Shear Wave Velocity Database and Its Application for Analysis of Non-Ergodic Site Amplification Effects

Pengfei Wang
Paolo Zimmaro, Ph.D.
Department of Civil and Environmental Engineering
University of California, Los Angeles

Sean K. Ahdi, Ph.D.
Exponent
Los Angeles, CA

Dong Youp Kwak, Ph.D.
Hanyang University
Ansan, South Korea

Principal Investigator: Jonathan P. Stewart, Ph.D., PE
Department of Civil and Environmental Engineering
University of California, Los Angeles

A report on research conducted with support from the US Geological Survey, California Strong Motion Instrumentation Program, and Consortium of Organizations for Strong Motion Observation Systems

Civil & Environmental Engineering Department
University of California, Los Angeles
December 2019

ABSTRACT

This project was comprised of two distinct yet mutually beneficial tasks. The first was the establishment of a United States Community V_S Profile Database (PDB). The second involved the use of that database to support ground motion studies that establish observation-based site response at ground motion recording stations and then seek to establish the degree to which it can be estimated using alternate prediction approaches.

The development of the PDB has been, and continues as, a major multi-institutional effort to develop an open-access V_S profile database for sites in the United States. The data described herein was collected from diverse sources that include consulting engineering reports from private industry, university research reports and other documents, federal open-file and similar reports, California state agency documents, and reports provided by electric utilities for selected sites. All data are strictly within the public domain, but much of it was for practical purposes inaccessible to most potential users. The V_S data sources encompass a wide array of geophysical techniques, are presented in many different formats, and are accompanied by widely divergent supplementary data, including P-wave velocities, geotechnical logs and other data, and penetration test data. A relational database schema of sufficient breadth and flexibility was developed to accommodate this diverse data set. The data are digitized and otherwise prepared in the standardized format specified by the database schema. A web interface (www.uclageo.com/VPDB) was developed for data query, visualization, and download. This resource is anticipated to be useful to geotechnical engineers and engineering seismologists for diverse applications in research and industry practice.

We derive non-ergodic site response for California sites using an expanded version of the NGA-West2 database. We then investigate the degree to which different site response analysis methods capture observations. An ergodic site term conditioned on time-averaged shear wave velocity in the upper 30 m of sites (V_{S30}) and basin depth provides a baseline against which other models are compared. In this study, we have investigated two site-specific models: ground response analysis (GRA) and square root impedance (SRI) method. These analyses are performed for sites meeting certain data availability requirements pertaining to number of recorded events and availability of site-specific V_S measurements from the PDB. We describe procedures developed to implement each approach, including protocols developed to assign unknown soil parameters. We describe a series of representative results for site with and without good fits of model predictions to observations. We compile results across sites to identify parameters that influence the effectiveness of ground response analyses and the epistemic uncertainty associated with different methods of site response prediction.

ACKNOWLEDGMENTS

Funding for this study was provided by the US Geological Survey under contract number G17AP00018, as well as the California Strong Motion Instrumentation Program, California Geological Survey, under Agreement Number 1016-985. Profile database work was also partially funded by the Consortium of Organizations for Strong Motion Observation Systems. Partial support for the first and third authors was also provided by the UCLA Civil & Environmental Engineering Department.

This support is gratefully acknowledged. The work presented here represents the views and opinions of the authors and does not reflect the policy, expressed or implied, of the US Government.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

1.1 PROJECT OVERVIEW

This project was comprised of two distinct yet mutually beneficial tasks. The first was the establishment of a United States Community V_S Profile Database (PDB). The PDB was developed by a community of researchers, with initial results presented by Ahdi et al. (2018) and Sadiq et al. (2018). Chapter 2 of this report summarizes the current status of the PDB, which is an operational resource as a result of the support of this project.

The second task involved the use of the PDB to support ground motion studies that establish observation-based site response at ground motion recording stations and then seek to establish the degree to which it can be estimated using alternate prediction approaches. Preliminary results of this work were presented by Wang et al. (2019). Chapters 3-0 of this report present the database used for these studies, the manner in which inputs to different site response models were developed, example data-to-model predictions for several sites, and a synthesis of results across the full inventory of 145 sites.

The specific motivations for the two major project tasks are described in the remainder of this chapter.

1.2 PROJECT MOTIVATION OF VS PROFILE DATABASE

Shear wave velocity is a critical parameter for ground motion prediction and many other applications including liquefaction. As such, large numbers of V_S profiles have been measured in the US, typically at sites in seismically active parts of the country. Prior to this project, there was no publically-accessible repository for these profiles that could be accessed by researchers as well as practicing engineers, geologists, and geophysicists.

Sources of profiles include USGS Open File Reports at strong motion stations, reports from industry, documents from state and local agencies, research publications, and documents provided by owners of major facilities such as power plants and dams. We focused on the western US in the development of the database. Future work will likely broaden this effort to include other regions, including the Pacific Northwest, the Intermountain West, the central and eastern US, and other regions.

The work established a formal relational database structure (i.e., a schema) for data archiving. We also established a GIS-based web portal to enable public access to the data. We manage the project through a Project Coordination Committee (PCC) with members from the USGS and others with a strong interest in this initiative.

1.3 PROJECT MOTIVATION OF SITE RESPONSE ANALYSIS

Ergodic models for site response provide a mean estimate conditioned on certain site parameters (typically the time averaged shear wave velocity in the upper 30 meters of the site, V_{S30}, and basin depth). The ergodic estimate of site response includes all site amplification mechanisms (impedance, nonlinearity, resonance, two- and three-dimensional wave propagation in basins, etc.), but these effects are smoothed over a large number of sites with different characteristics. As such, the associated site-to-site uncertainties (denoted ϕ_{S2S}) are substantial, increasing mean or >50%tile ground motions at long return periods as derived from probabilistic seismic hazard analyses (PSHA) relative to what would be obtained with more accurate methods.

Site-specific or *non-ergodic* site response is intended to account for wave propagation mechanisms at a specific site that control site response. An unbiased estimate of site-specific site response, for example as derived from analysis of earthquake recordings, substantially reduces ϕ_{S2S} (e.g., Rodriguez-Marek et al., 2014; Stewart et al. 2017). For sites without recordings, many projects seek to estimate site response using ground response analyses, which consider the effects of one-dimensional (1D) shear wave propagation and soil nonlinearity. Another method used for some applications include the square-root impedance (SRI) method (firstly introduced by Joyner et al, 1981, and later renamed by Boore, 2013). Open questions related to these practices are: (1) How effective are such methods at capturing observed behavior, and how does this change with period?; and (2) What levels of epistemic uncertainty (ϕ_{S2S}) should be used in PSHA when these site-specific site response methods are used?

A sensible means by which to answer these questions is through comparisons of predictions of ground response analysis results to data. Not surprisingly, this general line of research contains numerous contributions over many years, with a typical application taking various input motions, running them through 1D soil columns, and comparing resulting response spectra to those from recordings (e.g., Chang, 1996; Dickenson, 1994; Idriss, 1993). However, with the exception of vertical arrays, this research approach has a limited ability to answer the above questions, because predicted ground surface motions are strongly dependent on input motions, which are often highly uncertain. As a result, the effectiveness of the site response prediction is somewhat obscured.

The use of vertical arrays overcomes this problem because of the availability of recorded input motions, and has produced interesting findings that illustrate limitations, biases, and uncertainties associated with ground response analyses (e.g., Kaklamanos et al. 2013; Zalachoris and Rathje, 2015; Kaklamanos and Bradley, 2018; Afshari and Stewart, 2019). However, there are limitations associated with the use of vertical arrays to validate ground response analyses. First, the number of vertical arrays with sufficient ground motion recordings and site characterization is limited (but certainly growing with time). Second, vertical arrays only measure site response over the length domain of the array; as such they are not useful for evaluating long-period features that involve wavelengths longer than array dimensions. Third, the within-motion boundary condition that is used in analysis of vertical array data does not match that used in typical forward applications, in which outcropping input motions are selected.

To address these limitations, we apply here an alternative method for validating ground response analyses and other methods using data from surface-only instruments. The concept is to use recordings to infer the non-ergodic site response over a wide frequency range. The effectiveness of ground response analysis, and other methods, is then assessed by comparing predicted levels of site response against observation. This departs from the aforementioned prior work in that model effectiveness is not based on ground motions from a particular event (or series of events), but on the site amplification relative to a reference condition. This work was comprised of four components, as described in the following paragraphs.

The first component is assembling the required data. If not already available from another project (such as NGA projects, which involve substantial data collection and synthesis), this is a substantial task. The information required is identical to that needed for ground motion model development, namely, a database that includes information on source attributes, site conditions at recording stations, and ground motions (with record-specific processing details). In this project, we supplemented the NGA-West2 database with additional sites and events, as described in the next section. A need for the present work that is not shared in ground motion model development projects, is seismic velocity profiles at recording stations (particularly shear-wave velocity, V_S, vs depth profiles).

The second component consists of ground motion analyses targeted at extracting information on site responses at recording stations. The steps involved in developing these results are descried elsewhere (Stewart et al. 2017), so the procedure is not repeated here. What these analyses provide is an estimate of a site term, denoted η_s, for each site and response spectral oscillator period. This site term represents the mean difference between a regionally-unbiased ground motion model and observed motions at the site. For weak shaking conditions that do not induce soil nonlinearity, the sum of η_s and the ergodic site term for the site (F_S), comprises the mean non-ergodic site response (μ_{lnY}) relative to the ground motion model's reference condition:

$$\mu_{lnY} = F_S + \eta_S \qquad [1.1]$$

The third component consists of predicting site response for each site in the data inventory using available information on site conditions. In the case of GRA, a V_S profile is required, and borehole data indicating soil stratigraphy and soil type characteristics for each layer is also useful (for estimation of modulus reduction and damping relations). SRI also requires soil unit weight and V_S profiles.

The fourth component involves model-to-data comparisons in the form of residuals analyses for estimating model bias and uncertainty and the metrics for quantifying the overall model performance. These will be discussed in Chapter 5 and 6.

2 A United States Community Shear-Wave Velocity Profile Database

2.1 INTRODUCTION AND PROJECT MOTIVATION

Shear wave velocity (V_S) is a commonly used parameter for analysis in the fields of geotechnical earthquake engineering and engineering seismology. The shear modulus (G) of an earth material such as soil or rock relates directly to V_S and mass density (ρ):

$$G = \rho \cdot V_S^2 \quad [2.1]$$

As such, V_S can be used to describe material behavior in response to seismic wave propagation. Common analyses utilizing V_S include semi-empirical ground motion modeling (e.g. NGA West2 ground motion models described in Bozorgnia et al. 2014), site amplification studies (e.g., Joyner et al. 1981), assessment of soil ageing (Ohta and Goto 1978), semi-empirical models to assess liquefaction triggering potential (Andrus & Stokoe 2000, Kayen et al. 2013), and soil-structure interaction models (e.g., NIST 2012).

V_S profiles as a function of depth are obtained using geophysical methods applied in the field or, in relatively rare cases, using dynamic laboratory experiments on specimens retrieved from the field. Such measurements are commonly performed as part of seismic site characterization for research purposes (typically at strong motion recording stations) and for critical projects where design ground motions are to be developed, typically using site-specific procedures (e.g., Chapter 21 of ASCE 7; ASCE 2016). Examples of such projects include seismic assessments of nuclear power plants (e.g., Final Safety Analysis Reports submitted to the U.S. Nuclear Regulatory Commission, NUREG), dams, tall buildings, bridges, hospitals, and other critical infrastructure. Individual state and federal agencies, researchers, and engineering companies have, in many cases, published reports containing V_S profiles and related geotechnical data. However, in few cases has this information been synthesized and compiled (exceptions include Boore (2003a) for USGS open-file reports before 2003, and SW-AA (1980) for preceding NUREG research reports). In other cases, the data is maintained in internal databases that are not privileged (i.e., it is public domain), but are for practical purposes inaccessible to most potential users. In short, a vast amount of useful information has been measured and documented in some form within the Unites States, but for many potential users, most of this information is effectively inaccessible and is therefore not useful. This project was conceived as a means by which to "bring the data to the people" in the form of an open-access community V_S profile database (PDB).

The V_S PDB encompasses and extends existing databases in the United States for V_{S30} (the time-averaged V_S in the upper 30 m of the Earth's crust) (Yong et al. 2016) and cone penetration test (CPT) soundings that include V_S measurements (Holzer et al. 2010). Additionally, V_S data compilations have been prepared for a few individual states including Washington (Bilderback et al. 2008) and Oregon (Roe and Madin, 2013). A 3D geotechnical database primarily consisting of geotechnical unit descriptions from 1000 boring logs but with an unidentified number of V_S profiles was developed by Doroudian and Vucetic (1997) for Los Angeles after the Northridge earthquake, and a similar database of geotechnical data was created for Palo Alto, CA region by Iskandar et al. (1996). Those studies integrated the database with proprietary geographic information systems (GIS) software. The present effort is different from this prior work in that (1) full V_S profiles are collected and presented, not only V_{S30}, (2) the focus is primarily on collecting V_S profiles and including borehole and other geotechnical data to augment the geophysical test results, (3) V_S measurements are included from many geophysical methods and diverse data sources, and (4) data is organized according to a uniform schema within a formal relational database (RDB) accessible via a web interface. Table 2.1 summarizes web-accessible compilations of either V_S profiles or V_{S30} within the U.S. and globally. Some global data sets are similar in concept to what is described here (e.g., Japan, Taiwan, Italy), a principle difference being those datasets are focused solely on ground motion recording sites, whereas all sites with available V_S information that has been accessed by the project team are considered here. The New Zealand Geotechnical Database (NZGD 2017) is similar with respect to diversity of the sites incorporated, but is different in its inclusion of many sites without V_S profiles and its narrow geographic extent.

In this chapter the various data types that are included in the database are described. While all sites have a V_S profile, the geophysical methods used to obtain them vary as does the presence of additional data and metadata. Key statistics of the database are summarized at its current stage of development, in which the principal focus has been on California sites. Major data sources are described, some of which had particular features that impacted the database structure. A major component of this work is the development of the RDB schema, the structure of which is described herein, along with the front-end user interface with the database. An early version of the profile database was presented by Ahdi et al. (2018), which described initial database development and data collection efforts. In addition, Sadiq et al. (2018) presented the initial database schema.

2.2 DATA TYPES

The main criteria for including a site in the database is availability of a V_S profile measured in situ using one or more seismic geophysical methods at known locations. Published locations take many forms (e.g., maps, street addresses, Cartesian coordinates such as UTM, or geodetic coordinates using various datums); for this project, it is required that the geodetic coordinates (i.e. latitude and longitude) can be identified by some means, and as needed, coordinates are converted to the World Geodetic System (WGS84) coordinate reference system standard. The accuracy of geodetic coordinates for sites and locations of specific data is paramount for this project, given that the end user primarily interacts with the database through a map interface. The process of ensuring location accuracy for data sets where the interpretation of measurement locations is needed is provided in Section 2.3. The locations of sites in California that are currently represented the PDB are shown in Figure 2.1 (for a small number of the sites shown in the figure, the data has been obtained but has not yet been uploaded to the database).

Table 2.1 Existing web-accessible databases containing V_S profile information

Name/Host Organization	Location	Contents	Primary Focus	Web Link/Reference
Compilation of V_{S30} Data for the United States, USGS	United States	V_{S30}, site metadata	V_{S30} for strong motion recording stations	Yong et al. 2016; https://earthquake.usgs.gov/data/vs30/us/
Cone Penetration Testing (CPT) Data, USGS	United States	CPT data, V_S profiles, site metadata	Collection of CPT measurements	Holzer et al. 2010; https://earthquake.usgs.gov/research/cpt/data/
Geo-Station, National Institute for Earth Science and Disaster Resilience (NIED)	Japan	Station metadata; boring logs and geophysical data	V_S for strong motion recording stations	http://www.geo-stn.bosai.go.jp/jps_e/
National Center for Research on Earthquake Engineering (NCREE)	Taiwan	V_{S30}, $z_{1.0}$, and κ_0	Site parameters for Taiwan Strong Motion Instrumentation Program	http://egdt.ncree.org.tw/
Engineering Strong Motion Database, INGV and Orfeus	Italy/ Europe	Station metadata including V_{S30} and V_S profiles	Strong motion database with event and station list	https://www.orfeus-eu.org/stationbook/
European Geotechnical Database, Aristotle University of Thessaloniki and European Plate Observing System	Greece/ Europe	Station metadata including V_{S30} and f_0	V_S for strong motion recording stations	http://egd-epos.civil.auth.gr/
Earthquake Commission and Ministry of Business, Innovation, and Employment, New Zealand Government	New Zealand	Geotechnical data (SPT and CPT); V_S profiles	Geotechnical database developed following Canterbury earthquake sequence of 2010-11	NZGD 2017; https://www.nzgd.org.nz/

Note: $z_{1.0}$ = depth to 1.0 km/s V_S horizon; κ_0 = high-frequency attenuation of the acceleration Fourier amplitude spectrum in a log–linear space (Anderson and Hough 1984); f_0 = site fundamental frequency.

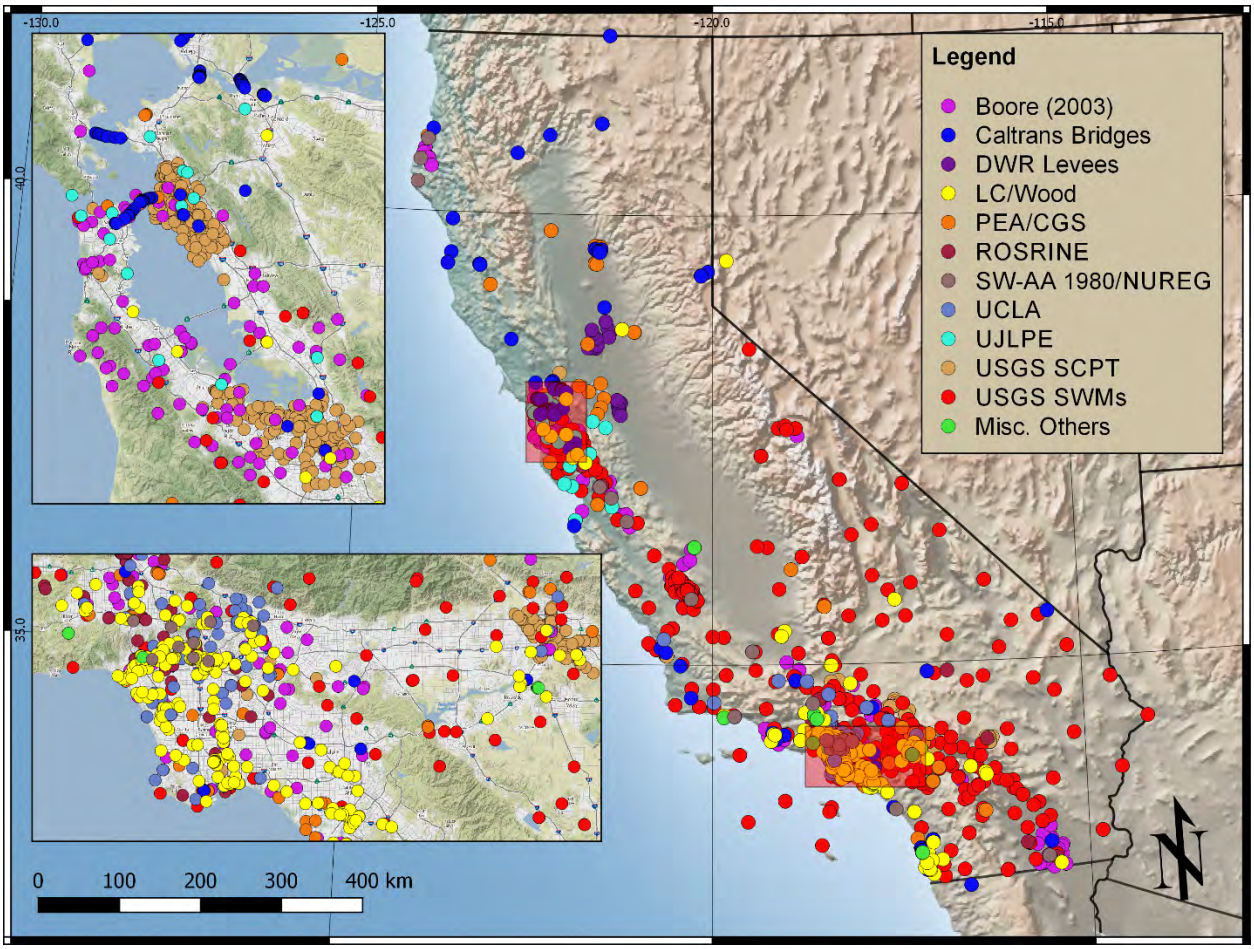


Figure 2.1 Map of California with major V_S PDB data sources (descriptions and definitions of abbreviations in Section 2.3).

V_S profiles derived from correlations to standard penetration testing (SPT) blow counts (e.g., Brandenberg et al. 2010) or CPT tip resistance (e.g., Wair et al. 2012) are not considered for inclusion, as these are not direct seismic velocity measurements. Additionally, V_S profiles constructed from dynamic laboratory tests are not considered for inclusion, as sample disturbance effects will typically produce an under-prediction bias of V_S (Anderson and Woods, 1975; Ishihara, 1996). Additional data that was added to the database when available include P-wave velocity (V_P) profiles, geotechnical/geological boring logs, co-located SPT and CPT measurements, laboratory test data, and horizontal-to-vertical spectral ratios (HVSR).

Although the format of the PDB differs from source documents, it was sought to present data as close as possible to how it was presented in source documents. Exceptions related to V_S profile data are rare and are discussed below (e.g., capping of "infinite" profile depths). The project team's intent was to serve as *purveyors* of the existing data, rather than analysts.

2.2.1 Metadata

Compiled metadata for each site includes a unique site number, site name, geodetic coordinates, description of geographic locality of the site (city, state, county) for data query purposes, various morphological attributes obtained from digital elevation models (DEMs), such

as elevation, topographic slope gradient, and geomorphic terrain classes (e.g., per Iwahashi and Pike 2007 and Iwahashi et al. 2018), and surficial geological units and descriptions from digital geological maps. This metadata is uniformly assigned to all sites in California, using the site geodetic coordinates. The quantitative terrain morphology information is obtained from the Shuttle Radar Topography Mission DEM at 30 arcsec resolution (~1 km grid spacing) (Farr and Kobrick 2000). The surficial geological information is obtained from the compilation map of California geological units in Wills et al. (2015). Spatial data is obtained in digital shapefiles for vector data or raster files for gridded data such as DEMs, and a spatial join is performed using the R packages `'raster'`, `'RGDAL'`, and `'sp'` (Hijmans 2018; Bivand et al. (2018); Bivand et al. 2013).

In many cases similar metadata exists in source documents, links to which are provided in the database. The uniform retrieval of this information reduces manual work and potential for data entry errors, and ensures consistency across all sites. Examples of other information that may be useful to users and can be found in source documents include top-of-borehole elevation or field-identified surface geology.

2.2.2 Geophysical Testing Methods

Sites with V_S profiles measured using either invasive or non-invasive geophysical methods are included in the database. Invasive methods include downhole (Warrick 1974) and crosshole testing (ASTM 2014), P-S suspension logging (Thiel & Schneider 1993), and seismic CPT (SCPT; Lunne et al. 1997). Non-invasive methods include active-source surface wave methods (SWMs) such as the spectral analysis of surface waves (SASW, Stokoe et al. 1994) and multi-channel analysis of surface waves, using both Rayleigh waves (i.e. MASW, Park et al. 1999a) or Love waves (i.e. MALW, Mari 1984, Yong et al. 2013); 1D and 2D passive-source SWMs (i.e. microtremor array measurements [MAM]) including spatial autocorrelation (SPAC) and extended spatial autocorrelation (ESAC) methods (Kanai et al. 1954; Aki 1957, Horike 1985; Okada et al. 1990); and body wave methods such as P- and S-wave seismic refraction (Redpath 1973, Telford et al. 1990).

Figure 2.2 breaks down V_S profile data in the database by geophysical method, both including and excluding a large subset of SCPT data from the USGS CPT database (described in Section 2.3) to emphasize the distribution of the other methods; the USGS CPT database contains 1221 sites and otherwise dominates the distribution of data by geophysical method. An important feature of Figure 2.2b is the significant fraction (56%) of V_S data obtained using non-invasive surface- and body-wave methods, which are becoming increasingly acquired and used in the field of engineering seismology (e.g., COSMOS, 201x).

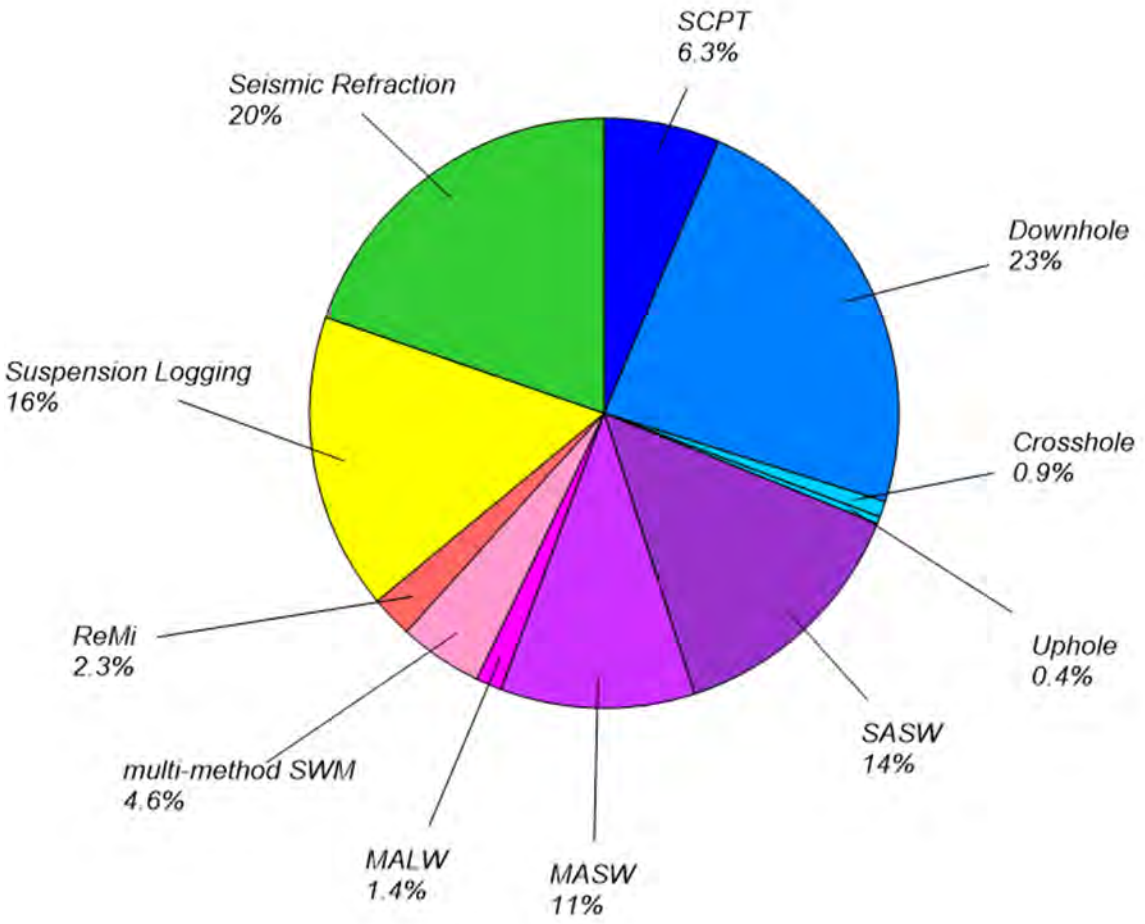


Figure 2.2 Distribution of V_S profiles included in database to date by geophysical method

In populating the database, V_S profiles were collected and digitized from geophysical methods that were deemed to be credible. Priority was not given to the inclusion of data derived from certain methods that have been shown to provide biased V_S measurements under typical conditions. One such method is controlled-source surface-wave dispersion measurement (CXW), a SWM developed by Poran et al. (1994). While there is a sizable study (91 sites) that utilized this method in southern California (Rodriguez-Ordoñez 1994), as well as another study (Dutta et al. 2000) that provides a large fraction of the available V_S profiles in the Anchorage, Alaska, region, studies by Boore & Brown (1998) and Wills (1998) have shown the CXW method to produce biased V_S profiles compared to profiles derived from invasive geophysical methods such as downhole and crosshole at a subset of overlapping sites in Los Angeles. While Boore & Brown (1998) and Wills (1998) found that V_{S30} specifically is not significantly biased, given that the present project is focusing on V_S profiles, these profiles have been excluded.

A second method given lower priority for inclusion in the database at this time is Refraction Microtremor (ReMi™), developed by Louie (2001). While the method is widely used by industry practitioners and comprises a significant portion of data available in some regions (e.g., Nevada), studies have shown potential errors in low-frequency phase velocities from passive-source SWMs with linear array geometry (Cox and Beekman 2011, Strobbia and Cassani 2011). As ongoing research investigates the proper application and potential issues and pitfalls in using ReMi™, such

as those described in Yong et al. 2013, the potential inclusion of this data in the V_S PDB will be reconsidered. Presently, the database contains 34 ReMi^{TM}-derived profiles located at hospital and school facilities that are part of the dataset provided by the California Geological Survey staff (C. Wills 2015, *pers. comm.*; see Section 2.3.3.4).

Aside from CXM and ReMi^{TM}, all data sources were accepted, and data presented in the database was not otherwise screened. It is acknowledged that some individual profiles may be considered problematic given the evolution in the state of knowledge since the work was completed. Examples of such concerns include the geometry of SWM arrays (Zhang et al. 2004), consistency of making travel-time picks for body-wave methods (e.g., Boore & Thompson 2007), or matching of experimental and theoretical dispersion curves using different inversion methods (Foti et al. 2011). Rather than check the data in relation to such criteria, the necessary information is provided, as available from source documents, to allow users to assess data quality and usefulness. The development of quantitative metrics for data quality has been discussed, but no such consensus metrics have been defined by the community or implemented here.

2.2.3 Computed Metadata

The database includes some computed parameters that are widely used. Primary among these is V_{S30} and its more general form, V_{SZ}, which is the time-averaged V_S to the maximum profile depth (i.e., z_p). V_{SZ} is computed as:

$$V_{SZ} = \frac{z_P}{\int_0^{z_P} \frac{\mathrm{d}z}{V_S(z)}} \qquad [2.2]$$

where $V_S(z)$ is the shear wave velocity at depth z from the profile.

The computation of V_{S30} requires substitution of z_p with 30 m in Eq. [2.2]. Note that V_{S30} is only computed for V_S profiles having $z_p \geq 30$ m; V_{S30} is not estimated for profiles with shallower maximum depths, as the choice of extrapolation method is subjective. There exists a body of literature that discusses this extrapolation, to which the database user is referred; a summary of V_{SZ}-to-V_{S30} extrapolation methods and a statistical comparison of their relative performance considering a global collection of V_S profile data is provided by Kwak et al. (2017). Time-averaged P-wave velocities have not been computed.

A problem arose in V_{SZ} computation for a subset of profiles (4%) where the original investigator presented the bottommost layer as having infinite thickness representing a semi-infinite half-space; in such cases, a value of z_p was assumed based on the measurement method and available information. For V_S profiles derived from SWMs, a general rule of thumb is that the maximum resolvable depth of investigation is the maximum measured wavelength multiplied by 2 (Heisey et al. 1982) and this assumption was used for 10 V_S profiles. For measurement methods besides SWMs, simple protocols for assigning z_p were developed. For invasive methods except suspension logging, z_p was simply taken as the deepest value at the bottom of the borehole or SCPT measurement. Because suspension logs generally have overlapping measurements at regular intervals at half the distance between the two receivers in the device's probe (e.g., if the receiver spacing is 1 m, and a measurement is taken every 0.5 m, there will be two overlapping measurements at each depth), z_p was taken as the deepest point plus the thickness of an additional

half layer. For body-wave methods such as S-wave refraction/reflection, the half-space layer was assumed to have double the thickness of the layer above it, and z_p is measured to that depth.

Figure 2.3(a) and Figure 2.3(b) show distributions of V_{SZ} and V_{S30} (for sites with $z_p > 30$ m) and z_p for profiles in the database. A common feature of V_{S30} datasets such as this is a histogram which takes the shape of a positively-skewed lognormal distribution (e.g., Seyhan et al. 2014, Figure 2; Ahdi et al. 2017). The reason for this is a sampling bias towards profiles measured in urban areas, which are more commonly located in deeper sedimentary basin environments that have lower V_{S30} than shallow-rock or hard-rock sites. Included here are multiple studies (e.g., Yong et al. 2013, Boore, 2003a) which made concerted efforts to also collect V_S data at stiffer sites.

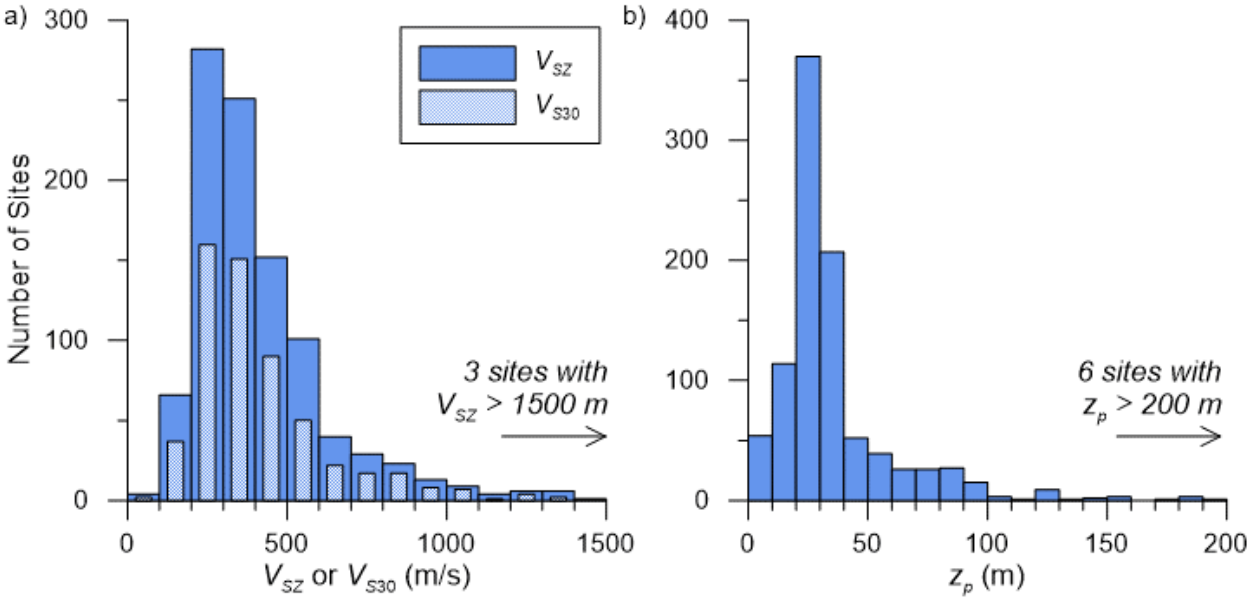


Figure 2.3 Distributions of (a) time-averaged V_S to z_p (V_{SZ}) and to 30 m depth (V_{S30}) and (b) maximum V_S profile depth (z_p) in the V_S PDB at time of writing.

2.3 DATA SOURCES

The process of collecting data form various sources is an extensive and ongoing effort. Data was collected primarily from published studies, including grey literature, from federal and state agencies and geological surveys, university research, and public data and reports from organizations and private firms that are otherwise relatively inaccessible to most potential database users. Whenever possible, the project team reached out to original investigators to confirm the data as public and to request the data in a digital form to expedite its integration into the RDB. This interaction was also useful in that any issues and inconsistencies with the dataset could be resolved with the guidance of the original investigator(s). Efforts were also made to connect with private consulting firms and utility companies who were willing to share parts of their internal databases. This section discusses individual data sources, the process of acquiring, digitizing, and unifying their formats for incorporation into the database, and quality checks that were performed in some cases.

2.3.1 Data Access Policies

Public datasets are those that are freely available in the public domain, such as data obtained for research purposes and data from reports that have been submitted for building construction and reviewed by state and/or municipal jurisdictions. While these data are considered as part of the public record (e.g., from reports for public projects like bridges, dams, hospitals, etc.), from a practical standpoint they have limited accessibility outside of the regulatory agencies that retain the documents containing the data. In contrast, *private* datasets are owned by a particular person or entity for its own use. Such data is most commonly developed by consulting firms for projects not subject to review by regulatory agencies, or less commonly, by researchers not funded by public agencies. The proprietary nature of these data may preclude their inclusion into the database. However, some data sharing among private industry may occur if firms and owners identify a business justification for data sharing. For example, in New Zealand, firms can utilize a public repository for geotechnical data provided that they also contribute to it (NZGD 2017).

To a limited extent, we have pursued the collection of data held by private consulting firms that maintain large libraries of V_S profile data, with attention paid to forming agreements to enable the inclusion of such data in the database. To date, this has occurred with Wood plc. (formerly LeRoy Crandall and Law/Crandall); their dataset is described in Section 2.3.5.1 below. While this data has been collected by a private firm, it is public in a way, as it is part of projects that were reviewed by public agencies.

The use of truly private data, as described above, would potentially introduce issues with accessibility and distribution. It is also recognized that participation of the USGS in this effort requires a data product consistent with USGS Fundamental Science Practices (USGS Fundamental Science Practices Advisory Committee 2011) to have all data freely accessible. As a result, the collection of strictly private data has not occurred to this point. Inclusion of private data in the future of the database project will be contingent upon successfully reaching agreements with data owners to include and publish their data, and a special flag will be provided to such data to initially exclude them from the public database.

2.3.2 Federal Data Sources

The original geographic focus of the V_S PDB was the state of California, but that focus soon expanded to include data from the entire U.S. Nevertheless, California is where the bulk of the current database resides, with data included from both federal sources (as described in this section) and state agencies (described in Section 2.3.3 below).

2.3.2.1 USGS Open File Reports

V_S profile data was considered from a variety of measurements performed mainly in California but also throughout the U.S. Data from 26 USGS open-file reports were incorporated, subsets of which have been reprocessed in two major data synthesis efforts for downhole logging (Boore, 2003a) and SCPT (Holzer et al. 2010). These reports are described in the subsections below.

USGS CPT Database

Holzer et al. (2010) presents CPT soundings performed across the U.S. The data is grouped by geographic region and were measured primarily to evaluate the liquefaction potential of different types of surficial geologic deposits in different regions of the U.S. The database contains 1807 total CPT investigations, including for most sites the tip resistance, sleeve friction, and inclination angle measurements, soil behaviour type (Robertson et al. 1986), and site-level metadata. SCPT travel time logs are included for 1221 sites (68%). Data is presented in a website with links to either PDF images or ASCII text files for each site. In a departure from the minimum criteria for data inclusion for this project, V_S profiles are *not* presented for these sites, but rather only the travel time logs provided from the SCPT downhole measurements were presented. This method of presentation is consistent with the intent of not analyzing the data within the PDB. To date this dataset is one of the major sources of V_S data included in the U.S. outside of California.

USGS Compendium of Downhole Data

Starting in 1975 the USGS began a campaign of performing geotechnical and geophysical field investigations in California at seismic stations that recorded strong ground motions in recent earthquakes or locations with documented shaking intensities in the 1906 San Francisco earthquake. Major focuses were clusters of sites in and around San Francisco and the southern San Francisco Bay Area, many of which recorded the 1979 Coyote Creek, 1984 Morgan Hill, and 1989 Loma Prieta earthquakes. Investigations were also performed in the greater Los Angeles region, including surrounding mountains, valleys, and coastal plains, many of which recorded the 1971 Sylmar, 1987 Whittier, or 1994 Northridge events. Additional investigations have been performed in the Imperial Valley (1979 Imperial Valley and 1987 Superstition Hills earthquakes, among others), Antelope Valley, Ventura basin in vicinity of Oxnard and Ventura, the Santa Barbara coastal plane, Parkfield and surround areas, and other parts of the state. The only geophysical method used for these studies was seismic downhole, recording both P- and S-wave velocities. Geologic logs with varying degrees of detail in lithostratigraphic descriptions are also provided, along with SPT blow counts and laboratory index test results in some cases.

Boore (2003a) reprocessed all pre-existing USGS downhole data, comprising 277 sites originally presented in 20 open-file reports, five journal papers, and data from 14 previously unpublished sites. The reprocessed data is presented in digital files available at his website (http://daveboore.com/data_online.html). For the V_S PDB, V_S and V_P profiles, stratigraphic information, SPT blow counts, and laboratory test results were included. Travel time data are presented in the original reports, but in most cases they are not tabulated, so they were not digitized for inclusion in the V_S PDB.

Aside from a formidable data entry task (including boring logs, SPT blow counts, and laboratory data), the principle challenge in preparing this data for the PDB was associated with the location of the borehole in which the seismic velocity measurements were performed. Among the main problems were the reprojection of datums used for assigning site coordinates and lack of precision in latitude/longitude coordinates that resulted in the improper plotting of the data on a modern map. (*Note*: in this project, the aim was to have 0.00001-degree precision for geodetic coordinates, which is on the order of ~1 m accuracy. Some older data had only 0.01-degree or 0.001-degree precision, resulting in a potential error of up to ~1 km.) The following procedure was applied to ensure locations are as accurate as possible for this dataset:

- Reprojection of all coordinates from the NAD27 datum to WGS84. Three of the most recent open-file reports (00-470, 01-506, and 02-203 [Gibbs et al. 2000, 2001, and 2002, respectively]) explicitly stated that the datum used was NAD27; older reports did not state the datum used, but Dave Boore (2018, *pers. comm.*) confirmed that the intent of reprocessing site information in the 2003 compendium was to use NAD27 for all sites. The general result of the reprojection is that the site location moved west ~100 m (at the northernmost site near Eureka, CA) to ~80 m (at the southernmost site in Calexico, CA), with no appreciable shift in the north-south direction.
- Plotting of all sites on Google Earth, using two points for each site: one for each coordinate based on the NAD27 and WGS84 datums. Each site in most of the original open-file reports summarized in the 2003 compendium was plotted on a historical 7.5-minute (1:24,000-scale) USGS topographic base map, an ongoing USGS effort that has been maintained for many years by the USGS (Fishburn et al. 2017). The georeferenced KMZ files for these maps are available from the USGS TopoView online database (https://ngmdb.usgs.gov/topoview/); the sites are overlain onto these historical maps. The site plan in the original USGS open-file reports has an icon indicating the approximate location of the site on the map; this location is compared to NAD27 and WGS84 points. In general, the coordinate reprojection to WGS84 was successful: for 157/277 sites (57%), the new value was used, and for only 2 sites the NAD27 coordinate was retained. This perhaps occurs because the original site plan's marker was indeed closer to this coordinate.
- Often, the reprojected (WGS84) coordinate did not line up properly on the historical map when compared to the site plan. In these cases, judgment was used to select new latitude/longitude coordinates to 0.00001-degree precision using Google Earth/Google Maps based on the plotted location in the original site plan. In some cases, the coordinate was moved hundreds of meters away: for example, Figure 2.4 shows the example of the “Quintara” site (USGS Open-File Report 77-850 [Gibbs et al. 1977]), which was moved approximately 1.8 km south based on the combination of comparing the site plan with the historical USGS Topo map, and by identifying street names and cultural markers near the locations of the original markers (on Lincoln Way at the southern end of Golden Gate Park) with Quintara Street (nine blocks south of Lincoln Way). Most cases were less egregious than Qunitara, where small adjustments on the order of tens of meters were made to result in a plot marker location in Google Maps satellite imagery. Examples include ensuring the site would be located in the right type of property based on a descriptive site name, not located in the middle of a street, or not located on the side of steep hill slopes where drilling would have been unlikely. All in all, judgment was used to relocate 105/227 (38%) of sites in this manner.
- Sometimes, the previous location was simply not sensible: for example, the site “Brentwood VA Hospital” plotted approximated 83 m east of its actual location, resulting in the location marker to fall in a residential area. Because in this case this would have been presented at an obviously erroneous location, the effort given to adjust coordinates is justified.
- For the remaining 13 sites (5%) there exist no site plans in their original reports, or their locations were plotted on overall regional maps rather than the more descriptive

7.5-minute-scale topographic maps used for verification. The lack of this site information precluded us from performing the same level of diligence in verifying the coordinate locations. As such, the reprojected WGS84 coordinate was used as described in the first bullet above.

Figure 2.4 Google Earth screen capture (above) of process used to relocate Quintara site in western San Francisco. The original location, based on the NAD27 datum, is represented by a red square; the location based on reprojected coordinates into the WGS84 datum is represented by a green circle; and the final location, represented by the yellow pin, was selected based on the original site plan (below) in USGS Open-File Report 77-850 (Gibbs et al. 1977). Note the location of the "Robert Louis Stevenson School" just east of the site, a cultural marker used to help identify the actual location in Google Earth using the georeferenced USGS historical topographic map.

USGS Studies Utilizing Surface-Wave Methods

Two main groups of researchers at the USGS have published reports to date utilizing surface wave methods to characterize strong-motion sites, most of which are in California with the remainder in CENA. One such group is Yong et al. (2013), who characterized 191 sites (187 in California and 4 in CENA) using multi-method surface wave testing, including combinations of active-source (e.g., SASW, MASW, P- and S-wave seismic refraction) and passive-source (1D, 2D, and single-station ambient/microtremor array) methods. Known as the ARRA project (due to funding from the American Recovery and Reinvestment Act), this study advanced the capabilities and state of knowledge in utilizing multiple SWMs and processing methodologies (including both Rayleigh- and Love-wave dispersion curve modeling) in both challenging and "well-behaved" geological environments. This project focused on stiffer and hard-rock sites in addition to the more common softer sites where most V_S data exists in the literature.

The ARRA project data presented a number of unique attributes for which judgment had to be exercised when incorporating the data into the V_S PDB. The two main issues were multiple measurements at the same site and the assignment of a location to different measurement methods. The multiple measurements issue is well illustrated in Figure 2.5 for ARRA site "CI.SLB," where two co-located SWMs (SASW and MASW) were used to obtain a combined experimental dispersion curve. Yong et al. (2013) considered the dispersion data both separately and together as a composite dispersion curve and thus obtained three different V_S profiles (one each from SASW, MASW, and a combination of the two methods), as shown in Figure 2.6. For this site, the V_{S30} values for the each of the three profiles were computed as 339, 339, and 336 m/s, respectively, and the ARRA project authors chose the latter because it combines SASW and MASW. However, for the V_S PDB, the project team's primary interest was to isolate the profiles from different measurement methods. Accordingly, three co-located V_S profiles were presented for this site in the V_S PDB. More generally, all the profiles developed for a site were not included if the authors indicate their preference against using a specific profile for characterization of a given site. Notes from the original ARRA report's Appendix were included to clarify pertinent issues regarding individual sites and profiles and how these data were incorporated into the V_S PDB. Also, V_S profiles computed from P-wave refraction methods using V_P/V_S ratios were explicitly excluded, as these are not a direct measurement of shear-wave propagation. Accounting for these considerations, it was ultimately determined that 61/191 sites (32%) have multiple V_S profiles that are valid for presentation in the V_S PDB; another 127 sites (66%) have only one V_S profile that is either presented or recommended among multiple profiles by the original authors; two sites do not present valid V_S profiles (CI.TOR and CI.WNS), and one site is a duplicate of another and was combined with its partner site (CI.BLA2 into CI.BLA).

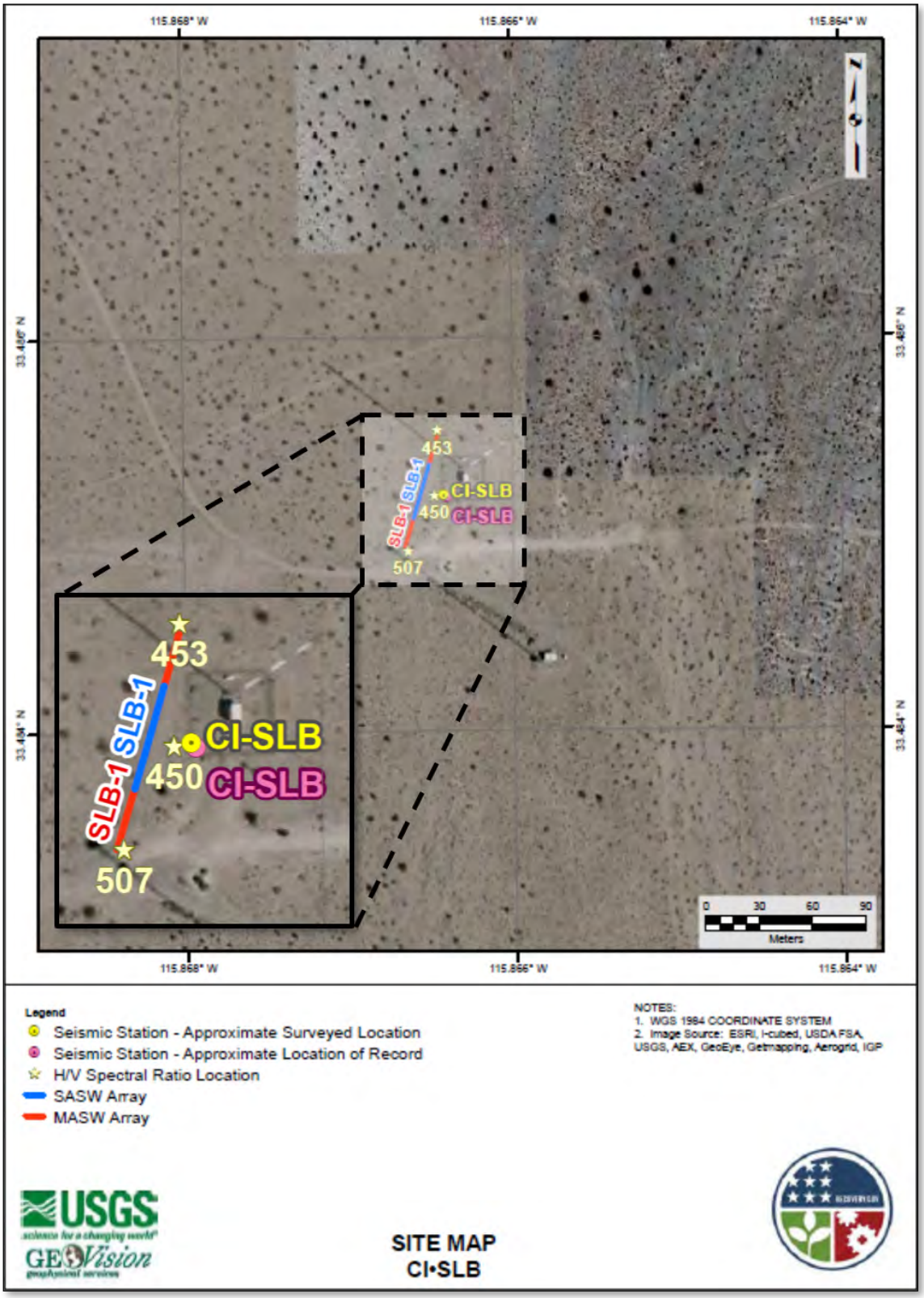


Figure 2.5 Site Map for site CI.SLB from Appendix of Yong et al. (2013). Area of zoom detail displays locations of collinear SWM testing arrays for SASW and MASW methods.

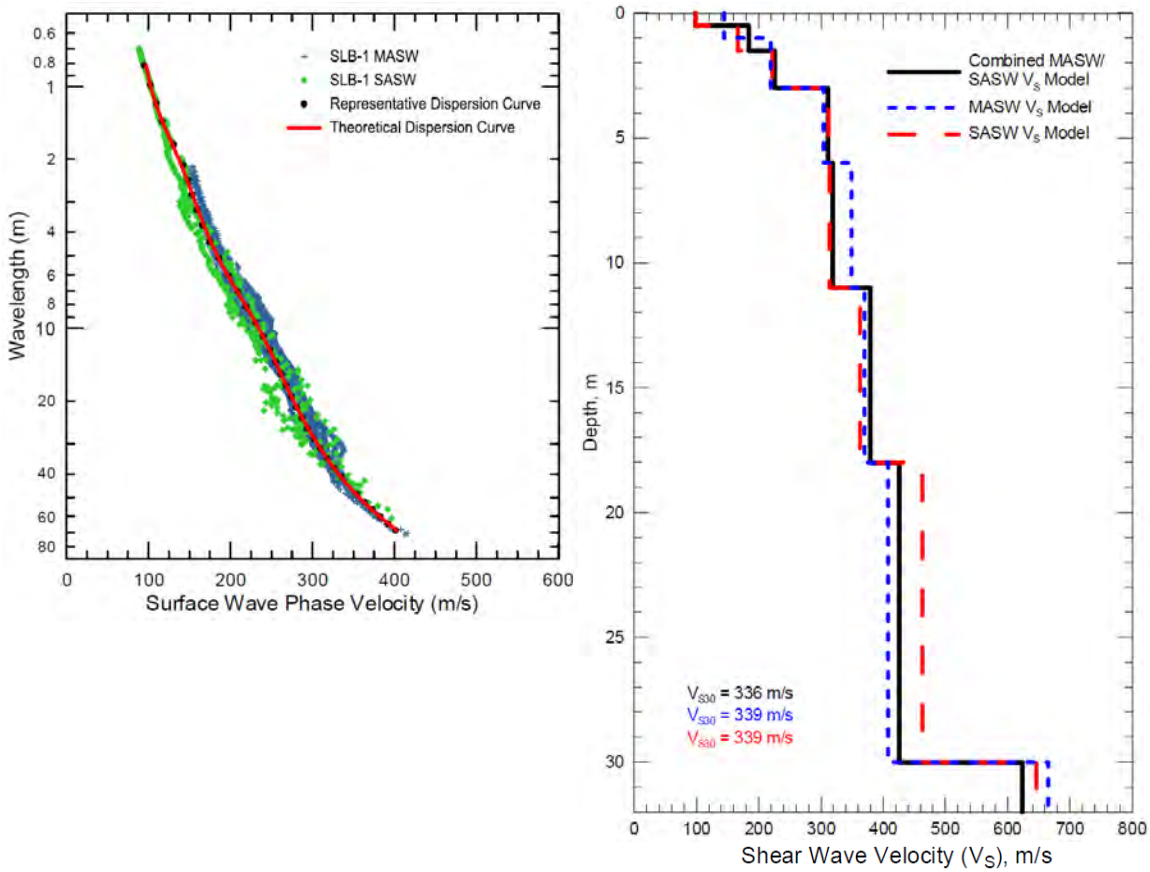


Figure 2.6 Dispersion curves (left) and inverted V_S profiles (right) for site CI.SLB from the Appendix of Yong et al. (2013). Two collinear SWM arrays are deployed at the same site, resulting in three potential modeling routines based on dispersion data from either or both MASW or SASW.

The measurement location issue pertains mainly to the spatially-distributed nature of measurements made using SWMs: sensor arrays are not typically a point in plan view, as is the case with borehole methods. For linear arrays, such as those used in SASW, MASW, seismic refraction, etc., the midpoint of the array was taken as the location of the obtained velocity profile; this approach is in line with that of Yong et al. (2013). For more complex 2D array geometries, such as L- or T-shaped arrays, or circular or triangular multi-station microtremor array measurements for SPAC, the identification of a representative single point is less straightforward. In general, wherever the level of detail of reporting is sufficient to isolate dispersion curves from different "legs" in L- and T-shaped arrays, the midpoint of each leg was taken as the profile location. If this decomposition was not possible, then coordinates at the intersection point of the two "legs" was taken as the profile's location. For circular and triangular arrays, the location of the center-most sensor was used to mark the profile location. Continuing the example of site CI.SLB, Figure 2.5 shows that the collinear MASW and SASW arrays, although of different lengths, happen to have the same midpoint, and as such, all three V_S profiles for this site are assigned the same coordinates. Overall, the aforementioned 61 sites with multiple V_S profiles yielded 166 V_S profiles in total, with the number of profiles per site ranging from 2 to 7. Comments are included where appropriate in the metadata for each site to explain these differences. Counting

these with sites that have one V_S profile, in total 290 V_S profiles are presented in the V_S PDB from the ARRA project.

Besides the ARRA effort, another group of USGS researchers have published three USGS OFRs that present data collected from harmonic-wave source modified SASW methodologies in California. (Kayen et al. 2005a,b; Thompson et al. 2010). These three reports present V_S profiles and dispersion curves at 125 strong motion stations that produced recordings from the 2004 Parkfield earthquake and other events throughout California. V_S profiles, dispersion curves, and site-level metadata were presented for each site included in the database. In some cases, draft versions of the open-file reports from these studies showed comparisons of measured to theoretical dispersion curves where occasional mismatches occurred; these were updated in the final versions of the reports and the information reported in the PDB reflects the final, corrected, values.

Thompson et al. (2010) present two V_S profiles for each site to demonstrate non-uniqueness in the dispersion curve inversion. They also present comparisons of their SASW-derived V_S models with downhole measurements at four sites; this is reproduced in Figure 2.7 below. Three of the four sites have favorable comparisons, while for one site (840RFU/Red Hills) the SASW models significantly underpredict V_S at nearly all depths compared to the downhole profile, except in the near surface where SASW did not capture a shallow thin layer presented in the downhole model. This feature is intriguing because it is generally understood that SWMs capture thinner layers at shallower depths while losing thin-layer resolving capabilities as longer surface-wave wavelengths, which image deeper parts of the profile by averaging larger volumes of earth material (Sanchez-Salinero et al. 1987). This example illuminates a benefit of the V_S PDB: clusters of profiles in close proximity and from different measurement methods allow users to data quality. In such cases of discrepancies at a given site or between data sources such as this one, judgment is deferred to the database user in selecting a V_S model.

2.3.2.2 United States Nuclear Regulatory Commission Reports

A series of studies in the 1970s commissioned by the U.S. Nuclear Regulatory Commission (NUREG) investigated strong motion sites that recorded the 1971 Sylmar earthquake and a few other selected events from that period. They also investigated other sites of interest in high seismic hazard regions of the U.S., including Montana and New Madrid. Reports were published by a consortium of the private firms Shannon & Wilson and Agbabian Associates (hereafter SW-AA). The studies focused mainly on the geotechnical, geological, and seismic site characterization of strong motion recording stations. While multiple reports were published, a compilation report (SW-AA 1980) summarizes data from all sites, and the reader is referred to their Table 2.1 and Table 2.2 for references to the individual constituent reports, which tend to contain more detailed site-specific information, such as site plans that the compilation report excluded.

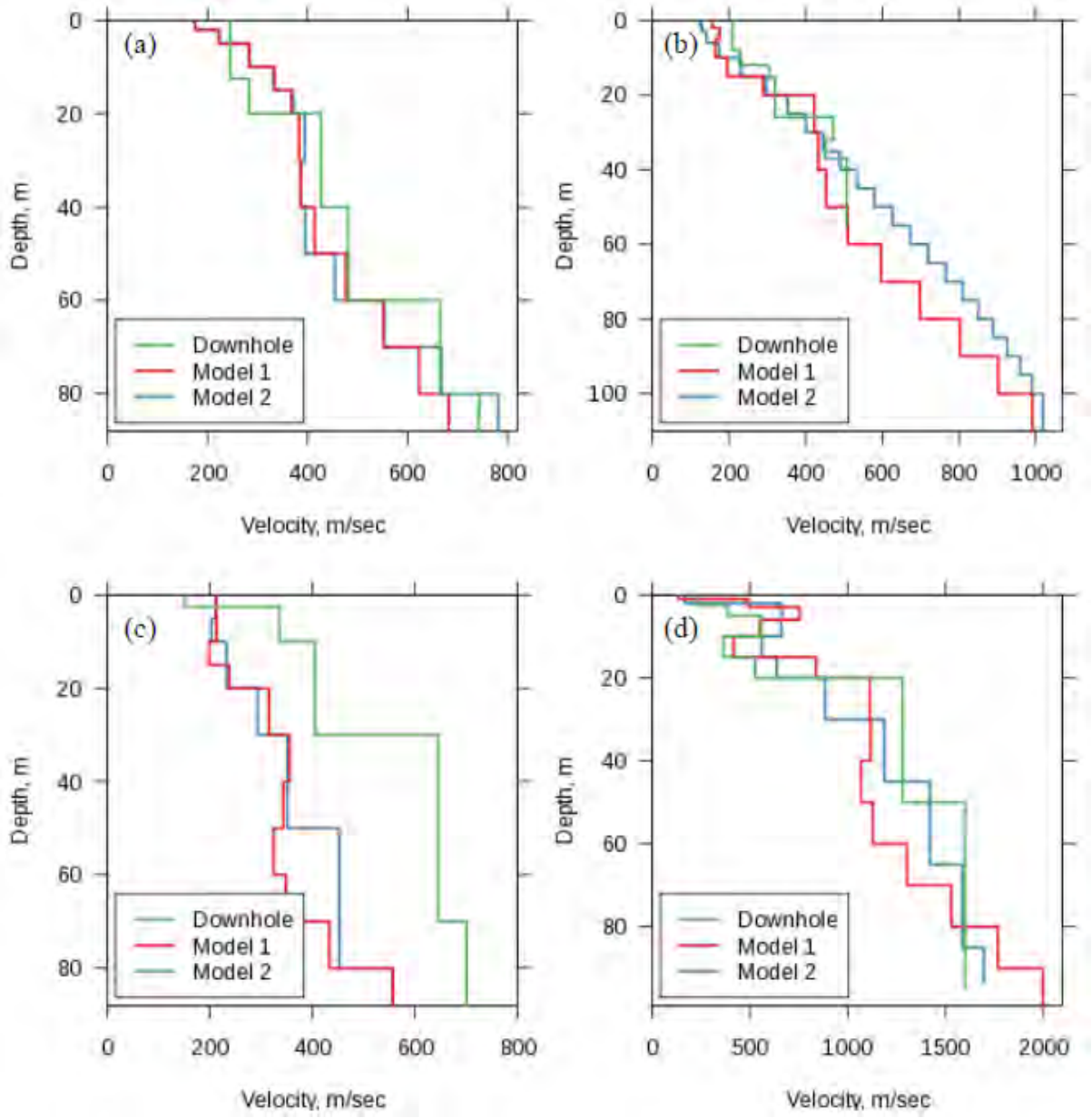


Figure 2.7 (from Thompson et al. 2010)**.** Comparison of V_S profiles derived from SASW and downhole methods from two different USGS open-file reports at four sites in central California (first site code is from Thompson et al. 2010, second site name is from previous USGS downhole studies summarized by Boore [2003a]): (a) 814VIN/Vineyard Canyon, (b) 815CHO/Cockrums Garage, (c) 840RFU/Red Hills, and (d) 841KFU/Jack Canyon. A favorable comparison exists for three of the sites, but for the 840RFU/Red Hills site both SASW profiles differ significantly from the downhole V_S profile.

The SW-AA consortium prioritized site investigations at locations of USGS seismographs, and they present data from their own site investigations and those of others at 83 stations, 76 of which are located in California. Of these 83 stations, 30 sites have geophysical testing performed by SW-AA, 20 sites lack V_S measurements (i.e., only boring logs or site descriptions from field reconnaissance are presented), and 15 sites utilize data from USGS or UCLA research reports (see Section 2.3.4.1 below) or data from LeRoy Crandall & Associates (see Section 2.3.5 below). Of

the remaining 18 sites, 16 were considered duplicates, as V_S testing was not performed at the site, but rather a V_S profile was assigned from a neighboring site (these 16 sites are characterized using 4 off-site profiles). This was possible because these sites are closely spaced in structures in dense urban areas of Los Angeles. There are two similar cases in which two sites share one borehole: at Caltech in Pasadena, CA, and at Portland State University in Portland, OR. Where SW-AA take information from other sources, since this project had access to those same sources, the data was reported as being derived from the original source. As a result, 31 V_S profiles at 30 sites (one site, Terminal Substation, has two measurements) were added to the PDB based on SW-AA reports, 22 of which are in California and the rest are located around the U.S.

For the V_S PDB, site level metadata such as geodetic coordinates (confirmed by cross-checking printed site plans and Google Earth), V_S profiles collected using downhole and cross-hole methods, and stratigraphy from borehole geologic logs were extracted from the SW-AA reports. Where available and the quality of the report allowed, lab test data was digitized as well; however, data of certain formats, such as CPT traces, are difficult to digitize because of their small as-printed size. Because tabulated digital CPT data were not provided in these reports, this data was not included in the database.

2.3.3 California State Data Sources

The initial emphasis in V_S PDB development was focused on California, where the project team had access to a numerous published and otherwise available V_S studies and where the PI had amassed a personal data collection. A systematic review of existing data resources was performed, which included community input facilitated by several public workshops. The process of identifying and accessing useful data was undertaken in coordination with public agencies in the state, namely the California Department of Transportation (Caltrans); the California Department of Water Resources (DWR), and its subsidiary the DWR Division of Safety of Dams (DSOD); and the California Geological Survey (CGS; formerly called the California Division of Mines and Geology, or CDMG). Data collected from these sources are described in this section.

2.3.3.1 California Department of Transportation Bridge Sites

Following the 1989 Loma Prieta and 1994 Northridge earthquakes, Caltrans performed investigations of the seismic response of many highway bridges. As part of these studies, Caltrans engineers performed 294 geophysical measurements at 91 bridge sites throughout the state from 1994 to 2007. This dataset spans many geographic regions and various geological settings, from soft Bay mud clay deposited in shallow waters of the San Francisco Bay, to very stiff volcanic lahar deposits in Butte County. All V_S and V_P data were measured using suspension logging, so resolution and quality are high for the velocity profiles. All Caltrans velocity profile data was obtained in digital form from Bill Owen (2015, *pers. comm.*).

Geologic logs from the boreholes in which the suspension log measurements were collected, including stratigraphic descriptions and lab test results, and were included in the V_S PDB where available. A challenge with this dataset is that the Caltrans logs and site plans are provided on scans of large plan sheets that are difficult to read in some cases. Soil descriptions, SPT N-values, and lab test results from the scanned logs were manually read and entered whenever possible. For 35 sites the boring logs were unable to be matched from the plans to V_S logs. In these cases, the V_S data is provided without geotechnical metadata. There were numerous sites where

the scans of as-built bridge plans and sections provided by Caltrans to UCLA (T. Shantz and B. Owen 2015, *pers. comm.*) were of low resolution and unreadable. With the exception of the "unpaired" sites noted above, these gaps were later filled by information accessed at GeoDOG, the Digital Archive of Geotechnical Data (https://geodog.dot.ca.gov/) maintained by Caltrans. Using this online database, gaps of poor-quality or missing data were filled in.

Geodetic coordinates are provided in the Caltrans documents that appear to use the WGS84 datum; this assessment is based on location checks performed in Google Earth and compared to locations marked on site plans. As a result, the Caltrans coordinates were entered into the PDB without adjustment.

2.3.3.2 Department of Water Resources Levee Sites

The California Department of Water Resources (DWR) provides regulatory oversight for 1330 miles of levees distributed throughout the state, with a heavy concentration in the Sacramento-San Joaquin River Delta region (CA DWR 2009). The stability of these levee systems is key to both protecting urban and non-urban areas from flooding, and to the effective operation of the California State Water Project. From 2006–2015 DWR embarked on the Levee Evaluation Program, which performed geotechnical and seismic evaluations for certain levee reaches. Excerpts of six reports were obtained from DWR staff (A. Balakrishnan 2015, *pers. comm.*) that present data from 29 SCPT investigations along urban levees in Sacramento, San Joaquin, Solano, Sutter, and Yolo counties. For the V_S PDB, SCPT-derived V_S profiles, CPT tip resistance, sleeve friction, and pore water pressure data, as well as site-level metadata, were included. DWR logs included NAD83 geodetic coordinates that were converted to WGS84.

Additional data was acquired in February 2019 from DWR staff (N. Novoa, 2019, *pers. comm.*) for 15 sites around the Delta region, a total of 66 V_S and V_P profiles. This data consists of 46 V_S and 20 V_P profiles derived from SCPT and suspension logging, geotechnical borehole and CPT data. The data were entered into the PDB by Tristan Buckreis.

2.3.3.3 Division of Safety of Dams Sites

In the summer of 2015 a preliminary list of sites that potentially contained V_S profiles at dam sites regulated by DSOD was obtained (Bill Frazier, 2015, *pers. comm.*). Members of the project team then visited the DSOD internal library to scan sections of reports. Reports for 60 dam sites were reviewed, 27 of which had V_S data. An additional 30 were missing V_S data, and three sites did not have data in DSOD files but are part of a proprietary dataset (see Section 2.3.5 below). For sites for which V_S does exist, V_S and V_P data, boring logs and lab testing data for invasive geophysical measurements, and site-level metadata were collected.

A challenge with this dataset was large variability of digital data quality, ranging from photographs of pages of reports as old as 1934, to high-quality scans of more recent reports that including digital or tabulated data files. Moreover, the available information for the 27 sites varied widely, from one site (St. Helena Dam) presenting eight coarse-resolution two-layer seismic refraction profiles that are ambiguous as to whether the velocity data is P- or S-wave, to another site (Perris Dam) having profiles from six SCPTs, two suspension logs, three combined active- and passive-source SWMs, three P- and S-wave refraction profiles. This data has not been incorporated into the PDB at this time. It is planned to systematically reassess this important

dataset and meet again with DSOD officials to ascertain what data is available with greater certainty.

2.3.3.4 California Geological Survey

The CGS reviews geotechnical reports for new construction and seismic retrofit of hospitals and medical centers (along with the Office of Statewide Health Planning and Development) and schools and community colleges (along with the Division of the State Architect). CGS staff (C. Wills 2017, *pers. comm.*) provided scans of reports for 38 medical facilities and 59 schools. These reports vary in age (1983-2017) and in the quality and types of data included. Each site in the V_S PDB includes, where available, V_S and V_P profiles, stratigraphic logs and lab test data for the boreholes where invasive geophysical measurements were performed, and CPT data where SCPT investigations were performed.

It is noted here that CGS staff provided data and reports (C. Wills 2015, *pers. comm.*) for an additional 141 V_S profile locations, gathered largely from documents filed with public agencies, that are part of a proprietary database maintained by Pacific Engineering and Analysis (PEA). This data will be discussed in Section 2.3.5 below.

The California Strong Motion Instrumentation Program, which is part of CGS, has funded several studies to characterize ground motion recording sites in California. Results from surface wave tests (MASW) at 41 sites are presented in reports by C.H.J. Inc (2008), Geovision (2016 and 2018), and Petralogix (2017). These data were recently obtained and are not yet incorporated into the database.

2.3.4 University Research and Other Reports

Projects performed by university research units or private companies under contract with public agencies or research organizations are important data sources. This section describes important datasets in which seismic velocity investigations were performed at strong motion stations, downhole arrays and other field test sites, sites on selected University of California campuses, and various other sites investigated in research studies.

2.3.4.1 University of California Earthquake Engineering Research Reports

After the 1971 Sylmar earthquake produced useable records from 208 strong motion recording stations located in structural basements or in the free field around the southern California area (Duke et al. 1971), C.M. Duke led an effort to characterize the seismic site conditions at 111 of these stations, results of which are presented in three data reports (Duke et al. 1971, Eguchi et al. 1976, Campbell et al. 1979). These reports followed up on an earlier study by Duke and Leeds (1962) that had characterized 63 sites in southern California. Excluding 7 sites with repeated measurements, this data set provides V_S profiles for 167 sites geographically spanning from the Imperial Valley to Kern County. The V_S profiles were measured using S-wave seismic refraction, downhole, and crosshole methods. For the V_S PDB, V_S profiles, V_P profiles where available (but excluding those derived from V_P/V_S ratios), stratigraphic information from boreholes used for invasive methods, and site-level metadata were included. Data from 81 of these sites are currently included in the database (i.e., the sites from Eguchi et al. 1976 and Campbell et al. 1979).

Some of the V_S profiles in the reports by Duke and others were critiqued by SW-AA (1980) for being improperly attributed to multiple strong motion recording stations. These critiques were made generally on the basis of the measurements being located too far away from the station, being located on different surficial geology, or penetrating into a different geological formation at depth. Another practical problem with a subset of the site data presented in Duke et al. (1971) is that the "Subsurface Models" (labeled "A" through "I") were largely based on correlations to geology at greater depths (usually below the upper 2–5 layers) rather than on in situ V_S measurements. Therefore, velocities from deeper layers were excluded from the V_S PDB.

Several individual research studies by University of California (UC) investigators have produced V_S profiles in California, including:

- Four sites in the east San Francisco Bay Area that experienced liquefaction during the 1989 Loma Prieta earthquake, in the cities of Richmond, Oakland, and Alameda (Mitchell et al. 1994). Site investigation methods included SCPT, crosshole, and SWMs. Additional testing including borings with SPT and flat plate dilatometer.
- The Santa Monica City Hall ground motion station (Chang 1996). Data from the site includes a boring log with soil layer descriptions and SPT N-values, lab test results, and V_S and V_P data from suspension logs.
- A joint UCLA-Caltrans test site for foundation testing in Hawthorn, CA (Wallace et al. 2001). Data from the site includes three boring logs with soil layer descriptions and SPT N-values, lab test results, and V_S and V_P data from three suspension logs. Five SCPT profiles are also available.
- The Pleasant Valley Pumping Plant site near Coalinga, CA (Stewart and Sholtis 2005). Data from the site is provided in two locations, above and below a cut slope. Above the cut slope, a strong motion site is located in the electrical switchyard. This location includes a boring log with soil layer descriptions and SPT N-values, and V_S and V_P data from suspension logs. Below the base of the cut slope near the pumping plant structure, CPT soundings and SCPT measurements were performed.
- Two sites with compacted fills that were impacted by the 1994 Northridge earthquake (Stewart et al. 2004). Data from a school site in Santa Clarita includes a boring log with soil layer descriptions and SPT N-values, lab test results, V_S and V_P data from suspension logs, and SCPT profiles. Data from a post office near the Santa Clara River in Santa Clarita includes a boring log with downhole V_S measurements, SPT N-values, and lab data. SCPT logs at the site are also available.
- The site of a landslide near the base of the Santa Monica Mountains during the 1994 Northridge earthquake (Pradel et al. 2005). Data from the site includes a boring log with soil layer descriptions, lab test results, and V_S and V_P data from suspension logs.
- A site on Sherman Island, CA that was used for field testing of model levees (Reinert et al. 2014). Data from the site includes a boring log with soil layer descriptions, lab test results and V_S data from SCPT.
- As p art of the NSF-funded Center for Embedded Networked Sensing at UCLA, a deep (100 m) borehole was drilled adjacent to the 17-story instrumented Factor Building on the UCLA campus (Steidl et al. 2004). The objective of this deep borehole was to install

an accelerometer at depth to study ground response of the soil column beneath the building. For the V_S PDB, V_S and V_P profiles from suspension logging were collected, as presented in a consulting report (GEOVision 2004). Geotechnical data is not available.

2.3.4.2 NEES@UCSB Sites

The George E. Brown, Jr. Network for Earthquake Engineering and Simulation (NEES) was a consortium of 15 research universities that operated from 2004-2014, each with different capabilities for experimental and numerical earthquake engineering applications. The UC Santa Barbara NEES group constructed and maintains two instrumented field test sites—the Wildlife Liquefaction array site and the Garner Valley Downhole Array site. Five other sites, including the Borrego Valley Field Site in San Diego County, CA; the Delaney Park Array in Anchorage, Alaska; the Hollister Earthquake Observatory near Hollister, CA; the San Jose US-101/I-280 Interchange site; and the Seattle Liquefaction Array, are maintained by UCSB staff. Data for these 7 sites from the NEES@UCSB website (http://nees.ucsb.edu/) was incorporated into the V_S PDB. The sites were characterized using an array of geophysical methods, including downhole, suspension logging, and active and passive SWMs. Information to fill gaps in site metadata was provided by J. Steidl (2018, *pers. comm.*).

2.3.4.3 University of California—Campus Earthquake Program

Starting in 1996, seven campuses of the University of California partnered with the Lawrence Livermore National Laboratory to perform site-specific analyses of seismic hazard for three campuses: UC Riverside, UC San Diego, and UC Santa Barbara. These included an array of geotechnical and geophysical field investigations and dynamic laboratory testing, including drilling deep boreholes for installing sensors at depth to create vertical arrays, and performing probabilistic seismic hazard analyses for the three campuses. For the V_S PDB, V_S and V_P profiles were included from suspension logs for all three campuses (including three boreholes for UCR), and stratigraphic logs and laboratory index test data for the UCSB borehole and one of the UCR boreholes (UCR-5, near the Rivera Library). The overall project is described in Heuze et al. (2004), and the data is presented in reports for Phases I and II of field investigations for each campus (UCR: Park et al. 1999b, Archuleta et al. 2000a; UCSD: Minster et al. 1999, Day et al. 2002; UCSB: Archuleta et al. 1997 and 2000b). Digital data for suspension logs were provided by GEOVision employees (R. Nigbor and R. Steller, 2018, *pers. comm.*) and UCSB faculty (J. Steidl 2015, *pers. comm.*). This data results in eight velocity profiles across four sites within the PDB.

2.3.4.4 Resolution of Site Response Issues from the Northridge Earthquake (ROSRINE)

The 1994 Northridge earthquake produced the largest strong motion data set for its time, with recordings from 263 stations (Chang et al. 1996). The ROSRINE project was organized by a consortium of the University of Southern California and the PEER Center, which sought to provide high-quality subsurface data at strong motion stations to improve GMMs (Nigbor et al. 2001). ROSRINE pioneered the use of relational databases integrated with GIS software on an online platform to organize, manage, and disseminate data in a timely manner. ROSRINE was organized into multiple phases of data collection, including field investigations and laboratory testing, each having a report and a data release.

While the original project website and data repository has been discontinued, the project data, including digital data and reports, was obtained from ROSRINE project PI Robert Nigbor (2016, *pers. comm.*). For the V_S PDB, the project team received data for 53 sites with measured V_S and V_P profiles. Of these, 33 are based solely on suspension logs and stratigraphic logs from borings, while 20 overlapped with sites investigated using downhole methods in USGS open-file reports (Gibbs et al. 1999 and 2000). Fourteen of the overlapping sites were characterized both with downhole measurements and suspension logs. For such cases, data and metadata from both sources were included. Later phases of the ROSRINE project (Phase 5b) characterized sites outside of the Los Angeles area, including three sites in the Imperial Valley, two in the Mojave Desert, and three in Northern California.

2.3.4.5 United States/Japan Loma Prieta Earthquake (UJLPE) Project

The Electric Power Research Institute (EPRI), the consortium of California Universities for Research in Earthquake Engineering (CUREe), and the Building Contractors Society of Japan undertook the UJLPE project to characterize sites that recorded the 1989 Loma Prieta earthquake (Thiel & Schneider 1993). Downhole (22 V_S profiles) and suspension log (27 profiles) data were collected at 33 sites (18 sites with both measurement types), which are compiled in Thiel & Schneider (1993) from multiple individual reports (Redpath 1991, Gibbs et al. 1992, AA 1993). Boring logs with soil descriptions and site plans are also available for all sites (Fumal 1991, Gibbs et al. 1992, Powers & Fumal 1993, WCC 1993).

A separate study was undertaken for six additional sites that included V_S profiles from suspension logging. Two of these that recorded the Loma Prieta earthquake, and four are in the Imperial Valley. V_S profiles from suspension logs and boring logs with soil unit descriptions are available for these sites. The data was presented in a report to the Kajima Corporation of Japan, which was also part of the UJLPE project (GEOVision 2000b).

2.3.4.6 University of University of Kentucky, Lexington/Kentucky Geological Survey—Seismic velocities for vicinity of New Madrid seismic zone

As part of a joint effort between the University of Kentucky, Lexington and the Kentucky Geological Survey, shear wave velocity profile data were collected in the broader New Madrid seismic zone and its vicinity. Li et al. (2013) published a relational database that stores all of the data collected as part of this multi-year effort. The database contains a total of 643 sites. For 227 of them, shear wave velocity profiles are available. All of the profiles have been characterized by means of *S*-wave seismic refraction surveys. This data set results in 227 profiles in the PDB.

2.3.5 Industry Data Sources

Throughout Section 2.3, "public" datasets have been described. While these range from easily accessed data downloaded from a website, to data sets from public agencies that were not readily accessible, all data were considered publicly accessible. Serious consideration was given to the treatment of data derived from private entities for non-research purposes at the start of the V_S PDB project. A concern expressed by USGS representatives was that they may be unable to participate in the project if private data linked to commercial interests were to be part of the V_S PDB.

It is important to explain the interpretation made here of the terms "proprietary" and "public" in the context of data that is collected by a private entity (generally a geotechnical

engineering or geophysics consulting firm). In many cases, data collected by these firms is contained in reports that are submitted for review to public agencies such as municipal or county building departments or state agencies (for hospitals or public schools). Upon submittal, these reports become public record. It is acknowledged that the accessibility of these documents is variable, with some entities maintaining libraries of reports and others not. Regardless, if the data is public record, it was treated as such whether the transfer of information comes from the public agency or the firm that prepared the report.

The V_S PDB project team sought to formulate agreements with private firms who maintain their own libraries of data collected on prior projects. In discussions with these agencies, the interest in public data was described, per the definition above. Some have been receptive and expressed interest in contributing to the project, as elaborated below.

2.3.5.1 LeRoy Crandall/Wood plc. Dataset

In southern California, a large internal database of V_S profiles has been developed over approximately five decades by LeRoy Crandall & Associates (L/C), a legacy consulting firm located in Los Angeles. L/C and its later incarnations (Law/Crandall & Associates, Mactec, AMEC, AMEC Foster-Wheeler, and now Wood plc.) collected and maintained V_S profile data, based mainly on downhole testing in boreholes.

The L/C internal database consists of reports scanned into digital format (e.g. TIFF and PDF files). Their database includes all prior projects, only some of which incorporated V_S profiling into the site characterization. From this dataset L/C staff (M. Hudson & M. Lew 2018, *pers. comm.*) identified project sites that contain in situ measured V_S profiles and for which the data could be released to the PDB project based on the "public" definition given above. The data date as far back as 1972, when L/C began to measure V_S profiles using downhole methods for projects. The geophysical measurements were made by L/C staff rather than being subcontracted to other firms (M. Lew 2018, *pers. comm.*). The L/C data set contains measured V_S and V_P profiles at more than 350 project sites throughout Southern California, including Imperial, Kern, Los Angeles, Orange, Riverside, Santa Barbara, San Bernardino, San Diego, and Ventura Counties. There are also seven sites in the San Francisco Bay area, and one site each in Roseville, CA, Lemont, IL, Reno, NV, and Cincinnati, OH.

The information provided for this dataset consisted of report excerpts that are pertinent to the V_S profile, including the travel time plots, V_S and V_P data (in tabulated or plotted format), and the geotechnical boring log only for the borehole in which the downhole measurement was performed. The boring logs generally contain soil unit descriptions, SPT blow counts, and laboratory test results, as shown in Figure 2.8. As the data consists of scanned files, it required manual entry into the digital PDB format. The project name is not released, but the L/C original project number (generally a 5-digit number, where the first two digits represent the year of the project, and the last three are the project number within that year) is included as site-level metadata.

This approach provides a potential blueprint for other organizations to share V_S profile data without losing competitive advantage in their practice. Because the data is public in the sense that it had been submitted in the past to a municipal building department, it also satisfies USGS Fundamental Science Practices (see Section 2.3.1 above).

2.3.5.2 Pacific Engineering & Analysis Dataset

Pacific Engineering & Analysis (PEA), a consulting firm that specializes in seismic ground motion hazard analysis, maintains a data set of V_S profiles. This database is proprietary and is generally not available outside of PEA. One exception was that the NGA-West2 project site database (Seyhan et al. 2014) contains V_{S30} values from measured V_S profiles where those profiles are near (within about 300 m) a strong motion station. NGA-West2 investigators did not have access to the V_S profiles (only V_{S30} was provided).

The V_S PDB project does not have access to the full PEA data set. However, as part of a prior collaboration between PEA and CGS, the PEA database was shared with Chris Wills of CGS for work investigating proxy-based models for V_{S30} estimation (Wills and Silva, 1998). A portion of the data provided to CGS is public, as briefly described in Section 2.3.3.4 above (C. Wills 2015, *pers. comm.*). Accordingly, this data was included in the V_S PDB. The data set encompasses 141 V_S profiles as follows:

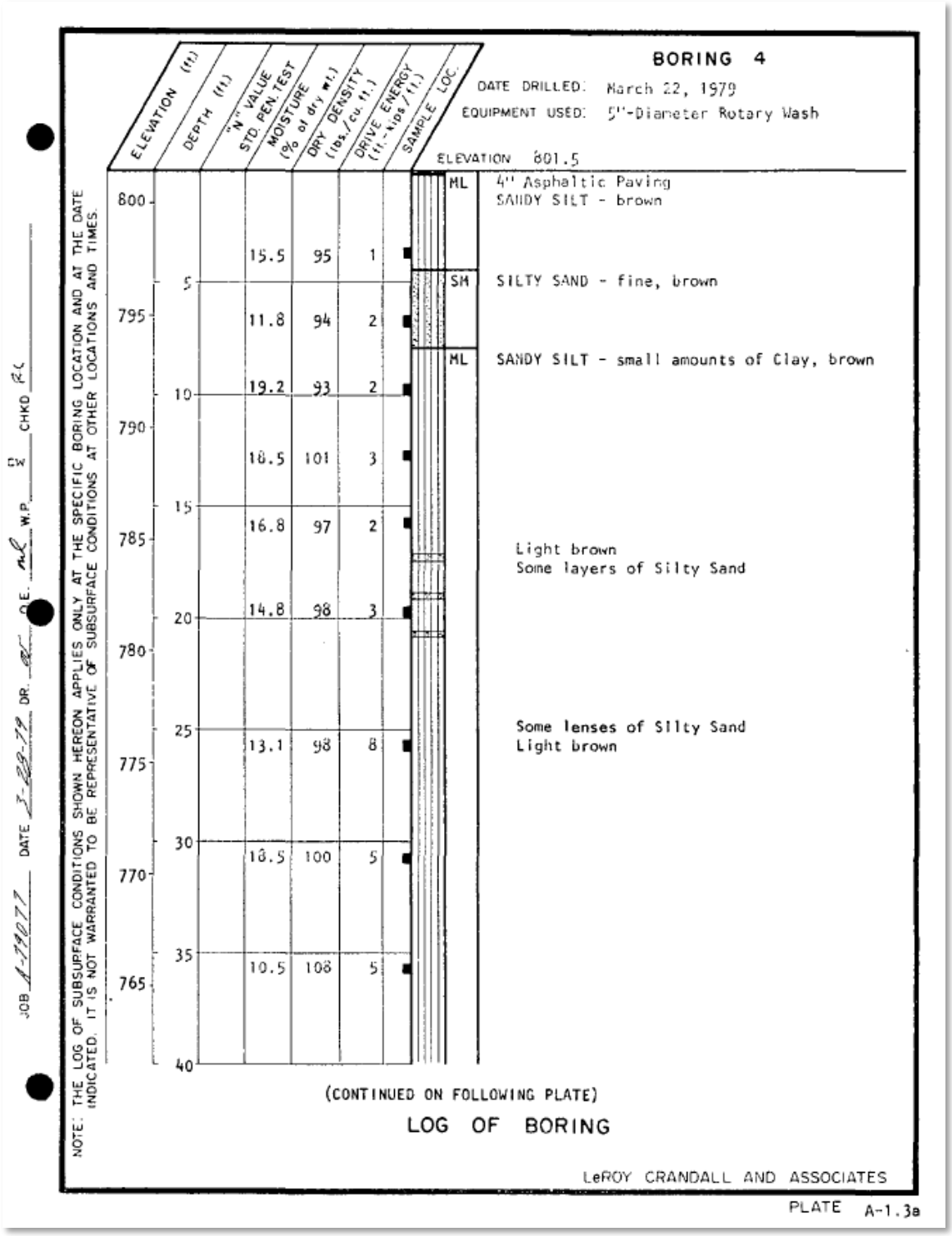


Figure 2.8 Example excerpt of boring log from L/C Project No. 79077.

- 21 are located at California hospital sites reviewed by CGS
- 68 are located at dams and reservoirs under the jurisdiction of the U.S. Army Corps of Engineers, CA DWR, or other agencies
- 45 are located at bridge sites that are not included in the Caltrans dataset described previously
- 4 are NEES@UCSB sites previously described in Section 2.3.4.2.
- The final 3 sites are Santa Monica City Hall, an elementary school in San Francisco, and a park in Orange County.

Of these 141 sites, only 30 overlap with those in previously described datasets (3 CGS hospitals; 22 Caltrans bridge sites; four NEES@UCSB sites, and SMCH). These data were presented in a uniform format, with digital lists of both V_S profile locations and V_S profiles, making their integration into the V_S PDB straightforward. Particular challenges included filling in gaps in metadata by cross-checking with files and notes provided by CGS staff that compile short excerpts of original data reports (C. Wills, 2017, pers. comm.) and cross-checking for duplicate profiles in other datasets.

2.3.5.3 Utility Companies

Pacific Gas and Electric Company (PG&E) measured V_S profiles for the power block location at the currently operating Diablo Canyon Nuclear Power Plant and at the site of the now-inactive Humboldt Bay Nuclear Power Plant. The Diablo Canyon data is presented in PG&E (2010) and the Humboldt Bay data is presented in Stewart and Stewart (1997). The Diablo Canyon data consists of suspension logs in bedrock materials. The Humboldt Bay data includes downhole V_S data and a boring log.

2.3.5.4 Miscellaneous Studies and Reports

A report by Woodward-Lundgren Associates for a study commissioned by the National Oceanic and Atmospheric Administration (Hansen et al. 1973) compiled V_S and geotechnical information at 17 sites using a variety of geophysical measurements; these data are incomplete, and also overlap in part with sites from Duke & Leeds (1962) and other original source documents. As such, the compilation in this report are not included in the database (although the data is provided, with attribution to source documents).

2.4 DATABASE STRUCTURE

A relational database (RDB) was adopted as the means by which to organize and archive the information in the PDB. This differs from the classical “flat file” structure of data collections that have been previously termed “databases” in geotechnical and earthquake engineering. Brandenberg et al. (2018) describe the benefits of using an RDB relative to spreadsheet tables for application to the Next Generation Liquefaction (NGL) project (Stewart et al. 2016). The portion of the NGL database structure (known as a “schema”) that relates to the geophysical and geotechnical data was adapted with some modification for use in this project. This section

describes the methods used to prepare data for processing and integration into the V_S PDB, the RDB schema developed for this project, procedures for organizing and transferring data into the online RDB, reasoning for use of methodologies presented here versus other existing data organization methods and file formats, and other elements of the back-end of the database that help it to function and be backed up.

2.4.1 Overview of Data Preparation and Processing Methods

Previously, Section 2.3 described individual data sets and the original formats in which data were obtained. Here the aim is to outline the overall procedure for integrating the data into the V_S PDB. Generally, data were acquired in formats of either a paper report, electronic scan in PDF or image file formats, or in ASCII/Excel digitized data files. To utilize the available data efficiently, it is necessary to collect and store it in a unified structured format. Major advantages of placing the data in a hierarchal structured format include (i) removal of the need for data normalization (i.e. formatting data in a tabular structure), (ii) dynamic data updating and expansion without corruption of the data structure, and (iii) rapid data querying. As such, substantial resources were utilized to digitize analog data and to organize it in a uniform manner. To facilitate efficient data entry and organization, a graphical user interface (GUI) program called "Unify" was utilized (Kottke 2012), which outputs a JavaScript Object Notation (JSON) file that both contains all data for a site or data source of interest and mirrors the V_S PDB schema (described in Section 2.4.3 below). Once the data is in a machine-readable format, codes were written that read and automatically parse the JSON files into tables in the Structured Query Language (SQL) database format, which comprises the "database". The MySQL RDB management system (RDBMS) was then used to manage and interact with the database. These steps are described in greater detail the following sections and in Brandenberg et al. (2018).

2.4.2 Description of V_S Profile Database Schema

The V_S PDB contains a diverse array of data and metadata pertinent to sites with V_S measurements. The current V_S PDB schema is the result of extensive discussion among the project team members, with community input via two public workshops. The schema describes the tables, fields, and relationships among tables in the RDB. The V_S PDB schema is comprised of 26 distinct tables. A list of the table names is provided in Table 2.2, which are grouped into three categories: general information, geophysical data, and geotechnical data. The table names in Table 2.2 have meanings and are described subsequently in this section.

Table 2.2 List of table names in the V_S PDB schema.

Group Type	Table Name	No. Fields
General	user	12
	authors	3
	site	18
	file	5
Geophysical	velocityProfileMeta	14
	velocityProfileArray	6
	dispersionCurveMeta	13
	dispersionCurveArray	6
	spectralRatioMeta	9
	spectralRatioArray	4
	travelTimeMeta	15
	travelTimeArray	4
Geotechnical	boringMeta	14
	boringArray	4
	standardPenetrationTestMeta	9
	standardPenetrationTestArray	8
	stratigraphySetMeta	4
	stratigraphySetArray	7
	labTest	9
	indexProperty	9
	grainSizeDistributionMeta	5
	grainSizeDistributionArray	4
	nonlinearTestMeta	19
	nonlinearTestArray	4
	conePenetrationTestMeta	16
	conePenetrationTestArray	6

Relationships among entries in tables are facilitated by primary and foreign keys. A *primary* key is assigned to an entity in the primary table where it is defined using an identifying integer number. For example, the metadata for the first two sites in the database would be listed in the `site` table (described below) as having primary keys `'site_ID' = 1` and 2, respectively, which are unique identifiers for those sites. The same item that has a primary key in one table may be referenced in a different table, in which case the same identifying number is provided as a *foreign* key. For example, if the first two sites each have two V_S profiles, `'site_ID' = 1` will be utilized as a foreign key for the first two entries in the `VelocityProfileMeta` table (described below), and `'site_ID' = 2` will be assigned to the third and fourth entries in that table.

The database schema is extensible, meaning that additional tables and fields may be added to the RDB structure over time as warranted. This feature is beneficial in cases where active or future research will encourage including additional information for specific geophysical methods, e.g., sensor spacing in SWMs (see Zhang et al. 2004); while this field is not currently part of the database schema, it is relatively easy to add if the community comes to an agreement necessitating its inclusion.

Figure 2.9 defines the content of tables containing general information in the RDB. The primary nexus in the RDB that organizes all related data is the `site` table, with `'site_ID'` serving as the primary key. A *site* in the V_S PDB is defined as a broad area of interest which may contain one or more pieces of information within an area; it may be a single point (for sites with a single profile) or an area in which several profiles are available. The geodetic coordinates of a site

are used only to plot the site on a map; these coordinates would be coincident with the physical location of a single geophysical or geotechnical profile if that is all that is available for a site, but is typically at a central location for sites with multiple profiles. The `site` table also contains high-level geographic, geologic, and geomorphic information, which are assigned using GIS and described in Section 2.2 above. It is noted here that in the V_S PDB, a *location* differs from a site in that it refers specifically to a piece of geophysical or geotechnical data.

The `user` and `authors` tables contain information about project members who uploaded data to the database and those who have created accounts to use the V_S PDB website. Additional files of interest, such as images or reports, may be uploaded to the database, and are stored as relative file path names in the `file` table within the database server.

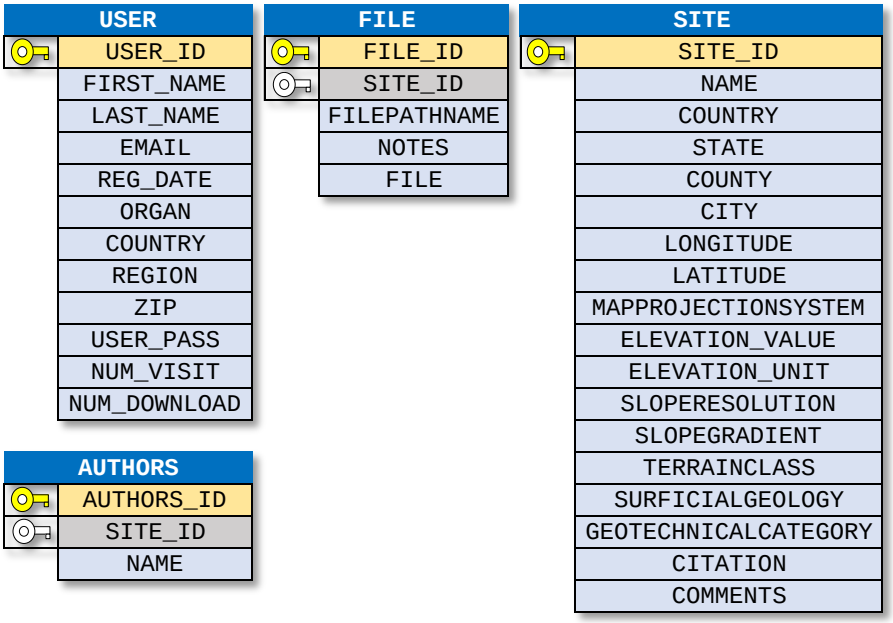


Figure 2.9 RDB tables containing general data.

Figure 2.10 and Figure 2.11 define the content of tables of geophysical and geotechnical information in the V_S PDB, respectively. Geophysical data includes velocity profiles, dispersion curves for SWMs, and travel time plots for SCPT measurements. HVSR data are planned to be included with a separate schema developed by Gospe et al. (2020) to be added to the V_S PDB. For each of these four, two tables are included: an `'X-Meta'` table, which includes the high-level metadata fields for each location, such as individual data location names, geodetic coordinates, geophysical method, and the units in which the data are presented; and an `'X-Array'` table, which actually stores the tabular data of interest, such as V_S versus depth, with individual primary keys of the format `'X-Meta_ID'` and `'X-Array_ID'`, and foreign keys for the `-Array` tables to link to the `-Meta` tables, and for the `-Meta` tables to link back to the `Site` table via the `'site_ID'`. For the `VelocityProfileMeta` table, a flag is included to indicate if the data for the particular location is a V_S or V_P profile, and for the `VelocityProfileArray` table, depth data can be provided in either stair-step format (e.g., using both depths to the top and bottom of layers) or as discrete depth points (one depth for one velocity, which is typical for suspension log data). The uncertainty in the velocity measurement can also be included.

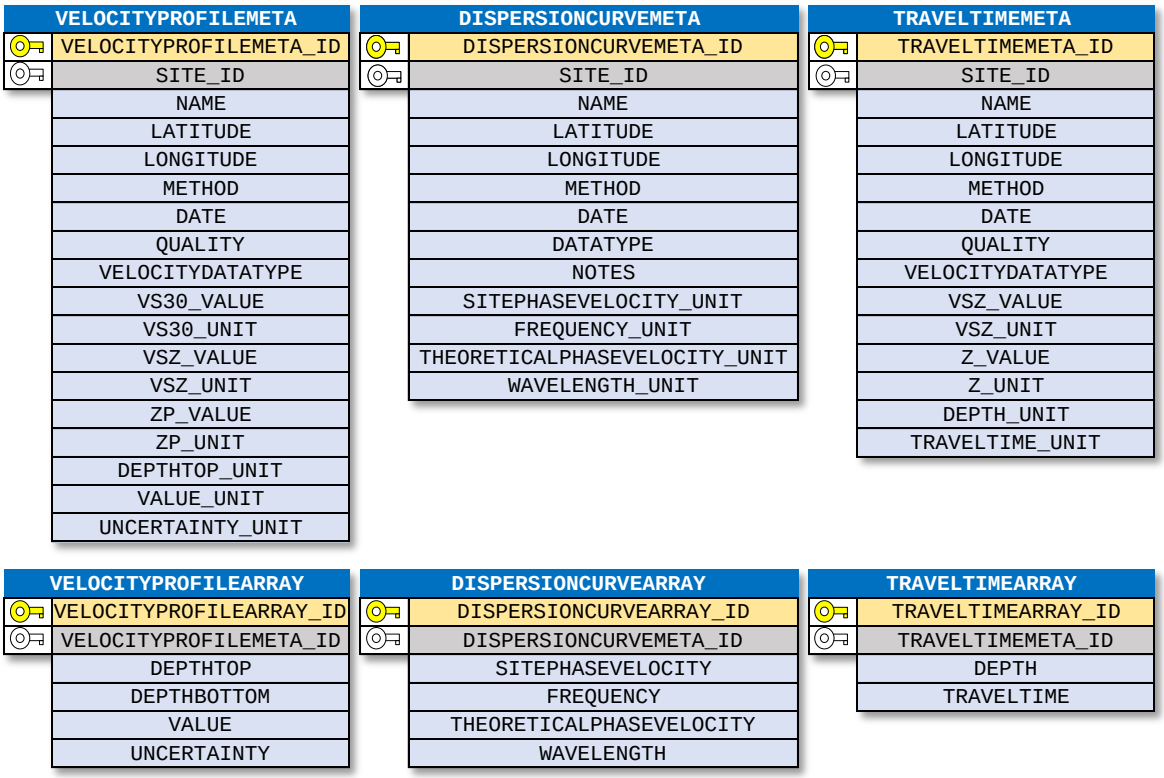


Figure 2.10 RDB tables containing geophysical data.

Geotechnical data are multi-tiered in the database hierarchy and are organized similarly to geophysical data with regard to including `-Meta` and `-Array` tables. The highest hierarchical level includes tables for geotechnical borehole and CPT data, which both link back to the `Site` table by using `'site_ID'` as the foreign key. Tables linking to the `Borings` table include SPT, stratigraphy, and laboratory test data; the latter is further subdivided into tables containing data for soil index property tests (e.g., density, water content, and Atterberg limits), grain size distribution data, and nonlinear test data (e.g., modulus reduction and damping curves). Each of the aforementioned geotechnical data contain `-Meta` and `-Array` tables except for the `'labTest'` and `'indexProperty'` tables. While the majority of sites that are and will be input into the V_S PDB will lack some, if not all geotechnical information, these elements were included in the schema because it is relevant for geotechnical applications and the information is available from some data sets (e.g., ROSRINE).

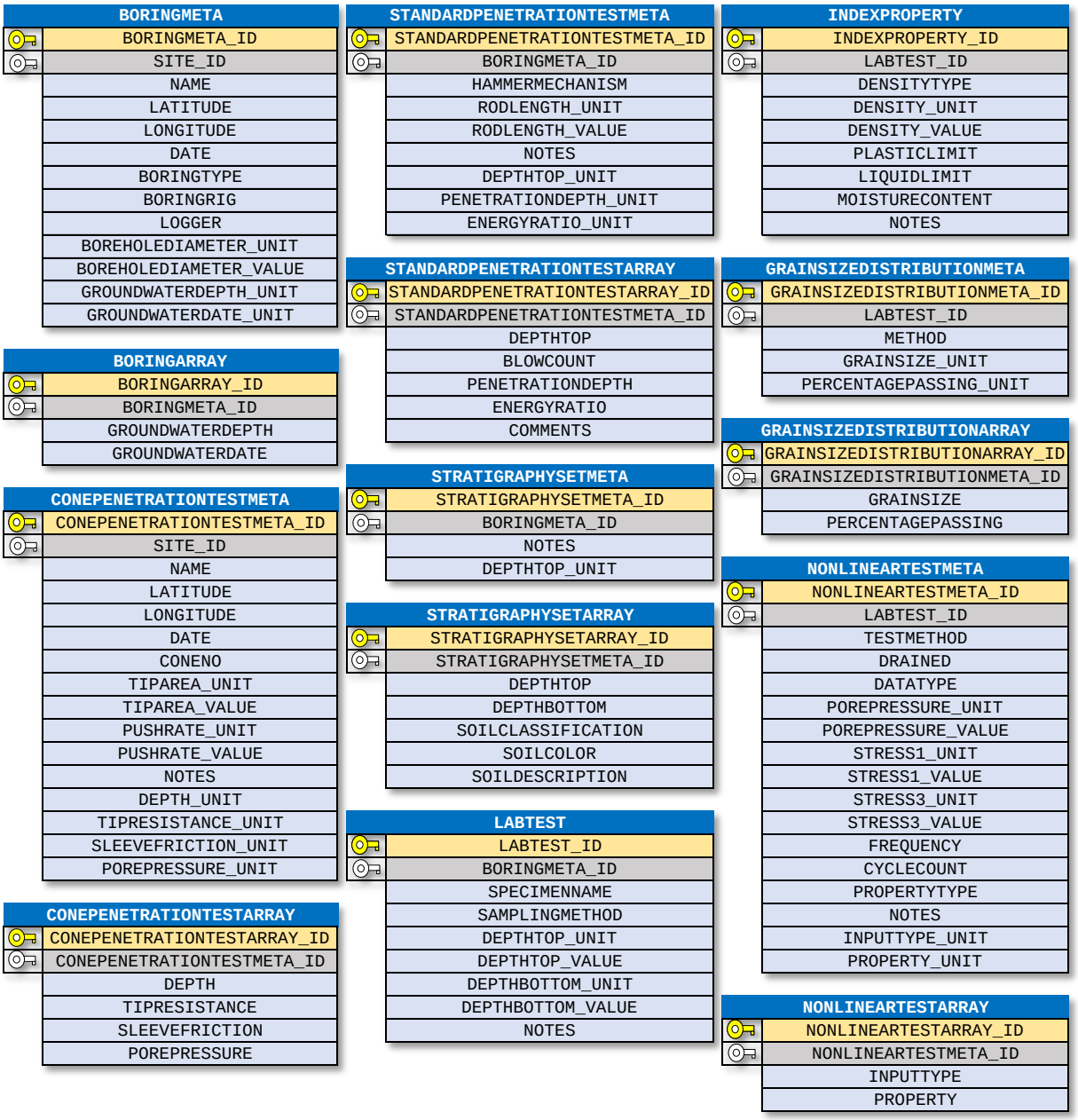


Figure 2.11 RDB tables containing geotechnical data.

2.4.3 Data Organization and Transfer to Relational Database

The RDB schema evolved over time, being influenced by the characteristics of different data sets as their information was incorporated. In parallel with this process, the Unify GUI was updated over time to align with the V_S PDB schema. Figure 2.12 shows screenshots of the current version of Unify GUI (version 1.11) for both an empty dummy site (with no data) and a populated example site. Unify was originally developed by the PEER-NGA-East Geotechnical Working Group for the purpose of storing V_S profile data into the JSON format. This compilation was originally prepared to support definition of reference-rock site conditions for central and eastern North America (Hashash et al. 2014) and development of preliminary proxy-based V_{S30} prediction models (Kottke

et al. 2012). JSON files have the attribute of both being human-readable ("self-describing") and easy for machines to parse and generate.

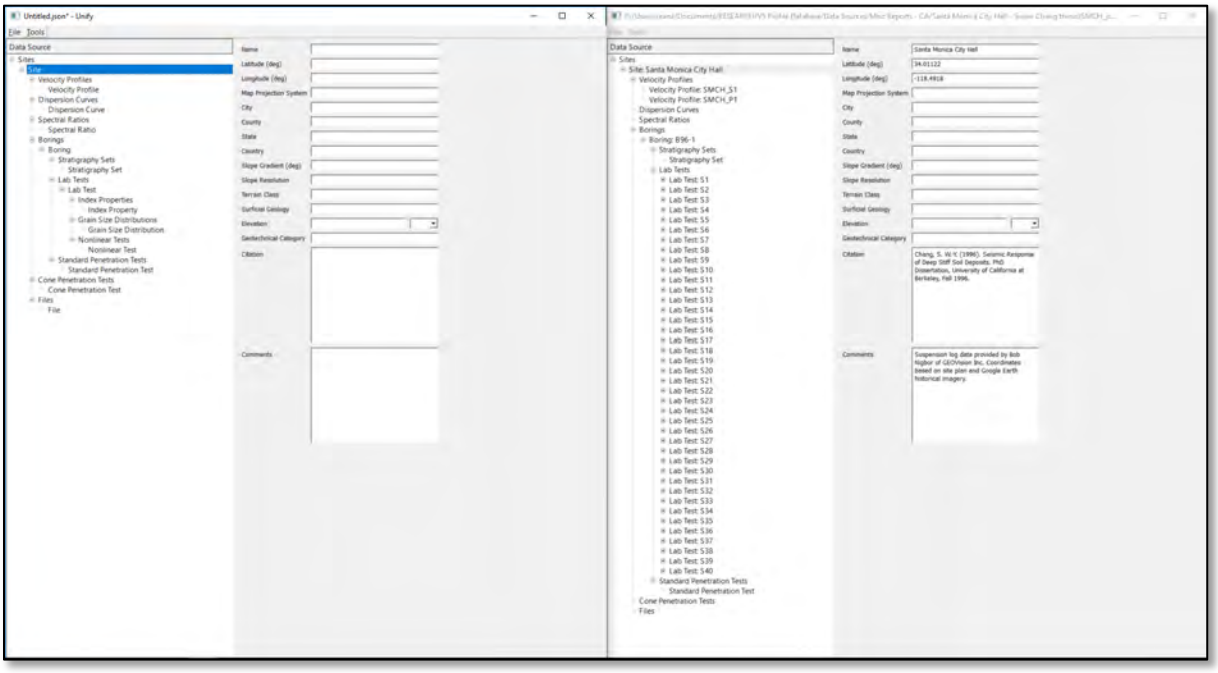


Figure 2.12 Screenshots of Unify GUI with example dummy unpopulated site (left) and complete example populated with data from Santa Monica City Hall site (right).

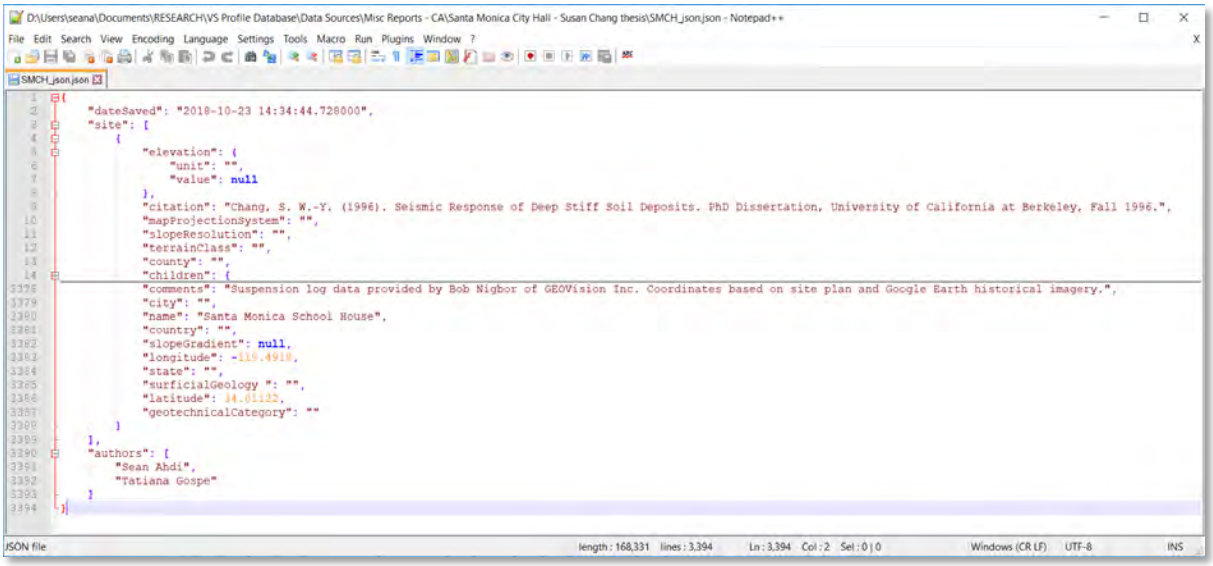


Figure 2.13 Screenshot of hierarchical data storage in JSON file format for Santa Monica City Hall example site (opened in Notepad++ text editor).

JSON files were found to be effective for storing the variously-formatted geophysical and geotechnical data for each site, and the metadata associated with all levels of the RDB schema. JSON files are lightweight ASCII (text) files that are computer language-independent but use

conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and others (Sadiq et al. 2018). While the JSON file structure is comparable to that of XML, JSON has added benefits of being more concise by not requiring end tags to close data groups and the ability to include array-based data. Figure 2.13 shows an example populated JSON file that is the output from the Unify screenshot in Figure 2.12. JSON files can store multiple location-level pieces of geophysical and/or geotechnical data and their associated metadata for a single site within a single file.

In the generation of JSON files, sites were generally grouped by the original project report or individual data sources. This practice was efficient for data entry, particularly because a given report generally contains uniform data formats and representative data types. It was considered to have one JSON file per site, but this would result in an inordinately larger number of files to work with. Organizing JSON files by data source is enabled by the extensibility of JSON files, and in some cases up to one hundred sites are included in a single JSON file. To date 1270 JSON files have been generated, representing 1807 sites in the V_S PDB.

The JSON files are parsed and data are rearranged into SQL formatted tables according to the schema described in the previous section. The primary keys are assigned in all RDB tables as automatically increasing integer values, and foreign keys are assigned accordingly. The database can be regenerated anew (i.e. "refreshed") with new keys if desired.

2.4.4 Other Data Organization Methods and Schemas

It is noted here that existing electronic data formats were considered for this project but for various reasons it was decided to not pursue their implementation for use in the V_S PDB. The Association of Geotechnical & Geoenvironmental Specialists (AGS) of the United Kingdom developed a text-file data exchange format in 1991 to transfer data seamlessly between organizations in the site investigation industry (AGS 2018). The main feature of their schema was to provide as consistently as possible four-letter table names and field codes. The NGL project RDB schema adheres to a similar pattern but using field codes more appropriate for their database. A problem with the AGS methodology is that the table names and field codes are not readily self-explanatory, which was avoided in the V_S PDB by foregoing brevity for clarity, as described in Section 2.4.2. Also, the AGS format did not contain tables in its schema for geophysical data, which was necessitated additional development for NGL and V_S PDB.

Other recent or concurrent projects to the V_S PDB have developed database schemas and data transfer methods. These projects were organized by the California Department of Transportation and the Swiss Seismological Service (SED) at ETH Zürich.

Caltrans developed a Geotechnical Virtual Data Center (GVDC) in the early 2000s which was used internally to manage paper reports and design documents for bridges (Shantz et al. 2015). The GVDC eventually was integrated with data formats from COSMOS, AGS, and the Federal Highway Administration (FHWA) to develop a standard under the working name DIGGS: Data Interchange for Geotechnical and Geoenvironmental Specialists. Caltrans passed DIGGS ownership to the Geo-Institute of ASCE in 2013, and are implementing a new version, DIGGS2.0 (Shantz et al. 2015, DIGGSML 2018). For the V_S PDB, DIGGS markup language (ML) formats were not chosen for implementation, as they merely constituted a data transfer protocol, rather than a database schema. The GVDC, which uses the DIGGS standard to unify data formats from

different sources, was originally designed as a data *broker*, rather than a data *repository* (Shantz et al. 2015), which is a primary goal of the V_S PDB.

The SED's QuakeML data interchange format is a markup language that is flexible, extensible and modular (similar to XML, see W3C 2018), and developed to organize and consolidate existing data formats for statistical seismology applications (Schorlemmer et al. 2011). QuakeML originally was designed to organize and store strong motion data for the SED's seismic network but was recently expanded in QuakeML2.0 to include site characterization information (Kästli & Euchner 2018). Until recently, the V_S PDB and QuakeML were developed in parallel without information exchange. The V_S PDB project team recently learned of the QuakeML effort at the COSMOS Workshop on International Guidelines for Applying Noninvasive Geophysical Techniques to Characterize Seismic Site Conditions at the 36th General Assembly of the European Seismological Commission in 2018. The V_S PDB project team looks forward to working with ETH and SED researchers to exchange ideas for best practices and data formats for future projects.

2.4.5 Other Back-End Considerations

The MySQL RDBMS is open-source, with online resources and documentation, and its use in NGL allows for seamless transition into the development of the V_S PDB. The RDBMS in particular was selected because of its widespread use (Pratt & Last 2014). The use of the NoSQL database format, which has the benefit of easier horizontal scaling for parallel computing, was briefly investigated, but ultimately it was decided against its use, for two reasons: (1) the V_S PDB is not foreseen to become too large to become unwieldy (currently the database is less than 10 MB), and (2) consistency with NGL was desired.

The V_S PDB is currently hosted on an online server managed under the uclageo.com domain, which is maintained by the geotechnical engineering group at the UCLA Department of Civil & Environmental Engineering. The V_S PDB project team have also worked with technical staff of the Natural Hazards Engineering Research Infrastructure (NHERI) DesignSafe Cyber-Infrastructure platform (Rathje et al. 2017) to replicate the V_S PDB onto servers at the Texas Advanced Computing Center (TACC), which provides a convenient backup. Replication onto DesignSafe also allows for improved user interaction with the database, a topic discussed in Section 2.5 below.

2.5 ONLINE INTERFACE

Convenient and accurate dissemination of data to users is essential for the V_S PDB project. This requires a rapid-rendering, accurate, and intuitive website that interfaces with an efficient database. The back-end MySQL RDBMS functions in conjunction with a web services interface developed using Hypertext Markup Language 5 (HTML5) and JavaScript to display the V_S PDB online. PHP-Hypertext Processor (PHP) is used to communicate between the HTML website and the RDB and perform queries on the RDB. The project website is located at https://uclageo.com/VPDB/, is publicly accessible, but requires registration to download data. The website is saved on the same server in which the RDB resides, which increases speed in fulfilling database user requests. The Leaflet API, an open-source JavaScript library, is used to render the georeferenced data in a map interface on the website. Figure 2.14 shows a screenshot of the V_S PDB website's map interface, zoomed into the Quintara site described in Section 2.3.2.1. It is also noted here that the issue of

location accuracy that originally existed for Quintara and other sites has been resolved, with all updated data stored online in the RDB.

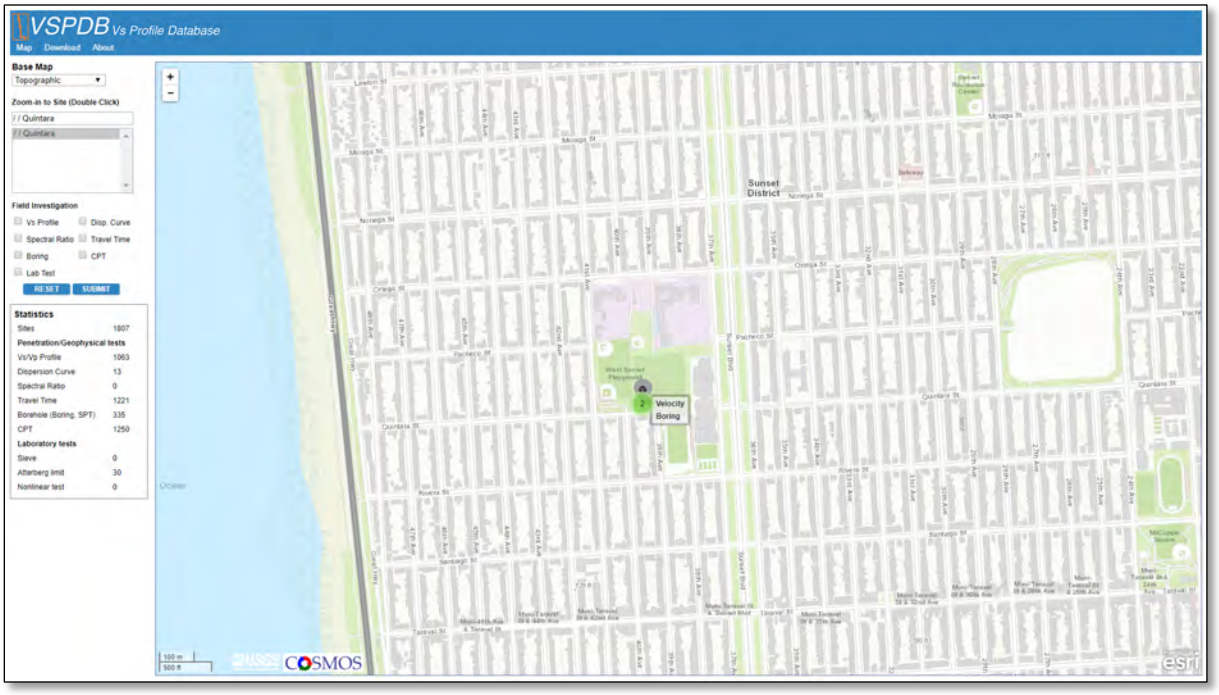


Figure 2.14 Screenshot V_S PDB website, zoomed into Quintara site in San Francisco.

The website contains three main sections: Map, Download, and About. The About section describes the project background, researchers involved in project development, sponsoring agencies, and links to publications. The main user interface with the database resides on the Map page and is split into two main sections: the map, which renders the data on an ESRI ArcGIS Online base map, and the information and query panel on the left. The ESRI base map defaults to a topographic map, but also contains options for street maps and other physical and cultural maps. The map automatically updates data points as the user zooms in and out of the page, using Leaflet Marker Clustering when the map is zoomed out, and allowing individual sites and locations within sites to pop out from a central location (if co-located) or spread out when the map zoom is greater. The left panel includes a statistical summary of the various types of data included in the V_S PDB and allows basic user queries for rendering sites on the map: users can filter what is plotted by selecting one or more data types, or alternatively the user can use the search bar to zoom into specific sites of interest by name.

The icons representing sites on the map are clickable and will display a pop-up balloon. This balloon contains the high-level metadata for the site and displays links to individual data sets at different locations within the site. The user can select one of three command buttons for each data type: “data”, “plot”, and “location”. Clicking the “data” option sends a request to the server to download a CSV file generated from the data in the RDB for the site of interest, including site-level metadata and the data itself (e.g., a V_S Profile). Clicking “plot” runs plotting routines to enable data visualization within a popup window; this option is available for V_S and V_P profiles, dispersion curves, HVSR ordinates as a function of frequency, CPT traces and travel time plots, and boring

logs (including stratigraphic and SPT information). Figure 2.15 shows the Quintara site with the pop-up window plotting the V_S profile from a downhole measurement, and also displays the velocity uncertainty as a function of depth (this plot region will remain blank if not populated in the database). The "location" button functions to zoom in to the data location more closely, a useful feature for sites that have multiple measurement locations.

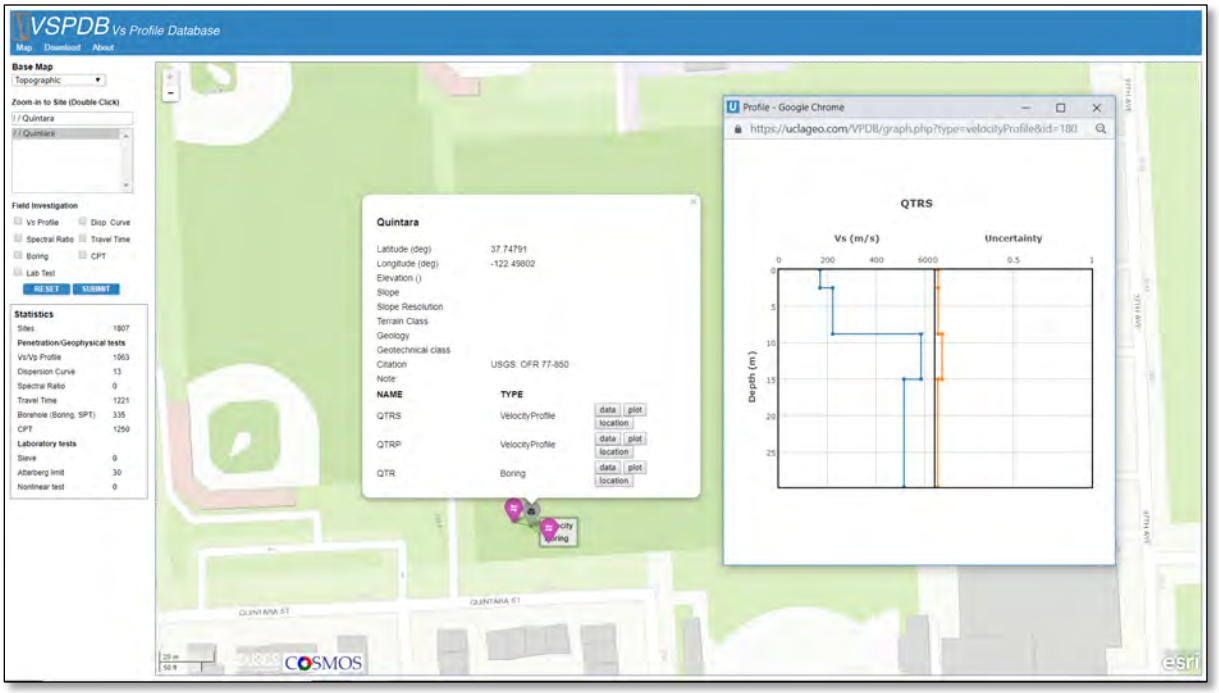


Figure 2.15 Screenshot V_S PDB website, zoomed into Quintara site with metadata pop-up balloon and example of V_S profile data and associated uncertainty from downhole measurements.

The Download page of the website allows users to perform refined queries to search for data using user-specified parameters. For example, users can search for V_S profiles based on ranges of V_{S30} or maximum profile depth, which can be useful if the user knows their intended depth of investigation or NEHRP site classification required (i.e. per BSSC 2009) for their project. Users can also search using geographic ranges, such as a box on the map or using the Point & Range feature to identify a maximum radius to search around a point of interest. Figure 2.16 shows an example search using a point in San Francisco and a maximum search radius of 20 km; this query returns 77 different V_S and V_P profiles. This search tool will be useful for users seeking data near a site of interest.

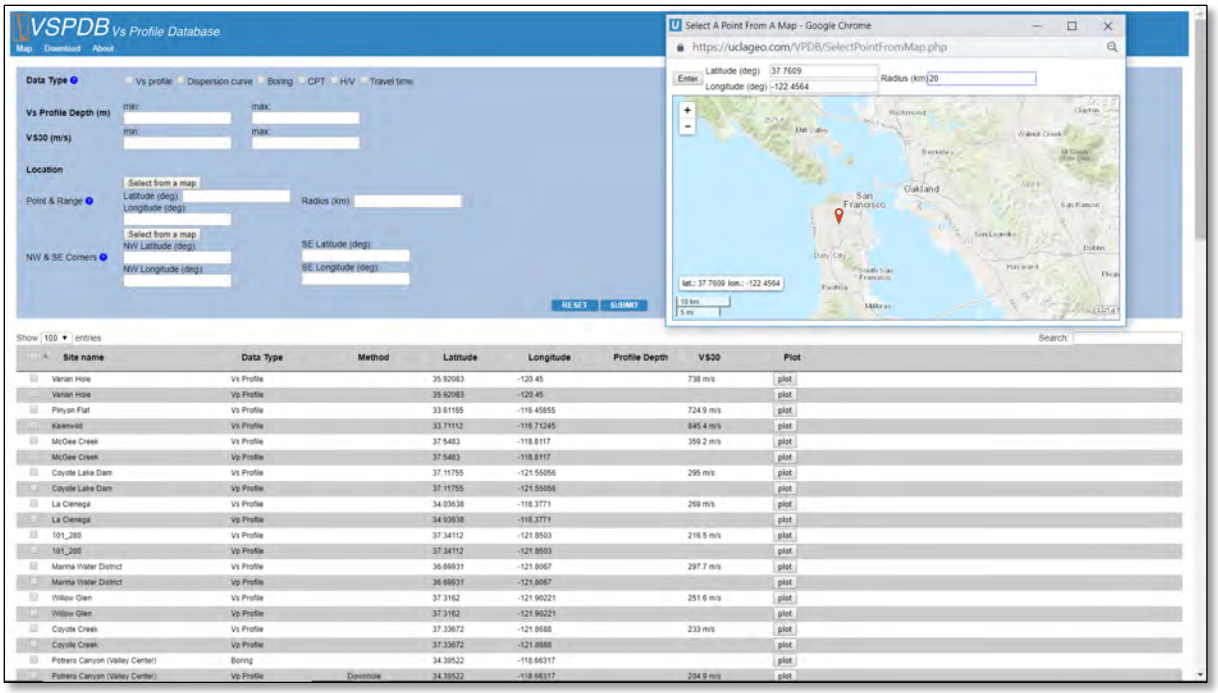


Figure 2.16 Screenshot V_S PDB website's data Download page, with Point & Range feature engaged for user query.

2.6 SUMMARY AND CONCLUSIONS

The United States Community V_S Profile Database is a resource to provide convenient access to public-domain V_S profiles and other pertinent data (e.g., V_P profiles and geotechnical conditions) and metadata for sites in the U.S. V_S data spanning five decades, multiple measurement techniques, accompaniment with varying levels of other data, and widely disparate formats are prepared in a uniform manner within a formal relational database. A relational database schema has been developed, with the benefit of extensive input from an advisory panel and the public (via workshops), to ensure the data organization will meet end user needs. Site-level metadata attributes (location, geology, and geomorphological parameters) are assigned to all sites in the database using uniform protocols to ensure consistency. The relational database is accessible through a convenient web interface that dynamically retrieves data according to user searches, enabling targeted data downloads or plotting.

It is anticipated that the database will be useful to researchers investigating earthquake ground motions, site effects, ground failure, and soil-structure interaction. For practitioners, the query-able database will facilitate office-based reconnaissance and improve site characterization.

The profile database is envisioned as a living tool that will grow over time. The growth can be in additional data, but also in the database schema itself, which is expandable to accommodate new data types and new data/metadata fields.

The V_S PDB project team envisions that eventually that members of the community will be able to log in and upload data to the database, with the source of the data being clearly identified. Uploaded data would automatically be formatted to the JSON format and then placed in queue for

review. Upon review and acceptance, the data would be automatically parsed into SQL tables for inclusion in the database and representation on the website. It is hoped that this effort will motivate similar efforts elsewhere, particularly in seismically active regions.

3 Database for Ground Motion Studies

The non-ergodic ground motion analyses described in Section 1.3 require a large database with many recordings for the second component (referred to here as *Full Database*). The database requirements for this component match those for ground motion model development projects. A subset of that database is used in components three and four (*Database Subset for Site Response Studies*).

3.1 FULL DATABASE

The database used in this study draws from an expanded version of 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 California to the NGA-West2 database (373 events, 1463 stations, 14,231 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. 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 tables within a relational database, which is housed on a local server. Data modifications and additions 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. Figure 3.1 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. We focus here on the Bay Area and southern California regions. Moreover, since difficulties can be encountered in the analysis of site terms using small magnitude data, we only consider **M** >= 4.0 events (Stafford et al. 2017). The data from events within the Bay Area and southern California regions in Figure 3.1 is derived from 25 earthquakes that have produced about 9,300 three-component recordings within the distance cutoffs suggested by Boore et al. (2014). These data are screened for magnitude (requiring **M** >= 4), to remove duplicate recordings (e.g., seismometers and accelerometers at the same location), and to remove recordings that appear to be unreliable from instrument malfunctions or similar. This leaves about 5873 usable three-

component records. Figure 3.2 shows the locations of these events and of the 1185 recording stations with recordings.

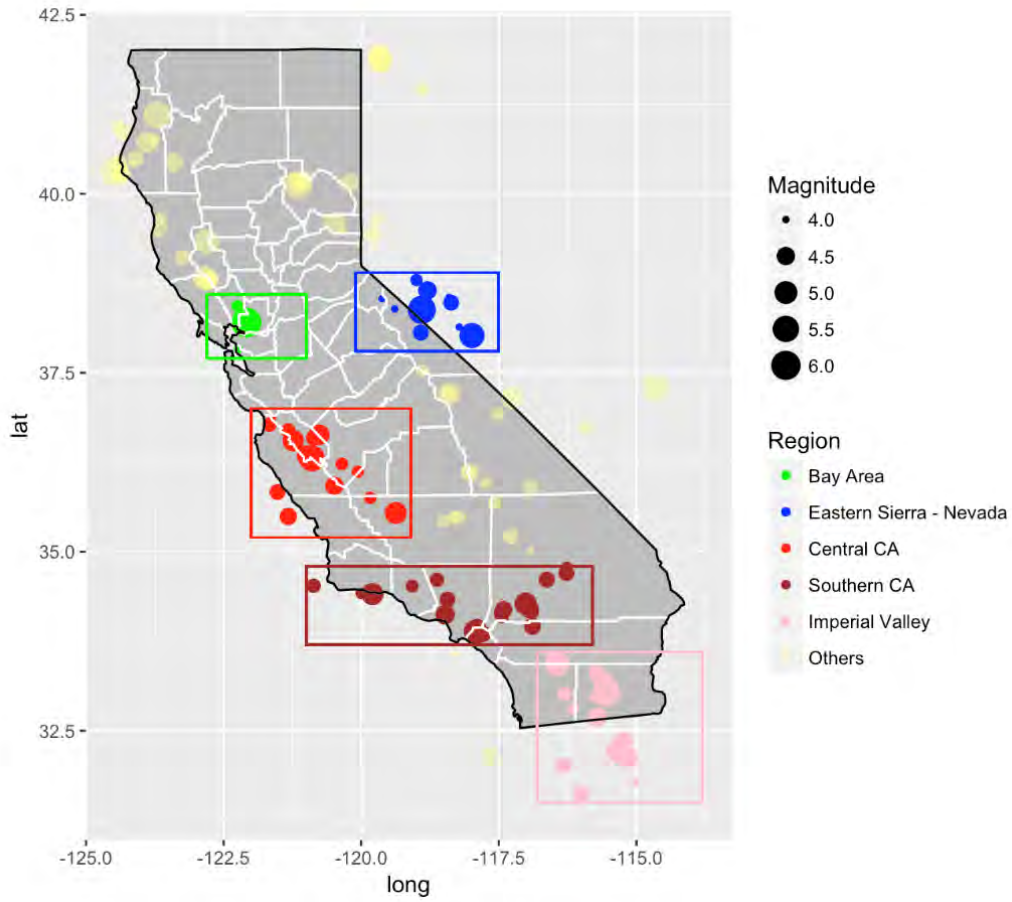


Figure 3.1 Locations of earthquakes in California and northern Mexico with **M** >= 4.0 since 2011 for which ground motion data has been compiled for addition to the NGA-West2 database

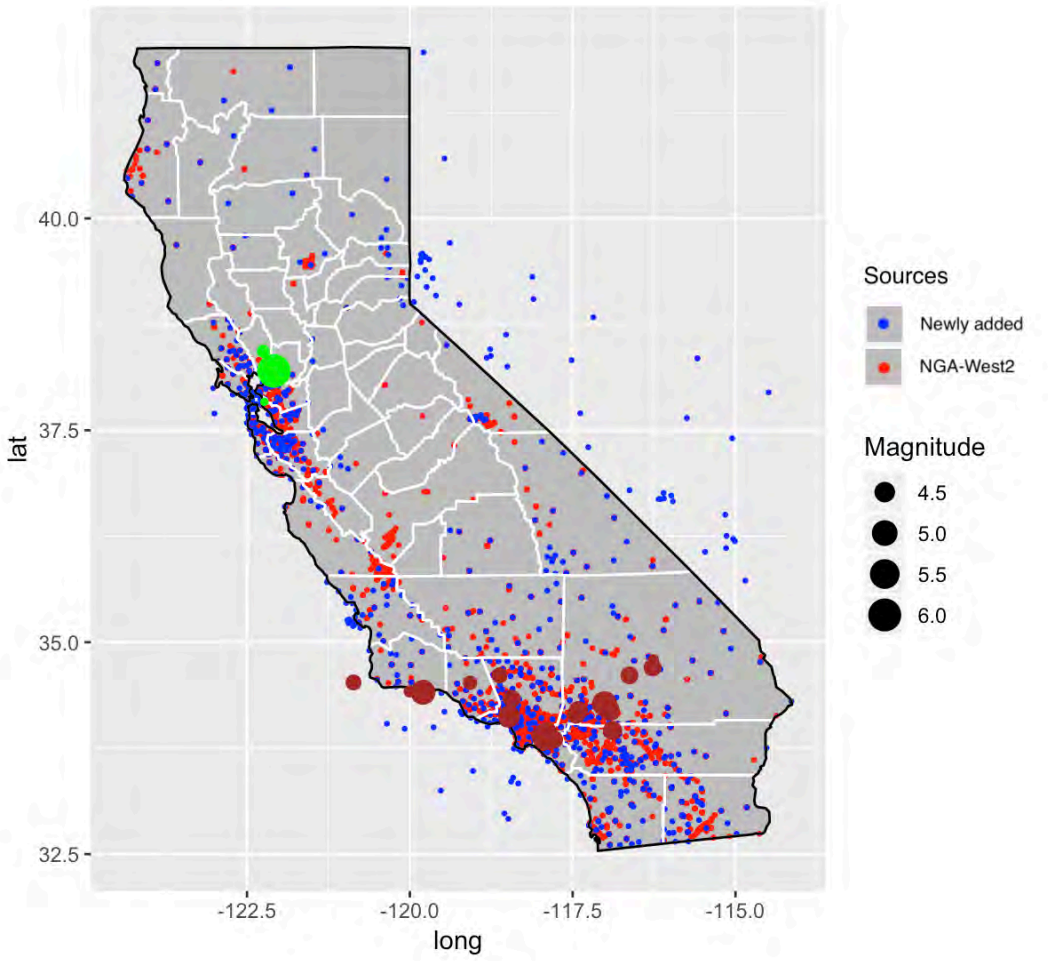


Figure 3.2 Map of California showing locations of considered earthquakes with **M** >= 4.0 since 2011 and locations of stations that recorded the events (blue – new stations, red – stations in NGA-West2 database)

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 and highest 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 3.3 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% at 2.5 sec.

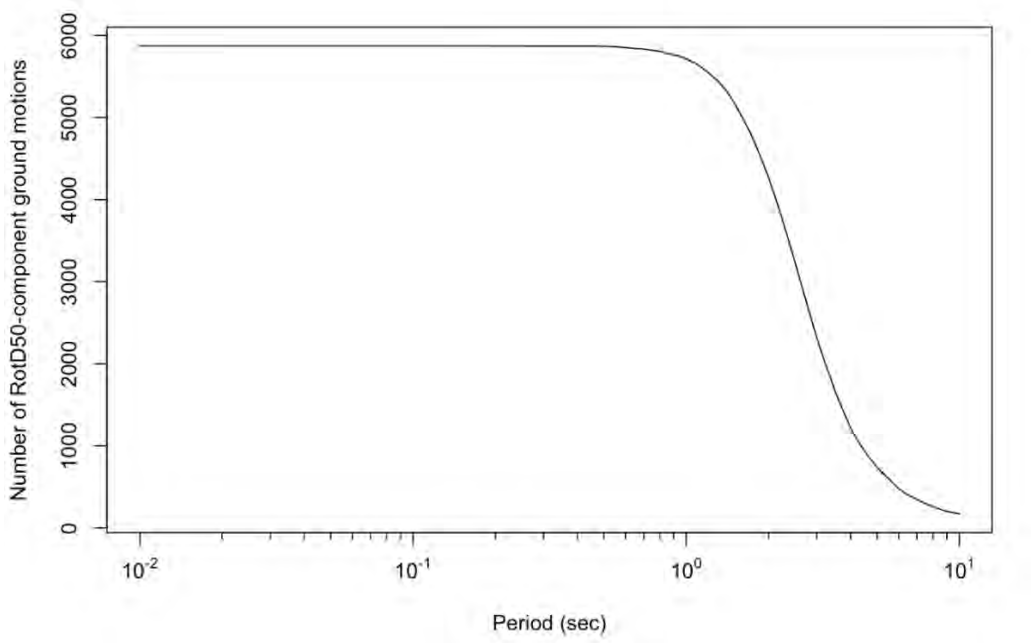


Figure 3.3 Number of usable RotD50-component ground motions as a function of oscillator period for the data added for the Bay Area and southern California regions

The assignment of site parameters is described in the next section.

3.2 ASSIGNMENT OF SITE PARAMETERS FOR FULL DATABASE

Considering both the NGA-West2 data and new data, there are 1818 recording sites shown in Figure 3.2. Of those, 1340 are sites that were included in the NGA-West2 site database. Hence, there are 478 new sites that require assignment of site parameters. We use measured V_S profiles to compute V_{S30} when available, and in the absence of this data, we use proxies (slope gradient – Wald and Allen (2007); terrain category – Yong et al., 2012 and Yong 2016; surface geology – originally by Wills and Clahan (2006) and Kriging interpolated by Thompson et al 2014, and later updated by Wills et al. 2015 and Thompson 2018).

When more than one proxy-based model for V_{S30} estimation is available, it is customary to employ multiple models to compute a weighted estimate of the mean V_{S30}. In the NGA-West2 project, the mean misfit of a proxy model relative to observations (μ_{lnV}) and its aleatory dispersion (σ_{lnV}) were used to compute relative weight as (Seyhan, et al. 2014):

$$\text{relative weight} = \frac{1}{\mu^2_{\ln V} + \sigma^2_{\ln V}} \qquad [3.1]$$

Actual weights were adjusted from relative weights ensure they sum to be one. However, that method does not consider correlations between proxy models. Kwok et al. (2018) described an approach to assign weights that accounts for correlations between the inputs used in two proxy-based models. Here, we extend that approach to consider an arbitrary number of *n* models.

As a problem of linear optimization with linear constraints, it can be formulated in matrix form as,

$$\text{Minimize} \quad \mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w} \tag{3.2}$$

$$\text{Subject to} \quad \mathbf{0} \leq \boldsymbol{I}\mathbf{w} \leq \mathbf{1} \text{ and } \mathbf{1}^{\mathrm{T}}\mathbf{w} = 1 \tag{3.3}$$

where $\mathbf{w}$ is a column vector of length n, $\boldsymbol{\Sigma}$ is a covariance matrix of dimension $n \times n$ of all available proxy models, $\mathbf{0}$ and $\mathbf{1}$ represent column vectors of length n, and $\boldsymbol{I}$ is an identity matrix (1 along diagonal, 0 otherwise). The i-th row and j-th column entry of $\boldsymbol{\Sigma}$ is the covariance of proxy models i and j, computed as:

$$\boldsymbol{\Sigma_{ij}} = \sigma_{\ln V,i}\sigma_{\ln V,j}\rho_{ij} \tag{3.4}$$

where $\sigma_{\ln V,i}$ and $\sigma_{\ln V,j}$ are standard deviation terms representing the aleatory uncertainty of models i and j, and ρ_{ij} is the correlation coefficient between the two models.

Eq. [3.2] is an objective function that, when minimized, provides the weights $\mathbf{w}$ for the considered proxy models that minimize the aleatory uncertainty of the combined (weighted) model. The first linear constrain in Eq. [3.3] requires the estimated weights to be within the range 0 and 1. The second linear constrain in Eq. [3.3] ensures that the sum of estimated weights is one. The mean prediction of the combined model is,

$$\hat{V}_{S30} = \mathbf{w}^{\mathrm{T}}\overline{\boldsymbol{V}}_{\boldsymbol{S30}} \tag{3.5}$$

where $\overline{\boldsymbol{V}}_{\boldsymbol{S30}}$ contains column vectors of mean V_{S30} estimates for each proxy model.

The optimization process requires a "training" dataset. Our dataset includes VS30 values from measured profiles in California from the VS profile database (Chapter 2) and additional VS30 values at a USGS1 web site (the additional sites have a measured VS30 but not a measured profile, which is why they do not appear in the profile database). We excluded VS30 values measured using the Remi method due to large uncertainty (and potential bias) associated with that approach (Cox and Beekman, 2011). This results in 853 VS30 values from measurements. We considered three proxy models: (1) surface geology with local data adjustment (Wills and Clahan, 2006 and Thompson 2018), (2) terrain proxy model (Yong 2016), and (3) surface gradient model (Wald and Allen, 2007). The calculation was performed using the constrOptim function in R. The optimization procedure produces the weights for the three models as follows:

- Surface geology with local data adjustment: 0.665
- Terrain categories: 0.323
- Surface gradient: 0.012

The relatively low weight for the surface gradient model is caused by the strong correlation of gradient with terrain categories, whereas the gradient model has larger dispersion (σ_{lnV}).

[1] https://earthquake.usgs.gov/data/vs30/us/

3.3 DATABASE SUBSET FOR SITE RESPONSE STUDIES

A subset of the full database is applied for site response studies. The criteria used to define this subset are: (1) a minimum number of recordings per site of 10 (applied to ensure statistically robust estimates of site term, η_S); (2) availability of a V_S profile for the site.

Figure 3.4 shows a histogram of the number of recordings at stations in the full database. Of the 1818 sites in the full database, 366 meet the minimum recordings/site criterion.

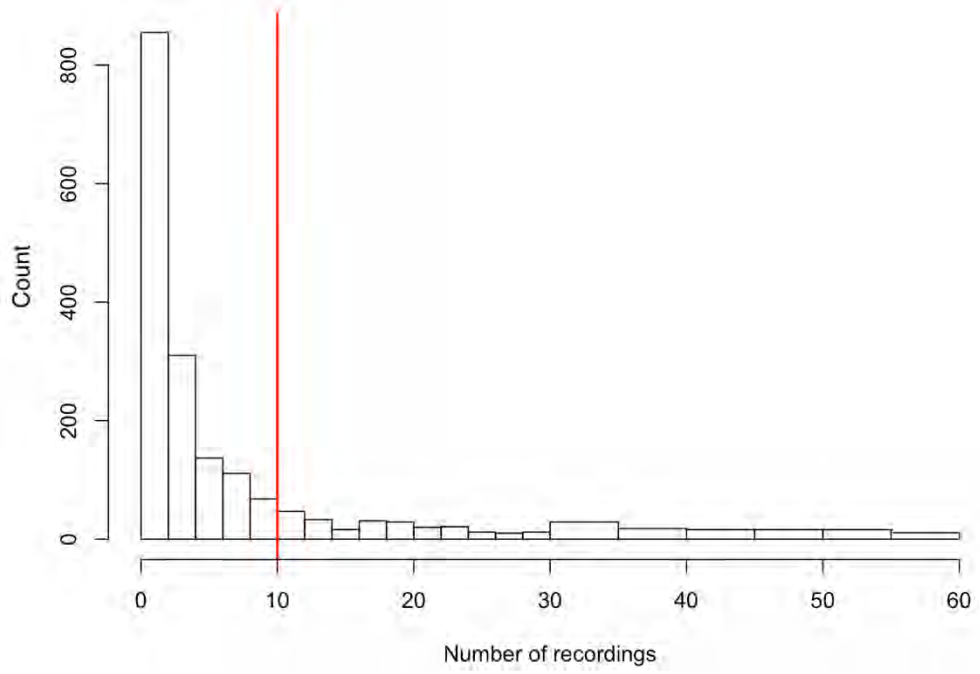


Figure 3.4 Histogram of number of recordings at stations in the full database. Ten is the minimum number of records/station for sites considered in the present research

We performed a search for V_S profiles for each of the sites meeting the first criterion. This was done using the shear wave velocity profile database compiled for California (Chapter 2). We find 145 sites with a V_S profile more than 30 m depth and within around 200 m of the strong motion site. Many of these profiles are from Yong et al. (2013), which provides V_S profiles from various surface wave tests and H/V spectral ratios from microtremors. Of the 145 sites with V_S profiles, only 3 have a boring log that indicates stratrigraphic details and soil/rock layer descriptions. This geotechnical data is needed to apply models for modulus reduction and damping as a function of shear strain.

Most of the recordings used in this research involve low ground motion amplitudes. Figure 3.5 shows a histogram of the ratio (strain index):

$$I_\gamma = \frac{PGV}{V_{S30}} \qquad [3.6]$$

where *PGV* is from the surface recording and is taken from the RotD50 component. This ratio provides an index related to shear strain (Idriss, 2011; Kim et al, 2016), and can be used to judge the degree to which soil responses are likely to be affected by nonlinearity. As shown in Figure

3.5, 97% of ground motions in the subset have $I_\gamma < 0.03\%$. We conclude that the soil responses are predominantly in the linear range, meaning that modulus reduction is unity and damping is at the minimum value. We refer to the minimum damping from geotechnical models (Darendeli 2001 for soils with fines; Menq 2003 for granular soils) as D_{min}^{L}. As a result, the primary need for stratigraphic and material description information is to define D_{min}^{L} as a function of depth.

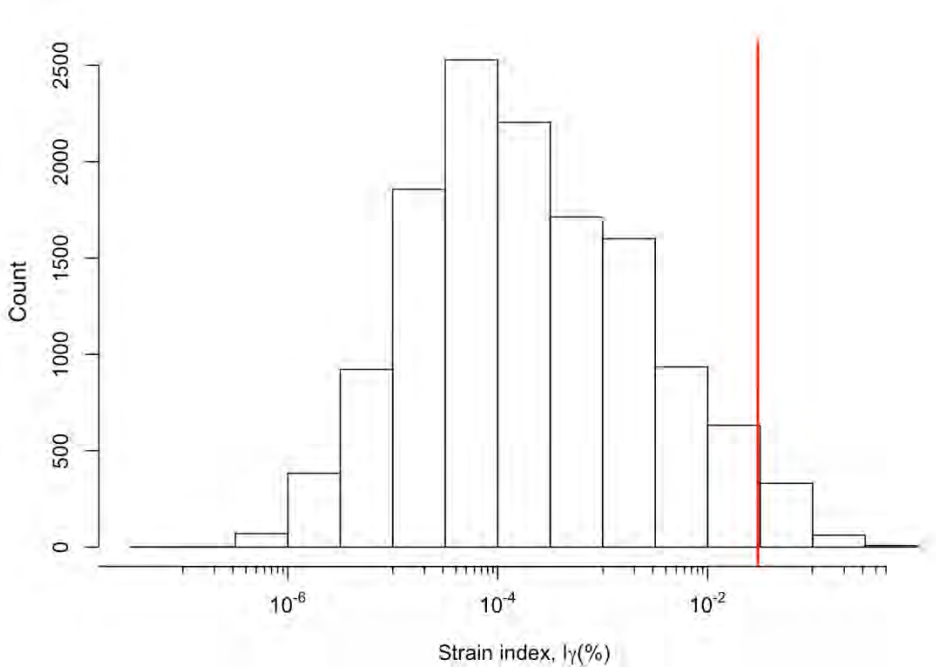


Figure 3.5 Histogram of strain index number of recordings at stations in the full database

To derive D_{min} profiles for use in ground response analyses, the next chapter describes (1) how stratigraphy was inferred to enable estimates of unit weight and D_{min}^{L} (for sites without borehole logs); and (2) how site spectral amplitude decay parameter (κ) was measured from recordings and then interpreted to constrain small-strain damping. As such, that chapter supports the development of alternative damping profiles, each of which are being considered in the validation analyses.

4 Model Inputs

In this chapter, we describe protocols used to assign unit weight and material damping if there is no available boring logs information and also describe the development of a κ-informed damping model for a site based only upon surface ground motion recordings.

4.1 INFERENCE OF UNIT WEIGHT AND MATERIAL DAMPING

Ground response analyses for linear conditions require shear wave velocity, unit weight, and D_{min} profiles. Shear wave velocity profiles are measured at each of the sites in the *Database Subset for Site Response Studies*. In most forward applications, geotechnical site characterization provides borehole logs that describe site stratigraphy and soil type information, which can be used to derive the input parameters used to predict unit weight and D_{min}^{L}. As described above, this is not the case for many of the sites considered in this research. This section describes how we estimate unit weight and soil parameters used to estimate material damping.

4.1.1 Unit Weight

For soil units, we estimate unit weight using phase relationships, which relate unit weight to void ratio, specific gravity, and saturation. Void ratio is taken from an empirical relationship with V_S shown in Figure 4.1 and given as (Rogers et al., 1985):

$$V_S = 42.9 + 94.1/e^2 \quad [4.1]$$

where V_S is in units of m/s.

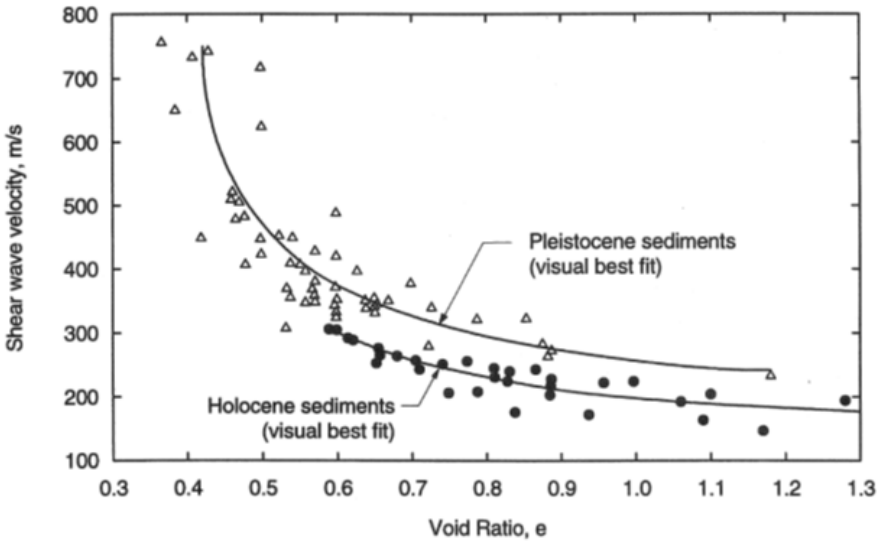


Figure 4.1 Empirical relationships between void ratio, age, and shear-wave velocity for alluvial sediments in southern California (Fumal and Tinsley 1985). Eq. [4.1] fits the combined data (Rogers et al 1985)

Specific gravity is commonly taken as $G_S = 2.7$. Saturation (S) is taken as 1.0 below the first depth where V_P exceeds 1500 m/s. Above that depth, or over the full depth where V_P data is absent, saturation is assumed as 50%. Total unit is then computed as:

$$\gamma = \frac{G_S \gamma_w}{1+e}\left(1 + \frac{eS}{G_S}\right) \qquad [4.2]$$

where γ_w is the unit weight of water (10 kN/m^3).

For rock units, we assigned unit weight based on V_S as follows:

$$\gamma = \begin{cases} 20\ \text{kN/m}^3, & \text{if } 450 < V_S < 700\ \text{m/s} \\ 22\ \text{kN/m}^3, & \text{if } V_S > 700\ \text{m/s} \end{cases} \qquad [4.3]$$

4.1.2 Stratigraphy and Soil Type to Estimate $D_{\min}$ in Soil Layers

Stratigraphic and soil type information is needed to apply the geotechnical model for $D_{\min}$ estimation by Darendeli (2001), which is conditioned on plasticity index (PI), over-consolidation ratio (OCR), and mean effective stress. Effective stress can be calculated using unit weights from the prior section and water table depth (as applicable). PI and OCR are generally derived from laboratory tests on samples retrieved from the field.

We consider two types of available information as potentially useful to assign stratigraphy and soil type information – the mapped surface geology and the V_S profile. Surface geology is used to estimate soil type near the ground surface. The V_S profile is used in combination with the surface unit assignment to estimate variations with depth.

Surface geology is taken from state-wide geologic maps by Wills and Clahan (2006) and Wills et al. (2015). We assume relationships between surface geological unit and PI/OCR, with details indicated in Table 4.1. Considerations in the development of the relationships in Table 4.1 include:

- Geologically young sediments (Holocene) are assumed to have low OCR, and older units are assumed to have relatively high OCRs. The rationale is that young deposits have relatively limited pseudo-overconsolidation from ageing and are unlikely to have experienced significant unloading from natural geological processes.
- Young sediments deposited in quiescent environments (e.g., bays, lakes, central/flat portions of alluvial basins) are assumed to be relatively fines- and clay-rich, thus having high PI. Young alluvial sediments deposited on steeper gradients are assumed to be relatively granular (PI = 0).
- Tertiary sedimentary bedrock units often carry information on rock type (e.g., shale, sandstone, etc.). We assume the bedrock units are similar to corresponding soil units (i.e., shale and sandstone interpreted as clay and sand, respectively).
- For pre-Quaternary units without information on material type or depositional environment, there is no basis for assuming relatively coarse- or fine-grained behavior. We assume an intermediate condition in this case (roughly corresponding to low-plasticity silt).

Table 4.1 is organized in reference to 12 geological units, recommended by Wills and Clahan (2006), that are encountered for the stations in the full database: Qal1, Qal2, and Qal3 are relatively young alluvial sediments likely to be of Holocene age; Qoa is older alluvium of Pleistocene age; QT describes sediments in the early Pleistocene to Pliocene periods, for which the method of deposition is unknown; Tsh, Tss, and Tv comprise Tertiary age bedrock consisting of shale, sandstone, or volcanic-origin materials (typically basalt or rhyolite), respectively; serpentine is a metamorphic rock of Tertiary age largely comprised of the clay mineral serpentinite; and Kss, Kjf, and crystalline are hard rock, typically of Cretaceous age.

Before assigning one of the hard rock classes (Kss, Kjf, crystalline), we perform a visual check of morphology using Google Earth™. When this check indicates that the surface appears to be soil, and if the velocity of the nearest-surface layer is compatible with soil, we assign a soil surficial unit and assign rock at greater depths where velocities become fast.

The soil property assignments in Table 4.1 apply for ground surface layers. The assignment of properties at depth is made in consideration of gradients in the V_S profile. If the surface layer consists of sediments or Tertiary rock, the soil index properties are not changed in successive layers absent sudden changes in velocity with depth. Sudden changes can trigger soil type changes – for example, when a granular layer is underlain by a much slower layer, the underlying unit is taken as clay. Similarly, when a fine-grained surface layer is underlain by a much stiffer layer, the underlying material is taken as granular. When a layer velocity exceeds 760 m/s, it is taken as rock. Figure 4.2 shows the flow chart used to assign soil type information as a function of depth.

Table 4.1 The list of 12 geological units and their corresponding PI and OCR. Ma indicates million years.

Geological age	Geol. unit	Description	Estimated Parameters
Holocene (< 0.011 Ma)	**Qal1**	Quaternary Holocene alluvium with flat gradients (< 0.5%).	PI = 30 OCR = 1.2
	Qal2	Quaternary Holocene alluvium with moderate gradients (0.5 - 2.0%).	PI = 10 OCR = 1.2
	Qal3	Quaternary Holocene alluvium with steep gradients (> 2%).	PI = 10 OCR = 1.2
Pleistocene (< 2.6 Ma)	**Qoa**	Quaternary Pleistocene alluvium. Soil composition unknown.	PI = 10 OCR = 2
Pliocene (2.6-5.3 Ma). Young era within the Tertiary.	**QT**	Quaternary to Tertiary deposits, including Saugus Fm. in So. CA, Paso Robles Fm. in central Coast Ranges, and Santa Clara Fm. in San Francisco Bay area. Soil composition unknown.	PI = 10 OCR = 2.5
Tertiary (2.6-66 Ma).	**Tsh**	Shale and siltstone units, such as the Repetto, Fernando, Puente, and Modelo Fms in So. CA.	PI = 15 OCR = 3
	Tss	Sandstone units, such as the Topanga Formation in So. CA and Butano Formation in San Francisco Bay area.	PI = 0 OCR = 3
	Tv	Volcanic units including the Conejo Volcanics in Santa Monica Mtns and the Leona Rhyolite in East Bay Hills.	PI = 15 OCR = 3
	Serpentine	Serpentine rock is clay-rich.	PI = 15 OCR = 3
Cretaceous	**Kss**	Cretaceous sandstone of the Great Valley Sequence	NA
	Kjf	Franciscan complex rocks, including mélange, sandstone, shale, chert, and greenstone.	NA
	crystalline	Crystalline rocks, including Cretaceous granitic rocks, Jurassic metamorphic rocks, schist, and Precambrian gneiss.	NA

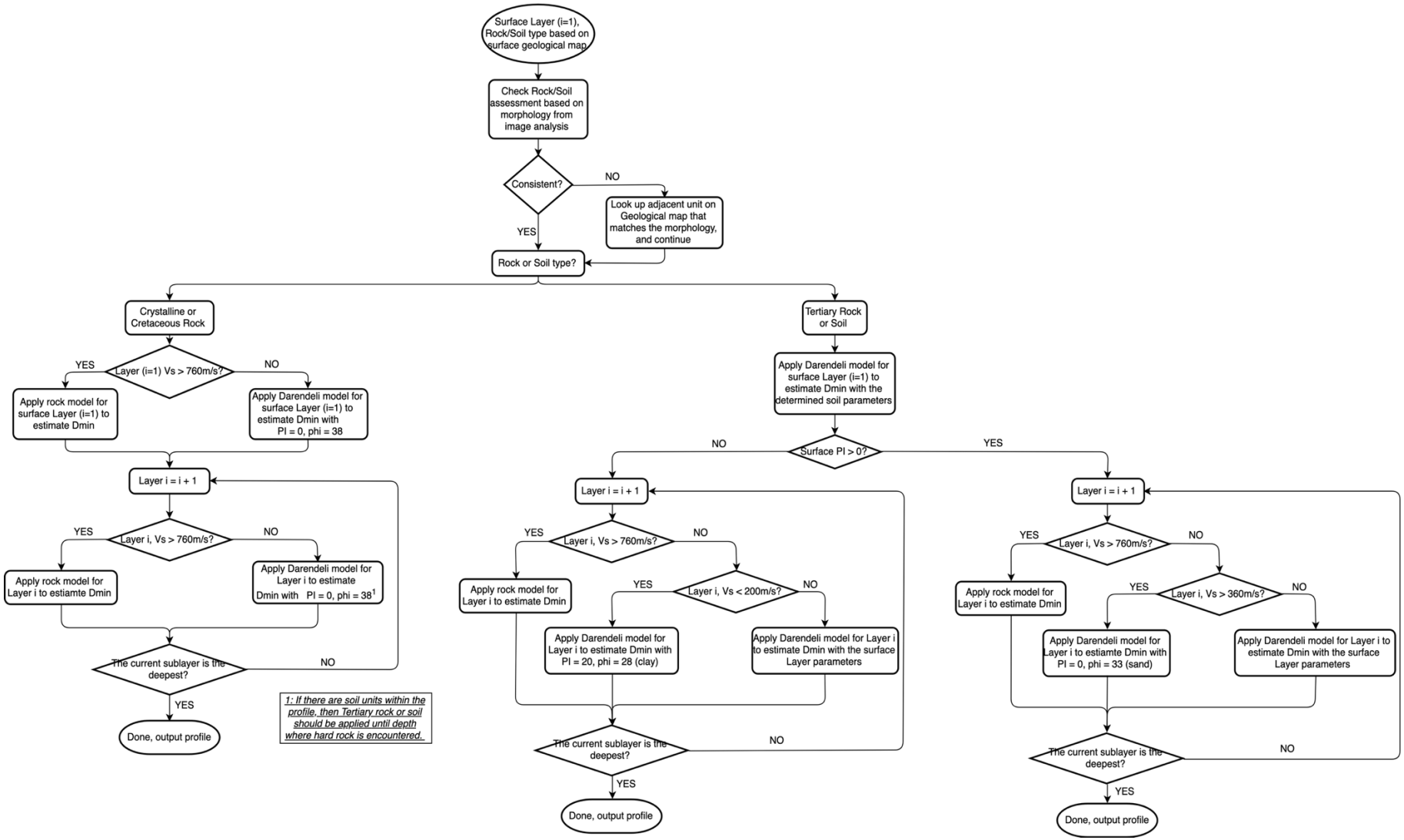


Figure 4.2 Flow chart used to assign soil type information as function of depth

4.1.3 $D_{\min}$ in Firm Rock Layers

The Darendeli (2001) model cannot be used for pre-Tertiary rock (units Kss, Kjf, crystalline). Laboratory data on material damping for such materials is limited. A presumably judgement-based model was presented by Schnabel (1973) and has been widely used since that time. Choi (2008) performed testing on welded Bandelier Tuff and Topopah Spring Tuff and developed damping models. Models from these two sources are compared in Figure 4.3. The $D_{\min}$–component from Topogah Spring Tuff is considered more representative of bedrock materials in our study region based on its unit weight (Bandelier Tuff has low unit weights). The D_{min} range for this material is about 0.2 – 1.0% (average = 0.3%). We have used the Choi model for the present work, but acknowledge that its use carries large uncertainty.

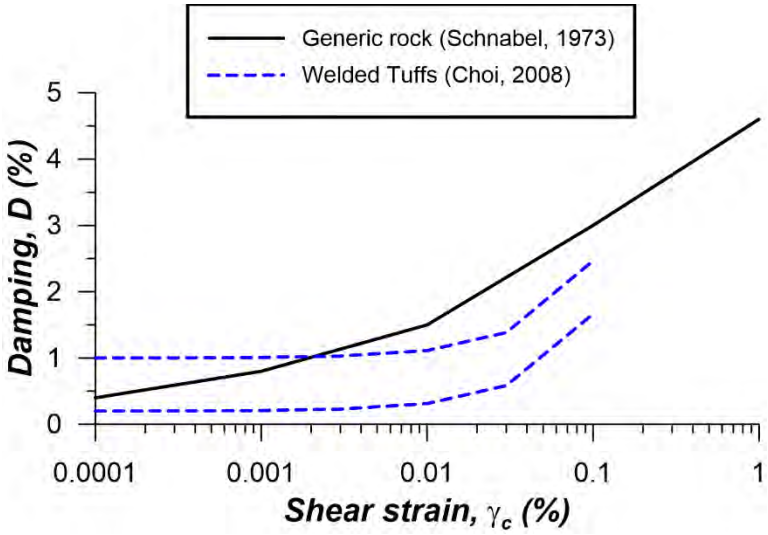


Figure 4.3 Comparison of rock damping model from Schnabel (1973) and range from Choi (2008)

4.2 κ-INFORMED DAMPING MODEL

4.2.1 Approach

Figure 4.4 shows Fourier amplitude spectra for ground motions at two examples sites in our database. The spectra show a characteristic feature, which is decay of Fourier amplitudes with increasing frequency for frequencies beyond the peak in the spectrum. This frequency-dependent decay can be described as:

$$D(f) = \exp(-\pi\kappa f) \qquad [4.4]$$

where f is frequency in Hz and κ is a decay parameter that can be established through fits to data (e.g., Anderson and Hough, 1984).

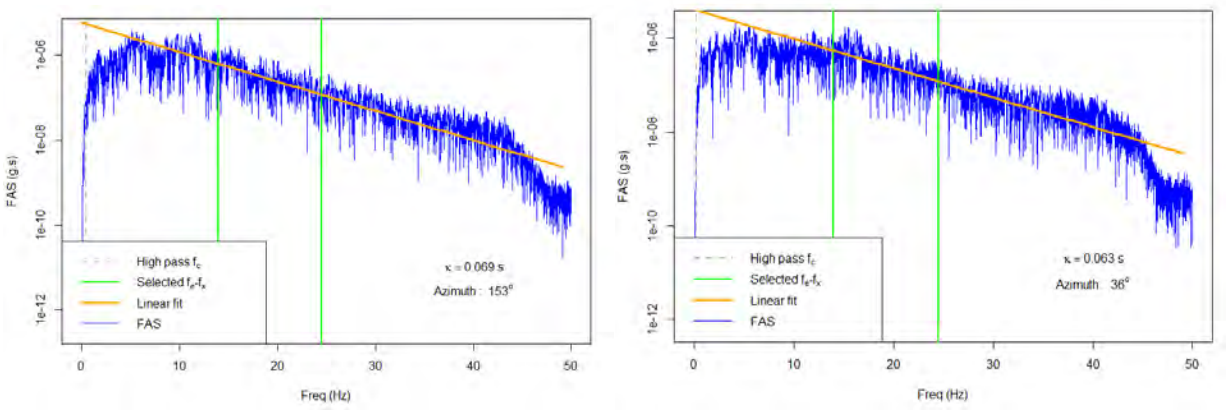


Figure 4.4 Analysis of κ from recordings at two example sites (left with Station Sequence Number (SSN): 100173, right with Station Sequence Number (SSN): 3046)

The decay parameter arises from material damping and wave scattering that occurs on the wave path from source-to-site, often including appreciable contributions from site response. The path and site response contributions to κ combine as (adapted from Anderson 1991):

$$\kappa = \kappa_0 + \kappa_R R \qquad [4.5]$$

where R is site-to-source distance, κ_R is the slope by which κ increases with distance, and κ_0 represents the cumulative effect of damping and wave scattering through the soil column.

The relationship between κ_0 and profile attributes can be expressed as (Hough and Anderson 1988; Chapman et al. 2003; Campbell 2009):

$$\kappa_0 = \int_0^{z_s} \frac{dz}{Q_{ef}(z) V_S(z)} \qquad [4.6]$$

where z_s is the site column thickness (depth to reference crustal rock) and $Q_{ef}(z)$ is the depth-dependent effective material quality factor, representing both the effects of frequency-dependent wave scattering and frequency-independent soil damping. Q_{ef} can be converted to an effective soil damping as follows (Campbell, 2009):

$$D_{ef}(\%) = \frac{100}{2Q_{ef}} \qquad [4.7]$$

Measurements of κ from recordings can, in principal, inform levels of damping applied in ground response analyses as follows:

1. Measure κ for a set of sites from multiple earthquakes, as shown for example in Figure 4.4.
2. Develop a regionally appropriate model for κ_R.
3. Adjust each measured value of κ, for each event recorded at a given site, to estimate κ_0 by re-arranging Eq. [4.5] as $\kappa_0 = \kappa - \kappa_R R$.

4. Since the soil/rock column thickness analyzed in ground response analysis is typically smaller than the full profile to reference crustal rock ($V_S \approx$ 2.5-3 km/s), adjust κ_0 from Step (3) as,

$$\Delta\kappa = \kappa_0 - \kappa_{0,b} = \int_0^{z_p} \frac{2D_{ef}(z)}{100} \frac{dz}{V_S(z)} \quad [4.8]$$

where $\kappa_{0,b}$ is the site decay parameter at the base of the profile and z_p is the depth of the analyzed soil column ($z_p < z_s$).

5. Modify the laboratory damping with a profile-specific adjustment factor F_D to match $\Delta\kappa$ from Step (4), which can be represented by re-writing Eq. [4.8] as:

$$\Delta\kappa = \int_0^{z_p} \frac{2D_{min}^{L}(z) * F_D}{100} \frac{dz}{V_S(z)} \quad [4.9]$$

The depth-invariant value of F_D represents the means by which the field observations of kappa inform the damping model. In some cases, F_D may be unreasonably high. To constrain F_D so that it provides damping values within a realistic range, we have enforced a maximum value of F_D = 10. The approach maintains the scaling of damping with soil type and depth in the laboratory models, while adjusting for other effects encountered in field conditions (scattering). The maximum value of F_D is taken by considering two references Afshari and Stewart (2019) and Tao and Rathje (2019). The former paper studied 21 vertical array sites and estimated F_D and the latter paper studied four vertical array sites and calculated the best multipliers (which has the same meaning as F_D) of $D_{\min}$ to fit observed site responses. The range of the recommended values for F_D and multipliers are from 1.5 up to 9.15, so we take the maximum value as 10.

Implementation of the above procedure requires several model components – distance correction term κ_R and site decay parameter for the base of profile condition $\kappa_{0,b}$. The following sub-sections describe the calculation of κ from recordings, models used for these components, and example results.

4.2.2 Fitting of κ from Ground Motions

We apply the κ-fitting procedures described in Afshari and Stewart (2019), which were adapted from Cabas et al. (2017) and Xu et al. (2020). The fit occurs over a range of frequencies from f_e to f_x (upper and lower bounds, respectively) that is selected for each record.

Search ranges for f_e and f_x are taken as 10-18Hz and 22-28Hz, respectively, each with 0.5Hz increments. For each possible combination of f_e and f_x, κ is computed for combinations of the two horizontal components rotated to various azimuths. The variability of κ with azimuth is computed for each f_e-f_x combination, which is expressed as a coefficient of variation (COV). We seek the combination of f_e and f_x that minimizes the azimuthal variability, and then take κ as the median. The Fourier amplitude spectra for two example sites shown in Figure 4.4 are for the azimuths and frequency ranges identified using this process.

4.2.3 Analysis of Path- and Site Contributions to κ

Rates of crustal attenuation vary spatially due to variations in geologic conditions. Conditions producing relatively fast ground motion attenuation rates (i.e., low crustal quality factor, Q) would be expected to increase κ_R. Insight into spatial variations of attenuation rates are provided by maps of frequency-independent Q (denoted Q_s) by Eberhart-Phillips (2016) for Northern California and Hauksson and Shearer (2006) for Southern California. Figure 4.5 shows maps of California indicating variations of Q_s at a depth of 10 km from the two sources. We estimate Q_s at a depth of 10 km from Eberhart-Phillips (2016) by linear interpolation between depths (the model does not provide Q_s at depth of 10 km). There are systematic differences between Q_s from the two sources, with values of southern California model being higher.

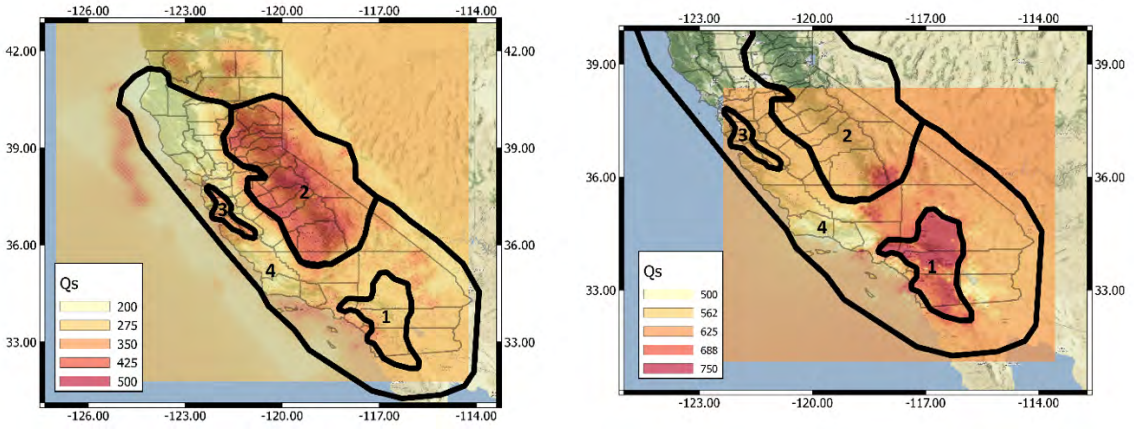


Figure 4.5 Spatial variation of frequency-independent quality factor (Q_s) for California as derived from two models at a depth of 10 km. The heat maps are generated by Inverse Distance Weighting (IDW) in QGIS. Zones considered in mixed-effects analysis of path κ are shown.

Allowing for differences in Qs between the two sources, considering both maps and calculating the best fit, we have assigned four zones of approximately uniform Qs, with the intention of computing κ_R separately for each zone. If the value of κ_R for zone i is taken as $\kappa_{R,i}$, then the $\kappa_R R$ term in Eq. [4.5]is computed as:

$$\kappa_R R = \sum_i \kappa_{R,i} R_i \qquad [4.10]$$

where R_i is the path length (between source and site) through zone i. Distance R_i is zero if the path does not go through the zone i.

We use mixed-effects regression (more specifically, random intercept model) to obtain $\kappa_{0,j}$ for each station j and $\kappa_{R,i}$ for each zone i, by adapting Eq. [4.5] as follows:

$$\kappa_{k,j} = \kappa_{0,j} + \sum_{i=1}^{n} \kappa_{R,i} R_{k,i} \qquad [4.11]$$

where $\kappa_{k,j}$ is the measured κ from recording k at station j, $\kappa_{0,j}$ is the site-specific decay parameter at station j, n is the number of zones in California, and $R_{k,i}$ is the source-to-site path length for recording k that goes through zone i. Station terms $\kappa_{0,j}$ are taken as random effects and path terms $\kappa_{R,i}$ as fixed effects. Eq. [4.11] is solved using an equivalent matrix form. The matrix form can be expressed explicitly as follows, the an example consisting of three sites,

$$\begin{bmatrix} \kappa_{1,1} \\ \kappa_{2,1} \\ \vdots \\ \kappa_{n_1,1} \\ \kappa_{n_1+1,2} \\ \kappa_{n_1+2,2} \\ \vdots \\ \kappa_{n_1+n_2,2} \\ \kappa_{n_1+n_2+1,3} \\ \kappa_{n_1+n_2+2,3} \\ \vdots \\ \kappa_{n_1+n_2+n_3,3} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \kappa_{0,1} \\ \kappa_{0,2} \\ \kappa_{0,3} \end{bmatrix} \qquad [4.12]$$

$$+ \begin{bmatrix} R_{1,1} & R_{1,2} & R_{1,3} & R_{1,4} \\ R_{2,1} & R_{2,2} & R_{2,3} & R_{2,4} \\ \vdots & \vdots & \vdots & \vdots \\ R_{n_1,1} & R_{n_1,2} & R_{n_1,3} & R_{n_1,4} \\ R_{n_1+1,1} & R_{n_1+1,2} & R_{n_1+1,3} & R_{n_1+1,4} \\ R_{n_1+2,1} & R_{n_1+2,2} & R_{n_1+2,3} & R_{n_1+2,4} \\ \vdots & \vdots & \vdots & \vdots \\ R_{n_1+n_2,1} & R_{n_1+n_2,2} & R_{n_1+n_2,3} & R_{n_1+n_2,4} \\ R_{n_1+n_2+1,1} & R_{n_1+n_2+1,2} & R_{n_1+n_2+1,3} & R_{n_1+n_2+1,4} \\ R_{n_1+n_2+2,1} & R_{n_1+n_2+2,2} & R_{n_1+n_2+2,3} & R_{n_1+n_2+2,4} \\ \vdots & \vdots & \vdots & \vdots \\ R_{n_1+n_2+n_3,1} & R_{n_1+n_2+n_3,2} & R_{n_1+n_2+n_3,3} & R_{n_1+n_2+n_3,4} \end{bmatrix} \begin{bmatrix} \kappa_{R,1} \\ \kappa_{R,2} \\ \kappa_{R,3} \\ \kappa_{R,4} \end{bmatrix}$$

where each site has n_1, n_2, and n_3 records respectively, then there are $(n_1 + n_2 + n_3)$ measured κ. There are three unknown random intercepts $\kappa_{0,1}$, $\kappa_{0,2}$, and $\kappa_{0,3}$, and four unknow slope parameters

$\kappa_{R,1}$, $\kappa_{R,2}$, $\kappa_{R,3}$, and $\kappa_{R,4}$. Two matrices, indicator matrix with dimension ($n_1 + n_2 + n_3$) by 3 and distance matrix with dimension ($n_1 + n_2 + n_3$) by 4, can be easily calculated by the corresponding positions and source-site locations.

The mixed effect regressions are then performed in R [packages *nlme* (Pinheiro et al. 2019) or *lme4* (Bates et al., 2015)] using the full database (i.e., NGA-West2 stations and records in California as augmented here). The resulting κ_R values are shown for each zone in Figure 4.6, where they are plotted against the Q_s values from the two references. The error bars shown in the figure indicate the estimation error for κ_R from the regressions and the within-zone ranges of Q_s. Regressions provided negative κ_R in Zone 2. This is because the data is not adequate to constrain κ_R. As κ_R physically cannot be negative and the negative estimate of κ_R is not statistically significant (with p-value > 0.05), we take it as zero for subsequent use in this study.

Another problem happens to Zone 3. Higher *Qs* values indicate lower attenuation and lower κ_R. However, our regression shows the opposite. Considering multiple studies (Kuehn, et al., 2019) have showed a much faster attenuation in Zone 3 than other regions, we will keep our results.

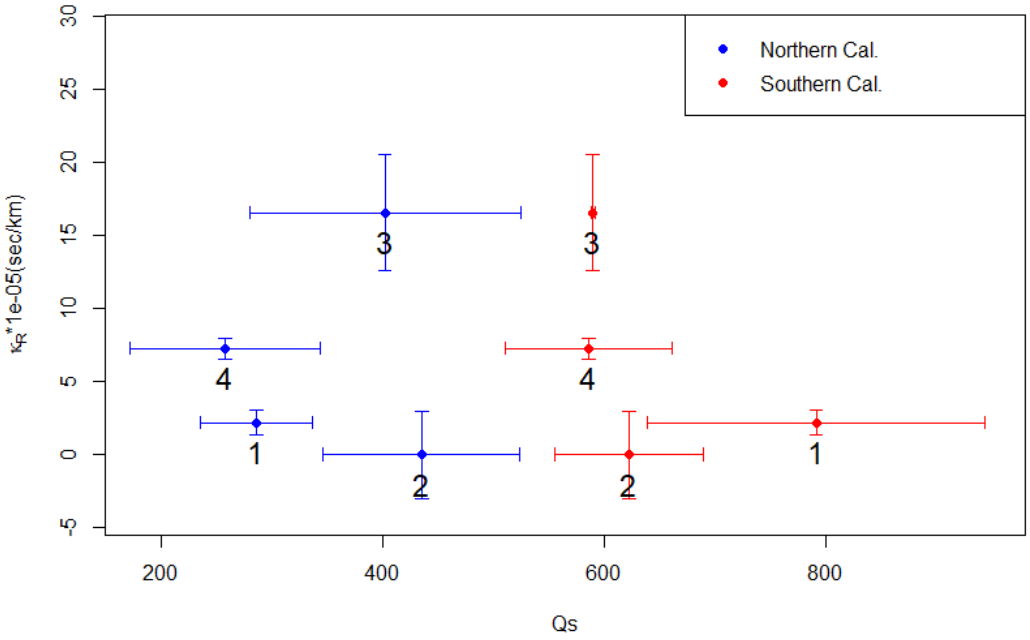


Figure 4.6 Variation of κ_R with average Qs within the four zones shown in Figure 4.5. Average Q_s is taken from both Eberhart-Phillips (2016) for Northern California and Hauksson and Shearer (2006) for Southern California

4.2.4 Base of Profile Site Decay Parameter, $\kappa_{0,b}$

The base of profile site decay parameter $\kappa_{0,b}$ is needed to estimate the change in site kappa over the profile depth ($\Delta\kappa$) using Eq. [4.8]. Because the sites considered in this research are surface-only instruments, $\kappa_{0,b}$ cannot be measured (i.e., from a downhole instrument) but instead is estimated from models. Several such models were considered.

Silva et al. (1998) used California data to relate κ_0 to V_{S30},

$$\kappa_0 = \begin{cases} 0.008 \text{ sec}, & V_{S30} > 1500 \text{ m/s} \\ 0.020 \text{ sec}, & 760 < V_{S30} < 1500 \text{ m/s} \\ 0.030 \text{ sec}, & 360 < V_{S30} < 760 \text{ m/s} \end{cases} \quad [4.13]$$

Van Houtte et al. (2011) and Xu et al. (2020) used larger databases from the KiK-net array in Japan and NGA-West data to derive empirical relationships between κ_0 and V_{S30}. Van Houtte et al. (2011) proposed:

$$ln\kappa_0 = 3.490 - 1.062 lnV_{S30} \quad [4.14]$$

Xu et al. (2020) proposed:

$$ln\kappa_0 = \begin{cases} k_1(lnV_1)^2 + k_2 lnV_1 + k_3, & 100\text{m/s} < V_{S30} < V_1 \\ k_1(lnV_{S30})^2 + k_2 lnV_{S30} + k_3, & V_1 < V_{S30} < V_2 \\ k_1(lnV_2)^2 + k_2 lnV_2 + k_3, & V_2 < V_{S30} < 3000\text{m/s} \end{cases} \quad [4.15]$$

where k_1=0.18, k_2=1.816, k_3=-7.38, V_1=155 m/s, and V_2 = 2000 m/s. The units of κ_0 are sec in both Eq. [4.14] and [4.15].

We apply the Xu et al. (2020) relationship in the present work. To obtain $\kappa_{0,b}$, we estimate the V_{S30} corresponding the base of the soil column by projecting vertically (constant velocity) the V_S at the deepest portion of the profile. We then enter this value into Eq. [4.15] to compute $\kappa_{0,b}$. Results of this process for the two example sites are shown in Table 4.2, as are derived values of $\Delta\kappa$ and F_D using Eq. [4.8] and Eq. [4.9].

Table 4.2 Site kappa results for the SSN = 100173 and SSN = 3046 sites

Site	κ_0 (sec)	$\kappa_{0,b}$ (sec)	$\Delta\kappa$ (sec)	F_D
Keenwild Fire Station, Mountain Center, CA (SSN: 100173)	0.054	0.018	0.036	10
Borrego Springs, CA (SSN : 3046)	0.044	0.026	0.018	10

4.3 Q-V_S DAMPING MODEL

There is also an empirical relation between Q_{ef} and V_S proposed by Campbell (2009),

$$Q_{ef} = 7.17 + 0.0276 * V_S \quad [4.16]$$

Afshari and Stewart (2019) has implemented this with Eq. [4.7] to derive a formula for soil damping that is only dependent on V_S,

$$D_{ef}(\%) = \frac{100}{2Q_{ef}} = \frac{100}{2(7.17 + 0.0276 * V_S)} \quad [4.17]$$

The performance relative to data of ground response analyses with this damping model will be compared in Chapter 5 and 6 to ground response analysis performance when a geotechnical damping model and κ-informed damping model are used.

5 Observation-to-Model Comparisons

The objective of the ground motion analyses conducted in this research was to examine the effectiveness of alternate site response prediction methods at predicting observed site responses. This occurs by comparing observations-to-models for individual sites, and by assembling results across groups of sites (including the full inventory). This chapter provides the former examination by illustrating representative example results for individual sites. The upcoming doctoral thesis of the first offer will include results for all sites in an electronic form.

Three site response analysis methods are considered. The first is the ergodic model (conditioned on V_{S30} and sediment depth) that comprises the reference model. The second is ground response analysis (GRA), which simulates wave propagation through a damped, elastic medium. The third is the square-root impedance model (SRI), which approximates the impedance components of ground response with simple expressions based on the principle of wave energy conservation across layer interfaces.

In this chapter, we first describe how site response effects are evaluated from recordings. Then, we describe below both site response models and present example results with variable levels of model-to-data fit.

5.1 EVALUATION OF OBSERVED SITE RESPONSE

Site amplification is represented by variable Y (in arithmetic units), which is the ratio of a ground motion intensity measure on the ground surface (Z) to the intensity measure on the reference condition (typically rock), X:

$$Y = \frac{Z}{X} \text{ or } \ln Y = \ln Z - \ln X \qquad [5.1]$$

All three variables (X, Y, Z) are assumed to be log-normally distributed with standard deviations ϕ_{lnX}, ϕ_{lnY}, and ϕ_{lnZ}, respectively. The use of variable ϕ for standard deviation indicates that it is associated with the within-event dispersion of earthquake ground motions. The total standard deviation (σ) also includes contributions from between-event variability τ.

Earthquake ground motions are affected by source, path, and site effects, each of which has corresponding terms in ground motion models (GMMs). Each of those terms may be systematically in error for a particular earthquake source, wave path, and site. Provided sufficient data exists, those systematic component errors can be estimated through mixed effects methods of

residuals analysis (e.g., Pinheiro et al. 2019). A general expression to help visualize such effects is as follows (adapted from Al Atik et al., 2010):

$$lnZ_{ij} = (\mu_{lnX})_{ij} + (F_S)_{ij} + \eta_{E,i} + \eta_{P,ij} + \eta_{S,j} + \varepsilon_{ij} \quad [5.2]$$

where Z_{ij} represents a recorded ground motion for event i and site j, $(\mu_{lnX})_{ij}$ represents the mean from a GMM for reference site conditions, $(F_S)_{ij}$ represents the mean site amplification from an ergodic site term, and $\eta_{E,i}$, $\eta_{P,ij}$, and $\eta_{S,j}$ represents event, path, and site terms, respectively. The term ε_{ij} represents the remaining residual when each of the above so-called fixed effects are removed. For the present work, we are not interested in non-ergodic path effects; hence, we use a slightly simplified form of Eq. (5.2):

$$lnZ_{ij} = (\mu_{lnX})_{ij} + (F_S)_{ij} + \eta_{E,i} + \eta_{S,j} + \varepsilon_{ij} \quad [5.3]$$

Note that the sum $(\mu_{lnX})_{ij} + (F_S)_{ij}$ is simply the GMM log mean $(\mu_{lnZ})_{ij}$. The difference between the observation and the mean prediction comprises the total residual R_{ij}:

$$R_{ij} = lnZ_{ij} - \left[(\mu_{lnX})_{ij} + (F_S)_{ij}\right] \quad [5.4]$$

The *Full Database* (Section 3.1) is significantly expanded (in California) relative to that used to develop NGA-West2 models. Accordingly, we recomputed residuals for the full data set used in the derivation of the Boore et al. (2014) model supplemented with the additional California data. Event terms and site terms were evaluated by partitioning total residuals as:

$$R_{ij} = c_k + \eta_{E,i} + \eta_{S,j} + \varepsilon_{ij} \quad [5.5]$$

where c_k is the overall model bias for the model k. The partitioning in Eq. [5.5] used mixed effects analysis procedures (*lmer* function in library *lme4* created by Bates et al. 2015). With the residuals having been derived in this manner, the true (non-ergodic) site response relative to a reference condition (typically V_{S30} = 760 m/s) is computed as:

$$(\mu_{lnY})_{ij} = (F_S)_{ij} + \eta_{S,j} \quad [5.6]$$

As a first approximation, the mean site response $(\mu_{lnY})_{ij}$ is independent of event i if the magnitude is greater than 4 and the amplitude of shaking is sufficiently weak than the effects of nonlinearly are modest. As a result, mean site amplification is denoted $(\mu_{lnY})_j$.

Site terms in the Full Database are based on a wide range of available site recordings, including some with as few as one recording. In contrast, the Database Subset for Site Response Studies (Section 3.3) only considers sites with at least 10 recordings along with meeting other site characterization criteria. As a result, the uncertainty in site terms and mean site responses is much lower for the subset than for typical sites in the larger database. However, as period increases, some motions for individual sites within the subset are beyond their usable range (i.e., when period exceeds the inverse of the lowest usable frequency, as defined in Section 3.1). To illustrate this effect, Figure 5.1 shows the decay of the number of sites with usable site terms with period, where a site is no longer considered usable once the available number of recordings at that site is less

than 5. Fall-off in the number of available sites occurs beyond a period of 2 sec, but the data loss is modest even out to 10 sec.

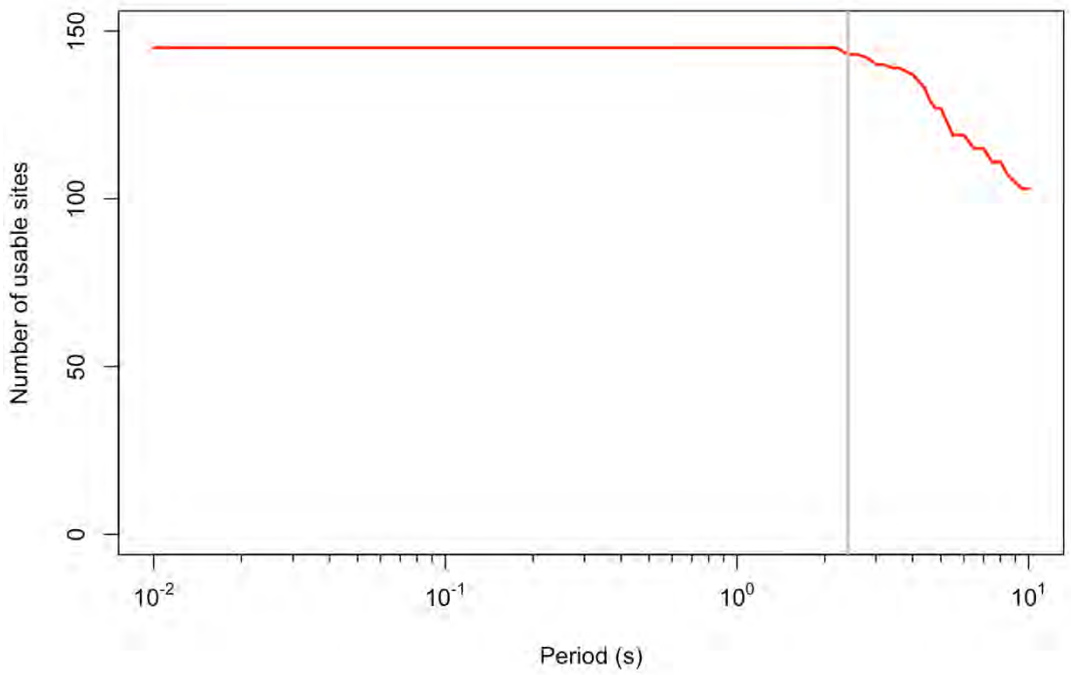


Figure 5.1 Decay of number of usable sites with oscillator period due to exceedance of usable period range. For this plot, the minimum number of records for a site to be considered usable is 5.

Each of the residuals terms in Eq. [5.3] and [5.5] has an accompanying standard deviation. The standard deviation of $\eta_{E,i}$ is denoted τ and represents the event-to-event variability. The standard deviation of $\eta_{S,j}$ is denoted ϕ_{S2S} and represents the site-to-site standard deviation. The standard deviation of ε_{ij} is denoted ϕ_{SS} and represents the "single station" within-event dispersion that combines effects of path-to-path variability and event-to-event variability in site response.

Figure 5.2 shows histograms of site terms for two intensity measures (peak acceleration and velocity) and their correlation. The degree of correlation is strong, as reflected by a correlation coefficient of 0.822. Figure 5.3 shows the dispersion terms evaluated from this process as a function of oscillator period.

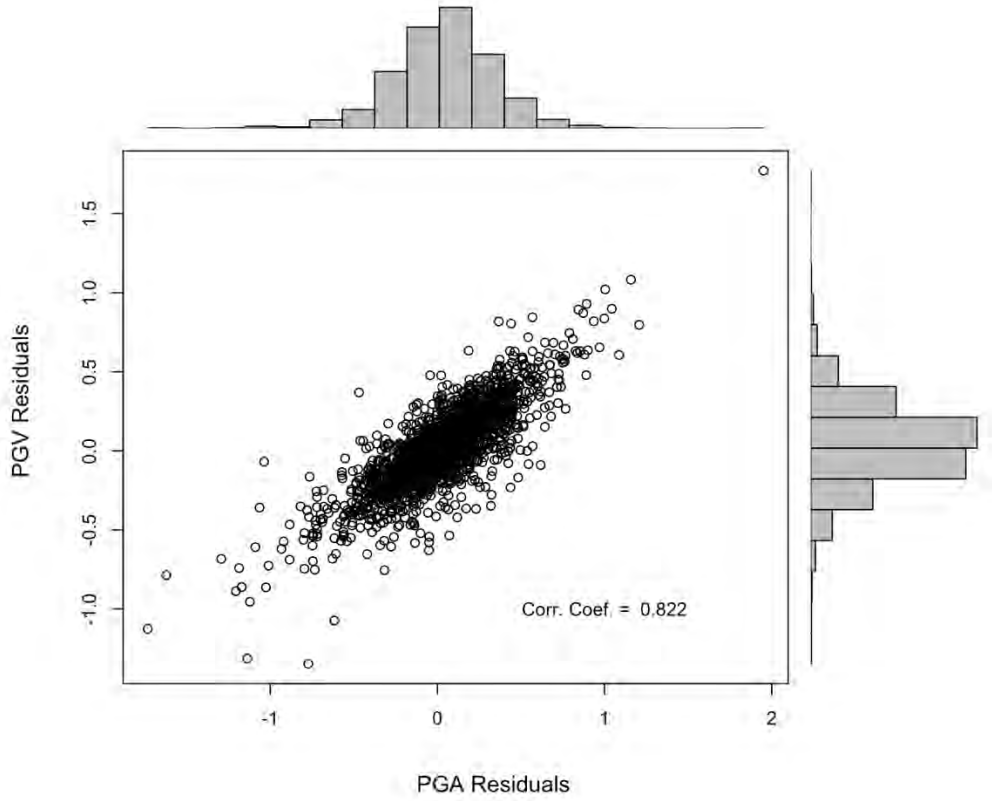


Figure 5.2 Histograms of PGV and PGA and their scatter plot

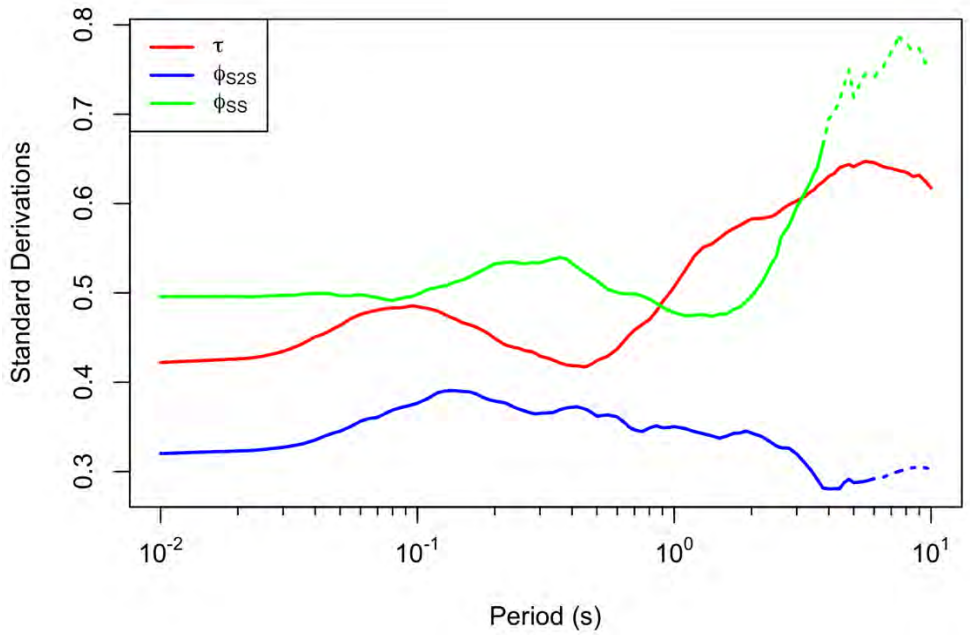


Figure 5.3 Standard derivations of partitioned residuals τ, ϕ_{S2S}, ϕ_{SS}. Shown as dotted when data population is reduced by 25% from the number for PGA.

By default, the above residuals analyses and dispersion terms apply to ground motion predictions using the ergodic site amplification model. Chapter 6 describes how the residuals analyses are modified for cases where site-specific models are used in lieu of ergodic site amplification models, which produces different estimates of the site-to-site variability ϕ_{S2S}.

5.2 GROUND RESPONSE ANALYSIS

One-dimensional ground response analyses (GRAs) model shear wave propagation through horizontal soil layers. As such, these analyses capture impedance and resonance effects on site response, and through the use of equivalent-linear or nonlinear methods, can also capture the effects of stiffness decrease and damping increase with increasing shear strain (e.g., NCHRP 2012, Stewart et al. 2014). However, additional aspects of wave propagation that contribute to site response are not captured, including basin effects.

Given the partial consideration of the full wave field in GRA, and its widespread utilization in engineering practice, validation is an important topic. Section 3.1 briefly reviewed prior validation studies, which are largely based on the utilization of vertical array data. This study differs from that body of work by using data from surface-only instruments. By so doing, we significantly increase the number of observations that can be considered, but at the "cost" of the input motion being uncertain. Moreover, the present approach defines long-period site response features for consideration in the validation, which is not possible with vertical array data.

5.2.1 Approach

There are three principle considerations in running GRA: (1) selection of an analysis platform, (2) selection of material properties, and (3) development of input motions. We use the analysis platform DeepSoil Ver 7 (Hashash et al., 2016), which we run in the linear mode. The assignment of soil parameters for sites with a V_S profile and with or without boring logs to identify soil types was presented in Section 4.1. Here we address the remaining issue of specifying input motions.

For a linear system, site amplification in the frequency-domain (i.e., the site-modification of Fourier coefficients) is independent of input motion attributes. However, the amplification of response spectral ordinates is dependent on the characteristics of input motions, due to sensitivity of oscillator response to the shape of the input spectrum (Stafford et al. 2017). As a result, we gave careful consideration to the means by which the input motion was derived, which is explained in this section.

Our approach is to utilize the recorded ground motion at the soil surface as a starting point for derivation of the input motion. We utilize a deconvolution approach, termed here the "inverse transfer function method" (ITF), to convert the surface records to estimates of the equivalent outcropping motion at the reference depth.

Figure 5.4 illustrates the ITF method, which has the following steps:

1. Fourier coefficients (conjugate complex-valued) of the recorded motion are computed, $Z(f)$;

2. The site transfer function amplitude, $|H(f)|$, under linear conditions is derived using DeepSoil and the site soil properties (the program does not provide transfer function phase);
3. We divide the surface record Fourier amplitudes by transfer function amplitudes for each frequency to estimate Fourier amplitudes of the input motion, $|X(f)|$. The phase of the surface record is combined with this amplitude to evaluate Fourier coefficients for the input, $X(f)$. After this calculation, an inverse Fourier transform (IFFT) is performed to compute the estimated input motion in the time domain.

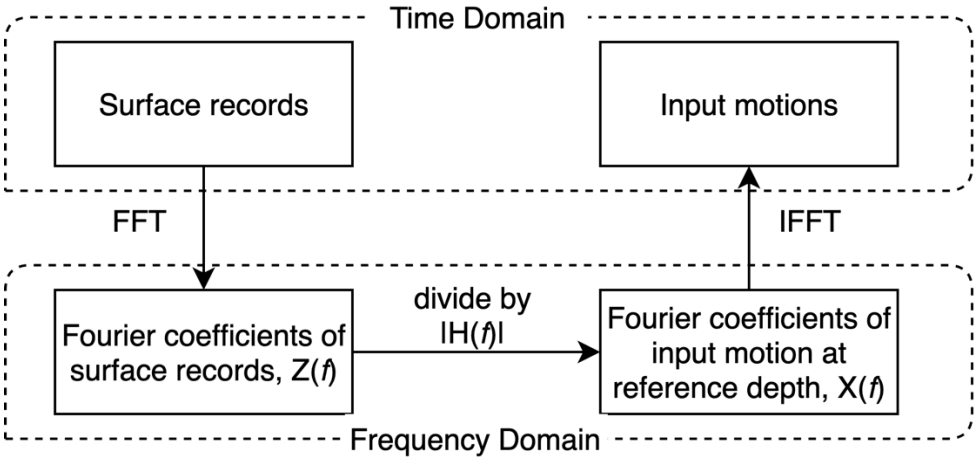


Figure 5.4 Flow chart for the ITF method

Figure 5.5 - Figure 5.10 show for six example sites the V_S profile, measured (or assumed) soil parameters (unit weight and damping) evaluated using procedures from Section 4.1, surface ($|Z(f)|$) and derived input Fourier amplitudes ($|X(f)|$) for a site recording, and response spectra for that same recording. Damping profiles are provided as derived from the geotechnical model (D_{min}^{L}), the κ-informed model, and the Q-V_S model. As shown in the figures, the changes in Fourier amplitudes and response spectra between the surface and the reference condition are limited to frequencies higher that the fundamental site frequency (f_0). Figure 5.11 compares transfer functions to ratios of response spectra for each site; in each case there is only a single transfer function, but the ratios of the response spectra vary from motion-to-motion. Note that $|H|$ and R.R.S. represent transfer function amplitude and ratio of response spectra of surface records to converted input motions.

There are some sites where the results of the deconvolution process produce unrealistic spectral shapes for input, as shown for example in Figure 5.12a. This tends to occur when soil damping is high, specifically when with high damping scaling factors are applied in the κ adjustment (Section 4.2). When this occurs, forward applications in which site amplification is computed using GRA or SRI may be unreliable at high frequencies, leading to unusual shapes of the site amplification (Figure 5.12b). We performed visual screening to identify these problems, and eliminate problematic sites for considering in summary statistics (Chapter 6). These removals only apply to one site with κ-informed damping (the site is retaining for use with other damping models).

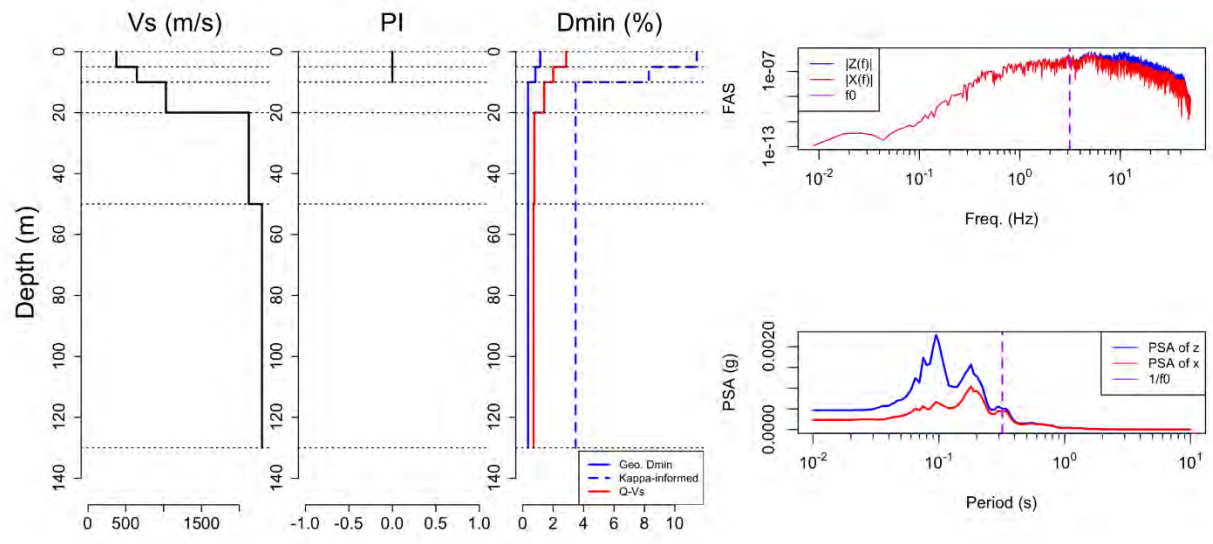


Figure 5.5 Site properties and modification of example surface recording for the site SSN = 100173

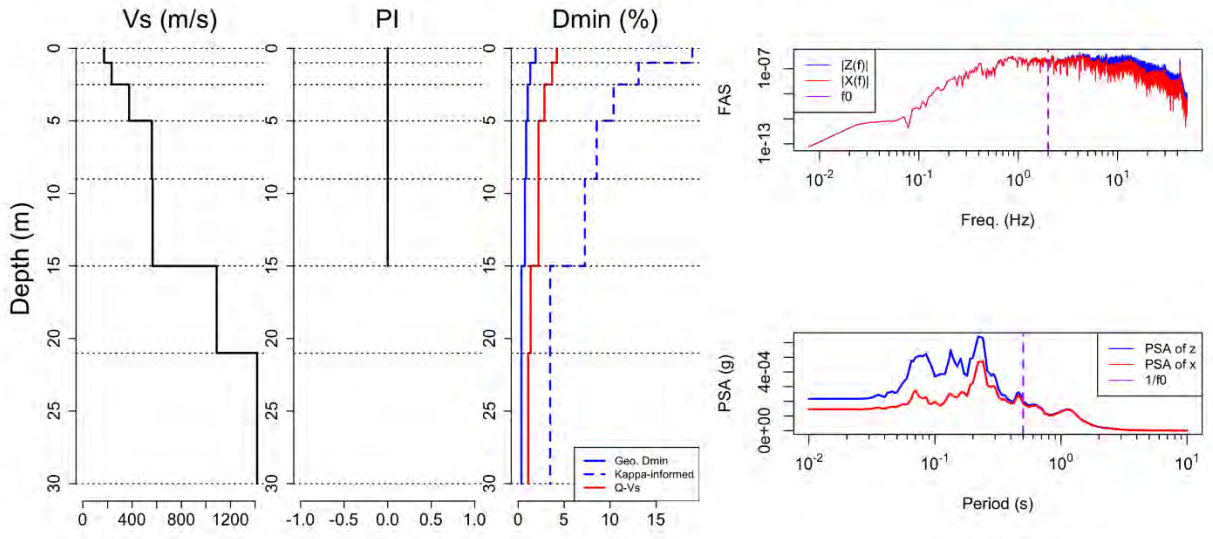


Figure 5.6 Site properties and modification of example surface recording for the site SSN = 3046

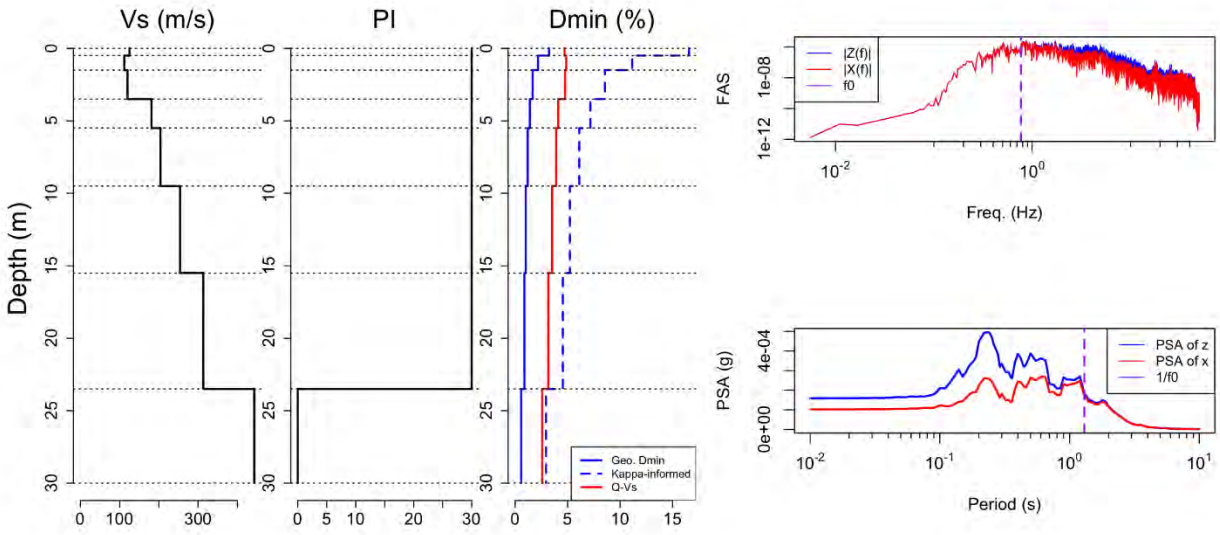


Figure 5.7 Site properties and modification of example surface recording for the site SSN = 100135

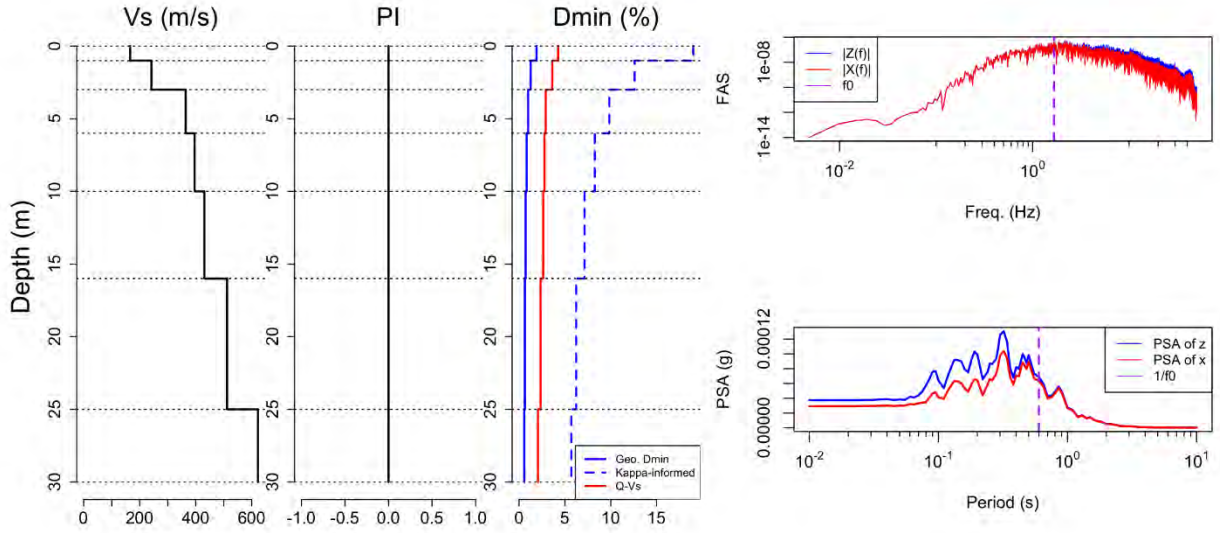


Figure 5.8 Site properties and modification of example surface recording for the site SSN = 3058

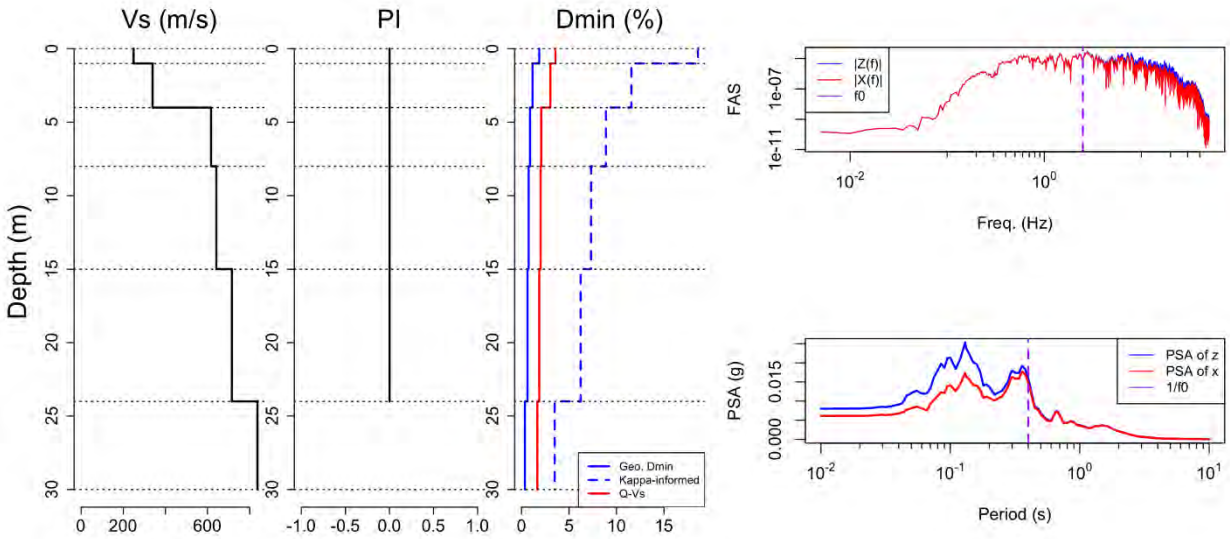


Figure 5.9 Site properties and modification of example surface recording for the site SSN = 100047

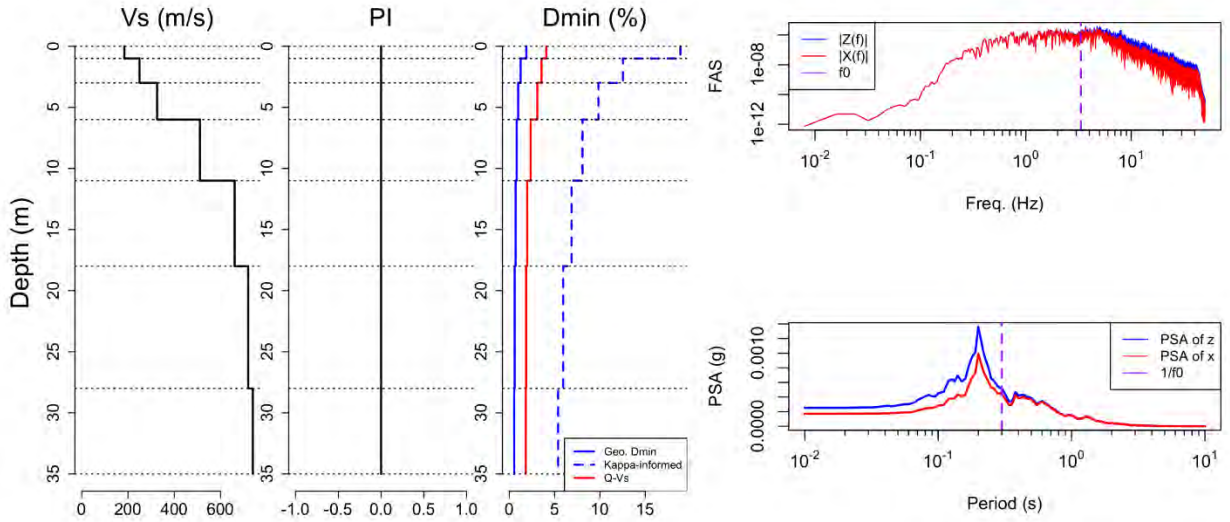


Figure 5.10 Site properties and modification of example surface recording for the site SSN = 3089

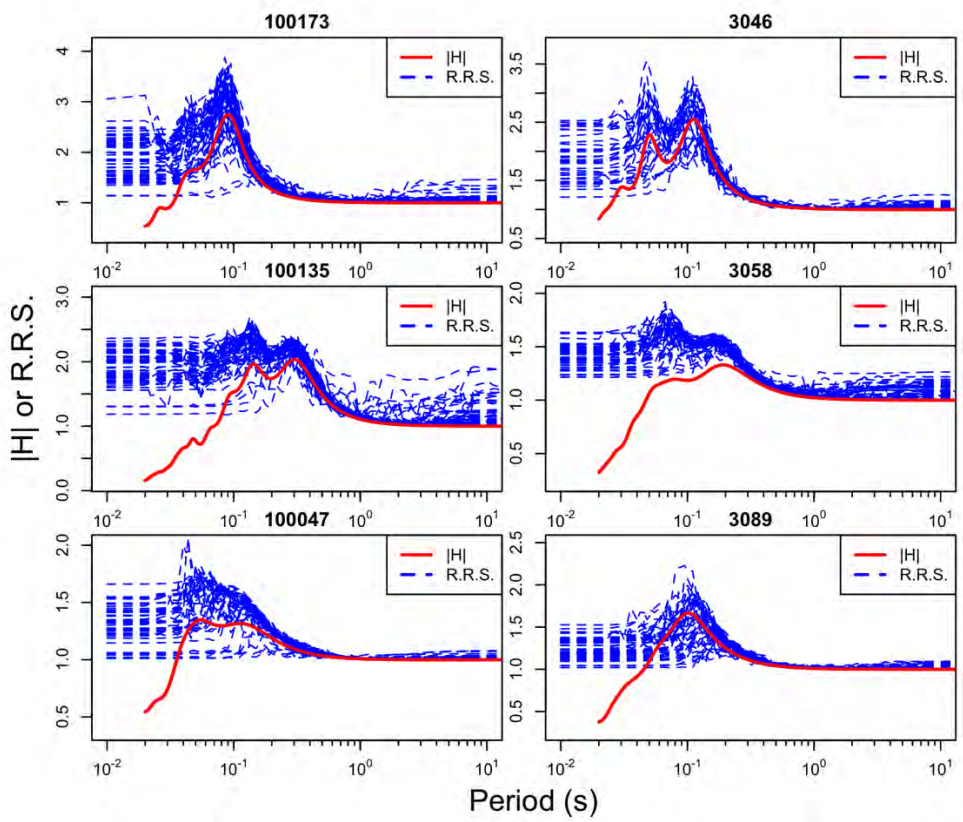
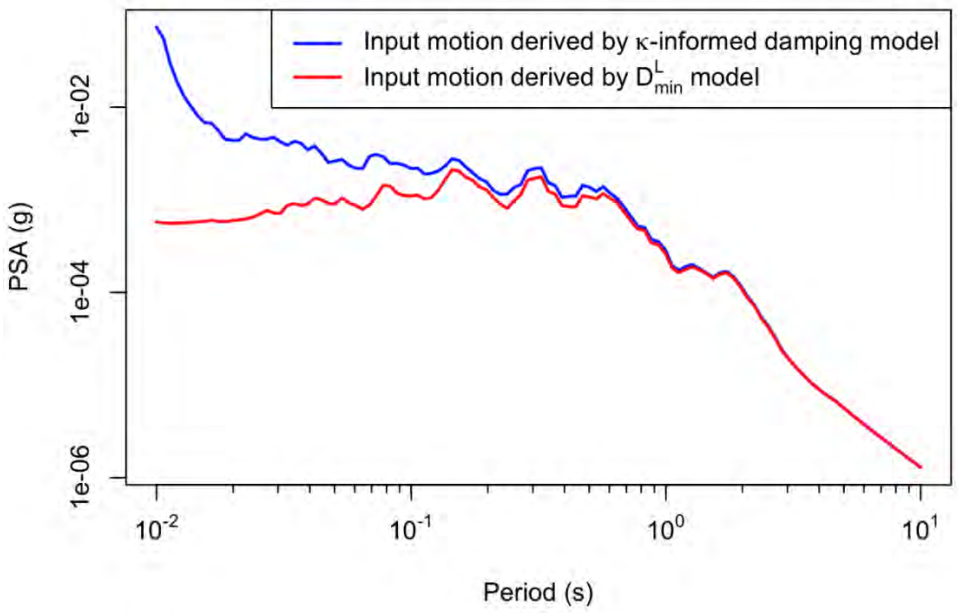
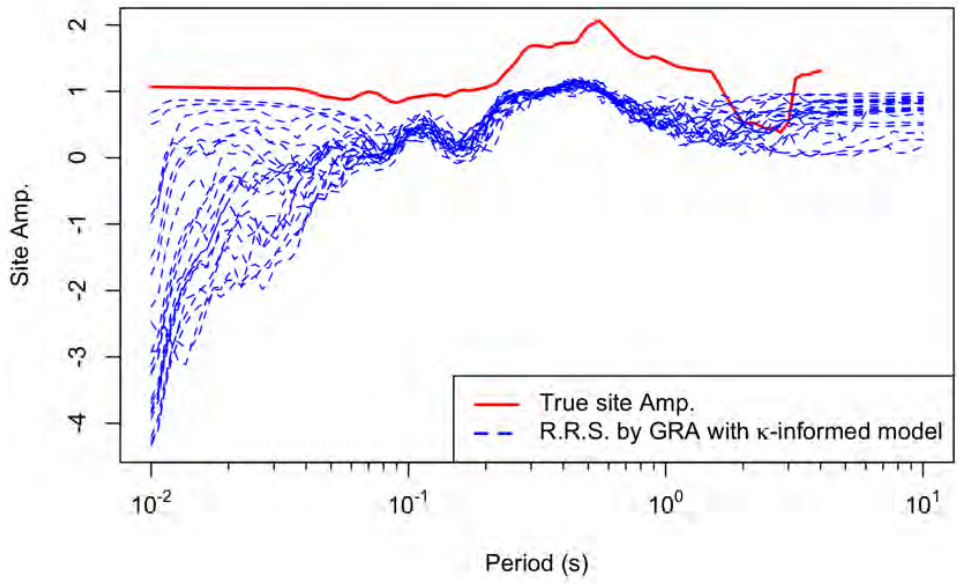


Figure 5.11 Comparison of surface/input transfer functions and response spectral ratios for sites SSN = 100173, SSN = 3046, SSN = 100135, SSN = 3058, SSN = 100047, SSN = 3089.



(a)

80000050



(b)

Figure 5.12. (a) Response spectra for input motions derived from deconvolution procedures; where large soil damping is used (κ-informed model), the spectra become unstable at short periods. (b) Site amplifications computed from ground response analysis using the input motions derived using κ-informed damping model; the rapid fall off at short periods is unrealistic and caused us to screen such sites when compiling summary statistics. The results in this figure apply for site SSN = 80000050.

Similar to the present work, the variation of surface/reference response spectral ratios was previously investigated by Stafford et al. (2017). We describe that work here to place the present results in context. Stafford et al. (2017) relate response spectra to the zeroth spectral moment, m_0 (Boore, 2003b; Bore, et al., 2016):

$$S_a \propto \sqrt{m_0} \quad [5.7]$$

where S_a is the response spectral ordinate. The zeroth spectral moment of surface motion can be expressed as (Parseval's formula),

$$m_{0,\mathrm{Z}}(f_n, \xi_n) = 2\int_0^\infty |X(f|\mathbf{M}, r)|^2 |H(f)|^2 |H_{osc}(f|f_n, \xi_n)|^2 \mathrm{d}f \quad [5.8]$$

where f_n and ξ_n are the oscillator natural frequency and damping, f is frequency in Hz, **M** is the earthquake moment magnitude, r is the source-site distance. The input motion Fourier amplitude is expressed as $|X(f|\mathbf{M}, r)|$. The conditioning on magnitude and distance is provided because Stafford et al. (2017) predict it using point source simulations. $|H(f)|$ is the soil column transfer function amplitude, as before. $H_{osc}(f|f_n, \xi_n)$ is the transfer function of the SDOF oscillator with natural frequency f_n and damping ratio ξ_n. Similarly, the zeroth spectral moment for the reference site condition can be expressed as,

$$m_{0,\mathrm{X}}(f_n, \xi_n) = 2\int_0^\infty |X(f|\mathbf{M}, r)|^2 |H_{osc}(f|f_n, \xi_n)|^2 \mathrm{d}f \quad [5.9]$$

Therefore, the site response expressed as ratio of response spectra is

$$\mathrm{Amp}(f_n, \xi_n) = \frac{S_{a,\mathrm{Z}}(f_n, \xi_n)}{S_{a,\mathrm{X}}(f_n, \xi_n)} = \sqrt{\frac{m_{0,\mathrm{Z}}(f_n, \xi_n)}{m_{0,\mathrm{X}}(f_n, \xi_n)}} \quad [5.10]$$

If the reference site input motions are converted surface motions as descried above, then the zeroth spectral moments for surface and at reference site would be

$$m_{0,\mathrm{Z,im}}(f_n, \xi_n) = 2\int_0^\infty \left|\frac{X(f|\mathbf{M}, r)}{H(f)}\right|^2 |H(f)|^2 |H_{osc}(f|f_n, \xi_n)|^2 \mathrm{d}f \quad [5.11]$$

$$m_{0,\mathrm{X,im}}(f_n, \xi_n) = 2\int_0^\infty \left|\frac{X(f|\mathbf{M}, r)}{H(f)}\right|^2 |H_{osc}(f|f_n, \xi_n)|^2 \mathrm{d}f \quad [5.12]$$

And Eq. [5.11] can be simplified to

$$m_{0,\mathrm{Z,im}}(f_n, \xi_n) = 2\int_0^\infty |X(f|\mathbf{M}, r)|^2 |H_{osc}(f|f_n, \xi_n)|^2 \mathrm{d}f \quad [5.13]$$

since $|H(f)|$ terms are canceled out.

We compare the amplification computed from using surface motions directly applied as input motions ($\mathrm{Amp}_{\mathrm{sr}}(f_n,\xi_n)$) with amplification computed using converted input motions ($\mathrm{Amp}_{\mathrm{im}}(f_n,\xi_n)$) by take the ratio of the two,

$$Ratio(f|f_n,\xi_n)=\frac{\mathrm{Amp}_{\mathrm{sr}}(f_n,\xi_n)}{\mathrm{Amp}_{\mathrm{im}}(f_n,\xi_n)}=\frac{\frac{m_{0,\mathrm{Z}}(f_n,\xi_n)}{m_{0,\mathrm{X}}(f_n,\xi_n)}}{\frac{m_{0,\mathrm{Z,im}}(f_n,\xi_n)}{m_{0,\mathrm{X,im}}(f_n,\xi_n)}}$$
$$=\frac{\left(\int_0^\infty\left|\frac{X(f|\mathbf{M},r)}{H(f)}\right|^2|H_{osc}(f|f_n,\xi_n)|^2\mathrm{d}f\right)*\left(\int_0^\infty|X(f|\mathbf{M},r)|^2|H(f)|^2|H_{osc}(f|f_n,\xi_n)|^2\mathrm{d}f\right)}{\left(\int_0^\infty|X(f|\mathbf{M},r)|^2|H_{osc}(f|f_n,\xi_n)|^2\mathrm{d}f\right)^2} \quad [5.14]$$

Let $|Z_{osc}(f|f_n,\xi_n)|^2=|X(f|\mathbf{M},r)*H_{osc}(f|f_n,\xi_n)|^2$ then Eq. [5.14] can be simplified as,

$$=\frac{\left(\int_0^\infty\frac{|Z_{osc}(f|f_n,\xi_n)|^2}{|H(f)|^2}\mathrm{d}f\right)*\left(\int_0^\infty|Z_{osc}(f|f_n,\xi_n)|^2|H(f)|^2\mathrm{d}f\right)}{\left(\int_0^\infty|Z_{osc}(f|f_n,\xi_n)|^2\mathrm{d}f\right)^2} \quad [5.15]$$

Approximating integration by summation, then Eq. [5.15] becomes

$$Ratio(f|f_n,\xi_n)=\frac{\left(\sum_{i=0}^\infty\frac{|Z_{osc}(f_i|f_n,\xi_n)|^2}{|H(f_i)|^2}\right)*\left(\sum_{i=0}^\infty|Z_{osc}(f_i|f_n,\xi_n)|^2|H(f_i)|^2\right)}{\left(\sum_{i=0}^\infty|Z_{osc}(f_i|f_n,\xi_n)|^2\right)^2} \quad [5.16]$$

Per the Cauchy-Schwarz inequality[2], this ratio is greater than or equal to one. The ratio is unity for the case of $H(f)=1$ for the full frequency range. The ratio increases if $H(f)$ departs from 1. Therefore, use of converted input motions will always decrease the site response computed from GRA relative to direct use of surface records as input motions. This conclusion hold as long as the transfer function $H(f)$ is non-unity for some frequencies within the considered range. As a result, GRAs would over-estimate site response if surface records were used as input motions.

5.2.2 Representative Results

We provide here example results for the same set of six example sites introduced in Section 5.2.1 (Figure 5.13-Figure 5.15). In each case, the observed site response (Eqs [1.1] and [5.6]) is compared to the site response predicted from an ergodic model (Seyhan and Stewart, 2014) and site-specific GRA. The three damping models described in Sections 4.2-4.3 (geotechnical, κ-informed, Q-V_S) are applied in GRA, such that their relative levels of effectiveness can be assessed. Site fundamental periods are shown in the figures as estimated by the simplified Rayleigh method suggested by Urzúa, et al. (2017).

[2] https://en.wikipedia.org/wiki/Cauchy%E2%80%93Schwarz_inequality

The results in Figure 5.13-Figure 5.15 show that the relatively smooth variations of site amplification with period that are predicted by the ergodic model seldom align with observed site responses, which often have local peaks and troughs. Figure 5.13 shows results for two sites (SSNs 100173 and 3046) that visually appear to exhibit good "fit" over a broad period range (i.e., the shapes of the site amplification vs period plots are similar, even if they are shifted relative to each other). Both are predominantly rock sites with shallow surficial layers of soil or weathered rock and large impedance contrasts (apparent from V_S profiles in Figure 5.5-Figure 5.6). These conditions give rise to large amplification near short fundamental periods (~0.1-0.2 s), and rapid fall-off of amplification for longer periods. This condition is well represented with the GRAs, with the κ-informed $D_{\min}$ model best representing short period amplifications for the two sites.

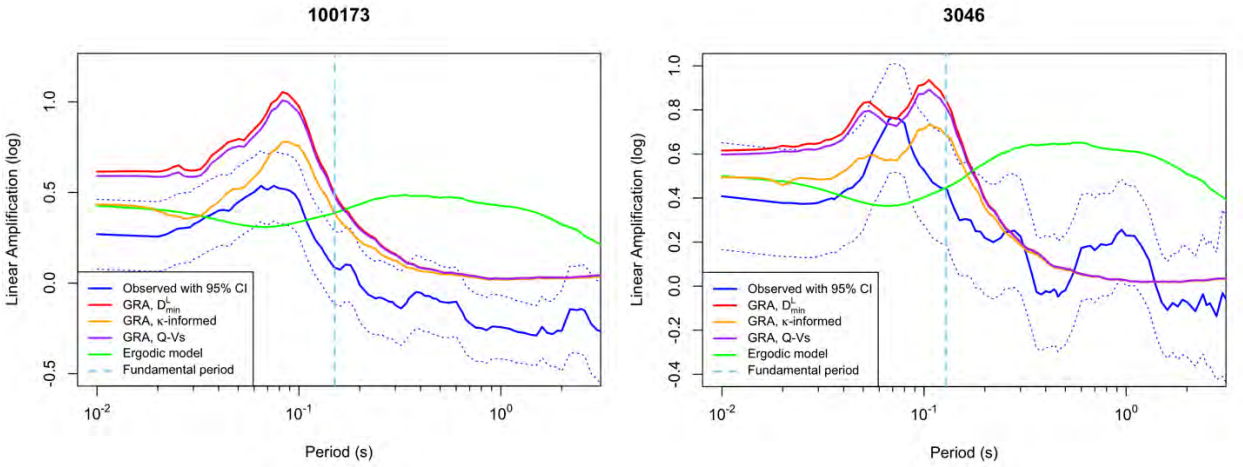


Figure 5.13 Non-ergodic site responses at the site SSN = 100173 and site SSN = 3046, compared with site response predictions obtained with use of ground response analysis and an ergodic model. The maximum period used in the plots is the median of the maximum usable periods from data processing. The ground response model provides a good estimate of the shape of the amplification function in these cases.

The favorable conditions for the two sites depicted in Figure 5.13 are not typical (similar findings of good fit over the full period range occur for around 10% of the 145 sites in the inventory). More commonly, comparisons indicate significant misfits over some or all of the period range. The sites in Figure 5.14 (SSNs 100135 and 3058) exhibit generally good fit of GRA to observation at short periods and poor fit at long periods. These sites have profiles of gradually increasing stiffness with depth to 25 m. A modest impedance contrast occurs for SSN 100135 (Figure 5.7), whereas 3058 has no interface with a large contrast (Figure 5.8). The modelled soil columns produce resonant periods near 0.35 s and 0.25 s, respectively, with amplification shapes up to resonance captured well by GRAs, although the amount of amplification is over-estimated. The misfits occur at longer periods, where the data indicate significant amplification continuing up to overall peaks near about 2.0 sec (100135) and 0.45 sec (3058), whereas the GRAs fall off to zero amplification beyond the modelled soil column periods of ~0.25-0.35 sec. This indicates that the soil portions of these profiles may continue beyond the maximum depths of the site models (30 m). This highlights both a benefit and a limitation of GRA:

- *Benefit*: Despite the incompleteness of the profile, the use of GRA is nonetheless beneficial within the period range of the soil column, relative to the ergodic model
- *Limitation*: GRA cannot capture site response for periods beyond the soil column period, which instead is better represented by the ergodic model (especially for 100135). This conforms with recommendations of Stewart et al. (2014) to only consider results of GRA up to the soil column period, and to transition site amplification to ergodic prediction at longer periods.

Among three $D_{\min}$ models, again κ-informed damping performs best at short periods.

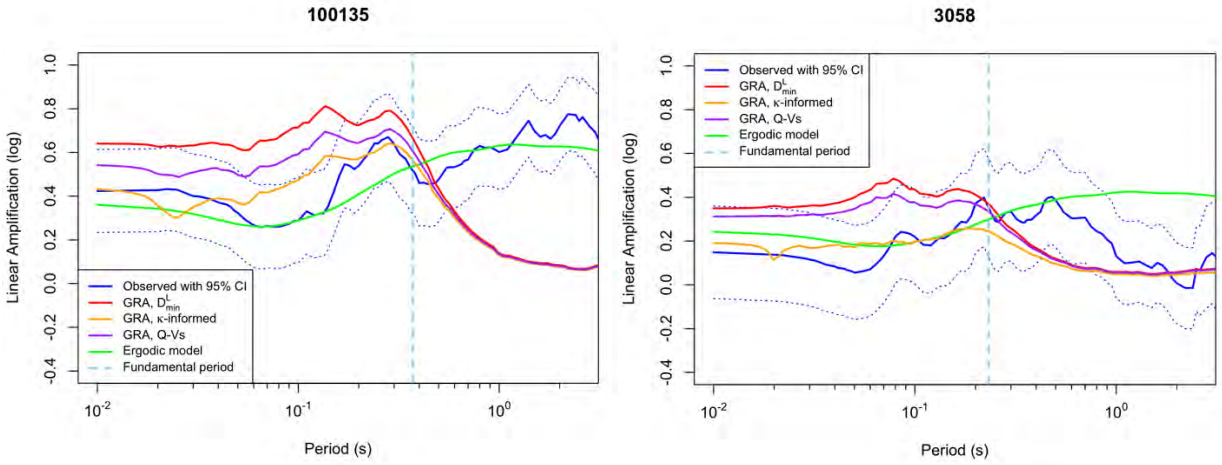


Figure 5.14 Non-ergodic site responses at the sites with SSN = 100135 and 3058, compared with site response predictions obtained with use of ground response analysis and an ergodic model.

The sites in Figure 5.15 (SSNs 100147 and 3089) exhibit poor fit of GRA to observation for all periods. Other than a large velocity step at 4 m for SSN 100147, these sites have profiles of gradually increasing stiffness with depth and no significant impedance contrasts. In both cases, the sites de-amplify ground motion (relative to the reference site condition of V_{S30} = 760 m/s) at short periods (i.e., T < 0.4-1.0 sec), but amplify ground motions at long period with peaks near 2.0 sec. The inability of GRA to model long period site response is similar to that shown in Figure 5.14, and as before, the ergodic model provides improved performance. What distinguishes these sites is the poor GRA performance at short periods, which is caused the de-amplification. De-amplification over at least a log cycle of period (as in Figure 5.15) occurs for around 37% of sites in the inventory, and thus is a relatively common condition. In the case of the sites in Figure 5.15, the reduction may be influenced by attenuation of high frequency motions from damping in deep sediments (e.g., SSN = 100047 site has $z_{1.0}$ = 567.5 m and $z_{2.5}$ = 840.2 m given in NGA-West2 flatfile).

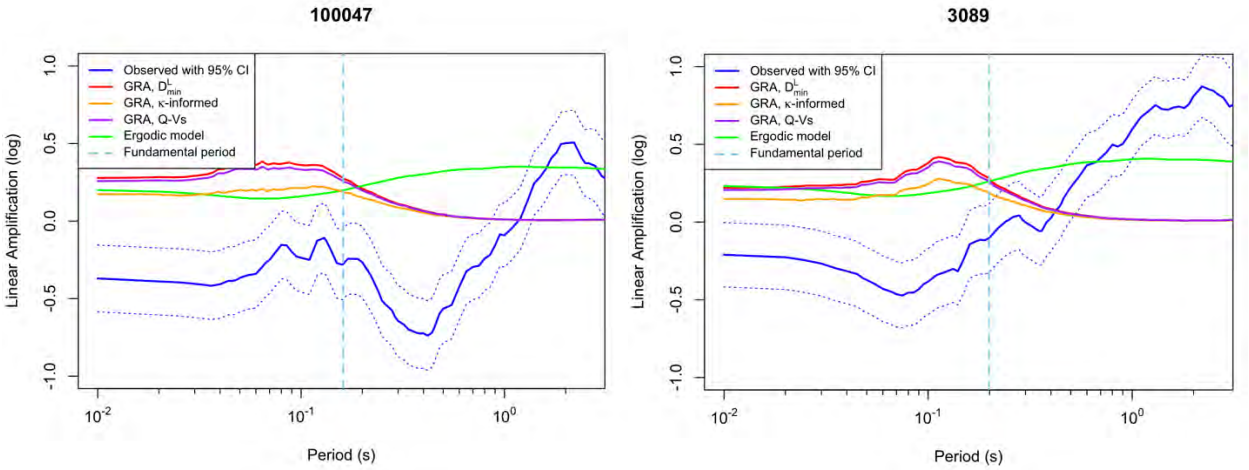


Figure 5.15 Non-ergodic site responses at sites with SSN = 100047 and 3089, compared with site response predictions obtained with use of ground response analysis and an ergodic model.

5.2.3 Goodness of Fit

In last section, we judged goodness-of-fit visually. In judging the quality of the fit, there are two considerations: (1) the general proximity of two site amplification vs period plots and (2) similarity of shapes of the plot. Visual judgements of goodness-of-fit have the drawbacks of being subjective and non-automated (time-intensive). Here we describe alternative quantitative metrics to measure the goodness-of-fit.

Residuals of site amplification defined similarly to Eq. [5.4] (difference between observation and model prediction in natural log units) are useful for individual intensity measures, but not for judging the relative positions of period-dependent quantities. Pearson's sample correlation coefficient r has been used to quantify the overall goodness-of-fit of observed and simulated transfer functions from vertical arrays (Thompson et al., 2012; Afshari and Stewart, 2019). Here we apply the same concept to compare oscillator period-dependent amplification functions. The correlation is computed for a given site j and recording i as,

$$r_{ij} = \frac{\sum_{t\in T}\left(\left(F_S^k\right)_i(t) - \overline{\left(F_S^k\right)_i}\right)\left(\mu_{lnY}(t) - \overline{\mu_{lnY}}\right)}{\sqrt{\sum_{t\in T}\left(\left(F_S^k\right)_i(t) - \overline{\left(F_S^k\right)_i}\right)^2}\sqrt{\sum_{t\in T}\left(\mu_{lnY}(t) - \overline{\mu_{lnY}}\right)^2}} \qquad [5.17]$$

Subscript j applies to all terms and is not shown. Subscript k is an index representing model type (ergodic, GRA with alternate damping models). T represents the period range, t represents a particular period, $t \in T$ (i.e., t can be any possible value in T). $\left(F_S^k\right)_i(t)$ is the site amplification for model k at period t, $\overline{\left(F_S^k\right)_i}$ is the average of $\left(F_S^k\right)_i(t)$ over T. $\mu_{lnY}(t)$ is the observed site amplification for site j (Eq. [5.6]) and $\overline{\mu_{lnY}}$ is the average of $\mu_{lnY}(t)$ over T. For a given site j, there is only one $\mu_{lnY}(t)$, while there are multiple $\left(F_S^k\right)_i(t)$ (one for each recorded earthquake.

The mean value of r across all events, denoted as $\bar{r}$, is used to measure goodness-of-fit. Large positive values indicate good fit.

Table 5.1 shows $\bar{r}$ values for the six example sites over the full period range and for periods up to the fundamental soil column period in both cases of ergodic and GRA-based predictions using the κ-informed D_{min}^{L} damping model. In the case of GRA, for the sites visually identified as "good fit" (Figure 5.13), the correlation coefficients for full range of periods (scenario one) range from 0.86 to 0.94 and for the periods up to fundamental periods (scenario two) range from 0.48 to 0.87, whereas "poor fit" sites (Figure 5.14 and full period range for Figure 5.15) have negative correlation over the full period range and correlations less than 0.24 over the more restrictive (short period) range. In the case of the ergodic model, correlation coefficients for the first two sites are less than those from GRA but are larger at the other sites. In summary, the mean correlation coefficients reflect good performance of GRA for good fit sites and the poor performance of poor fit sites. Correlation coefficient does not reflect what was visually judged as good fit at short periods for the intermediate sites.

Table 5.1 The summary of correlation coefficients for six example sites for both ergodic and GRA

Site	$\bar{r}$ for ergodic over the full period range	$\bar{r}$ for ergodic for the periods up to fundamental periods	$\bar{r}$ for GRA over the full period range	$\bar{r}$ for GRA for the periods up to fundamental periods
SSN = 100173	0.335	-0.948	0.943	0.872
SSN = 3046	-0.205	-0.854	0.862	0.476
SSN = 100135	0.827	0.768	-0.700	0.176
SSN = 3058	0.129	0.755	-0.099	-0.020
SSN = 100047	0.351	0.022	-0.386	0.242
SSN = 3089	0.896	0.974	-0.831	-0.160

A limitation of mean correlation coefficient is that it does not capture misfit related to vertical offsets between amplification functions. For example, the short-period, GRA mean correlation coefficient for site SSN = 100047 are relatively high, whereas vertical offsets are large (about one in ln units). An alternative metric used to quantify the similarity of two curves is Fréchet distance3. As illustrated in Figure 5.16, if a person and a leashed dog walk along separate curved paths, each can change velocity to maintain slack in the leash. The Fréchet distance between the two curves is the length of the shortest leash sufficient for both to traverse their separate paths, under the condition that neither is allowed to move backwards. If the velocities of the person and dog were the same, Fréchet distance would be the longest distance between their position over the duration of the walk. Fréchet distance is reduced by allowing velocities to change.

[3] https://en.wikipedia.org/wiki/Fr%C3%A9chet_distance

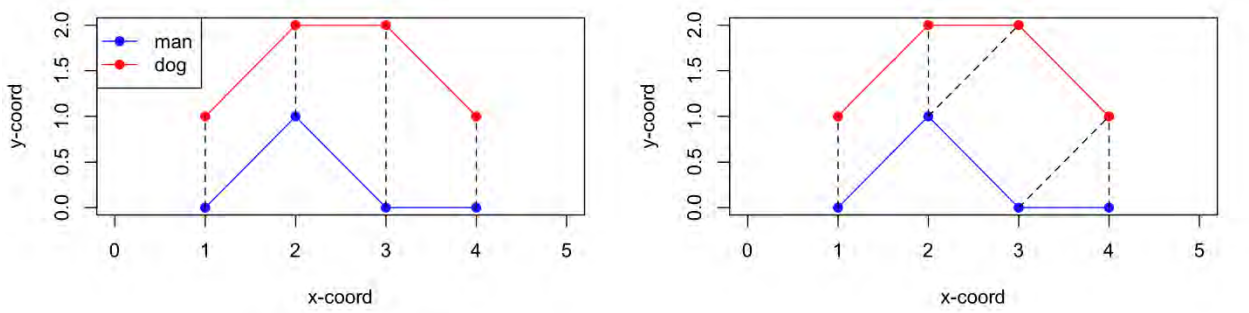


Figure 5.16 Schematic comparison of vertical distance between paths (left, max of 2) and Fréchet distance (right, max of $\sqrt{2}$).

The application of Fréchet distance as a means by which to compare alternate site amplifications can be expressed as,

$$\Psi(\bar{F}_S^k, \mu_{lnY}) = \min_{t_1,t_2 \in T} \max\left\{|\bar{F}_S^k(t_1) - \mu_{lnY}(t_2)|\right\} \quad [5.18]$$

where $\bar{F}_S^k$ is the mean linear site amplification from model k across all events, and t_1 and t_2 are the periods of $\bar{F}_S^k$ and μ_{lnY}. As before, all of the arguments in Eq. [5.18] apply for a given site (i.e., site j). We allow t_1 and t_2 to not match, but the calculation starts from short period and progresses to long periods without moving backwards. The calculation is made using function *Frechet* as provided in library *SimilarityMeasures* (Toohey, 2015) in R. In addition, we also compute Fréchet distance normalized by the range of μ_{lnY} ($\max(\mu_{lnY}) - \min(\mu_{lnY})$).

Table 5.2 and Table 5.3 show Fréchet distances and normalized Fréchet distances for the six example sites for the same model-to-observation combinations considered in Table 5.1. Low Fréchet distances, or normalized Fréchet distances, indicate better fit than large values. Site 100173 has large Fréchet distances, despite the similarity of shapes, due to the vertical shift. For the other sites, Fréchet distances, especially over the period range below the site period, are lower for good fit sites than for poor fit sites (Table 5.2). This trend also holds for normalized Fréchet distances (Table 5.3), except that even 100173 has relatively small values due to its large range. The range of normalized Fréchet distances for "good fit" sites is less than 0.8, whereas the normalized Fréchet distances for "intermediate fit" is around 1, and "poor fit" is greater than 1.5.

Table 5.2 The summary of Ψ for six example sites for both ergodic and GRA

Site	Ψ for ergodic over the full period range	Ψ for ergodic for the periods up to fundamental periods	Ψ for GRA over the full period range	Ψ for GRA for the periods up to fundamental periods
SSN = 100173	0.688	0.602	0.555	0.556
SSN = 3046	0.691	0.392	0.294	0.269
SSN = 100135	0.263	0.168	0.693	0.410
SSN = 3058	0.431	0.127	0.492	0.337
SSN = 100047	1.017	0.581	0.780	0.713
SSN = 3089	0.640	0.640	0.861	0.746

Table 5.3 The summary of normalized Ψ for six example sites for both ergodic and GRA

Site	Normalized Ψ for ergodic over the full period range	Normalized Ψ for ergodic for the periods up to fundamental periods	Normalized Ψ for GRA over the full period range	Normalized Ψ for GRA for the periods up to fundamental periods
SSN = 100173	0.745	0.865	0.602	0.799
SSN = 3046	0.678	0.758	0.289	0.521
SSN = 100135	0.482	0.407	1.267	0.996
SSN = 3058	0.728	0.371	0.833	0.983
SSN = 100047	0.817	1.877	0.627	2.304
SSN = 3089	0.477	1.271	0.641	1.480

5.3 SQUARE ROOT IMPEDANCE

The square-root-impedance (SRI) method, introduced by Joyner et al. (1981) and subsequently revised and further described by Boore (2013), applies principles of wave energy conservation to estimate changes in wave amplitude across layer boundaries. It is widely used in engineering seismology, particularly in combination with semi-stochastic ground motion simulation routines (e.g., Boore, 2003b). Prior validation studies have considered the SRI method in combination with

simulations, but seldom has the method been validated as a stand-alone model. Such validation is undertaken here; we first summarize the methodology and its implementation for this project, and then we present example results for the same set of 6 sites considered in Section 5.2.

5.3.1 Approach

SRI is based on ray theory, whereby conservation of energy requires that the amplitude of motion along a ray path to be inversely proportional to the square root of local seismic impedance ρV, where ρ and V are density and seismic wave velocity, respectively. Therefore, the amplification ratio at a soil site can be expressed as (notation taken from Joyner et al., 1981),

$$A = \left(\frac{\rho_R V_R}{\rho_S V_S}\right)^{1/2} \left(\frac{\cos i_R}{\cos i_S}\right)^{1/2} \quad [5.19]$$

where V_R and V_S are the near-surface velocities for rock and soil respectively, ρ_R and ρ_S are the rock and soil mass densities, and i_R and i_S are vertical angles of incidence (0 indicates vertical propagation). It is common to assume a vertical angle of incidence, which causes the cosine terms to cancel out (per Snell's Law, this may be a reasonable assumption near the ground surface for body waves; e.g., Kramer, 1996).

For a given frequency, amplification is computed using Eq. [5.19] by taking reference rock properties as fixed (independent of frequency) and taking the seismic impedance (product of velocity and density) of surface materials as depth-averaged values measured from the ground surface to a depth corresponding to a quarter wavelength (Boore and Joyner, 1997). Because wavelength is frequency-dependent, depth-averaged impedance and amplification are also frequency-dependent. As in Boore and Joyner (1997), we average seismic velocity and density separately instead of their product being averaged. Site amplification for a frequency of interest f is computed as follows:

1. Identify the reference condition for the calculation at depth z_R; taken here as the top of the last layer in the profile (usually having the fastest velocity). This establishes ρ_R and V_R;
2. Select an arbitrary depth $z < z_R$, and for this depth compute the shear wave travel time $\Delta t_{Vs}(z)$ from surface to depth z;
3. Compute the average density $\overline{\rho_S(z)}$, the average velocity $\overline{V_S(z)} = \frac{z}{\Delta t_{Vs}(z)}$, and the corresponding frequency $f(z) = \frac{1}{4 \times \Delta t_{Vs}(z)}$, which is based on the quarter wavelength;
4. Compute site amplification associated with frequency $f(z)$ as

$$A[f(z)] = \sqrt{\frac{\rho_R V_R}{\overline{\rho_S(z)} * \overline{V_S(z)}}} \quad [5.20]$$

5. Repeat steps 2-4 for different depths $z < z_R$ to develop amplification estimates over the largest possible frequency range. The lowest frequency corresponds to a quarter wavelength equivalent to z_R, which is close to the fundamental-mode resonant frequency of the soil column.

The amplification $A(f)$ obtained from this procedure applies in the frequency domain for the case of zero material damping. The effects of damping are accounted for through application of decay parameter κ (Boore, 2003b), specifically the change in κ from the surface to depth z_R, denoted $\Delta\kappa$. The correction is applied as

$$A_c[f(z)] = A[f(z)] * \exp(-\pi f \Delta\kappa) \quad [5.21]$$

We take $\Delta\kappa$ from the same three sources considered in GRA: (1) empirically estimated $\Delta\kappa$, (2) inferred $\Delta\kappa$ from Eq. [4.8] using $D_{min}^{L}(z)$ in place of $D_{ef}(z)$, and (3) inferred $\Delta\kappa$ from Eq. [4.8] using $\frac{1}{Q_{ef}(z)}$ in place of $\frac{2D_{ef}(z)}{100}$, with $Q_{ef}(z)$ computed from Eq. [4.16].

These procedures estimate the site transfer function amplitude (i.e., ratio of surface-to-reference Fourier amplitudes). The process by which the transfer function is combined with a phase spectrum to compute time series, from which R.R.S. can be computed, is identical to that described in Section 5.2.1.

5.3.2 Representative Results

Example results from SRI calculations are presented here for the same six example sites considered in Section 5.2. Results are presented in Figure 5.17-Figure 5.19 following the same format used for GRA comparisons -- observed site response (Eqs. [1.1] and [5.6]) is compared to an ergodic model prediction and site-specific SRI. The three damping models described in Sections 4.2-4.3 (geotechnical, κ-informed, Q-V_S) are applied in GRA, such that their relative levels of effectiveness can be assessed.

Figure 5.17 shows results for the "good fit" sites (SSNs 100173 and 3046). The fit of the SRI results over the full period range appears to be quite good, similar to the findings from GRA. Figure 5.18-Figure 5.19 show results for the second and third groups of sites. As before, the fits of SRI results to observation generally mirror the trends encountered for GRA. In particular, the fits are uniformly poor for these sites at long periods where SRI (and GRA) does not provide information on site amplification. However, a distinction from GRA is that the κ-informed damping model overestimates short period attenuation relative to alternative damping models, whereas that damping model provided the best outcomes with GRA.

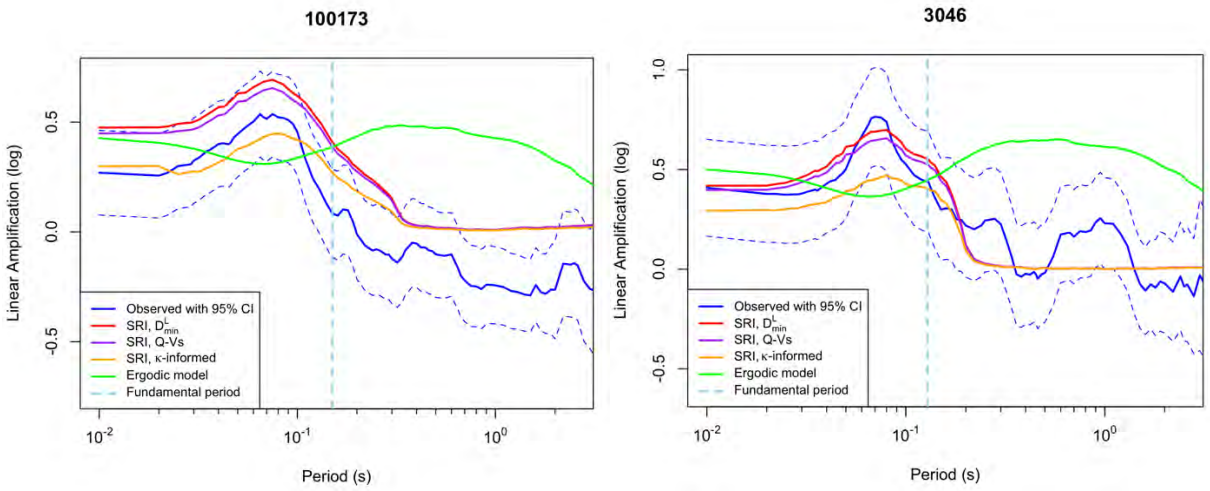


Figure 5.17 Non-ergodic site responses for sites with SSN = 100173 and 3046, compared with site response predictions from ergodic model and from site-specific SRI.

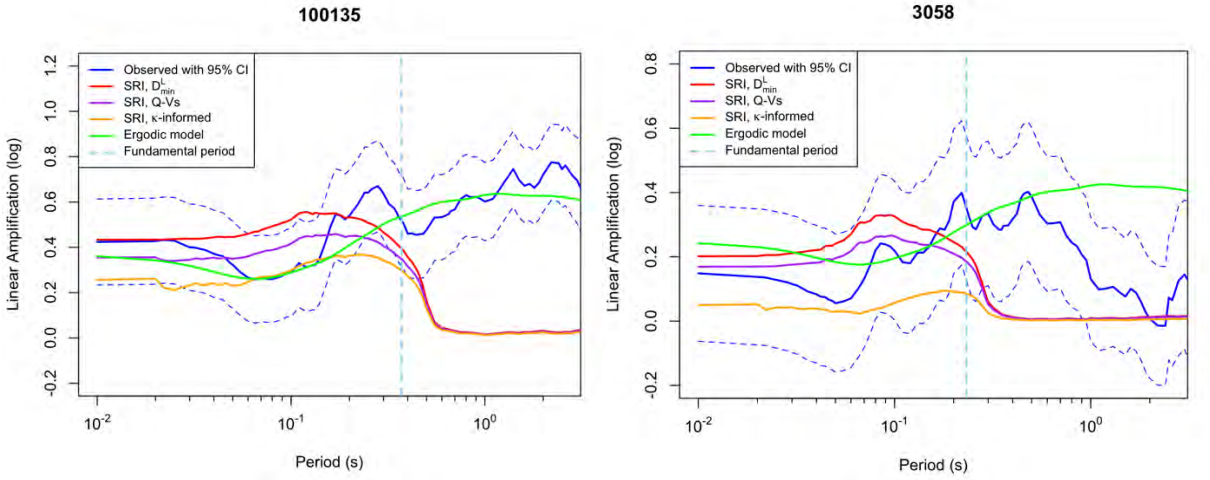


Figure 5.18 Non-ergodic site responses for sites with SSN = 100135 and 3058, compared with site response predictions from ergodic model and from site-specific SRI.

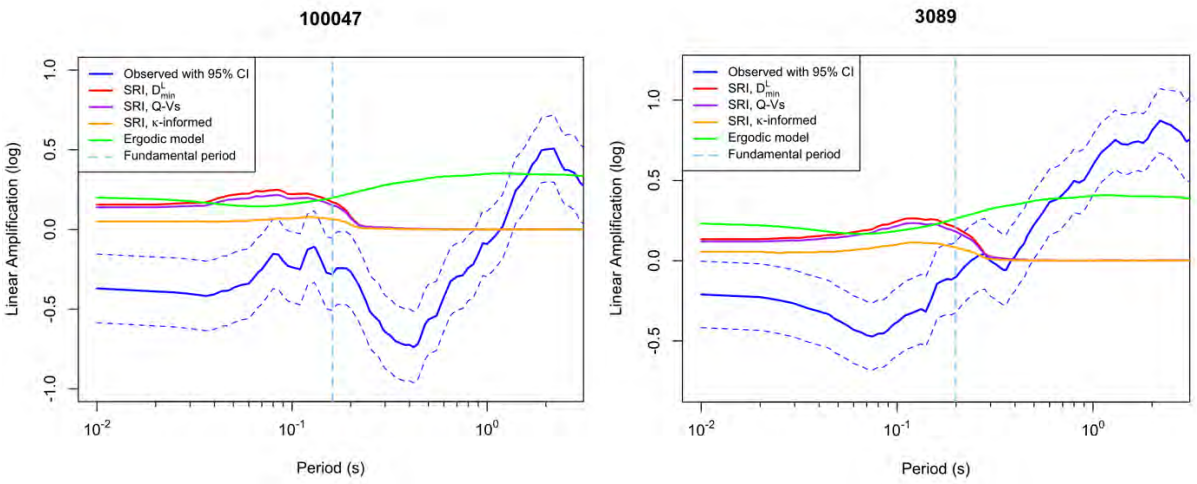


Figure 5.19 Non-ergodic site responses for sites with SSN = 100047 and 3089, compared with site response predictions from ergodic model and from site-specific SRI.

Table 5.4 lists goodness-of-fit metrics of SRI-based predictions using the D^L_{min} damping model for the six example sites for the same observation-to-model combinations considered previously for GRA. Interestingly, shape-based metrics for the good-fit sites are improved for SRI relative to GRA, and Fréchet distances are reduced. The trends of comparisons across sites and for different period ranges are as discussed in Section 5.2.3.

Table 5.4 The summary of Fréchet distance, normalized Fréchet distance, and correlation coefficients for SRI

Site	$\bar{r}$ for SRI over the full period range	$\bar{r}$ for SRI for the periods up to fundamental periods	Ψ for SRI over the full period range	Ψ for SRI for the periods up to fundamental periods	Normalized Ψ for SRI over the full period range	Normalized Ψ for SRI for the periods up to fundamental periods
SSN = 100173	0.967	0.984	0.466	0.350	0.504	0.503
SSN = 3046	0.888	0.909	0.266	0.104	0.260	0.202
SSN = 100135	-0.745	-0.077	0.756	0.236	1.382	0.574
SSN = 3058	-0.195	0.036	0.557	0.209	0.942	0.609
SSN = 100047	-0.257	0.467	0.743	0.586	0.596	1.893
SSN = 3089	-0.838	-0.245	0.870	0.656	0.648	1.301

6 Synthesis of Results Across Sites

6.1 APPROACH

The results for the six sites presented and discussed in Chapter 5 are useful to illustrate the proposed approach of using non-ergodic site response as a means to validate site response models, and to show examples of "good" and "poor" predictions. To fully realize the benefits of this approach for site response model validation, it is necessary to assimilate results over many sites, since many more sites can be used than with the more traditional method of analyzing vertical arrays.

The assimilation of results across multiple sites is undertaken with residuals analyses, using procedures described in Section 5.1. Two of the terms on the right side of Eq. [5.5] carry particular meaning regarding the effectiveness of ground response analysis. Term c_k represents overall model bias of model k (e.g., ergodic models, or site-specific models using GRA or SRI with different damping formulations), which indicates the average misfit across all sites. Term $\eta_{S,j}$ is a random effect indicating how well ground response analyses predict site response for site j relative to the overall model bias. If ground response analyses were unbiased for all sites, both overall bias c_k and site-specific random effects $\eta_{S,j}$ would be zero. Conversely, statistically significant absolute values of $c_k + \eta_{S,j}$ indicate biased predictions of site response. Sites for which problems were identified in deconvolved input motions are removed from the calculation of these summary statistics; this only affects results from the κ-informed damping model (Section 5.2.1).

An important role of site-specific analysis is to predict differences in site response between sites with different properties; this is quantified using the standard deviation of $\eta_{S,j}$. As discussed further below, this dispersion includes effects from modeling uncertainty and soil property uncertainty, and its evaluation is an important objective of this research.

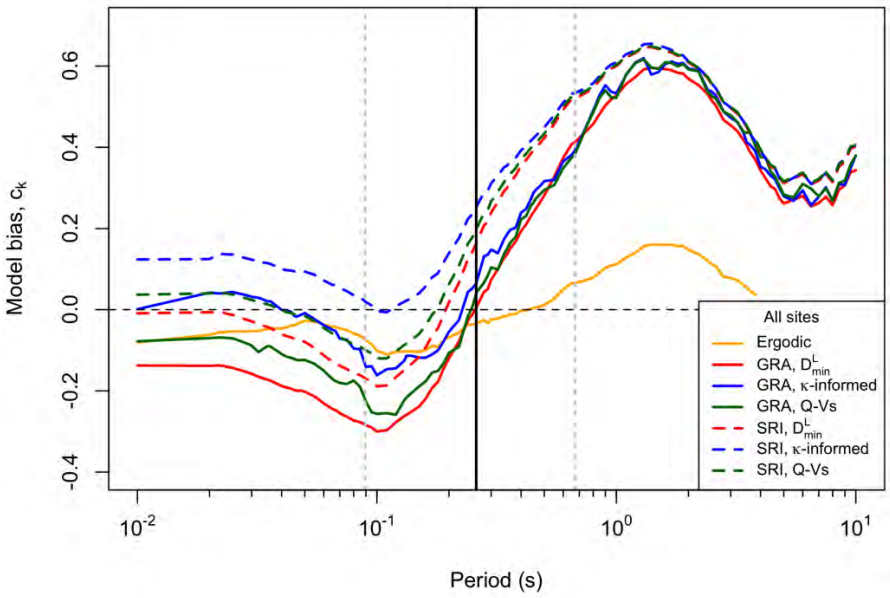
Term ε_{ij} represents the remaining variability after fixed event and site effects are subtracted. The standard deviation of ε_{ij} reflects the effects of path-to-path variability and event-to-event variability in site response for a given site. This variability is commonly referred to as single-station within-event variability, ϕ_{SS} (e.g., Al Atik et al. 2010).

6.2 OVERALL BIAS

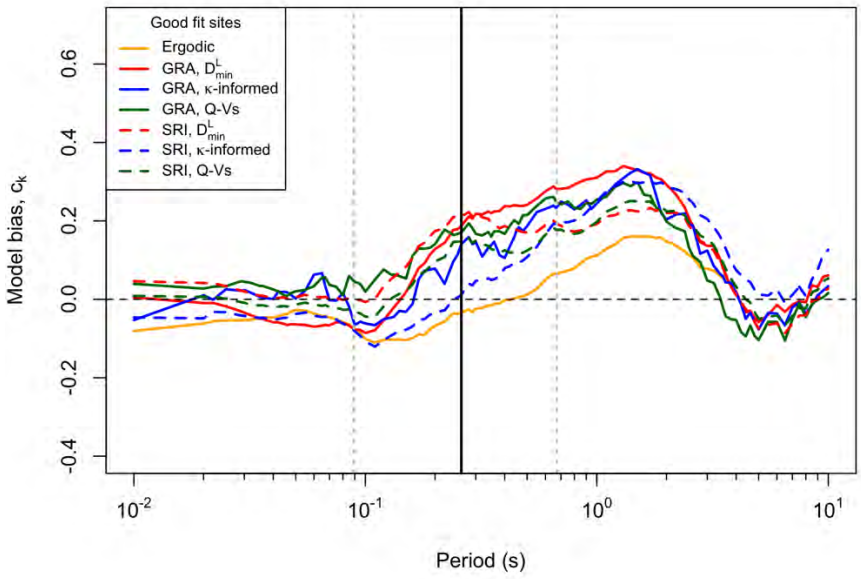
Overall model bias is plotted in Figure 6.1 for GRA and SRI, each with the three considered damping models (geotechnical, κ-informed, Q-V_S), and the Seyhan and Stewart (2014) ergodic model. The 5-95%tile range and median site periods are marked on the figure, which corresponds to the maximum usable period for such analyses. The results in Figure 6.1(a), which apply for all 145 sites, indicate that the site-specific methods exhibit small biases over their usable period range (< ~0.2). Biases for the ergodic model are also small, generally being closer to zero than the site-specific methods. These small biases indicate that the 145 sites considered in this project are effectively unbiased with respect to the much larger data set used to establish the ergodic model.

Among the results in Figure 6.1(a), the κ-informed damping model generally produces the smallest biases, which is consistent with results from Section 5.2. Another notable feature in Figure 6.1(a) is large positive bases for periods beyond the usable range, indicating under-prediction. This is expected, as most sites have amplification at these long periods, whereas GRA and SRI predict effectively zero site response for such conditions. The ergodic model maintains small biases at long periods.

A substantial unanswered questions in ground motion prediction is – *under what conditions do ground response analyses provide reliable estimates of site effects?* This question is addressed in the next section. Another interesting subject, addressed here and subsequent sections, is how aggregated statistics for "good fit" sites compare to those for the data population as a whole. For this purpose, we define "good fit" as normalized Fréchet distances < 0.6 over the usable period range, which applies to about 25 of the 145 sites. As shown in Figure 6.1(b), model biases for these good fit sites are smaller within the usable period range (i.e., T < ~ 0.3 sec).



(a)



(b)

Figure 6.1 Comparison of model bias using GRA and SRI with three damping models and ergodic model for (a) full database of 145 sites and (b) for subset of 25 sites with "good fit", as defined by normalized Fréchet distances < 0.6 over the usable period range. The vertical black solid line represents the median of soil column fundamental periods of 145 sites (as estimated using simplified Rayleigh method in Urzúa, et al., 2017). The two gray dotted lines represent 5 and 95% percentiles.

6.3 PREDICTABILITY OF SITE RESPONSE EFFECTIVENESS

We evaluate four parameters for their potential to identify sites for which ground response analyses reliably predict observed site response. Prediction quality is judged from between-site residuals for individual intensity measures (η_S) and normalized Fréchet distance. Low absolute values of η_S and low values of normalized Fréchet distance indicate good fit. Parameters considered for their ability to identify good fit sites are:

- Velocity ratio, R_V (Afshari et al. 2019): the ratio of time-averaged shear wave velocities over 5 m intervals at the base and top of each array ($R_V = V_{SDH}/V_{S5}$).
- Maximum between-layer amplification over profile, A_{max}: The square root of the impedance ratio (Eq. [5.19]) is computed across all interfaces in an array, and the maximum value is taken as A_{max}.
- Time-averaged shear wave velocity for profile, V_{SZ}
- Shear wave velocity at base of profile, V_{Sb}.
- Profile thickness, z_p.

Figure 6.2 shows trends of between-site residuals as derived from GRA with the geotechnical damping model with each of these parameters. Figure 6.3 shows the same information, but with the κ-informed damping model. Site residuals η_S are shown for PSA at oscillator periods of 0.01 sec (PGA), 0.2 sec, and 0.5 sec. The data are shown with symbols (one per site). We do not necessarily expect a strong trend of the residuals with the parameters on the abscissa, hence no fit line is provided. Rather, we provide binned means and binned standard deviations. A parameter could be considered as having predictive power for GRA effectiveness if either the mean of site residuals trends towards zero for a certain range of that parameter, or if the dispersion of site residuals appreciably decreases for a certain range of the parameter.

Examining the results in Figure 6.2, we do not find an appreciable trend of binned means with any of the considered parameters because the means are generally near zero, hence that indicator of parameter effectiveness does not apply in this case. Moreover, we find no trend of decreasing dispersion with any of the considered parameters. Figure 6.3 illustrates the effectiveness of GRA with κ-informed damping model against the same five parameters. The trends are similar to those shown in Figure 6.2.

One limitation of using site residuals η_S to evaluate model effectiveness is that they only quantify misfits at individual periods, and as such, do not describe fit across a range of periods (i.e., compatibility of shapes). Goodness-of-fit metrics (Section 5.2.3) provide a means by which to describe such cross-period fits. In next section, we show that normalized Fréchet distance represents well curve similarity, so this metric is used here.

Figure 6.4 shows the variation of normalized Fréchet distance with the five site parameters. The Fréchet distance used here is computed over the usable period range for each site (i.e. periods less than site period). Because Fréchet distance is always positive, trends in binned means occur that were not evident with residuals. The trends are towards reduced normalized Fréchet distance as R_V, V_{SZ}, and V_{Sb} increase, indicating improved fit for higher-impedance conditions. As shown in Figure 6.4, log-linear fit lines trend downward, with the slopes and p-values as shown in the figure. All of the slopes are statistically significant, but the slope for the V_{Sb} parameter has the smallest P-value, indicating the most significance.

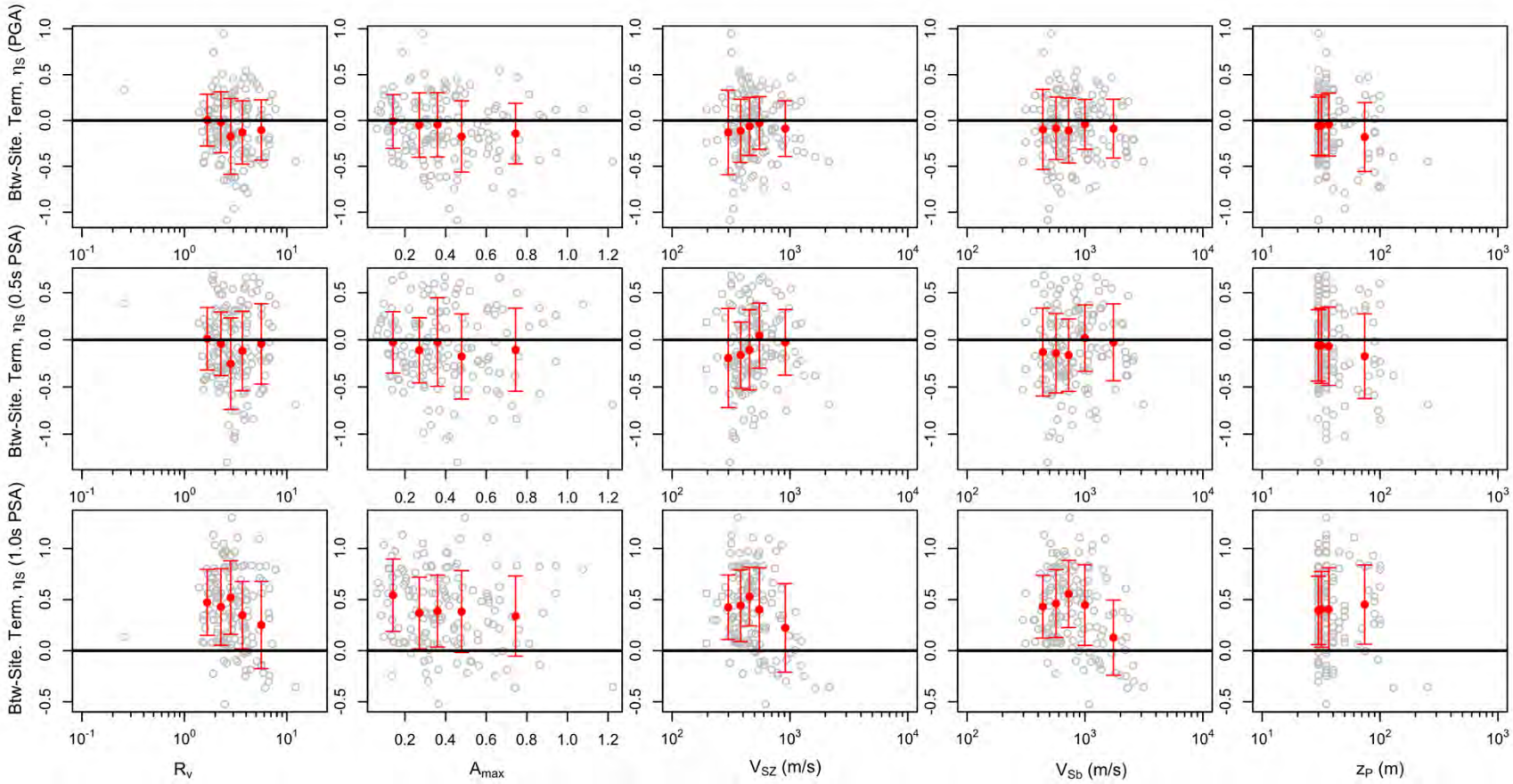


Figure 6.2 Trends of between site residuals (η_S) computed using geotechnical damping model with site parameters R_V, A_{max}, V_{SZ}, V_{Sb}, and z_p

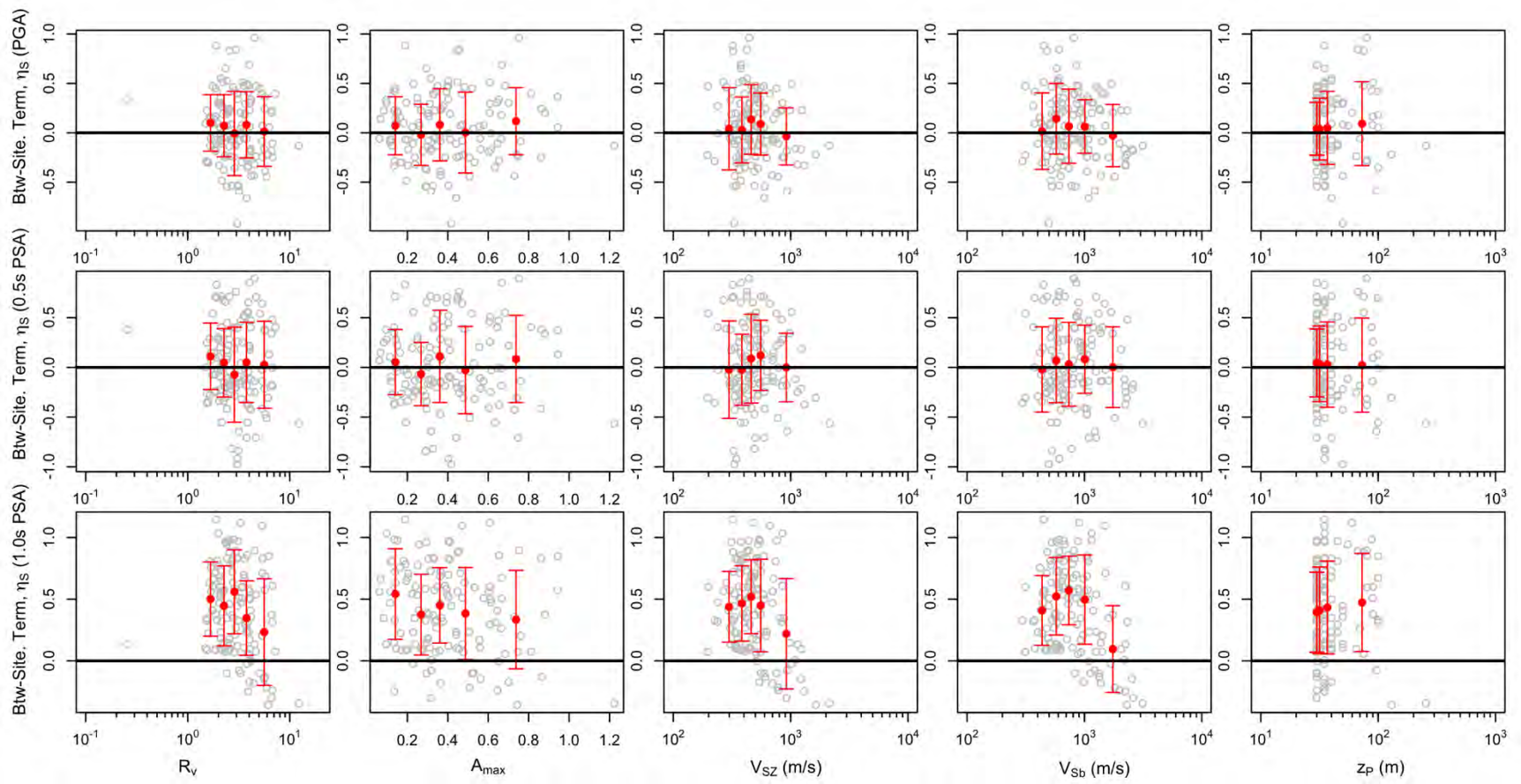


Figure 6.3 Trends of between site residuals (η_S) computed using κ-informed damping model with site parameters R_V, A_{max}, V_{SZ}, V_{Sb}, and z_p.

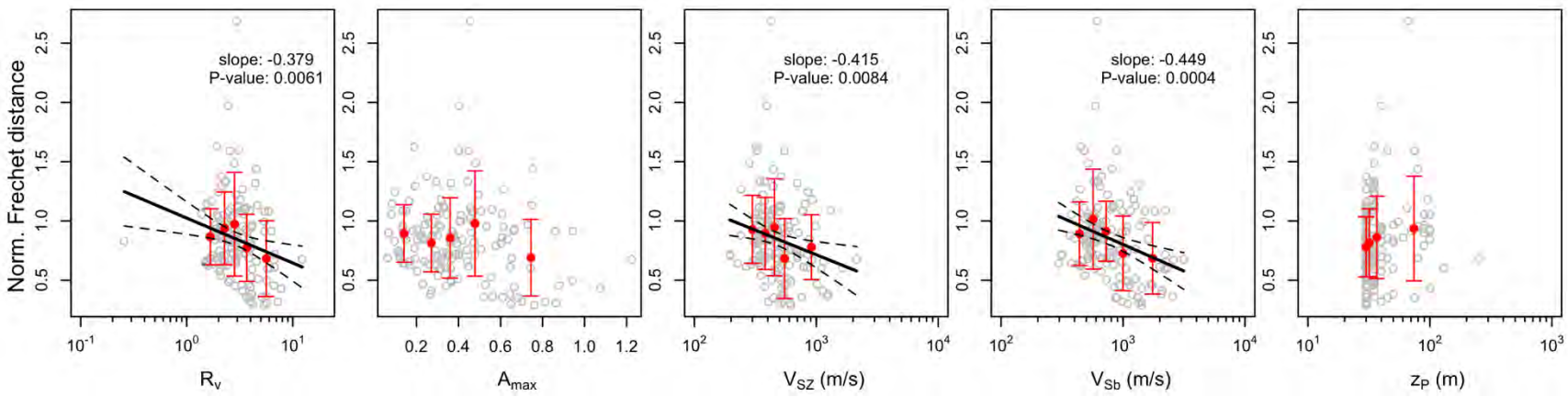


Figure 6.4 Trends of normalized Fréchet distances (Ψ) computed using geotechnical damping model with site parameters RV, V_{S30}, V_{SZ}, and V_{Sb}.

6.4 FULL-POPULATION COMPARISONS OF GOODNESS OF FIT METRICS

As described in Section 5.2.3, we have considered three metrics of multi-period goodness of fit, which are mean correlation coefficient, Fréchet distance, and normalized Fréchet distance. In this section, we compare the effectiveness of these metrics for the full data population.

Our approach is based on a classification of each site as having good, intermediate, or poor fit based on visual inspection. Then we identify the degree to which the various goodness-of-fit metrics are able to distinguish between the three qualitative categories.

Figure 6.5 shows histograms of the three goodness-of-fit metrics. Green bars indicate good fit, blue intermediate fit, and red poor fit. All three metrics perform are able to distinguish the qualitative categories, in the sense that green sites are clustered at low Fréchet distance and high correlation coefficient, whereas red sites have demonstrably higher Fréchet distances and lower correlation. However, neither of the goodness-of-fit metrics are able to perfectly separate the three categories.

Normalized Fréchet distance exhibits the best performance, because it most clearly distinguishes performance categories. Whereas correlation coefficient has mixing of green and blue bars with red bars, normalized Fréchet distance better distinguishes green and blue bars from red. It is also helpful that the range of the parameter is relatively broad.

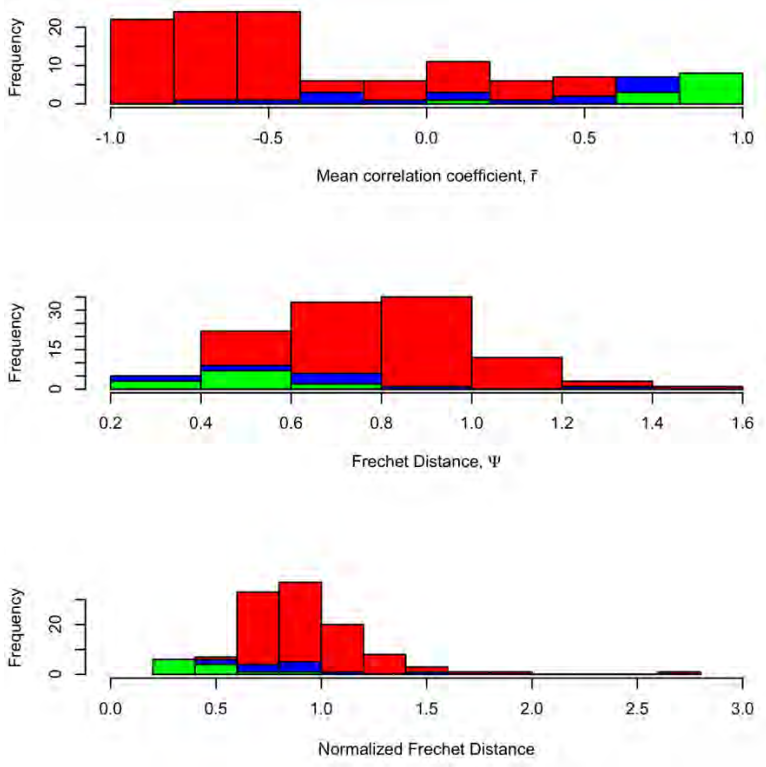


Figure 6.5 Histograms of three metrics for goodness of fit

We use the F-test to evaluate if the three goodness-of-fit metrics are statistically distinct between the three categories. Distinction can be quantified from p-values, in which values less than 0.05 indicate that the three categories exhibit statistically distinct values of the goodness-of-fit metrics. As shown in Table 6.1, none of the metrics provide for statistically distinct categories, but normalized Fréchet distance gives the smallest p-value.

Table 6.1 Summary of F-test for each metric

Metric	*p-value*
Correlation coefficient	0.303
Fréchet distance	0.418
Normalized Fréchet distance	0.180

6.5 SITE RESPONSE UNCERTAINTY AND VARIABILITY

The standard deviations of the random variables (non-constant) terms in Eq. [5.5] are related as follows:

$$\sigma = \sqrt{\tau^2 + (\phi_{S2S})^2 + (\phi_{SS})^2} \qquad [6.1]$$

where σ, τ, ϕ_{S2S}, and ϕ_{SS} are the standard deviations of R_{ij}, η_E, η_S, and ε_{ij}, respectively. Our principal interest is in ϕ_{S2S}, which represents the between-site variability of the misfits in predicted ground motion using GRA or SRI approaches. There are two sources of this variability: (1) modeling uncertainty, which is caused by the analysis method not representing accurately the site response physics, and (2) soil property uncertainty, which occurs if the measured V_S profile and inferred damping profile for the sites have some deviations from actual properties.

Figure 6.6 shows the period-dependence of ϕ_{S2S} when site response is evaluated using site-specific models (GRA and SRI) and the ergodic model. The site specific results are shown for the three considered damping models (geotechnical, κ-informed, Q-V_S). As in Figure 6.1, the range of site periods is shown to indicate useful period range. Figure 6.7 shows the same information but for the subset of "good fit" sites with normalized Fréchet distance < 0.6.

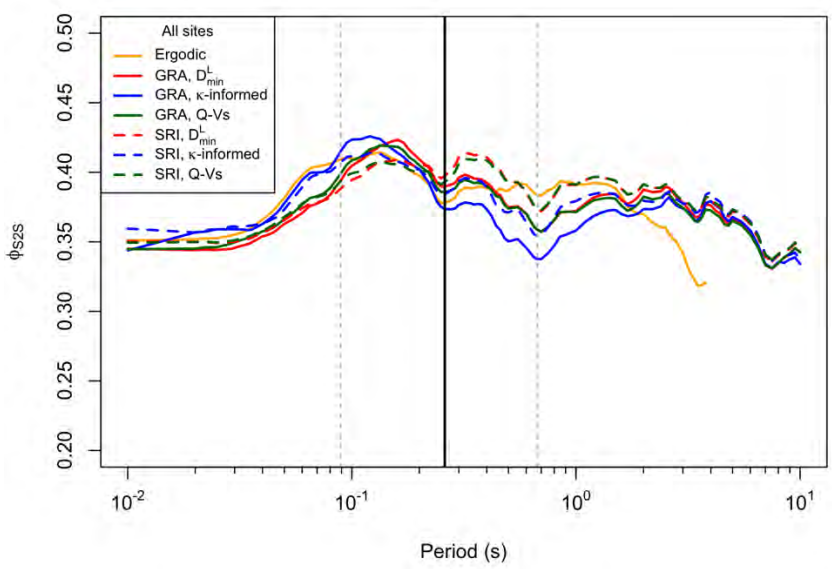


Figure 6.6 Comparison of between-site standard deviations (ϕ_{S2S}) from site-specific analyses (GRA and SRI) using three damping models and ergodic model. Statistics represent results across all sites.

A general observations from Figure 6.6 is that ϕ_{S2S} is not appreciably different for the ergodic model and the various site-specific models. The lack of appreciable ϕ_{S2S} reduction through the use of site-specific methods indicates that such analyses are, on average, not improving ground motion estimates significantly relative to the use of an ergodic model.

Examining Figure 6.6 in greater detail, we find:

(1) With the exception of the κ-informed models, site-specific ϕ_{S2S} values for $T < 0.2$ sec are lower than the ergodic model; these short periods are within the usable period ranges for the sites.

(2) In the period range of 0.2-1 sec, where the data "fall off" as the maximum usable period is exceeded, but also where many peak site responses occur, there are relatively large differences between results among damping models, with κ-informed damping models providing the smallest ϕ_{S2S} values.

(3) At long periods ($T > 1.5$ sec), the ergodic model has lower ϕ_{S2S} than all of the site-specific methods. We do not expect good results from GRA or SRI in this period range.

Part of the reason for the low ϕ_{S2S} reduction in Figure 6.6 is that site responses for many of the sites are poorly represented by ground response analyses, even over the applicable period range of the modelled soil column. However, as shown in Figure 6.7 for the subset of "good fit" sites where relatively shallow ground response is an important driver of site response, we find reductions of ϕ_{S2S} on the order of 0.05-0.10 within the usable period range. These reductions are most consistently achieved from the use of GRA, whereas SRI provides some relatively high dispersions.

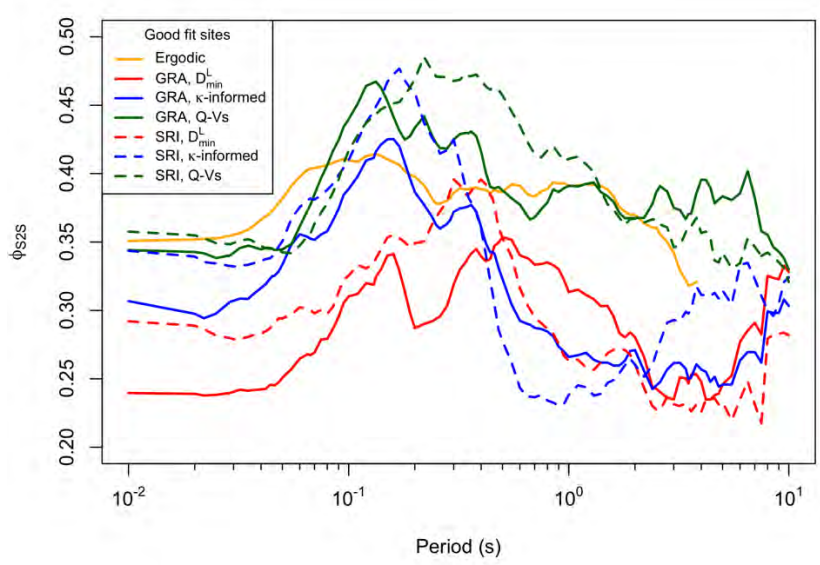


Figure 6.7 Comparison of between-site standard deviations (ϕ_{S2S}) from site-specific analyses (GRA and SRI) using three damping models and ergodic model. Statistics represent results across "good fit" sites with normalized Fréchet distance less < 0.6.

7 Conclusions

7.1 SCOPE OF INVESTIGATION

In this report, we present results of two different, but complimentary lines of research. The first was to establish a United States Community V_S Profile Database (PDB). Primary support for this work during the project period was from USGS, with supplemental support from the Consortium of Organizations for Strong Motion Observation Systems (COSMOS) and the UCLA Civil & Environmental Engineering Department. The second line of research was directed towards developing and applying a proposed new approach for empirical analysis of site effects, and then using those results for validation of various methods of site response prediction. Primary support for this work was jointly from USGS and the California Strong Motion Instrumentation Program, with supplementary support from the UCLA Civil & Environmental Engineering Department.

In Chapter 2 of this report, we have described the PDB schema and the various data types that are included in the PDB. While all sites have a V_S profile, the geophysical methods used to obtain them vary as does the presence of additional data and metadata. Key statistics of the database are summarized at its current stage of development, in which the principal focus has been on California sites. Major data sources are described, some of which had particular features that impacted the database structure. A major component of this work is the development of the PDB schema, the structure of which is described, along with the user interface with the database.

In Chapter 3-6 of this report, we use the PDB to support ground motion studies that establish observation-based site response at ground motion recording stations and then seek to establish the degree to which it can be estimated using alternate prediction approaches. To conduct the study, we began by converting the NGA-West2 data tables to a relational database format, and then supplemented that data with recordings from 25 earthquakes since late 2011 in California. Based on this expanded database, we identified 145 sites that have ample recordings (at least 10 events with **M** >= 4) and an on-site measured V_S profile for use in analysis. We developed protocols for assigning soil parameters when borehole data is limited (including unit weight, geotechnical damping, κ-informed damping, and Q-V_S damping). We then conducted site response analysis using site-specific methods (ground response analysis and square-root impedance analysis), each with three damping models, and with an ergodic model. We present and interpret results for individual sites and synthesize results for the full inventory of sites to investigate the predictability of site response with GRA and SRI, and to investigate epistemic uncertainties associated with the application of GRA or SRI methods.

7.2 SUMMARY OF MAJOR FINDINGS

The development of the United States Community V_S Profile Database (PBD) was a major multi-institutional effort aimed at developing an open-access online and query-able relational database for the uniform dissemination of V_S profile data and additional related information to the public. We have implemented the computer science definition of a "database", consisting of a relational database with tables of data organized into fields (i.e., columns). Each entry in the tables (i.e., in a given row) is assigned a key; the host table assigns primary keys, which can then be found within the database by referring to that entry using the same number as a foreign key in other tables. The relationships between tables, and fields within tables, through primary and foreign keys comprises the database schema.

V_S profile data and other supporting data and metadata was converted to a standardized JavaScript Object Notation (JSON) file format that facilitated efficient data entry and subsequent incorporation into the database. As of December 31 2019, the size of the database stands at 1479 V_S profile sites (1246 in California), with ongoing digitization and integration into the database expected to grow the project to over 5000 V_S profiles (discussed in Section 7.3). Additional information included in the database includes V_P profiles, geotechnical boring logs, penetration resistance measurements, and geospatial data. A website interface allows end-users to visualize data in real time or download it in CSV format, either on an individual site-by-site basis or via queries to the remote server to retrieve data based on criteria specified by the user, such as V_{S30} ranges or data within a specified radius of a location of interest. This project is expected to have a significant impact on both research and practice. The database was used in the ground motion studies that comprises the second phase of work.

The ground motion analysis portion of this research has the broad objective of investigating the effectiveness of ground response analysis, and other methods of site response analysis, through comparisons to observed site response as established from analysis of recordings. Effectiveness is judged, in this context, through bias and site-to-site uncertainty of predicted site response. We describe and illustrate the application of a new approach using non-ergodic site responses derived from surface-only instruments as the basis for validation studies.

The technical contributions of the ground motion analysis work fall into two general categories. First, several procedural matters had to be developed to implement the work, and second, the results provide new and valuable insights into the effectiveness of site-specific site response analysis methods and their associated epistemic uncertainties.

Significant original procedural elements include the following:

i. We have developed procedures to implement ground response analyses and square-root impedance methods given limited available site information (i.e., no boring logs). As described in Chapter 4, we considered two types of available information to assign layer boundaries (stratigraphy) and soil type information – the mapped surface geology and the V_S profile. Surface geology is used to estimate soil type near the ground surface. The V_S

profile is used in combination with the surface unit assignment to estimate variations with depth.

ii. We implemented a κ-informed damping model to leverage ground motion observations from sites as a means to constrain soil material damping. The procedure involves analysis of κ for each recording at a site, followed by subtraction of path κ derived on a record-specific basis given spatially variable path κ coefficients developed in this study. The result is a site κ, which provides a basis for adjustment of soil damping.

iii. We propose new metrics to quantify goodness-of-fit for site response predictions relative to observations. We propose and apply a normalized version of Fréchet distance to supplement mean correlation coefficient. We demonstrate that this metric is effective at identifying sites with and without "good" data to model fits.

iv. We propose "inverse transfer function method" (ITF) to deconvolve surface records to estimate input motions for a reference soil layer for use in GRA (Section 5.2.1). We demonstrate that GRA would over-estimate site amplification if the surface records had been used.

The insights into the effectiveness of site-specific site response analysis methods and their associated epistemic uncertainties can be summarized as:

1. Despite the depth of the profiles considered in this work being relatively small (30 to 255 m; site period ranging from 0.06 to 1.02 sec), ground response analyses (or square-root-impedance analyses) are able to improve site response predictions relative to ergodic models for approximately 17% of sites.
2. The inability of site-specific methods to improve prediction accuracy for the 83% sites could stem from three potential sources: (1) simulations of one-dimension wave propagation do not accurately characterize the physics of site response; (2) the measured V_S profile from the site does not accurately represent site conditions, either because of strong site heterogeneity or inaccurate measurements; (3) portions of the site profile beneath the profile depth significantly impact the site response in the frequency range of the measured profile.
3. The three problems identified in (2) above are common to some extent in virtually all site response simulations, so understanding their collective impact is of practical importance. The unknown influence of these factors introduces epistemic uncertainties, which we quantify. Lacking any knowledge of whether a given site is well represented with one-dimensional simulations, this epistemic uncertainty is similar to that of the site-to-site variability in ergodic models (indicating practically no benefit of site specific analysis). However, for the subset of sites whether this modeling is effective, the epistemic uncertainty is substantially reduced by amounts ranging from 0.05-0.10 in natural log units.
4. Given the significant differences in epistemic uncertainties for sites with unknown vs. known applicability of one-dimensional methods as an effective means of representing site response, the ability to identify this condition a priori is of substantial practical importance. We investigated five potential predictors of ground response analysis effectiveness, most of which represent impedance effects in different ways. Contrary to prior work using a much more limited dataset from vertical arrays, we find that ranges of these parameters

indicating high impedance conditions provide for more effective site response predictability, with arguably the velocity of the base layer (V_{Sb}) having the strongest predictive power. These features were not observable with single-period residuals, but were visible with the Fréchet distance parameter.

5. Damping models informed by site-specific κ perform better than alternative models, as indicated by reduced bias relative to alternative damping models.

Although there is some sensitivity to damping models, in general site-specific ground response analysis is more effective at minimizing epistemic uncertainty than square-root impedance methods.

7.3 RECOMMENDATIONS FOR FUTURE RESEARCH

The V_S PDB is designed and expected to be a living database, not a static repository of data. We do not anticipate stagnation in the growth of the database, and as such there must be a path forward for continuing to obtain and include high-quality V_S data and site metadata to ensure that this product remains vibrant and useful. Besides including additional data, improvements can be made to the web user interface of the V_S PDB. Python scripts have been developed and implemented to rapidly query the database online and perform relevant calculations using Jupyter Notebooks within the NHERI DesignSafe Cyber-Infrastructure (Pérez and Granger, 2007). This is an inherently faster means for users to visualize and manipulate data in the PDB because both the Jupyter Notebooks and the database remain on the server, allowing users to interact with the server computer through cloud services rather than using their own computers or requiring the user to download any data. These could be used to, for example, compute interval velocities, V_{SZ}, and V_{S30} from travel times presented for the USGS SCPT dataset.

Because this project has demonstrated a new means by which to evaluate site response and use it for model validation purposes, the potential future implementations for other data sets are nearly limitless. Additional future work that would benefit calculations of the type performed here include:

1. Validate and improve methods developed in this study to assign soil parameters where the sites do not have boring logs. This can be done by applying the procedures to sites with boring logs.
2. Through more comprehensive data analyses, refine spatial estimates of path κ.
3. Investigate alternative goodness of fit parameters, borrowing from the Fréchet distance concept, but considering an average distance instead of a peak value and considering alternate normalization strategies.
4. Consider an additional site response model that includes site frequency and evaluate its performance in a similar manner to those for GRA and SRI.

Studies of the type performed here for California would benefit from improved site characterization. There has been a trend in recent years for site characterization to focus on the upper 30 m, but for site response studies of the type performed here, deeper characterization to

firm rock layers is especially useful. These investigations should ideally include the development of stratigraphic logs that include detailed soil type descriptions.

References

Afshari, K., Stewart, J.P. (2019) Insights from California Vertical Arrays on the Effectiveness of Ground Response Analysis with Alternative Damping Models. *Bull. Seismol. Soc. Am.*, **109**, 1250-1264.

Afshari, K, Stewart, J.P., Steidl, J.H. (2019). California ground motion vertical array database. *Earthquake Spectra*, **35**, 2003-2015.

Agbabian Associates, Inc. (AA, 1993). Borehole Geophysical Measurements at 24 Soil and Rock Sites, *Report No. 9225-6427 for Electric Power Research Institute*, March 11, 1993.

Ahdi, S.K., Stewart, J.P., Ancheta, T.D., Kwak, D.Y., and Mitra, D. (2017). Development of V_S Profile Database and Proxy-Based Models for V_{S30} Prediction in the Pacific Northwest Region of North America**.** *Bull. Seismol. Soc. Am.*, **107**(4), 1781–1801.

Ahdi, S.K., Sadiq, S., Ilhan, O., Bozorgnia, Y., Hashash, Y.M.A., Kwak, D.Y., Park, D., Yong, Y., and Stewart, J.P. (2018). Development of a United States Community Shear Wave Velocity Profile Database, in *Seismic Hazard Analysis, Earthquake Ground Motions, and Regional-Scale Assessment, Geotechnical Special Publication No. 291 of the Geo-Institute of the American Society of Civil Engineers, Selected Papers from Sessions of the 5th Conference on Geotechnical Earthquake Engineering and Soil Dynamics (GEESD-V)*, eds. S.J. Brandenberg and M.T. Manzari, Austin, TX, USA, 10–13 June 2018, 330-339.

Aki, K. (1957). Space and time spectra of stationary stochastic waves, with special reference to microtremors. *Bulletin of the Earthquake Research Institute*, **35**, 415–457.

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

Anderson, J. G. (1991). A preliminary descriptive model for the distance dependence of the spectral decay parameter in southern California, *Bull. Seism. Soc. Am*. **81**, 2186–2193.

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, D.G. and Woods, R.D. (1975). Comparison of field and laboratory shear modulus. *Proc. ASCE Conf. on In Situ Measurements of Soil Properties*, Vol 1, 69–92.

Anderson, J.G. and Hough, S.E. (1984). A model for the shape of the fourier amplitude spectrum of acceleration at high frequencies, *Bull. Seismol. Soc. Am.*, **74**(5), 1969–1993.

Andrus, R.D. and Stokoe II, K.H. (2000). Liquefaction Resistance of Soils from Shear-Wave Velocity. *J. Geotech. Geoenviron. Eng.*, **126**(11), 1015-1025.

Archuleta, R., Nicholson, C., Steidl, J., Gurrola, L., Alex, C., Cochran, E., Ely, G., and Tyler, T. (1997). UC/CLC Campus Earthquake Program – Initial Source and Site Characterization Studies for the UC Santa Barbara Campus, *Institute for Crustal Studies University of California, Santa Barbara Report No. UCRL-ID-129196*, 92 p.

Archuleta, R., Elgamal, A., Heuze, F., Lai, T., Lavallée, D., Lawrence, B., Liu, P.-C., Matesic, L., Park, S., Riemer, M., Steidl, J., Vucetic, M., Wagoner, J., and Yang, Z. (2000a). UC/CLC Campus Earthquake Program – Strong Earthquake Motion Estimates for Three Sites on the U.C. Riverside Campus, *Report No. UCRL-ID-140522*, 68 p.

Archuleta, R., Bonilla, F., Doroudian, M., Elgamal, A., Heuze, F., Hoehler, M., Lai, T., Lavallée, D., Lawrence, B., Liu, P.-C., Martin, A., Riemer, M., Steidl, J., Vucetic, M., and Yang, Z., (2000b). UC/CLC Campus Earthquake Program – Strong Earthquake Motion Estimates for the UCSB Campus, and Related Response of the Engineering I Building, *Report No. UCRL-ID-138641*, 86 p.

Association of Geotechnical & Geonvironmental Specialists (AGS, 2018). Data Format webpage, available at https://www.ags.org.uk/data-format/ (last accessed 8 November 2018).

ASCE (2016). Minimum Design Loads and Associated Criteria for Buildings and Other Structures, *American Society of Civil Engineers Standard ASCE/SEI 7-16*, Reston, VA.

ASTM (2014). ASTM D4428/D4428M-14: Standard Test Methods for Crosshole Seismic Testing, *ASTM International*, West Conshohocken, PA.

Bates D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 67, 1-48.

Bilderback, E.L., Palmer, S.P., Folger, D.S., Poelstra, J.L., Magsino, S.L., and Niggemann, R.A. (2008). Shear-wave Database for Quaternary and Bedrock Geologic Units, Washington State, *Open File Report 2008*-2, Washington State Department of Natural Resources, Division of Geology and Earth Resources, Olympia, Washington, USA.

Bivand, R.S., Pebesma, E., and Gomez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*, 2nd ed., Springer: New York.

Bivand, R.S., Keitt, T. and Rowlingson, B. (2018). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.3-4. https://CRAN.R-project.org/package=rgdal.

Boore, D.M. (2003a). A compendium of P- and S-wave velocities from surface-to-borehole logging: summary and reanalysis of previously published data and analysis of unpublished data. *U.S. Geologic Survey Open-File Report 2003*–191, 13 p.

Boore, D.M. (2003b). Simulation of ground motion using the stochastic method. *Pure and Applied Geophysics*, **160**, 635-675.

Boore, D.M. (2004). Estimating V_{S30} (or NEHRP site classes) from shallow velocity models (depths < 30 m), *Bull. Seismol. Soc. Am.*, **94**, 591–597.

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. (2013). The uses and limitations of the square-root- impedance method for computing site amplification. *Bull. Seismol. Soc. Am.*, **103**(4), 2356-2368.

Boore, D.M., and Brown, L.T. (1998). Comparing shear-wave velocity profiles from inversion of surface-wave phase velocities with downhole measurements: systematic differences between the CXW method and downhole measurements at six USC strong-motion sites, *Seism. Res. Lett.* **69**, 222–229.

Boore, D.M., Joyner, W.B. (1997). Site amplifications for generic rock sites. *Bull. Seismol. Soc. Am.*, **87**(2), 327-341.

Boore, D.M., Stewart, J.P., Seyhan, E., and Atkinson, G.M, (2014). NGA-West 2 equations for predicting PGA, PGV, and 5%-damped PSA for shallow crustal earthquakes, *Earthquake Spectra*. **30**, 1057–1085.

Boore, D.M. and Thompson, E.M. (2007). On using surface-source downhole-receiver logging to determine seismic slownesses, *Soil Dynamics and Earthquake Engineering* **27**, 971–985.

Boore, D.M., Watson-Lamprey, J., & Abrahamson, N.A. (2006). Orientation-Independent Measures of Ground Motion. *Bull. Seismol. Soc. Am.*, **96**(4), 1502-1511.

Bora, S.S., Scherbaum, F., Kuehn, N., and Stafford, P.J. (2016). On the re- lationship between Fourier and response spectra: Implications for the adjustment of empirical ground-motion prediction equations (GMPEs), *Bull. Seismol. Soc. Am.* **106**(3), 1235–1253.

Bozorgnia, Y., Abrahamson, N.A., Atik, L.A., Ancheta, T.D., Atkinson, G.M., Baker, J.W., Baltay, A., Boore, D.M., Campbell, K.W., Chiou, B.S.-J., Darragh, R., Day, S., Donahue, J., Graves, R.W., Gregor, N., Hanks, T., Idriss, I.M., Kamai, R., Kishida, T., Kottke, A., Mahin, S.A., Rezaeian, S., Rowshandel, B., Seyhan, E., Shahi, S., Shantz, T., Silva, W., Spudich, P., Stewart, J.P., Watson-Lamprey, J., Wooddell, K., Youngs, R. (2014). NGA-West2 Research Project. *Earthquake Spectra*, **30**(3), 973-987.

Brandenberg, S.J., Bellana, N., and Shantz, T. (2010). Shear wave velocity as function of standard penetration test resistance and vertical effective stress at California bridge sites. *Soil Dyn. Earthq. Eng.*, **30**(10), 1026-1035.

Brandenberg, S.J., Kwak, D.Y., Zimmaro, P., Bozorgnia, Y., Kramer, S.L., and Stewart, J.P. (2018). Development of a United States Community Shear Wave Velocity Profile Database, *in Liquefaction Triggering, Consequences, and Mitigation, Geotechnical Special Publication No. 290 of the Geo-Institute of the American Society of Civil Engineers, Selected Papers from Sessions of the 5th Conference on Geotechnical Earthquake Engineering and Soil Dynamics (GEESD-V)*, eds. S.J. Brandenberg and M.T. Manzari, Austin, TX, USA, 10–13 June 2018, 426-433.

Building Seismic Safety Council (BSSC, 2009). *NEHRP Recommended Seismic Provisions for New Buildings and Other Structures (FEMA P-750)*, 2009 edition, Report prepared for the Federal Emergency Management Agency (FEMA), National Institute of Building Sciences, Washington, D.C.

Cabas, A., Rodriguez-Marek, A. and Bonilla, L.F. (2017). Estimation of site-specific Kappa (κ0)-consistent damping values at KiK-net sites to assess the discrepancy between laboratory-based damping models and observed attenuation (of seismic waves) in the field, *Bull. Seism. Soc. Am.* **107**, 2258-2271

California Department of Water Resources (CA DWR, 2009). *Delta Risk Management Strategy, Phase 1 – Executive Summary*. California Department of Water Resources, Sacramento, CA, 32 p.

Campbell, K.W., (2009). Estimates of shear-wave Q and κ0 for unconsolidated and semiconsolidated sediments in Eastern North America. *Bull. Seismol. Soc. Am.*, **99**, 2365-2392.

Campbell, K.W., Chieruzzi, R., Duke, C.M., and Lew, M. (1979). Correlations of Seismic Velocity with Depth in Southern California, *UCLA Engineering Report No. 7965*, 50 p.

Chang, S.W. (1996). Seismic response of deep stiff soil deposits, *Ph.D. Thesis*, Univ. of California, Berkeley.

Chapman, M. C., Talwani, P., and Cannon, R. C. (2003). Ground-motion attenuation in the Atlantic Coastal Plain near Charleston, South Carolina, *Bull. Seismol. Soc. Am.* **93**, 998 – 1011.

Choi, W.K. (2008). Dynamic properties of ash-flow tuffs. *Ph.D. thesis*, Department of Civil Engineering, University of Texas, Austin, Texas.

COSMOS (201x). COSMOS International Guidelines for Applying Noninvasive Surface-wave Methods to Characterize Seismic Site Conditions, in preparation.

Cox, B.R. and Beekman, A.N. (2011). Intramethod variability in ReMi dispersion measurements and V_S estimates at shallow bedrock sites, *J. Geotech. Geoenviron. Eng.*, **137**(4), 354-362.

Darendeli, M.B., (2001). Development of a new family of normalized modulus reduction and material damping curves, *PhD Thesis*, Department of Civil Engineering, University of Texas, Austin, TX.

Data Interchange for Geotechnical and Geoenvironmental Specialists Markup Language (DIGGSML, 2018). Welcome to DIGGS webpage, available at https://www.diggsml.org/ (last accessed 8 November 2018).

Day, S., Doroudian, M., Elgamal, A., Gonzales, S., Heuze, F., Lai, T., Minster, B., Oglesby, D., Riemer, M., Vernon, F., Vucetic, M., Wagoner, J., and Yang, Z. (2002). UC/CLC Campus Earthquake Program – Strong Earthquake Motion Estimates for Three Sites on the U.C. San Diego Campus, *Report No. UCRL-ID-140523*, 59 p.

Dickenson, S.E. (1994). The dynamic response of soft and deep cohesive soils during the Loma Prieta earthquake of October 17, 1989. *Ph.D. Thesis*, Univ. of California, Berkeley.

Doroudian, M. and Vucetic, M. (1997). Development of 3-D Geotechnical Database and its Application to the Evaluation of Nonlinear Site Response and Seismic Zonation, *UCLA Engineering Report No. ENG-97-186*, 355 p.

Duke, C.M. and Leeds, D.J. (1962). Site Characteristics of Southern California Strong-Motion Earthquake Stations, *UCLA Engineering Report No. 62-55*, 134 p.

Duke, C.M., Johnson, J.A., Kharraz, Y., Campbell, K.W., and Malpiede, N.A. (1971). Subsurface Site Conditions and Geology in the San Fernando Earthquake Area, *UCLA Engineering Report No. 7206,* 124 p.

Dutta, U., Biswas, N., Martirosyan, A., Nath, S., Dravinski, M., Papageorgiou, A., and Combellick, R. (2000). Delineation of spatial variation of shear wave velocity with high-frequency Rayleigh waves in Anchorage, Alaska, *Geophys. J. Int.*, **143**, 365-375.

Eberhart-Phillips, D. (2016). Northern California seismic attenuation: 3D QP and QS models. *Bull. Seismol. Soc. Am.* **106**, 2558-2573.

Eguchi, R.T., Campbell, K.W., Duke, C.M., Chow, A.W., and Paternina, J. (1976). Shear Velocities and Near-Surface Geologies at Accelerograph Sites that Recorded the San Fernando Earthquake, *UCLA Engineering Report No. 7653,* 72 p.

Farr, T.G., and Kobrick, M. (2000). Shuttle Radar Topography Mission produces a wealth of data. *EOS* **81**, 583-585.

Fishburn, K.A., Davis, L.R., and Allord, G.J. (2017). Scanning and georeferencing historical USGS quadrangles: *U.S. Geological Survey Fact Sheet 2017–3048,* 2 p., https://doi.org/10.3133/fs20173048.

Foti, S., Parolai, S., Albarello, D., and Picozzi, M. (2011). Application of Surface-Wave Methods for Seismic Site Characterization, *Surveys in Geophysics* **32**, 777-825.

Fumal, T.E. (1991). A Compilation of the Geology and Measured and Estimated Shear-Wave Velocity Profiles at Strong-Motion Stations that Recorded the Loma Prieta, California, Earthquake, *U.S. Geological Survey Open-File Report 91-311*, 174 p.

Fumal, T.E. and Tinsley, J.C. (1985). Mapping shear wave velocities of near surface geologic materials, *Prof. Paper 1360,* US Geological Survey.

GEOVision, Inc. (2000a). UCLA/Caltrans shaft test site, boreholes B-1, B-2, and B-3, Suspension P & S velocities. *GEOVision Project No. 0368. Report to Jonathan Stewart*, 9 October 2000.

GEOVision, Inc. (2000b). Near-Field Strong Motion Site Investigations at Selected California Sites, Volume 1 Data Report *GEOVision Report No. 9257-001. Prepared for Kajima Corp.*, 27 January 2000, 157 p.

GEOVision, Inc. (2004). UCLA Botanical Garden, Boring B-1, Suspension P & S velocities. *GEOVision Report No. 4438-01. Prepared for Kinemetrics, Inc.*, 30 July 2004.

Gibbs, J.F., Fumal, T.E., and Borcherdt, R.D. (1977). In-situ measurements of seismic velocities in the San Francisco Bay region...Part III, *U.S. Geological Survey Open-File Report 77-850,* 143 p.

Gibbs, J.F., Fumal, T.E., Boore, D.M., and Joyner, W.B. (1992). Seismic velocities and geologic logs from borehole measurements at seven strong-motion stations that recorded the Loma Prieta earthquake, *U.S. Geological Survey Open File Report 92-287,* 142 p.

Gibbs, J.F., Tinsley, J.C., Boore, D.M., and Joyner, W.B. (1999). Seismic Velocities and Geological Conditions at Twelve Sites Subjected to Strong Ground Motion in the 1994 Northridge, California, Earthquake: A Revision of OFR 96-740, *U.S. Geological Survey Open-File Report 99-446,* 148 p.

Gibbs, J.F., Tinsley, J.C., Boore, D.M., and Joyner, W.B. (2000). Borehole velocity measurements and geological conditions at thirteen sites in the Los Angeles, California region, *U.S. Geological Survey Open-File Report 00-470*, 118 p.

Gibbs, J.F., Boore, D.M., Tinsley, J.C., and Mueller, C.S. (2001). Borehole P- and S-wave velocity at thirteen stations in southern California, *U.S. Geological Survey Open-File Report 01-506*, 117 p.

Gibbs, J.F., Tinsley, J.C., and Boore, D.M. (2002). Borehole velocity measurements at five sites that recorded the Cape Mendocino, California earthquake of 25 April, 1992, *U.S. Geological Survey Open-File Report 02-203*, 48 p.

Gospe, T., Ahdi, S.K., Yong, A., Zimmaro, P., Wang, P., Buckreis, T., and Stewart, J.P. (2020). Supplementing Shear Wave Velocity Profile Database with Microtremor-Based H/V Spectral Ratios, *Proc. 17th World Conf. on Earthquake Eng. (17WCEE)*, September 13–18, 2020, Sendai, Japan, *abstract accepted*.

Hansen, W.R., Weiss, R.B., Idriss, I.M., and Cluff, L.S. (1973). Geotechnical data compilation for selected strong-motion seismograph sites in California, *Report to National Oceanic and Atmospheric Administration, Woodward-Lundgren & Assoc*., Oakland, CA, 368 p.

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

Hashash, Y.M.A., Musgrove, M.I., Harmon, J.A., Groholski, D.R., Phillips, C.A., and Park, D. (2016). DEEPSOIL 6.1, User Manual. Urbana, IL, Board of Trustees of University of Illinois at Urbana-Champaign.

Hauksson, E., and Shearer, P. (2006). Attenuation models (QP and QS) in three dimensions of the southern California crust: Inferred fluidsaturation at seismogenic depths, *J. Geophys. Res.* **111** (B05302).

Heisey, J.S., Stokoe II, K.H., Hudson, W.R., and Meyer, A.H. (1982). Determination of in situ shear wave velocities from Spectral Analysis of Surface Waves, *Research Report No. 256-2, Center for Transportation Research*, The University of Texas at Austin, 277 p.

Heuze, Archuleta, R., Bonilla, F., Day, S., Doroudian, M., Elgamal, A., Gonzales, S., Hoehler, M., Lai, T., Lavallee, D., Lawrence, B., Liu, P.-C., Martin, A., Matesic, L., Minster, B., Mellors, R., Oglesby, D., Park, S., Riemer, M., Steidl, J., Vernon, F., Vucetic, M., Wagoner, J., and Yang, Z. (2004). Estimating Site-Specific Strong Earthquake Motions, *Soil Dynamics and Earthquake Engineering* **24**, 199–223.

Hijmans, R.J. (2018). raster: Geographic Data Analysis and Modeling. R package version 2.7-15. https://CRAN.R-project.org/package=raster.

Holzer, T.L., Noce, T.E., and Bennett, M.J. (2010). Maps and documentation of seismic CPT soundings in the central, eastern, and western United States, *U.S. Geological Survey Open-File Report 2010*-1136, 10 p.

Horike, M. (1985). Inversion of the phase velocity of long-period microtremors to the S-wave-velocity structure down to the basement in urbanized areas. *J. Phys. Earth*, **33**, 59-96.

Hough, S.E., and Anderson, J.G. (1988). High-frequency spectra observed at Anza, California: Implications of Q structure. *Bull. Seismol. Soc. Am.,* **78**, 692–707.

Idriss, I.M. (1993). Assessment of site response analysis procedures. *Report to National Institute of Standards and Technology*, Gaithersburg, Maryland, Dept. of Civil & Environmental Eng., Univ. of California, Davis.

Idriss, I.M., (2011). Use of Vs30 to represent local site conditions, in *4th LASPEI/IAEE International Symposium Effects of Surface Geology on Strong Ground Motions*, Santa Barbara, CA.

Ishihara, K. (1996). *Soil Behaviour in Earthquake Geotechnics*, Clarendon Press.

Iskandar, V.E., Doroudian, M., and Vucetic, M. (1996). Development of geotechnical database for Palo Alto and its utilization for seismic microzoning, *Proceedings of the Eleventh World Conference on Earthquake Engineering (11WCEE)*, Acapulco, Mexico, paper no. 1277.

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

Iwahashi, J., Kamiya, I., Matsuoka, M., and Yamazaki, D. (2018). Global terrain classification using 280 m DEMs: segmentation, clustering, and reclassification, *Progress in Earth and Planetary Science*, **5**(1), DOI 10.1186/s40645-017-0157-2.

Joyner, W.B., Warrick, R.E., Fumal, T.E. (1981). The effect of Quaternary alluvium on strong ground motion in the Coyote Lake, California, earthquake of 1979. *Bull. Seismol. Soc. Am.,* **71**(4), 1333-1349.

Kaklamanos, J., Bradley, B.A., Thompson, E.M., Baise, L.G.. (2013). Critical parameters affecting bias and variability in site-response analyses using KiK-net downhole array data. *Bull. Seismol. Soc. Am.,*, **103**, 1733–1749.

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

Kanai, K., Tanaka, T., and Osada, K. (1954). Measurement of the microtremor. *Bulletin of the Earthquake Research Institute of Tokyo*, **32**, 199-209 (*in Japanese*).

Kästli, P. and Euchner, F. (2018). QuakeML 2.0: Data model on site characterization, presentation at the *COSMOS Guidelines Workshop at the 36th General Assembly of the European Seismological Commission (36th ESC-GA)*, Valletta, Malta, 6 September 2018.

Kayen, R.E., Carkin, B., Minasian, D., and Tinsley, J. (2005a). Shear wave velocity of the ground near southern California TRINET sites using the Spectral Analysis of Surface Waves Method (SASW) and parallel-arrayed harmonic-wave sources. *U.S. Geological Survey Open-File Report 2005*-1365, 47 p.

Kayen, R.E., Thompson, E.M., Minasian, D., and Carkin, B. (2005b). Shear-wave velocity of the ground near sixty California strong motion recording sites by the Spectral Analysis of Surface

Waves (SASW) Method and harmonic-wave sources. *U.S. Geological Survey Open-File Report 2005*-1366, 132 p.

Kayen, R.E., Moss, R.E.S., Thompson, E.M., Seed, R.B., Cetin, K.O., Der Kiureghian, A., Tanaka, Y., and Tokimatsu, K. (2013). Shear-wave velocity–based probabilistic and deterministic assessment of seismic soil liquefaction potential. *J. Geotech. Geoenviron. Eng.*, **139**(3), 407-419.

Kim, B., Hashash, Y.M.A., Stewart, J.P., Rathje, E.M., Harmon, J.A., Musgrove, M.I., Campbell, K.W., and Silva, W.J. (2016). Relative differences between nonlinear and equivalent-linear 1D site response analyses. *Earthq. Spectra*, **32**, 1845–1865.

Kottke, A.R. (2012). *Unify Manual*, last modified 11 April 2012.

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

Kramer, S.L. (1996). *Geotechnical Earthquake Engineering*. Prentice Hall.

Kuehn, N.M., Abrahamson, N.A., Walling, M.A. (2019). Incorporating Nonergodic Path Effects into the NGA-West2 Ground-Motion Prediction Equations. *Bull. Seismol. Soc. Am.,* 109 (2), 575–585.

Kwak, D.Y., Ancheta, T.D., Mitra, D., Ahdi, S.K., Zimmaro, P., Parker, G.A., Brandenberg, S.J., and Stewart, J.P. (2017). Performance evaluation of V_{SZ}-to-V_{S30} correlation methods using global V_S profile database, *3rd International Conference on Performance-based Design in Earthquake Geotechnical Engineering (PBD-III)*, Vancouver, Canada, July 16-19, 2017, paper No. 399.

Kwok, A.O.L., Stewart, J.P., Kwak, D.Y., Sun, P.-L. (2018). Taiwan-specific model for VS30 prediction considering between-proxy correlations, *Earthquake Spectra*, **34**, 1973-1993.

Li, Q., Woolery, E.W., Crawford, M.M., Vance, D.M. (2013). Seismic Velocity Database for the New Madrid Seismic Zone and Its Vicinity, *Kentucky Geological Survey* 0Information Circular 27, Series 12, University of Kentucky, Lexington, 19 p.

Louie, J.N. (2001). Faster, better: Shear-wave velocity to 100 meters depth from Refraction Microtremor Arrays, *Bull. Seimol. Soc. Am.*, **91**(2), 347-364.

Lunne, T., Robertson, P.K., and Powell, J.M.M. (1997). *Cone Penetration Testing in Geotechnical Practice*, Spon Press, Abingdon, U.K., 352 p.

Mari, J.L. (1984). Estimation of static corrections for shear-wave profiling using the dispersion properties of Love waves, *Geophysics*, **49**, 1169–1179.

Menq, F., (2003). Dynamic properties of sandy and gravelly soils, *PhD Thesis*, Department of Civil Engineering, University of Texas, Austin, TX.

Minster, B., Wagoner, J., Mellors, R., Day, S., Park, S., Elrick, S., Vernon, F., and Heuze, F. (1999). UC/CLC Campus Earthquake Program – Initial Source and Site Characterization

Studies for the U.C. San Diego Campus, *Institute of Geophysics and Planetary Physics, University of California, Riverside, Report No. UCRL-ID-134785*, 81 p.

Mitchell, J.K., Lodge, A.L., Coutinho, R.Q., Kayen, R.E., Seed, R.B., Nishio, S., Stokoe II, K.H. (1994). In situ Test Results from Four Loma Prieta Earthquake Liquefaction Sites: SPT, CPT, DMT, and Shear Wave Velocity, *UC Berkeley Earthquake Engineering Research Center, Report No. UCB/EERC-94/04, 186 p.*

NCREE (2017). Site Database for Taiwan Strong Motion Stations, *National Center for Research on Earthquake Engineering, Taiwan, Report No. NCREE-17-004*, 86 p (in Chinese).

NIED (2009). A study on subsurface structure model for deep sedimentary layers of Japan for strong-motion evaluation, *National Research Institute for Earth Science and Disaster Prevention Technical Note 337*, 222 p.

Nigbor, R.L., Swift, J.N., and Diehl, J.G. (2001). Resolution of Site Response Issues in the Northridge Earthquake (ROSRINE) – Data Collection, Processing and Dissemination from Phase 5b Field and Laboratory Investigations, *University of Southern California Report No. ROSRINE/USC-01-Rev. 0.9,* 26 November 2001, 144 p.

NIST (2012). Soil-structure interaction for building structures, Report No. NIST GCR 12-917-21, National Institute of Standards and Technology. U.S. Department of Commerce, Washington D.C. Project Technical Committee: Stewart, J.P. (Chair), C.B. Crouse, T. Hutchinson, B. Lizundia, F. Naeim, and F. Ostadan.

NZGD (2017). New Zealand Geotechnical Database webpage, https://www.nzgd.org.nz/ (last accessed 8 November 2018).

Ohta, Y. and Goto, N. (1978). Empirical Shear Wave Velocity Equations in Terms of Characteristic Soil Indexes. *Earthquake Engineering and Structural Dynamics*, **6**, 167-187.

Okada, H., Matsushima, T., Moriya, T., and Sasatani, T. (1990). An exploration technique using long-period microtremors for determination of deep geological structures under urbanized areas. *Butsuri-Tansa (Geophys. Explor.)*, **43**, 402-417 (*in Japanese*).

Park, C.B., Miller, R.D., and Xia, J. (1999a). Multichannel analysis of surface waves. *Geophysics*, **64**(3), 800-808.

Park, S.K., Elrick, S., and Funk, R. (1999b). UC/CLC Campus Earthquake Program – Initial Source and Site Characterization Studies for the U.C. Riverside Campus, Institute of Geophysics and Planetary Physics, University of California, Riverside, *Report No. UCRL-ID-134610*, 81 p.

Pérez, F. and Granger, B.E. (2007). IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering*, **9**(3), 21-29.

PG&E (2010). Evaluation of Shear-Wave Velocity at the DCPP ISFSI. *Pacific Gas and Electric Company, Geosciences Department, Calculation Document, Calc. Number: GEO.DCPP.10.02, 27* April 2010, 11 p.

Pinheiro J., Bates D., DebRoy S., Sarkar D., R Core Team (2019). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-140.

Poran, C.J., Rodriguez, J.A., Satoh, T., and Borden, R. (1994). A new interpretation method of surface wave measurements to obtain representative shear wave velocity profiles of soils, in: T.D. O'Rourke and M. Hamada, eds, *Proceedings from the fifth U.S.- Japan workshop on earthquake resistant design and countermeasures against soil liquefaction*, National Center for Earthquake Engineering Research Technical Report NCEER-94-0026, 425-443.

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

Pradel, D., Smith, P.M., Stewart, J.P., and Raad, G. (2005). Case History of Landslide Movement during the Northridge Earthquake, *Journal of Geotechnical and Geoenvironmental Engineering*, **131**(11), 1360-1369.

Pratt, P.J. and Last, M.Z. (2014). *Concepts of Database Management*, 8th ed., Cengage Learning: Boston, MA, 432 pp.

R Core Team (2018). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0. Available online at: http://www.R-project.org.

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**(3).

Redpath, B.B. (1973). Seismic refraction exploration for engineering site investigations, U. S. Army Engineer Waterway Experiment Station (USAEWES) Explosive Excavation Research Laboratory (EERL), Livermore, California, *Technical Report E-73-4*, 51 p.

Redpath, B.B. (1991). "Seismic Velocity Logging in the San Francisco Bay Area", *EPRI Agreement No. RP3014-06, for Electric Power Research Institute, Redpath Geophysics*, Murphys, California, August, 1991.

Reinert, E., Stewart, J.P., Moss, R.E.S., and Brandenberg, S.J. (2014). Dynamic response of a model levee on Sherman Island peat: A curated data set, *Earthquake Spectra,* **30**(2), 639-656.

Robertson, P.K., Campanella, R.G., Gillespie, D., and Greig, J. (1986). Use of Piezometer Cone data. *In-Situ'86: Use of In-situ testing in Geotechnical Engineering, Geotechnical Special Publication (GSP) 6*, American Society of Civil Engineers, Reston, VA, 1263-1280.

Rodriguez-Marek, A, Rathje, E.M., Bommer, J.J., Scherbaum, F., Stafford, P.J. (2014). Application of single-station sigma and site response characterization in a probabilistic seismic hazard analysis for a new nuclear site. *Bull. Seismol. Soc. Am.*, **104**, 1601-1619.

Rodriguez-Ordoñez, J.A. (1994). *A new method for interpretation of surface wave measurements in soils. Ph.D. Dissertation*, Engineering Department, North Carolina State University, 424 p.

Roe, W.P., and Madin, I. (2013). 3D geology and shear-wave velocity models of the Portland, Oregon, Metropolitan Area, *Oregon Department of Geology and Mineral Industries (DOGAMI), Open-File Report O-13-12*, 48 p.

Rogers, A.M., Tinsley, J.C., and Borcherdt, R.D. (1985). Predicting relative ground response, *U.S. Geol. Surv. Profess.* Paper 1360, 221-248.

Sadiq, S., Ilhan, O., Ahdi, S.K., Bozorgnia, Y., Hashash, Y.M.A., Park, D., Yong, A., and Stewart, J.P. (2018). A proposed seismic velocity profile database model. *11th National Conf. on Earthquake Engineering (11NCEE)*, Earthquake Engineering Research Institute, Los Angeles, USA, 25–29 June 2018.

Sanchez-Salinero, I., Roesset, J.M., Shao, K.-Y., Stokoe II, K.H., and Rix, G.J. (1987). Analytical Evaluation of Variables Affecting Surface Wave Testing of Pavements, *Transportation Research Record 1136,* Washington, D.C., 86-95.

Schnabel, P.B. (1973). Effects of local geology and distance from source on earthquake ground motions. *Ph.D. thesis*, University of California, Berkeley.

Schorlemmer, D., Euchner, F., Kästli, P., and Saul, J. (2011). QuakeML: status of the XML-based seismological data exchange format, *Annals of Geophysics*, **54**(1), doi: 10.4401/ag-4874.

Seyhan, E. and Stewart, J.P. (2014). Semi-empirical nonlinear site amplification from NGA-West 2 data and simulations. *Earthquake Spectra,* **30**, 1241-1256.

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

Shannon & Wilson–Agbabian Associates (SW-AA, 1980). Geotechnical data from accelerograph stations investigated during the period 1975-1979, *Summary Report No. NUREG/CR-1643*, Prepared for Division of Reactor Safety Research, Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commission, Washington, D.C., 165 p.

Shantz, T., Turner, L., and Roblee, C. (2015). Geotechnical Virtual Data Center (GVDC), presentation at *First Community V_S Profile Database Workshop*, PEER Center, Berkeley, CA, 27 May 2015.

Silva, W., Darragh, R.B., Gregor, N., Martin, G., Abrahamson, N., and Kircher, C. (1998). Reassessment of site coefficients and near-fault factors for building code provisions, Technical Report Program Element II: 98-HQ-GR-1010, Pacific Engineering and Analysis., El Cerrito, California USA.

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.

Steidl, J.H., Biswas, N., Davis, P.M., and Savage, W.U. (2004). ANSS Downhole Arrays at Instrumented Buildings, *in Proceedings of the International Workshop for Site Selection, Installation, and Operation of Geotechnical Strong-Motion Arrays – Workshop 1: Inventory of Current and Planned Arrays*, 14 and 15 October 2004, *eds.* P. de Alba, R.L. Nigbor, J.H. Steidl, and J.C. Stepp, *COSMOS Publication No. CP-2004/01*, Richmond, CA.

Stewart, J.P., Abrahamson, N.A., Atkinson, G.M., Baker, J.W., Boore, D.M., Bozorgnia, Y., Campbell, K.W., Comartin, C.D., Idriss, I.M., Lew, M., Mehrain, M., Moehle, J.P., Naeim, F., Sabol, T.A. (2011). Representation of bidirectional ground motions for design spectra in building codes. *Earthquake Spectra*, **27**(3), 927-937.

Stewart, J.P., K. Afshari, Y.M.A. Hashash (2014). Guidelines for performing hazard-consistent one-dimensional ground response analysis for ground motion prediction. *PEER Report 2014/16*, Pacific Earthquake Engineering Research Center, Berkeley, CA.

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

Stewart, J.P. and Stewart, A.F. (1997). Analysis of soil-structure interaction effects on building response from earthquake strong motion recordings at 58 sites, *Rpt. No. UCB/EERC-97/01, Earthquake Engineering Research Center, University of California, Berkeley*, 742 p.

Stewart, J.P., Chang, S.W., Bray, J.D., Seed, R.B., Sitar, N., and Riemer, M.F. (1995). A Report on Geotechnical Aspects of the January 17, 1994 Northridge Earthquake, *Seismological Research Letters*, **66**(3), 7-19.

Stewart, J.P., Smith, P.M., Whang, D.H., and Bray, J.D. (2004). Seismic compression of two compacted earth fills shaken by the 1994 Northridge earthquake, *J. Geotech. & Geoenv. Eng.*, **130**(5), 461-476.

Stewart, J.P. and Sholtis S.E. (2005). Case study of strong ground motion variations across cut slope, *Soil Dyn. Earthquake Eng.*, **25**(7-10), 539-545.

Stewart, J.P., Taciroglu, E., Wallace, J.W., Ahlberg, E.R., Lemnitzer, A., Rha, C., Khalili-Tehrani, P., Keowen, S., Nigbor, R.L., and Salamanca, A. (2007). Full Scale Cyclic Large Deflection Testing of Foundation Support Systems for Highway Bridges. Part I: Drilled Shaft Foundations, *UCLA Structural & Geotechnical Engineering Laboratory, Report No. 2007/01*, 228 p.

Stewart, J.P. (2016). "Joyner Lecture: Site response uncertainty and its implications for seismic risk characterization," EERI 2016 Annual Meeting, San Francisco, CA; SSA 2016 Annual Meeting, Reno, NV *Seismological Research Letters*, **87(2B)**, pp 516 (abstract).

Stewart, J.P., Kramer, S.L., Kwak, D.Y., Greenfield, M.W., Kayen, R.E., Tokimatsu, K., Bray, J.D., Beyzaei, C.Z., Cubrinovski, M., Sekiguchi, T., Nakai, S., and Bozorgnia, Y. (2016). PEER-NGL project: Open source global database and model development for the next-generation of liquefaction assessment procedures, *Soil Dyn. Earthq. Eng.*, **91**, 317-328.

Stokoe II, K.H., Wright, S.G., Bay, J.A., and Roesset, J.M. (1994). Characterization of geotechnical sites by SASW Method, *Geophysical Characterization of Sites*, R.D. Woods, ed., A.A. Balkema, Rotterdam, 15-25.

Strobbia, C. and Cassiani, G. (2011). Refraction microtremors: data analysis and diagnostics of key hypotheses, *Geophysics*, **76**(3), 11–20.

Tao, Y., Rathje, E.M. (2019). Insights into modeling small-strain site response derived from downhole array data, *J. Geotech. Geoenv. Eng.*, 145, 04019023.

Telford, W.M., Geldart, L.P., and Sheriff, R.E. (1990). *Applied Geophysics* (2nd Ed.): New York, Cambridge University Press, 770 p.

Thiel Jr., C.C. and Schneider, J.F. (1993). Investigations of Thirty-Three Loma Prieta Earthquake Strong Motion Recording Sites, *Report for United States/Japan Loma Prieta Earthquake Project, with reporting information collected with the support of California Universities for Research in Earthquake Engineering (CUREe) and Building Contractors Society of Japan*, 537 p.

Thompson, E.M., (2018), An Updated Vs30 Map for California with Geologic and Topographic Constraints: U.S. Geological Survey data release. DOI: 10.5066/F7JQ108S.

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

Thompson, E.M., Kayen, R.E., Carkin, B., and Tanaka, H. (2010). Surface-wave site characterization at 52 strong-motion recording stations affected by the Parkfield, California, M6.0 earthquake of 28 September 2004, *U.S. Geological Survey Open-File Report 2010–1168*, 120 p.

Thompson E.M., Wald, D.J., Worden, C.B. (2014) A Vs30 Map for California with Geologic and Topographic Constraints. *Bull. Seismol. Soc. Am.*, **104**, 2313–2321.

Toohey, K. (2015). SimilarityMeasures: Trajectory Similarity Measures. R package version 1.4. https://CRAN.R-project.org/package=SimilarityMeasures.

UCR (1989). Laboratorio de Ingeniería Sísmica. *International Federation of Digital Seismograph Networks, Universidad de Costa Rica*. https://doi.org/10.7914/SN/MF.

Urzúa, A., Dobry, R., and Christian, J.T. (2017). Is Harmonic Averaging of Shear Wave Velocity or the Simplified Rayleigh Method Appropriate to Estimate the Period of a Soil Profile?. *Earthquake Spectra*, **33**(3), 895-915.

USGS Fundamental Science Practices Advisory Committee (2011). U.S. Geological Survey Fundamental Science Practices: U.S. Geological Survey Circular 1367, 8 p.

Van Houtte, C., Drouet, C., Cotton, F. (2011). Analysis of the origins of κ (kappa) to compute hard rock to rock adjustment factors for GMPEs. *Bull. Seismol. Soc. Am.*, **101**, 2926–2941.

Wair, B.R., DeJong, J.T., and Shantz, T. (2012). Guidelines for estimation of shear wave velocity profiles, *Pacific Earthquake Engineering Research Center Report No. 2012/08*, 68 p.

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

Wallace, J.W., Fox, P.J., Stewart, J.P., Janoyan, K., Qiu, T., and Lermitte, S. (2001). Cyclic large deflection testing of shaft bridges. Part I-Background and field test results, *Report to California Department of Transportation, Civil & Environmental Engineering Dept., UCLA.*

Wang, P., and Stewart, J.P. (2019). Data-derived site response and its predictability using ergodic and site-specific methods. UCLA. https://escholarship.org/uc/item/3wb9h9fq.

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, Berkeley, CA.

Wair, B.R., DeJong, J.T., and Shantz, T. (2012). Guidelines for estimation of shear wave velocity profiles, *Pacific Earthquake Engineering Research Center Report No. 2012/08*, 68 p.

Warrick, R.E. 1974. Seismic investigation of a San Francisco bay mud site. *Bulletin of the Seismological Society of America*, **64**(2), 375–385.

Wills, C.J. (1998). Differences in shear-wave velocity due to measurement methods: A cautionary note, *Seismol. Res. Lett.*, **69**(3), 216-221.

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

Wills, C.J. and Silva, W.J. (1998). Shear-Wave Velocity Characteristics of Geologic Units in California. *Earthquake Spectra*, **14**(3), 533-556.

Wills, C.J., Gutierrez, C.I., Perez, F.G., and Branum, D.M. (2015). A Next Generation VS30 Map for California Based on Geology and Topography. *Bull. Seismol. Soc. Am.*, **105**, 3083–3091.

Woodward-Clyde Consultants (WCC), 1993. Field Investigations at Selected Strong Motion Accelerograph Sites, *Project No. 9000588, for Electric Power Research Institute*, Oakland, California, January 29, 1993.

World Wide Web Consortium (W3C, 2018). Extensible Markup Language (XML) webpage, available at https://www.w3.org/XML/ (last accessed 8 November 2018).

Xu, B., Rathje, E.M., Hashash, Y.M.A., Stewart, J.P., Campbell, K.W., and Silva, W.J. (2020). κ_0 for soil sites: Observations from Kik-net sites and their use in constraining small-strain damping profiles for site response analysis, *Earthquake Spectra*, https://doi.org/10.1177/8755293019878188.

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

Yong, A, Hough, S.E., Iwahashi, J., and Braverman, A. (2012). A terrain based site conditions map of California with implications for the contiguous United States. *Bull. Seismol. Soc. Am.*, **102,** 114-12.

Yong, A., Martin, A., Stokoe II, K.H., and Diehl, J. (2013). ARRA-funded VS30 measurements using multi-technique approach at strong-motion stations in California and central-eastern United States, *U.S. Geol. Survey Open-File Rept. 2013*–1102, 65 pp.

Yong, A., Thompson, E.M., Wald, D.J., Knudsen, K.L., Odum, J.K., Stephenson, W.J., and Haefner, S. (2016). Compilation of VS30 Data for the United States: U.S. Geological Survey Data Series 978, 8 p., http://dx.doi.org/10.3133/ds978.

Zalachoris, G, EM Rathje, 2015. Evaluation of one-dimensional site response techniques using borehole arrays, *J. Geotech. Geoenviron. Eng.*, **141**, 04015053.

Zhang, S.X., Chan, L.S., and Xia, J. (2004). The Selection of Field Acquisition Parameters for Dispersion Images from Multichannel Surface Wave Data, *Pure and Applied Geophysics*, **161**(1), 185–201.