived from mHVSR. The aim is to model the site response peaks for use as a supplement to the V_{S30}-scaling model.

I model site response peaks using a function consistent with Equation (5.17), but with two changes: (1) the a_0 coefficient is dropped and (2) the amplitude is scaled by P_{peak}:

$$f_{1,peak}(T, f_p, P_{peak}) = \begin{cases} P_{peak}\frac{2a_1}{\sqrt{3a_2}\pi^{\frac{1}{4}}}\left[1-\left(\frac{\ln(T\hat{f}_p)}{a_2}\right)^2\right]\exp\left[-\frac{1}{2}\left(\frac{\ln(\hat{f}_p)}{a_2}\right)\right]^2 & T \le \hat{f}_p^{-1} \\ P_{peak}\left[f_1(\hat{f}_p^{-1}, \hat{f}_p) - a_3\right]\exp\left[-\frac{1}{2}\left(\frac{\ln\left((T\hat{f}_p)^{-1}\right)}{2a_2/3}\right)^2 + a_3\right] & T > \hat{f}_p^{-1} \end{cases} \quad (5.20)$$

where $\hat{f}_p$ is the fitted peak frequency of site amplification, and $a_1 - a_3$ are regression coefficients.

Figure 5.26 shows the relationships between mHVSR peak frequency (f_p) and each of the regression coefficients. As expected, the strongest correlation is with $\hat{f}_p$ but correlations with a_1 and a_2 are also observed. Parameter c_3 has weaker correlation with f_p and was set to a constant value of zero. Least squares regression is used to relate $\hat{f}_p$ to f_p as a linear function in natural log – natural log space, and a_1 and a_2 to f_p as bilinear functions in semi – natural log space (these functions saturate for relatively high f_p).

$$\hat{f}_p(f_p) = \exp\left[m \ln(f_p) + b\right] \tag{5.21}$$

$$a_1(f_p) = \begin{cases} m \ln(f_p) + b & f_p < f_{lim} \\ m \ln(f_{lim}) + b & f_p \geq f_{lim} \end{cases} \tag{5.22}$$

$$a_2(f_p) = \begin{cases} m \ln(f_p) + b & f_p < f_{lim} \\ m \ln(f_{lim}) + b & f_p \geq f_{lim} \end{cases} \tag{5.23}$$

Estimates and their associated standard errors for the coefficients in Equations (5.21) – (5.23) are given in Table 5.10.

Table 5.10: Estimates and standard errors for coefficients in Equations (5.21) – (5.23).

Equation	$\boldsymbol{m}$	$\boldsymbol{b}$	$\boldsymbol{f_{lim}}$ (Hz)
(5.21)	0.9265 ± 0.0212	0.0978 ± 0.0235	-
(5.22)	-0.1790 ± 0.0420	0.2355 ± 0.0157	2.08 ± 0.49
(5.23)	-0.3378 ± 0.1467	0.5213 ± 0.0344	1.55 ± 0.42

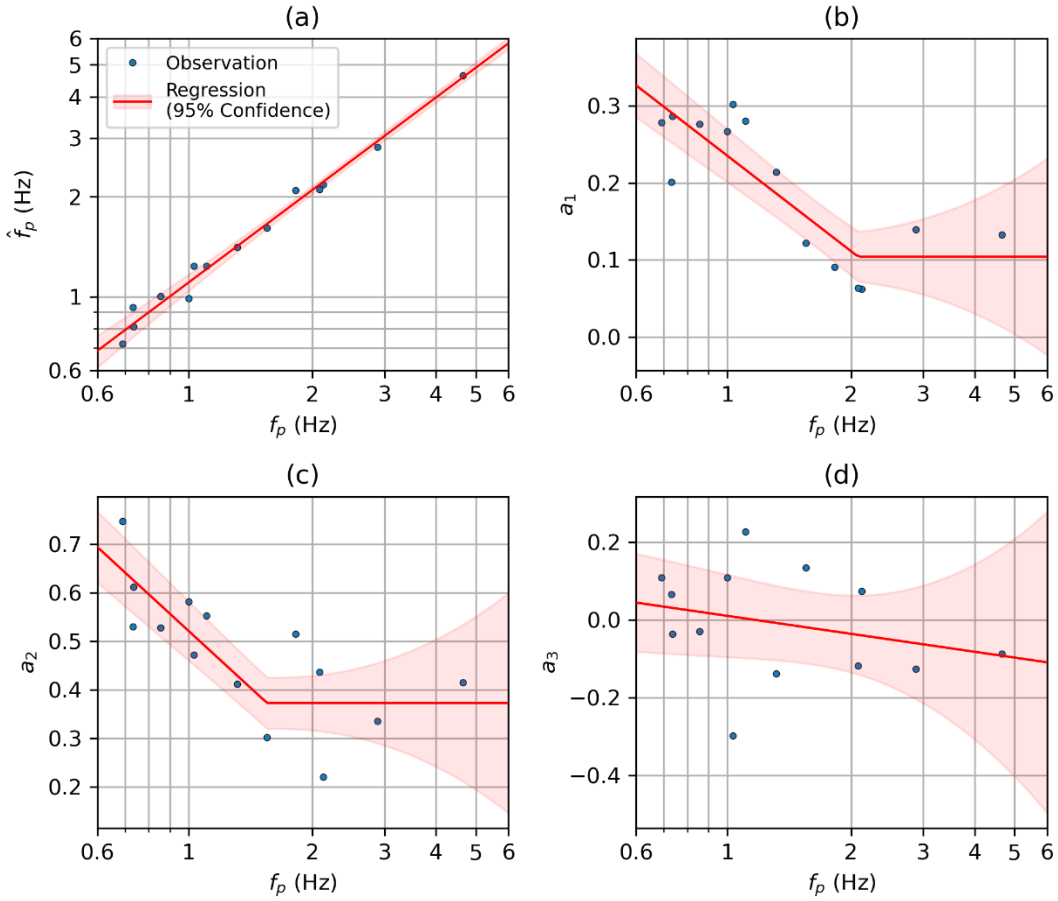


Figure 5.26: Relationships of mHVSR peak frequency (f_p) and coefficients of Equation (5.20): (a) $\hat{f}_p$; (b) a_1; (c) a_2; and (d) a_3.

The updated GMM uses the V_{S30}-scaling model in Equation (5.9)and the mHVSR-based model in Equation (5.20),

$$\delta W_{ij}^{v,p} = \ln(Y_{ij}) - [(c_0 + \Delta c_{0,r} + Ie_I) + F_E + F_P + F_{lin,ij}(V_{S30}) + f_{1,peak,j} + F_{nl,ij} + \eta_{E,i}] \quad (5.24)$$

where the linear site response is now a function of V_{S30}, f_p, a_p, and c_0 [i.e., $F_{lin}(V_{S30}, f_p, a_p, c_0) = F_{lin}(V_{S30}) + f_{1,peak}(f_p, a_p, c_0)$]. Using this model and the previously-derived event terms, within-event residuals ($\delta W_{ij}^{v,p}$) are computed and subsequently partitioned into site terms ($\eta_{S,j}^{v,p}$).

$$\delta W_{ij}^{v,p} = \eta_{S,j}^{v,p} + \epsilon_{ij} \quad (5.25)$$

Superscripts 'v' and 'p' indicate that that site terms are computed with a model that considers V_{S30} and mHVSR peak parameters. The remaining residual after the site term is accounted for is ϵ_{ij}.

Figure 5.27 shows the variation of $\eta_{S,j}^{v}$ and $\eta_{S,j}^{v,p}$ with oscillator period for Delta sites. As expected, the trend of $\eta_{S,j}^{v,p}$ is flatter than $\eta_{S,j}^{v}$ especially for periods between 0.4 and 2.5 s at sites with P_{peak} > 50% (where site resonance effects are the most pronounced). Additionally, the variability between 0.4 and 2.5 s is significantly reduced. These observations indicate that lowest-frequency mHVSR peaks ($f_{1,peak}$) improve the modeling of site effects related to resonance features. However, relatively large variability at short periods remains, and this model produces no modification of the V_{S30}-scaling model for sites with low P_{peak}.

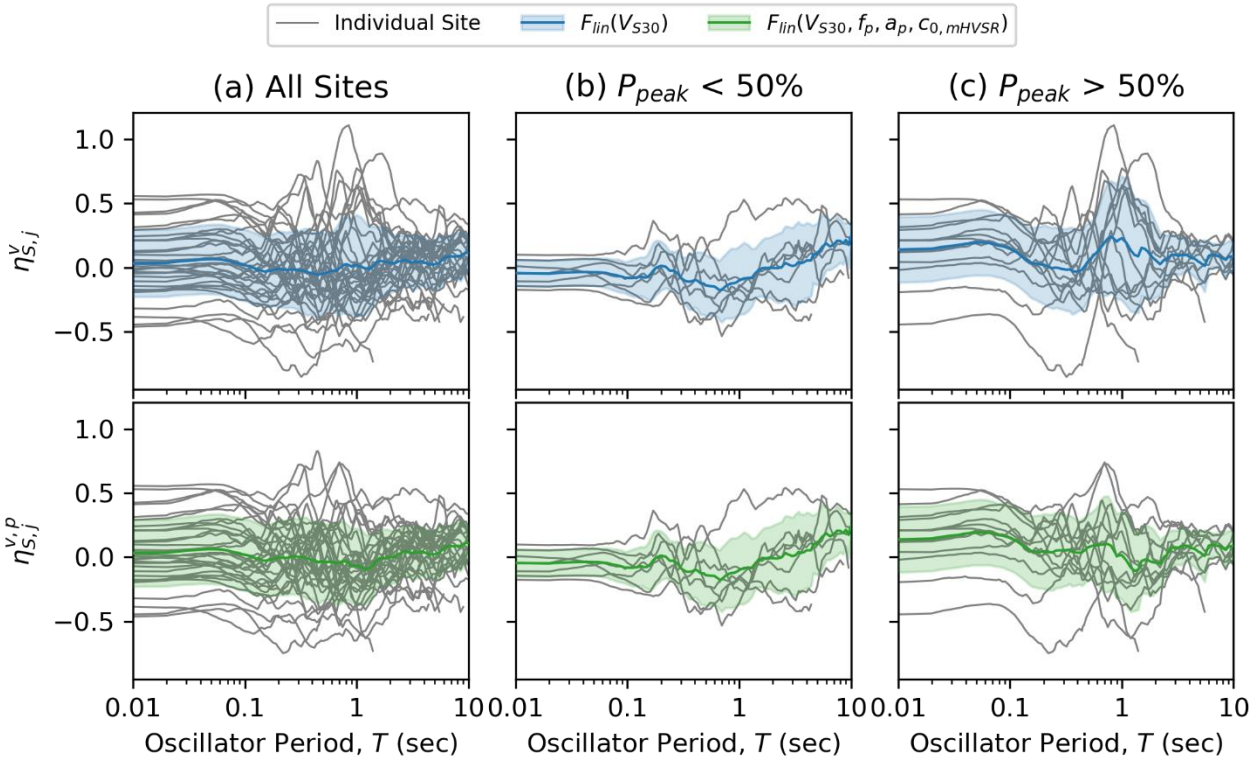


Figure 5.27: Comparison of bias in average site amplification predicted by $F_{lin}(V_{S30})$ and $F_{lin}(V_{S30}, f_p, a_p, c_0)$ for (a) all Delta sites, (b) sites predicted to have minimal site resonance (P_{peak} < 50%), and (c) sites predicted to have impactful site resonance (P_{peak} > 50%).

(c) Using mHVSR to Predict General Amplification

The effect of including site parameters derived from mHVSR peak features is a reduction in bias and variability within the period range affected by site resonance effects (periods between 0.4 and 2.5 s). As shown in Figure 5.27, the average remaining bias ($\bar{\eta}_{S,j}^{v,p}$) is period dependent, with relatively flat trends for any site condition at periods shorter than about 1 s. Furthermore, sites predicted to have site response peaks (P_{peak} > 50%) have a positive bias, while sites with P_{peak} < 50% have a slightly negative bias within this period range. To address these biases, we investigate their association with mHVSR amplitudes.

Precedent has been set by previous research efforts to incorporate mHVSR amplitudes into site response models. Senna et al. (2008) proposed a site response model for Japan conditioned on mHVSR and geologic or topographic units, where the shape of a reference spectrum is modified by a period dependent factor based on the mHVSR shape. Pinilla-Ramos et al. (2022) propose a mHVSR-based site response model for California, which similarly uses the whole period-dependent spectrum to predict amplification. These approaches differ from the work presented in this chapter because they do not systematically separate site resonance effects from general amplification. Pinilla-Ramos et al. (2022) present a model conditioned on mHVSR and V_{S30}, where mHVSR was found to have the greatest influence for periods between 0.5 and 4 s. These findings suggest that use of mHVSR amplitudes may have value for site response prediction, which is examined next using the Delta subregional data set.

Given that trends of $\eta_{S,j}^{v,p}$ are relatively flat for short-to-intermediate periods, I seek to capture the general level of site amplification with mHVSR amplitudes via a constant adjustment which goes to zero at long periods. Equation (5.26) presents the function used to model general

amplification ($f_{1,mean}$) conditioned on the mean mHVSR amplitude computed over some frequency bandwidth (μ_{mHVSR}).

$$f_{1,mean}(T, \mu_{mHVSR}) = \frac{a_4}{1+\exp\left[1.702\left(\frac{ln(T/T_1)}{0.385}\right)\right]} \tag{5.26}$$

where T_1 is the period on which the ramp function is centered (1.5 s), and

$$a_4(\mu_{mHVSR}) = \begin{cases} m_1\,\mu_1 - m_0 & \mu_{mHVSR} < \mu_1 \\ m_1\,\mu_{mHVSR} - m_0 & \text{otherwise} \\ m_1\,\mu_2 - m_0 & \mu_{mHVSR} > \mu_2 \end{cases} \tag{5.27}$$

where m_1 describes the slope of amplification scaling with respect to μ_{mHVSR}, m_0 is a constant, and μ_1 and μ_2 represent limiting values below- and above-which amplification scaling saturates, respectively. The logistic relation in Equation (5.26)produces constant amplification for $T < T_1 =$ 1.5 s followed by a transition to zero, as shown in Figure 5.28. The constant values in Equation (5.26)were selected to provide a smooth transition to zero amplification at periods greater than about 3 s.

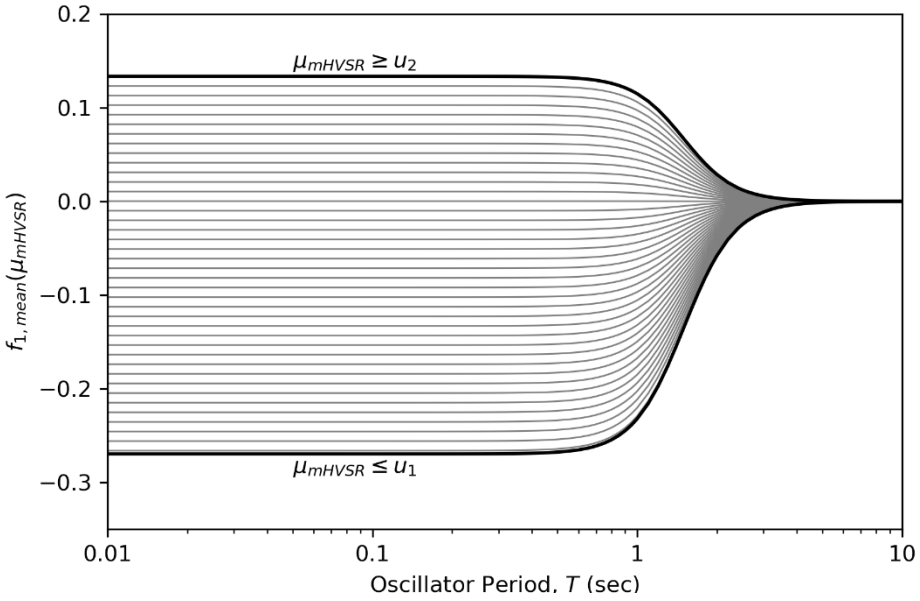


Figure 5.28: Schematic illustrating the period dependence of the mHVSR-based constant amplification adjustment ($f_{1,mean}$).

To estimate the coefficients in Equation (5.27), I compute the average values of $\eta_{S,j}^{v,p}$ (denoted $\mu_{\eta_{S,j}^{v,p}}$) and μ_{mHVSR} over different period bandwidths. μ_{mHVSR} is computed from lognormally spaced amplitudes (i.e., constant number of samples per log cycle) to reduce sampling bias. Figure 5.29 presents example plots of $\mu_{\eta_{S,j}^{v,p}}$ versus μ_{mHVSR} calculated over different period ranges, and the results of regressing Equation (5.27). In cases where there are meaningful trends, there is a positive correlation ($\mu_{\eta_{S,j}^{v,p}}$ increases as μ_{mHVSR} increases). Inspection of similar plots for many period ranges and consideration of the coefficients of determination (R^2), led me to conclude that the combination of $\mu_{\eta_{S,j}^{v,p}}$ computed over the period range of 0.01 to 3 s and μ_{mHVSR} computed from 0.33 to 50 Hz provides the strongest correlation as shown in Figure 5.29(f) (50 Hz is used as the upper frequency limit because it coincides with the Nyquist frequency of many permanent seismometers). The recommended values for m_0, m_1, μ_1, and μ_2 are given in Table 5.11.

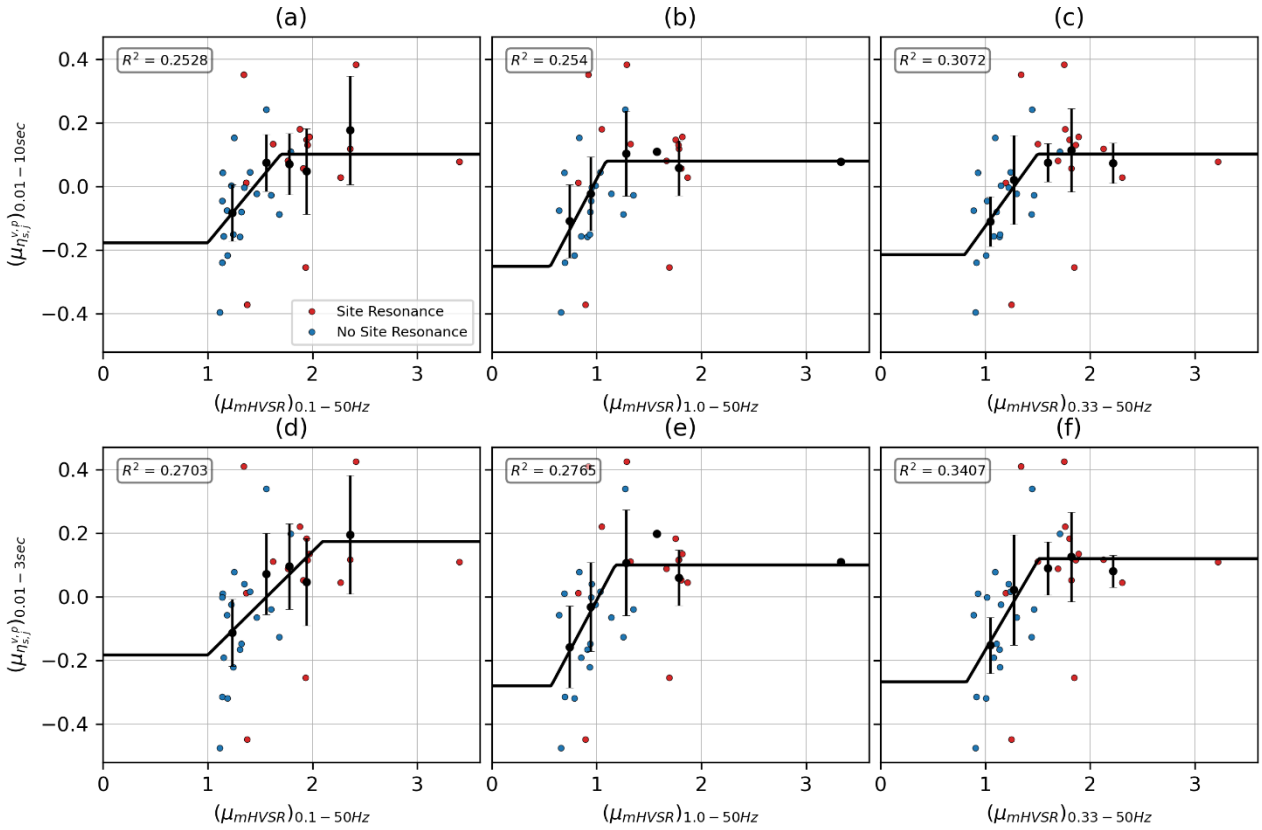


Figure 5.29: Relationship between average unmodeled site response after considering V_{S30}-scaling and peak-resonance effects ($\mu_{\eta_{S,j}^{v,p}}$) computed for period ranges of 0.01-10 sec (a-c) and 0.01-3 sec (d-f) with mean mHVSR amplitudes (μ_{mHVSR}) computed over frequency ranges of 0.1-50 Hz (a and d), 1-50 Hz (b and e), and 0.33-50 Hz (c and f).

Table 5.11: Estimates and standard errors for coefficients in Equation (5.27).

Coefficient	Value	Standard Error
$\boldsymbol{m_0}$	0.5760	0.2122
$\boldsymbol{m_1}$	-0.7435	0.2516
$\boldsymbol{\mu_1}$	0.8228	0.0000
$\boldsymbol{\mu_2}$	1.5224	0.1728

(d) Remaining Site Response Features

The additional amplification contributed by mHVSR-based site parameters to the regional site response model, which is added to $F_{lin}(V_{S30})$, may be described as the sum of Equations (5.20) and (5.26):

$$F_{lin}(f_p, a_p, c_0, \mu_{mHVSR}) = f_{1,peak}(T, f_p, a_p, c_0) + f_{1,mean}(T, \mu_{mHVSR}) \quad (5.28)$$

Using this model and the previously-derived event terms, within-event residuals ($\delta W_{ij}^{v,H/V}$) are computed

$$\delta W_{ij}^{v,H/V} = \ln(Y_{ij}) - [(c_0 + \Delta c_{0,r} + Ie_I) + F_E + F_P + F_{nl,ij} + F_{lin,j}(V_{S30}) + F_{lin,j}(f_p, a_p, c_0, \mu_{mHVSR}) + \eta_{E,i}] \quad (5.29)$$

where the linear site response is a function of V_{S30}, f_p, a_p, c_0, and μ_{mHVSR}:

$$F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR}) = F_{lin}(V_{S30}) + F_{lin}(f_p, a_p, c_0, \mu_{mHVSR}) \quad (5.30)$$

Those residuals are subsequently partitioned into site terms ($\eta_{S,j}^{v,H/V}$).

$$\delta W_{ij}^{v,H/V} = \eta_{S,j}^{v,H/V} + \epsilon_{ij} \quad (5.31)$$

Superscripts 'v' and '*H/V*' indicate that that site terms are computed with a model that considers V_{S30}, mHVSR peak parameters, and mHVSR amplitude parameters. The remaining residual after the site term is accounted for is ϵ_{ij} Delta sites can be broadly grouped into four categories based on remaining $\eta_{S,j}^{v,H/V}$ features: (1) sites with relatively flat trends with respect to period; (2) sites with uncaptured peak-site resonance; (3) sites with upward trending $\eta_{S,j}^{v,H/V}$ with increasing period; and (4) sites with downward trending $\eta_{S,j}^{v,H/V}$ with increasing period, as shown in Figure 5.30.

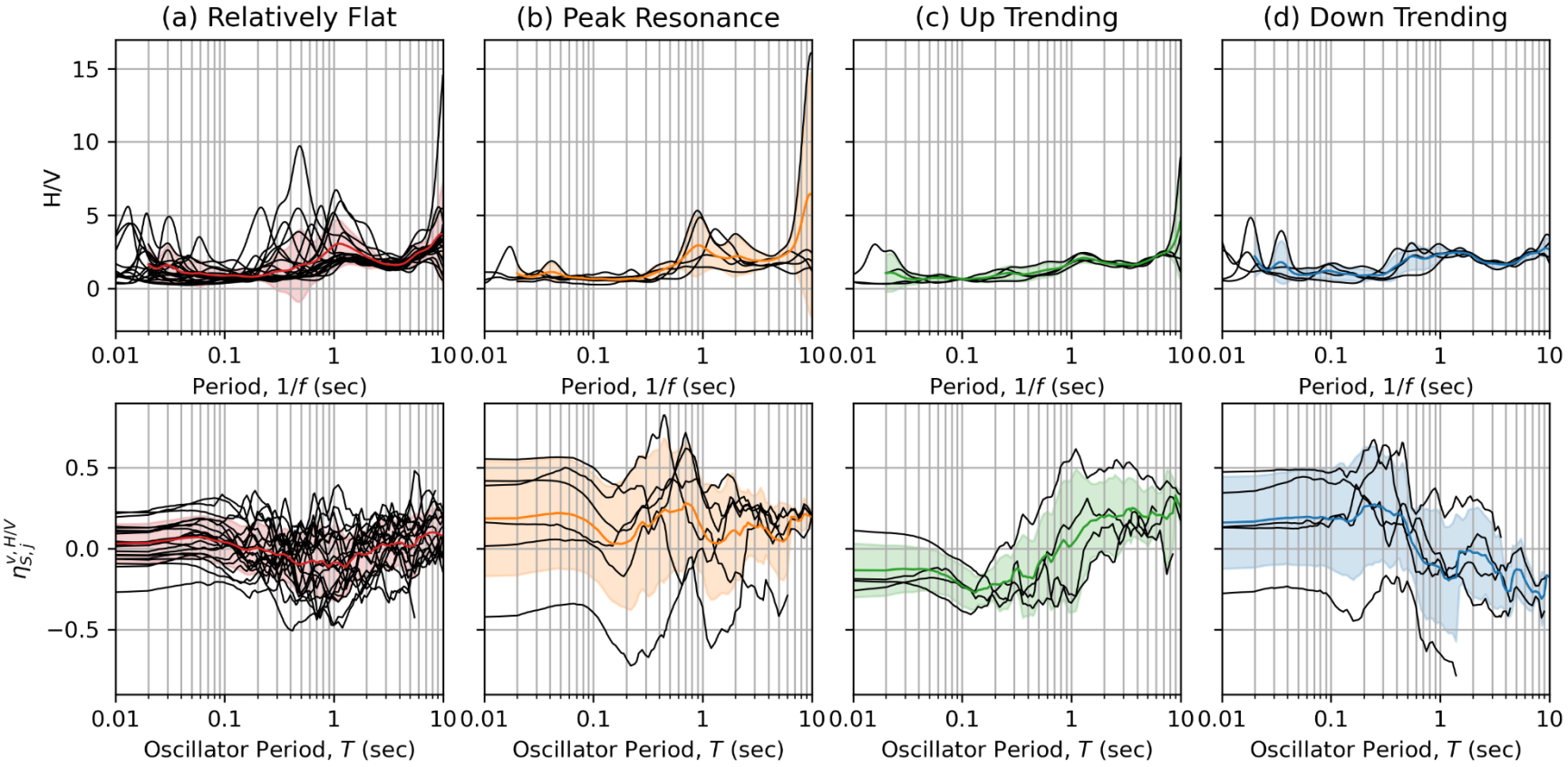


Figure 5.30: Plots of grouped mHVSR and $\eta_{S,j}^{v,H/V}$ for Delta sites which exhibit (a) relatively flat $\eta_{S,j}^{v,H/V}$ trends, (b) uncaptured peak resonance, (c) upward trending $\eta_{S,j}^{v,H/V}$ with increasing period, and (d) downward trending $\eta_{S,j}^{v,H/V}$ with increasing period.

Sites with relatively flat $\eta_{S,j}^{v,H/V}$ trends exhibit amplification which is more or less captured by $F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR})$ over the full period range being modeled. Sites with uncaptured peak resonance (BK_TWIT, CE_67587, CE_67615, NP_EMR, NP_LVB4, and WR_CLFN) fall into one of three categories: (1) sites with no mHVSR available (NP_LVB4); (2) sites with no mHVSR peak (CE_67615 and WR_CLFN); or (3) sites with exceptionally large site resonance effects which are not fully accounted for by $f_{1,peak}$. The uncaptured peak response manifests from a lack of mHVSR data or poor fit prediction by $f_{1,peak}$ which represents the average regional behavior conditioned on f_p. Therefore, regardless of which category these sites fall into, little can be done to improve the site response predictions.

Sites that exhibit upward- and downward-trending $\eta_{S,j}^{v,H/V}$ with increasing period exhibit characteristic signatures of basin effects with positive and negative $\delta z_{1.0}$, respectively. Recall, USGS SFCVM does not provided detailed $z_{1.0}$ estimates in the Delta region, so traditional basin effect modeling approaches are not currently viable in the Delta (Section 5.3.2). Pinilla-Ramos et al. (2022) postulate that mHVSR amplitudes provide more information about long-period site response than basin depths (e.g., $z_{1.0}$), since they inherently capture long-period amplification associated with the deeper velocity structure. Therefore, I compare trends of mHVSR amplitudes between these groups and those of relatively flat-trending sites to assess the predictive power of mHVSR for basin effects. Plots of average mHVSR are included for each group in Figure 5.30, and are plotted together in Figure 5.31.

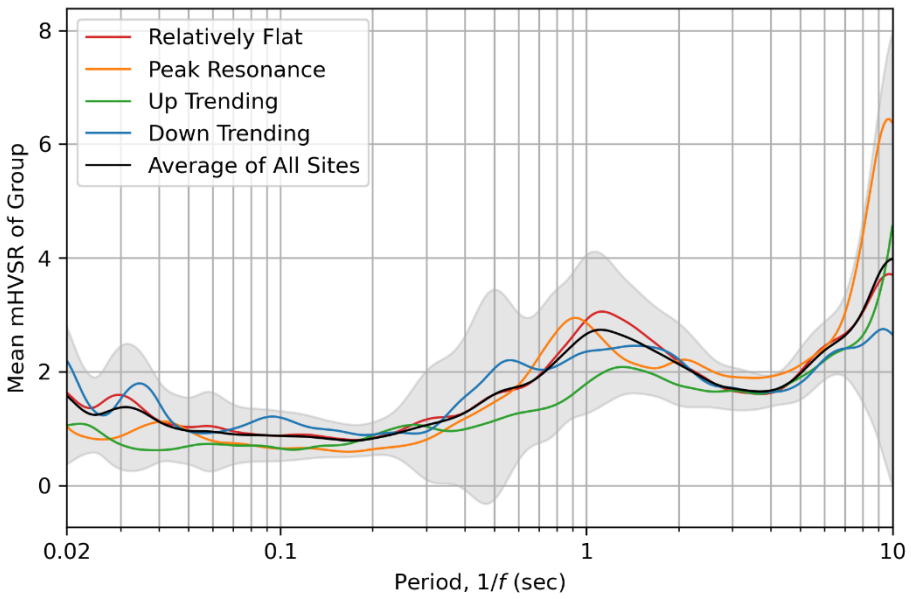


Figure 5.31: Average mHVSR for all Delta sites (black), sites with relatively flat $\eta_{S,j}^{v,H/V}$ (red), sites with uncaptured peak resonance (orange), sites with upward-trending $\eta_{S,j}^{v,H/V}$ (green), and sites with downward trending $\eta_{S,j}^{v,H/V}$ (blue).

The average mHVSR for each group exhibit similar trends, and all are within one standard deviation of the average mHVSR for all Delta sites (Figure 5.31) for all periods. Furthermore, when examining the spatial distribution of these sites (as shown in Figure 5.32), there appears to be no systematic clustering of any group. These observations suggests that mHVSR amplitudes may not significantly improve predictions of long-period (basin) amplifications within the current modeling context. However, these observations do not provide substantial evidence on their own that mHVSR amplitudes do not provide more information about long-period site response than basin depths, because $F_{lin}\left(V_{S30}, f_p, a_p, c_{0.mHVSR}, \mu_{mHVSR}\right)$ describes site effects conditioned on V_{S30} and mHVSR, both of which have been shown to correlate with $z_{1.0}$.

The remaining unmodelled site response effects are quantified in the aleatory variability model presented in Section 5.4.2.

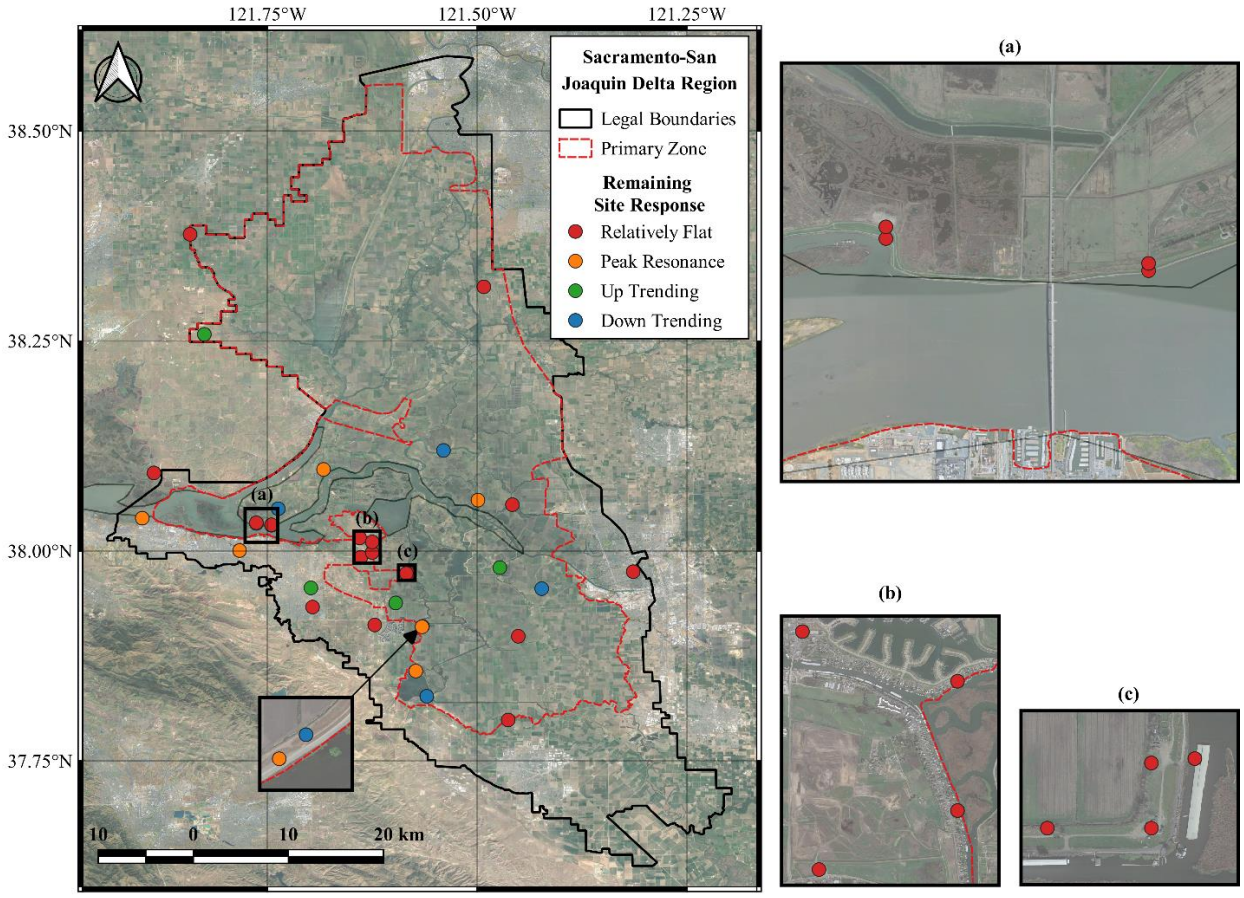


Figure 5.32: Map of locations of Delta sites grouped by trend of remaining site response: sites with relatively flat $\eta_{S,j}^{v,H/V}$ (red), sites with uncaptured peak resonance (orange), sites with upward-trending $\eta_{S,j}^{v,H/V}$ (green), and sites with downward trending $\eta_{S,j}^{v,H/V}$ (blue).

5.3.5. Effect of Nonlinearity

The models presented in Sections 5.3.1 – 5.3.4 were developed from relatively weak ground motions and are intended to predict linear site amplification. To investigate the potential effects of nonlinearity, I examine $\delta W_{ij}^{v,H/V}$ relative to *PGA* for the reference rock condition (PGA_r; for V_{S30} = 760 m/s) predicted using the GMM with the linear site terms in Equation (5.28) (Figure 5.33). A flat trend is expected if the site response from the events summarized in Table 5.2 is effectively linear, whereas nonlinearity would produce a downward trend for short period intensity measures.

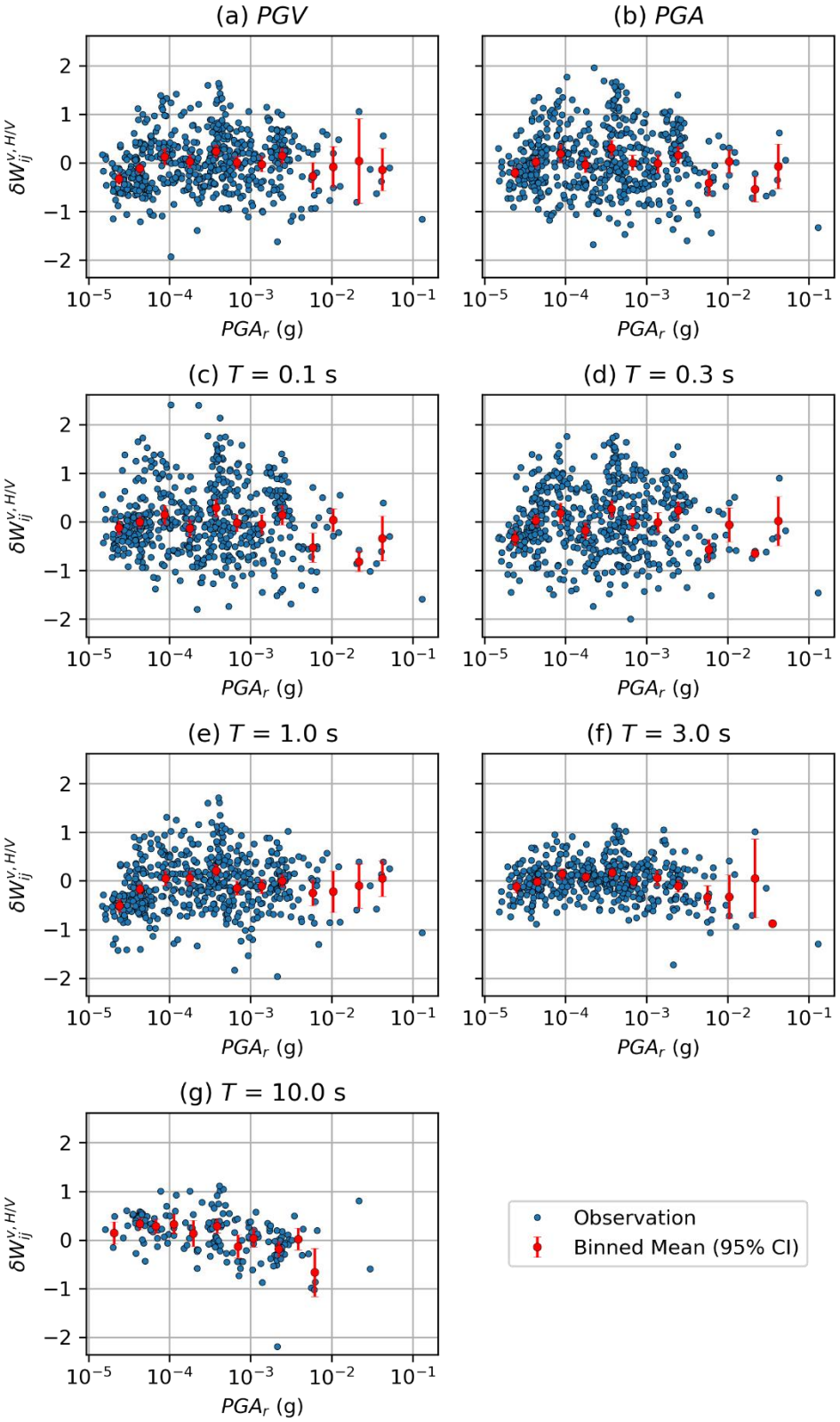


Figure 5.33: Plots of within-event residuals (δW_{ij}) versus PGA_r for (a) *PGV*, (b) *PGA*, and *PSA* at (c) T = 0.1s, (d) T = 0.3s, (e) T = 1.0s, (f) T = 3.0s, and (g) T = 10.0s.

The data are shown in Figure 5.33 along with binned means and their 95% confidence intervals (roughly three bins per log cycle of PGA_r). The largest PGA_r value is 0.13 g and 90% of the observations have PGA_r < 0.0025 g. The data trends are generally flat, being nearly perfectly so for PGA_r < 0.005 g. For stronger motions, there is an apparent downward trend at short periods, particularly for Sa at 0.1 sec, although the statistical significance of the trend is arguable because the data for stronger shaking are limited and the confidence intervals overlap zero. If the downward trends is interpreted as nonlinearity, it occurs at relatively low PGA_r levels relative to existing models for non-peat sits (e.g., 0.1 g; SS14). A relatively low transitional ground motion level for nonlinear response might be anticipated because of the exceptionally soft conditions in the Delta. Peaty organic soils are relatively soft, meaning that smaller ground motion levels are required to elicit large strains (shear strain is roughly proportional to PGV/V_S, and V_S is low for peats).

These observations suggest that nonlinearity may have influenced a small fraction (< 5%) of the ground motions used to estimate the nonergodic site response. As a result, my conclusion is that nonlinear effects are not expected to have a significant impact on the presented models.

5.4. Subregional Model Performance and Comparison to Global Ergodic Model

The recommended subregional model begins with the BSSA14 GMM, updates the anelastic path model as presented in Chapter 4, and updates the site response using one of two potential subregional models. The first subregional model requires V_{S30} only and is denoted $F_{lin}(V_{S30})$ (Equation 5.9). The second subregional model uses V_{S30} in combination with mHVSR parameters and is denoted $F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR})$ (Equation 5.30). This section presents a comparison of the predictions of three site response models: (1) SS14, (2) $F_{lin}(V_{S30})$, and (3) $F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR})$. Figure 5.34 presents individual plots for each Delta sites showing

amplification versus oscillator period as observed from non-ergodic site response and as predicted using each site response model.

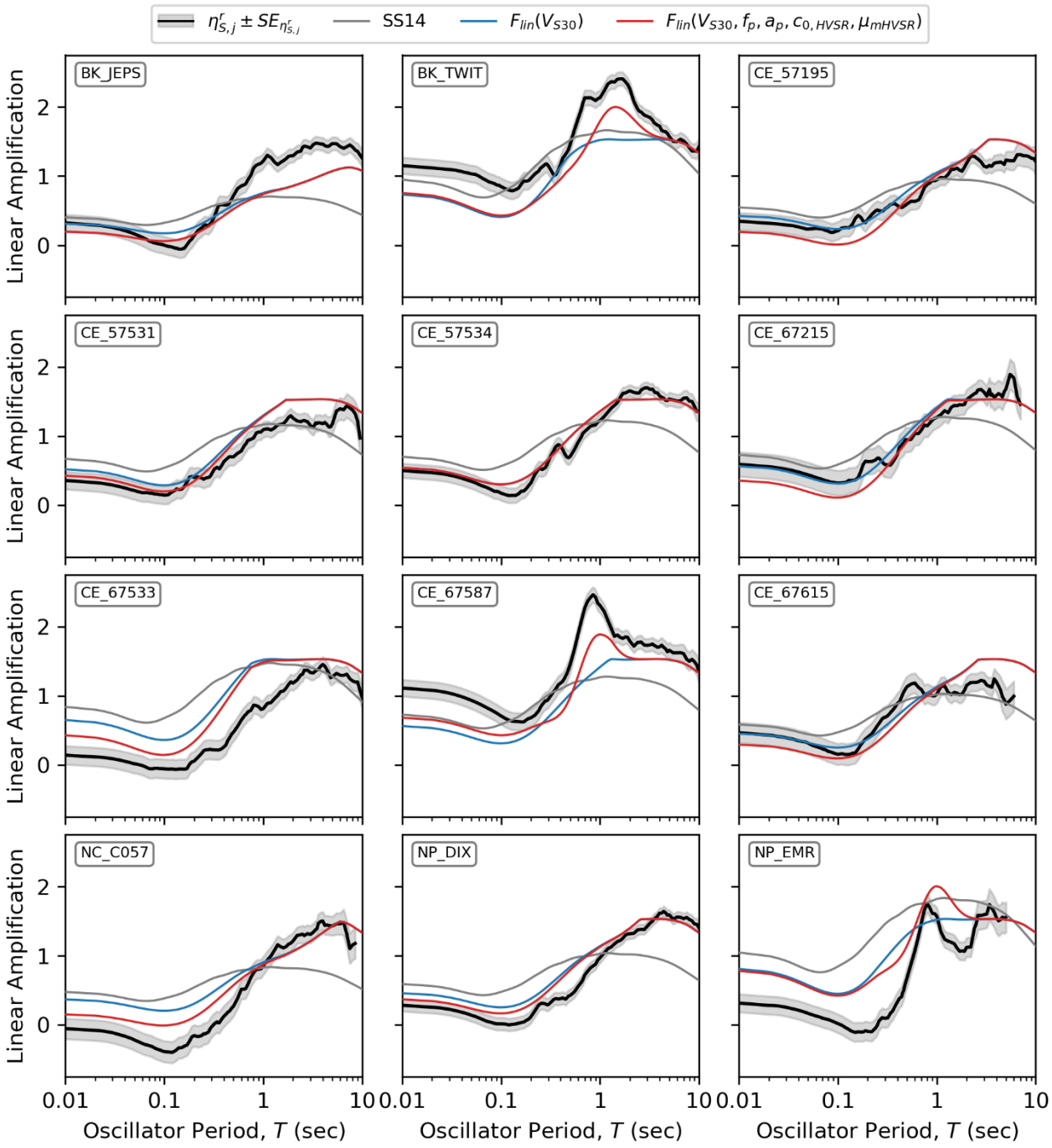


Figure 5.34: Comparisons of observed mean linear amplification $[(f_1)_j^o]$ with 95% confidence intervals and model predictions provided from SS14 (ergodic model) and two proposed regional models for the Delta $[F_{lin}(V_{S30})$ and $F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR})]$ for individual sites.

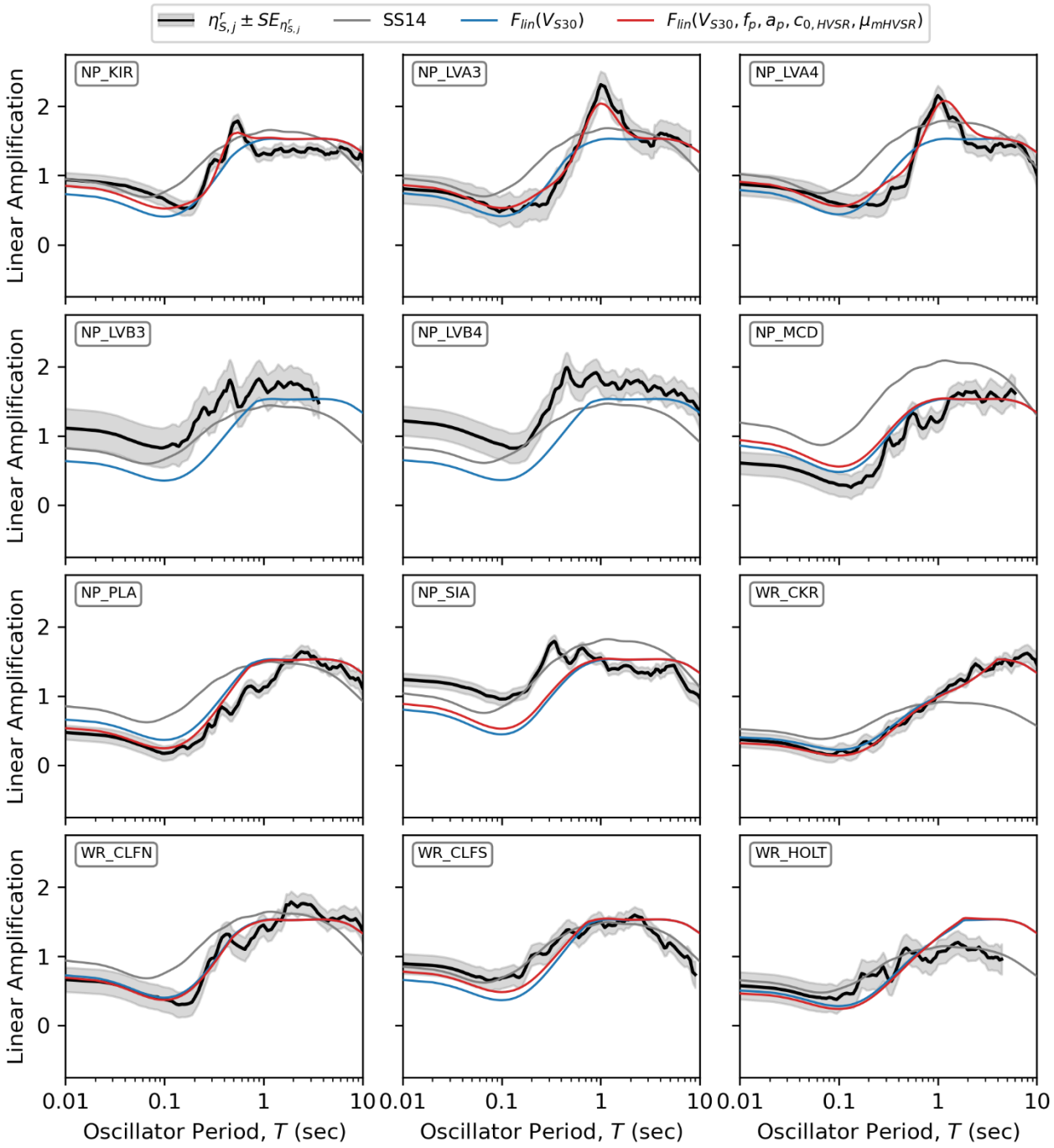


Figure 5.34: *Continued.*

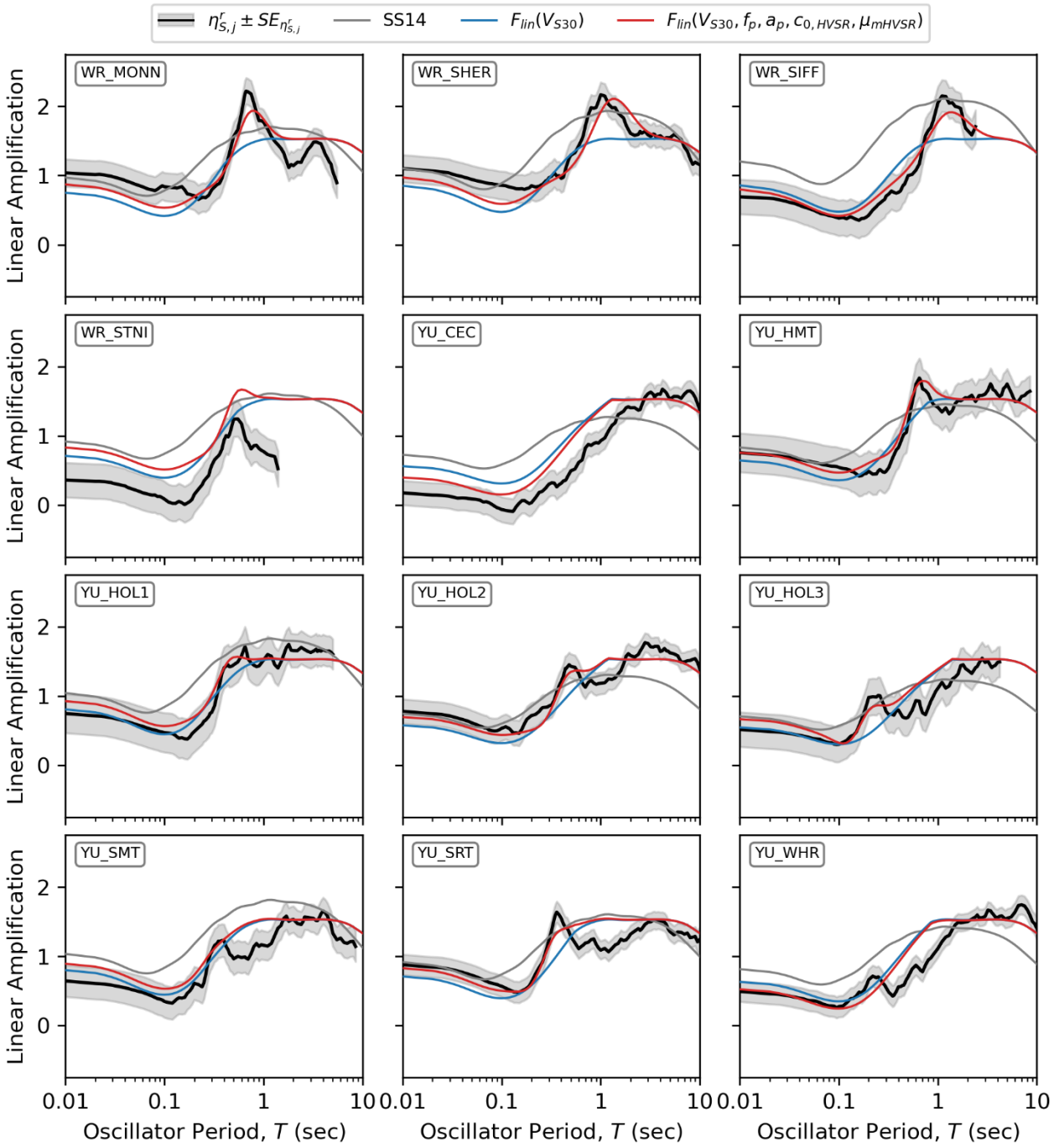


Figure 5.34: *Continued.*

The results shown in Figure 5.34 illustrate the significant bias when extrapolating SS14 to the soft-soil conditions encountered in the Delta, and the improved fits that are realized when using

the regionally calibrated site response models [$F_{lin}(V_{S30})$ and $F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR})$]. The following subsections quantitatively compare model bias and variability, and present aleatory variability models for forward application of the regional site response models.

5.4.1. Model Bias

Model bias is evaluated as the mean misfit of site terms computed when implementing each of the site response models. Figure 5.35 presents plots of $\eta_{S,j}$ versus oscillator period for all Delta sites, sites likely without peaks (P_{Peak} < 50%), and sites likely with peaks (P_{peak} > 50%). As expected, the regionally calibrated models [$F_{lin}(V_{S30})$ and $F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR})$] perform better than SS14 (i.e., are less biased). The substantial bias of SS14 demonstrates the need for region-specific site factors for exceptionally soft soils. When examining biases of the two subregional models (Figure 5.35a), there are negligible differences between $F_{lin}(V_{S30})$ and $F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR})$, both of which are effectively zero.

When examining P_{Peak} < 50% and P_{peak} > 50% sites, $F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR})$ generally performs better than $F_{lin}(V_{S30})$, except for periods > 3 s where the results are identical. The $F_{lin}(V_{S30})$ model under-predicts on average ground motions at short- to intermediate-periods (T < 3 s) for non-peak sites and over-predicts ground motions for this same period range for peak sites. When averaged across all sites, the model is unbiased. The $F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR})$ model removes these biases for both types of sites.

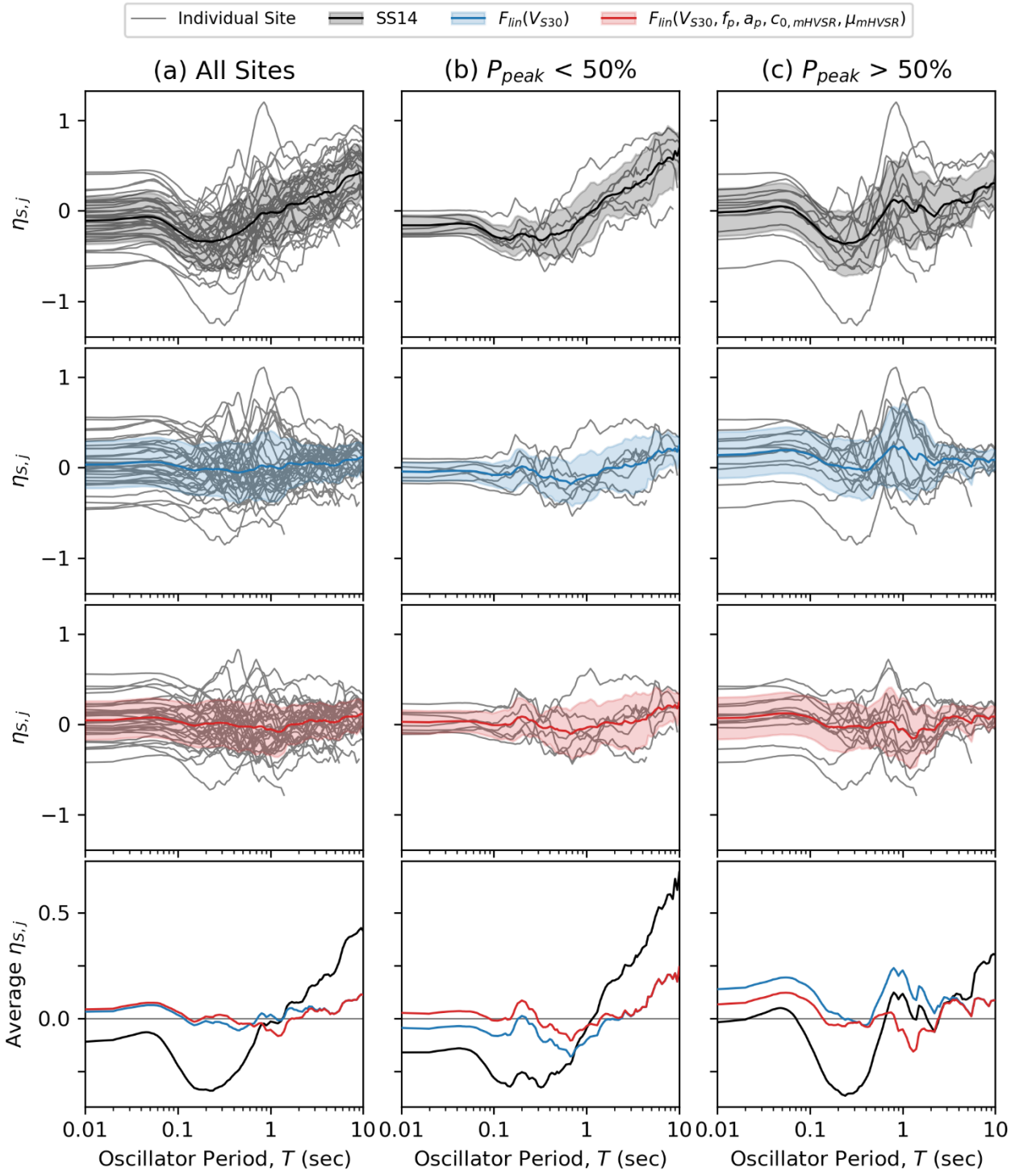


Figure 5.35: Comparison of bias in average site amplification predicted by SS14 and the two proposed regional site response models [$F_{lin}(V_{S30})$ and $F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR})$] for (a) all sites, (b) sites predicted to have minimal site resonance (P_{peak} < 50%), and (c) sites predicted to have impactful site resonance (P_{peak} > 50%).

5.4.2. Aleatory Variability

Residuals are computed using the BSSA14 GMM with the updated the anelastic path model as presented in CHAPTER 4:Chapter 4 and using one of three site response models: (1) SS14 (applied inside and outside of Delta), (2) subregional V_{S30}-based site response model (within-Delta only), and (3) subregional combined V_{S30}- and mHVSR-based site response model (within-Delta only). Aleatory variability terms that are computed include total (standard deviation of total residuals, σ), between-event (standard deviation of event terms, τ), within-event (standard deviation of within-event residuals, ϕ), site-to-site (standard deviation of site terms, ϕ_{S2S}), and single-station (standard deviation of remaining residuals ϵ_{ij}, ϕ_{SS}). These standard deviations are computed as part of the mixed effects analyses, not strictly as the standard deviations of the respective terms (the above descriptions are therefore qualitative). The within-event variability is related to ϕ_{S2S} and ϕ_{SS} as,

$$\phi = \sqrt{\phi_{S2S}^2 + \phi_{SS}^2} \tag{5.32}$$

The single-station variability (ϕ_{SS}) includes path-to-path variability, which was examined previously in Section 4.6, however it is revisited again here now considering data from Delta sites.

Figure 5.36 shows each of the standard deviation terms for the total California dataset and the Delta subset. The regionally calibrated site response models do not affect the results of the total California dataset. However, there are significant changes of within-event variability and its components (ϕ_{S2S} and ϕ_{SS}) for the Delta subset for most oscillator periods.

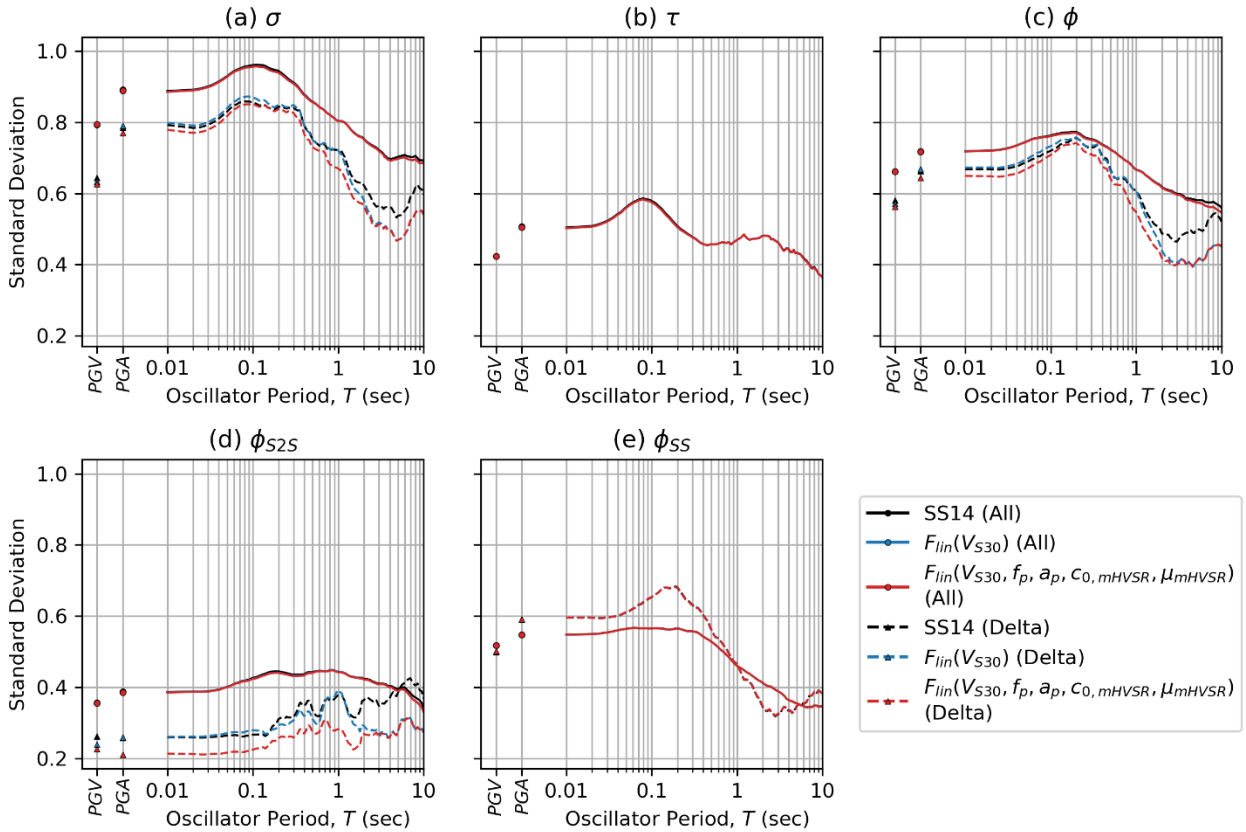


Figure 5.36: Standard-deviations calculated using SS14 and the proposed Delta-specific linear site response models which represent (a) total variability (σ); (b) between-event variability (τ); (c) within-event variability (ϕ); (d) site-to-site variability (ϕ_{S2S}); and (e) single-station within-event variability (ϕ_{SS}).

Goulet et al. (2018; Gea18) provides an ergodic ϕ_{SS} model conditioned on **M**, which is shown in Figure 5.37 with computed mean ϕ_{SS} values and their 95% confidence intervals from the Delta sites. The 2019 Ridgecrest **M**7.1 mainshock is the only event in the Delta subset with **M** $\geq$ 6.5, hence the large uncertainty and between-period variability in Figure 5.37(b). Focusing on the **M** < 5 data, the Gea18 model generally falls within the confidence intervals of the Delta results. Accordingly, I do not consider a Delta-specific ϕ_{SS} model to be justified. This conclusion is consistent with that reached in Section 4.6, which was also to recommend use of the Gea18 ϕ_{SS} model.

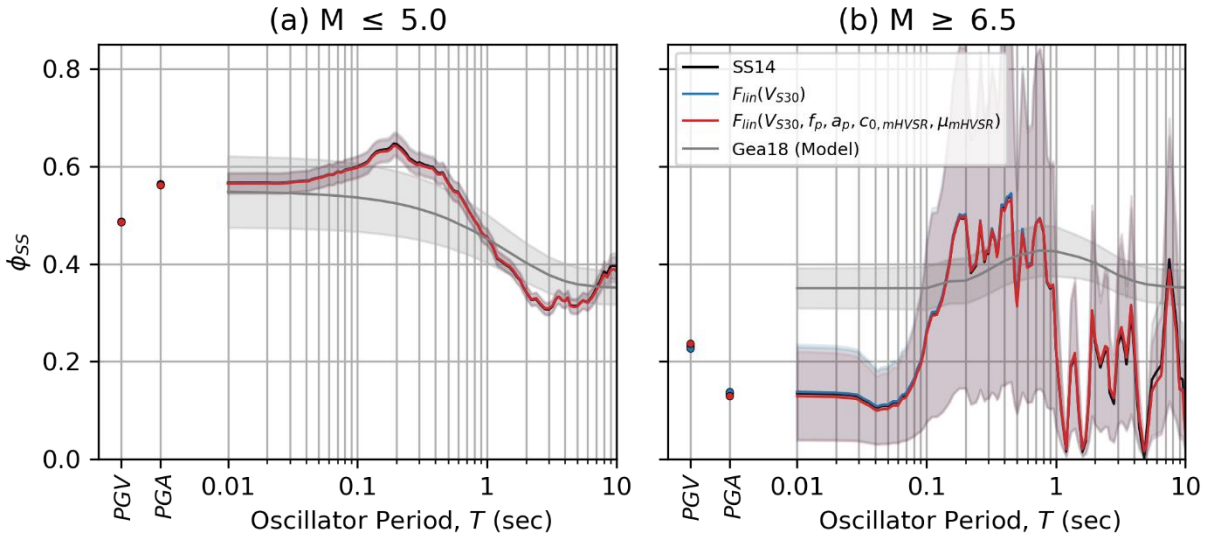


Figure 5.37: Single-station standard deviations (ϕ_{SS}) for the Goulet et al. (2018; Gea18) ϕ_{SS} model and the Delta data calculated using SS14 and the proposed Delta-specific linear site response models for (a) **M** ≤ 5.0 and (b) **M** ≥ 6.5.

Site-to-site variability (ϕ_{S2S}) represents uncertainty in site response modeling. It is null when site response is non-ergodic, but is non-zero for ergodic, regional, or subregional models. A reference global model for ϕ_{S2S} is provided by Al Atik (2015). Her model is based on residuals analyses of the NGA-West2 dataset, which was subsequently refined by Gea18 for BSSA14. ϕ_{S2S} was found to be magnitude-dependent, as shown in Figure 5.38. I compare computed mean values of ϕ_{S2S} and their 95% confidence intervals to the NGA-West2 results reported by Gea18. The subregional ϕ_{S2S} for the Delta are significantly lower than those in the global model (even when using SS14). $F_{lin}(V_{S30})$ (Equation 5.9) does not reduce variability when compared to SS14 for periods shorter than about 1.5 s, however significant improvement is observed at long periods. V_{S30} is the only dependent variable used in both models, however the multilinear form of $F_{lin}(V_{S30})$ is able to better capture the trends observed in site response at long periods. $F_{lin}(V_{S30}, f_p, a_p, c_0, \mu_{mHVSR})$ (Equation 5.30) reduces variability with respect to SS14 and

$F_{lin}(V_{S30})$ for T < 3.0 s, where for longer periods the reduction is identical to that achieved by $F_{lin}(V_{S30})$. These observations are consistent for both small and large **M** (at different levels), and provide evidence warranting a region-specific ϕ_{S2S} model.

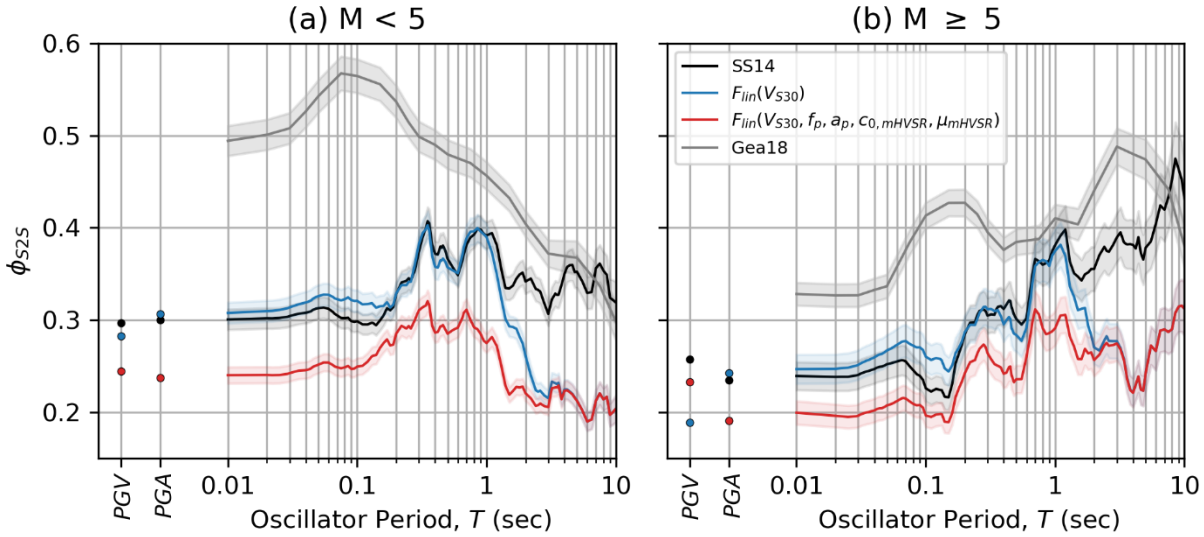


Figure 5.38: Site-to-site standard deviations (ϕ_{S2S}) for global data calculated using BSSA14 (Goulet et al., 2018: Gea18) and the Delta data calculated using SS14 and the proposed Delta-specific linear site response models for (a) **M** < 5 and (b) **M** ≥ 5.

I formulate a model for ϕ_{S2S} that is **M**-dependent for each of the two regionally calibrated site response models, which is consistent with the functional form proposed by Nweke et al. (2022).

$$\phi_{S2S}(\boldsymbol{M}) = \begin{cases} \phi_{S2S,1} & \boldsymbol{M} \leq 5 \\ \sqrt{\phi_{S2S,1}^2 - \Delta Var(\boldsymbol{M} - 5)} & 5 < \boldsymbol{M} < 6 \\ \sqrt{\phi_{S2S,1}^2 - \Delta Var} & \boldsymbol{M} \geq 6 \end{cases} \tag{5.33}$$

where $\phi_{S2S,1}$ is the site-to-site standard deviation for small **M**, and is modeled for each site response model as:

$$\left[\phi_{S2S,1}\right]_{V_{S30}}(T) = \begin{cases} 0.1793\,T + 0.3211 & T < 0.35 \text{ s} \\ 0.3839 & 0.35 \leq T < 0.96 \text{ s} \\ -0.1736\ln(T) + 0.3809 & 0.96 \leq T < 2.54 \text{ s} \\ 0.2151 & 2.54 \text{ s} \leq T \end{cases} \quad (5.34)$$

and

$$\left[\phi_{S2S,1}\right]_{V_{S30},f_p,a_p,c_0,\mu_{mHVSR}}(T) = \begin{cases} 0.2731\,T + 0.2558 & T < 0.27 \text{ s} \\ 0.3144 & 0.27 \leq T < 0.68 \text{ s} \\ -0.0964\ln(T) + 0.2981 & 0.68 \leq T < 1.95 \text{ s} \\ 0.2127 & 1.95 \text{ s} \leq T \end{cases} \quad (5.35)$$

where ΔVar represents the change in variance from small-to-large magnitudes, and is modeled as:

$$\Delta Var(T) = \begin{cases} 0.2012\,T + 0.0489 & T < 0.16 \text{ s} \\ 0.0817 & 0.16 \leq T < 0.57 \text{ s} \\ -0.0750\ln(T) + 0.0399 & 0.57 \leq T < 1.69 \text{ s} \\ 0.0007 & 1.69 \text{ s} \leq T \end{cases} \quad (5.36)$$

The dependencies of ϕ_{S2S} on the underlying site response model are contained within the $\phi_{S2S,1}$ terms, which are plotted in Figure 5.39(a). The models described by Equations (5.34) and (5.35) are shown by the black curves. Computed ΔVar values from both models suggest similar **M**-dependence effects, as shown in Figure 5.39(b). Accordingly, a single smoothed ΔVar model is recommended, which is described by Equation (5.36). The patterns of both models suggest that variability is the greatest at intermediate periods, which coincides with the range of fundamental site periods at Delta sites. ΔVar in the Delta is less than what Nweke et al. (2022) found in southern California.

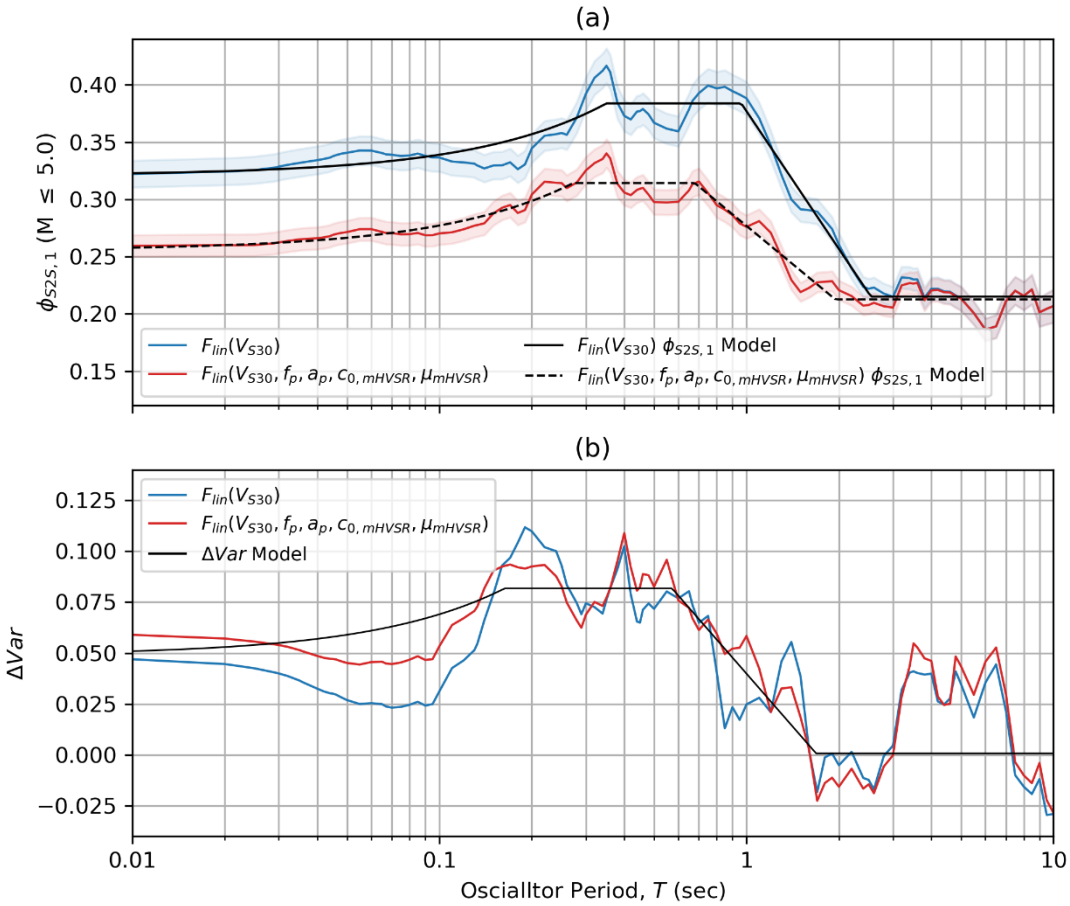


Figure 5.39: (a) Site-to-site standard deviations (ϕ_{S2S}) for Delta data with **M** ≤ 5.0 calculated using the proposed Delta-specific linear site response models, with their recommended $\phi_{S2S,1}$ models; and (b) ΔVar results and recommended model.

5.5. Conclusions

Site conditions in the Delta region of California are often much softer than the recommended range from global ergodic site response models, such as SS14. Not surprisingly, extrapolation of global models to these softer site conditions leads to strongly biased predictions and is subject to large epistemic uncertainty. In this chapter, a subregional model specific to Delta sites and immediately surrounding regions is developed to facilitate more reliable ground motion predictions in this critical region.

Nonergodic site response analyses methods are applied using ground motion data recorded at 36 seismic stations in the Delta region. These site responses comprise the "observations" (or data) from which regionally-calibrated linear site response models are developed. The range of V_{S30} in the study region is 100 – 390 m/s, with 22 sites (~61%) having peaty-organic deposits with peat thicknesses ranging from 0.4 m to 10.1 m.

I demonstrated that V_{S30}-scaling effects in the Delta and surrounding region are different than what is observed globally (i.e., SS14). Most notably, levels of amplification are lower for short periods, higher at long periods, and V_{S30}-scaling saturates for soft site conditions (V_{S30} < 150-200 m/s). The proposed V_{S30}-scaling model (Equation 5.9) can be used for any site conditions encountered in the Delta, and is a standalone model. We have not extended the V_{S30}-scaling model to incorporate sediment depth effects because the available basin depth model provides uniform depth estimates across the study region, and therefore cannot distinguish different site conditions.

After removing first-order site effects (i.e., V_{S30}-scaling), peak features related to site resonance effects were observed in the remaining site response for many sites (about 47%). An automated algorithm was presented to objectively identify the presence of these features for modeling purposes. The correlation between the presence of peaks in mHVSR and site response is good but arguably insufficient (only about 70% of sites with mHVSR peaks also have site response peaks). To improve predictions of site response peaks, a logistic function conditioned on mHVSR peak attributes (f_p, a_p, and c_0) was developed, which improved site response peak identification (91% success rate). A hybrid Riker wavelet and Gaussian pulse parametric model was then formulated to predict attributes of site response peaks given the peak probability and mHVSR peak attributes (Equations 5.18 – 5.20). These models are additive to the V_{S30}-scaling model.

Third-order site effects (i.e., general levels of amplification that remain after V_{S30}-scaling and mHVSR peak effects have been accounted for) were found to correlate with the average mHVSR amplitudes computed over the frequency range from 0.33 to 50 Hz (μ_{mHVSR}). A site response adjustment factor that is constant at short periods and smoothly transitions to zero at periods longer than about 3 s is used to model this effect (Equation 5.26). The complete mHVSR-informed contribution for predicting site amplification is the sum of the peak and constant models (Equation 5.28). Equation (5.28) can only be used when mHVSR is available at a site, and is additive to the V_{S30}-scaling model (Equation 5.9).

The regionally-calibrated V_{S30}-scaling model and mHVSR-informed variant were shown to significantly reduce bias when compared to the predictions provided by SS14. The V_{S30}-scaling model does not appreciably change the aleatory variability (ϕ_{S2S}) for periods shorter than about 1.5 s, however significant reductions around the order of 0.1 (natural log units) are observed at long periods. When the mHVSR-informed model components are used, ϕ_{S2S} is reduced by about 0.05 to 0.1 (natural log units) for short-to-intermediate periods.

Aleatory variability models for ϕ_{S2S} were proposed that should be used in conjunction with the appropriate site amplification model (V_{S30}-scaling only or mHVSR-informed variant). These models can be used with existing models from literature for ϕ_{SS} (Goulet et al. 2018) and τ (BSSA14).

The subregional models presented in this chapter are intended to predict linear site response in the Delta region of northern California. These models were developed using weak ground motions. Nonlinearity is expected for hazard level analyses, and is expected to have non-insignificant impacts on site amplifications in the Delta. Ongoing work is investigating nonlinear

effects for the purpose of developing a comprehensive regional site response model for forward application.

The models were developed using data from 36 seismic stations with V_{S30} between approximately 100 to 390 m/s, peat thicknesses (t_p) between 0.4 to 10.1 m, and mHVSR peak frequencies (f_p) between approximately 0.6 to 4.0 Hz. Bias could be expected for sites in the study region if they possess site characteristics significantly different than those used during model development. The Delta-specific subregional models should not be considered as applicable to other soft-soil regions without proper validation. However, the modeling approach outlined herein can be applied to other regions with unusual geologic conditions that may substantially impact site response.

CHAPTER 6

Conclusions and Recommendations

6.1. Conclusions

This dissertation presents a comprehensive description of procedures that can be applied to investigate regional source, path, and site response using empirical data. The collection, organization, and management of ground motion data is presented in Chapter 2. The geotechnical data and analysis methods used to assign site parameters for use as inputs in site response models is discussed in Chapter 3. Chapter 4 focuses on correcting for regional source and subregional anelastic path effects, in effect removing source- and path-bias from site terms for use in non-ergodic methods to facilitate the investigation of regional site response in Chapter 5. Although the regionally-calibrated site response models presented in Chapter 5 were the primary research objective, several insights and conclusions were reached during each phase of the research.

The relational ground motion database (GMDB), which was discussed in Chapter 2, provides a much needed augmentation of ground motion data to the California subset of the NGA-West2 database. The GMDB is envisioned to grow as additional ground motion data become available. This increase of available data has the potential to greatly impact the earthquake engineering and seismology communities of California, as more and more researchers and practitioners utilize the data. The GMDB schema has been set up to be able to store ground motion

data from any tectonic regime, which is a novel framework. The relational structure provides several advantages over traditional flatfiles:

1. Tables in relational databases reduce the potential for inconsistent metadata since metadata are entered only once.
2. As the amount of data grows, flatfiles become rather unwieldly whereas the serviceability and performance of a relational structure is able to accommodate growth.
3. Database queries are generally significantly faster than file input/output operations because databases store much of their data in random access memory (RAM) rather than on disk storage.
4. Specific fields are easily extracted in a relational structure, while the entire flatfile must be read to extract specific fields.
5. Flatfiles are generally published as *versions*, in which the burden falls onto the user to ensure they have the most up-to-date data. In contrast, data are made publically accessible through the act of uploading to the GMDB, so users can assume they have the most up-to-date data if they include a "get request" through the API in their codes (discussed in Section 2.4.3).

Pertaining to item 5 above, flatfiles do have certain advantages over dynamic databases. Specifically, a snapshot of the database at a specific time may be stores in a flatfile such that a particular study is reproducible. For this reason, the GMDB is configures to output a flatfile upon user request through an application programming interface.

In addition to organizational, management, and dissemination advantages summarized above, the effort of creating the GMDB resulted in the development of screening protocols which helped to

resolve metadata inconsistencies between separate data collection efforts. These protocols can be easily adapted for other applications.

The site characterization data discussed in Chapter 3 was used to assign site parameters for Delta ground motion recording sites co-located with measurements, and also facilitated the development of predictive models for site parameters when site velocity profile data is not available. In general, peat is widely encountered across the Delta, however usually with relatively modest thickness. A contributing factor to the relatively thin deposits of peat is subsidence, for which I developed an adjustment to the peat thickness (t_p) raster published by Deverel and Leighton (2010) (Equation 3.1). Peat thickness was found to have strong correlation to V_{S30} and the presence of peak features in mHVSR. I presented a t_p-based V_{S30} proxy model for the Delta (Equation 3.5), where V_{S30} is modeled to be inversely proportional to t_p. Using the mHVSR peak fitting algorithm proposed by Wang et al. (2022), I found that mHVSR peak features manifest at approximately 70% of Delta sites, which is significantly more often than what is observed across California as a whole.

In Chapter 4, I proposed a subregional anelastic path model which uses nine physiographical subregions for California derived from prior work (Hauksson and Shearer 2006; Chiou et al. 2010; Eberhart-Phillips 2016). I demonstrated that Geyser (induced) events possess different source effects when compared to other north coast tectonic events. I find path effects between induced and tectonic events to not be statistically distinct, therefore my analyses of path effects combined them. I introduce an induced constant adjustment term (e_I) to remove bias related to induced source effects so that they do not inflate between-event variability (τ), which is tabulated in Table 4.1.

I presented an iterative approach that identified subregions with path biases (relative to a California-wide model), adjusts subregional anelastic path coefficients (Δc_3^*; Equation 4.13), and adjusts constant term ($\Delta c_{0,r}$), which are tabulated in Tables 4.2 and 4.3. Other than the adjusted constant and anelastic attenuation terms, other terms were not modified from the Boore et al. (2014; BSSA14) GMM. The aleatory variability was generally unchanged from prior models (BSSA14 for τ and ϕ; Goulet et al. 2018 for ϕ_{SS}), although higher between-event variability was found for North Coast events.

The soft soil site conditions encountered in the Delta were shown to produce site effects that are not properly captured by current site response models (e.g., Seyhan and Stewart 2014; SS14). Nonergodic site response analyses methods were applied using ground motion data to develop estimates of period-dependent linear site response at 36 locations across the Delta and surrounding areas. Based on these observations, a V_{S30}-scaling model (Equation 5.9) was developed. Relative to SS14, the subregional model predicts lower levels of amplification at short periods and higher levels at long periods. Unlike SS14, V_{S30}-scaling saturates for soft site conditions (V_{S30} < 150-200 m/s).

The presence of peak features in mHVSR was found to correlate with the presence of peaks in site response for a majority of Delta sites. A predictive model was developed to estimate the presence of peaks in site response, which consists of a logistic function conditioned on mHVSR peak attributes (Equations 5.18 and 5.19). A Hybrid Ricker wavelet and Gaussian pulse parametric model is used to model site response peaks conditional on mHVSR peak attributes (Equation 5.20). An additional adjustment to site response, to correct for general levels of amplification that remain after V_{S30}-scaling and mHVSR peak effects are accounted for, is provided at short periods and smoothly transitions to zero at periods longer than about 3 s (Equation 5.26). These models can

only be used when mHVSR is available at a site. They are additive to the V_{S30}-scaling model (Equation 5.9), which combine to form the combined model described by Equation (5.30).

The regionally-calibrated site response models were shown to significantly reduce bias and variability when compared to the predictions provided by a global-ergodic model (e.g., SS14). The V_{S30}-scaling model reduces bias over all period ranges and variability at long periods (T > 1.5 s). The mHVSR-informed model further reduces bias for sites likely without site response peaks (P_{peak} < 50%) and sites likely to have peaks (P_{peak} > 50%), and variability at short-to-intermediate periods (T < 1.5 s). Aleatory variability models for ϕ_{S2S} are proposed for the V_{S30}-scaling model and V_{S30}-scaling plus mHVSR-informed models in Equations (5.33) – (5.36). These models are recommended for use with existing models from literature for ϕ_{SS} (e.g., Goulet et al. 2018) and τ (BSSA14).

6.2. Recommendations for Future Research

The research presented in this thesis provides a launch point for a number of additional studies that would benefit ground motion modeling in the Delta subregion and across California. Future work can be divided into three broad categories: (1) work related to advancing the GMDB, (2) work related to empirical ground motion studies which will benefit from the expanded GMDB, and (3) seismic related studies in the Delta.

6.2.1. Updates to the Ground Motion Database (GMDB)

As described in Chapter 2, the GMDB is designed and expected to be a living database, not a static repository of data. The California subset of the GMDB is well populated, however there are a significant number of **M** ≥ 3.0 earthquakes that have occurred since the completion of NGA-West2 data collection efforts that have yet to be assembled at the time of writing this dissertation. With

the advancement of semi-automated ground motion processing tools (e.g., Ramos-Sepulveda et al. 2023), ground motions from these earthquakes can relatively efficiently be processed and reviewed.

Future work is necessary to update the Jupyter Notebook described in Section 2.4.2 to streamline data transfer into the GMDB. In particular, the foreseen necessary updates include:

1. Integration of recent and future schema updates. The current version of the Notebook was developed to be consistent with the initial GMDB schema, however many fields and several tables have since been added to store additional metadata. Continuous updating of the Notebook is necessary to ensure that all GMDB fields get populated.
2. For sites located in California, add the capability to automatically assign site parameters (e.g., V_{S30} from proxy-based methods, geological units, isosurface depths, etc.). The project team includes experts who are knowledgeable of the agreed upon methods used to assign site parameters in California, which are currently performed through external scripts in R and Python. These scripts can be incorporated into the Notebook to assign all site parameters for California sites.
3. Sites located outside of California require consultations with experts familiar in those environments, so fully automated site parameter assignments are not yet realizable in these regions. Once experience is obtained and community consensus is reached, these methods can be integrated into the Notebook.
4. Recent additions to the GMDB schema include distance metrics and derived site parameters included in NGA-West2. Codes need to be written and vetted to compute these values for all recently added ground motions and sites, and for any new data added in the future.

5. Lastly, the functions to compute the RotDXX intensity measures need to be incorporated. The scripts which are currently used to compute these metrics are inefficient and take a substantial amount of time. Work should be done to improve these codes to increase efficiency, so that RotDXX metrics can be computed and uploaded at the time the time-series are uploaded.

Similarly, the API and URL builder tool may require updating as changes are made to the GMDB schema. Additional tools need to be created to allow users to efficiently obtain time-series data and associated metadata for forward applications, and to upload user-defined *collections*. These tools will likely come in the form of additional API *resources* and interactive web portals.

The schema has been meticulously developed to be able to accommodate ground motion data from any tectonic regime (i.e., NGA-West2, NGA-East, and NGA-Sub type data in one unified database). Currently the GMDB only stores ground motion data in California and CENA (i.e., NGA-East). As data is added from other regions and other tectonic regimes (i.e., subduction type events), changes to the schema are likely to be warranted.

Lastly, it is anticipated that future NGA projects (i.e., West3, Sub2, and East2) will use a relational database structure, as done for example with NGA-Sub (Mazzoni et al. 2022). I anticipate that some aspects of the GMDB framework will be adopted and improved upon for NGA-West3 and subsequent NGA projects. Additionally, the work performed by the GMDB team serves as a precursor to the types of issues which should be addressed during these future projects, namely:

1. Translating fields from a flatfile structure into a relational schema, which can be easily navigated by users.

2. Resolving metadata inconsistencies between separate NGA products (e.g., station and site databases).
3. Dissemination as a queryable database in addition to flatfile releases for version control. NGA-Sub tables were released as a series of CSV files which mirror a relational database, but is in principle not a publically queryable database.

The lessons learned and protocols developed by the GMDB project team will prove to be beneficial to future NGA and other ground motion database projects.

6.2.2. Applications of the GMDB

Applications of ground motion data in current and future versions of the GMDB are nearly limitless. However, I focus in this subsection on general applications of ground motion data to the advancement of modeling path effects and site response.

(a) Path Effect Modeling

The modeling framework presented in Chapter 4 is an adaptation of the cell-specific method presented by Dawood and Rodriquez-Marek (2013) to consider physical constraints during discretization, which I refer to as a "subregional method". The results of my work suggest that subregionalization leads to similar results as a model developed using the cell-specific method with significantly more zones (e.g., Kuehn et al. 2019). An obvious follow-up is to perform a rigorous quantitative comparison between Kuehn et al. (2019) and the model presented in Chapter 4, both derived from an identical dataset. I suspect that the Keuhn et al. (2019) model (with updated coefficient values based on the expanded dataset) will perform better than the model presented in Chapter 4. What we should seek over the long term are an appropriate number geology-based subregions that can produce similar performance to that achieved by arbitrary cells. These

subregions must be constrained by some logical means, therefore advanced studies of Q_S would likely prove to be invaluable.

Work should be conducted to determine the level of subregion resolution that can be used to practically quantify non-ergodic path effects. As discussed in Section 4.2.2, a fully non-ergodic path model is a conceptualization that is difficult to verify because path effects represent a coupling of a particular source and site, and cannot be modeled as a feature at a point in space. However, a path model can be assumed to be practically non-ergodic once no additional reduction in aleatory variability is achieved when increasing the complexity of the model. This is a topic which I suspect to be the subject of many investigations in the upcoming years, weather modelers choose to use subregional, cell-specific, or some yet-to-be presented method.

A separate follow-up is to perform similar studies using identical subregions with different base GMMs (e.g., Abrahamson et al. 2014; Campbell and Bozorgnia 2014; Chiou and Youngs 2014; and Idriss 2014). I suspect similar conclusions, however the coefficient estimates will be different. The combination of these models can then be used to perform probabilistic seismic hazard analysis, to assess the impact of non-ergodic path effects on hazard. Hazard levels at most sites are controlled by nearby sources, so I anticipate minor impacts from updating the anelastic path models. However, sites where distant sources contribute non-insignificant levels (e.g., megathrust events for sites in northern California; Figure 6.1) are likely to be affected by updating anelastic path models.

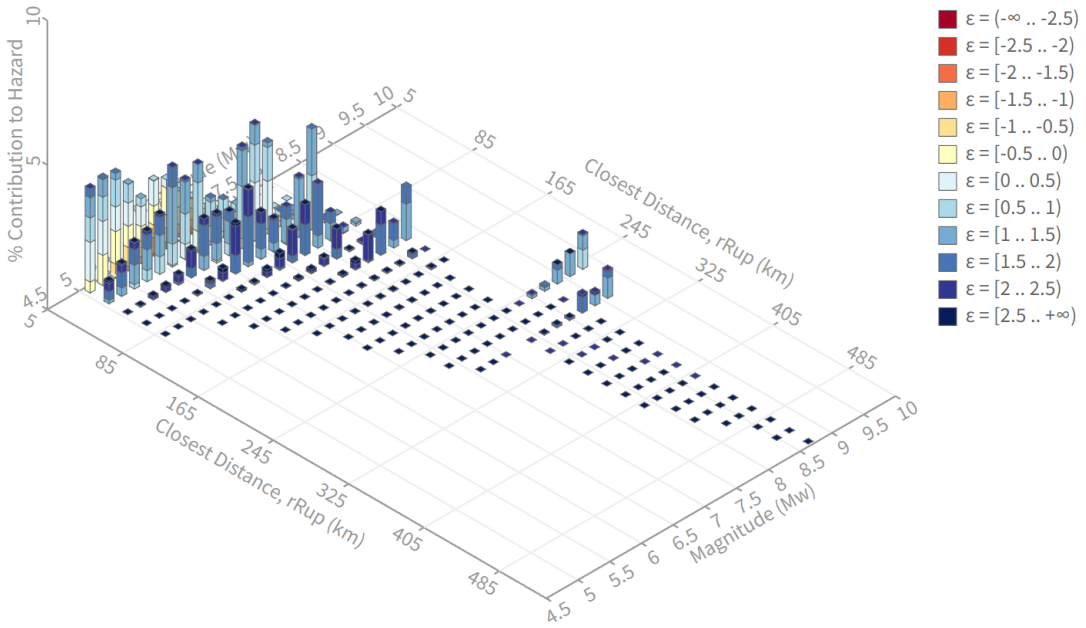


Figure 6.1: Disaggregation results for a site in Sacramento, California and a return period of 475 years. Disaggregation performed using the USGS Unified Hazard Tool (Dynamic: Conterminous U.S. 2014 update – v4.2.0).

One of the primary limitations of non-ergodic path modeling methods (i.e., cell-specific or subregional) is data availability. These types of models can only be developed for regions with a sufficiently dense distribution of data to confidently constrain anelastic coefficients (e.g., California, Japan, New Zealand, etc.). There exist many regions affected by seismic demands, which do not have sufficient instrumentation to develop non-ergodic path models. This is where I think subregional methods can provide some insight over cell-specific counterparts.

Fundamentally, subregionalization is the act of discretizing some domain with consideration of physical attributes. Just as early site response models separated amplification response for generic *soil* and *rock* sites (e.g., Abrahamson and Silva 1997), subregionalization can be performed using generic crustal features (e.g., accretionary, batholiths, pull-apart, etc.). If the anelastic coefficients for these generic subregions are found to exhibit similar characteristics, then an ergodic but subregionalized modeling approach can be pursued. In other words, a generic

subregional path model can be developed and subsequently applied in regions of sparse data. The challenging aspect of this proposal is developing the subregionalization, which will require collaboration of experts across different fields from around the world to develop generic categories which are internally compatible.

My final recommendation for future work related to modeling path effects is to consider the spatial variability of geometric spreading and the depth of wave propagation (i.e., 3D subregionalization). Current path modeling approaches, including the subregional method presented in this dissertation, assume a 2D-straight path between the earthquake and the site, which is an over simplification. Rates of anelastic attenuation are usually less for more competent materials (higher V_S), which are generally found at greater depths. True ray paths will take the fastest route, which will result in deviations from a straight line in 2D and 3D space. It follows that 3D seismic velocity models (e.g., Lee et al. 2014; Shaw et al. 2015; Hirakawa and Aagaard 2021) can be used to help define subregionalization with depth constraints. However, estimation of depth-dependent subregional anelastic coefficients ($\Delta c_{3,r}$) will be more challenging than for 2D coefficients. Seismic ray-tracing methods which define rules about how seismic waves propagate through 3D domains (e.g., Julian and Gubbins 1977; Nakanishi and Yamaguchi 1986; Vidale 1988; Rawlinson et al. 2008) can be adopted by modelers. This style of modeling is likely overly optimistic, however as more and more ground motion data becomes available it becomes increasingly possible.

(b) Site Response Modeling

Nonergodic site response methods which utilize ground motion data are only possible at sites which have recorded many events. Prior to the development of the GMDB, only 508 sites in California had five or more ground motion records available. That number has since grown to

1,292 sites in the GMDB at the time of writing this dissertation, illustrating the impact of recent data collection efforts. The increase of the number of records per site is accompanied by an increase in the range of site conditions [Section 2.3.1(b)]. These two factors combined mean that state-wide and subregional assessments (e.g., Parker and Baltay 2022; Chapter 5) of site-response using nonergodic methods may provide significant improvements over existing ergodic site response models.

In Section 5.3.2, I examined the dependence of site response on isosurface depth (e.g., $z_{1.0}$) for the Central Valley and observed significant basin effects (Figure 5.15), but did not recommend a model for application in the Delta region due to the uniform $z_{1.0}$ estimates provided by USGS SFCVM. A follow-up study with the aim of improving SFCVM (i.e., use of seismic tomography) followed by a rigorous study of basin effects in the Central Valley, would prove to be beneficial. Furthermore, recent work has been done to re-evaluate basin effects in southern California basins (e.g., Nweke et al. 2022), therefore similar studies in northern California (i.e., Bay Area basins) should be conducted.

The application of mHVSR data for predicting site response has received attention in recent years (e.g., Wang et al. 2022a,b; Pinilla-Ramos et al. 2022). The work presented in Section 5.3.4 demonstrates the potential for site parameters derived from mHVSR to improving site response predictions. A promising approach, applied previously by Kwak et al. (2017), Hassani and Atkinson (2018a), and Wang et al. (2022a), is to separate the effects of site-resonant peak features from general levels of amplification from V_{S30}-scaling models. Peak features are observed less frequently across the state as a whole (about 25-35% of sites with data have peak features; Wang et al. 2021), so a state-wide investigation could provide significant insight into the utility of mHVSR on improving site response predictions at sites without prominent mHVSR peak features.

Specifically, investigations into the utility of mHVSR to improve long-period site response predictions (i.e., basin effects), and comparisons to conventional site parameters (e.g., $z_{1.0}$ and $z_{2.5}$), are needed.

6.2.3. Path Forward for Research in the Delta

The site response models presented in Chapter 5 are only applicable to the elastic, or linear, site response. In the absence of strong ground motion records, simulations via ground response analyses (GRA) can be used to assess the effects of nonlinearity. I am currently working with a colleague, Pengfei Wang (Old Dominion University), to perform GRA to develop a large simulated dataset from which a nonlinear model will be developed. The nonlinear model is expected capture typical strain-dependent effects (i.e., soil softening which increases damping and shifts amplification to progressively lower frequencies), in addition to modifications of site-resonance effects (shortening peak frequencies, while widening peaks and decreasing their amplitudes). Following the conclusion of this work, a comprehensive site response model calibrated for the Delta subregion will be complete.

Seismic hazards pose a major risk for critical infrastructure (e.g., levees, gas pipelines) in the Delta (DRMS 2009). The work presented in Chapters 3, 4, and 5 provide significant advancements to the characterization and prediction of seismic ground motions in the Delta region. Updated regional seismic hazard analysis should be performed to re-assess hazard levels and identify the contributing sources using methods outlined by Wang et al (2022c). Furthermore, the subregional anelastic path (Equation 4.11) and regionally-calibrated site response models (Equations 5.9 and 5.30) can be used to analyze the system reliability of Delta levees using the framework proposed by Zimmaro et al. (2019). The results of these types of studies will likely

have high-impact, given the potential state-wide repercussions if significant damage to the vital Delta infrastructure were to occur.

REFERENCES

Aagaard, BT, RW Graymer, CH Thurber, AJ Rodgers, T Taira, RD Catchings, CA Goulet, and A Plesch. (2020). Science plan for improving three-dimensional seismic velocity models in the San Francisco Bay region, 2019–24, *U.S. Geol. Surv. Open-File Rept. 2020-1019*, 37. doi: 10.3133/ofr20201019.

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

Abrahamson, NA and WJ Silva. (1997). Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seismological Research Letters*, 68 (1), 94-127. doi: 10.1785/gssrl.68.1.94.

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

Abrahamson, NA, and RR Youngs, (1992). A stable algorithm for regression analyses using the random effects model, *Bull. Seismol. Soc. Am*. 82 (1), 505–510. doi: 10.1785/bssa0820010505.

Ahdi, SK, TD Ancheta, V Contreras, T Kishida, DY Kwak, AO Kwok, GA Parker, Y Bozorgnia, and JP Stewart. (2017). NGA-Subduction site database. *Proc. 16th World Conference on Earthquake Engineering*, Paper No. 4926, Santiago, Chile.

Ahdi, SK, DY Kwak, TD Ancheta, V Contreras, T Kishida, AOL Kwok, S Mazzoni, F Ruz, JP Stewart. (2022). Site parameters applied in NGA-Sub database, *Earthq. Spectra*, 38 (1), 494-520. doi: 10.1177/87552930211043536.

Ahdi SK, Mazzoni S, Kishida T, Wang P, Nweke CC, Kuehn NM, Contreras V, Rowshandel B, Stewart JP, and Bozorgina Y. (2020). Engineering characteristics of ground motions recorded in the 2019 Ridgecrest Earthquake Sequence, *Bull. Seismol. Soc. Am.*, 110 (4), 1474-1494. doi: 10.1785/0120200036.

Ahdi, SK, S Sadiq, O Ilhan, Y Bozorgnia, YMA Hashash, DY Kwak, D Park, A Yong, and JP Stewart. (2018). Development of a United States Community Shear Wave Velocity Profile Database. *Geotech. Eng. and Soil Dynamics V*, Austin TX, 330-339. doi: 10.1061/9780784481462.032.

Ahdi, SK, JP Stewart, TD Ancheta, DY Kwak, D Mitra. (2017). Development of V_S profile database and proxy-based models for V_{S30} prediction in the Pacific Northwest region of North America, *Bull. Seism. Soc. Am.*, 107 (4), 1781-1801. doi: 10.1785/0120160335.

Al Atik, L and N Abrahamson. (2010). Nonlinear site response effects on the standard deviations of predicted ground motions. *Bull. Seismol. Soc. Am.* 100 (3), 1288-1292. doi: 10.1785/0120090154.

Ancheta, TD, RB Darragh, JP Stewart, E Seyhan, WJ Silva, BS-J Chiou, KE Wooddell, RW Graves, AR Kottke, DM Boore, T Kishida, and JL Donahue. (2013). PEER NGA-West2 Database, PEER Report No. 2013/03, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, 134 pp.

Ancheta, TD, RB Darragh, JP Stewart, E Seyhan, WJ Silva, BS-J Chiou, KE Wooddell, RW Graves, AR Kottke, DM Boore, T Kishida, and JL Donahue. (2014). NGA-West2 database, *Earthq. Spectra*, 30 (3), 989 – 1005. doi: 10.1193/070913EQS197M.

Anderson, JG, and JN Brune. (1999). Probabilistic seismic hazard analysis without the ergodic assumption. *Seismological Research Letters*, 70 (1), 19-28. doi: 10.1785/gssrl.70.1.19.

Arias, A. (1970). A measure of earthquake intensity, in Hansen, R., ed., Seismic design for nuclear power plants: Cambridge, Mass., Massachusetts Institute of Technology Press, p. 438-483.

ASTM Standard D1586/D1586M-18. (2018). Standard test method for standard penetration test (SPT) and split-barrel sampling of soils. ASTM International, West Conshohocken, PA. doi: 10.1520/D1586_D1586M-18E01.

ASTM Standard D3441. (2016). Standard test method for mechanical cone penetration testing of soils. ASTM International, West Conshohocken, PA. doi: 10.1520/D3441-16.

ASTM Standard D4428/D4428M-14. (2014). Standard test methods for crosshole seismic testing. ASTM International, West Conshohocken, PA. doi: 0.1520/D4428_D4428M-14.

ASTM Standard D5753. (2018). Standard guide for planning and conducting geotechnical borehole geophysical logging. ASTM International, West Conshohocken, PA. doi: 10.1520/D5753-18.

ASTM Standard D5777. (2018). Standard guide for using the seismic refraction method for subsurface investigation. ASTM International, West Conshohocken, PA. doi: 10.1520/D5777-18.

ASTM Standard D7400/D7400M-19. (2019). Standard methods for downhole seismic testing. ASTM International, West Conshohocken, PA. doi: 10.1520/D7400_D7400M-19.

Atkinson, GM, and M Morrison. (2009). Observations on regional variability in ground-motion amplitudes for small-to-moderate earthquakes in North America, *Bull. Seismol. Soc. Am.* 99 (4), 2393–2409. doi: 10.1785/0120080223.

Atwater BF. (1982). Geologic maps of the Sacramento–San Joaquin Delta. U.S. Geological Survey Miscellaneous Field Studies Map MF–1401. doi: 10.3133/mf1401.

Babyak, MA. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models, *Psychosomatic Medicine*, 66 (3), 411-421. doi: 10.1097/01.psy.0000127692.23278.a9.

Bates D, Mächler M, Bolker B, Walker S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67 (1), 1–48. doi: 10.18637/jss.v067.i01.

Bay Delta Live. Last access June (2021). https://www.baydeltalive.com/maps/20803.

Benson, RC and LB Yuhr. (2016). What is site characterization. In: Site characterization in Karst and Pseudokarst Terraines. Springer, Dordrecht. doi: 10.1007/978-94-017-9924-9_11.

Biek, RF, PD Rowley, JM Hayden, DB Hacker, GC Wills, LR Hintze, RE Anderson, and KD Brown. (2010). Geologic map of the St. George and East part of the Clover mountains 30' x 60' quadrangles, Washington and Iron counties, Utah. Utah Geological Survey Open-File Report 242DM, Utah Geological Survey.

Boore, DM. (2003). Simulation of ground motion using the stochastic method, *Pure and Applied Geophysics* 160, 635-675.

Boore DM. (2004). Estimating $V_S(30)$ (or NEHRP site class) from shallow velocity models (depths < 30 m), *Bull. Seismol. Soc. Am.*, 94 (2), 591-597. doi: 10.1785/0120030105.

Boore, DM. (2010). Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. *Bull. Seismol. Soc. of Am*. 100 (4), 1830-1835. doi: 10.1785/0120090400.

Boore, DM and GM Atkinson. (2008). Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s, *Earthq. Spectra*. 24 (1), 99–138. doi: 10.1193/1.2830434.

Boore, DM, JP Stewart, E Seyhan, and GM Atkinson. (2014). NGA-West2 equations for predicting PGA, PGV, and 5%-Damped PSA for Shallow Crustal Earthquakes. *Bull. Seismol. Soc. Am.*, *Earthq Spectra*, 30 (3), 1057-1085. doi: 10.1193/070113EQS184M.

Boore, DM, J Watson-Lamprey, and NA Abrahamson. (2006). Orientation-independent measures of ground motion, *Bull. Seism. Soc. Am.* 96 (4A), 1502-1511. doi: 10.1785/0120050209.

Borcherdt, RD. (1994). Estimates of Site-Dependent Response Spectra for Design (Methodology and Justification). *The International Journal of Robotics Research*. 10 (4), 224-235. doi: 10.1177/027836492092747 6.

Borcherdt, RD and JF Gibbs. (1976). Effects of local geological conditions in the San Francisco Bay region on ground motions and the intensities of the 1906 earthquake. *Bull. Seismol. Soc. Am.*, 66 (2), 467-500. doi: 10.1785/BSSA0660020467.

Borcherdt, RD, and G Glassmoyer. (1994). Influences of local geology on strong and weak ground motions recorded in the San Francisco Bay region and their implications for site specific building-code provisions, *The Loma Prieta, California Earthquake of October 17, 198*9 *Strong Ground Motion*, U. S. Geological Survey Professional Paper 1551-A, A77-A108.

Bozorgnia, Y, NA Abrahamson, SK Ahdi, TD Ancheta, L Al Atik, RJ Archuleta, GM Atkinson, DM Boore, KW Campbell, BSJ Chiou, V Contreras, RB Darragh, S Derakhshan, JL Donahue, N Gregor, Z Gulerce, IM Idriss, C Ji, T Kishida, AR Kottke, N Kuehn, DY Kwak, AOL Kwok, P Lin, J Macedo, S Mazzoni, S Midorikawa, S Muin, GA Parker, S Rezaeian, H Si, WJ Silva, JP Stewart, M Walling, K Wooddell, RR Youngs. (2022). NGA-Subduction research program, *Earthq. Spectra*, 38 (2), 783-798. doi: 10.1177/87552930211056081.

Bozorgnia, Y, NA Abrahamson, L Al Atik, TD Ancheta, GM Atkinson, JW Baker, A Baltay, DM Boore, KW Campbell, BS-J Chiou, R Darragh, S Day, J Donahue, RW Graves, N Gregor, T Hanks, IM Idriss, R Kamai, T Kishida, A Kottke, SA Mahin, S Rezaeian, B Rowshandel, E Seyhan, S Shahi, T Shantz, W Silva, P Spudich, JP Stewart, J Watson-Lamprey, K Wooddell, and R Youngs. (2014). NGA-West2 research project, *Earthq. Spectra*, 30 (3), 973-987. doi: 10.1193/072113EQS209M.

Breiman L, JH Friedman, RA Olshen, and CJ Stone. (1984). Classification and Regression Trees. Wadsworth. doi: 10.1201/9781315139470.

Brocher, TM. (2005a). Compressional and shear wave velocity versus depth in the San Francisco Bay area, California: Rules for USGS Bay area velocity model 05.0.0, *U.S. Geol. Surv. Open File Rept. 2005-1317*.

Brocher, TM. (2005b). Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. Seismol. Soc. Am*., 95 (6), 2081–2092. doi: 10.1785/0120050077.

Brocher, T. (2006). The new USGS 3D seismic velocity model for northern California, *Seismol. Res. Lett*., 77, 271.

Buckreis, TE, A Winders, P Wang, SJ Brandenberg, and JP Stewart. (2021). Microtremor data collected in sacramento-san joaquin delta region of california. doi: 10.17603/ds2-dk6t-8610. (DesignSafe-CI).

Buckreis, TE, JP Stewart, SJ Brandenberg, P Wang, and C Nweke. (2022). Data Files for California Subregional Path Effect Study. doi: 10.17603/ds2-mbee-8941. (DesignSafe-CI).

California, (1992). *Seismic Stability Evaluation of the Sacramento-San Joaquin Delta Levees*, Phase I Report: Preliminary Evaluations and Review of Previous Studies, Division of Design and Construction, California Department of Water Resources, 203 pp.

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

Castellaro, S, and F Mulargia. (2009). VS30 estimates using constrained H/V measurements, *Bull. Seismol. Soc. Am.* 99 (2A), 761–773. doi: 10.1785/0120080179.

CGIAR-CSI. (2006). NASA Shuttle Radar Topographic Mission (SRTM). The SRTM data is available as 3 arc second (approx. 90m resolution) DEMs. The dataset is available for download at: http://srtm.csi.cgiar.org/

Chatelain, JL, B Guillier, BF Cara, AM Duval, K Atakan, PY Bard, and The WP02 SESAME team (2008). Evaluation of the influence of experimental conditions on H/V results from ambient noise recordings, *Bull. Earthq. Eng*. 6, 33–74, doi: 10.1007/s10518- 007-9040-7.

Chen, P, and EJ Lee. (2017). UCVM 17.3.0 Documentation. Retrieved from http://hypocenter.usc.edu

Cheng, T, B Cox, JP Vantassel, and L Manuel. (2020). A statistical approach to account for azimuthal variability in single-station HVSR measurements, *Geophys. J. Int.* 223 (2), 1040–1053, doi: 10.1093/gji/ggaa342.

Chiou, B, R Darragh, N Gregor, and W Silva. (2008). NGA project strong-motion database. *Earthq. Spectra*, 24 (1), 23 – 44. doi: 10.1193/1.2894831.

Chiou, B and RR Youngs. (2008). An NGA model for the average horizontal component of peak ground motion and response spectra, *Earthq. Spectra*, 24 (1), 173–215, doi: 10.1193/1.2894832.

Chiou, B-SJ and RR Youngs. (2013). Update of the Chiou and Youngs NGA ground motion model for average horizontal component of peak ground motion and response spectra, PEER Report 2013/07. Pacific Earthquake Engineering Research Center, UC Berkeley (headquarters).

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

Chiou, B, R Youngs, N Abrahamson, K Addo. (2010). Ground motion attenuation model for small-to-moderate Shallow Crustal Earthquakes in California and its implications on regionalization of ground-motion prediction models. *Earthq. Spectra*. 26 (4), 907-926. doi: 10.1193/1.3479930.

Clayton, R, T Heaton, M Chandy, A Krause, M Kohler, J Bunn, R Guy, M Olson, M Faulkner, MH Cheng, L Strand, R Chandy, D Obenshain, A Liu, and M Aivazis. (2011). Community Seismic Network, *Annals of Geophysics*, 54 (6), 738-747. doi: 10.4401/ag-5269.

Cleveland, WS. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74 (368), 829-836. doi: 10.2307/2286407.

Contreras V, JP Stewart JP, T Kishida, RB Darragh, BSJ Chiou, S Mazzoni, N Kuehn, SK Ahdi, K Wooddell, RR Youngs, Y Bozorgnia, R Boroschek, F Rojas, and J Órdenes. (2020). Chapter 4: Source and Path Metadata, in Data Resources for NGA-Subduction Project, PEER Report 2020/02, J.P. Stewart (editor), Pacific Earthquake Engineering Research Center, UC Berkeley (headquarters).

Contreras, V, JP Stewart, T Kishida, RB Darragh, BSJ Chiou, S Mazzoni, RR Youngs, NM Kuehn, SK Ahdi, K Wooddell, R Boroschek, F Rojas, and J Ordenes. (2022). NGA-Sub source and path database. *Earthq. Spectra*, 38 (2), 799–840. doi: 10.1177/87552930211065054.

Davis-Foster, D. (2022). Pypi-json 0.3.0.

Dawood, HM, A Rodriguez-Marek. (2013). A method for including path effects in ground-motion prediction equations: An example using the Mw 9.0 Tohoku Earthquake aftershocks. *Bull. Seismol. Soc. Am.* 103 (2B), 1360–1372. doi: 10.1785/0120120125.

Day, SM., R Graves, J Bielak, D Dreger, S Larsen, KB Olsen, A Pitarka, and L Ramirez-Guzman. (2008). Model for basin effects on long-period response spectra in southern California. *Earthq. Spectra*, 24 (1), 257-277. doi: 10.1193/1.2857545.

Department of Water Resources (DWR). (2012). "Delta habitat conservation program, Delta seismicity study; Results of deep seismic program well installations." Project Geology report No. 80-10-36., April 2012.

Department of Water Resources (DWR). (2013a). "Geologic report: Curtis Landing Fish Release site modification." Project Geology report No. 80-10-38., January 2013.

Department of Water Resources (DWR). (2013b). "2009 through 2012 geotechnical data report, pipeline/tunnel option." Delta Habitat and Conveyance Program. April 2013.

Department of Water Resources (DWR). (2014a). "Delta seismic stability study, Deep Hole Program; Final results of the 1994 exploration program at the Clifton Court site." Project Geology report No. 80-10-07, April 2014.

Department of Water Resources (DWR). (2014b). "Delta seismic stability study, Deep Hole Program; Final results of the 1994 exploration program at the Montezuma Slough Control Structure site." Project Geology report No. 80-10-08, April 2014.

Department of Water Resources (DWR). (2014c). "Delta seismic stability study, Deep Hole Program; Final results of the 1994 exploration program at the Sherman Island site." Project Geology report No. 80-10-09, April 2014.

Department of Water Resources (DWR). (2014d). "Delta seismic stability study, Deep Hole Program; Final results of the 1994 exploration program at the Staten Island site." Project Geology report No. 80-10-10, April 2014.

Department of Water Resources (DWR). (2014e). "Sherman Island new salvaged fish release sites exploration geology report." Project Geology report No. 80-10-39., January 2014.

Department of Water Resources (DWR). (2015a). "Clifton Court Forebay Dam embankment foundation, three-dimensional stratigraphic model." Project Geology report No. 51-10-25. April 2015.

Department of Water Resources (DWR). (2015b). "Geologic data report: Delta flood emergency preparedness, response, and recovery program, Rio Vista and West Weber South facilities." Project Geology report No. 80-10-42, December 2015.

Department of Water Resources (DWR). (2016). "West False River temporary drought barrier geologic exploration report." Project Geology report No. 80-10-46, October 2016.

Department of Water Resources (DWR). (2017). "Geologic data report: Delta flood emergency preparedness, response, and recovery program, West Weber Avenue North facility, Stockton, California." Project Geology report No. 80-10-44, February 2017.

Department of Water Resources (DWR). (2018). *i03 LegalDeltaBoundary*. 1:24,000, California Natural Resources Spatial Data. The dataset is available for download at: https://gis.data.cnra.ca.gov/datasets/57b02f8a5e77465f902376dbd9522585_0/

Department of Water Resources (DWR), (2022). Delta conveyance and flood protection, https://water.ca.gov/Programs/Flood-Management/Delta-Conveyance-And-Flood-Protection.

Deverel, SJ, S Bachand, SJ Brandenberg, CE Jones, JP Stewart, and P Zimmaro. (2016). Factors and processes affecting Delta levee system vulnerability, *San Francisco Estuary and Watershed Science*, 14 (4). doi: 10.15447/sfews.2016v14iss4art3.

Deverel, SJ and DA Leighton. (2010). Historic, recent and future subsidence, Sacramento-San Joaquin Delta, California, USA, *San Francisco Estuary and Watershed Science*, 8 (2). doi: 10.15447/sfews.2010v8iss2art1.

Dobry, R, I Oweis, and A Urzúa. (1976). Simplified procedures for estimating the fundamental period of a soil profile, *Bull. Seismol. Soc. Am.*, 66 (4), 1293-1321. doi: 10.1785/BSSA0660041293.

Dobry R, RD Borcherdt, CB Crouse, IM Idriss, WB Joyner, GR Martin, MS Power, EE Rinne, and RB Seed. (2000). New Site Coefficients and Site Classification System Used in Recent Building Seismic Code Provisions. *Earthq. Spectra*. 16 (1), 41-67. doi: 10.1193/1.1586082.

Drexler, JZ. (2011). Peat formation processes through the millennia in tidal marshes of the Sacramento–San Joaquin Delta, California, USA. *Estuaries and Coasts*. 34, 900-911. doi: 10.1007/s12237-011-9393-7.

Drexler, J.Z., C.S. DeFontain, and D.L. Knifong. (2007). Age Determination of the Remaining Peat in the Sacramento-San Joaquin Delta, California, USA, *U.S. Geol. Survey Open-File Rept. 2007-1303*.

[DRMS] California Department of Water Resources. (2009). Executive summary. Delta Risk Management Strategy.

Eberhart-Phillips, D. (2016). Northern California seismic attenuation: 3D Q_P and Q_S models. *Bull. Seismol. Soc. Am.* 106 (6), 2558-2573. doi:10.1785/0120160060.

Ekström G, M Nettles M, and AM Dziewonski. (2012). The global CMT project 2004-2010: Centroidmoment tensors for 13,017 earthquakes, *Phys. Earth Planet. Inter*., 200-201, 1-9. doi: 10.1016/j.pepi.2012.04.002.

Electrical Power Research Institute (EPRI). (1988). A Criterion for Determining Exceedance of the Operating Basis Earthquake, EPRI NP-5930, Palo Alto, CA.

Erdem, JE, J Boatwright, and JB Fltecher. (2019). Ground-motion attenuation in the Sacramento-San Joaquin Delta, California, from 14 Bay Area earthquakes, including the 2014 M 6.0 South Napa earthquake. *Bull. Seis. Soc. Of Am.*, 109 (3), 1025-1033. doi: 10.1785/0120180182.

Galloway, M., D.R. Jones, and S.E. Ingebritsen. (1999). Land Subsidence in the United States: Part II Drainage of Organic Soils. 79-94.

Gardner, JK and L Knopoff. (1974). Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian? *Bull. Seismol. Soc. Am.*, 64 (5), 1363–1367. doi: 10.1785/BSSA0640051363.

Gautier, L. (2009). Rpy2: Interface to use R from Python. GitHub repository. https://github.com/rpy2/rpy2.

Gelman, A, JB Carlin, HS Stern, DB Dunson, A Vehtari, and DB Rubin. (2014). Bayesian data analysis, 3rd edition, CRC Press.

Ghofrani, H and GM Atkinson. (2014). Site condition evaluation using horizontal-to-vertical response spectral ratios of earthquakes in the NGA-West 2 and Japanese databases, *Soil Dyn. Earthq. Eng.*, 67, 30-43. doi: 10.1016/j.soildyn.2014.08.015,

Ghofrani, H, GM Atkinson, and K Goda. (2013). Implication of the 2011 M9.0 Tohoku Japan earthquake for the treatment of site effects in large earthquakes, *Bull. Earthq. Eng.*, 11 (1), 171-203. doi: 10.1007/s10518-012-9413-4.

Goulet, C, Bozorgnia, Y, Abrahamson, N, Kuehn, NM, Al Atik, L, Youngs, R, Graves, RW, Atkinson, G. (2018). Next generation attenuation relations for the central and eastern United States (NGA-East) (No. DOE-UC-08314). PEER report no. 2018/08. Berkeley, CA: Pacific Earthquake Engineering Research Center (PEER), University of California, Berkeley, December.

Goulet CA, Bozorgnia Y, Kuehn N, Al Atik L, Youngs RR, Graves RW, and Atkinson GM. (2021a). NGA-East Ground-Motion Characterization model part I: Summary of products and model development *Earthq. Spectra*, 37 (1), 1231-1282. doi: 10.1177/87552930211018723.

Goulet CA, Kishida T, Ancheta TD, Cramer CH, Darragh RB, Silva WJ, Hashash YMA, Harmon J, Parker GA, Stewart JP, and Youngs RR. (2021b). PEER NGA-East database, *Earthq. Spectra*, 37 (1), 1331-1353. doi: 10.1177/8755293021101569.

Graves, RW. (1993). Modeling three-dimensional site response effects in the Marina District, San Francisco, California, *Bull. Seism. Soc. Am.* 83 (4), 1042–1063. doi: 10.1785/BSSA0830041042.

Hardebeck, J L, AJ Michael, and TM Brocher. (2007). Seismic velocity structure and seismotectonics of the eastern San Francisco Bay region, California, *Bull. Seismol. Soc. Am.*, 97 (3), 826–842. doi: 10.1785/0120060032.

Harder, LF and HB Seed. (1986). Determination of penetration resistance for coarse-grained soils using the Becker Hammer Drill. College of Engineering, University of California, Berkeley, Calif. Rep. No. UCB/EERC-86/06. May 1986.

Harmon J, YMA Hashash, JP Stewart, EM Rathje, KW Campbell, WJ Silva, and O Ilhan. (2019). Site amplification functions for central and eastern North America—Part II: Modular simulation based models. *Earthq. Spectra*, 35 (2), 815–847. doi: 10.1193/091117EQS179M.

Harris, CR, KJ Millman, SJ van der Walt, et al. (2020). Array programming with NumPy. *Nature*. 585, 357-362. doi: 10.1038/s41586-020-2649-2.

Hashash, YMA, O Ilhan, JA Harmon, GA Parker, JP Stewart, EM Rathje, KW Campbell, and WJ Silva. (2020). Nonlinear site amplification model for ergodic seismic hazard analysis in Central and Eastern North America. *Earthq. Spectra*, 36 (1), 69-86. doi: 10.1177/8755293019878193.

Hassani, B and GM Atkinson. (2016). Applicability of the site fundamental frequency as a V_{S30} proxy for Central and Eastern North America, *Bull. Seismol. Soc. Am.*, 106 (2), 653-664. doi: 10.1785/0120150259.

Hassani, B and GM Atikinson. (2018a). Application of site-effects model based on peak frequency and average shear-wave velocity to California. *Bull. Seismol. Soc. Am.*, 108 (1), 351-357. doi: 10.1785/0120170062.

Hassani, B and GM Atkinson. (2018b). Site-effects model for central and eastern North America based on peak frequency and average shear-wave velocity. *Bull. Seismol. Soc. Am.*, 108, 338-350.

Hauksson, E, and PM Shearer. (2006). Attenuation models (Q_P and Q_S) in three dimensions of the southern California crust: Inferred fluid saturation at seismogenic depths, *J. Geophys. Res*., 111, B05302, doi: 10.1029/2005JB003947.

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

Héloïse, C, PY Bard, AM Duval, and E Bertrand. (2012). Site effect assessment using KiK-net data: Part 2 – Site amplification prediction equation based on f0 and Vsz, *Bull. Earthq. Eng.*, 10, 451-489. doi: 10.1007/s10518-011-9298-7.

Hirakawa, ET and BT Aagaard. (2021). Evaluation of updates for the USGS San Francisco Bay region 3D seismic velocity model in the east and north bay portions. *Bull. Seismol. Soc. Am.*, 112 (4), 2070-2096. doi: 10.1785/0120210256.

Horike, M. (1985). Inversion of the phase velocity of long-period microtremors to the S-wave velocity structure down to the basement in urbanized areas, *Journal of Physics of the Earth*, 33 (2), 59–96. doi: 10.4294/jpe1952.33.59.

Howard, AK. (1986). Soil classification handbook: unified soil classification system. Denver, CO.: Geotechnical Branch, Division of Research and Laboratory Services, Engineering and Research Center, Bureau of Reclamation.

Hunter JD. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9 (3), 90-95.

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

Idriss, IM. (2014). An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthq. Spectra*, 30 (3), 1155-1177. doi: 10.1193/070613EQS195M.

Idriss, IM, and HB Seed. (1968). Seismic response of horizontal soil layers. *Journal of the Soil Mechanics and Foundations Division*, 94 (4), 1003-1031. doi: 10.1061/jsfeaq.0001163.

Inc, S. (2014). *fuzzywuzzy: Fuzzy String Matching in Python*.

Iwahashi, J and R Pike. (2007). Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology (in press). *Geomorphology*. 86 (3). 409-440. 10.1016/j.geomorph.2006.09.012.

Jennings, CW, RG Strand, and TH Rogers. (1977). Geologic map of California: California Division of Mines and Geology, scale 1:750,000.

Johnson, LR and EL Majer. (2017). Induced and triggered earthquakes at The Geysers geothermal reservoir. *Geophysical Journal International*, 209 (2), 1221–1238. doi: 10.1093/gji/ggx082.

Julian, BR and D Gubbins. (1977). Three-dimensional seismic ray tracing. *J. Geophys*. 43 (1), 95-113.

Jupyter Development Team. (2022). Ipywidgets 8.0.2: Interactive HTML widgets.

Kanai, K, T Tanaka, and K Osada. (1954). Measurement of the microtremor. *Bull. Earthq. Research Institute of Tokyo*, 32, 199–209 (in Japanese).

Kazemian, S, HBK Bujang, A Prasad, and B Maassoumeh. (2011). A state of art review of peat: Geotechnical engineering perspective. *International Journal of Physical Sciences*. 6 (8), 1974-1981. doi: 10.5897/IJPS11.396.

Kim, B, YMA Hashash, EM Rathje, JP Stewart, S Ni, PG Somerville, AR Kottke, WJ Silva, and KW Campbell. (2016). Subsurface Shear Wave Velocity Characterization Using P -Wave Seismograms in Central and Eastern North America. *Earthq. Spectra*. 32 (1), 143-169. doi: 10.1193/123013EQS299M.

Kishida, T, RW Boulanger, NA Abrahamson, MW Driller, and TM Wehling. (2008). Seismic response of levees in the Sacramento-San Joaquin Delta. *Earthq. Spectra*. 25 (3), 557-582. doi: 10.1193/1.3157259.

Kishida, T, RW Boulanger, NA Abrahamson, TM Wehling, and MW Driller. (2009). Regression models for dynamic properties of highly organic soils. *Journal of Geotechnical and Geoenvironmental Engineering*. 135 (4), 533-543. doi: 0.1061/(ASCE)1090-0241(2009)135:4(533).

Kishida, T, S Wang, S Mazzoni, C Markham, Y Lu, Y Bozorgnia, S Mahin, J Bray, M Panagiotou, J Stewart, R Darragh, N Abrahamson, J Hollenback, C Gutierrez, B Chiou, S Muin, and DS Dreger. (2014). Geotechnical engineering reconnaissance of the August 24, 2014 M6 South Napa Earthquake, Section 2, Geotechnical Extreme Events Reconnaissance Association (GEER).

Kluyver, T, B Ragan-Kelley, F Pérez, B Granger, M Bussonnier, J Frederic, K Kelley, J Hamrick, J Grout, S Corlay, P Ivanov, D Avila, S Abdalla, C Willing, and 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. pp. 87-90. doi: 10.3233/978-1-61499-649-1-87.

Konno, K and T Ohmachi. (1998). Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor, *Bull. Seismol. Soc. Am.*, 88 (1), 228–241, doi: 10.1785/BSSA0880010228.

Kotha, SR, G Weatherill, D Bindi, and F Cotton. (2020). A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe. *Bull. Earthquake Eng.*, 18, 4091–4125. 10.1007/s10518-020-00869-1.

Kottke, A. (2020). arkottke/pykooh: v0.3.1 (2020-07-17). Available at: https://doi.org/10.5281/zenodo.3950071.

Kottke, AR, NA Abrahamson, DM Boore, Y Bozogrnia, CA Goulet, J Hollenback, T Kishida, OJ Ktenidou, EM Rathje, WJ Silva, EM Thompson, and X Wang. (2021). Selection of random vibration theory procedures for the NGA-East project and ground motion modeling, *Earthq. Spectra*, 37 (1), 1420-1439. doi: 10.1177/87552930211019052.

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

Kuehn, NM, NA Abrahamson, and MA Walling. (2019). Incorporating nonergodic path effects into the NGA-West2 ground-motion prediction equations. *Bull. Seismol. Soc. Am.*, 109 (2), 575-585. doi: 10.1785/0120180260.

Kuehn, NM, T Kishida, M AlHamaydeh, G Lavrentiadis, and Y Bozorgnia. (2020). A Bayseian model for truncated regression for the estimation of empirical ground-motion models. *Bull. Seismol. Soc. Am.*, 18 (14), 6149-6179. doi: 10.1007/s10518-020-00943-8.

Kuyuk, SH, Allen RM. (2013). Optimal seismic network density for earthquake early warning: A case study from California, *Seismol. Research Letters*, 84 (6), 946–954. doi: 10.1785/0220130043.

Kwak DY, Ahdi SK, Wang P, Zimmaro P, Brandenberg SJ, and Stewart JP. (2021). Web portal for shear wave velocity and HVSR databases in support of site response research and applications. UCLA Geotechnical Engineering Group.

Kwak, DY, TD Ancheta, D Mitra, SK Ahdi, P Zimmaro, GA Parker, SJ Brandenberg, and JP Stewart. (2017a). Performance evaluation of V_{SZ}-to-V_{S30} correlation methods using global V_S profile database, *Proc. 3rd Int. Conf. on Performance-Based Design in Earthquake Geotechnical Engineering,*, M Taiebat, D Wijewickreme, A Athanasopoulos-Zekkos, and RW Boulanger (editors), Paper No. 399, Vancouver, BC.

Kwak, DY, JP Stewart, SJ Mandokhail, and D Park. (2017b). Supplementing V_{S30} with H/V spectral ratios for predicting site effects. *Bull. Seismol. Soc. Am.*, 107 (5), 2028-2042, doi: 10.1785/0120160353.

Landwehr, N, NM Kuehn, T Scheffer, and N Abrahamson. (2016). A nonergodic ground-motion model for California with spatially varying coefficients, *Bull. Seismol. Soc. Am.*, 106 (6), 2574–2583.

Lee EJ, P Chen, TH Jordan, PB Maechling, MAM Denolle, and GC Beroza. (2014). Full 3-D tomography for crustal structure in Southern California based on the scattering-integral and adjoint-wavefield methods, *J. Geophys. Res. Solid Earth*, 119 (8), 6421-6451, doi: 10.1002/2014JB011346.

Li, Z. (2021). Recent advances in earthquake monitoring I: Ongoing revolution of seismic instrumentation. *Earthq. Science*, 34 (2), 177-188. doi: 10.29382/eqs-2021-0011.

Lottes, IL, A DeMaris, and MA Adler. (1996). Using and interpreting logistic regression: A guide for teachers and students. *Teaching Sociology*, 24 (3), 284-298. doi: 10.2307/1318743.

Ludington, S, BC Moring, RJ Miller, KS Flynn, and MJ Hopkins. (2005). Preliminary integrated databases for the United States - Western States: California, Nevada, Arizona, and Washington: U.S. Geological Survey Open-File Report OFR 2005-1305, U.S. Geological Survey, Reston, Virginia, USA.

Lund, J., Hanak, E., Fleenor, W., Howitt, R., Mount, J., and Moyle, P., (2007). *Envisioning Futures for the Sacramento-San Joaquin Delta*, Public Policy Institute of California, San Francisco, CA.

Mai, PM and KKS Thingbaijam. (2014). SRCMOD: An online database of finite-fault rupture models. *Seismological Research Letters*, 85 (6), 1348-1357. doi: 10.1785/0220140077.

Mazzoni S, Kishida T, Stewart JP, Contreras V, Darragh RB, Ancheta TD, Chiou BSJ, Silva WJ, and Bozorgnia Y. (2022) Relational database used for ground-motion model development in the NGA-sub project. *Earthq. Spectra*. 38 (2), 1529-1548. doi: 10.1177/87552930211055204.

McDonald, GN and FX Ashland. (2008). Earthquake site-conditions map for the Wasatach Front urban corridor, Utah, Utah Geological Survey Special Study 125, Utah Geological Survey.

Menke, W. (1989). Geophysical data analysis: Discrete inverse theory. San Diego, Academic Press.

Mihaylov, D, MH El Naggar, and S Dineva. (2016). Separation of high- and low-level ambient noise for HVSR: Application in city conditions for greater Toronto area. *Bull. Seismol. Soc. Am.* 106 (5), 2177-2184, doi: 10.1785/0120150389.

Mishra, P, CM Pandey, U Singh, A Keshri, and M Sabaretnam. (2019). Selection of appropriate statistical methods for data analysis. *Ann. Card/ Anaesth.* 22 (3), 297-301. doi: 10.41-3/aca.ACA_248_18.

Molnar, S, JF Cassidy, S Castellaro, C Cornou, H Crow, JA Hunter, S Matsushima, FJ Sanchez-Sesma, and A Yong. (2018). Application of microtremor horizontal-to-vertical spectral ratio (MHVSR) analysis for site characterization: State of the Art, *Surv. Geophys*. 39, 613–631, doi: 10.1007/s10712-018-9464-4.

Molnar, S, A Sirohey, J Assaf, PY Bard, S Castellaro, C Cornou, B Cox, B Guillier, B Hassani, H Kawase, S Matsushima, F Sánchez-Sesma, and A Yong. (2022). A review of the microtremor horizontal-to-vertical spectral ratio (MHVSR) method. *Journal of Seismology*. 26. 653-685. doi: 10.1007/s10950-021-10062-9.

More, JJ. (1977). The Levenberg-Marquardt Algorithm: Implementation and theory, Numerical Analysis, ed. G. A. Watson, Lecture Notes in Mathematics 630, Springer Verlag, pp. 105-116.

Nakamura, Y. (1989). A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface, *Q. Rep. Railway Tech. Res. Inst*. 30 (1), 25–33.

Nakanishi, I and K Yamaguchi. (1986). A numerical experiment on nonlinear image reconstruction from first-arrival times for two-dimensional island arc structure. *J. Phys. Earth*, 34 (2), 195–201. doi: 10.4294/jpe1952.34.195.

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

Okada, H, T Matsushima, T Moriya, and T Sasatani. (1990). An exploration technique using long-period microtremors for determination of deep geological structures under urbanized areas, *Butsuri-Tansa (Geophys. Explor.)*, 43, 402–417 (in Japanese).

Pandas Development Team. (2020). Pandas-dev/pandas: Pandas 1.4.4. *Zenodo*.

Panzera, F, J Alber, W Imperatori, P Bergamo, and D Fäh. (2021). Reconstructing a 3D model from geophysical data for local amplification modelling: The study case of the upper Rhone valley, Switzerland. *Soil Dynamics and Earthquake Engineering*, 155, 107-163. doi: 10.1016/j.soildyn.2022.107163.

Park, CB, RD Miller, and J Xia. (1999). Multichannel analysis of surface waves, *Geophysics*, 64 (3), 800-808. doi: 10.1190/1.1444590.

Parker, GA and AS Baltay. (2022). Empirical map-based nonergodic models of site response in the Greater Los Angeles Area, *Bull. Seismol. Soc. Am.*, 112 (3), 1607-1629. doi: 10.1785/0120210175.

Parker, GA, JA Harmon, JP Stewart, YMA Hashash, AR Kottke, EM Rathje, WJ Silva, and KW Campbell. (2017). Proxy-Based VS30 Estimation in Central and Eastern North America, *Bull. Seism. Soc. Am.,* 107 (1), 117-131. doi: 10.1785/0120160101.

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

Parolai, S and JJ Galiana-Merino. (2006). Effect of transient seismic noise on estimates of H/V spectral ratios. *Bull. Seis. Soc. Am.,* 96 (1), 228–236. doi: 10.1785/0120050084.

Pedregosa, F, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot, and E Duchesnay. (2011). Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, 12, 2825-2830.

Pinilla-Ramos, C, N Abrahamson, and R Kayen. (2022). Estimation of site terms in ground-motion models for California using horizontal-to-vertical spectral ratios from microtremor. *Bull. Seismol. Soc. Am.*, doi: 10.1785/0120220033.

Power, M, B Chiou, NA Abrahamson, Y Bozorgnia, T Shantz, C and Roblee. (2008). An overview of the NGA project, *Earthq. Spectra*, 24 (1), 3 –21. doi: 10.1193/1.2894833.

Python-Visualization. (2020). *Folium*. Retrieved from https://python-visualization.github.io/folium/.

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramos-Sepulveda, ME, GA Parker, EM Thompson, SJ Brandenberg, O Ilhan, YMA Hashash, EM Rathje, and JP Stewart. (2022). High-pass corner frequency selection for implementation in the USGS automated ground motion processing tool. *SSA 2022 Annual Meeting*, Bellevue, WA, April 19-23, 2022.

Ramos-Sepulveda, ME, 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*, Los Angeles, California, March 26-29, 2022.

Rawlinson, N, J Hauser, and M Sambridge. (2008). Seismic ray tracing and wavefront tracking in laterally heterogeneous media, *Advances in Geophysics*, 49, 203-273. doi: 10.1016/S0065-2687(07)49003-3.

Reitz K. (2022). Requests 2.28.1: Python HTTP for humans.

Riley, SJ, SD DeGloria, and R Elliot. (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences*, 5 (1-4), 23-27.

Robertson, PK. (2016). Cone penetration test (CPT)-based soil behaviour type (SBT) classification system — an update. *Canadian Geotechnical Journal*. 53 (12), 1910-1927. doi: 10.1139/cgj-2016-0044.

Rowley, PD, VS Williams, GS Vice, DJ Maxwell, DB Hacker, LW Snee, and JH Mackin. (2006). Interim geologic map of the Cedar City 30' x 60' quadrangle, Iron and Washington counties, Utah, Utah Geological Survey Open-File Report 476DM, Utah Geological Survey.

Ryan, H. (1994). Ricker, Ormsby, Klauder, Butterworth—A choice of wavelets. *CSEG Recorder*, 19 (7), 8–9.

Seed, HB, C Ugas, and J Lysmer. (1976). Site-dependent spectra for earthquake-resistant design. *Bull. Seismol. Soc. Am.*, 66 (1), 221-243. doi: 10.1785/BSSA0660010221.

Senna, S, S Midorikawa, and K Wakamatsu. (2008). Estimation of spectral amplification of ground using H/V spectral ratio of microtremors and geomorphological land classification. *J. Jpn. Assoc. Earthq. Eng.*, 8 (4), 1-15. doi: 10.5610/jaee.8.4_1.

Seyhan, E and JP Stewart. (2014). Semi-empirical nonlinear site amplification from NGA-West2 data and simulations. *Earthq. Spectra*, 30 (3), 1241-1256. doi: 10.1193/063013EQS181M.

Seyhan, E, JP Stewart, TD Ancheta, RB Darragh, and RW Graves. (2014). NGA-West2 site database, *Earthq. Spectra*, 30 (3), 1007-1024. doi: 10.1193/062913EQS180M.

Shafiee, A, JP Stewart, and SJ Brandenberg. (2015). Reset of secondary compression clock for peat by cyclic straining. *J. Geotech. Geoenviron. Eng.*, 141 (3), 02815001. doi: 10.1061/(ASCE)GT.1943-5606.0001286.

Shaw, JH, A Plesch, C Tape, MP Suess, TH Jordan, , G Ely, E Hauksson, J Tromp, T Tanimoto, R Graves, K Olsen, C Nicholson, PJ Maechling, C Rivero, P Lovely, CM Brankman, and J Munster. (2015). Unified Structural Representation of the southern California crust and upper mantle, *Earth Planet. Sci. Lett.*, 415, 1-15, doi: 10.1016/j.epsl.2015.01.016.

Site EffectS assessment using AMbient Excitations (SESAME). (2004). Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations—Measurements, processing and interpretation, *European Commission Project No. EVG1-CT2000-00026, 62*.

Smith RL and WP Roe. (2015). Oregon geologic data compilation (OGDC) - Release 6, Oregon Department of Geology and Mineral Industries (DOGAMI), Open-File Rept. O-13-12, electronic dataset.

Snow AD, J Whitaker, M Cochran, et al. (2021). Pyproj4/pyproj: 3.0.1rc0. *Zenodo*.

Squires, GL. (2001). *Practical Physics* (4th ed.). Cambridge: Cambridge University Press. doi: 10.1017/CB)9781139164498.

Stafford, JS, A Rodriguez-Marek, B Edwards, PP Kruiver, and JJ Bommer. (2017). Scenario dependence of linear site-effect factors for short-period response spectral ordinates. *Bull. Seismol. Soc. Am.*, 107 (6), 2859–2872. doi: 10.1785/0120170084.

Stewart, JP, K Afshari, and CA Goulet. (2017). Non-Ergodic Site Response in Seismic Hazard Analysis. *Earthq. Spectra*, 33 (4), 1385-1414. doi: 10.1193/081716eqs135m.

Stokoe II, KH, SG Wright, JA Bay, and JM Roesset. (1994). Characterization of geotechnical sites by SASW Method, *Geophysical Characterization of Sites*, R.D. Woods, ed., A.A. Balkema, Rotterdam, 15–25.

Strutz, T. (2016). *Data Fitting and Uncertainty (A practical introduction to weighted least squares and beyond)*. Springer Vieweg. ISBN 978-3-658-11455-8, Chapter 3.

Thompson, EM. (2018). An updated Vs30 map for California with geologic and topographic constraints: U.S. Geological Survey data release. doi: 10.5066/F7JQ108S.

Thurber, CH, TM Brocher, H Zhang, and VE Langenheim. (2007). Three-dimensional P wave velocity model for the San Francisco Bay region, California, *J. Geophys. Res*. 112, B07313. doi: 10.1029/2006JB004682.

URS Corporation/Jack R. Benjamin & Associates, Inc. (2008). *Delta Risk Management Strategy* (DRMS), Phase 1, Risk Analysis Report, Prepared for the California Department of Water Resources (DWR).

Urzúa, A. (1974). Determinación del Período Fundamental de Vibración del Suelo, Thesis presented to the University of Chile for the title of Ingeniero Civil and the degree of Master in Soil Mechanics (in Spanish).

Vidale, JE. (1988). Finite-difference calculations of traveltimes. *Bull. Seism. Soc. Am.*, 78, 2062–2076.

Viegas, G, and L Hutchings. (2010). Source characteristics of micro-earthquakes at the Northwest Geysers Geothermal Field, California. Transactions - Geothermal Resources Council. 34. 1265-1272.

Virtanen, P, R Gommers, TE Oliphant, M Haberland, T Reddy, D Cournapeau, E Burovski, P Peterson, W Weckesser, J Bright, SJ van der Walt, M Brett, J Wilson, KJ Millman, N Mayorov, ARJ Nelson, E Jones, R Kern, E Larson, CJ Carey, İ Polat, Y Feng, EW Moore, J VanderPlas, D Laxalde, J Perktold, R Cimrman, I Henriksen, EA Quintero, CR Harris, AM Archibald, AH Ribeiro, F Pedregosa, P van Mulbregt, and SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17 (3), 261-272. doi: 10.1038/s41592-019-0686-2.

Wald, DJ and TI Allen. (2007). Topographic slope as a proxy for seismic site conditions and amplification, *Bull. Seismol. Soc. Am.*, 97 (5), 1379–1395. doi: 10.1785/0120060267.

Wang, P. (2020). Predictability and Repeatability of Non-Ergodic Site Response for Diverse Geological Conditions. Ph.D. thesis. UC Los Angeles.

Wang, P, Z Liu, SJ Brandenberg, P Zimmaro, and JP Stewart. (2022c). Regression-based event selection for hazard-consistent seismic risk assessment. *12th National Conference on Earthquake Engineering*, EERI, Salt Lake City, UT, Jun 27 – Jul 1, 2022.

Wang, P, JP Stewart, Y Bozorgina, DM Boore, and T Kishida. (2017). R Package for Computation of Earthquake Ground Motion Response Spectra. Report No. 2017/09. PEER, UC Berkeley.

Wang, P, YT Tsai, JP Stewart, A Mikami, and SJ Brandenberg. (2022a). Region-specific linear site amplification model for peaty organic soil sites in Hokkaido, Japan, *Earthq. Spectra*, 38 (3), 2207-2234, doi: 10.1177/87552930221082965.

Wang, P, P Zimmaro, T Gospe, SK Ahdi, A Yong, and JP Stewart. (2021). Horizontal-to-vertical spectral ratios from California sites: Open-source database and data interpretation to establish site parameters, *Report GIRS-2021-06*, B. John Garrick Risk Institute, Natural Hazards Risk and Resiliency Research Center, UCLA (Center Headquarters), 64 pages.

Wang, P, P Zimmaro, TE Buckreis, T Gospe, SJ Brandenberg, SK Ahdi, A Yong, and JP Stewart. (2022b). Relational database for Horizontal-to-Vertical Spectral Ratios, *Seism. Research Letters*, 93 (2A), 1075-1088. doi: 10.34948/N3KW20.

Water Education Foundation. (2010). C-WIN: Peripheral Canals: Way past, past, and present, https://www.watereducation.org/aquafornia-news/c-win-peripheral-canals-way-past-past-and-present.

Wathelet, M, JL Chatelain, C Cornou, G Di Giulio, B Guillier, M Ohrnberger, and A Savvaidis. (2020). Geopsy: A user-friendly open-source tool set for ambient vibration processing, *Seismol. Res. Lett.* 91 (3), 1878–1889, doi: 10.1785/0220190360.

Wells DL and KJ Coppersmith. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bull. Seismol. Soc. Am.*, 84 (4), 974-1002. doi: 10.1785/BSSA0840040974.

Wheling, TM, RW Boulanger, R Arulnathan, LR Harder, and MW Driller. (2003). Nonlinear dynamic properties of a fibrous organic soil. *Journal of Geotechnical and Geoenvironmental Engineering*, 129 (10), 929-939. doi: 10.1061/(ASCE)1090-0241(2003)129:10(929).

Wills, CJ, CI Gutierrez, FG Perez, and DM Branum. (2015). A Next Generation VS30 Map for California Based on Geology and Topography. *Bull. Seismol. Soc. Am*., 105 (6), 3083–3091. doi: 10.1785/0120150105.

Wong, LS, R Hashim, and FH Ali. (2009). A review on hydraulic conductivity and compressibility of peat. *Journal of Applied Sciences*, 9 (18), 3207-3218. doi: 10.3923/jas.2009.3207.3218.

Yong, A. 2016. Comparison of measured and proxy-based VS30 values in California, *Earthq. Spectra*. 32 (1), 171-192. doi: 10.1193/013114EQS025M.

Yong, A, SE Hough, J Iwahashi, and A Braverman. (2012). A terrain based site conditions map of California with implications for the contiguous United States. *Bull. Seismol. Soc. Am.*, 102 (1), 114-122. doi: 10.1785/0120100262.

Yong, A, EM Thompson, D Wald, KL Knudsen, JK Odum, WJ Stephenson, and S Haefner. (2015). Compilation of Vs30 data for the United States: U.S. Geological Survey Data Series 978, 8 p.

Zadeh, LA. (1988). Fuzzy logic. *Computer*. 21 (4), 83-93. doi: 10.1109/2.53.

Zhao, JX and H Xu. (2013). A comparison of V_{S30} and site period as site-effect parameters in response spectral ground-motion prediction equations, *Bull. Seismol. Soc. Am.*, 103 (1), 1-18. doi: 10.1785/0120110251.

Zimmaro, P, JP Stewart, SJ Brandenberg, DY Kwak, and R Jongejan. (2019). Multi-hazard system reliability of flood control levees, *Soi. Dyn. Earthq. Eng.*, 124, 345-353. doi: 10.1016/j.soildyn.2018.04.043.