Title
Reconsidering basin effects in ergodic site response models

Authors
Nweke, Chukwuebuka C
Wang, Pengfei
Brandenberg, Scott J
et al.

Publication Date
2018-10-01

RECONSIDERING BASIN EFFECTS IN ERGODIC SITE RESPONSE MODELS

Chukwuebuka C. Nweke, Pengfei Wang, Scott J. Brandenberg, and Jonathan P. Stewart[1]

Department of Civil & Environmental Engineering

University of California, Los Angeles

Abstract

We investigate benefits of regionalizing basin response in ergodic ground motion models. Using southern California data, we find average responses between basin structures, even when the primary site variables used in ground motion models (V_{S30} and depth parameters) are controlled for. For example, the average site response in relatively modestly sized sedimentary structures (such as Simi Valley) are under-predicted at short periods by current models, whereas under-prediction occurs at long periods for larger sedimentary structures. Moreover, site-to-site within-event standard deviations vary appreciably between large basins, basin edges, smaller valleys, and non-basin (mountainous) locations. Such variations can appreciably impact aleatory variability.

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), these effects are averaged over many sites globally with conditioning on time-averaged 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 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 V_{S30}-based site amplification models form the primary basis for ergodic site effect modeling in the development of the USGS seismic hazards mapping program in the western US (Petersen et al. 2015). Many other site-specific applications, as well as ongoing work related to the 2018 version of the USGS maps, consider basin effect modeling using the NGA-West2 depth terms. This work has caused a number of important questions to be raised. There are two principle considerations related to the prediction of mean amplification in basins:

[1] Corresponding author and presenter. Email: jstewart@seas.ucla.edu

- 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.
- Amplification function: Basin amplification models were derived using data from basins in Japan and California. Variability in basin-related amplification between regions, and between basins within a given region, is likely present but is not captured with current procedures.

We are in the midst of a long-term research effort in which these and other issues pertaining to mean site amplification are being addressed. As part of this work, we are also investigating 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}).

$$\sigma_{ln} = \sqrt{\tau_{ln}^2 + \phi_{ln}^2} \tag{1}$$

Within-event variability has contributions from region-to-region and site-to-site variations in path and site effects. 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):

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

where ϕ_{lnY} is the remaining variability when path- and site-specific models are used, which appears to be principally associated with event-to-event variations in site response at a particular site (Stewart et al., 2017).

Given the limited information on basins that is considered in current GMMs (depth only), we investigate here the potential for regional variations in site response associated with particular basin structures. Likewise, given the limited information on site condition (V_{S30}) that is considered in models of aleatory variability, we investigate variations of site-to-site variability between site categories selected to reflect different morphological conditions. Our study region is southern California, which was selected due to large volumes of ground motion data and the availability of models describing the velocity structure in sedimentary basins.

Following this introduction, we describe the database compiled for the present study. We then present a site categorization scheme intended to distinguish sites having different levels and types of basin 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. The data

analysis examines residuals of the data set relative to NGA-West2 models. These residuals analyses investigate model bias with respect to site categories and specific geologic structures such as the Los Angeles basin. The dispersion of residuals is used to investigate changes in site-to-site variability between categories and between specific basins. The results are interpreted to provide insights into how basin models can be improved for ground motion modeling in southern California.

Database

We begin 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 (191 events, 898 stations, 8245 recordings) over the time period 1938 to 2010. The site portion of the database (Seyhan et al. 2014) was developed to provide the principle 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 mostly 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. As part of this project and other complimentary projects, we converted the spreadsheet files that comprised the original NGA-West2 flatfile (pertaining to sources, sites, and ground motions) into a formal 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 on DesignSafe (Rathje et al. 2017).

We have identified earthquakes and recordings since 2011 in California, which significantly extend the NGA-West2 database. In this extension of the database, we only consider **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 1 shows the locations of events sorted by magnitude, most of which occur in five main regions: Bay Area, Eastern Sierra and Nevada, central California, southern California, and Imperial Valley and northern Mexico. These five zones incorporate most of the urban areas in the state, and contain a large fraction of the ground motion stations. There are over 33,000 three-component recordings from 179 events. As explained further in the next section, we focus here on the southern California region. The data from events within the Southern California region in Figure 1 is derived from 22 earthquakes that have produced about 9,300 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 4260 usable three-component records. Figure 2 shows the locations of these events and of the 362 recording 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 Ancheta et al. (2014). This processing provides a lowest usable frequency for each ground motion component. Horizontal ground motion components are combined to median-component (RotD50) as defined by Boore (2010) using the routines given in Wang et al. (2017). We take the lowest useable frequency for RotD50 as the higher of the two as-recorded values. Figure 3 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 2.5 sec.

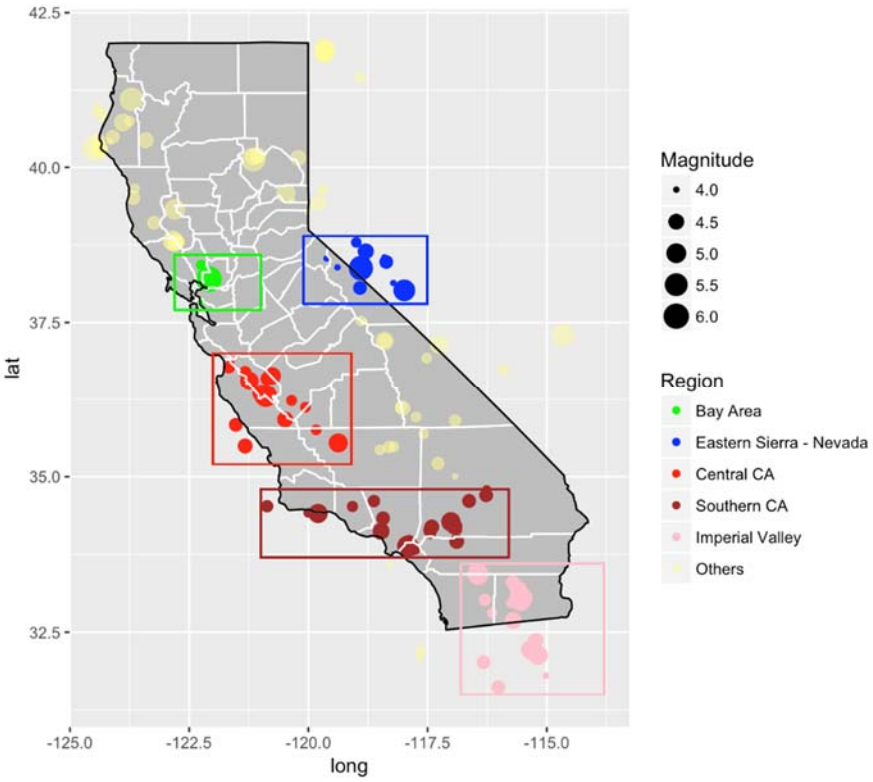

Figure 1. Locations of earthquakes in California and northern Mexico with **M** > 4 since 2011 that have for which ground motion data has been compiled for addition to the NGA-West2 database

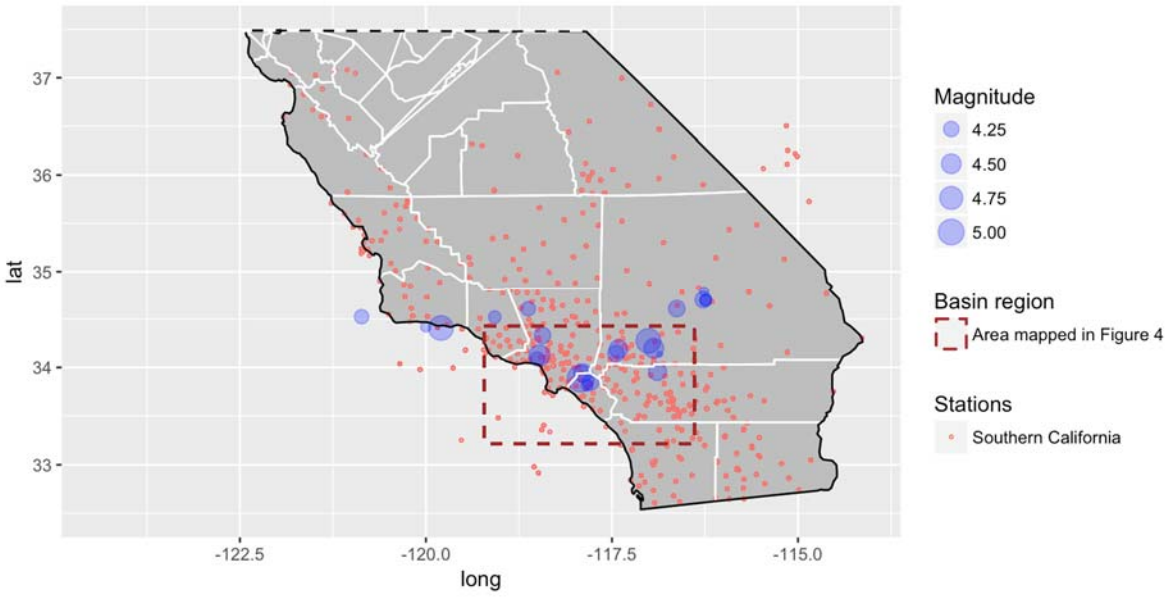

Figure 2. Map of southern California region showing locations of considered earthquakes with **M** > 4 since 2011 and locations of stations that recorded the event

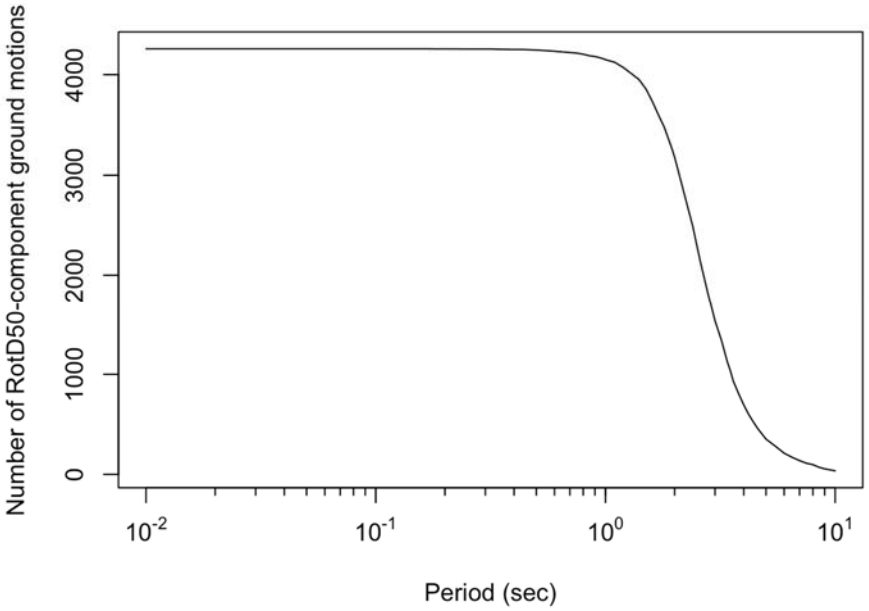

Figure 3. Number of usable RotD50-component ground motions as a function of oscillator period for the data added for the southern California region.

Considering both the NGA-West2 data and new data, there are 777 recording sites within the rectangular area shown in Figure 2, which is shown in greater detail in Figure 4. Of those, 736 are sites that were included in the NGA-West2 site database. Hence, there are 41 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 four sites (data obtained from USGS V_{S30} Compilation Database[2]). For sites without V_{S30} 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 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.

[2] https://earthquake.usgs.gov/data/vs30/

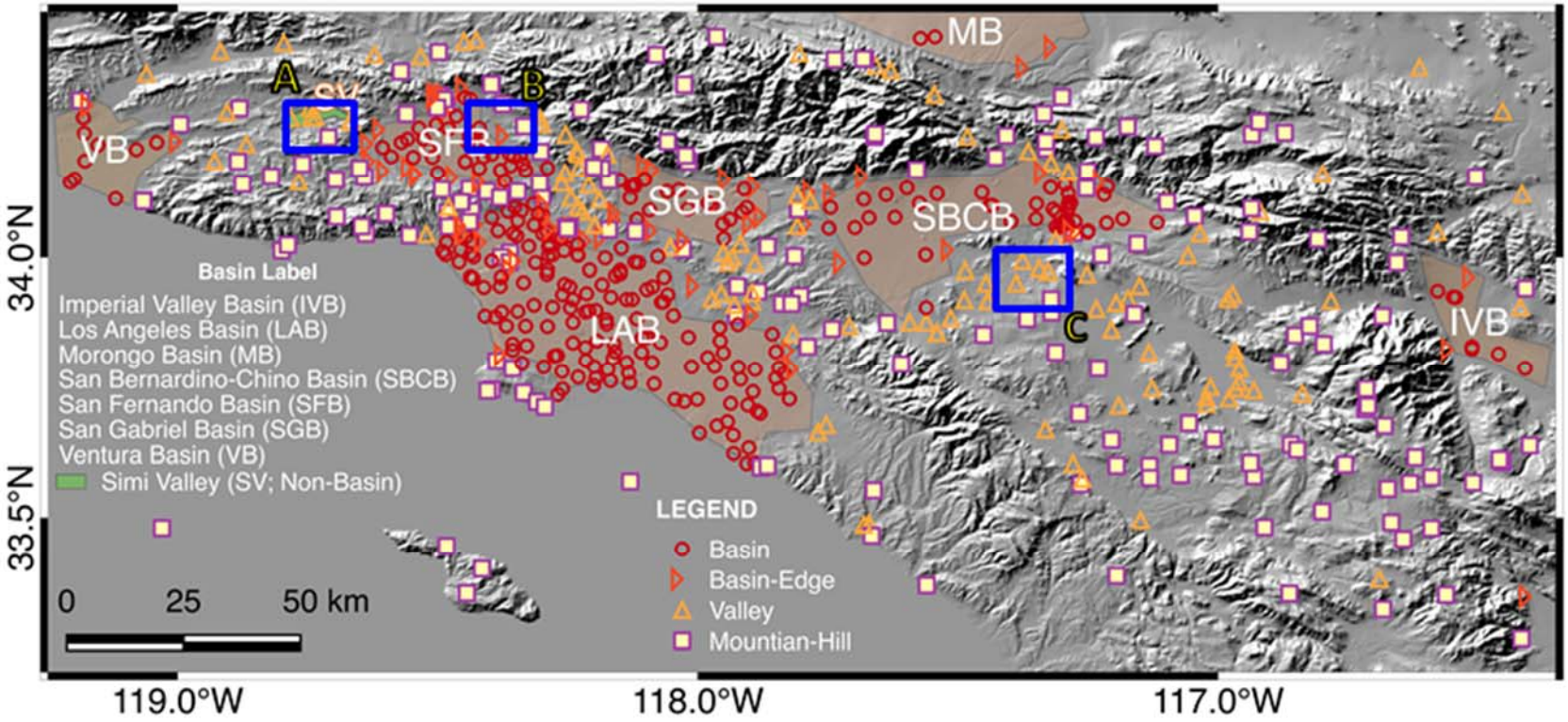

Figure 4. Detail map of southern California showing ground motion stations and sedimentary basins and related features considered in this paper. Ground motion sites are plotted according to a morphology-based site categorization scheme proposed in this paper. Boxes A, B, and C are detailed in subsequent figures in this paper.

Table 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

Source parameters were compiled for each of the 22 new events. The range of moment magnitudes is 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. Finite fault models are not available for any of the considered events, to our knowledge. 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 (2017).

Figure 5 summarizes attributes of the compiled data. Figure 5a shows the newly added data in magnitude distance-space in comparison to the NGA-West2 data. Figure 5b shows the distribution of site data in V_{S30}-$z_{1.0}$ space, with data from particular basins (as defined in the next section) delineated. The plots in Figure 5 show that that data set has been significantly expanded. This 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 110 stations with 10 or more recordings in the study region; whereas the current data set now has 174 such stations.

Figure 5b shows the model for predicting $z_{1.0}$ given V_{S30} proposed by Chiou and Youngs (2014) along with the southern California data. The Chiou and Youngs model is meant to apply for all of California, including San Francisco Bay Area sites. Comparing the binned means of the data 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 model. As a result, the NGA-West2 basin terms may not be optimally centered. The functional form for the Chiou and Youngs (2014) model is,

$$\ln(\overline{z_{1.0}}) = \mathrm{v}_0 ln\left(\frac{V_{S30}^4+v_1^4}{1360^4+v_1^4}\right) - \ln(1000) \quad (3)$$

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 2.

Table 2. Basin depth predictive model coefficients (Eqs. 3-4)

Parameters (CY14)	Value	Parameters (this study)	Value
v_0	-1.7875	c_0	1.02
v_1	570.94	c_1	-0.5
		v_μ (m/s)	266.4
		v_σ	0.20

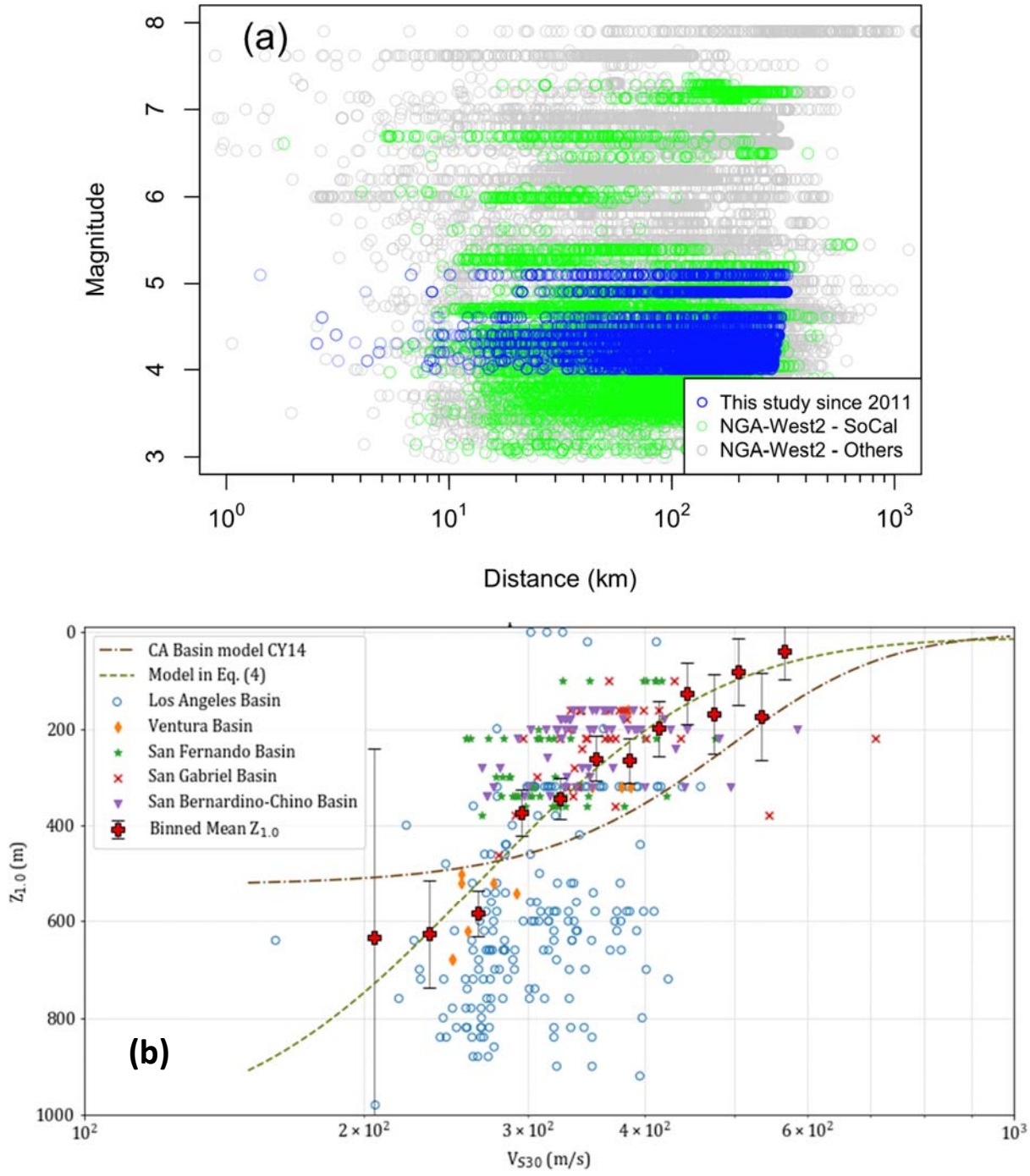

Figure 5. (a) Added data points (from southern California region as shown in Figure 2) in magnitude-distance space; (b) distribution of new and previous data in V_{S30}-$z_{1.0}$ space, including model for the relationship between these parameters by Chiou and Youngs (2014) for California.

We sought to develop an improved fit to the data. Our objective was to fit the trend shown by the binned mean while also enforcing physical bounds at the limits, whereby the depth scaling would flatten with respect to V_{S30}. We suggest the following function to provide the desired shape:

$$\overline{z_1} = c_1 \left[1 + erf\left(\frac{log(V_{S30}) - log(v_\mu)}{v_\sigma \sqrt{2}}\right)\right] + c_0 \quad (4)$$

where v_μ defines the center of the scaling relationship where the slope is steepest and v_σ defines the width of the ramp. Eq. (4) returns the mean $z_{1.0}$ in units of km. Values for all coefficients are given in Table 2. 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 model to the southern California data is given in Figure 5b. Even with this improved fit, it is apparent that the model fits some basins better than others. The fit is good for the relatively deep near-coast basins (Los Angeles, Ventura), whereas depths are generally over-predicted for inland basins (San Fernando, San Gabriel, San Bernardino-Chino).

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 (i.e., different users would make identical assignments) basin classification system useful for ground motion amplification purposes. Such classifications have not been provided in prior work, to our knowledge.

Southern California Study Region

The present research on basin response effects is in its early stages. While we ultimately anticipate considering several regions with pronounced basin features and ample earthquake recordings, we have initially focused on the southern California region shown in Figure 4. The approximate limits of the region are (from west to east) Ventura to Landers and (from south to north) Borrego Springs 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 1).

We have identified five major basin structures within the study region, the approximate outlines of which are shown in Figure 4. These are the Los Angeles basin, the Ventura basin, the San Fernando basin, the San Gabriel basin, and the San Bernardino-Chino basin.

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). Intermitted 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 eastern-most basins within the study area (San Gabriel, San Bernardino-Chino) are continental pull-apart/graben basins that formed as a result of regional faulting. The San Gabriel and San Bernardino-Chino basins are adjacent alluviated lowlands with sediments deposited via erosion from the San Gabriel and San Bernardino Mountains. Both basins are separated by the Glendora Volcanics in the San Jose Hills as well as the Cucamonga Fault Zone (Anderson et al. 2004; Yeats 2004).

The intermitted subsidence and uplift experienced by the western-most basins likely led to the large depths that exists in those basins compared to the shallower depths observed in the eastern-most basins which formed from transform-graben induced valleys adjacent to uplifted blocks. As a result of these differences in geologic history, differences in site response might reasonably be expected. This hypothesis is tested in the present study.

Basin Categorization

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 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 represent basin size in the site categories.

The proposed categorization scheme is given in Table 3. Two of the categories are obvious – representing 'within basin' and 'outside basin' conditions (Categories #3 and #0, respectively). The valley category (#1) is intended to introduce lateral dimension to the categorization. We

considered Simi Valley (identified in Figure 4 as Box A, detail in Figure 6) to be a good example of a sedimentary depression of modest dimension that should be differentiated from those of large dimension, like the Los Angeles basin. Driven by this admittedly arbitrary example, 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 4) were manually classified according to the categories in Table 3. The manual classification was performed using terrain maps from Google Maps™, where a visual assessment of slope and terrain roughness/texture were used along with information on 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 7a (detail of Box B from Figure 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 a basin is located within a region that is relatively flat. The basin edge site in Figure 7a 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 7b (detail of Box C from Figure 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.

Table 3. Proposed basin classification criteria for Southern California

Category	Description	Criteria	Cat. #	Number of Sites
Basin	Site location in basin interior	Basin width in short direction > 3 km	3	281
Basin Edge	Sites along basin margin	Within 300 m of basin edge[1]	2	71
Valley	Site location in 'small' sedimentary structure	Valley width in short direction < 3 km	1	125
Mountain-Hill	Sites without significant sediments, generally having topographic relief	Generally identified on basis of appreciable gradients and/or irregular morphology	0	190

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

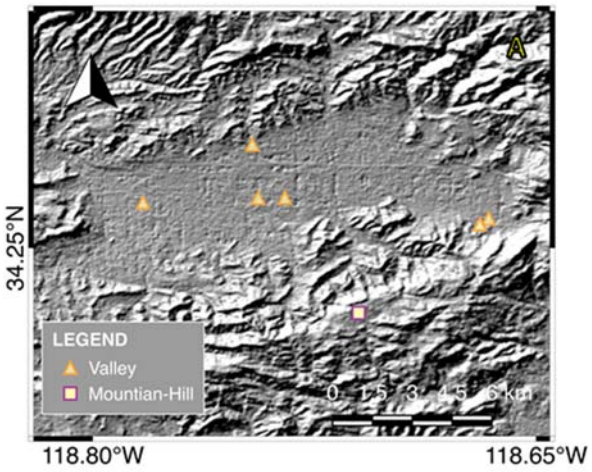

Figure 6. Simi Valley region (Box A in Figure 4)

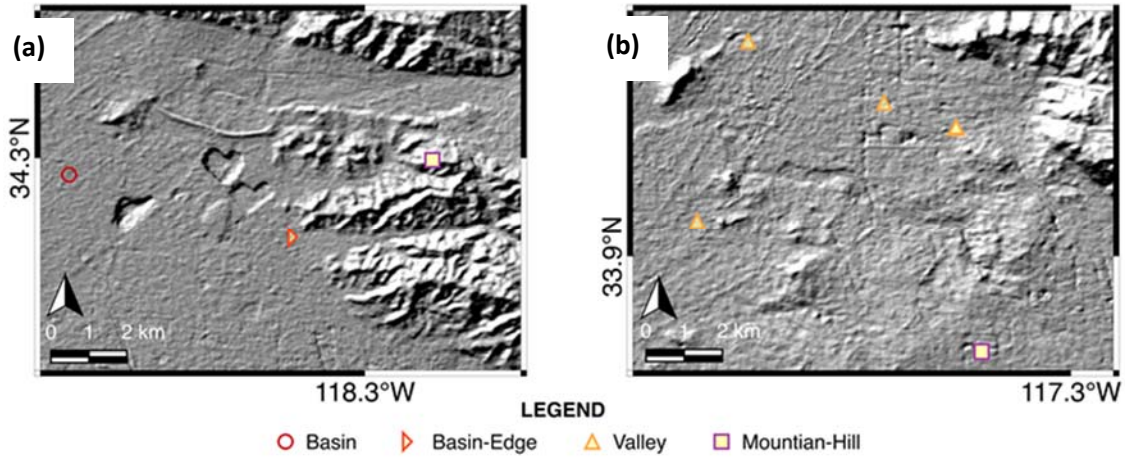

Figure 7. (a) Example location in north-eastern San Fernando Basin with relatively unambiguous site categorizations (Box B in Figure 4); (b) example location in Riverside for which the site classification was more challenging (Box C in Figure 4).

Ground Motion Analysis

Ground motion analyses were undertaken to investigate whether the site categories in Table 3 are useful, in combination with current basin effect models, for differentiating site effects in basins and other areas. This was investigated using a subset of database described previously (i.e., only the NGA-West2 data are currently considered, the newly added data is being considered in ongoing work). Our analysis uses residuals of NGA-West2 models. We focus on one such model in this paper (Boore et al. 2014), but other models are being considered in the project.

Data selection

We use a subset of the NGA-West2 database applicable to events in the Southern California region shown in Figure 2. The data added for the time period since 2011, as shown in Figure 5a, is contained without our data set but has not yet been analyzed. At this time, events in the Imperial Valley region and northern Mexico have not been considered, although that is being investigated in ongoing work.

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 a range of oscillator periods shorter than $1/(1.25f_{hp})$, where f_{hp} is the high-pass frequency selected during component-specific data processing. This frequency is provided in the NGA-West2 flatfile, and was developed in the present work for the added recordings.

As shown in Figure 5a, the data set spans a magnitude range of about 3 to 7 and a closest distance range of about 1 to 400 km. Figure 5b shows that the range of V_{S30} is about 200-600 m/s and the range of $z_{1.0}$ is about 0 to 1000 m.

Residuals analysis

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

$$R_{ij} = \ln(Z_{ij}) - \mu_{ln}(\mathbf{M_i}, F_i, R_{ij}, V_{S30}, z_{1.0}) \quad (5)$$

where index i refers to an earthquake and index j refers to a recording. The quantity Z_{ij} is a ground motion observation expressed as an intensity measure. The term μ_{ln} is the mean prediction in natural log units of a ground motion model, which uses the arguments in the parenthesis in Eq. (5). We use the Boore et al. (2014) model, which has the arguments listed in Eq. (5), where *F* is a style of faulting parameter (reverse, strike-slip, etc.), *R* is the Joyner-Boore distance, and other parameters are as defined previously.

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:

$$R_{ij} = \eta_{E,i} + \delta W_{ij} \quad (6)$$

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

$$\delta W_{ij} = \eta_S + \varepsilon_{ij} \quad (7)$$

where ε_{ij} is the remaining residual when the event and site terms have been removed. Recalling the standard deviation terms from the *Introduction*, the standard deviation of η_E terms is τ_{ln}, the standard deviation of δW terms is ϕ_{ln}, the standard deviation of η_S terms is ϕ_{S2S}, and the

standard deviation of ε_{ij} is $\sqrt{\phi_{P2P}^2 + \phi_{lnY}^2}$ (the P2P term appears because we are not accounting for non-ergodic path effects).

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

$$R_{ij} = c_k + \eta_{E,i} + \eta_S + \varepsilon_{ij} \tag{8}$$

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.

Our analysis of site effects from the data is based principally on the interpretation of site terms η_S. By using these results, we have removed from the residuals systematic effects associated with the earthquake events (i.e., η_E), which are expected to be unrelated to site response. Another effect that needs to be checked before interpreting site effects is path-scaling. This can be done by checking for trends of δW_{ij} with distance, which is shown in Figure 8 for the intensity measures of PGA and $S_a(2.0)$ (pseudo-spectral acceleration at a period of 2.0 sec). The lack of trend suggests that the path scaling in the model is unbiased for the data set, and hence the model is suitable for analysis of site effects.

By removing event-related effects, and checking for path effects, we improve the likelihood that trends observed in the data are principally related to site response. These are important checks to perform when analyzing site effects using residuals, which is also known as a *non-reference site* approach (Field and Jacob, 1995).

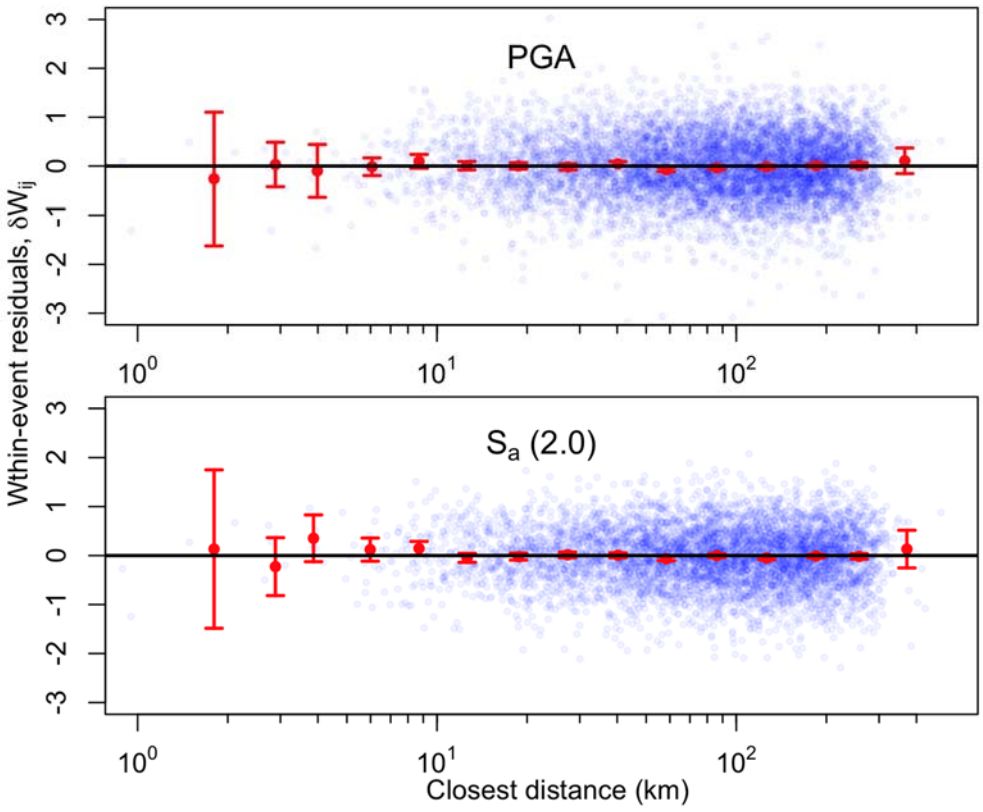

Figure 8. Within-event residuals for southern California data plotted as a function of distance. Binned means and their 95% confidence intervals are shown. There is no trends in the data with distance, indicating that the path scaling in the ground motion model is unbiased for the region.

Mean site response

For the intensity measures of PGA and $S_a(2.0)$, Figure 9 shows the trend of site terms η_S with differential depth, defined as:

$$\delta z_1 = z_{1.0} - \overline{z_{1.0}} \tag{9}$$

For the present analysis, we take $\overline{z_{1.0}}$ using the CY14 model (Eq. 3). We will investigate the impact of updates to the centering model in future work. Figure 9 shows results for all data combined (i.e., all site categories). There is no appreciable trend in the site terms, which indicates that the site terms in the ground motion model (V_{S30}-scaling term and depth term) are capturing average regional trends.

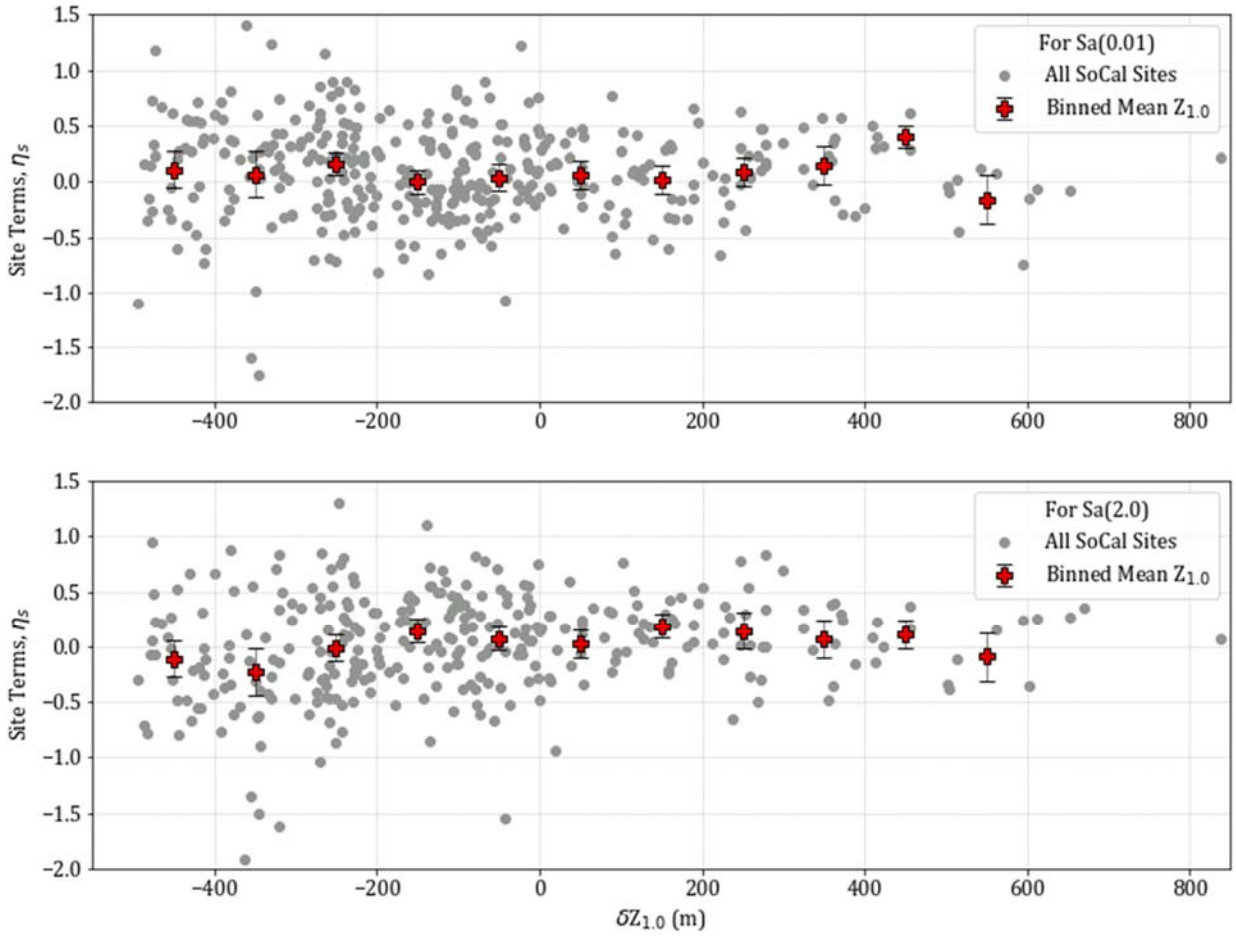

Figure 9. Variation of site terms with differential depth δz_1 for all considered sites in southern California region for intensity measures of PGA and $S_a(2.0)$.

Figure 10 shows trends of $S_a(2.0)$ data with differential depth in the site categories in Table 3. The differential depth range -300 to -100 m (for basin, valley, and mountain-hill sites) has an upward trend. The basin term in the ground motion model has a ramp in this range, so the residuals suggest this ramp could be steeper. Interestingly, the basin edge data indicate a weakly negative trend over this same depth range, indicating that the ramp in the models should be flatter for this category.

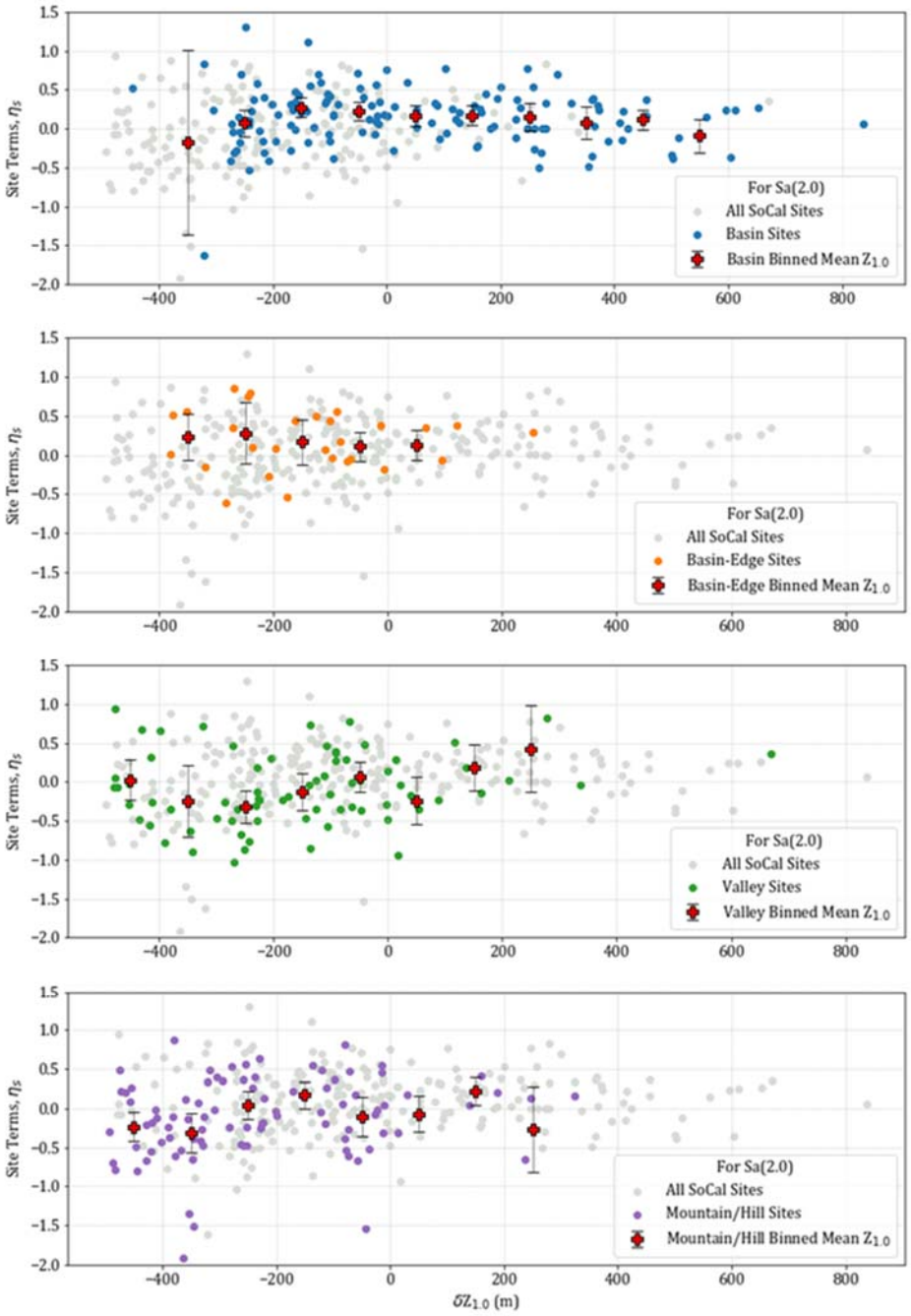

Figure 10. Variation of site terms with differential depth δz_1 for sites within the four site categories for intensity measure $S_a(2.0)$.

Mean biases are plotted as a function of period in Figure 11. The means are generally small (less than about 0.1), the main exceptions being basin sites at long periods (> 1.0 sec), valley sites at short periods (< 0.5 sec), and basin edge sites over the full period range, each of which have positive biases (ground motions under-predicted). Mountain-hills sites have negative bias (ground motions are over-predicted).

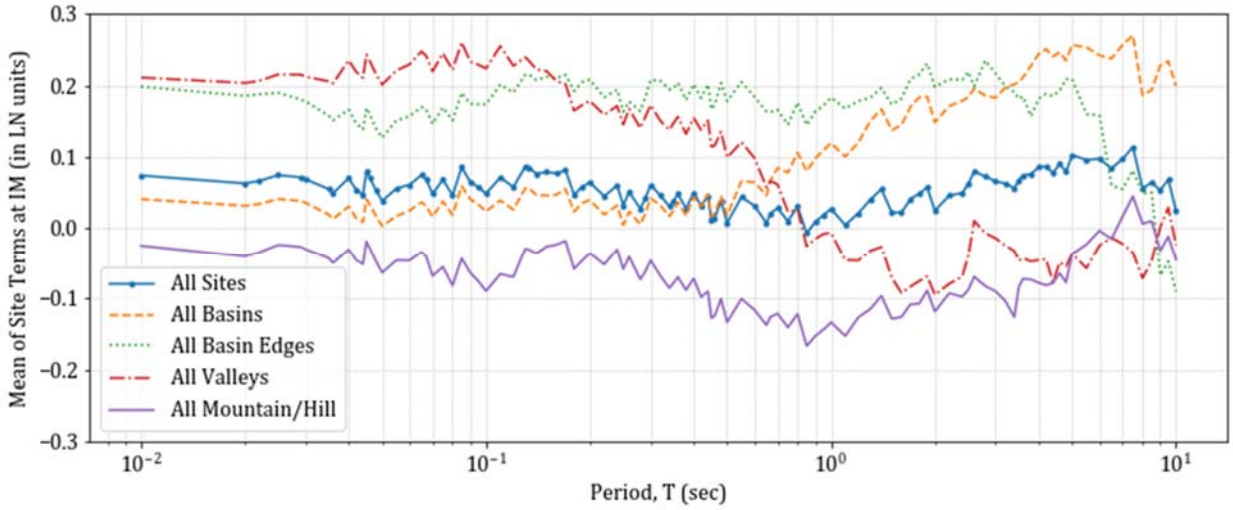

Figure 11. Period-dependence of mean of site terms the four site categories.

Figure 12 shows trends for basins (i.e., sites in the basin and basin edge categories) for four basin structures with significant information: Los Angeles, San Fernando, San Bernardino-Chino, and San Gabriel. The trends with differential depth are generally flat for the Los Angeles basin, which is not surprising because this basin dominates the data set. The San Fernando data has a weak downward trend with differential depth for negative δz_1, suggesting that the reduction of ground motion for $\delta z_1 < 0$ may produce under-prediction bias for this basin (although the data is sparse). The results for the San Gabriel and San Bernardino-Chino basins show some evidence of upward trends, suggesting that the differential scaling could be slightly stronger for these structures than the ergodic model.

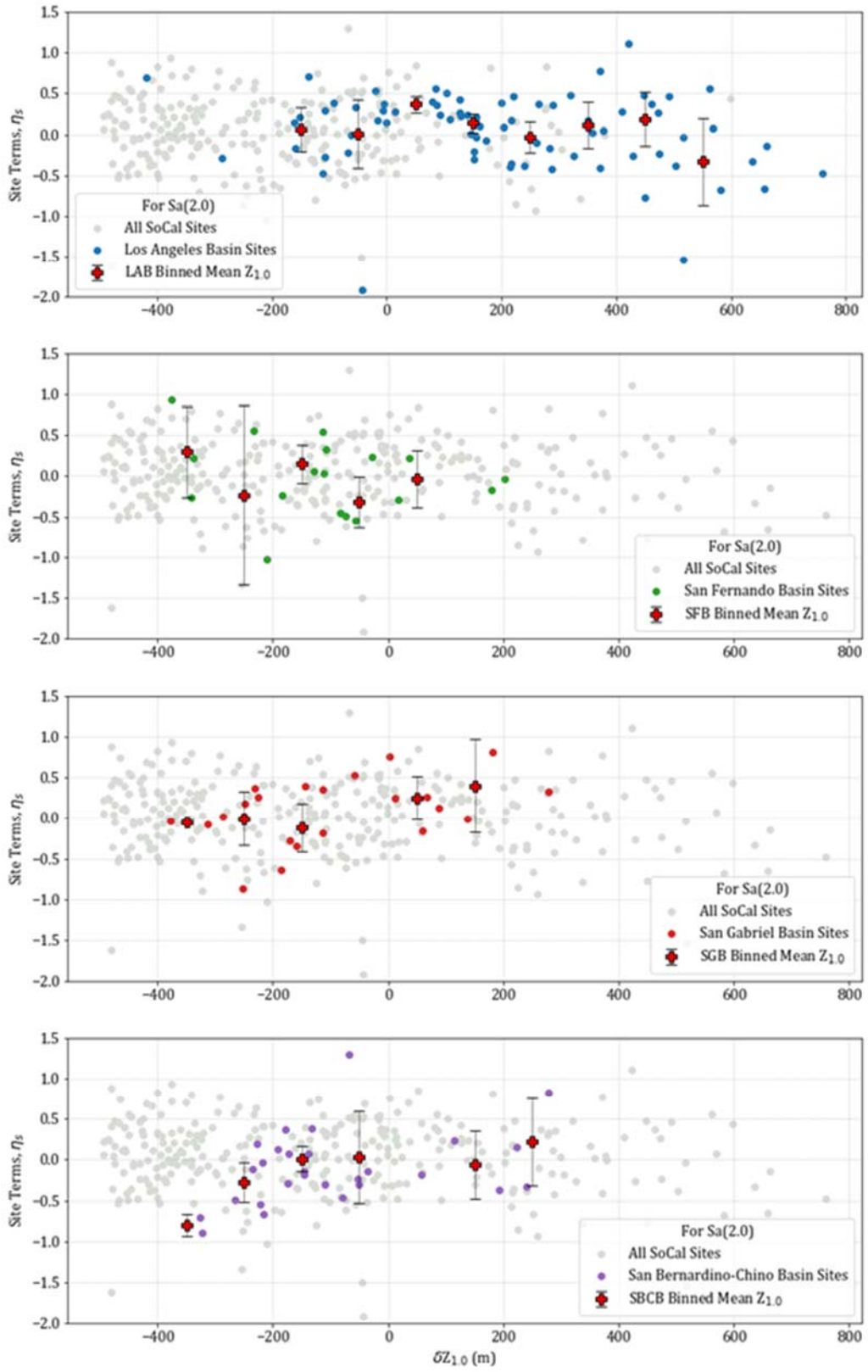

Figure 12. Variation of site terms with differential depth δz_1 for basin and basin edge sites within four basin structures, intensity measure $S_a(2.0)$.

Site-to-site variability

Al Atik (2015) performed residuals analyses similar to those presented here for the full NGA-West2 data set, and based on those analyses, proposed models for site-to-site standard deviation 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 present results for the subset of events with **M** < 5.5 in order to illustrate the effects of site condition on ϕ_{S2S}. Similar trends were observed for the larger magnitude data.

Figure 13 compares ϕ_{S2S} for the full data set with the findings of Al Atik (2015). The two sets of standard deviations are nearly identical, indicating that the Southern California data is consistent with global data regarding site-to-site standard deviations.

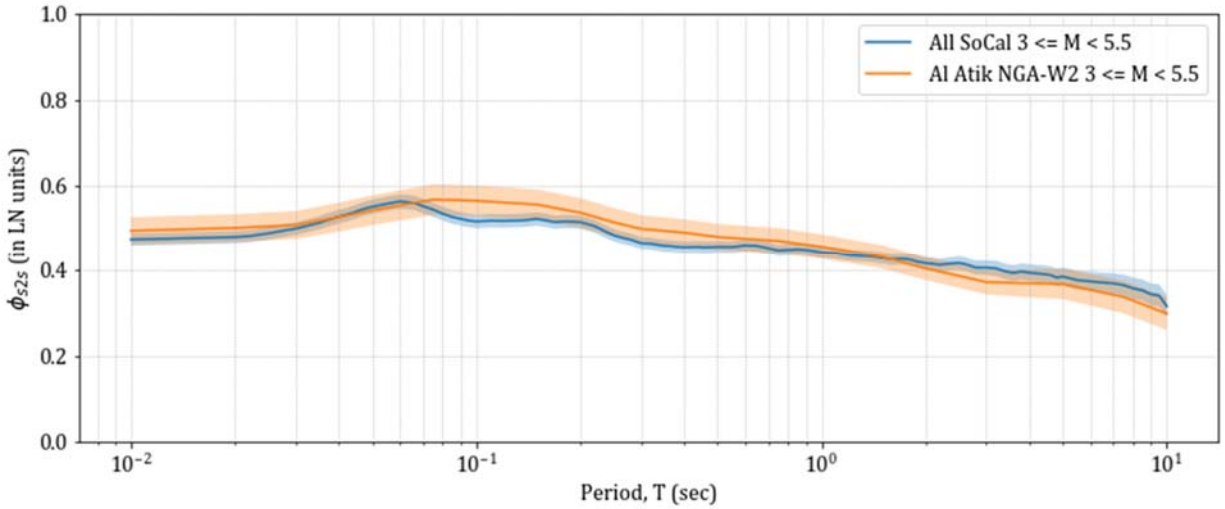

Figure 13. Site-to-site standard deviations, and their 95% confidence intervals, as a function of period for global data (Al Atik 2015) and Southern California data considered in this study. **M** < 5.5 events.

Figure 14 compares ϕ_{S2S} for sites within the proposed site categories in Table 3, with the overall ϕ_{S2S} (across all sites) shown as a baseline for comparison. The variations with site condition are appreciable. Basin sites have lower than baseline ϕ_{S2S} over nearly the full period range. In contrast, basin edge sites have higher variability at short periods and lower at period > 1 sec. Valleys have nearly the opposite trend, with low variability at short periods and high variability at long periods. Mountain-hill sites follow a similar trend to basin edge sites, although with consistently higher variability across all periods.

Figure 15 compares site-to-site standard deviations for all basin sites (as shown in Figure 14) to results for individual basins. The variations in ϕ_{S2S} between basins are small relative to the variability between site categories (Figure 14).

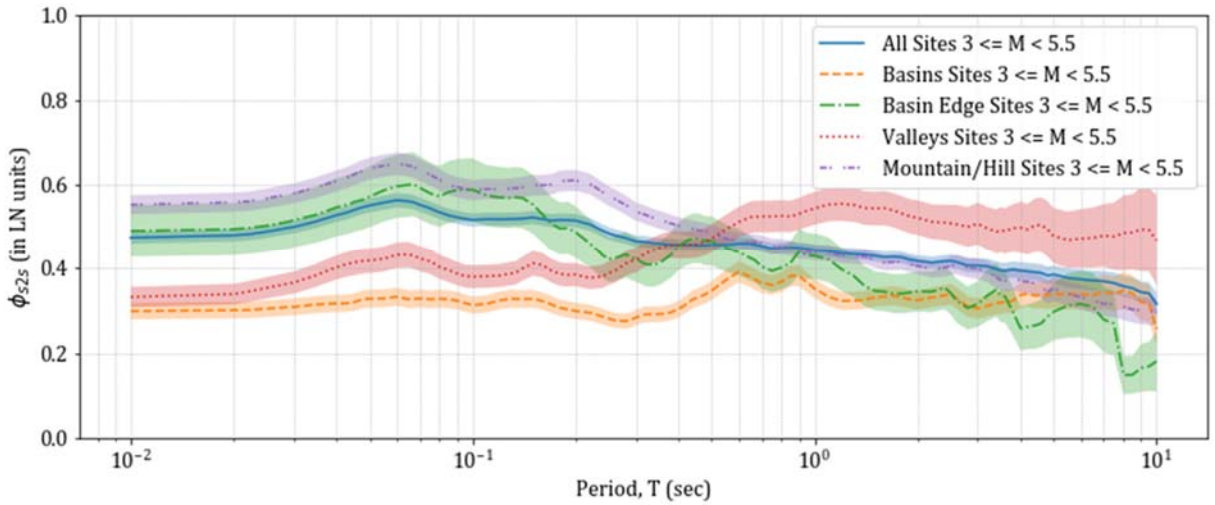

Figure 14. Site-to-site standard deviations, and their 95% confidence intervals, as a function of period for Southern California data sorted by site category. **M** < 5.5 events.

Summary and Interpretation of Results

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, it has not previously been considered for basin depth. In this paper, we present preliminary results of an ongoing study investigating the performance of these global models with respect to southern California data, with particular attention placed on various sedimentary basin structures in the region.

We propose 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. Introducing this scheme to ground motion modeling reveals some features that to some extent might be expected. For example, relatively small sedimentary structures have stronger ground motions at short periods than provided by the ergodic models. This is expected because predominant periods for such sites would also be expected at short periods. Likewise, the ergodic models under-predict long-period ground motions in larger sedimentary structures (basins), which would be expected to have, on average, long predominant periods. When specific basin structures are investigated, we see some differences between deep coastal basins (e.g., Los Angeles) and generally shallower, graben-type interior basins (San Bernardino-Chino, San Gabriel, San Fernando). This suggests that the geologic histories of basins may have a quantifiable impact on site effects beyond their effect on V_{S30} and sediment depth. This hypothesis will be explored further in future work.

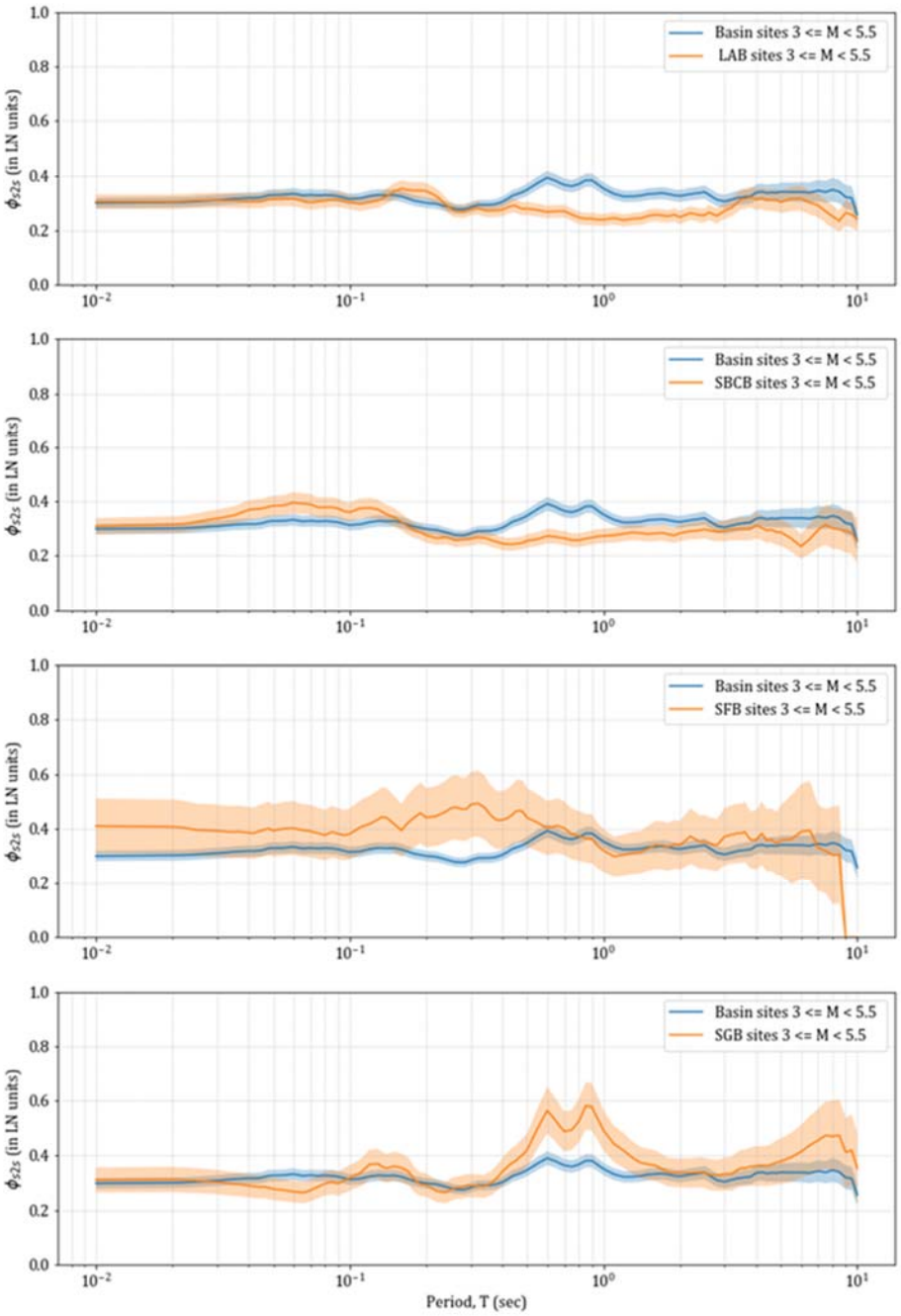

Figure 15. Variation of site-to-site standard deviations, and their 95% confidence intervals, as a function of period for four basin structures. LAB = Los Angeles basin, SBCB = San Bernardino-Chino basin, SFB = San Fernando basin, SGB = San Gabriel basin.

The morphology-based site categories have an appreciable impact on site-to-site standard deviation, which is a significant contributor to overall within-event dispersion. These effects are arguably the most impactful finding of this study. Dispersion is much lower for basin sites than for other site categories, whereas mountain-hill sites have relatively high site-to-site dispersion.

Individual basins have relatively minor variations in site-to-site variability, suggesting that a single model could be used to represent basins collectively in southern California.

Acknowledgments

Funding for this study is provided by California Strong Motion Instrumentation Program, California Geological Survey, Agreement Number 1016-985. This support is gratefully acknowledged. Partial support was also provided under award number G17AP00018 from the U.S. Geological Survey and from NSF-AGEP California Alliance Fellowship (awards 1306595, 1306683, 1306747, 1306760). The views and conclusions presented in this paper are those of the authors, and no endorsement is implied on the part of the State of California or the U.S. Federal government. We thank Yousef Bozorgnia of UCLA for providing access to data processing codes. We thank Phil Maechling, Edric Pauk, and Mei-Hui Su of the Southern California Earthquake Center for providing access to community velocity models.

References

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.

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.

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.

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

Contreras, V., 2017. Development of a Chilean ground motion database for the NGA-Subduction project, *MS Dissertation*, Univ. California, Los Angeles.

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.

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., 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.

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.

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.

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.

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.

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.

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*, https://doi.org/10.1193/083117EQS170M.

Petersen, M.D., Moschetti, M.P., Powers, P.M., Mueller, C.S., Haller, K.M., Frankel, A.D. Zeng, Y. Rezaeian, S., Harmsen, S.C., Boyd, O.S., Field, N., Chen, R., Rukstales, K.S., Luco, N., Wheeler,

R.L., Williams, R.A., and Olsen, A.H., 2015. The 2014 United States national seismic hazard model. *Earthquake Spectra*, **31**, S1-S30.

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**.

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.

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.

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.

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

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**, 2313-2321

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.

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.