Title

Enhancement of Shear Wave Velocity Profile Database and Application for ??30 Estimation in Utah

Author

Choi, Joseph

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Enhancement of Shear Wave Velocity Profile Database and Application for V_{S30} Estimation in Utah

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Civil Engineering

by

Joseph Choi

2024

ABSTRACT OF THE THESIS

Enhancement of Shear Wave Velocity Profile Database

and Application for V_{S30} Estimation in Utah

by

Joseph Choi

Master of Science in Civil Engineering

University of California, Los Angeles, 2024

Professor Jonathan Paul Stewart, Chair

The research presented in this thesis was directed toward improvement of the shear wave velocity profile database (VSPDB) in two principal respects and its application for sites in Utah.

The first improvement of the VSPDB was related to quality assurance checks of information that has been uploaded to the database since about 2016. A tool was created in python

to plot profiles and each of 4225 profiles were visually examined to identify errors. These errors include zero velocity values, depth errors, and incorrect data formatting. In many cases, potential errors could only be confirmed or corrected by reviewing source documents. Following these reviews, profiles for 621 sites were corrected (15%).

The second improvement of the VSPDB involved replacing the original database structure, or schema, with a new version. The new schema includes a test table that allow multiple data types associated with a given test, such as a seismic cone penetration test, to be properly linked to each other. To implement the new schema, an entirely new version of the database was created. Data from the original database has been migrated to the new database, although this is not the case for all data fields. While all the V_S fields have been migrated, additional work will be required to migrate other data types. The website to access the database (https://www.vspdb.org) will still access the old database until the remaining data types have been fully migrated.

The VSPDB was supplemented with 259 profiles from 205 unique sites in Utah. This study region is of interest because ground motions from Utah are being considered in the NGA-West3 project, but prior to the present work, no information on site conditions was available in NGA databases. Three of the profiles were paired with Utah ground motion stations but 92 other stations lack a co-located profile. For this reason, a proxy-based model to predict the time-averaged velocity in the upper 30 m of the site (V_{S30}) was developed that considers surface geology and surface gradient for Quaternary units in Utah. The available profile data did not allow the development of models for rock sites, for which V_{S30} was estimated using models for other regions.

The thesis of Joseph Choi is approved.

Idil Deniz Akins

Scott Joseph Brandenberg

Jonathan Paul Stewart, Committee Chair

University of California, Los Angeles

2024

ACKNOWLEDGMENTS

The research presented in this thesis was funded during the Summer of 2024 by a scholarship in the name of Alan and Kimberly Yourman.

I would like to personally thank my advisors and professors for their support during my time at UCLA. Thank you Professors Scott J. Brandenberg and Jonathan P. Stewart for teaching me during my undergraduate and graduate studies. I would have never thought about going into geotechnical engineering if it weren't for your classes. Thank you for your patience and wisdom.

Next, I would like to thank Tristan Buckreis. From being my T.A. in undergraduate to being a mentor during the completion of this thesis, it's not an understatement to say that you have been a part of my geotechnical journey since the beginning. Thank you for your support. I appreciate the data shared with me on VS profiles in Utah by Ivan Wong and Brady Cox.

I would also like to thank my family and friends who have supported me this entire time. Without them this academic journey would not have been the same.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Scope of Work

1.1 BACKGROUND AND MOTIVATION

The shear-wave profile velocity database (VSPDB) is an open-access database that archives, organizes, and makes available to the public shear-wave velocity (V_S) data and related metadata for use by professional engineers and researchers. V_S measures how fast shear waves travel through the ground surface. It is also related to the shear modulus of soil as,

$$G_{max} = \rho V_S^2 \tag{1.1}$$

where ρ is the soil mass density. V_S is a useful site characterization parameter because the speed at which the waves propagate is dependent on the material it passes through. This parameter is commonly used in geotechnical earthquake engineering applications, including for the derivation of a stress-strain backbone curve (this initial modulus is G_{max}) and ground response analyses in which the nature of ground motion change through a profile is dependent on changes in V_S with depth.

Shear wave velocity is also used when no ground response analyses are to be performed but semi-empirical ground motion models (GMMs) are to be applied. The primary parameters used to represent site condition is the time-average velocity in the upper 30 m of the site, V_{S30}. The V_{S30} parameter is computed from the V_S profile as the ratio of 30 m to the shear wave travel time from the ground surface to a depth of 30 m.

There are a variety of methods that are used to measure V_S profiles, which can be categorized as either invasive or non-invasive. Invasive procedures involve penetration into the ground with a borehole or cone penetration test (CPT) sounding, deployment of geophones below ground, and measurement of travel times directly from a source to receiver, and calculation of velocity as the ratio of travel distance to travel time. These types of procedures include the following borehole methods: downhole, crosshole, and suspension logging, as well as seismic cone penetration test (SCPT) (Kramer and Stewart, 2024; Section 6.5.2). Non-invasive procedures include utilizing actively produced surface waves or measuring natural microtremors using sensors deployed on the ground surface. The measured quantity is the phase velocity vs. frequency (or wavelength) for a surface wave, generally a Rayleigh wave, which is then converted to one or a series of V_S profiles through an inversion process. Noninvasive procedures include the following tests: Multichannel Analysis of Surface Waves (MASW), Spectral Analysis of Surface Waves (SASW), Seismic Refraction (Refraction), and Passive Seismic Methods.

Although V_S is widely measured and used, prior to the development of the VSPDB there was no centralized space for the data to be stored or accessed, which motivated the creation of VSPDB in approximately 2018 as an online relational database (Ahdi et al. 2018). Since then, VSPDB has grown significantly, with 4223 velocity profiles added along with 481 active, registered users. As the database has developed and grown, a series of problems with the data have been reported by users that can be attributed to input errors that were not caught due to limited and inconsistent quality assurance protocols. Advanced users in the UCLA Geotechnical Engineering Group, such as Dr. Tristan Buckreis, have also identified deficiencies in the schema for the relational database for VSPDB. To address both concerns, it was decided to create a new version of the database, to improve the structural problems, and to perform the necessary quality assurance on the current data before importing it to the new database. By fixing current data issues and updating the schema to improve the data integrity of the database, the database can accept new data and users can be confident in the reliability of the database.

The work on the VSPDB was conducted coincident with work on the Next Generation Attenuation (NGA)-West3 a ground motion database development effort (Buckreis et al. 2024). The intersections between the projects are many, because the VSPDB is used to provide V_S profiles for ground motion recording sites and to provide V_S profiles that can be used to develop proxy-based models for V_{S30} estimation. Within the NGA-West3 project, there was a particular need to develop such proxy-based models for Utah. There are currently 95 ground motion stations in Utah that have produced usable ground motion records; V_{S30} values are needed for these stations and measured V_S profiles are not available. To address this need, I uploaded available V_S profile data from Utah to the VSPDB and the data to develop a geology-based V_{S30} proxy model.

1.2 SCOPE

The scope of this project can be summarized as follows:

1. Perform quality assurance on existing velocity profiles in VSPDB. By looking meticulously at each profile and the original source document the data came from, errors in the data were identified and fixed.

2. Assess and update the VSPDB schema. With a better understanding of how the VSPDB relational database has been used in recent work, the needs for an improved schema were developed. The schema organizes the data so that it is more intuitive for users to access, and provides a data structure that reduces the potential for data entry errors.

3. Identify and download data from sites in Utah that include V_S profiles. Digitize and prepare the data for uploading to the VSPDB. Upload the data. Assemble information that can be

used for proxy-based models, including different digital elevation models (DEM) (corresponding to different levels of spatial resolution) and geologic maps.

4. Develop a regional V_{S30} proxy model for Utah. Using site characterization data, identify trends between slope, geology, and V_{S30}. Develop appropriate prediction equations for the mean and standard deviation of V_{S30} conditional on slope and a distilled list of geologic categories. Use the model to provide V_{S30} estimates for ground motion stations.

1.3 ORGANIZATION

The thesis is structured as follows:

Chapter 2 discusses the errors found in velocity profiles that were in the database and the process applied to identify them. An online coded tool was created to facilitate visual checks. Profiles were screened for errors and fixed. The types of errors and the methods by which they were fixed are described in detail. Examples are shown to illustrate the types of errors found.

Chapter 3 discusses what a relational database is and why it is appropriate for the new database. A comparison between the old and new database is made to show how the changes made in the structural organization are beneficial. A detailed description of the new schema is provided. Examples are given to highlight advantages of the new schema. The design of the website is shown, as well as a walkthrough is provided on how to access and navigate the database. Since the database is still in development and is expecting updates, future changes are described.

Chapter 4 describes the newly added data to the VSPDB for Utah and proxy model development. The types of data include velocity, CPT, travel time, and borehole logs. Using the newly added velocity data, a slope-geology proxy model is developed to estimate V_{S30}. The model's parameters, slope and geology, are obtained with digital elevation models (DEM) and geologic maps. By combining these two and the measured V_{S30} data, V_{S30} can be estimated for a given geology and slope. Using the proposed models, V_{S30} is estimated and assigned to Utah ground motion stations in the NGA-West3 database.

2 Database Errors

2.1 INTRODUCTION

Users of the VSPDB have reported that some of the velocity profiles did not match with the profiles provided in source documents, thus motivating an investigation into the data and screening for errors. Not only is fixing these errors critical for accurate analysis to be conducted, it also provides insight into how the schema can be improved to reduce the potential for future errors. This chapter documents the quality assurance process, the errors found in the data for some sites in the VSPDB, and how they were fixed. I also describe what percentage of sites were affected by different types of errors.

2.2 ERRORS AND RESOLUTION

Using DesignSafe's Jupyter Notebook tool, a graphical user interface (GUI) displaying the velocity profile's metadata, table of the data itself, and a plot of the profile was created to analyze the existing profiles and find profiles with significant data errors. This tool was developed by Dr. Tristan Buckreis and tested and applied by me. Figure 2.1 shows the user interface window for profile selection and commenting; the reviewer can identify profiles based on their velocityProfileMeta_ID or different profile attributes such as location. Once a profile is identified, it was plotted for user assessment using a python function that is separate from the GUI. An example of a plotted profile is provided in Figure 2.2; in this case no errors were found. Returning to the interface from Figure 2.1, options are provided for comments and the reviewer can accept, reject, or skip the profile.

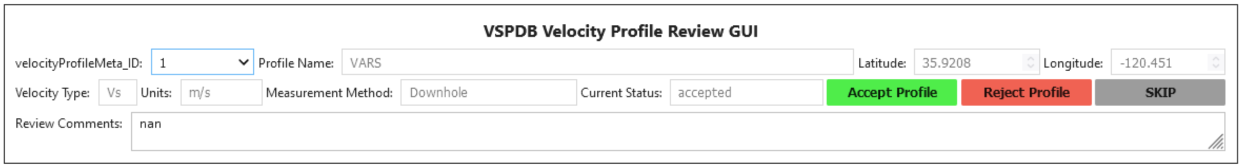
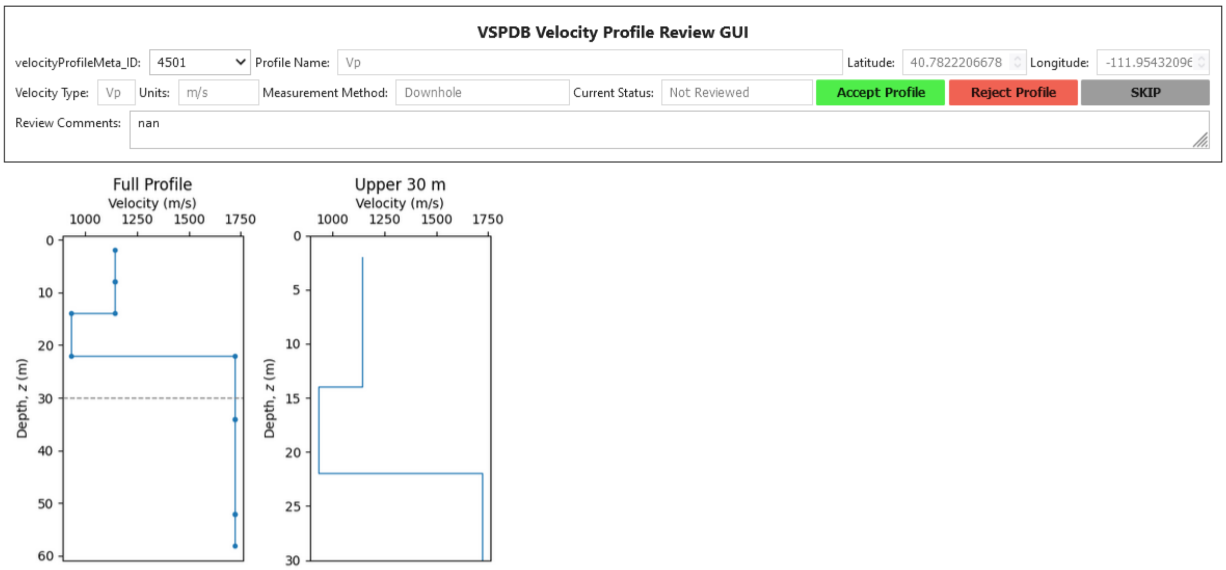


Figure 2.1. GUI used to review profiles



	velocityProfileMeta_ID	velocityProfileArray_ID	depthTop	depthBottom	value	uncertainty	depthTop_unit	value_unit	vs30_value	vs30_unit
0	4501	101376	2	8	1141	NaN	m	m/s	NaN	NaN
1	4501	101377	8	14	1141	NaN	m	m/s	NaN	NaN
2	4501	101378	14	22	930	NaN	m	m/s	NaN	NaN
3	4501	101379	22	34	1721	NaN	m	m/s	NaN	NaN
4	4501	101380	34	52	1721	NaN	m	m/s	NaN	NaN
5	4501	101381	52	58	1721	NaN	m	m/s	NaN	NaN

Figure 2.2. Example of a plotted profile without apparent errors. velocityProfileMeta_ID 4501.

The GUI was created using ipywidgets, a Python library, to create text boxes that show a profile's ID, name, velocity type, method, and coordinates. Matplotlib was used to visually represent the profiles. Using this GUI, each velocity profile was examined and marked as "accepted" if there were no issues or "rejected" otherwise. As profiles were rejected, review comments were written as to why the profile was rejected. Of the database's 4223 profiles, 2667 profiles were accepted, 1514 profiles were rejected, and 42 profiles were undecided. The undecided profiles require a confirmation with their relevant source document to determine if they have no errors, but the required documents are unavailable at this time. These source documents are listed in Table 2.1. Among the 1514 rejected profiles, 2716 possible errors were found. Although review comments were fairly unique to each respective profile, they have been binned into three groups: 0 values, data entry, or literature check (Figure 2.3).

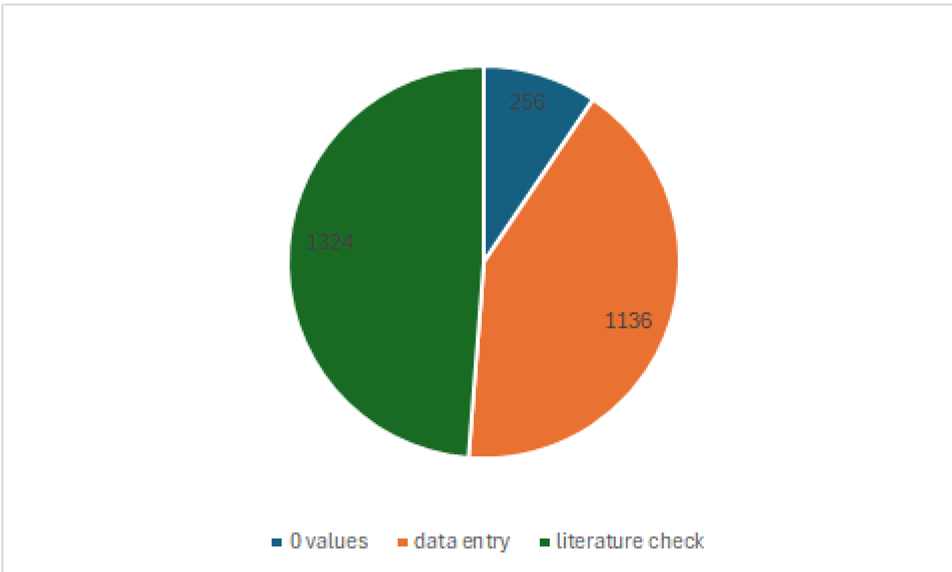


Figure 2.3. Distribution of types of review comments for rejected profiles

Among rejected profiles, the "0 values" group was assigned when a profile has an assigned velocity of 0 at one or more depths, which most often occurred at the end of the profile but occasionally occurred at seemingly random depths within the profile. Rejected profiles in the "data entry" group had one of the following problems:

1. 0 assigned as the final depth in the profile
2. Negative velocity values
3. Layer-based profiles entered as a point-based profile (or vice versa). Note that a layer-based profile is a profile for which a constant velocity is assigned over various depth

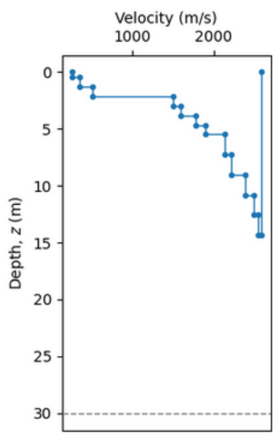
intervals (e.g., a single value over the depth range of 5-10 m, another value for 10-15 m, etc.). These are typical for downhole and surface wave tests. A point-based profile has velocities assigned at particular depths, which is typical for suspension logging.

4. Null data values (shown as “NaN” by Python)
5. Repeated data values
6. Layer-based profiles not properly layered (i.e. the final depth of the 1st layer was not the initial depth of the 2nd layer)
7. Unit issues (profiles that were incorrectly assigned as meter or feet in velocity or depth fields)
8. V_S entered as V_p (and vice-versa)

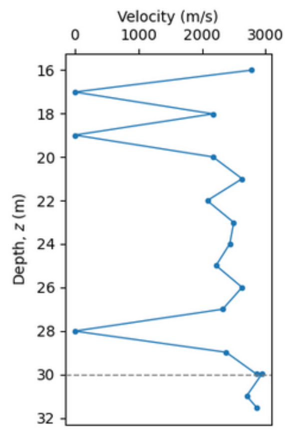
Rejected profiles in the “0 values” or “data entry” groups typically had obvious errors that could be resolved without consulting source documents. Rejected profiles in the “literature check” group may or may not have problems and the analyst (me) felt that it would be advisable to check with the source document.

Profiles were rejected if the graphical plot of the velocity profile looked unconventional or if the data appeared to contain potential errors. Examples of typical problems are shown in Figure 2.4. Part (a) shows a profile with a base depth of zero; Part (b) shows an example where three depths have zero velocity; Part (c) shows an example where one depth has a negative velocity of -10,000; Part (d) shows a table for a layer-based profile in which the base depth for each layer is given as 0; Part (e) and (f) show tables for layer-based profiles in which the top depth and bottom depth, respectively, are not provided; Part (g) shows an example of a profile with a non-physical feature. Given the amount of potential errors found, the majority (90%) of the rejected profiles were compared to profiles from source documents. Looking at the rejected profiles original source, 621 profiles were confirmed to be data entry errors that differed from what was presented in the literature. 892 profiles were found to be fine as-is.

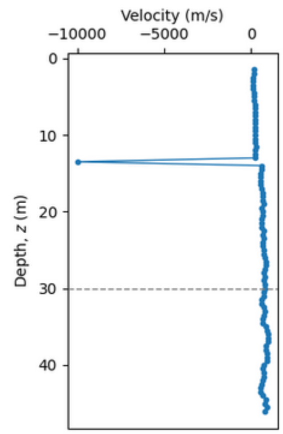
(a)



(b)



(c)



(d)

depthTop	depthBottom	value
0.00	0	326.1
2.75	0	326.1
2.75	0	293.1
5.75	0	293.1
5.75	0	311.1
9.25	0	311.1
9.25	0	530.5
13.25	0	530.5
13.25	0	649.4
17.75	0	649.4
17.75	0	707.5
30.00	0	707.5

(e)

depthTop	depthBottom	value
NaN	0.688848	243.840
NaN	9.332976	1539.240
NaN	17.285208	5733.288

(f)

depthTop	depthBottom	value
5.0	NaN	360
5.0	NaN	392
5.5	NaN	330
5.5	NaN	297
6.0	NaN	289
...	...	...
39.0	NaN	450
39.0	NaN	448
39.5	NaN	450
40.0	NaN	424
40.5	NaN	495

(g)

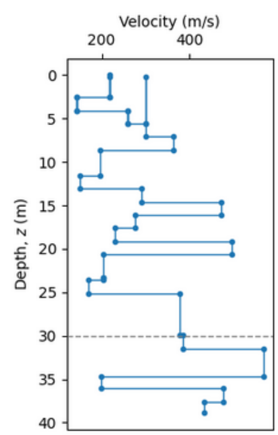


Figure 2.4. Tables and plots illustrating typical data entry errors

Returning to the list of errors below Figure 2.3, the 0 values and data entry errors 2, 4, 5, and 9 were simple data entry errors from users that uploaded the data. These were resolved by looking at the source document and entering the correct values. For data entry error 7, all the data was converted to meters for consistency. Error 3 was resolved by removing the extra depthTop entries and making sure the depthBottom was consistent with depthTop going from layer to layer. Errors 1, 3, 6, and 8 were found to be preventable with an improved schema. This is discussed in detail in Section 3.5.

Non data-entry related errors included:

- 291 profiles with missing citations
- 296 profiles with missing measurement methods

Priority was given to profiles with data-entry related errors. In cases where a profile had data entry and non data-entry related errors, non data-entry related errors were updated as well. If a profile only had non data-entry related errors, the profile was left as-is.

As noted previously, of the 1514 initially rejected profiles, 892 turned out to be accurate as originally provided. Among these 892 initially rejected profiles, two common observations (which are not mutually exclusive to a single profile) that caused the initial rejection to occur were:

1. repeated depths with different velocities
2. null, 0 m/s velocity data, or -9999 m/s velocity data

These observations were commonly found in suspension logs from Caltrans bridge sites. While data errors are clearly present, the data is consistent with the original source. The following decisions were made to reconcile these observations:

1. Repeated depths with different velocities were left as is because it is best to retain that there is uncertainty in the velocity if that is what was reported in the original source.

2. When null, 0 m/s, or -9999 m/s velocities were observed, null velocity values were concluded to be the best representation of these velocity values. Although keeping 0 m/s and -9999 m/s velocity values would be the best way to keep the data true to its original source, it is unreasonable to have these values and would incorrectly affect V_{S30} calculations. Null velocity values can be readily ignored for V_{S30} calculations.

3 Updated VSPDB Schema

3.1 INTRODUCTION

The structure for the original VSPDB was developed in 2016-2017 before the database was populated and began to be used for major projects like NGA-West3. In the intervening years, some opportunities for improvement of the VSPDB schema were identified. This chapter describes the previous VSPDB schema, highlights its shortcomings, presents a new schema and how it overcomes these shortcomings, and describes various issues related to its continued development and use.

3.2 RELATIONAL DATABASES

The VSPDB is a relational database that stores data in tables and uses structured query language (SQL) to manage the data. By organizing the data in tables, a definitive structure is established and the relationships between tables are clear. In relational databases, tables have different fields (or columns) that are used to store data. An example is the case of storing velocity data. A velocity data table is created with two main components to document the V_S profile; the measured velocity values and the depths they are measured at. Therefore, two fields that would be required are a "depth" and "velocity" field. An additional feature of fields is that datatypes can be assigned to a field. In the example, data relating to depth and velocity are real numbers that can vary in decimal digits, so the fields should be a FLOAT type. If a "name" field is used to label the profile, a VARCHAR data type would be assigned because letters are expected.

The organizational structure of a relational database is provided by its *schema*. A database's schema can be thought of as a map of the database that shows how the data are related to each other. The previous VSPDB schema is provided in Figure 3.1 (adapted from Ahdi et al. 2018). The velocity profile data, which has two tables (velocityProfileMeta and velocityProfileArray), are linked to the site, as are a series of other data types including dispersion curves, authors, and various metadata. The other data types are not directly linked to the velocity tables.

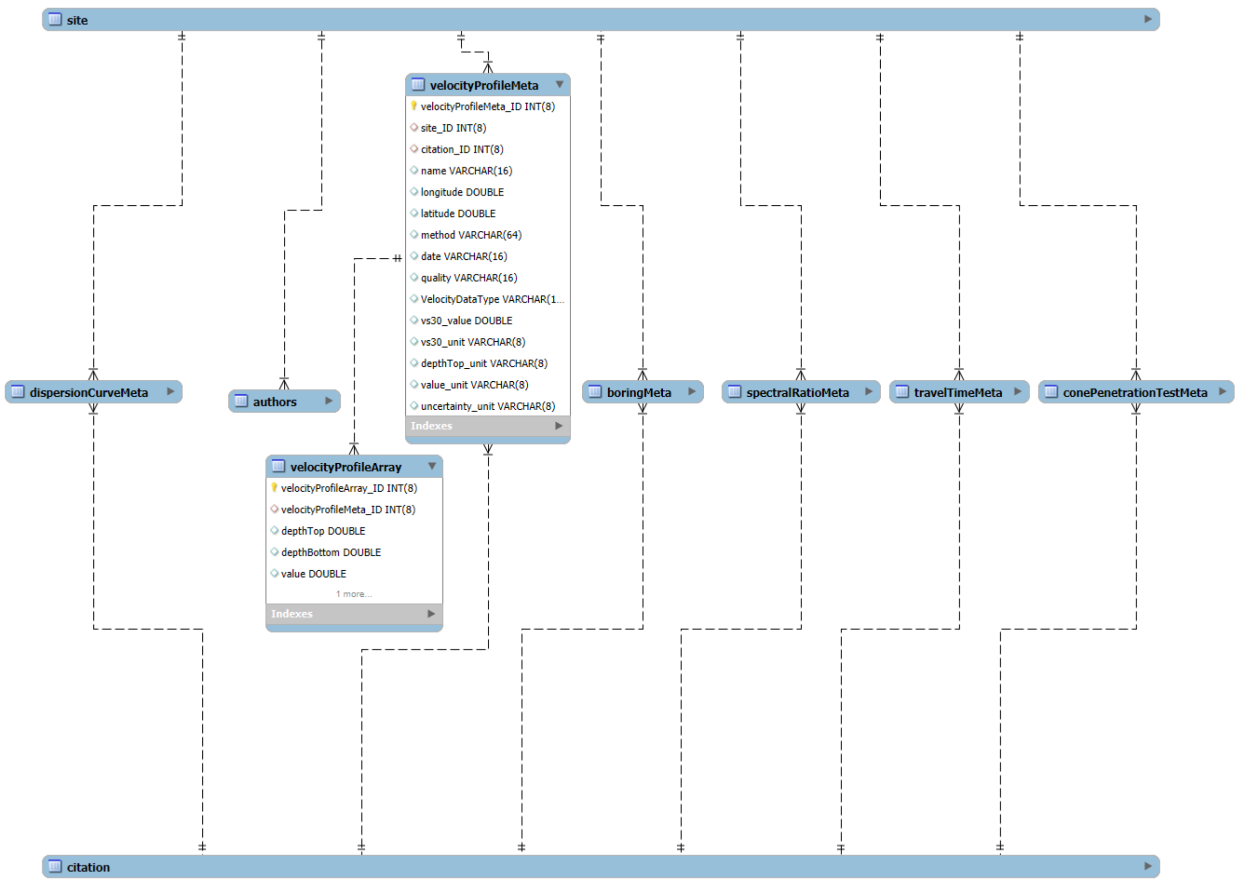


Figure 3.1. Simplified schema diagram of the old database with field names for velocity related tables

For the new database, a relational database will continue to be used as it is able to efficiently store, visualize, and upload/download data. Although the scope of this research applies to velocity profiles, VSPDB also hosts other data such as dispersion curves, boring logs, CPT soundings, travel time, and HVSR. To accommodate the variety of data and keep the relationships between the data clear, the use of a relational database is retained but the schema is modified.

3.3 SHORTCOMINGS

A major update to the database is provided in the form of an adjusted schema. The motivation for creating a new database schema comes from structural inefficiencies of the old database schema for velocity profiles and developing ways to reduce the potential for errors in entry of new data by users. As VSPDB expanded, the relationships between data became disjointed.

As noted in Section 3.2 and shown in Figure 3.1, the old schema has different data directly connected to the site table. This schema structure fails to distinguish the relationships between various data types within a single site. For instance, consider the following scenario: a SCPT is conducted at a site that includes conventional CPT sounding information (tip and sleeve resistance), V_S profile information, and travel time data. In the old schema, CPT data would go in conePenetrationTestMeta, V_S would go in velocityProfileMeta, and travel time would go in travelTimeMeta. All these metadata tables share the same site table, but it is not apparent that they come from the same SCPT. A user using this data would need to match the coordinates and citation of each metadata to see that they are from the same SCPT. To extend the scenario, consider an additional suspension logging test is performed at the same site and V_S and V_p data is collected. To add the suspension logging results to VSPDB, the V_S and V_p data would go into velocityProfileMeta. The site now has 3 velocity profiles, 2 of which are from a suspension log test and 1 is from a sCPT. From the perspective of the schema, these three velocity profiles are not related in any way and it would be up to users to compare the coordinates of the velocity profiles to find that two of the profiles have the same coordinates and are a V_S- V_p pair.

3.4 NEW SCHEMA

These issues are addressed in the new schema by introducing a tests table to relate all dependent information. In this context a "test" indicates a particular exploration and its associated elements. For example, a boring log with downhole or suspension logging that was performed in the borehole would be considered to be a single test. A boring log with similar information at different locations within the same site would have a different test ID. Figure 3.2 shows a simplified diagram of the new schema.

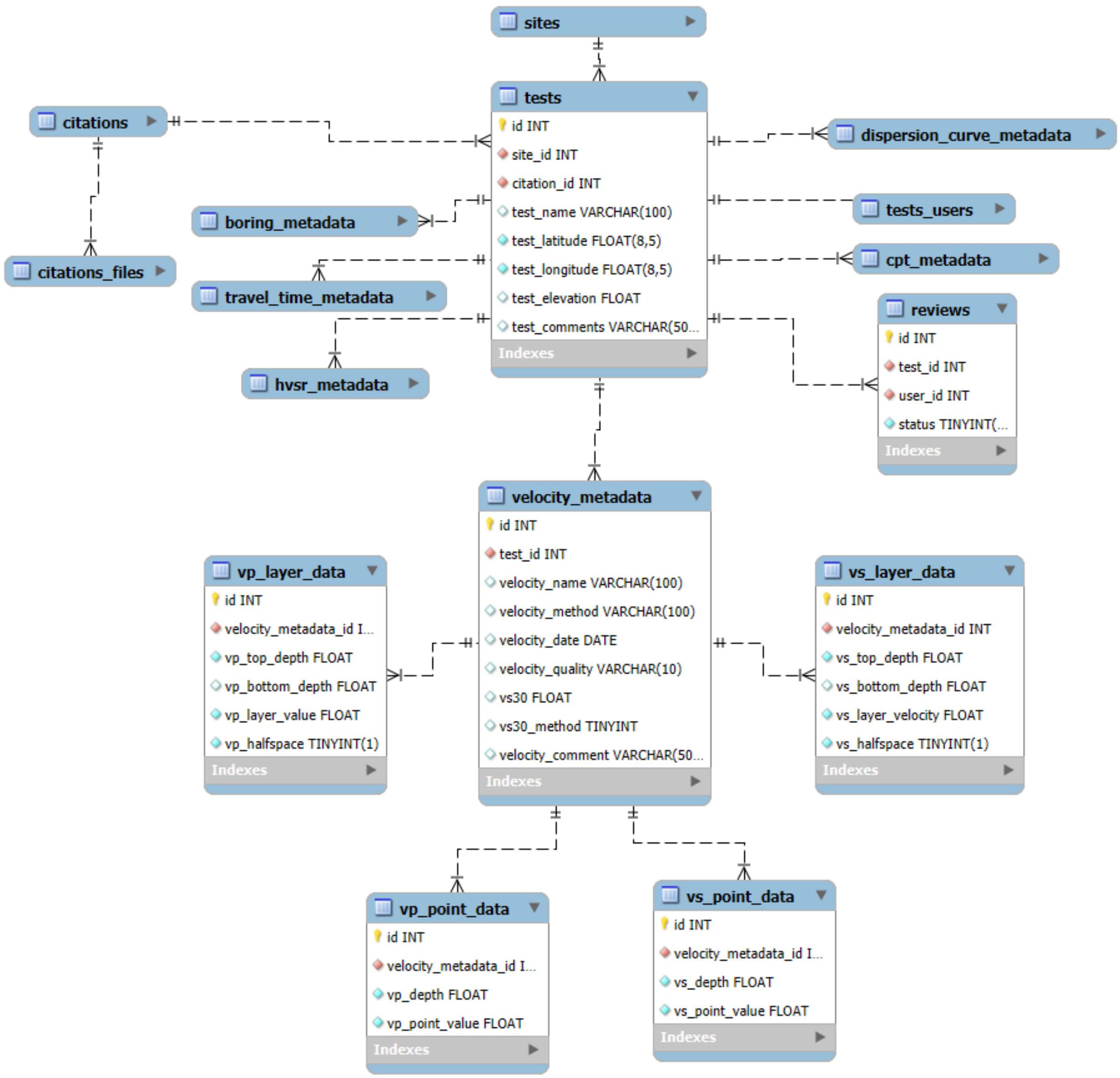


Figure 3.2. Simplified schema diagram of the new database with field names for velocity related tables

Using the test table, the issues brought up in the example scenario are solved. For the SCPT data, the cpt_metadata, velocity_metadata, and travel_time_metadata would be related to the same test. This clearly defines a relationship between the different SCPT-related data and metadata. As for the suspension log data mentioned in the example, the data would be assigned to another test at the same site. Assigned to the same test would be the borelog associated with the suspension log data. With these changes, the velocity profiles for the site are separated between two different types of tests and a clear relationship between the profiles exists. Another benefit to this schema is the way that it can handle multiple profiles inverted from the same dispersion curve. Different inversions can be assigned different velocityMetadata IDs to distinguish them from each other. However, the velocityMetadata can all be connected to the same test, which shows that the profiles are related to each other (in this case, they are derived from the same dispersion curve). A description of all the tables in the new schema is presented in Table 3.1.

Table 3.1. New schema table descriptions

Group Name	Table Name	Number of fields	Description
Miscellaneous	site_geologies	4	Surficial geologic information
	sites	11	Site information
	tests	8	Test (or exploration) information
	tests_users	3	Information about who organized the data for upload
	users	16	User information
	reviews	4	Review information
Citation	citations	4	Citation information
	files	5	File information
	citations_files	3	Join table between "citations" and "files"
Borehole Data	boring_metadata	9	Boring information
	samples	7	Sample information
	index_properties	7	Index property information (e.g., PL, LL, density)
	stratigraphies	7	Stratigraphic information
	ground_waters	4	Ground water table information
	spt_metadata	5	Standard penetration test (SPT) information
	spt_data	7	Standard penetration test (SPT) data points
	gradation_metadata	4	Gradation test information
	gradation_data	4	Gradation data points
CPT data	cpt_metadata	8	Cone penetration test (CPT) metadata information
	cpt_data	6	Cone penetration test (CPT) data points
Dispersion Curve Data	dispersion_curve_metadata	6	Dispersion curve metadata information
	dispersion_curve_data	5	Dispersion curve data points
Travel Time Data	travel_time_metadata	7	Travel time metadata information
	travel_time_data	6	Travel time data points
HVSR Data	hvsr_metadata	24	Spectral ratio (HVSR) information
	mean_curves	5	Mean spectral ratio (HVSR) ordinates
	polar_curves	6	Spectral ratio (HVSR) polar-curve ordinates
	azimuth_variations	3	Spectral ratio (HVSR) polar-curve azimuth information
	hv_processings	29	Spectral ratio (HVSR) processing information
	files_hv_processings	4	Join table between "files" and "hv_processings"
Velocity Data	velocity_metadata	9	Velocity profile metadata information
	vp_layer_data	6	Compression-wave velocity (Vp) layer-based data
	vp_point_data	4	Compression-wave velocity (Vp) point-based data
	vs_layer_data	6	Shear-wave velocity (Vs) layer-based data
	vs_point_data	4	Shear-wave velocity (Vs) point-based data

Given the large number of data-entry related errors caused by inexperienced but well-intentioned users, a reviews table was added. The reviews table has a status field that indicates if the data has been reviewed and is used to review test table data. This documents quality assurance for information in the new database. By implementing a review process, future data entry related errors are more likely to be found early (more discussion about a potential review process in Section 3.7).

A few more details were updated. The nomenclature has been updated to avoid duplicate field names across different tables. Rather than having a sites table with "*name*" as a field, we now have a sites table with "*site_name*" as a field. This makes discussing the new database more coherent and adds clarity as to which tables and fields we are referring to. Additional changes were made to some of the assigned datatypes. For fields referring to dates, we assigned a DATE datatype instead of a string to standardize date formats. In vs_layer and vs_table, a *halfspace* field is introduced that indicates which layers are halfspace layers. Halfspaces are derived from V_S profiles that are inverted from dispersion curves and used to indicate that V_S is constant below the halfspace. *Halfspace* fields can have a value of "0" or "1", "0" indicating the layer is not a halfspace and "1" indicating the layer is a halfspace.

Another significant change is the way velocity data is stored. Table 3.2. compares velocity related tables between the old and new schema.

Table 3.2. Comparison of velocity related tables between the old and new schema

Schema	Table Name	Number of Fields
Old Schema	velocityProfileMeta	5
	velocityProfileArray	6
New Schema	velocity_metadata	9
	vs_point_data	6
	vs_layer_data	4
	vp_point_data	6
	vp_layer_data	4

Compared to the old schema, where velocity data was handled by one table with fourteen fields, we now have four velocity tables with four to six fields. By having four distinct tables, users do not need to specify whether a profile is a V_S versus V_P or layer-based versus point-based. Details of the new velocity tables and their fields are provided in Table 3.3. Nullable refers to if the field can take NULL values. Halfspace fields have a default value of "0".

Table 3.3. Velocity related tables and relevant information

Table Name	Field (Column Name)	Data Type	Nullable	Default Value
vp_layer_data	id	int	NO	NULL
	velocity_metadata_id	int	NO	NULL
	vp_bottom_depth	float	YES	NULL
	vp_halfspace	tinyint	NO	0
	vp_layer_value	float	NO	NULL
	vp_top_depth	float	NO	NULL
vp_point_data	id	int	NO	NULL
	velocity_metadata_id	int	NO	NULL
	vp_depth	float	NO	NULL
	vp_point_value	float	NO	NULL
vs_layer_data	id	int	NO	NULL
	velocity_metadata_id	int	NO	NULL
	vs_bottom_depth	float	YES	NULL
	vs_halfspace	tinyint	NO	0
	vs_layer_velocity	float	NO	NULL
	vs_top_depth	float	NO	NULL
vs_point_data	id	int	NO	NULL
	velocity_metadata_id	int	NO	NULL
	vs_depth	float	NO	NULL
	vs_point_value	float	NO	NULL

3.5 MITIGATE FUTURE ISSUES

The schema was updated to not only improve on the organizational structure of the database, but also to reduce future data entry errors. Of the data entry errors presented in Section 2.2, errors 1 (0 assigned as the final depth in the profile) and 3 (layer/point based data being entered incorrectly) have a lower chance of occurring when new data is added within the framework provided by the new schema. Error 1 in particular comes from the old schema's limitation of addressing halfspaces, as noted in Section 3.4. Error 3 came from the old schema's structure of only having one table to store velocity points.

3.6 USER INTERACTION

With these changes to the schema, working with the velocity data has become simpler for users. In order to compute V_{S30} in the old schema, specific details had to be considered by users because all velocity profiles were stored in a single velocity table. If a user wanted to compute V_{S30} using the old schema, they would have to do the following: (1) query velocityProfileMeta, (2) filter for V_S profiles only, and (3) check if the data was entered as layer-based or point-based. In Step 3, if V_S data in a layer-based data structure was mistakenly entered as a point-based or vice-versa (one of the common errors mentioned in Section 2.2) it affected the calculation process. On the other hand, the steps required to compute V_{S30} are simplified with the new schema. Now users only need to query two tables, vs_layer and vs_point, where the data in both tables follow a standardized format. The only additional detail that users would need to check is if the final layer in vs_layer is a halfspace, which is noted in the *halfspace* field.

3.7 UPDATES TO OTHER ASPECTS: WEBSITE, UPLOADING, FUTURE ACCESSIBILITY

The new database will eventually replace the existing database and be accessible through https://www.vspdb.org. A development website has been set up for testing purposes, and screenshots of the development site are provided below to show how the site will work in the future. Users will be able to login if they were also a user of the old database because users' account information was migrated. Currently the website is still in the early development stage and only the velocity tables are accessible, including the relevant parent tables: sites and tests. The data can be accessed by adding the table name in front of the URL. For example, to view the data stored in the velocityMetadata table, the URL would look like “https://www.vspdb.org/velocityMetadata”. On this page, new data can be added and current data can be viewed, edited, or deleted. When viewing a velocityMetadata entry, the related velocity tables are present. Screenshots of the webpage are shown in Figure 3.3. and Figure 3.4.

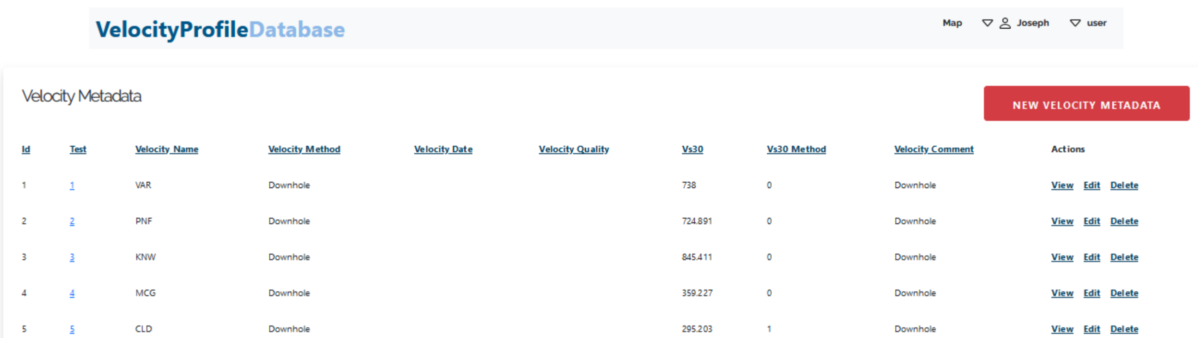


Figure 3.3. Screenshot of velocityMetadata page. New metadata can be added using the red “NEW VELOCITY METADATA”. Metadata can be interacted with underneath the “Actions” column.

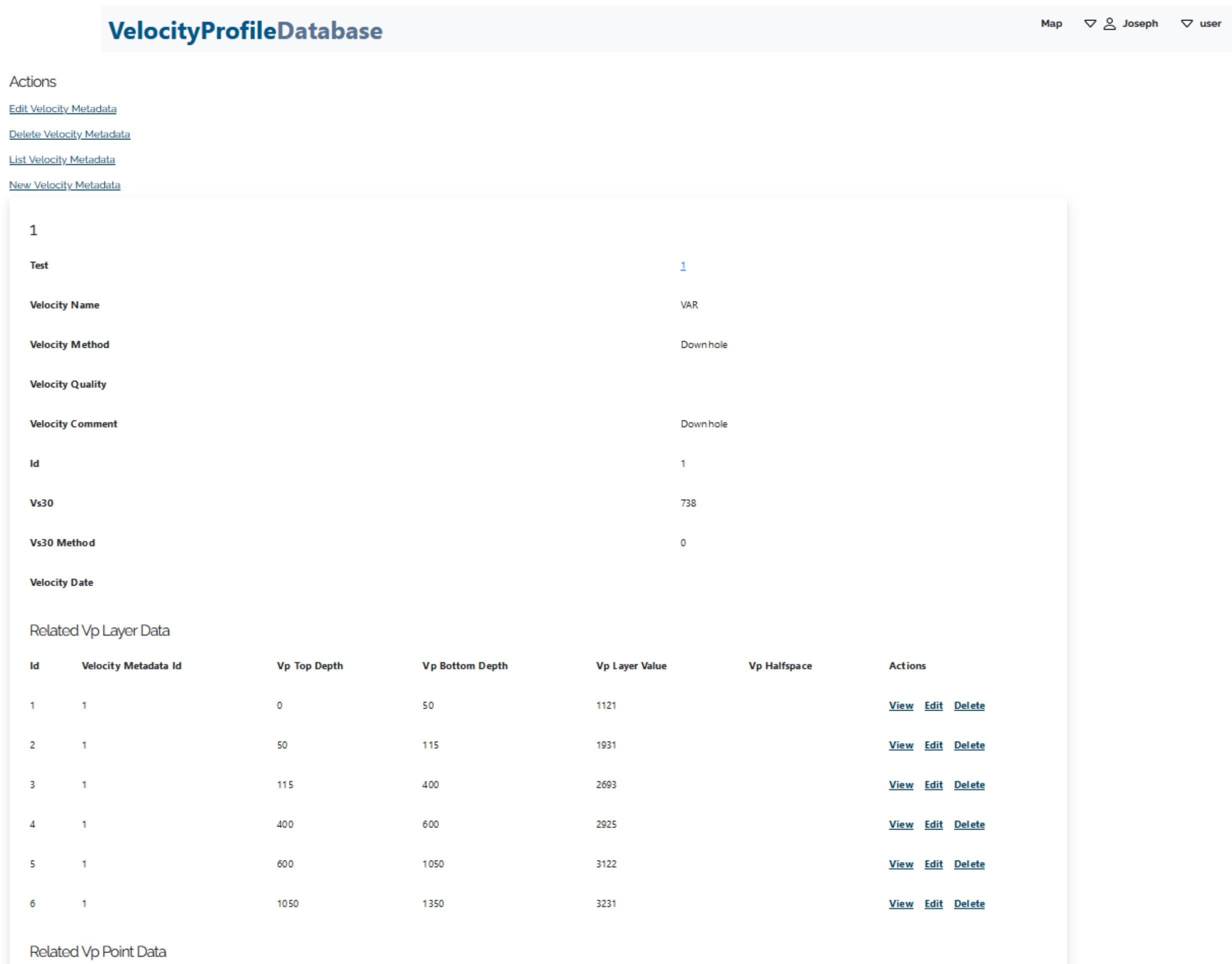


Figure 3.4. Screenshot of velocityMetadata information page. The fields in velocityMetadata can be viewed as well as any related V_S and V_p data.

Given that the website is still in the development phase, a list of modifications that still need to be implemented are the following:

1. Plotting velocity profiles
2. Adding a navigation tree so that tables can be accessed easily
3. Adding multiple data at once (expanded on below)
4. Adding a map feature that shows the locations of all the available data. Displaying all the data at once would be overwhelming, so filters will be put in place so that users can select the type of data they want to see.

To add velocity data users would need to add each point or layer one at a time. Since velocity data can span hundreds of points or have multiple layers, currently any data entry via the website would be very time consuming. For this project, data was added to the new database with a code-based solution through the backend. Having only admin people be responsible for uploading data would be unsustainable, thus adding a way for users to add data efficiently will be implemented. One way this could be achieved is by having a “upload by file” option, where users can upload a csv file. The data in the upload file would need to be organized the same way as how the schema maps the relationships between data. Users would be responsible for assigning IDs for each data and relating tables. To simplify this process, a template would be available on the website for users to download. This would not only allow users to add velocity data, but data regarding the other tables as well. When users upload velocity profiles, a code to compute V_{S30} will be implemented as well. The way in which V_{S30} will be computed based on the availability of data is presented in section 4.2. Based on my experience working with the Next Generation Liquefaction (NGL) database (Ulmer et al. 2023), uploading a csv file is convenient and the new database would greatly benefit from having this option.

Another process that is going to be implemented is a review process, as noted previously in Section 2.2. When the accuracy of the data has been confirmed by a documented review, the reviews table would indicate that the profile has been reviewed. Profiles that were reviewed would be filterable by users if they would like to use only reviewed data. All the profiles that are currently in the database indicate they have been reviewed because they have been reviewed by me.

4 Seismic Velocity Data and V_{S30} Proxy-Based Prediction Model for Utah

4.1 INTRODUCTION

One of the major applications of the VSPDB is to provide site condition information for the locations of ground motion stations for use in ground motion modeling projects like NGA-Sub (Bozorgnia et al. 2022) and NGA-West3. The primary site parameter used in these research projects is V_{S30}, as introduced in Section 1.1. Typical protocols for V_{S30} assignments (Ahdi et al. 2022) are to use a profile at the instrument site where available, which is checked using look-up functions that query the VSPDB. Such estimates carry low uncertainty (typically taken as 0.1 natural log standard deviation; Seyhan et al. 2014). For sites without a measured V_S profile, V_{S30} is estimated using proxy relationships. Such proxy relationships are most useful when they are developed using data from the target region. Relations of this type are available for a number of regions in the US including California (Wills et al. 2015), central and eastern North America (Parker et al. 2017), and the Pacific Northwest / Cascadia (Ahdi et al. 2017).

As noted in Section 1.1, the work documented in this dissertation occurred coincident with the development of the NGA-West3 project. That project is significantly enhancing the ground motion database for portions of the US that experience active tectonic region earthquakes, including California, Nevada, Arizona, New Mexico, Utah, and Idaho. Outside of California, one of the states with the greatest concentrations of earthquake data is Utah, which is the location of the highly active Wasatch Fault located near major urban centers like Salt Lake City. Prior to the present work, there was no representation of V_S profiles from Utah in the VSPDB and there were no proxy-based V_{S30} prediction models for the region that meet NGA quality standards.

To meet this need, this research included two major phases concentrated on Utah. First, we worked with the Utah Geological Survey, consulting geotechnical engineers active in the state, and Brady Cox at the Utah State University to identify 259 available V_S profiles. Second, we used this information to develop a geology and slope proxy-based V_{S30} prediction model. The first half of this chapter presents the data available in Utah and the sources they came from. The second half shows how the data was analyzed to create the regional V_{S30} proxy-based prediction model taking into account the region's geologic conditions and topographic gradients. The results were applied to assign V_{S30} to 95 ground motion stations that have produced usable ground motion recordings in the NGA-West3 project.

4.2 UTAH DATA

The majority of the velocity data comes from Utah Geological Survey's (UGS) shear-wave-velocity database (Stephenson et al. 2007), which compiled information from a combination of professional engineering reports for projects and research reports. This database contains a total of 253 V_S profiles, which come from 203 distinct sites (due to some sites having multiple profiles). Data was also obtained for four additional sites from Wong et al. (2024) and two profiles developed at or near ground motion stations (Cox 2023, Jackson 2024). The velocity data was obtained through multiple measurement methods consisting of downhole, SCPT, surface wave dispersion, and SASW. Aside from V_S profiles, additional data that the documents included are: p-wave velocity (V_p), CPT soundings, borelogs, SPT N-values, and travel time data. The data is centralized around Salt Lake City and spans from Utah Lake to the Great Salt Lake region. A summary of the data is shown in Table 4.1 and a map of the measurement locations is shown in Figure 4.1. The V_S data and other data from these sources were provided in either tabulated form or as a plot. When plots were presented, tools were used to extract the V_S and V_p, such as WebPlotDigitizer (https://automeris.io/). Dispersion curves associated with velocity data obtained through SASW methods were fully digitized. For CPT logs, tip resistance (q_c) and sleeve friction (f_s) versus depth were digitized. For borelogs, blow count and stratigraphy were digitized.

Table 4.1. Summary of sources and relevant data

Reference	Method	Vs count	Vp count	CPT count	Boring Log Count	Travel Time count	Dispersion Curve Count
AGEC, 1997, Geotechnical data for LDS Assembly Building, North Temple Street to 200 North and West Temple Street to Main Street, Salt Lake City, Utah: Unpublished consultantÆs report, project no. 973042, variously paginated.	Unspecified	1			1		
AGRA Earth and Environmental, 1997, Geoseismic and geotechnical evaluation, existing Ogden City municipal building remodel/seismic study, Ogden, Utah: Unpublished consultant's report, project no. 7-817-000786, variously paginated.	Downhole	1			1		
Bay, J.A., Gilbert, J., and Sasanakul, I., 2003, SASW testing at the Metropolitan Water District, Little Cottonwood Canyon water treatment plant, Salt Lake City, Utah: Unpublished consultant's report, 28 p.	SASW	9	9				
Chen and Associates, Inc., Date unknown, Geotechnical data for proposed 43-story tower at 400 West South Temple: Unpublished consultant's report, project no. 608U, unknown pages.	Unspecified	1					
Conetec/Bischoff	SCPT	9					
Conetec/Earthtec, 2001, CPT results, Morinda site, American Fork, Utah: Unpublished consultant's report, 6 p.	SCPT	1		1		1	
Conetec/Earthtec, 2001, CPT results, Tank center, Springville, Utah: Unpublished consultant's report, 6 p.	SCPT	1		1		1	
Conetec/Kleinfelder, 1996, Shear wave velocity and geotechnical data for the I-15 reconstruction project, Salt Lake County, Utah: Various unpublished consultantÆs reports, variously paginated	SCPT	36		36			
Conetec/Kleinfelder, 2000, Shear wave velocity and geotechnical data for the proposed Legacy Highway project, Salt Lake County, Utah: Various unpublished consultantÆs reports, variously paginated	SCPT	3					
Dames and Moore, 1996, Stage 1 I-15 seismic hazard analysis: Salt Lake City, data from unpublished consultantÆs report for Utah Department of Transportation, variously paginated.	Downhole	10			10		
LGS Geophysics Inc., 1996, Results of downhole seismic shear wave survey at the Statnamic test site, Salt Lake International Airport, Salt Lake City Utah: Unpublished consultant's report, 4 p.	Downhole	1	1				
LGS Geophysics Inc., 1998, Shear wave velocity profile for proposed American Stores office tower, 300 South and Main Street, Salt Lake City, Utah: Unpublished consultant's report, pages unknown.	Downhole	1			1		
LSG Geophysics, 1997, Results of downhole seismic shear wave survey at the Intermountain Health Care site, 2800 S. 4600 W., West Valley City, Utah: Unpublished consultantÆs report, 4 p.	Downhole	1			1		
Schuster, G.T., and Sun, Yonghe, 1993, Surface wave inversion of near surface shear velocities in Salt Lake Valley: unpublished Final Technical Report to the U.S. Geological Survey, NEHRP contract no. 1434-92-G-2175, 31 p.	Surface Wave Dispersion	69					
Tinsley, J.C., King, K.W., Trumm, D.A., Carver, D.L., and Williams, Robert, 1991, Geologic aspects of shear-wave velocity and relative ground response in Salt Lake Valley, Utah, in McCalpin, J.P., editor, Proceedings of the 27th symposium on engineering	Downhole	23	23				
Youd and Briggs (2003)	Downhole/Crosshole	2	2	1			
Bay, J.a., 2008, Shallow Shear Wave Velocity Profiling of Poorly Characterized Earthquake Site-Response Units in Urban Utah, Davis, and Weber Counties, Utah	SASW	43					43
Wong, 2024 Personal Communications		4					
USGS		4					
No Citation (Source document provided by UGS)	SASW	44					44

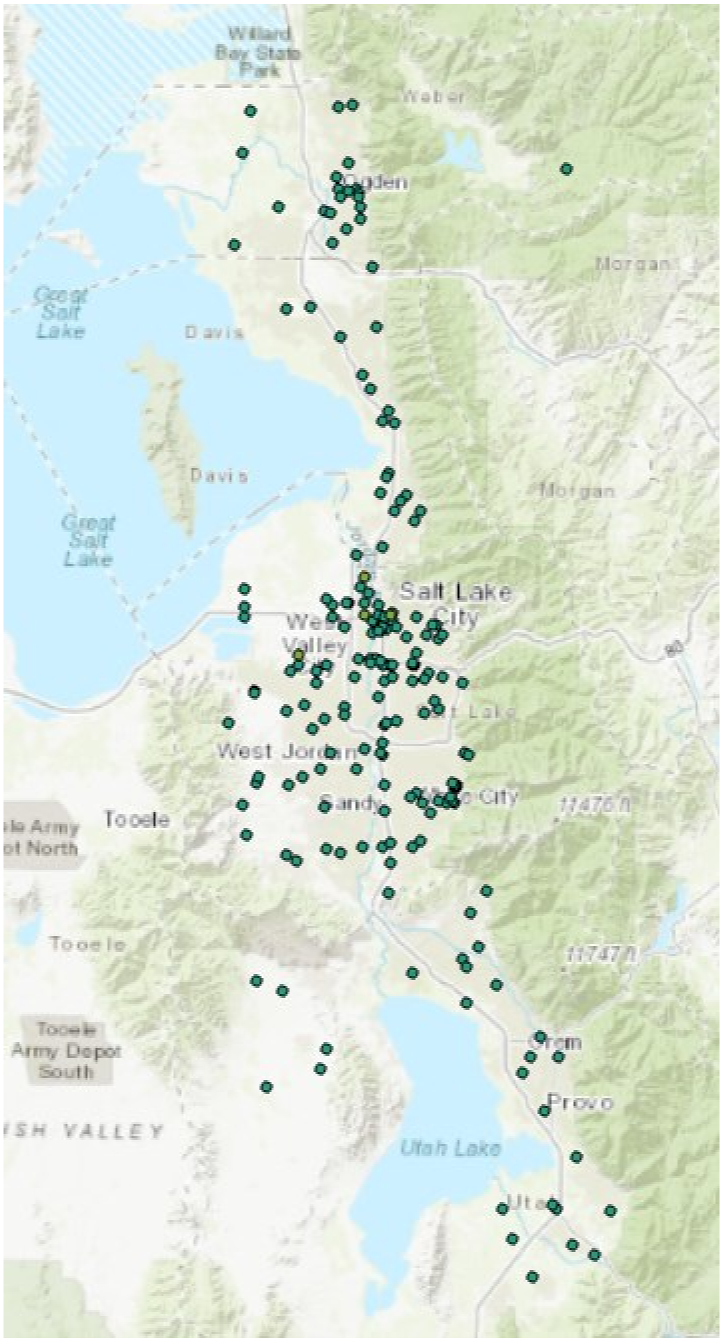


Figure 4.1. Locations of Measured Velocity Data in Utah

Using the profiles in the database, V_{S30} was calculated for each profile. V_{S30} was calculated based on available data from the profile, where the different data conditions are defined by a vs30_method code:

Table 4.2. *VS*30_method codes and descriptions

VS30_method code	Description
0	computed from a deep profile (maximum depth ≥ 30m
1	extrapolated from a shallow (maximum depth < 30m) V_S profile using Dai et al. (2013) model with region specific coefficients from Kwak et al. (2017)
2	insufficient data to compute V_{S30} (e.g., shallowest depth > 5m, deepest depth < 30m)
3	no V_S available (Vp measurement only)

For profiles assigned a code (0), V_{S30} was computed using the following equation:

$$V_{S30} = \frac{30}{\Sigma\left(\frac{\Delta z_i}{v_i}\right)} \tag{4.1}$$

where Δz_i is the layer's thickness, v_i is the layer's velocity, and the summation is across the number of layers required to reach 30 m depth. For codes (2) and (3), V_{S30} is not computed due to the lack of V_S data. For code (1), a correlation model proposed by Dai et al. (2013) is used to estimate V_{S30},

$$V_{S30} = \frac{30}{\left(tt(0,z)+\frac{30-z}{V_{S[z-30]}}\right)} \tag{4.2}$$

where $V_{S(z)}$ represents the V_S at depth $= z$ and $tt(z)$ represents travel time from surface to depth $= z$. The travel time is computed using a correlation between V_S and $logV_{S(z)30}$

$$logV_{S30} = b_0 + b_1 logV_{s(z)} \tag{4.3}$$

The model is derived from a correlation between V_{S30} and $logV_{S(z)30}$ done on datasets from California, Kik-net (Japan), and Turkey. According to a performance evaluation by Kwak et al (2017), Dai et al. (2013) models provided lower model bias and dispersion relative to measured V_{S30} values than other such models.

For Utah profiles, 248 profiles were computed by vs30_method code (0) and 11 profiles by vs30_method code (1). A histogram of V_{S30} for the Utah region is shown in Figure 4.2.

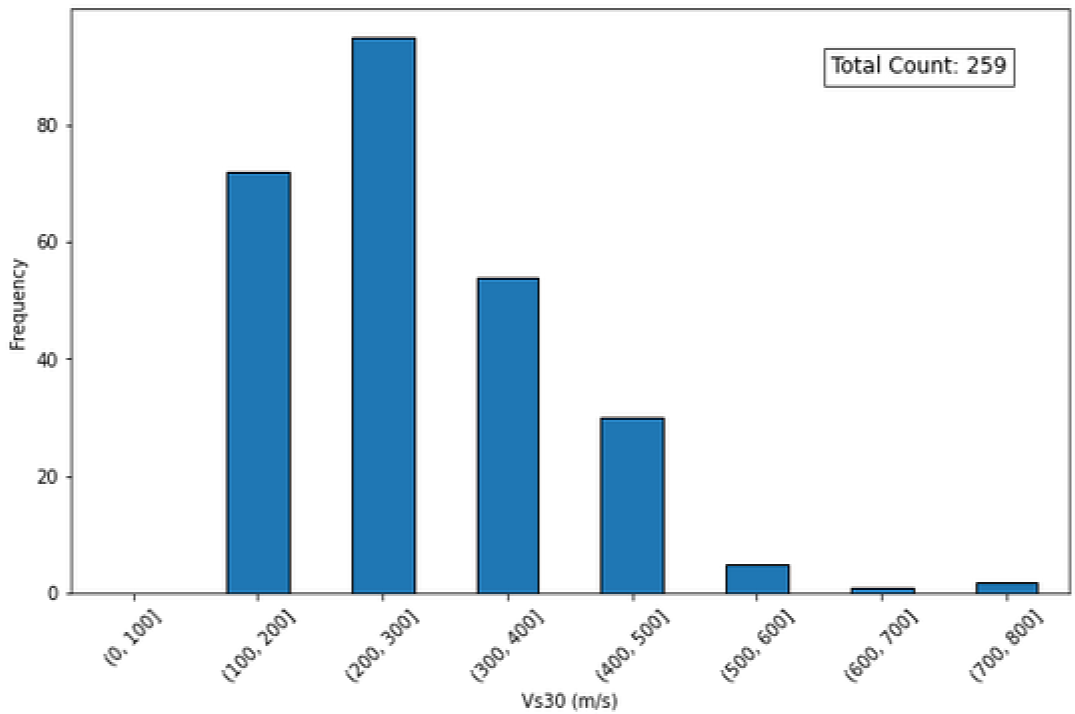


Figure 4.2. V_{S30} Histogram for V_S data in Utah

Looking at the range of computed V_{S30}, approximately 80% of the V_{S30} profiles are less than 360 m/s, indicating that the data collection was mainly done for soil sites. Although 259 profiles are available, profiles from Schuster and Sun (1993) are a combination of Rayleigh, Love, and Refraction wave profiles. Only Rayleigh profiles were considered when using profiles from Schuster and Sun (1993).

4.3 TOPOGRAPHIC GRADIENT

To determine the topographic gradient, GTOPO30, a 30-arcsec resolution digital elevation model (DEM) provided by USGS's Earth Resources Observation and Science (EROS) center was used (https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30). The DEM was provided as a raster file. To compute topographic gradients, ArcGIS Pro's 3D Analyst Tool - slope, was used to identify the slope from each cell of the raster. Other resolutions, such as 1/3-arcsec, 1-arcsec, 7.5 arcsec, and 15 arcsec were also considered. The use of very small cell sizes ultimately was not considered due to past research having found them to be less effective (Allen and Wald 2009). The 7.5, 15, and 30 arc sec results were considered in the development of proxy models. Figure 4.3 shows a histogram of topographic gradient for the 203 profile sites.

4.4 SURFACE GEOLOGY

Surface geologic conditions were obtained from geologic maps sourced from UGS (Utah Geological Survey). A combination of 7.5' and 30' x 60' maps were used, where the 7.5' maps had a resolution of 1:24,000 and the 30' x 60' maps ranged from a resolution of 1:62,000 to 1:100,000. Table 4.2 lists each of the maps that were used to look up surface geology for the 203 profile sites. The lookup process utilized the Utah Geological Survey (UGS) geologic web portal (https://geology.utah.gov/apps/intgeomap/). As shown in Table 4.2, some of the maps provided on the portal are in digital form while others are raster. The digital maps were imported to GIS software to look up geologic units for the site locations. For the raster maps, the units were looked up manually.

Table 4.3. Geologic maps used to look up surface geology for the 203 V_S profile sites.

Map	Publication Date	Number of Profiles	Map Format	Size	Resolution	Citation
Draper Quad	2018	24	GIS	7.5'	1:24,000	McKean (2018)
Midvale Quad	2024	23	RASTER	7.5'	1:24,000	McKean (2024)
Ogden Quad	2016	22	GIS	30' x 60'	1:64,000	Coogan (2016)
Plain City Quad	2012	2	GIS	7.5'	1:24,000	Harty (2012)
Provo Quad	2011	25	GIS	30' x 60'	1:64,000	Constenius (2011)
Clearfield Quad	2005	2	RASTER	7.5'	1:24,000	Sack (2005)
Roy Quad	2023	2	RASTER	7.5'	1:24,000	Sack (2023)
Rush Valley Quad	2023	3	GIS	30' x 60'	1:64,000	Clark (2023)
Salt Lake City Quad	2003	67	GIS	30' x 60'	1:100,000	Bryant (2003)
Salt Lake City South Quad	2019	40	GIS	7.5'	1:24,000	McKean (2019)
Sugar House Quad	2020	21	GIS	7.5'	1:24,000	McKean (2020)
Tooele Quad	2020	26	GIS	30' x 60'	1:64,000	Clark (2020)

Once the as-mapped surface geology was identified for each velocity location, initial geologic groupings were made in consideration of the geologic age and depositional environments. This was necessary because using the unit descriptions on the maps, 67 different geologic units were identified for the 205 sites. Geologic group names were assigned in consideration of age and depositional environment in the case of soil sites and age and rock type in the case of rock sites. When geological conditions were first assessed, 203 of the velocity profile sites were found to be Quaternary in age while only two profile sites are rock sites. The profiles in rock were not considered for this analysis due to the lack of data. Geologic age was split into three groups: (H) Holocene, (P) Pleistocene, and (Q) Quaternary. Where Q is assigned, the age was ambiguous between the Holocene and Pleistocene epochs. Depositional environment was split into eight possible categories: (a) alluvial, (c) colluvial, (d) deltaic, (e) eolian, (f) fill, (g) glacial, (l) lacustrine, or (v) mass-movement. In locations where multiple depositional environments existed, multiple letters were used to assign an environment. A “mass-movement” depositional environment is a broad category encompassing lateral spread and landslide deposits. With this procedure, a total of 19 geologic groups were formed, with ten groups having five or less measured V_S data. A histogram of geologic units is shown in Figure 4.3.

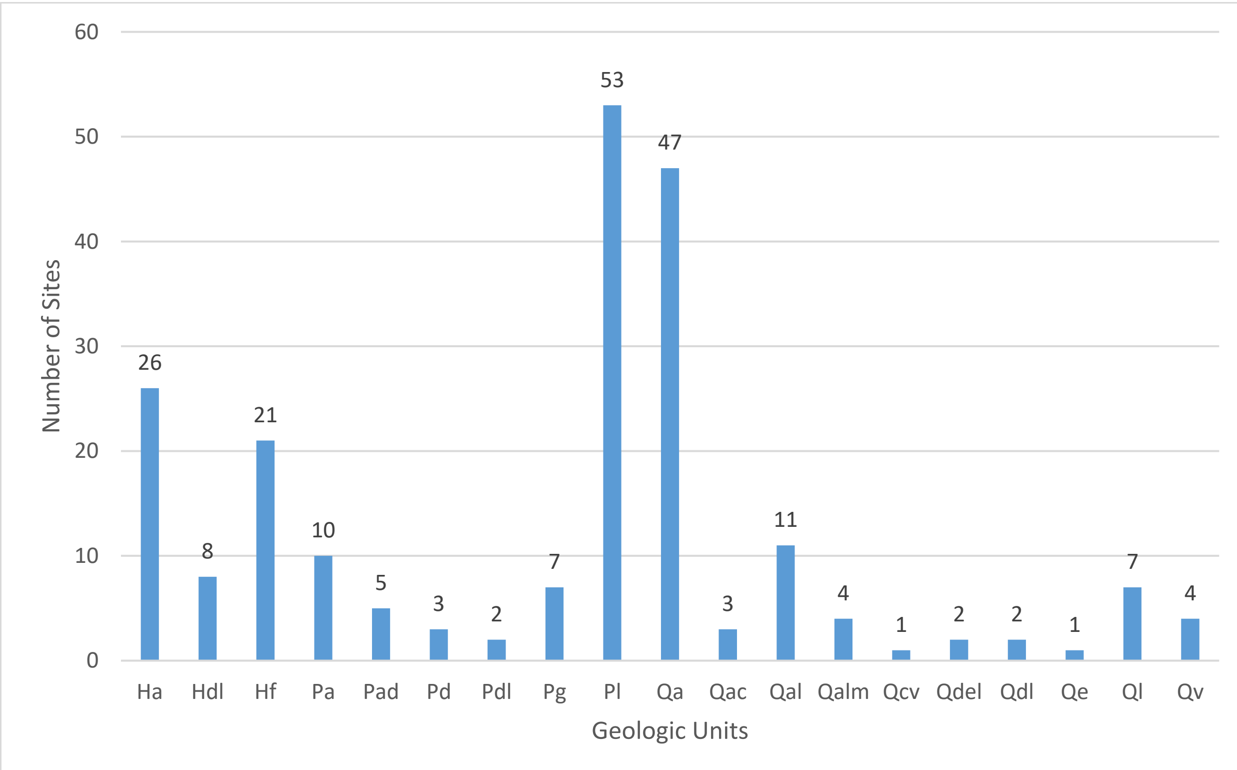


Figure 4.3. Geologic unit histogram using proposed naming convention

Many of the geologic groupings have small sample sizes, especially for Quaternary units. Depositional environments alluvium and lacustrine have the most sites with measured velocities.

4.5 REFINEMENT OF GROUPING BASED ON DATA ATTRIBUTES

Since multiple geologic maps of varying accuracy were used to assess geologic conditions, some of the groups identified in Section 4.4 may be redundant. To determine if the geologic groups have statistically unique mean V_{S30} values, I plot in Figure 4.4 the natural logarithmic means with error bars indicating ± one standard error of V_{S30} for each geologic group. Natural logarithmic means were used because V_{S30} has generally been found to be approximately lognormal (Figure 4.2). In Figure 4.4., groups Qcv and Qe have no standard error because there is only one profile in those units. Group Pdl has a large standard error range, but there are only two profiles in the group. Groups with more profiles (Ha, Hf, Pl, and Qa) have a small standard error range despite the large number of observations.

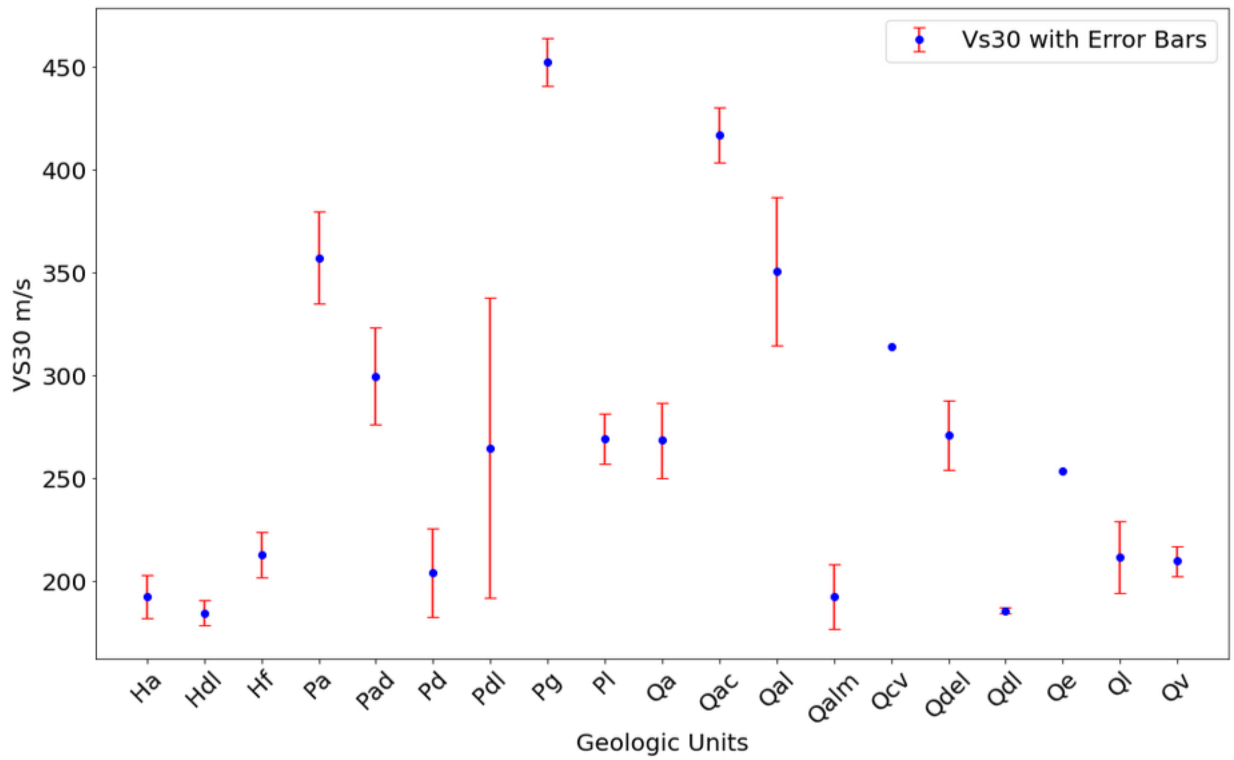


Figure 4.4. Natural log means for each geologic groups and their standard Errors shown as a range

Where multiple categories have significant overlap of the velocity ranges indicated by the standard error bounds, categories can be combined, especially if the geologic descriptions are similar (i.e., a rock unit would not be combined with Pleistocene soil). Examples of similar units in Figure 4.4 are Pa with Qal and Qa with Qdel and Qe. For a more rigorous statistical analysis to see if groups were distinct, F-tests were used. Using SciPy, a Python library, and its ‘stats.f_oneway’ function (SciPy, 2024), a one-way Analysis of Variance (ANOVA) (Snedecor and Cochran, 1989) was performed to evaluate whether different potential groups (e.g., Pa and Qal) should be retained as separate sub-models or combined into a grouped model. The calculation produces an F_1 statistic,

$$F_1 = \frac{\left[RSS_f - (RSS_1 + RSS_2)\right] / [(df_1 + df_2) - df_f]}{\sigma^2} \tag{4.4}$$

in which RSS_f is the residual sum of squares of a data set for a pair of geologic groups, RSS_i is the residual sum of squares for individual group *i*, df_i is the degree of freedom for model or submodel *i* (one if the model is mean only and two if the model includes a slope gradient term), and

$$\sigma^2 = \frac{RSS_1 + RSS_2}{N_f - (df_1 + df_2)} \tag{4.5}$$

where N_f is the number of data points in the combined data set. The F_1 statistic is compared with the F distribution to evaluate the significance level for the test. If the p-values are > 0.05, the sub-groups can be considered to be not distinct. The results of the F-tests are shown in Figure 4.5.

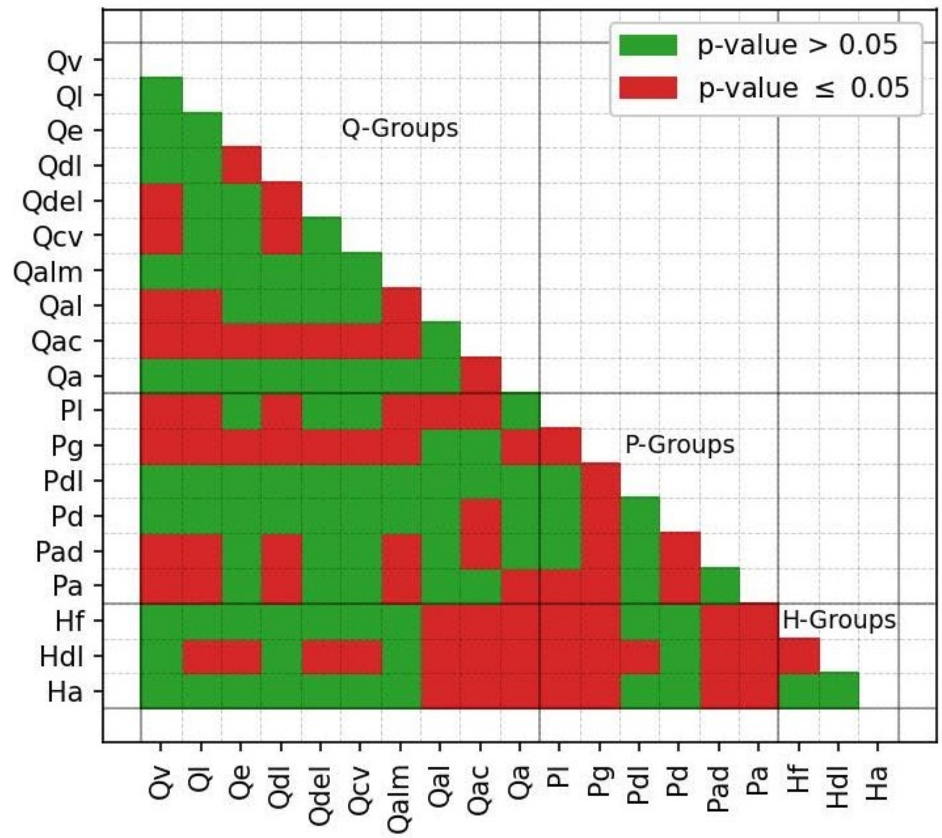


Figure 4.5. Heatmap showing which combinations of geologic units are distinct (p<0.05) or not distinct (p>0.05) based on the F1 statistic

The F-tests reveal that there are more non-distinct groups (which can be combined) than distinct groups (that need to be separated). The results of the F-tests generally conform with expectation (e.g., high p-values for Pa/Qal and Pdl/Pl indicate these groups are not distinct and can be combined).

Using both the binned mean plot (Figure 4.4) and the F-test results (Figure 4.5), geologic groups were further combined from the 19 identified in Section 4.4. In this evaluation, I gave less weight to the F-test results when the sample sizes were small (<10). This process resulted in 11 groups, five of which represent combinations of groups from Section 4.4 as follows: Pa/Qal, Pdl/Pl, Qa/Qdel/Qe, Qac/Qcv/Qv and Qalm/Qdl/Hdl. The sample sizes for the final groups are shown in Figure 4.6.

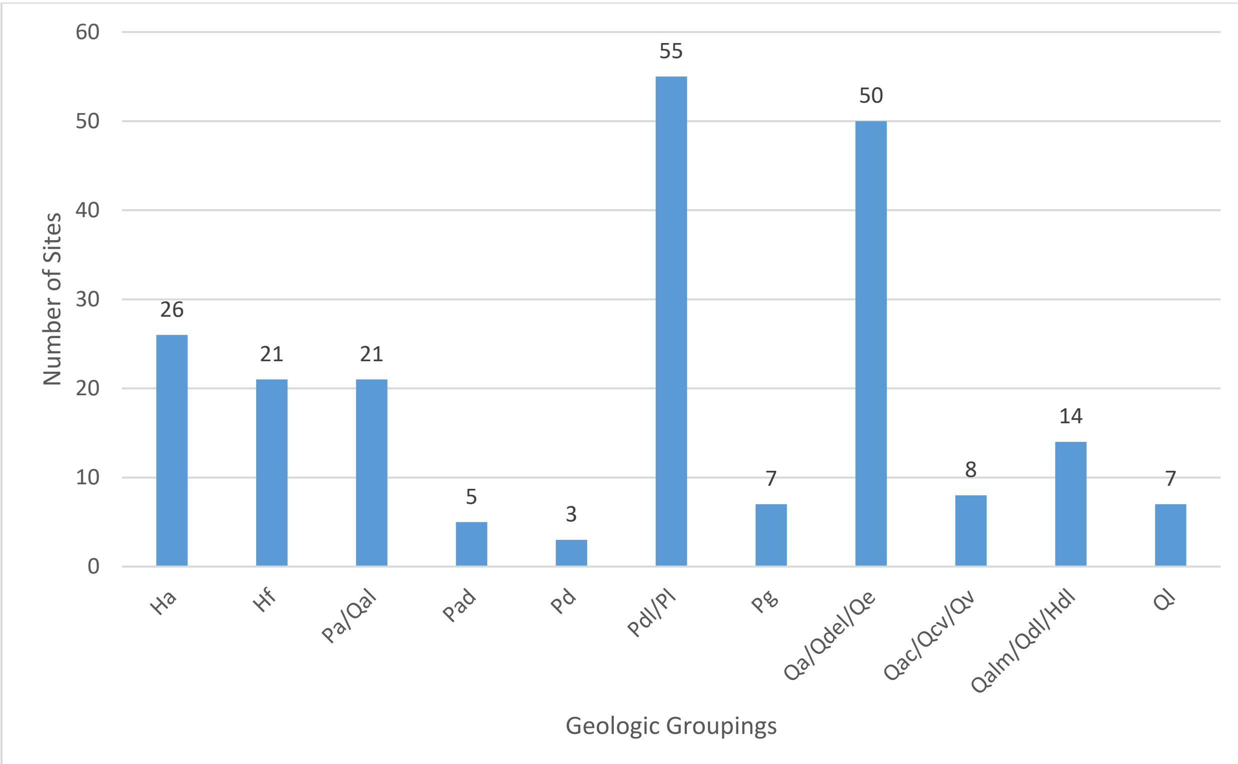


Figure 4.6. Geologic Histogram of Combined Groups and the available data for each combined group

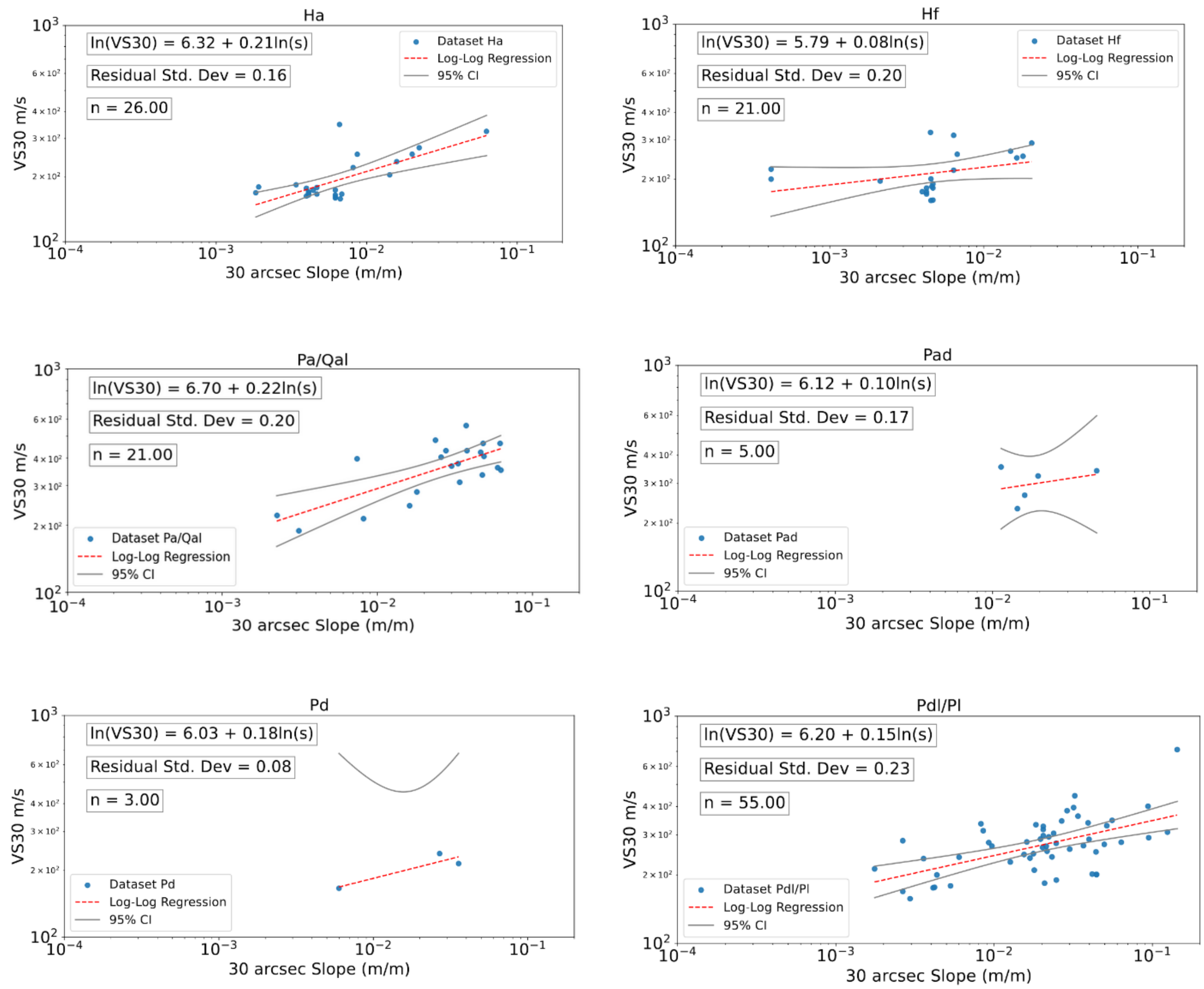
4.6 GEOLOGY-SLOPE PROXY DEVELOPMENT

Within the finalized geologic groups from Section 4.5, I examined trends between V_{S30} and 30 arc sec surface gradient (Section 4.3) using an Ordinary Least Squares (OLS) regression. The regressions were conducted with the 'statsmodel' library in Python (SciPy). Regressions were investigated in both semilog and log-log space to see which regression possessed a better visual fit,

$$\ln(V_{S30}) = c_0 + c_1 s \quad (4.6a)$$

$$\ln(V_{S30}) = c_2 + c_3 \ln(s) \quad (4.6b)$$

where V_{S30} has units of meters per second and slope is in meters per meter. The log-log regression results were selected due to their better visual fit to the data and its lower residual standard deviations. The results of the regression are shown in Figure 4.7.



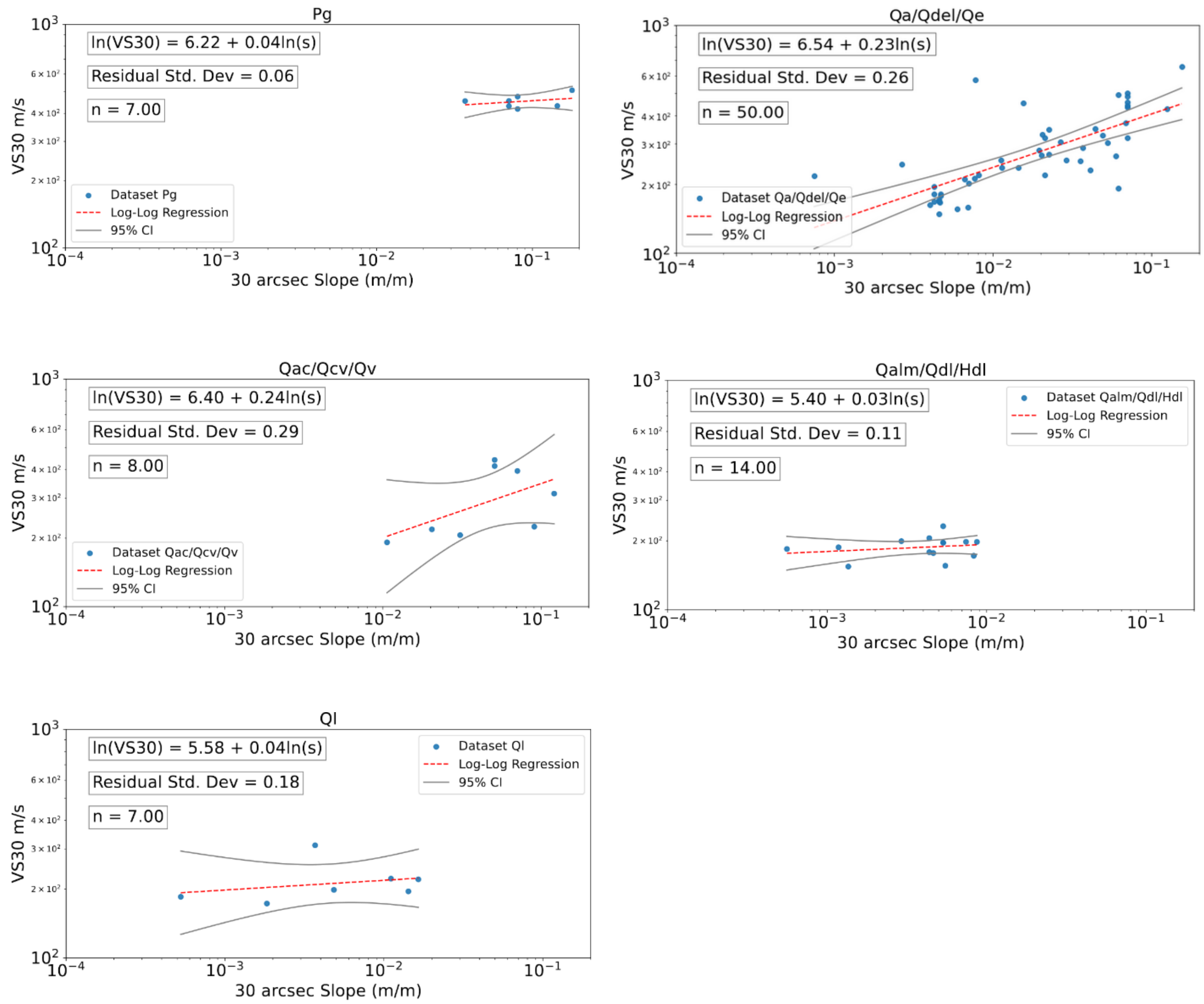


Fig 4.7. Plots of V_{S30} as a function of slope for all combined geologic groups

If the 95% confidence interval for slope parameter c_3 did not include zero, it was interpreted that slope has a statistically significant effect on V_{S30} and the ln(V_{S30}) mean estimate can be calculated using Equation (4.6b). The regression results indicate the effects of gradient on V_{S30} exist for only five geologic groupings: Ha, Hf, Pa/Qal, Pdl/Pl, Qa/Qdel/Qe. Note that these groups with gradient-dependence include all three age categories. The finding of gradient dependence of V_{S30} in geologically young units is consistent with V_{S30} proxy models developed in other regions (e.g., Wills et al. 2015, Parker et al. 2017). Gradient dependence also occurs for different depositional environments, including alluvium, lacustrine, and fill. From a visual inspection of Figure 4.7, it could be argued that all groups but Qalm/Qdl/Hdl and Ql exhibit some trends with slope of varying strength, even though the regression results say otherwise. This is mainly due to the limited data in these geologic groups. Small sample sizes lead to wider confidence intervals thus increasing the likelihood that zero will fall within the interval. My conclusion is that while a trend with slope may exist for these units, the available data is insufficient to demonstrate its statistical significance.

For the groups that did not exhibit a statistically significant trend with slope, the exponent of the natural log mean of the group is selected to represent V_{S30}. Table 4.2 summarizes the geology-slope proxy models proposed to estimate V_{S30} as well as groups that will use a mean.

Table 4.4. Summary of Proposed V_{S30} Estimation models based on geologic conditions and slope

Category		Group Moments		Gradient Relationship		
				Log-Log (Eq. 4.6b)		
Group	N	Mean V_{S30} (m/s)	Standard Deviation, $\sigma_{\ln V}$	C2	C3	Standard Deviation, $\sigma_{\ln V}$
Ha	26	192.65	0.23	6.32	0.21	0.16
Hf	21	213.10	0.22	5.79	0.08	0.20
Pa/Qal	21	345.89	0.28	6.70	0.22	0.20
Pad	5	299.77	0.18			
Pd	3	204.25	0.19			
Pdl/Pl	55	269.19	0.27	6.20	0.15	0.23
Pg	7	452.74	0.07			
Qa/Qdel/Qe	49	271.08	0.38	6.54	0.23	0.26
Qac/Qcv/Qv	8	285.64	0.35			
Qalm/Qdl/Hdl	14	187.15	0.11			
Ql	7	211.65	0.19			

4.7 ASSIGNING V_{S30} TO GROUND MOTION STATIONS

Of the 95 ground motion station sites in the NGA-West3 project, only 3 have a V_{S30} value derived from an on-site profile. For those three sites, the profile was used. One of the three profiles at ground motion sites is a suspension log performed by GEOVision Geophysical Services in August of 2003 (Jackson 2024). For the other 92 sites, the use of proxy-based models are required. This was done using a combination of the regional proxy for Quaternary units as presented in Section 4.6 and proxies for other regions for rock geologic units.

For the Quaternary units, the proposed proxy models for V_{S30} estimation are presented in Table 4.2 and Eq. (4.6b). To apply the models, surface geology and surface gradient were looked up using the same map resources described in Sections 4.3-4.4 for the ground motion sensor site locations. Of the 92 sites requiring a proxy-based estimate, 42 are on Quaternary units and were evaluated using these procedures.

Many of the ground motion stations are on rock sites, with 50 of the 92 stations requiring proxy-based estimates being pre-Quaternary. The geologic era distribution for ground motions stations on rock sites is shown in Figure 4.8.

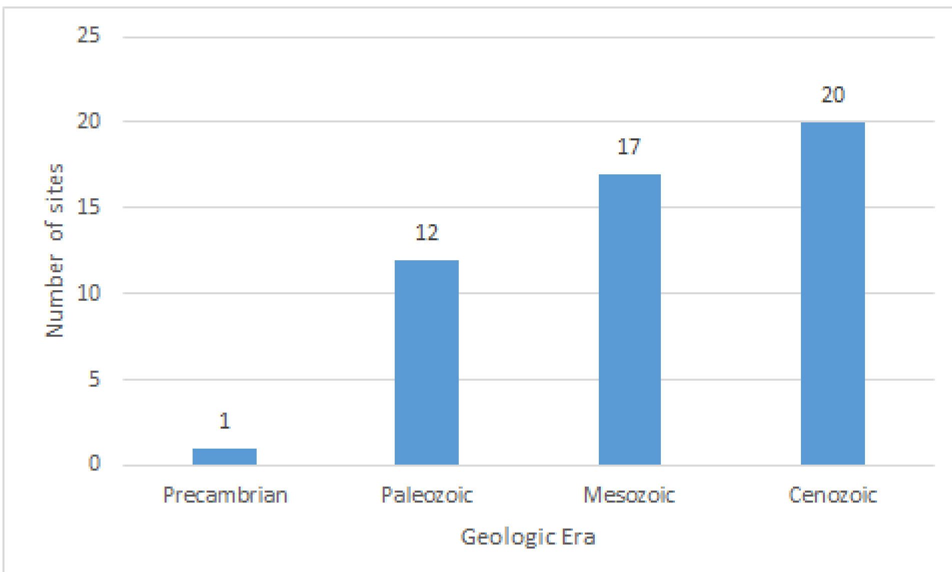


Figure 4.8. Geologic era histogram for ground motion stations on rock sites

Due to the lack of measured velocity data for rock sites in Utah, local proxy relationships could not be developed and mean V_{S30} for geologic units of different age were borrowed from California and central and eastern North America (Wills et al. 2015 and Parker et al. 2017, respectively). Wills et al. and Parker et al. do not provide V_{S30} means for every geologic unit that exists for Utah, which is why both regions were required. Wills et al. (2015) was used to estimate median V_{S30} for Cenozoic and Mesozoic units and Parker et al. (2017) was used to estimate median V_{S30} for Paleozoic and Precambrian sites. A summary of the borrowed rock units and mean V_{S30} values is shown in Table 4.4. Of the 12 stations located on Paleozoic Era rock, one had a co-located V_S profile that was assigned to the station (V_{S30} = 360.82 m/s).

Table 4.5. Description of how geologic units from Parker et al. 2017 and Wills et al. 2015 were used to assign mean V_{S30} to ground motion stations.

Geologic Era	Source of Borrowed Value	Source Geologic Units	Description	V_{S30} (m/s)
Cenozoic	Wills et al. 2015	Tss	Tertiary Sandstone	468.4
		Tsh	Tertiary Shale	385.1
		Tv	Tertiary Volanic	518.9
Mesozoic	Wills et al. 2015	Kss	Cretaceous Sandstone	502.5
Paleozoic	Parker et al. 2017	Group 15	Non-glaciated Paleozoic related formations	684
Precambrian	Parker et al. 2017	Group 17	Formations from Precambrian Era	699

Figure 4.9 shows a histogram of the assigned V_{S30} values for the Utah stations. Table 4.6 provides the following for each NGA-West3 recording station site in Utah: site ID as used in the ground motion database, mean V_{S30} value, $\sigma_{\ln V}$, and assignment code (0 or 1 for sites with profiles, 2 for Quaternary sites, 3 for pre-Quaternary sites). For the 42 sites with V_{S30} estimated from Sec 4.6, $\sigma_{\ln V}$ is as given in Table 4.3. For the rock sites, the $\sigma_{\ln V}$ is derived using the standard deviations for the respective units in the source documents (Wills et al. 2015 and Parker et al. 2017) combined with additional epistemic uncertainty as described in Ahdi et al. (2022).

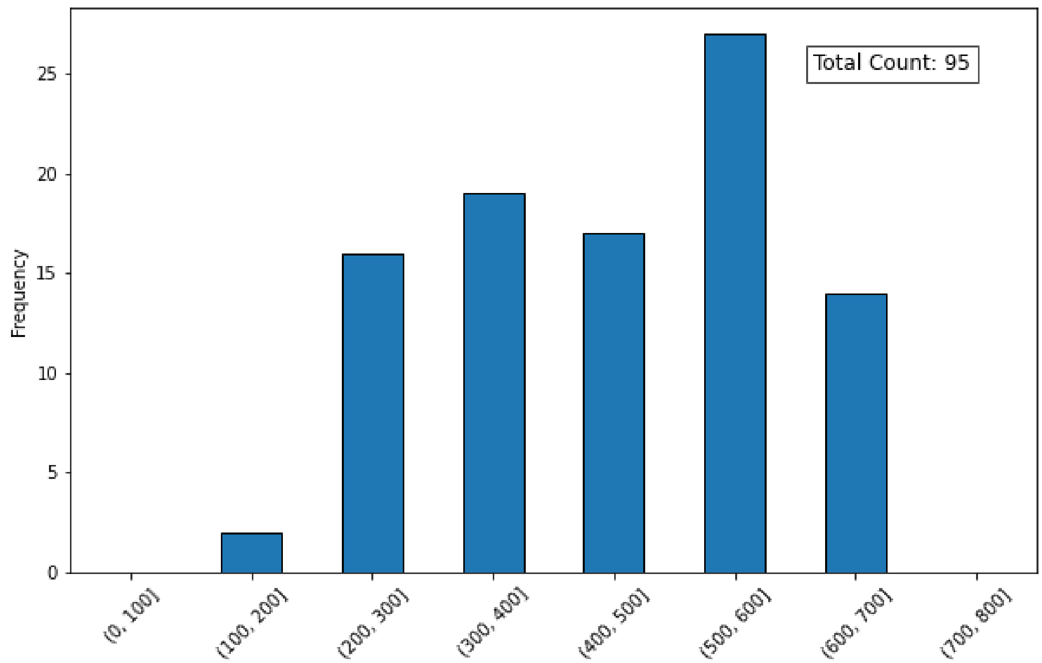


Figure 4.9. V_{S30} Histogram for Assigned GMPD Ground Motion Stations

Table 4.6. Assigned V_{S30} and $\sigma_{\ln V}$ for Utah Stations

site_id	V_{S30}	$\sigma_{ln(V)}$	Assignment Code
4723	518.9	0.39	3
4724	468.4	0.5	3
4725	385.1	0.39	3
7964	258.55	0.26	2
7974	468.4	0.5	3
8016	502.5	0.5	3
8017	502.5	0.5	3
8087	197.84	0.27	2
8660	502.5	0.5	3
8670	288.28	0.26	2
8705	234.48	0.27	3
8706	502.5	0.5	3
8707	502.5	0.5	3
8708	468.4	0.5	3
8709	502.5	0.5	3

8710	518.9	0.39	3
8711	502.5	0.5	3
8712	502.5	0.5	3
12040	502.5	0.5	3
12056	475.64	0.26	2
12078	502.5	0.5	3
12082	518.9	0.39	3
12102	684	0.64	3
12104	684	0.64	3
12105	684	0.64	3
12106	684	0.64	3
12108	684	0.64	3
12109	352.23	0.26	2
12110	331.12	0.26	2
12111	699	0.87	3
12113	300.9	0.36	2
12114	502.5	0.5	3
12115	684	0.64	3
12116	385.1	0.39	3
12117	486.86	0.22	2
12118	360.82	0.1	3
12119	468.4	0.5	3
12120	347.88	0.22	2
12121	293.82	0.23	2
12122	222.35	0.26	2
12123	251.52	0.26	2
12124	358.48	0.26	2
12125	186.42	0.16	2
12126	366	0.22	2
12127	202.4	0.22	2
12128	165.67	0.1	0
12129	684	0.64	3
12131	502.5	0.5	3
12132	468.4	0.5	3
12211	285.64	0.35	0
12212	296.95	0.26	2
12213	227.78	0.16	2
12214	300.9	0.36	2
12253	502.5	0.5	3
12254	502.5	0.5	3
12530	468.4	0.5	3

12531	468.4	0.5	3
12532	401.47	0.23	2
12533	684	0.64	3
12653	316.07	0.27	2
12654	312.2	0.27	3
14266	518.9	0.39	3
14280	518.9	0.39	3
14281	385.1	0.39	3
14282	518.9	0.39	3
14283	518.9	0.39	3
14284	354.96	0.22	2
14285	256.7	0.22	2
14339	402.55	0.22	2
14569	444	0.56	3
14570	468.4	0.5	3
14575	502.5	0.5	3
14576	468.47	0.22	2
14577	502.5	0.5	3
14583	349.33	0.22	2
14584	561.06	0.22	2
14590	319.18	0.22	2
14591	518.9	0.39	3
14592	341.55	0.16	2
14593	351.5	0.16	2
14596	205.42	0.16	2
14597	326.1	0.22	2
14598	255.89	0.16	2
14621	369.53	0.16	2
14622	254.85	0.26	2
14623	283.62	0.23	2
14632	245.41	0.23	2
14635	211.65	0.19	0
15053	393.28	0.22	2
15054	684	0.64	3
15055	684	0.64	3
15060	684	0.64	3
15065	502.5	0.5	3
15066	368.43	0.16	2
15067	419.31	0.16	2

5 Conclusion

5.1 SCOPE OF RESEARCH

This thesis goes through the process of screening VSPDB and rectifying errors found in the data. Once the data was cleaned up, it was migrated to a different database with a new schema that improves on the organizational structure of VSPDB. New data was added to the database for sites in Utah, which was then used to develop a V_{S30} proxy model conditioned on topographic gradient and surface geology. Using the proxy models, V_{S30} was assigned to ground motion stations.

5.2 MAJOR FINDINGS

Following a screening of the velocity profiles in the database, approximately 621 out of 4225 (~15%) profiles were found to have data entry errors. These data entry errors included:

1. 0 assigned as the final depth in the profile
2. Negative velocity values
3. Layer-based profiles entered as a point-based profile (or vice versa)
4. Null data values
5. Repeated data values
6. Layer-based profiles not properly layered
7. Unit issues
8. V_S entered as Vp (and vice-versa)

Error 4 was resolved by setting the bottom depth of the layer equal to the depth of the previous layer. The other errors required reviewing original source documents to resolve them.

The organization structure of VSPDB has been improved by updating the schema by including a test table that allows multiple data types to be properly linked to each other. The new schema was implemented by creating an entirely new version of the database. All V_S data from the original database has been migrated to the new database.

Using new data added to the database, a slope-geology proxy model was developed and used to estimate V_{S30} for ground motion stations in Utah. In cases where trends with slope were not statistically significant, the median V_{S30} of the geologic group was used. Amongst the geologic groupings created, 13 out of 19 geologic groups were found to have indistinct natural log mean V_{S30} and were simplified to a single group. The finalized groups were

1. Ha
2. Hf
3. Pa/Qal
4. Pad
5. Pd
6. Pdl/Pl
7. Pg
8. Qa/Qdel/Qe
9. Qac/Qcv/Qv
10. Qalm/Qdl/Hdl
11. Ql

Out of the finalized groups, (1) Ha, (2) Hf, (3) Pa/Qal, (6) Pdl/Pl, and (8) Qa/Qdel/Qe were statistically determined to have trends with slope. As for the other groups, there was not enough data to determine if a slope with trend existed and exponent of the natural means of V_{S30} were used as the representative V_{S30} for the group.

Applying the proposed models to assign V_{S30} for ground motion stations in Utah, it was found that the models were able to be used for approximately 47% of stations. Of the stations in which models were applied, approximately 91% of stations utilized the linear regression models to estimate V_{S30}. For the 50 stations with rock geology, V_{S30} was assigned from rock formations in the same era from Wills et al. (2015) and Parker et al. (2017). The two measured velocity profiles in rock have lower V_{S30} values than the borrowed median velocities.

5.3 FUTURE WORK

The findings in this thesis suggest that future work can be done to enhance the database and develop more accurate V_{S30} estimate models. Regarding the database, I recommend receiving feedback from VSPDB users to improve user interaction with the website. Although more features are already planning to be implemented, the interaction thus far with the database is from a developer's point of view. Once VSPDB users have had time to use the website, they can offer practical suggestions that would improve the overall experience.

Before the new database can be accessed, work needs to be done to migrate the remaining data. The remaining data is: HVSR, travel time, dispersion curves, borings, and CPTs. The work that needs to be done is assigning proper test IDs to the remaining data. For test IDs created for the velocity data, there are some relationships between the current tests and the remaining data that need to be established. Some examples include linking the correct dispersion curve to the derived velocity profiles and assigning the correct borelog to a suspension logging test.

Another recommendation is to collect more measured velocity data and investigate trends with geology and slope using the procedure in this thesis. One of the limiting factors to using statistical methods to analyze geologic groups was the small sample size. The accuracy and range of the proxy models developed to estimate V_{S30} can be greatly improved with additional measured velocity data. With more data, it would be clearer to identify which geologic groups truly have no trend with slope. Data collection over rock sites is especially recommended. Only two profiles over rock exist, whereas over 50% of ground motion stations are over rock. To accurately assign V_{S30} to these stations, measured velocities from rock sites are required.

REFERENCES

Ahdi S.K., Kwak D.Y., Ancheta T.D., Contreras V., Kishida T., Kwok A.O., Mazzoni S., Ruz F., Stewart J.P. (2022). Site parameters applied in NGA-Sub database. *Earthquake Spectra*, **38(1)**, 494-520. doi:10.1177/87552930211043536

Ahdi S.K., Sadiq S., Olhan O., Bozorgnia Y., Hashash Y.M.A., Kwak D.Y., Park D., Yong Y., Stewart J.P. (2018). "Development of a United States community shear wave velocity profile database," in Geotechnical Engineering and Soil Dynamics V: Seismic Hazard Analysis, Earthquake Ground Motions, and Regional-Scale Assessment, June 10-13, 2018, Austin, TX, *ASCE Geotechnical Special Publication* No. 291, S.J. Brandenberg and M.T. Manzari (eds.), 330-339.

Ahdi S.K., Stewart J.P., Ancheta T.D., Kwak D.Y., Mitra D. (2017) Development of Profile Database and Proxy-Based Models for Prediction in the Pacific Northwest Region of North America. *Bulletin of the Seismological Society of America 2017*, **107 (4)**, 1781–1801. doi:https://doi.org/10.1785/0120160335

Allen T.I., 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 2019*, **99 (2A**): 935–943. doi:https://doi.org/10.1785/0120080255

Bozorgnia Y., Abrahamson N.A., Ahdi S.K., Ancheta T.D., Atik L.A., Archuleta R.J., Atkinson G.M., Boore D.M., Campbell K.W., Chiou B.S.J., Contreras V., Darragh R.B., Derakhshan S., Donahue J.L., Gregor N., Gulerce Z., Idriss I.M., Ji C., Kishida T., Kottke A.R., Kuehn N., Kwak D., Kwok A.O., Lin P., Macedo J., Mazzoni S., Midorikawa S., Muin S., Parker G.A., Rezaeian S., Si H., Silva W.J., Stewart J.P., Walling M., Wooddell K., Youngs R.R. (2022). NGA-Subduction research program. *Earthquake Spectra*, **38(2)**, 783-798. doi:10.1177/87552930211056081

Bryant B. (2003) Geologic map of the Salt Lake City 30' x 60' quadrangle, north-central Utah and Uinta County, Wyoming. M-190dm. UGS. 1:100,000 scale.

Buckreis T. E., Stewart J. P., Shams R. R., Mohammed S., Meibai L., Kishida T., Lanzano G., Luzi L., Zimmaro P., Bradley B. A., Huang J., Nweke C. C., Brandenberg S. J., Donahue J. L., Bozorgnia Y. (2024). Overview of the NGA-West3 Ground Motion Database. Poster Presentation at 2024 SCEC Annual Meeting.

Clark D.L. (2020) Geologic Map of the Tooele 30' x 60' Quadrangle, Tooele, Salt Lake, and Davis Counties, Utah. M-284DM. UGS. 1:62,000 scale.

Clark D.L. (2023) Geologic Map of the Rush Valley 30' X 60' Quadrangle, Tooele, Utah, and Salt Lake Counties, Utah. M-294DM. UGS. 1:62,000 scale.

Constenius K.N. (2011) Interim geologic map of the Provo 30' x 60' quadrangle, Utah, Wasatch, and Salt Lake Counties, Utah. OFR-586dm. UGS. 1:62,000 scale.

Coogan J.C. (2016) Interim geologic map of the Ogden 30' x 60' quadrangle, Box Elder, Cache, Davis, Morgan, Rich, and Summit Counties, Utah, and Uinta County, Wyoming. OFR-653dm. UGS. 1:62,000 scale.

Cox B.R. (2023) Deep Shear Wave Velocity Profiling Using MASW and MAM Surface Wave Methods: Valley Emergency Communication Center, Salt Lake City, Utah

Harty K.M. (2012) Geologic map of the Plain City quadrangle, Weber and Box Elder Counties, Utah. M-253DM. UGS. 1:24,000 scale.

Jackson T.S. (2024) "Evaluating 1D and 2D Small-Strain Ground Response Analyses at the I-15 Downhole Array Using Recorded Aftershocks from the M5.7 2020 Magna, Utah Earthquake". https://digitalcommons.usu.edu/etd2023/197

Kramer, S.L., & Stewart, J.P. (2024). *Geotechnical Earthquake Engineering* (2nd ed.). CRC Press. https://doi.org/10.1201/9781003512011

McKean A.P. (2018) Interim geologic map of the Draper quadrangle, Salt Lake and Utah Counties, Utah. OFR-683DM. UGS. 1:24,000 scale.

McKean A.P. (2019) Geologic Map of the Salt Lake City South Quadrangle, Salt Lake County, Utah. M-283DM. UGS. 1:24,000 scale.

McKean A.P. (2020) Geologic Map of the Sugar House Quadrangle, Salt Lake County, Utah. M-285DM. UGS. 1:24,000 scale.

McKean A.P. (2024) Interim Geologic Map of the Midvale Quadrangle, Salt Lake County, Utah. OFR-761. UGS. 1:24,000 scale.

Parker G.A., Harmon J.A., Stewart J.P., Hashash Y.M.A., Kottke A.R., Rathje E.M., Silva W.J., Campbell K.W. (2017) Proxy-Based VS30 Estimation in Central and Eastern North America. *Bulletin of the Seismological Society of America 2017*, **107 (1)**, 117–131. doi:https://doi.org/10.1785/0120160101

Sack D.I. (2005) Geologic map of the Clearfield 7.5' quadrangle, Davis County, Utah. MP-05-4. UGS. 1:24,000 scale.

Sack D.I. (2005) Geologic map of the Roy 7.5' quadrangle, Weber and Davis Counties, Utah. MP-05-3. UGS. 1:24,000 scale.

Seyhan E., Stewart J.P., Ancheta T.D., Darragh R.B., Graves R.W. (2014) NGA-West2 Site Database. *Earthquake Spectra*, **30(3)**, 1007-1024. doi:10.1193/062913EQS180M

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

Stephenson, W.J., Williams, R.A, Odum, J.K., and Worley, D.M. (2007) Miscellaneous high-resolution seismic imaging investigations in Salt Lake and Utah Valleys for earthquake hazards: Utah Geological Survey

Wills C. J., Gutierrez C. I., Perez F. G., 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. doi: https://doi.org/10.1785/0120150105

Wong I., Darragh R.B., Wu Q., Chowdhury I. (2024) Analyses of site effects in the 2020 M 5.6 Magna, Utah, Earthquake and development of shallow site amplification factors for Salt Lake Valley, Report to US Geological Survey, NEHRP Grant G22AP00234.