Cal Poly

Caltech

SCEC
AN NSF+USGS CENTER

UC Irvine

UCLA

UC Santa Barbara

USC

Site Response of Southern California Sedimentary Basins and Other Geomorphic Provinces

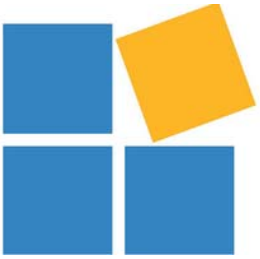
Chukwuebuka C. Nweke
Jonathan P. Stewart
Scott J. Brandenberg
Civil & Environmental Engineering Department
University of California, Los Angeles

A report on research supported by United States Geological Survey, Southern California Earthquake Center, California Department of Transportation, and National Science Foundation

Report GIRS-2020-12

DOI: 10.34948/N3159F

University of California, Los Angeles (headquarters)



Natural Hazards Risk & Resiliency Research Center

B. John Garrick Institute for the Risk Sciences

Site Response of Southern California Sedimentary Basins and Other Geomorphic Provinces

Chukwuebuka C. Nweke, Ph.D., Jonathan P. Stewart, Ph.D., and Scott J. Brandenberg, Ph.D.

Civil and Environmental Engineering Department
University of California, Los Angeles

A report on research conducted with support from US Geological Survey (Contract G19AP00025), Southern California Earthquake Center (Subawards 104715500 and 118062755), the California Department of Transportation, and a National Science Foundation AGEP Fellowship program.

Report GIRS-2020-12
DOI: 10.34948/N3159F

Natural Hazards Risk and Resiliency Research Center
B. John Garrick Institute for the Risk Sciences
University of California, Los Angeles (Headquarters)

October 2020

ABSTRACT

Seismic site response can be influenced by a variety of physical mechanisms that include amplification due to resonance, nonlinearity, topographic effects, impedance contrasts, and contributions from two- or three-dimensional wave propagation in sedimentary basins. Current ground motion models use ergodic procedures that average these effects over many sites globally by conditioning on the time-averaged shear wave velocity in the upper 30 m (V_{S30}), and in some cases, on the depth to a shear wave velocity isosurface (z_x) that is also known as the basin depth parameter.

Current site amplification models conditioned on V_{S30} reflect, in an average sense, most of the aforementioned physical mechanisms, including basin effects. The site response contributions from basin effects are associated with a differential depth parameter taken as the difference between depth for a particular site and average basin depth conditioned on V_{S30}. The basin amplification models are "centered", in the sense that they predict changes in ground motion amplification for non-zero differential depths. The changes in ground motion amplification are positive and negative for sites with positive and negative differential depths, respectively. The models predicting this behavior are derived using data from both northern and southern California, and for sites situated within sedimentary basin structures but also other geomorphic provinces (e.g., sedimentary structures of different scales and sites with shallow soil overlying rock).

We investigate the benefits of regionalizing basin response in ergodic ground motion models. Using southern California data we consider the following questions: (1) how should basin and non-basin locations be classified?; (2) how does mean site response and the associated ground motion variability differ for basins compared to non-basin geomorphic provinces?; and (3) what are the variations in basin response between different basin structures and how can this be modelled for predicting ground motion intensity measures?

We recommend a site classification scheme that distinguishes basins, basin edges, valley (sedimentary structures smaller in scale than basins), and mountain/hill areas. Moreover, we distinguish basins in southern California based on their geologic origin: coastal basins with varied depositional histories and large depths (e.g., Los Angeles Basin); inland, fault-bounded basins with relatively shallow sediments derived from neighboring mountains (e.g., Chino Basin); Imperial Valley, a basin on the transform fault plate boundary at the location of a graben located in the step-over between the San Andreas and Imperial Faults.

To address the second and third questions, we compile a large ground motion database for southern California that significantly expands upon the data available in the NGA-West2 project, and which has the benefit of significantly increasing the number of recorded events per site. We verify that an NGA-West2 ground motion model has unbiased source and path terms relative to the dataset, and we make minor modifications to the global V_{S30}-scaling function to fit the mean

of southern California data. Using the slightly adjusted ground motion model, we compute site terms for 670 sites, which approximately represent the mean difference between the actual site response and the ground motion model prediction.

We find that the combined data (i.e., site terms) from all sites exhibit trends with differential depth that are qualitatively similar to those in NGA-West2 models (de-amplification for negative differential depth, amplification for positive). We find basin and basin edge categories to be similar to each other, but different from the combined data set in the sense that de-amplification at negative differential depths is generally absent. Moreover, the depth-invariant mean amplification for this condition is positive, indicating under-prediction from the V_{S30}-scaling model. We find the valley and mountain/hill categories to exhibit similar trends in which amplification scales with differential depths and has both positive and negative values. The depth-invariant mean amplification for these conditions is negative.

Among basin sites, we find differences for coastal and inland basins. The response of coastal basins essentially matches that for the overall basin category (amplification scales up with increasing differential depth). A similar pattern is followed by the Imperial Valley Basin. However, the response of other inland basins is different, with no appreciable dependence on differential depth, but apparent uniform shifts that are specific to individual basins (but which are poorly constrained by the data).

Models are proposed to capture the mean behavior of the recommended groups -- coastal basins, inland basins, and Imperial Valley Basin. Site-to-site standard deviation terms (ϕ_{S2S}) are found to vary strongly across geomorphic provinces, with basins and valleys having notably lower dispersions than mountain/hill sites and the reference ergodic model.

ACKNOWLEDGMENTS

Support for this research project was provided by from USGS Contract G19AP00025, SCEC Subawards 104715500 and 118062755, and the California Department of Transportation (Caltrans) Task Order 3, administered through the Pacific Earthquake Engineering Research Center. Limited support was also provided by NSF-AGEP California Alliance Fellowship (awards 1306595, 1306683, 1306747, 1306760). The support of these organizations is appreciated.

Pengfei Wang performed processing of ground motions recorded since 2012 that were used in this research and assisted with the development and maintenance of the site database used in this study.

The opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the study sponsors, the B. John Garrick Risk Institute, the Pacific Earthquake Engineering Research Center, or the Regents of the University of California.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

Seismic site response can be influenced by a variety of physical mechanisms, including amplification above impedance contrasts, resonance, nonlinearity, topographic effects, and amplification related to two- or three-dimensional wave propagation in sedimentary basins. For the purposes of site response modeling using ergodic procedures (including the site terms in NGA-West2 ground motion models, GMMs), these effects are averaged over many sites globally with conditioning on time-averaged shear wave velocity in the upper 30 m (V_{S30}) and, in some cases, on basin depth parameters.

The portion of the site amplification model conditioned on V_{S30} reflects, in an average sense, all of these physical mechanisms, including basin effects to the extent they are present in the empirical data from which the V_{S30} term is derived. The contribution of basin amplification can be loosely associated with an average depth conditional on that V_{S30}. The basin amplification models incorporated into NGA-West2 GMMs are 'centered', in the sense that they predict changes in amplification at long periods for depths different from that average. For long-period ground motions, such models predict de-amplification (less than provided by the V_{S30}-scaling function) for shallower depths, and amplification for larger depths.

The NGA-West2 site amplification models (V_{S30}-based and basin depth-based components) form the primary basis for ergodic site effect modeling in the development of the 2018 US National Seismic Hazard Model (Petersen et al. 2020) for sites in the western U.S.. The USGS implementation of the basin models applies the basin depth term only when it produces amplification relative to the V_{S30}-based model. Deamplification effects at sites with relatively shallow depths were not applied.

There are several critical elements of the NGA-West2 basin amplification models that we have interrogated in this study using data from southern California::

- Centering: Because the basin amplification model operates on a depth difference (depth minus V_{S30}-conditioned mean), it is sensitive to the mean depth model. Current relations for the mean depth apply for broad regions (California, Japan) and have large scatter. Here we re-evaluate the mean depth model for southern

California based on additional information and an alternative model for basin geometry to that originally considered.

- Site Geology: While the NGA-West2 models are referred to as "Basin amplification models", in fact they operate solely on sediment depth without explicit consideration of whether the sites are in basins or other types of geological features. Here we evaluate whether a site's geological setting (e.g., mountainous region, small sedimentary feature, broad sedimentary basin) affects the relationship between site response and differential depth.
- Amplification Function: The NGA-West2 basin effects model de-amplifies ground motion for negative differential depths and amplifies ground motion for positive differential depths. These effects occur at long-periods only (T > 0.65 sec). Here we re-evaluate each model element by answering the following questions for each geological setting: (1) does site amplification depend on differential depth?; (2) does de-amplification occur for some range of differential depths?; and (3) does amplification occur for some range of differential depths?

As part of this work, we also investigated the dispersion of ground motion, also known as aleatory variability. This variability is represented in seismic hazard analyses using a total standard deviation (σ_{ln}), which has contributions from between-event variability (τ_{ln}) and within-event variability (ϕ_{ln}).

$$\sigma_{ln} = \sqrt{\tau_{ln}^2 + \phi_{ln}^2} \tag{1.1}$$

Within-event dispersion has contributions from path-to-path and site-to-site variabilities. Regional and azimuthal variations in path effects account for different attenuation rates as ground motions propagate from source to site along different paths. Ground motion models provide average attenuation rates, and the aleatory variability associated with variations from that average is denoted ϕ_{P2P}. Similarly, regional and site-to-site variations in geologic structure cause variable levels of site amplification, even when 'primary' site variable V_{S30} and basin depth terms are specified. Regional variations are accounted for in region-specific ergodic models, which may have different levels of ground motion scaling with V_{S30} (e.g., Parker et al. 2019). Site-to-site variations in site response relative to regional models is appreciable, due to the many aforementioned factors not considered in ergodic models; the dispersion associated with these variations is denoted ϕ_{S2S}. Assuming statistical independence, these different sources of within-event variability combine as follows (modified from Al Atik et al. 2010):

$$\phi_{ln} = \sqrt{\phi_{P2P}^2 + \phi_{S2S}^2 + \phi_{lnY}^2} \tag{1.2}$$

where ϕ_{lnY} is the remaining variability when path- and site-specific models are used.

Chapter 2 of this report describes the database compiled for the present study. Chapter 3 presents a site categorization scheme intended to distinguish sites having different geological settings that might be anticipated to affect site response (e.g., basins, mountain/hill areas, etc.). All sites in the ground motion database are classified following this scheme for use in ground motion data analysis. A basin classification algorithm is presented for the purposes of executing the site categorization scheme.

Chapter 4 describes data analysis procedures that are designed to estimate site response effects from recorded ground motions using an adaptation of the classical non-reference site approach (Field and Jacob, 1995) in which the "true" (or non-ergodic) site response is computed. As part of these analyses, we examine residuals of the data set relative to NGA-West2 GMMs to ensure that potential source and/or path model biases are not mapped into site response estimates.

Because this study seeks to identify contributions to site response beyond what is captured in V_{S30}-scaling models, in Chapter 5 we critically evaluate the suitability of an NGA-West2 V_{S30}-scaling model relative to the data and make minor adjustments for the southern California study region. In Chapter 6 we investigate variations in site response with differential depth for different geological settings and different basin structures. The dispersion of residuals is used to investigate changes in site-to-site variability between categories and specific basins. We develop models that predict site response as a function of geologic setting and differential depth. The report is concluded in Chapter 7 with a summary of principal findings, statement of model limitations, and suggestions for future research.

2 Ground Motion Database

A database of ground motion recordings for the southern California study region was compiled. The database consists of recordings from the NGA-West2 database and data added as part of this study and related efforts (Wang and Stewart 2019; Ahdi et al. 2020). All newly added data is processed using PEER-NGA procedures and metadata are compiled for each new event and station.

2.1 NGA-WEST2 DATASET

We began with the NGA-West2 database (Ancheta et al., 2014), which is a global database for active tectonic regions. There is a significant contribution of data from southern California to the NGA-west2 database (189 events, 789 stations, 6946 recordings) over the time period 1938 to 2010. The site portion of the database (Seyhan et al. 2014) was developed to provide the principal site parameters used in model development − V_{S30} and various depth parameters denoted as z_x. These depths indicate the vertical distance from the ground surface to the first crossing of a shear wave velocity isosurface; the most widely used values are $z_{1.0}$ and $z_{2.5}$ for depths to the 1.0 km/s and 2.5 km/s isosurfaces. Section 2.4 below describes the assignment of V_{S30} and depth parameters to the database.

As part of this project and Wang and Stewart (2019), we converted the spreadsheet files that comprised the original NGA-West2 flatfile (pertaining to sources, sites, and ground motions) into a relational database, which is housed on a local server. Additions of data are made within the relational database. The database is accessed using Python scripts within Jupyter notebooks.

2.2 EXPANDED DATASET

We have identified earthquakes and recordings since 2011 in California, including the recent Ridgecrest Earthquake sequence, which significantly extend the NGA-West2 database. In this extension of the database, we only consider events that satisfy a magnitude criterion (**M** > 4 events) due to difficulties that can be encountered in the analysis of site terms using smaller magnitude data (Stafford et al., 2017). Figure 2.1 shows the locations of events sorted by magnitude, most of

which occur in southern California, the Bay Area and north coast, the eastern Sierra and Nevada, central California and Imperial Valley and northern Mexico. These areas incorporate most of the urban areas in the state, and contain a large fraction of the ground motion stations. There are over 29,330 three-component recordings from 403 events in California.

We focus here on the southern California region. The data from events contributing to the southern California region, shown in Figure 2.1 as a dotted red box, is derived from 29 earthquakes that have produced about 10,700 three-component recordings within the distance cutoffs suggested by Boore et al. (2014). The data are screened to remove duplicate recordings (e.g., seismometers and accelerometers at the same location) and recordings that appear to be unreliable from instrument malfunctions or similar, which leaves about 8106 usable three-component records. Within the study region, there are 1004 stations that have provided recordings.

Each of the three-component records has been processed according to standard protocols developed during Pacific Earthquake Engineering Research center (PEER)-NGA projects, as described in Kishida et al. (2020). This processing provides a lowest usable frequency for each ground motion component. PEER-NGA procedures take the lowest usable oscillator frequency as 1.25 times the low-cut corner frequency used in record processing. Horizontal ground motion components are combined to median-component (RotD50) intensity measures as defined by Boore (2010) using the routines given in Wang et al. (2017). We take the lowest usable oscillator frequency for RotD50 as the higher of the two single-component values. Figure 2.2 shows the number of usable RotD50 horizontal-component ground motions as a function of oscillator period. The fall-off begins at about 1.0 sec and the data is reduced by 50% by 3 sec.

Figure 2.3 shows the newly added data in magnitude-distance space in comparison to the NGA-West2 data. The combined data set for California now has 29,375 recordings from 2673 stations and 403 events. For Southern California, the dataset now has 11,819 recordings (previously 6946 recordings) from 1004 stations (previously 789 stations) and 220 events (previously 188 events). The expansion of the data evident in Figure 2.3 was critical for the present study because our analysis of site terms (defined below) becomes increasingly robust as stations have more usable records. Prior to the present work, there were 139 stations with 10 or more recordings in the study region; whereas the current data set now has 250 such stations.

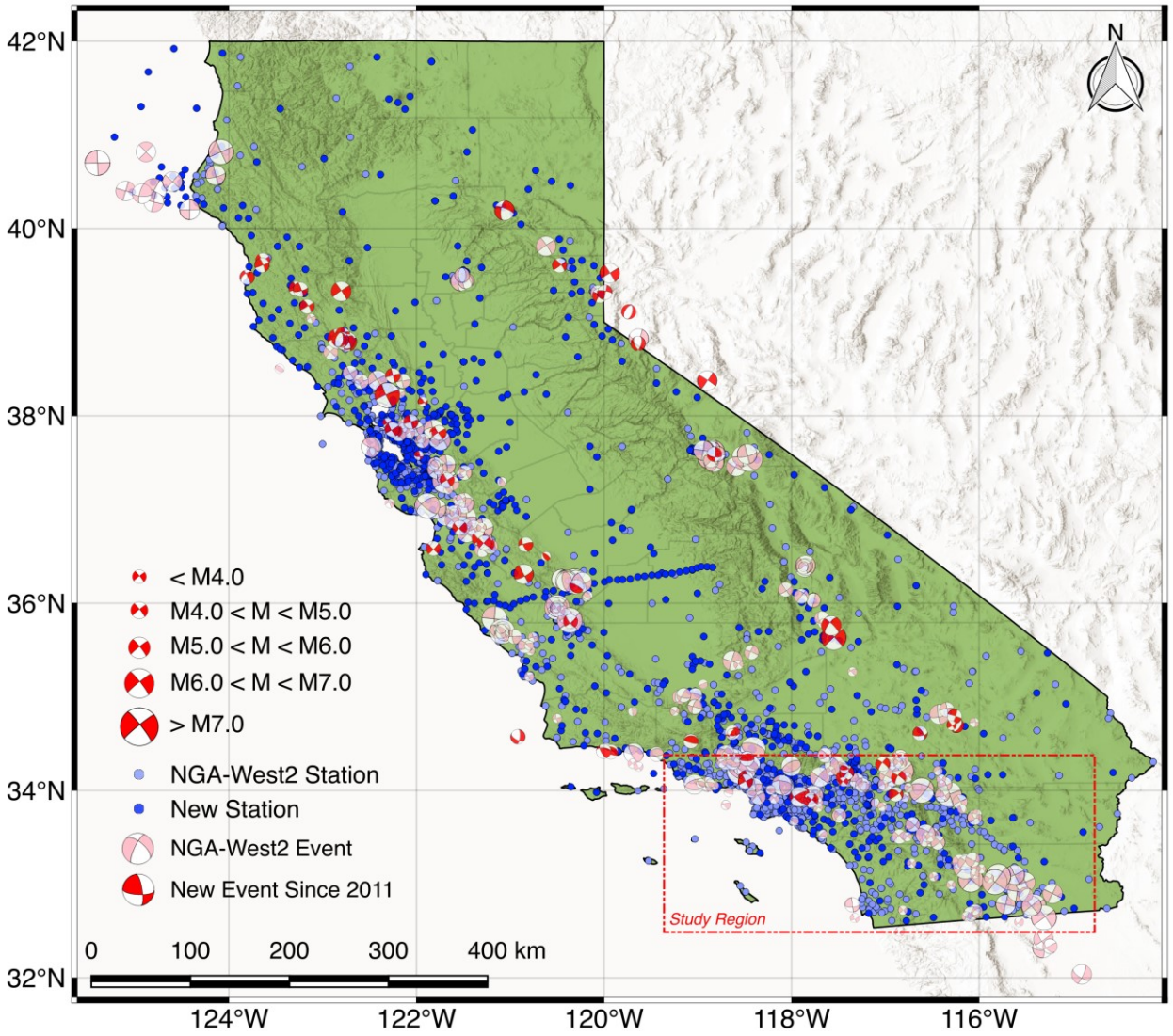


Figure 2.1. Locations of earthquakes in California and Northern Mexico from NGA-West2 and since 2011 for which ground motion data has been compiled, as well as stations that recorded the events.

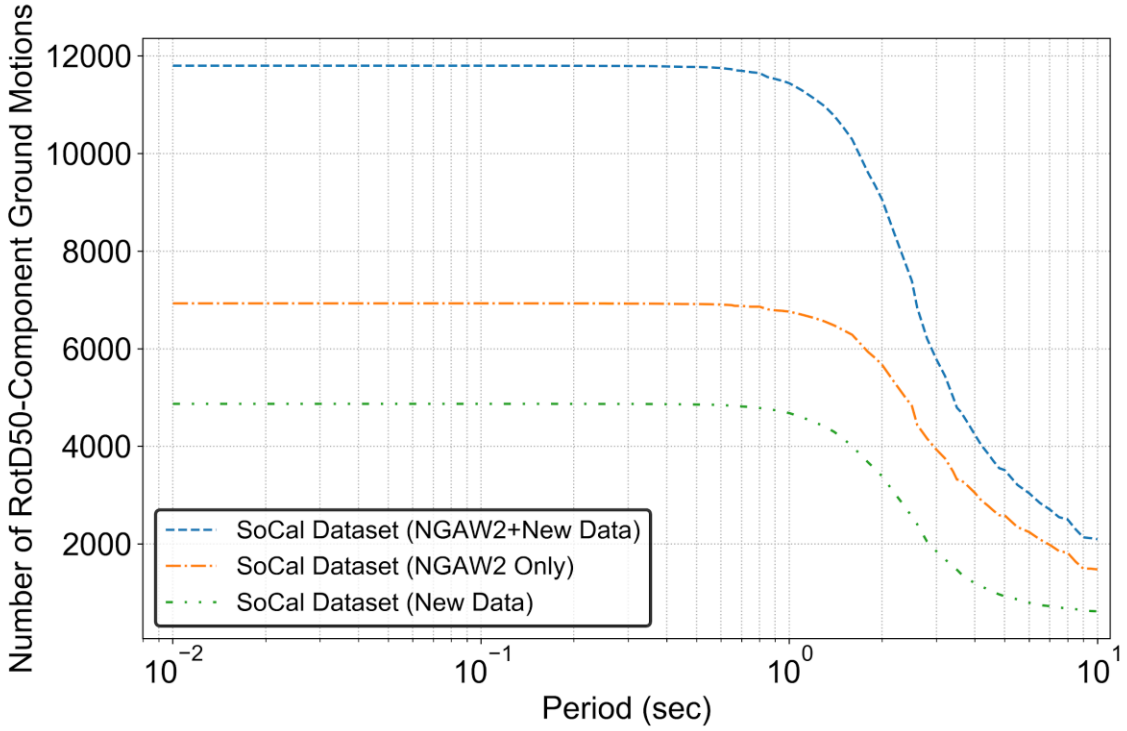


Figure 2.2. Number of usable RotD50-component ground motions as a function of oscillator period for the data added from the southern California region.

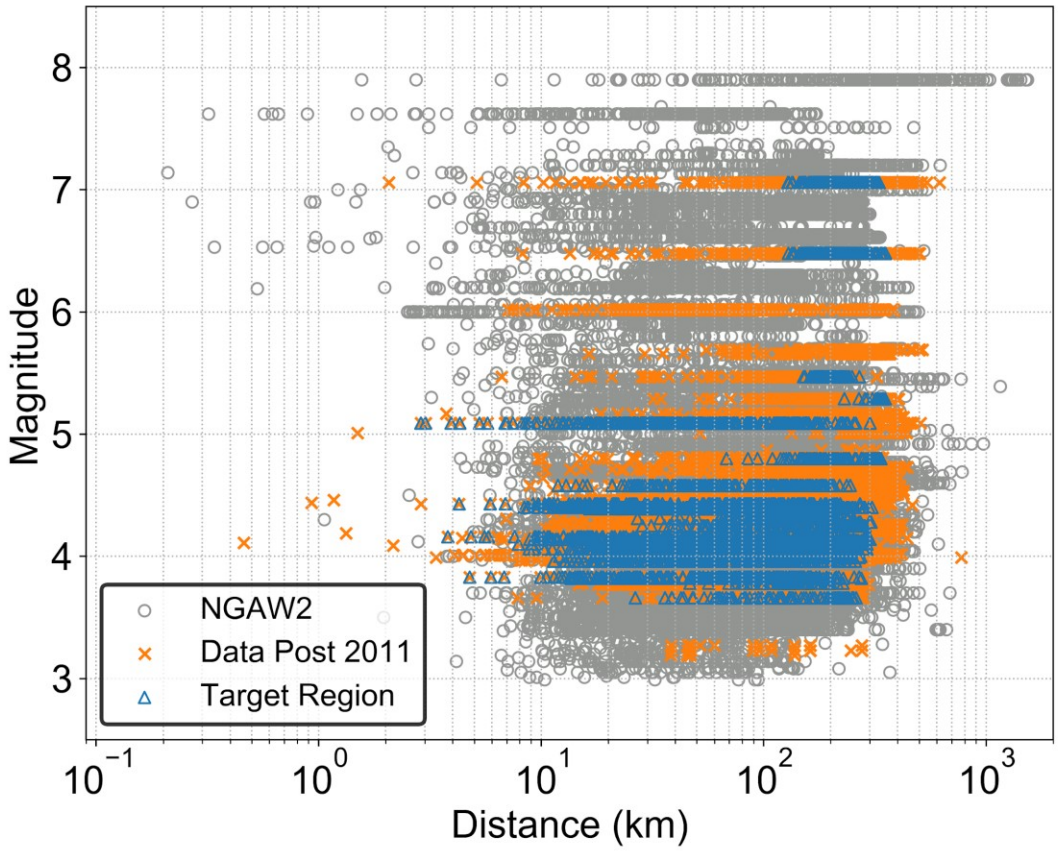


Figure 2.3. Visualization of the updated database in magnitude-distance space including the NGA-West2 contribution, the newly added data from events since 2011, and highlighting the data used in this study for southern California.

2.3 ASSIGNMENT OF SITE PARAMETERS

Considering both the NGA-West2 data and new data, there are 1004 recording sites within the rectangular area shown in Figure 2.1, which is shown in greater detail in Figures 2.4 and 2.5. Of those, 789 are sites that were included in the NGA-West2 site database. Hence, there are 215 new sites that require assignment of site parameters. Following protocols given in Seyhan et al. (2014), V_{S30} was assigned using local shear wave velocity measurements where available – this applies to 21 sites (data obtained from V_S profile database; Ahdi et al. 2018). For sites without V_{S30} from measurements, we use the V_{S30} map derived from geologic- and topographic-based proxy relationships by Thompson et al. (2014), as updated by Thompson (2018) (2/3 weight). We also

consider the terrain-based proxy model of Yong et al. (2012), as updated by Yong (2016) (1/3 weight).

Basin depth parameters $z_{1.0}$ and $z_{2.5}$ were obtained for all of the considered sites, including the NGA-West2 sites and the newly added sites. Older values were replaced because of updates, and expansion, of the southern California basin models. Table 2.1 shows the basin models, including version numbers, used in this compilation. Regions for which basin models have been developed since the close of the NGA-West2 project include the central valley region of California (San Joaquin valley and Santa Maria River valley) and Mojave Desert region. There are 136 sites for which basin depths had not been assigned in NGA-West2 (depths marked as -999) for which depths are now provided.

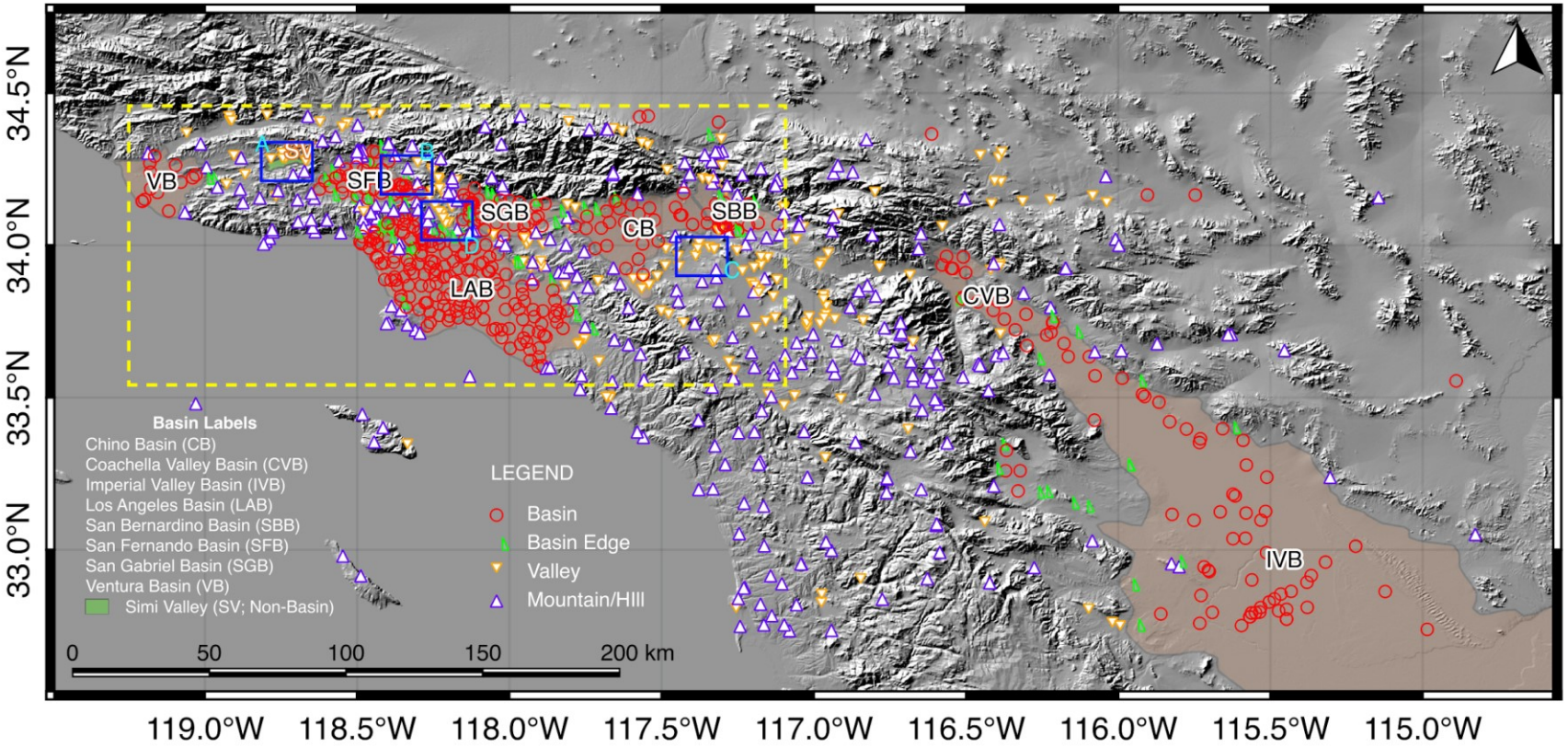


Figure 2.4. Detailed map of southern California showing ground motion stations and sedimentary basins and related features considered in this paper. Ground motion sites are plotted according to a morphology-based site categorization scheme proposed in this paper. Boxes A, B, and C are detailed in subsequent figures in this paper. The yellow dotted box is detailed in Figure 2.5.

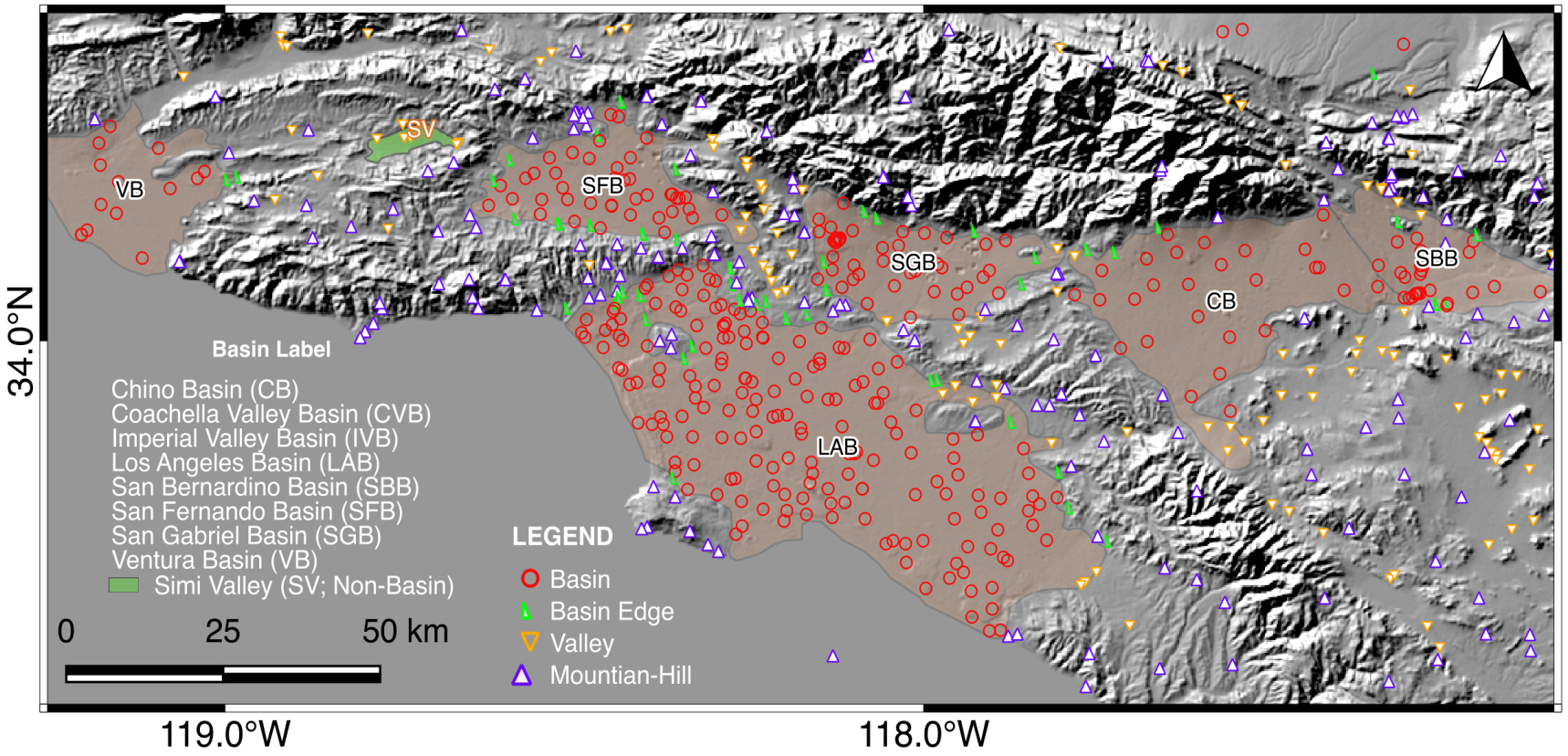


Figure 2.5. Expanded detail of the southern California map showing ground motion stations and basins.

Table 2.1. Seismic velocity models registered into the Unified Community Velocity Model (UCVM) modified from Small et al. (2017)

Model Name UCVM Abbreviation	Description	Region, Coverage Coordinates	References
SCEC CVM-H, v.15.1, (cvmh)	3D velocity model defined on regular mesh, no geotechnical layer. Based on 3D tomographic inversions of seismic reflection profiles and direct velocity measurements from boreholes	So. CA; −120.8620, 30.9565; −113.3329, 30.9565; −113.3329, 36.6129; −120.8620, 36.6129	Süss and Shaw 2003; Shaw et al. 2015
SCEC CVM-S4, (cvms)	3D velocity model defined as rule-based system with a geotechnical layer. Uses query of velocity by depth using empirical relationships from borehole sonic logs and tomographic studies	Irregular area in So. CA	Kohler et al. 2003
SCEC CVM-S4.26, (cvms5) SCEC CVM-S4.26.M01, (cvmsi)	3D velocity model defined on regular mesh, no geotechnical layer. Uses query of velocity by depth based on CVM-S4 as starting model, improved using full 3D tomography. 3D velocity model defined on regular mesh with query by depth that adds a GTL to CVM-S4.26	So. Central CA, So. CA; −116.0000, 30.4499; −122.3000, 34.7835; −118.9475, 38.3035; −112.5182, 33.7819	Lee et al. 2014
USGS Hi-res and Lo-res etree v.08.3.0, (cencal)	3D velocity model defined on regular mesh with geotechnical layer that uses velocity query by depth	Bay Area, No. & Central CA; −126.3532, 39.6806; −123.2732, 41.4849; −118.9445, 36.7022; −121.9309, 35.0090	Brocher et al. 2006
Central CA model, SCEC CCA06, (cca)	3D tomographic inversions done on a coarse mesh (500 m), trilinear interpolation between nodes	Central CA; -122.9362, 36.5298; -118.2678, 39.3084; -115.4353, 36.0116; -120.0027, 33.3384	Still in beta; Chen & Lee (2017)
SCEC CS17.3, (cs173) SCEC CS17.3-H, (cs173h)	CyberShake 17.3 velocity model with added geotechnical layer (UCVMC18.5) 17.3 model integrated with Harvard Santa Maria and San Joaquin basin models with geotechnical layer	Central CA; -127.6187, 37.0453; -124.5299, 41.3799; -112.9435, 35.2956; -116.4796, 31.2355	Still in beta; Ely et al. 2010, 2017
Mod. Hadley Kanamori (1d)	1D velocity model in nine layers that defines V_p and scaling relationship for V_s. Non-basin areas.	So. CA, irregular boundary	Hauksson 2010
Northridge region (bbp1d)	1D velocity model defined in 18 layers, derived from velocity profiles at SCSN stations. Non-basin areas	Northridge region, irregular boundary	Graves and Pitarka 2010

2.4 SOURCE PARAMETERS AND DISTANCE CALCULATION

Source parameters were compiled for each of the 32 new events. Most of the events have moment magnitudes ranging from 4.0 to 5.1, and as such finite fault effects are not considered to be significant for the derivation of site-to-source distances. For the Ridgecrest Earthquake sequence foreshock (**M**6.5) and mainshock (**M**7.1), finite fault effects were considered in distance calculations as described in Ahdi et al. (2020). Other than the Ridgecrest events, parameters compiled for each event include hypocenter location (latitude, longitude, depth), focal mechanism, moment magnitude, and rake angle. Focal mechanisms were assigned from rake angles (λ) as follows (e.g., Campbell and Bozorgnia, 2014):

- Reverse, λ = 30 to 150 deg
- Normal, λ = -150 to -30 deg
- Strike-slip, otherwise

Site-to-source distances were computed using the CCLD5 program that was updated as part of the NGA-Subduction project, as described by Contreras et al. (2020).

3 Basin Classification

A basin is a depression in the earth's surface filled by deep deposits of soft sediments that decrease in thickness towards their margins (Allen and Allen 2013). Two major types of basins are those formed in continental and oceanic settings. Further classifications have been proposed by Dickinson (1974, 1976) and Kingston et al. (1983) that consider tectonic setting (divergent, convergent, subduction) and the state of the deposited sediments (i.e., environment present at time of sediment deposition, which can change over time). Our objective is a simple and repeatable basin classification system useful for ground motion amplification modelling. Such classifications have not been provided in prior work, to our knowledge. We also develop V_{S30}-conditioned mean depth models in this chapter.

3.1 SOUTHERN CALIFORNIA STUDY REGION

The present research on basin response effects has focused on the southern California region shown in Figure 2.4. The approximate limits of the region are (from west to east) Ventura to Vidal Junction by the Colorado River and (from south to north) San Diego to Phelan. Several factors motivated our selection of this region:

- Ground motion data is abundant, both in terms of the number of earthquakes and the average number of recordings per event.
- The region spans a range of geological conditions, including regions with basins of different sizes and origins, and mountainous non-basin regions.
- There is a large body of work, spanning several decades, to develop seismic velocity models for the region's sedimentary basin structures (i.e., Magistrale et al. 2000; other documents cited in Table 2.1).

We have identified eight major basin structures within the study region, the approximate outlines of which are shown in Figures 2.4 and 2.5. These are the Los Angeles basin (LAB), the Ventura basin (VB), the San Fernando basin (SFB), the San Gabriel basin (SGB), the Chino basin (CB), the San Bernardino basin (SBB), the Coachella Valley basin (CVB), and the Imperial Valley basin (IVB).

The three western-most basins (Ventura, Los Angeles, and San Fernando) have experienced a complex evolution associated with the transformation of the southern California region from a convergent plate boundary to a transform plate boundary (Ingersoll and Rumelhart 1999). Intermittent uplift and subsidence of mountains and basin floors provided continental and oceanic sediment depositional environments. Moreover, the three basins were connected at some points in their history, later becoming separated by uplifts of the Santa Monica and Santa Susana Mountains in conjunction with formation of complex fault systems (Langenheim et al. 2011).

The San Gabriel basin lies between the Verdugo mountain range, the Sierra Madre fault, the Glendora Volcanics in the San Jose Hills, and the Cucamonga Fault Zone (Anderson et al. 2004; Yeats 2004). Surface geology within the San Gabriel basin and the orientation of alluvial deposits suggests that it may be a section of the western-most basins that was separated by uplift of the tertiary sedimentary rocks (Fernando formation, Monterey formation and other members of the Puente formation; Dibblee and Ehrenspeck 1998).

Some of the eastern basins within the study area (Chino, San Bernardino, Coachella Valley) are continental pull-apart/graben basins that formed as a result of regional faulting. The Chino and San Bernardino basins are adjacent alluviated lowlands with sediments deposited via erosion from the San Gabriel and San Bernardino Mountains. The San Jacinto fault zone defines the boundary separating these basins (Graves 2008).

The eastern-most basin, Imperial Valley, is distinct in its characteristics and origin. The basin was formed by a sequence of step-overs in the strike-slip plate boundary transform fault between the southern terminus of the San Andreas fault (roughly coincident with the north-east margin of the Salton Sea) and the Imperial Fault, which begins south of the Salton Sea. These step-overs produce crustal tension that is expressed as a down-dropped graben. The region is also affected by volcanic and geothermal activity associated with the local crustal spreading (Lizarralde et al. 2007; Alles 2011).

In summary, there are three general types of basins in the study region. Starting from the west, the first type are deep, formerly connected basins formed by intermittent subsidence and uplift (Ventura, Los Angeles, San Fernando). The second type have shallower depths and were formed as transform-graben induced valleys adjacent to uplifted blocks. The San Gabriel basin has attributes similar to the second type (its inland location and shallow depth) but also some geological similarity to the coastal type. Third, the Imperial Valley formed as a graben induced by a step-over in the plate boundary. As a result of these differences in geologic history, differences in site response might reasonably be expected. This hypothesis is tested through ground motion data analysis (Chapter 4) and the development of the basin effects model (Chapter 6).

3.2 CATEGORIZATION OF GEOMORPHIC PROVINCES

In this study, we investigate the impact of information beyond V_{S30} and sediment depth in the analysis of ground motions in basins. This requires a site categorization scheme to indicate whether a site is located within or outside of a basin. Because basin effects tend to occur at long periods, which is presumably related to the approximate alignment of long wavelengths with the large dimensions of many of these sedimentary structures, we consider basin size in the development of site categories.

The proposed categorization scheme is given in Table 3.1. Two of the categories are obvious – representing 'within basin' and 'outside basin' conditions (Categories #3 and #0, respectively). The valley category (#1) introduces a lateral dimension to the categorization. We considered Simi Valley (identified in Figure 2.4 as Box A, detail in Figure 3.1) to be an example of a sedimentary depression of modest dimension that should be differentiated from those of large dimension, like the Los Angeles basin. In consideration of this and other similar examples, we selected a limiting width of 3 km to differentiate basins (larger dimensions) from valleys (smaller dimensions). The basin edge category (#2) is intended to account for physical processes known to occur at basin edges, including basin edge generated surface waves (e.g., Graves, 1993; Graves et al. 1998; Kawase, 1996; Pitarka et al., 1998), and in some cases, focusing effects associated with lens-like structures (Baher and Davis, 2003; Stephenson et al., 2000). By differentiating basin edge sites from interior basin sites, we enable investigation of potential differences between ground motions in these domains.

Ground motion recording sites within the study area (i.e., Figure 2.4) were manually classified according to the categories in Table 3.1. These classifications are provided in an Electronic Supplement. The manual classifications were performed using terrain maps from Google Maps™, where visual assessments topography were made along with evaluation of the short dimension of the sedimentary structure and (as applicable) distance from edge. These classifications are admittedly subjective, although we sought to be as systematic as possible in the process.

Figure 3.2 (detail of Box B from Figure 2.4) shows an example of three sites comprising mountain-hill, basin edge, and basin conditions located near the northern edge of the San Fernando basin. The eastern-most site categorized as mountain-hill is located on an outcrop rock mass, while the western-most site categorized as basin is located within a region that is relatively flat. The basin edge site in Figure 3.2 is just west of an adjacent break in slope between the basin and non-basin areas. These classifications were relatively straightforward based on the differences in morphology in this region. Figure 3.3 (detail of Box C from Figure 2.4) shows a more ambiguous case, consisting of a mountain-hill site and several valley sites located in Riverside. The combination of basin and non-basin features (i.e., sites located in modestly-sized or narrow flat areas surrounded by rock outcrops or hills) at these sites partially motivated establishment of the valley category.

Figure 3.4 (detail of Box D from Figure 2.4) shows the downtown Los Angeles area consisting of three mountain-hill sites, three basin edge sites, and several basin sites. The mountain-hill sites in downtown Los Angeles are Puente formation materials (sandstone and siltstone) that were uplifted due to the folded blocks from the Elysian Park Anticline. Subsequent erosion and terrain morphological changes led to the formation of areas with steep relief relative to adjacent valley and basin locations. The valley sites shown in the northeast portion of Figure 3.4 are located in northeast Los Angeles and are adjacent to the Los Angeles River.

Table 3.1. Proposed geomorphic provinces for Southern California

Province	Description	Criteria	Cat. #	Number of Sites
Basin	Site location in basin interior	Basin width in short direction > 3 km	3	426
Basin Edge	Along basin margin	Within 300 m of basin edge[1]	2	75
Valley	'Small' sedimentary structure	Valley width in short direction < 3 km	1	178
Mountain-Hill	Sites without significant sediments, generally having topographic relief	Generally identified on basis of appreciable gradients and/or irregular morphology	0	329

[1] Basin edge defined visually from break in slope (topographic features)

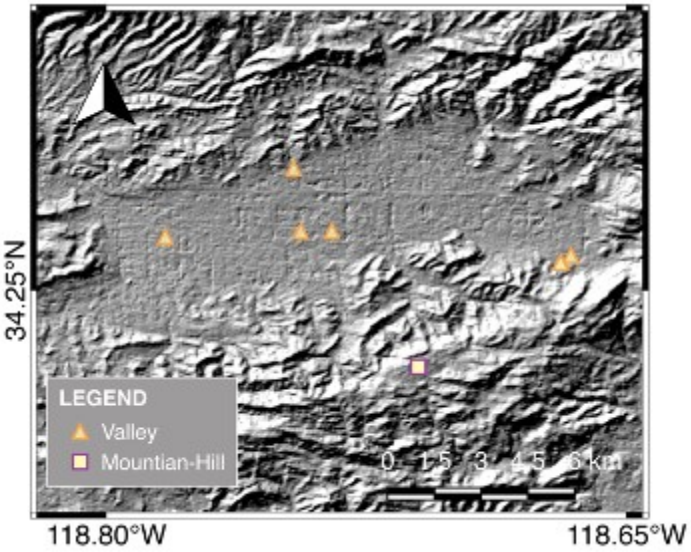


Figure 3.1. Simi Valley region (Box A in Figure 2.4)

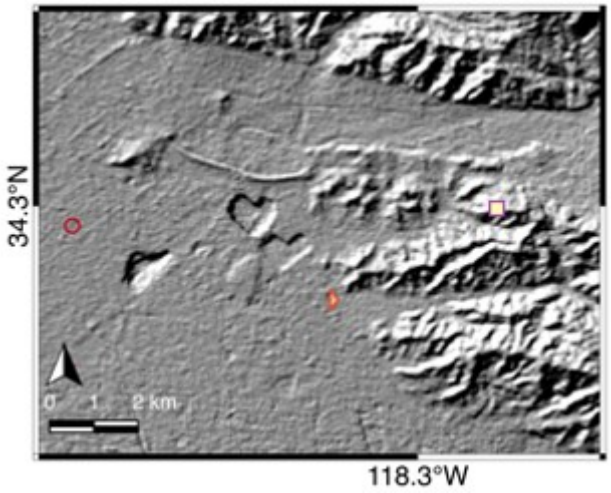


Figure 3.2. Example location in north-eastern San Fernando Basin with relatively unambiguous site categorizations (Box B in Figure 2.4).

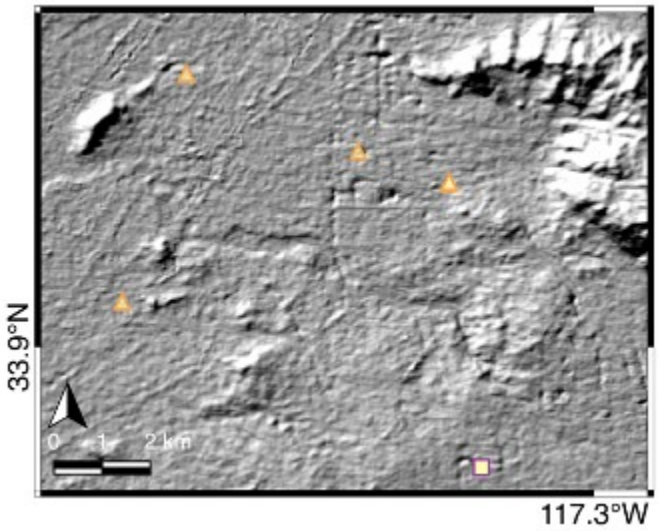


Figure 3.3. Example location in Riverside for which site classification was challenging (Box C in Figure 2.4).

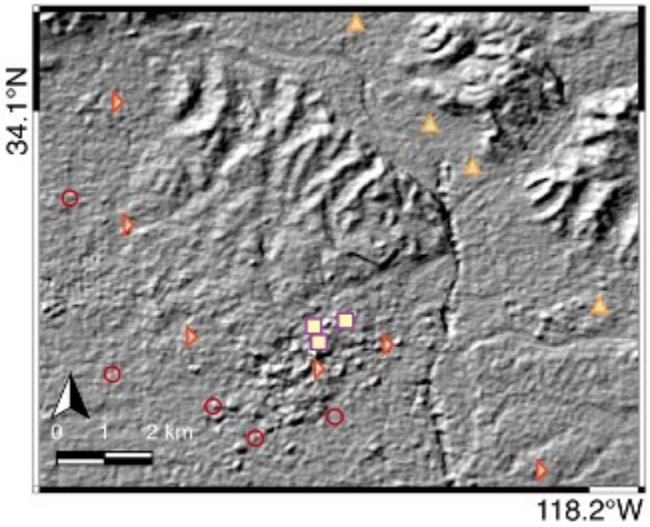


Figure 3.4. Detail map of Downtown Los Angeles (Box D in Figure 2.4).

3.3 GEOMORPHIC PROVINCE ASSIGNMENT GUIDELINES

The basin categorization process described in Section 3.2 was undertaken by interpreting morphology maps, and was also informed by our understanding of the regional geology. This process is necessarily interpretive and subjective. As such, it is not necessarily repeatable, which presents some challenges. First, as with any subjective process, different analysts might reach different conclusions on important details like the location of a basin boundary or the interpretation of a particular feature as a "basin" or "valley". Second, questions will naturally arise regarding the appropriate interpretation of other sedimentary structures outside of our study region (e.g., Central

Valley, Mojave, eastern Sierra Nevada ranges), for which assignments according to the classification scheme in Table 3.1 may be needed.

In this section, we develop an objective basin classification model using digital elevation model (DEM) terrain parameters. The approach we adopt is to develop a "training" dataset based on regularly-spaced grid points over the southern California region. We then classify each grid point as either basin, non-basin, or ambiguous. For the basin and non-basin grid points, we then use logistic regression to quantify the probability of a grid point being classified as basin or non-basin conditional on terrain parameters. We then further refine the model by mapping probability ranges to basins, mountain-hill, valleys, and basin edges based on the "test" dataset corresponding to ground motion sensor locations.

The morphology metrics considered were elevation, slope, curvature, and surface texture. These were evaluated using the 7.5 arc-sec Digital Elevation Model (DEM) (Danielson and Gesch 2011). Curvature was calculated as the second derivative of the surface elevation. Surface texture was calculated as the standard deviation of the ground surface slope for a radius of 1610 m around the site of interest, with each individual slope value determined using the Spatial Analyst Slope tool in ArcMap.

The training dataset consists of grid points spaced 4.78 km in the north-south direction and 11.34 km in the east-west direction, as shown in Figure 3.5. The subjective methodology described in Section 3.2 was used to assign to each grid point one of the following categories: basin, non-basin, or ambiguous. The "ambiguous" category is applied to sites that are basin edges, valleys, or some combination of those features as shown in Figures 3.1 - 3.4, and these points were not used in developing the basin classification model.

Figure 3.6 (top panel) shows examples of basin edge sites taken as ambiguous for the present analysis. These include areas such as Downtown Los Angeles (DTLA), Baldwin Hills (BH), Irvine (IRV), Pomona (PM), Woodland Hills (WH), and Alhambra (ALH). Figure 3.6 (top panel) also shows examples of valley sites taken as ambiguous in the eastern portion of the study area such as Riverside (RVS), Moreno Valley (MV), and Lake Elsinore (LKEL). Figure 3.6 (bottom panel) shows examples of both valley and basin edge sites taken as ambiguous in the region west and southwest of the Salton Sea. This area includes Ocotillo Wells (OW), the eastern margins of Borrego Springs (BRS), and Superstition Hills (SPH).

Logistic regression was used to evaluate the relationships between these morphological parameters and classification of sites as either basin or non-basin. Using maximum likelihood estimation, logistic regression determines an unknown probability for a given linear combination of independent variables. In the case of basin categorization, the estimated probability indicates the likelihood that a site is in either the basin [p(basin) = 1 or 100%] or non-basin [p(basin) = 0 or 0%] categories. As previously stated, ambiguous sites are not included in the logistic regression, but the resulting model is evaluated for effectiveness on a "test" dataset consisting of ambiguous site assignments.

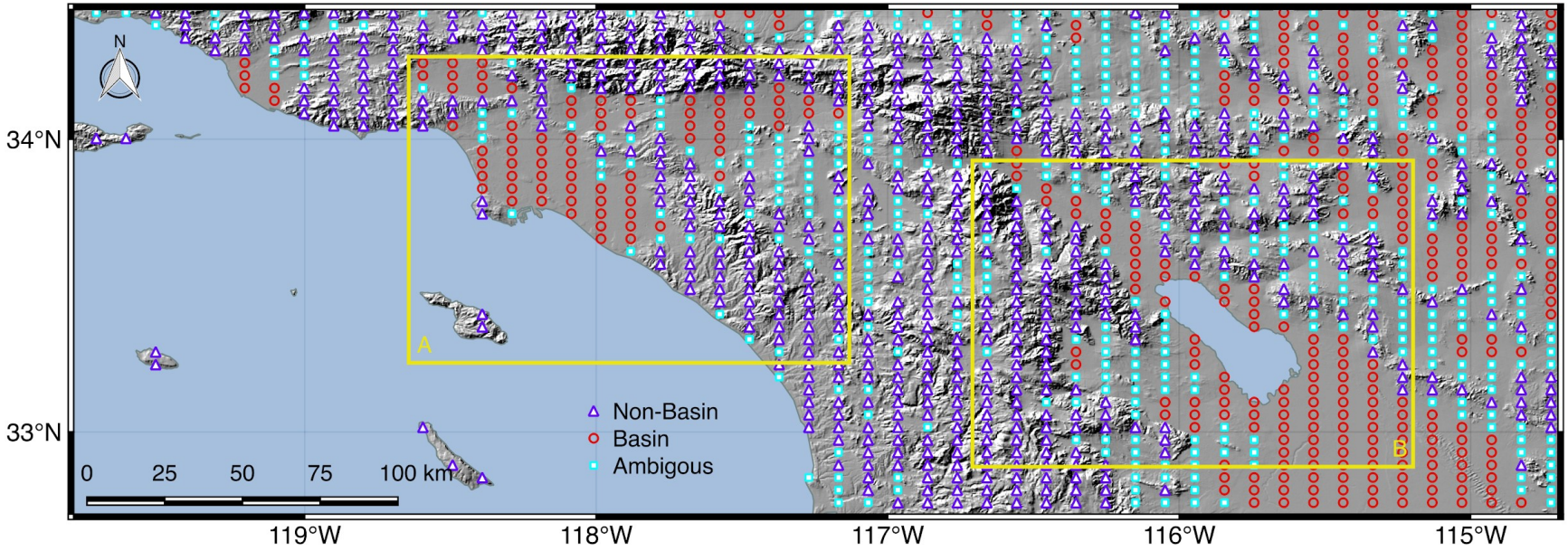


Figure 3.5. Map of categorized sites in the study region for basin algorithm development. Box A and B are detailed in Figure 3.6

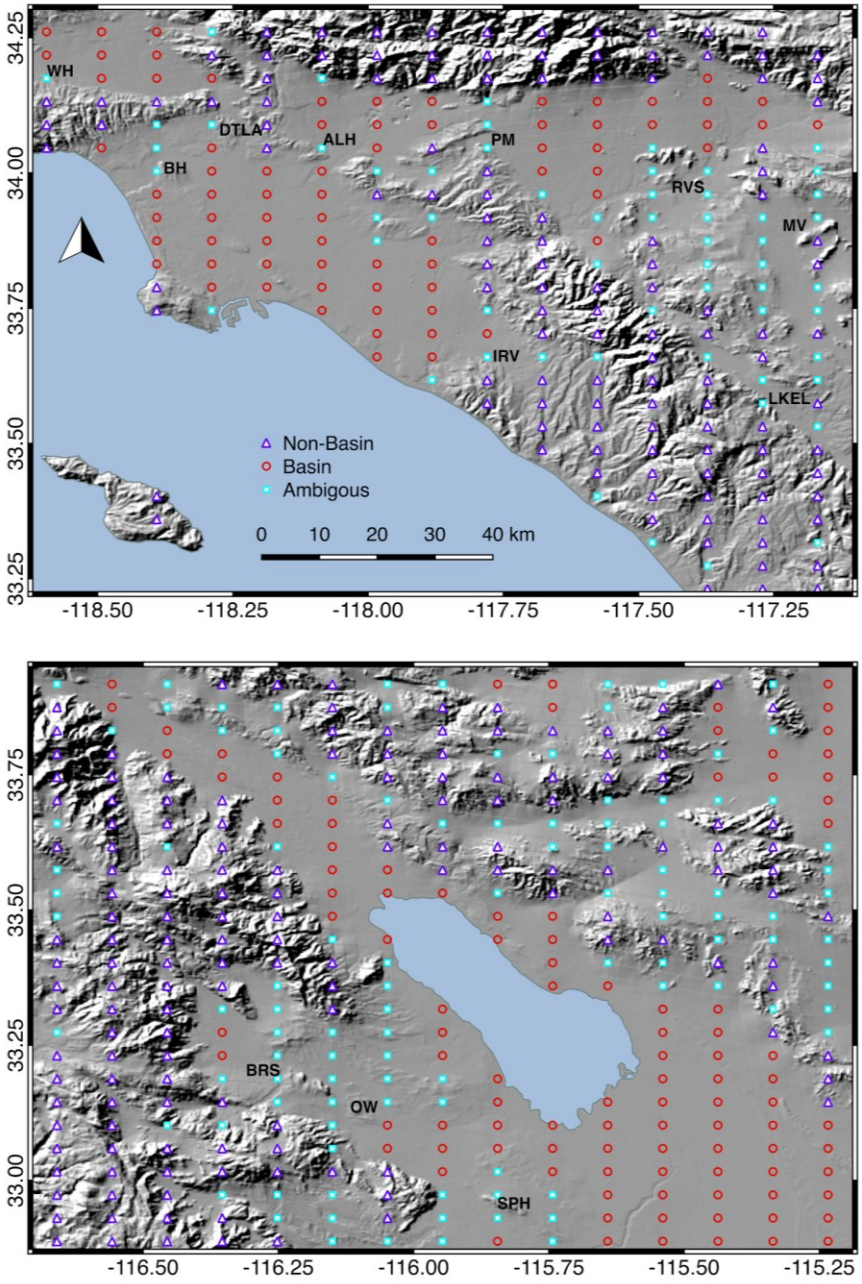


Figure 3.6. Detailed map of basin classification of sites used for algorithm development. Top panel (Detail Box A from Figure 3.5) focuses on the Los Angeles basin, San Bernardino basin and adjacent basins. Bottom panel (Detail Box B from Figure 3.5) focuses on the Imperial Valley basin and adjacent regions.

Multicollinearity (i.e., high correlation among predictor variables) can cause unreliable estimates of regression coefficients. Topographic morphology parameters do not meet this requirement (i.e., they are correlated; Iwahashi and Pike 2007), so we do not use the parameters in combination, but individually. In effect, we seek to identify the morphological parameter best suited to basin classification. Figure 3.7 shows distributions of the morphological parameters for basin, non-basin, and ambiguous categories. Surface texture (Fig. 3.7d) most clearly delineates non-basin from basin categories, as quantified by the small overlap between the basin and non-basin distributions, and is therefore considered to be the best predictive parameter for basin categorization.

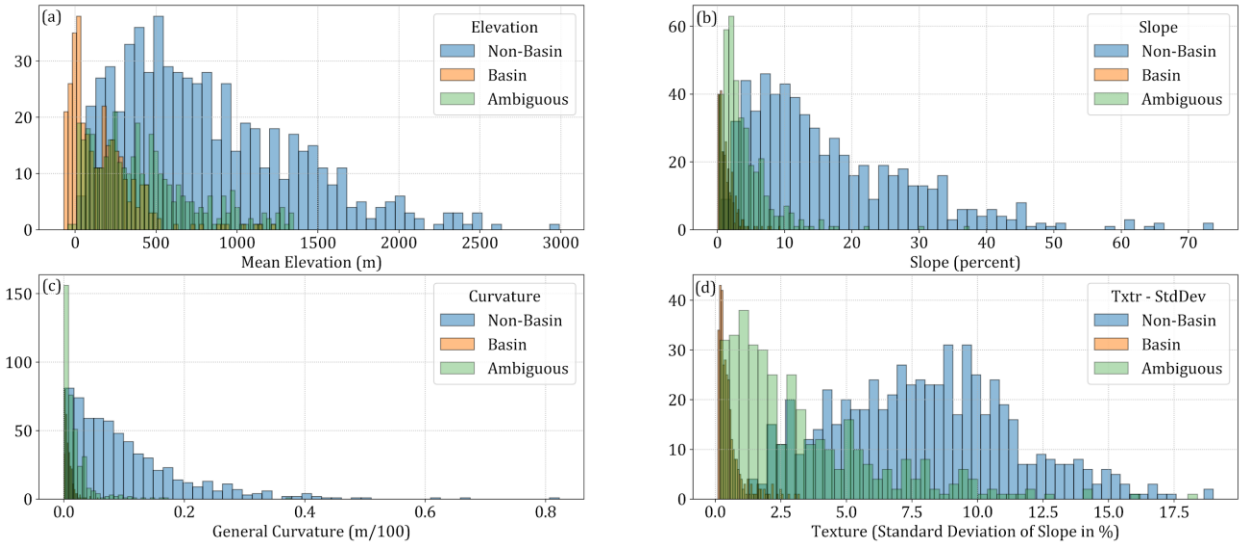


Figure 3.7. Distribution of the morphological data queried at the grid coordinates partitioned by basin category (basin, non-basin).

We use logistic regression to derive a probability of a site being in the basin category based on texture, $\mathcal{T}$, in units of percent. This regression only considers sites that are basin or non-basin (ambiguous sites are removed). It is expected that regression will indirectly identify ambiguous sites by estimating intermediate probabilities between 0.0 and 1.0. The functional form of the regressed model is:

$$P(basin) = \frac{1}{1+e^{\left(-(\beta_0+\beta_1 \cdot \mathcal{T})\right)}} \tag{3.1}$$

Where β_0 and β_1 are regressed coefficients and P is given as a fraction (not percent). Values for the coefficients are given in Table 3.2.

Table 3.2. Basin Categorization Algorithm coefficients (Eq. 3.1)

Coefficient	Value	95% Confidence Interval
β_0 (dimensionless)	6.8537	[5.234 , 8.473]
β_1 (1/%)	-3.3605	[-4.139, -2.582]

We applied the logistic regression model from Eq. 3.1 to the training dataset defined in Figure 3.5 (without the ambiguous sites) to assess and verify its capabilities. Table 3.3 shows a confusion matrix for the model on the training data, which serves as a performance measure. The number of true positives (correctly estimating basin sites) and true negatives (correctly estimating non-basin sites) indicates generally good performance. This is supported by the performance scores shown in Table 3.4. The overall model accuracy is high at 98%. The precision for both classes (basin and non-basin) are high, as indicated by the sensitivity (i.e., basin recall rate; how often it predicts basin for basin cases) and specificity (i.e., non-basin recall rate; how often it predicts non-basin for non-basin cases), which both have values above 97%. These scores relative to the training dataset serve as a verification that the regression model from Eq. 3.1 is a capable classifier for the basin and non-basin geomorphic categories in Table 3.1.

Table 3.3. Confusion Matrix for the training dataset comparing predicted basin categories with the manually assigned station categories stations from Figure 3.5.

	Predicted Positive	Predicted Negative
True Positive	621	8
True Negative	11	322

Table 3.4. Basin algorithm accuracy evaluation for the training dataset in Figure 3.5

	Basin	Non-Basin
Precision	0.98	0.98
Recall	0.97	0.99
Accuracy	0.98	

We now turn our attention to assessing the performance of Eq. 3.1 by applying it to the "test" dataset, defined as the locations of the ground motion records defined in Figure 2.4. Table 3.5 shows the confusion matrix for the model on the test data for the basin and mountain-hill categories, and Table 3.6 shows the performance scores. The number of true positives and true negatives are high. The precision, sensitivity, specificity, and overall accuracy are also high with values above 93%. This indicates that the logistic regression model from Eq. 3.1 is a good

geomorphic classifier as it is capable of correctly categorizing sites as basin and mountain-hill at a high rate.

Table 3.5. Confusion Matrix for the test dataset comparing between predicted basin categories and the categorized stations from Figure 2.4.

	Predicted Positive	Predicted Negative
True Positive	381	27
True Negative	22	280

Table 3.6. Basin algorithm accuracy evaluation for the test dataset in Figure 2.4

	Basin	Non-Basin (Mountain-Hill)
Precision	0.95	0.93
Recall	0.93	0.91
Accuracy	0.93	

Figure 3.8 shows the geospatial basin probability map that results from application of Eq. 3.1. Stations in the study region from Figure 2.4 are overlaid on the probability map in Figure 3.8. The shading in Figure 3.8 ranges from basin to non-basin probabilities, where basins (and basin related features like basin edges) have probabilities of 0.5 or higher, as listed in Table 3.7. Whereas non-basins have probabilities of 0.25 or lower. For probabilities between 0.25 and 0.5, we suggest the use of judgement to discern between valley, basin edge, and non-basin as the assigned category.

Table 3.7. Basin Category Probability Ranges

Category	**Description**	**Basin Probability Range**
Basin	Site location in basin interior	$0.5 < P < 1$
Non-Basin (Mountain-Hill)	Sites without significant sediments, generally having topographic relief	$0 < P < 0.25$

We justify the assignment of the probability ranges to the basin categories (as listed in Table 3.7) based on qualitative and quantitative analyses as previously outlined in part through Tables 3.3 - 3.6. From a qualitative perspective, the category classification for the stations in Figure

3.8 generally seem to align with the predicted basin categorization from the algorithm. The quantitative assessment above using confusion matrices and performance scores show that the model is an adequate classifier for basin and non-basin categories. The probability ranges for the ambiguous sites such as basin edge and valley categories are assigned based on the possible prediction gradient with basin edges closer to basins (higher probability of being a basin location) and valley closers to mountain-hill (lower probability of being a basin location). However, the model shows a reduction in performance when classifying ambiguous sites using the intermediary probability ranges from Table 3.7.

From Figure 3.8 (detail map), it is evident that the algorithm exhibits difficulties in assessing sites on cemented playas (located immediately southwest of Baldwin Hills, “BH”) and complex valley systems such as in the Glendale (GLD) area north of Downtown Los Angeles. As a result, we advise that the proposed algorithm (Eq. 3.1) be used solely to categorize basin and non-basin sites, and judgement be applied for intermediary categories such as valleys and possibly basin edges.

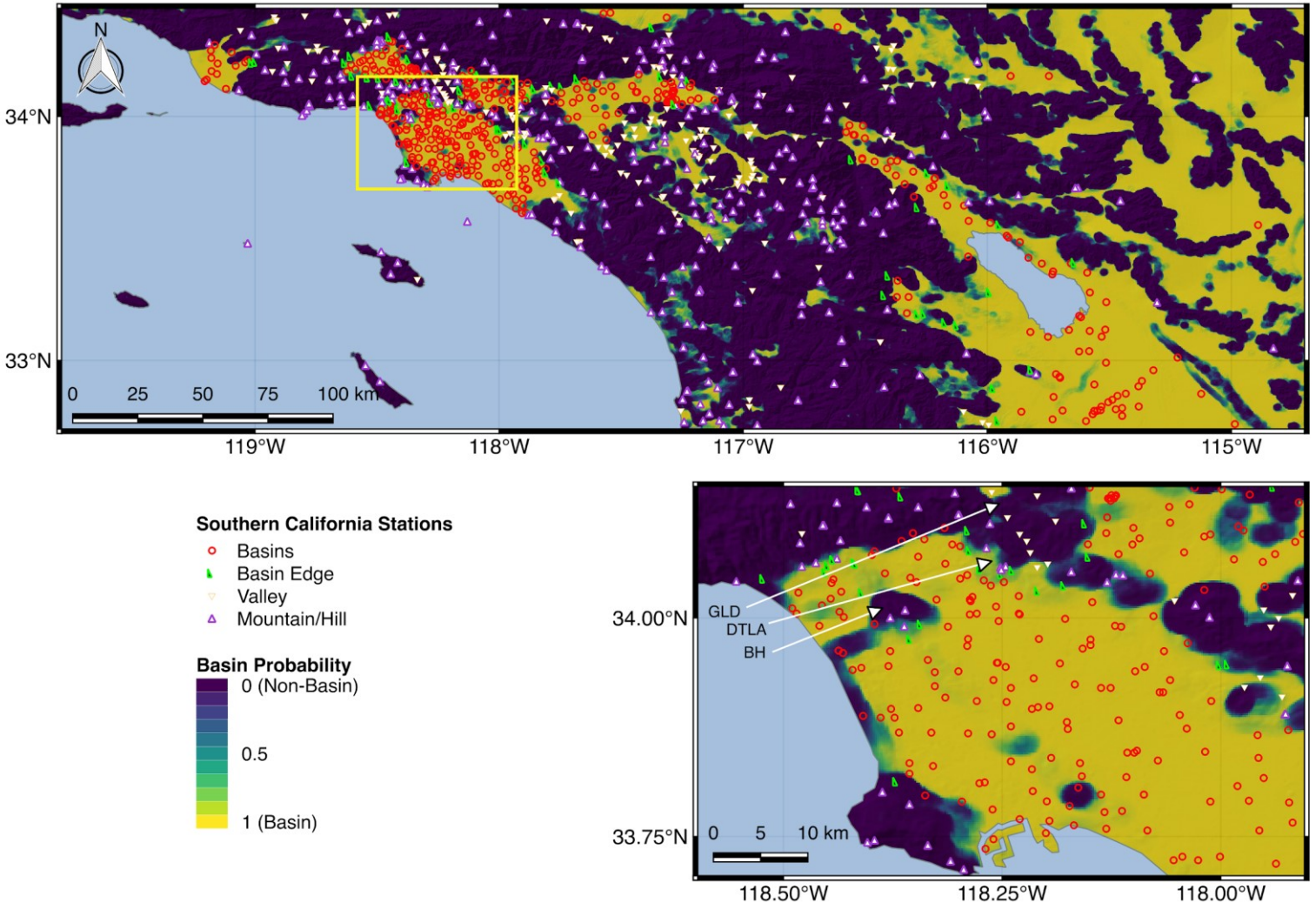


Figure 3.8. Basin prediction map in the form of a geospatial probability distribution based on Eq. 3.1. Categorized stations from Figure 2.4 are overlain to evaluate the algorithm performance. Second panel is a detailed enhancement of the yellow box in the larger panel. That highlights the performance of the basin classification algorithm through overlap comparison

3.4 MEAN DEPTH MODEL FOR SOUTHERN CALIFORNIA

Basin depth and V_{S30} are correlated, which introduces potential for regression problems associated with multicollinearity (e.g., Curtis and Ghosh 2011) since least squares regression assumes predictors are uncorrelated. To reduce multicollinearity we apply a "centering" procedure to define the mean value of basin depth, $\overline{z}_{1.0}$, conditional on V_{S30}, and subsequently compute the difference between basin depth and mean basin depth (i.e., $\delta z_{1.0} = z_{1.0} - \overline{z}_{1.0}$), and use $\delta z_{1.0}$ in regression. Current relations for the mean depth in ground motion models apply for broad regions (California, Japan) and have large scatter. In addition, the basin depth values used to develop the current relations for mean depth were estimated from an older version of CVM-S4 (v4). There is an updated version of CVM-S4.26.M01 and it is one of two primary velocity models for Southern California with the other being CVM-H v15.1 as shown in Table 2.1. Therefore, we define here two southern California-specific centering models.

After updating basin depths within the southern California study region using both CVM-S4.26.M01 and CVM-H v15.1, there were a number of sites with $z_{1.0}$= 0. A few of these sites were located in areas that contradicted the assigned zero basin depth value (basin margins, valleys, soft rock). These sites were identified by screening for zero basin depths located on surface geology known to not consist of hard rock. These included Quaternary alluvium (Holocene, Pleistocene, or Undivided), and Tertiary sedimentary rocks. The basin depths were then updated for a subset of the aforementioned screened sites by projecting outcrop slopes below the ground surface as a means to estimate a relative depth of the deposited sediments. This applied solely to sites located on basin margins adjacent to outcrops with appreciable slopes. Sites that did not fall into this category were assigned a value of -888 which indicates that estimates of ground motion using the GMM should be computed without the contributions from the basin component. Zero basin depth sites located on older formations (Cretaceous), volcanic rock, or crystalline bedrock were accepted without adjustments.

In the Imperial Valley basin, it was discovered that the estimates of basin depth $z_{1.0}$ derived from the current CVM-S4.26.m01 model deviate significantly from prior version (CVM-S4 v4.). While newer versions would generally be preferred, we noticed that CVM-S4 v4 produced depths that are more similar to depths from CVM-H (v15.1 and v11.1). For example, some sites in Imperial Valley previously assigned $z_{1.0}$ = 1400 m are currently estimated to have $z_{1.0}$= 200 m. As a consequence of this, as well as conversations with experts on different basin models (R. Graves, personal communication, 2020), we elected to not use CVM-S4.26.M01 $z_{1.0}$estimates for the Imperial Valley basin area. Instead, the CVM-S4 v4 values in Imperial Valley are taken from the NGA-West2 database where available and are unassigned for newer sites (given a value of -888 in database).

Figure 3.9 a-b shows the distribution of site data in V_{S30} - $z_{1.0}$space for the adjusted CVM-S4 estimates and CVM-H v15.1, respectively, with data from particular basins (as defined in

Section 3.2) delineated with colored markers. This figure also includes depth-V_{S30} points (in gray) for sites that are not inside of basins as defined in Chapter 3 as long as an actual depth value is available for the location. Figure 3.10a-b are nearly identical renditions of the prior two figures but with sites delineated by geomorphological category instead of basin label. Also shown in both figures is the model for predicting $z_{1.0}$ given V_{S30} proposed by Chiou and Youngs (2014), which was developed to apply for all of California, including San Francisco Bay Area sites. The functional form for the Chiou and Youngs (2014) model is,

$$ln(\overline{z}_{1.0}) = \upsilon_0 ln\left(\frac{V_{S30}^4+\upsilon_1^4}{1360^4+\upsilon_1^4}\right) - ln(1000) \quad (3.2)$$

Where V_{S30} is in m/s and $z_{1.0}$ is in km. The coefficients recommended by Chiou and Youngs (2014) are listed in Table 3.8.

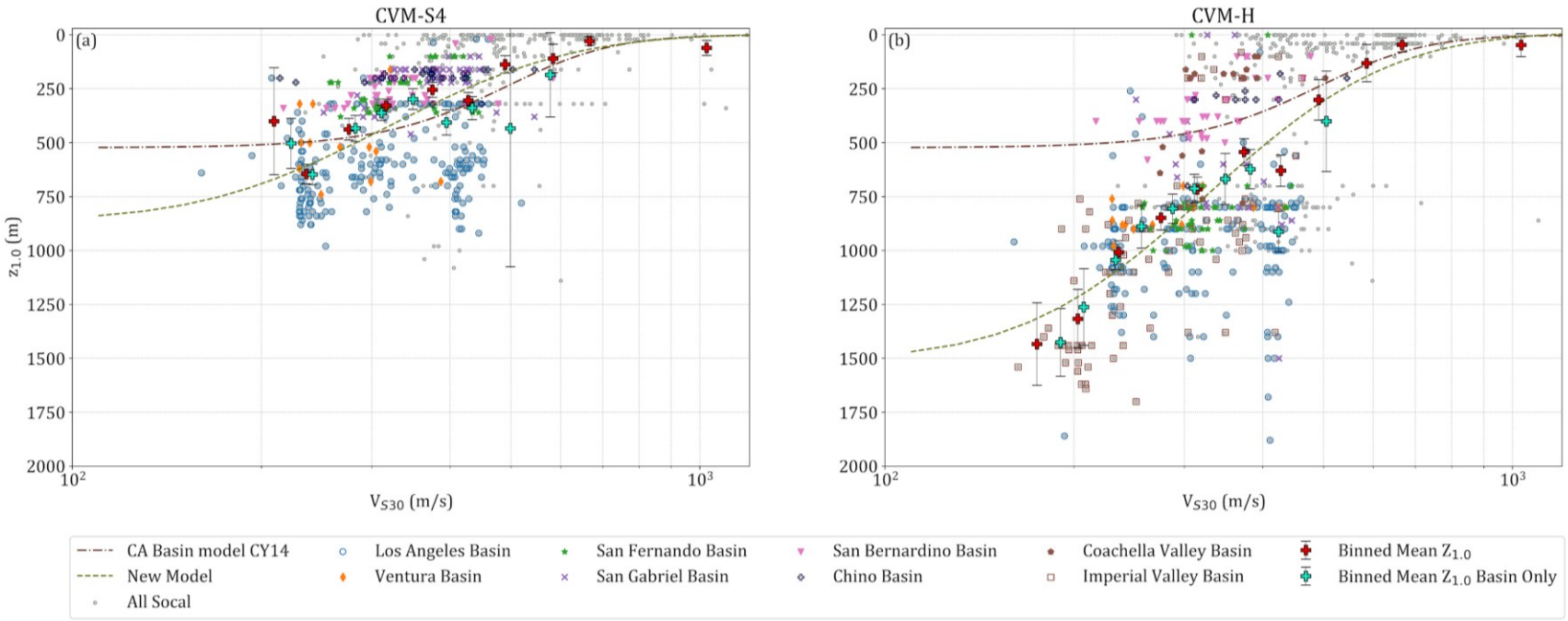


Figure 3.9. Distribution of southern California data in V_{S30}-$z_{1.0}$ space, including the model for the relationship between these parameters by Chiou and Youngs (2014) for California and proposed in Eq. 3.3 for southern California delineated by basin label.

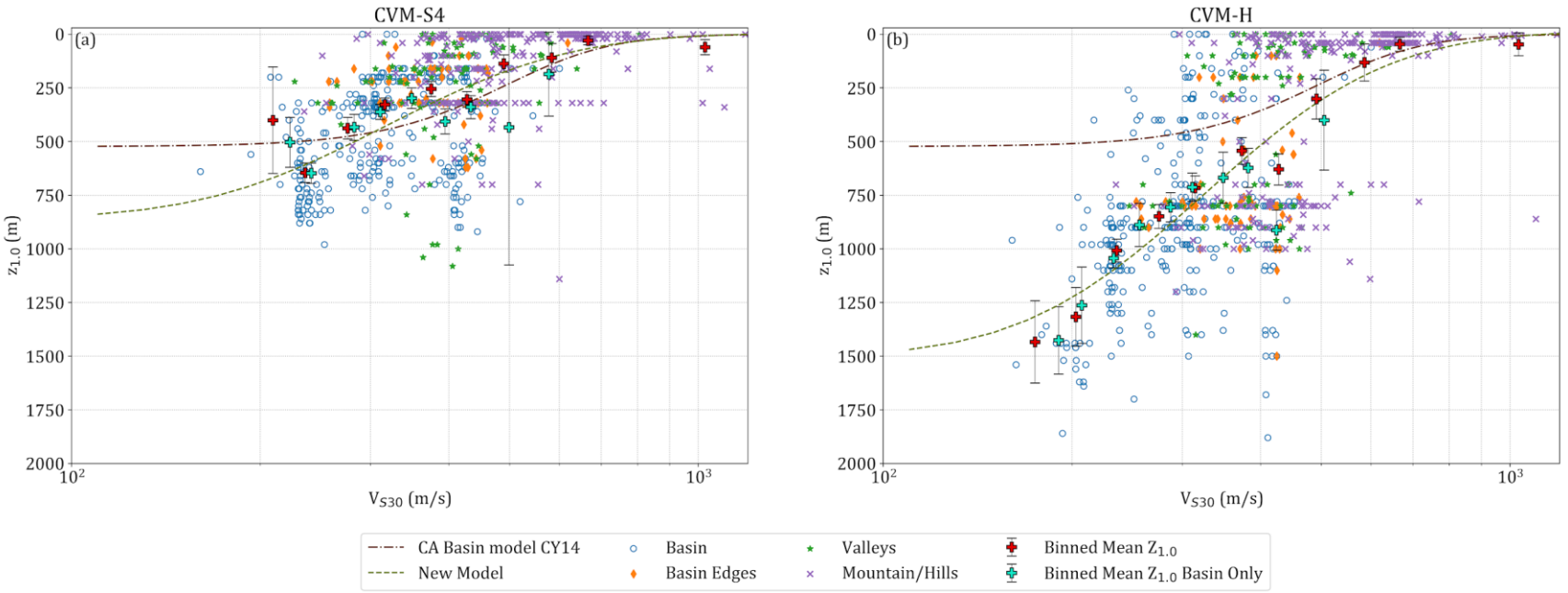


Figure 3.10. Distribution of southern California data in V_{S30}-$z_{1.0}$ space, including the model for the relationship between these parameters by Chiou and Youngs (2014) for California and proposed in Eq. 3.3 for southern California delineated by basin geomorphological category.

Table 3.8. Basin depth predictive model coefficients (Eqs. 3.2-3.3)

Parameters (CY14)	Value	Parameters (this study - CVM-S4)	Value	Parameters (this study - CVM-H)	Value
v_0	-1.7875	c_0	0.8596	c_0	1.502
v_1 (m/s)	570.94	c_1	-0.4324	c_1	-0.7582
		υ_μ (m/s)	317.46	υ_μ (m/s)	325.74
		υ_σ	0.2334	υ_σ	0.2324

Binned mean markers are shown in Figure 3.9 and 3.11 both for the data set as a whole and the basin-only sites. As expected, the basin-only binned means have larger depths, although the differences are modest for V_{S30} < 400 m/s. Comparing either set of binned means to the model, it is apparent that the increase in depth as V_{S30} decreases is stronger in the southern California data than in the Chiou and Youngs (2014) model. Accordingly, we sought to develop an improved fit to the data by fitting the trend shown by the "all sites" (red) binned means while also enforcing physical bounds at the limits, where the depth scaling should flatten with respect to V_{S30}. We suggest the following function to provide the desired shape:

$$\overline{z}_{1.0} = c_1 \left[1 + erf\left(\frac{log(V_{S30}) - log(\upsilon_\mu)}{\upsilon_\sigma \sqrt{2}}\right)\right] + c_0 \qquad (3.3)$$

where υ_μ defines the center of the scaling relationship where the slope is steepest and υ_σ defines the width of the ramp. Eq. (3.3) returns the mean $z_{1.0}$ in units of km. Values for all coefficients are given in Table 3.8. The *erf* function can be solved for in most numerical software packages. In Excel, *erf*(x) is given by ERF(x).

The fit of the proposed models to the southern California data is shown in Figure 3.9 - 3.11. Even with this improved fit, it is apparent that the model fits some basins better than others. In the case of CVM-S4.26.M01, the fit is good for the relatively deep near-coast basins (Los Angeles, Ventura) and Coachella, whereas mean depths are generally larger than actual depths for inland basins (San Fernando, San Gabriel, San Bernardino, Chino, and Imperial Valley). Similarly, the fit is good for all basin sites, but mean depths are larger than actual depths for basin-edge sites and valley sites. Some mountain/hill sites (and a portion of valley sites) have estimated depths that we consider less accurate; this occurs in regions where sedimentary basin models are limited and do not capture the geologic complexity (smaller features and variations), including some mountainous

areas, portions of the Mohave desert, some portions of the Inland Empire (e.g., Anza and Moreno Valley areas), and the Imperial Valley region.

In the case of CVM-H v15.1, the fit is good for the relatively deep near-coast basins (Los Angeles, Ventura), San Fernando, and Imperial Valley, whereas mean depths are larger than actual depths for the inland basins located in-between fault systems (San Bernardino, Chino, San Gabriel, and Coachella). The fit with regards to basin categories is good except for valleys and mountain/hill sites with most of their sites partitioned between approximately 100 m and 800 m.

The significant depths reported for some mountain/hill sites in both velocity model V_{S30}-$z_{1.0}$ plots may be attributed to the weak/soft rocks that are prevalent in the southern California region. In particular, areas such as Palos Verdes Hills, Baldwin Hills, Glendale Hills, Chino Hills, and others are composed of uplifted weak sedimentary rocks that may have relatively low shear wave velocities and thus large $z_{1.0}$ values.

3.5 CONFIDENCE IN BASIN DEPTH PARAMETERS

The basin depth parameters used in this research are taken from CVM-S4 and CVM-H (Table 2.1). These versions of the basin models are updated from relatively familiar versions of basin models that were used in the NGA-West2 project (CVM-S4.0, CVM-H v11.1). In the case of the CVM-S4 model, in most cases, the updates from the earlier version to the current version produced no change or small changes in basin depths $z_{1.0}$. However, in the case of the Imperial Valley Basin (IVB), there were substantial and consistent reductions of basin depths across the inventory in the CVM-S4 model, as shown in Figure 3.11. The updated basin depths appear to be unreliable – given the geology of the basin we anticipate large depths, which are no longer encountered. As a result of this feature of the models and the large between-version changes, we lack confidence in the CVM-S4 model for IVB and CVB, where similar problems were encountered. This problem is not encountered with the CVM-H model. As a result, site amplification models for IVB and CVB are only derived using CVM-H and not CVM-S4.

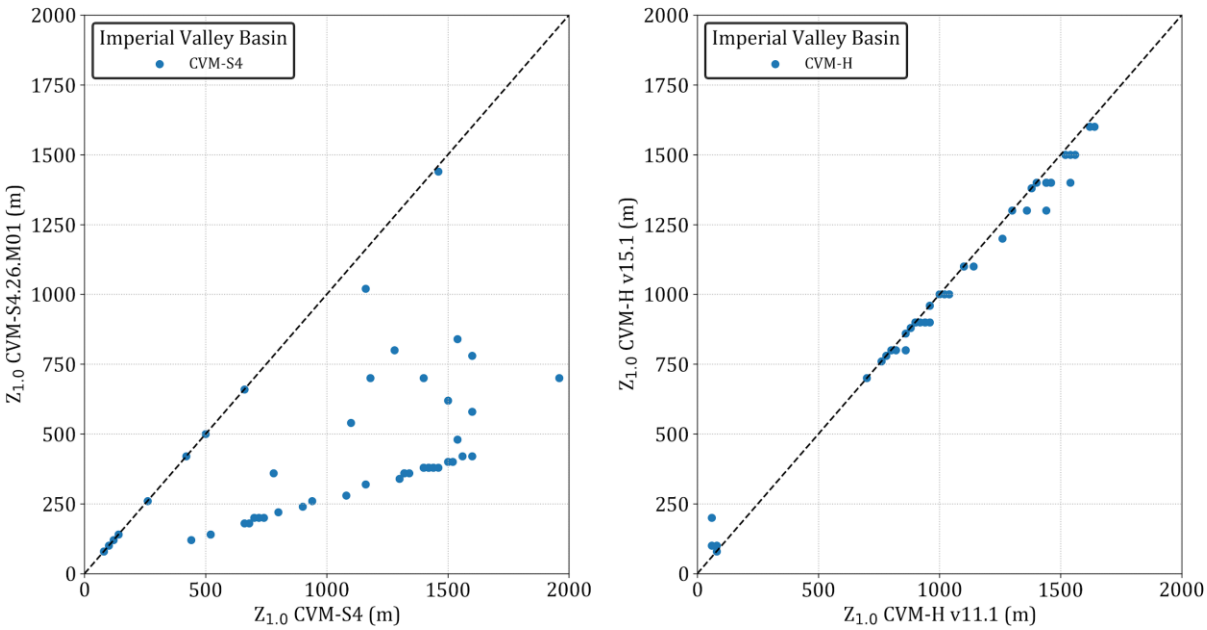


Figure 3.11. Comparison of basin depths from version used in NGA-West2 project and current version for IVB. The depths are generally consistent for CVM-H but have decreased significantly in CVM-S4.

4 Ground Motion Analysis

Ground motion analyses were undertaken with the objective of evaluating the ability of the proposed geomorphic site categories (Table 3.1) in combination with differential depth to explain site response effects beyond that which is captured in V_{S30}-scaling models. These analyses use the database presented in Chapter 2 and the mean depth models in Section 3.4. We apply data analysis procedures that estimate site response effects from recorded ground motions. Our approach is an adaptation of the classical non-reference site approach (Field and Jacob, 1995) in which the "true" (or non-ergodic) site response is computed. This chapter focuses on residual analyses of the data set relative to an NGA-West2 GMM. The objective is to ensure that source and path models are unbiased, because if such biases were present, they would map into site response estimates. An earlier version of the analyses here was presented by Nweke et al. (2018).

4.1 DATA SELECTION

We use a subset of the NGA-West2 database applicable to events in the southern California region as well as the newly added data developed in the present study, as shown in Figure 2.1. Using this subset of events, we apply the data screening criteria of Boore et al. (2014). Particularly important elements of those criteria include (1) the use of magnitude and instrument-dependent distance cut-offs that are intended to minimize sampling bias and (2) only using recordings over their usable oscillator period range as described in Section 2.3.

As shown in Figure 2.3, the data set spans a magnitude range of about 3 to 7 and a closest distance range of about 1 to 600 km. The range of V_{S30} is about 160-800 m/s and the range of $z_{1.0}$is about 0 to 2000 m, with basin depths derived from CVM-H being deeper on average.

4.2 RESIDUAL ANALYSIS PROCEDURE AND LIMITATIONS

The difference between a recorded ground motion intensity measure and a model prediction is referred to as a residual, R:

$$R_{ij} = ln(Y_{ij}) - \left[\mu^{r}_{ln,ij}(\mathbf{M}_i, F_i, (R_{jb})_{ij}) + F_S\right] \quad (4.1)$$

where index i refers to an earthquake and index j refers to a particular station that provides a recording. The quantity Y_{ij} is a ground motion observation expressed as an intensity measure. The term $\mu^r_{ln,ij}$ is the mean ground prediction for reference rock site conditions in natural log units from a GMM. We use the BSSA14 GMM, which has the arguments listed in Eq. 4.1, where F is a style of faulting parameter (reverse, strike-slip, etc.), R_{jb} is the Joyner-Boore distance, and other parameters are as defined previously. The term F_S is a site amplification model. For the residuals analyses presented below (Section 4.3) the amplification model is modified as in Chapters 5-6 for southern California sites. For all other regions, the site amplification model as given in BSSA14 is used.

Non-zero residuals occur for a variety of reasons. A portion of the data-model differences are purely random, having no known associations. Other portions of the residuals are more systematic. For example, the ground motions for a particular event or a particular site may be systematically high or low relative to the global average. These systematic differences are referred to as event terms and site terms, η_E and η_S, respectively. As a result of these systematic effects, residuals can be partitioned as:

$$R_{ij} = \eta_{E,i} + \delta W_{ij} \tag{4.2}$$

where δW_{ij} is the within-event residual, which can be further partitioned as,

$$\delta W_{ij} = \eta_{S,j} + \varepsilon_{ij} \tag{4.3}$$

where ε_{ij} is the remaining residual when the event and site terms have been removed. Recalling the standard deviation terms from Chapter 1, the dispersion of η_E is τ_{ln}, the dispersion of δW_{ij} is ϕ_{ln}, the dispersion of η_S is ϕ_{S2S}, and the dispersion of ε_{ij} is $\sqrt{\phi^2_{P2P} + \phi^2_{lnY}}$. The P2P term appears because we are not accounting for non-ergodic path effects as investigated for example by Landwehr et al. (2016) and Kuehn et al. (2019).

Event and site terms are computed using mixed effects analyses (Gelman et al. 2014) :

$$R_{ij} = c_k + \eta_{E,i} + \eta_{S,j} + \varepsilon_{ij} \tag{4.4}$$

where c_k is an overall model bias for ground motion model k. For a given intensity measure, the mixed effects analysis provides estimates of c_k, η_E for all events, and η_S for all sites. The mixed-effects analyses are performed in *R* [R Core Team, 2019; Bates et al. 2015] as a subroutine exercised from the main coding script in Python [Gautier 2009].

The effectiveness of the partitioning process represented by Eq. 4.2 and Eq. 4.3 at distinguishing source, path, and site effects depends on the ability of the underlying ground motion model to capture data trends in the study region. For example, Figure 4.1 shows distance attenuation trends of site-corrected data (to a reference condition of 760 m/s) from events with attenuation rates that do and do not follow the trend from a path model. Where the trend is followed but the data have a mean shift as in Figure 4.1(a), the event term represents a source effect. Where the data and model path trends are different as in Figure 4.1(b), the event term represents a combination of source and path effects. To the extent that different earthquakes may have different attenuation rates, this causes the τ_{ln} term to have some "contamination" from variable path effects. In a similar manner, event-to-event variations in path terms that are not captured by the GMM would be expected to contaminate site terms derived using Eq. 4.3, and the site-to-site variability term ϕ_{S2S}. The magnitude of these contamination effects, while not formally quantified, are expected to be modest provided that the underlying GMM's path term produces a globally (i.e., across all events in the region) unbiased representation of the distance attenuation of ground motion. We check this feature in the next section.

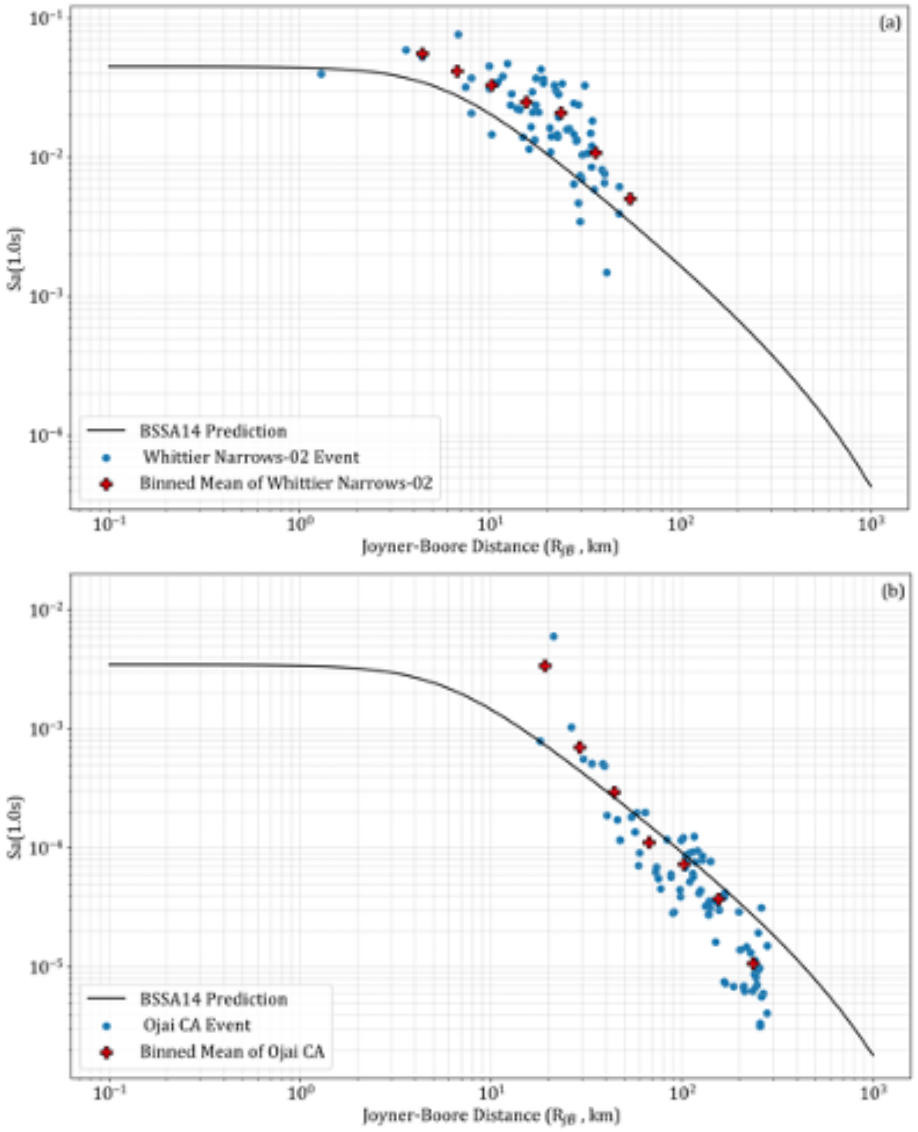


Figure 4.1. Path attenuation of S_a (1.0 s) data following site-correction to 760 m/s (using Seyhan and Stewart 2014 model) as compared to global ergodic GMM (BSSA14): (a) **M** 5.27 1987 Whittier-Narrows, CA event and (b) **M** 4.13 2016 Ojai, CA event.

4.3 RESIDUAL ANALYSIS RESULTS

In this section, we investigate the performance of the GMM with respect to the database developed for the study region (Chapter 2). The goal is to check model bias through the c_k term, the scaling of event terms with magnitude, and the scaling of within-event residuals with distance. What is most crucial for the analyses that follow in Chapters 5 and 6 is lack of bias in the path model, for reasons described previously.

Figure 4.2 plots bias term c_k with period. The data used in the plot is the NGA-West2 data set with the new southern California data added. As described in Section 4.2, the GMM that is used is BSSA14 with the site amplification model modified as in Chapters 5-6 for southern California sites. In the process of deriving the new site amplification model, residuals were also obtained using the BSSA14 site amplification model for all sites; the results are qualitatively similar to those presented here.

Two results are shown corresponding to basin depths derived from the CVM-S4 and CVM-H models. The bias is non-zero because the data used in the residuals analysis is not the same as the data used to derive the GMM. Nonetheless, the bias terms are small (absolute values generally less than 0.05), which indicates general compatibility of the model with the data.

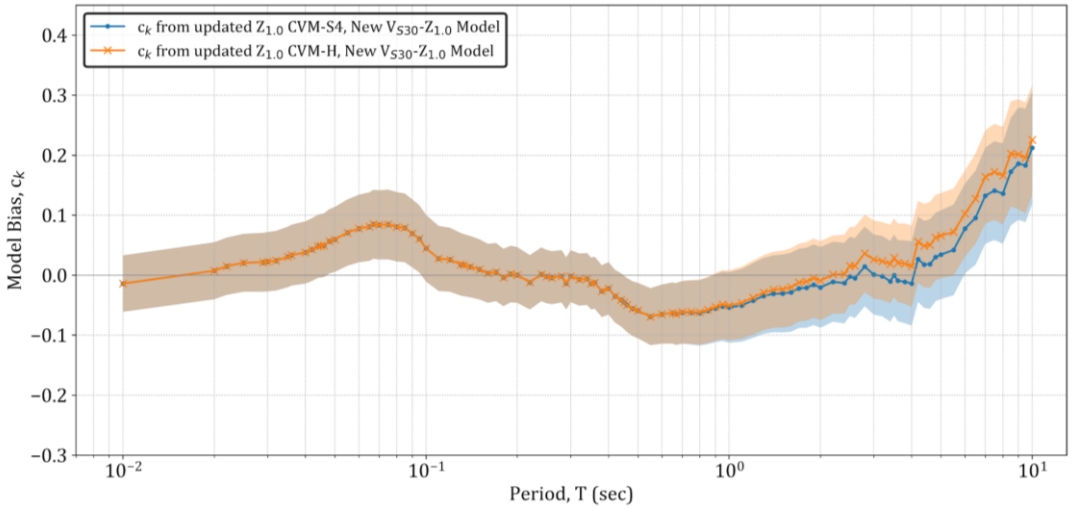


Figure 4.2. Model Bias and the 95% confidence interval vs period using data set from Chapter 2. The two trends represent basin depth estimate differences from CVM-S4 and CVM-H.

Figure 4.3 shows event terms as a function of period, both for the full NGA-West2 data set and the subset of data from southern California. The mean is consistently close to zero as a function of period. This is required by mixed effects analysis for the data set as a whole. However, the mean for the subset of southern California data is slightly positive, indicating that there is a subtle regional bias relative to the GMM. Figure 4.4 shows event terms as a function of magnitude for PGA and $S_a(3.0)$ (pseudo-spectral acceleration at a period of 3.0 sec). There is no perceptible trend for $\mathbf{M} > 4$ events, although an upward trend exists for $\mathbf{M} < 4$. The general lack of trend in the event terms in Figure 4.4 indicates that the magnitude scaling model in the GMM captures the data trends for the southern California database used in this study.

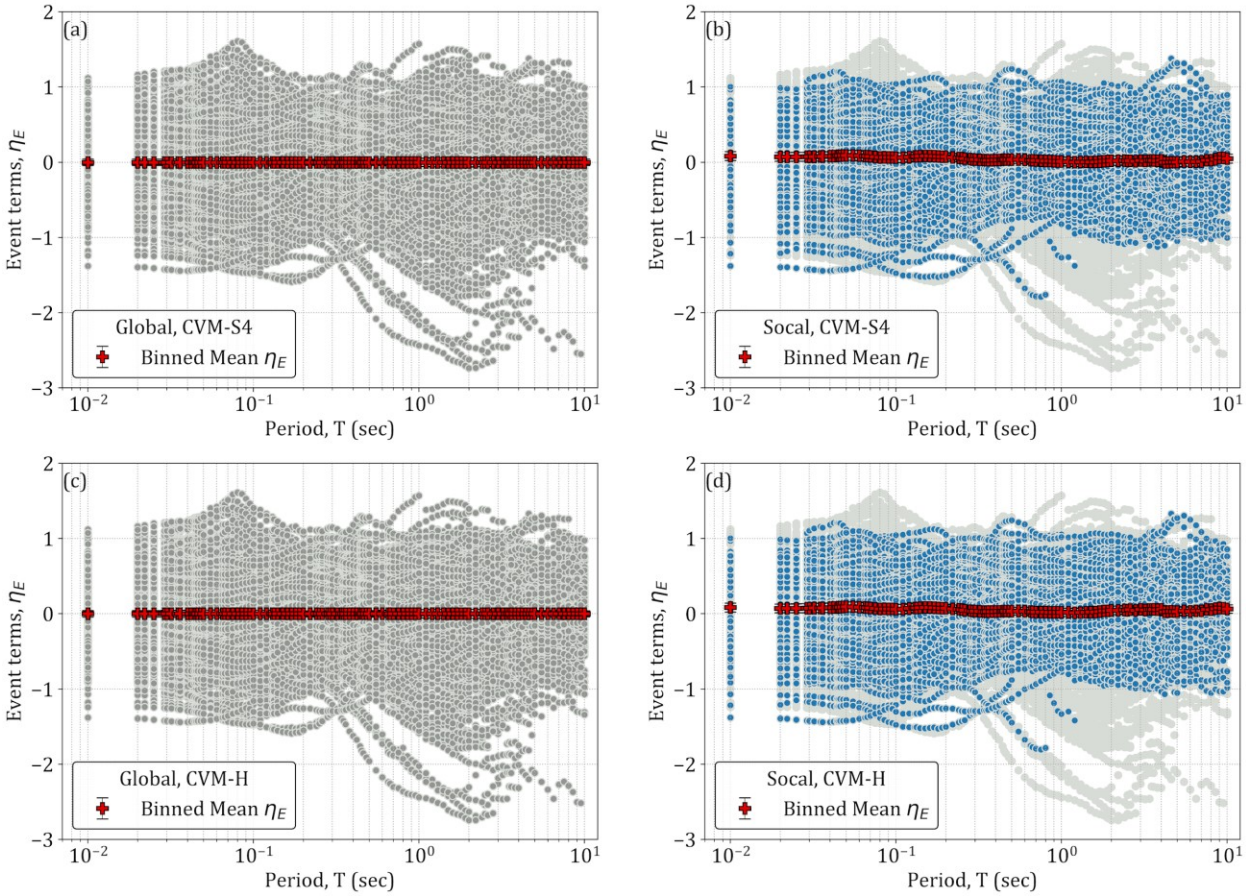


Figure 4.3. Event terms as a function of period using full NGA-West2 data set (left) and subset of southern California data (right). Binned means and their 95% confidence intervals are shown. Top panel is based on basin depths estimated from CVM-S4, while the bottom panel is based on basin depths from the CVM-H model.

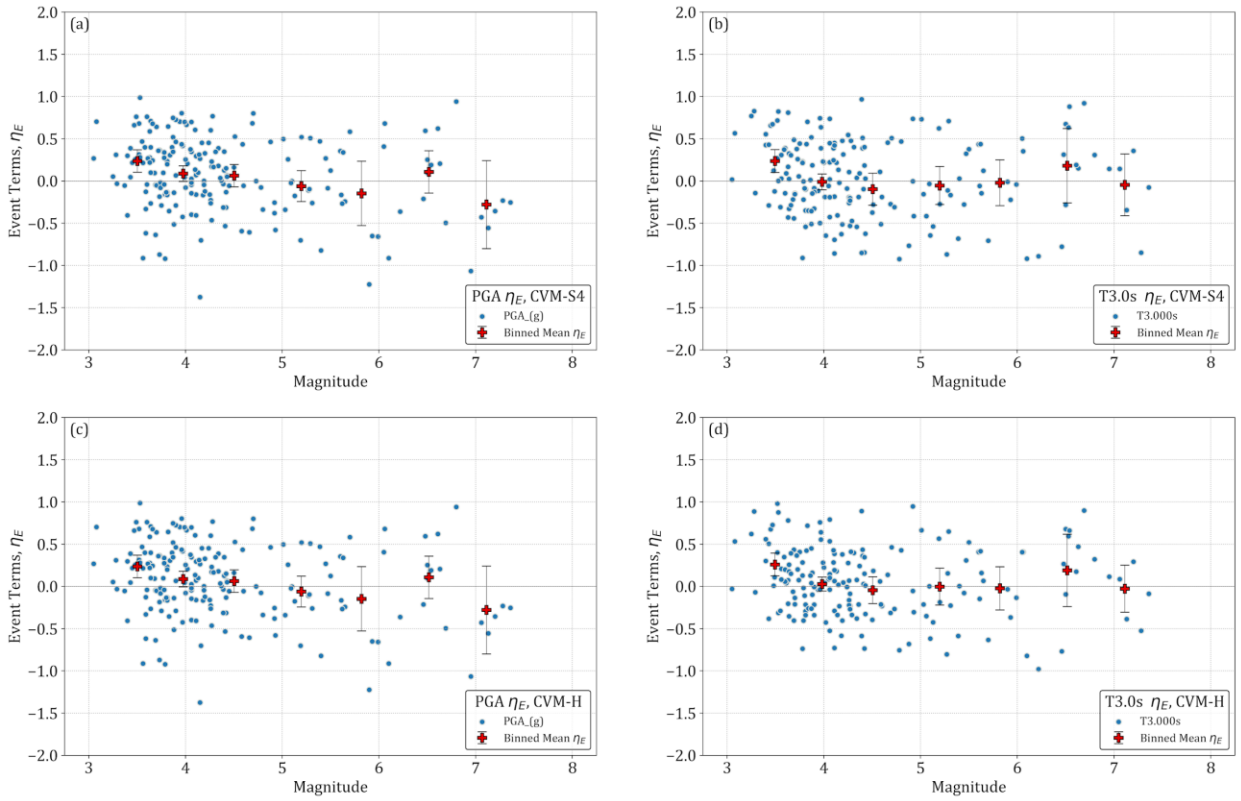


Figure 4.4. Event terms for PGA and S_a (3.0) as a function of magnitude for southern California data. Binned means and their 95% confidence intervals are shown.

Figure 4.5 shows the trends of δW_{ij} with distance for the PGA and S_a(3.0) intensity measures and the subset of the broader database from southern California. The lack of trend suggests that the path scaling in the model is unbiased for the data set, and hence the model is generally suitable for analysis of site effects. However, this lack of overall bias does not indicate a lack of path bias for individual events. Figure 4.6 shows trends of δW_{ij} with distance for six earthquakes of **M** = 4.45, 5.03, 5.09, 5.39, 5.99, and 7.06 respectively, using S_a(3.0) data. Event 10370141 arguably has an upward trend of residuals with distance (attenuation rate in model is too fast), Event 14095628 has a downward trend for distances under 200 km, the Whittier-Narrows-01 event has a subtle upward trend over the distance range, while the others have generally flat trends. Such differences are expected as a result of region- or path-specific variations in attenuation rates that are not captured by ergodic path models. As a consequence of these features being present in the data, some of the within-event variability attributed to site response in this study is actually caused by path-to-path variability. An underlying assumption of the non-reference site approach for estimating site effects is that these path biases average approximately to zero, and hence do not bias the estimate.

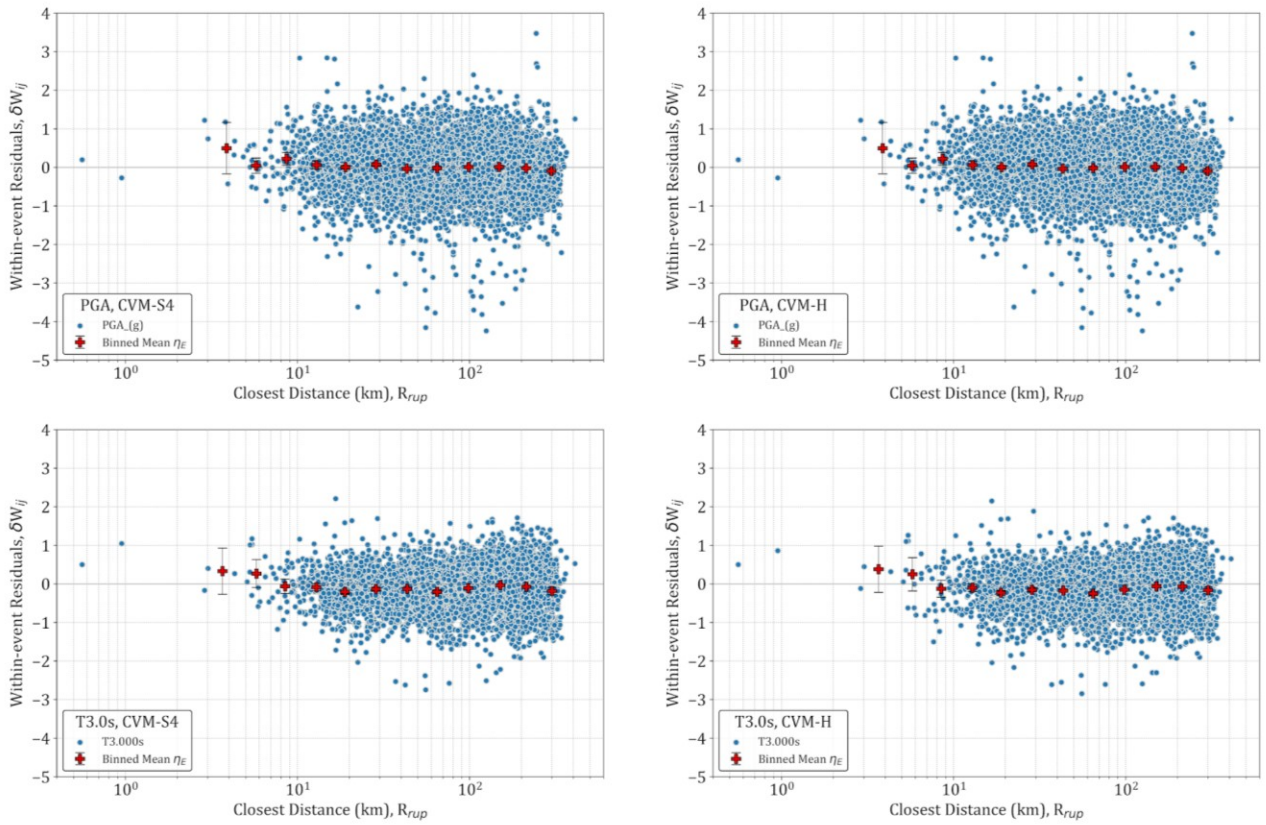


Figure 4.5. Within-event residuals for southern California data plotted as a function of distance. Binned means and their 95% confidence intervals are shown. There are no trends in the data with distance, indicating that the path scaling in the ground motion model is unbiased for the region.

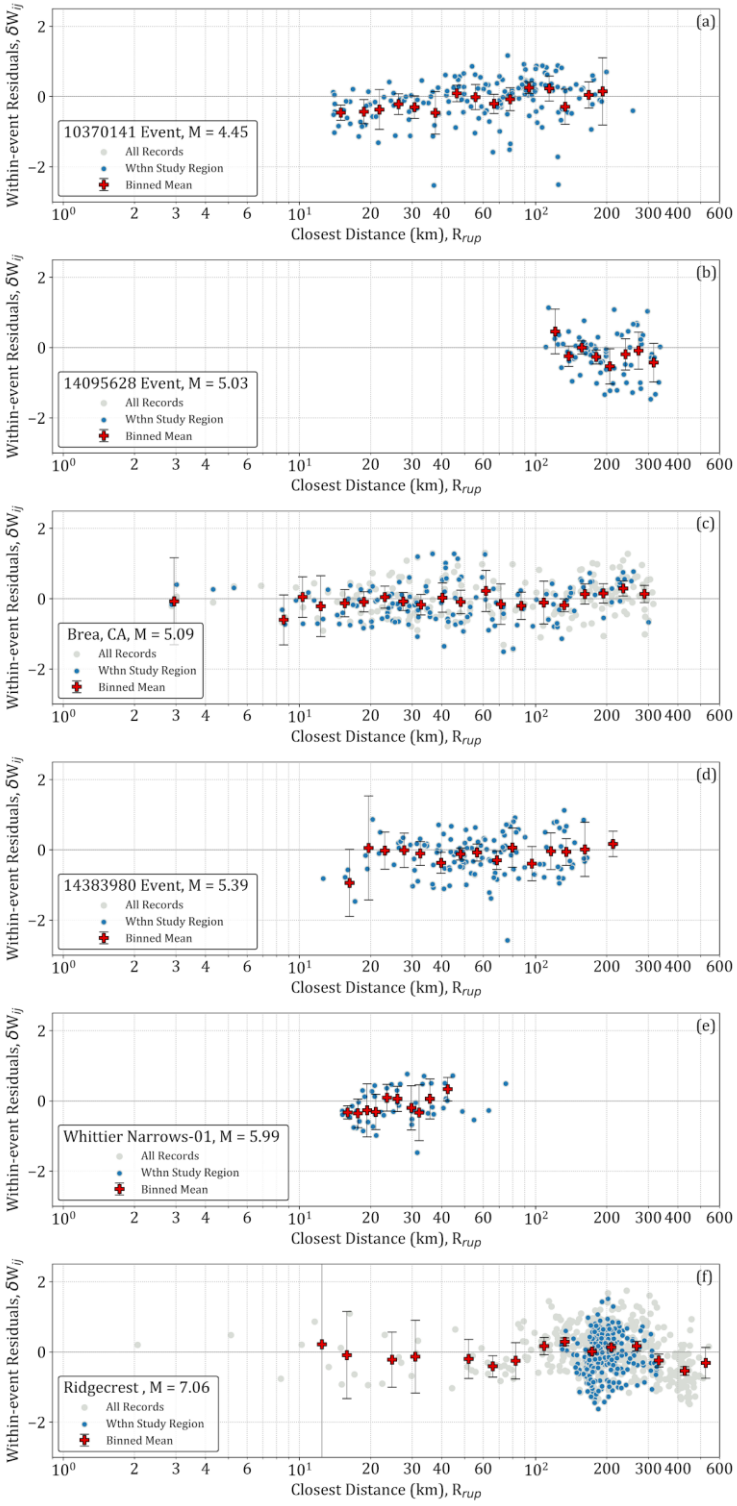


Figure 4.6. Within-event residuals for six southern California events with different magnitudes. Binned means and their 95% confidence intervals are shown.

4.4 SYNTHESIS OF SITE EFFECTS ANALYSIS APPROACH

In this chapter, we have shown that the source and path components of the BSSA14 model are unbiased with respect to the southern California portion of the database assembled in Chapter 2. With that having been established, the non-reference site approach is applied to estimate site response effects at sites in the study region and to develop site amplification models based on those results.

The analysis approach used to develop the site amplification models in Chapters 5 and 6 can be summarized as follows:

1. Using c_k terms and η_E terms from Section 4.3 in combination with the BSSA14 median model, derive data residuals relative to a reference rock model (i.e., with the site amplification model set to zero). These site terms are denoted $\eta^r_{S,j}$, and their physical meaning is site amplification relative to a V_{S30} = 760 m/s reference condition.
2. Check the BSSA14 V_{S30}-scaling model against the $\eta^r_{S,j}$ results, and adjust the model as needed so that it is centered with respect to the southern California data.
3. Compute site terms (η_S) from residuals calculated using the c_k terms and η_E terms from Section 4.3 and the BSSA14 median model modified as needed to account for the southern California V_{S30}-scaling from Step (2). No basin response model is used in these computations.
4. Investigate the trends of η_S with respect to differential depth for the different geomorphic provinces (categories) given in Table 3.1. Decide which categories produce distinct results and which (if any) can be combined.
5. Using the subset of sites from the overall database that are located in basins, evaluate the trends of η_S with differential depth for different basin structures. Evaluate whether particular combinations of basins (e.g., coastal vs. inland) produce distinct trends in site amplification.
6. Based on the median basin model derived in Steps (4)-(5), investigate variations of ground motion dispersion for different geomorphic categories and/or basins.

Steps (1) and (2) are the subject of Chapter 5. Steps (3) to (6) are the subject of Chapter 6.

5 Southern California V_{S30} Scaling Model

5.1 INTRODUCTION

While the objective of this research is to provide insight into the influence of sedimentary basins and related geologic features on site response, our approach is to characterize such effects relative to average site response conditioned on V_{S30}. As such, it is important for the V_{S30}-scaling model used as a reference to be unbiased relative to the southern California data set.

We begin with the linear V_{S30} scaling model developed by Seyhan and Stewart (2014) (hereafter SS14), and adjust that model as needed to fit the data. We do not anticipate significant bias, as the SS14 model was developed from a data set rich with California recordings. Nonetheless, SS14 is re-evaluated and adjusted to fit southern California data in the remainder of this chapter.

5.2 MODEL DEVELOPMENT

Site amplification models used in modern GMMs are the sum of three terms in natural log units:

$$F_s = F_{lin} + F_{nl} + F_b \tag{5.1}$$

where F_{lin} is the linear site-amplification term, F_{nl} is the nonlinear site-amplification term, and F_b is the basin-depth term. The F_{lin} term describes the variation of site amplification with V_{S30} for small-strain (linear) conditions, and is commonly referred to as a V_{S30}-scaling term. In Eq. 5.1, F_{nl} is nearly zero for recordings at low amplitudes (which is a significant majority of the dataset). For the purpose of this chapter, F_b reverts to its default value of zero because site response is only modelled with respect to V_{S30} (the F_b term is the subject of Chapter 6).

To develop a linear V_{S30}-scaling term specific to southern California, we begin by computing within-event rock residuals (δW_{ij}^r) using a reference rock conditioned GMM (μ_{ij}^r) and event terms ($\eta_{E,i}$),

$$\delta W_{ij}^{r} = ln(Y_{ij}) - \left[\mu_{ln,ij}^{r} + F_{nl,ij} + \eta_{E,i}\right] \quad (5.2)$$

where Y_{ij} is the ground motion intensity measure from event i as recorded at station j. The non-linear site effect (F_{nl}) is subtracted from ln(Y_{ij}) to isolate the linear component of the site response, and is taken from SS14. The reference rock conditioned GMM ($\mu_{ln,ij}^{r}$) applies for a reference site condition of 760 m/s. The event term is computed from mixed effects analysis as described in Section 4.2.

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

$$\delta W_{ij}^{r} = \eta_{S,j}^{r} + \varepsilon_{ij} \quad (5.3)$$

The partitioning is done using mixed-effects analysis routines in *R* [R Core Team, 2019; Bates et al. 2015], which are used as a subroutine in Python [Gautier 2009]. The superscript *r* indicates the site term $\eta_{S,j}^{r}$ is computed relative to the reference-rock velocity of 760 m/sec.

In Figure 5.1, the $\eta_{S,j}^{r}$ terms from southern California sites are plotted against V_{S30}. The SS14 global linear V_{S30}-scaling model is shown with a blue line for reference. Because the data exhibit appreciable scatter, binned means are added with their confidence intervals to facilitate visualization of data trends. The V_{S30} scaling trends in Figure 5.1 generally consist of a flat region at low V_{S30}, a downward slope that indicates reduced site amplification as V_{S30} increases, and a second flat region at high V_{S30}. The SS14 model produces these same trends except for the flat portion at low V_{S30}.

We elected to modify SS14 both because of the flat region at low V_{S30} and steeper slopes for intermediate V_{S30} at some periods (e.g., 1.0 and 3.0 sec in Figure 5.1). As done previously for central and eastern North America (Parker et al. 2019), the F_{lin} function is taken as tri-linear,

$$F_{lin} = \begin{cases} c\ ln\left(\frac{V_{C1}}{V_{ref}}\right) & V_{S30} \leq V_{C1} \\ c\ ln\left(\frac{V_{S30}}{V_{ref}}\right) & V_{C1} < V_{S30} \leq V_{C2} \\ c\ ln\left(\frac{V_{C2}}{V_{ref}}\right) & V_{S30} > V_{C2} \end{cases} \quad (5.4)$$

where V_{C1} and V_{C2} are the lower and upper limits of the sloping portion of the function, V_{ref} = 760 m/s, and c is the slope of the middle portion. The model development process began by selecting limiting velocities as follows:

- V_{C2} was not modified from SS14, because the data did not require changes
- V_{C1} was determined by computing the mean of site terms with V_{S30} < 220 m/s and finding the intersection of that mean with the sloping middle portion. This process was iterative, because the slope of the middle portion depends to a limited degree on V_{C1}.

The slope c of the middle portion was then set by least squares regression using data between V_{C1} and V_{C2}. Finally, initial values of parameters developed using this process were smoothed by eye with respect to period. The resulting model coefficients are given in Table E1. We excluded sites from the Imperial Valley Basin in the fitting process because the trends of this data are distinct from other portions of southern California (shown by green symbols in Figure 5.1).

The fit obtained from Eq. 5.4 is shown with the black line in Figure 5.1. The low-V_{S30} plateau is primarily driven by soft soil sites in the San Fernando Valley and Los Angeles Basin. Although Imperial Valley sites (green symbols) were not used in the regression, they also support the presence of a plateau, although generally at a lower level than the other southern California sites. As shown in SS14, other regions with soft soils (including the Bay Mud sites in the San Francisco Bay Area) lack the plateau feature. Accordingly, we interpret the plateau as a regional feature.

As with the SS14 model, the V_{S30}-scaling parameter c is largest for oscillator periods T between about 1 and 5 secs and smallest at short periods. The wavelengths associated with these long periods exceed 30 m, so the strength of this trend is expected to be a result of correlation between V_{S30} and the deeper velocity structure as shown in Boore et al. [2011].

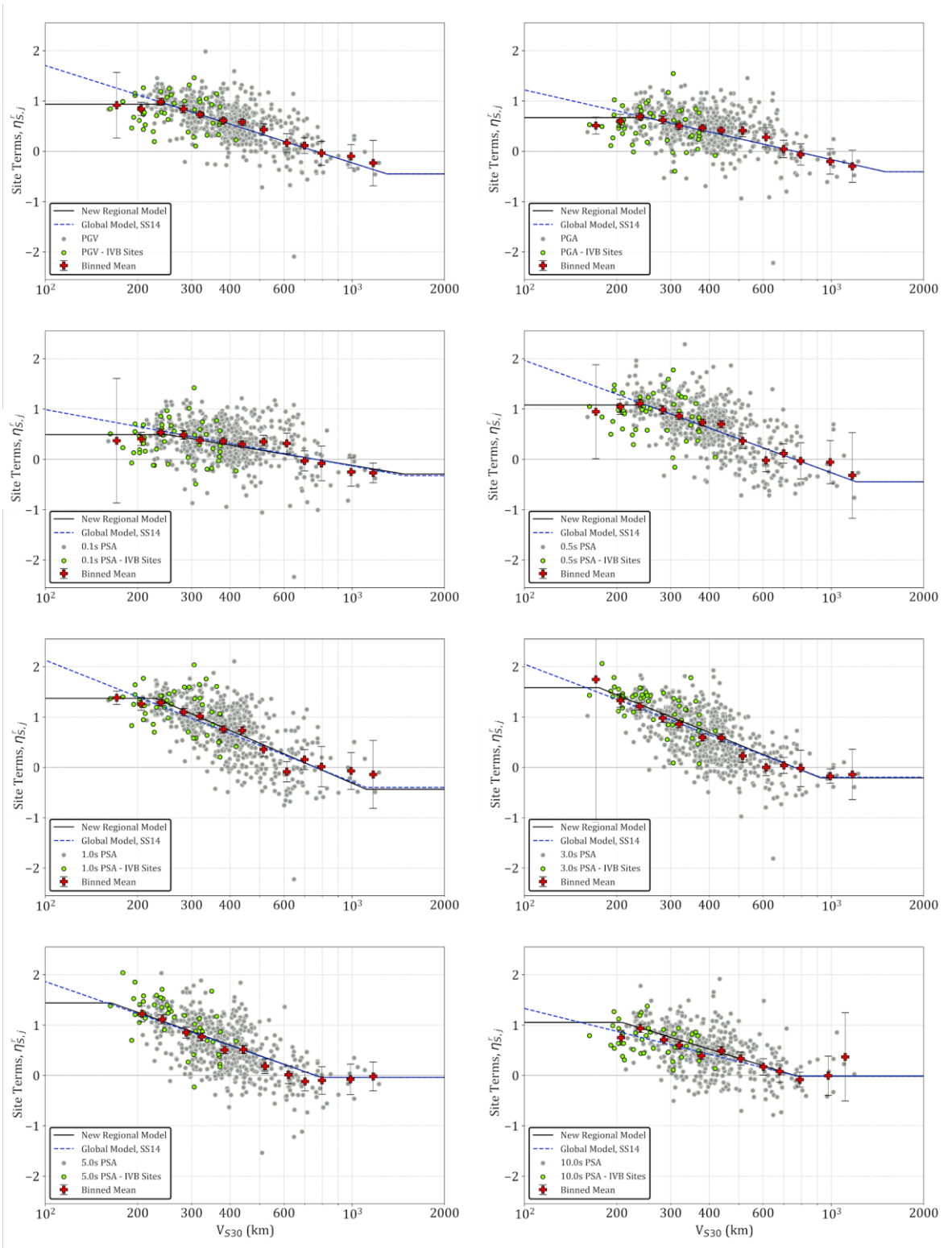


Figure 5.1. Comparison of the current global and new proposed southern California V_{S30}-scaling model for PGV, PGA, and a range of PSA oscillator periods between 0.1-10 sec.

Figure 5.2 compares the slope parameter c for the global SS14 model to the southern California model developed in this study. At shorter periods (< 0.45 sec) the southern California V_{S30}-scaling is weaker than in SS14 to a minor extent. For $T > 0.5$ sec, the southern California V_{S30}-scaling is stronger, with the largest differences occurring between approximately 1 and 3 sec.

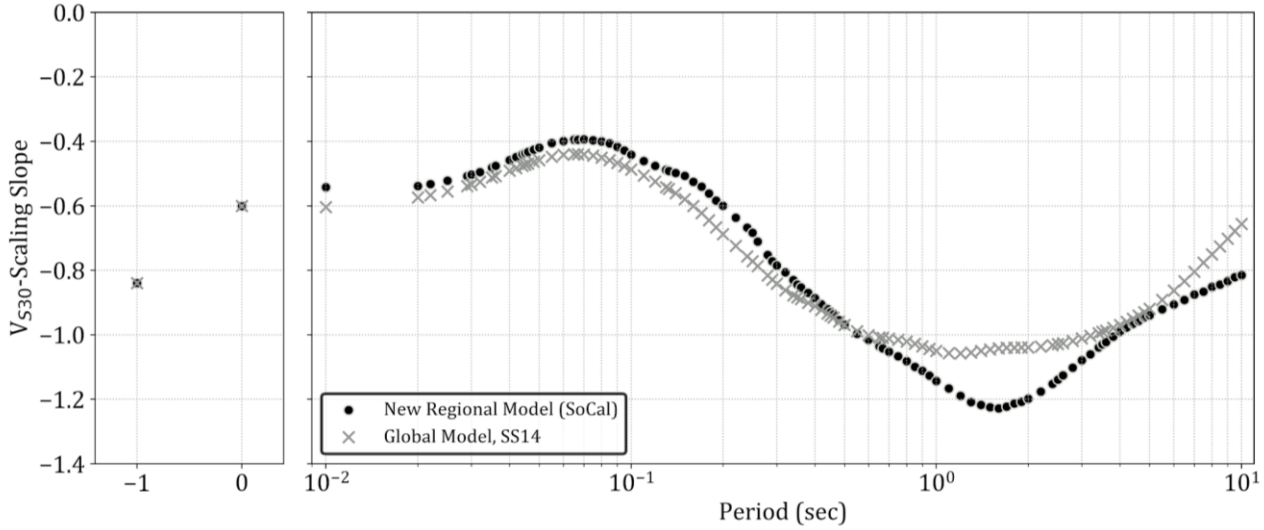


Figure 5.2. Comparison of the VS30-scaling slope between the Seyhan and Stewart (SS14) model used in BSSA14 GMM and the slope for the southern California study region.

5.3 RESIDUALS ANALYSIS

We adapt Eqs. (4.1) and (4.2) to compute within-event residuals using a GMM that includes the regional V_{S30}-scaling model,

$$\delta W_{ij}^{v} = ln(Y_{ij}) - \left[\mu_{ln,ij}^{r} + F_{lin} + F_{nl,ij} + \eta_{E,i}\right] \qquad (5.5)$$

where F_{lin} is the V_{S30}-scaling model from Section 5.2, F_{nl} is the nonlinear model from SS14, and all other terms are as defined previously. The event term is computed from mixed effects analysis as described in Section 4.2. As described in Section 4.2, the analysis of event terms is performed using a version of the GMM that includes basin response (models developed here in southern California; BSSA14 basin model elsewhere). For this reason, the development of the basin amplification models and the analysis of site terms is iterative; the results evaluated here and in Chapter 6 are based on event terms computed using final versions of the site amplification model following this iterative procedure. Moreover, because there are two models for basin depths in the study region (CVM-S4 and CVM-H) there are two sets of event terms.

The V_{S30}-dependent within-event residuals are partitioned into V_{S30} dependent site terms ($\eta_{S,j}^{v}$), which approximately represent the difference between average site amplification and the model defined within the brackets in Eq. (5.5) for many events at each recording station, and the remaining residual (ε_{ij}),

$$\delta W_{ij}^{v} = \eta_{S,j}^{v} + \varepsilon_{ij} \tag{5.6}$$

As in Section 5.2, the partitioning is performed using mixed-effects analysis routines in *R* [R Core Team, 2019; Bates et al. 2015], which are used as a subroutine in Python [Gautier 2009]. The superscript *v* indicates the site term $\eta_{S,j}^{v}$ is computed using a GMM that includes V_{S30}-scaling as provided in Section 5.2.

Figure 5.3 shows the variation of $\eta_{S,j}^{v}$ with V_{S30} for the data set as a whole and the four geomorphic provinces listed in Table 3.1. The trends for the data set as a whole are flat with a mean of zero, as expected. For the basin, basin edge, and mountain hill provinces there are no appreciable trends of residuals with V_{S30}, which indicates that the V_{S30}-scaling model captures first-order site effects for those conditions. Valley sites exhibit a modest downward trend with V_{S30}, indicating stronger V_{S30}-scaling than the overall southern California model. We do not attempt to capture this effect in our recommended models. The mean of $\eta_{S,j}^{v}$ for all sites is also shown in the plots of Figure 5.3; these features of the data are analyzed further in Chapter 6.

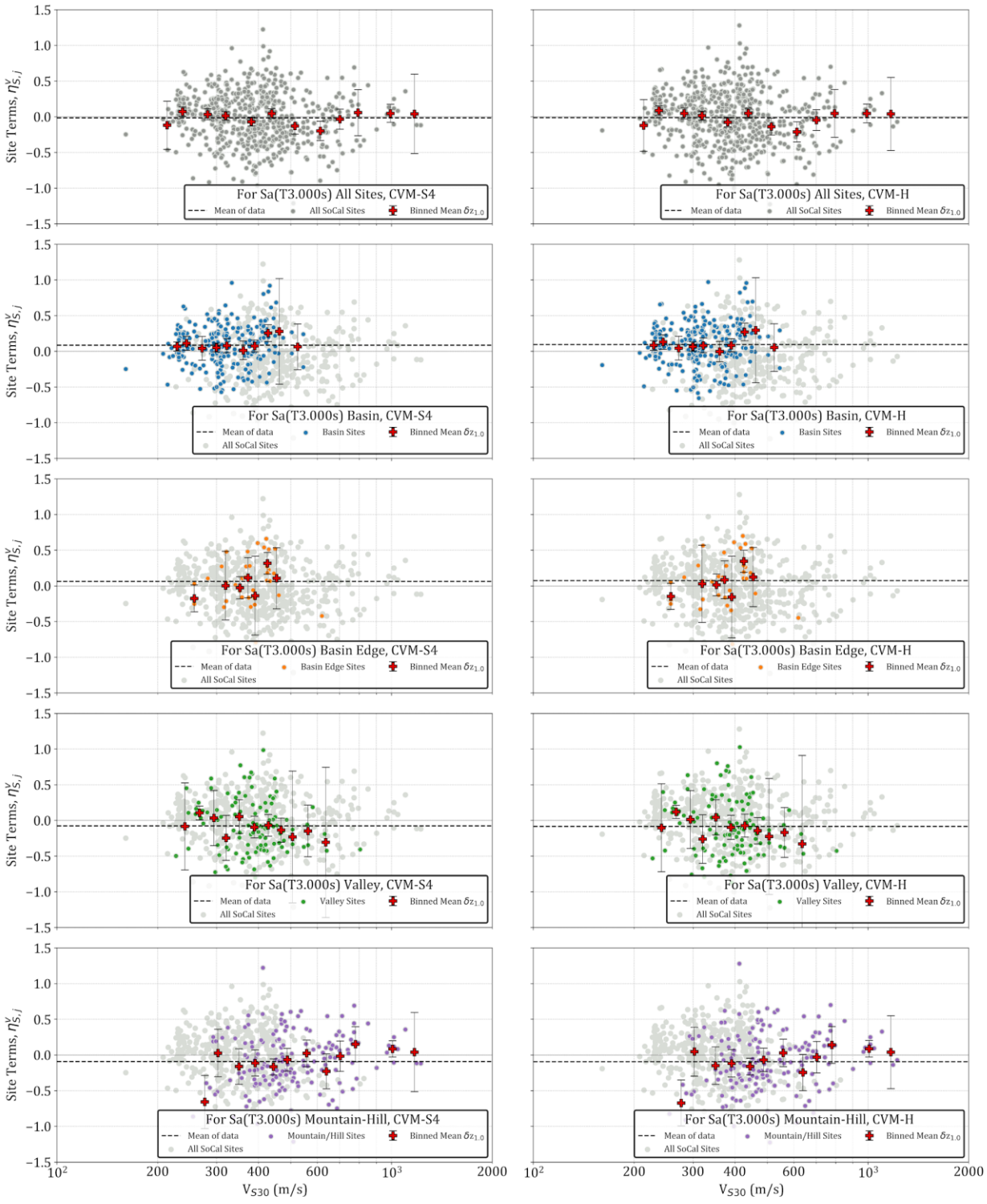


Figure 5.3. Variations of southern California site terms with V_{S30} for the intensity measure of $S_a(3.0)$ for the site categories proposed in Table 3.1. The dashed line indicates the mean of the data subset display in the plot.

6 Southern California Basin Amplification Model

6.1 INTRODUCTION

In this chapter, we develop models that are intended to supplement the V_{S30}-scaling model in Chapter 5 based on geomorphic province and basin depth. For brevity, these are referred to as "basin amplification models," but it should be understood that their applicability is not restricted to sedimentary basins. The models developed here improve upon an existing basin depth model incorporated into the BSSA14 GMM, the attributes of which are reviewed in Section 6.2. Section 6.3 describes data trends for different geomorphic provinces and Section 6.4 describes model development based on those trends. Section 6.5 describes various attributes of model performance, including residuals plots and site-to-site aleatory variability. A summary of our recommendations for ergodic site response modeling is presented in Chapter 7.

The work presented here has focused on southern California for reasons outlined in Sections 2.2 and 3.1 (ample data and variety of basin geologic features). The objectives of the data analyses and model development in this chapter are to address the following questions:

- How does mean site response and the associated ground motion variability differ for basins compared to non-basin geomorphic provinces?
- What are the variations in basin response between different basin structures and how can this be parameterized for predicting ground motion intensity measures?

6.2 NGA-WEST2 BASIN MODEL

The NGA-West2 project introduced the concept of conditioning "basin" amplification models not on sediment depth itself (as had been done in NGA-West1), but on the differential depth defined as,

$$\delta z_{1.0} = z_{1.0} - \bar{z}_{1.0} \tag{6.1}$$

where $z_{1.0}$ is the depth to the 1.0 km/sec shear wave isosurface and $\bar{z}_{1.0}$ is the mean depth to this isosurface conditional on the site V_{S30}. The concept behind the use of this parameter is connected with "centering" -- namely, the site response from the V_{S30}-scaling model represents the mean of the data for a given V_{S30}, and a model conditioned on the differential depth in Eq. (6.1) represents site conditions that depart from centered conditions. The centering principle has no direct association with the 1.0 km/sec isosurface; the same principle can be applied to alternate isosurfaces such as 2.5 km/sec.

The NGA-West2 GMMs that include a basin term use differential depths (Eq. 6.1) as the independent variable. The data used to constrain the models are mainly from southern California, the San Francisco Bay Area, and Japan, where basin models were available at the time of that project (described in Seyhan et al. 2014). However, data from sites located outside of basins (as defined in Section 3.2) were also included, including Mountain/Hill, Valleys, and Basin Edge sites. In some cases these depths came from the "basin" models available at that time, whereas in other cases the depths came from V_S profiles that cross the 1.0 km/sec horizon.

As illustrated in Figure 6.1, the main features of the basin models in NGA-West2 are: (1) no change relative to V_{S30}-based site amplification for short periods (generally < 0.7 sec); (2) de-amplification of long-period motions for negative $\delta z_{1.0}$; and (3) amplification of long-period motions for positive $\delta z_{1.0}$. In the application of these models, there has been some debate on whether to allow de-amplification for negative $\delta z_{1.0}$ sites (e.g., Petersen et al. 2020).

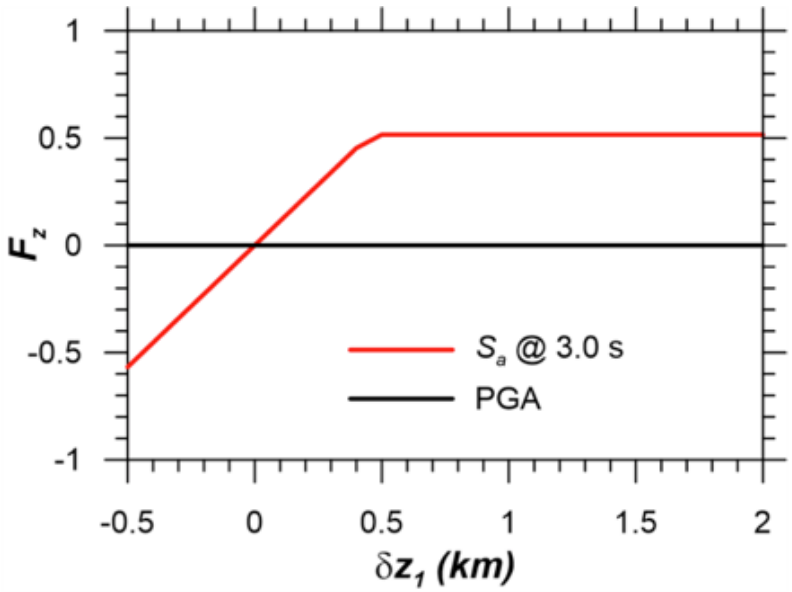


Figure 6.1. Short period (PGA) and long period (S_a(3.0)) basin amplification features as implemented in the BSSA14 GMM from NGA-West2.

6.3 BASIN MODEL PARAMETER AND SITE CATEGORIZATION ASSESSMENT

The data analysis and model development approach adopted in this research was described in Section 4.4. We have previously established that the southern California data considered in this research is unbiased *en masse* relative to path effects (Section 4.3), and we have developed a (slightly) modified version of a previous V_{S30}-scaling model that is customized for the study region (Section 5.2). In the remainder of this chapter, we investigate the influence of geomorphic provinces (Table 3.1) on the portions of site response not explained by the V_{S30}-scaling model. As described in Section 5.3, those portions of site response are quantified by site terms $\square_{\square,\square}^{\square}$, which are examined within different geomorphic provinces (basins, basin edges, valleys, mountain-hills) and individual basin structures (LAB, SFB, VB, SGB, SBB, CB, CVB, IVB).

The site amplification effects examined here, while derived from a variety of geomorphic provinces, are referred to as *basin effects*. These basin effects are expressed in natural log units using the F_b term in Eq. 5.1. In this section, we investigate features of the southern California data that should be captured in models for F_b. A model for F_b is developed in Section 6.4.

6.3.1 Site Responses within Categories

Overall Mean

Figure 6.2 shows the mean of southern California $\eta_{S,j}^{v}$ terms plotted against oscillator period for each geomorphic province. Due to the differences in event terms, results are shown separately for the two basin models (CVM-S4 and CVM-H). The means across all categories ("All sites", inclusive of basin and non-basin locations) are approximately zero. However, individual categories have non-zero means that in some cases depart significantly from zero. To provide an indication of statistical significance, 95% confidence intervals are shown for two categories with the greatest (basin) and smallest (basin edge) numbers of sites.

Category Means

Figure 6.2 shows that mountain-hill sites exhibit negative biases over the full period range indicating over-prediction (de-amplification relative to the mean response). For periods less than about 0.32 sec., basin sites show no bias, but for longer periods basin sites have positive bias (indicating under prediction) that peaks between 4-6 sec. Valley and basin edge sites exhibit positive biases at short periods but transition to negative biases at long periods (> 0.75 sec. for valleys and > 5 sec. for basin edges). This behavior is consistent for the two CVMs.

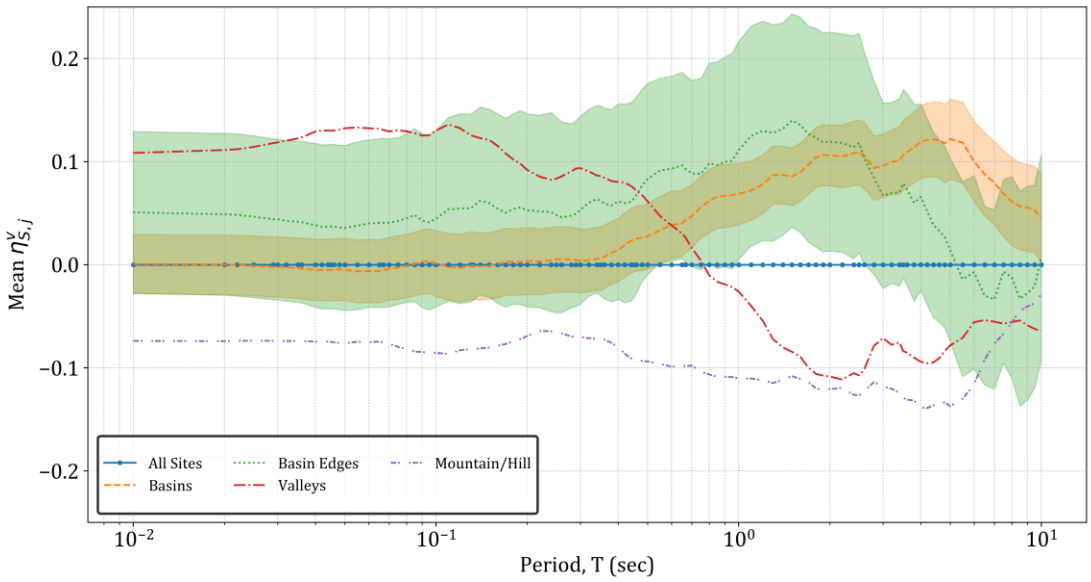


Figure 6.2. Mean of V_{S30} dependent site terms ($\eta^{v}_{s,j}$) as a function of oscillator period for the four geomorphic provinces with 95% confidence interval show for the basin and basin edge category.

Figure 6.3 shows mean $\eta^{v}_{S,j}$ terms plotted against oscillator period for individual southern California basins for both CVMs. These individual basin subsets contain sites within the basin and basin edge geomorphic provinces (denoted BBE). At long periods (> 1 sec) the coastal basins (LAB, VB, SFB) and IVB exhibit positive biases. Inland basins have varied behavior, with SBB, CB, and CVB exhibiting minimal or negative biases while SGB has positive bias. Basins for which the mean exceeds BBE have a stronger-than-average site response; this occurs for IVB, LAB, and VB. The remaining basins have a weaker-than-average response, which is particularly pronounced for inland basins other than IVB and SGB.

The trends in Figures 6.2-6.3 can be interpreted on the basis of site response physics. Geomorphic provinces and basins with especially large depths would be expected to have low resonant frequencies (whether from shear waves or surface waves), thus producing larger than average long-period site responses that manifest as positive means in these figures (e.g., as observed for the basin category and deep basins such as LAB, VB, and IVB). Geomorphic provinces with shallow depths (mountain/hill and valley) or basins of modest depth (SBB, CB, CVB) would be expected to have relatively high resonant frequencies and essentially no low-frequency amplification, which would manifest as positive averages at short periods that transition to zero or negative averages at long periods.

The very large bias shown in Figure 6.3 for the VB carries large uncertainty due to limited data. The IVB means shown in Figure 6.3 are of comparable reliability to those for other basins. However, due to limited confidence in the CVM-S4 basin model (discussed in Section 3.5), the

interrogation of IVB sites with respect to differential depth is only carried out using the CVM-H model subsequently in this report.

Overall and Category Site-to-Site Dispersions

Site-to-site dispersion ϕ_{S2S} represents the standard deviation of site terms, η_S (Section 4.2). The GMM used to compute the residuals from which ϕ_{S2S} is derived influences the resulting dispersion. Within the context of this study, the largest dispersion would be expected from the V_{S30}-based ergodic model, in which geomorphic provinces are not considered. This dispersion is plotted in Figure 6.4 as a reference value (labelled as "ergodic"). Also shown in the figure for all sites in aggregate when the site response model is amended to include the mean adjustments for geomorphic provinces described above (labelled as "ergodic + provinces"). The incorporation of province means into the site response model reduces the aggregated ϕ_{S2S} by amounts ranging from 0.01 at short periods to 0.03-0.04 at periods near 4-6 sec.

Lastly, standard deviations of site terms within geomorphic provinces are shown with confidence intervals marked for the largest and smallest numbers of sites (basin and basin edge, respectively). These results show substantial variations, with basins generally having the lowest dispersion and mountain/hill and basin edge sites having the largest dispersions. The province-specific dispersions in Figure 6.4 provide a point of comparison for models developed in Section 6.4 in which site response scaling with different depth is considered within provinces and individual basins. The effects of this additional model complexity on site-to-site dispersion is investigated in Section 6.5.

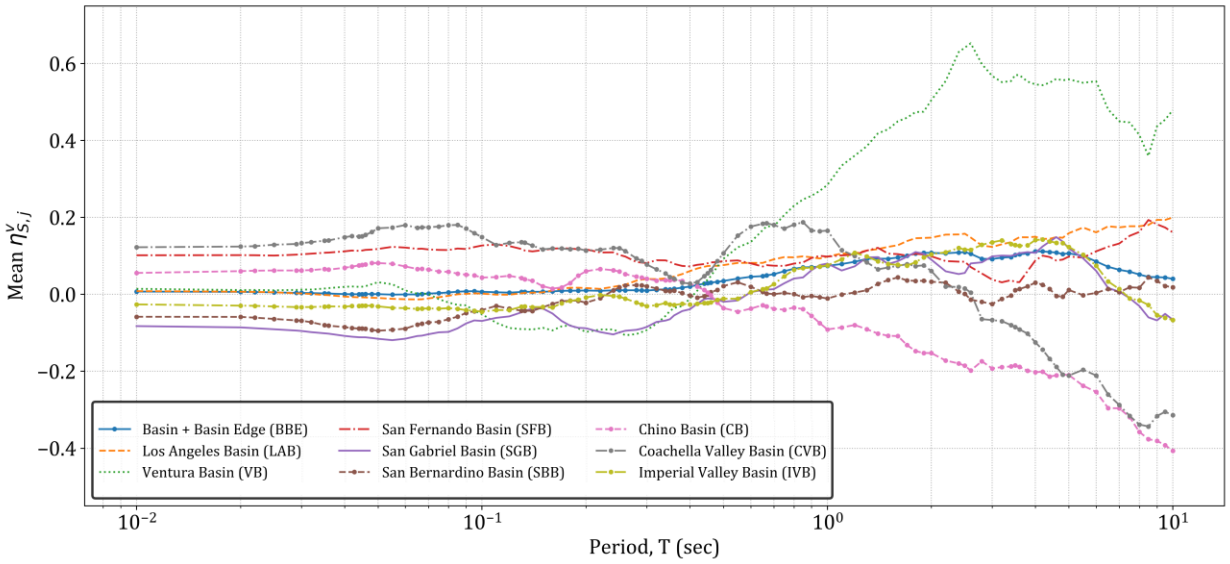


Figure 6.3. Mean of V_{S30} dependent site terms ($\eta^{v}_{S,j}$) as a function of oscillator period for the individual basins in southern California.

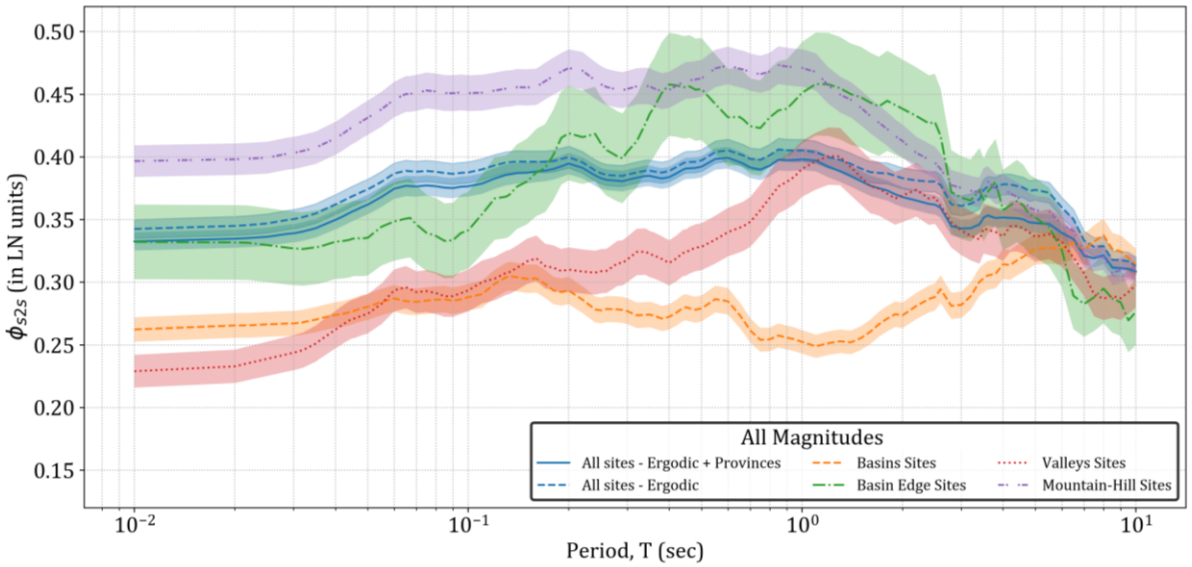


Figure 6.4. Site-to-site dispersions (ϕ_{S2S}) for aggregated dataset without and with consideration of geomorphic province means (ergodic and ergodic+province, respectively) and for individual provinces prior to the fitting of models conditioned on differential depth.

6.3.2 Scaling of Site Response with Differential Depth

Here we investigate trends of $\eta^{v}_{S,j}$ (and hence site amplification) with $\delta z_{1.0}$ as defined in Eq. (6.1), using the mean depths given by Eq. (3.3). Figure 6.5 shows the individual site terms and binned means for the entire southern California data set (labelled as "All sites") for the intensity measures of PGA and $S_a(3.0)$. This form of the data is comparable to what was considered in NGA-West2, and the trends are similar. Specifically, we again see no appreciable trend in the site terms at PGA for both CVMs, whereas trends emerge at longer periods.

The long-period data exhibit amplification for positive differential depths ($\delta z_{1.0} > 0$) with both basin models, whereas negative differential depths ($\delta z_{1.0} < 0$) produce variable amounts of deamplification. These trends mirror those found in NGA-West2 (Figure 6.1). The large depth amplification is stronger in CVM-S4 than in CVM-H, whereas the deamplification is much weaker in CVM-S4 than in CVM-H.

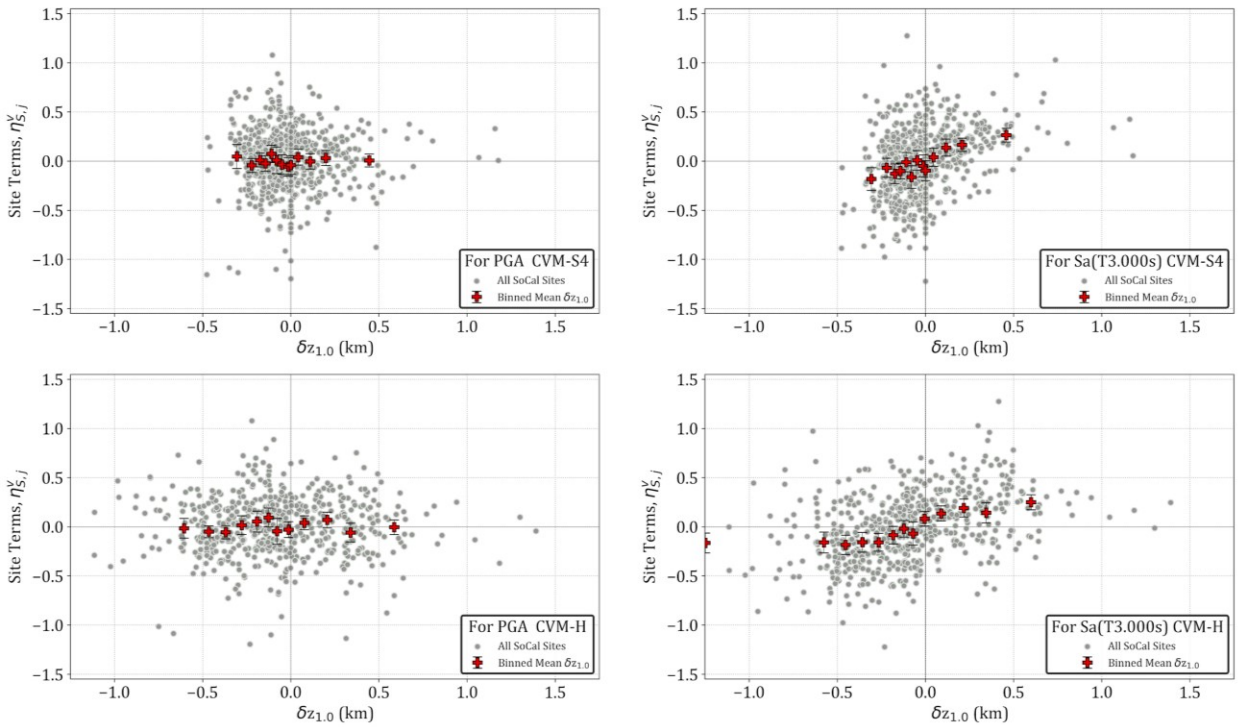


Figure 6.5. Variation of site terms with differential depth $\delta z_{1.0}$ for all considered sites in the southern California region for intensity measures of PGA and $S_a(3.0)$.

Figure 6.6 shows the trends of $S_a(3.0)$ site terms with differential depth for the site categories in Table 3.1 for both CVM-S4 and CVM-H. Elements of these data that are important for model building (Section 6.4) include:

- Mean offset from zero, which indicates bias in the data relative to the V_{S30}-scaling model
- Slope of the data with respect to $\delta z_{1.0}$, which if present, indicates scaling of site response with differential depth
- For ground motion parameters and categories that produce scaling with $\delta z_{1.0}$, limits at low and/or high $\delta z_{1.0}$ beyond which scaling with $\delta z_{1.0}$ is not supported by the data.

Each of these features is evident in the basin category, which is well populated for $\delta z_{1.0} \approx$ -0.5 − 1.0 km. Within this category, data from sites with $\delta z_{1.0} < 0$ has binned means near zero; this is a significant finding, suggesting that basins on average do not produce de-amplification for relatively shallow depths. Scaling of site amplification occurs for the approximate differential depth range of $\delta z_{1.0} \approx$ -0.2 − 0.5 km. For $\delta z_{1.0} >$ 0.5 km, the mean site amplification is approximately 0.4 in natural log units, corresponding to approximately 50% amplification relative to the V_{S30}-scaling model. As a result of the lack of bias for negative differential depth sites, and the positive bias for positive differential depth sites, the overall population average for basins is positive, as shown in Figure 6.2.

While each of the four categories exhibit scaling with differential depth, the limiting values (if any) of $\delta z_{1.0}$ for the scaling are not always apparent. For the basin edge category, $\delta z_{1.0}$ ranges from about -0.3 − 0.3 km (CVM-S4) and -0.5 − 0.5 (CVM-H) and neither depth limit is defined from the data. Over this limited data range, the mean levels of amplification and the scaling with $\delta z_{1.0}$ are arguably similar for the basin and basin edge categories. For the valley category, most sites have negative $\delta z_{1.0}$ and the lower $\delta z_{1.0}$ limit is reasonably well defined at approximately -0.2 km whereas the upper limit is undefined. The mountain/hill category similarly has mainly negative $\delta z_{1.0}$ sites; in this case the negative limit is poorly defined whereas the positive limit is reasonably well established at about 0.2 − 0.3 km. For the negative $\delta z_{1.0}$ condition, the mountain/hill category produces the largest mean de-amplification among the four categories (approximately -0.2 − -0.3 natural log units, or 20-25%).

Viewing the data as a whole, the de-amplification features at negative differential depths that are modelled in NGA-West2 and are present in the current "All data" set are mainly associated with valley and mountain/hill sites, whereas the amplification features are mainly associated with basin sites. Prior research on site amplification at shallow soil and rock sites has postulated topographic effects, edge/boundary effects, and rock property variations (discontinuities, mineral compositions related to strength and stiffness, weathering) as mechanisms that might control site response (e.g., Siddiqqi and Atkinson 2002, Pischiutta et al. 2017). While we recognize that these mechanisms do not have a clear physical association with differential depth, the scaling evident in the data nonetheless suggest the parameter has predictive power.

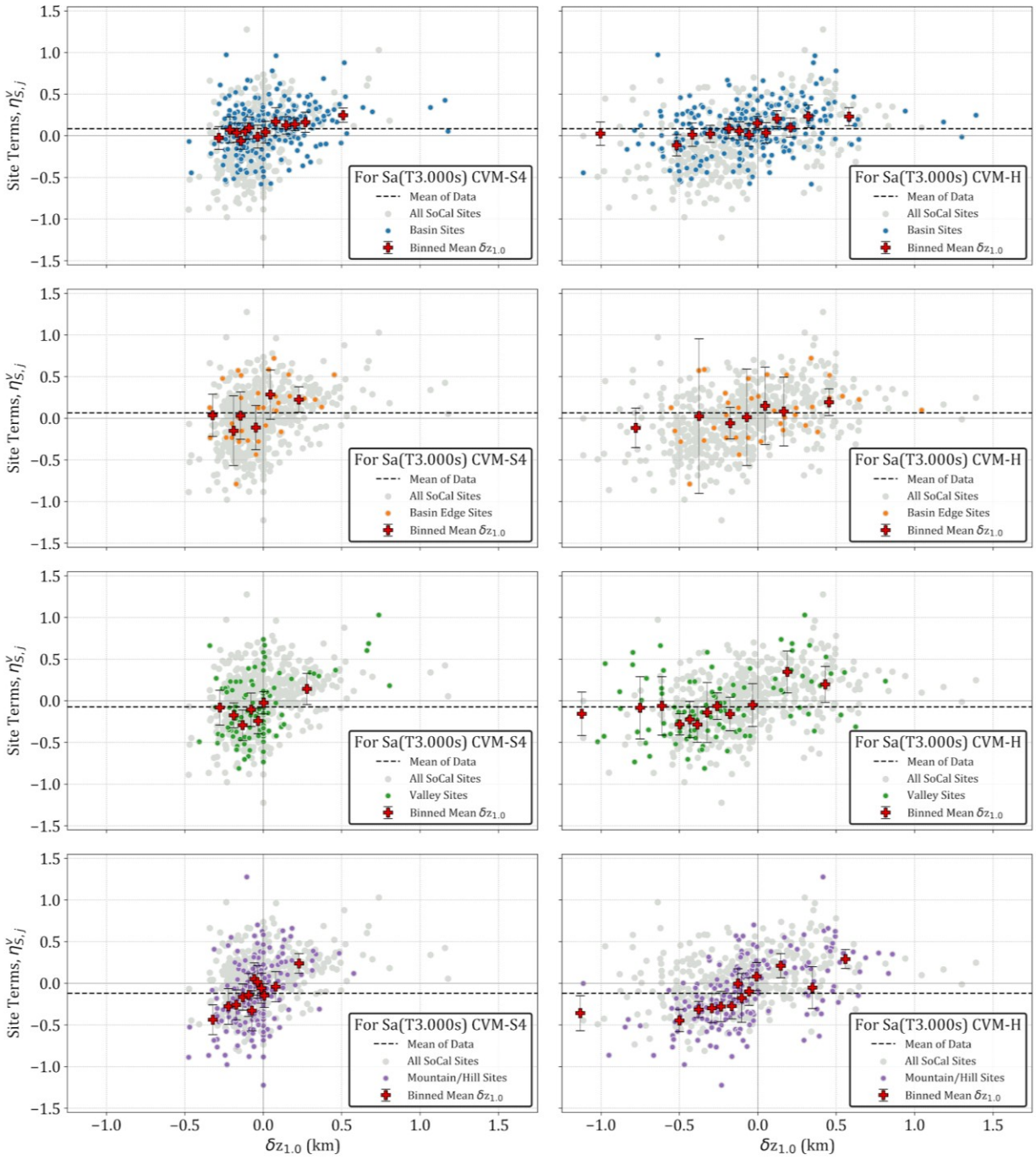


Figure 6.6. Variations of southern California $S_a(3.0)$ site terms with differential depth $\delta z_{1.0}$ for the geomorphic provinces proposed in Table 3.1. The dashed line indicates category means

Figure 6.7 shows the trend of site terms with differential depth for the various individual basins in the southern California study region (Section 3.1). The subsets of data in these individual basins are derived from basin and basin edge categories. Among the seven basin structures considered, only LAB has enough data to clearly identify the three features described previously (mean, trend with $\delta z_{1.0}$, range of $\delta z_{1.0}$ with the trend). The data trends for LAB in Figure 6.7 are similar to those for the basin category as a whole, suggesting that a model developed for the basin category would largely reflect LAB response.

For the other six basin structures, data trends are visually examined for consistency with those in LAB as follows:

- *Coastal basins*: Data trends for SFB have different features when depths are derived from CVM-S4 vs. CVM-H. In CVM-S4, most of the data is between $\delta z_{1.0}$ = -0.2 − -0.3 km with a nearly zero mean, although small a positive mean cluster at $\delta z_{1.0}$ = 0.1 km implies an upward trend consistent with LAB. In CVM-H, the data occupy a much wider range of $\delta z_{1.0}$, with negative binned means for negative $\delta z_{1.0}$ (different from LAB) and positive or zero binned means for positive $\delta z_{1.0}$ (similar to LAB). VB is also coastal, but the data is too limited to identify trends. Overall, the data from coastal basins are sufficiently similar that we consider combining them in the modeling process (Section 6.4).
- *Imperial Valley Basin (IVB)*: As described in Section 3.1, this inland basin has a distinct geologic origin from other basins in our study region, and therefore is considered separately from coastal and other inland basins. Because we lack confidence in the CVM-S4 depths for IVB (as discussed in Section 3.5), results are shown only for CVM-H in Figure 6.7. The trends in the data can be described as essentially zero mean for $\delta z_{1.0}$ < -0.2 km, climbing to a plateau at about 0.25 for $\delta z_{1.0}$ > 0.1 km. These features are broadly consistent with those for coastal basins.
- *Inland basins:* The inland basins considered in our study are SBB, SGB, CB, and CVB. In the case of SBB, the apparent slopes are negative for CVM-S4 and positive for CVM-H. Given this inconsistency and the weakness of the slopes, a $\delta z_{1.0}$-dependent model for SBB may not be justified. In the case of CB, residuals have a negative overall mean, a positive slope for $\delta z_{1.0}$ < 0, and are approximately zero for $\delta z_{1.0}$ > 0. The amount of data available for CVB is limited, but the trends for the CVM-H model generally appear to be compatible with those for CB (results not shown for CVM-S4 for the same reasons as IVB). In the case of SGB, the two models have different depth ranges (-0.3 − 0.1 km for CVM-S4; -0.6 − 0.4 for CVM-H), but in both cases there is no clear trend with $\delta z_{1.0}$ and the mean is positive. Since these inland basins exhibit distinct features from LAB and the basin category as a whole (general lack of trend with $\delta z_{1.0}$ and variable means), separate models for inland basins may be justified.

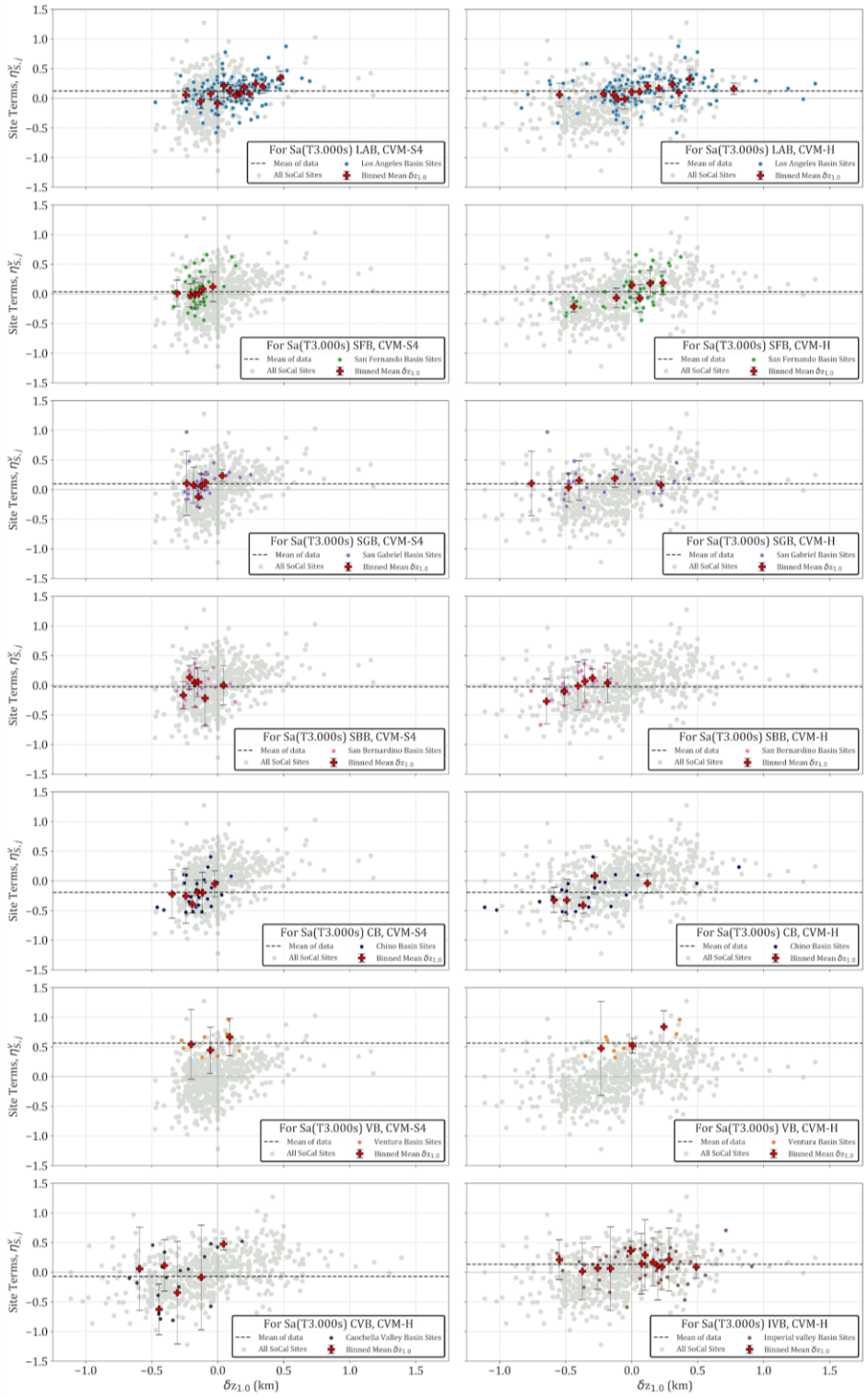


Figure 6.7. Variations of southern California $S_a(3.0)$ site terms with differential depth $\delta z_{1.0}$ for individual basins described in Section 3.1. The dashed line indicates the mean of the data subset display in the plot.

6.4 MEAN MODEL DEVELOPMENT

The $\eta_{S,j}^{v}$ plots presented in Section 6.3 demonstrate that both geomorphic provinces and the geological origins of individual basins affect site response, and that for several provinces/basins, site response scales with differential depth, $\delta z_{1.0}$. This section describes the development of models that capture essential elements of the data trends. There are two main considerations in developing models for the mean response:

1. Select a function that captures trends of $\eta_{S,j}^{v}$ with $\delta z_{1.0}$, where they exist
2. Group conditions that exhibit similar responses

The following subsections describe both considerations as well as coefficient smoothing.

6.4.1 Functional Form

We propose a tri-linear form for basin term F_b in Eq. (5.1), which is flat for low and high values of $\delta z_{1.0}$ and sloped between these plateaus for intermediate differential depths, as follows:

$$F_b = \begin{cases} f_7 + f_6 f_8 & \delta z_{1.0} < f_8 \\ f_7 + f_6 \delta z_{1.0} & otherwise \\ f_7 + f_6 f_9 & \delta z_{1.0} > f_9 \end{cases} \tag{6.2}$$

This form is similar to that used in BSSA14, except that non-zero F_b is now allowed at $\delta z_{1.0} = 0$. Per centering requirements, $F_b(\delta z_{1.0} = 0)$ should be zero for the data as a whole, but this requirement does not apply to individual provinces or basins. Parameter f_6 represents the slope of the differential depth scaling (units of 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 in units of km between which site response scales with $\delta z_{1.0}$.

For conditions without differential depth scaling (e.g., inland basins other than IVB), f_6 is set to zero and f_7 is taken as mean $\eta_{S,j}^{v}$ terms from Section 6.3.1. For such cases, only the middle term in Eq. (6.2) is required.

For conditions with differential depth scaling, we first select limiting differential depths (f_8 and f_9) based on visual inspection. As described in Section 6.3.2, well-populated geomorphic provinces (mountain/hill, valley, basins) and basin structures (LAB) enable such identifications. Where this is not possible due to sparse data, we apply judgement in the selection of these scaling limits based on the data trends for other, geologically similar categories with more data.

Next, the slope f_6 and offset f_7 was set by least squares regression using data between f_8 and f_9. For reasons provided in Section 3.5, data from CVB and IVB is excluded from these evaluations for the case of the CVM-S4 model. Example fits are shown in the next subsection based on individual and grouped datasets.

6.4.2 Data Grouping by Site Condition

As described in Section 6.3.2, the overall (aggregated) dataset and each of the individual geomorphic provinces demonstrate dependencies of site amplification ($\eta_{S,j}^{v}$) on differential depth, $\delta z_{1.0}$, in a manner that is generally compatible with Eq. (6.2). The basin and basin edge categories are grouped together for these reasons:

- While the trends in Figure 6.6 are not identical, the sparsely populated basin edge data does not reject the basin model, as shown subsequently.
- The "boundary" between basin and basin edge sites is somewhat arbitrary, and keeping these as separate entities would produce a step of site amplification that is difficult to justify given the arbitrariness of the boundary.

Figure 6.8 (PGA) and 6.9 (3.0 sec Sa) shows the data from Figure 6.6 for each geomorphic province (including the "all data" group) for which models are provided for both the CVM-S4 and CVH-H basin models. These conditions are all data, basin and basin edge (BBE), valley, and mountain/hill. Model coefficients are provided in Table E1 in the electronic supplement.

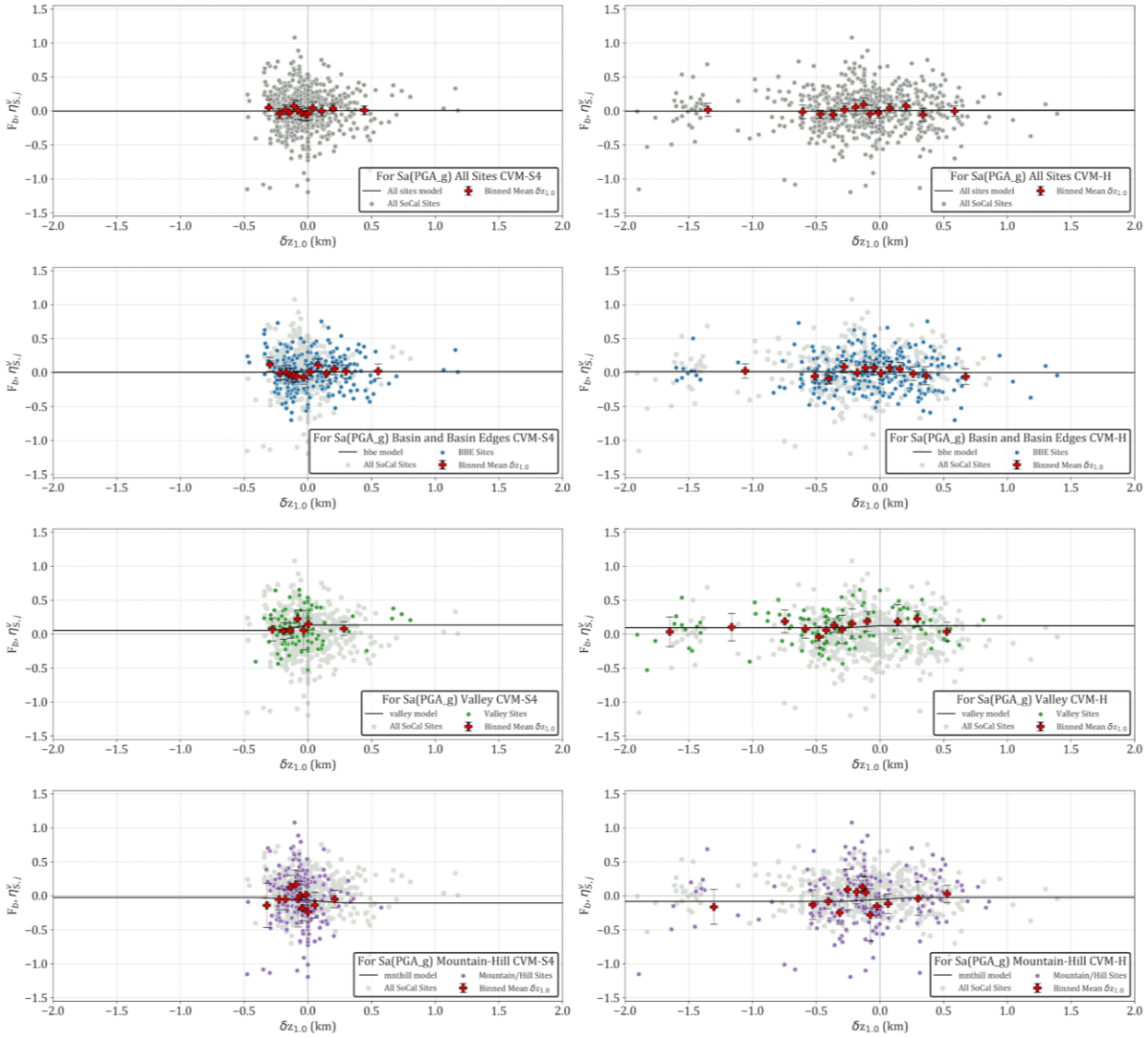


Figure 6.8. Category $\eta^{v}_{S,j}$ PGA data and mean fit per Eq. 6.2 against differential depth, with basin and basin edge groups combined.

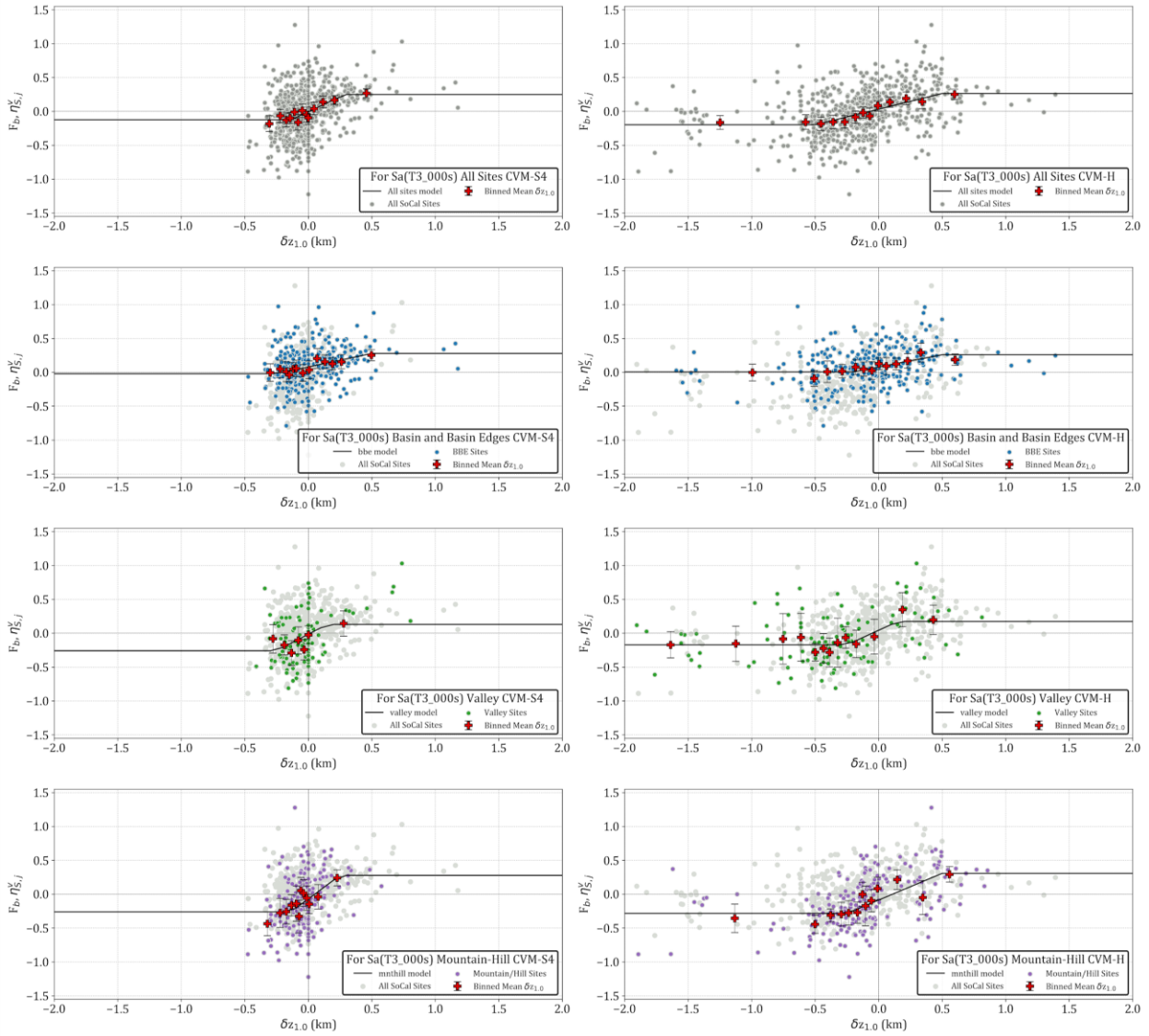


Figure 6.9. Category $\eta^v_{S,j}$ 3.0 sec Sa data and mean fit per Eq. 6.2 against differential depth, with basin and basin edge groups combined.

Figures 6.10 and 6.11 (same ground motion parameters) shows data and selected fits for basin groups, which BBE sites within a given sedimentary structure. The BBE sites are grouped into three sets:

1. Coastal basins, comprising LAB, SFB, and VB. These basins share a similar geologic origin.
2. IVB
3. Inland basins, generally bounded by faults and filled with sediments derived from surrounding mountains. These include SGB, CB, SBB, and CVB.

Set 1 is dominated by LAB and data for the other basins in this set do not reject the trend largely set by LAB. IVB (Set 2) has trends well represented by Eq. (6.2) and with some similarity to those for coastal basins in the case of the CVM-H model (no model is provided for CVM-S4; Section 3.5).

Set 3 is in many respects the most complex of the southern California basin groups. The common attribute of these basins, aside from their geology, is a lack of clear trends of $\eta_{S,j}^{v}$ with $\delta z_{1.0}$. As a result, we set f_6 to zero and take f_7 from the category mean $\eta_{S,j}^{v}$. Figure 6.12 shows these resulting basin amplifications along with the mean across the four inland basins (three for CVM-S4). For basins in the western U.S. having similar geologic characteristics to these inland basins, but limited data upon which to set f_7, we recommend using the mean values in Figure 6.12 and representing epistemic uncertainty in the mean amplification using the between-basin standard deviation of f_7 illustrated in Figure 6.12.

Figure 6.13a shows the variation of slope parameter f_6 and intercept parameter f_7 vs oscillator period for "all data" and the geomorphic provinces for which models are provided. At short periods (< 0.35 sec), the model provides no scaling for BBE, negative scaling for mountain/hill, and positive scaling for valleys. At long periods (> 0.65 sec), each group shows positive scaling. Figure 6.13b provides f_7 for the six modelled basin structures. Smoothed versions of the coefficients in Figures 6.13a-b are also shown, as described in the Section 6.4.4.

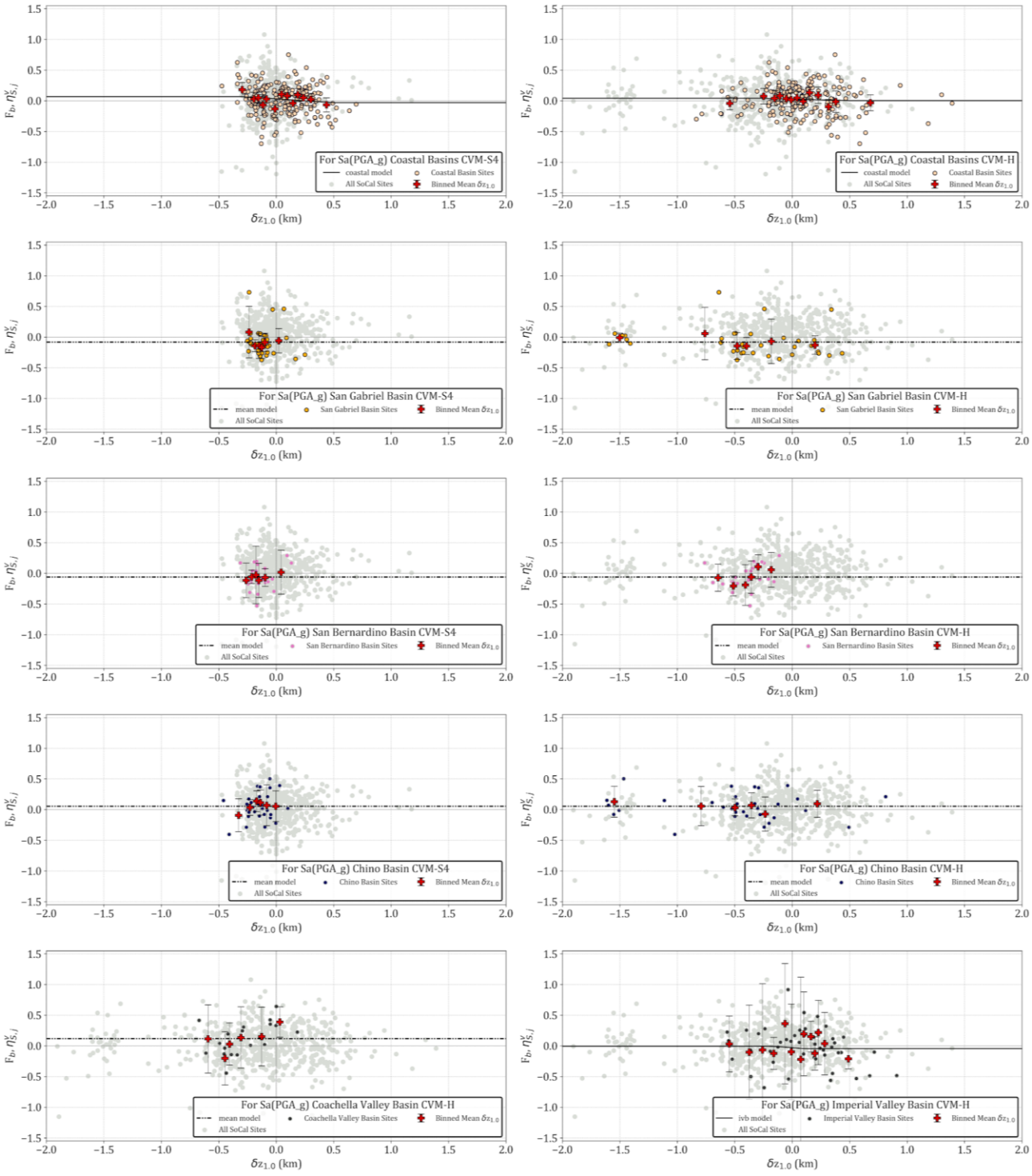


Figure 6.10. Basin category $\eta^{v}_{s,j}$ PGA data and mean fit per Eq. 6.2 for coastal, IVB, and inland basins.

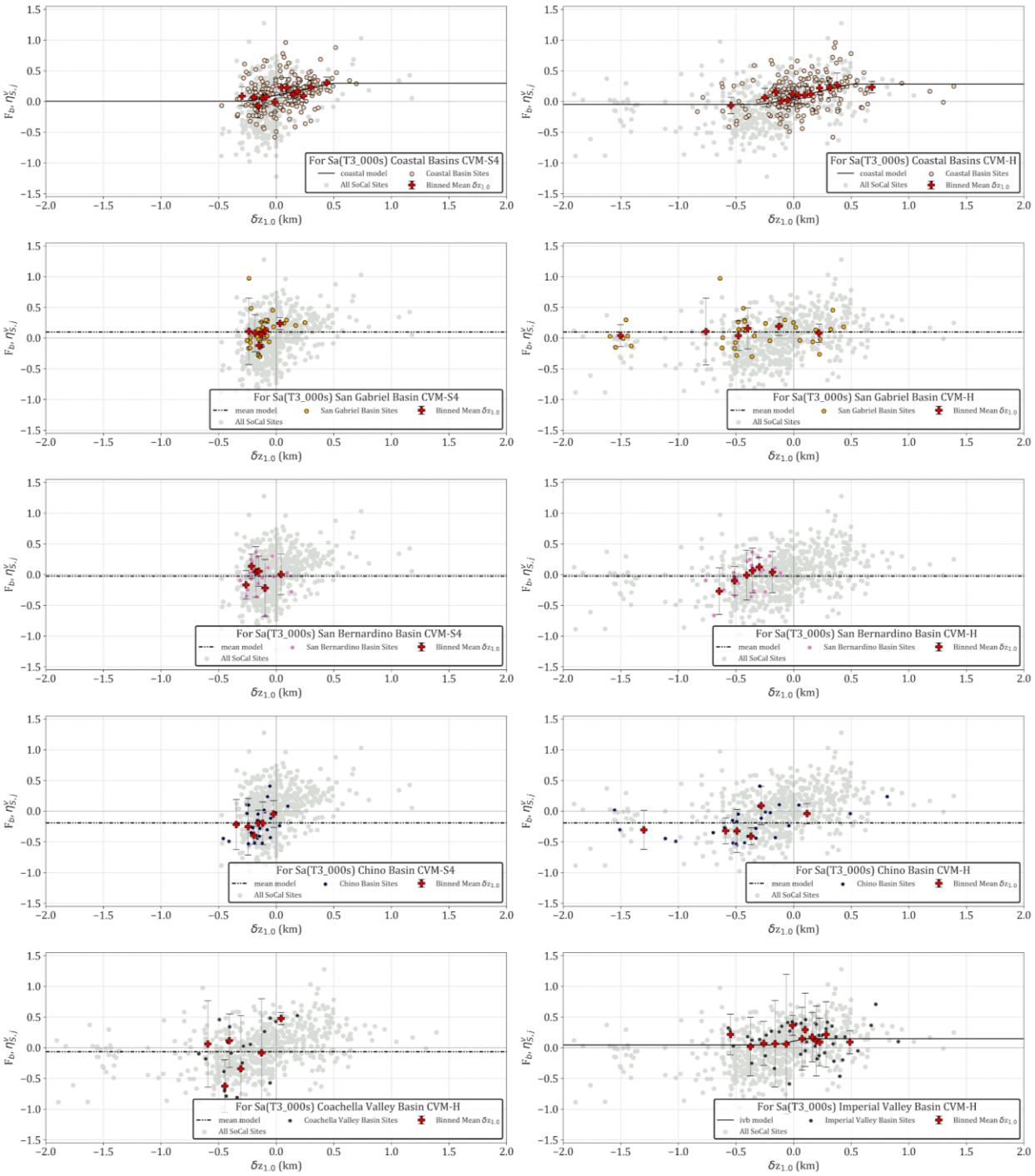


Figure 6.11. Basin category $\eta_{s,j}^{v}$ 3.0 sec Sa data and mean fit per Eq. 6.2 for coastal, IVB, and inland basins.

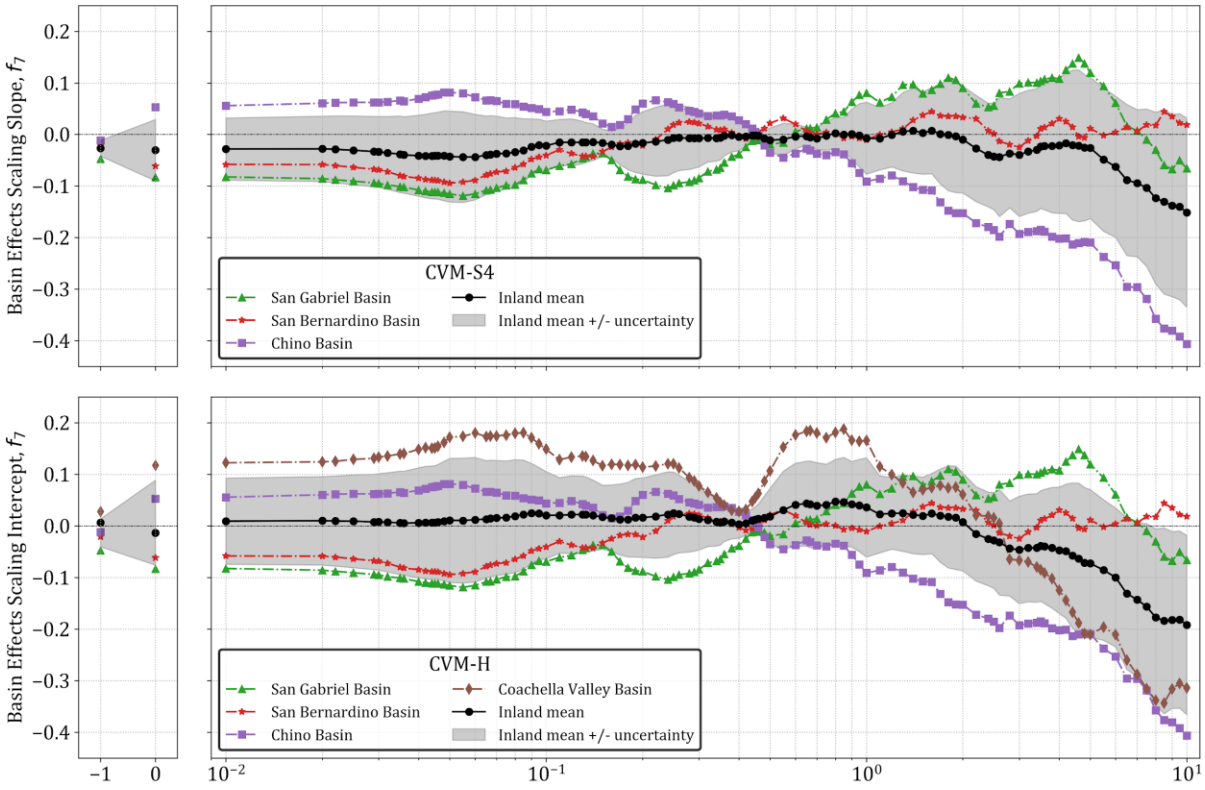
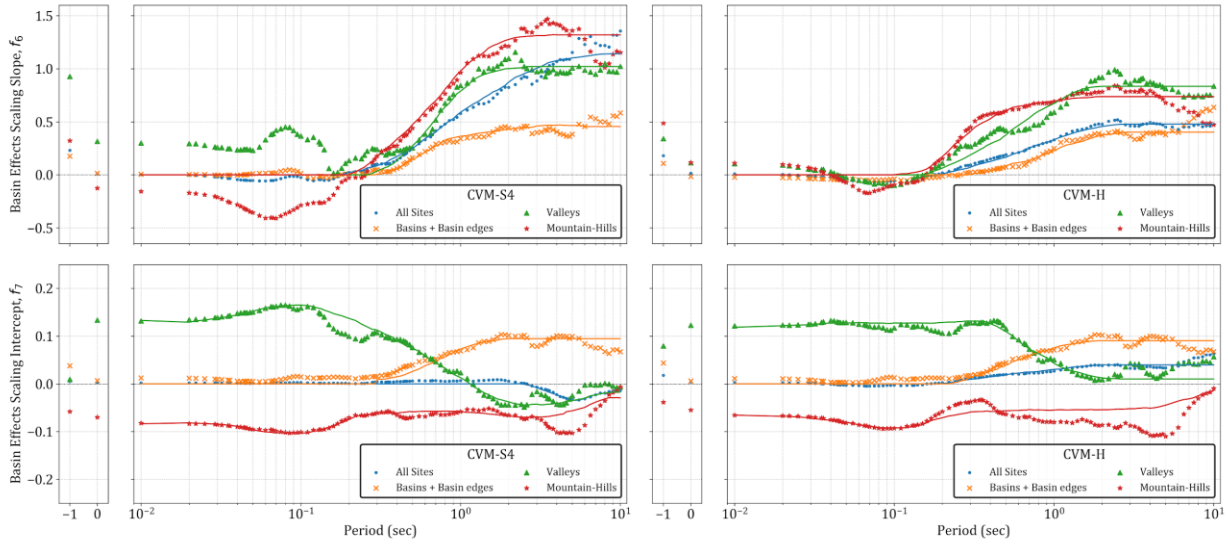


Figure 6.12. Comparison of depth-independent basin amplification modifiers (f_7) for inland basins SGB, CB, SBB, and CVB. Mean ± standard deviation of f_7 represent the epistemic uncertainty of this parameter

(a): f_6 and f_7 for geomorphic provinces



(b): f_6 and f_7 for individual and groups of basins

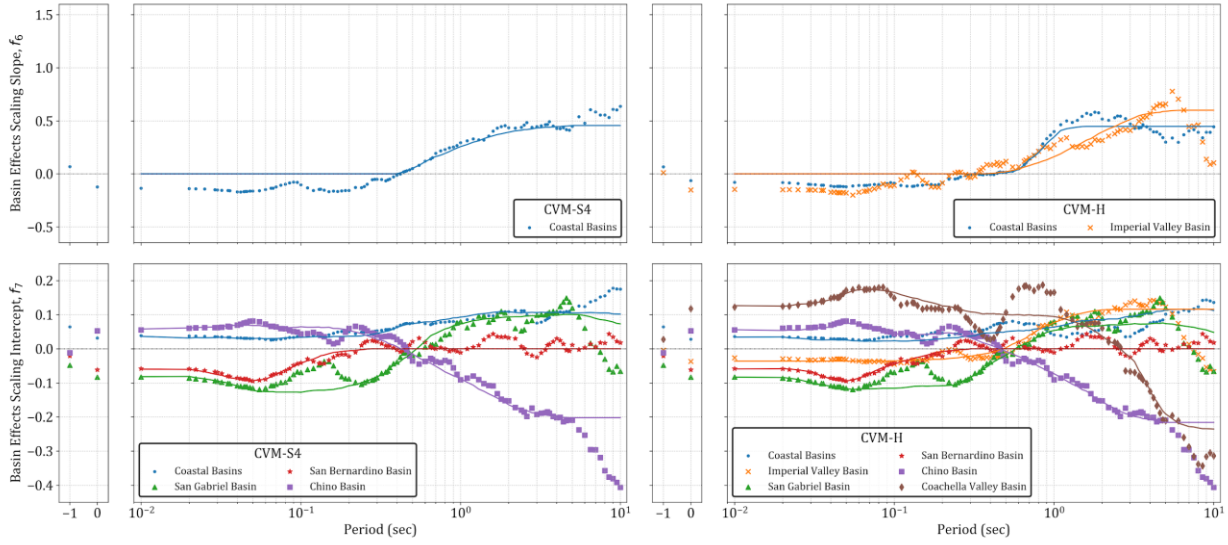


Figure 6.13. Comparison of depth-dependent basin amplification modifiers (f_6 and f_7) for (a) All site, BBE, Valley, and Mountain/Hill, and (b) Coastal basins and the individual inland basins

6.4.3 Significance Testing

This section provides a quantitative assessment of significance of gradients in the $F_b - \delta z_{1.0}$ relationships and the grouping decisions made during model development (Section 6.4.2). The latter is accomplished using F-Tests, which are used to evaluate the degree to which two data subsets are distinct from each other. We apply F-Tests to grouping decisions that were made in model development.

Gradient significance

We first test if the sample population is approximately normally distributed, which is assumed in significance testing for gradients and in F-tests. Figure 6.14 shows histograms of $\eta_{S,j}^{v}$ for the overall sample population (All sites) and individual geomorphic provinces and basin structures (basin, basin edge, valley, mountain-hill, LAB, SFB, SGB, SBB, CB) satisfy the required assumption of normally distributed data. Based on visual inspection of histograms and Q-Q plots, the distributions are approximately normal.

Figure 6.15 shows the results of tests on the statistical significance of gradients in the F_b-$\delta z_{1.0}$ relations (e.g., as shown in Figures 6.8-6.11). The significance of the slope is evaluated relative to a null hypothesis of a zero slope (f_6=0) model. A level of significance (p-value) is computed that indicates the degree to which the null hypothesis can be rejected − low values of p (< 0.05) indicates that this is the case and the slope can be considered significant. The calculations are performed using the SciPy python package (Virtanen et al. 2020). In Figure 6.15, the p-values are plotted against oscillator period for different geomorphic provinces (Figure 6.15a) and different basin structures (Part b) for both CVM-S4 and CVM-H.

Consider for example the "all sites" results in Figure 6.15. P-values are close to 1 for short periods (< 0.4 sec) and drop appreciably and are substantially below the threshold of 0.05 for T > 0.5 sec. Similar results are obtained for each geomorphic province, although the transition to statistically significant slopes occurs at longer periods for the valley and mountain hill provinces than for BBE. Among basin sites, the coastal basin group and IVB (with CVM-H) are the only cases where p < 0.05 at long periods. For other inland basins, relatively high p-values above the threshold indicate no statistically significant differential depth dependence, which supports the modeling decision to neglect this effect for these basins.

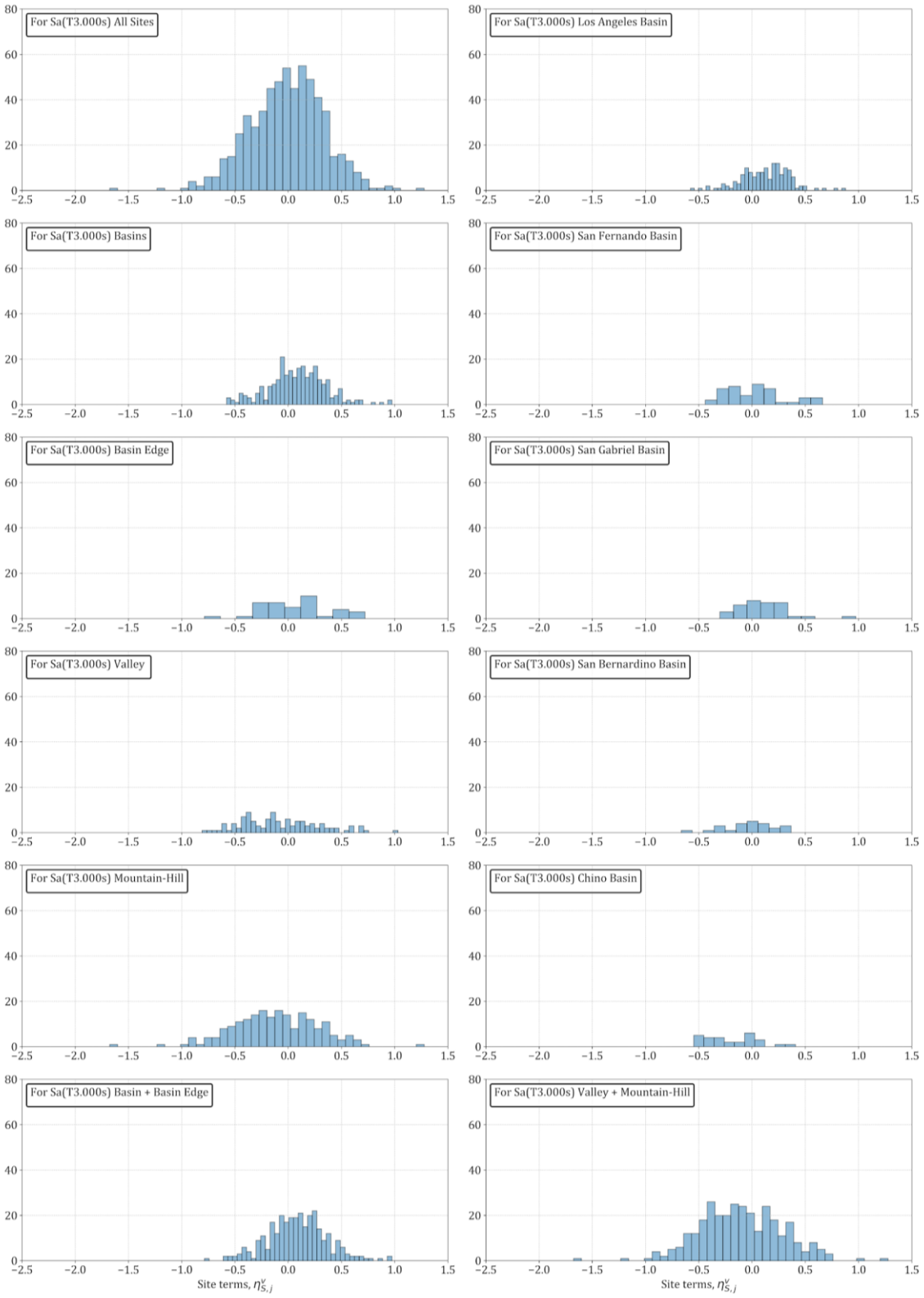
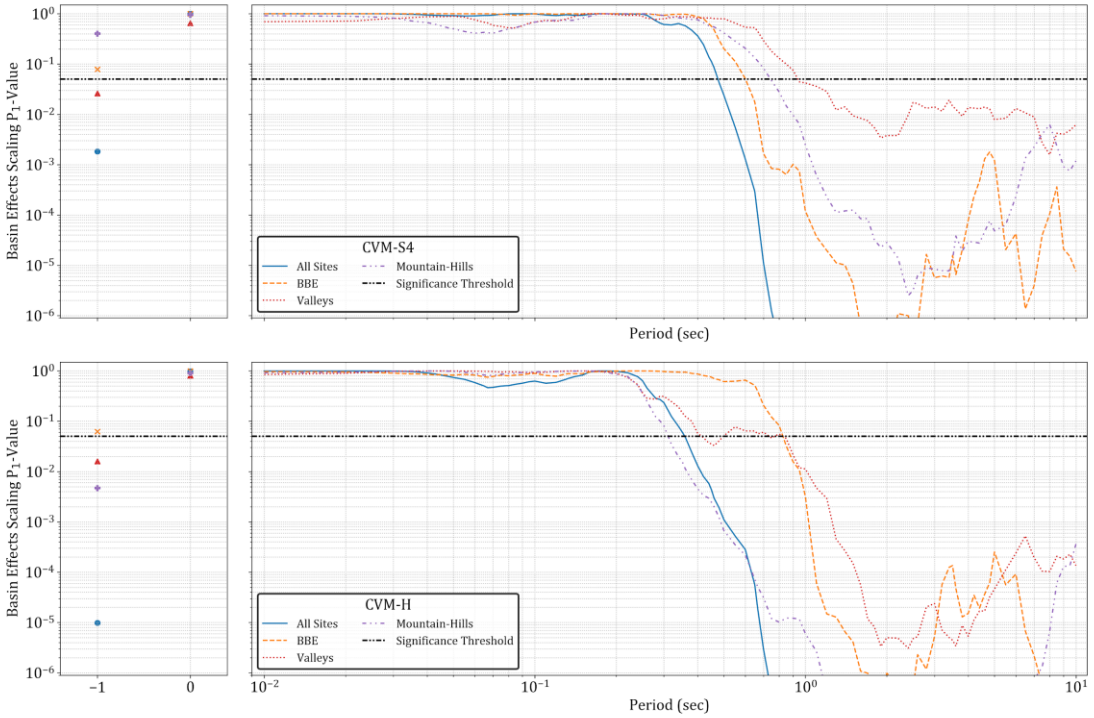


Figure 6.14. Distribution of $\eta^{v}_{s,j}$ for 3.0 sec S_a for all data and the different geomorphic provinces.

(a): Slope parameter significance tests, geomorphic provinces



(b): Slope parameter significance tests, individual and groups of basins

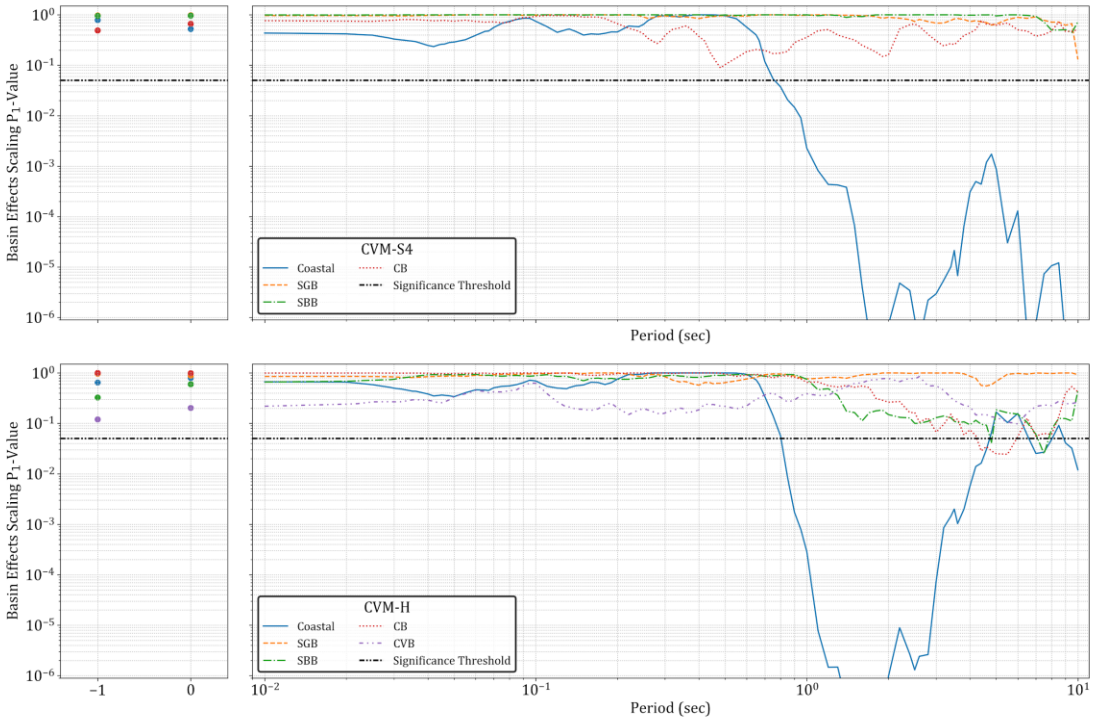


Figure 6.15. Significance of slope parameter f_6 for (a) geomorphic provinces for which models are developed and (b) combined basin groups and individual basins for which models are developed.

Sub-Category Distinction

We perform statistical testing to guide modeling decisions on whether regressed models from combined categories (e.g., BBE) are significantly different from alternate models fit to individual sub-categories (basin and basin Edge in this example). We consider these tests for the following grouping decisions:

- Geomorphic Provinces
    - Basin and basin edge vs BBE
    - Valley and mountain/hill vs combined group, VMH
- Basin structures
    - LAB and SFB vs coastal
    - Individual inland basins vs combined group

We applied a systematic process for deciding when data from sub-groups are statistically distinct; this same process was applied previously by Parker et al. (2017) and is summarized briefly here. We use an *F*-test (Snedecor and Cochran, 1989) that compares the statistical performance of submodels with that of a full model for a common data set. For example, BBE represents a full model, whereas submodels would comprise the basin and basin edge bins. The one-way analysis of variance (ANOVA) *F* statistic is computed as (adapted from Snedecor and Cochran, 1989):

$$F_1 = \frac{\left(RSS_f - (RSS_1 + RSS_2)\right) / \left((df_1 + df_2) - df_f\right)}{\hat{\sigma}^2} \tag{6.3}$$

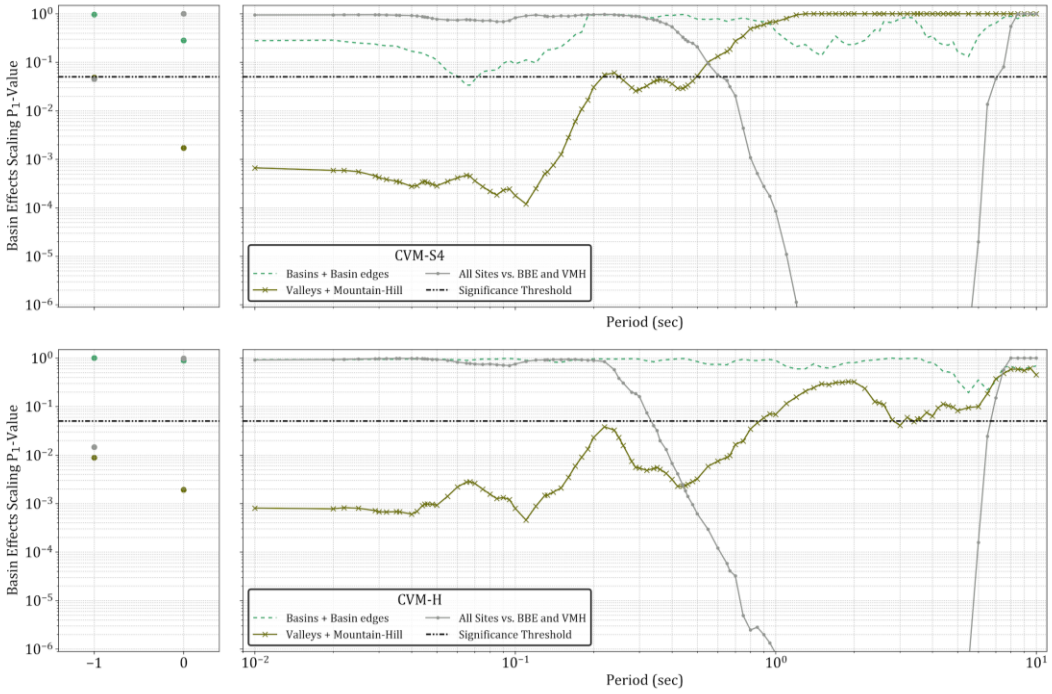
where *RSS* refers to residual sum of squares (based on misfits from mean model predictions) for submodels (subscripts *1* and *2*) and the full model (*f*), *df* refers to model degree of freedom (one if the model consists of a simple mean, four if the model uses the three parts of Eq. 6.2, and

$$\hat{\sigma}^2 = \frac{RSS_1 + RSS_2}{N_f - (df_1 + df_2)} \tag{6.4}$$

where N_f is the number of data points in the full model. By comparing an *F* statistic to the *F* distribution, significance level (*p*) is computed, which is used to judge potentially indistinct submodels. If F_1 has a *p* value ≤ 0.05, the subgroups are considered distinct.

Figure 6.16 shows the results of F tests of the first type vs oscillator period. Part (a) shows p-values for the proposed BBE group and the not-selected VMH group. Part (b) similarly shows results for the basin structure groupings.

(a): Subgroup distinction tests, geomorphic provinces



(b): Subgroup distinction tests, basin groups

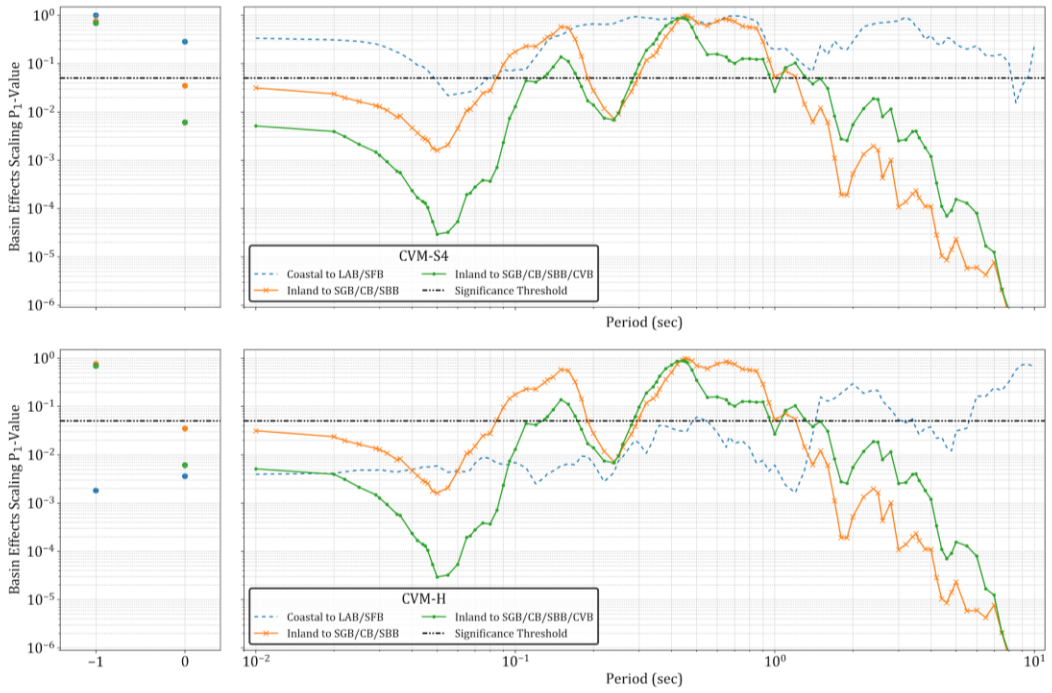


Figure 6.16. Trend of F-Test p-value as a function of oscillator period for (a) combined vs individual geomorphic provinces and (b) combined vs individual basins. -1 and 0 indicate PGV and PGA, respectively.

The results in Figure 6.16(a) show that p-values for the BBE tests are generally well above 0.05, indicating that the subgroups are not distinct. This supports the grouping that we have applied. P-values for the VMH group fall below the 0.05 threshold for short-periods (< 0.5 sec for CVM-S4 and < 0.8 sec for CVM-H). The distinction for these periods is mostly related to different mean amplification levels, and supports the use of separate models for the valley and mountain/hill categories.

The results in Figure 6.16(b) show that the p-values for the coastal group are generally above 0.05, which indicates that distinct models for LAB and SFB are not required by the data. P-values for the inland group are below 0.05 for significant period ranges (including T > 1 sec), which supports the use of distinct, basin-specific means for this group.

6.4.4 Coefficient Smoothing

The model coefficients in Eq. 6.2 were smoothed so as to avoid abrupt transitions ("corners") in predicted response spectra. The smoothed coefficients are those set by regression (f_6 and f_7). The parameters for limiting differential depths (f_8 and f_9) were selected by visual inspection in a manner than ensured smooth transitions between periods.

Figure 6.13 shows regressed values of f_6 and f_7 as discrete points vs oscillator period, and the smoothed representation of these parameters that are tabulated in Table E1 in the electronic supplement to this report. The smoothing was performed by visual inspection (essentially, drawing the lines in Figure 6.13 through the points).

6.5 MODEL PERFORMANCE

6.5.1 Residuals Analysis

We perform residuals analyses to confirm that the mean models described in Section 6.4 remove data trends with respect to the considered independent variables (i.e., provinces, basin designations, $\delta z_{1.0}$). Total residuals are computed using Eq. 4.1 with site term F_S taken as in Eq. 5.1. The basin term F_b in Eq. 5.1 is given in Eq. 6.2 and described in Section 6.4. Total residuals are partitioned into event terms (η_E), site terms (η_S), and remaining residual (ε) as shown in Eq. 4.4 using mixed effects analyses.

Figure 6.17 shows mean site terms for sites within the geomorphic provinces in Table 3.1. The means are near zero, and notably are much smaller than those computed without use of the basin model (Figure 6.2). Figure 6.18 shows mean site terms for sites within individual basins. These too are close to zero with the exception of Ventura.

Figure 6.19 shows S_a(3.0 s) site terms for all sites versus differential depth, $\delta z_{1.0}$, which can be compared to the result without the basin model in Figure 6.5. The trends with differential

depth are not completely removed at long periods. This occurs because of the weak trends with $\delta z_{1.0}$ for inland basins that are not modelled. Figures 6.20 and 6.21 similarly show site terms vs. $\delta z_{1.0}$ for the geomorphic provinces and individual basins/basin groups, respectively, that were considered in model development. The previously observed trends (Figures 6.6-6.7) are generally removed, although some features present over narrow ranges of $\delta z_{1.0}$ (i.e., the positive bias for mountain/hill near $\delta z_{1.0} = 0$) remain. An exception is the Ventura basin, for which a large positive residual remains; the limited available data indicates a very large basin effect that we did not attempt to capture in our model.

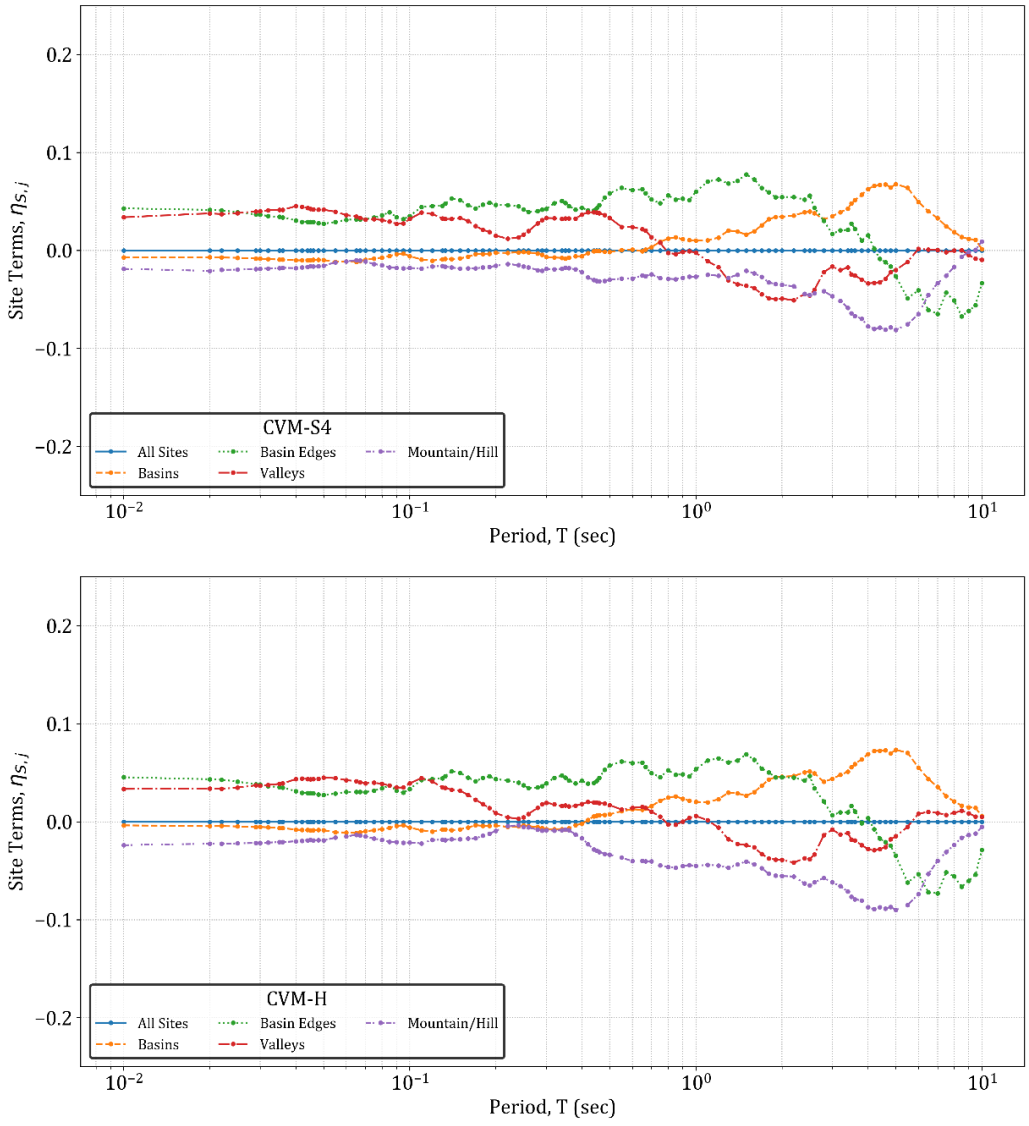


Figure 6.17. Mean of southern California site terms computed using the full site amplification model (including V_{S30} and basin components) as a function of oscillator period for the four geomorphic provinces in Table 3.1. A similar plot without consideration of the basin model appears in Figure 6.2.

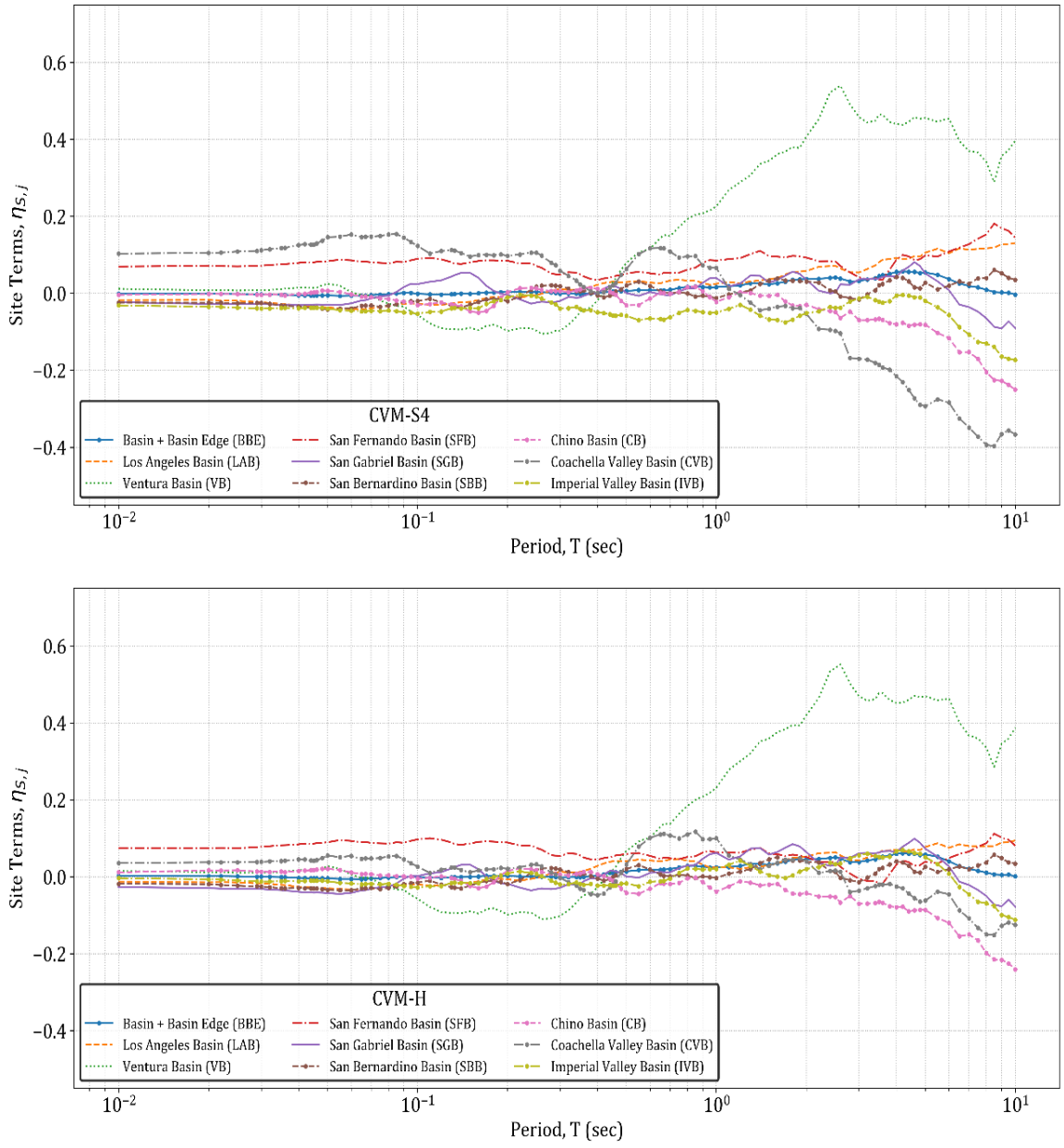


Figure 6.18. Mean of southern California site terms compute using the full site amplification model for individual basins

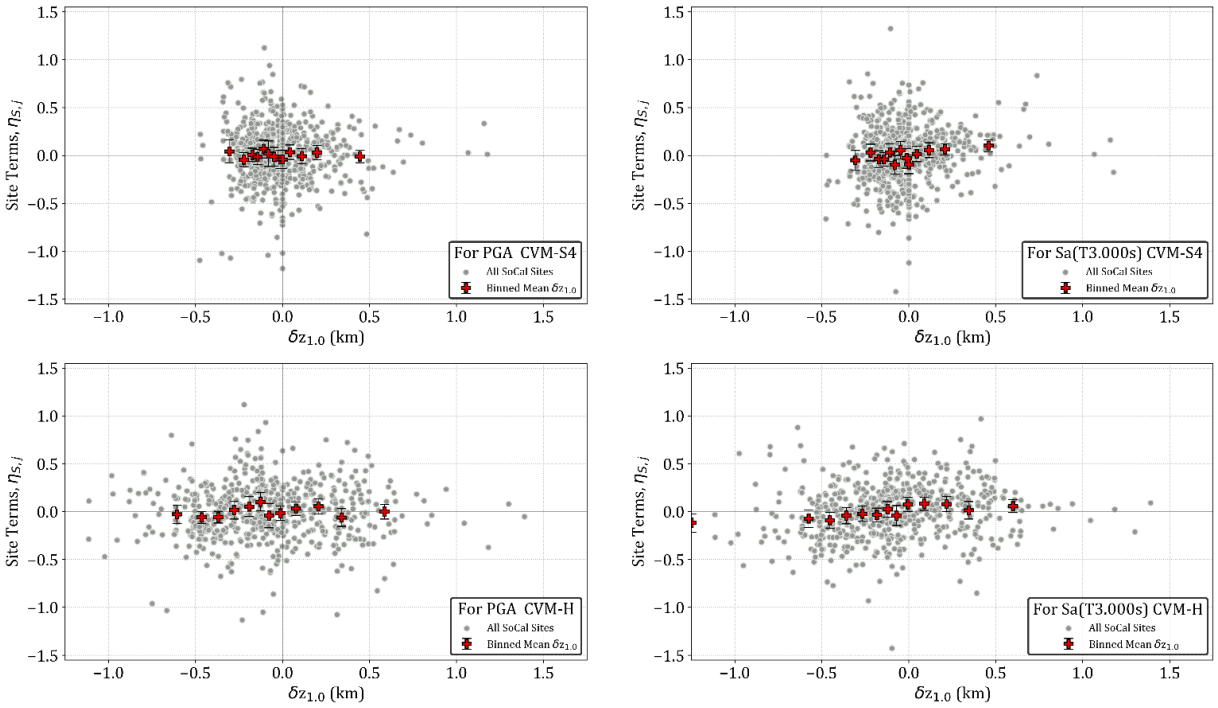


Figure 6.19. Variation of southern California site terms computed using the full site amplification model with differential depth $\delta z_{1.0}$ for all considered sites in the southern California region for intensity measures of PGA and $S_a(3.0)$. A similar plot without consideration of the basin model appears in Figure 6.5.

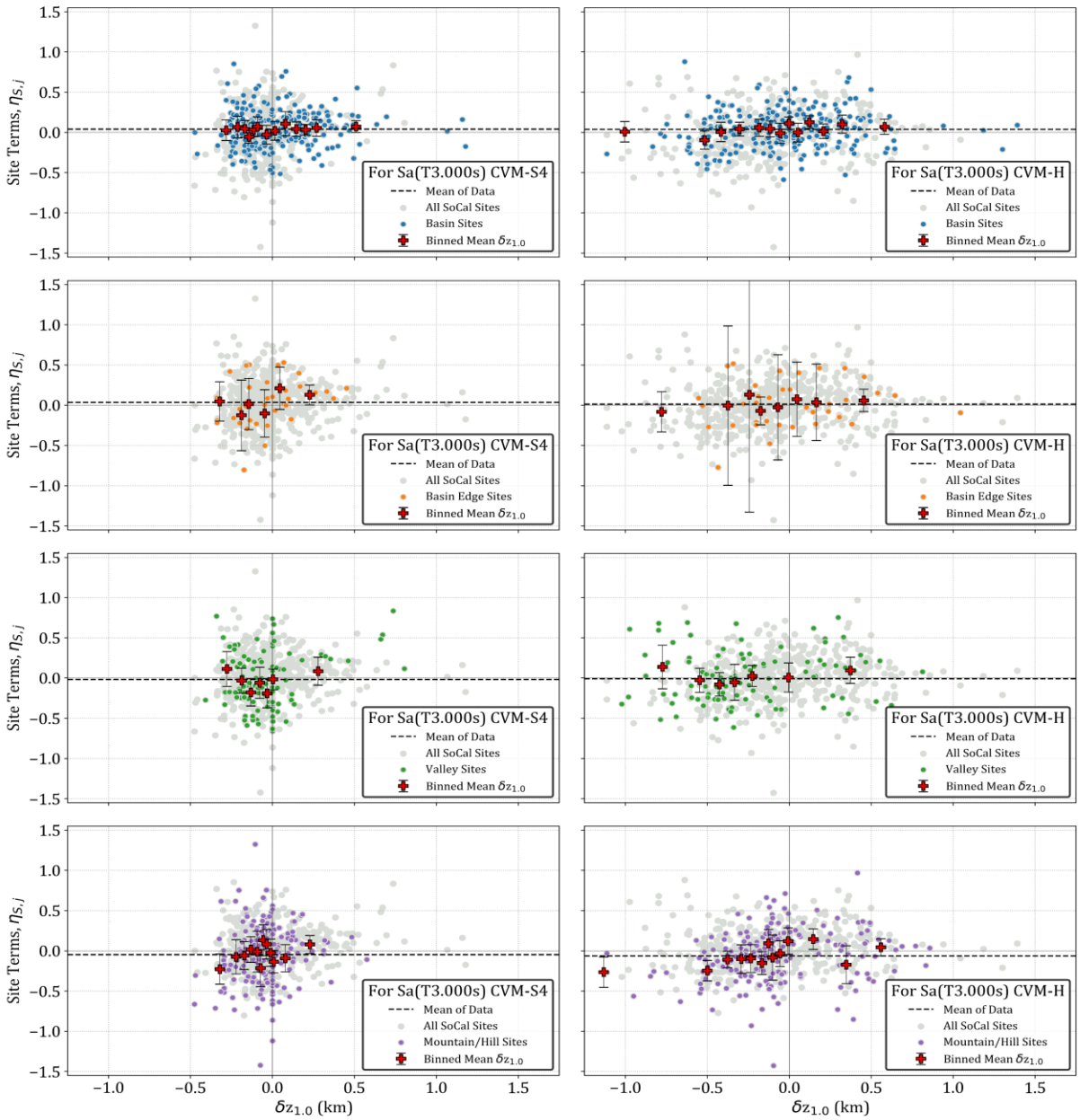


Figure 6.20. Variations of southern California $S_a(3.0)$ site terms computed using the full site amplification model with differential depth $\delta z_{1.0}$ at an intensity measure of $S_a(3.0)$ for the geomorphic provinces proposed in Table 3.1. A similar plot without consideration of the basin model appears in Figure 6.6.

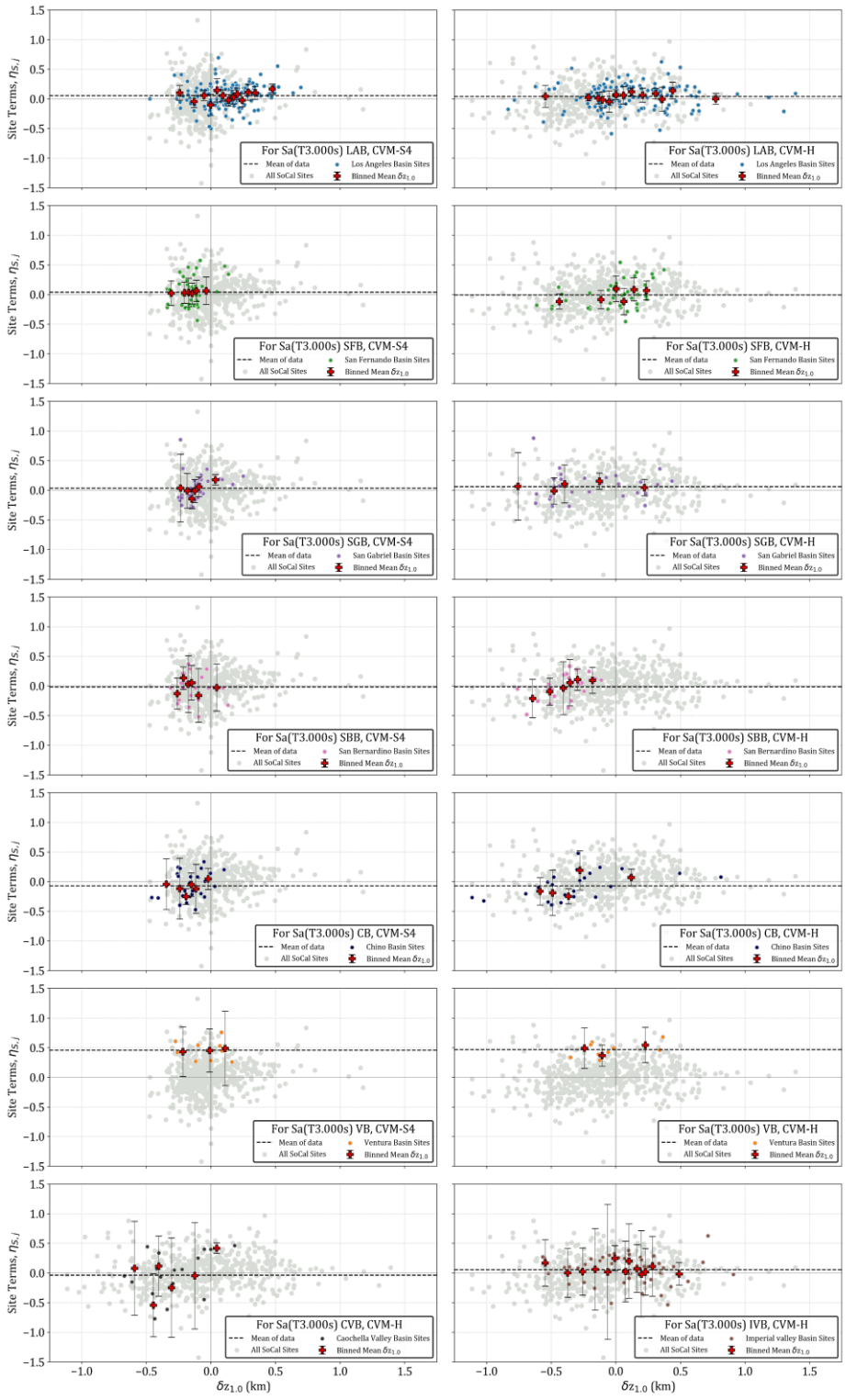


Figure 6.21. Variations of southern California $S_a(3.0)$ site terms computed using the full site amplification model with differential depth $\delta z_{1.0}$ for individual basins described in Section 3.1. A similar plot without consideration of the basin model appears in Figure 6.7.

6.5.2 Site-to-Site Variability

Site-to-site variability is quantified as the standard deviation of site terms and denoted as ϕ_{S2S}. Al Atik (2015) performed residuals analyses, similar to those employed for this study, for the full NGA-West2 data set, and based on those analyses, proposed a model for ϕ_{S2S}. Her analyses showed that ϕ_{S2S} is magnitude-dependent, with higher variability for oscillator periods < 1.0 sec for **M** < 5.5 events than for **M** > 5.5 events. At periods > 1.0 sec, the reverse was true (higher ϕ_{S2S} for larger **M** events). These results provide a useful baseline against which to compare our results. Here we illustrate ϕ_{S2S} features of the southern California dataset for comparison to the global model of Al Atik (2015), and develop an **M**- and site-dependent model.

Figure 6.22 compares ϕ_{S2S} for the expanded data set (described in Chapter 2) to the Al Atik (2015) model. The regional standard deviations for southern California are consistently lower than in the global model for both CVMs. This is expected, because ϕ_{S2S} reflects the effects of geologic variability on site response, after conditioning on V_{S30}. It is not surprising that this geological variability would be lower for a specific region than for a global model. For both the global model and the southern California data in this study, ϕ_{S2S} decreases with **M** at short periods (T < 1.5-2 sec) and increases with **M** at longer periods.

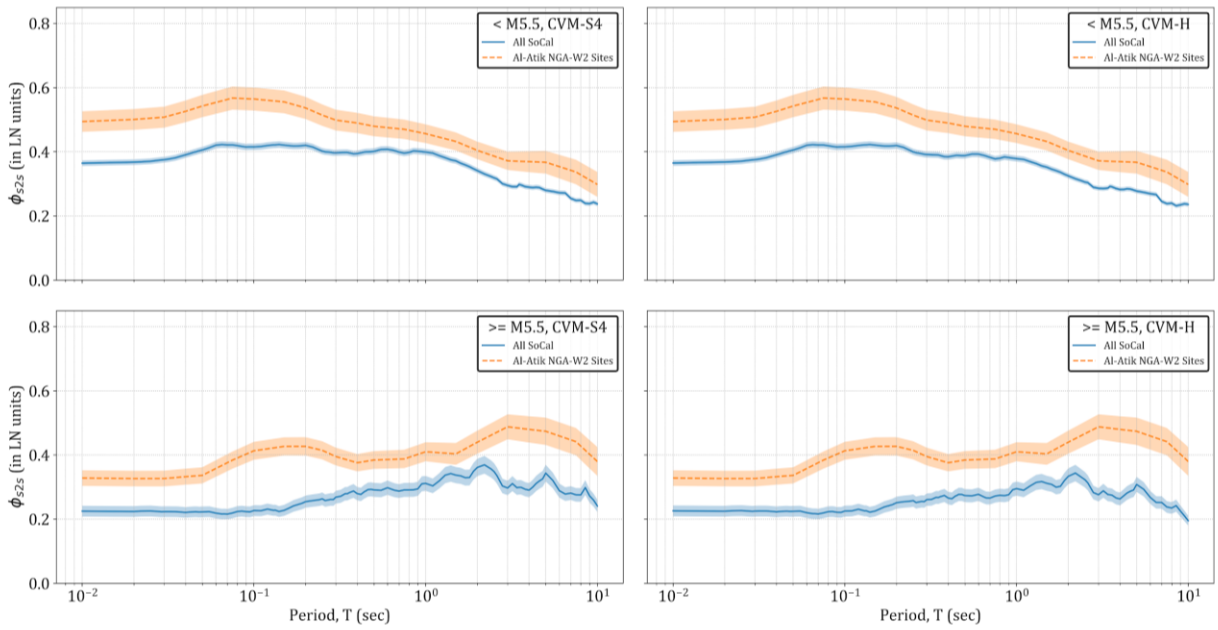


Figure 6.22. Site-to-site standard deviations, and their 95% confidence intervals, as a function of oscillator period for global data (Al Atik 2015) and the expanded data set for the southern California study region based on CVM-S4 (left panels column) and CVM-H (right panels column). **M** < 5.5 events (top panels row), **M** > 5.5 events (bottom panels row).

Figure 6.23(a) compares ϕ_{S2S} for sites within the geomorphic provinces outlined in Table 3.1 with the overall ϕ_{S2S} (across all sites) shown as a baseline for comparison. The results in Figure 6.23(a) apply to all magnitudes -- the use of all magnitudes is important to have sufficient amounts of data within each province to define statistically robust within-province ϕ_{S2S}. The variations among geomorphic provinces are appreciable. Basin (BBE) and valley sites have lower ϕ_{S2S} than mountain/hill sites. Based on this observation, and the magnitude-dependence of ϕ_{S2S} evident in Figure 5.18, we formulate a model for ϕ_{S2S} that is site condition- and **M**-dependent as follows:

$$\phi_{S2S} = \begin{cases} \phi_{S2S,1} & \mathbf{M} \leq 5 \\ \sqrt{\phi_{S2S,1}^2 - \Delta Var(M-5)} & 5 < \mathbf{M} < 6 \\ \sqrt{\phi_{S2S,1}^2 - \Delta Var} & \mathbf{M} \geq 6 \end{cases} \qquad \text{Eq. 6.3}$$

where $\phi_{S2S,1}$ is the site-to-site standard deviation for small magnitudes and ΔVar is the change in variance from small-to-large magnitudes.

The site-dependence of ϕ_{S2S} is contained in the $\phi_{S2S,1}$ terms, which are plotted in Figure 6.23(b). As shown in the figure, we proposed smoothed models for $\phi_{S2S,1}$ for BBE, valley, and mountain/hill sites. The **M**-dependence of ϕ_{S2S} is represented by the ΔVar term. We compute this term for "all sites" and individual geomorphic provinces by separately computing site terms for subsets of the data with **M** < 5.5 and **M** > 5.5 and differencing the variances, with the results shown in Figure 6.23(c). While the results for some categories are poorly constrained because of limited data, they nonetheless follow a similar pattern as represented by the recommended ΔVar model. The model was set by eye and is intended to be close to the "all data" result and intermediate between the mountain/hill and BBE results.

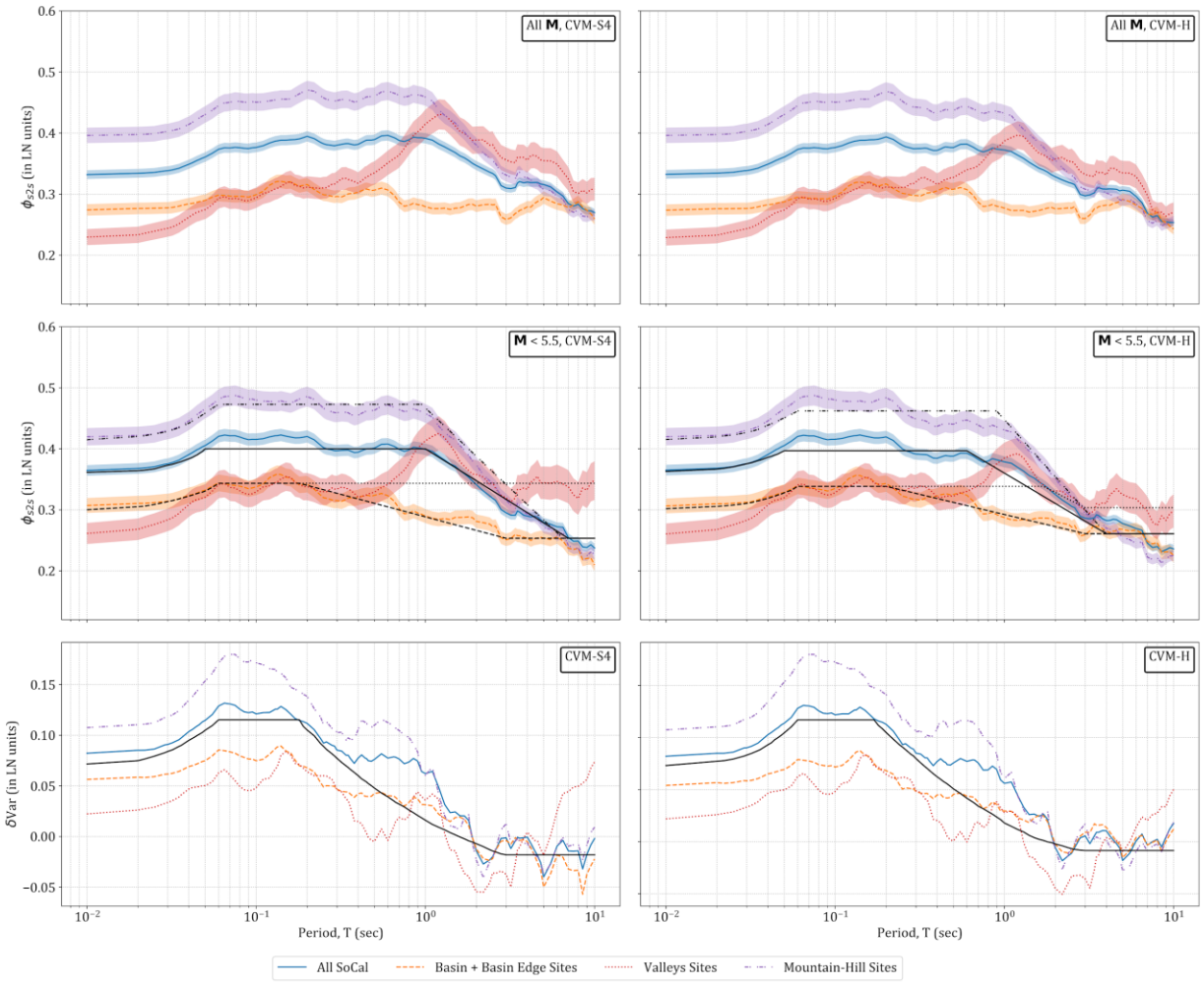


Figure 6.23. Site-to-site standard deviations, and their 95% confidence intervals, as a function of oscillator period for (a) all **M**, to show dependence on geomorphic province; (b) **M** < 5.5, to show the computed dispersions and the fitted model for $\phi_{S2S,1}$; and (c) ΔVar results by province and recommended model.

Figure 6.24 compares the all-**M** site-to-site variability for BBE sites (same as shown in Figure 6.23a) to all-**M** results for individual basins. The variations in ϕ_{S2S} between most basins are small relative to the variability between geomorphic provinces with the exception of the CVB and IVB. The deviation of the CVB at long periods is likely due to sparse data. The differences observed in the IVB variations is likely due to the distinct geology and subsurface geomorphic characteristics (Section 3.1). For all other basins, the variability is generally lower than the baseline BBE sites over the full period range. Overall, the results in Figure 6.24 are judged to not require ϕ_{S2S} models specific to individual basins.

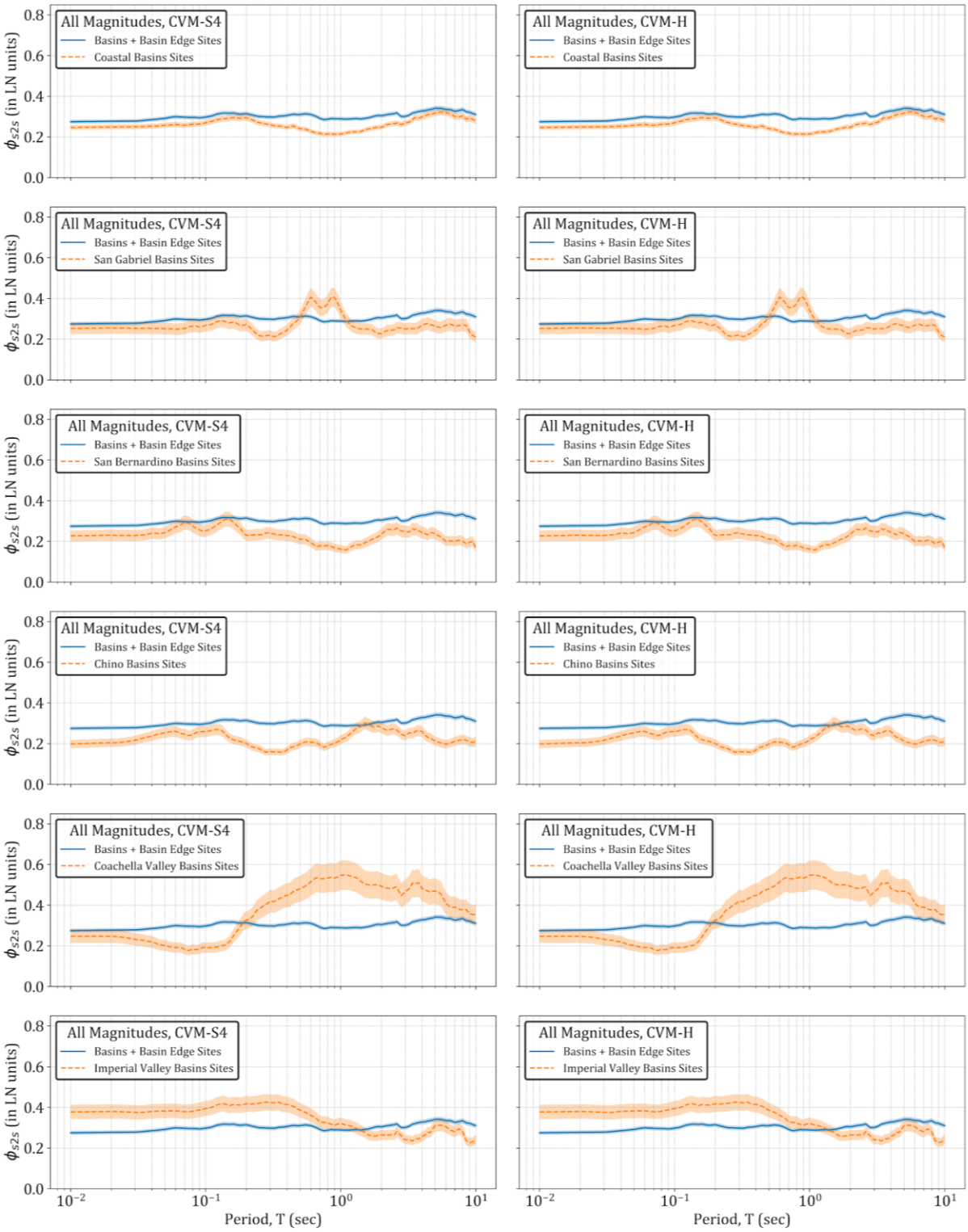


Figure 6.24. Variations of site-to-site standard deviations, and their 95% confidence intervals, as a function of period for all the southern California basin structures described in Section 3.1.

6.5.3 Site Amplification Plots

In this section we demonstrate the features of the proposed site amplification models, which combine V_{S30}-scaling with the F_b models in Eq. 6.2 with coefficients in Table E1. To begin, consider three sites with V_{S30}=300 m/s, weak shaking conditions (such that $F_{nl} = 0$), and $\delta z_{1.0} = f_8$ (lower limit of scaling), 0, and f_9 (upper limit of scaling) for the "all sites" model derived for CVM-S4. Figure 6.25 shows site amplification (ln units) vs oscillator period for these sites, from which the modification of site amplification by the F_b model is apparent. This version of the model is most directly comparable to the basin model in BSSA14, which is shown for $\delta z_{1.0}$ = -0.5, 0, and 0.5 km. The model from this study (labelled as "SoCal") is very similar to that from BSSA14 for the case of $\delta z_{1.0} = 0$, which reflects the similarity of the two V_{S30}-scaling models. For the case of finite differential depths, the SoCal models produces smaller changes at the limits of $\delta z_{1.0}$-scaling than does BSSA14 (less de-amplification for negative $\delta z_{1.0}$, less amplification for positive $\delta z_{1.0}$).

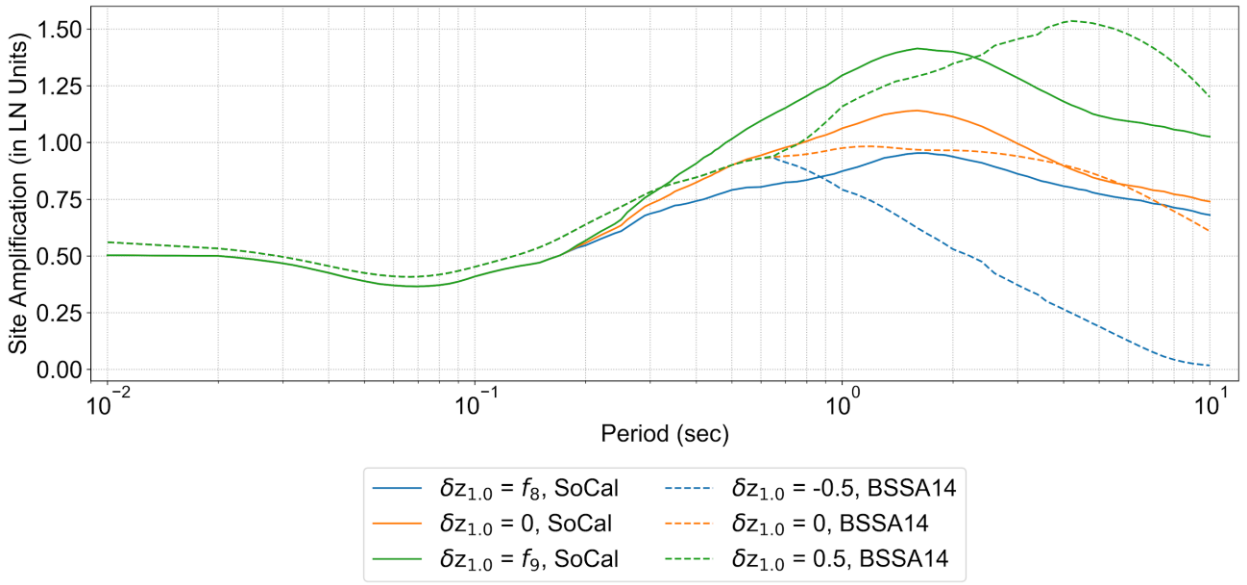


Figure 6.25. Total site amplification vs. oscillator period for sites with V_{S30} = 300 m/s, weak input motions such that nonlinear effects vanish (F_{nl} = 0), and variable levels of differential depth ($\delta z_{1.0}$), for "all sites" model (labelled as "SoCal") and BSSA14 model. $\delta z_{1.0}$ is sampled from lower limit, zero, and upper limit of scaling

Figure 6.26(a) shows amplification vs period for the $\delta z_{1.0} = f_8$ case (lower limit of scaling) for "all sites", mountain/hill, valley, and BBE. The soil and "all sites" cases are shown with V_{S30}=300 m/s and the rock case (mountain/hill) are shown with V_{S30}=600 m/s. Figures 6.25(b) and (c) provide similar comparisons across provinces for the cases of $\delta z_{1.0} = 0$ and $\delta z_{1.0} = f_9$ (upper limit of scaling). Features of the model revealed by these plots is that the largest deamplification for negative differential depths occurs for the mountain/hill category whereas the largest amplification for positive differential depths occurs for the BBE category. Comparing the valley

and BBE categories, the former has more short-period amplification, and the latter has more long-period amplification.

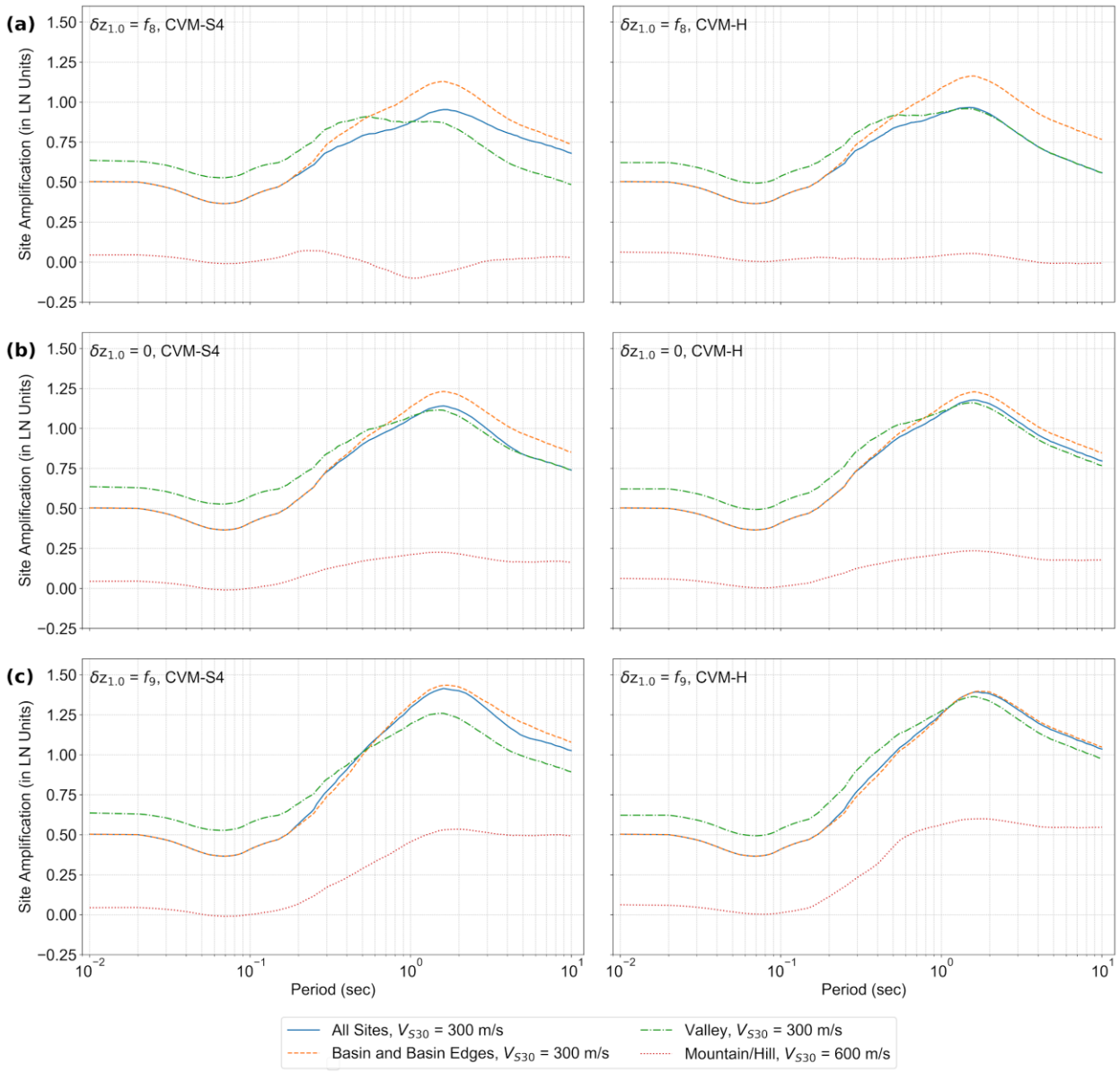


Figure 6.26. Total site amplification vs. oscillator period for sites with V_{S30}=300 and/or 600 m/s, weak input motions such that nonlinear effects vanish (F_{nl}=0), and variable levels of differential depth ($\delta z_{1.0}$), for following cases: (a) sites at lower limit of scaling $\delta z_{1.0}=f_8$ for various geomorphic provinces, with V_{S30}=300 and 600 m/s applied to soil/all and rock provinces, respectively; (b) sites at centered condition $\delta z_{1.0}$=0 for various geomorphic provinces, with V_{S30}=300 and 600 m/s applied to soil/all and rock provinces, respectively; and (c) sites at upper limit of scaling $\delta z_{1.0}=f_9$ for various geomorphic provinces, with V_{S30}=300 and 600 m/s applied to soil/all and rock provinces, respectively.

Figure 6.27 shows amplification vs period for different basin structures, using V_{S30}=300 m/s in each case. Shown are the coastal basin model for $\delta z_{1.0} = f_8$, 0, and f_9, and the inland basin models (individual basins, inland mean, and inland mean ± standard deviation). The inland basins other than IVB (i.e., SGB, CB, SBB, CVB) are unchanged across the three plots because the models are independent of depth. The band of amplifications for these inland basins (e.g., natural log amplification ranging from 0.9-1.2 at T=2 sec) provides a frame of reference for visualizing the responses of coastal basins and IVB, which are $\delta z_{1.0}$-dependent. The coastal basins and IVB have large $T > 2$ sec amplification for all differential depths, but especially for the $\delta z_{1.0} = f_9$ case. This amplification exceeds the range from inland basins, which might be expected given the large sizes/depths of the coastal and IVB cases.

The coastal and IVB results are similar in Figure 6.27, although this is potentially misleading because a consistent V_{S30} = 300 m/s was used for all basins, whereas velocities for the IVB case are lower for many sites (Figure 5.1). Among the inland basins, CB produces low amplification, SGB high amplification, CVB high amplification for a relatively narrow period range (peak near 1.5 sec), and SBB intermediate levels of amplification.

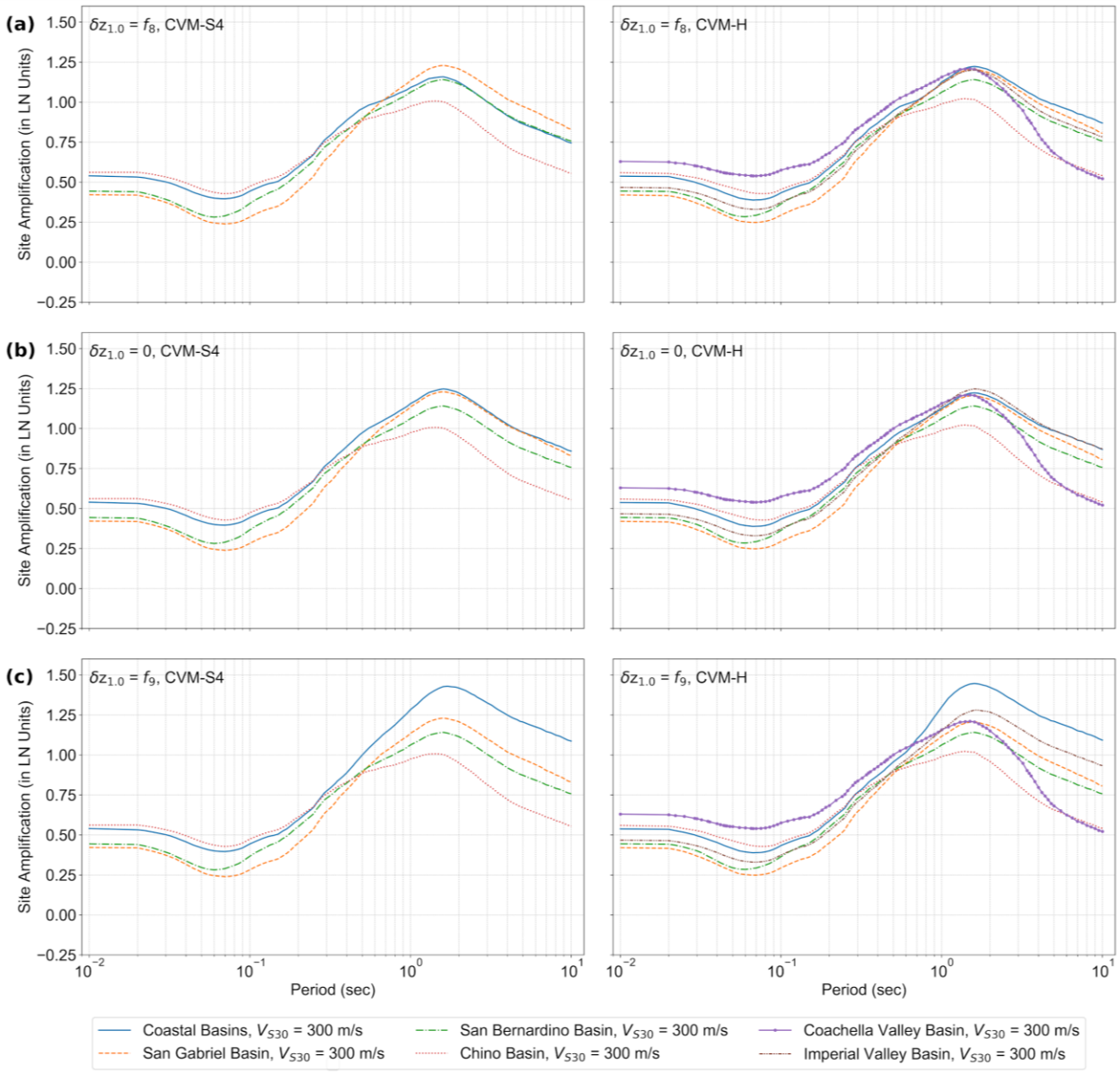


Figure 6.27. Total site amplification vs. oscillator period for sites with V_{S30}=300 m/s in different basin structures and weak input motions such that nonlinear effects vanish (F_{nl}=0): (a) $\delta z_{1.0}=f_8$, (b) $\delta z_{1.0}= 0$, (c) $\delta z_{1.0}=f_9$.

7 Conclusions and Recommendations

7.1 SUMMARY OF STUDY

The V_{S30}-scaling and basin differential depth-scaling relations in some of the NGA-West2 ground motion models are ergodic, meaning that they are intended to represent average site response for a large data set. While regionalization of site response was checked for V_{S30}-scaling in the development of the original NGA-West2 models, regionalization has not previously been considered for basin depth models. In this study we have investigated regional site response in terms of V_{S30}-scaling and differential depth scaling for different geomorphic provinces and sedimentary basin structures.

To facilitate this investigation, we have proposed a morphology-based site categorization scheme intended to distinguish sites in large sedimentary basins from sites in smaller sedimentary structures (valleys), along basin edges, and in non-basin areas. We also classify different basin structures having a variety of geological origins, including coastal basins, inland fault-bounded basins, and plate-boundary basins associated with fault stepover. These site classifications are provided for a large number of ground motion recording sites in southern California. These sites are included within a ground motion and meta-data relational database that significantly expands the NGA-West2 database for applications in southern California.

We develop a regional site amplification model with two components. The first is a regional V_{S30}-scaling relation that only slightly modifies an NGA-West2 model (SS14). This model is used for "centering" purposes -- namely, subsequent model components modify the predictions of this model. We provide models for mean depth to 1.0 km/s shear wave isosurfaces conditioned on V_{S30}, from which a differential depth can be defined for any particular site. These mean depth models differ for the CVM-S4 and CVM-H SCEC basin models.

The second site amplification model component provides domain-specific models for modifying site amplification relative to the V_{S30}-scaling relation. These domains include various geomorphic provinces inclusive of basin and non-basin conditions, as well as different basin structures having a variety of geological origins. Through these analyses, we seek to gain insight into attributes of basins beyond depth that affect site response and its variability.

7.2 RECOMMENDED MODEL

7.2.1 Model Summary

We take site amplification in southern California as the mean change in ground motion relative to a 760 m/s reference condition. We recommend to compute site amplification defined in this manner from Eq. (5.1), where the F_{lin} term (linear V_{S30}-scaling) is taken from Eq. (5.4), the F_{nl} term (for nonlinearity) is taken from literature (SS14), and the F_b term (so-called basin effect) is taken from Eq. (6.2). The F_b term is conditioned on geomorphic domains (Table 3.1, Figure 2.4) and $\delta z_{1.0}$, where the latter is depth to 1.0 km/s shear wave isosurface minus mean depth conditioned on V_{S30} (Eq. 6.1). Recommended mean depth models are given in Eq. (3.3).

Application of the mean site amplification models described above affects site-to-site standard deviation (ϕ_{S2S}). We recommend a magnitude-dependent ϕ_{S2S} model that varies with period and geomorphic domain, as given by Eq. (6.3).

All model coefficients are provided in Table E1 (electronic supplement to this report), with the exception of the mean depth model, for which coefficients are provided in Table 3.5.

7.2.2 Model Attributes

The linear V_{S30}-scaling model developed in this project is similar to prior models (e.g., SS14). The main changes in the proposed model relative to prior models concern the basin component and site-to-site standard deviation terms (ϕ_{S2S}).

When data sites across all geomorphic provinces are combined, we do not find significant differences from current basin models. However, the present analyses have elucidated the source of certain features of those models, namely:

- De-amplification for negative differential depths does not occur in basins or basin edge regions, but rather are mainly derived from mountain/hill regions.
- Amplification for positive differential depths occurs in all geomorphic provinces.
- Amplifications features in sedimentary structures have substantial variations, being different for:
    - Large sedimentary structures (basins) vs smaller structures (valleys). The main impact is in the period range of the amplification, being longer in basins than in valleys
    - Coastal basins with complex depositional processes that include marine sediments, vs inland basins that tend to be fault-bounded and are filled with alluvial sediments. With the exception of IVB, we do not find a dependence of amplification on

differential depth for inland basins, whereas these effects are present at long periods for coastal basins and IVB.

As in a prior global model (Al Atik 2015), we find ϕ_{S2S} to be **M**-dependent. Our overall regional model for ϕ_{S2S} has lower variability than in the global model. However, there are variations in ϕ_{S2S} by geomorphic province, being lower in basins than other provinces.

7.3 LIMITATIONS

The applicability of the proposed basin categorization scheme presented in Chapter 3 is limited to the region of southern California, or possibly regions with similar geologic history and surface features. Potential regions include Mojave desert, central California, Baja California, southern Arizona, and southwestern New Mexico. The procedures are not expected to be applicable to regions that have experienced glaciation, due to the alteration of the basin surface topography.

The ground motion models developed here are applicable to the southern California region shown in Figure 2.4 and the geomorphic domains therein (Table 3.1). The data used to develop these models is mostly of modest magnitude (**M** = 4-6). As with the current basin components of NGA-West2 models, it is possible that there are magnitude dependencies in site amplification that are not captured here.

The linear V_{S30}-scaling model is applicable for V_{S30} = 200-1500 m/s.

The models are strictly applicable to the seven basin structures depicted in Figure 2.4. We anticipate that there will be applications for other structures, especially for inland basins, that were not considered in this study. Mean amplification for such basins could, in principle, be derived from within-basin recordings. If such studies are not available, the mean amplification model for inland basins can be applied with consideration of its epistemic uncertainty using logic trees.

7.4 FUTURE WORK

The research approach developed in this study can be applied to any active region with basin structures and ample ground motion recordings. By developing regional basin models, it will be possible to improve the accuracy of mean site response and reduce site-to-site standard deviations as applied in ground motion hazard analyses. Some particularly important regions that could be targeted in future work are the San Francisco Bay Region, the California Central Valley, and basin structures beneath urban centers in the intermountain west region of the U.S (e.g., Las Vegas, Salt Lake City, Reno, Denver.

While the parameterization of basin effects considered here improves upon prior models, it is still a relatively simple model of a more complex phenomenon. For example, the present model does not consider the location of the source relative to the basin or direction from which seismic

waves enter the basin, which prior work has found to be influential (e.g., Choi et al. 2005; Graves et al. 2008). Validated ground motion simulations, for which there is no practical limitation in the amount of data that can be considered, could inform the development of alternative model framework that consider these effects.

REFERENCES

Ahdi, S. K., Sadiq, S., Ilhan, O., Bozorgnia, Y., Hashash, Y. M., Kwak, D. Y., Park, D., Yong, A., & Stewart, J. P. (2018). Development of a United States community shear wave velocity profile database. In *Geotechnical Earthquake Engineering and Soil Dynamics V: Seismic Hazard Analysis, Earthquake Ground Motions, and Regional-Scale Assessment* (pp. 330-339). Reston, VA: American Society of Civil Engineers.

Ahdi, S. K., Mazzoni, S., Kishida, T., Wang, P., Nweke, C. C., Kuehn, N. M., Contreras .V, Rowshandel B., Stewart J.P., & Bozorgnia, Y. (2020). Engineering characteristics of ground motions recorded in the 2019 Ridgecrest earthquake sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1474-1494.

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

Al Atik, L., (2015). NGA-East: Ground motion standard deviation models for Central and Eastern North America, *PEER Report No. 2015/09,* Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Allen, P. A., and Allen, J. R., (2013). *Basin Analysis: Principles and Application to Petroleum Play Assessment*. John Wiley & Sons.

Alles, D. L. (2011). Geology of the Salton Trough. *Edited by DL Alles, Western Washington University. http://fire. biol. wwu. edu/trent/alles/GeologySaltonTrough. pdf on February*, *25*, 2012.

Ancheta, T.D., Darragh, R.B., Stewart, J.P., Seyhan, E., Silva, W.J., Chiou, B.S.-J., Wooddell, K.E., Graves, R.W., Kottke, A.R., Boore, D.M., Kishida, T., and Donahue, J.L., (2014). NGA-West2 database*, Earthquake Spectra*, 30, 989-1005.

Anderson, M., Matti, J., and Jachens, R., (2004). Structural model of the San Bernardino basin, California, from analysis of gravity, aeromagnetic, and seismicity data. *J. Geophy. Res.*: *Solid Earth*, 109(B4).

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

Bates D, 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.

Boore, D.M., (2010). Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion, *Bull. Seismol. Soc. Am*., 100, 1830–1835.

Boore, D. M., Thompson, E. M., & Cadet, H. (2011). Regional correlations of VS 30 and velocities averaged over depths less than and greater than 30 meters. *Bulletin of the Seismological Society of America*, *101*(6), 3046-3059.

Boore, D. M., Stewart, J. P., Seyhan, E., & Atkinson, G. M. (2014). NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthquake Spectra*, *30*(3), 1057-1085.

Brocher, T. M., Aagaard, B.T., Simpson, R.W., and Jachens, R.C., (2006). The USGS 3D seismic velocity model for northern California, presented at the 2006 Fall Meeting, AGU, San Francisco, California, 11–15 December, Abstract S51B–1266.

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

Chen, P., and Lee, E. J., (2017). UCVM 17.3.0 Documentation. Retrieved from http://hypocenter.usc.edu

Chiou, B. S. J., & Youngs, R. R. (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, *30*(3), 1117-1153.

Choi, Y, JP Stewart, and RW Graves (2005). "Empirical model for basin effects that accounts for basin depth and source location," Bull. Seism. Soc. Am., 95 (4), 1412-1427.

Contreras, V., Stewart, J. P., Kishida, T., Darragh, R. B., Chiou, B. S. J., Mazzoni, S., Kuehn, N., Ahdi, S.K., Wooddell, K., Youngs, R.R., & Bozorgnia, Y. (2020). Chapter 4: source and path metadata. *Data Resources for NGA-Subduction Project, PEER Report*, *2*.

Curtis, S.M. and Ghosh, S.K. (2011). A Bayesian approach to multicollinearity and the simultaneous selection and clustering of predictors in linear regression, *J. Statistical Theory and Practice*, 5:4, 715-735, DOI: 10.1080/15598608.2011.10483741

Danielson, J. J., & Gesch, D. B. (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)* (p. 26). US Department of the Interior, US Geological Survey.

Dibblee, T. W., and Ehrenspeck, H. E. (1998). Geologic map of the Mt. Wilson and Azusa quadrangles, Los Angeles County, California, Dibblee Geological Foundation, Dibblee Foundation Map DF-67, scale 1:24,000.

Dickinson W. R., (1974). Plate tectonics and sedimentation: Society of Economic Paleontologists and Mineralogists Special Publication 22, p. 1-27.

Dickinson, W. R., (1976). Plate tectonic evolution of sedimentary basins: American Association of Petroleum Geologists Continuing Education Course Notes Series 1, 62 p.

Ely, G., P. Small, T. Jordan, P. Maechling, and F. Wang. (2016). A V_{S30}-derived near-surface seismic velocity model. In preparation. http://elygeo.net/2016-Vs30GTL-Ely+4.html.

Ely, G. P., Jordan, T. H., Small, P., & Maechling, P. J. (2010). A V_{S30}-derived near-surface seismic velocity model. In *Abstract S51A-1907, Fall Meeting*. San Francisco, CA: AGU.

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

Gautier, L. (2009). Rpy2: Interface to use R from Python. *GitHub repository*, https://github.com/rpy2/rpy2.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B., (2014). *Bayesian Data Analysis*, 3rd edition, CRC Press.

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

Graves, R. W. (2002). The seismic response of the San Bernardino basin region. *AGUFM*, *2002*, S21A-0968.

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

Graves, RW, BT Aagaard, KW Hudnut, LM Star, JP Stewart, and TH Jordan (2008). Broadband simulations for Mw 7.8 southern San Andreas earthquakes: Ground motion sensitivity to rupture speed" Geophysical Research Letters, Vol. 35, L22302, doi:10.1029/2008GL035750

Graves, R. W., & Pitarka, A. (2010). Broadband ground-motion simulation using a hybrid approach. *Bulletin of the Seismological Society of America*, *100*(5A), 2095-2123.

Hauksson, E., (2010). Crustal structure and seismic distribution adjacent to the Pacific and North America plate boundary in southern California, *J. Geophys. Res.*, 105 (B6), 13,875–13,903.

Ingersoll, R. V. and Rumelhart, P. E., (1999). Three-stage evolution of the Los Angeles basin, southern California. *Geology*, 27, 593-596.

Iwahashi, J., & Pike, R. J. (2007). Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology, 86(3-4), 409-440.

Landwehr, N, NM Kuehn, T Scheffer, NA Abrahamson, 2016. A Nonergodic Ground-Motion Model for California with Spatially Varying Coefficients, *Bull. Seism. Soc. Am.*, 106, 2574-2583

Langenheim, V. E., Wright, T. L., Okaya, D. A., Yeats, R. S., Fuis, G. S., Thygesen, K., & Thybo, H., (2011). Structure of the San Fernando Valley region, California: Implications for seismic hazard and tectonic history. *Geosphere*, 7, 528-572.

Lee, E.J., Chen, P., Jordan, T.H., Maechling, P.J., Denolle, M., and Beroza, G.C., (2014). Full-3-D tomography for crustal structure in southern California based on the scattering-integral and the adjoint-wavefield methods, *J. Geophys. Res.*, 119, 6421–6451.

Lizarralde, D., Axen, G. J., Brown, H. E., Fletcher, J. M., González-Fernández, A., Harding, A. J., Holbrook, W.S., Kent, G.M., Paramo, P., Sutherland, F., & Umhoefer, P. J. (2007). Variation in styles of rifting in the Gulf of California. *Nature*, *448*(7152), 466-469.

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

Kingston, D. R., Dishroon, C. P., & Williams, P. A., (1983). Global basin classification system. *AAPG Bulletin*, 67, 2175-2193.

Kishida, T., Di Giacinto, D., & Iaccarino, G. (2017). Comparison of manual and automated ground motion processing for small-to-moderate-magnitude earthquakes in Japan. *Earthquake Spectra*, *33*(3), 875-894.

Kohler, M. D., Magistrale, H., and Clayton, R.W., (2003). Mantle heterogeneities and the SCEC reference three-dimensional seismic velocity model version 3, *Bull. Seismol. Soc. Am.*, 93, 757–774.

Kuehn, NM, Abrahamson, N.A., and Walling, M.A., (2019). Incorporating nonergodic path effects into the NGA-West2 ground-motion prediction equations, *Bull. Seismol. Soc. Am.*, 109, 575-585.

Magistrale, H., Day, S., Clayton, R., Graves, R.W., (2000). The SCEC southern California reference three-dimensional seismic velocity model version 2, *Bull. Seism. Soc. Am.*, 90, S65–S76.

Nweke, C. C., Wang, P., Brandenberg, S. J., & Stewart, J. P. (2018). Reconsidering basin effects in ergodic site response models.

Parker, G. A., Harmon, J. A., Stewart, J. P., Hashash, Y. M., Kottke, A. R., Rathje, E. M., Silva, W. J. & Campbell, K. W. (2017). Proxy-Based VS30 Estimation in Central and Eastern North AmericaProxy-Based VS30 Estimation in Central and Eastern North America. *Bulletin of the Seismological Society of America*, *107*(1), 117-131.

Parker, G.A., Stewart, J.P., Hashash, Y.M.A., Rathje, E.M., Campbell, K.W., Silva, W.J., (2019). Empirical linear seismic site amplification in Central and Eastern North America, *Earthquake Spectra*, 35, 849-881.

Petersen, M. D., Shumway, A. M., Powers, P. M., Mueller, C. S., Moschetti, M. P., Frankel, A. D., Rezaeian, S., McNamara, D. E., Luco, N., Boyd, O. S., & Rukstales, K. S. (2020). The 2018 update of the US National Seismic Hazard Model: Overview of model and implications. *Earthquake spectra*, *36*(1), 5-41.

Pischiutta, M., Cara, F., Di Giulio, G., Vassallo, M., & Cultrera, G. (2017). Ground motion amplification at rock sites: the competing role of topography and fractured rocks in the San Giovanni fault, central Italy. *AGUFM*, *2017*, S22A-08.

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

Rathje, E.M., Dawson, C., Padgett, J.E., Pinelli, J.-P., Stanzione, D., Adair, A., Arduino, P., Brandenberg, S.J., Cockeril, T., Esteva, M., Haan, F.L. Jr., Hanlon, M., Kareem, A., Lowes, L., Mock, S., and Mosqueda, G., (2017). DesignSafe: A new cyberinfrastructure for natural hazards engineering. *Natural Hazards Review*. 18.

Seyhan, E., & Stewart, J. P. (2014). Semi-empirical nonlinear site amplification from NGA-West2 data and simulations. *Earthquake Spectra*, *30*(3), 1241-1256.

Shaw, J. H., Plesch, A., Tape, C., Suess, M.P., Jordan, T.H., Ely, G., Hauksson, E., Tromp, J., Tanimoto, T., Graves, R.W. et al., (2015). Unified structural representation of the southern California crust and uppermantle, *Earth Planet. Sci. Lett.*, 415, 1, doi: 10.1016/j.epsl.2015.01.016.

Siddiqqi, J., & Atkinson, G. M. (2002). Ground-motion amplification at rock sites across Canada as determined from the horizontal-to-vertical component ratio. Bulletin of the Seismological Society of America, 92(2), 877-884.

Small, P., Gill, D., Maechling, P. J., Taborda, R., Callaghan, S., Jordan, T. H., Olsen, K.B., Ely, G.P., and Goulet, C., (2017). The SCEC unified community velocity model software framework. *Seismological Research Letters*, 88(6), 1539-1552.

Snedecor, G. W., & Cochran, W. G. (1989). Statistical Methods, eight edition. *Iowa state University press, Ames, Iowa*.

Stafford, J.S., Rodriguez-Marek, A., Edwards, B., Kruiver, P.P., and Bommer, J.J., (2017). Scenario dependence of linear site-effect factors for short-period response spectral ordinates. *Bull. Seismol. Soc. Am.*, 107, 2859–2872.

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

Süss, M. P., and Shaw, J.H. (2003). P wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, Calif. *J. Geophys. Res.*, 108, 2170, doi: 10.1029/2001JB001628.

Thompson, E.M., (2018). An updated Vs30 Map for California with geologic and topographic constraints: U.S. Geological Survery data release. https://doi.org/10.5066/F7JQ108S.

Thompson, E.M., Wald, D.J., and Worden, C.B., (2014). A V_{S30} map for California with geologic and topographic constraints, *Bull. Seismol. Soc. Am.*, 104, 2313-232.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J.,

Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., van der Plas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3), 261-272.

Wang, P., Stewart, J.P., Bozorgnia, Y., Boore, D.M., and Kishida, T., (2017). "R" Package for computation of earthquake ground-motion response spectra, *PEER Report No. 2017/09*, Pacific Earthquake Engineering Research Center, UC Berkeley, CA.

Wang, P., & Stewart, J. P. (2019). *Data-derived site response and its predictability using ergodic and site-specific methods*. eScholarship, University of California.

Yeats, R.S., (2004). Tectonics of the San Gabriel Basin and surroundings, southern California. *GSA Bulletin*, 116(9-10), 1158-1182.

Yong, A.K., (2016). Comparison of measured and proxy-based V_{S30} values in California, *Earthquake Spectra*, 32, 171-192.

Yong, A.K., Hough, S.E., Iwahashi, J., and Braverman, A., (2012). Terrain-based site conditions map of California with implications for the contiguous United States. *Bull. Seismol. Soc. Am.* 102, 114-128.