Title
An Improved Framework for the Analysis and Dissemination of Seismic Site Characterization Data at Varying Resolutions

Author
Ahdi, Sean Kamran

Publication Date
2018

UNIVERSITY OF CALIFORNIA
Los Angeles

An Improved Framework for the Analysis and Dissemination of Seismic Site Characterization Data at Varying Resolutions

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

Civil Engineering

by

Sean Kamran Ahdi

2018

ABSTRACT OF THE DISSERTATION

An Improved Framework for the Analysis and Dissemination of Seismic Site Characterization Data at Varying Resolutions

by

Sean Kamran Ahdi

Doctor of Philosophy in Civil Engineering

University of California, Los Angeles, 2018

Professor Jonathan Paul Stewart, Chair

The most commonly used parameter for representing site conditions for ground motion studies is the time-averaged shear-wave velocity in the upper 30 m, or V_{S30}. While it is preferred to compute V_{S30} from a directly measured shear-wave velocity (V_S) profile using in situ geophysical methods, this information is not always available. One major application of V_{S30} is the development of ergodic site amplification models, for example as part of ground motion model (GMM) development projects, which require V_{S30} values for all sites.

The first part of this dissertation (Chapters 2-4) addresses the development of proxy-based models for estimation of V_{S30} for application in subduction zone regions. These procedures are applied at 6433 strong motion recording stations (SMSs) for the NGA-Subduction project, which

has the goal of developing GMMs for global subduction zone (SZ) earthquakes. Relatively detailed V_{S30} prediction models are developed in this thesis for application to the Cascadia SZ in the Pacific Northwest (PNW) region and the Alaska/Aleutian SZs. These are the portions of the United States at greatest risk to seismic hazards from subduction zone earthquakes. In these regions, only 8% of SMSs have measured V_{S30} values.

Proxy-based V_{S30} statistical models based on in situ measurements were developed from information including surficial geology, topographic gradient (i.e. "slope"), and geomorphic terrain categories. The PNW and Alaska studies result in proxy-based V_{S30} models based on (1) a hybrid of generalized surficial geologic groups conditioned on topographic slope, and (2) geomorphic terrain categories. With 928 measured V_{S30} values available in the PNW, statistically robust proxy models were developed, with 18 generally well-populated geology groups assigned logarithmic mean and standard deviation V_{S30} values, six of which are conditioned on topographic slope, and certain groups reflecting glacial and volcanic geology. Additionally, 13 of a possible 16 terrain classes were well-populated, and these were also assigned V_{S30} statistical moments. Ultimately, due to strong correlation between the two proxies but an overall lower dispersion of model residuals for the hybrid geology-slope proxy compared to the terrain proxy, use of the hybrid slope-geology proxy model was recommended. For Alaska, a different approach was required, as most geology groups were not well populated. In these cases, V_{S30} data from the PNW were borrowed for similar geologic groups, residuals analysis for Alaska-only and combined Alaska-PNW group moments were computed, and bias was considered in model prediction. The standard deviations of such groups' model predictions were increased by adding an epistemic uncertainty in the mean to reflect the uncertainty in adopting a proxy for use outside of its original intended region of application.

Other SZ regions included in NGA-Subduction project rely on regional V_{S30} prediction models developed previously or concurrent with this project by others (i.e., Japan, Taiwan, New Zealand, and Chile), or required development of procedures as part of the present work to assign V_{S30} where regional models are unavailable (i.e., Central America/Mexico and South America outside of Chile). The V_S data collection effort for the latter two regions resulted in a general lack of measured V_{S30} data and proxy information, precluding robust proxy model development. As such, an existing geomorphic terrain class proxy model was borrowed from California, with an additional assigned epistemic uncertainty in the mean to account for increased uncertainty in implementing proxy-based models outside of their original intended region of use. Basin depth terms are also provided for V_S profiles that exceed velocity thresholds (e.g., 1.0 or 2.5 km/s) or estimated for regions where 3D seismic velocity models exist (e.g., Cascadia, Japan, New Zealand, and Taiwan).

A similar study was undertaken in Iran to populate a site database with V_{S30} values for a ground motion modeling. Analysis of measured Iranian V_S data and comparison of within-group moments for geology and terrain proxies in other regions around the world showed that the Iranian V_{S30} values did not vary much across different geology groups, a possible sign that the seismic refraction velocity data lacked adequate resolution in the upper 30 m to provide accurate V_{S30}. To mitigate this, a third approach for proxy development was formally developed, in which moments for similar geologic groups were borrowed from the PNW and California, averaged, and used for assignment to Iranian strong motion stations, again with care to factor in inter-regional epistemic uncertainty. This work defined the framework for the assignments of V_{S30} to the aforementioned data-poor regions of Central America, Mexico, and South America.

The second part of this dissertation (Chapter 6) concerns the development of the United

States Community V_S Profile Database (PDB), 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.

The dissertation of Sean Ahdi is approved.

Yousef Bozorgnia

Scott Joseph Brandenberg

Paul M. Davis

Robert E. Kayen

Jonathan Paul Stewart, Committee Chair

University of California, Los Angeles

2018

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

The research presented in this dissertation was sponsored by multiple agencies. The work of Chapters 2, 3, and 4 was sponsored by the Pacific Earthquake Engineering Research (PEER) Center through the larger NGA-Subduction project under contract number FMI-20122689, FM Global, the U.S. Geological Survey (USGS), California Department of Transportation, and Pacific Gas & Electric Company (PG&E), and the Bonneville Power Administration. The work in Chapter 2 was presented in multiple conference papers with numerous co-authors, including Ahdi et al. (2017c, 16WCEE), Kishida et al. (2017, 16WCEE), and Kishida et al. (2018, 11NCEE), and has been updated and expanded upon here; Chapter 2 will be included as Chapter 5, "NGA-Subduction Site Database," of a report published by PEER (2019) describing the NGA-Subduction Database, with co-authors Victor Contreras, Dong Youp Kwak, Annie Kwok, and Jonathan P. Stewart. The work in Chapter 3 was published in a journal paper in the Bulletin of the Seismological Society of America (Ahdi et al. 2017a) and modified for presentation here. The work in Chapter 4 was presented in a conference paper (Ahdi et al. 2017b, PBD-III) and modified for presentation here. I would like to thank the entire NGA-Subduction project team for all of their advice and patience over the duration of the project, particularly Yousef Bozorgnia, Tadahiro Kishida, and Silvia Mazzoni. I am grateful for my BSSA co-authors Tim Ancheta, Dong Youp Kwak, and Devjyoti Mitra for numerous meetings and assistance with database development and coding for the proxy-based V_{S30} model development project early in my Ph.D. My coauthors and I are grateful to Susan Chang, Robert Kayen, and Sheri Molnar for providing unpublished V_S profile data in the PNW and Alaska; Xavier Vera-Grunauer for providing unpublished V_S profile data in Ecuador; Bill Stephenson for providing information about the USGS Cascadia basin model and at the time

unpublished $z_{2.5}$ values in the PNW; Chris Wills and Alan Yong for many helpful discussions on development of V_{S30} proxy models; Tadahiro Kishida for assembling the list of strong motion recording stations that recorded Cascadia subduction zone events; and Grace Parker for providing insight gained from a similar V_{S30} proxy development study performed for NGA-East. My co-authors and I would like to thank Alan Yong and two anonymous reviewers of the BSSA manuscript for suggestions that led to the splitting of Alaska from the PNW and thus formulation of three approaches to proxy-based V_{S30} modeling described in Chapter 2.

The work of Chapter 5 was supported by the Pacific Earthquake Engineering Research Center (PEER). Scientific collaboration with the Building & Housing Research Center (BHRC) of the Ministry of Road and Urban Development of Iran is gratefully acknowledged. I would personally like to thank Kioumars Afshari, Samantha Hangsan, Grace Parker, and Yi Tyan Tsai for their significant contributions in assigning and processing metadata to strong-motion station sites in a timely manner.

The work in Chapter 6 was partially supported by the USGS External Grants program under award number G17AP00018. The views and conclusions contained in Chapter 6 are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. A consortium of the Consortium of Organizations for Strong Motion Observation Systems (COSMOS), Pacific Gas & Electric (PG&E), and Southern California Edison (SCE) provided supplementary grant funding for Dong Youp Kwak to develop the V_S PDB project's relational database and online interface. Institutional support was provided by COSMOS, the PEER Center at UC Berkeley, and the B. John Garrick Institute of Risk at UCLA. The work in this chapter was presented in two conference papers (Ahdi et al. 2018, GEESD-V; Sadiq et al. 2018, 11NCEE). I would like to thank my co-authors and project leadership of the V_S PDB project,

including Yousef Bozorgnia, Youssef Hashash, Okan Ilhan, Dong Youp Kwak, Duhee Park, Shamsher Sadiq, Alan Yong, and Jonathan Stewart; this project truly involves a community effort, and I am proud to have been part of this excellent team. Funding for the in-person community V_S PDB project planning workshops in 27 May 2015 at UC Berkeley and 29 June 2017 at UCLA were provided by the Pacific Earthquake Engineering Research (PEER) Center, The B. John Garrick Risk Institute at UCLA, Caltrans, and BC Hydro. I would like to thank all those who participated in and presented during these workshops. I would like to thank the following (in no particular order) for contributing in-kind V_S profile data and associated reports: Chris Wills, Jennifer Thornburg, and Christopher Tran (CGS); Tom Shantz and Bill Owen (Caltrans); Scott Brandenberg (digitized Caltrans data); Ariya Balakrishnan (CA DWR); Bob Nigbor (ROSRINE and CUREE digital data); Jamie Steidl (NEES@UCSB site data); Bob Nigbor and John Diehl (EPRI/Kajima Inc. reports; GEOVision consulting reports); Rob Steller (GEOVision/Agbabian suspension log data for UC-CEP); Robert Kayen (digitized data for USGS open-file reports); and Susan Chang (City of Seattle). I thank David Teague, Brady Cox, Robb Moss, Mitchell Craig, Gabriel Toro, and Annie Kammerer among many who have directed me to potential future data sources. I would like to thank Loralinda Bustamante, Tina Glorioso, and Shawn Jones of the DWR DSOD for granting access to their library, helping identify dam sites with V_S profile data and finding their associated reports, and allowing use of their conference room to scan these reports. I thank Tadahiro Kishida for helping to scan these reports (and for bringing the scanners!). I would like to thank Martin Hudson and Marshall Lew of Wood plc, who provided access to their company's internal data archive to extract relevant publicly-reviewed data from the former LeRoy Crandall & Associates and Law/Crandall reports. I thank Scott Brandenberg and Paolo Zimmaro for many discussions regarding the organization and structure of the database schema, Scott's

assistance with hosting the VS PDB website on the uclageo.com domain and allocating appropriate server space, and Paolo's assistance and interaction with Chris Jordan from the Texas Advanced Computing Center (TACC), who mirrored our database onto the NHERI DesignSafe web interface. I thank Drew Gomberg for developing a Python script and GUI for populating JSON files with geospatial information, and Pengfei Wang for performing quality checks on inconsistent data entered in the database. I thank all participants of the two community workshops who provided their input in project development. Finally, I thank Alisa Stewart, Drew Gomberg, Annie Huang, Kenneth Hudson, Stefano Ingargiola, Jason Kim, Avery Kirschbaum, Justin Qiu, Andy Shang, Nikolai Ramalingam, Jonathan Song, Craig Sugiyama, Bailey Uy, and Tristan Whisenant, Hyewon Choi, Inhyeok Choi, Taehun Han, Jiseok Kim, Sangbaek Park, Daeun Song, and Seunghun Yang, who all helped with digitization and organization of data. Without their dedicated efforts, the success of project would not have been possible.

On a personal note, I would first and foremost like to thank my advisors, Jonathan Stewart and Scott Brandenberg, who have provided *much* time and guidance throughout my undergraduate studies, masters coursework, and Ph.D. research; without their support and belief in me, I might have departed UCLA far too soon. It has been an immense privilege to work with those who demand excellence and truly desire to see you succeed. I would also like to express my gratitude to the other members serving on my committee: Yousef Bozorgnia, for endless support and advice; Robert Kayen, for excellent discussions and tips for doing geophysical fieldwork (particularly in Japan), and always looking to integrate geological principles into geotechnical engineering; and Paul Davis, for numerous discussions of the applicability of scientific principles to earthquake engineering research. I would also like to thank Don Browne, Marty Hudson, Tony Martin, Bob Nigbor, Jamie Steidl, Mladen Vucetic, An Yin, and Alan Yong, for helping shape my research and

academic interests into what they are today. There are far too many people that positively influenced and impacted me during my time at UCLA to count here, but I truly am grateful for all their support. I do give special thanks to my geotechnical friends and officemates for everything from giving technical advice to keeping me sane, including in particular: Pavlo Chrysovergis, for starting me on the geotechnical dirt trail many years ago; Kioumars Afshari, for his infinite knowledge on ground motions and on life; Estefan Garcia, for being "that friend" at "that other campus" who reminded me that the Ph.D. is difficult no matter where you are; Dong Youp Kwak, for all the quick help in times of need and for being a source of inspiration with his countless abilities to excel; Grace Parker and Yi Tyan Tsai, for their level-headed insight from undergraduate classes to Ph.D. research, and willingly working together from day one until the end; and Jason Buenker, Victor Contreras, Maria Giovanna Durante, Mandro Eslami, Bahar Heidarzadeh, Ebuka Nweke, Claudia Rangel, Ali Shafiee, Benjamin Turner, Pengfei Wang, Yang Yang, Samuel Yniesta, Paolo Zimmaro, and many more, for their endless conversations and insights, helping development of research ideas and life goals, and allowing me to gain diverse viewpoints and joining me in life's experiences. A sincere thank you also to past UCLA Geotechnical Engineering Ph.D. students, particularly Tim Ancheta, Joe Coe, Anne Lemnitzer, Dong Youp Kwak, Ahmadreza Mortezaie, Ted Reinert, Emel Seyhan, and Lisa Star; your advice and support before and throughout my Ph.D. has been immense. I thank you all for your wonderful friendship.

Go Bruins!

VITA

Education

M.S., 2014, Civil (Geotechnical) Engineering, University of California, Los Angeles

B.S., 2013, Civil Engineering, University of California, Los Angeles

B.S., 2013, Geology/Engineering Geology, University of California, Los Angeles

Professional Experience

Teaching Associate, Henry Samueli School of Eng. & Applied Science, UCLA (2014-2018)

Graduate Research Assistant, Dept. of Civil & Env. Engineering, UCLA (2014-2018)

Geophysical Technician/Intern, GEOVision, Inc., Corona, CA, (2011)

Civil Engineering Technician/Intern, PLSA, Inc., Solana Beach, CA, (2010)

Selected Publications

Ahdi, S.K., O. Ilhan, S. Sadiq, Y. Bozorgnia, Y.M.A. Hashash, D. Park, A. Yong, J.P. Stewart (2018). Development of a United States Community Shear Wave Velocity Profile Database, *5th Conference on Geotechnical Earthquake Engineering and Soil Dynamics (GEESD-V)*, June 10-13, 2018, Austin, Texas, *accepted*.

Ahdi, S.K., J.P. Stewart, T.D. Ancheta, D.Y. Kwak, D. Mitra (2017). Development of *VS* profile database and proxy-based models for *VS*30 prediction in the Pacific Northwest Region of North America. *Bulletin of the Seismological Society of America*, **107**(4), 1781–1801.

Ahdi, S.K., J.P. Stewart, T.D. Ancheta, D.Y. Kwak, D. Mitra (2017). Proxy-based *VS*30 prediction in Alaska accounting for limited regional data, *Proc. 3rd Int. Conf. on Performance-Based Design in Earthquake Geotechnical Engineering (PBD-III)*, Vancouver, Canada, July 16-19, 2017, Paper No. 482.

Ahdi, S.K., T.D. Ancheta, V. Contreras, T. Kishida, D.Y. Kwak, A.O. Kwok, G.A. Parker, Y. Bozorgnia, J.P. Stewart (2017). NGA-Subduction Site Database. *Proc. 16th World Conf. on Earthquake Eng. (16WCEE)*, January 9-12, 2017, Santiago, Chile. Paper No. 4926.

1 Introduction

1.1 MOTIVATION OF RESEARCH

Engineering estimates of earthquake ground shaking potential are frequently provided by ground motion models (GMMs), which estimate distributions of ground motion intensity measures (GMIMs). GMM developers utilize recorded acceleration time series from earthquake events and relate their associated GMIMs to independent variables related to: (1) the earthquake source (e.g., magnitude), (2) the path of propagating seismic waves (such as distance to the earthquake hypocenter or ruptured fault), and (3) near-surface ground conditions at the recording site of interest. Recently developed GMMs generally focus on different tectonic settings, intended for use globally or for specific geographic regions. These include the Next Generation Attenuation (NGA)–West 2 (NGA-W2) project GMMs (e.g. Boore et al. 2014; Campbell & Bozorgnia 2014) developed for active tectonic regions such as the Western United States, Japan, and Taiwan, and other regions worldwide; the NGA-East models (e.g. Yenier and Atkinson, 2015) developed for Central and Eastern North America (CENA) and representing stable continental regions; and the NGA-Subduction (NGA-Sub) models (e.g., Abrahamson et al. 2018), which cover subduction zones around the Pacific Rim.

While an understanding of each of the three aforementioned components of GMMs are important, this dissertation focuses on the characterization of near-surface site conditions, which is required for the development of site amplification models as used in GMMs. The most

commonly used parameter for representing site conditions is the time-averaged shear wave velocity in the upper 30 meters (V_{S30}) of a site, as the shear wave velocity (V_S) of a material relates directly to its shear modulus (Equation 6.1). While the best practice for obtaining V_{S30} is to compute it from a V_S profile directly measured at the site of interest to a depth of 30 m or greater, this data is not always available. For the purpose of GMM development, large-scale geophysical field-testing campaigns to measure V_S profiles at ground motion recording stations, such as the study performed by Yong et al. (2013), are beneficial, but are expensive and may require years to complete. As such, it becomes important to be able to estimate V_{S30} for a given site using secondary information, or "proxies", that are more readily available, such as surficial geology, geotechnical descriptors, or various morphological parameters derived from digital elevation models (DEMs, e.g., Farr and Kobrick 2000) such as elevation, topographic gradient (slope), or geomorphic terrain classes (e.g., Iwahashi and Pike 2007). Issues such as differences in geological interpretations among map publishers and coarseness of resolution of geologic maps and DEMs contribute to the uncertainty inherent in proxy information (and thus in V_{S30} estimation).

Beyond V_{S30}, two parameters that are sometimes used in GMMs are basin depth terms and site fundamental frequency (f_0, or its inverse, period, T_0). Basin depth is defined as the depth to V_S exceeding certain velocity horizons, and has been shown to represent amplification from long-period waves. Site fundamental frequency is related to the natural modes of vibration of the soil column and can be predictive of local peaks in ground motion spectra.

The balance between attaining site-specific V_{S30} accuracy versus the relative infeasibility of directly measuring V_{S30} leaves the engineering seismology community with a philosophical question, one in which I am personally interested: how can one advocate for making in situ geophysical measurements using credible data acquisition and processing techniques, while

simultaneously creating generalized models for V_{S30} prediction using these various proxies? Through this research, I have explored this spectrum of V_{S30} estimation, investigating various proxy-development techniques from V_S source data, and quantifying the uncertainty associated with various methodologies.

1.2 OVERVIEW OF V_{S30} PROXY DEVELOPMENT METHODS

As described in Section 1.1, the development of proxy-based V_{S30} models requires a database of measured V_S profiles and proxy information both widely available across a study region. In recent years, the latter are most readily publicly available in electronic formats; for example, the United States Geological Survey (USGS) hosts the National Geologic Map Database (NGMDB), an open-access tool to allow look-ups of surface or bedrock geology from published geological maps across the entire country. However, obtaining V_S profile data generally requires significant effort in data collection and unification of data formats, and often V_S profiles are considered to be proprietary data if collected by commercial entities. Data and metadata quality are also important, such as having depth of profile (z_p) greater than 30 m, knowing accurate *and* precise latitude and longitude coordinates of measurement locations, and using data from credible geophysical methods, among many other factors. The present discussion does not intend to outline how to best vet all of these parameters before deciding if a certain profile warrants inclusion or exclusion in a profile database (PDB); this is to some degree discussed in the following chapters.

While we can sometimes use judgment to include or adjust V_S profiles or V_{S30} data that do not meet all criteria for data quality, an issue that cannot be overcome is minimal or a complete lack of measured V_S data in the study area of interest. Proxy-based V_{S30} models are as robust as their constituent V_S profile data. Particularly in the case of geology- and terrain-based V_{S30} estimation models, enough measured V_{S30} data must exist within each model group to ensure

statistical significance of model predictions. When either an entire region, or a subset of category groups (e.g., geology or terrain) within a region, are underpopulated, proxy models from other regions can be borrowed and implemented in the region of interest, while taking care to recognize the increased uncertainty stemming from the application of a model outside its intended region of use. Three approaches for V_{S30} proxy development are proposed, which are based on the quantities and quality of available V_{S30} and proxy data in a given region. Following this discussion, the subsequent chapters of this dissertation present examples for implementation of these approaches in specific regions around the world based on V_S PDB source data of differing sizes and quality.

The general framework for developing category-based V_{S30} proxy models consists of three approaches, each of which require relatively large-scale (<1:250,000, preferably <1:50,000) geologic maps across the entire study region. These maps can be in query-able geographic information systems (GIS) files or can more simply be scanned and georeferenced images of geologic maps. If such maps are unavailable, globally available DEM models can be used to compute geomorphic terrain categories at sites of interest (e.g., Iwahashi and Pike 2007, Iwahashi et al. 2018). Other region-specific and global proxies may exist and can be implemented using different weights for V_{S30} estimation based on the predictive power of each proxy-based model.

Approach I is the most desirable, wherein the study region has a substantial database of measured V_S profiles, and the model developer has access to relatively large-scale electronic geologic maps. Simplified geologic groups based on map information are assigned to all sites in the profile database, and these are used to compute the first and second statistical moments (i.e., the mean and standard deviation) of V_{S30}, which are then applied to sites of interest. Preferably, within-group dependency on topographic slope (which can be computed from DEMs) would also be investigated and included in the model for appropriate groups. Conversely, *Approach III* is the

least desirable scenario, where virtually no V_S profile data is available in the study region, or wherein such data exists but the model developer deems it to be of insufficient quality or quantity to develop a robust proxy-based statistical V_{S30} model. In such cases, region-specific geologic groups for the study area are developed, but proxy group moments are borrowed from other regions located in similar tectonic settings. This approach necessarily involves increased epistemic uncertainty in estimates in the mean and standard deviation of V_{S30}.

Approach II falls between these two end-member cases, for which there might only be limited available V_S profile data of good quality within the study region. Generalized geologic groups that are well-populated with V_{S30} data are developed for specific application to the study region, as for the other approaches. These groups are linked to similar groups in existing proxies in other regions, from which group moments are borrowed. Then, residuals between each model's within-group moments and measured V_{S30} values are computed and compared, and the target group's region-specific moments are adjusted as needed. This process allows for available region-specific data to be applied directly in development of the proxy-based V_{S30} prediction model.

1.3 SCOPE OF DISSERTATION

Chapter 2 of this dissertation outlines the development of the site database (SDB) for the NGA-Subduction project, a large multi-national effort aimed at developing GMMs for seven global subduction-zone (SZ) regions. The population of site characterization information at each strong motion recording station (SMS), particularly V_{S30} for all sites in the NGA-Sub SDB, comprised a large fraction of my Ph.D. research. A portion of sites (39%) in the SDB were populated using measured data, but the rest were populated using V_{S30} proxies that are either (a) region specific, most of which were developed as part of NGA-Sub (an exception is Japan), or (b) relatively general as a result of limited local data—these estimates are described in Chapter 2. The

former group includes sites that recorded events from the Cascadia and Alaska/Aleutian SZs, and the latter group includes sites in Central America & Mexico (CAM) that recorded events in the Cocos SZ, and sites in South America (outside Chile) that recorded Nazca SZ events. Proxy-based V_{S30} models were adopted from the literature for Japan and New Zealand or developed by other NGA-Sub researchers for Taiwan and Chile; these are reviewed in Chapter 2.

Chapter 3 presents a study in the Pacific Northwest (PNW) region of North America to develop proxy-based V_{S30} estimation models using a substantive assembled database of measured V_S and V_{S30} data, which NGA-Sub GMM developers have incorporated in their modeling of ground motions from the Cascadia SZ. The relatively large database (928 V_S/V_{S30} measurement locations) allows for implementation of Approach I, described in Section 1.2, for proxy-based V_{S30} development. It should be noted that the work and data presented in Chapter 3 and Appendix A of this dissertation was published as an article in the Bulletin of the Seismological Society of America (Ahdi et al. 2017a).

Chapter 4 of this dissertation focuses on proxy-based V_{S30} prediction model development in Alaska, which was a spinoff study of the PNW PDB and model development described in Chapter 3. Alaska is a region rich in ground motion recordings used in NGA-Sub, but possesses a markedly smaller database of measured V_S data (126 profiles), which prompted the implementation of Approach II (described in Section 1.2) to estimate V_{S30} and its associated uncertainty for SMSs. A feature of this study required borrowing data from (and lending data to) the PNW and the V_{S30} proxy developed for that region. Information in Chapter 4 is described in a paper presented at the 3rd International Conference on Performance-Based Design in Earthquake Geotechnical Engineering (PBD-III) (Ahdi et al. 2017b).

Chapter 5 focuses on a similar study in Iran that was undertaken in conjunction with the

Building & Housing Research Center (BHRC) of the Iranian Ministry of Road & Urban Development, where a proxy-based V_{S30} model was developed using a national Iranian V_S profile dataset of uncertain data quality and relatively small-scale geologic maps. This necessitated the development and formalization of Approach III, as described in Section 1.2. In Iran, models were adopted from other regions in the world (namely California and the PNW) with similar tectonic settings to assign V_{S30} and its uncertainty to SMSs. Again, an increase in uncertainty was considered given the use of a proxy model outside of its intended region of application.

Chapter 6 describes the United States Community V_S Profile Database (PDB), which was developed in the present work in collaboration with an advisory committee consisting of researchers from UCLA, University of Illinois, and Hanyang University in Korea, and with institutional support from COSMOS, USGS, and UCLA. The V_S PDB is a major multi-institutional effort that has developed an open-access V_S profile database for sites in the United States. Data was collected for this project from a wide variety of sources, ranging from federal government reports, California state agencies, university research reports, reports from private industry, and utility companies. Data collection primarily began in California, where over 2300 profiles exist, and expanded to encompass the entire U.S. V_S profiles obtained from a wide variety of geophysical methods are included in this database, along with additional geotechnical data and metadata pertaining to the measurement site. Due to the expansive nature of data types encountered during data collection, a relational database schema specific to this project has been developed, and novel data organization protocols were utilized to handle and store different data formats. Finally, a web-based interface was developed to allow easy access, querying, visualization, and downloading of the data. The work described herein has been presented widely at both domestic and international conferences and has garnered interest from researchers and industry practitioners in both the

engineering seismology and geotechnical earthquake engineering communities. It is anticipated the V_S PDB will be both a useful tool for analysis and act as a guide for similar database development in the earthquake engineering field worldwide. The data and information presented in this chapter are included as part of an Earthquake Data Paper manuscript that is in preparation for submission to *Earthquake Spectra*, the professional journal of the Earthquake Engineering Research Institute (EERI).

Chapter 7 summarizes the research performed during my Ph.D. with the goal of looking towards potential future research work that can be performed particularly with relation to the data collected in the V_S PDB. Future work can be divided into tasks related to NGA-Sub and to work aimed at expanding the V_S PDB with new data sources (Section 7.2.1) and applications of the V_S PDB (7.2.2). Specific applications include investigating non-ergodic (site-specific) site response utilizing V_S PDB data (Section 7.2.2.1), studying the predictability of site response with 1D models, which depends on the assumption of lateral homogeneity of soil strata (Section 7.2.2.2), studying the consistency of the use of horizontal-to-vertical spectral ratio (HVSR) of recorded ground motions or ambient microtremors to estimate site fundamental frequencies (Section 7.2.2.3), and an effort aimed at improving the modeling of V_S structure in sedimentary basins and their implementation in GMMs (Section 7.2.2.4). It is hoped that the work of this dissertation underscores the importance of the implementation of rigorous geological and tectonic principles to explain physical processes that are manifest in engineering seismology and geotechnical earthquake engineering—while creating a framework for future studies to consider this important, sometimes overlooked, and often oversimplified, information.

2 NGA-Subduction Site Database

2.1 OVERVIEW AND ORGANIZATION OF SITE DATABASE

NGA-Subduction is a multi-year, multi-institution international effort to develop ground motion models (GMMs) for subduction zones around the world. The principle regions from which data has been collected are: Alaska, the Pacific Northwest of North America, Central America & Mexico (CAM), Japan, New Zealand (NZ), South America, and Taiwan. A global map of stations and earthquake epicentral locations included in the NGA-Sub project database is presented in Figure 2.1. Along with ground motion data, database development included major efforts related to systematic development of supporting metadata related to the seismic source and recording stations. I was the lead student investigator on a team responsible for the site component, with many collaborators and coordinated parallel studies (identified in Acknowledgments). This chapter presents the results of that work.

The words "site" and "station" are used somewhat interchangeably, with "station" generally referring to the actual strong motion instrument, and "site" generally referring to a more general description of the location of interest (which for this project coincides with the strong motion stations). The NGA-Sub site database (SDB) contains 6433 sites, and for each site the following data is provided:

- site name and station ID, often adopted from the original strong motion network's

station code/name;

- a unique station sequence number (SSN) that acts as the identifier for every site in the database;
- information about station location, such as latitude, longitude, depth below ground surface, elevation, and in some cases, information on sensor housing;
- recommended V_{S30} values and aleatory variabilities linked to codes that describe V_{S30} assignment protocols, along with associated NEHRP site classes (Dobry et al. 2000);
- where epistemic uncertainty in the mean V_{S30} values is appreciable, this uncertainty is reported in the form of a standard deviation term;
- details related to measured V_S profiles when available, such as the maximum depth of the profile (z_p) and time averaged V_S to z_p (V_{SZ});
- site information used to predict V_{S30} from proxy-based models when measured V_{S30} values are absent;
- basin depth information such as the depth to a particular V_S horizon (i.e., z_x = the depth to the x km/s iso-surface) where available from measurements or regional 3D velocity models; and
- indicators of whether a station is located in the forearc or backarc of the particular subduction zone region for which it recorded data.

A breakdown of the number of sites in the SDB by region is presented in Figure 2.2. The region assigned to sites in Figure 2.2 is based principally on the region that produced the earthquake, not necessarily the site location. In most cases, these coincide. For example, Japan contributes the most sites (35%), and those sites have collectively produced 57% of the NGA-Sub ground motions, virtually all of which are from earthquakes in Japan. Conversely, the Cascadia group in Figure 2.2,

which comprises 18% of the SDB (second highest after Japan), includes sites in the Pacific Northwest region (British Columbia, Washington, Oregon, and the region of northern California north of Cape Mendocino) as well as other, more distant regions including other parts of California, inter-mountain west states, and central and eastern North America (generally east of the Rocky Mountains). Despite the large number of sites, the number of ground motion records from Cascadia events is relatively small (only 3% of the ground motion inventory).

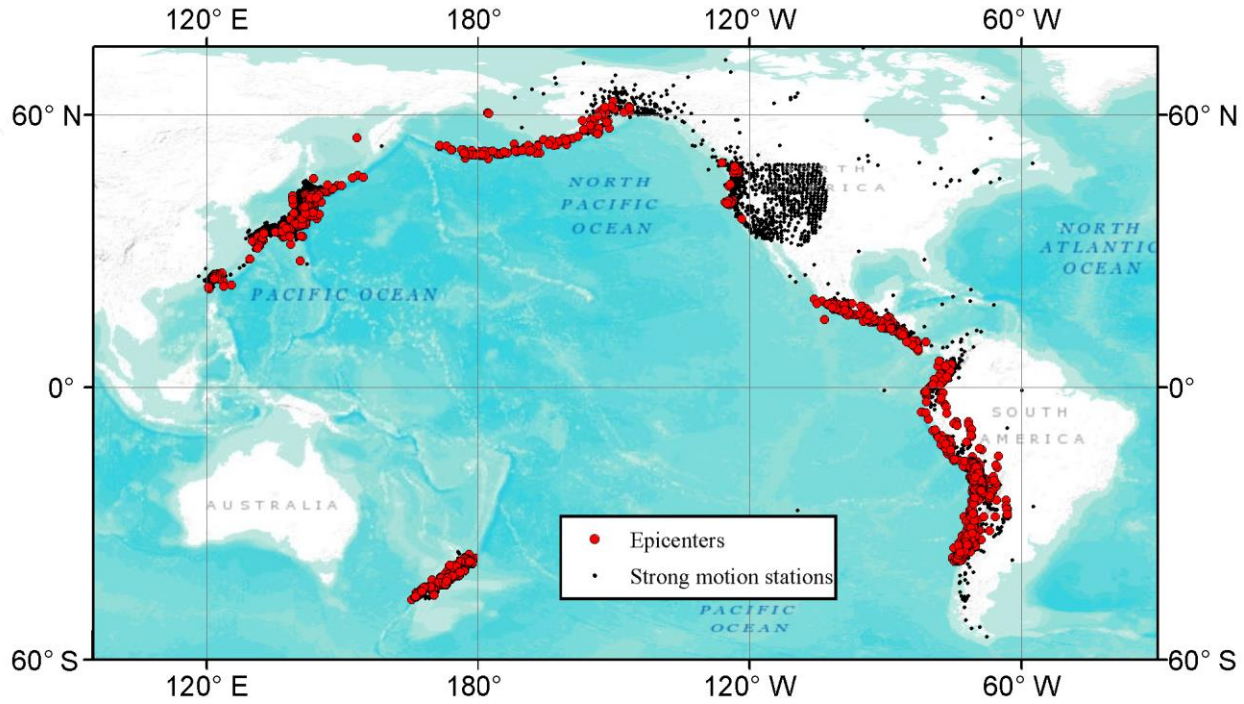

Figure 2.1. Locations of epicenters and strong motion recording stations in NGA-Sub database (source: Kishida et al. 2018).

Some strong motion stations included in the NGA-Sub SDB, particularly in California, Alaska, Taiwan and Japan, have also recorded shallow crustal events. As such, the sites are included in the NGA-West2 and NGA-East project databases. Where overlap exists, station metadata and any assigned V_{S30} and basin depth values that existed in those databases were generally sought to be adopted for assignment in the NGA-Sub SDB, except for sites in California that recorded Cascadia events, for which V_{S30} was updated using more recent methods (see Section

2.3.3a). However, this was not performed systematically because of complexities related to inconsistent station naming conventions, station numbering, and other problems.

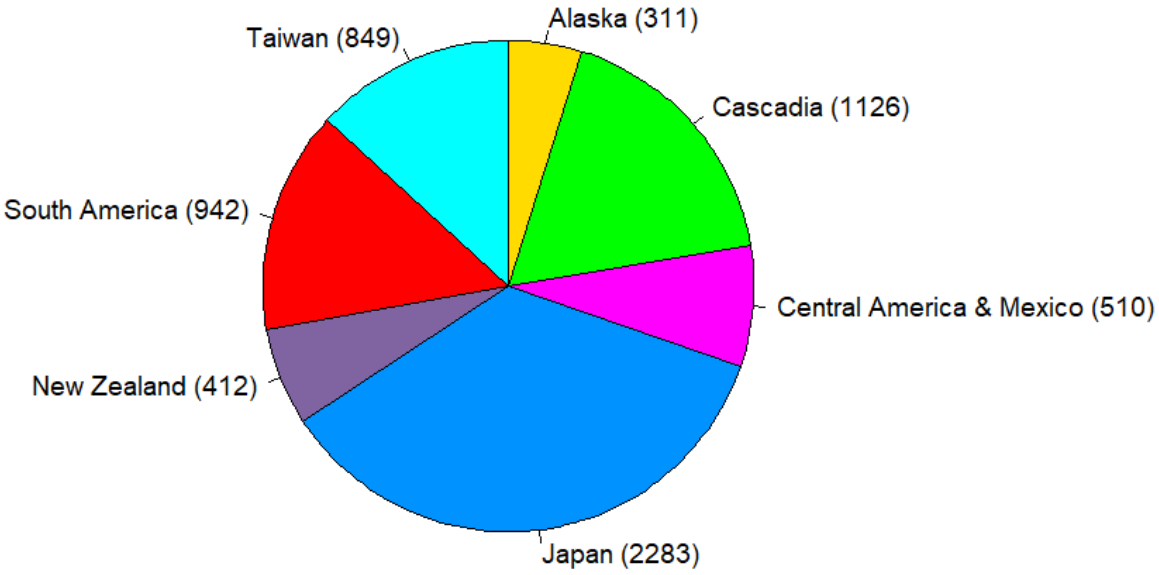

Figure 2.2. Breakdown of number of sites in NGA-Sub site database by region (N = 6433).

2.1.1 Approach for SDB Development

In past NGA projects, the site "database" took the form of a spreadsheet (i.e., a Microsoft Excel spreadsheet [XLS] or comma-separated values [CSV] file), with each entry (row) corresponding to an individual site. In NGA-Sub, the project team has organized the data into a relational database, where interrelated tables of data and metadata communicate using various unique keys, called primary keys, to ensure consistency between and across tables. A more in-depth description of the relational database is provided in Chapter 2 of PEER (2019). Chapter 6 of this dissertation discusses the development of a relational database for the United States Community V_S profile database. Within the NGA-Sub relational database, there is an SDB table that is similar in content to the previous spreadsheet files. The contents of this table can be exported to spreadsheet format (.XLS or .CSV).

The SDB is organized by region, with the unique SSNs assigned as sequential integers of the pattern $N \times 10^6$ within each region, where N is an integer from one to seven accounting for each of the seven subduction zone regions shown in Figure 2.2, in alphabetical order (i.e., Alaska stations are numbered 1000001, 1000002, …; Taiwan stations are numbered 7000001, 7000002, …). This methodology was adopted to (1) facilitate using the SSN as a primary key for the SDB table within the NGA-Sub relational database and (2) allow the database to grow in the future when new stations are inevitably added, or when SDBs from NGA-West2 and NGA-East are merged with this one, such that each region will not feasibly run out of integer values for SSN. A similar numbering scheme is implemented for other data tables in the NGA-Sub database. The extensibility of the relational database prevents the need for maintaining sequential SDB spreadsheet files (up to 32 versions of which were used in NGA-West and NGA-West2; Seyhan et al. 2014).

To avoid redundant listings of sites in the SDB, a hierarchy for removal of duplicate sites was applied, whereby sites were combined when they had identical latitude and longitude coordinates (within 0.0001-decimal degree precision), as well as reasonably consistent station names and/or identification numbers from the station network. Separate SSNs are assigned for cases when the station network changed but the instrument ostensibly remained the same, or the instrument itself was changed at the same location. In the case of vertical arrays, with multiple sensors at the same latitude and longitude but different depths, multiple "sites" (with distinct SSNs) are provided in the SDB, each having a different sensor depth.

2.1.2 Sources of Station data

Ground motion data for each region was generally obtained from the web sites of organizations that operate local accelerograph or seismograph networks. These same web sites

often have some basic station information, typically including station locations, instrument information, and in some cases, geotechnical data and seismic velocity profiles. Details on data sources is provided in Chapter 3 of PEER (2019).

Most sites in Japan are part of the KiK-Net, K-NET, Port and Airport Research Institute (PARI), Japan Meteorological Agency (JMA), or Tokyo Electric Power Company (TEPCO) networks. Taiwan station data is derived from the Taiwan Strong Motion Instrumentation Program, managed by the Central Weather Bureau (CWB), as well as the broadband seismic observation network, co-managed by CWB and the Institute of Earth Sciences, Academic Sinica in Taiwan (NCREE, 2017). Station information from Alaska was obtained from the Alaska Earthquake Center at the University of Alaska at Fairbanks (AEC, 2018). In Cascadia, data was accessed from sites belonging to numerous networks, including the IRIS Transportable Array (IRIS 2003), the Pacific Northwest Seismic Network (UW 1963), and the Canadian National Seismograph Network (GSC 1989), among multiple others across California and other western states. In Mexico, station data were mainly obtained from the Guerrero Network operated by University of Nevada Reno (Anderson et al. 2006) and COSMOS. In Central America, major networks exist for most countries, such as the Laboratorio de Ingeniería Sísmica in Costa Rica (UCR 1989), the Medio Ambiente y Recursos Naturales (MARN) network in El Salvador (SNET 2018), and the Tomography Under Costa Rica and Nicaragua (TUCAN) network (Abers & Fischer 2003). In South America, station data comes from 36 different networks; particularly important networks within this region include the Chilean Seismological Network and RENADIC, operated by the University of Chile, the Red Nacional Acelerógráfos de Colombia, the PerU Lithosphere and Slab Experiment (Wagner et al. 2010), and the Ecuador Seismic Network. The data from New Zealand was taken from Kaiser et al. (2016, 2017), which contains site information for stations included in

a ground motion flat file described in Van Houtte et al. (2017).

Each network has varying levels of existing site characterization information, details of which are discussed in the following sections. Table 2.1 lists all networks in the NGA-Sub SDB and the number of sites from each. There are 92 networks represented in the SDB. Of the 6433 stations, 417 (6%) lack network assignments, with nearly all (412) coming from NZ.

Table 2.1. Summary of strong motion station networks in NGA-Sub SDB.

Network	Count	Network	Count	Network	Count	Network	Count
K-NET	1021	COSMOS	39	INETER	20	MX	5
CWB	802	ZD	39	ESCIGSMN	18	EPCO	4
KiK-net	702	XS	38	XH	17	IGP	4
TA	414	AV	37	CGS	13	ONA	4
JMA	314	NSMP	36	NU	13	CERESIS	3
UW	191	Historic	35	TEPCO	13	G	3
CSN	149	IU	35	XV	13	II	3
AK	146	NP	34	CDMG	11	NR	3
CN	136	OV	34	KNET	11	OO	3
RENADIC	132	BK	33	XJ	11	IE	2
LIS	122	RENAC	33	GI	10	NCSN	2
RNAC	121	C	32	USGS	10	NV	2
MARN	114	JP	31	CM	9	TC	2
PARI	102	US	30	Idel	9	UU	2
NC	75	DGG	29	CC	8	ZX	2
BO	69	CX	27	ICE	8	AX	1
PB	64	NN	27	SV	8	BDSN	1
XY	64	C1	25	AE	7	GE	1
CI	63	TW	24	ZW	7	MG	1
USGSH	61	CISMID	23	CU	6	PA	1
ZA	61	2B	22	EC	6	PUCP	1
YO	49	Y9	22	GT	6	WC	1
UNR	45	YC	22	UO	6	YJ	1

The remainder of this chapter describes site data and metadata used for NGA modeling purposes. Emphasis is placed on proxy models used to estimate V_{S30} and its aleatory variability and uncertainty for regions for which related studies were not already published. This includes

Central America & Mexico (events from the Cocos subduction zone), and South America (Nazca subduction zone) excluding Chile. Regions for which V_{S30} prediction models have been developed and are already published are reviewed in this chapter, but the reader is referred to the original sources or subsequent chapters in this dissertation for more details. This applies to Taiwan, Cascadia, Alaska, and New Zealand. For Japan, data and proxy models were already published as part of NGA-West2 and other projects. Updates were applied in NGA-Sub that are presented here.

2.2 MEASURED V_S DATA FOR V_{S30} EVALUATION

2.2.1 Data Sources

Wherever possible, V_S profiles developed from in situ geophysical testing are identified for use in the characterization of site conditions at a ground motion instrument site. The profile is used to compute and assign V_{S30} and also, where applicable, to assign depths to 1.0 and 2.5 km/s shear wave velocity iso-surfaces. In general, V_S profile data is considered if it reflects direct measurements (from in situ geophysical testing), the profile extends to a profile depth z_p of at least 5 m, the profile begins within 5 m of the ground surface, and the profile location (geodetic coordinates) is known. Data derived using a wide array of geophysical measurement techniques was used (Chapter 6 describes geophysical testing methods considered to be appropriate for measuring V_S profiles). One technique that is not considered credible (CXW, Poran et al. 1994) was nonetheless used for estimates of V_{S30} for sites in Alaska, as explained in Chapter 4. In general, profiles from the ReMi^{TM} method (Louie 2001) were not compiled for use in the PDBs, with the exception of a small number of sites in Chile.

Of the 6433 sites in the SDB, 2514 have an assigned V_S profile. As part of the NGA-Subduction project, considerable effort was put into identifying suitable profiles. Correspondence with national and regional agencies, as well as individual researchers and geotechnical/geophysical

consulting firms, was undertaken. Ultimately all of the data used in the project is public domain. The profiles and accompanying metadata had disparate formats in source documents. The V_S profiles were digitized (if not already in digital form) and assembled into "profile databases" (PDBs) for individual SDB regions. Further information on these PDBs are described in Chapter 3 for Cascadia (also presented in Ahdi et al. 2017a) and Chapter 4 for Alaska (also in Ahdi et al. 2017b). For Taiwan, profile data was obtained from a web site maintained by the CWB, as described in NCREE (2017) and Kwok et al. (2018). For Chile, the PDB is derived from a variety of university reports and profiles from the personal files of a consulting firm, as described by Contreras et al. (2018). For Japan, V_S data for stations that are part of the KiK-Net and K-net networks was obtained from a web site maintained by National Research Institute for Earth Science and Disaster Resilience (NIED: http://www.kyoshin.bosai.go.jp/). Site data for stations in the Port and Airport Research Institute network were obtained from the PARI website (https://www.eq.pari.go.jp/kyosin/). We did not identify a source of V_S profile data for sites in the JMA network. In the case of New Zealand, we did not compile a PDB, but rather relied on site metadata compiled for strong motion stations by Kaiser et al. (2016, 2017), some of which is measurement-based (their quality factor Q1, and in some cases, Q2). This is described in further detail in Section 2.3.3.

Figure 2.3 shows the distribution of all V_{S30} data, both measurement- and proxy-based, in the NGA-Sub SDB, showing a commonly observed lognormal distribution skewed towards high V_{S30} (stiffer) sites. Table 2.2 shows the breakdown of measured versus proxy-based V_{S30} data in the SDB by region. The regions with the highest percentages of V_{S30} values based on measurements are Japan (76%) and Taiwan (56%); these regions also had the highest such percentages in NGA-West2, with Taiwan at 53% and Japan at 34% (Ancheta et al., 2013). This shows that proxy-based

models were needed in all regions, and that the need was greatest in regions other than Japan and Taiwan. The substantial need for V_{S30} assignments from proxy-based models motivated a series of studies to develop such models on a region-specific basis, as described further in Sections 2.3.3 and 2.3.4. Figure 2.4 shows the distribution of profile depth z_p for measured V_S profiles in the SDB for which this information exists (i.e., the total number of V_S profiles is less than the number of measurement-based V_{S30} values in the SDB because for some sites the original source documentation provides only a V_{S30} value).

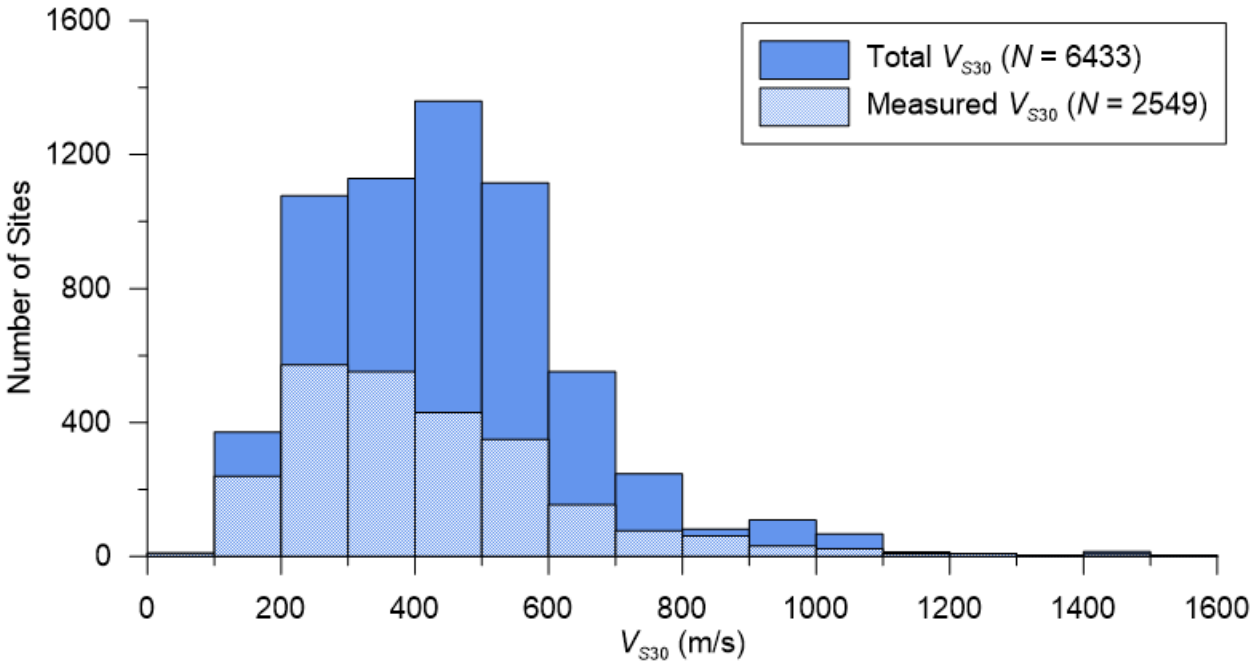

Figure 2.3. Distribution of all V_{S30} assigned to SMSs in the NGA-Sub SDB, with a histogram for the subset of sites with assignments from measured in situ V_S profiles.

Limited measured V_S profile/V_{S30} data were acquired for the stations in regions that recorded events in the Cocos subduction zone (i.e., Central America & Mexico) and the Nazca subduction zone (i.e., South America, but excluding Chile). Fifteen V_S profiles from geophysical testing using downhole and suspension log methods were collected in and around Mexico City at

strong motion stations and ground failure sites after the 1985 Mexico City Earthquake (Ohta et al. 1986; Seed et al. 1987). These were applicable to 11 strong motion stations in the NGA-Sub SDB, with two profiles applicable to six different stations (based on the criteria of proximity within 300 m; Borcherdt 2002; Seyhan et al. 2014). Arango et al. 2010 compiled V_{S30} values from measured V_S profile data acquired for past microzonation studies in El Salvador (Faccioli et al. 1988, five profiles) and Nicaragua (Faccioli et al. 1973, two profiles); V_{S30} from these profiles were assigned to seven and two strong motion stations in each country, respectively. In South America, 27 V_S profiles measured using SASW (Vera-Grunauer, 2014) and combined MASW & SASW (Nikolau et al. 2016). A subset of these were applied to 13 strong motion stations in Ecuador, primarily concentrated in the Guayaquil region and measured during reconnaissance after the 2016 Musine earthquake. The digital V_S profile data were provided by X. Vera-Grunauer (2017, *pers. comm.*). Sixteen V_S profiles were collected in Peru (Cortez-Flores, 2004), one of which could be matched to a strong motion station for V_{S30} assignment. Finally, four measured V_{S30} values obtained from the flat file of the South America Risk Assessment (SARA) project of the Global Earthquake Model (GEM) yielded V_{S30} assignments at five strong motion station sites in Colombia (Castillo et al. 2016).

Table 2.2. Breakdown of measured- and proxy-based estimated V_{S30} data by NGA-Sub region.

Region	Total	Measured	Estimated
Alaska	311	16 (5%)	295 (95%)
Cascadia	1126	80 (7%)	1046 (93%)
Central America & Mexico	510	24 (5%)	486 (95%)
Japan	2283	1731 (76%)	552 (24%)
New Zealand	412	29 (7%)	412 (100%)
South America	942	190 (20%)	772 (82%)
Taiwan	849	473 (56%)	376 (44%)
Total	6433	2494 (39%)	3939 (61%)

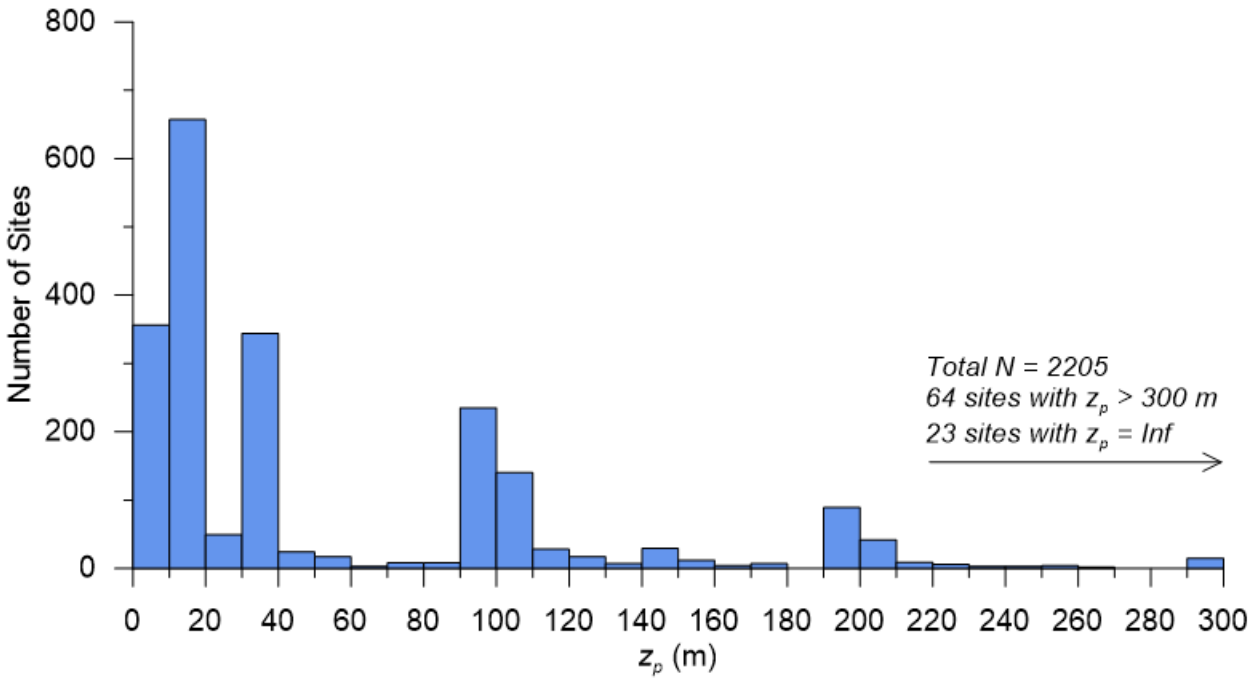

Figure 2.4. Distribution of z_p for measured V_S profiles assigned to SMSs in the NGA-Sub SDB.

2.2.2 V_{S30} Computation

The time-averaged V_S to the maximum profile depth z_p is computed as

$$V_{SZ} = \frac{z_p}{tt_z} \tag{2.1}$$

in which

$$tt_z = \int_0^{z_p} \frac{dz}{V_S(z)} \tag{2.2}$$

where tt_z is the travel time for shear waves to travel from depth z_p to the ground surface. In practice the integral is taken as a summation across depth intervals with constant velocities. When $z_p \geq 30$ m, which occurs for 1490 of 2514 SDB sites with assignment of V_{S30} from measured V_S profiles, V_{S30} is computed by replacing z_p with 30 m.

For the 1024 sites having $z_p < 30$ m, V_{S30} must be estimated by extrapolation. There exist numerous extrapolation schemes in the literature, which are described and compared in Chapter 3

and Appendix A of this dissertation. Statistical analysis on five different V_{SZ}-to-V_{S30} extrapolation schemes (Boore 2004; Boore et al. 2011; Dai et al. 2013; Midorikawa and Nogi 2015; and Wang and Wang 2015), presented in Appendix A and expanded upon in Kwak et al. (2017a), demonstrates that in general, the model framework developed by Dai et al. (2013), which relies on regressions on V_S profiles for individual regions or datasets to predict the time-averaged V_S from z_p to 30 m, provides the lowest uncertainty in its application to particular regions of interest when conditioned with measured V_{S30} data for that region. Dai et al. (2013) was used for shallow profiles in Japan, with regression coefficients provided in Kwak et al. (2017a, their Table 2), and in the PNW, as described in Chapter 3 of this dissertation, while similar but Taiwan-specific extrapolation V_{SZ}-to-V_{S30} procedures discussed in Kuo et al. (2012) were used in Taiwan, as described in Kwok et al. (2018).

2.3 PROXY-BASED ESTIMATION OF V_{S30}

A substantial effort was made to develop region-specific, proxy-based V_{S30} prediction models for application to key NGA-Sub regions for which V_S profile data was accessible. This differs from the NGA-West2 project, for which proxy-based V_{S30} prediction models from prior literature were generally used, with some exceptions (details in Seyhan et al. 2014). The region-specific prediction models used in NGA-Sub are:

- Pacific Northwest: Chapter 3 of this dissertation and Ahdi et al. (2017a),
- Alaska: Chapter 4 of this dissertation and Ahdi et al. (2017b),
- Taiwan: Kwok et al. (2018),
- Chile: Contreras et al. (2018),
- New Zealand: Kaiser et al. (2016, 2017),

- Japan: Not previously published, presented below.

Following a review of methods used for V_{S30} prediction in Section 2.3.1, Section 2.3.2 describes a general framework for development and application of models that reflects regional data availability and associated uncertainties. Region-specific models are then briefly summarized for NGA-Sub applications that are published elsewhere and describe (in more detail) approaches used for regions for which proxies were not already developed (i.e., Central America & Mexico, and South America outside Chile).

2.3.1 Methods used for Proxy-Based V_{S30} Prediction

Proxy-based models can be categorized in different manners. One is on the basis of region of applicability, with global models distinguished from local models. Global models require two attributes—the proxy itself must be globally available and the predictive model of V_{S30} given the proxy should be based on a geographically diverse data set. Based on this definition, the principle global model is that of Wald and Allen (2007) and Allen and Wald (2009), which use the proxy of topographic slope gradient at 30 arc-sec resolution and collections of V_{S30} measurements from California, Italy, and Taiwan (model for active tectonic regions) and Australia and Tennessee (model for stable continental regions). The same DEM used for topographic slope can also be used to define geomorphic terrain classes on the basis of slope gradient and metrics of convexity and texture (e.g., Iwahashi and Pike, 2007; Iwahashi et al., 2018). Hence, while a global model is possible using these approaches, to date the applications have been local, specifically California (Yong et al., 2012; Yong, 2016), Greece (Stewart et al., 2014a), Taiwan (Kwok et al. 2018), and Cascadia (Chapter 3 of this dissertation). Different resolutions of DEMs can also be used, as discussed in in Allen and Wald (2009) and Stewart et al. (2014a) but may result in a bias in slope estimates for a given location based on canopy effects due to vegetation or presence of buildings

at higher DEM resolutions (Stewart et al. 2014a).

Local models are applicable to a particular domain, typically defined on the basis of political boundaries or changes in the predominant crustal structure. For a given domain, a second level of categorization concerns the type of proxies considered for use in the correlation model. These include surface geology, geotechnical descriptors, slope gradient, geomorphic terrain class, elevation, and hybrids of more than one proxy. Table 2.3 summarizes some existing V_{S30}-prediction relationships, including the region of applicability and the proxies used. Several recent models use a combination of surface geology and ground slope, an approach introduced by Wills & Gutierrez (2008) and advanced by Thompson & Wald (2012), Thompson et al. (2014), and Parker et al. (2017), among others. The Thompson et al. (2014) study, later updated by Thompson (2018), begins with a geology-slope approach and then computes residuals between V_{S30} data at profile locations and the model, which are then mapped using a kriging approach. For application to the USGS ShakeMap product, V_{S30} estimates are provided by combining the model prediction with location-specific residuals. The Japan Engineering Geomorphologic Classification Map (JEGM) provides an effective category-based proxy that reflects geological and morphological conditions; means and standard deviations of V_{S30} are provided by category, which is similar to the terrain-based methods. The other method listed in Table 2.3 is based on geotechnical descriptors (Chiou and Youngs, 2008) and applied in California. This approach was not used in the present work.

Table 2.3. Literature summary for proxy-based methods for V_{S30} estimation.

Proxies Considered	Region	Parameterization (Category/Equation)	No. of Groups	References	Description/Notes
Surface Geology	California Beijing	Categories, Equations	19 4	Wills and Clahan (2006) Xie et al. (2016)	Xie et al. (2016): Bilinear model with two equations covering 4 surface geological units.
Topographic Gradient	Global	Categories	8	Wald and Allen (2007), Allen and Wald (2009)	Slope gradient computed from 30 arc-sec (~1 km) resolution grid spacing from SRTM.
Terrain Categories	California Greece, Taiwan, PNW	Categories	16	Iwahashi and Pike (2007); Yong et al. (2012), Yong (2016), Stewart et al. (2014a), Kwok et al. (2018), Ahdi et al. (2017)	Surface morphology categorized by slope gradient, local convexity, and surface texture. SRTM DEM at 30 arc-sec grid spacing.
Geotechnical Descriptors	California Japan	Categories	5	Chiou and Youngs (2008), Seyhan et al. (2014)	Geotechnical site categories, from Geomatrix 3rd letter scheme.
Geomorphic/Geologic maps	Japan	Categories	22	Matsuoka et al. (2006), Matsuoka and Wakamatsu (2008), Wakamatsu and Matsuoka (2013)	National geomorphic/geologic maps digitized at 7.5 arc-sec. V_{S30} predicted from JEGM category, slope gradient, elevation, and distance from mountain/hill.
Hybrid: Geology & Topographic Gradient	California CENA Greece, PNW, Alaska	Categories, Equations	15 14 5	Wills et al. (2015) Parker et al. (2017) Stewart et al. (2014a) Ahdi et al. (2017a,b)	Geologic units from various maps grouped into categories based on descriptions of lithology/ depositional environment. For certain groups, slope-depended regression equations presented.
Hybrid: Geotechnical & Elevation	Taiwan	Categories, Equations	5	Chiou and Youngs (2008), Ancheta et al. (2013)	Grouped by GMX 3rd letter geotechnical descriptors combined with station elevation within each GMX category.
Hybrid: Surface Geology, Slope, Elevation	Taiwan	Categories, Equations	3	Ahdi et al. 2017b Kwok et al. 2018	Surface geology classified using 1:50,000-scale maps (otherwise 1:250,000). SRTM DEM at 30 arc-sec grid spacing for gradient.

2.3.2 Proposed V_{S30} Prediction Framework

For large, global projects like NGA-Sub, there is a need to estimate the site parameter V_{S30} for regions with highly variable levels of data availability and quality. Some regions, like California, Taiwan, and Japan, have relatively extensive V_S data and map resources (geology, etc.) that provide relevant proxies at high resolution. Other regions largely lack V_S profile data and may or may not have reliable maps for proxies other than global 30 arc-sec DEM maps (e.g., SRTM30, Farr & Kobrick 2000). Project requirements dictate that values of V_{S30} are needed for all sites, so a framework is needed to provide this, along with appurtenant uncertainties.

The framework described here distinguishes between variability and uncertainty. *Variability* here refers to the standard deviation representing the dispersion of V_{S30} data relative to a mean estimate for a given set of predictor variables, and is denoted as $\sigma_{\ln V}$. *Uncertainty* here refers to lack of knowledge of the appropriate value of statistical moments (mean and standard deviation), also known as epistemic uncertainty, and is denoted as σ_{ep}. Emphasis is placed on epistemic uncertainty in mean estimates.

Approach I: Good quality V_S data and proxy maps

Regions for which Approach I applies have ample V_S profile data and geological map resources that allow relevant proxies to be mapped at high spatial resolution. The development of appropriate V_{S30}-prediction models for such regions begins by assembling a V_S PDB. Preferably, this V_S data is of high quality, dense spatial resolution, and spans a wide array of geological and geomorphological environments. In most recent models, primary proxy that is considered is surface geology, which should be presented at high resolution (ideally 1:50,000 or larger scale). Larger map scales provide more confidence of mapping accuracy with respect to the geological units present at a site of interest. Morphological information such as topographic gradient and/or

elevation is often combined with mapped geologic category.

Judgement is used to group categories from various geological maps of different scales and potentially from different authors or institutions. Next, V_{S30} moments are computed for each category, usually under the assumption of a log normal distribution. The aleatory uncertainty is taken as category standard deviation ($\sigma_{\ln V}$), the mean is the category mean in natural log units. By convention, the exponent of that mean is denoted $\mu_{\ln V}$, and has units of m/s. Where justified by the data, the mean within a category may be dependent on slope and possibly elevation. V_{S30} moments are developed in a like manner for multiple groups, which taken together, constitute a proxy-based V_{S30} prediction model. In some cases, additional factors, such as influence of basins or prior glaciation on V_{S30} (e.g., Parker et al. 2017 and Chapter 3 of this dissertation), can be investigated using residuals analyses to find particular groupings that can improve model predictive power.

The epistemic uncertainty in the means developed for Approach I can be represented by the standard error of the mean, which decreases the as standard deviation decreases and the number of data points used to compute the moments increases. This uncertainty is generally small and is not reported.

For NGA-Sub, region-specific proxies developed in this manner were prepared for the Pacific Northwest (Chapter 3) and Taiwan (Kwok et al. 2018). A prior model meeting this general description was updated for Japan. A proxy-based model conditioned on terrain categories instead of geology was developed for Chile (Contreras et al. 2018).

Approach II: Limited V_S data, good quality maps

Approach II is applicable to regions where some V_S profile data is available, but the amount of information is not adequate to develop models for geologic categories in the manner described for Approach I. These regions generally have good quality geologic maps, as with Approach I

regions.

The concept behind this approach is to apply a proxy-based V_{S30} prediction model for a source (Approach I) region to a target (Approach II) region, and then to assess the applicability using residuals analysis. This allows variations between geologic categories, established from the data-rich source region, to be applied to the target region. If residuals analyses reveal bias, this bias should be removed by adjusting the source model for application to the target region.

The geologic categories used for the target region should be appropriate for the regional geologic conditions (e.g., accounting for local features such as glacial or volcanic deposits). A source region with categories appropriate for comparison to the target must be selected carefully. Prior to NGA-Sub, the primary example of Approach II was the use of a California geology-based model (Wills and Clahan, 2006) for alluvial sites in Italy (Scasserra et al. 2009). Approach II was used in NGA-Sub for some geology groups in Alaska, with the PNW taken as the source region (described in Chapter 4 of this dissertation).

Aleatory uncertainties for Approach II are generally taken from the source region. Epistemic uncertainties can be estimated from the standard error of the bias computed during validation.

Approach III: V_S data absent or of low quality, variable access to geologic maps

Approach III is applied when little to no measured V_S/V_{S30} data is available for a region, or for a specific geologic group of interest within a region. Approach III can also apply when data is available, but it is judged to be unreliable. These regions may or may not have reliable geologic maps. If geologic maps are available, region-specific geological groups are identified as in Approaches I and II. V_{S30} moments for similar groups are then assigned from other (source) regions. This is similar to Approach II, but without the validation step. If geologic maps are not

available, global slope or terrain class models may be applied. This approach is not preferred if geologic maps are available, because several studies have found stronger predictive power from geology-based proxies or hybrid geology-slope proxies (Seyhan et al., 2014, Ahdi et al. 2017a, Parker et al. 2017).

Approach III involves larger epistemic uncertainty than other approaches. To estimate this uncertainty, $\mu_{\ln V}$ values are assembled from Approach I studies for California (Wills et al. 2015), Taiwan (Kwok et al. 2018), Greece (Stewart et al. 2014a), the Pacific Northwest (Chapter 3 and Ahdi et al. 2017a), and CENA (Parker et al. 2017), as shown in Table 2.4. The standard deviations of the natural logs of these means provides an estimate of epistemic uncertainty. These standard deviations are estimated separately for Holocene sediments (principally alluvium), Pleistocene sediments (principally older alluvium and terrace deposits), and Tertiary-aged sedimentary bedrock materials. In each case, as shown in Table 2.4, the epistemic uncertainty (σ_{ep}) is approximately 0.2 in natural log units. Accordingly, σ_{ep} = 0.2 is assigned as the epistemic uncertainty.

Table 2.4. Computation of average epistemic uncertainty for similar V_{S30} geological age groups across multiple study regions.

Age	Mean (ln) V_{S30} Values Across Regions (m/s)					Group Moments	
	PNW	CENA	CA	Greece	Taiwan	$\mu_{\ln V}$ (m/s)	$\sigma_{\ln V}$
H	277	210	278	327	320	**279**	**0.158**
Pl	458	271	362	471	508	**404**	**0.229**
T	455	351	405	456	702	**460**	**0.232**

Note: H = Holocene, Pl = Pleistocene, T = Tertiary

Examples where Approach III was implemented in NGA-Sub include geology groups in Alaska that had little to no measured V_{S30} data (Chapter 4 of this dissertation) and geomorphic terrain classifications per Iwahashi & Pike (2007) for Central America, Mexico, and South America outside of Chile. Similar procedures were applied for Iran for a project unrelated to NGA-

Sub (Chapter 5 of this dissertation).

2.3.3 Application of Existing Regional Proxy-Based V_{S30} Prediction Models

Over the four-year duration of the NGA-Sub project, regional proxy-based V_{S30} prediction models were developed for the Pacific Northwest, Alaska, Taiwan, New Zealand, and Chile. Those models are published elsewhere. In this section, those models are briefly reviewed, and comments are provided regarding their application to NGA-Sub sites.

(a) Pacific Northwest

The Pacific Northwest (Cascadia) model is presented in Chapter 3 and Ahdi et al. (2017a), and formally encompasses Oregon, Washington, and southwestern British Columbia. The V_S profile dataset gathered in the PNW allowed for development of 18 well-populated surficial geology categories based on geologic map units largely at 1:24,000 to 1:100,00 scale, providing a high level of resolution in geologic units, encompassing a range of alluvial and glaciation-related sedimentary depositional environments, and three types of rock (sedimentary, igneous, and metamorphic). As such, this is an Approach I model.

Six of the 18 groups were found to correlate to topographic slope and were provided a power-law model to condition assignments based on slope within each group. A geomorphic terrain proxy-based model, following the 16 classes prescribed by Iwahashi and Pike (2007) was also developed with 13/16 classes having well-populated groups (greater than 3 data points).

An electronic supplement to Ahdi et al. (2017a) contains V_{S30} assignments and supporting metadata for PNW sites. In a few cases, sites missing from that supplement have assignments added in the current version of the SDB using the procedures described in Chapter 3. Sites that recorded Cascadia events but that are located in California have V_{S30} assignments from the Wills

et al. (2015) V_{S30} prediction model, which is based on surficial geology and three slope bins for alluvial categories (254 sites). For 314 sites located in the intermountain west of the U.S., an Approach III framework was utilized whereby the terrain classification-based V_{S30} prediction model from Yong (2016) was used to assign mean and standard deviation V_{S30} values, with an epistemic uncertainty assigned (σ_{ep} = 0.2). Sites that recorded Cascadia events that are located in CENA have V_{S30} assignments from the Parker et al. (2017) prediction model.

(b) Alaska

The V_{S30} prediction model for Alaska was developed in coordination with model development for the PNW. The model is described in Chapter 4 of this dissertation and in Ahdi et al. (2017b). The V_S profile dataset in Alaska is limited in number, and the measurements that are available are clustered in relatively few areas (Anchorage, Fairbanks, Seward, Valdez, and areas considered in post-earthquake reconnaissance following the 2002 Denali event). For much of Alaska, large-scale geologic maps are not available, which necessitated the use of very small-scale regional maps (e.g., the statewide map of 1:584,000, Wilson et al. 2015).

As a result, much of the Alaska-specific regional model was developed following Approach II, but with some geological groups utilizing Approaches I and III as well. Comparing the 18 geological groups from the PNW to the site conditions in Alaska resulted in an implementation of five different categories of proxy attribution. Approach I was used for one group (alluvium) which had enough Alaska-specific V_{S30} measurements. A combination of Approaches I and II was used for three groups (lacustrine, alluvial fan, and loess deposits) that had similar data populations in Alaska and PNW without appreciable inter-region bias. Approach II was used for two groups (artificial fill and glacigenic sediments) where bias was checked against PNW group moments and was found to be negligible. Approach III was used for other groups that were

underpopulated with respect to Alaskan data or had site conditions not present in the PNW proxy (e.g., tectonic mélange). The source region used to assign moments to Alaska was generally the PNW, although California (Wills et al. 2015) was used for the tidal-flat and mélange groups.

Metadata and V_{S30} assignments for Alaska sites were not presented in Ahdi et al. (2017b), but they are provided in the SDB based on the protocols in that paper and Chapter 4.

(c) Chile

Chile is among the most seismically active countries in the world and has contributed substantially to the world-wide subduction event inventory (Chapter 4 of PEER 2019). The density of recording stations was relatively low up through the time of the 2010 **M**8.8 Maule earthquake, with the available networks primarily being operated by two academic departments in the University of Chile. Since that time, the number of strong motion stations has increased significantly (Leyton et al. 2018), as has the amount of seismic site characterization performed at stations and for engineering projects.

As part of NGA-Sub, a profile database was assembled for Chile, drawing heavily upon university and industry contacts. Contreras et al. (2018) describes the V_S dataset (492 V_{S30} measurements) developed for Chile. To date, geologic maps have not been accessed to provide surface geology metadata. As a result, geomorphic terrain classes (based on Iwahashi & Pike 2007) were used as the proxy for V_{S30} prediction. All but two of the 16 categories were well populated with profile data. In three cases, terrain classes of similar description were grouped. Aside from terrain class, the model considers regional effects caused by differences in climate in the arid north and more fertile regions to the south. The model is presented in Contreras et al. (2018) and was applied here for metadata and site class assignments in the SDB. Future work for Chile entails development of a hybrid geology/slope V_{S30} model, once suitable geology maps are accessed.

(d) New Zealand

As noted in Section 2.2.1, Kaiser et al. (2016, 2017) assembled site metadata for strong motion stations in New Zealand. Each site has an assigned value of V_{S30} along with an indicator of Q1, Q2, or Q3. These indicators have qualitative descriptions of the quality and uncertainty of the site parameters. They define Q1 (assigned to 29/412 [7%] of New Zealand sites in the SDB) as "well-constrained measurements of V_{S30} from non-invasive surface-wave methods or borehole Seismic Cone Penetrometer Testing (SCPT)," indicating that Q1 corresponds with in situ measurement-based V_{S30} values. Q2 (33 sites) is defined as being one or more of the following: 1) "estimates based on partly constrained near-surface V_S structure (i.e. well-constrained to depths less than 30 m)"; 2) "estimates from known local strata and V_S approximated using established correlations"; or 3) "well-constrained measurements at nearby geologically similar sites." This description makes it difficult to determine on a site-specific basis which Q2 sites are measurement- or proxy-based. Q3 (350 sites) is defined as one or both of: 1) "Estimates from broad-scale national V_{S30} maps" or 2) "estimates at sites with poor constraints." Sites with Q2 and Q3 estimates of V_{S30} are considered to be based on Approach I, given the local attributes of the estimates.

V_{S30} values provided by Kaiser et al. (2016) are adopted for NGA-Sub without modification. For Q1 sites, an aleatory variability is assigned that is appropriate for V_{S30} as established from a V_S profile, which is $\sigma_{\ln V} = 0.1$ (Seyhan et al. 2014). For Q2 sites, $\sigma_{\ln V}$ is assigned as 0.25, which is a typical value for V_{S30} uncertainty as derived from profiles developed using geotechnical data (Kwak et al. 2015). For Q3 sites, $\sigma_{\ln V}$ is assigned as 0.4, which is a typical value for V_{S30} uncertainty as derived from surface geology proxies (e.g., Figure 3.13).

Ongoing work by researchers in New Zealand is developing a V_{S30} map of the entire country based on geological and topographic constraints, similar to work done by Thompson et al.

(2014) in California, and uses Bayesian inferencing to condition statistical groups based on *a priori* group moments obtained from the hybrid geology/slope proxy model developed for the PNW, and updates posterior distributions based on added New Zealand V_{S30} data (Foster et al. 2017).

(e) Taiwan

Kwok et al. (2018) assembled a V_S profile database from site data on the CWB website and used this data to develop a Taiwan-specific V_{S30} prediction model conditioned on geological age categories in combination with slope and elevation. The geological categories were derived using large-scale maps (1:50,000). A terrain classification proxy (based on Iwahashi & Pike 2007 classes) was also developed, with 15 of 16 classes being well-populated with measured V_{S30} data. These models are best described as Approach I. An electronic supplement to Kwok et al. (2018) provides V_{S30} assignments and supporting metadata, and this information was transferred to the SDB for NGA-Sub.

2.3.4 Previously Unpublished Proxy-Based V_{S30} Assignments

This section describes V_{S30} assignments for NGA-Sub sites in regions where previous prediction models are not used in their existing form. In the case of Japan, two previous Approach I models are adapted for use in NGA-Sub. For the remaining regions (Central America/Mexico, South America other than Chile, western Canada and Alaska panhandle), Approach III methods are applied.

(a) Japan

In Japan, V_S profiles were compiled from strong motion recording sites within the K-NET and KiK-Net networks (Aoi et al. 2004; http://www.kyoshin.bosai.go.jp) and the PARI network (Ichii et al. 1999; http://www.eq.pari.go.jp/kyosin/). There are 1667 profiles with profile depths z_p

≥ 10 m. In the K-NET network, typical profile depths are 10-20 m, whereas for KiK-Net and PARI these are 100-200 m and < 200 m, respectively. If geotechnical investigations (e.g., SPT or CPT) at a site are available but geophysical investigations are not, models correlating V_S with penetration resistance and effective stress are used to estimate V_S data (Kwak et al. 2015). This method is applied to 42 of the PARI sites.

Each of the 1667 sites in the profile database were assigned one of the 16 terrain categories from the Iwahashi and Pike (2007) classification scheme. As was done for California initially by Yong et al. (2012) and then updated by Yong (2016), category moments $\mu_{\ln V}$ and $\sigma_{\ln V}$ were computed using the Japanese data. Figure 2.5 shows the Japan category means along with 95% confidence intervals and means from Yong (2016). Data for categories 2, 10, and 14 are too few to compute reliable statistics (≤ 4 observations). For those categories results from Yong (2016) were adopted as indicated in Figure 2.5. For all other categories, Japan moments were used. Based on this rationale, recommended moments for each category are given in Table 2.5.

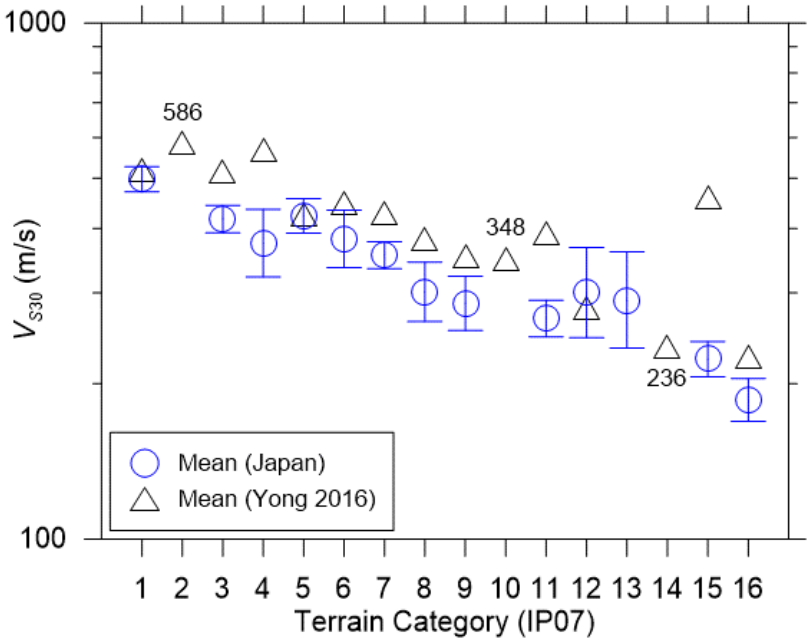

Figure 2.5. Comparison of V_{S30} means for Iwahashi and Pike (2007) terrain classes for Japan (Yong 2016) and Japan. Classes 2, 10, and 14 are poorly populated in the Japan data set, and the values written in the figure from California are used for application in Japan.

Table 2.5. Moments for Iwahashi and Pike (2007) terrain classes for application in Japan. Moments adopted from California are shown in parenthesis.

IP07 Terrain Class	N	μ_{lnV} (m/s)	σ_{lnV}	μ_{lnV} of Y16 (m/s)	σ_{lnV} of Y16
1	205	498	0.411	519	0.38
2 (CA)	4	(586)	(0.16)	586	0.16
3	212	416.8	0.456	517	0.38
4	29	374.6	0.415	568	0.46
5	120	422.8	0.43	425	0.37
6	6	381.6	0.16	448	0.14
7	245	354.9	0.479	429	0.38
8	16	301.4	0.27	382	0.32
9	46	286.4	0.421	353	0.16
10 (CA)	0	(348)	(0.09)	348	0.09
11	98	267.6	0.412	392	0.48
12	12	300.2	0.355	281	0.20
13	22	290.5	0.513	NA	NA
14 (CA)	2	(236)	(0.14)	236	0.14
15	83	223.3	0.365	460	0.52
16	40	186.1	0.309	225	0.20

The other proxy that was considered was introduced by Matsuoka et al. (2006) for categories within the "Japan Engineering Geomorphologic Classification Map" (JEGM). The JEGM utilizes geomorphology, surface geology, slope angle, and relative relief to classify locations into geomorphologic units. The empirical correlations are based on shear-wave velocity profiles from 1937 sites (this is a different data set than that compiled for NGA-Sub). The categories have been modified somewhat since 2006 (new categories added), with the list as of 2013 provided in Table 2.6 (from Wakamatsu and Matsuoka 2013; three categories that do not have stations are omitted). Also shown in the table are: (1) category means and standard deviations as provided by Matsuoka et al. (2006) and (2) the minimum, maximum, and median of the mapped V_{S30} values for the category, which differ from the Matsuoka et al. (2006) category mean due to varying morphological influences within the categories (from changes in slope angle and relative relief). Categories 1-4 correspond approximately to rock conditions, 5-7 are transitional categories,

and categories of 8 and above represent various soil conditions. Matsuoka et al. (2006) provide intra-category regressions against elevation for categories 8-13, against slope for categories 3, 5, and 8-11, and against distance from hills for categories 8, 10, 13, 15, and 18-19. We used JEGM maps and associated values of V_{S30} at 7.5 arc-sec grid-size resolution by Wakamatsu and Matsuoka (2013). No modifications to these V_{S30} values were applied.

Table 2.6. JEGM site categories, within-category moments from Matsuoka et al. (2006, "Mea06"), and attributes of mapped V_{S30} within categories from Wakamatsu and Matsuoka (2013) JEGM.

Cat.	Description	μ_{lnV} (m/s) (Mea06)	σ_{lnV} (Mea06)	Max. V_{S30} (m/s)	Med. V_{S30} (m/s)	Min. V_{S30} (m/s)
1	Mountain	707.5	0.295	775.5	708.4	641.3
2	Mountain footslope	400	0.212	400.3	400.3	400.3
3	Hill	428	0.403	526.1	408.3	294.7
4	Volcano	509	0.373	510.4	510.4	510.4
5	Volcano footslope	302	0.23	361.3	294.5	226.9
6	Volcanic hill	405	0.136	405.6	405.6	405.6
7	Rocky strath terrace	351	0.216	351.4	351.4	351.4
8	Gravelly terrace	418	0.281	589.2	466.6	252.9
9	Terrace covered with volcanic ash soils	269	0.265	418.5	270.5	197.4
10	Valley bottom lowland	345	0.364	544.2	394.5	191.5
11	Alluvial fan	323	0.267	436.2	337.5	253.1
12	Natural levee	198	0.286	267.1	201	185.2
13	Back marsh	160	0.267	192.9	166.55	140.6
14	Abandoned river channel	183	0.21	183.8	183.8	183.8
15	Delta and coastal lowland	171	0.246	207.5	168.15	141.6
16	Marine sand and gravel bars	258	0.262	260.2	260.2	260.2
17	Sand dune	194	0.283	194.5	194.5	194.5
18	Lowland between coastal dunes and/or bars	*NA*	*NA*	*NA*	*NA*	*NA*
19	Reclaimed land	182	0.283	236.3	173	149.9
20	Filled land	*NA*	0.276	253.3	188.3	152.6
21	Rocky shore, rock reef	*NA*	*NA*	429.1	429.1	429.1

Because two models (using different proxies) have been developed for prediction of the natural log mean and standard deviation of V_{S30}, a weighted combination of the two estimates is needed. An approach that has the objective of minimizing the standard deviation of the estimate

of V_{S30} that results from the combination of the two proxies was applied (mathematical formulation given in Kwok et al. 2018). This approach was originally developed as part of NGA-Sub for V_{S30} assignments in Japan by D.Y. Kwak (2016, *pers. comm.*). The weights applied to the two proxies depend on the standard deviation of residuals for the respective categories used in each proxy ($\sigma_{\ln V}$) and the degree of correlation between proxies. Using the Japan PDB compiled for NGA-Sub, the standard deviations for the dataset as a whole are 0.44 for the prediction model based on terrain categories, and 0.30 for the JEGM-based model; however, different combinations of categories between proxies can lead to more similar dispersions. For example, JEGM category 3 (Hill) has $\sigma_{\ln V} = 0.40$ and is shared for some sites with terrain class 15 ($\sigma_{\ln V} = 0.37$).

The correlation coefficient is calculated using normalized residuals:

$$\varepsilon_i = \frac{\ln\left(V_{S30}\right) - \mu_{\ln Vi}}{\sigma_{\ln Vi}} \tag{2.3}$$

where index *i* refers to model 1 (terrain) or 2 (JEGM) and $\ln(V_{S30})$ is a measured value from the profile database. Figure 2.6a shows that the two sets of residuals are relatively weakly correlated with $\rho_{12} = 0.68$ for the data set taken as a whole. This correlation coefficient is used for all category combinations. Based on this correlation coefficient and the respective $\sigma_{\ln V}$ values, the JEGM model typically receives higher weights. For example, the combination of category 15 for both JEGM and terrain provides $\sigma_{\ln V} = 0.25$ and 0.37, respectively, which gives a weight of essentially unity to JEGM with a combined standard deviation of 0.25. On the other hand, for sites with JEGM category 3 and terrain class 15, the weights are 0.35 (JEGM) and 0.65 (terrain). Each combination of categories in the application of the two models receives a unique set of weights.

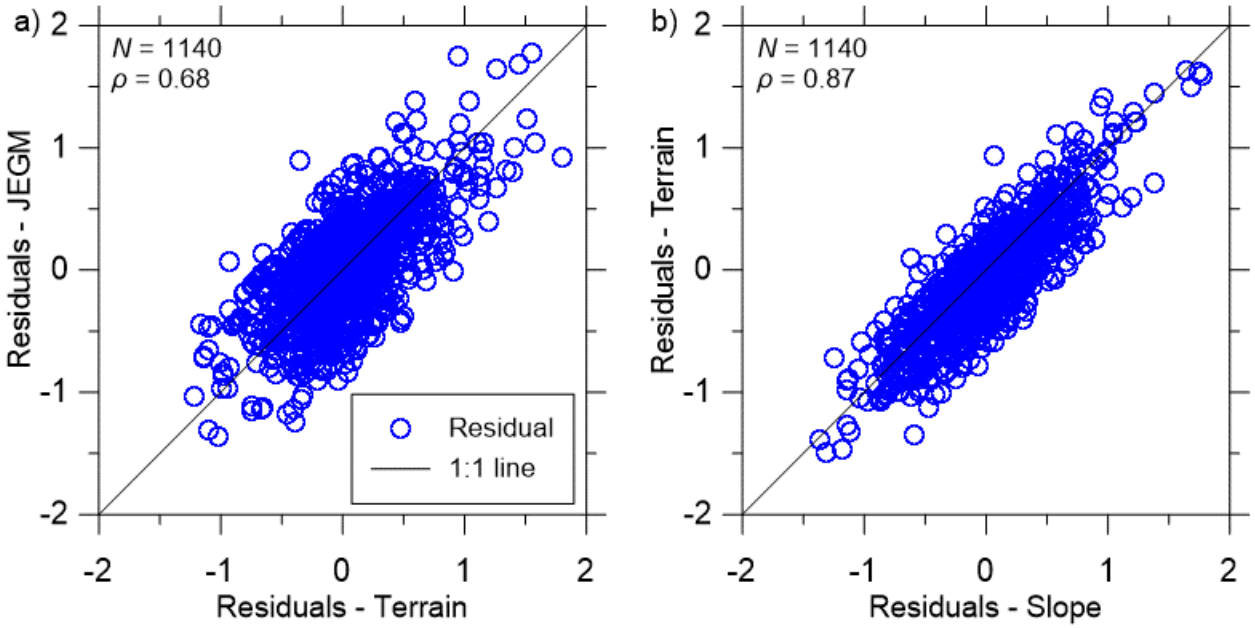

Figure 2.6. Correlations of residuals from proxy-based V_{S30} estimates using data from Japan. (a) modest correlation using JEGM and terrain proxies; (b) strong correlation using terrain- and slope-based proxies.

Additional prediction models based on other proxies were considered for use in Japan, including a geotechnical classification scheme (Chiou and Youngs 2008, as updated in Seyhan et al. 2014) and topographic slope (Wald and Allen 2007; Allen and Wald 2009). The geotechnical scheme was not used because category assignments are subjective, and because the dispersion ($\sigma_{\ln V}$) is larger than for the conceptually-similar JEGM approach. Topographic slope was not used because of strong correlation with the terrain-based approach (ρ = 0.87), which is shown in Figure 2.6b.

(b) Central and South America, other than Chile

As discussed in Section 2.2, V_S profile data is limited in Central and South American countries for which ground motions have been recorded from Cocos and Nazca subduction zone events (excluding Chile, Section 2.3.3c). This applies to Mexico, all of Central America, and all

of South America outside Chile.

There was limited access to geologic maps for these regions. A series of small-scale (1:250,000) maps of Mexico from the Servicio Geológico Mexicano were identified (SGM 2017). Geologic maps for other regions in Central and South America (aside from Chile) were not accessed.

As a result, Approach III was applied to these regions using the Iwahashi and Pike (2007) terrain categories as the proxy. The selected source region for the model is California, using category moments from Yong (2016). An epistemic uncertainty of $\sigma_{ep} = 0.2$ is assigned.

(c) Western Canada and Alaska Panhandle

Some stations, particularly in the majority of British Columbia, the Yukon Territory, and the Alaska Panhandle (seven sites, south of 60°N latitude), recorded events in both the Cascadia and Alaska subduction zones but were beyond the geographic extent of applicability of either the PNW- or Alaska-specific V_{S30} prediction models. The line of latitude at 55°N as depicted in Figure 2.7 (white line) is used to divide these stations by region (Alaska to the north, Cascadia to the south) for the purpose of creating a unique list of stations across the SDB. Hence, a regional flag for western Canada and the Alaska panhandle is not present in the SDB, but procedures were needed to assign site parameters in these sections, which are given here.

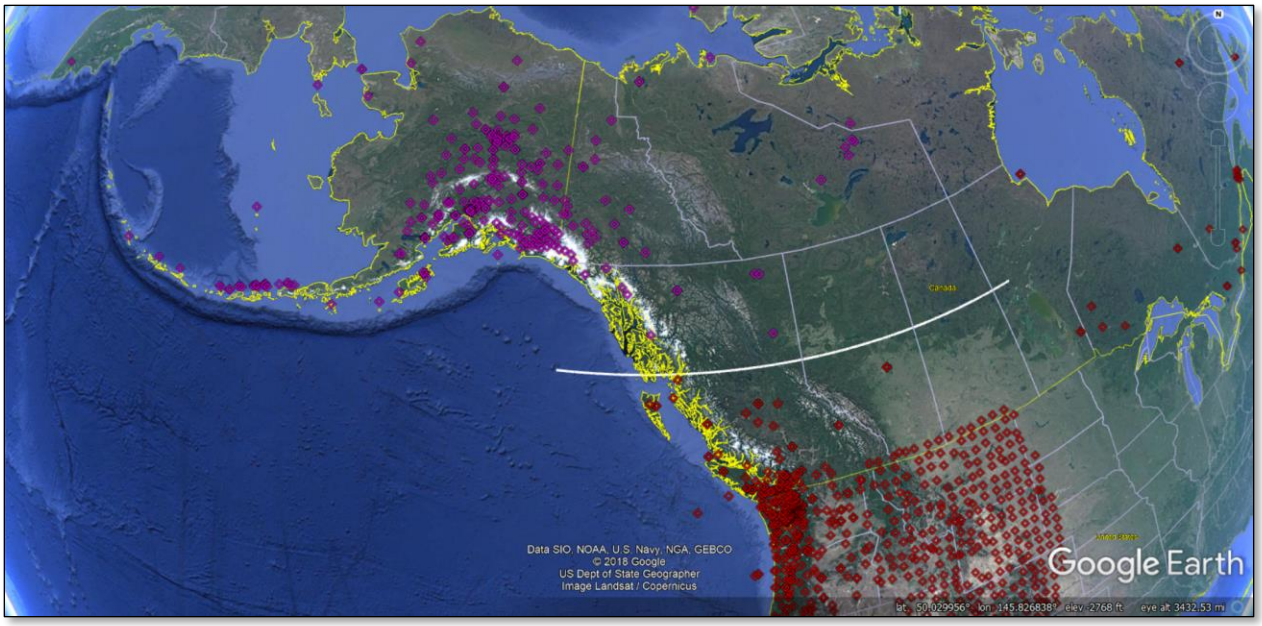

Figure 2.7. Map of stations in North America that have recorded Alaska events (purple icons) and Cascadia events (red icons), divided by the 55°N parallel (white line).

The region encompassing western Canada and the Alaskan Panhandle was divided into different tectonic regimes based on the bedrock geological map shown in Figure 2.8. A thrust fault marks the eastern edge of deformation in the Canadian Cordillera that was associated with the Laraimde Orogeny of the Late Cretaceous period, east of which lies the relatively-undeformed Canadian Shield geographic province. V_{S30} is assigned to all stations southwest of this thrust fault with a geometric mean of the PNW and Alaska models, for groups for which the V_{S30} moments differ; otherwise, the PNW V_{S30} moments are assigned. An epistemic uncertainty is also applied (σ_{ep} = 0.2). Sites northeast of this fault lie in the Canadian Shield, and we assign V_{S30} moments from the model developed for CENA by Parker et al. (2017).

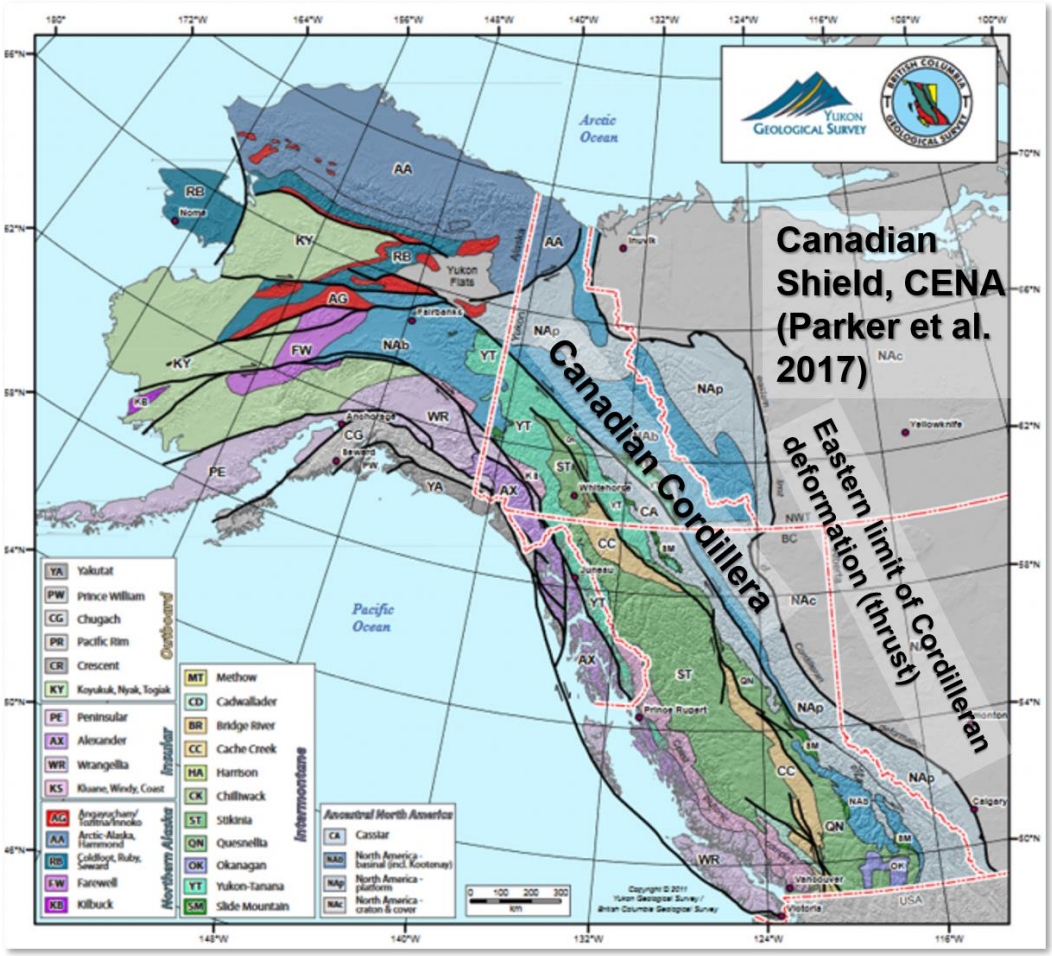

Figure 2.8. Compilation bedrock geological map of terranes comprising the Canadian Cordillera in western Canada and Alaska. A thrust fault (thick black line) marks the "eastern limit of Cordilleran deformation"; the Canadian Shield lies east of this fault. Figure modified from Colpron and Nelson (2001).

2.4 V_{S30} ASSIGNMENTS

described in preceding sections, assignments of V_{S30} can be made using a wide range of methods with variable levels of associated variability and uncertainty, depending on the availability of V_S measurements and locally-calibrated proxy-based V_{S30} prediction models. For each site in the NGA-Sub SDB, a preferred median V_{S30} ($\mu_{\ln V}$) and an associated variability ($\sigma_{\ln V}$) were assigned. In cases where mean estimates carry large epistemic uncertainty, a standard deviation on the mean (σ_{ep}) was also assigned. The process by which these assignments are made

is shown in Table 2.7 (the codes are given in the site database file).

Table 2.7. Protocol used in NGA-Sub for assignment of preferred V_{S30} and related parameters.

Code	Description
0	V_{S30} computed using profile with $z_p \geq 30$ m. Standard deviation taken as $\sigma_{\ln V} = 0.1$. Epistemic uncertainty on mean not assigned.
1	Profile is available but maximum V_S profile depth $z_p < 30$ m. V_{S30} is estimated using an extrapolation relationship, preferably with region-specific regression coefficients, e.g. those in Kwak et al. (2017a) for each region, for use with the method described in Dai et al. (2013). $\sigma_{\ln V} = \sqrt{\sigma_e^{\,2} + 0.1^2}$ (Equation 3.10). Values of σ_e given in Kwak et al. (2017a).
1.5	Estimate V_S profile from standard penetration test blow counts and local correlations between V_S and penetration resistance/effective stress (this correlation is only used in Japan; Kwak et al. 2015). V_{S30} computed from estimated profile. $\sigma_{\ln V} = 0.25$.
2	No profile available. Mean V_{S30} estimated using region-specific models based on geology or hybrid geology-morphology proxies. This code applied in PNW, portions of Alaska and Canada, Japan, Taiwan, and New Zealand. $\sigma_{\ln V}$ assigned based on category statistics.
3	No profile available. V_{S30} and its variability are estimated using region-specific models based on geomorphic terrain categories (Iwahashi and Pike, 2007). This code used in Chile, Japan, and California (for Cascadia events). $\sigma_{\ln V}$ assigned based on category statistics.
4	No profile available. Mean V_{S30} estimated using models developed for source region other than the target region. Source region models can be based on geology, hybrid geology-terrain proxies, or geomorphic terrain categories. This code applied in portions of Alaska, Central and South America (outside of Chile), and western Canada and the Alaska Panhandle. $\sigma_{\ln V}$ assigned based on source region category statistics. Epistemic uncertainties (σ_{ep}) assigned.

2.5 BASIN DEPTH TERMS

2.5.1 Overview

Basin depth terms as used in GMMs are defined as vertical distances from the ground surface to the first occurrence of a particular V_S horizon. These depths are used to provide a first-order representation of basin geometry in alluvial or sedimentary basin environments. Commonly used basin depth terms are $z_{1.0}$. and $z_{2.5}$, which are depths to the $V_S = 1.0$ km/s and 2.5 km/s velocity horizons, respectively, and were used in four of the five NGA-West2 GMMs (Gregor et al. 2014).

Depth parameters are assigned for a site in the SDB from a measured in situ V_S profile that meets or exceeds the specified velocity horizon, or when such profiles are not available, from a 3D seismic velocity model for a particular region. An exception is New Zealand, where Kaiser et

al. (2016, 2017) provide depth $z_{1.0}$ from profiles where available, and otherwise estimate it using various methods unrelated to a 3D seismic velocity models (resolution indicated by quality flags Q1-Q3). Table 2.8 summarizes the number of sites assigned basin depth terms in the NGA-Sub SDB from both of these assignment protocols. A significant portion of sites (2350/6433, 39%) do not have an assignment of a basin depth term. Such terms are not compiled in Alaska and Central America & Mexico. Only three sites have this parameter assigned in South America.

Table 2.8. Summary of basin depth terms included in NGA-Sub SDB for various regions. "Estimated" depths are from 3D models, with exception of New Zealand.

Region	Total Sites in SDB	$z_{1.0}$			$z_{2.5}$		
		Assigned	Measured	Estimated	Assigned	Measured	Estimated
Cascadia	1126	15	15	*NA*	458	1	457
Japan	2283	2228	609	1619	2021	113	1908
New Zealand*	412	412	29	383	0	0	*NA*
South America	942	3	3	*NA*	0	0	*NA*
Taiwan	849	802	0	802	*NA*	*NA*	*NA*

2.5.2 Cascadia 3D Velocity Models

The USGS has developed a 3D seismic velocity model for the PNW, as first presented by Stephenson (2007) and recently updated by Stephenson et al. (2017). The model was developed to support seismic hazard studies and ground motion simulations. The model encompasses a region from approximately 40.2°N to 50°N latitude, and from about 122°W to 129°W longitude, and 0-60 km depth.

As described by Stephenson (2007), the backbone of the velocity model is a geologic model encompassing six units, as shown in Figure 2.9: (1) continental sedimentary basins (subdivided into Quaternary and Tertiary basin units), (2) continental crust, (3) continental mantle, (4) oceanic sediments, (5) oceanic crust, and (6) oceanic mantle. Some details of the geologic structure related to the Seattle fault are included in the 2017 update. The structure of sedimentary basins is described

separately for Quaternary and Tertiary basins.

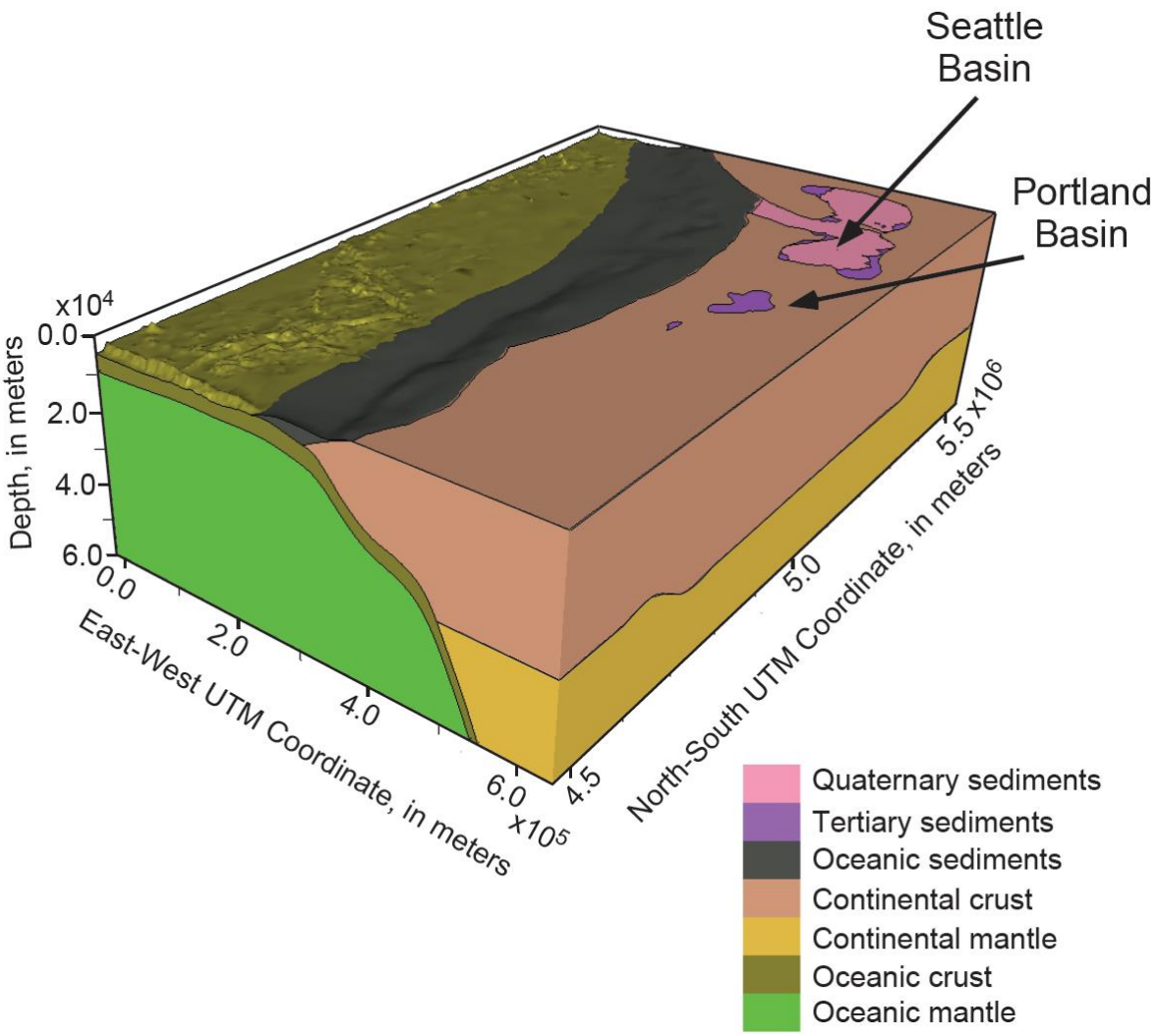

Figure 2.9. 3D representation of Cascadia geology as employed in velocity model of Stephenson (2007) and Stephenson et al. (2017) (source: Stephenson et al. 2017).

The lateral limits of Quaternary basins (all of which are in the Puget lowlands) are based on a smoothed representation of the Quaternary-Tertiary contact from the Schuster (2005) surface geologic maps for Washington state (1:500,000-scale), The lateral limits of Tertiary basins in the Puget lowlands are based on the V_P = 4.5 km/s isocontour from the Seismic Hazards Investigations in the Puget Sound (SHIPS, Brocher et al. 2001) and P-wave seismic tomography (Ramachandran

et al. 2006). The 4.5 km/s 3D isosurface was projected to the ground surface to define the basin boundary. The Portland area has only Tertiary basins in the USGS model, and the boundaries of these basins are based on depth to bedrock constrained by well data intersecting the boundary of crystalline rocks under Tertiary sedimentary deposits (Yeats et al. 1996; Gannett et al. 1998).

The thickness of Quaternary basins is constrained from borehole and seismic refraction data. Basins with Quaternary sedimentary cover less than 30 m in thickness are not included in the model, such as for the Portland and Tualatin basins or the Willamette Valley. Within Quaternary basins, a uniform V_P profile is used with values of 1,500, 1,905, and 1,980 m/s at depths of 0, 200, and 1,000 m, respectively. V_S is derived from the V_P profile using a V_P/V_S ratio of 2.5, with some exceptions at depths beyond 150 m. Velocity limits within Quaternary units are 600 m/s (minimum) and 900 m/s (maximum).

The thickness of Tertiary basins is taken as the 4.5 km/s V_P contour as derived from oil industry borehole data in the Puget Lowlands (Ramachandran et al. 2006) and from well data in the Portland area (Yeats et al. 1996; Gannett et al. 1998). V_P profiles within the Puget Lowland basins was derived from tomographic studies (Ramachandran et al. 2006). Variations on this process were used for other basins, include Willamette Valley, where V_S is evaluated using a constant V_P/V_S ratio of 2.

Within these 3D velocity models, the 1.0 km/s V_S horizon occurs within Quaternary or Tertiary basins structures, but is not considered to be particularly meaningful or useful for ground motion modeling. Rather, the 2.5 km/s V_S horizon is preferred in this region, which does not occur within these basins, but will typically occur at or near the base of basins of either age group.

As shown in Figure 2.9, surrounding and underlying basins in the 3D model is the continental crust unit. Seismic velocity structure in this unit in the Puget Lowland region is based

on V_P as derived from SHIPS tomographic data (Ramachandran et al. 2006). For the remainder of the model, seismic velocities are based on V_S derived from tomographic data from Moschetti et al. (2007). V_P and V_S are related to each other in both tomographic datasets using empirical relationships from Brocher (2005).

Within the Cascadia region, only 15 sites had a measured $z_{1.0}$ from a V_S profile and only one site had a measured $z_{2.5}$ from a profile (Table 2.8). Accordingly, almost all based depths for Cascadia is based on the USGS model. Bill Stephenson provided $z_{2.5}$ values at strong motion sites located within Quaternary and Tertiary basins (2016, *pers. comm.*). Outside of these basin structures, basin depths are given in the SDB as zero.

Figure 2.10 shows $z_{2.5}$ values at strong motion sites in the Cascadia region. Overall, there is no particular relationship between V_{S30} and $z_{2.5}$, but there are strong differences in depths within specific basin structures. The Seattle basin (southern lobe of structure defined as Seattle basin in Figure 2.9) has the largest depths. Portland has a very consistent distribution of depths between about 1.5-2.0 km. The Everett basin (northern lobe of structure defined as Seattle basin in Figure 2.9) is much shallower than the Seattle basin, as is the Georgia basin near Vancouver (not marked in Figure 2.9).

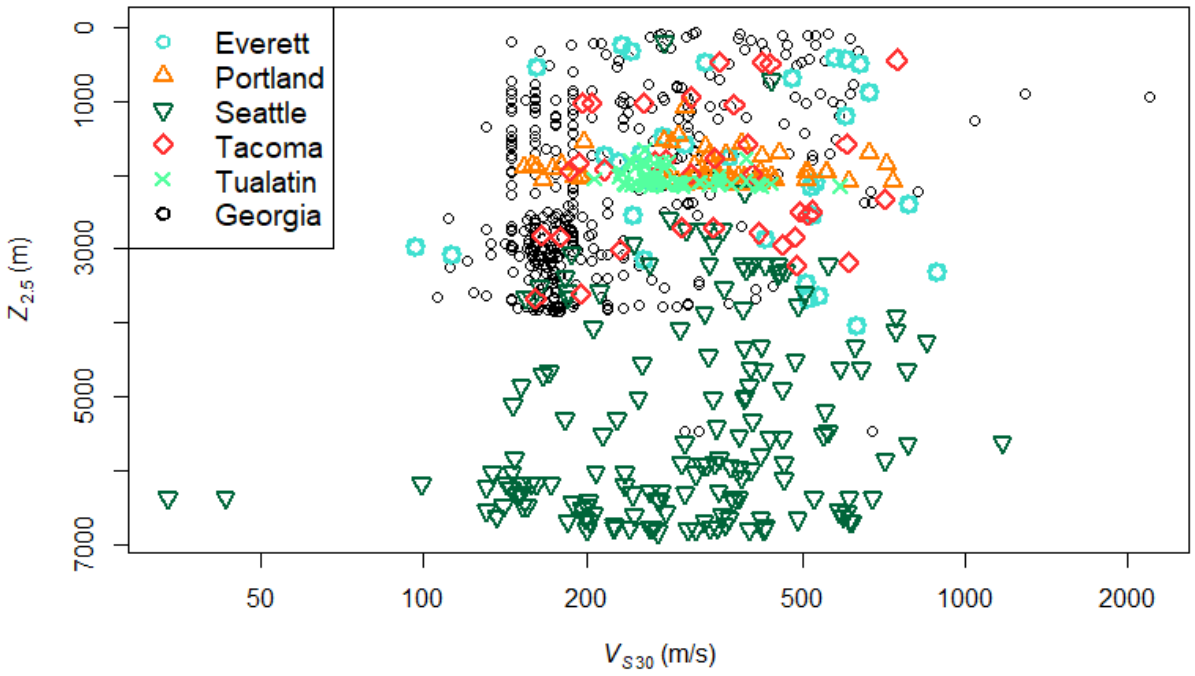

Figure 2.10. Basin depths ($z_{2.5}$) from Stephenson et al. (2017) as a function of V_{S30} for various basin structures in the Cascadia region.

2.5.3 Taiwan 3D Velocity Model

Seismic velocity models of the Western Plain and Taipei Basin of Taiwan have been developed by the Taiwan National Center for Research in Earthquake Engineering (NCREE). A Taipei basin model is presented by Lin et al. (2014) and a Western Plain basin model is presented by Kuo et al. (2016). Most of the Taiwanese sites in the SDB (94%) have been assigned $z_{1.0}$ from models presented in these studies.

Figure 2.11 shows the location of the Taipei basin, which can be identified based on elevation. For Taipei, recordings of local events at ground motion instruments were interpreted using the receiver function method (i.e., peaks in H/V spectral ratios used to infer depth to velocity contrasts). The observations were used in an inversion procedure to identify depths in an assumed six-layer profile. The *P*- and *S*-velocities of the layers were fixed to avoid trade-offs between the velocity and layer thickness. The result is a six layer, spatially variable velocity model.

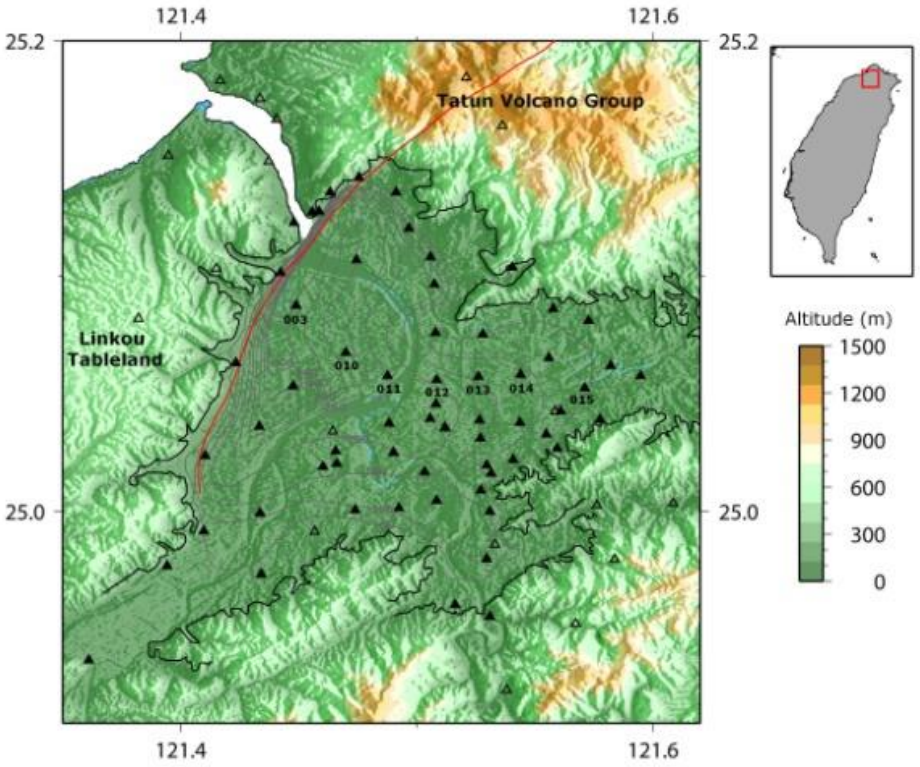

Figure 2.11. Map showing Taipei basin, as identified from topography (source: Lin et al. 2014).

Figure 2.12 shows the limits of the Western Plain. In this region, passive circular arrays were used to record surface waves, along with H/V spectra from microtremors. A frequency-wavenumber approach was used with the circular array data to provide frequency-phase velocity dispersion curves. These curves were jointly inverted with the H/V spectra to estimate shear wave velocity structure at measurement locations. These results were combined to form the seismic velocity model.

The Taipei and Western Plains velocity models were queried by C.-K. Kuo (2017, *pers. comm.*) for Taiwan strong motion sites. These results were added to the SDB for use in NGA-Subduction and are also being used in a Taiwan seismic hazard study (NCREE, 2017).

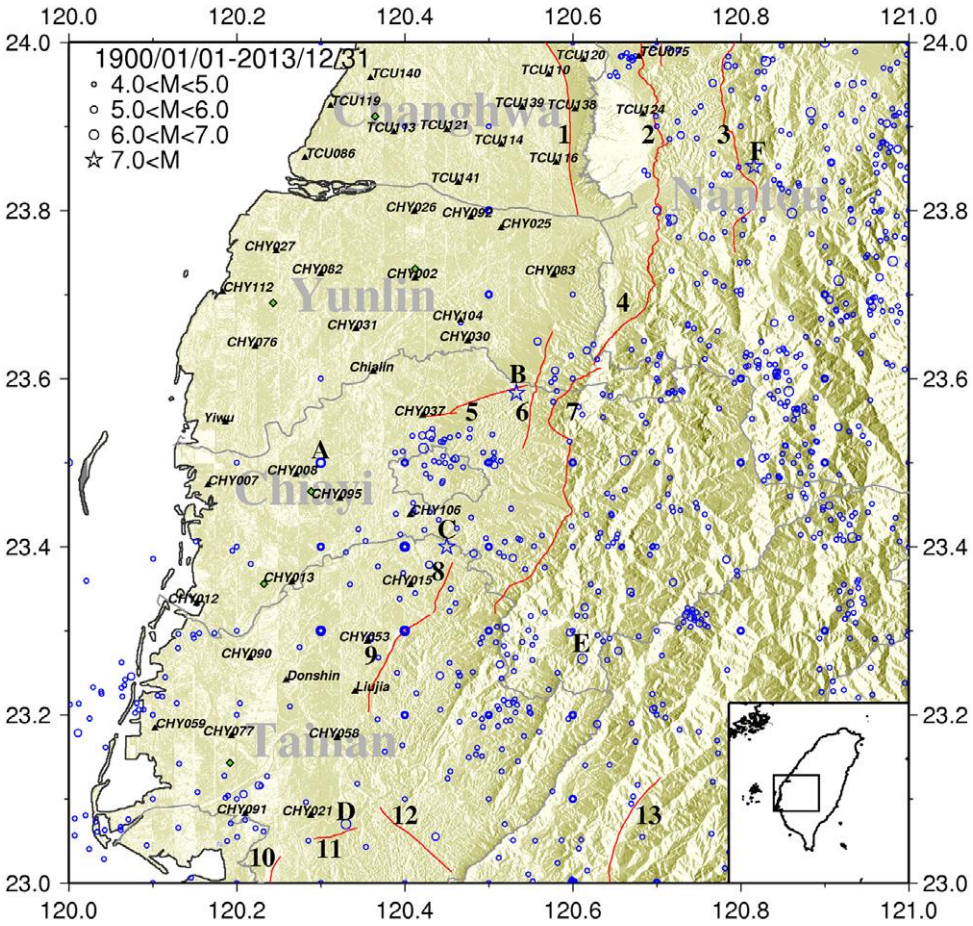

Figure 2.12. Map showing Western Plain in Taiwan (source: Kuo et al., 2016).

2.5.4 Japan 3D Velocity Model

For Japan, the NIED (National Research Institute for Earth Science and Disaster Prevention) seismic velocity model was utilized. The basin model was developed through a combination of deep boreholes, reflection and refraction surveys, micro-tremor surveys, and gravity surveys (Fujiwara et al., 2009, 2012). Tomography was used to verify and refine the basin structure. The model covers depths for x = 0.35 to 3.0 km/s. The basin depth lookups from the NIED model were performed by Dongyoup Kwak (2016, *pers. comm.*) using files accessible at http://www.j-shis.bosai.go.jp/en/. Figure 2.13 shows depth distributions in Japan based on these models. This same Japan basin model was used in NGA-West2 (Ancheta et al. 2013).

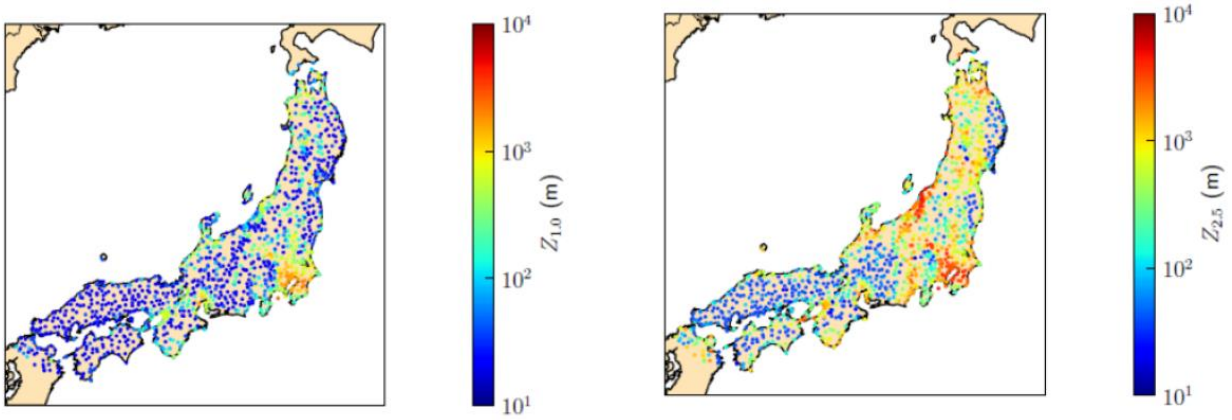

Figure 2.13. Spatial distributions of the $z_{1.0}$ and $z_{2.5}$ basin depth parameters in Japan based on NIED models (source: Ancheta et al. 2013).

2.6 VOLCANIC ARC FLAGS

Subduction zone plate boundaries produce geologic structures in the crust and upper mantle that affect seismic wave propagation. Many subduction zones are associated with a volcanic arc, where the down-going oceanic slab begins to melt, and plumes of magma rise to form volcanoes on the surface of the overriding slabs. For NGA-Sub, the volcanic arc location on land was used to categorize the forearc (trench-side) and backarc of each subduction zone region.

The delineation of the volcanic arc allows both epicentral locations and strong motion sites to be classified as forearc or backarc. Most subduction zone events occur either at the interface or within the subducting slab. Interface events are by definition at the trench and located in the forearc. Slab events would only fall in the backarc if it were located far down the slab (e.g., Figure 2.14). As a result, most slab events are in the forearc, but especially deep events can be in the backarc. Source metadata including volcanic arc flags are described in Chapter 4 of the NGA-Sub Database report (PEER 2019).

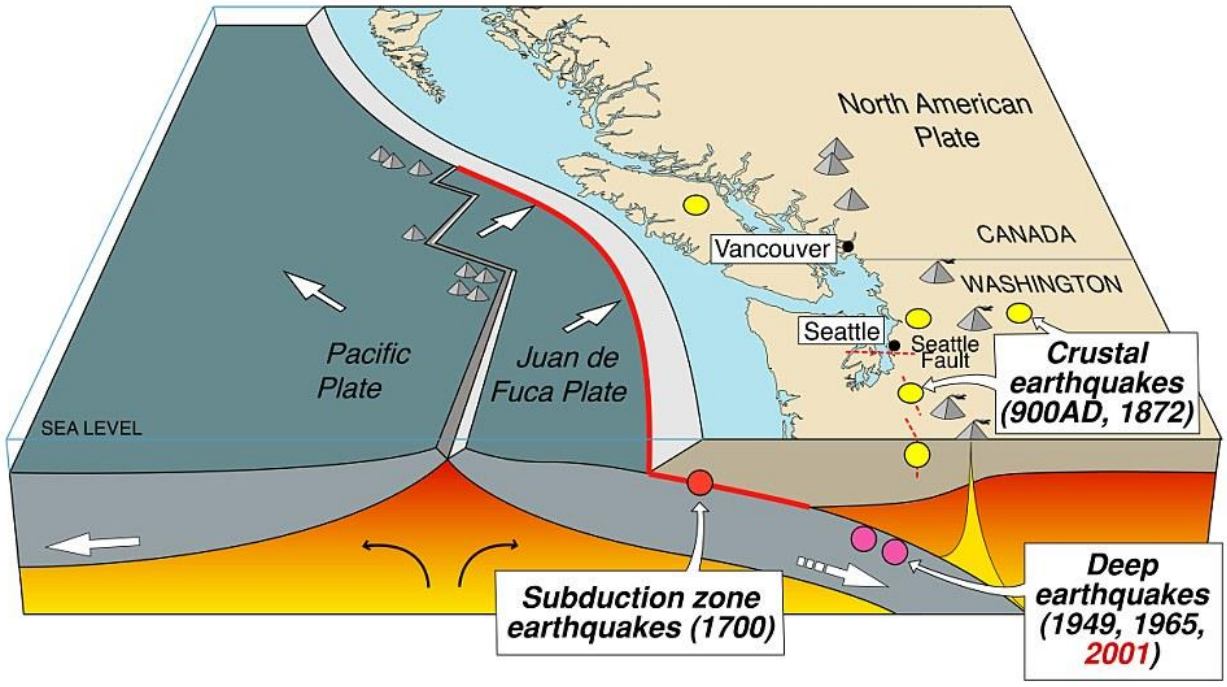

Figure 2.14. Cascadia subduction zone geometry, displaying different earthquake sources. Interface earthquakes are labeled as "Subduction zone earthquakes" and intraslab earthquakes are labeled as "Deep earthquakes". After Wells et al. (2000).

Volcanic arcs were determined for each subduction zone in the NGA-Sub database by drawing a line by eye through the average trend of volcanic peaks. The locations of volcanic peaks were obtained from the Smithsonian Institute's Global Volcanism Program (2013). Judgment was used to average out a smooth line for the volcanic arc, rather than to represent the arc as a jagged piecewise line connecting volcanic peaks. All sites in the SDB were assigned a volcanic flag using the schema provided in Table 2.9. Most backarcs correspond to Flag 1 and most forearcs to Flag 2. Japan's complex tectonic geometry required it to be separated into multiple forearc and backarc regions. There are two forearcs, relating to subduction of both the Pacific and Philippine Sea plates under Japan (Figure 2.15). The Philippine Sea plate subducts beneath the southern portion of Japan at the Ryukyu Trench; the associated forearc is designated with Flag 3.

Table 2.9. Description of Volcanic Flags.

Flag	Count	Description
0	561	Outside of volcanic arc zone
1	1927	Backarc
2	3169	Forearc
3	708	Forearc, Ryukyu Trench

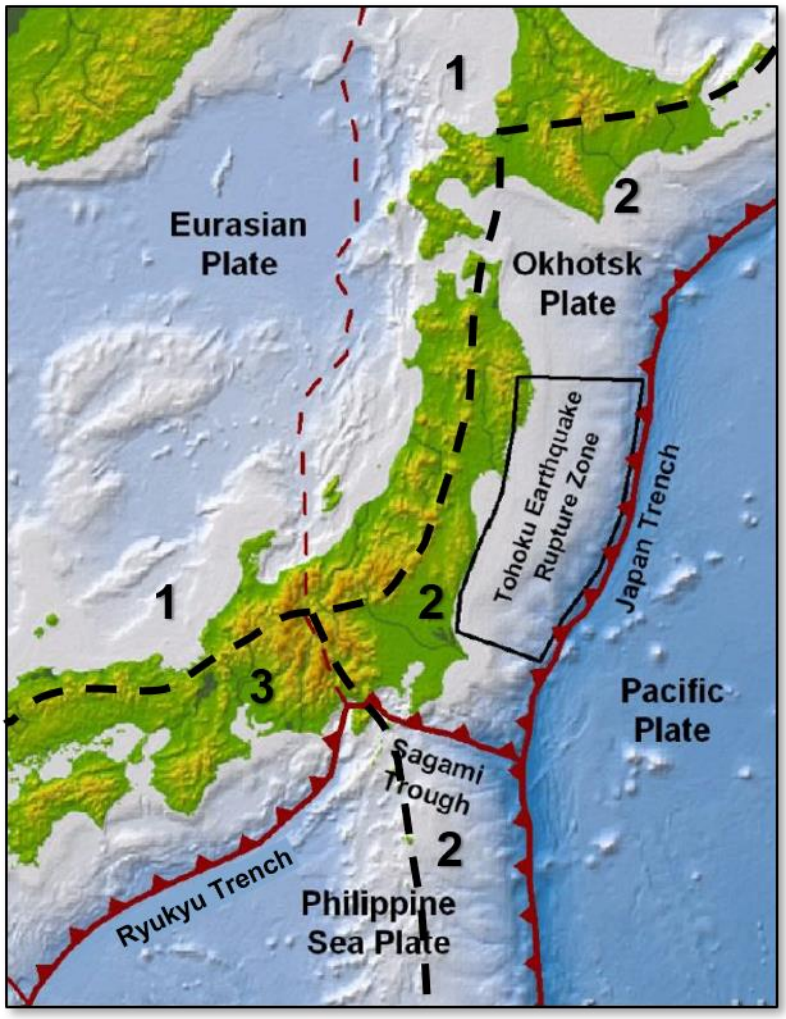

Figure 2.15. Japan's multiple subduction zones with volcanic flag regions indicated (modified from Apel et al. 2012).

In general, the edges of the subducting slab were defined by the limits of slab depth contours as defined by the USGS's Slab1.0 model (Hayes et al. 2012). Figure 2.16 shows how the edges of the slab, as well as the edges of the forearc and backarc zones, were defined for the example of the Alaska subduction zone. Areas deemed to be outside of the subduction zone (east

of the edge in Figure 2.16) are assigned Flag 0. Also shown in Figure 2.16 are locations of volcanoes from the Global Volcanism Program (2013) (orange symbols) and the volcanic arc passing through them (white line between zones 1 and 2). This task was repeated for all of the NGA-Sub regions except Taiwan. All Taiwan sites received a forearc flag assignment of 2, as the island largely consists of forearc accretionary sediments uplifted due to the oblique northeastward subduction of the Philippine Sea plate with respect to the Ryukyu trench, flipping to eastward subduction of the Eurasian plate in the South China Sea (Ho 1986 [Fig. 3], Chemenda et al. 1997, Lundberg et al. 1997).

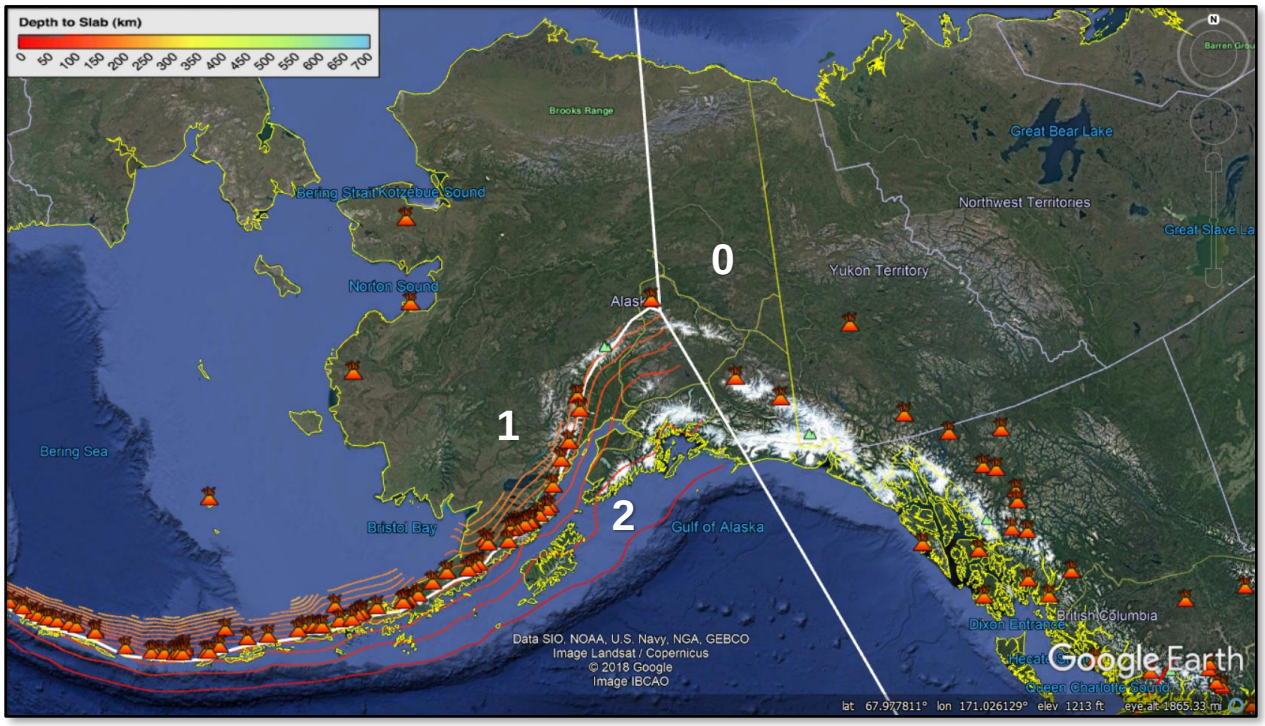

Figure 2.16. Example of volcanic flag region extents in Alaska/Aleutians subduction zone. The white lines follow the trend of the eastern edge of the subducting slab, for which contours of depth are obtained from the USGS Slab1.0 project (Hayes et al. 2012).

The volcanic arc regions described in this section were used to compute portions of path lengths within each zone for use in GMM development. The computation of path lengths is described in Chapter 4 of the NGA-Sub Database report (PEER 2019).

3 Development of V_S Profile Database and Proxy-Based Models for V_{S30} Prediction in the Pacific Northwest Region of North America

3.1 INTRODUCTION

This chapter presents methods and recommendations for estimating V_{S30} in the Pacific Northwest (PNW) region of North America using Approach I as described in Section 2.3.2. The major initial application of these methods was for estimating site parameters at SMSs that have contributed data to the NGA-Sub project (Chapter 2 of this dissertation and in Kishida et al. 2017) and for which geophysical data are not available. Additional applications include estimation of site parameters for forward predictions of ground motions in the PNW and potentially other regions given the robustness of the dataset collected and described herein. The natural log means and standard deviations of V_{S30} were estimated from secondary information derived from surface geologic maps and digital elevation models, which is generally available at large scales across the study region. These maps provide site-specific proxies (further described in Chapter 2 of this dissertation), which include surface geology categories, topographic slope, and geomorphic terrain classes, which in turn describe spatial variations of V_{S30} moments. A crucial element of this effort was the compilation of a database of V_S profiles for the PNW, which contains information from 928 sites in Oregon, Washington, and British Columbia.

Following this introduction, previous studies on proxy-based V_{S30} estimation for other

geographic regions are briefly reviewed (see 2.3.1 for further detail). No such prior studies have been completed for the PNW region, which has unique conditions associated with its history of glaciation and volcanic activity. The profile dataset and the metadata compiled for proxy development are then described, including information sources and the level of detail/resolution. The development of proxy-based models for the region is then described, and results of which are compared to findings from studies in other regions. Next, the procedure by which the assembled geophysical data and proxies were used to populate the NGA-Subduction project site database for ground motion recording sites in the PNW regions of the U.S. and Canada is described. Similar protocols to those for NGA-West 2 and NGA-East are followed (e.g., Seyhan et al. 2014; Parker et al. 2017).

3.2 PREVIOUS STUDIES

An overview of proxy-based V_{S30} prediction models using global and regional data sets and predictor variables is provided in Section 2.3. Similar models were not found in the literature for the PNW region. However, two closely related studies are of interest. Bilderback et al. (2008) assembled 243 V_S profiles, mostly in Washington State, but also including data from Portland, OR. For each of 48 geologic units spanning from Quaternary sediments to bedrock, they develop mean V_S values binned by within-layer depths. Models of this sort provide insight into the velocity characteristics of specific units but cannot be used to estimate V_{S30} in a generic sense, because stratigraphy is unknown for most locations. Roe and Madin (2013) performed a similar study for Portland, OR. They developed mean V_S-depth distributions for 23 geologic units spanning from young sediments to rock. These models are used with a three-dimensional geologic model for Portland (also from Roe and Madin 2013), which allows look-up of unit thicknesses for any geodetic coordinate. In this manner, they develop V_S profiles across the study region, from which

V_{S30} values are computed. Their results are presented in the form of site category maps derived from V_{S30} using building code site category definitions (Dobry et al. 2000).

The work described in this chapter was undertaken because: (1) while many proxy-based V_{S30} estimation procedures are present in the literature (Table 2.3), they were developed for other regions having different geology from the PNW, where glaciation and volcanism are notable factors; (2) prior studies in the PNW region were intended for within-layer V_S estimation, which is different from the present objective of V_{S30} estimation directly from broadly-available geospatial information (e.g. surface geology and DEMs); and (3) methods for estimation of V_{S30} from proxies are needed for applications including GMM development (e.g., the NGA-Subduction project) and regional ground motion hazard and risk studies.

3.3 V_S PROFILE DATABASE (PDB) FOR PNW

3.3.1 Database Attributes

The V_S profile database (PDB) is a digitized collection of 917 V_S profiles from various sites throughout the PNW, with an additional 11 sites for which a measured V_{S30} is available, but not a V_S profile. (It is noted here that in this chapter, "PDB" refers to a data compilation of V_S profiles, not a relational database as described in Chapter 6 of this dissertation.) V_S profile measurements were not performed as part of the present work; rather, results of prior studies are utilized. Source documents include reports from national and state geologic surveys (USGS; Geological Survey of Canada, GSC; Washington Division of Natural Resources, DNR; Oregon Department of Geology and Mineral Industries, DOGAMI) and university research units, among others (see Section 3.10, *Data and Resources*). Extensive correspondence was undertaken with numerous agencies and researchers to secure, interpret, and digitize this data, all of which is non-proprietary and had disparate formats in source documents. Source details for each site in the PNW PDB are given in

Ahdi et al. 2017a (available in Table S2 in the electronic supplement to that article). Site data was included in the PNW PDB if the V_S profile is based on geophysical testing, extends to a maximum (profile) depth $z_p \geq 6$ m, begins within 5 m of the ground surface, geodetic coordinates (i.e., latitude and longitude) are known, and the geophysical measurement technique is considered credible based on experience and judgment, at least for obtaining V_{S30}. As mentioned previously, the present study does not formally encompass Alaska, although it is noted here that an additional 126 sites in Alaska with measured V_S or V_{S30} data were compiled (see Chapter 4 of this dissertation and Ahdi et al. 2017b).

The contents of the PDB are twofold. First is a data file containing the as-reported V_S profile information for all 928 sites, consisting of site identification number, interval depth ranges, and interval velocities (Table S1 in the electronic supplement to Ahdi et al. 2017a). Second is a summary file containing:

1. Information on the measurements themselves, including geophysical method, profile depth (z_p), site location, and data source;
2. Summary data on the V_S profile, including time-averaged velocities to various horizons (V_{S6}, to V_{S30} at 2 m intervals, V_{S30} to V_{S70} at 10 m intervals, and V_{SZ});
3. Metadata related to site location, consisting mainly of proxies and their sources.

The V_S profile data will be disseminated as part of the U.S. Community V_S Profile Database project, described in Chapter 6 of this dissertation. Both PNW PDB files are included as electronic appendices to Ahdi et al. 2017a (Tables S1 and S2, available in the electronic supplement).

Geophysical methods represented in the PNW PDB include invasive methods such as borehole-based downhole and crosshole (152 profiles), seismic CPT (SCPT, 165), and P-S suspension logging (40). Also included are results from non-invasive methods (632 profiles total)

such as spectral analysis of surface waves (SASW, Stokoe et al., 1994; 43 profiles), multi-channel analysis of surface waves (MASW, Park et al., 1999; 67 profiles), microtremor-array and ambient noise measurements (MAM, 78 profiles), and P- and/or S-wave seismic reflection/refraction (Telford et al., 1990; 490 sites). MAM methods represented in the compiled reports include various 1D and 2D approaches using the Spatial Autocorrelation (SPAC) or the Extended Spatial Autocorrelation (ESAC) analysis methods (Kanai and Tanaka, 1954; Horike, 1985; Okada et al., 1990). Within the PNW PDB, 100 sites have profiles derived from multiple non-invasive measurement methods that were used to develop composite site dispersion curves, which in turn were used for the inversion of site V_S profiles. Sites where MAM or other ambient measurement geophysical methods were used all coincided with sites with multiple measurements, except for 11 sites utilizing 2D MAM arrays in southern British Columbia (S. Molnar 2016, *pers. comm.*). Among the PDB V_S profiles, 197 have first-layer thicknesses greater than 30 m, of which 192 are derived from seismic refraction/reflection. There was less confidence in these profiles, as they may not accurately represent the near-surface velocity gradient with depth. As such, some adjustments were made to the V_{S30} values computed from these profiles, as described further in Section 3.3.2 below.

Figure 3.1 shows the locations of PDB sites, which occur throughout Washington, Oregon, and southern British Columbia, but have concentrations in Portland, OR, Seattle, WA, and the Fraser River Delta region south of Vancouver, BC. Most measurements are located within alluvial or basin regions, which provides sampling bias for the V_{S30} values towards Quaternary geology at the expense of Tertiary and Mesozoic units. Figure 3.2 shows a histogram of V_{S30} values from the PNW PDB, which illustrates the concentration of data at slow velocities typical of soil sites. The number of sites used in Figure 3.2 and elsewhere in this paper do not match the total number of

sites in the PNW PDB, as some profiles were deemed unsatisfactory for computing V_{S30} (indicated in the "V_{S30} Comment" column of the PDB, Table S2 of the electronic supplement to Ahdi et al. 2017a). Figure 3.1 also shows locations of SMS sites used in the NGA-Sub project (Chapter 2 of this dissertation) from which it is evident that most PDB sites are not co-located with SMS sites (only 79 of 459 SMS sites in the PNW region have a measured V_S profile within 300 m of the instrument).

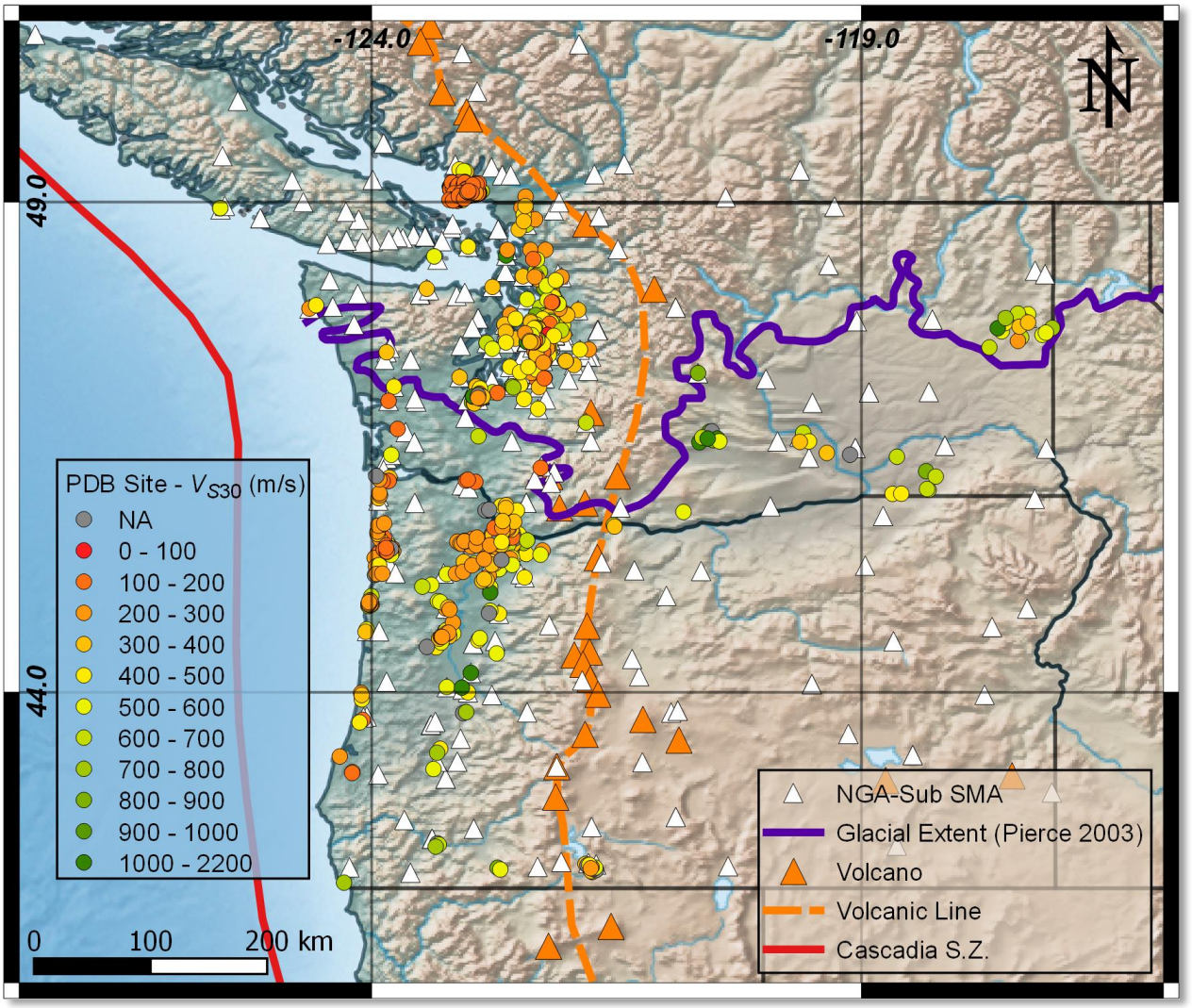

Figure 3.1. Map of PNW region showing locations of PDB and SMS sites with respect to the glacial extent as mapped by Pierce (2003) and the volcanic line separating forearc (west) from backarc (east) sites. NA assigned to PDB sites with V_{S30} not used in analyses.

3.3.2 V_{S30} Computation

V_{S30} is computed using Equations (2.1) and (2.2), as described in Section 2.2.2. When $z_p \geq$ 30 m, which occurs for 567 of 910 PDB sites having useable V_S profiles, V_{S30} is computed by replacing z_p with 30 m in Equations (2.1) and (2.2). Figures 3.3 shows a histogram of the distribution of z_p for sites with V_S profiles not derived solely from seismic refraction or reflection methods (182 of which have much deeper profiles, with z_p between 2500 and 5500 m).

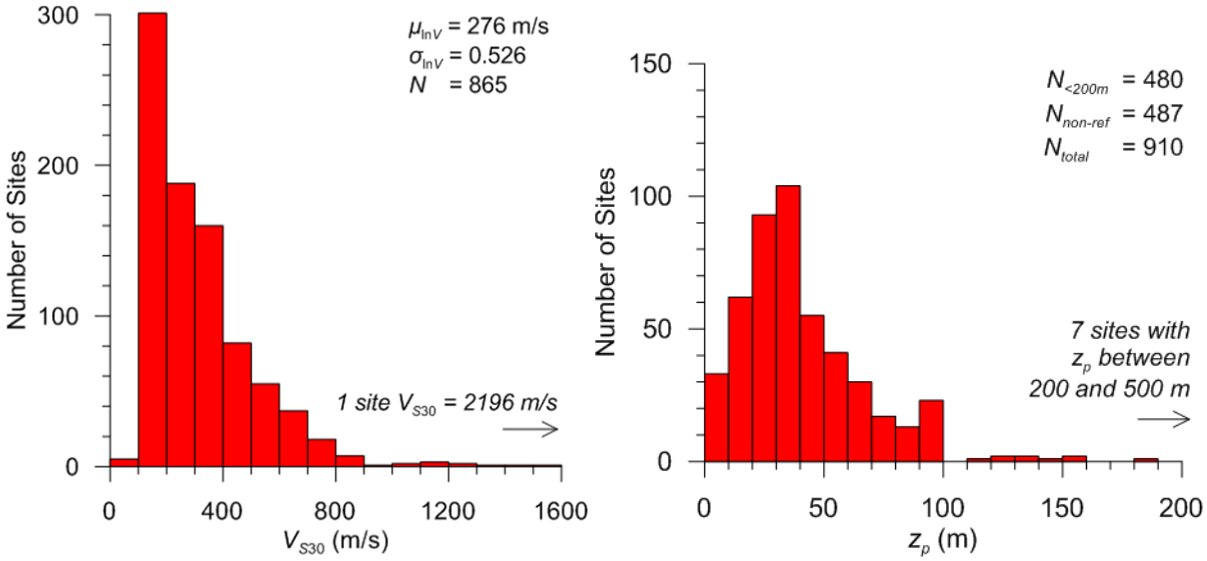

Figure 3.2 (left): Histogram of V_{S30} for all sites in PNW PDB (BC, OR, WA).
Figure 3.3 (right): Histogram of z_p for sites in PNW PDB. Sites with profiles measured using only seismic refraction/reflection geophysical methods are excluded. For sites with a V_S profile that includes a halfspace, z_p is taken as the depth to the halfspace.

As noted previously, 197 sites have thick first layers (> 30 m, ranging from 70 to 240 m) and were mainly derived from seismic reflection or refraction methods, which are coarse near the ground surface. As elaborated in the *Residuals Analysis* section (3.7.1) below, V_{S30} estimates from these sites were found to be generally biased towards faster velocities than other sites sharing similar proxy attributes but with higher-resolution profiles. These sites, mostly in the Frasier River Delta region of British Columbia, are principally on Holocene soils of various types (177 profiles), with 15 on older sediments and 5 lacking geology assignments. The bias occurs for Holocene or

recent sediments, is smaller magnitude for older deposits, and is not differentiated by depositional environment within the Holocene age group. Moreover, the mean bias is nearly constant with respect to first-layer thickness over the range of ~70–175 m, and markedly increases for thicker first layers. The mean bias for Holocene sites with first layer thickness < 175 m is approximately 0.23 in natural log units, and the mean bias for sites on older deposits is approximately 0.07 in natural log units. These biases can be understood to result from gradients in the V_S profiles that are present at the sites, but not reflected in the profiles. For example, if the first layer is 100 m in thickness, the first layer velocity in effect is V_{S100}, which is a poor estimate of V_{S30} for sites with velocity gradients. While the challenges associated with the use of these profiles was recognized, the data was retained because (1) the first-layer velocities represent a site property (time-averaged velocity over the first-layer thickness); (2) the difference between measured first-layer velocity and V_{S30} has a clear physical basis (profile gradient effects) that can be corrected for in an approximate manner; and (3) the value of these observations, which enrich the data set, over-rides the uncertainty associated with the applied correction. Accordingly, for the development of final proxy relations, this study (1) did not consider 17 profiles with first layer thicknesses > 175 m (such sites had appreciably larger bias); (2) for Holocene sites, V_{S30} values computed directly from reported profiles were reduced by multiplying by exp(-0.23) (arithmetic units); and (3) for sites on materials older than Holocene, a similar reduction of exp(-0.07) was applied.

For sites having z_p < 30 m, V_{S30} must be estimated by extrapolation. Procedures for such extrapolations have been proposed by Boore (2004) and Boore et al. (2011) based on analysis of data from California and Japan, respectively. Other proposed procedures include those of Dai et al. (2013), Wang and Wang (2015), and Midorikawa and Nogi (2015). Figure 3.4 shows plots of V_{SZ} against V_{S30} for the 565 PBD sites with $z_p \geq 30$ m for values of z_p = 10 and 20 m. The results

generally indicate data trends closer to the Boore (2004) model than the Boore et al. (2011) model. Because the Boore (2004) model is based on California data having modest V_S-depth gradients, whereas the Boore et al. (2011) model is based on data from KiK-Net Japan strong motion stations having steeper gradients, these results imply relatively modest gradients at the PDB sites consistent with other parts of the western U.S.

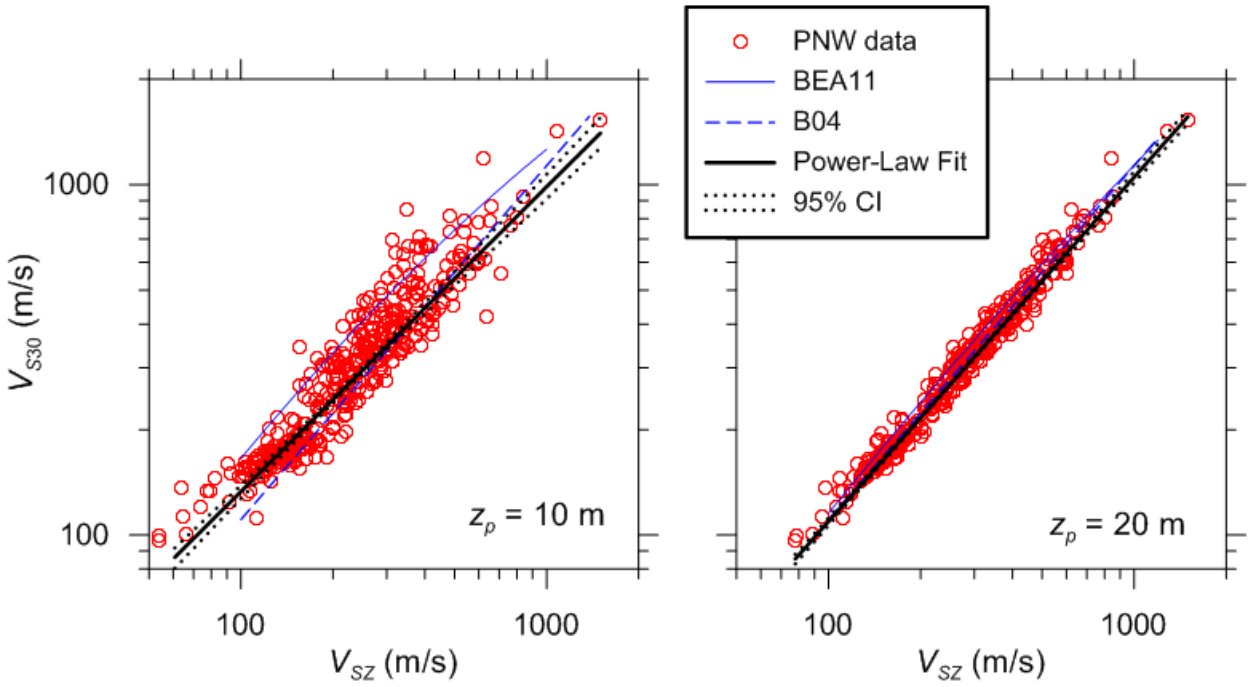

Figure 3.4. Trend of V_{S30} versus V_{SZ} for z_p = 10 and 20 m. V_{S30} extrapolation schemes for Boore (2004) and Boore et al. (2011) are shown for reference.

Rather than adopting an existing model, a process of comparing alternative models was undertaking after customizing their coefficients for the PNW PDB velocities, the results of which are presented in Appendix A. The Dai et al. (2013) model was selected, with coefficients customized for the PNW PDB sites, based on lack of bias and small dispersion. This model is given by:

$$\widehat{V_{S30}} = \frac{30m}{\frac{z_p}{V_{SZ}} + \frac{30 - z_p}{\widehat{V_{SZ,30}}}} \tag{3.1}$$

where $V_{SZ,30}$ is the mean time-averaged velocity for shear wave travel from depth z_p to 30 m when z_p < 30 m, as given by:

$$\widehat{V_{SZ,30}} = \exp\left[d_0(z_p) + d_1(z_p) \ln\left(V_S(z_p)\right)\right] \tag{3.2}$$

where $V_S(z_p)$ is the velocity at depth z_p and $d_0(z_p)$ and $d_1(z_p)$ reflect the outcome of the PNW-specific regressions, and are given by:

$$d_0 = \alpha_0 + \alpha_1 \ln\left(z_p\right)^{\alpha_2} \tag{3.3}$$

$$d_1 = \beta_0 + \beta_1 \ln\left(z_p\right)^{\beta_2} \tag{3.4}$$

The α and β coefficients are given in Table 3.1. A model for the depth-dependent standard deviation of V_{S30} predictions using this approach (σ_e) is given in Appendix A and further elaborated upon in Kwak et al. (2017a).

Table 3.1. Coefficients for PNW-specific V_{SZ} to V_{S30} extrapolation following the procedure of Dai et al. (2013).

i	0	1	2
α_i	33.89	-1.4545	0.7878
β_i	0.23	0.3939	0.5252

3.4 PROXY ATTRIBUTION

Metadata for PNW PDB sites were compiled, as well as for sites targeted for application of proxies (e.g., NGA-Sub SMS sites). Compiled data includes:

- surficial geologic site conditions;
- an indicator regarding site location within or beyond the extent of the Cordilleran ice sheet that was present during the late Pleistocene;
- indicators of site locations within mapped basins (Seattle, Tacoma, Everett, Georgia, Portland, Tualatin, etc.);
- indicators of site location with respect to the volcanic front separating forearc from backarc regions (for ground motion applications)
- topographic gradient (slope) from the Shuttle Radar Topography Mission (SRTM) digital elevation models (DEMs) at 30 arc-sec resolution; and
- geomorphic terrain categories based on procedures in Iwahashi and Pike (2007).

These proxies are listed in Table S2 for PDB sites and Table S3 for SMS sites (both available in the electronic supplement to Ahdi et al. 2017a). Results for Alaska sites are given in Chapter 4 of this dissertation and in Ahdi et al. (2017b).

The PNW study region has three digital compilations of surface geologic maps. Attributes of these compilations are shown in Table 3.2, which include map scale (the largest available scale is preferred); period of original mapping among source documents; and compilation period. The surficial geology maps delineate Quaternary sediments (e.g., alluvium, till, loess) and outcropping rock units. Maps of relatively large scale typically delineate more refined age distinctions, particularly among Quaternary sediments. Information extracted for each location includes the geologic unit name, symbol, age, and description. In addition to the map information, the surface geologic conditions identified by geologist site visits or as attributed to uppermost layers in a V_S profile were recorded where available.

Porter et al. (1983) mapped the extent of the Late Wisconsin Cordilleran ice sheet, which

was present during the Late Pleistocene, in the PNW/Rocky Mountains region. Pierce (2003) provides an updated map which was used in this study. These limits are shown in Figure 3.1 and were used to populate tags regarding prior glaciation for PDB and SMS sites.

Table 3.2. Summary of surface geology map resources utilized for study region.

Region	Scales	Mapping Period	Compilation Period	References
Southern British Columbia	1:50K	1979-1980	1994	Dunn and Rickets, 1994
Washington	1:24K and 1:100K	1959-2012	2005-present	WA-DNR-GER,2010; WA-DNR-GER, 2014
Oregon	1:24K to 1:500K	1937-2013	2004-present	Smith and Roe, 2015

The geographic limits of basins for the study region are defined by both a seismic velocity-depth model and the surficial geologic map. The existing seismic velocity-depth model for the Seattle basin was developed by Delorey and Vidale (2011) and by Stephenson (2007) for the broader Cascadia region, and is described in detail in Section 2.5.2. The update to the latter study in Stephenson et al. (2017) includes velocity variations in the 3D models for the following basin structures: Seattle, Tacoma, Everett, Portland, and Tualatin. The definition of a basin boundary for this study is the margin of Quaternary (Holocene, Pleistocene, or undivided) sediments within the Puget Lowland and greater Portland area basin structures. Sites outside of these boundaries are considered as non-basin and have 0 as their basin flags in Ahdi et al. (2017a) Tables S2 and S3 (available as electronic supplements to that article). Sites within the basin structures considered by Stephenson et al. (2017) have depths to the 2.5 km/s shear wave horizon ($z_{2.5}$) populated in Tables S2 and S3 of Ahdi et al. (2017a) (W. Stephenson 2016, *pers. comm.*).

Figure 3.1 shows an interpretation of the volcanic line separating forearc and backarc regions (forearc to the west; backarc to the east). The volcanic line was drawn “by eye” through volcano locations obtained from the Global Volcanism Program (2013), as described in Section

2.6. This boundary separates regions of relatively fast vs. slow crustal deformation according to GPS velocity vectors plotted by McCaffrey et al. (2007, their Figure 2). Tables S2 and S3 in Ahdi et al. (2017a) include tags indicating whether sites are located in the forearc or backarc regions.

Additional metadata compiled for PDB and SMS sites is based on SRTM30 raster files (Farr and Kobrick, 2000); parameters compiled include geomorphic terrain categories based on procedures in Iwahashi and Pike (2007), topographic gradient, and elevation relative to sea level.

3.5 PROXY CONDITIONAL ON GEOLOGIC CATEGORIES

3.5.1 Grouping of Geologic Categories

The 928 PDB locations were grouped to identify features that produce distinct V_{S30} distributions. As has been common in prior work (e.g., Seyhan et al. 2014), a log-normal V_{S30} distribution was assumed, not because the data passes formal tests of normality, but because this distribution provides a better fit by eye than alternatives such as normal or beta distributions. As such, the data distribution was represented within groups by their mean V_{S30} (taken as the exponent of the natural log mean, and denoted $\mu_{\ln V}$, which has units of m/s) and log standard deviation ($\sigma_{\ln V}$, dimensionless). Within each group, the mean trend of V_{S30} with 30 arcsec topographic gradient was considered.

Using geologic maps (per Table 3.2), the PDB sites fall in 124 geologic units, which are too many for development of a V_{S30} prediction model. A significant majority of the sites are located on Quaternary sediments, and hence the 124 units mostly describe various soil conditions. The grouping process occurred in two phases. The first was judgment-based, in consideration of lithological unit descriptions, depositional environment, and age; this resulted in 42 categories. For data within individual categories and various combinations of categories, moments $\mu_{\ln V}$ and $\sigma_{\ln V}$ are

computed and mean trend with topographic gradient (s) is evaluated as:

$$\overline{\ln\left(V_{S30}\right)} = c_0 + c_1 \ln\left(s\right) \tag{3.5}$$

where V_{S30} is in m/s, slope gradient s is expressed as a decimal (meters per meter), and c_0 and c_1 are regressed coefficients. The gradient effect is considered statistically significant when the null does not fall within the 95% confidence intervals for c_1. This approach of using within-category regressions (Equation 3.5) has been applied previously by Stewart et al. (2014a) and Parker et al. (2017); Wills and Gutierrez (2008) and Wills et al. (2015) apply a different methodology in which gradient is only considered only for the young alluvium category, and mean V_{S30} is taken within three slope bins instead of using a continuous line.

As described further in Section 3.5.2, examinations of V_{S30} distributions within and across categories supported combining some of the 42 categories, such that 18 groups remain (Table 3.3), each having a geologic description and unique attributes ($\mu_{\ln V}$, $\sigma_{\ln V}$, and trend with gradient). As shown in Table 3.3, geologic age was not used as the primary (first-order) group discriminator, which is different from other regional studies utilizing geology for V_{S30} prediction (see Table 2.3). Rather, the primary discriminator considered is the description of lithology and depositional environment, which in many cases can be associated with multiple ages (e.g., alluvium, Group 6, has three age bins—Holocene, Pleistocene, and undivided Quaternary). The lack of further discrimination by age in these cases results to some degree from limited data in certain age bins, but given the data that is available, age was not observed to have predictive power for V_{S30} within these bins (according to data analysis procedures given in the Section 3.5.2).

Table 3.3. Summary of geology-based and hybrid geology-slope-based V_{S30} proxy.

Group No.	Description	Age[1]	N[2]	$\mu_{\ln V}$ (m/s)	$\sigma_{\ln V}$	c_0[3]	c_1[3]	Constituent Categories[4]
1	Peat	H	68	161	0.348	--	--	peat
2	Fraser River: overbank silt/clay	H	74	182	0.259	5.520	0.0506	fr-ov-sc
3	Fraser River: overbank sand/silt, sandy/clayey loam, channel deposits	R,H	122	198	0.263	--	--	fr-ob-ss, fr-loam-sc, fr-sl, fr-chnl
4	artificial fill	H	89	198	0.314	5.625	0.0762	af
5	fluvial + estuarine deposits	Qu	31	239	0.578	--	--	fluv
6	alluvium & valley sediments	R,H,Qu	90	249	0.496	5.976	0.1002	al, val
7	flood deposits: sands, fines, floodplain, undifferentiated	R,H,P, Qu	91	322	0.243	5.904	0.0275	fld-s, fld-f, al-fp, fld
8	lacustrine (incl. glaciolacustrine)	H,P,Qu	10	326	0.135	6.057	0.0657	lac, glac
9	beach, bar, dune deposits	R,H,P, Qu	20	339	0.431	6.326	0.1264	bbd, bch, dune
10	fan deposits	Qu	37	360	0.338	--	--	fan
11	Loess	P,Qu	22	376	0.38	--	--	loess
12	glacigenic sediments (drift & outwash)	H,P	68	399	0.305	--	--	gdc, gdm, go, gos, goa, gor
13	flood deposits: channel, gravel, coarse	P	37	448	0.288	--	--	fld-chnl, fld-gr, fld-co
14	glacial moraines & till	H,Pl,Qu	66	453	0.341	--	--	mor, till
15	undifferentiated sediments & sedimentary rocks	H,P,Qu, T+	42	455	0.363	--	--	c-sed, m-sed, m-Msed, ns-seds, seds
16	terrace deposits & old alluvium	H,P,Qu, T+	21	458	0.507	--	--	terr, oal
17	volcanic rocks & deposits	H,P,T+	14	635	0.663	--	--	Volc
18	crystalline rocks (igneous & metamorphic)	T+	5	750	0.427	--	--	xln

***** = no dependency on topographic gradient

Note 1: R = Recent, H = Holocene; P = Pleistocene, Qu = Quaternary-undivided, T+ = Tertiary & older.

Note 2: N = number of profiles. All data from PNW except groups 8, 10, & 11, which include data from Alaska.

Note 3: Dashes "--" indicate groups with no dependency on topographic gradient.

Note 4: Categories are given in Tables S2 and S3 (in the electronic appendix to Ahdi et al. 2017a) for each site. af = artificial fill. fr- = Fraser River Delta deposits; -ov-sc = silty/clayey overbank deposits; fr-ob-ss = sandy/silty overbank deposits; -loam-sc = silty-clay loam; -sl = silty loam; -chnl = channel facies. fluv = fluvial & estuarine deposits; al = alluvium; val = valley sediment. fld = flood deposits; -s = sand facies; -f = fine-grained facies; -gr = gravelly facies; -co = coarse-grained facies. al-fp = alluvial floodplain deposits. lac = lacustrine deposits; glac

= glaciolacustrine deposits; bbd = beach, bar, & dune deposits; bch = beach deposits. gdc = continental glacial drift; gdm = glaciomarine drift; go = glacial outwash (undifferentiated); gos = sand facies; goa = advanced glacial outwash; gor = recessional glacial outwash. c-sed = continental sedimentary rocks; m-sed = marine sedimentary rocks; m-Msed = marine metasedimentary rocks; ns-seds = near-shore sediments and sedimentary rocks; seds = undivided sediments and sedimentary rocks. mor = moraines. terr = terrace deposits; oal = old alluvium. volc = volcanic rocks and deposits. xln = crystalline rock.

The effect of site location within or beyond the extent of the Cordilleran ice sheet was investigated separately from lithology. While some of the group descriptions in Table 3.3 are clearly associated with glaciation (e.g., Groups 8, 12, 14) or are specific to regions known to have been within the glacial extent (e.g., Fraser River, Groups 2-3), other groups (e.g., alluvium) could occur in both glaciated and non-glaciated regions. By separating sites within these groups on the basis of glacial history, the potential effects on V_{S30} of relatively recent sediments deposited on top of stiffer glacigenic deposits were investigated. Such conditions have been found to produce distinct trends in the mean and standard deviation of V_{S30} in glaciated vs. non-glaciated regions of central and eastern North America (CENA; Parker et al. 2017).

Because the Fraser River Delta region in southern Vancouver has a high data concentration, statistics for this region were also investigated separately from those of otherwise similar geology to identify potentially distinct regional features.

3.5.2 Statistic Testing in Grouping Process

A systematic process for deciding when groups of V_{S30} values are statistically distinct was applied; this same process was presented previously by Parker et al. (2017) and is summarized briefly here. Two types of *F*-tests were used (Snedecor and Cochran, 1989), which compare the statistical performance of submodels with that of a full model for a common data set. For example, a full model may comprise sediments of alluvial origin (Group 6), whereas submodels could represent age bins. The one-way analysis of variance (ANOVA) *F* statistic is computed as (adapted from Snedecor and Cochran, 1989):

$$F_1 = \frac{\left(RSS_f - (RSS_1 + RSS_2)\right) \Big/ \left((df_1 + df_2) - df_f\right)}{\hat{\sigma}^2} \tag{3.6}$$

where *RSS* refers to residual sum of squares (based on misfits from median model predictions) for submodels (subscripts *1* and *2*) and the full model (*f*), *df* refers to model degrees of freedom (one if the model consists of a simple mean, two if the model includes a slope gradient term), and

$$\hat{\sigma}^2 = \frac{RSS_1 + RSS_2}{N_f - \left(df_1 + df_2\right)} \tag{3.7}$$

where N_f is the number of data points in the full model. To better distinguish data groups having similar means but differing dispersion, a second *F* statistic was computed (Snedecor and Cochran, 1989):

$$F_2 = \frac{\sigma_1^2}{\sigma_2^2} \tag{3.8}$$

where σ_1 and σ_2 are the dispersions of residuals in the subgroups. The F_2 statistic is used to test the null hypothesis that two normal populations from which samples are drawn have the same variance. By comparing an *F* statistic to the *F* distribution, a significance level (*p*) is computed, which is used to judge potentially indistinct submodels. If F_1 has a *p* value ≤ 0.05, or F_2 has a *p* value ≤ 0.05 or ≥ 0.95, the subgroups are considered distinct. To meet the requirement of normally populated data populations, both *F*-tests were performed on residuals in natural logarithmic units.

As described above, data analysis using *F*-tests was part of the culling process to reduce 42 geologic categories to the 18 groups shown in Table 3.3. This process involved computing $\mu_{\ln V}$, $\sigma_{\ln V}$, and gradient trends for the 42 categories. Based on visual inspection of the data, including consideration of low data populations in some categories, various combinations of categories were identified for *F*-testing to judge distinctness. Figure 3.5 shows an example of distinct categories retained as different groups—flood deposits segregated into groups of sands, fines, and floodplain

sediments, and undifferentiated flood deposits (Group 7), and channel, gravel, and coarse deposits (Group 13). Results of *F*-tests for these categories are shown in Table 3.4, which shows distinction per the F_1 statistic. As shown in Figure 3.5, the distinction in this case is caused by different group means. Several other factors can produce a conclusion of distinct groups, including different levels of gradient-dependence, and different data dispersions. Figure 3.6 shows an example of not-distinct categories—Fraser River overbank sand and silt and Fraser River sandy/clayey loam. These two categories, along with two other Fraser River Delta-specific sedimentary categories, were combined into a single group (3) based on *F*-test results showing that the distributions are not distinct (Table 3.4) and visual inspection of the data (Figure 3.6 and corollary plots for other categories).

Figure 3.7 shows histograms and V_{S30}-gradient plots for each of the 18 groups in Table 3.3. Table 3.4 lists *F*-test results demonstrating between-group distinction for combinations of groups for which the geologic descriptions were sufficiently similar that testing was considered to be warranted. Some relevant findings from this process include:

- Two Fraser River groups are provided, which have different material textures for a common depositional environment (fine vs. coarse; Groups 2 vs. 3). As expected, seismic velocities are slower for fine-grained overbank deposits than for coarse-grained materials.
- Four groups are provided for non-glacigenic moving-water-deposited sediments not located within the Fraser River delta: Group 5 (fluvial & estuarine deposits), Group 6 (alluvium and valley sediments), Group 10 (alluvial fan deposits), and Group 16 (terrace deposits and old alluvium). The different depositional processes and age effects in these cases lead to distinct velocities, with $\mu_{\ln V}$ increasing in the order of 5-6-10-16 and gradient-dependence for Group 10.

- **Table 3.4.** Summary of selected *F* statistic test results used in geology proxy group development.

Tested Groups or Categories*	F_1	p_1	F_2	p_2	Distinct Means?	Distinct Sigmas?	Distinct (0) or Non-Distinct (1)
2 vs 3	5.008	3.24E-15	0.969	0.554	0	1	0
5 vs 6	0.134	1	1.353	0.140	1	1	1
6 vs 9	6.630	1.35E-05	1.328	0.246	0	1	0
6 vs 10	17.230	1.86E-15	2.163	0.005	0	0	0
6 vs 15	49.797	3.28E-26	1.869	0.014	0	0	0
7 vs 9	0.503	0.983	0.318	1	1	1	1
7 vs 10	4.321	2.41E-06	0.518	0.993	0	1	0
7 vs 13	43.214	2.65E-22	0.713	0.899	0	1	0
7 vs 16	22.131	1.51E-10	0.230	1	0	1	0
8 vs 7	0.016	1	0.307	0.971	1	1	1
8 vs 9	0.080	1	0.098	0.999	1	1	1
9 vs 11	0.679	0.800	1.285	0.287	1	1	1
10 vs 14	10.759	2.60E-16	0.979	0.517	0	1	0
10 vs 16	4.686	2.73E-04	0.444	0.983	0	1	0
12 vs 13	3.638	3.17E-05	1.121	0.360	0	1	0
13 vs 14	0.027	1	0.711	0.865	1	1	1
13 vs 16	0.045	1	0.323	0.998	1	1	1
14 vs 16	0.013	1	0.453	0.991	1	1	1
(col+ls) vs 16	0.248	0.861	0.467	0.709	1	1	1
seds vs 6	17.138	4.72E-16	0.566	0.821	0	1	0
seds vs. sed. rocks	0.435	0.906	1.057	0.420	1	1	1
fr-ob-ss + fr-loam-sc	0.140	1	1.457	0.106	1	1	1

***Note**: Category abbreviations are same as those in Note 3 of Table 3.3, above.

- Several categories involve unique material types or depositional processes, including Group 1 (peat), Group 4 (fill), Group 5 (fluvial and estuarine), Group 8 (lacustrine and glaciolacustrine), Group 10 (fans), and Group 11 (loess).
- Two groups are provided for sediments derived directly from glacial processes: glacial moraines and till (Group 14) and glacigenic sediments (drift and outwash; catch-all group for glacially-derived, non-lacustrine sediments; Group 12).
- Groups 8, 10, and 11 were sparsely populated with PNW profiles (5, 8, and 5, respectively).

Visual inspection of these PNW within-group data distributions and similar distributions for data from Alaska revealed similar features. As a result, the PNW and Alaska data for these three groups were combined. Further discussion of the Alaska data can be found in Chapter 4 of this dissertation (Ahdi et al. 2017b).

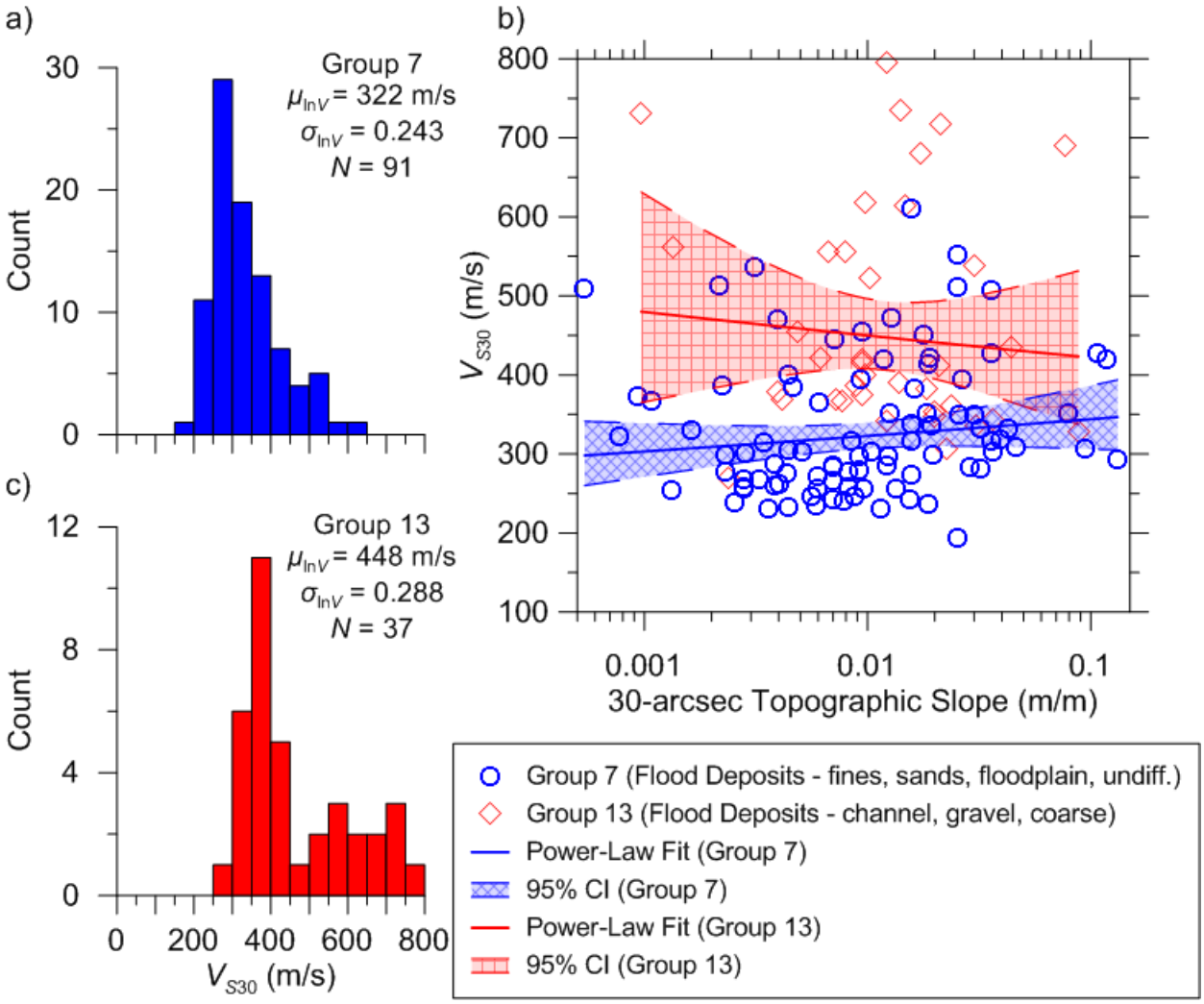

Figure 3.5. Comparison of two similar geologic groups (Groups 7 and 13) with respect to lithology/depositional environment that were ultimately deemed to be statistically distinct; a) and b) V_{S30} distributions for the two groups; c) V_{S30}-slope plots.

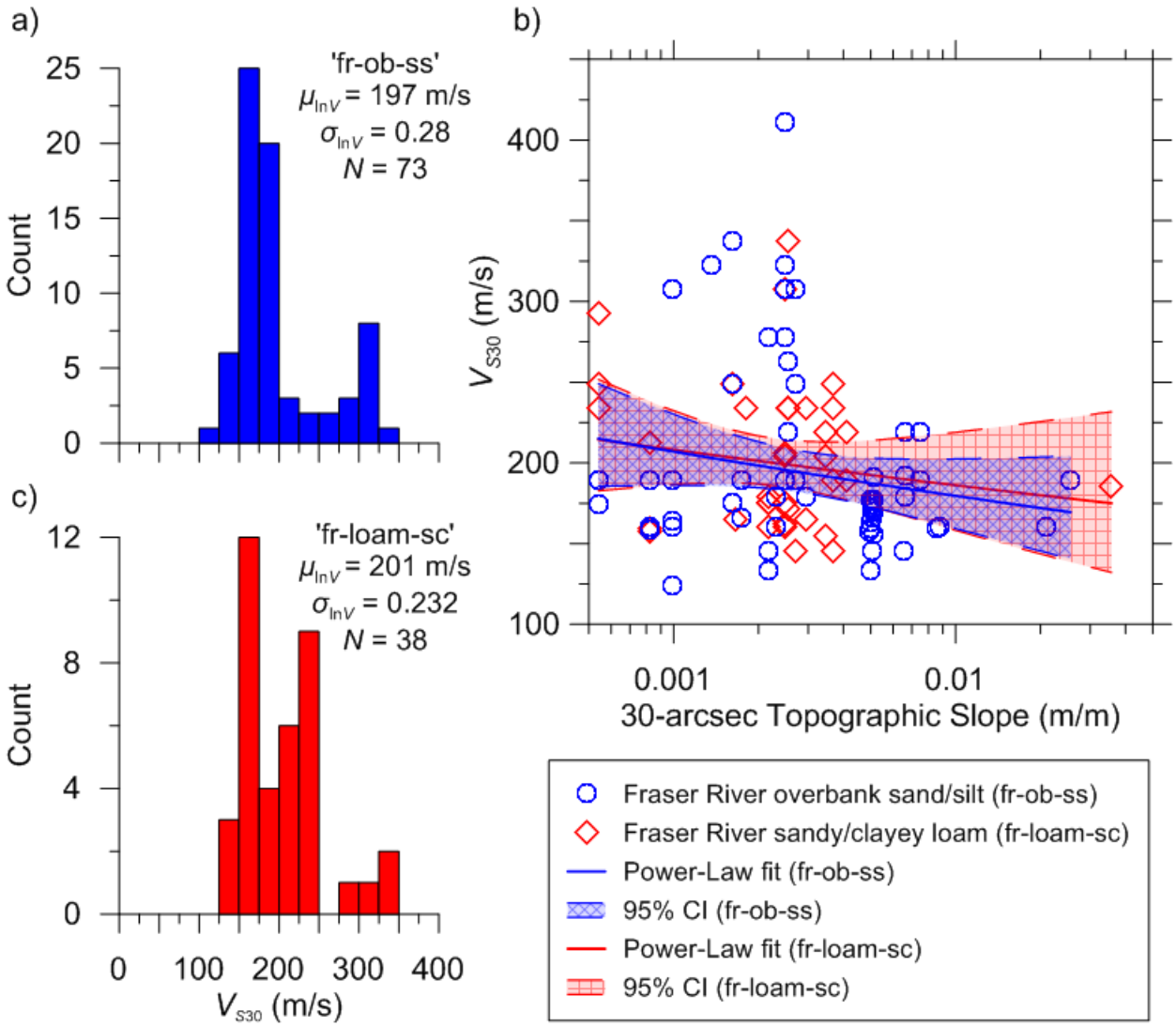

Figure 3.6. Comparison of two statistically-similar geologic categories (Fraser River overbank sand and silt, and Fraser River sandy/clayey loam) that were ultimately combined within a group (Group 3) due to lack of V_{S30} distinction. a) and b) V_{S30} distributions for the two categories; c) V_{S30}-slope plots.

- Three categories descriptive of alternate bedrock configurations are provided: sedimentary (sometimes grouped with undifferentiated sediments on geologic maps) (Group 15), volcanic (Group 17), and crystalline igneous and metamorphic (Group 18). Per Table 3.3, statistical testing does not support combining undifferentiated with other unconsolidated sediments, such as those of Group 6 (alluvial and valley sediments). Rather,

undifferentiated sediments group with sedimentary rocks. This is perhaps due to coarse geologic mapping, which can result in undivided sedimentary units being lumped together with various sedimentary rocks in some cases.

- The data include four profiles of colluvium and landslide deposits. The data from these sites are similar to those for Group 16 (see Table 3.4), which includes terrace deposits and older alluvium. However, these data were not combined, which would have required broadening the description of Group 16, due to the different depositional environments. As a result, these four profiles are not present in any of the 18 groups in Table 3.3. Wills et al. (2015) faced a similar problem with the California data; they chose to group colluvium and landslide deposits with the adjacent bedrock units.
- From a purely statistical standpoint, the following groups could have been combined Groups 5 and 6; Groups 7, 8, and 9; Groups 9 and 11; and Groups 13, 14, and 16. In each case, the *F* tests do not support distinction, but the groups were kept separate so that the geologic classifications retain physical meaning. I admit that this represents judgement and is subjective.

A general observation from the PNW data is that most category standard deviations fall within the range $\sigma_{\ln V} \approx 0.25$–$0.45$, which is typical of prior results for sediment categories in active regions such as California, Japan, and Greece (references in Table 2.3). Initially, higher standard deviations were expected as a result of the glacial history in portions of the PNW, which has previously been shown to produce high standard deviations in CENA (Parker et al., 2017). It was fortunate that this finding was not repeated here, as it indicates greater predictive power for the proposed model than in some CENA categories. An exception is Group 17 (volcanic rocks and deposits), which has a high standard deviation ($\sigma_{\ln V} = 0.66$). Figure 3.8 shows profiles for this

category, many of which have relatively low gradients typical of soils, while others have steep gradients typical of weathered rock or surficial sediments overlying more intact rock such as basalt. The grouping of these disparate conditions is a result of their being combined on many geologic maps in the region.

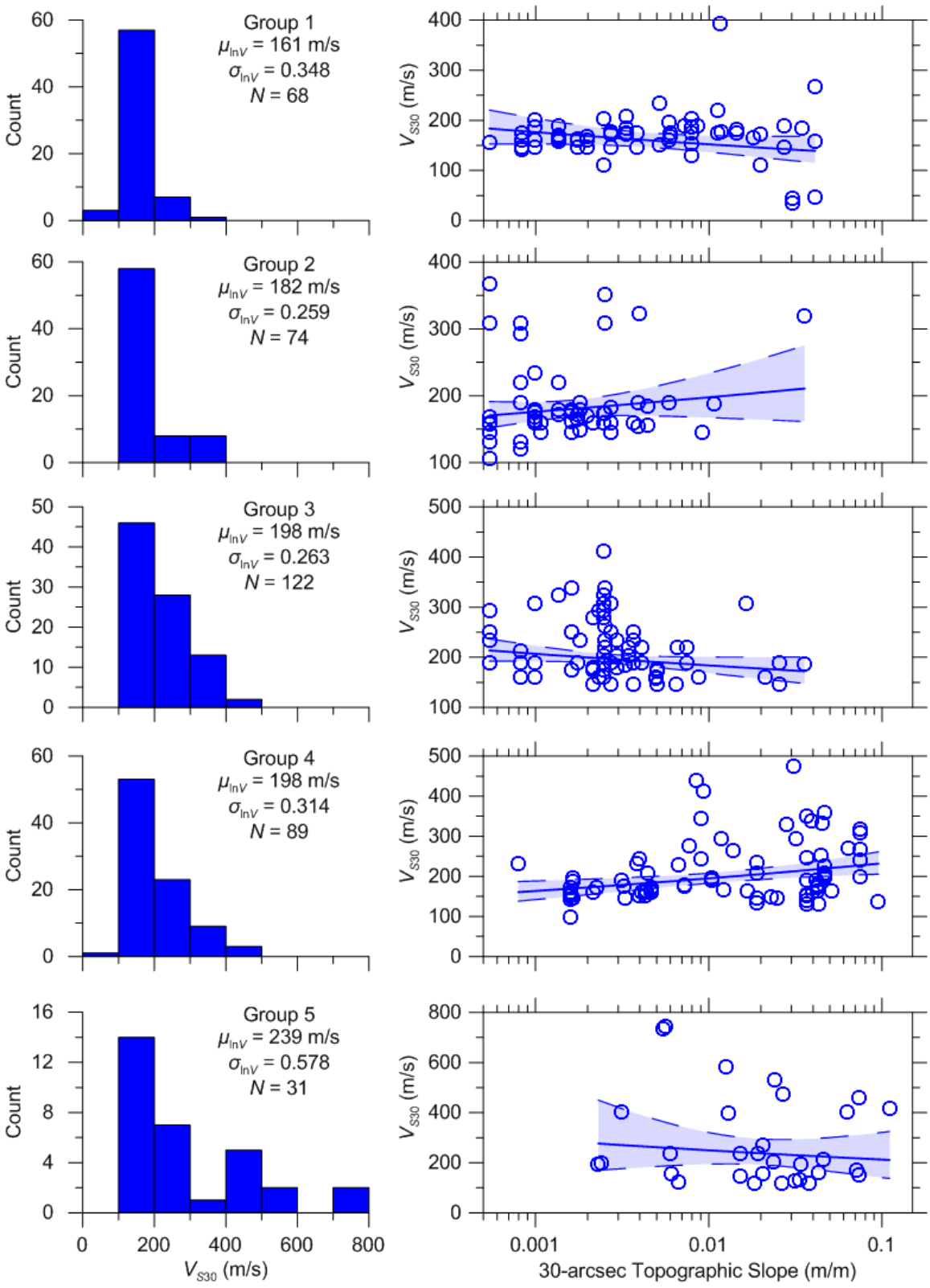

Figure 3.7. Geology proxy-based group V_{S30} distributions (left column) and topographic gradient dependency (right column) for the 18 recommended geology groups, as defined in Table 3.3. Plots in right column show power-law fits and 95% CI of data.

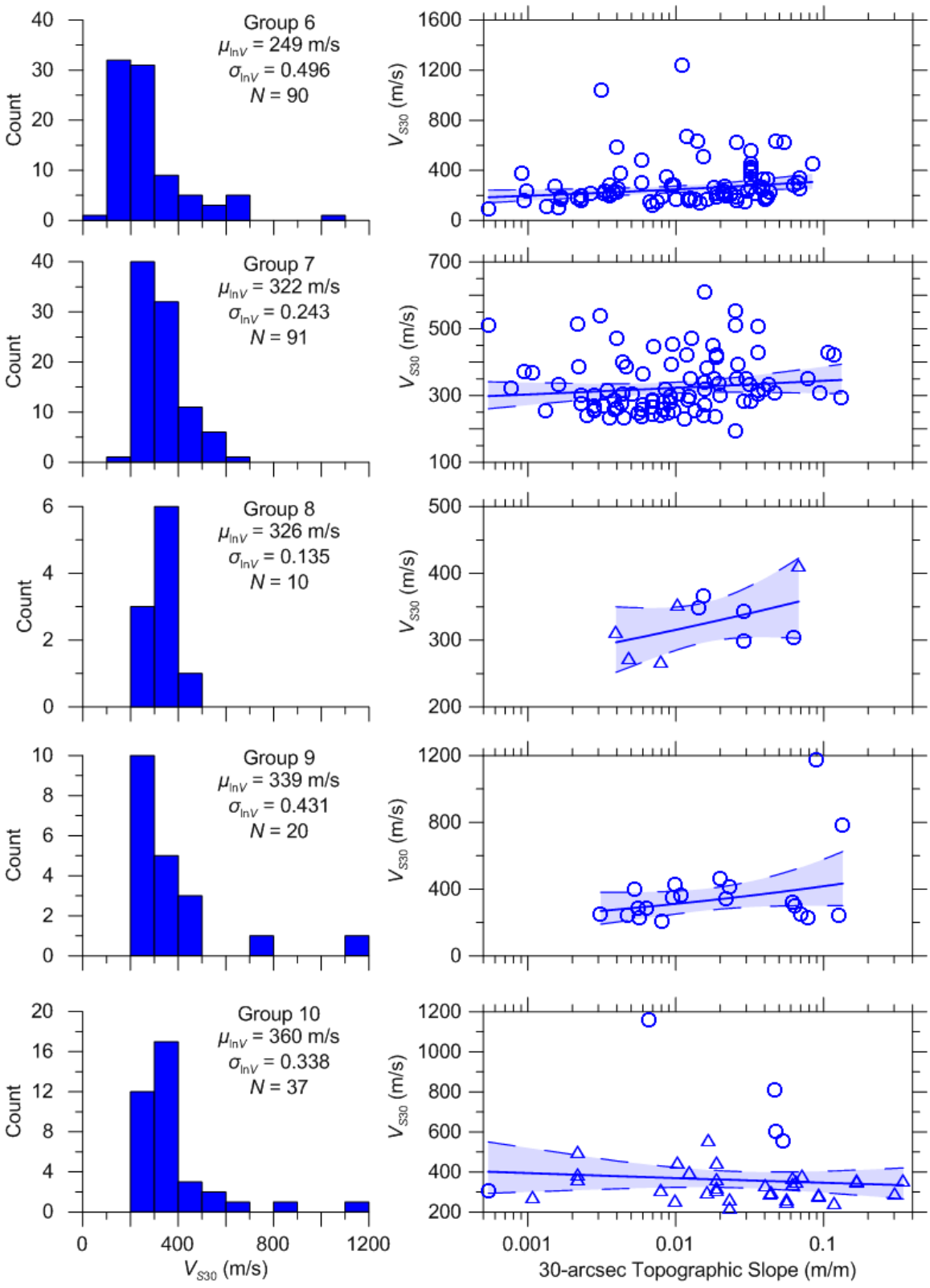

Figure 3.7 continued.

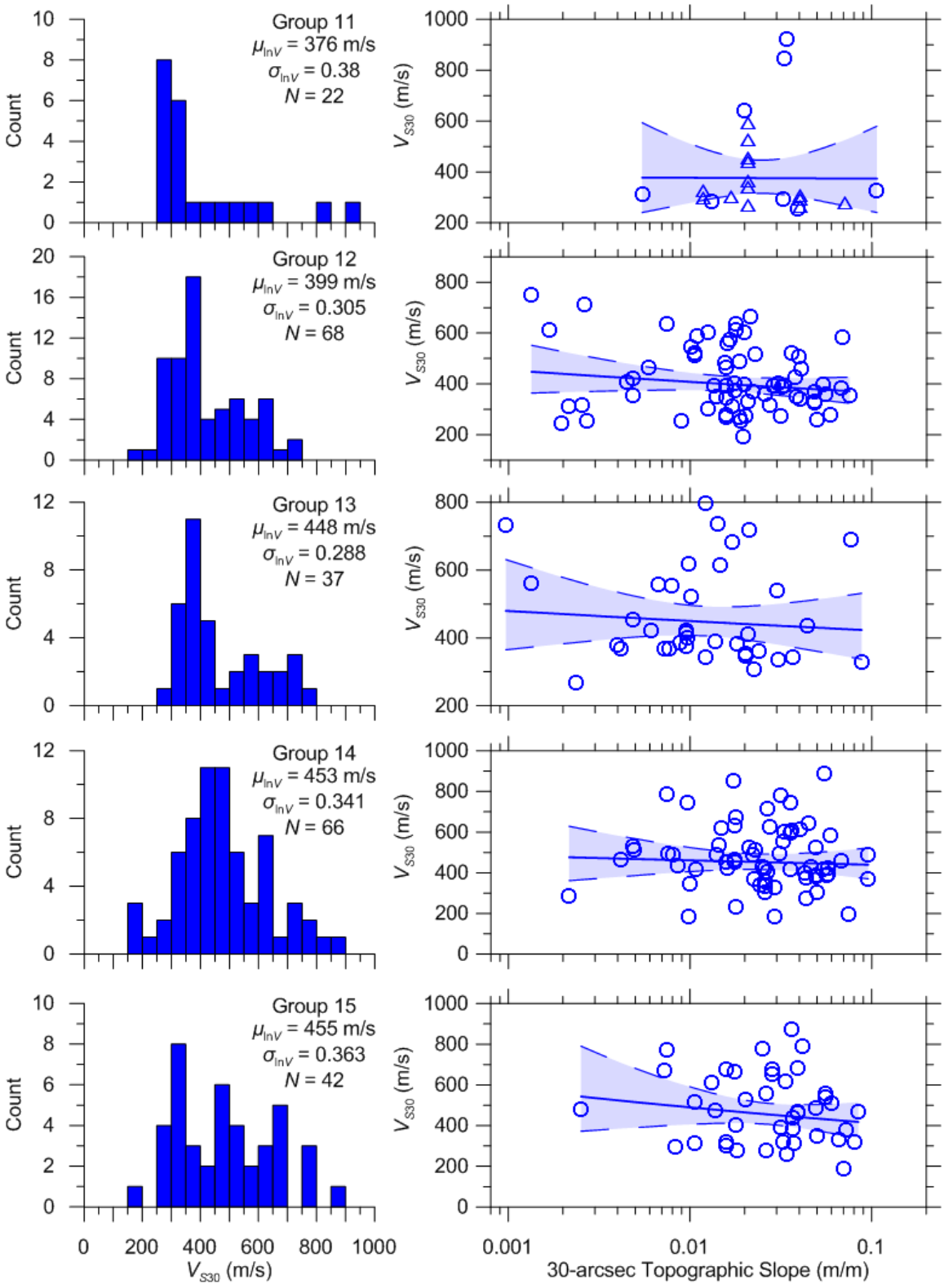

Figure 3.7 continued.

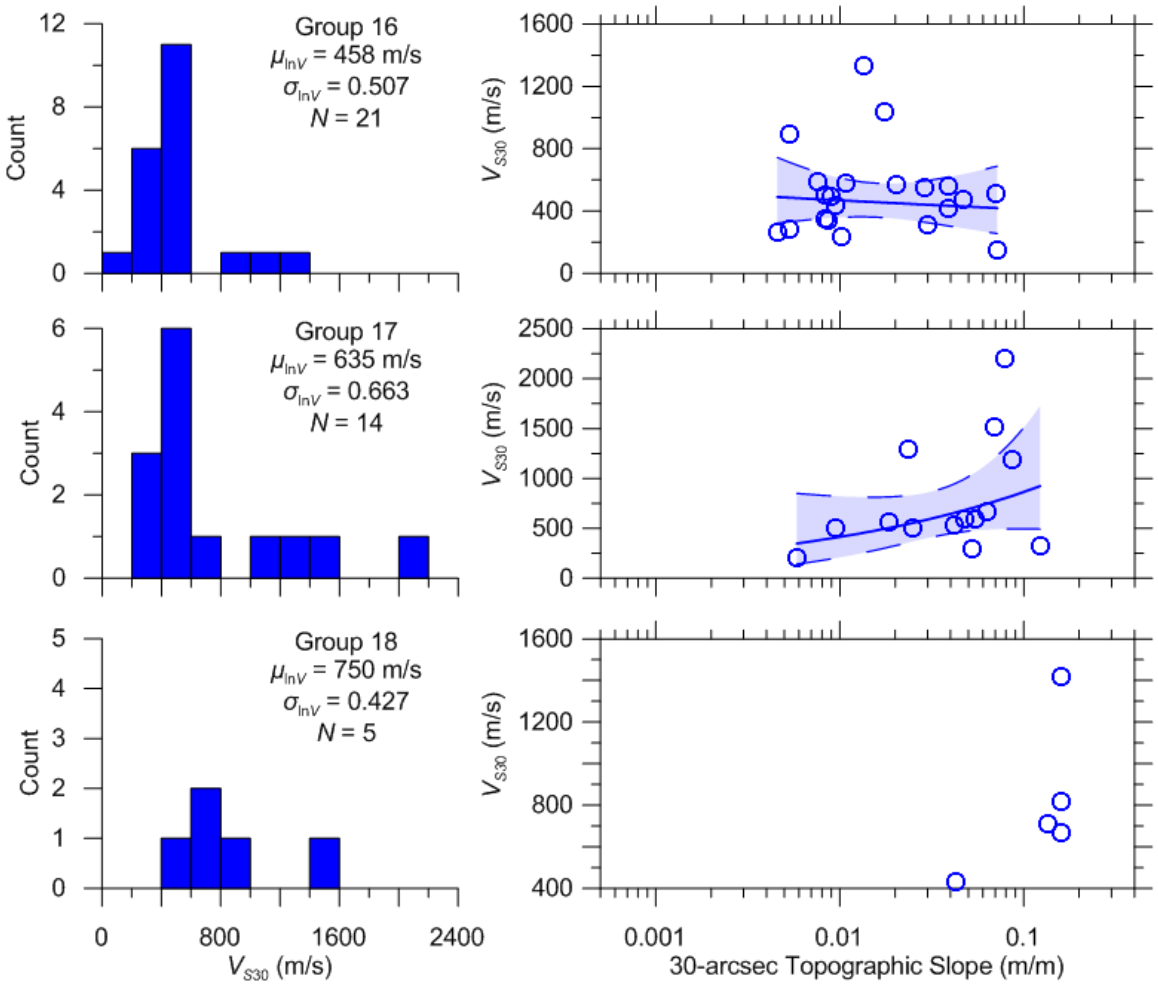

Figure 3.7 continued.

3.5.3 Proposed Model

The proposed model provides estimates of $\mu_{\ln V}$ and $\sigma_{\ln V}$ conditional on group number and, in some cases, gradient s. For groups without provided values of c_0 and c_1 (1, 3, 5, 10–18), the mean should be taken as the gradient-independent $\mu_{\ln V}$ value from Table 3.3. For other groups, the mean is taken from Equation (3.5) using the coefficients in Table 3.3. For both model types, $\sigma_{\ln V}$ is provided in Table 3.3.

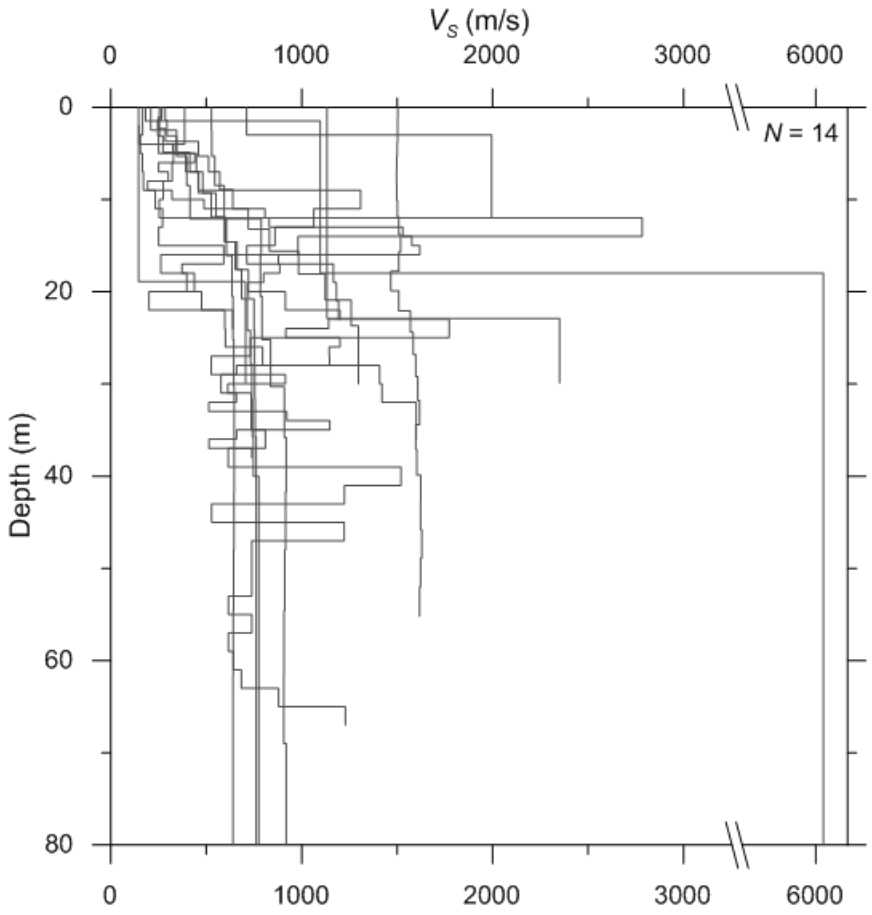

Figure 3.8. V_S profiles comprising Group 17 (volcanic rocks and deposits).

3.6 PROXY CONDITIONAL ON GEOMORPHIC TERRAIN CATEGORIES

As noted in Section 2.3.1, Yong et al. (2012) introduced the use of terrain classes as a proxy for V_{S30} estimation, using data from California. Yong et al. (2012) do not compute simple category moments, instead adopting a sampling approach similar to bootstrapping for cross-validation purposes. The California model was subsequently updated by Yong (2016).

The statistical moments $\mu_{\ln V}$ and $\sigma_{\ln V}$ were computed using full within-group populations, which is different from the prior work of Yong (2016) and others. Of the 928 PDB sites, 876 have assigned terrain classes. Results are presented in Table 3.5 and Figure 3.9 for each of the 16 geomorphic terrain classes of Iwahashi & Pike (2007). Note that terrain classes 2, 6, and 10 were

poorly populated (< 3 V_{S30} values per category), and as such were not assigned V_{S30} moments in this model. For categories for which the PNW $\mu_{\ln V}$ is derived from poorly-populated data sets (2, 6, 10), results by Yong (2016) from California were used here. For these categories, the associated $\sigma_{\ln V}$ was increased due to the epistemic uncertainty of applying a model in a target region outside of that for which it was developed (additional details provided in Section 2.3.2).

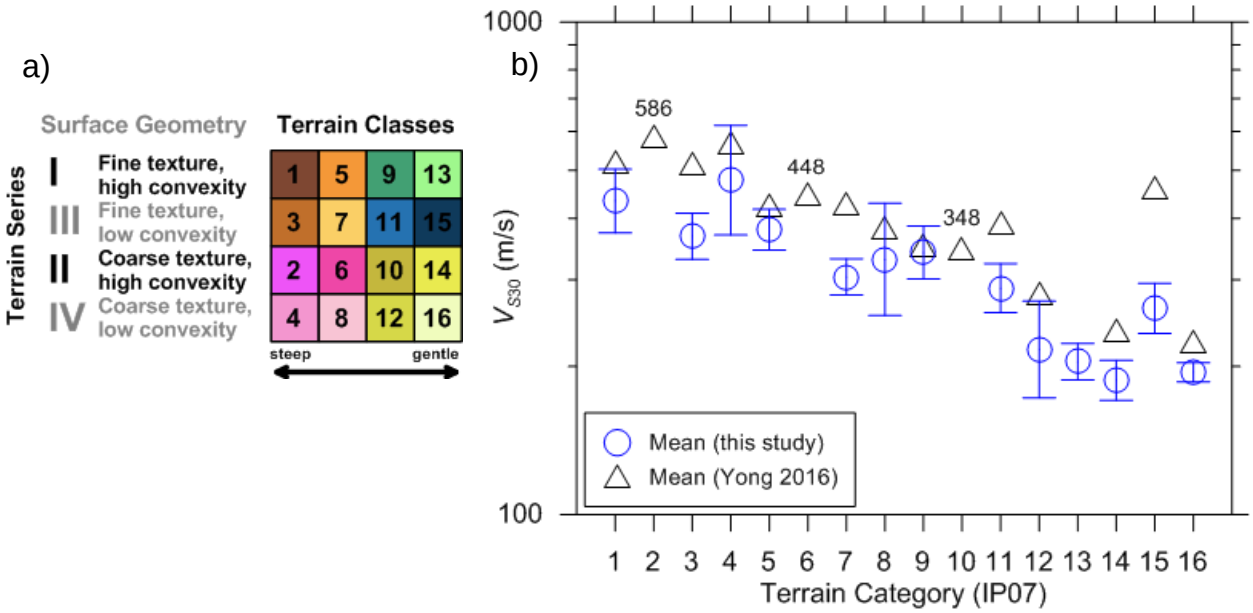

Figure 3.9. (a) Depiction of Iwahashi and Pike (2007) terrain classes; (b) category mean V_{S30} values as given by Yong (2016) and from the PNW PDB (with confidence intervals).

Table 3.5. Summary of terrain class-based V_{S30} proxy for the PNW. Moments adopted from California are shown in parentheses.

IP07 Terrain Class	N	$\mu_{\ln V}$ (m/s)	$\sigma_{\ln V}$	$\mu_{\ln V}$ of Y16 (m/s)	$\sigma_{\ln V}$ of Y16
1	30	433	0.417	519	0.38
2 (CA)	0	(586)	(0.16)	586	0.16
3	59	368	0.424	517	0.38
4	13	478	0.471	568	0.46
5	76	379	0.421	425	0.37
6 (CA)	3	(448)	(0.14)	448	0.14
7	177	304	0.574	429	0.38
8	26	330	0.684	382	0.32
9	50	341	0.445	353	0.16
10 (CA)	2	(348)	(0.09)	348	0.09
11	73	288	0.501	392	0.48
12	23	216	0.553	281	0.20
13	62	204	0.343	NA	NA
14	29	187	0.256	236	0.14
15	70	262	0.502	460	0.52
16	171	194	0.297	225	0.20

3.7 PROXY PERFORMANCE

3.7.1 Residuals Analysis

Proxy-based estimates of V_{S30} were assigned to the 928 profiles in the database using the protocols given in Table 3.3 (geology-slope proxy) and Table 3.5 (terrain proxy). Residuals in natural log units were calculated as:

$$R_i = \ln(V_{S30})_i - \overline{\ln(V_{S30})_i} \tag{3.9}$$

where $\ln(V_{S30})_i$ is the natural log of the V_{S30} calculated from V_S profile i, and $\overline{\ln(V_{S30})_i}$ is the proxy-based natural log mean for profile i. Means of residuals ($\overline{R}$) are near zero because the performance is evaluated using the same data set used in model development; accordingly, the log standard deviations of the residuals ($\sigma_{\ln V}$) is of primary interest here.

Figure 3.10 shows histograms of residuals computed from both proxies using all profiles,

previously glaciated profiles, and non-glaciated profiles. The metrics for overall proxy performance are mean residual $\overline{R} = 3.8 \times 10^{-3}$ and $\sigma_{\ln V} = 0.351$ (geology-slope proxy, Fig 10a) and $\overline{R} = -7.4 \times 10^{-2}$ and $\sigma_{\ln V} = 0.465$ (terrain proxy, Fig 3.10b). The difference in $\sigma_{\ln V}$ between these proxies is larger than has been observed for active crustal regions (Seyhan et al. 2014). In a preliminary version of this model in which V_{S30} values for thick first-layer sites were not adjusted as described previously, a non-zero mean residual was observed for that subset of sites. This motivated the aforementioned adjustments of the data for such sites as described in Section 3.3.2.

The glaciated and non-glaciated sub-groups have similar distributions, indicating that site location relative to the glacial extent (Figure 3.1) does not carry predictive power. In the case of the geology-slope proxy, the lack of effect on the mean is due to glacial effects having been accounted for by the geologic classification scheme, as two categories are associated with glacigenic deposits (Groups 12, 14) and two are in a location that was glaciated (Groups 2, 3). The lack of a glacial effect on $\sigma_{\ln V}$ was unexpected and contrary to experience in CENA (Parker et al. 2017). As shown in Figure 3.10, the mean bias for glacial and non-glacial sites is larger for the terrain proxy than for the hybrid slope/geology proxy. This likely occurs because terrain categories do not directly distinguish glaciated and unglaciated regions, and hence the effect of glaciation on V_{S30} was not considered.

Figure 3.11a shows the distribution of geology-slope proxy residuals with basin depth $z_{2.5}$ (described in Section 2.5.2). Residuals for sites mapped outside of basins are shown at depth zero. The results indicate no trend with depth, suggesting that the aforementioned geologic categories sufficiently account for basin effects on V_{S30}. Residuals of the terrain proxy also lack a depth trend (Figure 3.11b).

Two geologic groups, 6 and 15, have significant fractions of both glaciated and non-

glaciated sites. For Groups 6 and 15, residuals indicate a bias (in natural log units) of –0.12 and –0.056 for glaciated sites (proxy over-predicts), and 0.12 and 0.040 for non-glaciated sites (proxy under-predicts), respectively. Partitioning these groups in consideration of glaciation was considered, but this modification was not made because of the relatively small impact on $\mu_{\ln V}$ and the sparseness of the data.

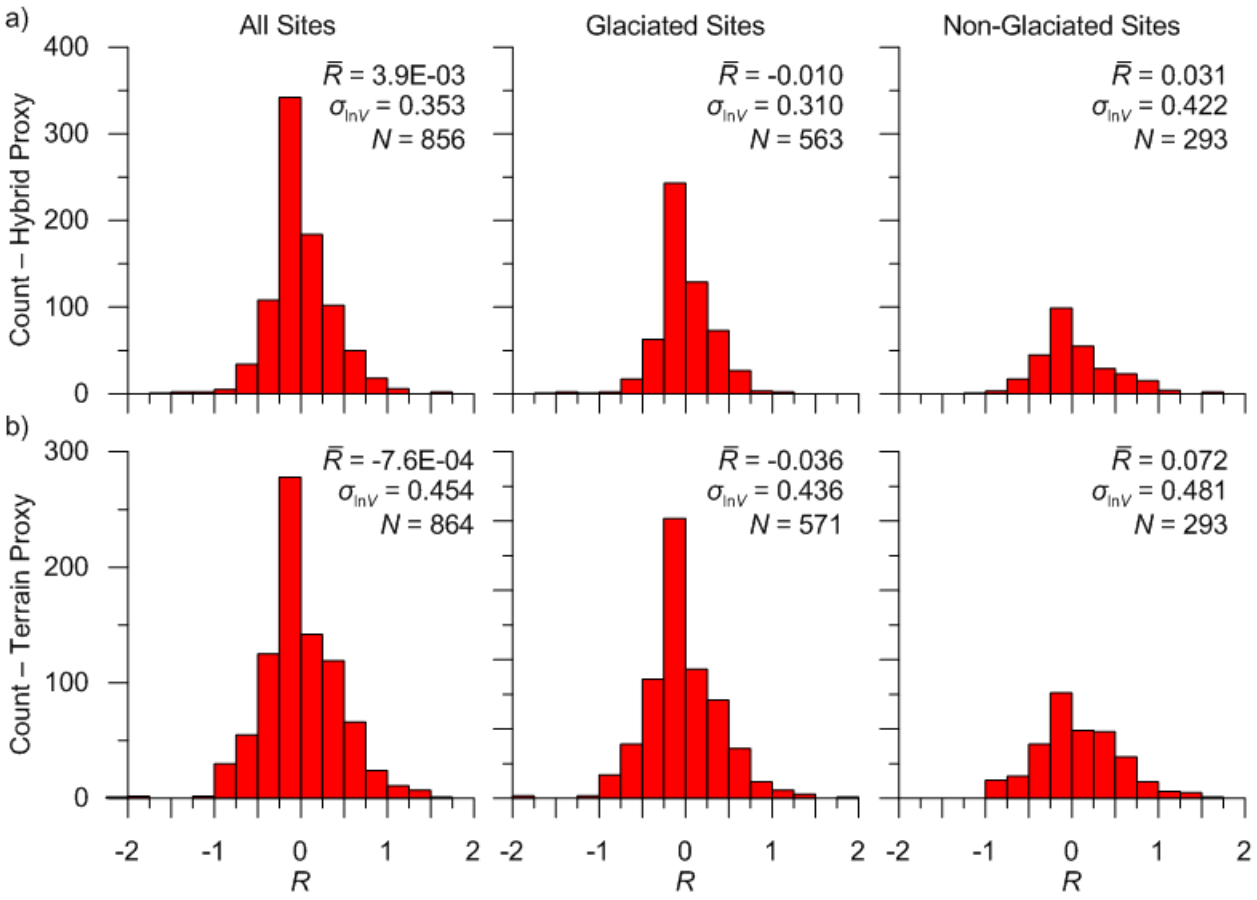

Figure 3.10. Analysis of log-V_{S30} residuals for (a) Geology/topographic gradient and (b) terrain proxies for all (left), glaciated (middle), and non-glaciated (right) sites.

The correlation of the two proxies (hybrid slope-geology and terrain) was investigated through analysis of residuals for the 855 sites that have predictions by both methods. Figure 3.12 shows the relationship between the two sets of residuals, which are indeed well-correlated with ρ = 0.78 (ρ = 1.0 would indicate perfect correlation). The strong correlation of the PNW data is

consistent with the conclusion of Yong (2016) that there are only marginal gains to using multiple proxies, such as using both the hybrid geology-slope and geomorphic terrain proxies.

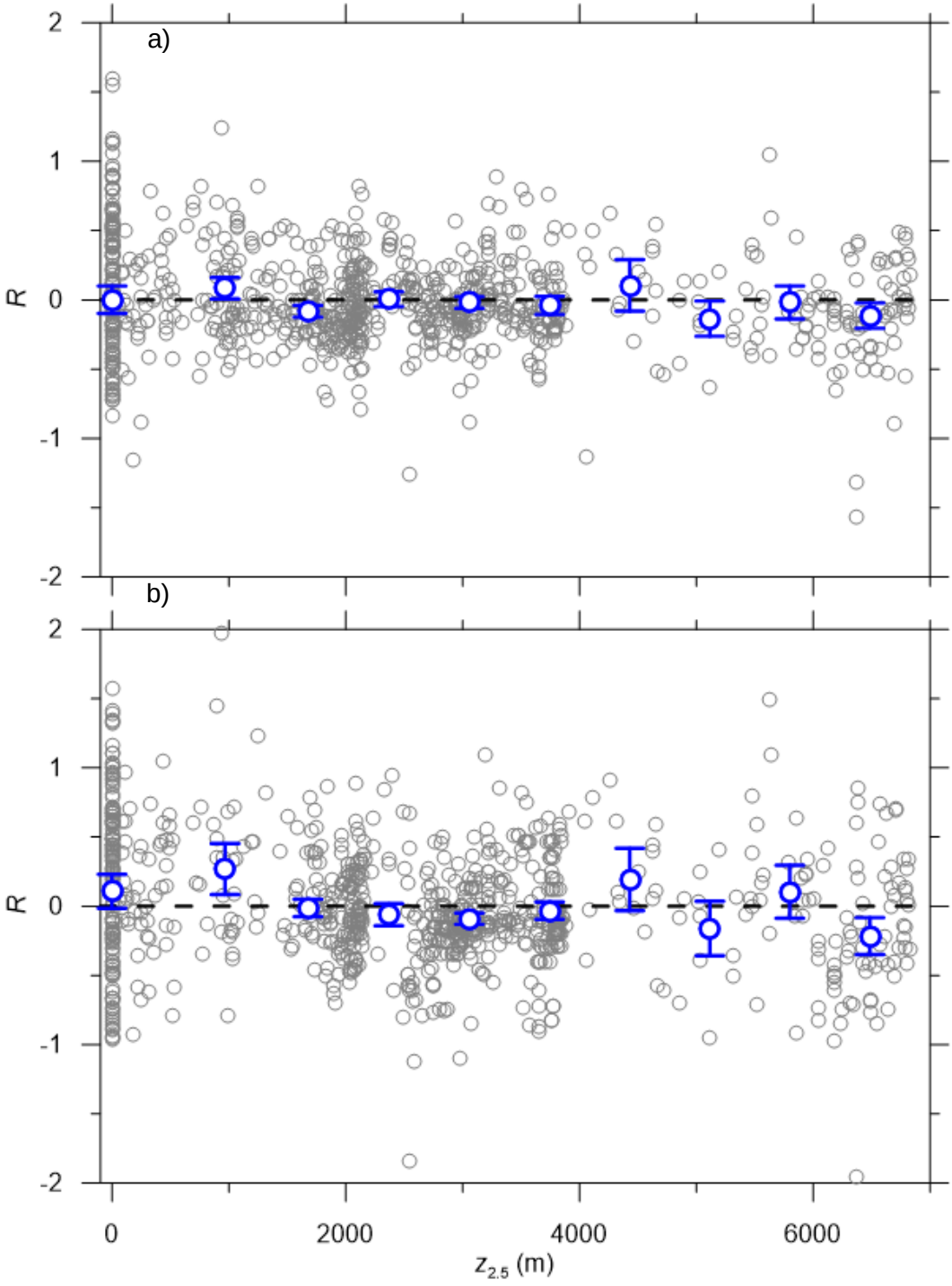

Figure 3.11. Trend of log-V_{S30} residuals with basin depth $Z_{2.5}$ for the (a) geology proxy and (b) terrain proxy. Sites located outside of basins shown at depth 0.

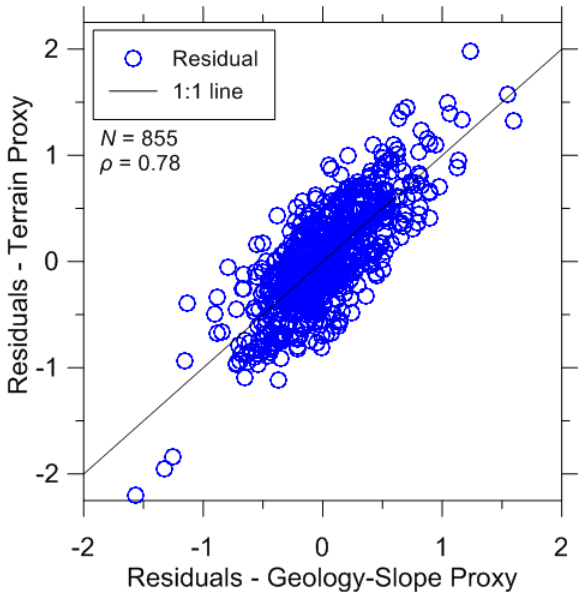

Figure 3.12. Correlation of hybrid geology-slope and terrain proxy V_{S30} residuals.

3.7.2 Comparison to Proxy Estimates for Similar Geologic Categories in Other Regions

This section presents comparisons of V_{S30} predictions for similar geologic groups among three general regions of North America: California, which is an active crustal region (Wills et al. 2015); the present study for the PNW (subduction region); and CENA, which is a stable continental region (Parker et al. 2017). While each region has some unique groups, others are sufficiently general that they are present in two or three of the studies. Comparisons in this section are made for these common groups, given in Table 3.6.

Figure 3.13 shows group means for the three regions (two in some cases) for the six general geologic groups in Table 3.6. When a group has a gradient-dependent model, the variation of the mean velocity over the topographic slope range from the source data is shown using symbols extended over the applicable range on the ordinate axis. The mean values for the PNW categories are shown with 95% confidence intervals, so that the statistical significance of differences between other regions and the PNW data can be judged (where the mean is gradient-dependent, Groups 4,

Table 3.6. Common geologic groups in geology-based proxies for regions of North America.

Geology	California Categories (Wills et al. 2015)	PNW Group No. (this study)	CENA Group No. (Parker et al. 2017)
Artificial fill	af/qi	4*	NA
Alluvium & Alluvial Fans	Qal 1-3*	6*, 10*	1
Pleistocene alluvium (without glaciation)	Qoa	16	5*
Tertiary rock (generally sedimentary)	Tsh, Tss	15	12*
Volcanics	Tv	17	NA
Older, crystalline rock	Xtaline	18	13, 15, 17

* Gradient-dependent model (Equation 3.5 or similar)

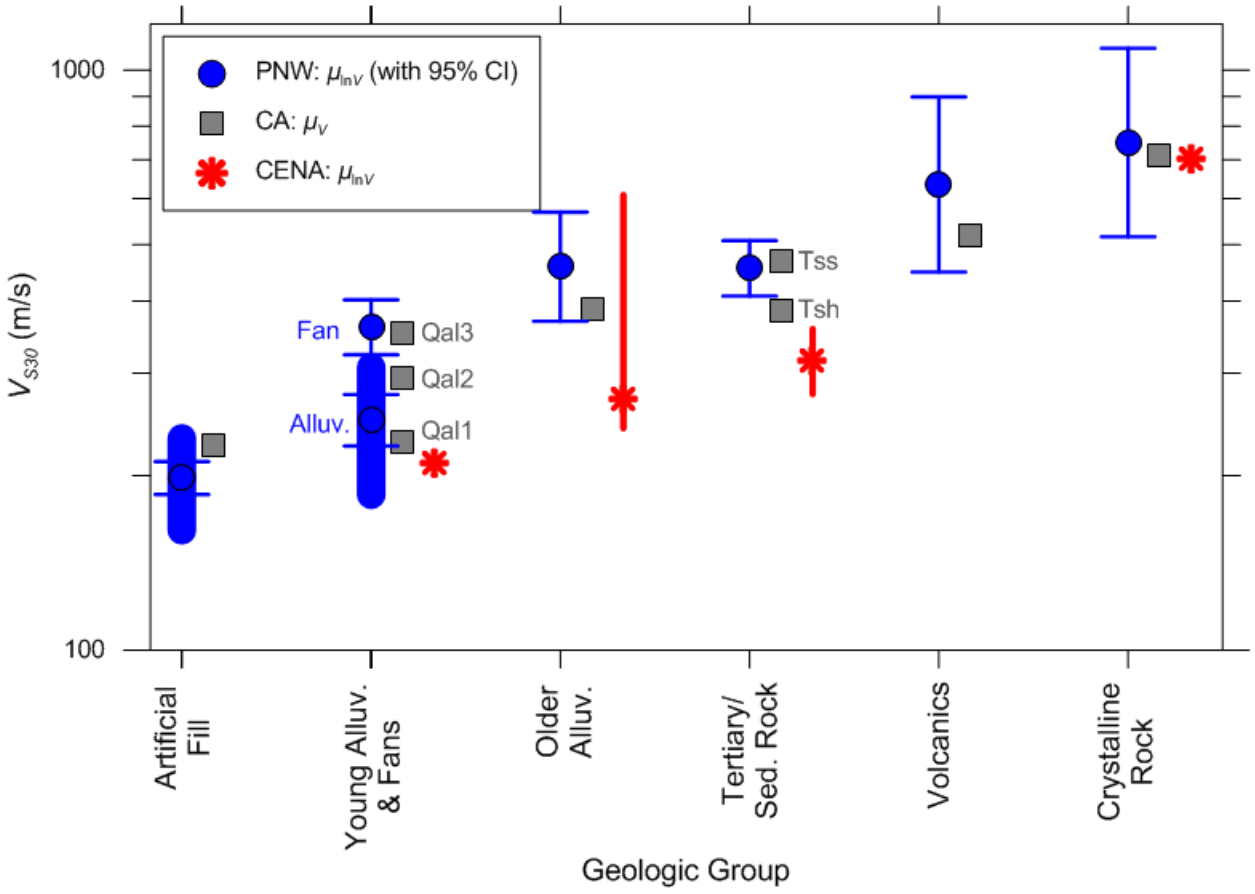

Figure 3.13. Comparison of geology proxy-based group natural log mean V_{S30} values from three different tectonic regions of North America. The means for CA are arithmetic. Elongated symbols for PNW and CENA span the range of topographic gradient data within applicable groups, per Table 3.3. PNW data presented with 95% confidence intervals (horizontal bars).

6, and 10, the confidence interval range is plotted for the median slope). Overall, there is a surprising level of similarity between the mean statistics, especially between California and the PNW. Some notable differences include much stronger gradient-dependence for California alluvium than the other regions, and lower Pleistocene alluvium and Tertiary means in CENA. Note that *it would be misleading to conclude from this figure that regional geology-based proxies are not needed*, due to features not shown in Figure 3.13, including groups unique to the respective regions and variable standard deviations (especially in CENA relative to other regions).

3.8 RECOMMENDATIONS FOR V_{S30} ASSIGNMENTS

It is well-established that best practices in site characterization for seismic analysis include direct measurement of seismic velocities, preferably extending to firm materials such as rock. The site parameter V_{S30} can be readily computed from the V_S profile, or in the case that the profile depth is < 30 m, estimated from shallower profiles. The proxy relationships developed in this chapter are recommended when seismic velocity measurements are unavailable, and it is necessary to estimate V_{S30}. The potential for relationships between the frequency of the peak in horizontal-to-vertical spectral ratios and V_{S30} is acknowledged as an alternative approach (Hassani and Atkinson, 2016), although such relationships have not been developed and calibrated in the PNW region. For application to V_{S30} assignments for PNW sites in the NGA-Sub SMS database (described in Chapter 2 of this dissertation and presented in Table S3 in the electronic supplement to Ahdi et al. 2017a), the following protocols were applied (in order of preference), which are similar to those described in Section 2.4 and Table 2.7:

0. Assign mean V_{S30} as computed using profile with z_p > 30 m. Standard deviation taken as $\sigma_{\ln V}$ = 0.1 per Seyhan et al. (2014).
1. For sites with a V_S profile that extends to depth z_p < 30 m, estimate V_{S30} using the

extrapolation relationship given by Equations (3.1–3.4) with the coefficients in Table 3.1. The associated $\sigma_{\ln V}$ can be taken as the square-root sum of variances associated with the depth extrapolation (σ_e^2) (Equation A.10) and 0.1^2, as follows:

$$\sigma_{\ln V} = \sqrt{\sigma_e^{\,2} + 0.1^2} \tag{3.10}$$

2. Estimate V_{S30} and its uncertainty using the predictive models developed in this chapter:
    a. Use the hybrid slope-geology proxy when applicable. Take the mean and standard deviation from Table 3.3, except when coefficients c_0 and c_1 are given in Table 3.3, in which case take the mean using those coefficients with Equation (3.5).
    b. Use the IP07 terrain category proxy only when geology is unavailable.

Under Code 2 above, the geology-slope proxy is given preference over the terrain proxy, based on lower $\sigma_{\ln V}$ and strong correlation between predictions from the two models as documented for PNW sites in Section 3.7.1 above.

For sites outside of the PNW region but in the NGA-Subduction site database (generally California, the inter-mountain west, and CENA regions; 691 sites in total), Protocols 0 and 1 are applied where V_S profile data is available. Where such data is missing, V_{S30} was estimated as follows:

3. Estimate V_{S30} and its uncertainty using proxies for regions outside of PNW, as follows:
    a. Sites in California (253): Use the average of $\mu_{\ln V}$ and $\sigma_{\ln V}$ values from Wills et al. (2015) (surface geology proxy) and Yong (2016);
    b. Sites in inter-mountain west (367): Use category $\mu_{\ln V}$ values from Yong (2016). An epistemic uncertainty σ_{ep} = 0.2 is assigned on the mean value.

c. Sites in CENA (71): Apply protocols given in Parker et al. (2017).

The numbers in the above list are codes provided in the SMS database (described in Chapter 2 of this dissertation and presented in Table S3 in the electronic supplement to Ahdi et al. 2017a). Of the 1126 SMS sites in the NGA-Sub SDB that have recorded Cascadia subduction zone earthquakes, 79 (7.1%) are Code 0 (on-site V_{S30} measurement to depth $z_p \geq 30$ m), 6 (0.6%) are Code 1 (V_{S30} measurement to depth $z_p < 30$ m), 374 (29.9%) are Code 2 (assigned based on the proxies developed in this paper), and 653 (62.5%) are Code 3 (assigned based on various previously developed proxies). Fourteen sites have no V_{S30} assignment.

3.9 SUMMARY AND CONCLUSIONS

A large majority of seismic recording stations in the Pacific Northwest (PNW) region of the U.S. lack geophysical measurements, which necessitated the estimation of site parameters from suitable proxies to facilitate the development and application of GMMs and site amplification models. In this chapter, the development of proxy-based models for estimation of the site parameter V_{S30} was presented, the immediate application of which was in the NGA-Subduction project (Chapter 2). The basis for estimating basin depth parameters was also described, using available basin models for the PNW region.

A V_S measurement database was compiled that draws upon a variety of public domain sources; information on each entry in the database is given in Table S2 (available in the electronic supplement to Ahdi et al. 2017a). Geologic information was utilized from map resources compiled for Oregon, Washington, and British Columbia, supplemented by mapping that indicates glaciation/non-glaciation and the presence of sedimentary basins.

Tables 3.3 and 3.5 present coefficients needed to apply the recommended proxy relationships, which are based on a combination of surface geology and topographic gradient

(hybrid slope-geology proxy) and categories associated with a geomorphic terrain classification scheme (terrain proxy). Proxies of this type have been applied previously elsewhere, but are applied to the subject region for the first time here.

Eighteen geologic groups were identified that encompass most conditions encountered in the study region, and models for V_{S30} estimation in each group were proposed. Some groups take the mean V_{S30} as a simple category natural log mean, whereas others take the mean from a gradient-based model specific to the category using Equation (3.5). Three groups borrow data from Alaska (Chapter 4 of this dissertation and Ahdi et al. 2017b) to improve the robustness of model predictions. Values of $\sigma_{\ln V}$ accompanying each mean estimate are given in Table 3.5. These estimates are used when more reliable, site-specific information is unavailable, as given by the implementation procedures in the previous section.

Sixteen geomorphic terrain classes that are globally available (Iwahashi and Pike, 2007) and that have been used previously for regional application (Yong et al. 2012; Yong 2016) are adopted herein. Natural log mean and standard deviation values are provided in Table 3.5 for classes that are well-populated in the PNW. These values show modest differences from previously proposed category moments in California (Yong 2016). A strong correlation was found between predictions from the hybrid geology-slope and geomorphic terrain proxy-based V_{S30} prediction models using PNW data. This correlation, combined with the lower standard deviation of residuals from the model that uses the hybrid geology-slope proxy, led to the recommendation to use this method over the terrain class model when both are available. This finding is specific to the PNW region and should not be assumed to apply elsewhere.

3.10 DATA AND RESOURCES

The velocity profile data utilized in this study were obtained from a variety of open sources,

none of which were proprietary. Main sources are described in the *Database Attributes* section (3.3.1) above. The information directly used in the analyses presented in this study is presented in its entirety in the electronic supplements to Ahdi et al. (2017a). A near-global, 30 arc-sec-resolution DEM from the Shuttle Radar Topography Mission (SRTM30; Farr & Kobrick, 2000) was used in this study to compute elevation and topographic gradients at PDB and SDB stations (http://www2.jpl.nasa.gov/srtm/; last accessed May 2016). Globally-available automated terrain classification data can be found at http://gisstar.gsi.go.jp/terrain/front_page.htm (last accessed May 2016).

Surface geological map data were obtained from the state geological surveys of Oregon (Smith & Roe 2015; http://www.oregongeology.org/pubs/dds/p-OGDC-6.htm) and Washington (WA-DNR-GER, 2010 & 2014; http://www.dnr.wa.gov/programs-and-services/geology/publications-and-data/gis-data-and-databases); and the Geological Survey of Canada for southern British Columbia (http://geoscan.nrcan.gc.ca/starweb/geoscan/servlet.starweb?path=geoscan/fulle.web&search1=R=194084); (all were last accessed July 2016). Locations of volcanoes were obtained from the Smithsonian Institute's Global Volcanism Program (http://dx.doi.org/10.5479/si.GVP.VOTW4-2013; last accessed July 2016). Analyses were performed using R (www.r-project.org), and maps created using Quantum GIS (http://www.qgis.org). Physical and hypsometric map imagery were obtained from http://www.naturalearthdata.com/.

4 Proxy-Based V_{S30} Prediction in Alaska Accounting for Limited Regional Data

4.1 INTRODUCTION

This chapter summarizes methods and recommendations for estimating V_{S30} for geologic conditions in Alaska. A major initial application of these methods is for estimating site parameters at SMS stations that have contributed data to the NGA-Sub project (Chapter 2 of this dissertation) and for which geophysical data are not available. The methodology, which mirrors that presented in Chapter 3 for V_{S30} proxy development in the PNW, involves estimating the natural log mean and standard deviation of V_{S30} from secondary information derived from surface geologic maps and digital elevation models, which is generally available across the study region.

A crucial element of this effort was the compilation of a database of seismic velocity profiles for the study region, which is termed a profile database (PDB). (It is noted here that in this chapter, "PDB" refers to a data compilation of V_S profiles, not a relational database as described in Chapter 6 of this dissertation.) The PDB for Alaska consists of 126 geophysical profiles and/or V_{S30} values, clustered in Anchorage and a few other major urban areas, and selected sites that were affected by the 2002 Denali earthquake. Because data for Alaska is relatively sparse, the Alaska PDB was supplemented with selected information from the PNW PDB as described in Chapter 3 (Ahdi et al. 2017a). While the scope of that study did not include the development of proxy-based

V_{S30} prediction models for Alaska, commonalities across geologic groups were leveraged for model development because of the generally similar tectonic regime and geology.

The development process of the proxy-based V_{S30} model for Alaska is described in this chapter. Approach II for model development (described in Section 2.3.1) was taken, in which (1) category models are developed using local (Alaska) data when justified by data quality and quantity, (2) adopted from a PNW proxy for poorly populated geologic conditions, and (3) adopted from California for certain conditions where necessary. Models conditioned on terrain classes were not utilized for Alaska; the reader is referred to Chapter 3 for the companion study in the PNW for a discussion of these models, which found them to have less predictive power than geology-based models.

The geophysical data and proxies assembled as a part of this study were used to populate the NGA-Sub SDB for ground motion recording sites in Alaska (Chapter 2 of this dissertation). Similar to the proxy-based V_{S30} model developed for Cascadia, model development presented in this chapter follows protocols that mirror similar efforts for NGA-West 2 and NGA-East (e.g., Seyhan et al. 2014; Parker et al. 2017) with some modifications.

4.2 V_S PROFILE DATABASE FOR ALASKA

4.2.1 Database Attributes

A V_S PDB was compiled for Alaska that consists of a digitized collection of 90 V_S profiles, with an additional 36 sites for which a measured V_{S30} is available, but not a V_S profile. The PNW PDB described in Chapter 3 is considered as part of this study, with 917 V_S profiles and 11 sites with a measured V_{S30} value. Profiles are considered when they are based on geophysical testing and extend to a maximum (profile) depth $z_p \geq 6$ m, have known geodetic coordinates (i.e., latitude and longitude), and are derived from geophysical measurement techniques that are considered

credible, at least for obtaining V_{S30}.

The contents of the Alaska PDB include V_S profile information (site identification number, time-averaged velocities to different depths, profile depth) and metadata related to site location (proxies and their sources). The PDB file for PNW sites is an electronic supplement to Ahdi et al. (2017a), which also contains a more detailed description of the database contents and the data sources. The PDB file for Alaska will be presented as part of the U.S. Community V_S Profile Database (described in Chapter 6 of this dissertation).

Sites included in the Alaska PDB are predominantly non-SMS sites in Fairbanks (Cox et al. 2012), SMS and other sites that recorded or were affected by the 2002 Denali earthquake (Kayen et al. 2004), a series of sites in Seward and Valdez (R. Kayen, 2016, *pers. comm.*), and two SMS sites in Anchorage (SW-AA 1980, Steidl et al. 2004). The Cox et al. (2012), Kayen et al. (2004) and Kayen (2016, *pers. comm.*) sites were investigated using non-invasive SWMs including SASW and MASW (88 sites total). The Anchorage sites were investigated using downhole (SW-AA 1980) and suspension logging (Steidl et al. 2004).

A project-level priority was to include as many profiles as possible, and thus discrimination based on geophysical measurement methods was not undertaken with one exception. V_S profiles derived using the controlled-source surface-wave dispersion measurement method (CXW) developed by Poran et al. (1994) were not considered to be credible. This method was used to collect data at 36 sites in Anchorage (Dutta et al., 2000). Judgment on the suitability of this method draws upon the recommendations of Boore and Brown (1998) and Wills (1998), who caution that V_S profiles derived from this method are biased compared to those derived from invasive methods such as downhole and crosshole. However, V_{S30} values were retained for these sites in the present work, as they do not vary significantly from V_{S30} values obtained from other methods, as

determined for sites in Los Angeles (Boore and Brown 1998). Further discussion of the consideration of the use of CXW and other potentially-biased methods is provided in Section 6.2 describing the U.S. Community V_S Profile Database project (Chapter 6 of this dissertation).

Figure 4.1 shows V_S/V_{S30} measurements and SMS locations in Alaska. The PDB sites are concentrated in Anchorage, Fairbanks, Seward, Valdez, and other locations affected by the 2002 Denali earthquake. Most measurements are located within alluvial or basin regions. Figure 3.1 shows the locations of PDB sites in southern British Columbia, Oregon, and Washington, where V_S measurements are less generally clustered than in Alaska, but is especially abundant in Portland, OR, Seattle, WA, and the Fraser River Delta region south of Vancouver, BC.

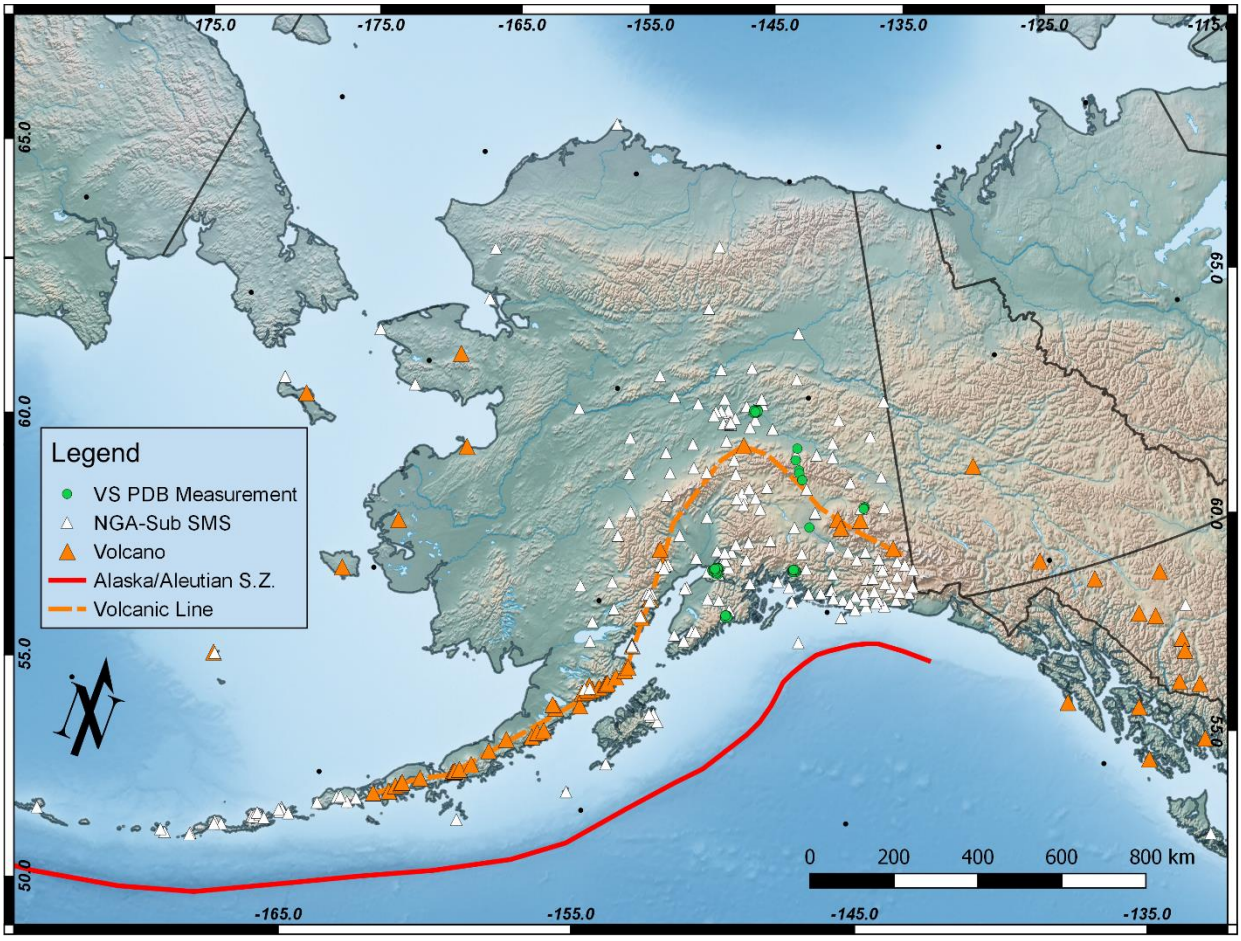

Figure 4.1. Map of Alaska showing locations of PDB and SMS sites with respect to the volcanic line separating forearc (south) from backarc (north) sites. Black dots mark latitude/longitude intersections at five-degree intervals.

4.2.2 V_{S30} Computation

V_{S30} is computed using Equations (2.1) and (2.2) as described in Section 2.2.2. When $z_p \geq$ 30 m, which occurs for 651 of 1007 PDB sites, V_{S30} is computed by replacing z_p with 30 m in Equations (2.1) and (2.2). For the sites having $z_p < 30$ m, V_{S30} was estimated by extrapolation using Dai et al. (2013), using PNW region-specific regression coefficients described in Chapter 3 and in Kwak et al. (2017a).

4.3 PROXY ATTRIBUTION

Metadata was compiled for PDB sites in Alaska as well as for NGA-Sub SMS sites. Compiled data includes surficial geologic site conditions, an indicator regarding site location within or beyond the extent of the Cordilleran ice sheet that was present during the late Pleistocene (this applies to all of the sites in Alaska), indicators of site location with respect to the volcanic front separating forearc from backarc regions (described in Section 2.6), and topographic slope from the SRTM30 DEM at 30 arcsec resolution. While the Iwahashi and Pike (2007) geomorphic terrain classes were compiled and use for proxy development in the PNW, this was not done for Alaska, for two reasons. First, much of Alaska was not covered by the original Iwahashi and Pike (2007) global raster map (their Figure 6); large stretches of Alaska and other high latitudes were marked with terrain class 17, signifying "no data". Second, the PNW study showed that the proxies of (a) hybrid geology/slope and (b) terrain were highly correlated, with the terrain proxy-based model performing more poorly based on residuals analysis of predicted to measured data (see Section 3.7.1).

Two compilations of geological maps for Alaska were considered for proxy model development. One is a digital compilation of surface geologic maps by Wilson et al. (2015). Preference was not given to use these maps because of their small scale (1:584,000), which results

in limited resolution of Quaternary unit descriptions and unit boundaries. The other compilation is the National Geologic Map Database (NGMDB) of the USGS, which was utilized to look up maps at larger scales, where available. Where large-scale maps are not available in Alaska, the NGMDB reverts to the Wilson et al. (2015) map. The surficial geologic maps delineate locations of Quaternary sediments (e.g., alluvium, till, loess) and outcropping rock units. Maps of relatively large scale typically delineate a more refined age distinction, particularly among Quaternary sediments. Information extracted for each location includes geologic unit, age, and description. In addition to the map information, the surface geologic conditions as identified by geologist's site visits or as attributed to uppermost layers in a V_S profile were recorded for each site in the PDB and NGA-Sub SMS list for Alaska, where available.

Porter et al. (1983) mapped the extent of the Late Wisconsin Cordilleran ice sheet, which was present during the Late Pleistocene, in the PNW/Rocky Mountains region. Pierce (2003) provides an updated map which was used in this study. As mentioned previously, essentially the entirety of Alaska plots within the glaciated region. The limits for the PNW are shown in Figure 3.1.

Figures 3.1 and 4.1 show the interpretation of the volcanic line separating forearc and backarc regions, with forearc to the south/west and backarc to the north/east, for Alaska and the PNW, respectively. The volcanic line was drawn "by eye" through volcano locations from the Global Volcanism Program (2013). This procedure is discussed in greater detail in Section 2.6. The PDB and SDB files include tags indicating whether sites are located in the forearc or backarc regions (Ahdi et al. 2017a).

4.4 PNW HYBRID GEOLOGY/SLOPE PROXY

Although the focus of the work in this chapter is on a proxy-based V_{S30} prediction model

for Alaska, the manner by which the work was undertaken was to first develop models for the PNW, and then to test the applicability of those models (with modifications as needed) for Alaska, i.e. per Approach II (described in Section 2.3.2). For this reason, this section reviews the proxy development process for the PNW, then in Section 4.6 the Alaska model development is explained in detail.

The 928 PDB locations in the PNW were grouped to identify features that produce distinct V_{S30} distributions, a process described in detail in Chapter 3 (Ahdi et al. 2017a). Assuming a log-normal V_{S30} distribution, data distributions were represented within groups by $\mu_{\ln V}$ and $\sigma_{\ln V}$. Within each group, the mean trend of V_{S30} with 30 arcsec topographic gradient was considered.

As described in Section 3.5, the collection of PNW PDB sites fall in 124 distinct geologic units. Most sites are located on Quaternary sediments. The initial grouping was judgment-based, in consideration of lithological unit descriptions, depositional environment, and age; this resulted in 42 categories. For data within individual categories and various combinations of categories, moments $\mu_{\ln V}$ and $\sigma_{\ln V}$ were computed and a mean trend with topographic gradient (s) was evaluated as per Equation (3.5). The gradient effect is considered statistically significant when null does not fall within the 95% confidence intervals for c_1.

Statistical F-testing (explained in Section 3.6.2) of V_{S30} distributions within and across categories supported combining many of the 42 categories, such that 18 groups remained (Table 4.1, part of which is a reproduction of Table 3.3 for comparative purposes), each having a geologic description and a unique set of attributes ($\mu_{\ln V}$, $\sigma_{\ln V}$, and trend with topographic gradient). Descriptions of lithology and depositional environment were used as the primary discriminator among groups, as opposed to using geologic age as the first order discriminator, as such descriptions can be associated with multiple ages (e.g., alluvium, Group 6, has three age bins—

Table 4.1. Summary of geology-based and hybrid geology-slope-based V_{S30} proxy for Alaska and PNW. For Alaska, the Class designation indicates the approach used to arrive at the recommended natural log mean model and standard deviation values (Table 4.2).

Group	Description	Alaska						PNW				
		N	$\mu_{\ln V}$ (m/s)	$\sigma_{\ln V}$	c_0	c_1	Class[1]	*N*	$\mu_{\ln V}$ (m/s)	$\sigma_{\ln V}$	c_0	c_1
1	Peat	0	161	0.522	*	*	IV	68	161	0.348	*	*
2	Fraser River: overbank silt/clay	0	182	0.395	5.520	0.0506	IV	74	182	0.259	5.520	0.0506
3	Fraser River: overbank sand/silt, sandy/clayey loam, channel deposits	0	198	0.263	*	*	IV	122	198	0.263	*	*
4	artificial fill	4	198	0.314	5.625	0.0762	III	89	198	0.314	5.625	0.0762
5	fluvial + estuarine deposits	0	239	0.867	*	*	IV	31	239	0.578	*	*
6	alluvium & valley sediments	45	323	0.365	5.928	0.0266	I	90	249	0.496	5.976	0.1002
7	flood deposits: sands, fines, floodplain, undifferentiated	2	322	0.243	5.904	0.0275	IV	91	322	0.243	5.904	0.0275
8	lacustrine (incl. glaciolacustrine)	10	326	0.135	6.057	0.0657	II	10	326	0.135	6.057	0.0657
9	beach, bar, dune deposits	0	339	0.647	6.326	0.1264	IV	20	339	0.431	6.326	0.1264
10	fan deposits	37	360	0.338	*	*	II	37	360	0.338	*	*
11	Loess	22	376	0.380	*	*	II	22	376	0.380	*	*
12	glacigenic sediments (drift & outwash)	17	399	0.305	*	*	III	68	399	0.305	*	*
13	flood deposits: channel, gravel, coarse	0	448	0.432	*	*	IV	37	448	0.288	*	*
14	glacial moraines & till	3	453	0.512	*	*	IV[2]	66	453	0.341	*	*
15	undifferentiated sediments & sedimentary rocks	1	455	0.545	*	*	IV	42	455	0.363	*	*
16	terrace deposits & old alluvium	0	458	0.761	*	*	IV	21	458	0.507	*	*
17	volcanic rocks & deposits	1	635	0.995	*	*	IV	14	635	0.663	*	*
18	crystalline rocks (igneous & metamorphic)	0	750	0.641	*	*	IV	5	750	0.427	*	*

[1] Class V assignments specific to the Alaska site database for NGA-Sub include 10 sites on mélange rocks. V_{S30} moments, borrowed from California (Wills et al. 2015) for geologic group KJf, are $\mu_{\ln V}$ = 665 m/s and $\sigma_{\ln V}$ = 0.662 (after increase by σ_{ep} = 0.2 to account for epistemic uncertainty).
[2] Group 14 has 66 profiles in the PNW, and 3 in Alaska, which meets the threshold. However, these 3 are closely concentrated and have V_{S30} values that are significantly lower than the mean of Group 14 in the PNW. Thus, these 3 were not used in model development for Alaska.

Holocene, Pleistocene, and undivided Quaternary). The lack of further discrimination by age in these cases resulted to some degree from limited data in certain age bins, but to the extent that the data is available, age was not observed to have predictive power for V_{S30} within these bins.

Site location within or beyond the extent of the Cordilleran ice sheet did not affect V_{S30} beyond the classifications present in Table 4.1. Site location within or outside of basins also did

not carry predictive power (Section 3.7).

The proposed model for the PNW provides estimates of $\mu_{\ln V}$ and $\sigma_{\ln V}$ conditional on generalized geologic groups and, in some cases, topographic slope s. For groups without provided values of c_0 and c_1, the mean is taken as the gradient-independent $\mu_{\ln V}$ value from Table 4.1. For other groups, the mean is taken from Equation (3.5) using the coefficients in Table 4.1. For both model types, $\sigma_{\ln V}$ is provided in Table 4.1.

4.5 ALASKA HYBRID GEOLOGY/SLOPE PROXY

The range of geologic conditions at PDB sites in Alaska is much more limited than those for the PNW, due to data concentrations in a few locations. Nearly all Anchorage sites are founded on overconsolidated sediments overlying variable thicknesses of clay of the Bootlegger Cove Formation (Updike 1985). These sites receive various alluvial, fan, lacustrine, and glacial till/moraine geologic classifications on the USGS NGMDB maps. In Fairbanks, all sites were mapped as alluvial, fan, or loess deposits. At other locations in Alaska, geologic maps generally have lower resolution (smaller scale), and as such many sites are mapped simply as alluvial, fan, or undifferentiated sedimentary deposits.

In general, the geologic conditions at PDB sites in Alaska can be associated with the same categories provided in Table 4.1, which were originally developed for the PNW. Accordingly, in many cases groups created for the PNW were applied to Alaska sites. While the geologic conditions present at strong motion sites are much more diverse than those for PDB sites, Table 4.1 nearly covers the range of conditions presented in the Alaska portion of the NGA-Sub SDB; for example, the Bootlegger Cove formation is classified in Group 8 (glaciolactustrine sediments). Exceptions include tectonic mélange rocks (10 sites in the Alaska SDB) and tidal-flat deposits (one site in the Alaska PDB), which are assigned V_{S30} moments from proxies developed for

California (i.e., Wills et al. 2015).

In developing guidelines for assigning V_{S30} to sites in Alaska, Approach II was developed (upon similar work by Scasserra et al. 2009) to be adaptable to groups with varying amounts of data. Where data is available and the velocities from Alaska are judged to be distinct from those in the PNW, an Alaska-derived model was used, comparable in form to those developed for PNW (but using different coefficients). This applies to Group 6 (alluvium) only, which is denoted Class I. Where data is present but limited (at least three profiles), the applicability of PNW models to Alaskan sites was tested using residuals analysis. In some of these groups, the proxy-based V_{S30} estimates for the PNW include Alaska data due to their similarly (Class II); in others they do not use Alaska data, but the group data across regions nonetheless appear to be similar (Class III). Additional groups accommodate cases with no Alaska data (Classes IV and V).

The residuals of log mean V_{S30} for candidate groups were calculated using Equation (3.9) using the procedure described in Section 3.7.1. If the mean ± the standard error of the residuals encompassed zero, that group's bias with respect to the PNW model was deemed to be statistically insignificant for Alaska data and the PNW model was adopted. Table 4.2 shows candidate groups that were tested in this manner, and resultant classes into which each group was placed for the purpose of V_{S30} assignment.

Table 4.2. Summary of classes for V_{S30} proxy development for Alaska based on number of profiles N in each geologic group.

Group	N	Bias, $\bar{R}$ (m/s)	Dispersion, $\sigma_{\ln V}$	Standard Error	Class
4	4	0.249	0.573	0.287	III
6	45	0.384	0.254	0.038	I
8	5	0.013	0.127	0.057	II
10	32	-0.086	0.211	0.037	II
11	14	-0.073	0.259	0.069	II
12	17	-0.012	0.309	0.075	III
14	3	-0.840	0.217	0.125	IV[1]

[1] See Note 2 from Table 1.

To summarize the recommended approach, the classes for V_{S30} moment assignment are as follows:

I. Use good quality Alaska data when available to develop Alaska-specific model (i.e. Group 6 – alluvium).

II. Use models developed jointly with PNW and Alaska data, as described in Chapter 3 (per Ahdi et al. 2017a); use these moments after checking the bias of the residuals and verifying that they are statistically insignificant (i.e. Groups 8, 10, 11).

III. Borrow as-published PNW proxy, but check the significance of bias of residuals; if bias insignificant, use the original $\sigma_{\ln V}$. (i.e. Groups 4 and 12).

IV. Borrow as-published PNW proxy without checking residuals due to lack of data; inflate by $\sigma_{ep} = 0.2$ (all other groups).

V. Borrow as-published proxy model from other regions and increase sigma by $\sigma_{ep} = 0.2$ (e.g. one SMS site on Qi, tidal-flat deposits, borrows V_{S30} moments from Wills et al. 2015).

4.6 RECOMMENDATIONS FOR V_{S30} ASSIGNMENTS

To echo Section 3.8, best practices in site characterization for seismic analysis include direct V_S measurement, preferably extending to firm materials such as rock. V_{S30} can be readily computed from the V_S profile, or in the case that the profile depth is < 30 m, estimated from shallower profiles. The proxy relationships developed in this chapter are recommended when seismic velocity measurements are unavailable for Alaskan sites, and it is necessary to estimate V_{S30}. For application to V_{S30} assignments in the NGA-Sub site database, the same protocols for measured sites (Codes 0 and 1) as those for the PNW (described in Section 3.8) are used, but Code

2 (proxy-based V_{S30} estimation) is modified to represent the proxies developed and used in this study. The Alaska-specific proxy-based V_{S30} assignment protocol for Code 2 is:

2. Estimate V_{S30} and its uncertainty using the models presented here. Take the natural log mean and standard deviation from Table 4.1, except when coefficients c_0 and c_1 are given in Table 4.1, in which case take the mean using those coefficients with Equation (3.5). Table 4.1 reflects class assignments I to V and their impact on uncertainty.

4.7 SUMMARY AND CONCLUSIONS

A large majority of seismic recording stations in Alaska lack geophysical measurements, which necessitates the estimation of site parameters from suitable proxies to facilitate the development and application of GMMs and site amplification models. This study focused on the development of proxy-based models for the estimation of V_{S30} for sites in Alaska, the immediate application of which was all SMSs in Alaska that have produced usable recordings from subduction events for the NGA-Sub project (as described in Chapter 2).

The main challenge that was faced in developing the models in this chapter is the lack of widely variable data availability across geology groups in the study region. A mixture of Approaches I, II, and III was implemented to use data to directly derive models that are justified by relatively abundant data (I); uses data to validate models borrowed from elsewhere where data is limited but finite (II); and simply borrows models for similar geologic conditions without validation when data is absent, but giving recognition to the additional uncertainty this process entails (III).

5 Proxy-based V_{S30} Estimates for Sites in the Iranian Strong Motion Database

5.1 INTRODUCTION

This chapter describes the process of compiling a site database for ground-motion recording station in Iran. A nation-wide strong-motion network has produced a dataset of 9756 unprocessed ground-motion records from 6203 events between 1973 and 2013. The focus of the discussion in this chapter is on Iranian SDB development, where a site (i.e., an SMS) was included and processed if it had produced at least one ground motion recording.

Figure 5.1 shows a map of Iran with the locations of 1135 SMSs considered in this study. Sites shown with a blue symbol (558 sites) are those that have been assigned a V_{S30} from previous studies, as described in Section 5.2. Seismic site conditions are unknown for the remaining 577 sites, which are marked with a red symbol.

Site conditions at ground-motion stations in Iran are sparse in the archival literature. Saffari et al. (2012) developed an earthquake ground-motion model for Iran where site conditions are represented as discrete site categories defined on the basis of V_{S30}, per the Iranian Code of Practice for the Seismic Resistant Design of Buildings (BHRC 2007); this is similar to NEHRP categories B-E (Dobry et al. 2000). Saffari et al. (2012) reference BHRC reports by Sinaeian et al. (2008; 2010a-f) as the source of their site data.

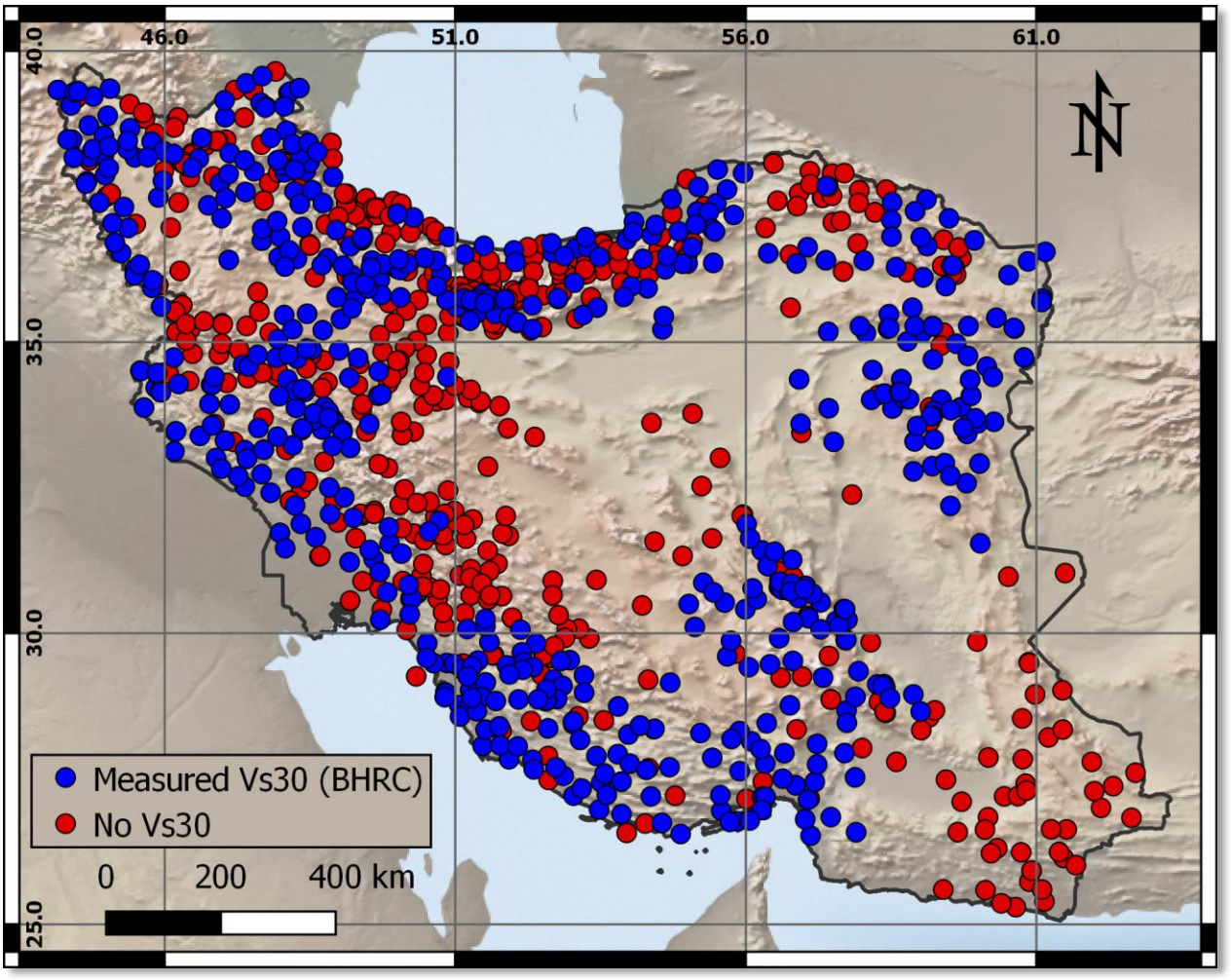

Figure 5.1. Map of Iran showing locations of earthquake ground-motion recording stations with and without reported site characterization, obtained from the BHRC.

Shafiee et al. (2016) have presented a model for extrapolating V_{SZ} to V_{S30}. The source of the data used in their study is indicated as "Zamin Physic Pouya Consulting Engineers" in reports dated 2005–2013. It is unknown if this data is different from the data used by Saffari et al. (2012).

Section 5.2 presents the following: (1) data sources considered in the present work; and (2) analysis of the V_{S30} data from the BHRC reports. Section 5.3 assesses the reliability of this data and a justification for the "preferred V_{S30} values" applied to the 1135 SMSs. The site database compiled through this work is presented in Appendix B of this dissertation.

5.2 DATA SOURCES

5.2.1 Shear-Wave Velocity

As mentioned in Section 5.1, 558 sites have assigned V_{S30} values from V_S measurements for ground-motion stations. It is assumed that these V_{S30} values were computed from V_S profiles in reports by BHRC; see Appendix B for V_{S30} values as assigned to applicable sites.

Seven reports by BHRC (Sinaeian et al. 2008; 2010a-f) contain station names and codes which often do not match those in the Iranian site database. Accordingly, it was generally not feasible to link entries in those reports to specific ground-motion recording sites. In some cases, it appears that the same measured V_{S30} value was applied to multiple ground motion stations. For example, three sites near Tehran, Qazvin1 (QazviX), Qazvin1 (Qazvi1Y), and Qazvin2 (Qazv2), are assigned the same V_{S30} value of 380 m/s. If these are from the same profile, which cannot be confirmed, the V_{S30} assignments are not in conformance with the standards of NGA-type projects. This is because the inter-station distances are large (2.5 km, 5.0 km, and 5.7 km), and V_S profiles should only be associated with an SMS if the separation distance is much smaller—typically < 300 m (Borcherdt 2002; Ancheta et al. 2013). Because it is not possible to associate profiles in the seven BHRC reports with ground-motion stations, these issues could not be checked systematically. It was apparent that relatively liberal profile-station site associations were made that increase uncertainty in the V_{S30} assignments.

All velocity measurements in the seven BHRC reports are based on seismic refraction. The reports cumulatively document velocity data for 350 sites. For 50 of these sites in the Sinaeian et al. report (2008), both *P*- and *S*-wave profile data are reported, both as velocity model cross sections from seismic refraction, and in tabular format. For the remaining 300 sites in the Sinaeian et al. reports (2010a-f), only profiles of *P*-wave velocity are provided, except for four sites that

also provide *S*-wave profiles, but with no tabular information. These reports show raw travel time-distance data for *S*-waves; thus, it may be that V_S profiles were developed as part of the work, but those profiles were not provided. As a result, it was uncertain whether the V_{S30} values were based on measurement-based V_S profiles for these sites. The alternative is that V_P values may have been converted to V_S values using an assumed Poisson's ratio, and V_{S30} computed from those inferred profiles, an approach that is not typically considered in larger compilation studies of V_S (e.g., Chapters 2 and 6 of this dissertation). For subsequent work, the V_{S30} values provided in the BHRC reports were adopted as it was impossible to verify for most sites.

5.2.2 Surface Geology

As shown in Table 5.1, surface geology maps were obtained from two sources, with near full coverage of the country (95% of land area, see Figure 5.2). Appendix B lists the geologic conditions for each site as written in the legends of the maps. Note that there is the potential for geologic category misclassifications due to the relatively small scale (1:125,000) of the maps that were used; generally maps on the order of 1:50,000 or larger are preferred, such as those used in for model development in the PNW (see Table 3.2)

Table 5.1. Sources of geologic maps in Iran.

Source	Scale	Date range	# maps obtained	Notes
Geological Survey of Iran	1:250,000	1967-1996	94	Differentiation among Quaternary units
National Iran Oil Company	1:250,000	1963-2012	24	Focus on oil-bearing strata, Quaternary units lumped into a single unit

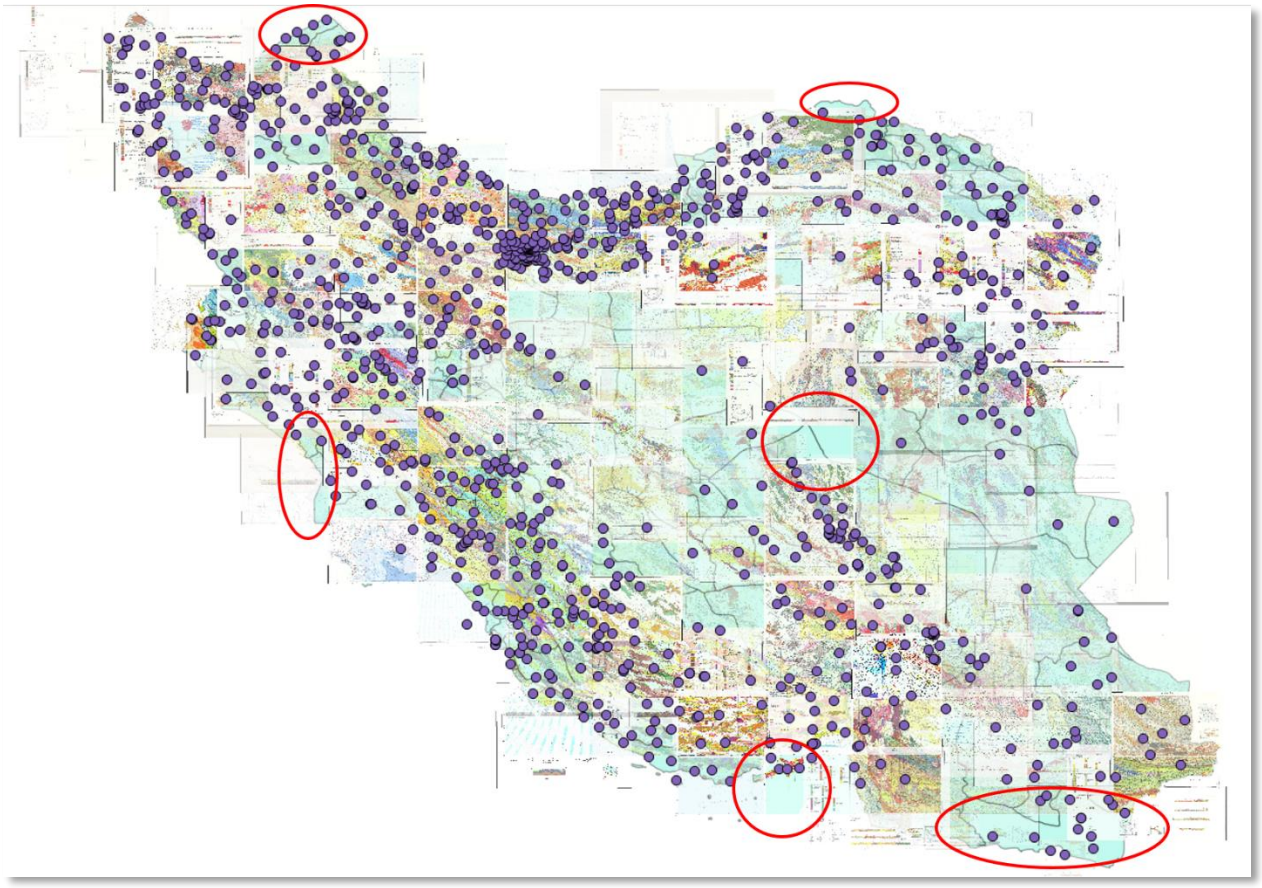

Figure 5.2. Coverage of geologic maps of Iran, with 12 sites located in small zones of land (red ovals, 5% of the country) lacking map coverage.

The range of geologic conditions at the 1135 sites was reviewed. It was necessary to apply judgment regarding conditions with distinct ages, depositional environments, and possibly soil types to distill the many conditions shown in Appendix B to the ten groups listed in Table 5.2. Similar grouping was applied in previous studies of this sort for other regions such as California (Wills et al. 2015); CENA (Parker et al. 2017); the Pacific Northwest portion of North America (Chapter 3 of this dissertation and Ahdi et al. 2017a); and Greece (Stewart et al. 2014a). The applicable unit designation from Table 5.2 is listed in Appendix B for each of the 1135 sites. Thirty sites have no group assignment for various reasons, as indicated in the Comment field in Appendix B.

The sedimentary Tertiary units in Table 5.2 may include Quaternary-mapped sites if: (1) the ages span into the upper Tertiary age (i.e., Plio-Pleistocene, Miocene/Pleistocene, Quaternary-Tertiary, etc.); and (2) if the lithological description of the Quaternary-mapped site indicates potential lithification. An example is site Kermanshah1 (KRM01), which is mapped as Quaternary marls and detritic deposits; marls warrant grouping under Tf.

Table 5.2. Unit descriptions adopted in this study to encompass common geologic conditions in Iran.

Group Designation	Description	No. sites
Qy	Quaternary: Active channel young alluvium	107
Qe	Quaternary: Eolian, including loess, dunes, windblown silt	22
Qc	Quaternary: Clay flat, mud flat, lacustrine	39
Qt	Quaternary: High/low terrace and fans and older alluvium	325
Qu	Quaternary: Undivided	328
Tc	Sedimentary Tertiary: Coarse – conglomerate, sandstone	48
Tf	Sedimentary Tertiary: Fine – limestone, marl, shale	97
Ms	Mesozoic and older sedimentary rocks	64
V	Volcanics: Tuff, basalt, lahars, etc.	61
Xln	Intrusive igneous and metamorphic rocks	14

5.3 TERRAIN AND SURFACE GRADIENT

The SRTM30 DEM at 30 arc-sec resolution (Farr and Kobrick 2000) was queried to extract topographic gradient, and elevation relative to sea level. Iwahashi and Pike (2007) terrain classes were also assigned, as described in Chapters 2 and 3. These results are listed for each of the 1135 sites in Appendix B.

5.4 EVALUATION OF SHEAR-WAVE VELOCITY DATA FROM BHRC REPORTS

Using data for the 558 ground motion recording sites with V_{S30} values, within-bin $\mu_{\ln V}$ and $\sigma_{\ln V}$ statistics were computed for each of the 10 geologic groups listed in Table 5.2 and each of the

16 geomorphic terrain classes recommended by Iwahashi and Pike (2007). The mean of the data was computed in natural log units. In the case of geologic groups, the dependence of V_{S30} on 30 arcsec gradient was examined using Equation (3.5).

Figure 5.3 shows histograms and V_{S30}-gradient plots for each of the 10 geologic groups. Table 5.3 summarizes the results for each group. Topographic slope-based regression coefficients c_0 and c_1 (used in Equation 3.5) were included only when the gradient was considered to be statistically significant based on visual inspection. Figure 5.4 presents the summary statistics of selected groups and compares them to generalized geology groups with similar descriptions that contain data from three other regions, all in North America (PNW, CENA, and California). A striking feature of the results in Figures 5.3 and 5.4 is the fast velocities for V_{S30} data from Iran, particularly in the Quaternary geologic units. For example, the mean velocity for young alluvium (Qy) is 652 m/s, while the similar geologic group (Group 6) in the PNW has $\mu_{\ln V}$ = 249 m/s (see Table 3.3). Also striking is the relatively small variations in mean velocity between geologic units, with all units other than Qe falling within the mean velocity range of 530–790 m/s. Figure 5.5 shows within-bin V_{S30} statistics for terrain classes along with the same results for California by Yong (2016). As with geologic classes, the results indicate much faster velocities in Iran than for California.

One possibility that might explain why the V_{S30} values for Iranian sites are high is shallow soil layers overlying rock. If this were the case, it could be that the measurements are accurate and reflect a geologic condition distinctly different from the averages elsewhere. To test this hypothesis, seven sites in the Tehran basin were selected, which are reported to have sediment depths ranging from 400–1400 m (Shirzad and Shomali 2014). Figure 5.6 shows the locations of those sites, which are all at least 5 km from the nearest bedrock outcrop, along with their V_{S30}

values, which range from 300–613 m/sec. Given that these sites are likely not affected by shallow bedrock, the V_{S30} values appear to be far too fast for the young sediments expected at these locations.

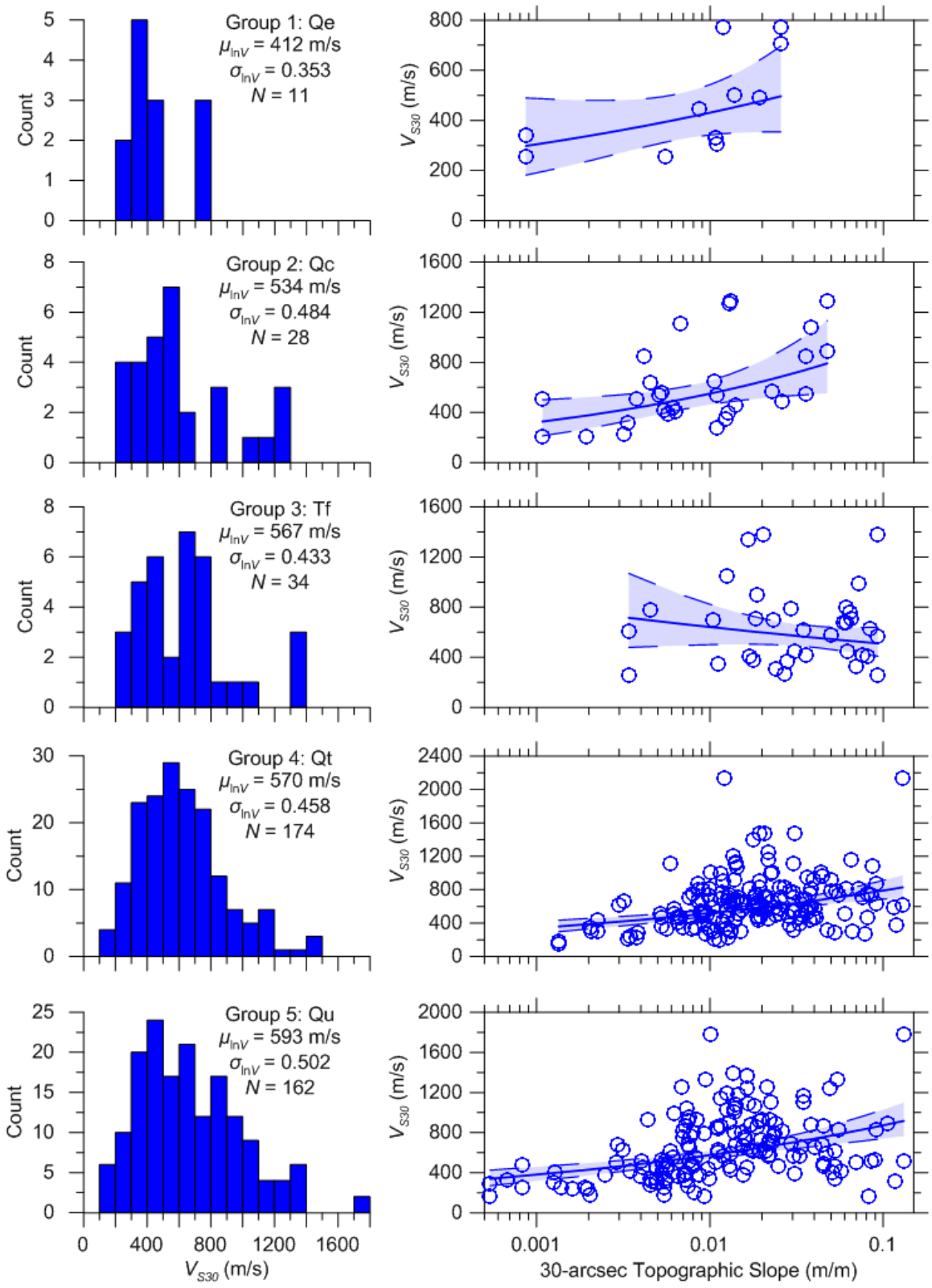

Figure 5.3. Geology proxy-based group V_{S30} distributions (left column) and topographic gradient dependency (right column). Plots in right column show power-law fits and 95% confidence intervals of data.

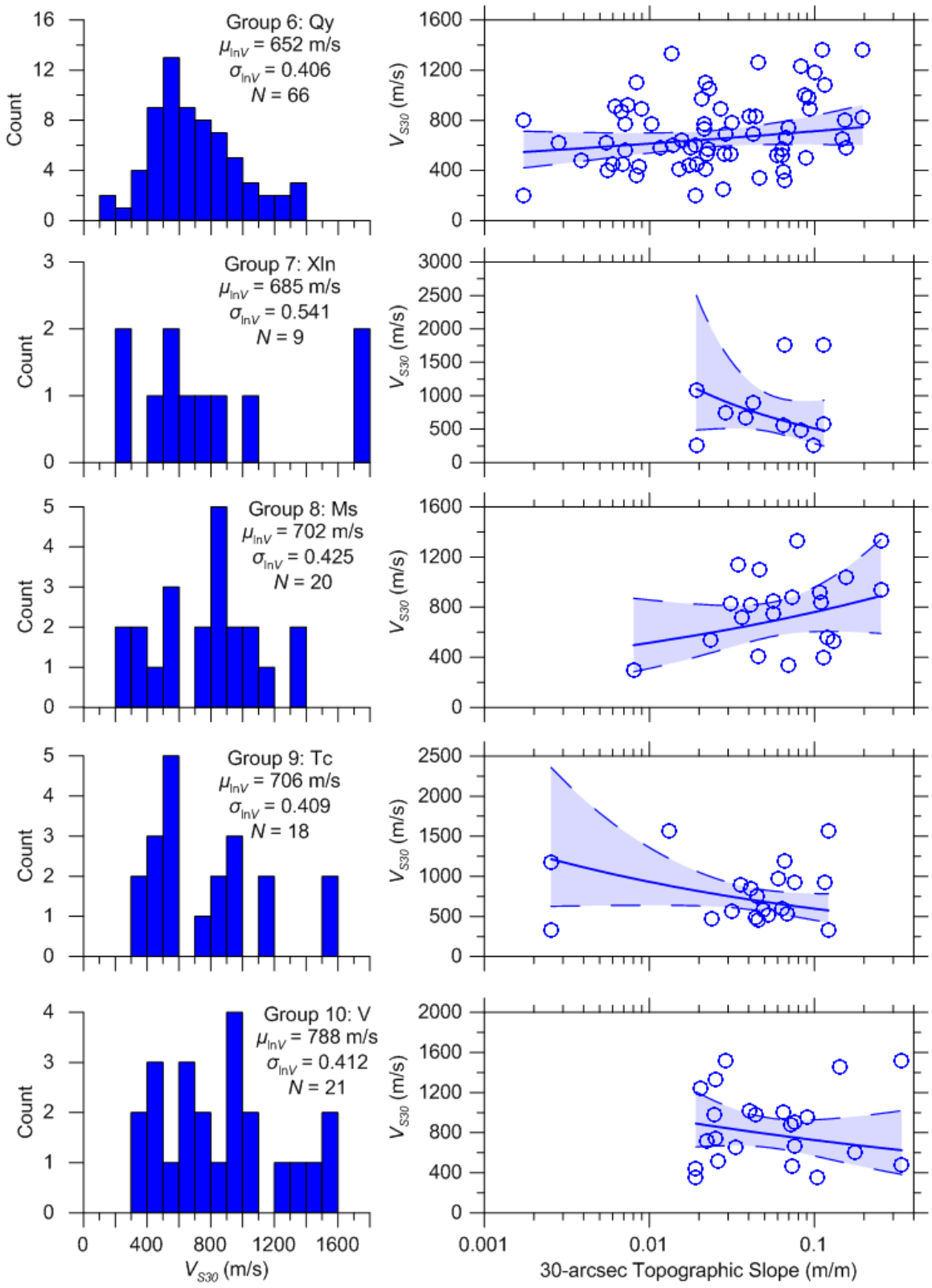

Figure 5.3 continued.

Table 5.3. Summary of geology-based and hybrid geology-slope-based V_{S30} proxy.

Group	Symbol	μ_{lnV} (m/sec)	σ_{lnV}	N	c_0	c_1
1	Qy	652	0.406	66	NA	NA
2	Qe	412	0.353	11	NA	NA
3	Qc	534	0.484	28	7.3855	0.2333
4	Qt	570	0.458	174	7.0923	0.1823
5	Qu	593	0.502	162	7.1887	0.1811
6	Tc	706	0.409	18	NA	NA
7	Tf	567	0.433	34	NA	NA
8	Ms	702	0.425	20	NA	NA
9	V	788	0.412	21	NA	NA
10	Xln	685	0.541	9	NA	NA

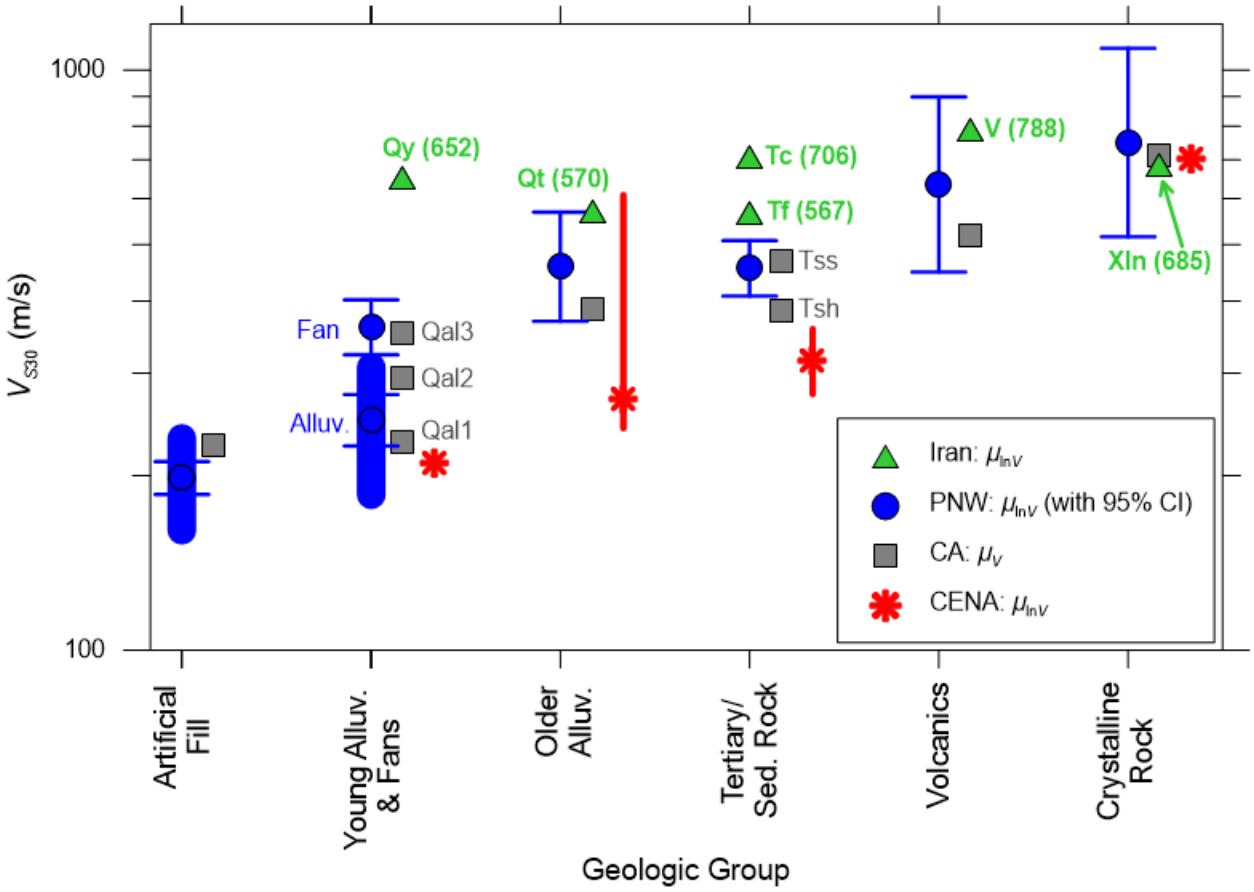

Figure 5.4. Comparison of geology proxy-based group natural log mean V_{S30} values from Iran with three different tectonic regions of North America. The means for CA are arithmetic. PNW data presented with 95% confidence intervals (horizontal bars). Figure modified from Figure 3.13 (Ahdi et al. 2017a, their Figure 13).

From the data interpretation presented in this section, it was concluded that the V_{S30} values provided in the BHRC reports (Sinaeian et al. 2008; 2010a-f) and assigned to SMSs are not reliable. The use of reflection and refraction methods for shallow (30 m) V_S measurement is not typical, and more conventional procedures should be undertaken in future work to further test the datasets provided in the Sinaeian et al. (2008; 2010a-f) reports. Until further verification work is undertaken, the existing data should not be used either for V_{S30} assignments at specific sites, or for the development of proxy-based predictive models.

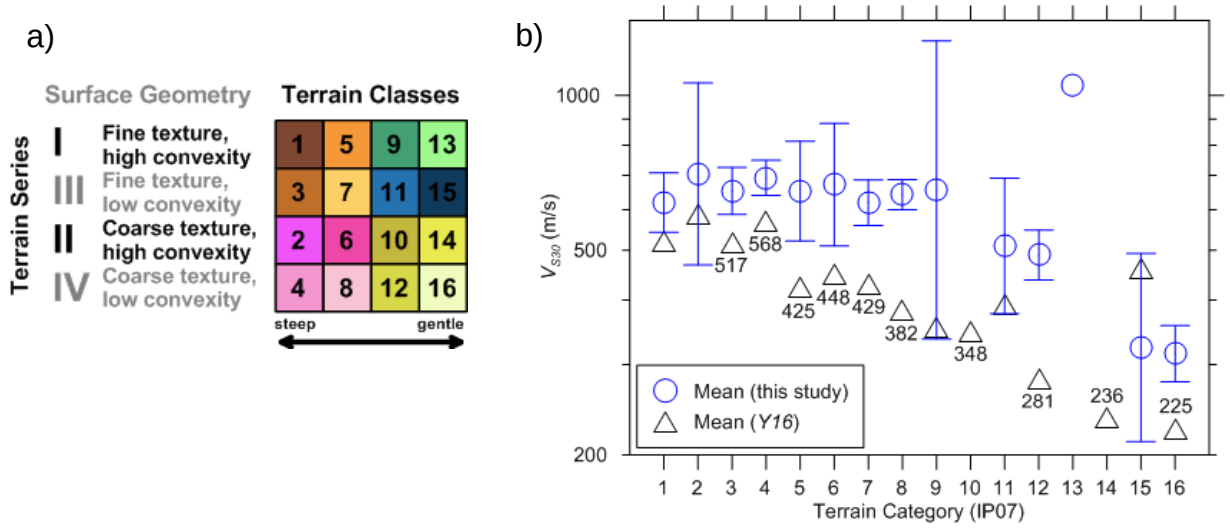

Figure 5.5. (a) Depiction of Iwahashi and Pike (2007) terrain categories; and (b) category mean V_{S30} values as given by Yong (2016) shown as *Y16*, and from the Iran site database (with confidence intervals).

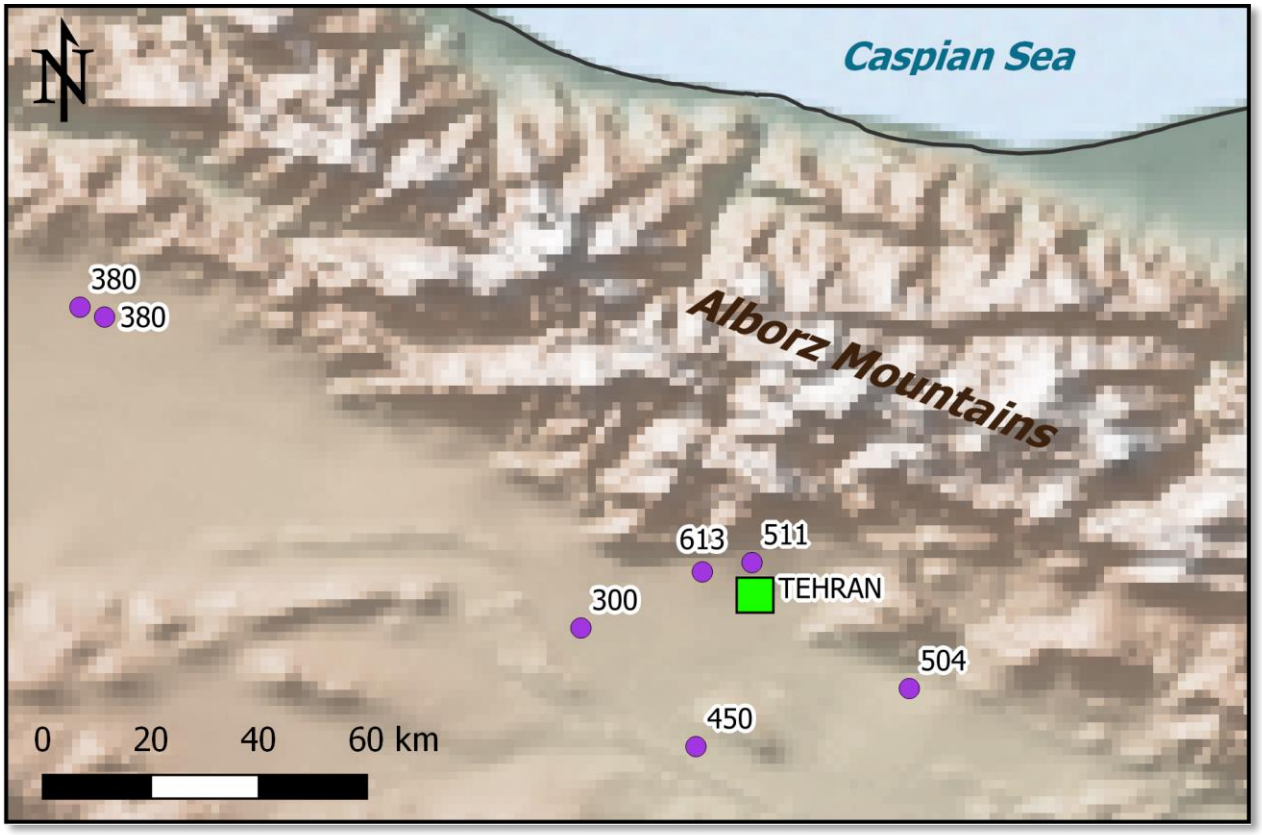

Figure 5.6. Selected sites in the Tehran basin and their reported V_{S30} values.

5.5 INFERENCE OF V_{S30} FOR IRAN SITES AND RECOMMENDED NEXT STEPS

It was required to implement Approach III for V_{S30} estimation, which is described in Section 2.3.3. For each of the ten geologic groups listed in Table 5.2, comparable groups from other regions, specifically California (CA) and the Pacific Northwest (PNW) region of North America, were identified (Wills et al. 2015 and Ahdi et al. 2017a, respectively). Those group associations are provided in Table 5.4 along with recommended V_{S30} moments for the Iranian groups. For assignment to SMSs, the standard deviation values reported in the "Preferred $\sigma_{\ln V}$" column of Table 5.4 should be increased by σ_{ep} = 0.2 (as described in Section 2.3.2) to reflect epistemic uncertainties in borrowing $\mu_{\ln V}$ values from a source region for use in a target region (i.e., Iran) without validation. Similarly, the use of California $\mu_{\ln V}$ values within geomorphic terrain

categories as provided by Yong (2016) was recommended; see Figure 5.4. As before, an increased standard deviation is recommended when using California values for Iranian sites. Table 5.5 lists the recommended $\mu_{\ln V}$ and $\sigma_{\ln V}$ values for the terrain proxy-based V_{S30} prediction model.

Table 5.4. Summary of recommended Iranian geologic group moments and comparable groups from other regions used as the basis for estimation. Preferred category mean and standard deviation values indicated for each Iranian group.

Recommended Group Moments, Iran geology				Comparable groups from other regions			
Group		Pref. $\mu_{\ln V}$ (m/sec)	Pref. $\sigma_{\ln V}$	Name	$\mu_{\ln V}$ (m/sec)	$\sigma_{\ln V}$	Region
1	Qy	277	0.25	Qal1	223	0.21	CA
				Qal2	285	0.25	
				Qal3	335	0.31	
2	Qe	352	0.40	beach, bar, dune deposits	334	0.44	PNW
				loess	372	0.37	
3	Qc	215	0.17	Qal1	223	0.21	CA
				Qal_deep Imp. Valley	207	0.14	
4	Qt	360	0.36	Qoa	360	0.36	CA
5	Qu	315	0.25	Qal1	223	0.21	CA
				Qal2	285	0.25	
				Qal3	335	0.31	
				Qoa	360	0.36	
				Qs	306	0.11	
				QT	418	0.35	
6	Tc	427	0.43	Tss	427	0.43	CA
7	Tf	365	0.33	Tsh	365	0.33	CA
8	Ms	552	0.43	Kss	458	0.43	CA
				KJf	665	0.44	
9	V	560	0.46	volc	635	0.66	PNW
				Tv	493	0.32	CA
10	Xln	621	0.52	Xtaline	621	0.52	CA

Table 5.5. Summary of recommendations of V_{S30} moments for terrain classifications for Iran (after Yong 2016).

IP07 Terrain Category	Preferred $\mu_{\ln V}$ (m/s)	Preferred $\sigma_{\ln V}$
1	519	0.38
2	586	0.16
3	517	0.38
4	568	0.46
5	425	0.37
6	448	0.14
7	429	0.38
8	382	0.32
9	353	0.16
10	348	0.09
11	392	0.48
12	281	0.20
13	*NA*	*NA*
14	236	0.14
15	460	0.52
16	225	0.20

Per Tables 5.4 and 5.5, the preferred V_{S30} values for each site in Appendix B are taken as the geometric mean of the values for the site's geology group and terrain class, respectively. Similarly, the preferred $\sigma_{\ln V}$ values are computed from geometric means, and an epistemic uncertainty $\sigma_{ep} = 0.2$ is presented as well.

Recommendations for improvement of V_{S30} assignments include:

1. Carrying out a limited campaign to determine V_S profiles using geophysical methods and operators with a proven record of reliability, such as those described in Section 3.3.1, preferably at some of the same locations as those described in Section 5.2.1.
2. Verification that existing results are reasonably consistent with the validation dataset from Recommendation 1, or confirmation that they are not. If not: (1) more reliable data should be compiled from other sources (including from consulting reports), as

applicable, and (2) new in situ V_S profiling should be carried out. Consideration of horizontal-to-vertical spectral ratio (HVSR) as part of a new effort of site characterization should be given.

3. Assignment of a profile to a site is not recommended unless it is truly proximate or represents a condition that strongly correlates to V_S (e.g., a rock outcrop) and is confirmed by a field visit by a geologist.
4. Proxies should be developed following Approach I standards, such as those described in this section, but with more reliable V_S data, to estimate V_{S30} for sites without measurements.

6 A United States Community Shear-Wave Velocity Profile Database

6.1 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):

$$G = \rho \cdot V_S^{\,2} \tag{6.1}$$

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 [2003] 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 geological 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 had the foresight to integrate 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 as an augment to 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 6.1 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.

Table 6.1. 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.

6.2 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 6.3. The locations of sites in California that are currently represented in the PDB are shown in Figure 6.1.

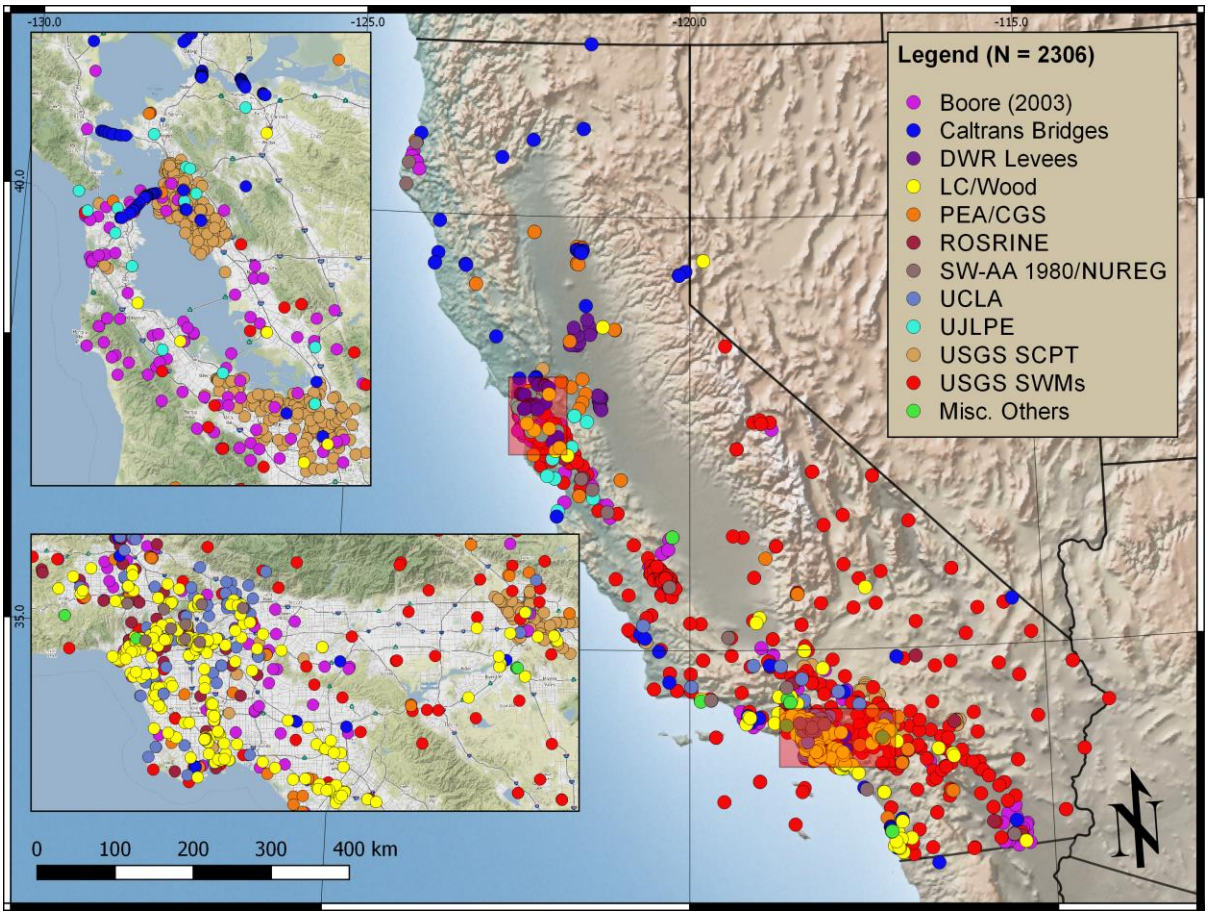

Figure 6.1. Map of California with major V_S PDB data sources (descriptions and definitions of abbreviations in Section 6.3).

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.

6.2.1 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.

6.2.2 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, Kobayashi 1959) 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 6.2 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 6.3) 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 6.2b is the significant fraction (30%) of non-invasive V_S data from both surface- and body-wave methods, which are becoming increasingly acquired and used in the field of engineering seismology (e.g., COSMOS, 201x).

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.

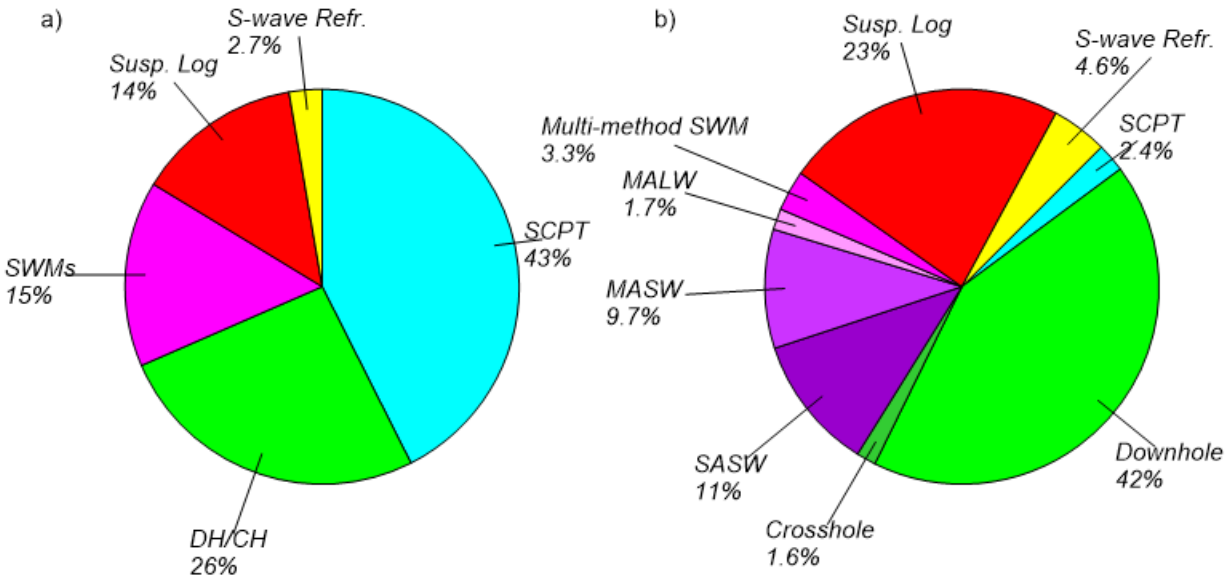

Figure 6.2. Distribution of V_S profiles included in database to date by geophysical method, both including (a) and excluding (b) the USGS CPT dataset (see Section 6.3). SWMs in Figure 1a are broken down into separate methods in Figure 1b.

A second method given lower priority for inclusion in the database at this time is Refraction Microtremor (ReMi^{TM}), 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^{TM}, such

as those described in Yong et al. 2013, the potential inclusion of this data in the V_S PDB will be reconsidered.

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.

6.2.3 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 using Equations (2.1) and (2.2), by dividing z_p by the total travel time through all layers of a V_S profile, as described in Section 2.2.2. The computation of V_{S30} requires substitution of z_p with 30 m in Equations (2.1) and (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. (2017a). 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.

Figures 6.3a and 6.3b 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. 2017a). 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 2003) which made concerted efforts to also collect V_S data at stiffer sites.

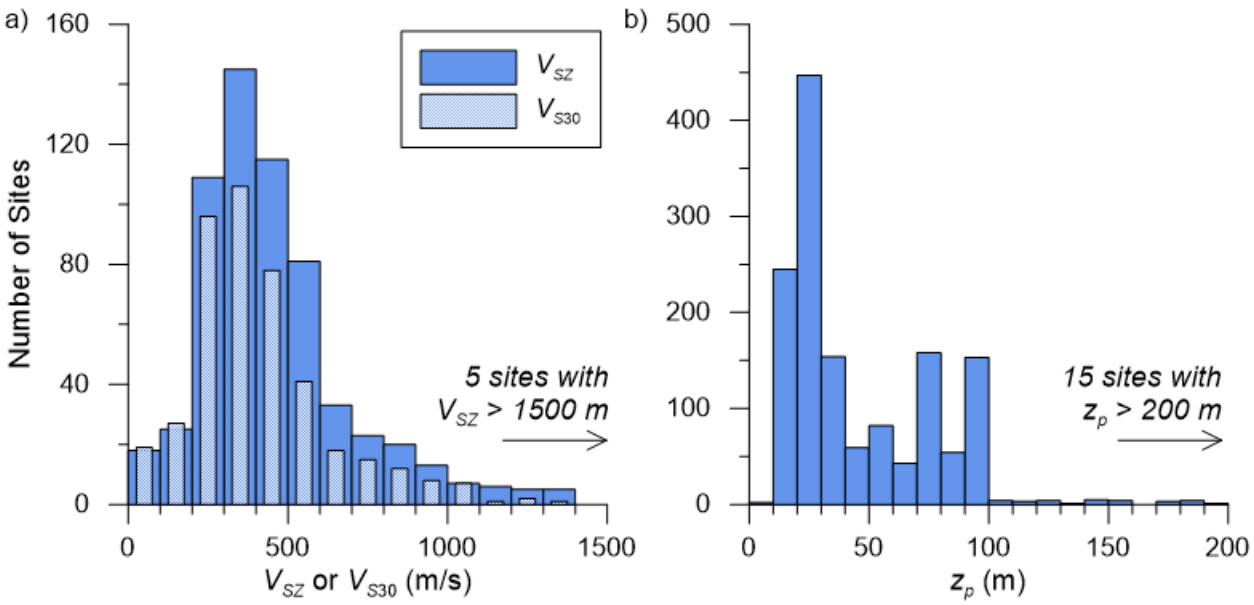

Figure 6.3. Distributions of (a) maximum V_S profile depth (z_p) and (b) time-averaged V_S to z_p (V_{SZ}) and to 30 m depth (V_{S30}) in the V_S PDB at time of writing.

6.3 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.

6.3.1 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 6.3.5.1 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.

6.3.2 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 6.3.3 below).

6.3.2.1 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 2003) 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), 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 (2003) 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 6.4 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

Figure 6.4. 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.

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.

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 6.5 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 6.6. 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).

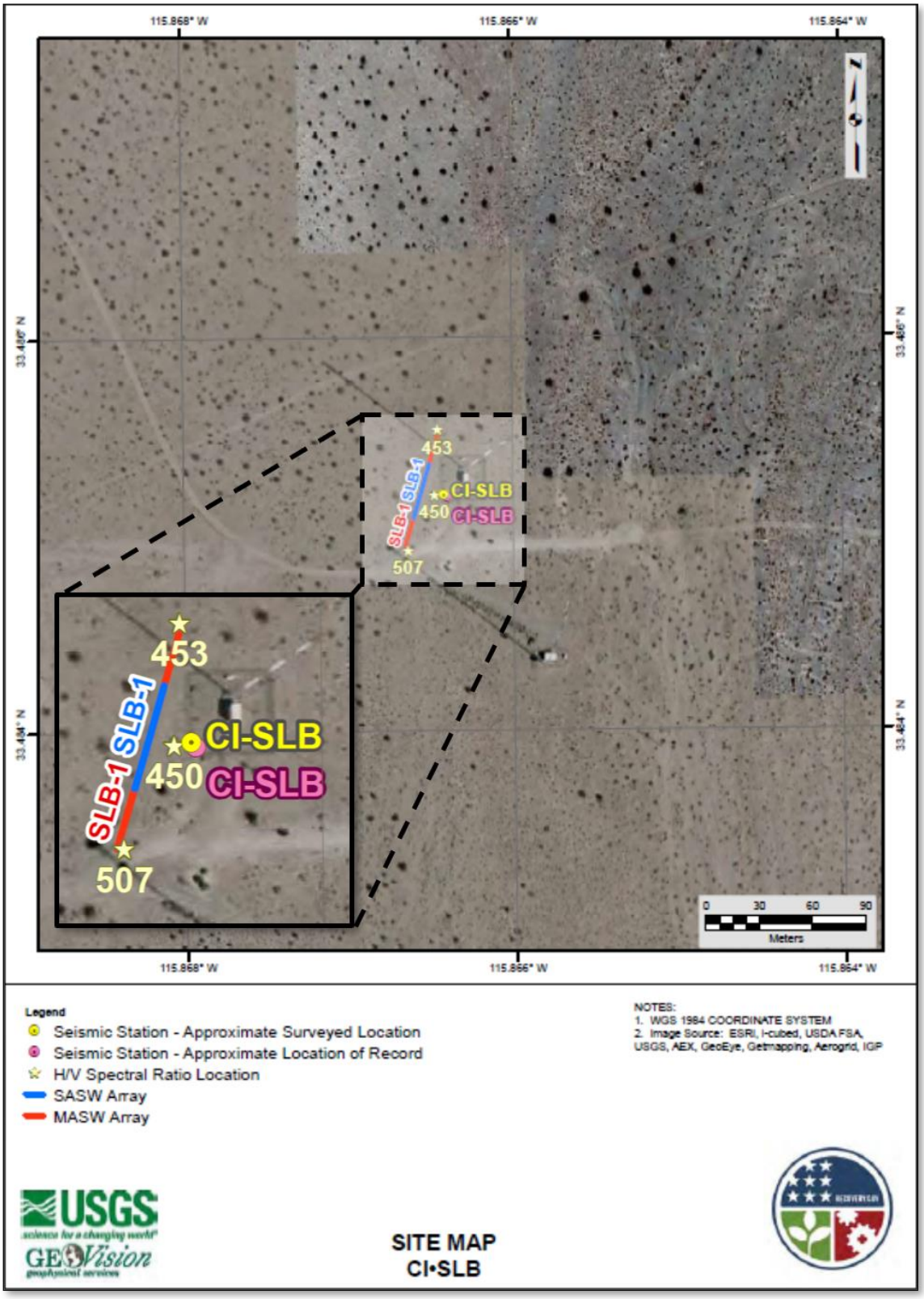

Figure 6.5. Site Map for site CI.SLB from Appendix of Yong et al. (2013). Area of zoom detail displays locations of collinear SWM testing arrays for SASW and MASW methods.

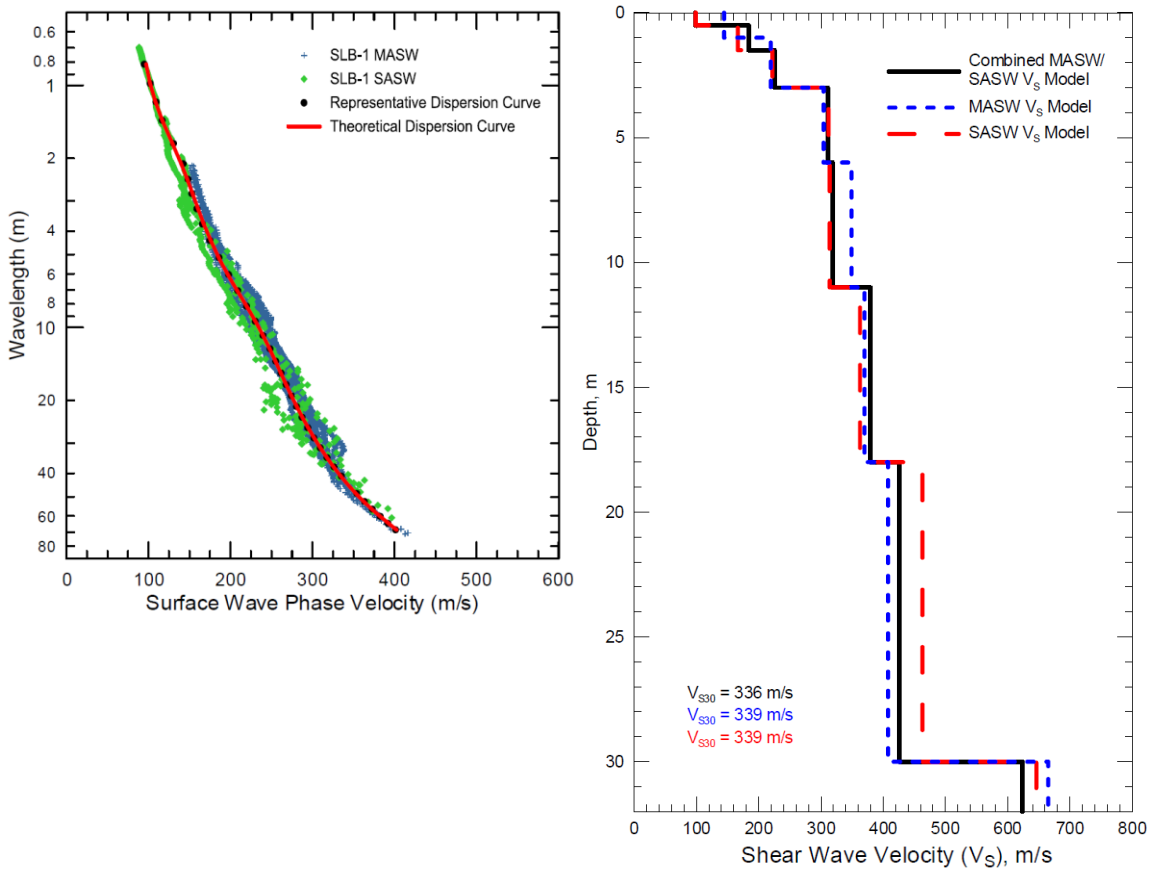

Figure 6.6. Dispersion curves (left) and inverted V_S profiles (right) for site CI.SLB from the Appendix of Yong et al. (2013). Two collinear SWM arrays are deployed at the same site, resulting in three potential modeling routines based on dispersion data from either or both MASW or SASW.

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 6.5 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 6.7 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.

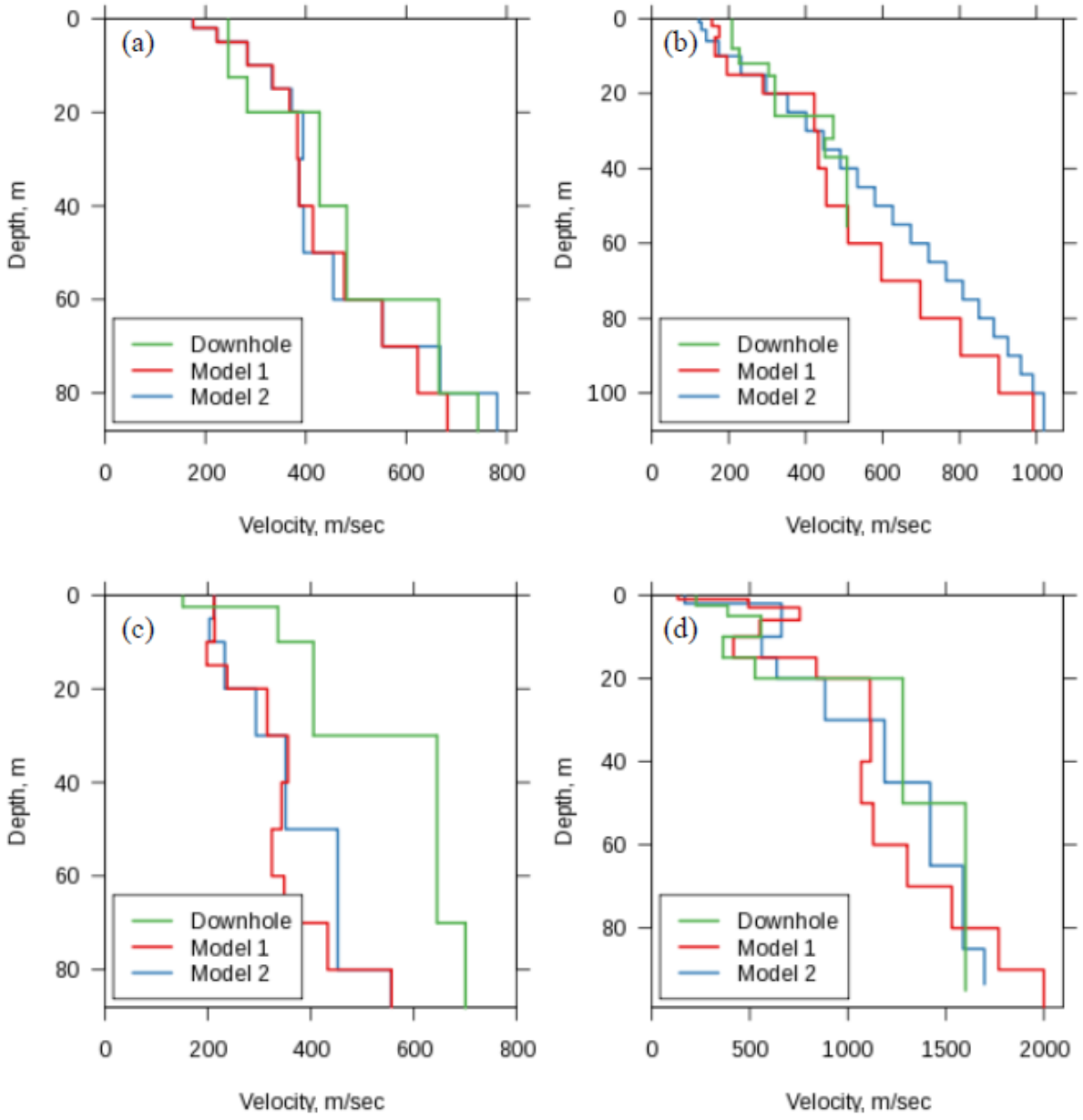

Figure 6.7. (from Thompson et al. 2010)**.** Comparison of V_S profiles derived from SASW and downhole methods from two different USGS open-file reports at four sites in central California (first site code is from Thompson et al. 2010, second site name is from previous USGS downhole studies summarized by Boore [2003]): (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.

6.3.2.2 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. Other sites of interest in high seismic hazard regions of the U.S. were also considered. 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 Tables 1 and 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.

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 6.3.4.1 below) or data from LeRoy Crandall & Associates (see Section 6.3.5 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.

6.3.3 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 Jonathan Stewart had amassed a personal data collection (2015, *pers. comm.*). 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.

6.3.3.1 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.

6.3.3.2 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.

6.3.3.3 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 6.3.5 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.

6.3.3.4 California Geological Survey – Compilation of Reviewed Sites

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 6.3.5 below.

6.3.4 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.

6.3.4.1 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 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.

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.
- As part 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.

6.3.4.2 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.*).

6.3.4.3 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.*).

6.3.4.4 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.

6.3.4.5 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 soils 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).

6.3.5 Industry Data Sources

Throughout Section 6.3, "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.

6.3.5.1 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 geotechnical design 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 6.8. 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.

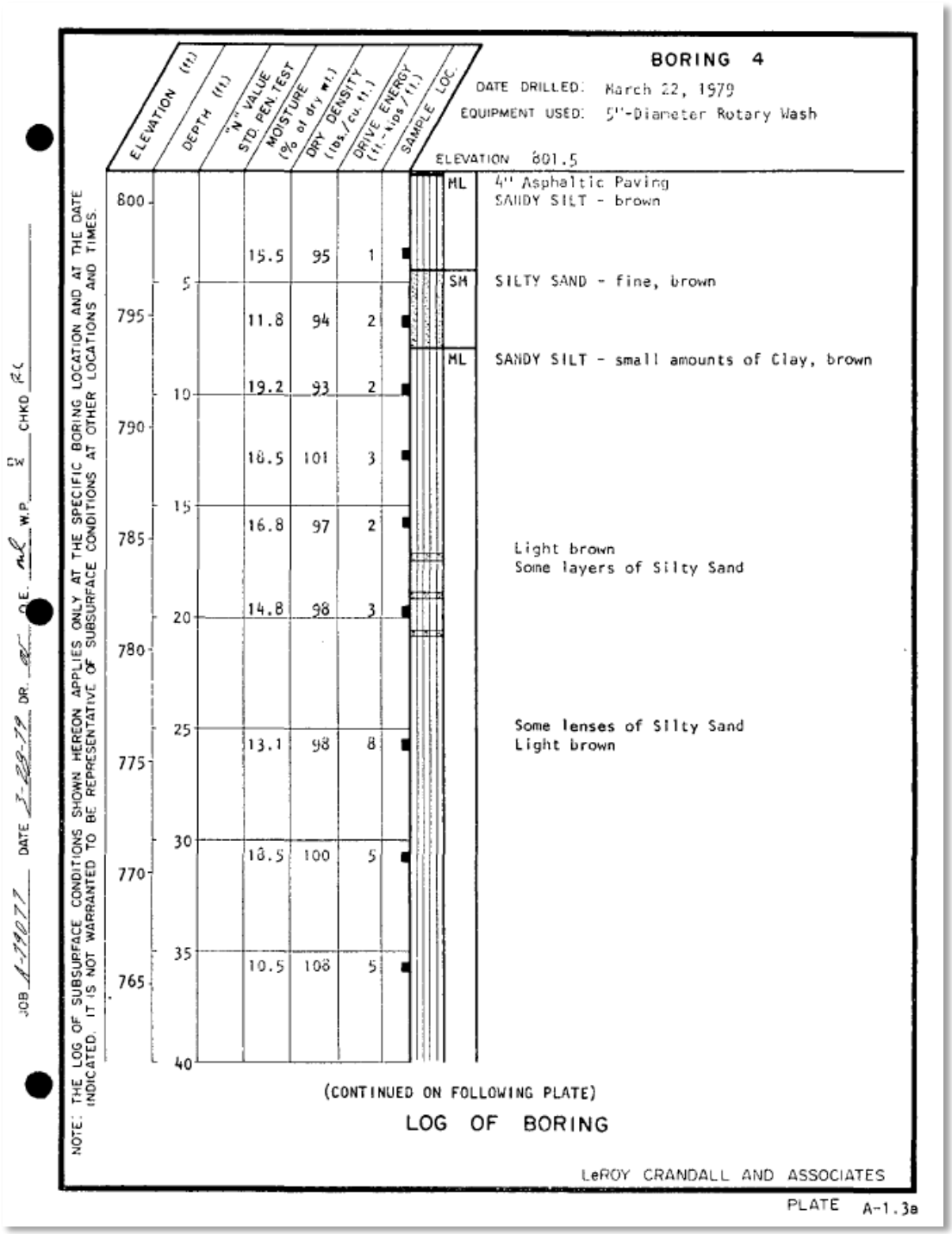

Figure 6.8. Example excerpt of boring log from L/C Project No. 79077.

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 6.3.1 above).

6.3.5.2 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 6.3.3.4 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:

- 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 6.3.4.2
- 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.

6.3.5.3 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.

6.3.5.4 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).

6.4 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 (in Chapters 3 and 4, the term "PDB" was used to remain consistent with terminology in the original publications). 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.

6.4.1 Overview of Data Preparation and Processing Methods

Previously, Section 6.3 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 6.4.3 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).

6.4.2 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 6.2, which are grouped into three categories: general information, geophysical data, and geotechnical data. The table names in Table 6.2 have meanings and are described subsequently in this section.

Table 6.2. 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 6.9 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 6.2 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.

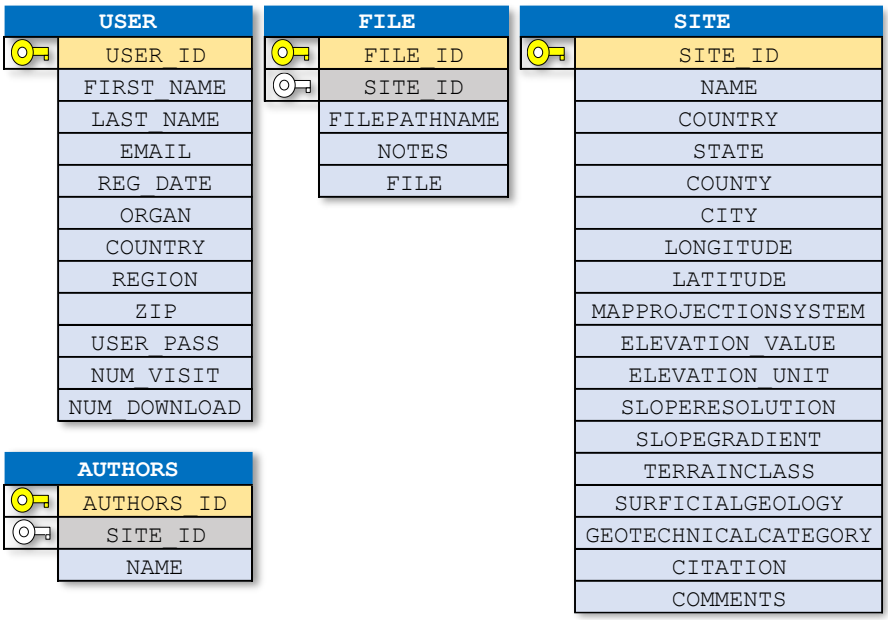

Figure 6.9. RDB tables containing general data.

Figures 6.10 and 6.11 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, HVSR data, and travel time plots for SCPT measurements. 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.

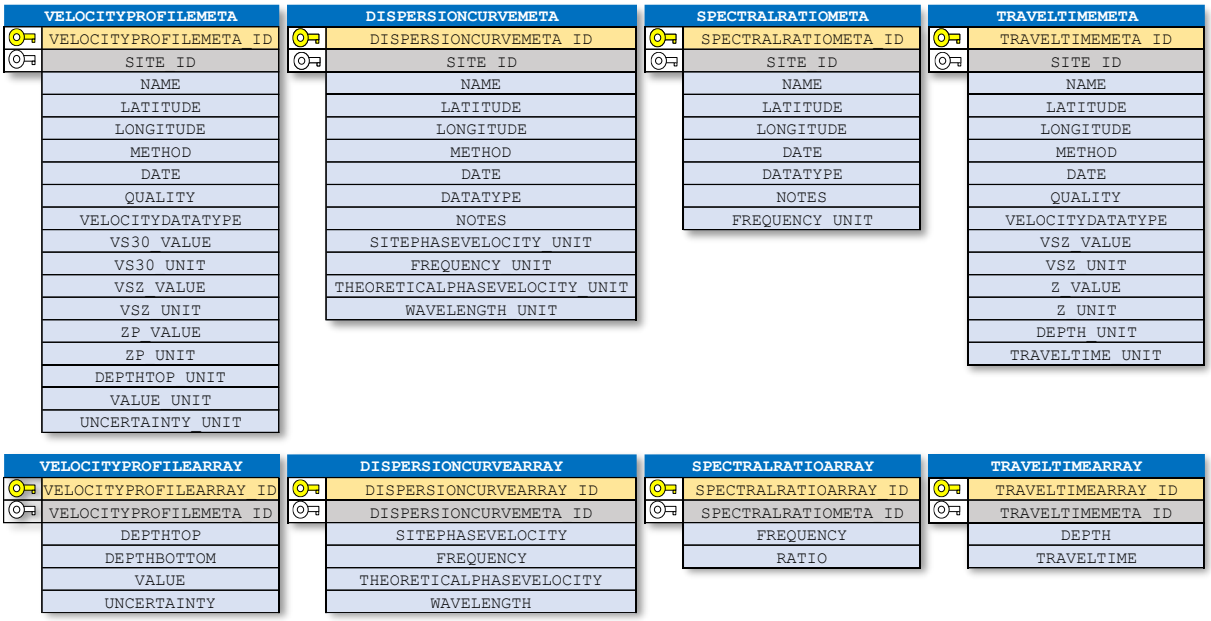

Figure 6.10. RDB tables containing geophysical data.

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).

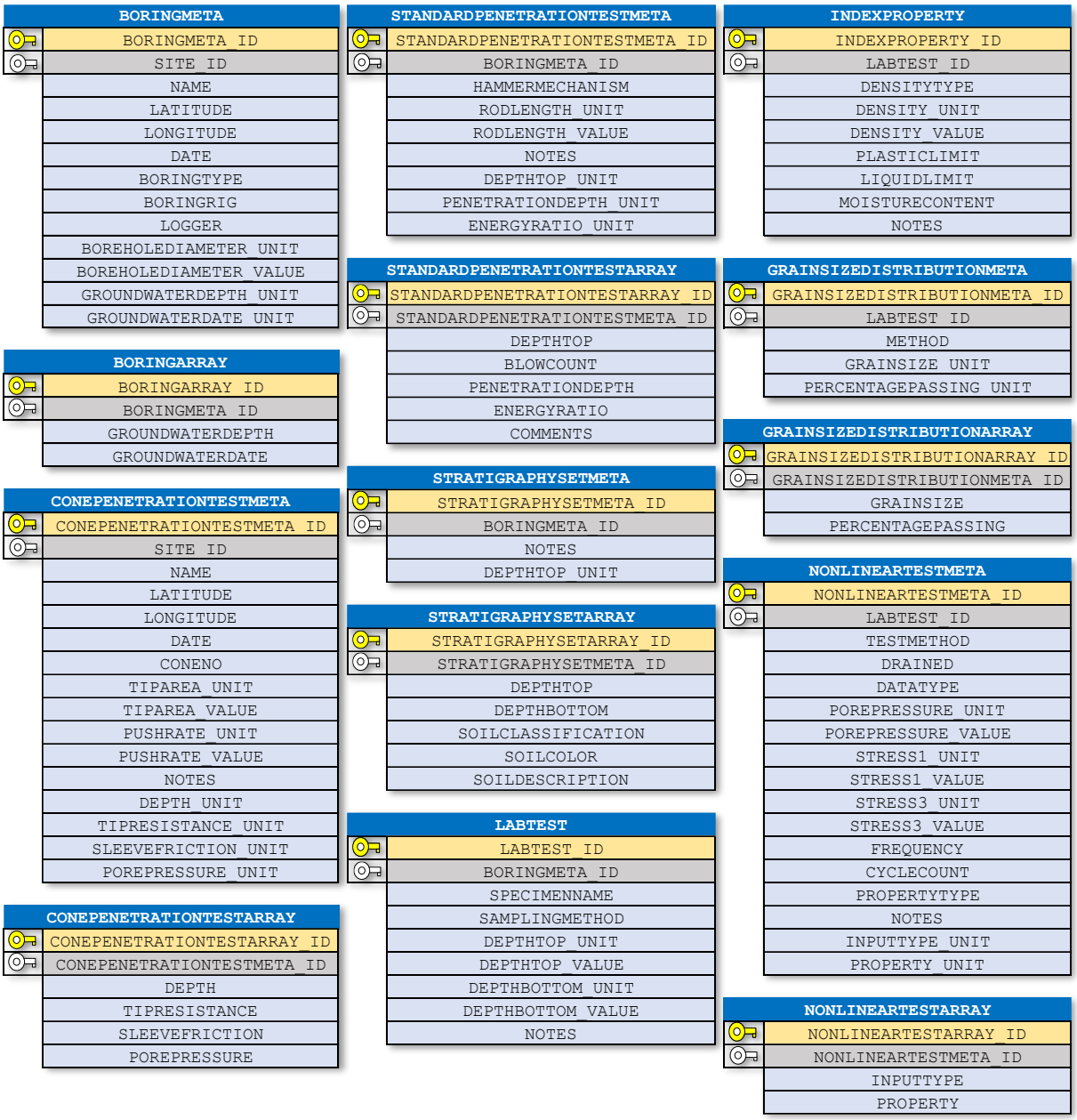

Figure 6.11. RDB tables containing geotechnical data.

6.4.3 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 6.12 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.

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 6.13 shows an example populated JSON file that is the output from the Unify screenshot in Figure 6.12. 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.

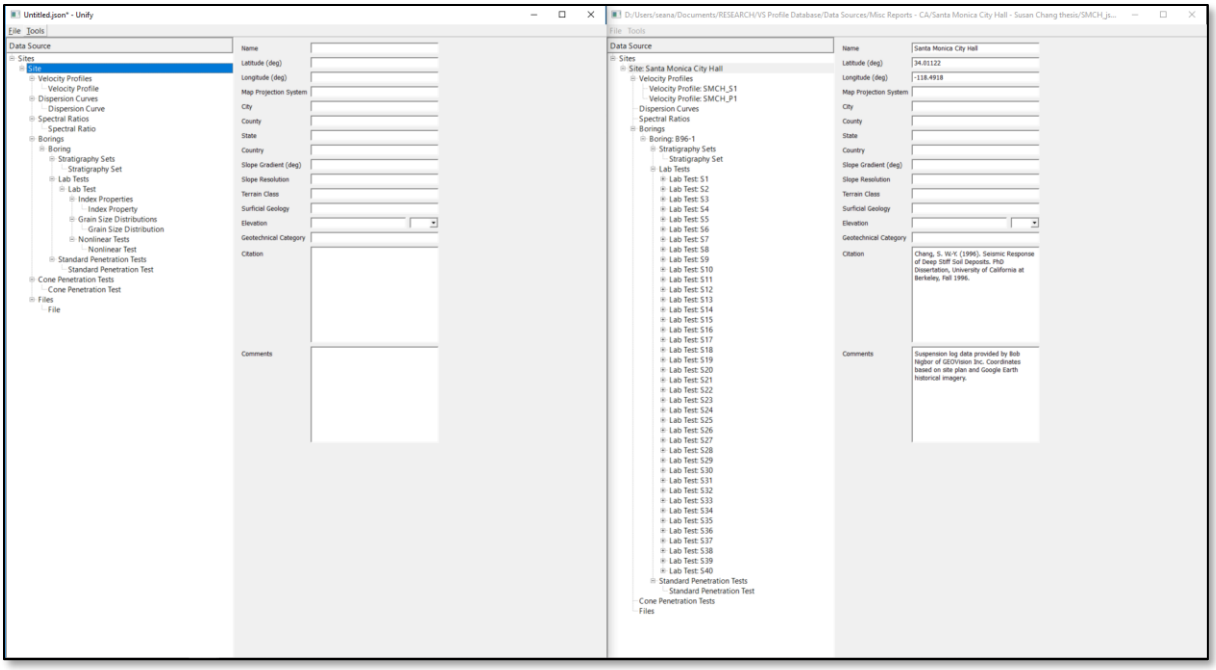

Figure 6.12. Screenshots of Unify GUI with example dummy unpopulated site (left) and complete example populated with data from Santa Monica City Hall site (right).

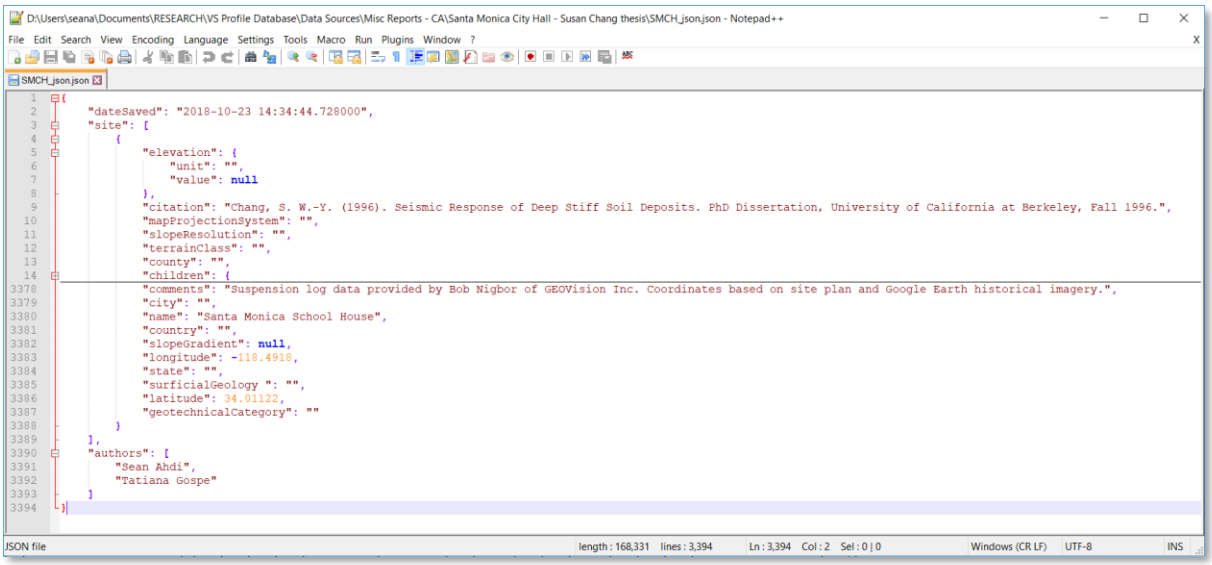

Figure 6.13. Screenshot of hierarchical data storage in JSON file format for Santa Monica City Hall example site (opened in Notepad++ text editor).

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.

6.4.4 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 6.4.2. 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.

6.4.5 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 mirror the V_S PDB onto servers at the Texas Advanced Computing Center (TACC), which provides a convenient backup. Mirroring onto DesignSafe also allows for improved user interaction with the database, a topic discussed in Section 6.5 below.

6.5 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 6.14 shows a screenshot of the V_S PDB website's map interface, zoomed into the Quintara site described in Section 6.3.2.1. 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.

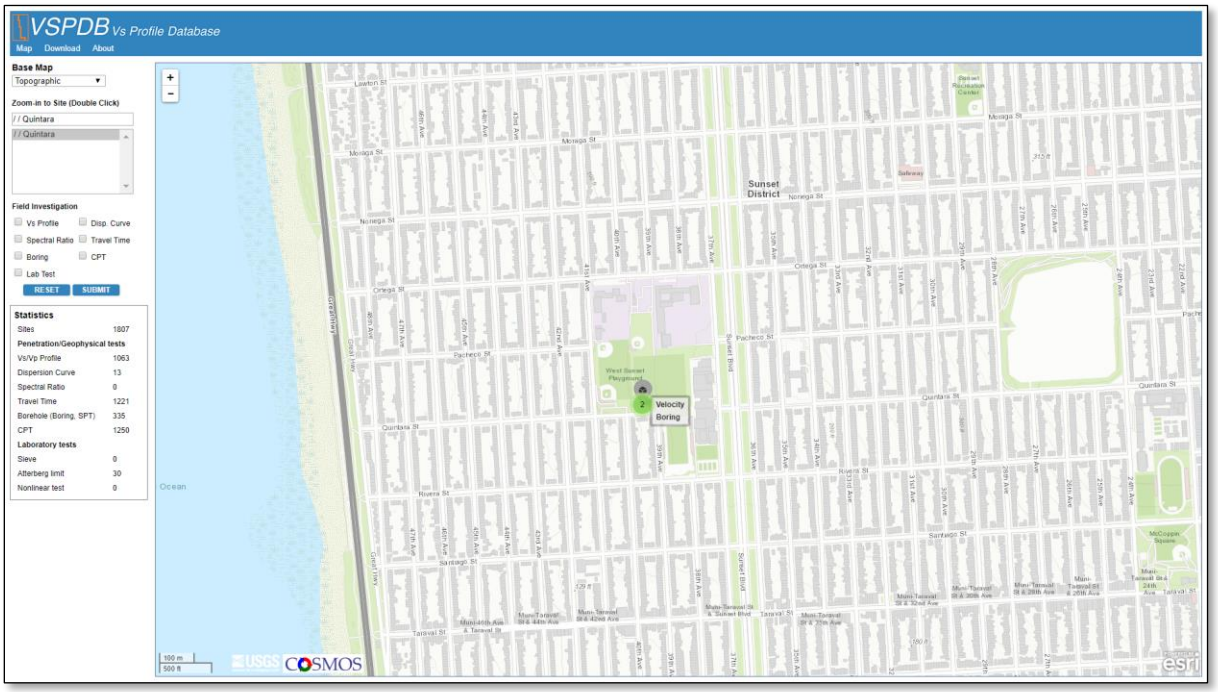

Figure 6.14. Screenshot V_S PDB website, zoomed into Quintara site in San Francisco.

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 6.15 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.

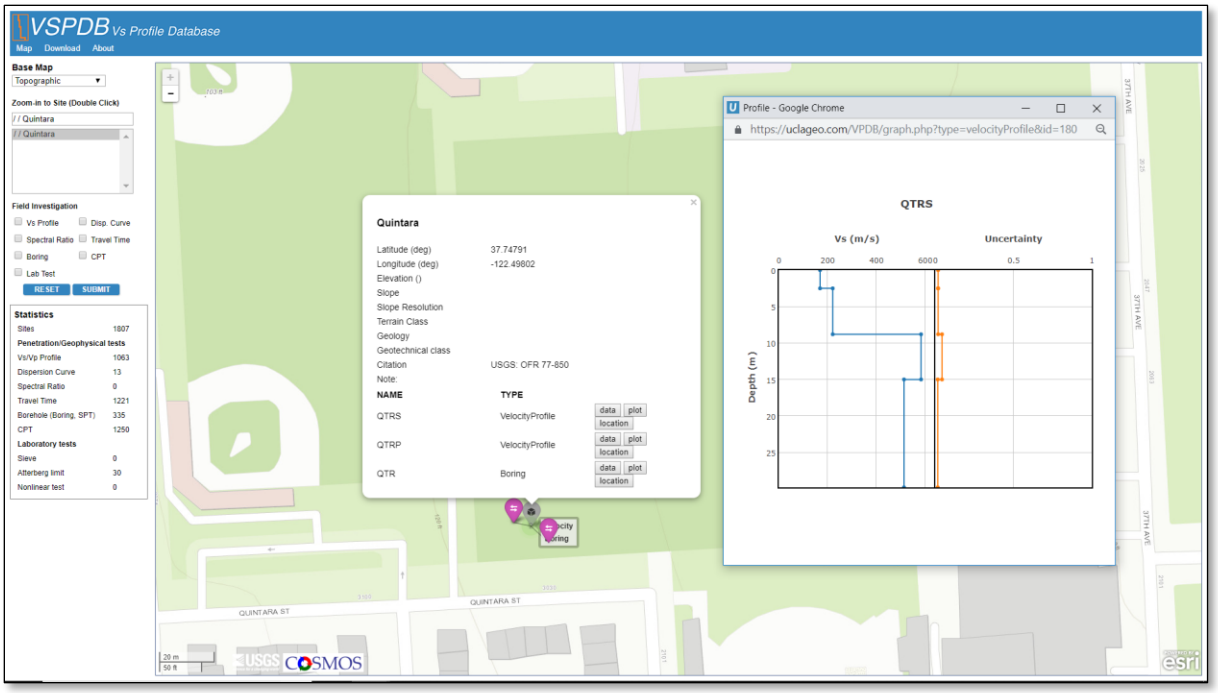

Figure 6.15. Screenshot V_S PDB website, zoomed into Quintara site with metadata pop-up balloon and example of V_S profile data and associated uncertainty from downhole measurements.

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 6.16 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.

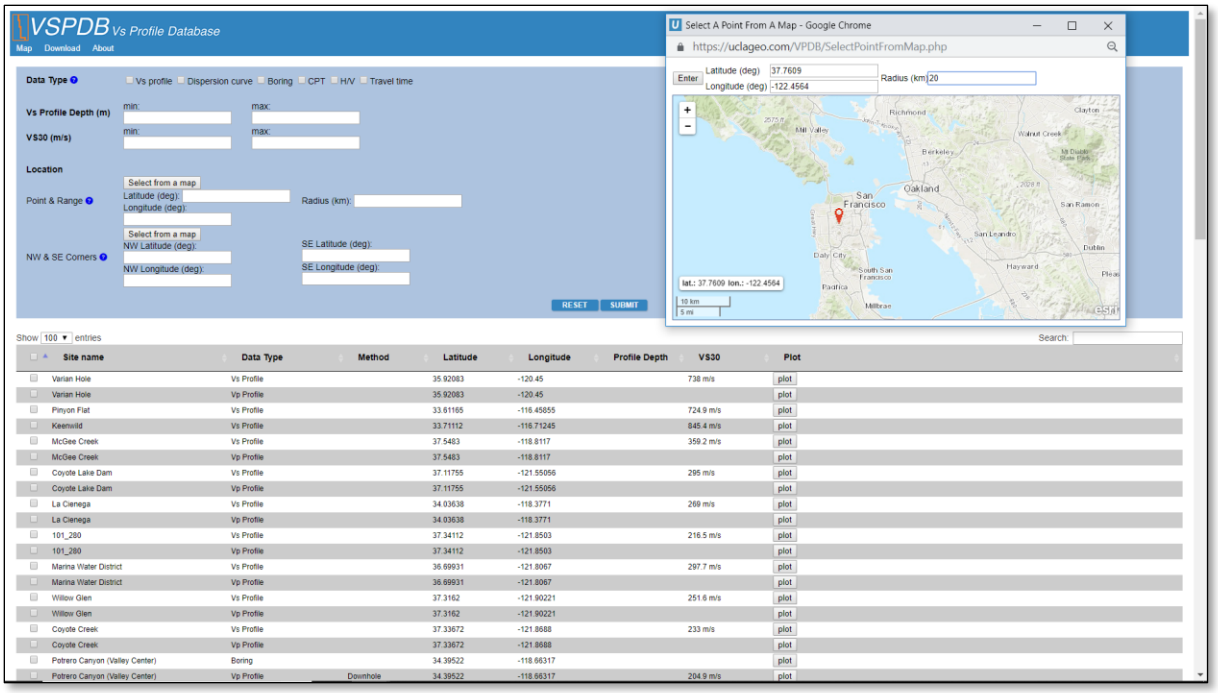

Figure 6.16. Screenshot V_S PDB website's data Download page, with Point & Range feature engaged for user query.

6.6 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.

7 Summary, Conclusions, and Future Work

7.1 SCOPE AND RESULTS

In this dissertation, seismic site characterization was studied from the perspective of utilizing in situ measured V_S data where available, and proxy-based modeling of V_S-based site parameters where no in situ measurements exist. The NGA-Subduction ground motion model (GMM) development effort necessitated the assignment of site parameters to strong motion recording sites at seven subduction zones worldwide. The time-averaged shear-wave velocity in the upper 30 m of the crust (V_{S30}) was selected to represent site conditions based on precedent work that has generally found it to be effective as a first-order site parameter, its frequent use in past large GMM development projects such as NGA-West2, and availability of numerous correlations for its estimation when measured data is unavailable.

In the NGA-Sub site database (SDB), which was discussed in Chapter 2, 6433 strong motion sites were assigned V_{S30} values from V_S measurements where available, and from an array of V_{S30} prediction models otherwise. Only 39% of sites in the SDB can be assigned a measured V_S profile from in situ geophysical testing in reasonable proximity to the strong motion station. Proxies used in the prediction models often included surficial geology units that were grouped based on geological interpretation and were conditioned on topographic slope and elevation in some cases. Additional proxies that were used include geomorphic terrain categories and other

region-specific proxies, such as the JEGM combined geomorphology and geology maps available across Japan. Many of the V_{S30} prediction models used in NGA-Sub were newly developed as part of the project. A tripartite framework for the development of proxy-based models was formulated, whereby the mean and standard deviations of V_{S30} are assigned with variable levels of confidence and uncertainty, depending on the availability of measured V_S data and local, large-scale geologic maps. The approaches in this framework range from models developed with abundant V_S measurements and good quality geological maps or other proxies (Approach I) to guidelines for borrowing models from other regions when a study region or particular geologic category lacks V_S data (Approach III); an intermediate approach (II) allows for a target region to utilize sparse V_S data to validate borrowing of a model from a host region. Epistemic uncertainties associated with the use of approaches involving relatively limited data are also defined.

Chapter 3 discussed the development of robust V_{S30} prediction models following Approach I based on (1) a hybrid slope/geology proxy, with 18 geologically informed categories, six of which were dependent on topographic slope, and (2) 13 of 16 geomorphic terrain categories (per Iwahashi and Pike 2007), similar to past work done in California (Yong 2016). A dataset of 928 measured V_S profiles was collected in Oregon, Washington, and British Columbia and analyzed for V_{S30} trends within the aforementioned proxy groups. Residuals analyses were performed to check for biases with respect to glaciation and basin effects; the former was found and identified for four geologic categories, and the latter was not found to be statistically significant. Because of a relatively strong correlation (ρ = 0.78) among the residuals of the geology/slope and terrain proxies, but an overall lower dispersion of residuals for the hybrid model (0.35) compared to the terrain model (0.45), the recommendation was made to the NGA-Sub working group that the PNW-specific hybrid model should be used to assign V_{S30} and its uncertainty wherever geological

maps are available, and the PNW-specific terrain model should be used when high-quality geologic information at large scales was not available.

Chapter 4 discussed the development of an Alaska-specific V_{S30} prediction model, which used the geological categories for the PNW model and various amounts of Alaska-based measured V_{S30} data (126 profiles total). This model development was a mixture of Approaches I, II, and III, where (1) an Alaska-specific V_{S30} group for alluvium was developed with 45 measurements (Approach I); (2) combined Alaska and PNW group moments were applied for three geology groups that had similar data populations in both regions without appreciable inter-region bias (combination of Approach I and II); (3) PNW group moments were applied for two groups with fewer V_{S30} measurements in Alaska, and bias was deemed insignificant for Alaska data (Approach II); and (4) group moments were applied from the PNW model of Chapter 3 and a California model (Wills et al. 2015) for Alaska groups with no measured V_{S30} values (Approach III). Epistemic uncertainties of 0.2 (natural log units) are assigned for mean V_{S30} values assigned using Approach III.

Chapter 5 discussed a similar study performed in Iran. In this case, the strong motion database consists of 1135 sites, 558 of which were assigned co-located V_S profiles for V_{S30} assignments. Those profiles were initially used to develop geology- and terrain-based V_{S30} prediction models following Approach I for application to the remaining sites without co-located V_S measurements. However, it was decided that Approach I prediction models could not be developed for the Iranian database, because of lack of confidence in the V_{S30} data, which stems in part from (1) the use of relatively-coarse seismic refraction methods lacking detail in the upper 30 m and (2) the lack of appreciable differences between rock and soil categories. Instead, Approach III was adopted, applying averages of both the PNW and California hybrid geology/slope proxy

and terrain proxy-based V_{S30} prediction models with equal weights to the Iranian sites. Recommendations moving forward were to perform new V_S profile measurements with higher granularity in the near-surface (e.g. MASW) and to acquire HVSR measurements, and check to see if the existing Iranian V_S profile dataset are reasonably consistent with these new measurements. If so, these supplemental measurements can warrant implementing Approach I for developing a region-specific V_{S30} proxy model in Iran.

Finally, Chapter 6 focused on the development of the United States Community V_S Profile Database (PBD), 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. This project has implemented the computer science definition of a "database", utilizing a relational database with tables of data that interact with each other using keys that are formally described using a schema, an approach that has not commonly been taken in earthquake engineering research or practice. Data was converted to a standardized JavaScript Object Notation (JSON) file format that allowed for lightweight handling and storage of the varied data types at 1800 V_S profile sites currently in the database, with ongoing digitization and integration into the database expecting to grow the project to over 5000 V_S profiles (discussed in Section 7.2.1 below). 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, with numerous future research directions already imminently evident and described in Section 7.2 below.

In conclusion, I hope that my work leads to more and stronger bridges between the research and practice communities in the earth sciences and in geotechnical and earthquake engineering and ground motion modeling. Physical processes that are quantified using seismic site characterization information, and the methods used to obtain such information, are better informed by incorporation of high-quality geological data. Approaches developed as part of this dissertation may be applied by researchers worldwide who either (a) seek to develop their own proxy-based V_{S30} models, rather than borrowing such models from other regions, or (b) need to borrow models when site characterization data is lacking across their study area or within specific geological environments. I want to emphasize that this work is *not* meant to undermine the many advances in the field of near-surface geophysical site characterization, and accordingly, *there is indeed no substitute for making high-quality in situ measurements of* V_S. Any project of importance should budget for such measurements and provide an estimate of their uncertainties.

7.2 FUTURE WORK

Future work relates both to NGA-Sub and the V_S PDB. For the former, it is anticipated that a future NGA-Sub2 project will commence, which will additionally focus on data collection and GMM development efforts for SZs in Indonesia, the Calabrian arc in southern Italy, and the subduction zone in the Aegean Sea south of Greece. Work will be required to develop proxy-based V_{S30} models for Indonesia and to build upon existing models for Italy (e.g. Scassera et al. 2009) and Greece (Stewart et al. 2014), incorporating new datasets by researchers in the latter two countries such as the Engineering Strong Motion Database in Europe (Luzi et al. 2016), maintained by the *Istituto Nazionale di Geofisica e Vulcanologia* (National Institute of Geophysics and Volcanology, Italy) and under the auspices of the Observatories & Research Facilities for European Seismology (ORFEUS) and the Seismology and Earthquake Engineering Research

Infrastructure Alliance for Europe, SERA).

Future work pertaining to the V_S PDB can be split into two categories: continuing data updates (Section 7.2.1), and applications of the V_S PDB (Section 7.2.2).

7.2.1 Continuing Updates to the V_S Profile Database

As described in Chapter 6, the V_S PDB is designed and expected to be a living database, not a static repository of data. The project leadership does 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. Plans for future data sources that may be included after the initial efforts described in Chapter 6 are discussed in the subsections below. One major note is that collaboration with the USGS V_{S30} Compilation project team (described in Section 6.3.2) has resulted in an agreement for their continuing to share measured V_S profile data into the future, providing the V_S PDB with another large potential source of data for years to come. Future potential data sources are described in the following subsections.

Besides including additional data, improvements can be made to the web user interface of the V_S PDB. In the future, Python scripts can be 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 RDB 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.

7.2.1.1 Site Databases of NGA Projects

Across the three major regions of NGA projects (NGA-West2, NGA-East, and NGA-Sub), there are 7262 sites with measured V_{S30} values from around the world; ostensibly these are derived from in situ measured V_S profiles. Within the U.S., there are 441 profiles in California (Seyhan et al. 2014), 1507 profiles in the central and eastern U.S., defined as CENA minus Canada (Parker et al. 2017), 608 profiles in the Pacific Northwest (i.e. Oregon and Washington, and excluding Canada; Ahdi et al. 2017a), and 90 profiles in Alaska (Ahdi et al. 2017b). Currently it is assumed that the NGA-West2 V_S measurements are fully encompassed in the aforementioned data sources in this report, but a more detailed investigation to cross-check all sites across both databases should be performed as part of future quality checks. For NGA-East, only 84 of the 1507 profiles were

Table 7.1. Summary of strong motion stations and available V_S data from NGA Projects

NGA Project	Total Sites in Each SDB	Sites with Measured V_{S30}	Measured V_S Sites in U.S.
NGA-West 2	4149	2014	441
NGA-East	1390	2754*	1507
NGA-Sub	6433	2494	698
Totals	11,972	7262	2646

*Number of V_S *profiles* in profile database used to develop V_{S30} proxies; number of measured V_{S30} assigned to stations is 84.

assigned to strong motion stations; the remainder were included in Parker et al. (2017) as part of their development of a hybrid geology-slope V_{S30} prediction model for assignment to stations that lacked measurements. While the remainder of the V_S profiles were not presented by Parker et al. (2017) because they were taken from proprietary sources, references to their original source documents are provided, and the project team plans to pursue those profiles in the future pursuant to the protocols for inclusion of private data as descried in Section 6.3.5. As the NGA projects

unfold and potentially advance into future iterations, we will re-connect with NGA-East project leadership to share V_S profile data as appropriate. NGA project station and V_S data is summarized in Table 7.1.

7.2.1.2 Outlook for Inclusion of Intermountain West Data

The Intermountain West (IMW) is defined as a 200,000 km^2 region of the western U.S. that encompasses zones of deformed crust in the North American tectonic plate, from the eastern edge of the Sierra Nevada to the eastern front of the Rocky Mountains and having moderate seismicity throughout this zone (Smith & Sbar 1974, Haller et al. 2015). For the purposes of the V_S PDB, we develop our own definition for IMW, as related to the NGA project existing geographic regions, to include the states of Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming. It can be argued that the eastern portions of California, Oregon, and Washington (i.e., east of the Sierra Nevada and Cascade Mountains) should be included in the IMW. V_S data for these regions have been included in the site databases for NGA-West 2 (Seyhan et al. 2014) and NGA-Subduction (Ahdi et al. 2017a,b). V_S profile data for the IMW region has been assembled by Yong et al. (2016) for the USGS V_{S30} compilation, with up to 814 V_S profiles in Idaho, Nevada, Utah, and Wyoming available from at least four publications. We anticipate that these and other data sources from will eventually be incorporated into the V_S PDB, as described in Section 7.2.1.4.

7.2.1.3 Geometrics, Inc./OYO Corporation Dataset

The research division of the private geophysical firm Geometrics, Inc. focuses on development of novel surface wave methods, particularly the SPAC method and other methods using passive-source energy, and production of the SeisImager software for analyzing SWM dispersion curve data and inverting for V_S profiles. The V_S PDB project leadership held a meeting

with the chief Research Geophysicist of Geometrics, who presented the company's internal database of MAM data (K. Hayashi 2018, *pers. comm.*), available at http://seisimager.esy.es/index.htm. An agreement was made to share the file organization methodology and database schema structure pertaining to the V_S PDB project (described in Section 6.4) as the beginning of an agreement for Geometrics to share V_S profiles and associated data and metadata from their database. Currently there are hundreds of V_S profiles, primarily from active and passive SWMs, and HVSR data collected from around the world, with a subset of 125 measurement sites in the USA (K. Hayashi 2018, *pers. comm.*). The data are presented online with a Google Maps interface, and data can be visualized on the website or downloaded in an XML (eXtensible Markup Language) file format. We look forward to continuing collaboration with Geometrics, Inc. moving forward.

7.2.1.4 Other Potential Future Datasets

While it would be impossible to truly collect every single published study containing V_S measurements in the U.S., it is anticipated that the database will grow in popularity and use as time goes on, such that adding more data from obscure data sources and owners will become increasingly possible. Having said this, a plan exists to target a number of specific data sources/owners. A list of these include: EPRI reports for the CEUS; large personal research collections of faculty members from various universities, including the Georgia Institute of Technology (SCPT), University of California at Davis (SCPT), University of Texas at Austin and the Texas Seismological Network (TexNet, SWMs and invasive methods), University of Nevada at Reno (ReMi^{TM}) and at Las Vegas (SWMs); the California Strong Motion Instrumentation Program (CSMIP), a division of the CGS; infrastructure related to dams and other utilities owned by Southern California Edison and Pacific Gas & Electric; local/municipal jurisdictions that

oversee building construction, such as major cities in regions with high seismic hazard (e.g., Los Angeles, San Francisco, Portland, Seattle); state departments of transportation around the U.S. outside California, which are located in seismically active regions; the California Department of Water Resources; and utility companies who have vested interest in ensuring the seismic resiliency of their distributed infrastructure that conveys water, electricity, and natural gas.

Data collection will not be limited geographically, but rather sources that will have the highest yield of V_S profiles for easy inclusion will be pursued. It would also be unwise to rule out the potential future inclusion of international data, given that political boundaries do not define geological boundaries, and that there already exists access to numerous international data sources through NGA project working groups, particularly in Canada and other countries where seismic hazard is high, and potential European partners via the COSMOS Facilitation Committee that is developing international guidelines for non-invasive seismic site characterization (COSMOS 201x). We look forward to the potential growth of the V_S PDB in ways in which we may not be presently aware.

7.2.2 Applications of the V_S PDB

General applications of the V_S PDB include updates and improvements to existing proxy-based V_{S30} prediction models, such as those in California (Wills et al. 2015, Yong 2016), the PNW and Alaska regions described in Chapters 3 and 4 of this dissertation, and NGA-East (Parker et al. 2017). Improvements to these models can stem from implementing more measured V_{S30} data for increased statistical robustness or utilizing better proxies, such as updating the terrain class proxy by replacing the 16 Iwahashi and Pike (2007) terrain classes at ~1 km DEM resolution to the Iwahashi et al. (2018) 15-class schema at ~250 m DEM resolution. More specific applications are described in the following subsections.

7.2.2.1 Investigation of Non-Ergodic Site Response

Seismic site response is the modification of earthquake ground motions due to the effects of local site geological conditions. Site response should not be confused with ground response; *site response* is taken as the cumulative effects of all modes of site amplification (basin effects, surface waves, and shallow ground response), whereas *ground response* is the simulation-based response of relatively shallow soil layers (10s to 100s of meters) as computed using one-dimensional (1D) geotechnical wave propagation analyses (e.g., DEEPSOIL, Hashash et al. 2015).

When a GMM is used to estimate ground motion intensity measures, generally the sum of three terms is computed: a seismic source term parameterized on source characteristics such as moment magnitude, a wave propagation path term parameterized on site-to-source distance, and a site term parameterized on site characteristics such as V_{S30}. If developed empirically, the GMM site term predicts the average site response within the empirical database for a given value of V_{S30} and basin depth terms (e.g., $z_{1.0}$ or $z_{2.5}$). Site response determined in this way is considered as *ergodic* (Anderson and Brune, 1999), which for the present application means that particular features of the site that may cause its site response to deviate from the global average are not considered. Ergodic site response has the benefit of being easy to evaluate, but there are two drawbacks: (1) the mean response for a given site may be biased and (2) because site-to-site variations are not considered in the mean estimate, the associated within-event standard deviation used by the GMM is relatively large. The alternative is *non-ergodic* site response (Stewart et al. 2017), which considers the particular geologic conditions at a site controlling site response and should produce an unbiased estimate of median site response. Non-ergodic site response can be evaluated through the use of multiple earthquake recordings at a site of interest (which is unlikely for common design projects) or through simulation-based GRA, which requires a full V_S profile as

opposed to only a V_{S30} value. Procedures for analysis of non-ergodic site response from recordings cannot be used for many practical situations when on-site ground motion recordings are unavailable.

Using GRA to predict site effects can be problematic. Using data from downhole arrays in Japan, Thompson et al. (2012) found that one-dimensional methods of GRA provide a reasonable prediction of observed surface/downhole transfer functions at only 18% of their investigated sites within the KiK-Net array. Further SASW testing at the sites showed that in poorly-behaved sites, measured Rayleigh-wave dispersion curves, which can be considered as representative of a "site signature", varied significantly, likely due to complex geological structure that affected Rayleigh wave propagation and dispersion characteristics. The opposite held true for well-behaved sites: SASW dispersion curves were more consistent within the vicinity of the same site. Similar work by Afshari and Stewart (2017) at California vertical array sites also finds appreciable mismatches at > 50% of sites.

Accordingly, the profession lacks a reliable way to identify sites for which the site response is likely to differ significantly from the ergodic estimate. The V_S profile database described in Chapter 6 of this dissertation can be utilized to develop procedures for evaluating the median and natural log standard deviation of V_S profiles conditional on V_{S30} (and possibly basin depth parameters, e.g., $z_{1.0}$ or $z_{2.5}$). This median profile is hypothesized to be roughly associated with the ergodic site response, and that by comparing the median and natural log standard deviation of that V_{S30}-conditioned V_S profile to the site-specific V_S profile, it may be possible to identify sites for which the non-ergodic median site response is similar to, or significantly different from, the ergodic estimate. In this regard, the V_S PDB effort would allow for multiple profiles to be queried that fit a target V_{S30}, or range of V_{S30} values, to develop this global-average V_S profile (in a natural

logarithmic sense, corresponding to the median) conditional on a V_{S30} value. Individual profile anomalies, or an average of those fitting certain criteria such as specific surficial geologic units, can be compared to the median profile by integration over the depth of the difference in the profiles. This concept is illustrated in Figure 7.1. If the median profile were assumed to be associated with the ergodic site term, then individual profiles that vary little from the median profile over their full depth would represent a condition where the non-ergodic site response term is approximately zero. Thus, the ergodic mean could be taken as the non-ergodic mean, which would significantly reduce the effort required for computing site-specific ground motion estimates and could also potentially decrease the resulting within-event standard deviation ($\phi_{\ln}$) of such estimates. Such a study would expand upon the results in Kwak et al (2017b), who used HVSR measurements to develop a model that is intermediate between ergodic and non-ergodic that reduced $\phi_{\ln}$ by an average of 0.04. This finding would also help to fill the knowledge gap for characterization of non-ergodic site response in the absence of on-site earthquake recordings.

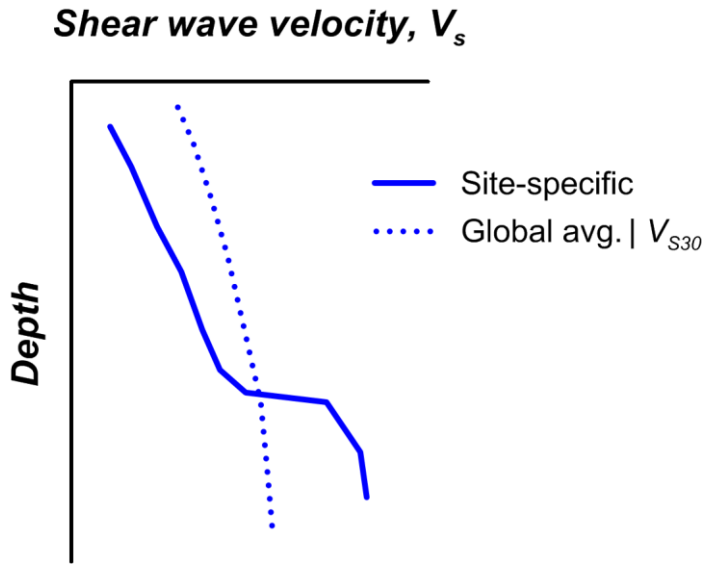

Figure 7.1. Schematic illustration of conditional global average V_S profile equivalent to conditional median for log normal distribution), which varies with depth more smoothly than a site-specific profile having the same V_{S30}. Figure courtesy Jonathan P. Stewart.

7.2.2.2 Predictability of Site Response with 1D Velocity Models

Ground response analyses are generally considered to improve upon the accuracy of ergodic predictions from GMMs. While 1D GRA cannot capture the true site response, which includes the 3-dimensional effects of surface waves, basins, and topography, it is an improvement over ergodic estimates of ground motion. A fundamental assumption of 1D GRA is horizontal layering of strata and the vertical propagation of shear waves, which can indeed capture impedance, nonlinear, and resonance effects (Stewart et al. 2014b).

Previous studies have investigated the accuracy of 1D GRA predictions. Baturay and Stewart (2003) found that in sites with large impedance contrasts, equivalent-linear 1D GRA models better predict recorded ground motions compared to the use of GMMs with generic site terms, but do not capture deep basin effects at sites in the San Fernando Valley that recorded the 1994 Northridge earthquake. Thompson et al. (2012) showed that linear 1D GRA could not capture site response from recorded motions at a majority of 100 sites across Japan, likely due to complex site geology producing 3D effects. These conclusions demonstrate a need for determining if a particular site is suitable for 1D GRA.

A study is planned to use HVSR as a diagnostic to quantify if sites are well- or poorly-behaved under the 1D assumption in GRA. Multiple HVSR measurements using ambient microtremors can be made at sites that have both surface and downhole recordings of earthquakes, such as those presented in Afshari and Stewart (2017). Sites where spatially-distributed HVSR measurements are not variable with respect to measured frequency of the HVSR amplitude peak (f_{peak}) indicate a resonant frequency that is consistent across the site. This lends to the assumption that 1D GRA would be appropriate for such a site; the opposite would hold true for sites where multiple, spatially-varying HVSR measurements would yield different values of f_{peak}. From these

measurements, a metric would be developed to quantify the amount of departure from horizontal layering for a given site, and as such determine if the particular site is amenable to 1D GRA. The practical impact of such a diagnostic factor would allow important projects to quickly determine if a site qualifies for 1D GRA, particularly at sites where ground motion observation stations do not exist. The field investigation component using HVSR measurements would likely follow established guidelines developed in Europe (SESAME, 2004).

The development of the V_S Profile Database described in Chapter 6 provides a repository for HVSR measurements at sites both with and without V_S measurements. The plan to populate the database with this data facilitates the study mentioned in this section and in the following section (7.2.2.3), where clusters of data can be easily identified via the online map interface. Additionally, because the V_S PDB schema is extensible, additional metadata pertinent to the HVSR method can be added in the future, increasing the value of the database.

7.2.2.3 Consistency of HVSR from Different Sources and its Effect on Site Amplification Models

Site amplification models used within ergodic GMMs often rely on simplified index parameters representing the geologic condition of a site, such as V_{S30} and basin depth parameters ($z_{1.0}$ or $z_{2.5}$). While these parameters represent the small-strain stiffness of the geologic materials at a site of interest, they do not consider the fundamental frequency (f_0) of the site, which may have a significant effect on estimating GMIMs during strong ground motions. Resonance effects have been observed in site amplification (Field and Jacob, 1993, 1995; Theodulidis et al., 1996; Bonilla et al., 1997; Satoh et al., 2001; Bonilla et al., 2002; Cadet et al., 2012), particularly for sites with high impedance contrasts at depth. Some researchers have questioned the validity of V_{S30} as a proxy for site amplification (Castellaro et al. 2008), and some site amplification studies (Cadet et

al. 2012, Zhao and Xu 2013, Ghofrani et al. 2013, and Braganza et al. 2016) have proposed replacement of V_{S30} or basin depth parameters to be replaced altogether by f_0.

The primary method for estimating f_0 for the aforementioned studies is to compute f_{peak}, the frequency of the highest-amplitude peak of the horizontal to vertical spectral ratio (HVSR) of recorded strong ground motions in the frequency-domain. This parameter has been shown in numerous studies to correlate to f_0 (Lermo and Chávez-García, 1993; Lachet et al., 1996; Theodulidis et al., 1996; Bonilla et al., 2002, Kawase et al., 2011; Cadet et al., 2012, Ghofrani et al., 2013, Hassani and Atkinson, 2015). The f_{peak} parameter can also be computed from the pre-event noise in the recorded seismic signal or from performing ambient-vibration (i.e. passive-source) surface wave testing (e.g., Nakamura 1989).

However, it is unclear from the literature if HVSR-derived amplitudes at f_{peak} from each of these methods are consistent. Amplitudes from ambient noise measurements have been found to be lower than those derived using earthquakes as the source (Mexico, Lermo and Chavez-Garcia, 1994; Armenia, Field et al., 1995; Iceland, Atakan et al., 1997; Thessaloniki, Greece, Lachet et al., 1996; Garner Valley, California, Theodulidis et al., 1996; southern Italy, Mucciarelli et al., 2003). Other studies (Moya et al. 2000, Satoh et al. 2001) have also shown that even frequency values of f_{peak} are not consistent between ambient-noise and earthquake sources.

The critical issue resulting from this discrepancy is that end-users of site amplification models developed for f_0 but based on analysis of HVSR from strong ground motion may not be able to obtain such information, as ground motion recording stations are not always available at the site of interest (a similar problem to predicting non-ergodic site response, discussed in Section 7.2.2.1). More commonly, users would rely on microtremor studies to estimate HVSR, but if the methods are inconsistent, the models are rendered less useful. A study that utilizes the V_S PDB

described in Chapter 6 to quantify and compare the uncertainties associated with HVSR derived from these various methods would be beneficial for future development of site amplification models that seek to use f_0 alone or in conjunction with V_{S30} and/or basin depth terms as the main site parameter. As proposed in Section 7.2.2.2, the addition of HVSR data added to the V_S PDB would be relatively straightforward after data sources are identified and data is acquired and would facilitate studies that can be used in future GMM development that consider various parameters that represent site conditions.

7.2.2.4 Improved Models of V_S Structure in Sedimentary Basins for Use in Ground Motion Models

Ground Motion Models (GMMs) are derived in part using databases of recorded ground motions and earthquake- and site-specific information to estimate ground motion intensity measures. Source, path, and site effects affect intensity measures, and each of these components has corresponding terms in a GMM. The site terms in GMMs often use the time-averaged V_{S30} as the principle site parameter, which therefore acts as the predictor of *average* ground response and basin response effects.

Three of the NGA-West2 GMMs were formulated to "center" basin amplification models with respect to the V_{S30}-scaling models that are used in combination. The concept is that the V_{S30}-scaling terms contain a certain degree of globally average (i.e., ergodic) amplification that includes basin response effects. For a given V_{S30}, those average basin effects can be associated with an average depth; in other words, the basin amplification model should be centered on that average depth. To achieve centering, the NGA-West2 basin models are derived not from the depth itself, but from the differential depth, defined as the actual or measured basin depth minus the average for a given value of V_{S30}. The centered basin models predict larger than average ground motions

at long oscillator periods for deeper-than-average sites, and the opposite for shallow sites.

As a result of this model formulation, the relationship between basin depth and V_{S30} is critical. The models derived in NGA-West2 can be improved, as they were derived for relatively large regions (i.e., California and Japan). Using the data from the United States Community V_S PDB allows for improved depth-V_{S30} models that may be regional (e.g., basin-specific) or which may be related to the geologic processes that lead to basin formation, such as differentiating typical alluvial fan basin edges from tectonically controlled pull-apart basins such as the Imperial Valley of California. Recent work by Nweke et al. (2018) articulates this framework, regionalizing basin response for seven different basins in southern California that have various tectonic and structural geologic controls on their shape and formation.

Preliminary work in the PNW shows that there does not seem to be a correlation between $z_{2.5}$ and V_{S30} for measured data from Ahdi et al. (2017a) across five basins in the Pacific Northwest region, as shown in Figure 7.2. However, Figure 2.8 shows that regionalization of the various basins in the PNW might hold the key to better parameterizing the V_{S30}-$z_{2.5}$ scaling models that can be used in NGA-Subduction GMM development. Work of this sort plays a critical role in localizing site response models to improve accuracy and reduce uncertainties.

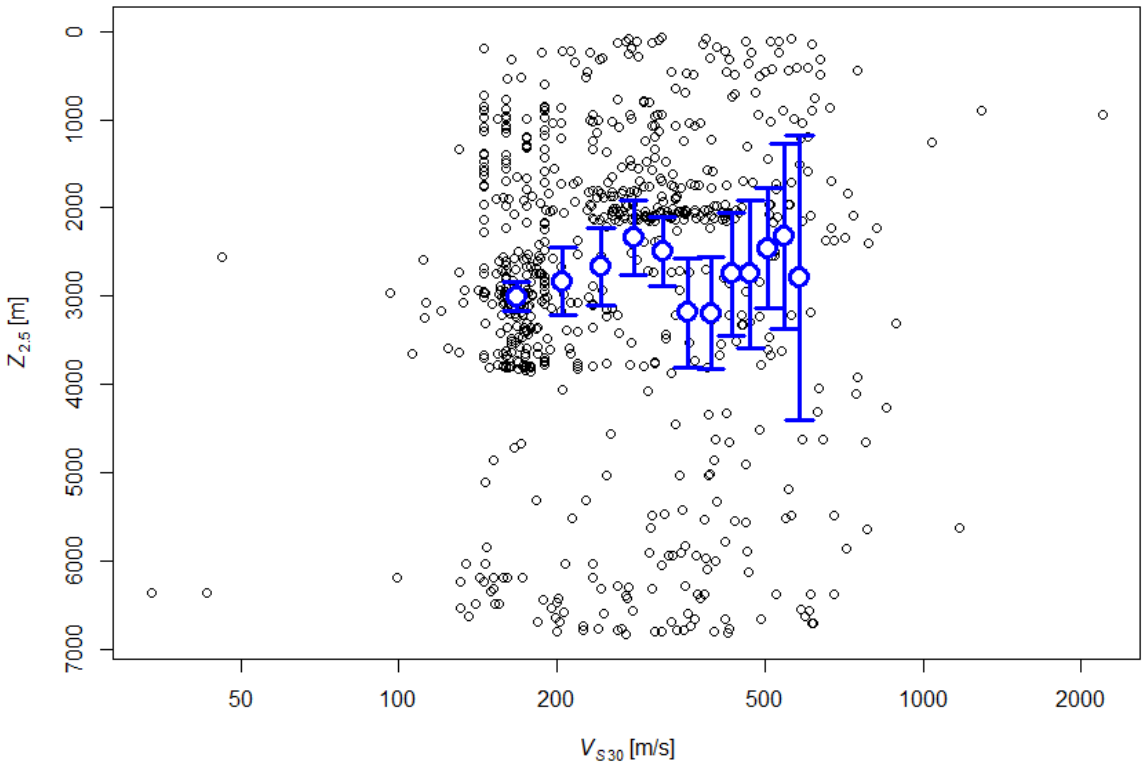

Figure 7.2. Depth to the V_S = 2.5 km/s velocity horizon versus V_{S30} in the Pacific Northwest.

Appendix A: Extrapolation Model from V_{SZ} to V_{S30}

Shear wave velocity (V_S) profiles are often limited to < 30 m depth. In such cases, V_{S30}, the time-averaged V_S to 30 m depth, cannot be measured directly, but is alternatively estimated from correlation models using time-averaged V_S down to the profile depth (V_{SZ}) and/or bottom-layer V_S at the maximum profile depth [$V_S(z_p)$]. Five models were considered from the literature and regression coefficients were computed for each using the Pacific Northwest (PNW) dataset. This Appendix presents the relative performance of each model.

The first model considered is Boore (2004, denoted B04), which was developed from 135 borehole geophysical measurements in California. The model form is:

$$\log\left(V_{S30}\right) = a_0 + a_1 \log\left(V_{SZ}\right) \tag{A.1}$$

where a_0 and a_1 are model coefficients. The second considered model (Boore et al. 2011, denoted Bea11) is parabolic in form, and is based on data from 638 KiK-net sites in Japan:

$$\log\left(V_{S30}\right) = b_0 + b_1 \log\left(V_{SZ}\right) + b_2 \log\left(V_{SZ}\right)^2 \tag{A.2}$$

Midorikawa and Nogi (2015, denoted MN15) use data from 2009 sites in Japan to extend the B04 model by considering the V_S at the profile depth, V_S (z_p), as follows:

$$\log\left(V_{S30}\right) = c_0 + c_1 \log\left(V_{SZ}\right) + c_2 \log\left(V_S\left(z_p\right)\right) \tag{A.3}$$

Dai et al (2013, denoted Dea13) present an approach in which the travel time from the base of the profile to 30 m is estimated from V_S (z_p), as follows:

$$V_{S30} = \frac{30}{\frac{z_p}{V_{SZ}} + \frac{30 - z_p}{V_{SZ30}}} \tag{A.4}$$

$$\log\left(V_{SZ30}\right) = d_0 + d_1 \ln\left(V_S\left(z_p\right)\right) \tag{A.5}$$

where V_{SZ30} represents the time-averaged V_S from the profile depth (z_p) to 30 m, which is used for the travel time computation over this depth range. Dea13 present the methodology, but do not provide coefficients.

Wang and Wang (2015, denoted WW15) use the following model:

$$\log\left(V_{S30}\right) = \log\left(V_{SZ2}\right) + \frac{\log(30) - \log\left(z_2\right)}{\log\left(z_2\right) - \log\left(z_1\right)}\left[\log\left(V_{SZ2}\right) - \log\left(V_{SZ1}\right)\right] \tag{A.6}$$

where z_1 and z_2 are two depths within the profile ($z_2 > z_1$) having corresponding time-averaged velocities to those depths of V_{SZ1} and V_{SZ2}. As z_1 and z_2 approach 30 m, the correlation to measured V_{S30} increases and variation of error decreases (Wang and Wang, 2015). Testing with PNW data, setting profile depth as z_2 and $z_1 = z_2 - 1$ m results in the least variation; whereas $z_1 = z_2 - 5$ m results in the minimum bias of errors. This method does not use regression coefficients.

For the first four methods, PNW-specific model coefficients were developed using the PNW V_S profile database described in the main text. For this purpose, the 450 profiles with $z_p \geq 30$ m were considered. Figure A1 shows model coefficients from the present analysis and the original study. The differences are large and statistically significant for the B04 and Bea11 models (Figures A1a-b), based on the published coefficients falling outside the 95% confidence intervals of the coefficients in the present study. The differences are much smaller for the MN15 model.

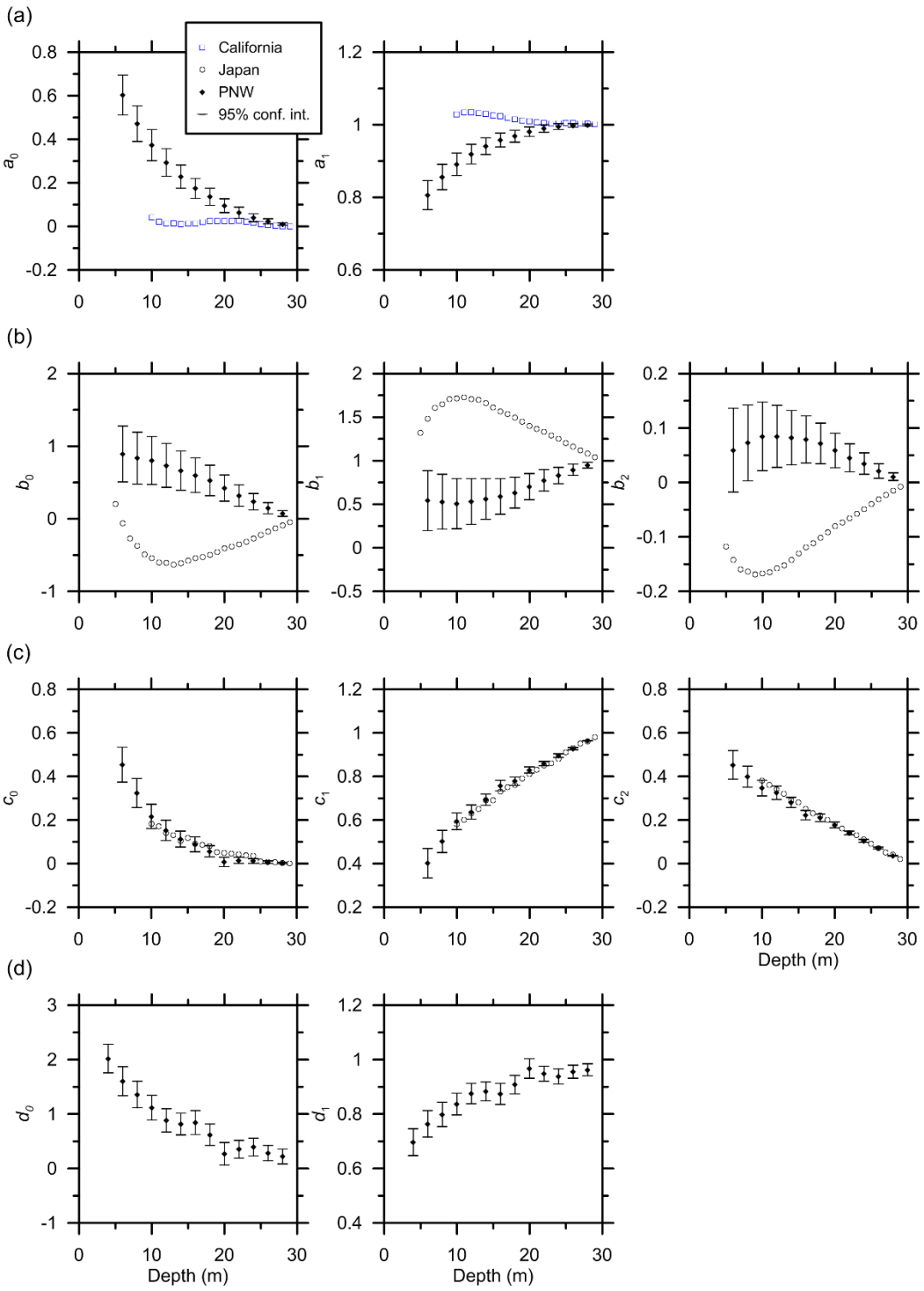

Figure A.1. Model coefficients regressed using PNW database and their 95% confidence intervals for z_p = 4–30 m. Coefficients from original studies (B04, Bea11, MN15) are also shown: (a) Boore (2004); (b) Boore et al. (2011); (c) Midorikawa and Nogi (2015); (d) Dai et al. 2013.

Model performance is validated by analyzing residuals (R) as follows:

$$R = \ln\left(V_{S30}\right) - \ln\left(V_{S30}\right) \tag{A.7}$$

where V_{S30} is measured from PDB data and V_{S30} is estimated using Equations (A.1-A.6). Figure A2 shows mean ($\overline{R}$) and standard deviation (σ_e) of R. Linear regression models directly estimating V_{S30} (i.e., B04, Bea11, MN15) have zero $\overline{R}$, whereas Dea13 has slightly negatively-biased $\overline{R}$ for shallow depths (z_p < 10 m). The $\overline{R}$ from WW15 has a significant positive bias for z_p < 18 m.

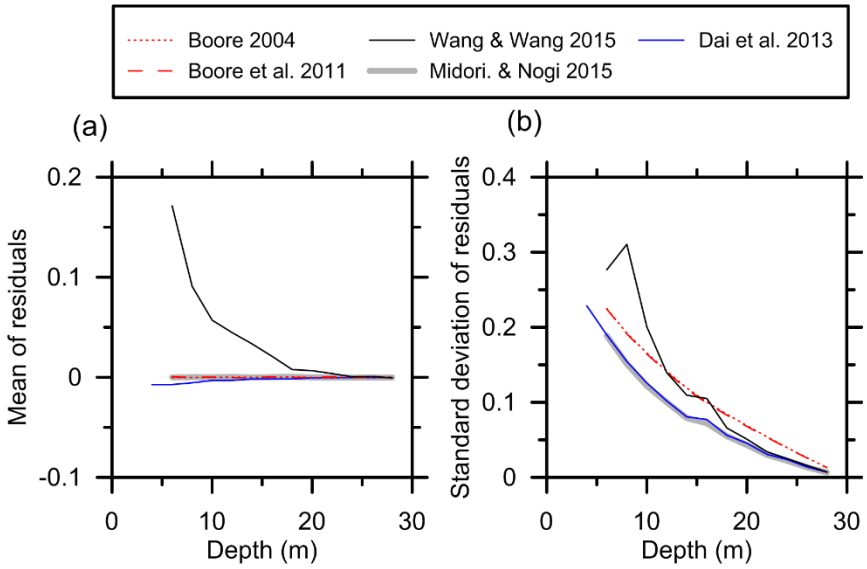

Figure A.2. (a) Mean and (b) standard deviation of residuals [i.e., $\ln\left(V_{S30}\right) - \ln\left(V_{S30}\right)$].

Dea13 and MN15 have comparable dispersion of residuals (σ_e) which are lower than those for other methods (B04, Bea11, and WW15). While both Dea13 and MN15 have low σ_e, the bias evident in MN15 residuals motivate the selection of the Dea13 methodology and coefficients, as mentioned in the main text.

Rather than providing coefficients of Dea13 (d_0 and d_1) for each depth, a nonlinear model

was fit to d_0 and d_1, making each as a function of depth z_p:

$$d_0 = \alpha_0 + \alpha_1 \ln\left(z_p\right)^{\alpha_2} \tag{A.8}$$

$$d_1 = \beta_0 + \beta_1 \ln\left(z_p\right)^{\beta_2} \tag{A.9}$$

where α_0, α_1, and α_2 are regression coefficients for d_0, and β_0, β_1, and β_2 are regression coefficients for d_1. Figure A3 shows d_0 and d_1 with the model fits, and Table A1 lists the fit model's coefficients. The models lie within the 95% confidence intervals of d_0 and d_1 at each depth. Thus, predicting V_{S30} using d_0 and d_1 from the fit model (fit line in Figure A3a and b) is similar to predicting V_{S30} using d_0 and d_1 developed for each depth (points in Figure A3a and b). A fit model to σ_e of Dea13 (Figure A3c) as a function of profile depth z_p is also provided:

$$\sigma_e = 0.394 - 0.117 \ln\left(z_p\right) \tag{A.10}$$

Table A1. Regression parameters for Dai et al. (2013) coefficients using PNW data.

Coefficient	Intercept	Slope	Power
d_0	α_0 = 3.892	α_1 = -1.451	α_2 = 0.777
d_1	β_0 = 0.228	β_1 = 0.394	β_2 = 0.524

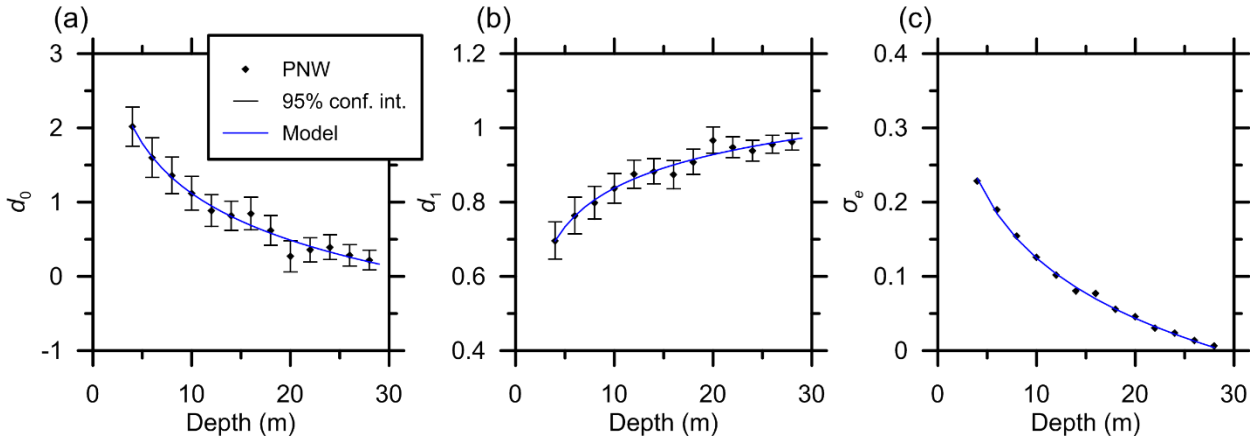

Figure A.3. Model fits to coefficients (a) d_0 and (b) d_1, and (c) standard deviation of Dea13 using PNW data.

Appendix B: Iranian Strong Motion Site Database

See attached spreadsheet file “Appendix_B_Iran_SDB_final.csv”. Contents are described in Chapter 5 of this dissertation.

REFERENCES

Abers, G.A. and Fischer, K.M. (2003). Tomography Under Costa Rica and Nicaragua. *International Federation of Digital Seismograph Networks*. https://doi.org/10.7914/SN/YO_2003.

Abrahamson, N.A., Kuehn, N., Gulerce, Z., Gregor, N., Bozorgnia, Y., Parker, G.A., Stewart, J.P., Chiou, B., Idriss, I.M., Campbell, K.W., and Youngs, R.R. (2018). Update of the BC Hydro subduction ground-motion model using the NGA- Subduction dataset, *Report No. 2018/02, Pacific Earthquake Engineering Research Center*, UC Berkeley.

Afshari, K. and Stewart, J.P. (2017). Implications of California Vertical Array Data for the Analysis of Site Response with 1D Geotechnical Modeling, *Report to California Geological Survey*, Civil & Environmental Engineering Department, UCLA.

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

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

Ahdi, S.K., Stewart, J.P., Kwak, D.Y., Ancheta, T.D., and Mitra, D. (2017b). Proxy-Based V_{S30} Prediction in the Pacific Northwest and Alaska, *3rd International Conference on Performance-based Design in Earthquake Geotechnical Engineering (PBD-III)*, Vancouver, Canada, July 16–19, 2017, paper No. 482.

Ahdi, S.K., Ancheta, T.D., Contreras, V., Kishida, T., Kwak, D.Y., Kwok, A.O., Parker, G.A., Bozorgnia, Y., and Stewart, J.P. (2017c). NGA-Subduction Site Database. *Proceedings of the 16th World Conference on Earthquake Engineering*, Santiago, Chile, January 9th to 13th 2017. Paper No. 4926.

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.

AEC (2018). Seismic Network webpage, Alaska Earthquake Center, University of Alaska Fairbanks, https://earthquake.alaska.edu/network.

Allen, T. I., and Wald, D.J. (2009). On the use of high-resolution topographic data as a proxy for seismic site conditions (V_{S30}), *Bulletin of the Seismological Society of America,* **99**(2A), 935–943.

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.

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. (2013). PEER NGA-West2 Database, *PEER Report 2013/03*, Pacific Earthquake Engineering Research Center, Berkeley, California, USA.

Anderson, D.G. and Woods, R.D. (1975). Comparison of field and laboratory shear modulus, *Proc. ASCE Conference 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, *Bulletin of the Seismological Society of America,* **74**(5), 1969–1993.

Anderson, J.G. and Brune, J. (1999). Probabilistic seismic hazard analysis without the ergodic assumption, *Seismological Research Letters*, **70**, 19-28.

Anderson, J., et al. (2006). Guerrero Accelerograph Network: Highlights From 20 Years of Operation, *8th U.S. National Conference on Earthquake Engineering (8NCEE), 18–22 April 2006, San Francisco, CA.*

Andrus, R.D. and Stokoe II, K.H. (2000). Liquefaction Resistance of Soils from Shear-Wave Velocity, *Journal of Geotechnical and Geoenvironmental Engineering*, **126**(11), 1015–1025.

Aoi., S., Takashi, K., and Fujiwara, H. (2004). Strong-Motion Seismograph Network Operated by NIED: K-NET and KiK-net, *Journal of the Japan Association of Earthquake Eng.,* **4**(3), 65–74.

Apel, T., Williams, C., and Grossi, P. (2012). The M9.0 Tohoku, Japan Earthquake: Short-Term Changes in Seismic Risk, *Risk Management Solutions (RMS) Inc. Special Report*, 30 p.

Arango, M.C., Strasser, F.O., Bommer, J.J., Hernandez, D.A., and Cepeda, J.M. (2010). A strong-motion database from the Central American subduction zone, *Journal of Seismology*, **15**(2), 261–294.

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 & Geoenvironmental 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 Internationa*l, West Conshohocken, PA.

Atakan, K., Brandsdóttir, B., Halldórsson, P., and Fridleifsson, G.O. (1997). Site response as a function of near-surface geology in the South Iceland seismic zone, *Nat. Hazards* **15**, 139–164.

Baturay, M.B. and Stewart, J.P. (2003). Uncertainty and bias in ground motion estimates from ground response analyses, *Bulletin of the Seismological Society of America*, **93**(5), 2025–2042.

BHRC (2007). *Iranian Code of Practice for Seismic Resistant Design of Buildings, Standard No. 2800*, 3rd ed., Building and Housing Research Center of Iran, Tehran, Iran.

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.

Bonilla, L.F., Steidl, J.H., Gariel, J.-C., and Archuleta, R.J. (2002). Borehole response studies at the Garner Valley downhole array, Southern California, *Bulletin of the Seismological Society of America*, **92**, 3165–3179.

Bonilla, L.F., Steidl, J.H., Lindley, G.T., Tumarkin, A.G., and Archuleta, R.J. (1997). Site amplification in the San Fernando Valley, California: Variability of site-effect estimation using the S-wave, coda, and H/V methods, *Bulletin of the Seismological Society of America*, **87**, 710–730.

Boore, D.M. (2003). 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. (2004). Estimating V_{S30} (or NEHRP site classes) from shallow velocity models (depths < 30 m), *Bulletin of the Seismological Society of America*, **94**, 591–597.

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, *Seismological Research Letters*, **69**, 222–229.

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., Thompson, E.M., and Cadet, H. (2011). Regional correlations of V_{S30} and velocities averaged over depths less than and greater than 30 m, *Bulletin of the Seismological Society of America*, **101**, 3046–3059.

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.

Borcherdt, R.D., and Gibbs, J.F. (1976). Effects of local geological conditions in the San Francisco Bay region on ground motions and the intensities of the 1906 earthquake. *Bulletin of the Seismological Society of America*, **66**, 467–500.

Borcherdt, R.D. and Glassmoyer, G. (1994). Influences of local geology on strong and weak

ground motions recorded in the San Francisco Bay region and their implications for site-specific building-code provisions, *The Loma Prieta, California Earthquake of October 17, 1989—Strong Ground Motion, U.S. Geological Survey Professional Paper 1551-A*, A77–A108.

Borcherdt, R.D. (2002). Empirical evidence for acceleration-dependent amplification factors, *Bulletin of the Seismological Society of America*, **92**(2), 761–782.

Bozorgnia, Y., Abrahamson, N.A., Al Atik, L., 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.R., Mahin, S.A., Rezaeian, S., Rowshandel, B., Seyhan, E., Shah, S., Shantz, T., Silva, W.J., Spudich, P., Stewart, J.P., Watson-Lamprey, J., Wooddell, K.E., and Youngs, R.R. (2014). NGA-West2 Research Project, *Earthquake Spectra*, **30**(3), 973–987.

Braganza, S., Atkinson, G.M., Ghofrani, H., Hassani, B., Chouinard, L., Rosset, P., Motazedian, D., and Hunter, J. (2016). Modeling Site Amplification in Eastern Canada on a Regional Scale. *Seismological Research Letters* **87**(4), 1008–1021.

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 Dynamics and Earthquake Engineering*, **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.

Brocher, T.M., Parsons, T., Blakely, R.J., Christensen, N.I., Fisher, M.A., Wells, R.E., and SHIPS Working Group (2001) Upper crustal structure in Puget Lowland, Washington—Results from the 1998 Seismic Hazards Investigation in Puget Sound, *Journal of Geophysical Research*, **106**(B7), 13541–13564.

Brocher, T.M. (2005). Empirical relations between elastic wave speeds and density in the Earth's crust, *Bulletin of the Seismological Society of America*, **95**(6), 2081–2092.

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.

Cadet, H., Bard, P.-Y., Duval, A.-M., and Bertrand, E. (2012). Site effect assessment using KiK-net data: Part 2—site amplification prediction equation based on f_0 and V_{SZ}, *Bulletin of Earthquake Engineering*, **10**, 451–489.

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., 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.

Campbell, K.W. and Bozorgnia, Y. (2014). NGA-West2 Ground Motion Model for the Average Horizontal Components of PGA, PGV, and 5% Damped Linear Acceleration Response Spectra. *Earthquake Spectra* **30**(3), 1087-1115.

Castellaro, S., Mulargia, F., and Rossi, P.L. (2008). Vs30: Proxy for Site Amplification? *Seismological Research Letters*, **79**(4), 540–543.

Castillo, L.F., Dimaté, C., Drouet, S., Fernandez, G.A., Montalva, G., Bastias, N., Morales, C., Pirchiner, M., Singaucho, J.C., and Weatherill, G. (2016). A South American Strong Motion Database and Selection of ground motion prediction equations (GMPEs) for seismic hazard analysis in South America, Research topic 6 (RT6). Wiki of the South America Risk Assessment (SARA) project of the Global Earthquake Model (GEM), available at https://sara.openquake.org/hazard_rt6 (last accessed May 2017).

Chang, S.W., 1996. *Seismic Response of Deep Stiff Soil Deposits*, *Ph.D. Dissertation*, U.C. Berkeley, 459 p.

Chemenda, A.I., Yang, R.K., Hsieh, C.-H., Groholsky, A.L. (1997). Evolutionary model for the Taiwan collision based on physical modelling, *Tectonophysics*, **274**, 253–274.

Chiou, B.S.-J., and Youngs, R.R. (2008). NGA Model for Average Horizontal Component of Peak Ground Motion and Response Spectra, *PEER Report 2008/09*, Pacific Earthquake Engineering Research Center, Berkeley, California, USA.

Chiou, B.S.-J., Darragh, R.B., Gregor, N., and Silva, W.J. (2008). NGA project strong-motion database, *Earthquake Spectra*, **24**(1), 23–44.

Colpron, M. and Nelson, J.L. (2011). A Digital Atlas of Terranes for the Northern Cordillera. Accessed online from Yukon Geological Survey (www.geology.gov.yk.ca), last visited 12 December 2018.

Contreras, V., Ruz, F., Ahdi, S.K., and Stewart, J.P. (2018). V_S Profile Database and Proxy-Based Model for V_{S30} Prediction in Chile for NGA-Subduction, *Proceedings of the 11th National Conference on Earthquake Engineering (11NCEE)*, Los Angeles, USA, 25–29 June 2018.

Cortez-Flores, A.M., 2004. *Site Response of The 2001 Southern Peru Earthquake, M.S. Thesis*, Washington State University, 291 p.

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, *Journal of Geotechnical and Geoenvironmental Engineering*, **137**(4), 354–362.

Cox, B.R., Wood, C.M., and Hazirbaba, K. (2012). Frozen and unfrozen shear wave velocity seismic site characterization of Fairbanks, Alaska, *Journal of Cold Regions Engineering*, **26**(3), 118–145.

Dai, Z., Li, X., and Hou, C. (2013). A shear-wave velocity model for V_{S30} estimation based on a conditional independence property, *Bulletin of the Seismological Society of America*, **103**(6), 3354-3361.

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.

Delorey, A.A., and Vidale, J.E. (2011). Basin Shear-Wave Velocities beneath Seattle, Washington, from Noise-Correlation Rayleigh Waves, *Bulletin of the Seismological Society of America*, **101**(5), 2162–2175.

Dobry, R., Borcherdt, R.D., Crouse, C.B., Idriss, I.M., Joyner, W.B., Martin, G.R., Power, M.S., Rinne, E.E., and Seed, R.B. (2000). New site coefficients and site classification system used in recent building seismic code provisions. *Earthquake Spectra*, **16**, 41–67.

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.

Dunn, D. and Ricketts, B. (1994). Surficial geology of Fraser Lowlands digitized from GSC Maps 1484A, 1485A, 1486A, and 1487A (92G/1,2,3,6,7; 92H/4), *Geological Survey of Canada, Open File 2894*, 4 maps at 1:50,000 scale.

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, *Geophysical Journal International*, **143**, 365–375.

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.

Faccioli, E.E., Battistella, C., Alemani, P., and Tibaldi, A. (1988). Seismic microzoning investigations in the metropolitan area of San Salvador, El Salvador, following the destructive earthquake of 10 October 1986. In: *Proceedings of the International Seminar on Earthquake Engineering, Innsbruck, Austria*, 28–65.

Faccioli, E.E., Santayo, V., and Leone, J.L. (1973). Microzonation criteria and seismic response studies for the city of Managua. In: *Proceedings of the Earthquake Engineering Research Institute Conference on the Managua, Nicaragua, Earthquake of December 23, 1972, Vol. 1*, 271–291.

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

Field, E.H. and Jacob, K.H. (1993). The theoretical response of sedimentary layers to ambient seismic noise, *Geophysical Research Letters*, **20**, 2925–2928.

Field, E.H. and Jacob, K.H. (1995). A comparison and test of various site response estimation techniques, including three that are not reference site dependent, *Bulletin of the Seismological Society of America*, **85**, 1127–1143.

Field, E.H., Clement, A.C., Jacob, K.H., Aharonian, V., Hough, S.E., Friberg, P.A., Babaian, T.O., Karapetian, S.S., Hovanessian, S.M., and Abramian, H.A. (1995). Earthquake site response study in Giumri (formerly Leninakan), Armenia, using ambient noise observations, *Bulletin of*

the Seismological Society of America, **85**, 349–353.

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.

Foster, K., Bradley, B.A., Wotherspoon, L.M., and McGann, C.R. (2017). A Vs30 map for New Zealand based on surficial geology, topography, and direct measurements. *QuakeCoRE Annual Meeting*, 3–6 September 2017, Taupo, New Zealand (poster).

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.

Fujiwara, H., Kawai, S., Aoi, S., Morikawa, N., Senna, S., Kudo, N., Ooi, M., Hao, K.X.-S., Hayakawa, Y., Toyama, N., Matsuyama, N., Iwamoto, K., Suzuki, H., and Ei, R. (2009). A study on subsurface structure model for deep sedimentary layers of Japan for strong-motion evaluation. *Technical note of the National Research Institute for Earth Science and Disaster Prevention, No. 337* (in Japanese).

Fujiwara, H., Kawai, S., Aoi, S., Morikawa, N., Senna, S., Azuma, H., Ooi, M., Hao, K.X.-S., Hasegawa, N., Maeda, T., Iwaki, A., Wakamatsu, K., Imoto, M., Okumura, T., Matsuyama, H., and Narita, A. (2012). Some Improvements of Seismic Hazard Assessment based on the 2011 Tohoku Earthquake. *Technical Note of the National Research Institute for Earth Science and Disaster Prevention, No. 379* (in Japanese).

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.

Gannett, M.W., and Caldwell, R.R. (1998). Geologic framework of the Willamette Lowland aquifer system, Oregon and Washington: *U.S. Geological Survey Professional Paper 1424–A*, 32 p., 8 plates.

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 Corporation*, 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.

Ghofrani, H., Atkinson, G.M., and Goda, K. (2013). Implications of the 2011 M9.0 Tohoku Japan earthquake for the treatment of site effects in large earthquakes, *Bulletin of Earthquake Engineering*, **11**, 171–203.

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.

Global Volcanism Program (2013). Volcanoes of the World, v. 4.5.0. Venzke, E (ed.). Smithsonian Institution. Downloaded 15 Jul 2016. http://dx.doi.org/10.5479/si.GVP.VOTW4-2013.

Gregor, N., Abrahamson, N.A., Atkinson, G.M., Boore, D.M., Bozorgnia, Y., Campbell, K.W., Chiou, B.S.-J., Idriss, I.M., Kamai, R., Seyhan, E., Silva, W.J., Stewart, J.P., and Youngs, R.R. (2014). Comparison of NGA-West2 GMPEs, *Earthquake Spectra*, **30**, 1179–1197.

GSC (1989). Canadian National Seismograph Network. *International Federation of Digital Seismograph Networks, Geological Survey of Canada*. https://doi.org/10.7914/SN/CN.

Haller, K.M., Moschetti, M.P., Mueller, C.S., Rezaeian, S., Petersen, M.D., and Zeng, Y. (2015). Seismic Hazard in the Intermountain West, *Earthquake Spectra* **31**(S1), S149–S176.

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, *Bulletin of the Seismological Society of America,* **104**, 684–701.

Hashash, Y.M.A., Musgrove, M.I., Harmon, J.A., Groholski, D., Phillips, C.A., and Park, D. (2015). *DEEPSOIL V6.1 User Manual*. Urbana, IL, Board of Trustees of University of Illinois at Urbana-Champaign.

Hassani, B. and Atkinson, G.M. (2015). Referenced empirical ground-motion model for eastern North America. *Seismological Research Letters*, ***86***(2A), 477–491.

Hassani, B. and Atkinson, G.M. (2016). Applicability of the site fundamental frequency as a V_{S30} proxy for central and eastern North America, *Bulletin of the Seismological Society of America,* **106**(2), 653–664.

Hayes, G.P., Wald, D.J., and Johnson, R.L. (2012). Slab1.0: A three-dimensional model of global subduction zone geometries, *Journal of Geophysical Research*, **117**, B01302, doi:10.1029/2011JB008524.

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.

Ho, C.S. (1986). A Synthesis of the Geologic Evolution of Taiwan, *Tectonophysics*, **125**, 1–16.

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, *Journal of Physics of the Earth* **33**,

59–96.

Ichii, K., Sato., Yu., Sato, Yo., Hoshino, Y., Iai, S. (1999). Site characteristics of strong-motion earthquake stations in ports and harbours in Japan (Part VI). *Technical note of the port and harbor research institute No 935*, Ministry of Transport, Japan.

Idriss, I.M. (1990). "Response of soft soil sites during earthquakes*," Proceedings of the H. Bolton Seed Memorial Symposium*, J.M. Duncan (ed.), Vol. 2, 273-290.

IRIS (2003). USArray Transportable Array, *International Federation of Digital Seismograph Networks*. https://doi.org/10.7914/SN/TA.

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., and Fumal, T.E. (1981). The effect of quaternary alluvium on strong ground motion in the Coyote Lake, California, earthquake of 1979, *Bulletin of the Seismological Society of America*, **71**(4), 1333–1349.

Kaiser, A., Van Houtte, C., Perrin, N., Wotherspoon, L., McVerry, G., Cousins, J., Dellow, S., Wotherspoon, L., Bradley, B., and Lee, R. (2016). Characterizing GeoNet strong motion sites: Site metadata update for the 2015 Strong Motion Database, *2016 New Zealand Society of Earthquake Engineering Conference.*

Kaiser, A., Van Houtte, C., Perrin, N., Wotherspoon, L., McVerry, G., (2017). Site characterisation of GeoNet stations for the New Zealand strong motion database, *Bulletin of the New Zealand Society for Earthquake Engineering*, **50**(1) 39–49.

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., Thompson, E.M., Minasian, D., Moss, R.E.S., Collins, B.D., Sitar, N., Dreger, D., and Carver, G. (2004). Geotechnical Reconnaissance of the 2002 Denali Fault, Alaska, Earthquake, *Earthquake Spectra*, **20**(3), 639–667.

Kawase, H., Sánchez-Sesma, F.J., and Matsushima, S. (2011). The optimal use of horizontal-to vertical spectral ratios of earthquake motions for velocity inversions based on diffuse-field theory for plane waves, *Bulletin of the Seismological Society of America*, **101**, 2001–2014.

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. *Journal of Geotechnical and Geoenvironmental Engineering*, **139**(3), 407-419.

Kishida T., Bozorgnia, Y., Abrahamson, N.A., Ahdi, S.K., Ancheta, T.D., Boore, D.M., Campbell, K.W., Chiou, B., Darragh, R., Gregor, N., Kamai, R., Kwak, D.Y., Kwok, A.O., Lin, P., Magistrale, H., Midorikawa, S., Parker, G., Si, H., Silva, W., Stewart, J.P., Tsai, C., Wooddell, K., and Youngs, R. (2017). Development of the NGA-Subduction Database. *Proceedings of the 16th World Conference on Earthquake Engineering*, Santiago, Chile, January 9–13 2017. Paper No. 3452.

Kishida T., Contreras, V., Bozorgnia, Y., Abrahamson, N.A., Ahdi, S.K., Ancheta, T., Boore, D.M., Campbell, K., Chiou, B., Darragh, R., Gregor, N., Kuehn, N., Kwak, D.Y., Kwok, A.O., Lin, P., Magistrale, H., Mazzoni, S., Muin, S., Midorikawa, S., Parker, G., Si, H., Silva, W., Stewart, J.P., Wooddell, K.E., and Youngs, R.R. (2018). NGA-Sub Ground Motion Database. *Proceedings of the 11th National Conference on Earthquake Engineering (11NCEE)*, Los Angeles, USA, 25–29 June 2018.

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, *Proceedings of the 15th World Conference on Earthquake Engineering (15WCEE)*, Lisbon, Portugal, 24–28 September 2012, Paper No. 4557.

Kuo, C.-H., Wen, K.-L., Hsieh, H.-H., Lin, C.-M., Chang, T.-M., and Kuo, K.-W. (2012). Site classification and Vs30 estimation of free-field TSMIP stations using the logging data of EGDT. *Engineering Geology*, **129**–**130**, 68–75.

Kuo, C.-H., Chen, C.-T., Lin, C.-M., Wen, K.-L., Huang, J.-Y., and Chang, S.-C. (2016). S-wave velocity structure and site effect parameters derived from microtremor arrays in the Western Plain of Taiwan, *Journal of Asian Earth Sciences*, **128**, 27–41.

Kwak, D.Y., Brandenberg, S.J., Mikami, A., and Stewart, J.P. (2015). Prediction Equations for Estimating Shear-Wave Velocity from Combined Geotechnical and Geomorphic Indexes Based on Japanese Data Set, *Bulletin of the Seismological Society of America*, **105**(4), 1919–1930.

Kwak, D.Y., Ancheta, T.D., Mitra, D., Ahdi, S.K., Zimmaro, P., Parker, G.A., Brandenberg, S.J., and Stewart, J.P. (2017a). 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.

Kwak, D.Y., Stewart, J.P., Mandokhail, S.J., and Park, D. (2017b). Supplementing V_{S30} with H/V Spectral Ratios for Predicting Site Effects, *Bulletin of the Seismological Society of America*, **107**(5), 2028–2042.

Kwok, A.O., Stewart, J.P., Kwak, D.Y., and Sun, P.-L. (2018). Taiwan-specific model for V_{S30} prediction considering between-proxy correlations, *Earthquake Spectra, in press,* https://doi.org/10.1193/061217EQS113M.

Lachet, D., Hatzfeld, C., Bard, P.-Y., Theodulis, N., Papaioannou, C., and Savvaidis, A. (1996). Site effects and microzonation in the city of Thessaloniki (Greece): comparison of different approaches, *Bulletin of the Seismological Society of America*, **86,** 1692–1703.

Lermo, J. and Chávez-García, F.J. (1993). Site effect evaluation using spectral ratios with only one station, *Bulletin of the Seismological Society of America*, **83**, 1574–1594.

Lermo, J. and Chávez-García, F.J. (1994). Are microtremors useful in site response evaluation? *Bulletin of the Seismological Society of America*, **84**, 1350–1364.

Leyton, F., Leopold, A., Hurtado, G., Pastén, C., Ruiz, S., Montalva, G., and Saéz, E., 2018.

Geophysical Characterization of the Chilean Seismological Stations: First Results, *Seismological Research Letters*, **89**(2A), 519–525.

Lin, C.-M., Wen, K.-L., Kuo, C.-H., and Lin, C.-Y. (2014). S-wave velocity model of Taipei Basin, *5th Asia Conference on Earthquake Engineering (5ACEE)*, 16–18 October 2014, Taipei, Taiwan.

Louie, J.N. (2001). Faster, better: Shear-wave velocity to 100 meters depth from Refraction Microtremor Arrays, *Bulletin of the Seismological Society of America*, **91**(2), 347–364.

Lundberg, N., Reed, D.L., Liu, C.-S., and Lieske Jr., J. (1997). Forearc-basin closure and arc accretion in the submarine suture zone south of Taiwan, *Tectonophysics*, **274**, 5–23.

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

Luzi, L., Puglia, R., Pacor, F., Gallipoli, M.R., Bindi, D., and Mucciarelli, M. (2011). Proposal for a soil classification based on parameters alternative or complementary to V_{S30}. *Bulletin of Earthquake Engineering*, **9**, 1877.

Luzi, L., Puglia, R., Russo, E., and ORFEUS WG5, (2016). Engineering Strong Motion Database, version 1.0., *Istituto Nazionale di Geofisica e Vulcanologia, Observatories & Research Facilities for European Seismology*, doi: 10.13127/ESM.

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

Matsuoka M., Wakamatsu, K., Fujimoto, K., and Midorikawa, S. (2006). Average shear-wave velocity mapping using Japan Engineering Geomorphologic Classification Map. *Structural Engineering & Earthquake Engineering*, **23**(1), 57–68.

Matsuoka, M. and Wakamatsu, K. (2008). Site amplification capability map based on the 7.5-arc-second Japan Engineering Geomorphologic Classification Map. *No. H20PRO-93*, National Institute of Advanced Industrial Science and Technology, Intellectual Property management.

McCaffrey, R., Qamar, A.I., King, R.W., Wells, R., Khazaradze, G., Williams, C.A., Stevens, C.W., Vollick, J.J., and Zwick, P.C. (2007). Fault locking, block rotation and crustal deformation in the Pacific Northwest, *Geophysical Journal International*, **169** (3), 1315-1340.

Midorikawa, S., and Nogi, Y. (2015). Estimation of V_{S30} from shallow velocity profile, *Journal of the Japan Association of Earthquake Engineers*, **15**(2), 91-96 (*in Japanese*).

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.*

Moschetti, M.P., Ritzwoller, M.H., and Shapiro, N.M. (2007). Surface wave tomography of the western United States from ambient noise—Rayleigh wave group velocity maps, *Geochemistry, Geophysics, and Geosystems,* **8**(Q08010), doi: 10.1029/2007GC001655.

Moya, A., Schmidt, V., Segura, C., Boschini, I., and Atakan, K. (2000). Empirical evaluation of site effects in the metropolitan area of San José, Costa Rica, *Soil Dynamics and Earthquake Engineering,* **20**, 177-185.

Mucciarelli, M., Gallipoli, M.R., and Arcieri, M. (2003). The stability of the horizontal-to vertical spectral ratio of triggered noise and earthquake recordings, *Bulletin of the Seismological Society of America,* **93**, 1407–1412.

Nakamura, Y. (1989). A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. *Quaterly Report Railway Tech. Research Institute*, 30-1, 25–30.

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.

Nikolau, S., Vera-Grunauer, X., and Gilsanz, R. (2016). *GEER-ATC Earthquake Reconnaissance: April 16th 2016, Musine, Ecuador, Version 1b*, 24 October 2016, 604 p.

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.

Ohta, T., Hiehata, S., Takahashi, K., Kitamura, E., Motosaka, M., Kamata, M., and Miyamura, M. (1986). Research on the Strong Ground Motion in Mexico City During the Earthquake of September 19, 1985 Michoacan-Guerrero, Mexico, *Kajima institute of Construction Technology, Kajima Corporation, KICT Report No. 68*, 42 p.

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.

Parker, G.A., Harmon, J.A., Stewart, J.P., Hashash, Y.M.A., Kottke, A.R., Rathje, E.M., Silva, W.J., and Campbell, K.W. (2017). Proxy-Based V_{S30} Estimation in Central and Eastern North America. *Bulletin of the Seismological Society of America,* **107**(1), 117-131.

PEER (2019). NGA-Subduction Database, *PEER Report 201x/xx*, Pacific Earthquake Engineering Research Center, Berkeley, CA.

Pérez, F. and Granger, B.E. (2007). IPython: A System for Interactive Scientific Computing, Computing in Science and Engineering, **9**(3), 21–29, doi:10.1109/MCSE.2007.53, available at http://ipython.org.

Petersen, M.D., Moschetti, M.P., Powers, P.M., Mueller, C.S., Haller, K.M., Frankel, A.D., Zeng, Y., Rezaeian, S., Harmsen, S.C., Boyd, O.S., Field, N., Chen, R., Rukstales, K.S., Luco, N., Wheeler, R.L., Williams, R.A., and Olsen, A.H. (2015). The 2014 United States National Seismic Hazard Model. *Earthquake Spectra*, **31**, S1–S30.

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.

Pierce, K.L. (2003). Pleistocene glaciations of the Rocky Mountains, *Development in Quaternary Science* **1**, 63–76.

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.

Porter, S.C., K.L. Pierce, and T.D. Hamilton, (1983). Late Pleistocene glaciation in the Western United States, *in*: Porter, S.C. (Ed.), *The Late Pleistocene*, Vol. 1, *of*: Wright, H.E., Jr. (Ed.), *Late Quaternary Environments of the United States*. Minneapolis, Minn., University of Minnesota Press, 71–111.

Powers, T.J. and Fumal, T.E. (1993). Geologic Logs from 25 Boreholes Near Strong Motion Accelerographs that Recorded The 1989 Loma Prieta, California, Earthquake, *U.S. Geological Survey Open-File Report 93-502*, 177 p.

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.

Ramachandran, K., Hyndman, R.D., Brocher, T.M. (2006). Regional *P* wave velocity structure of the Northern Cascadia Subduction Zone, *Journal of Geophysical Research*, **111**(B12301), doi:10.1029/2005JB004108.

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.

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-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.

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 Conference on Earthquake Engineering (11NCEE)*, Earthquake Engineering Research Institute, Los Angeles, USA, 25–29 June 2018.

Saffari, H., Kuwata, Y., Takada, T., Mahdavian, A. (2012). Updated PGA, PGV, and spectral acceleration attenuation relations for Iran, *Earthquake Spectra*, 28(1), 257–276.

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.

Satoh, T., Kawase, H., and Matsushima, S. (2001). Differences between site characteristics obtained from microtremors, S-waves, P-waves, and codas, *Bulletin of the Seismological Society of America,* **91**, 313–334.

Scasserra, G., Stewart, J.P., Bazzurro, P., Lanzo, G., and Mollaioli, F. (2009). A comparison of NGA ground-motion prediction equations to Italian data, *Bulletin of the Seismological Society of America,* **99**(5), 2961–2978.

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.

Schuster, J.E. (2005). Geologic map of Washington State, *Washington State Department of Natural Resources, Division of Geology and Earth Resources, Map GM-53*, scale 1:500,000, with pamphlet 44 p.

Seed, H.B., Romo, M.P., Sun, J., Jaime, A., and Lysmer, J. (1987). Relationships between soil conditions and earthquake ground motions in Mexico City in the earthquake of Sept. 19, 1985, *UC Berkeley Earthquake Engineering Research Center, Report No. UCB/EERC-87/15*, 112 p.

SESEAME (2004). Guidelines for the Implementation of the H/V Spectral Ratio Technique on Ambient Vibrations – Measurements, Processing and Interpretation, *Site EffectS assessment using AMbient Excitations (SESAME) European Research Project; European Commission – Research General Directorate, Project No. EVG1-CT-2000-00026*, December 2004, 62 p.

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

SGM (2017). Geological map series of Mexico, 1:250,000 scale, *Servico Geológico Mexicano*, https://www.sgm.gob.mx/Gobmx/en/About_SGM/Geology.html, last accessed Dec. 2018.

Shafiee, A.H., Zafarani, H., and Jahanandish, M. (2016). Model selection for correlating VS30 with average shear-wave velocities at lower depths based on the Iranian data, *Bulletin of the Seismological Society of America*, **106**(1), 289–299.

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.

Shirzad, T. and Shomali, Z.H. (2014). Shallow crustal structures of the Tehran basin in Iran resolved by ambient noise tomography, *Geophysical Journal International*, **196**(2), 1162–1176.

Sinaeian, F., Mirzaei Alavijeh, H., and Farzanegan, E. (2008). Site geology investigation in accelerometric stations using seismic refraction method, *BHRC Publication No. R-511*, Vol. 1, Building and Housing Research Center of Iran, Tehran, Iran.

Sinaeian, F., Mirzaei Alavijeh, H., and Farzanegan, E. (2010a). Site geology investigation in accelerometric stations using seismic refraction method in Ardebil and East Azarbayejan Provinces, *BHRC Publication No. R-532*, Vol. 2, Building and Housing Research Center of Iran, Tehran, Iran.

Sinaeian, F., Mirzaei Alavijeh, H., and Farzanegan, E. (2010b). Site geology investigation in accelerometric stations using seismic refraction method in Hormzgan and South of Fars Provinces, *BHRC Publication No. R-533*, Vol. 3, Building and Housing Research Center of Iran, Tehran, Iran.

Sinaeian, F., Mirzaei Alavijeh, H., and Farzanegan, E. (2010c). Site geology investigation in accelerometric stations using seismic refraction method in Lorestan and Hamedan Provinces, *BHRC Publication No. R-534*, Vol. 4, Building and Housing Research Center of Iran, Tehran, Iran.

Sinaeian, F., Mirzaei Alavijeh, H., and Farzanegan, E. (2010d). Site geology investigation in accelerometric stations using seismic refraction method in Fars Province, *BHRC Publication No. R-535*, Vol. 5, Building and Housing Research Center of Iran, Tehran, Iran.

Sinaeian, F., Mirzaei Alavijeh, H., and Farzanegan, E. (2010e). Site geology investigation in accelerometric stations using seismic refraction method in Kerman Provinces, *BHRC Publication No. R-536*, Vol. 6, Building and Housing Research Center of Iran, Tehran, Iran.

Sinaeian, F., Mirzaei Alavijeh, H., and Farzanegan, E. (2010f). Site geology investigation in accelerometric stations using seismic refraction method in Semnan, Golestan and Mazandaran Provinces, *BHRC Publication No. R-537*, Vol. 7, Building and Housing Research Center of Iran, Tehran, Iran.

Smith, R.B. and Sbar, M.L. (1974). Contemporary Tectonics and Seismicity of the Western United States with Emphasis on the Intermountain Seismic Belt, *Geological Society of America Bulletin*, **85**, 1205–1218.

Smith, R.L., and Roe, W.P. (2015). Oregon Geologic Data Compilation (OGDC)—Release 6, *Oregon Department of Geology and Mineral Industries (DOGAMI), Open-File Report O-13-12*, electronic dataset.

Snedecor, G.W. and Cochran, W.G. (1989). *Statistical Methods*, 8th edition, Ames, Iowa: Blackwell Publishing Professional. ISBN 978-0-8138-1561-9.

SNET (2018). Webpage of the Servicio Nacional de Estudios Territoriales (SNET) Seismic Network, Medio Ambiente y Recursos Naturales (MARN), El Salvador, available at http://www.snet.gob.sv/ver/sismologia/.

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.

Stephenson, W.J. (2007). Velocity and density models incorporating the Cascadia subduction zone for 3D earthquake ground motion simulations, Version 1.3, *U.S. Geological Survey Open-File Report 2007–1348*, Earthquake Hazards Ground Motion Investigations, 24 p.

Stephenson, W.J., Reitman, N.G., and Angster, S.J. (2017). P- and S-wave velocity models incorporating the Cascadia Subduction Zone for 3D earthquake ground motion simulations—Update for Open-File Report 2007–1348, *U.S. Geological Survey Open-File Report 2017-1152*, Earthquake Hazards Ground Motion Investigations, 28 p.

Stepp, C., Ponti, D.J., Turner, L.L., Swift, J.N., Devlin, S., Zhu, Y., Benoit, J., and Bobbitt, J. (2009). PEER Lifelines Geotechnical Virtual Data Center, *PEER Report 2009/108*, Pacific Earthquake Engineering Research Center, Berkeley, CA.

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. 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., 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., 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., Klimis, N., Savvaidis, A., Theodoulidis, N., Zargli, E., Athanasopoulos, G., Pelekis, P., Mylonakis, G., and Margaris, B. (2014a). Compilation of a Local V_S Profile Database and Its Application for Inference of V_{S30} from Geologic- and Terrain-Based Proxies, *Bulletin of the Seismological Society of America*, **104**(6), 2827–2841.

Stewart, J.P., K. Afshari, and Hashash, Y.M.A. (2014b). 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. (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 Dynamics and Earthquake Engineering*, **91**, 317–328.

Stewart, J.P., Afshari, K., and Goulet, C.A. (2017). Non-Ergodic Site Response in Seismic Hazard Analysis, *Earthquake Spectra*, **33**(4), 1385–1414.

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.

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

Theodulidis, N., Archuleta, R.J., Bard, P.-Y., and Bouchon, M. (1996). Horizontal to vertical spectral ratio and geological conditions: the case of Garner Valley downhole array in southern California, *Bulletin of the Seismological Society of America*, **86,** 306–319.

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., 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., Baise, L.G., Tanaka, Y. and Kayen, R.E. (2012). A taxonomy of site response complexity. *Soil Dynamics and Earthquake Engineering*, **41,** 32–43.

Thompson, E.M. and Wald, D.J. (2012). Developing Vs30 Site-Condition Maps by Combining Observations with Geologic and Topographic Constraints, *15th World Conference on*

Earthquake Engineering (15WCEE), Lisbon, Portugal, 14–28 September 2012.

Thompson, E.M., Wald, D.J., and Worden, C.B. (2014). A V_{S30} map for California with geologic and topographic constraints, *Bulletin of the Seismological Society of America*, **104**, 2313–2321.

Thompson, E.M. (2018). An Updated V_{S30} Map for California with Geologic and Topographic Constraints: U.S. Geological Survey data release, https://doi.org/10.5066/F7JQ108S.

Toro, G.R., Abrahamson, N.A., and Schneider, J.F. (1997). Model of strong ground motions from earthquake in central and eastern North America: Best estimates and uncertainties, *Seismological Research Letters*, **68**, 41–57.

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

Updike, R.G. (1986). Engineering Geologic Maps of the Government Hill Area, Anchorage, Alaska, *U.S. Geological Survey Miscellaneous Investigations Series Map I-1610*, scale 1:4,800.

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

UW (1963). Pacific Northwest Seismic Network, *International Federation of Digital Seismograph Networks, University of Washington*. https://doi.org/10.7914/SN/UW.

Van Houtte, C., Bannister, S., Holden, C., Bourguignon, S., and McVerry, G. (2017). The New Zealand Strong Motion Database, *Bulletin of the New Zealand Society for Earthquake Engineering*, **50**(1) 1–20.

Vera-Grunauer, X., 2014. *Seismic Response of a Soft, High Plasticity, Diatomaceous Naturally Cemented Clay Deposit, Ph.D. Dissertation*, U.C. Berkeley, 910 p.

Wagner, L., Beck, S., and Long, M. (2010) PerU Lithosphere and Slab Experiment. *International Federation of Digital Seismograph Networks*. https://doi.org/10.7914/SN/ZD_2010.

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.

Wakamatsu K. and Matsuoka, M. (2013). Nationwide 7.5-arc-second Japan Engineering Geomorphologic Classification Map and V_{S30} zoning. *Journal of Disaster Research*, **8**(5), 904–911.

Wald, D.J. and Allen, T.I. (2007). Topographic slope as a proxy for seismic site conditions and amplification. *Bulletin of the Seismological Society of America*, **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, H.-Y., and Wang, S.-Y. (2015). A New Method for Estimating V_S(30) from a Shallow Shear-Wave Velocity Profile (Depth < 30 m), *Bulletin of the Seismological Society of America,* **105**(3), 1359–1370.

Washington State Department of Natural Resources, Division of Geology and Earth Resources (WA-DNR-GER, 2010). Digital geology of Washington State at 1:100,000 Scale, June 2010: *Washington State Department of Natural Resources, Division of Geology and Earth Resources,* version 3.0.

Washington State Department of Natural Resources, Division of Geology and Earth Resources (WA-DNR-GER, 2014). Surface geology, 1:24,000--GIS data, June 2014: *Washington Division of Geology and Earth Resources Digital Data Series DS-10,* version 1.0.

Wells, R.E., Blakely, R.J., Simpson, R.W., Weaver, C.S., Haugerud, R., and Wheeler, K. (2000). Tectonic plate motions, crustal blocks, and shallow earthquakes in Cascadia, *Great Cascadia Earthquake Tricentennial Open House, 26 January 2000,* Burke Museum, University of Washington, Seattle, WA (poster).

Wills, C.J. (1998). Differences in shear-wave velocity due to measurement methods: A cautionary note, *Seismological Research Letters*, **69**(3), 216–221.

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. and Clahan, K.B. (2006). Developing a map of geologically defined site-condition categories for California, *Bulletin of the Seismological Society of America*, **96**, 1483–1501.

Wills, C.J. and Gutierrez, C.I. (2008). Investigation of Geographic rules for Improving Site-Conditions Mapping, *Final Technical Report, USGS/NEHRP Award Number 07HQGR0061*, California Geological Survey, Sacramento, California, USA.

Wills, C.J., Gutierrez, C.I., Perez, F.G., and Branum, D.M. (2015). A Next Generation V_{S30} Map for California Based on Geology and Topography, *Bulletin of the Seismological Society of America*, **105**(6), 3083–3091.

Wilson, F.H., Hults, C.P., Mull, C.G., and Karl, S.M. (2015). Geologic map of Alaska, *U.S. Geological Survey Scientific Investigations Map 3340*, pamphlet 196 p., 2 sheets, scale 1:1,584,000, http://dx.doi.org/10.3133/sim3340.

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

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

Xie, J., Zimmaro, P., Li, X., Wen, Z., and Song, Y. (2016). V_{S30} Empirical Prediction Relationships Based on a New Soil-Profile Database for the Beijing Plain Area, China, *Bulletin of the Seismological Society of America*, **106**(6), 2843–2854.

Yeats, R.S., Graven, E.P., Werner, K.S., Goldfinger, C., and Popowski, T.A. (1996). Tectonics of the Willamette Valley, Oregon, *in* Rogers, A.M., Walsh, T.J., Kockelman, W.J., and Priest, G.R., *eds., Assessing earthquake hazards and reducing risk in the Pacific Northwest, Volume 1, U.S. Geological Survey Professional Paper 1560*, 183–222.

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

Yong, A. (2016). Comparison of Measured and Proxy-Based V_{S30} Values in California. *Earthquake Spectra*, **32**(1), 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. *Bulletin of the Seismological Society of America*, **102,** 114–128.

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 V_{S30} Data for the United States, *U.S. Geological Survey Data Series 978*, 8 p., http://dx.doi.org/10.3133/ds978.

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.

Zhao, J.X., and Xu, H. (2013). A comparison of V_{S30} and site 660 period as site-effect parameters in response spectral ground-motion prediction equations, *Bulletin of the Seismological Society of America*, **103**, 1–18.