Improving Site Response Analysis for Earthquake Ground Motion Modeling

Thesis by
Jian Shi

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2019
Defended April 3rd, 2019

Jian Shi
ORCID: 0000-0002-1969-7579

Acknowledgements

First and foremost, I would like to thank my advisor, Professor Domniki Asimaki, for her guidance, support, and patience over the years. I have learned so many things from her, and the most important one of them is the ability of critical thinking: from the formulation of theories to the designing of experiments and ultimately to the rigorous validation of results. Additionally, she taught to me the importance of grasping the fundamental principles behind a phenomenon, and how to transfer those principles into tackling related problems. These skills have been invaluable to the work in this thesis, and they will continue to benefit my future careers and life. For this, I am immensely grateful.

Next, I would like to acknowledge the other members of my thesis committee, Professor Jose Andrade, Dr. Robert Graves, and Professor Nadia Lapusta. They have offered me valuable guidance along the way: their advice and perspectives directly helped the research work of this thesis, and the scientific rigor they have demonstrated will have a profound influence on me in the future.

I would also like to acknowledge the postdoctoral researchers and PhD students in my research group. I am fortunate to have such supportive and knowledgeable group mates for and bouncing ideas, and their suggestions and recommendations have been very helpful to this thesis. Apart from my own group, I have also learned a lot from other PhD students and professors in the MCE Department and the Seismological Laboratory.

I am also thankful to Caltech for providing such an excellent environment for my learning and research. I appreciate the financial stability and other non-academic resources that Caltech generously offers. More importantly, I appreciate the friendly and supportive research atmosphere here: the enthusiasm for discovery, the confidence, and the modesty of Caltech's professors and fellow PhD students have been a great inspiration and encouragement to me.

Lastly, I would like to thank my family and my friends—both in and outside Caltech—for their support throughout my time here. I would not have been able to go through those difficulties without it, and I will always remember that.

Abstract

The modeling of earthquake-induced ground motions plays an important role in the quantification of seismic hazards, which contributes to the ultimate goal of saving lives and reducing economic loss. Site response is a natural phenomenon in which soils in the earth's shallow crust alter the amplitude, frequency content, and duration of earthquake-induced ground motions. Therefore, improvements in the research of site response directly contribute to ground motion modeling, and eventually to seismic hazard quantification.

This thesis presents two models that advance the current research in site response.

The first model provides a tool to predict near-surface shear-wave velocity profiles from Vs30 (a proxy that represents the general stiffness of a site). This model bridges the gap between the lack of information about near-surface soil properties and the need to model site response on a regional scale (city, county, or above).

The second model is a stress-strain model for describing 1D shearing behaviors of soils. It is capable of capturing both the small-strain and the large-strain behaviors, which makes it suitable for modeling very strong ground motions. More importantly, this model enables seismologists to construct stress-strain curves from only shear-wave velocity information, again improving our ability to model site response on a regional scale. Our validation study shows that this model outperforms the prevalent stress-strain model (namely, the MKZ model) by a considerable margin.

Lastly, we demonstrate how the two models above can improve earthquake ground motion modeling: we develop an improved version of site factors for the Western United States. These site factors are provided as Fourier spectral ratios, and phase factors are provided for the first time, which enables the time delay of earthquake waves to be modeled. They can be used for incorporating site response in earthquake ground motion simulations, as well as for improving seismic hazard maps for the Western United States.

Published Content and Contributions

Shi, J., and D. Asimaki (2017), From stiffness to strength: Formulation and validation of a hybrid hyperbolic nonlinear soil model for site-response analyses, *Bulletin of the Seismological Society of America*, *107*(3), 1336–1355, doi:https://doi.org/10.1785/0120150287,
J.S. participated in the conception of the project, analyzed data, developed mathematical models, performed numerical simulations, and wrote the majority of the manuscript.

Shi, J., and D. Asimaki (2018), A generic velocity profile for basin sediments in california conditioned on Vs30, *Seismological Research Letters*, *89*(4), 1397–1409, doi:https://doi.org/10.1785/0220170268,
J.S. participated in the conception of the project, analyzed data, developed mathematical models, and wrote the majority of the manuscript.

Table of Contents

List of Illustrations

List of Tables

Nomenclature

In multiple chapters

Ground motion prediction The practice of modeling/predicting the intensity or the waveform of ground shaking induced by earthquakes.

Ground motion modeling Same meaning as "ground motion prediction".

Site response The phenomenon in which the soil layers near the ground surface alter the intensity, frequency content, and duration of the earthquake ground motions.

Site effect An equivalent term as "site response".

V_S Shear-wave velocity, a physical property of the soil that defines its stiffness under shearing.

V_S **profile** A series of different V_S values at different depths (below ground surface).

V_{S30} The reciprocal of travel time of shear waves from 30 m deep to the ground surface (0 m). It reflects the general stiffness of a site up to 30 m deep, and could be indicative of the stiffness of even deeper depths.

z_1 The depth that a V_S profile first reaches 1 km/s. Sometimes also known as the "basin depth".

1D One-dimensional.

2D Two-dimensional.

3D Three-dimensional.

GMPE Ground motion prediction equations, which describe the empirical correlation between various earthquake parameters (such as magnitude, epicentral distance) and the ground motion intensity level.

In Chapter 2

SVM The sediment velocity model: a 1D shear-wave velocity model for sedimentary soils, as presented in Chapter 2.

SCEC The Southern California Earthquake Center, a research institute that coordinates earthquake-related research with special focuses on, but not limited to, southern California.

CVM The Community Velocity Model, a set of models developed within SCEC, which provides 3D crustal S-wave and P-wave velocity information for southern California. It reports S- and P-wave velocity values at a given (latitude, longitude, depth) combination. Currently, the CVM has to sub-models: CVM-SCEC ("CVM-S") and CVM-Harvard ("CVM-H").

CVM-S Also known as "CVM-SCEC", which is one of the sub-models within the CVM.

CVM-H Also known as "CVM-Harvard", which is the other sub-model within the CVM. It reports slightly different results from CVM-S.

GTL The "Geotechnical Layer" model proposed by *Ely et al.* (2010), which is also a shear-wave velocity model for the near-surface soils. It is based only on mathematical constraints, not physics.

In Chapter 3

HH The hybrid hyperbolic model: a 1D stress-strain model for shearing behaviors of soils presented in Chapter 3.

MKZ The modified Kondner-Zelasko model: a 1D stress-strain model for shearing behaviors of soils, proposed by *Matasovic and Vucetic* (1993). To date, it is the most prevalent model in 1D site response analysis.

KiK-net The "Kiban Kyoshin Network", a strong-motion seismograph network in Japan, which consists of pairs of seismometers installed in a borehole and on the ground surface.

LN The linear method for site response analysis.

EQ_{HH} The equivalent linear method (proposed by *Seed and Idriss*, 1970) using the HH stress-strain model.

EQ_{MKZ} The equivalent linear method (proposed by *Seed and Idriss*, 1970) using the MKZ stress-strain model.

NL_{HH} The finite-difference nonlinear method using the HH stress-strain model.

NL_{MKZ} The finite-difference nonlinear method using the MKZ stress-strain model.

In Chapter 4

WUS The Western United States.

CEUS The Central and Eastern United States.

NGA-West2 The second iteration of the "Next Generation Attenuation Relationship for the Western United States" initiated by the Pacific Earthquake Engineering Research Center (PEER). Its products are five ground motion prediction equations (GMPEs) for WUS.

Site factors Quantitative descriptions of how site response alters earthquake ground motions at different frequencies. To completely describe site response, we need to have both amplification factors and phase factors.

Amplification factors Quantitative descriptions of how site response alters the amplitudes of earthquake ground motions at different frequencies.

Phase factors Quantitative descriptions of how site response causes time lags of different frequencies of earthquake ground motions.

SAG19 Site factors for WUS developed in Chapter 4.

KAS14 Site factors for WUS developed by *Kamai et al.* (2014).

HIH18 Site factors for WUS developed by *Hashash et al.* (2018).

NSHM National Seismic Hazard Maps developed by the U.S. Geological Survey.

BBP The Broadband Platform, a software package developed by the Southern California Earthquake Center (SCEC), which can generate 0-100 Hz seismograms for historical and scenario earthquakes in California, Eastern North America, and Japan using several alternative computational methods.

Chapter 1

Introduction

As the opening of the whole thesis, this chapter begins with some background information of earthquakes and earthquake-induced damages, which are the context of this thesis. It then introduces some details of the research in modeling earthquake-induced ground motions. Subsequently, site response, which is an important phenomenon in ground motion modeling and is the main focus of this thesis, is introduced. Lastly, the overall structure and scope of this thesis are briefly presented.

Contents of this chapter

1.1 Earthquakes

An earthquake is a phenomenon in which the sudden release of energy in the interior of the earth causes the ground to shake and displace. While weaker earthquakes can not be felt on the ground, stronger ones can cause severe ground shaking, posing great threats to human lives and properties.

The catastrophic nature of earthquakes necessitates the study of them, and the fact that earthquakes are unpreventable necessitates the study of mitigating their damages. The contribution of this thesis mainly lies in the latter, a discipline we refer to as engineering seismology.

1.2 Earthquake-induced hazards

Earthquakes-induced hazards, or seismic hazards, can endanger lives and cause economic losses. This section is a brief survey of the most important types of seismic hazards.

1.2.1 Ground motions

Ground motions are the shaking of the ground caused by earthquakes. Strong ground motions can cause structural damages, i.e., damages to buildings, roads, bridges, dams, etc. The main reason for such damages is that excess stresses within the structural components (columns, beams, walls, etc.) are accumulated during ground shaking, and they eventually exceed the strength of the materials (cement, concrete, steels, etc.), making them fail. In addition to the failure of structural components, the various auxiliary components (such as pipes, valves, wires, and decorations) can also incur damages during shaking.

Ground motions can also result in other secondary damages, which are presented below.

1.2.2 Liquefaction

The term liquefaction refers to phenomena where soils behave like a liquid rather than a solid, which are caused by the buildup of excess pore-water pressure within soils during ground shaking. Liquefaction can produce very large displacements in the ground, thus damaging structures or other properties.

1.2.3 Landslides

Landslides refer to the phenomena where soils or rocks move downhill under the force of gravity. The triggering factors of landslides include earthquakes, rainfall, construction activities, etc. During earthquakes, the shear stresses within the soils/rocks on the slopes exceed their shear strength, thus causing the soils/rocks to slide/fall down and bury people,

structures, and properties along the path. Landslides can also block river channels and form barrier lakes, which, once burst, can cause even more serious damages.

1.2.4 Tsunamis

Tsunamis are sea waves with long wavelengths that are caused by earthquakes, under-sea landslides, or other non-geologic mechanisms. In subduction-zone earthquakes, the abrupt vertical seafloor movements release energies into water, producing tsunami waves that propagate away. Upon reaching shorelines, the tsunami wave heights increase dramatically and rush onto land to cause devastating disasters. For example, the 2004 M_{w} 9.1 Sumatra earthquake in Indonesia caused tsunamis that propagated across and beyond the Indian Ocean, producing waves as high as 30 m (100 ft) when traveling inland. That tsunami caused 227,899 deaths (*National Oceanic and Atmospheric Administration*).

1.2.5 Fires, diseases, and more

Earthquakes can also cause other secondary damages such as fires, diseases, or other monetary loss. For example, earthquakes cause fires by rupturing gas lines and power lines, and further prohibits the extinction of fires by damaging the water pipes or sprinklers. Diseases can also spread in regions affected by earthquakes, due to the shortage of water and medical care.

1.3 Modeling ground motions

Among the various types of earthquake-induced damages mentioned above, earthquake ground motions are the most prevalent, since they are often the cause of the other damages. Also, except for some notable events (such as the 1906 San Francisco Earthquake and Fires and the 2004 Sumatra Earthquake and Tsunami), ground motions are usually the controlling factor for damages in earthquakes.

1.3.1 Importance of modeling ground motions

Modeling ground motions enables us to quantify the intensity of ground shaking. Its results can be used in designing buildings and other structures, or in estimating regional-scale seismic hazards and risks, which are the potential economic loss and life threats from earthquakes.

"Ground motion modeling" is sometimes referred to as "ground motion predictions", but it is imperative to note that ground motion predictions are not the same as earthquake predictions. Earthquake predictions attempt to report the time (accurate to a specific day or

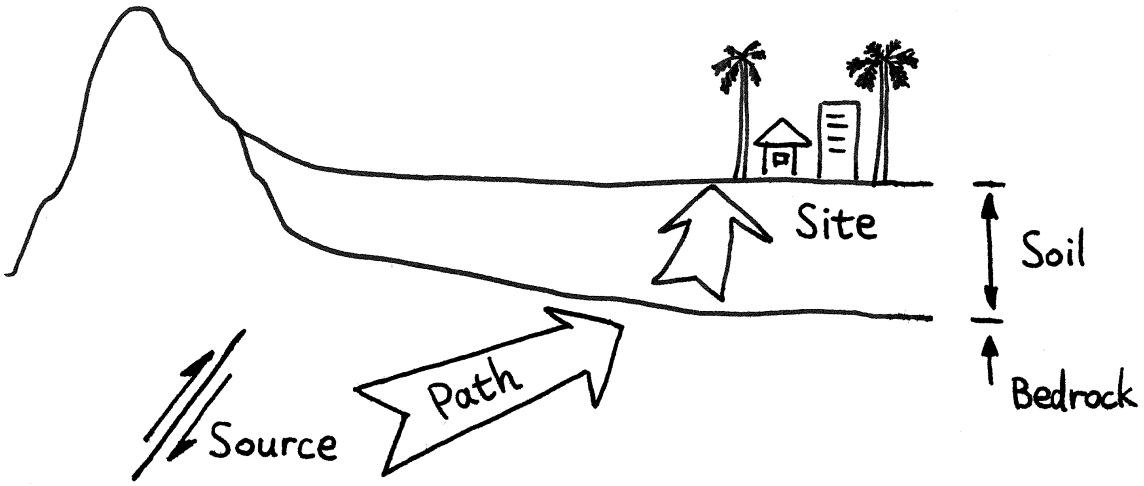


Figure 1.1: Three aspects that govern earthquake ground motions: source, path, and site. This illustration reflects the typical near-surface geology in many populous areas: sedimentary soils underlain by bedrocks. (Modified from a similar illustration in *Kramer*, 1996.)

week), the location, and the magnitude of earthquakes, and the seismological community has not yet been able to predict a major earthquake (*U.S. Geological Survey*). On the other hand, ground motion predictions involve modeling the characteristics of ground motions given an earthquake scenario, therefore a completely different concept.

1.3.2 Aspects affecting earthquake ground motions

To properly model earthquake ground motions, it is important to first identify different aspects that govern them, that is, aspects that determine the temporal frequency, spatial distribution, and specific characteristics of ground motions.

There are three main aspects that govern ground motions: source, path, and site. Figure 1.1 illustrates these three aspects.

1.3.2.1 Earthquake sources

Earthquake sources refer to the physical processes that generate earthquakes. Earthquakes can occur from natural or human-induced causes. The naturally occurring earthquakes are usually caused by the movement of the tectonic plates and the associated abrupt release of mechanical energies. The human-induced earthquakes have various causes, ranging from reservoirs, mining, hydraulic fracturing, carbon capture and storage, etc. Needless to say, there will be no earthquake ground motions without an earthquake source.

1.3.2.2 Paths of seismic waves

The energy released from earthquake sources propagates within the earth in the form of seismic waves, ultimately arriving at the ground surface to make an impact on human beings. The paths that seismic waves take can alter the intensity, duration, and frequency contents of earthquake ground motions, making it another important aspect in this list.

1.3.2.3 Site response

Site response refers to the phenomena that the very top of the earth's crust—soils, weathered rocks, etc.—alter the intensity, the duration, and the frequency contents of earthquake ground motions. Even though the length scale of concern to site response is only a few hundred meters, which is a small portion of the whole crust (25 to 40 km), the effects of site response to ground motions can be very significant.

(Site response is the main focus of this thesis, and more introductory materials are presented in Section 1.4.)

1.3.3 Current practice

To date, there are mainly two types of approaches to model/predict ground motions: the data-driven approach and physics-based approach, each having their own advantages and disadvantages.

1.3.3.1 The data-driven approach

The data-driven approach collects earthquake records (those that consist of information of earthquake magnitude, source type, source-to-site distance, local site condition, and recorded ground motion intensities, etc.) and utilizes statistical techniques (such as traditional linear regression or machine learning) to construct empirical correlations between ground motion predictors (magnitude, source type, ...) and ground motion intensities. Such correlations are often referred to as the "ground motion prediction equations" (GMPEs), or sometimes the "attenuation relationships".

The advantage of this approach is that it allows researchers to establish correlations from data, without fully knowing the causality behind the problems at hand. Naturally, its disadvantage comes from the fact that the amount and coverage of data is limited. For example, if a GMPE is developed with a data set without earthquakes of magnitudes exceeding 8.0, then it may not credibly predict ground motions from a magnitude 8.3 earthquake. Furthermore, current GMPEs do not predict ground motion waveform, only

spectral accelerations at different frequencies, which limits its use in civil engineering designs (*Yamamoto and Baker*, 2013).

1.3.3.2 The physics-based approach

The physics-based approach utilizes wave propagation theories and numerical simulation techniques to solve wave equations in 1D, 2D, or 3D.

This gives the physics-based approach a big advantage over the data-driven approach: researchers can potentially "extrapolate" physics-based models to make predictions beyond the range of recorded data (e.g., beyond the largest ground motion ever recorded) as long as all the relevant physical processes can be correctly modeled. However, the disadvantage of the physics-based approach is that some physical processes are very complex are not yet well understood, so various simplifications and assumptions have to be made, which limits its predictive capability. Another disadvantage of the physics-based approach is the long computational time and resources it requires.

The data-driven and physics-based approaches are not mutually exclusive. On the contrary, one benefits from the improvement of the other: GMPEs can often validate physics-based results, and credible physics-based simulations can augment the data for developing newer versions of GMPEs. (In fact, Chapter 4 of this thesis is an example of using the physics-based approach to improve the data-driven approach.)

1.4 Site response

The focus of this thesis is on improving the research of site response—the third aspect affecting earthquake ground motions (Section 1.3.2.3). By making improvements in site response analysis, this thesis contributes to improving ground motion modeling (both the data-driven and the physics-based approaches), and could ultimately contribute to improved engineering design and seismic hazard quantification.

1.4.1 What is site response?

In order to understand the concept of site response, we need to first take a look at the major landforms of the earth.

Among the earth's major landforms, basins and plains mainly consist of soils underlain by bedrocks, and highlands and plateaus mainly consists of bedrocks that are exposed at the ground surface (henceforth referred to as the "rock outcrop"). According to wave propagation theories, when seismic waves travel through rocks or soils, the wave amplitude

is lower in stiffer materials (i.e., rocks) and higher in softer materials (i.e., soils). This means that seismic waves are amplified when they travel from bedrocks into soils, or in other words, ground motion amplitudes are higher on soils than on adjacent rock outcrop (see the illustration in Figure 1.1). This phenomenon is referred to as "site response" (or "site effects", "site amplification").

Site response was observed in real earthquakes and qualitatively documented in the literature dating back to as early as the 1900s (such as *Milne*, 1908), and pioneering quantitative studies were carried out by *Reid* (1910); *Neumann* (1954); *Kanai et al.* (1956); *Gutenberg* (1957), etc. For example, *Gutenberg* (1957) collected ground motion recordings from 25 sites within 30 miles from the Seismological Laboratory of the California Institute of Technology, and compared them with the data recorded at the Seismological Laboratory site (on crystalline rock). Gutenberg found that at fairly dry alluvium sites, the ground motions (of frequency between 0.67 to 1.0 Hz) were amplified 5 times or more compared to those recorded at the Seismological Laboratory. He also observed that the frequency at which the amplification was greatest varied with the depth of the underlying alluvium.

A more recent example of site response is the different ground motions on Yerba Buena Island and Treasure Island during the 1989 M_s 7.1 Loma Prieta earthquake in northern California. These two islands are located in the San Francisco Bay; Yerba Buena Island is a rock outcrop and Treasure Island is an artificial island filled by soils (San Francisco Bay Mud). The two islands are merely several miles apart, so during the same earthquake, the influence from earthquake source and path was the same for both islands. Referring to Section 1.3.2, the only remaining factor is site response—and the ground motion level on the soil island is about 3 times as high as on the rock island[1].

1.4.2 Quantifying site response

The difference between soil stiffness and rocks stiffness determines the level of amplification: in general, the larger the difference, the larger the amplification. Therefore, to quantify site response, we need to quantitatively describe material stiffness at the locations of interest.

However, the stiffness of soils is not a constant: it reduces with soil deformation. As soil deformation undoubtedly happens during ground shaking, the soils becomes softer, thus can further amplify seismic waves. To make things more complex, the energies in the seismic waves are partially absorbed by soil particles during shaking (referred to as damping). Damping decreases the wave amplitudes, counteracting the effects from soil softening.

[1] On Yerba Buena Island: E-W peak ground acceleration: $0.06g$, N-S: $0.03g$; on Treasure Island: E-W: $0.16g$, N-S: $0.11g$. Data source: *Kramer* (1996).

We refer to soil softening and damping effects as nonlinear site response[2]. The study of nonlinear site response started in the 1960s, with studies by *Kondner and Zelasko* (1963); *Hardin and Drnevich* (1972a,b); *Seed and Idriss* (1970).

1.4.3 Challenges in site response analysis

As hinted in the paragraphs above, the keys to better site response analysis are mainly: (1) correctly describing soil and rock stiffness, and (2) correctly describing nonlinear soil behaviors. These two tasks may seem straightforward and achievable for specific locations, but on a regional scale (city, county, or above), they become quite challenging.

1.4.3.1 Linear material properties

The soil and rock stiffness—without deformation—is referred to as the linear material properties, and we use shear-wave velocities (V_S) and shear modulus as the metrics for stiffness.

The main challenge in studying linear material properties is not in the measuring of them in the fields, because there have been numerous well-established testing techniques for this purpose (such as the cross-hole test, cone penetration test, surface-wave dispersion techniques, etc.). However, these tests can only measure the linear properties of a single location at a time, and the regional-scale measurements are not financially feasible. And since the quantification of seismic hazards is always a regional-scale effort, researchers have to rely on easily accessible proxies, such as V_{S30} maps[3] or 3D crustal velocity models.

Therefore, the main challenge here is how to improve the accuracy and resolution of estimating linear material properties from those proxies, and our efforts in addressing this challenge is presented in Chapter 2.

1.4.3.2 Nonlinear soil behaviors

Nonlinear soil behaviors, namely the soil softening and damping effects, can be directly measured in the laboratory with well-established techniques such as the direct shear test, resonant column test, cyclic triaxial test, etc. But the regional-scale measurements of the nonlinear properties are even less feasible (technically and financially) than linear properties. Proxies do exist for estimating nonlinear properties without laboratory testing data, but they have some intrinsic flaws which limit their applicability for very strong ground motions.

[2] Rocks can also exhibit softening and damping, but at a much smaller level compared to soils.

[3] See the Nomenclature (page xix) for definition of V_{S30}.

Therefore, the main challenge here is how to improve these proxies, and it is what Chapter 3 addresses.

1.5 Structure of this thesis

The structure of the remainder of this thesis is fairly straightforward.

Chapter 2 focuses on addressing the challenge in estimating linear material properties. To do so, it proposes a data-driven model that correlates the available proxy—V_{S30}—to the linear properties (namely, V_S profile). And in Section 2.3, this model is benchmarked against measurements and shows satisfactory performance.

Chapter 3 focuses on addressing the challenge in estimating nonlinear material properties. To do so, it proposes 1D stress-strain model that explicitly incorporates shear strength of soils. More importantly, Chapter 3 also provides a pipeline for estimating the nonlinear soil properties from linear soil properties in the absence of laboratory test data, making it possible to simulate nonlinear site response on a regional scale. The model and the pipeline are benchmarked against earthquake recordings in Japan, and show satisfactory performance even for the strongest ground motions.

Chapter 4 demonstrates an important application of the models in Chapters 2 and 3: it provides an improved version of nonlinear site factors for the Western United States. Nonlinear site factors are an integral part for ground motion modeling, so they have the potential to ultimately help advance seismic hazard quantification.

Chapter 5 summarizes the whole thesis, and proposes potential future research directions.

1.6 Scope of this thesis

This thesis limits mainly on 1D site response. Some phenomena specific to 2D and 3D problems (such as topographic effects, surface wave amplification, basin effects) are therefore not discussed. Nevertheless, with some assumptions and simplifications, the results in this thesis can still be of use to some 2D/3D problems.

Chapter 2

A generic velocity profile for basin sediments in California conditioned on Vs30

This chapter presents the Sediment Velocity Model (SVM), a 1D shear-wave velocity model for basin sediments in California. It translates V_{S30}, a proxy that represents the stiffness of the near-surface sediments, into a 1D velocity profile suitable for ground motion simulations. We develop this model based on the statistics of 914 measured velocity profiles.

The need for such a model arises from the fact that near-surface soil layers play a critical role in modifying the amplitude, frequency, and duration of earthquake ground motions, and thus are an important factor in engineering design and seismic hazard quantification.

We present the procedures of developing this model (Section 2.2), the evaluation of its performance (Section 2.3), and some empirical correlations found in the data (Sections 2.4 and 2.5).

The contents of this chapter is adopted from our publication, *Shi and Asimaki* (2018):

J. Shi and D. Asimaki (2018), "A generic velocity profile for basin sediments in California conditioned on V_{S30}", *Seismological Research Letters*, 89(4), 1397–1409.

Contents of this chapter

2.1 Introduction

Recent advances in computational seismology have enabled physics-based simulations of earthquake ground motions on regional scales (i.e., hundreds of kilometers); for example, the works by *Taborda and Bielak* (2013); *Olsen and Takedatsu* (2015); *Lozos et al.* (2015); *Sun et al.* (2015); *Hartzell et al.* (2016); *Zhang et al.* (2017) and others. In seismic hazard analyses, simulated ground motions can be used to populate spatially sparse and scarce ground motion datasets (e.g., *Ramirez-Guzman et al.*, 2015) and to generate data from high-magnitude earthquake scenarios yet to occur—namely, the scenarios that govern risk assessment, loss prevention, and/or earthquake preparedness practices (e.g., *Jones et al.*, 2008). Physics-based simulations enable a wide range of applications, such as fault systems analysis, strong ground motion prediction, and earthquake hazard assessment. More importantly perhaps, these high-fidelity simulations shed light on complex regional 3D wave-propagation effects and improve our understanding and representation of phenomena such as basin edge wave focusing and surface wave generation, basin resonance, and topography effects.

An important aspect of physics-based ground motion simulations is the use of realistic crustal material properties, such as wave velocity and mass density. As crustal material properties are obviously region-specific, there exist various material property datasets specific to different regions. An example is the Community Velocity Models (CVMs) developed by the Southern California Earthquake Center (SCEC), which provide three-dimensional (3D) seismic velocity information for southern California. The purpose of the CVMs is "to provide a unified reference model for the areas of research that depend of the subsurface velocity structure, which include seismicity location, tomographic velocity modeling, and strong motion simulations" (*SCECpedia*, 2016). Currently, the CVMs have two sub models: CVM-SCEC ("CVM-S") and CVM-Harvard ("CVM-H").

CVM-S was originally developed by *Magistrale et al.* (1996) and *Magistrale et al.* (2000), (initial versions) and later *Kohler et al.* (2003) (version 3 and version 4, or "CVM-S4"). According to *SCECpedia* (2016), the deeper sediment velocities of CVM-S4 "are obtained from empirical relationships that take into account age of the formation and depth of burial; the coefficients of these relationships are calibrated to sonic logs taken from boreholes in the region. Shallow sediment velocities are taken from geotechnical borehole measurements." And hard-rock velocities are based on tomographic inversions. CVM-S4 was iteratively updated using full 3D tomography (by *Lee et al.*, 2014) to improve both the shallower sedimentary and the deeper crustal velocities. The resultant velocity model after 26 iterations was named CVM-S4.26. The most recent CVM-S version is CVM-S4.26.M01, which

differs from CVM-S4.26 in that the M01 version "preserves some of the geotechnical information in the original CVM-S4 model that was lost during the tomography improvements" (*SCECpedia*, 2016).

The other sub model, CVM-H, was originally developed by *Süss and Shaw* (2003) using industry seismic reflection profiles, direct borehole velocity measurements, and 3D tomographic inversions. After several updates, the most recent CVM-H version is CVM-H15.1.0.

Within such regional 3D velocity models, the realistic representation of shallow-sediment velocities (within the top 200 m or so) is becoming increasingly essential, because shallow sediments can have significant effects on the higher frequencies (> 1 Hz) of strong ground motions, and the state-of-the-art supercomputers are demonstrating the capability to simulate ground motions in such frequency ranges (such as the recent efforts by *Fu et al.*, 2017 that achieved an 18 Hz simulation at 8 m spatial resolution). However, the inversion methods used in CVM-S and CVM-H are limited by the spatial resolution of available data, and thus cannot resolve very fine velocity variations for the shallow sediments.

To address this issue, CVM-S4.26.M01 and CVM-H15.1.0 employ two different approaches: CVM-S4.26.M01 directly incorporates geotechnical measurements (i.e., measured shear wave velocities for sediments) at a few hundred sparse locations in southern California. However, the data at those sparse locations do not "spread" to adjacent areas, which limits the benefits of including actual measurements (as will be discussed in details in the section "Shallow crustal structure of SCEC CVM-S4.26.M01").

On the other hand, CVM-H15.1.0 has an optional "Geotechnical Layer" (GTL) model that replaces the soft sediments in the top 350 m of the crust. Developed by *Ely et al.* (2010), GTL uses an analytical formula to generate 1D velocity profiles and requires two input parameters: (1) the weighted average shear wave velocity of the top 30 m (V_{S30}), and (2) the CVM-H15.1.0 shear wave velocity at 350 m depth ($V_{S(350)}$). The formula of GTL results in a smooth velocity profile from 0 to 350 m, which is not realistic because strong velocity impedance (such as basin edge) often occurs above that depth. Therefore, the use of GTL can alter the shapes of basins unintentionally, and thus can negatively affect the overall quality of predicted strong ground motions (*Taborda et al.*, 2016).

Another shortcoming of GTL comes from its use of the V_{S30} value: the formulation of GTL only mathematically ensures that the V_S profile it produces has a resultant V_{S30} equal to the given V_{S30}, so there is no guarantee that the velocities between 0 and 30 m would be realistic. And as the predicted V_S extends deeper than 30 m, the potential inaccuracy becomes exacerbated, as will be shown in Section 2.3.

Prompted by the shortcomings of GTL and the importance of capturing site effects caused by soft sedimentary materials, we have developed a shallow crustal velocity model for California, henceforth referred to as the Sediment Velocity Model (SVM). The SVM is formulated as a function of V_{S30}, which can be queried from the California V_{S30} maps (such as those by *Yong et al.*, 2012; *Thompson et al.*, 2014; *Wills et al.*, 2015), and allows the depth of basins to be determined from crustal velocity models (e.g., the CVMs), hence preserving the shapes of basin edges. The SVM is based on statistical analyses of 914 V_S profile measurements that we obtained from various sources. In this chapter, we describe the development and validation of the SVM and demonstrate its predictive capabilities compared to CVM-S4.26.M01, CVM-H15.1.0, and GTL. For brevity, we will use "CVM-S" and "CVM-H" in subsequent texts to represent "CVM-S4.26.M01" and "CVM-H15.1.0" respectively. And the term "CVM-H" (or "CVM-H15.1.0") in subsequent texts refers to the velocities without the optional GTL layers.

2.2 Data analysis and model development

In this section, we present our procedures to develop the Sediment Velocity Model (SVM), including data source, pre-processing, parameterization, and calibration of the model. At the end of this section, we present the formulas of the SVM and the calibrated model parameters.

2.2.1 Data source

We obtain four datasets of V_S profile measurements for the development of the SVM: (1) 178 profiles measured by *Yong et al.* (2013), (2) 277 profiles documented in *Boore* (2003), (3) 137 profiles collected by Chris Wills from the California Geological Survey (personal correspondence), and (4) 322 profiles measured by LeRoy Crandall and Associates (personal correspondence). The total number of profiles is 914. The first three datasets are publicly accessible, and the last dataset is proprietary. For simplicity, we henceforth refer to these four datasets as AY, DB, CW, and LC, respectively.

The V_S profiles in all four datasets are measured within California, concentrated mostly in Los Angeles and San Francisco areas, as shown in Figure 2.1. The histograms of V_{S30} values calculated from the profiles of each dataset are shown in Figure 2.2. The vast majority of sites have V_{S30} between 200 and 500 m/s, and outside this range, the measurements are relatively scarce.

The V_S profiles were measured using two families of site characterization techniques: the DB, CW, and LC profiles were measured using invasive methods (e.g., suspension logging,

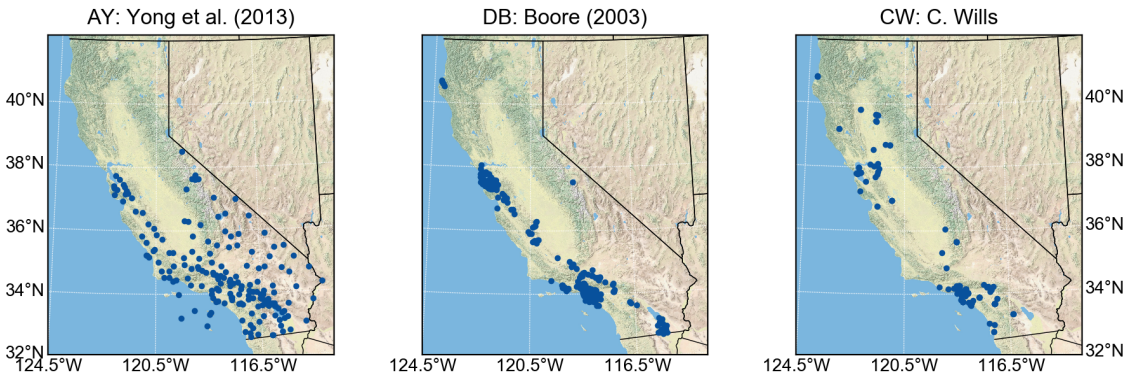


Figure 2.1: Locations of V_S profile measurements of three of the four datasets: AY (*Yong et al.*, 2013), DB (*Boore*, 2003), and CW (Chris Wills). Each dot on the map denotes the location of a V_S profile measurement. (The LC dataset is proprietary, so we do not show their locations here.)

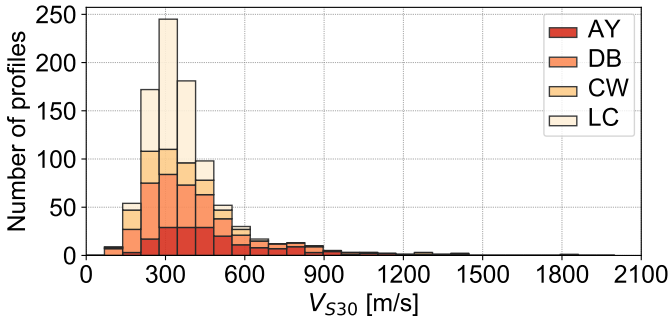


Figure 2.2: Histograms of V_{S30} values of the measured profiles. There is an abundance of measurement for V_{S30} between 200 and 500 m/s, and outside of this range, the measurements are relatively scarce.

cross-hole and down-hole tests), and the AY profiles were measured using non-invasive methods (based on the inversion of surface wave dispersion curves, such as SASW, MASW, and/or ReMi). A number of previous studies (*Boore and Asten* 2008; *Boore and Brown* 1998; *Brown et al.* 2002; *Rix et al.* 2002; *Stephenson et al.* 2005) have shown that non-invasive shear-wave velocity profiling techniques produced similar results as the invasive techniques. Thus in principle, the four datasets could have been merged directly for our subsequent analyses. However, we identified one discrepancy between the AY profiles and the DB+CW+LC profiles that can be traced back to the measurement techniques, which we discuss in the next section.

2.2.2 Merging invasive and non-invasive datasets

To produce an analytical V_S profile formula as a function of V_{S30}, we followed an approach similar to *Wills and Clahan* (2006), except instead of using geologic units in that study, we grouped our profiles in V_{S30} bins. Specifically, we first define a series of sliding V_{S30} bins with overlap (meaning that a given V_S profile could belong to more than one bin), and for each bin, we select the boundaries to ensure that the bins contain a statistically significant number of V_S profiles (ideally more than 40, if available), while making sure that the number of bins was adequately large for us to derive meaningful correlations between V_S profiles and V_{S30}. (Hence the bins are "wider" in larger V_{S30} ranges because of less available measurements.) Next, within each bin, we calculate an average V_S profile (by taking the arithmetic mean of V_S values) and the corresponding V_{S30} of each averaged profile.

It is during this process that we notice a disagreement between velocity measurements made via invasive and non-invasive techniques. To demonstrate this, we show four groups of averaged V_S profiles from each dataset that correspond to almost the same averaged V_{S30} (Figure 2.3, (a)–(d)). From left to right, the four averaged profiles in each subplot correspond to V_{S30} = 300 ± 6 m/s, 400 ± 12 m/s, 500 ± 10 m/s, and 560 ± 8 m/s (the uncertainty range of V_{S30} is due to the finite width of the V_{S30} bins containing the individual measurements). Each averaged V_S profile comes from averaging at least 14 (and up to 236) measured profiles around the specific V_{S30}.

Figure 2.3(a)–(d) shows that the average profiles of the four datasets are in very good agreement. However, by zooming in to the top 10 m (Figure 2.3(e)–(h)), we see that the average AY profiles (non-invasive method) give systematically lower velocity than the DB, CW, and LC profiles (invasive methods) for the topmost 2-3 m. We further notice that the average velocity of the AY profiles varies gradually with depth, whereas the average velocities of the top 2-3 m from the other three databases are approximately constant. We attribute this systematic discrepancy to the fact that in invasive methods, the downhole geophone is placed in the borehole at relatively large depth increments (for example, as documented in *Boore*, 2003, "at approximately every 2.5 m in depth"). Thus the layering resolution of the invasive measured profiles is 2.5 m in most cases, and the V_S of the topmost layer represents the average stiffness from 0 to 2.5 m depth (in the case of the DB dataset). On the other hand, non-invasive techniques rely on data inversion using software, whereby the depth increments are "virtual" and typically on the order of 1.0 m.

Therefore, to merge the four datasets for the development and calibration of the SVM, we calculate the average V_S for the top 2.5 m of every profile in AY (in the same manner as calculating V_{S30}), and assign the average velocity to the whole top 2.5 m, as if the top

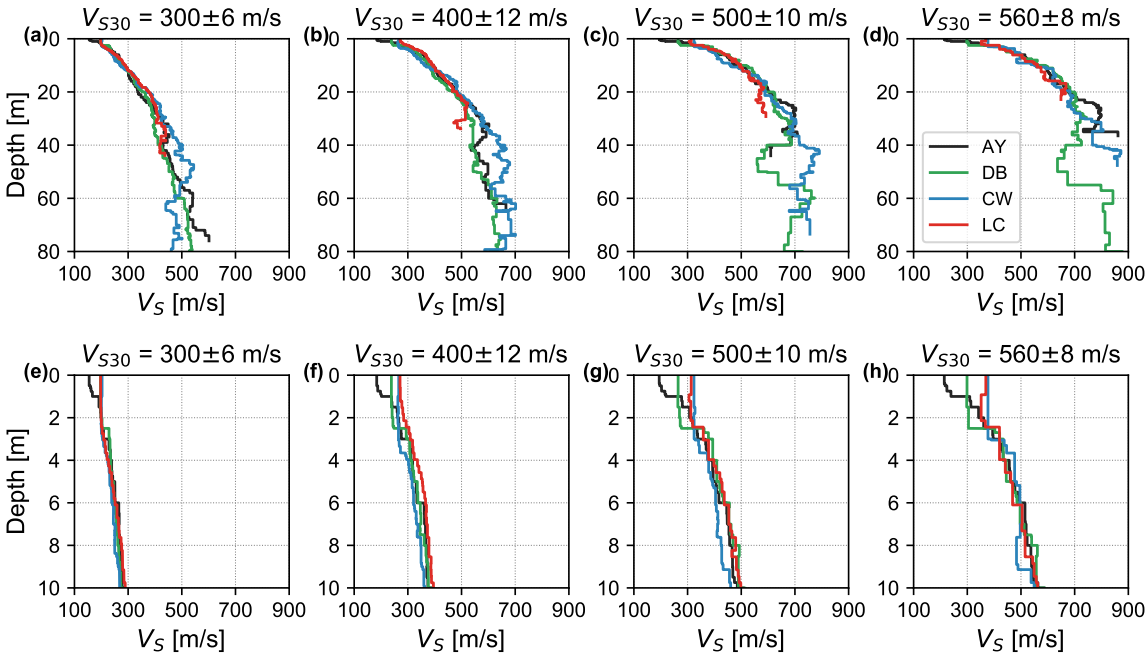


Figure 2.3: (a) to (d): Comparison of average V_S profiles of four different V_{S30} calculated from the four datasets. The V_{S30} value ranges are marked on top of each sub-figure. (e) to (h): The zoomed-in version of (a) to (d), only showing the top 10 m.

layer were homogeneous. Averaging velocities over such a short length only slightly affects high frequencies (above 15 Hz), as shown in *Boore and Thompson* (2007), and thus is less important even for engineering applications. With this adjustment, we circumvent the systematic discrepancy caused by different V_S profiling techniques, and merge the four datasets for our subsequent analyses.

2.2.3 Training versus test datasets

Next, we randomly reserve a small number of measured profiles as the "test set", and use the remaining as the "training set". We use the training set to develop our SVM, and use the test set to validate it. The purpose of the training/test split is to ensure that the model does not fit to the noises in the training data ("overfitting"), and thus can perform well in making future predictions. In other words, the test set acts as the "unseen data" to the model, and the model's performance on the test set should be close to its performance on any real-world input data.

Although there are no specific rules that constrain the relative size of the training and test sets, one can see that the test set needs to be large enough for the validation to be statistically informative, and the training set should be as large as possible, because the quality of

the predictive model depends on the amount and quality of the training data. With this consideration in mind, and via some trial and error, we choose the following criteria to pick the test set:

1. We group the profiles in AY, DB, and CW into V_{S30} bins defined by the following bin edges: [200, 300, 400, 550, 800, 1500] (unit: m/s). The widths of bins reflect the relative abundance/scarcity of measured profiles (refer to the histogram in Figure 2.2). Profiles outside of [200, 1500] m/s are not considered, because they are too few to be excluded from the training set. The LC database is proprietary, so we cannot show individual profiles from it, and thus we do not pick any LC profiles into the test set. Since we have already shown that the average profiles from the four datasets are virtually indistinguishable, excluding the LC profiles from the test set does not affect the quality of the test set.

2. Within each V_{S30} bin of AY, DB, and CW:

    a) If the bin contains more than 30 profiles (inclusive), we randomly pick four profiles into the test set (using the random number generator in MATLAB).

    b) If the bin contains less than 30 profiles (exclusive) and more than 15 profiles (inclusive), we randomly pick three profiles into the test set.

    c) If the bin contains less than 15 profiles (exclusive), no profile is picked from that bin into the test set.

Following the criteria above, our test set contains 17 profiles from AY, 15 from DB, 11 from CW, and none from LC, totaling 43 profiles, or approximately 10% of the qualifying profiles (i.e., satisfying criterion 2(a) or 2(b)). The training set, on the other hand, contains 914 – 43 = 871 profiles.

2.2.4 SVM parameterization and calibration

Using the 871 profiles of the training set, we then repeat the binning process described above (without distinguishing the four datasets) and calculate the averaged V_S profiles as well as the error bounds for each bin. We assume that V_S follows the log-normal distribution following the suggestion by *Toro* (1995), so the average V_S is the mean of the log-normal distribution, which is

$$\mathrm{AVG}_{Vs} = \exp(\mu_{\ln V_S} + \sigma^2_{\ln V_S}/2) \tag{2.1}$$

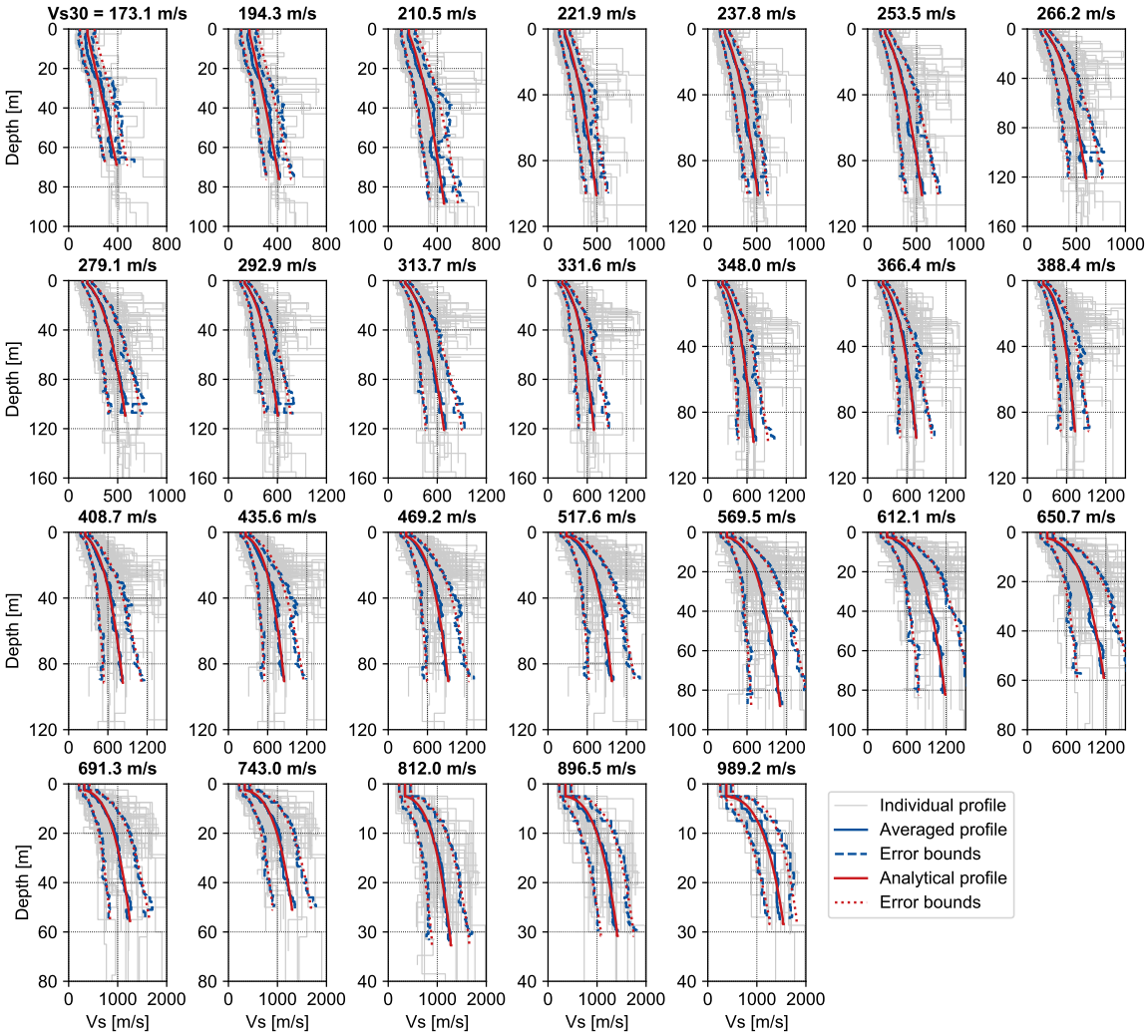


Figure 2.4: The individual V_S profiles (in gray), average V_S profiles (with error bounds, in blue) of each V_{S30} bin, as well as analytical V_S curves (in red) fitted to the average V_S profiles and the error bounds. Note that the LC profiles are not plotted here because they are proprietary (they are used in calculating the averaged profiles and the error bounds, though).

and the standard deviation (STD) of V_S is

$$\mathrm{STD}_{Vs} = \sqrt{\left[\exp(\sigma^2_{\ln V_S}) - 1\right] \exp(2\mu_{\ln V_S} + \sigma^2_{\ln V_S})} \tag{2.2}$$

where $\mu_{\ln V_S}$ and $\sigma^2_{\ln V_S}$ are the sample mean and sample variance of $\ln(V_S)$ (which is normally distributed). Also, because the depths of the measured profiles vary significantly, we only use the average V_S values at depths where we have more than 12 V_S measurements in order to avoid statistical biases in calculating the average. Figure 2.4 shows the averaged V_S profiles (as blue solid lines) and error bounds (as blue dash lines, calculated as $\exp(\mu_{\ln V_S} \pm \sigma_{\ln V_S})$) of each bin.

From Figure 2.4, we see that the average V_S increases monotonically with depth for all V_{S30} bins—with the exception of the top 2.5 m, where the average V_S is assumed to be constant (see previous sections for justification). We thus choose the following analytical expression to fit the average V_S profiles in each bin:

$$V_S(z) = \begin{cases} V_{S0} & , 0 \leqslant z < z^* \\ V_{S0}\left(1 + k\left(z - z^*\right)\right)^{1/n} & , z > z^* \end{cases} \tag{2.3}$$

where z is depth in meters, and z^* is chosen as 2.5 m; V_{S0} is the shear wave velocity (m/s) from $z = 0$ to 2.5 m, and k and n are two dimensionless parameters: k is analogous to the "slope" of the curve which describes how fast V_S increases with z, and n controls the degree of curvature ($n = 1$ is a straight line, $n > 1$ is "convex", and $n < 1$ is "concave"). The analytical expression for $z > z^*$ is adapted from *Vrettos* (1996). Even though soil profiles in reality are not necessarily completely smooth, both *Kaklamanos and Bradley* (2016) and *Shi and Asimaki* (2017) have shown that "smoother" soil profiles (i.e., with fewer large velocity contrasts) may be more realistic because they yield better 1D site response predictions.

The curve fitting results are also shown in Figure 2.4 as red solid and dash lines. Note that for each V_{S30} bin (i.e., each subplot), we fit both the averaged V_S profile and the error bound profiles, which yields three analytical curves corresponding to three sets of (V_{S0}, k, n) values.

We also observed that (V_{S0}, k, n) vary as a function of V_{S30} as shown in Figure 2.5. We specifically observed the following trends:

1. V_{S0} increases almost linearly with V_{S30}, namely stiffer sites (larger V_{S30}) are characterized by higher V_S of their topmost layer.

2. k increases with V_{S30}, which indicates that V_S increases faster with depth for stiffer sites.

3. n initially increases with V_{S30} and then gradually reaches $n = 4$. This trend indicates that the curvature of V_S profiles initially increases with increasing V_{S30}, and gradually becomes constant. Note that $n = 1.0644$ for the lowest V_{S30} in our dataset (173.1 m/s), meaning that the average V_S profiles are always convex.

4. The error bounds of (V_{S0}, k, n) increases with V_{S30} in general (which can also be seen in Figure 2.4), which reflects both the scarcity of measurements in the large V_{S30} regime and the complex geologic conditions that diverge from pure sediment deposition and consolidation that typically characterizes the softer sites.

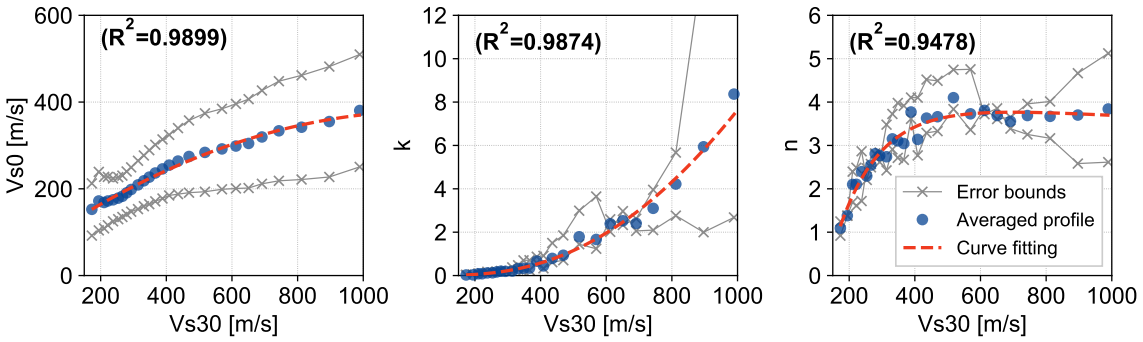


Figure 2.5: Relationships between V_{S0}, k, n and V_{S30}. Each "o" marker represents the curve-fitting result shown in a subplot of Figure 2.4. The red dashed curves are the fit to the general trend of the "o" markers. We also show the error bound values of V_{S0}, k, n, which come from the red dashed lines in Figure 2.4.

2.2.5 The SVM parameters as a function of V_{S30}

Next, we choose analytical functions to capture the variation of (V_{S0}, k, n) versus V_{S30} shown in Figure 2.5. The fit for each parameter is shown in the figure, with R^2 values marked in each subplot. The three analytical expressions are:

$$\begin{aligned} V_{S0} &= p_1 (V_{S30})^2 + p_2 (V_{S30}) + p_3 \\ k &= \exp\left(r_1 (V_{S30})^{r_2} + r_3\right) \\ n &= s_1 \exp(s_2 V_{S30}) + s_3 \exp(s_4 V_{S30}) \end{aligned} \tag{2.4}$$

where $p_1 = -2.1688 \times 10^{-4}$, $p_2 = 0.5182$, $p_3 = 69.452$, and $r_1 = -59.67$, $r_2 = -0.2722$, $r_3 = 11.132$, and $s_1 = 4.110$, $s_2 = -1.0521 \times 10^{-4}$, $s_3 = -10.827$, $s_4 = -7.6187 \times 10^{-3}$. Therefore, given a V_{S30} value, we can first use Eq (2.4) to calculate (V_{S30}, k, n), and then substitute them into Eq (2.3) to calculate an analytical V_S profile.

Note that the V_{S30} values of the data points in Figure 2.5 range from 173.1 to 989.2 m/s (see Figure 2.4). Therefore we recommend the valid range of V_{S30} for Eq (2.4) (and hence the SVM) to be [173.1, 1000] m/s. Additionally, our datasets contain mostly soil to shallow crust profiles, so we do not recommend using the SVM to generate shear-wave velocities higher than 1,500 m/s (see the last subplot of Figure 2.4). To be more conservative, we recommend using the SVM profile up to V_S =1,000 m/s.

2.3 Validation of the SVM

We validate the SVM against our test set, and compare it with CVM-S, CVM-H, and GTL. Specifically, we calculate V_{S30} of every test set profile, and use them to generate 43 SVM profiles. We then query CVM-S and CVM-H profiles based on the coordinates of the test set sites. We can also obtain $V_{S(350)}$ from the CVM-H profiles, and together with the V_{S30} values we can calculate GTL profiles. (We use $V_{S(350)}$ from CVM-H rather than CVM-S because this GTL model is intended to be used on top of CVM-H, not CVM-S.)

Note that 13 of the test set sites are in northern California, i.e., outside the geographic range of CVM-H and CVM-S. Therefore for these sites, their CVM-H or CVM-S goodness-of-fit scores will be set to "NaN" (not 0). And to calculate their GTL profiles, we use a generic $V_{S(350)}$ of 1,950 m/s (from the Bay Area velocity model proposed in *Aagaard et al.*, 2008). Also note that we compare the profiles down to the depth of measurements or the depth where measured V_S =1,000 m/s, whichever is shallower.

To evaluate the performance of the different velocity models, we define the following two metrics:

1. Goodness-of-fit of the predicted velocity profiles: $\mathrm{GoF_{Vs}}$.

   - Discretize the "predicted" (SVM, GTL, CVM-H or CVM-S) and measured profiles into layers of 0.1 m thickness;
   - At the midpoint of each layer, calculate the goodness-of-fit of the "predicted" V_S as follows:

   $$\mathrm{gof_{Vs}} = 10 - 10\mathrm{erf}\left(\frac{4\left|\ln V_S^{\mathrm{predicted}} - \ln V_S^{\mathrm{measured}}\right|}{\left|\ln V_S^{\mathrm{measured}}\right|}\right) \tag{2.5}$$

   where erf() is the error function. A perfect goodness-of-fit score is exactly 10, and scores closer to 0 are poor fits. (The factor 4 in the numerator prevents the scores from all being close to 10, while preserving the ranking of the scores.)
   - Compute the predicted profile's goodness of fit, $\mathrm{GoF_{Vs}}$, by averaging the $\mathrm{gof_{Vs}}$ across all layers.

2. Goodness-of-fit of the predicted profile's linear site amplification: $\mathrm{GoF_{AF}}$.

   - Calculate linear amplification factors ("AF") from the predicted and measured V_S profiles, denoted as $\mathrm{AF_{pred}}$ and $\mathrm{AF_{meas}}$, respectively. AF are calculated in

SeismoSoil (*Asimaki and Shi*, 2017) using the method proposed by *Thomson* (1950) and *Haskell* (1953).

- Calculate the goodness-of-fit between $\mathrm{AF}_{\mathrm{pred}}$ and $\mathrm{AF}_{\mathrm{meas}}$ within the following frequency bin edges: $[0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5, 10, 20]$ (unit: Hz). Within each bin, calculate the averaged amplification factors: $\overline{\mathrm{AF}}_{\mathrm{pred}}$ and $\overline{\mathrm{AF}}_{\mathrm{meas}}$, and then the goodness-of-fit:

$$\mathrm{gof}_{\mathrm{AF}} = 10 - 10\mathrm{erf}\left(\frac{\left|\ln \overline{\mathrm{AF}}_{\mathrm{pred}} - \ln \overline{\mathrm{AF}}_{\mathrm{meas}}\right|}{2\left|\ln \overline{\mathrm{AF}}_{\mathrm{meas}}\right|}\right) \tag{2.6}$$

  Similarly, the factor 2 in the denominator prevents the scores from all being close to 0, while preserving the ranking of the scores.

- Compute the profile's amplification goodness-of-fit, $\mathrm{GoF}_{\mathrm{AF}}$, by averaging $\mathrm{gof}_{\mathrm{AF}}$ across all frequency bins.

The goodness-of-fit of SVM, GTL, CVM-S, and CVM-H profiles against the test set profiles are shown in Figure 2.6 as histograms, where we observe the following patterns:

- The SVM yields satisfactory V_S profile and site amplification predictions at all 43 sites.
- GTL performs less satisfactorily than the SVM.
- CVM-S yields nearly perfect predictions ($\mathrm{GoF}_{\mathrm{Vs}} = 10$, and $\mathrm{GoF}_{\mathrm{AF}} = 9$ or 10) at 9 sites, and poor V_S predictions at about 10 sites ($\mathrm{GoF}_{\mathrm{Vs}}$ close to 0). We explain the reasons for this "binary" performance of CVM-S in the next section.
- CVM-H shows comparable performance as CVM-S, but it does not produce predictions with scores of nearly 10.

Figure 2.7 shows four example sites from the test set. Figure 2.7(a) shows an example where CVM-S yields perfect predictions (to be explained in the following section). Figure 2.7(b) is an example where CVM-S does not match the measurement perfectly. Figure 2.7(c) and (d) represent the case where CVM-S reports a homogeneous stiff layer near the ground surface. The GTL profile in Figure 2.7(c) starts from a reasonable velocity at the ground surface (note that both GTL and the SVM use the same input V_{S30}), but gets more erroneous towards deeper depth, which is due to the unrealistic assumption of GTL that there is no velocity contrast between 0 and 350 m, and is not an issue for the SVM.

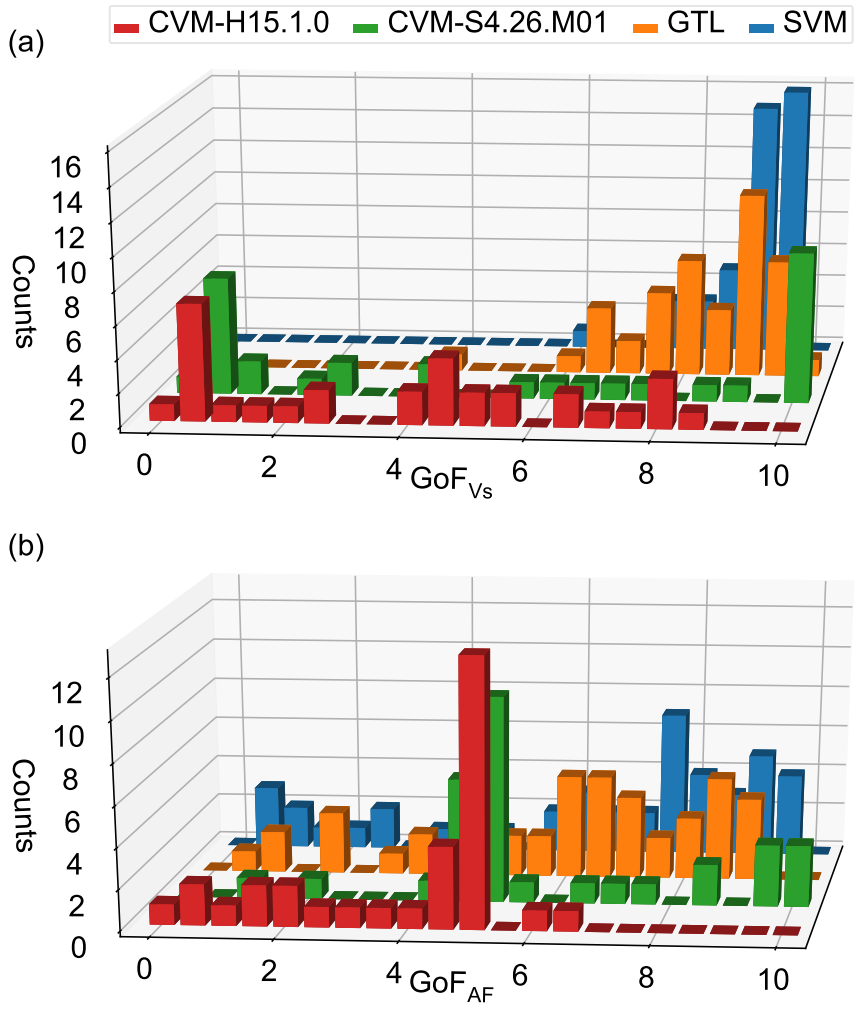


Figure 2.6: Histograms of goodness-of-fit scores, $\mathrm{GoF_{Vs}}$ and $\mathrm{GoF_{AF}}$, for four velocity models: CVM-H15.1.0, CVM-S4.26.M01, GTL, and SVM.

2.3.1 Shallow crustal structure of SCEC CVM-S4.26.M01

To further investigate the phenomenon of CVM-S in Figure 2.7(a), we investigate the locations of all 43 test set sites (Figure 2.8(a)), where we see the perfect V_S matches by CVM-S only occur at some (not all) sites within the Los Angeles basin, and only for profiles measured using invasive techniques (DB and CW, not AY). In other words, at these sites, CVM-S directly incorporates measured V_S profiles from DB and CW databases. To examine how the site-specific properties at these sites affect the neighboring locations in CVM-S, we query 2D cross sections from CVM-S that include these nine perfectly-matched sites. Their locations, cross-sections and individual velocity profiles are shown in Figures 2.8(b), 2.9, and 2.10.

Figure 2.8(b) shows the location of the nine sites (839, 1297, ALH, 929, LAD, RIN, 921,

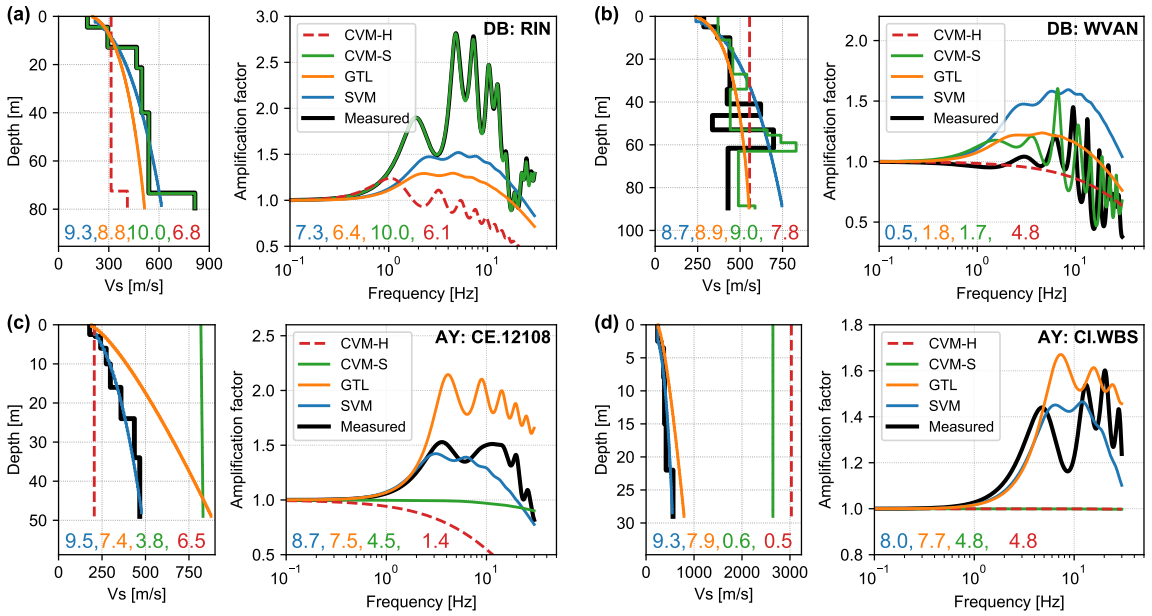


Figure 2.7: Predicted and measured V_S profiles and amplification functions for four of the test set sites: RIN, WVAN, CE.12108, and CI.WBS. The V_S profile comparison is shown on the left of each subplot, and the amplification factor comparison is shown on the right of each subplot. The $\mathrm{GoF_{Vs}}$ and $\mathrm{GoF_{AF}}$ scores corresponding to each V_S model are shown on the bottom of each subplot, using the same color as indicated in the legend.

842, and 861) in the greater Los Angeles area. Figure 2.9 shows 2D V_S profiles of cross sections that pass through these nine sites. Very distinctly, one can see a vertical "streak" at the location of each site, which indicates that at these locations the 1D CVM-S V_S profile is different from the neighboring V_S values. We then query the 1D CVM-S V_S profiles directly adjacent to these nine sites (on the left and right sides along the 2D cross section, and within a 75-meter radius), and plot the 1D profiles together in Figure 2.10. For all nine sites, we observe that the CVM-S profile matches the measured profile, but at only a short distance (< 75 m) away, the CVM-S profile becomes distinctly different from the measured profile. In other words, CVM-S cannot generate the measured (hence most accurate) V_S profiles unless queried exactly at certain grid points. As for 3D ground motion simulations using CVM-S, it is highly unlikely that such sparse inclusion of measured 1D profiles would improve the accuracy of 3D simulations, because such measured profiles would be considered numerical noises of very high frequencies by the wave propagation codes.

To obtain a more realistic estimate of the goodness-of-fit of the CVM-S profiles, we recalculate the goodness-of-fit scores at these nine sites using the adjacent V_S profiles shown in Figure 2.10. We show the updated $\mathrm{GoF_{Vs}}$ and $\mathrm{GoF_{AF}}$ distributions as two violin plots in Figure 2.11. We can see that the SVM yields the highest mean and median goodness-of-fit

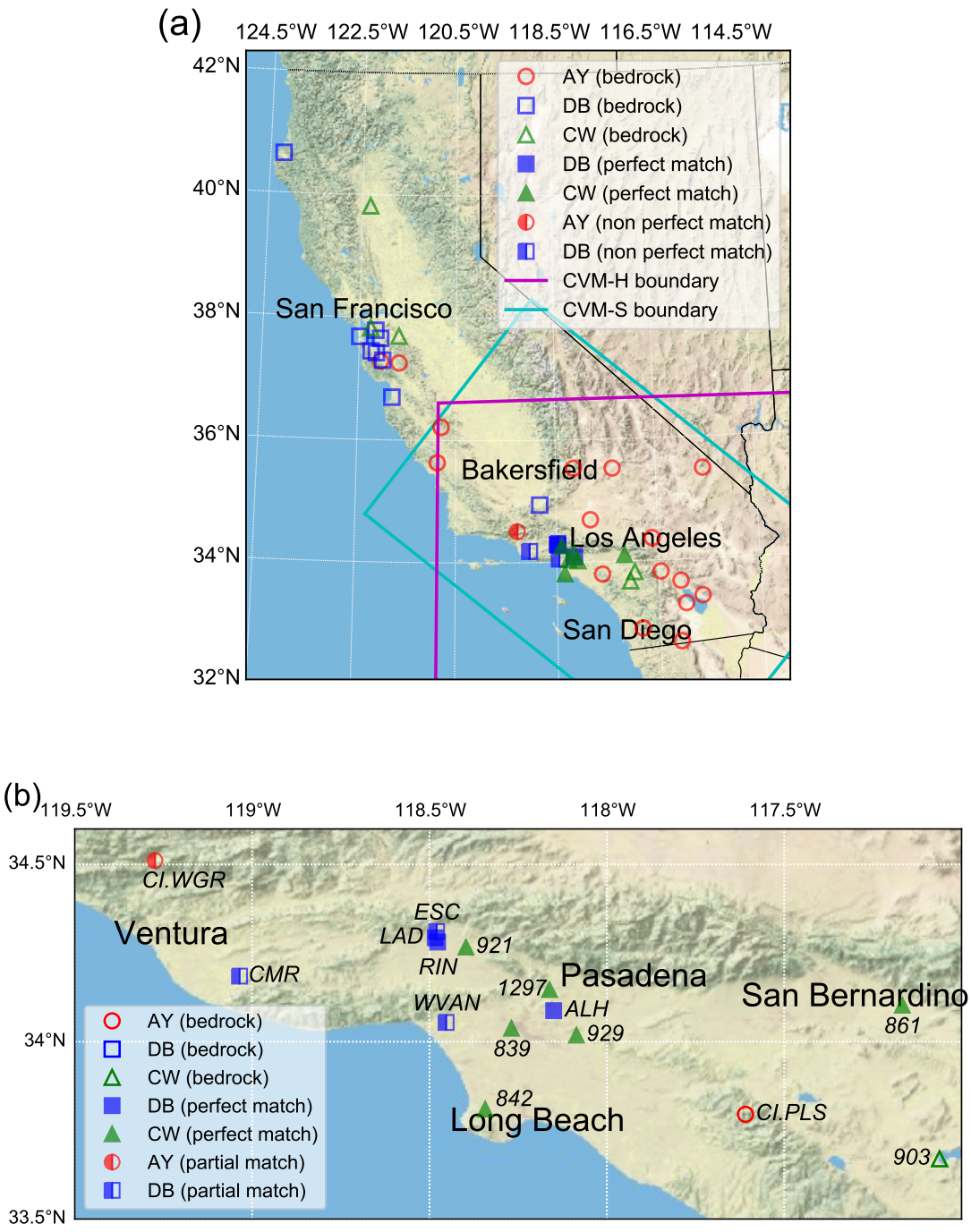


Figure 2.8: (a) The locations of all 43 test set sites. Solid markers indicate sites where CVM-S4.26.M01 produces perfect matches to the measurement, empty markers indicate sites where CVM-S4.26.M01 produces homogeneous stiff V_S profiles ("bedrock"), and half-full markers indicate the non-perfect-match-nor-bedrock case. The magenta and cyan boxes indicate the geographic boundaries of CVM-H15.1.0 and CVM-S4.26.M01, respectively. (b) The locations of the test set sites in the greater Los Angeles area, with site names shown beside the markers.

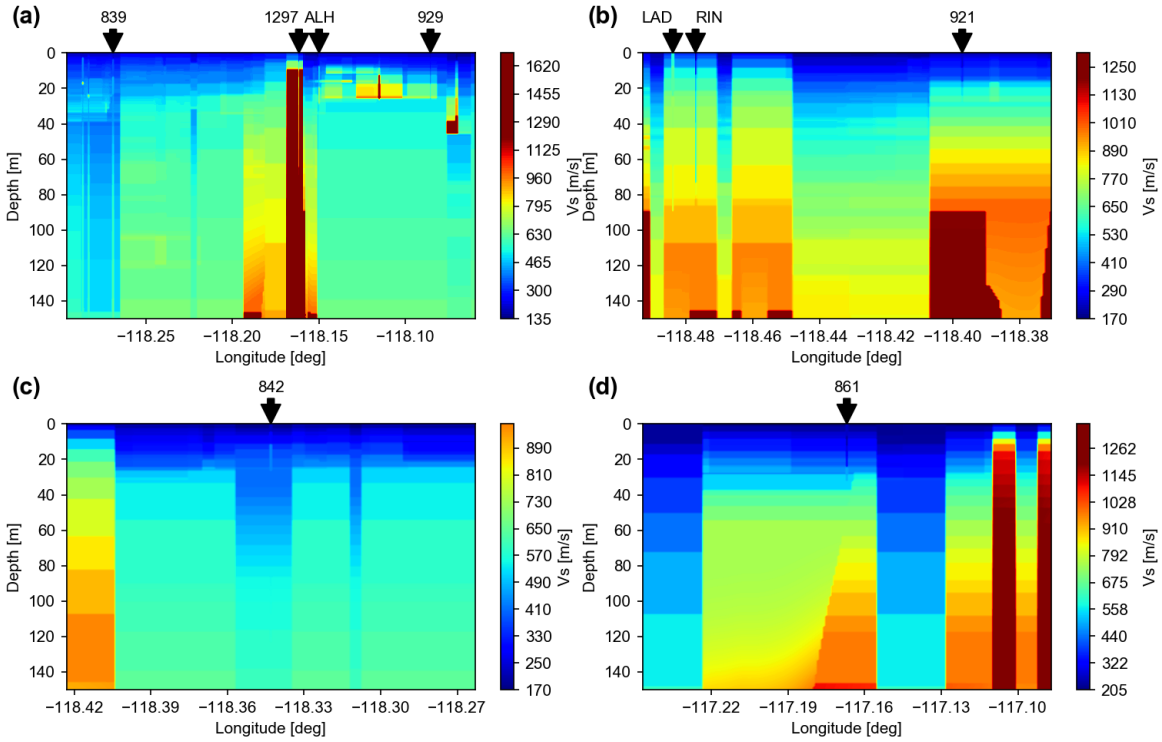


Figure 2.9: 2D V_S profiles of different cross sections that pass through the nine sites (where CVM-S4.26.M01 provides perfect fit to measurements). The nine sites are: 839, 1297, ALH, 929, LAD, RIN, 921, 842, and 861, and their locations are indicated as thick arrows above each color map. Very distinctly, one can see a vertical "streak" at the location of each site, which indicates that at these locations the 1D CVM-S4.26.M01 V_S profile is different from the neighboring V_S values.

scores among the four models (for both GoF_{Vs} and GoF_{AF}), which indicates that the SVM has the potential to provide better predictions of V_{S30}-dependent V_S profiles (GoF_{Vs}) and 1D site amplification, and by extension, improved high-frequency simulated ground motions when translated into a 3D sediment velocity model.

2.4 Stochastic properties of the SVM

In this section, we report some correlations that we find from the data: (1) layer thickness versus depth, and (2) the uncertainty bound of V_S. Admittedly, correlations may not indicate causality: these two correlations may reflect not only the true uncertainties of the sediments but also the "artificial" uncertainties introduced in the V_S profiling process (and it is quite unfeasible to separate them). But for lack of better available alternatives, they are still useful for generating randomized realizations of 1D velocity profiles (and potentially 2D and 3D profiles) based on the smooth SVM profiles (e.g., using the randomization method proposed by *Toro*, 1995).

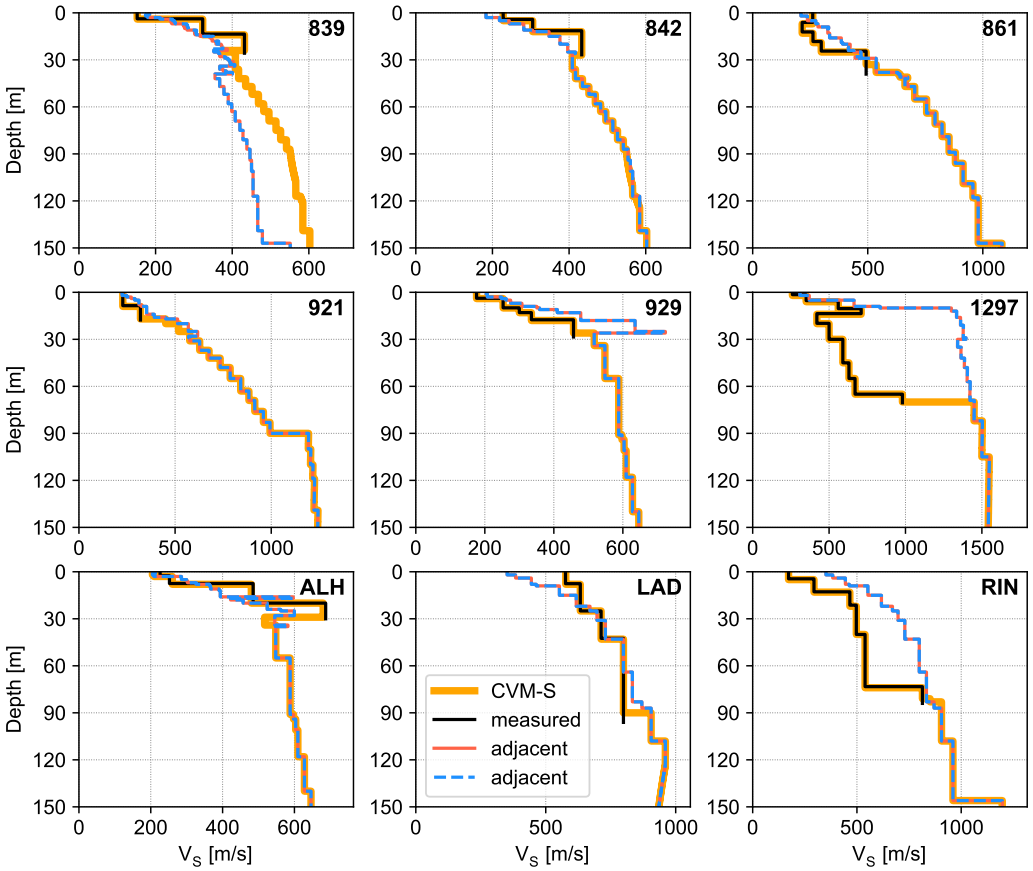


Figure 2.10: 1D V_S profiles at (and next to) the nine test set sites. In each subplot, we show the 1D V_S profile (1) as queried from CVM-S4.26.M01 at the exact latitude/longitude coordinate of the nine sites, (2) adjacent to the latitude/longitude coordinate (on both left and right sides along the 2D cross section, and no farther than 75 m away from the site), and (3) measured V_S profile. For all the nine sites, the CVM-S4.26.M01 profile matches the measured profile, but only a short distance away, the CVM-S4.26.M01 profile becomes distinctly different from the measurement.

2.4.1 Layer thickness versus layer depth

We collect layer thickness (denoted as h) and layer depth (at midpoint of layer, denoted as z_{mid}) of every layer of every measured V_S profile (including test set sites), and perform a "bin-and-fit" procedure:

1. Group the (h, z_{mid}) data points into these z_{mid} bins edges: [0, 10, 20, 30, 40, 50, 65, 85, 100, 120, 160, 200] (unit: m);
2. Calculate the mean values of h and z_{mid} within each bin, assuming h and z_{mid} are normally distributed;
3. Curve-fit the means of h against the means of z_{mid} using the least-square method.

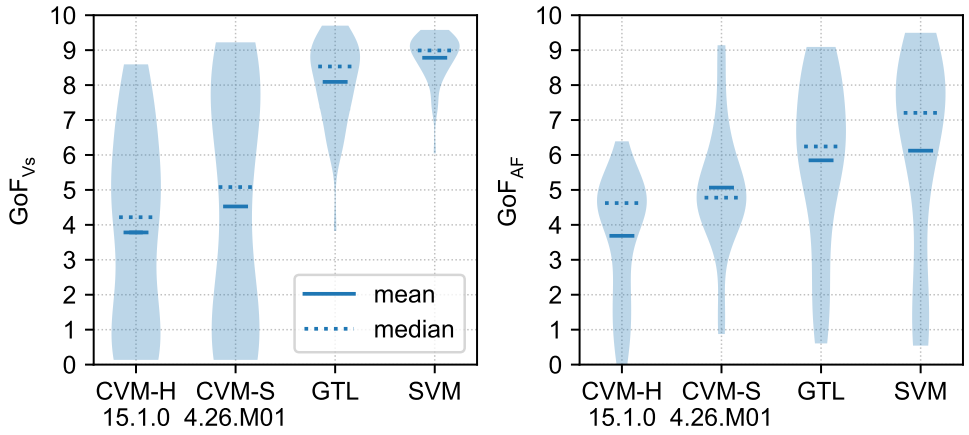


Figure 2.11: Violin plots of the goodness-of-fit scores ($\mathrm{GoF}_{\mathrm{Vs}}$ and $\mathrm{GoF}_{\mathrm{AF}}$) for CVM-H, CVM-S, GTL, and SVM. The shaded areas represent the distribution of scores, and the horizontal bars represent the mean and median scores.

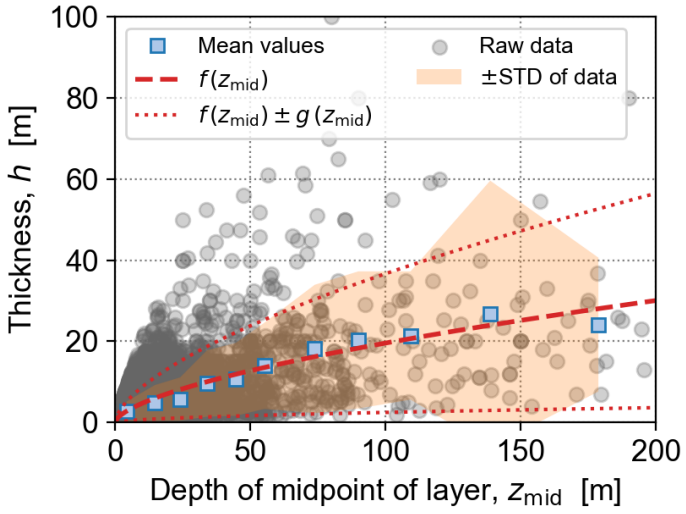


Figure 2.12: Correlation between layer thickness h and layer midpoint depth z_{mid}. The curve fitting results are also shown ($f(z_{\mathrm{mid}})$ and $g(z_{\mathrm{mid}})$ are defined in Eqs 2.7 and 2.8).

The raw data and curve fit are shown in Figure 2.12. The correlation between h and z_{mid} is

$$h = f(z_{\mathrm{mid}}) = 1.125\, z_{\mathrm{mid}}^{0.620} \tag{2.7}$$

with $R^2 = 0.942$, and the correlation between STD_h (standard deviation of h) and z_{mid} is

$$\mathrm{STD}_h = g(z_{\mathrm{mid}}) = 0.951\, z_{\mathrm{mid}}^{0.628} \tag{2.8}$$

with $R^2 = 0.694$. The units of h, STD_h, and z_{mid} are all m.

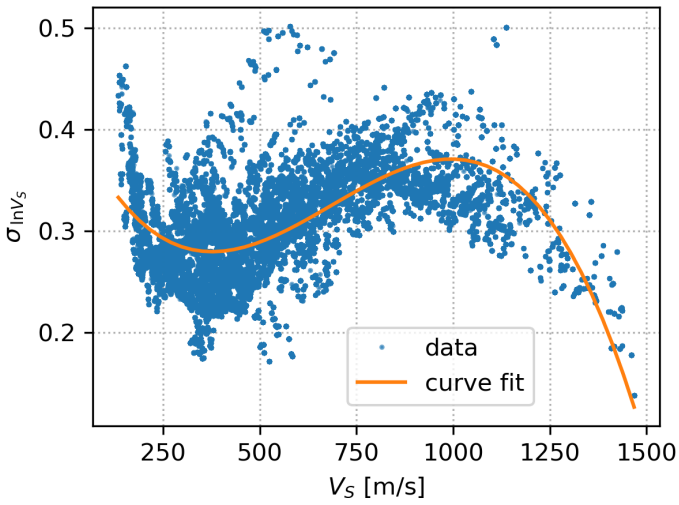


Figure 2.13: Curve fitting between $\sigma_{\ln V_S}$ and V_S. The R^2 score of the curve fit is 0.3747.

2.4.2 Uncertainty bound of V_S

In Figure 2.4, we show the standard deviation envelope of V_S for every V_{S30} bin. Here we gather the standard deviation of V_S (denoted as STD_{Vs}) and correlate it against V_{S30} and z_{mid} (layer depth, see Eq 2.7 for definition):

$$\text{STD}_{Vs} = -89.7085 + 1.6434 z_{\text{mid}} + 0.5204 V_{S30} \tag{2.9}$$

where the units are m, m, and m/s, respectively; and $R^2 = 0.772$. We can see that STD_{Vs} increases with both z_{mid} and V_{S30}, which is also evident in Figure 2.4. Since the valid range of V_{S30} is 173.1 m/s, Eq (2.9) does not produce negative STD_{Vs} values.

Another way to express the uncertainty bound is as follows:

$$\sigma_{\ln V_S} = \left(-7.769 \times 10^{-10}\right) V_S^3 + \left(1.597 \times 10^{-6}\right) V_S^2 - \left(8.7240 \times 10^{-4}\right) V_S + 0.4233 \tag{2.10}$$

where the unit of V_S is m/s. The curve-fitting result for this formula is shown in Figure 2.13.

Note that from the mean SVM velocity (as produced from Eqs 2.3 and 2.4) and STD_{Vs}, one can back-calculate $\mu_{\ln V_S}$ and $\sigma^2_{\ln V_S}$ (the mean and variance of $\ln(V_S)$) from solving the system of Eqs 2.1 and 2.2. This means that Eqs 2.9 and 2.10 reflect different aspects of the same trend.

Also, these two equations do not necessarily suggest any physically significant causality; they merely reflects the trend within the data.

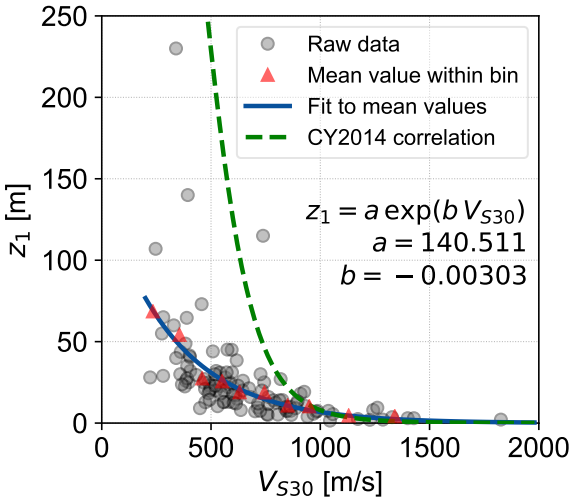


Figure 2.14: Relation between z_1 and V_{S30}. (blue solid line), as calculated from the Vs profile measurements. Mean z_1 value within each V_{S30} bin is calculated (assuming log-normal distribution). A curve can be fitted to the mean value (blue solid line), with $R^2 = 0.97$. We also plot the model proposed by *Chiou and Youngs* (2014) on the same plot, and we can see that their correlation does not match our data so well.

2.5 z_1-V_{S30} relationship

For each measured Vs profile, we collect their V_{S30} and z_1 information[1] and plot the data points in Figure 2.14. (Note that z_1 information is not always available for every measured Vs profile, because some measurements do not reach the depth where V_S exceeds 1000 m/s. Such profiles are naturally excluded from Figure 2.14.) In order to find a mathematical expression between them, we employ the the following "bin-and-fit" procedure:

1. Group the data points into V_{S30} bins: [0, 250, 400, 500, 600, 700, 800, 900, 1000, 1250, 1500];
2. Calculate the mean z_1 value within each bin, assuming the data points follow log-normal distribution;
3. Fit the mean z_1 values (against the arithmetic mean z_1 values of each bin) using least-square method.

The result is shown as a blue line in Figure 2.14, and the correlation is

$$z_1 = 140.511 \exp\left(-0.00303 V_{S30}\right) \tag{2.11}$$

[1] z_1: the depth that the V_S profile first reaches 1,000 m/s, sometimes also known as the "basin depth".

This result means that z_1 (depth to bedrock) decreases when V_{S30} increase (i.e., stiffer sites), which conforms with our intuition. We also plot the model proposed by *Chiou and Youngs* (2014) on the same plot, and we can see that their model does not match our data so well. We should note here that we only included 166 (out of 914) measured Vs profiles which reached 1000 m/s (so that we could infer z_1). For the remaining 748 profiles, the bedrock is deeper than the measurement, thus the "actual" z_1 may be larger than the prediction given in Equation (2.11).

2.6 Summary

In this chapter, we describe the development of a V_{S30}-based near-surface shear wave velocity model, namely the Sediment Velocity Model (SVM), based on 914 measured 1D V_S profiles in California. The data sources are: two publicly available USGS reports (*Boore* 2003; *Yong et al.* 2013), one California Geological Survey dataset (personal correspondence), and proprietary data from LeRoy Crandall and Associates (personal correspondence).

Using V_{S30} as input, the SVM can produce one-dimensional V_S profiles that are shown to be in good agreement with the measured data. The values of V_{S30} can be readily queried from V_{S30} maps (such as *Yong et al.*, 2012; *Thompson et al.*, 2014; *Wills et al.*, 2015). The reliable range for input for the SVM V_{S30} is 173.1 to 1,000 m/s (i.e., outside this range, we do not have statistically abundant data to develop or validate the predicted V_S profiles).

We also conducted a validation study, using 43 out of the 914 measured profiles as benchmark ("test set") to test the predictive performance of four velocity models: CVM-H15.1.0, CVM-S4.26.M01, GTL (by *Ely et al.*, 2010), and SVM. Our validation study shows the following:

1. In many cases within their geographic ranges, CVM-H15.1.0 and CVM-S4.26.M01 predict stiff rocks (V_S > 1000 m/s), whereas the measured profiles show soft sediments, indicating a lack of resolution for both models.

2. GTL requires both V_{S30} and $V_{S(350)}$ as input, which makes it unusable in cases where $V_{S(350)}$ is not readily available. And even with accurate $V_{S(350)}$ information, GTL can still predict unrealistic V_S profiles due to its inherent assumption of no velocity contrast between 0 and 350 m.

3. The goodness-of-fit of V_S profiles (GoF_{Vs}) of the SVM is the highest among the four models, suggesting that it has the potential to produce better 2D/3D velocities for

ground motions simulations than the available sediment velocity models in southern California.

4. The goodness-of-fit of amplification factors (the GoF_{AF} score) of the SVM is also the highest among the four models, suggesting that it will likely lead to improved 1D site amplification factors.

To take advantage of the improved prediction accuracy of the SVM, one can use the SVM-generated 1D profiles to populate the shallow depths of a 2D or 3D region. Specifically, one can use the SVM profiles to replace the velocities of the 3D regional velocity models (such as the CVMs) from the ground surface down to the depth where V_S reaches a value between 1,000 m/s to 1,500 m/s. However, for softer sites (with lower V_{S30}), estimating V_S values up to 1,000 m/s requires some extrapolation (as seen in Figure 2.4), which is due to the limited coverage of V_S values in the four datasets.

We should also note that the model parameters of the SVM are calibrated using V_S profile measurements from California. Hence, if it is used outside California, its predictions could be less accurate. However, the methodology presented in this chapter is not region-specific: with statistically abundant V_S profile measurements from other regions, one could follow the same procedures to develop a separate SVM. In a less fortunate scenario where there are not enough V_S data, the SVM parameters in this chapter can at least be used to give a crude estimate of the potential level of site amplification in regions characterized by soft basin deposits overlying stiff bedrock.

Chapter 3

A hybrid hyperbolic nonlinear soil model for site-response analyses

This chapter presents the hybrid hyperbolic (HH) model—a 1D stress-strain model that can capture both the small-strain stiffness and large-strain shear strength of soils. Its ability to capture soil strength enables it to satisfactorily model nonlinear site response even for very strong input motions: in our benchmarking study (Section 3.3), the HH model outperforms the currently prevalent MKZ model by a considerable margin. More importantly, in the absence of laboratory test data, the HH model parameters can be empirically calibrated with only V_S information, making it particularly useful on a regional scale, where geotechnical data are sparse.

The contents of this chapter is adopted from our publication, *Shi and Asimaki* (2017):

J. Shi and D. Asimaki (2017), "From stiffness to strength: Formulation and validation of a hybrid hyperbolic nonlinear soil model for site-response analyses", *Bulletin of the Seismological Society of America*, 107(3), 1336–1355.

Contents of this chapter

3.1 Introduction

Site response has long been known to play an important role in modifying the amplitude, frequency and duration of earthquake shaking. For non-liquefiable sites, studies have shown that site response frequently amplifies the low-frequency components of weak-to-medium-intensity motions, and introduces complex patterns of amplification and deamplification for higher intensity motions associated with extensive soil yielding (*Beresnev and Wen*, 1996; *Field et al.*, 1997; *Hartzell*, 1998; *Hartzell et al.*, 2004).

Near-surface site response primarily affects the high-frequency components of ground shaking (>1 Hz). Since these components are frequently captured by stochastic methods in earthquake simulations (*Liu et al.*, 2006; *Graves and Pitarka*, 2010; *Mai et al.*, 2010), site response analyses have been traditionally performed separately from the source and path components of simulated ground motions. For well-characterized soil sites, the ground surface motion can be computed using elaborate constitutive soil models (e.g., *Seidalinov and Taiebat*, 2014). For practical applications, however, where geotechnical site characterization data are scarce and sparse, the use of elaborate geotechnical models introduces unavoidably large uncertainties in the selection of input parameters. Simplified one-dimensional site-specific models with few input parameters do exist, but these models achieve simplicity by focusing on a specific strain range of soil behavior. For example, elastic-perfectly-plastic models such as the Mohr-Coulomb (*de Coulomb*, 1776) or the Drucker-Prager (*Drucker and Prager*, 1952) have been developed to match the material strength measured in quasi-static laboratory tests; while the Ramberg-Osgood (*Ramberg and Osgood*, 1943), the hyperbolic (*Kondner and Zelasko*, 1963) and the modified hyperbolic models (*Matasovic and Vucetic*, 1993), have been developed with emphasis on the low to medium strain range response of soils to cyclic loading (e.g., on the basis of resonant column test results).

In this chapter, we present a new one-dimensional (1D) total stress analysis model that addresses the limitations of existing simplified site-specific models to simultaneously capture the low-strain (stiffness) and the large-strain (strength) response of soils; and furthermore, it does so using the shear-wave velocity (V_S) profile as the only input parameter. Given V_S, the so-called hybrid hyperbolic model (HH) captures the fundamental physics of strain-dependent stiffness and strength by means of empirical correlations that have been validated through laboratory experiments and field tests.

Not limited to a specific strain range, our model is thus applicable for site response analyses to low, medium, and strong intensity shaking including near-field motions, provided that the site conditions and incident motions can be approximately captured by a 1D model. In addition to site-specific problems, HH can be used as part of hybrid ground motion prediction equations (GMPEs), to numerically extend site amplification factors to long return period events. In the following sections, we present the formulation of the model, and an extensive validation study of its performance using 2,756 seismic records from 9 KiK-net strong motion stations in Japan. Using this statistically significant database of ground motions, we compare HH's predictions to the widely used equivalent linear and a series of nonlinear constitutive models for site-specific response analyses.

3.2 A hybrid hyperbolic (HH) stress-strain model

3.2.1 Nonlinear stress-strain models for site response analyses: state of the art

Numerous one-dimensional shear stress-strain models have been proposed in the last 50 years. Among others, the hyperbolic model (a.k.a., the KZ model) originally proposed by *Kondner and Zelasko* (1963), has been extensively used because its formulation is simple and its parameters, A and B, reflect physical material properties:

$$\tau(\gamma) = \frac{\gamma}{\frac{1}{A} + \frac{\gamma}{B}} \tag{3.1}$$

where τ is the shear stress and γ is the shear strain. Figure 3.1 shows the example of a KZ stress-strain curve. In this formulation, the tangent slope of $\tau(\gamma)$ equals to A at $\gamma = 0$, and it asymptotically converges to B as $\gamma \to +\infty$. Thus, when used to approximate shear stress-strain soil behavior, A can be set equal to the initial (or maximum) shear modulus of the soil, $G_{\max}$, and B equal to the shear strength, τ_f. Equation (3.1) can then be rewritten as:

$$\tau(\gamma) = \frac{\gamma}{\frac{1}{G_{\max}} + \frac{\gamma}{\tau_f}} = \frac{G_{\max}\gamma}{1 + \frac{G_{\max}}{\tau_f}\gamma} \tag{3.2}$$

When calibrated to match both stiffness ($G_{\max}$) and strength (τ_f), however, KZ often lacks the necessary geometric flexibility to capture the intermediate strain range. To address this issue, *Matasovic and Vucetic* (1993) proposed a modified hyperbolic model (hereafter referred to as MKZ), with two additional "curvature parameters", β and s. The functional form of MKZ is shown in Equation (3.3):

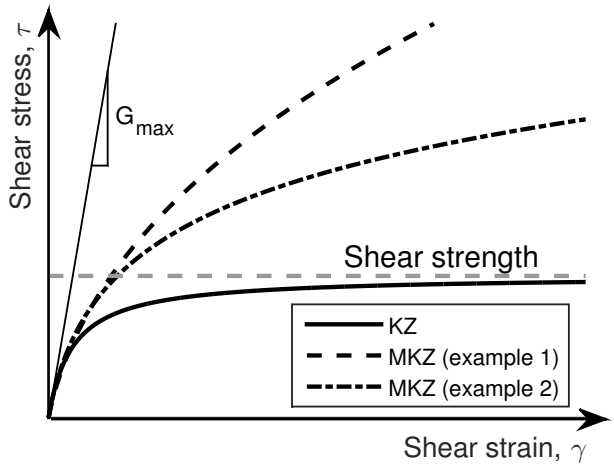


Figure 3.1: Examples of KZ and MKZ curves. The KZ and MKZ curves have the same initial tangent slope, i.e., $G_{\max}$. The KZ curve converges asymptotically towards the shear strength, while the MKZ curves do not have upper bounds and thus will increase to infinity.

$$\tau_{\mathrm{MKZ}}(\gamma) = \frac{G_{\max}\gamma}{1 + \beta\left(\frac{G_{\max}}{\tau_{\mathrm{f}}}\gamma\right)^{s}} \tag{3.3}$$

It is worth noting that the popular study by *Darendeli* (2001), which provided the functional form of modulus reduction and damping curves for generic soils as a function of soil properties (such as PI, the plasticity index, p'_0, the *in-situ* overburden stress, and OCR, the over-consolidation ratio, namely, the ratio between the maximum past overburden pressure and p'_0), is based on MKZ's functional form with $\beta = 1$ and $s = 0.9190$.

However, while the tangent slope of Equation (3.3) at $\gamma = 0$ is $G_{\max}$ (the low-strain soil stiffness), $\tau(\gamma)$ is unbounded for $\gamma \to +\infty$. Since Equation (3.3) does not converge to a finite strength at infinite strain, *Matasovic and Vucetic* (1993) introduced the soil shear strength through an auxiliary reference strain defined as $\gamma_{\mathrm{ref}} = \tau_{\mathrm{f}}/G_{\max}$, and Equation (3.3) becomes:

$$\tau_{\mathrm{MKZ}}(\gamma) = \frac{G_{\max}\gamma}{1 + \beta(\gamma/\gamma_{\mathrm{ref}})^{s}} \tag{3.4}$$

Recognizing that τ_{MKZ} is unbounded, *Matasovic and Vucetic* (1993) and *Darendeli* (2001) validated MKZ only in the low-to-medium strain range, namely, below 0.5%. (0.5% is approximately the largest strain that can be mobilized in resonant column (RC) tests, which were the kinds of tests used by *Darendeli* (2001) to derive empirical correlations for the MKZ parameters; *Matasovic and Vucetic* (1993), on the other hand, used data from cyclic direct simple shear (cDSS) tests that can mobilize up to 1% strain to validate their model.)

Medium to strong ground shaking, however, can induce strains beyond 0.5%. For example, as we will show in more detail later in this chapter, the 2003/9/26 Mw 8.3 Hokkaido earthquake induced more than 3% strain at seismic station KSRH10 (in Hamanaka, Hokkaido, Japan). To simulate motions of such intensity, the stress-strain model needs not only to be able to predict the small-strain soil behavior, but also: (a) the mobilized shear stress at larger strains (beyond the range of applicability of MKZ), and (b) the soil shear strength; thus imposing a physical upper bound to the maximum ground response to seismic shaking.

Recognizing these issues, several recent studies have proposed either entire new stress-strain models or modifications of previous formulations based on MKZ. For example, *Hashash et al.* (2010) proposed to manually adjust the functional form of MKZ beyond 0.1% strain so that it matches the correct soil shear strength. Although this approach does capture the correct soil strength, the stress-strain behavior at intermediate strains is artificially (geometrically) constructed to merge the low and large strain ends, and thus may not reflect the "true" soil behavior. *Stewart and Kwok* (2008) and *Yee et al.* (2013) proposed and employed a similar geometry-based modification to MKZ in the 0.3–0.5% transition strain zone between low-strain stiffness and material strength. This method also captures the soil strength, but its stress prediction for larger strains has not been validated against laboratory data. Additionally, by geometric construction, this method is prone to "discontinuities" (e.g., stress singularity, negative stress, or stress decreasing with strain), especially if the MKZ-predicted stress exceeds the target shear strength at very small strains (which could be, for example, the case for the more confined, deeper layers of soil profiles).

More recently, *Gingery and Elgamal* (2013) and *Groholski et al.* (2016) proposed two additional shear strength incorporated models, but neither study has shown that the models are flexible enough to fit laboratory data over the entire strain range. On the other hand, the strength-incorporated hybrid model proposed by *Hayashi et al.* (1994) was both geometrically flexible and was validated against laboratory stress-strain data; however, they did not provide correlations between the model parameters and physical soil properties. This limitation constrains the usability of their model to cases where laboratory stress-strain data are available and reliable. Then, when such data are not available (as is most frequently the case in site response analyses), we need a stress-strain model that is not only geometrically flexible enough, but also is previously validated against laboratory data, and formulated on the basis of parameters with physical meanings. This was the motivation for the development, calibration, and validation of the HH model presented next.

3.2.2 The formulation of the hybrid hyperbolic (HH) model

Motivated by the need to capture the dynamic soil behavior over the entire strain range (from stiffness to strength), and in light of the above-mentioned limitations of existing models, we have developed a nonlinear stress-strain model formulated as a composite of MKZ in the low-to-medium strain range, and a higher order, flexible KZ model (heretofore referred to as FKZ) in the large strain range. As such, the new model, coined hybrid hyperbolic (HH), takes advantage of the flexibility of MKZ in the low-to-medium strain range, gradually transitions into FKZ, and asymptotically converges to the shear strength of soil. The novelty of HH lies not only in the flexibility of its functional form to fit a wide range of laboratory stress-strain data, but also in its ability to be calibrated with only shear-wave velocity (V_S) information, which can yield realistic stress-strain curves when laboratory data are unavailable.

The functional form of the HH model is:

$$\tau_{\mathrm{HH}}(\gamma) = w(\gamma) \cdot \tau_{\mathrm{MKZ}}(\gamma) + [1 - w(\gamma)] \cdot \tau_{\mathrm{FKZ}}(\gamma) \tag{3.5}$$

where $\tau_{\mathrm{MKZ}}(\gamma)$ is the MKZ stress (defined in Equation (3.4)), $w(\gamma)$ is a transition function to be defined later, and $\tau_{\mathrm{FKZ}}(\gamma)$ is the aforementioned new "flexible hyperbolic" (FKZ) model, intended for modeling large-strain behaviors:

$$\tau_{\mathrm{FKZ}}(\gamma) = \frac{\gamma^d \cdot \mu}{\frac{1}{G_{\max}} + \frac{\gamma^d \cdot \mu}{\tau_{\mathrm{f}}}} \tag{3.6}$$

Comparing Equation (3.6) to Equation (3.2) (KZ model), we see that FKZ has two additional parameters, μ and d. FKZ is thus a generalized hyperbolic model that collapses to the original two-parameter KZ model when $\mu = d = 1$. The physical meaning and evaluation of μ and d will be presented in the next sub section.

The transition function from the stress-strain response of MKZ to FKZ is defined as:

$$w(\gamma) = 1 - 1/\left[1 + 10^{-a\left(\log_{10}(\gamma/\gamma_{\mathrm{t}}) - 4.039a^{-1.036}\right)}\right] \tag{3.7}$$

where γ_{t} and a are two parameters, the "transition strain" and "rate of transition" respectively. The formulation of $w(\gamma)$ comes from modifying an S-shape function, $s(x) = 1/(1 + 10^{-x})$, and fixing the point where $s(x)$ starts to deviate from the asymptote (and the two coefficients, 4.039 and 1.036, in Equation (3.7) come from this process).

Referring to Figure 3.2, $w(\gamma) = 1$ when $\gamma \geqslant \gamma_{\mathrm{t}}$, and it transitions to $w(\gamma) = 0$ at a rate controlled by a. This means that $\tau_{\mathrm{HH}} = \tau_{\mathrm{MKZ}}$ when $\gamma \leqslant \gamma_{\mathrm{t}}$, and transitions into

τ_{FKZ} when $\gamma > \gamma_t$. On the basis of the functional form of HH, γ_t could be physically interpreted as the strain beyond which MKZ is not considered a reliable representation of soil behavior (approximately 0.01%–3%). (The hybrid model by *Hayashi et al.* (1994) also has a transition function, $e^{-\alpha\gamma}$, but contrary to HH, it deviates from 1 at $\gamma = 0$.) Given the extensive literature available on the development and calibration of MKZ (*Matasovic and Vucetic*, 1993; *Darendeli*, 2001; *Vardanega and Bolton*, 2013), we will next focus on the derivation and calibration of FKZ and the transition function. We will base these derivations on laboratory tests on large-strain soil behavior, namely, direct simple shear (DSS) and triaxial (TX) tests, which usually measure shear strain from ~ 0.05% to 20%.

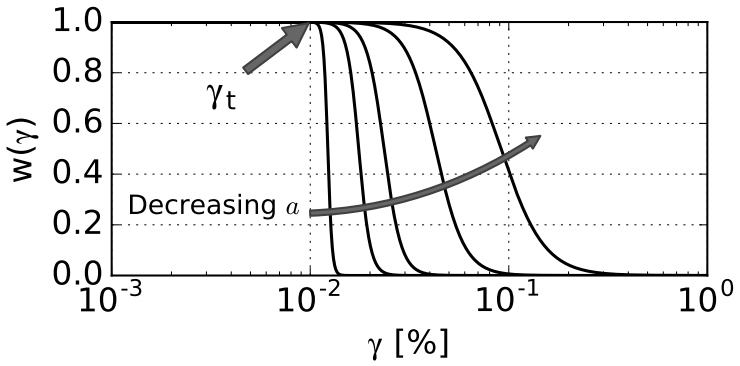


Figure 3.2: The transition function, $w(\gamma)$

Combining the functional forms of MKZ and FKZ, each formulated to capture a different range of shear strain, we circumvent the dilemma of flexibility versus simplicity without the parameters losing their clear physical meanings. Table 3.1 summarizes the nine (9) parameters of the HH model, and Figure 3.3 shows an example of the hybrid stress and hybrid modulus reduction curve predicted by HH.

3.2.3 HH prediction of large-strain shear stress and shear strength

As mentioned above, laboratory tests that measure low (e.g., RC test) and medium/large strain behavior (e.g., DSS, TX tests) only overlap over a very narrow strain range, in the vicinity of 0.5%. Furthermore, in RC tests, soil behavior is affected by large strain rate (i.e., dynamic) effects, whereas DSS tests are performed quasi-statically. Although RC and DSS data can be merged by careful consideration of rate effects (*Vardanega and Bolton*, 2013),

Symbol	Meaning
G_{max}	Initial stiffness
γ_{ref}	Reference strain, related to soil properties
β	Shape parameter of MKZ
s	Shape parameter of MKZ
γ_t	Transition strain from MKZ to FKZ
a	Rate of transition from 1 to 0
τ_f	Shear strength
μ	From FKZ, representing stress accumulation
d	Shape parameter of FKZ

Table 3.1: Summary of the nine (9) parameters of the hybrid hyperbolic (HH) model

it is rare to find RC and DSS test data performed on the same soil, subjected to the same state of stress and stress history. Given the constraints above, we initially validated HH by separately evaluating MKZ and FKZ in their corresponding strain ranges against laboratory experiments.

Figure 3.4 shows the calibration of FKZ using four sets of high quality stress-strain data measured in DSS tests (from *Ladd and Edgers*, 1972; *McCarron et al.*, 1995). Two of the FKZ parameters can be readily obtained from data: τ_f is the peak shear stress (note that FKZ and hence HH assumes no post-peak stress reduction), and G_{max} is calculated as follows,

$$G_{max} = \frac{G^*}{G^*/G_{max}} \tag{3.8}$$

where G^* is the initial secant modulus of the DSS test, which is different from the initial tangent modulus G_{max} measured in RC or TS tests. And we estimate G^*/G_{max} by the following MKZ formula:

$$\frac{G^*}{G_{max}} = \frac{1}{1+\beta\,(\gamma^*/\gamma_{ref})^s} \tag{3.9}$$

where γ^* is the smallest shear strain in the DSS test dataset, and γ_{ref} can be calculated using PI, OCR, and p'_0 (all provided in the DSS test dataset) from Darendeli's correlation (Equation (A5) in the Appendix).

The third parameter of FKZ, μ, can be calculated as follows: *Vardanega and Bolton* (2011) related the steepness of the DSS stress-strain curve to OCR and p'_{m0} (definition under

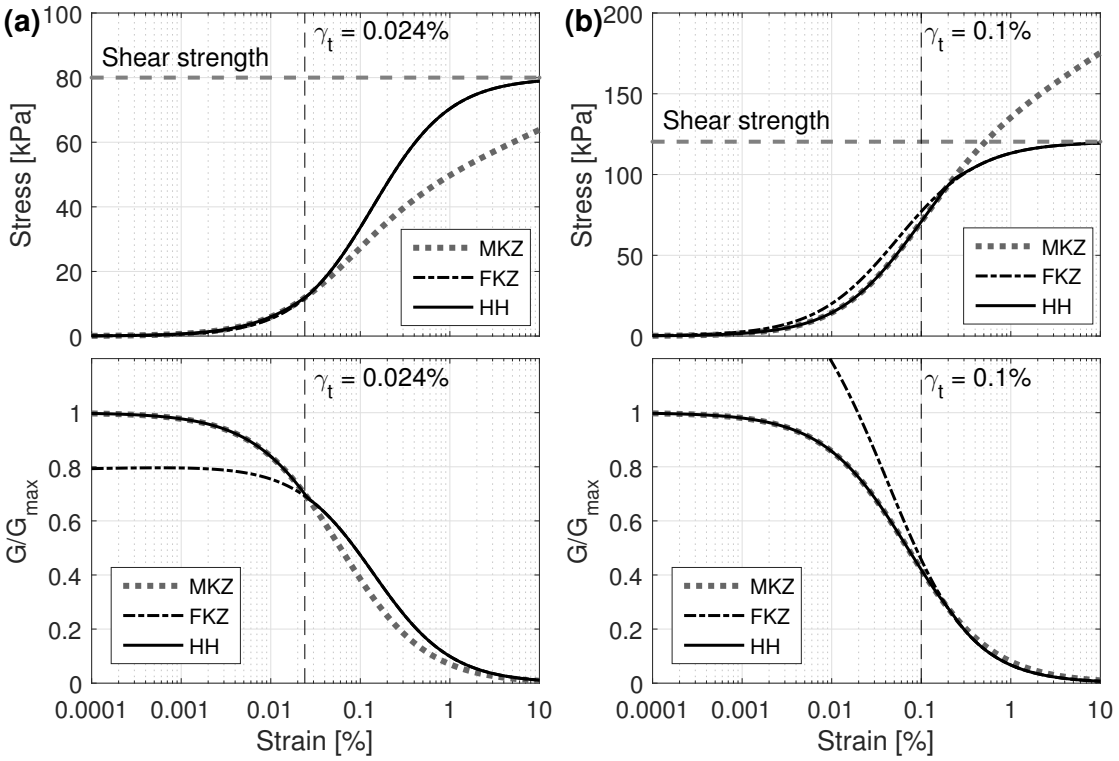


Figure 3.3: Hybrid Hyperbolic (HH) stress and modulus reduction curves, compared to MKZ. (a) Example for shallower layers, where Darendeli's MKZ curve usually underestimates shear stresses even at 10% strain. (b) Example for deeper layers, where Darendeli's MKZ curve usually overestimates shear stresses.

Equation (A5)),

$$\gamma_{M=2} = 0.000872(\mathrm{OCR})^{0.47}\left(p'_{\mathrm{m0}}\right)^{0.28} \tag{3.10}$$

where $\gamma_{M=2}$ is the strain at which 50% of the shear strength is mobilized. Combining Equations (3.10) and (3.6), we have:

$$\mu = \frac{1}{0.000872\left(G_{\max}/\tau_{\mathrm{f}}\right)(\mathrm{OCR})^{0.47}\left(p'_{\mathrm{m0}}\right)^{0.28}} \tag{3.11}$$

where p'_{m0} is in kPa, and $G_{\max}/\tau_{\mathrm{f}}$ and OCR are dimensionless. This indicates that μ controls how fast shear stress accumulates when shear strain increases. Then μ and τ_{f} represent, respectively, two important soil properties in higher strain range: (a) stress accumulation, (b) shear strength.

With $G_{\max}$, τ_{f}, and μ calculated as above, d remains the only free parameter in FKZ. And then we estimate d by least-squares curve fitting. As shown in Figure 3.4, the fit is very satisfactory. The values of d fall in a narrow range (mean = 1.03, standard deviation = 0.12), and within this range, d did not show clear correlations with any of the soil properties we

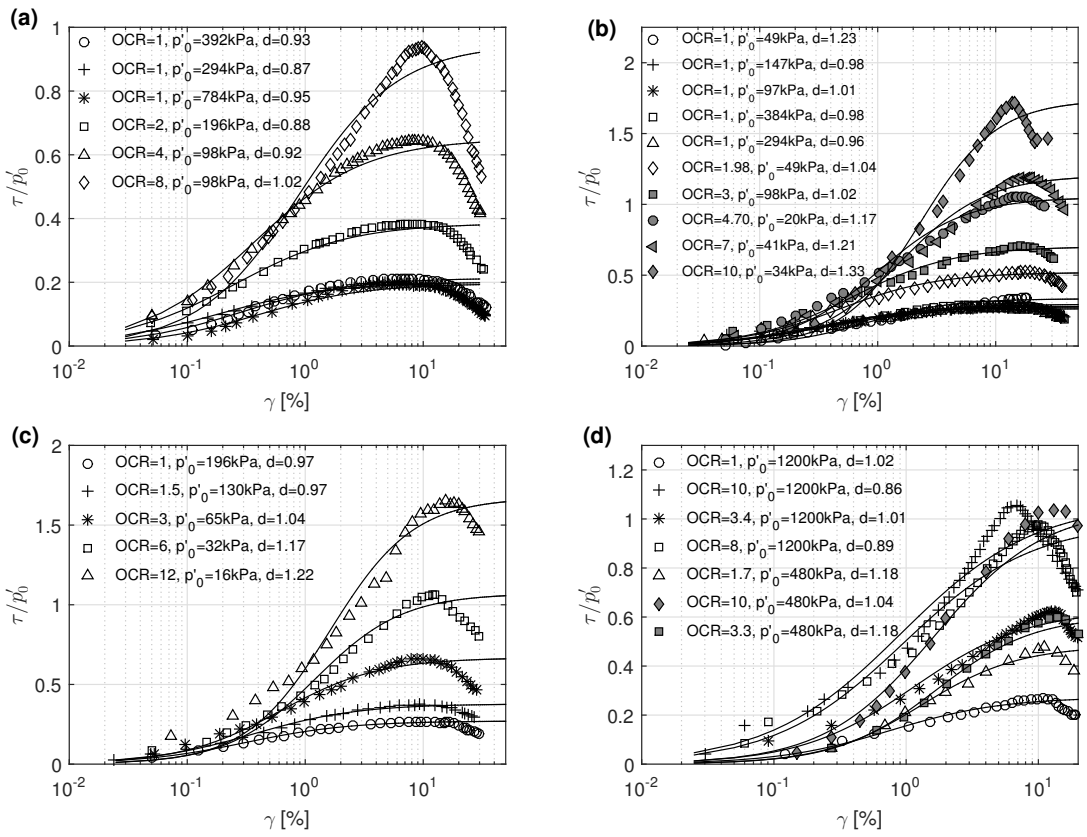


Figure 3.4: Curve fitting of FKZ to direct simple shear (DSS) testing data. (a) Boston blue clay (data from *Ladd and Edgers*, 1972). (b) Maine organic clay (*Ladd and Edgers*, 1972). (c) Bangkok clay (*Ladd and Edgers*, 1972). (d) Beaufort Sea clay (*McCarron et al.*, 1995).

considered. Since d is the power of μ in the FKZ formula, we associated the variation of d with either uncertainties not captured in Equation (3.11) or errors in the DSS test data.

To further test the validity of Equation (3.11), we perform a blind stress-strain prediction (fixing d at the mean value, 1.03) and compare the results with an independent set of DSS data (from *Koutsoftas*, 1978), as shown in Figure 3.5. Although these stress-strain curves have been constructed using the only $G_{\max}$, τ_f, PI, OCR, and p'_0 (all provided in *Koutsoftas*, 1978), the FKZ prediction compares very well with laboratory data.

When stress-strain laboratory data are available in the large strain range, one can directly obtain two of the FKZ parameters, τ_f and $G_{\max}$, while μ and d can be estimated from curve fitting of the same data in lieu of Equation (3.11). On the other hand, in absence of laboratory data, we propose an "HH calibration (HHC)" procedure that "constructs" empirical τ_{FKZ} and thus τ_{HH}. We should highlight here that HHC requires *only* the shear-wave velocity (V_S) profile of the site as input—a situation often encountered by engineers and seismologists

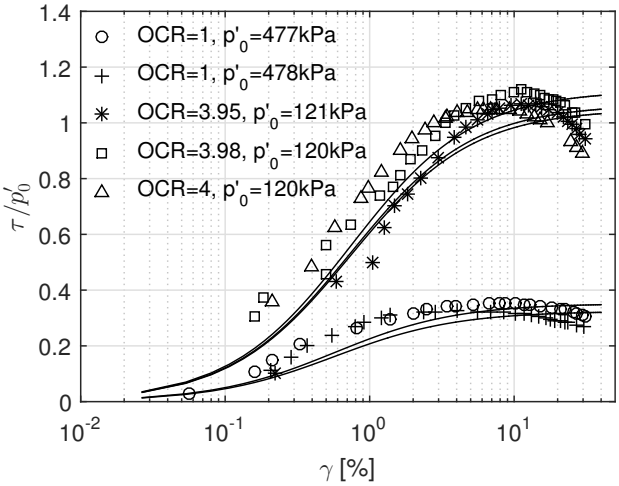


Figure 3.5: Reconstruction of stress-strain curves using the HHC procedure, and the comparison against actual DSS test data (test data from *Koutsoftas*, 1978). The value of μ is calculated using Equation (3.11), and d takes 1.03.

performing site response analysis, just as in the KiK-net validation study in the next section. The detailed formulas of the HHC procedure are presented in the Appendix.

We should also note that the data in Figure 3.4 are from static/quasi-static tests, with strain rate on the order of ~1% per hour. This strain rate is four orders of magnitude slower than the typical strain rate of dynamic site response analyses. To account for the effects of strain-rate in site response (*Richardson and Whitman*, 1963; *Ladd and Edgers*, 1972), *Vardanega and Bolton* (2013) suggested a simple rate-effect correction factor Z (see Appendix for formula) as part of the following correction scheme, which we adopted in the simulations presented in the following sections:

$$\tau_{\text{dynamic}} = \tau_{\text{static}} \cdot Z, \text{ and } G_{\text{dynamic}} = G_{\text{static}} \cdot Z \tag{3.12}$$

Applying this correction to FKZ, Equation (3.6) becomes

$$\tau_{\text{FKZ,dyn}} = Z \cdot \tau_{\text{FKZ,stat}} = \frac{\gamma^d \cdot \mu}{\frac{1}{Z \cdot G_{\max}} + \frac{\gamma^d \cdot \mu}{Z \cdot \tau_f}} \tag{3.13}$$

As we can see, only static $G_{\max}$ and static τ_f need to be scaled up by Z while μ and d are not affected. This means that we can directly use dynamic $G_{\max}$ and dynamic τ_f as HH parameters, without the need to recalibrate μ and d. The procedures for determining dynamic $G_{\max}$ and dynamic τ_f are also listed in the Appendix. In short, in the typical case where only V_S profiles are available, $G_{\max}$ need not be corrected by Z, while τ_f should be corrected by Z.

3.2.4 HH prediction of hysteretic soil damping

Another aspect of the validity of a dynamic stress-strain soil model is its capability to properly capture the hysteretic damping ratio as a function of shear strain. Hysteretic damping ratio (or intrinsic attenuation) is defined as the ratio between the energy absorbed in one loading cycle to the maximum elastic energy stored over the same cycle. Some recent studies, such as *Phillips and Hashash* (2009) and *Li and Assimaki* (2010), proposed hysteretic rules that can fit stress-strain models (such as MKZ) to actual damping data or empirically derived damping curves (e.g., *Seed and Idriss*, 1970; *Vucetic and Dobry*, 1991; *EPRI*, 1993; *Darendeli*, 2001). Such hysteresis rules are capable of simultaneously matching shear modulus and damping data, thus resolving the issue of damping overestimation at higher strains that has been the major drawback of traditional Masing rules (*Masing*, 1926).

Still, even when such non-Masing rules are implemented, the numerical scheme cannot accurately represent the actual damping behavior if the stress-strain model is not geometrically flexible enough to fit the damping data. An example is shown in Figure 3.6, where we observe the poor fit of MKZ to damping of sands (measured data from *Matasovic and Vucetic*, 1993) and clays (design curves from *EPRI*, 1993) at large strains ($> 1\%$). Shown in the same figure is the fit of HH to the same data/curve, which is significantly better than MKZ, particularly in the large strain range ($> 1\%$) where some soils exhibit damping reduction under certain combinations of soil type and overburden stress. This damping reduction has been documented in *EPRI* (1993), and is attributed by *Matasovic and Vucetic* (1993) to the dilative behavior of soils at higher strains. With a monotonically increasing damping prediction with strain, MKZ overestimates soil damping for certain soil types at larger strains, which in turn could lead to an underestimation of ground response to very strong input motions (namely, in the context of this study, motions that mobilize shear strains higher than 1%). In contrast, HH has more parametric flexibility in capturing damping reduction, which ensures a satisfactory damping representation in the numerical scheme.

We should note here that, in absence of measured data, we recommend the empirical formulas proposed by *Darendeli* (2001) (which are documented in the Appendix) to estimate the damping ratios and calibrate the HH model parameters.

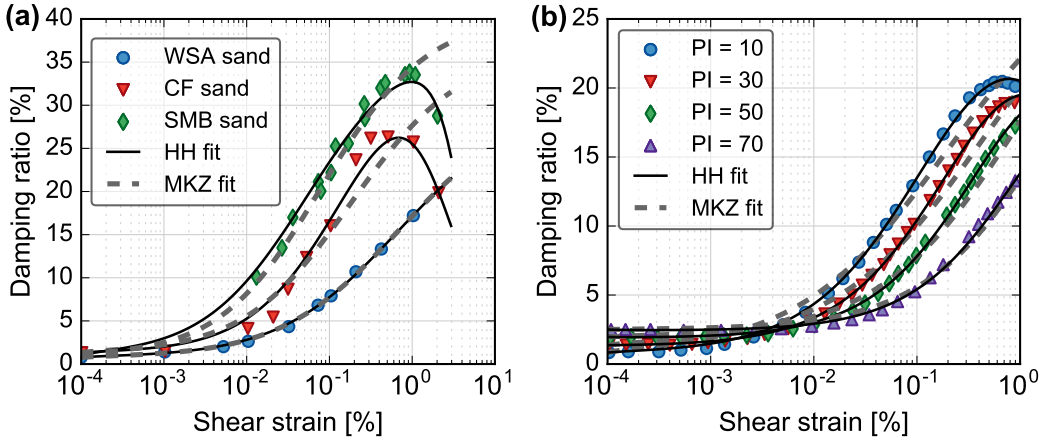


Figure 3.6: Comparison of curve-fitting performance of MKZ and HH models. The HH model shows better flexibility to describe the damping-strain behavior, especially for strains $> 0.1\%$, where soils would exhibit a damping reduction. The MKZ model usually overestimates damping at higher strains. (a) Measured data for different sands (*Matasovic and Vucetic*, 1993). (b) Design curves for different clays (*EPRI*, 1993).

3.3 A comparative study of 1D site response methodologies

In this section, we implement the MKZ and the HH models in nonlinear site-specific response analyses and evaluate their performance by comparing simulation results to surface strong motion recordings at nine downhole arrays of the Japanese strong-motion seismograph network, KiK-net (*Aoi et al.*, 2004). At the same time, we also perform a series of simplified site response analyses (linear and equivalent linear) and we quantify the strain beyond which their performance diverges from the nonlinear predictions and from the recorded ground motions. In each case, we use the subsurface shear wave velocity profile at each station as input and we employ the same empirical correlations to evaluate the modulus reduction and damping curves. Before we proceed with the quantitative comparative results, we first provide an overview of the characteristics of the three family of site response analysis methods used in this chapter:

In the linear method, the material properties remain constant during shaking. Our linear method is in the frequency domain with linear visco-elastic material behavior. It has been repeatedly shown that the linear method is not suitable for site response analyses to strong ground motions, with the possible exception of hard rock sites (NEHRP classes A and B) or very weak motions—PGA at rock outcrop smaller than $\sim 0.1g$ (*Hartzell et al.*, 2004; *Assimaki and Li*, 2012; *Kaklamanos et al.*, 2013).

The equivalent linear method, originally proposed by *Seed and Idriss* (1970), accounts for material yielding (modulus reduction) and hysteretic attenuation (damping) by iteratively matching the soil modulus and damping to a characteristic strain level. Nonetheless, this method essentially still a linear method because material properties remain constant throughout an iteration—although the stiffness is reduced and damping is increased compared to the linear method. This method yields satisfactory results for relative stiff sites subjected to intermediate levels of strain ($< 0.1\%$), but severely underestimates the high frequencies (> 5 Hz) of the ground motion—as will be shown in the following sections.

The nonlinear method is performed in the time domain, where the material properties are adjusted instantaneously to the strain level and loading path. The nonlinear scheme used in this chapter was developed by *Li and Assimaki* (2010) and *Assimaki and Li* (2012) and has the following features:

1. It uses a memory variable technique proposed by *Liu and Archuleta* (2006) to simulate the frequency-independent small-strain soil damping.

2. It can incorporate any stress-strain model with a closed-form expression, including (but not limited to) the MKZ and HH models implemented in this chapter.

3. It uses the *Li and Assimaki* (2010) hysteresis rule that is based on the hysteresis model proposed by *Muravskii* (2005), and can simultaneously match the $G/G_{\max}$ and damping curves. This goal is achieved by using two sets of stress-strain parameters as input for the nonlinear scheme: the first set describes loading/reloading (obtained from fitting a certain stress-strain model to $G/G_{\max}$ measurements, or directly from empirical correlations that generate the stress-strain model parameters), and the second set describes unloading (obtained from fitting the stress-strain model to damping data). Through the geometric representation of narrower and more realistic hysteresis loops, the *Li and Assimaki* (2010) rule yields a better fit than the extended Masing rules (*Pyke*, 1979; *Kramer*, 1996), which seek a "compromised" matching of $G/G_{\max}$ and damping curves. It should be pointed out that the *Li and Assimaki* (2010) hysteresis rule can incorporate any closed-form monotonic stress-strain model, including (but not limited to) MKZ and HH.

3.3.1 KiK-net strong motion stations

The KiK-net strong-motion seismograph network in Japan consists of approximately 670 stations with a ground surface and a downhole array instrument pair (*Aoi et al.*, 2004). To

evaluate the performance of the HH model relative to MKZ and the simplified linear and equivalent linear analyses methods, we identified nine (9) stations for which, according to the taxonomy proposed by *Thompson et al.* (2012), 1D wave propagation is a valid assumption for site response analyses. In Thompson's taxonomy, such an assumption is valid when the theoretical transfer function (TTF; using the Thomson-Haskell method (*Thomson*, 1950; *Haskell*, 1953)) of a horizontally layered profile is in good agreement with the average empirical transfer function (ETF) at the same site, evaluated as the spectral ratio between the surface and borehole measurements (from weaker motions with $\mathrm{PGA}_{\mathrm{surface}} < 0.1g$ only). The site characteristics of the nine stations are shown in Table 3.2, and their V_S profiles are shown in Figure 3.7.

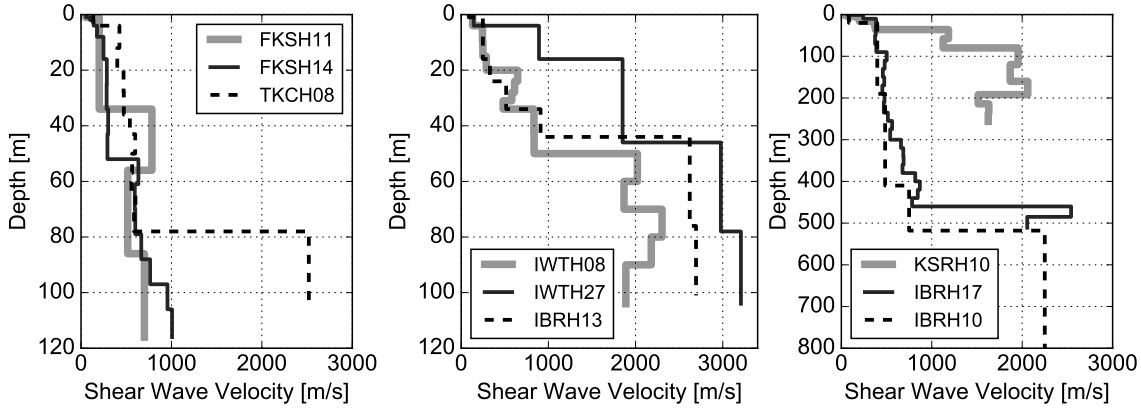


Figure 3.7: V_S profiles of all the nine stations, after waveform inversion

3.3.2 Soil profiles by waveform inversion

Although shear wave velocity profiles are available at the KiK-net stations from suspension logging and downhole tests performed at the time of installation, previous studies (e.g., *Assimaki et al.*, 2006; *Assimaki and Steidl*, 2007; *Thompson et al.*, 2009), have shown that these profiles are sometimes too coarse to capture the effects of site amplification on the higher frequency components of ground shaking even for very weak ground motions. To obtain a more detailed description of the soil profile at the selected stations, we employed the waveform inversion technique proposed by *Assimaki et al.* (2006), using the V_S profiles provided by KiK-net as initial trial profile and five weak ground motions ($\mathrm{PGA}_{\mathrm{surface}} < 0.01g$) across which we averaged the V_S results.

Station	V_{S30} [m/s]	NEHRP class	f_0 [Hz]	N
FKSH11	183.5	D/E	1.6	288
FKSH14	205.2	D	1.2	291
IBRH10	114.2	E	0.33	198
IBRH13	278.5	D	2.8	602
IBRH17	287.8	D	0.37	360
IWTH08	276.5	D	2.9	163
IWTH27	613.6	B/C	7.4	484
KSRH10	158.7	D/E	1.7	188
TKCH08	335.0	C/D	2.0	182

Table 3.2: Summary of the nine (9) KiK-net stations used in this chapter. f_0 is the fundamental frequency of the empirical transfer function (ETF), and N is the number of recorded events used in the numerical simulations.

Figure 3.8(a) shows an example at Station FKSH14 of the TTF before and after waveform inversion, and compares them to the corresponding ETF. We should note here that the ETF in Figure 3.8(a) is computed from all 1697 weak ground motions ($\mathrm{PGA}_{\mathrm{surface}} < 0.1g$) recorded at FKSH14, which is a statistically significant amount to demonstrate the effectiveness of the inversion technique. Figure 3.8(b) then compares the V_S profile before and after waveform inversion. Finally, Figure 3.9 shows the optimized TTF (after waveform inversion) versus the averaged ETF calculated from all weak ground motions at all nine stations.

3.3.3 Ground motion data and goodness-of-fit criteria

For the simulations presented in the following sections, we used ground motions with peak ground acceleration (PGA) higher than $0.01g$ recorded at the nine KiK-net stations between the year 2000 to 2013. The detailed number of records from each station is shown in Table 3.2. For each record, EW and NS components were first rotated into SH and SV components according to the azimuth between the station and the epicenter. Successively, the SH borehole and ground surface components were used as input motion and benchmark response, respectively.

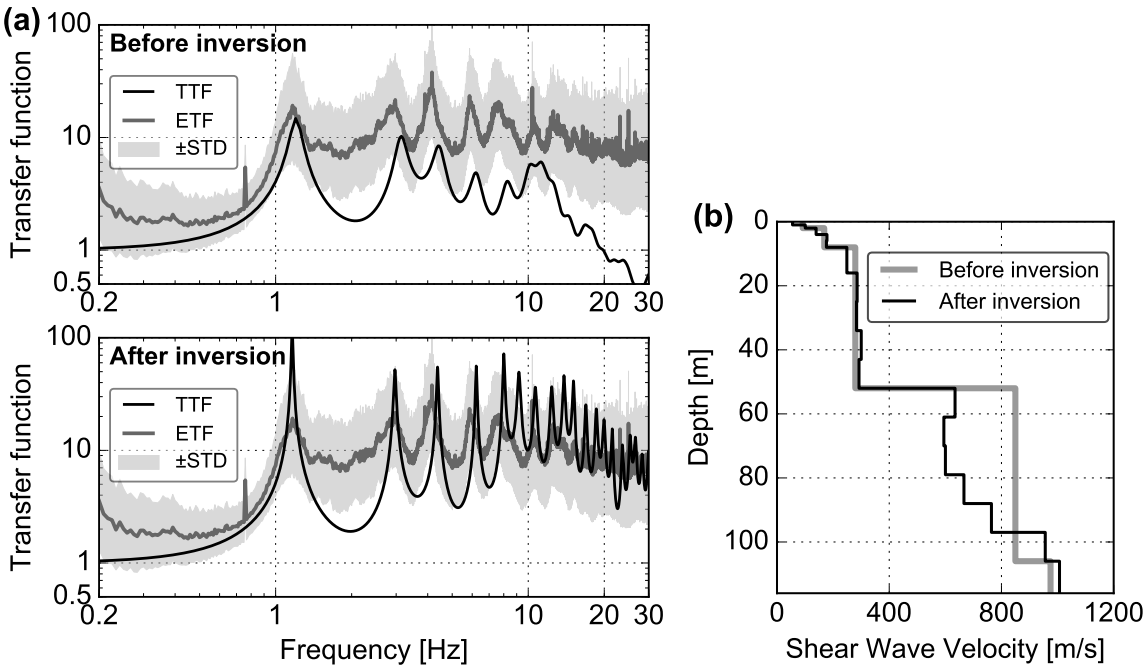


Figure 3.8: An example at FKSH14 which shows that waveform inversion yields a better match between TTF and ETF. (a) Comparison of TTFs (before and after inversion). (b) Comparison of V_S profiles (before and after inversion)

The simulated ground surface motions were quantitatively compared to the benchmark, using a goodness-of-fit (GoF) scheme synthesized from *Anderson* (2004). As in the work of *Kristeková et al.* (2009); *Olsen and Mayhew* (2010); *Taborda and Bielak* (2013) among others, our GoF gauntlet differs from Anderson's original scheme in the following three aspects:

1. We have omitted the cross-correlation between simulation and recording, due to the sensitivity of this GoF measure to small misalignments between waveforms.

2. Instead of comparing PGA, PGV, and PGD, we compare the root mean square (RMS) of the acceleration, velocity, and displacement respectively, namely the dominant rather than peak amplitude of the time series.

3. The range of the score is changed from $[0, 10]$ in the original scheme (with 10 representing perfect match) to $[-10, 10]$. In our scheme, 0 represents perfect fit between simulations and recordings, positive values indicate over-prediction, and negative values indicate under-prediction. The formulas for mapping differences to scores are also changed from an exponential-decay-type function to the error function

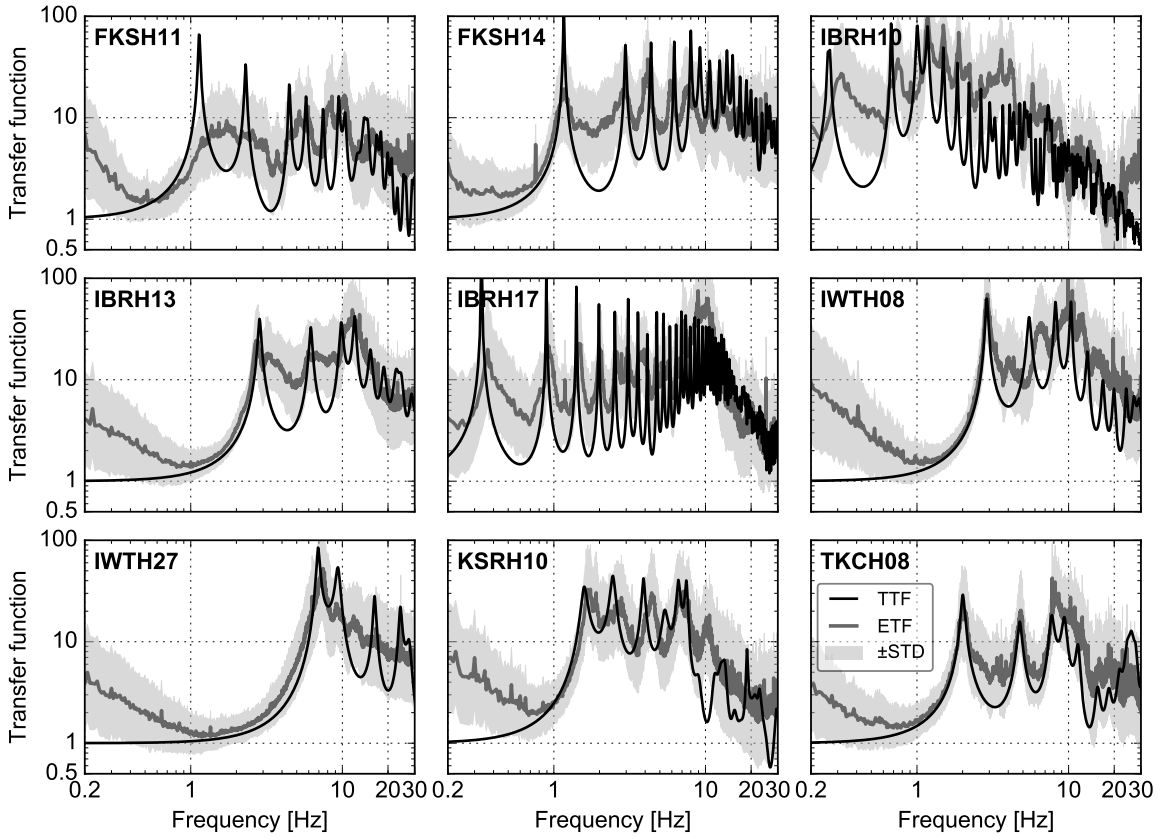


Figure 3.9: Theoretical transfer functions compared to empirical transfer functions (calculated from weak motions, with surface PGA$< 0.1g$) at each station. The shaded gray areas are the $\pm\sigma$ bounds of the averaged ETF.

scaled up by 10 (as shown in Table 3.3), because the error function is "directional" (i.e., under- and over-prediction distinguishable) and symmetric around zero.

The physical meanings of each score (S_1, S_2, $\cdots$, S_9) and the formulas are summarized in Tables 3.3 and 3.4. The average of S_1 through S_9, $\bar{S}_{[B]}$, can be calculated for one pair of simulated and measured waveforms band-pass-filtered through frequency band $[B]$. In other words, $\bar{S}_{[B]}$ provides GoF at frequency range $[B]$. The final score, $\bar{R}$, is the average over five different frequency bands:

$$\bar{R} = \frac{1}{5}\left(\bar{S}_{[0.5\text{-}25]\text{Hz}} + \bar{S}_{[0.5\text{-}2]\text{Hz}} + \bar{S}_{[2\text{-}5]\text{Hz}} + \bar{S}_{[5\text{-}10]\text{Hz}} + \bar{S}_{[10\text{-}25]\text{Hz}}\right) \tag{3.14}$$

Score #	Meaning: Similarity between	Formula
S_1	Normalized Arias intensity time histories, $\bar{A}_\mathrm{m}(t)$ and $\bar{A}_\mathrm{s}(t)$	$\Gamma\left(\bar{A}_\mathrm{m}(t), \bar{A}_\mathrm{s}(t)\right)$
S_2	Normalized energy integral time histories, $\bar{E}_\mathrm{m}(t)$ and $\bar{E}_\mathrm{s}(t)$	$\Gamma\left(\bar{E}_\mathrm{m}(t), \bar{E}_\mathrm{s}(t)\right)$
S_3	Arias intensity (peak values), A^*_m and A^*_s	$\Phi\left(A^*_\mathrm{m}, A^*_\mathrm{s}\right)$
S_4	Energy integral (peak values), E^*_m and E^*_s	$\Phi\left(E^*_\mathrm{m}, E^*_\mathrm{s}\right)$
S_5	RMS acceleration, $\mathrm{rms}(a_\mathrm{m})$ and $\mathrm{rms}(a_\mathrm{s})$	$\Phi\left(\mathrm{rms}(a_\mathrm{m}), \mathrm{rms}(a_\mathrm{s})\right)$
S_6	RMS velocity, $\mathrm{rms}(v_\mathrm{m})$ and $\mathrm{rms}(v_\mathrm{s})$	$\Phi\left(\mathrm{rms}(v_\mathrm{m}), \mathrm{rms}(v_\mathrm{s})\right)$
S_7	RMS displacement, $\mathrm{rms}(d_\mathrm{m})$ and $\mathrm{rms}(d_\mathrm{s})$	$\Phi\left(\mathrm{rms}(d_\mathrm{m}), \mathrm{rms}(d_\mathrm{s})\right)$
S_8	Spectral acceleration (5%), $\vec{R}_\mathrm{m}$ and $\vec{R}_\mathrm{s}$	$\Gamma\left(\vec{R}_\mathrm{m}, \vec{R}_\mathrm{s}\right)$
S_9	Fourier spectra, $\vec{F}_\mathrm{m}$ and $\vec{F}_\mathrm{s}$	$\Gamma\left(\vec{F}_\mathrm{m}, \vec{F}_\mathrm{s}\right)$

Table 3.3: Meanings and formulas of the modified goodness-of-fit (GoF) scores used in this chapter. The subscript "m" stands for "measurement", and the subscript "s" stands for "simulation". Definitions of functions Γ and Φ as well as Arias intensity and energy integral are listed in Table 3.4. rms: root mean square.

Function	Formula	Meaning
$\Phi\left(x_\mathrm{m}, x_\mathrm{s}\right)$	$10 \cdot \mathrm{Erf}\left(\frac{x_\mathrm{s}-x_\mathrm{m}}{x_\mathrm{m}}\right)$	Relative error, scaled to $[-10, 10]$
$\Gamma\left(\vec{y}_\mathrm{m}, \vec{y}_\mathrm{s}\right)$	$\frac{1}{N}\sum_{k=1}^{N}\Phi\left(y_{\mathrm{m},k}, y_{\mathrm{s},k}\right)$	Average relative error of all N points in $\vec{y}_\mathrm{m}$ and $\vec{y}_\mathrm{s}$
$A(t)$	$\frac{\pi}{2g}\int_0^t a^2(\tau)\,\mathrm{d}\tau$	Arias intensity time history, where a is acceleration
A^*	$A(T_\mathrm{d})$	Peak value of $A(t)$, where T_d is the duration of ground motion
$\bar{A}(t)$	$A(t)/A^*$	Normalized Arias intensity time history
$E(t)$	$\frac{\pi}{2g}\int_0^t v^2(\tau)\,\mathrm{d}\tau$	Energy integral time history, where v is velocity
E^*	$E(T_\mathrm{d})$	Peak value of $E(t)$, where T_d is the duration of ground motion
$\bar{E}(t)$	$E(t)/E^*$	Normalized energy integral time history

Table 3.4: Definition of functions Γ and Φ, as well as Arias intensity and energy integral, which appear in Table 3.3. Erf() is the error function.

3.4 Results and discussion

We performed site response predictions for all the motions with measured $\text{PGA}_{\text{surface}} > 0.01g$ at the nine KiK-net stations, using the following analyses: (a) linear (LN), (b) equivalent linear (EQ), and (c) nonlinear (NL). For the EQ and NL analyses, we employed two stress-strain relations: MKZ and HH. Because V_S profiles are the only available soil properties useful for site response analyses at the KiK-net stations, we employ the HHC procedure to obtain MKZ and HH parameters (note that MKZ parameters are a subset of the HH parameters). For the NL analyses, we provide MKZ and HH parameters to the numerical scheme, while for the EQ analyses, we construct MKZ and HH modulus reduction curves from their parameters and use them as reference points for 1D interpolation to iteratively determine the strain-compatible soil properties. Damping values come from Darendeli's empirical correlations (documented in the Appendix), and both the MKZ and the HH models are fitted to the same values, yielding different sets of input parameters. (As is shown earlier in Figure 3.6, HH's fit to the same damping data is better than MKZ's.)

Evidently, there are five different model-analyses pairs (henceforth referred to as "methods"):

1. LN: linear method
2. EQ_{MKZ}: equivalent linear method using the MKZ modulus curves and Darendeli's damping curves
3. EQ_{HH}: equivalent linear method using the HH modulus curves and Darendeli's damping curves
4. NL_{MKZ}: nonlinear method using the MKZ modulus and damping parameters
5. NL_{HH}: nonlinear method using the HH modulus and damping parameters

For each of the five methods above, we then calculated the goodness-of-fit (GoF) scores, $\bar{S}_{[B]}$'s and $\bar{R}$, as defined in Equation (3.14). In what follows, we synthesize and present the scores for each method at each station against two metrics: (1) measured surface PGA, (2) maximum shear strain within the soil column (as estimated by NL_{HH}), γ_{max}.

3.4.1 Goodness-of-fit scores versus PGA

An example of the goodness-of-fit score for each event plotted versus $\text{PGA}_{\text{surface}}$ (PGA recorded by seismometers on ground surface) at FKSH14 is shown in Figure 3.10 for the five methods mentioned above. Each point in the figure corresponds to the response of

the profile to one event, each calculated with a different method (indicated by different markers).

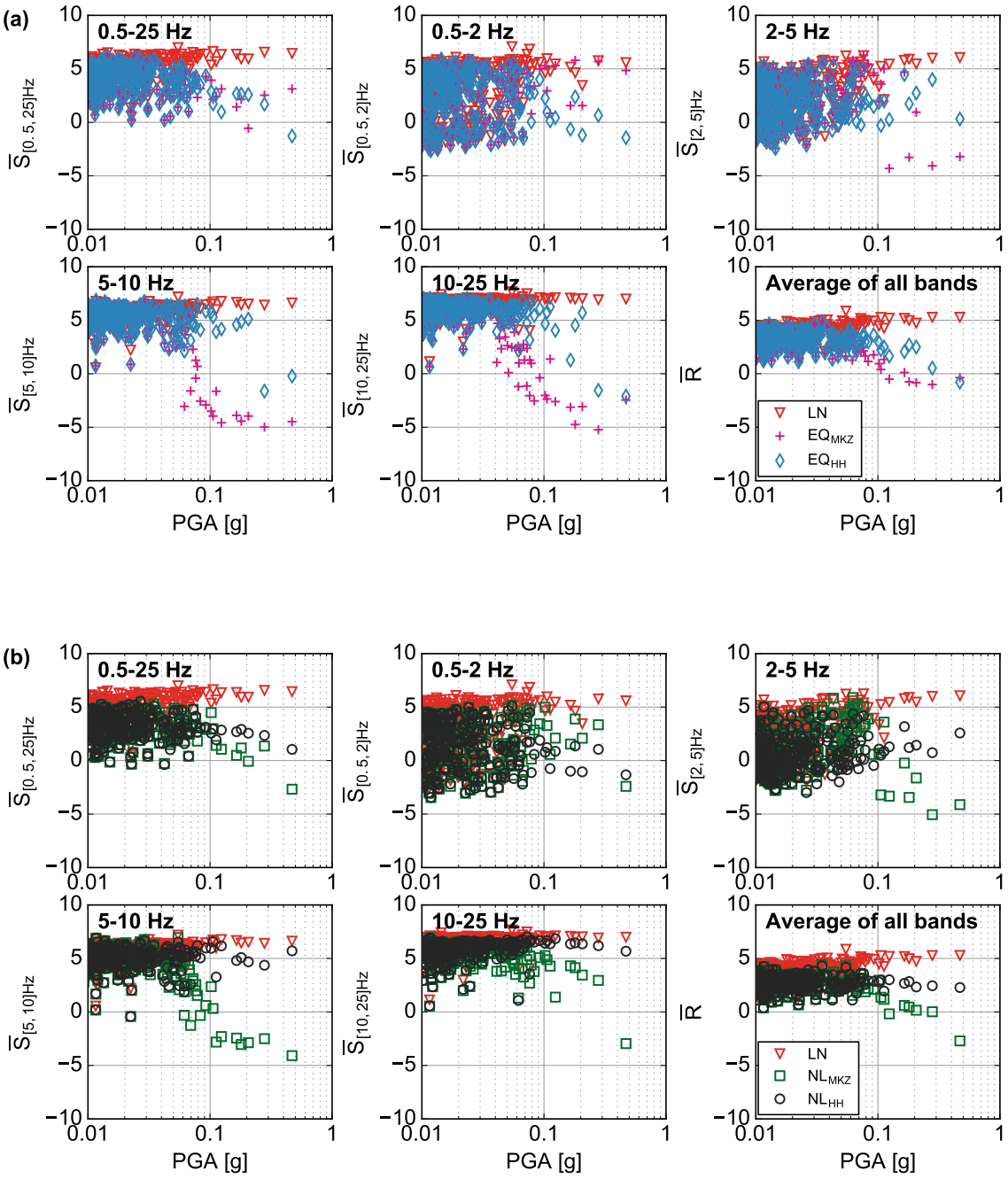


Figure 3.10: Goodness-of-fit scores versus PGA (recorded by the seismometer at ground surface) at station FKSH14, for all five methods. Positive score indicates over-prediction, negative score indicates under-prediction, and 0 indicates perfect fit. (a) LN, EQ_{MKZ} and NL_{MKZ}. (b) LN, EQ_{HH}, and NL_{HH}. (The scores of the other stations are provided in the electronic supplement.)

The LN scores are shown in both Figure 3.10(a) and (b). Clearly, the LN method significantly overestimates the surface ground motion for all frequency bands, a misfit that increases with increasing PGA. This comes hardly as a surprise since the LN method does not consider material degradation (strain-compatible or instantaneous), and thus it cannot reproduce the reduced site amplification or de-amplification observed in the records for stronger ground motions.

Figure 3.10(a) next shows the EQ_{MKZ} and EQ_{HH} scores. The $\bar{R}$ values of all three methods are similar for small PGAs, but start to deviate for PGA $\gtrapprox 0.05g$: $\bar{R}$ by EQ_{MKZ} and EQ_{HH} decrease with PGA, and the EQ_{MKZ} goodness of fit is for the most part lower than the EQ_{HH} fit to the ground surface observations. Also, we observe that for the higher frequencies, both $\bar{S}_{[5\text{-}10]Hz}$ and $\bar{S}_{[10\text{-}25]Hz}$ of EQ_{MKZ} decrease with PGA to almost -5; and $\bar{S}_{[10\text{-}25]Hz}$ and $\bar{S}_{[10\text{-}25]Hz}$ of EQ_{HH} at high PGAs also show a slightly decreasing trend with PGA. This observation confirms that the equivalent linear method can severely underestimate high frequency motions, especially when employed with non-strength-corrected modulus reduction and damping (MKZ in this case) curves.

Figure 3.10(b) then shows the scores of NL_{MKZ} and NL_{HH}. For PGA $< 0.05g$, the two NL methods yield slightly better predictions than LN, EQ_{MKZ}, and EQ_{HH}. For PGA $> 0.05g$, the advantage of NL_{HH} becomes more apparent: its $\bar{R}$ score remains close to 0 (perfect fit) and appears to be independent of PGA. The $\bar{R}$ score by NL_{MKZ} however, is worse than that of NL_{HH}, yet better than EQ_{MKZ} and similar to EQ_{HH}. Also, higher frequency (above 5 Hz) scores of NL_{MKZ} estimated ground motions still indicate an underprediction trend of the observed high frequency content.

Assimaki et al. (2008) proposed a threshold rock-outcrop PGA value of $0.2g$, beyond which nonlinear analyses should be performed in lieu of linear and equivalent linear methods. Similarly, *Kaklamanos et al.* (2013) reported the threshold observed PGA of $0.1g$. In the present study, the threshold PGA for FKSH14 is shown to be $0.05g$, significantly lower than the two previous studies: in this case, nonlinear site response clearly manifests at medium intensity shaking, namely at levels of PGA lower than previously expected.

We should emphasize here that we do not define the threshold PGA where the GoF score "crosses over" zero, but rather at the PGA where the score starts to deviate from its "linear baseline" (i.e., the "baseline" score for very small PGAs). We base this rationale on the fact that this "linear baseline" reflects all the factors contributing to the misfit in the linear range (for example, errors in the small-strain soil properties, deviation from 1D wave propagation conditions, etc.). And since the impact of these factors on the goodness-of-fit usually remains constant with ground motion intensity, the contribution of a specific site response

method to the goodness-of-fit starts to manifest when a "new trend" in the score starts to show. The same rationale applies to the threshold maximum strain (γ_{max}), which will be presented later.

Figure 3.11 plots the GoF score $\bar{R}$ versus $PGA_{surface}$ for all the events at all stations.

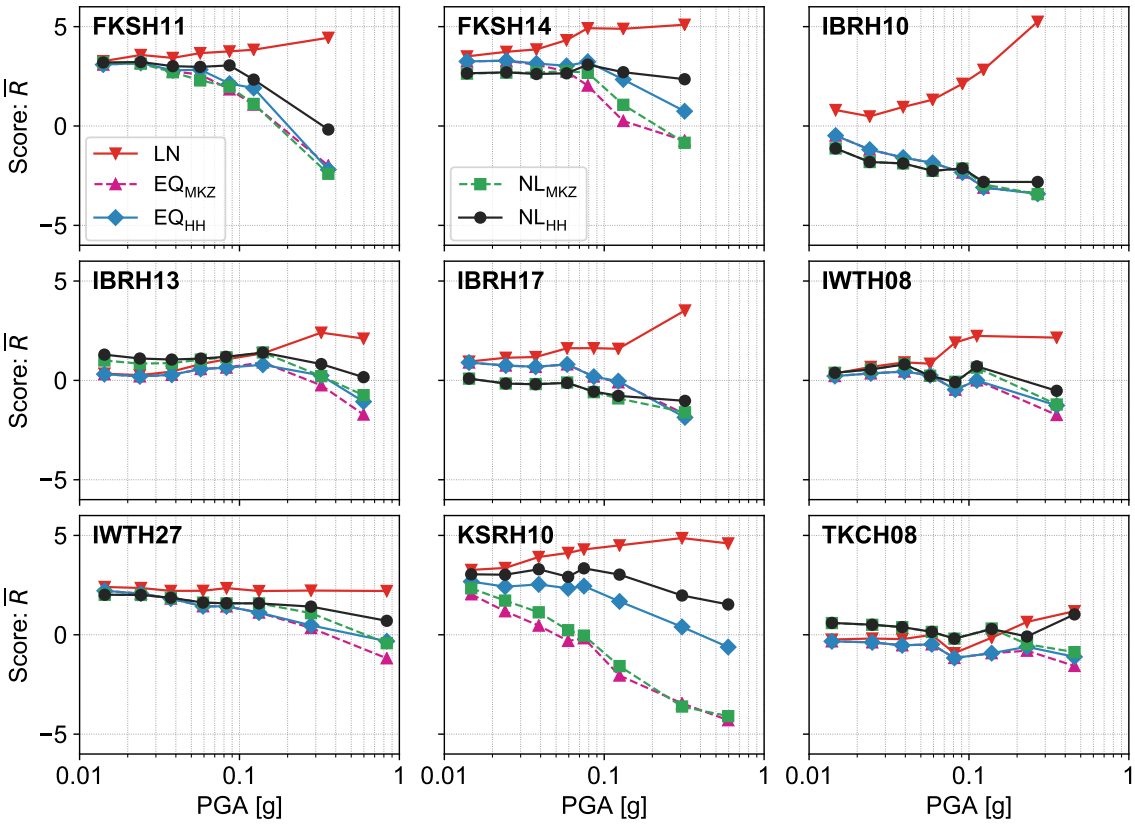


Figure 3.11: Goodness-of-fit scores, $\overline{R}$ (defined in Equation (3.14)), versus PGA (recorded by the seismometers at ground surface) at all nine stations, with all five methods. Positive score indicates over-prediction, negative score indicates under-prediction, and 0 indicates perfect fit. Each marker dot is the average of all the $\overline{R}$ values with a PGA bin (whose center is indicated by the marker). Thus marker dots represent the overall $\overline{R}$ at their respective PGA level, for that particular station.

Each marker symbol is the averaged value of all the $\bar{R}$ values within a PGA bin, representing the overall $\bar{R}$ for that specific PGA level. From this figure we can observe that the $\bar{R}$ by LN is generally the highest (namely, LN overestimates ground motions) except at station TKCH08. For the other four methods, at stations FKSH11, FKSH14, IBRH13, IWTH27, and KSRH10, $\bar{R}$ values are "sensitive" to increasing PGA and are quite different from method to method: LN overestimates ground motions, EQ_{MKZ} and NL_{MKZ} increasingly

underestimate ground motions, EQ_{HH} shows improvement over EQ_{MKZ} and NL_{MKZ}, and NL_{HH} shows even better performance than EQ_{HH}—qualitatively determined since it gives slightly higher $\bar{R}$ than EQ_{HH} and would be considered a conservative (safer) prediction in engineering design. We also observe that NL_{HH} is rather "insensitive" to the increase of PGA, which can be a significant advantage, because we can know *a priori* that the quality of high-PGA simulations by NL_{HH} is similar to the quality of low-PGA simulations, whereas the performance of EQ_{MKZ} or NL_{MKZ} for low-PGA motions is hardly indicative of that of stronger shaking (higher PGAs). For the other four stations (IBRH10, IBRH17, IWTH08, and TKCH08), $\bar{R}$ by all the methods (except LN) are rather "insensitive" to PGA and are "closer together" to each other. Hence, threshold PGA for these nine stations are different, which range from $0.02g$ to more than $0.5g$. This is because the level of soil nonlinearity at these four stations is not as significant as at others stations—and PGA alone is not always a good indicator of soil nonlinearity. This will be discussed in detail in the next subsection.

3.4.2 Goodness-of-fit scores versus maximum strain

If we consider the modulus reduction curves used in both EQ and NL methods, the maximum strain level, γ_{max}, induced by earthquake shaking is a strong indicator of the extent of nonlinearity experienced by the soil column during a given event. For example, a hard rock site may experience only low strains even when subjected to strong ground motions, indicating little (if at all) nonlinearity; whereas a soft site may exhibit rather large strains under moderate shaking, indicating that the site response was strongly nonlinear. To examine the relationship between the intensity of ground motion (quantified by PGA) and the extent of nonlinearity (quantified by γ_{max}), Figure 3.12 plots γ_{max} versus $PGA_{surface}$ for all stations and events presented above. For a given station, γ_{max}-PGA is a linear correlation with narrow scatter. One could thus conclude that plotting GoF vs γ_{max} instead of PGA would result in some scaled version of the GoF-PGA trend presented in the previous subsection. However with a closer look at the Figure 3.12, we observe that for a given PGA, γ_{max} can vary over a factor of 100 from station to station (e.g., compare KSRH10 to TKCH08). Therefore, for inter-station analysis, it is important to factor in different site conditions by using γ_{max} instead of PGA.

Figure 3.13 shows the GoF score ($\bar{R}$) plotted against γ_{max} for all nine stations and five site response methods. The diversity of site conditions provide a wide range of simulated γ_{max}, ranging from $4 \times 10^{-4}\%$ to 3%, which in turn enables us to better assess the performance of each method. Each subfigure corresponds to one method, and each marker corresponds

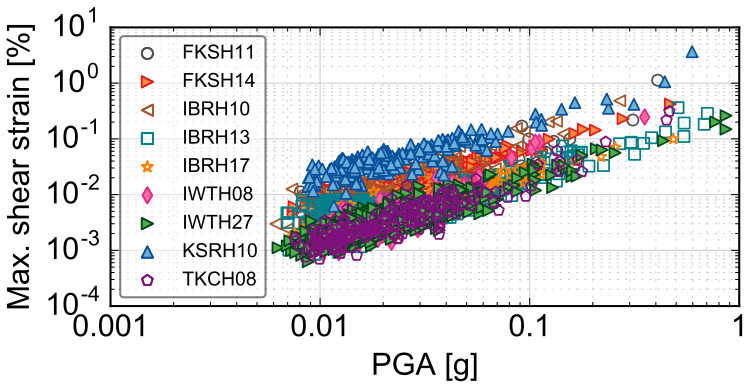


Figure 3.12: Maximum shear strain (γ_{max}) within the soil column versus the corresponding PGA (recorded by the seismometers at ground surface). Each marker symbol corresponds to a certain event at a certain KiK-net station.

to one of the 2,756 events. The black solid line is the average score within each γ_{max} bin, and the dashed lines are standard deviation bounds.

From this figure, we clearly observe that the $\bar{R}$ by LN increases from ~0 to ~5 with increasing strain. EQ_{MKZ} and NL_{MKZ} remain close to 0 until γ_{max} reaches 0.04%, beyond which the score rapidly decreases, indicating that the methods underpredict the ground surface motions for larger strains. Note that this threshold strain of ~0.04% is lower than the previously reported 0.1–0.3% threshold by *Kaklamanos et al.* (2013). EQ_{HH} and NL_{HH} show better performance than EQ_{MKZ} and NL_{MKZ}: not only $\bar{R}$ is close to 0, but also $\bar{R}$-γ_{max} is approximately constant over the entire strain range. Since the only difference between the HH- and MKZ-based methods is the stress-strain model (HH vs MKZ), one can conclude that the difference in the goodness-of-fit is the result of the better soil behavior representation over the entire strain range—from stiffness to strength—of the HH model.

A more detailed examination of Figure 3.13 reveals that $\bar{R}$ by NL_{HH} is better than the corresponding $\bar{R}$ by EQ_{HH} at high strains (> 0.5%). Specifically, while NL_{HH} slightly overpredicts the ground surface motions, EQ_{HH} underpredicts the observed ground shaking, which could lead to unsafe design in engineering practice. Figure 3.14(a) depicts only the salient features of Figure 3.13, by averaging all the $\bar{R}$ values within the each γ_{max} bin. One can clearly see the advantages of NL_{HH} over the other four methods in the large strain range, where their predictions become increasingly poor with increased ground motion intensity. The strain level where this prediction divergence between methods is defined as "threshold strain", γ_{thres}. Figures 3.14(b) and (c) show the score for higher frequencies, $\bar{S}_{[5,10]Hz}$ and $\bar{S}_{[10,25]Hz}$, where we see that even though EQ_{HH} has relative satisfactory performance for $\bar{R}$

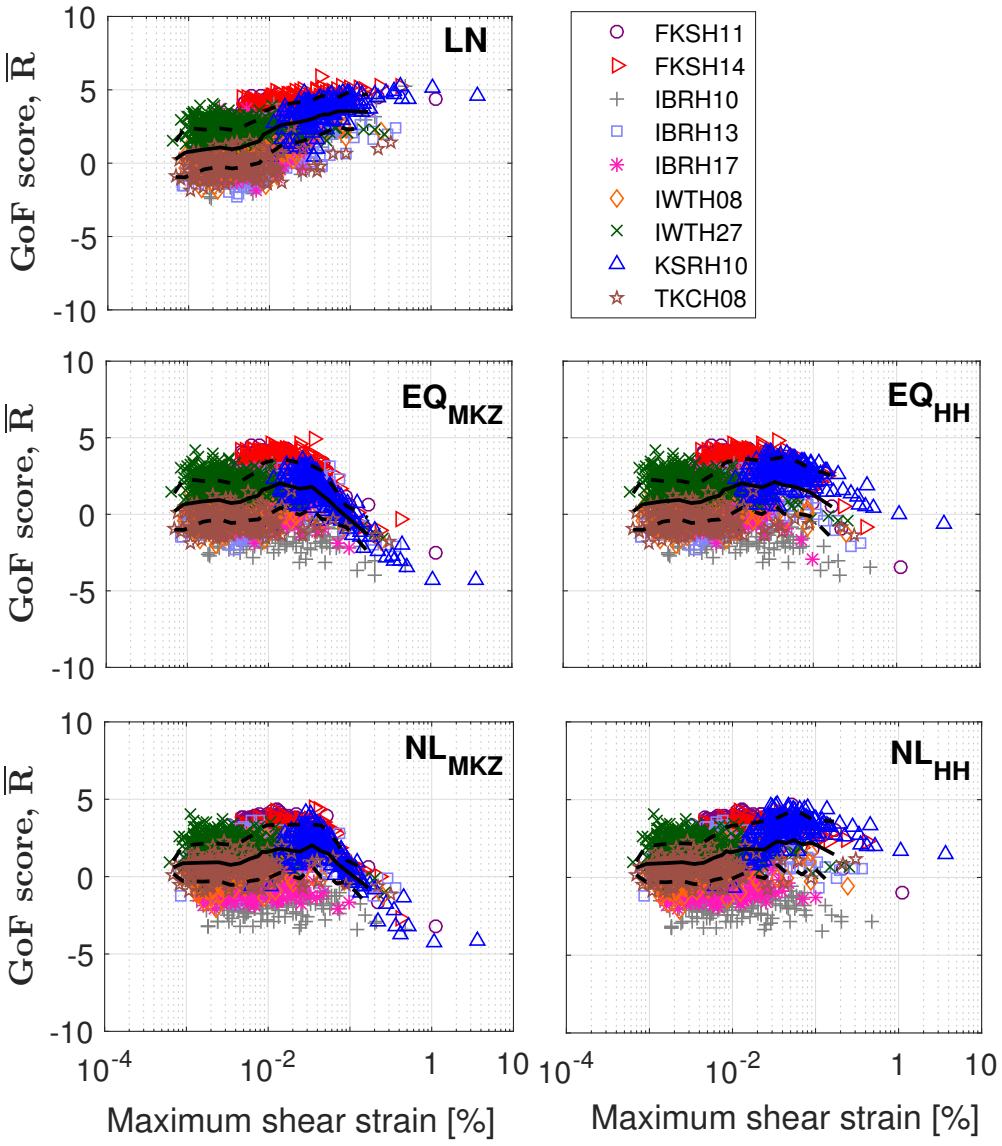


Figure 3.13: Goodness-of-fit score $\overline{R}$ versus $\gamma_{\max}$ within the soil column, for all nine stations and all five methods. Positive score indicates over-prediction, negative score indicates under-prediction, and 0 indicates perfect fit. NL_{HH} provides the best overall performance among the five methods (slight over-prediction is preferred in engineering design). Each marker dot is one event recorded at a certain station. The solid line is the average score within each $\gamma_{\max}$ bin, and the dashed lines are standard deviation bounds.

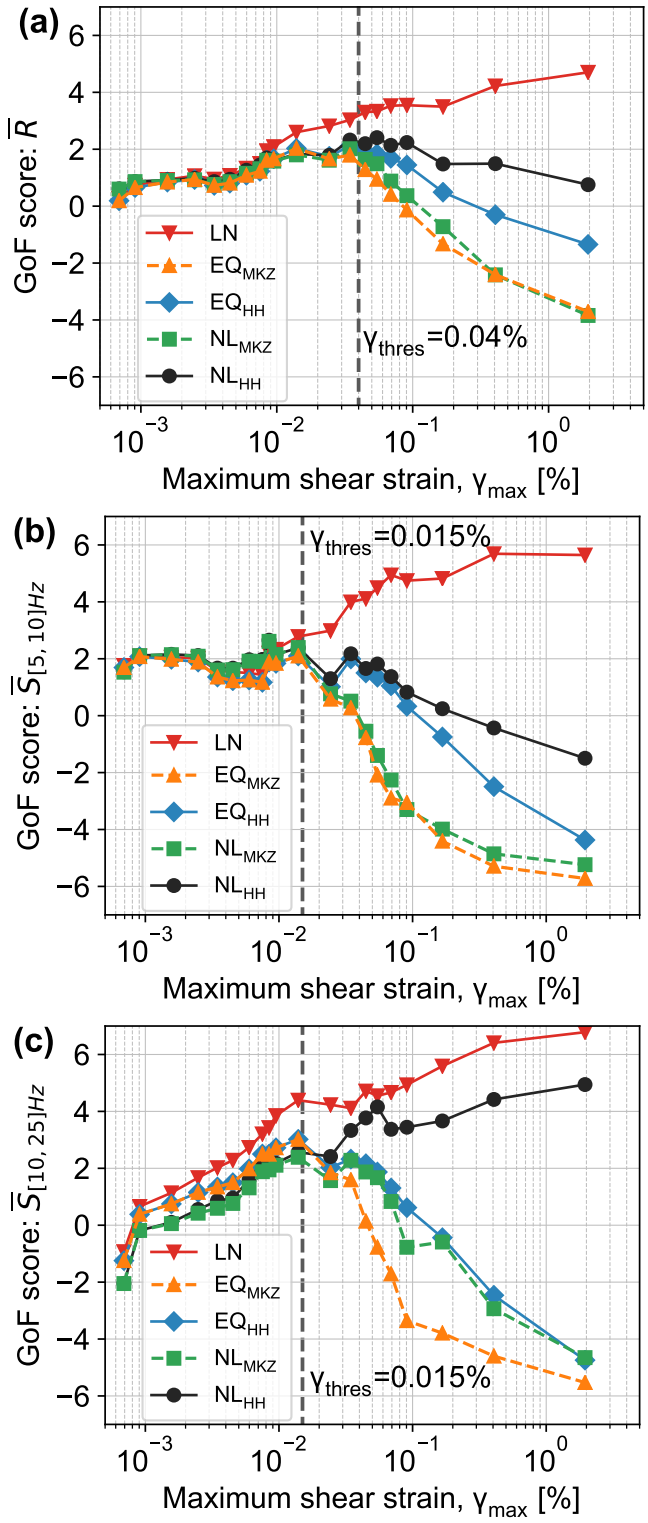


Figure 3.14: Goodness-of-fit score, $\overline{R}$, $\overline{S}_{[5,10]\text{Hz}}$, and $\overline{S}_{[10,25]\text{Hz}}$ versus maximum shear strain ($\gamma_{\max}$). Positive score indicates over-prediction, negative score indicates under-prediction, and 0 indicates perfect fit. NL_{HH} has the best overall performance among the five methods (slight over-prediction is preferred in engineering design). The threshold strain (beyond which all other methods are less reliable than NL_{HH}) is 0.04% for $\overline{R}$, and even lower (0.015%) for higher frequencies: $\overline{S}_{[5,10]\text{Hz}}$ and $\overline{S}_{[10,25]\text{Hz}}$.

(averaged across all frequency bands), it severely underestimates higher frequencies when $\gamma_{max} > \gamma_{thres}$, while NL_{HH} still has the best relative performance among five methods. One should note here that γ_{thres} for $f \geqslant 5$ Hz is even lower, on the order of 0.015%. These results reinforce as well as quantify the well known limitation of the equivalent linear method (regardless of whether the stress-strain curves used in the iterations are corrected for shear strength or not): the use of the equivalent linear method, especially for frequencies above 5 Hz, might lead to severe underestimation of the ground motion intensity even for events too weak to be associated with nonlinear site response in practice.

3.4.3 Case analysis of two strong events

We next present a detailed analysis of two strong events, to further shed light on the difference between the NL and EQ methods, and between HH and MKZ constitutive soil models. We specifically discuss the 2011/3/11 M_w 9.0 Tohoku-Oki Earthquake mainshock recorded at FKSH11 and the 2003/9/26 M_w 8.3 Hokkaido Earthquake mainshock recorded at KSRH10. The 2003 event corresponds to the largest calculated γ_{max} of all stations, 3.67%; the 2011 event corresponds to the second largest γ_{max} (1.13%) of all stations, and the largest γ_{max} at FKSH11. Figure 3.15 shows the waveforms and Fourier spectra (smoothed using a Konno-Ohmachi smoothing window by *Konno and Ohmachi*, 1998) of the ground surface recording and simulation results. For both events, and in line with our previous analyses, LN overestimates ground motion intensities. For the 2003 event, EQ_{MKZ} and NL_{MKZ} severely underestimate ground motions, while EQ_{HH} and NL_{HH} predict adequate intensity. Still, EQ_{HH} underestimates frequencies higher than 5 Hz (as shown in the Fourier spectra plot). For the 2011 event, all methods except LN underpredict the ground motion. The Fourier spectra plot, however, shows that EQ_{MKZ} and EQ_{HH} severely "overdamp" higher frequencies ($\geqslant$4 Hz), while NL_{HH} provides relatively satisfactory predictions over the full frequency range. (It is worth noting that a good PGA prediction does not necessarily indicate good match in all frequencies.)

The reason that NL_{HH} does not underestimate surface ground motions as much as NL_{MKZ} is that HH incorporates shear strength, which yields a "stiffer" response (more elastic) in the shallow layers and a "softer" response in the deeper layers than MKZ. Figure 3.16 shows the stress-strain loops calculated by NL_{MKZ} and NL_{HH} at a shallow and a deep layer, for both 2003 and 2011 events. From Figure 3.16(c) and (f), we can see that at larger depths, NL_{MKZ} and NL_{HH} follow almost the same stress path, with very similar peak stress and peak strain (much lower than γ_t). However, at shallow depths (Figure 3.16(a) and (d)),

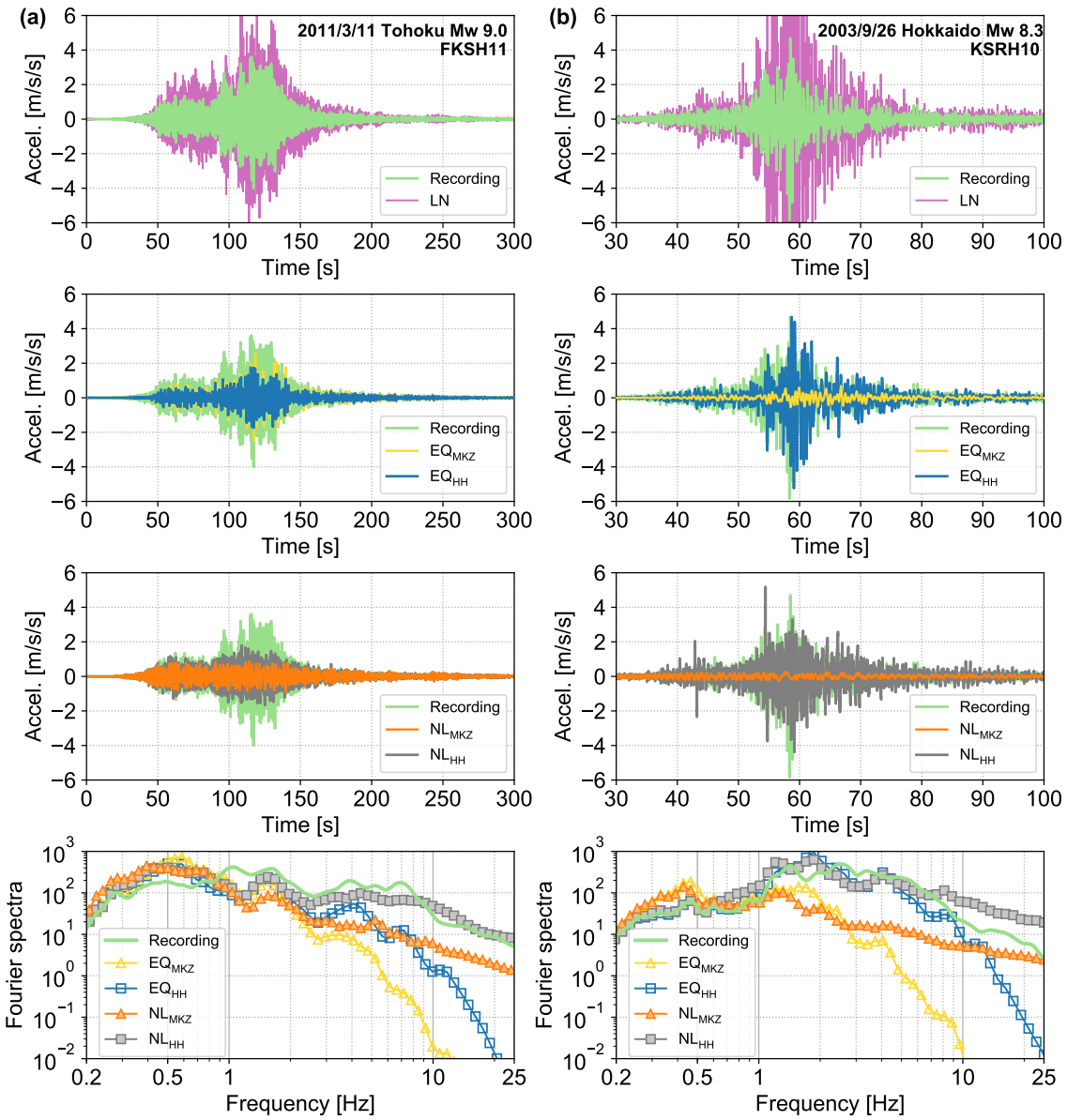


Figure 3.15: Time history and Fourier spectra (smoothed) for the (a) 2011/3/11 M_{w} 9.0 Tohoku Earthquake recorded at FKSH11 and (b) 2003/9/26 M_{w} 8.3 Hokkaido Earthquake recorded at KSRH10. Recording and simulations are plotted together for comparison. $\mathrm{NL_{HH}}$ offers the best overall prediction than the other methods, which is especially evident in the Fourier spectra plots.

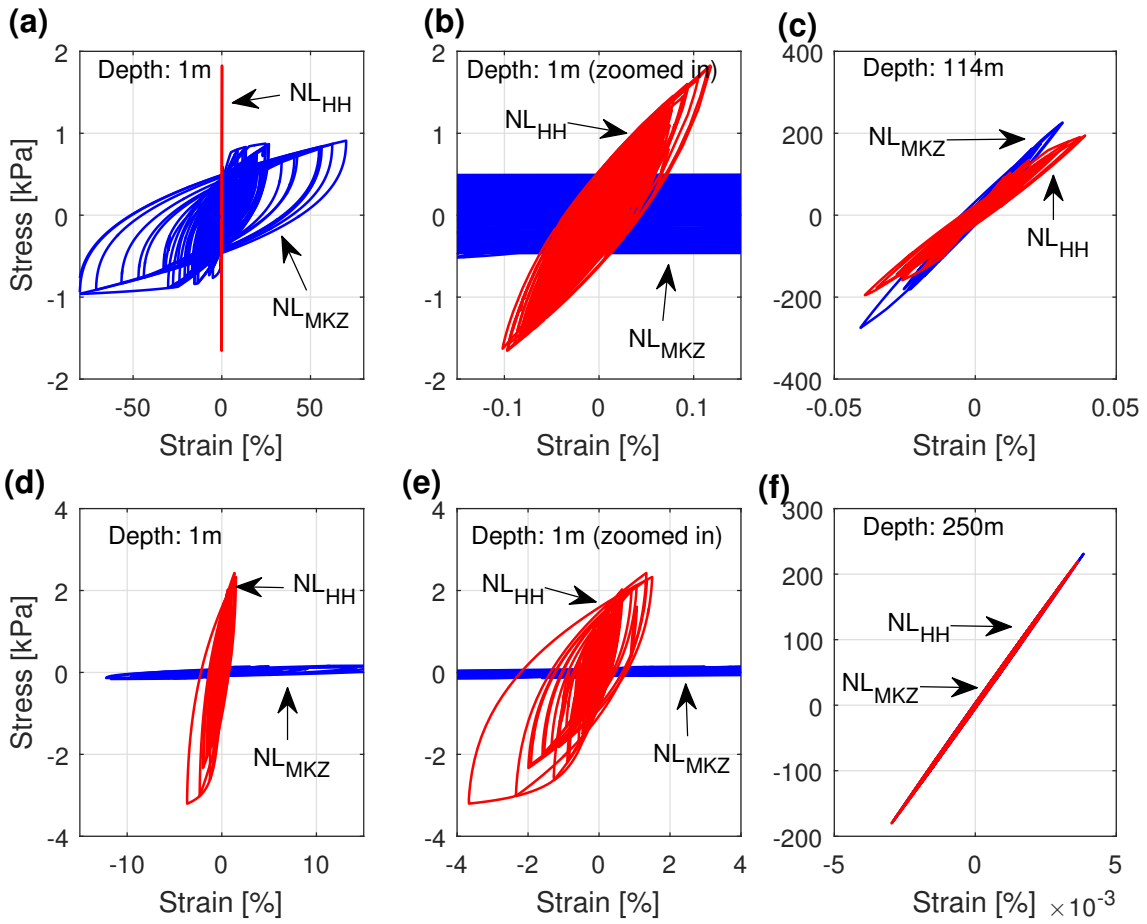


Figure 3.16: Stress-strain loops by NL_{MKZ} and NL_{HH} at a shallow and a deep layer at FKSH11 and KSRH10. (a)–(c): 2011/3/11 Tohoku Earthquake recorded at FKSH11; (d)–(f): 2003/9/26 Hokkaido Earthquake recorded at KSRH10. (a) and (d) are shallow layers of the two stations, and (b) and (e) are the horizontally zoomed-in versions of (a) and (d). (c) and (f) are deep layers of the two stations. At deep layers, stress-strain loops by NL_{MKZ} and NL_{HH} are almost identical, but at shallow layers, NL_{MKZ} underestimates stress and overestimates strain.

NL_{MKZ} erroneously produces excessively large strain level (over 50% in (a)) and severely underpredicts stress (nearly no stress in (d)), while NL_{HH} produces stiffer stress-strain loops with higher peak stress and lower peak strain levels. Thus it is mainly the shallower layers that contribute to the difference between NL_{MKZ} and NL_{HH}.

3.5 Conclusions

We presented a new stress-strain soil model for 1D total stress site response analysis, which we coined the Hybrid Hyperbolic (HH) model. The stress-strain curve of HH compared very well to data from resonance column tests in the small-strain range and direct simple shear tests in the medium-to-large strain range, indicative of the satisfactory geometric

flexibility of the functional form of the model. We also showed that with as little input as the shear wave velocity (V_S) profile of a site, HH can capture the shear strength of soils, and also realistically describe the accumulation of stress as a function of strain (before shear strength is reached), based on empirical correlations between V_S and HH's parameters.

Subsequently, we presented an extensive validation and comparative study of HH against a widely employed nonlinear model, MKZ, both as part of equivalent linear and nonlinear analyses. For this purpose, we used 2,756 downhole array recordings at nine KiK-net strong motion stations, where we used downhole recordings as input and surface recordings as benchmark. Across all stations and events, our findings are briefly summarized below:

1. The linear method (LN) overestimated the ground motion amplitude in the time and frequency domains increasingly with ground motion intensity.

2. The equivalent linear method with the MKZ model (EQ_{MKZ}) gave satisfactory predictions below $\gamma_{\text{thres}} = 0.04\%$ (or $\text{PGA}_{\text{surface}} \simeq 0.05g$), while yielding increasingly underestimated ground motions with ground motion intensity.

3. The equivalent linear method with the HH model (EQ_{HH}) performed better than EQ_{MKZ} due to the use of HH model; it underpredicted stronger ground motions (which induced larger strains) to some degree, and significantly underestimated high frequencies (above 5 Hz).

4. The nonlinear method with the MKZ model (NL_{MKZ}) had similar performance and similar trend as EQ_{MKZ}, only slightly better for $\gamma > \gamma_{\text{thres}}$.

5. The nonlinear method with the HH model (NL_{HH}) had the most satisfactory performance across all ground motion intensities (PGA up to $0.9g$, strain up to 3.67%) and for both broadband and high frequencies (unlike EQ_{HH}); perhaps most importantly, its goodness-of-fit appeared insensitive to the increase of ground motion intensity.

Through this chapter, we attempt to further advance our understanding of the predictive capabilities (as well as the limitations) of 1D site-specific response models: a topic that has become the focus of a series of recently published papers. For example, two recent studies, *Kaklamanos et al.* (2015) and *Zalachoris and Rathje* (2015), employed validation methodologies similar to this study, although both were based on significantly smaller datasets of events. Their goodness-of-fit metrics were also different from this chapter: they calculated the residue of response spectra or amplification function between simulations and recordings, whereas we included additional intensity measures such as RMS acceleration,

RMS velocity, Arias Intensity, etc., each correlating with different aspects of the destructive potential of seismic shaking. *Zalachoris and Rathje* (2015) concluded that EQ and NL both underpredicted ground motions for strains larger than 0.4% even when using stress-strain curves corrected for shear strength, which was not consistent with our results. The source of such discrepancy could be from (1) their use of the strength-incorporation scheme by *Hashash et al.* (2010), which, as pointed out earlier, employed "engineering judgment" rather than a rigorous calibration process, or from (2) their choice of validation sites: among the 11 sites used in that study, only four have been shown by *Thompson et al.* (2012) to conform to 1D wave propagation conditions. *Kaklamanos et al.* (2015), on the other hand, used KiK-net stations that satisfy the 1D wave propagation conditions, and concluded that the nonlinear method (equivalent to NL_{MKZ} in the present study) offered a slight improvement over equivalent linear method (equivalent to EQ_{MKZ} in the present study), which is in general consistent with our findings (see Figure 3.14(a)). They did not, however, use strength-incorporated models in their work.

Comparing the median GoF score of all methods, we determined that 0.04% is the threshold strain beyond which the performance of all the other four methods deteriorate except NL_{HH}. This value is lower than that reported in *Kaklamanos et al.* (2013) (0.1%) and only slightly lower than in *Kaklamanos et al.* (2015) (0.05%). The validation results in Section 3.4 suggests that the use of a strength-incorporated stress-strain soil model in a time-domain nonlinear scheme is not only important for site responses analyses of rare (and extreme) events, but also for medium-intensity events (with PGA as low as $0.05g$).

Notwithstanding the rigor of our analyses, the statistical significance of the dataset that we used could be challenged by the scarcity of very strong ground motion recordings. However, the factors that cause imperfect goodness-of-fit for weaker ground motions, such as errors in initial velocity and attenuation (or damping) profiles or the angle of wave incidence would become insignificant for stronger events. Thus we believe that the goodness-of-fit scores at higher strains—as few as they may be—are still statistically representative. Since global databases of strong motion downhole array recordings are limited, centrifuge model tests could complement the available strong motion field recordings with carefully calibrated nonlinear site response tests to be used as benchmarks for similar studies in the future.

Lastly, we would also like to note that the KiK-net ground motion dataset represents a group of sites with profiles generally stiffer than the conditions prevailing in highly seismically active regions in the US such as California and Central/Eastern US. Consequently, the maximum strain level presented here is most likely a lower bound of the strain that sediments in the Los Angeles Basin or the Mississippi Embayment would experience for the same

shaking. And since our analyses showed that the threshold strain beyond which nonlinear analyses become increasingly important is only 0.04%, we believe that the use of nonlinear analyses with strength-incorporated models such as HH should be considered instead, or at least in addition to equivalent linear analyses, even for ground motions that have been traditionally considered too weak to cause substantial soil yielding.

3.6 Appendix: the HH calibration procedure

This Appendix provides a guideline for empirically constructing modulus reduction and damping curves using only the V_S profile as input. We should remind the reader here that all rules of thumb for the selection of parameters should be reserved for the cases where there is no information on the site conditions, and that measured material properties, if available, are always preferable.

3.6.1 Constructing stress-strain curves

Stress-strain curves can be constructed in the following steps. This is referred to as the *HH calibration (HHC) procedure*:

1. Evaluation of p'_0, the vertical confining pressure, at a specific depth:

$$p'_0 = \sum_j \rho_j g h_j - u \tag{A1}$$

where g is gravitational acceleration, ρ_j and h_j are the mass density and thickness of each soil layer, respectively. The summation should be carried out from the ground surface to the layer of interest. And u is the water pressure at that specific depth. If no information about the water table is available, then the decision lies upon the engineer or scientist to assume dry or saturated soil conditions. (In the case of the present study, we assume the soils are dry.) The mass density can be evaluated, following *Mayne et al.* (1999), as

$$\rho = 1 + \frac{1}{0.614 + \frac{58.7}{V_S}(\ln z + 1.095)} \tag{A2}$$

where ρ is in g/cm^3, V_S is in m/s, and z is the depth of the soil layer in meters.

2. Evaluation of vertical pre-consolidation stress of soil, p'_{pre}, following *Mayne et al.* (1998),

$$p'_{\text{pre}} = 0.106 \cdot V_S^{1.47} \tag{A3}$$

where p'_{pre} is in kPa, and V_S is in m/s.

3. Evaluation of OCR (over-consolidation stress):

$$\mathrm{OCR} = p'_{\mathrm{pre}} / p'_0 \tag{A4}$$

4. Construct τ_{MKZ} using Equation (3.4), with $\beta = 1$, $s = 0.9190$, and γ_{ref} evaluated as follows (recommended by *Darendeli*, 2001),

$$\gamma_{\mathrm{ref}} = \left[\phi_1 + \phi_2 \left(\mathrm{PI}\right) \left(\mathrm{OCR}\right)^{\phi_3}\right] \cdot \left(p'_{\mathrm{m0}}\right)^{\phi_4} \tag{A5}$$

where p'_{m0} is the mean effective confining pressure (unit: atm), to be defined in the next step; $\phi_1 = 0.0352$, $\phi_2 = 0.0010$, $\phi_3 = 0.3246$, and $\phi_4 = 0.3483$ are calibrated by *Darendeli* (2001); and PI is the plasticity index. With only V_S information available, we use a rule of thumb to determine PI, which is similar in principle to that of *Zalachoris* (2014).

$$\mathrm{PI} = \begin{cases} 10, & \text{if } V_S \leqslant 200\,\mathrm{m/s} \\ 5, & \text{if } 200 < V_S \leqslant 360\,\mathrm{m/s} \\ 0, & \text{if } V_S > 360\,\mathrm{m/s} \end{cases} \tag{A6}$$

5. The mean effective confining pressure, p'_{m0} is the average of three stress components (two horizontal, p'_{h0}, and one vertical, p'_0),

$$p'_{\mathrm{m0}} = \frac{2p'_{\mathrm{h0}} + p'_0}{3} = \frac{2K_0 + 1}{3} p'_0 \tag{A7}$$

where K_0 is the ratio of horizontal and vertical confining pressure, and is evaluated as (following *Mayne and Kulhawy*, 1982)

$$K_0 = (1 - \sin\phi') \cdot \mathrm{OCR}^{\sin\phi'} \tag{A8}$$

where ϕ' is the effective friction angle of soils, and is taken as 30° (without better available information).

6. Evaluation of dynamic shear strength of soil, τ_{f}. We choose the formula of undrained shear strength (s_{u})—following *Ladd* (1991)—for all soil types with $V_S \leqslant 760\,\mathrm{m/s}$, because earthquakes loading is imposed so quickly that "even coarse-grained soils do not have sufficient time to dissipate excess porewater pressure, and thus undrained condition applies" (*Budhu*, 2011, page 267). And for materials with $V_S > 760\,\mathrm{m/s}$ (rocks or very stiff soils), we use Mohr-Coulomb criterion to determine the shear strength. And then we will apply the rate-effect correction factor, Z.

$$\tau_{\mathrm{f}} = \begin{cases} Z \cdot 0.28 \left(\mathrm{OCR}\right)^{0.8} p'_0, & \text{if } V_S \leqslant 760\,\mathrm{m/s} \\ Z \cdot p'_{\mathrm{n}} \tan\phi', & \text{if } V_S > 760\,\mathrm{m/s} \end{cases} \tag{A9}$$

where p'_n is the normal effective stress on the failure plane, calculated as

$$p'_n = \frac{\sigma'_1 + \sigma'_3}{2} - \frac{\sigma'_1 - \sigma'_3}{2} \sin\phi' \tag{A10}$$

where σ'_1 and σ'_3 are the larger and smaller one, respectively, between p'_0 (vertical confining stress) and p'_{h0} (horizontal confining stress), and ϕ' is 30° as before. And Z, the rate-effect correction factor, has the form (following *Vardanega and Bolton*, 2013),

$$Z = 1 + \frac{1}{20} \log_{10}\left(\frac{\dot{\gamma}}{10^{-6}}\right) \tag{A11}$$

where $\dot{\gamma}$ is the strain rate of the cyclic loading (unit: s^{-1}), and 10^{-6} s^{-1} is the strain rate of the static tests like DSS. A typical earthquake frequency is taken as 1 Hz, then a simplistic shear strain rate is 10^{-2} s^{-1} (*Vardanega and Bolton*, 2013), yielding $Z = 1.20$. We use 1.20 for for all ground motions.

7. Evaluation of G_{max}, the initial shear modulus, from wave propagation theory,

$$G_{max} = \rho V_S^2 \tag{A12}$$

where ρ is in kg/m^3, and V_S is in m/s. Note that this G_{max} from Equation (A12) is the dynamic shear modulus, thus rate-effect correction is not necessary. However, when a G_{max} is calculated as the initial slope of a static stress-strain test, it needs to be multiplied by Z to be used as a HH model parameter.

8. Evaluation of μ: the following formula (identical to Equation (3.11)) is used

$$\mu = \frac{1}{0.000872\,(G_{max}/\tau_f)\,(\mathrm{OCR})^{0.47}\left(p'_0 \text{ in kPa}\right)^{0.28}} \tag{A13}$$

Because the original study by *Vardanega and Bolton* (2011) that provided Equation (3.10) was for clays and silts, we apply Equation (A13) only to soils with $V_S \leqslant 760\,m/s$. And for rocks or very stiff soils ($V_S > 760\,m/s$), we simply use $\mu = 1$ for lack of more information. Inaccurate as this may seem, the shear strain in such stiff layers seldom reaches γ_t, thus the FKZ part does not play a role in the simulation for these layers.

9. Finally, use an optimization algorithm to find d, γ_t, and a, so that τ_{HH} is continuous and monotonically increasing, and the difference (or the "area") between τ_{MKZ} and τ_{FKZ} is minimized (in order for MKZ and FKZ to correspond to similar soil types at smaller strains). The search range for d is set to $1.03 \pm (3 \times 0.12)$ (using the "3σ

rule"), where 1.03 and 0.12 are the mean and standard deviation of the curve fitting in Figure 3.4. And the search range for γ_t is $[0.01\%, 3\%]$, which is slightly wider than $[0.1\%, 1\%]$ to ensure MKZ and FKZ can merge. In practice, γ_t usually takes the strain where $\tau_{MKZ} = \tau_{FKZ}$, a usually takes 100 (quick transition). Admittedly, uncertainties would be introduced in this step, but this is unavoidable since we do not have any actual stress-strain measurements, only a shear wave velocity value.

3.6.2 Constructing damping-strain curves

Damping-strain curves are calculated following *Darendeli* (2001).

$$\xi = b\,(G/G_{max})^{0.1} \cdot \xi_{Masing} + \xi_{min} \ \text{(Unit: \%)} \tag{A14}$$

where $G/G_{max} = \tau_{MKZ}/(\gamma \cdot G_{max})$, and

$$\xi_{Masing} = 1.0222\xi_* - 0.00676\xi_*^2 + \left(6.1519 \times 10^{-5}\right)\xi_*^3 \ \text{(Unit: \%)} \tag{A15}$$

where

$$\xi_* = \frac{100}{\pi}\left[4\frac{\gamma - \gamma_{ref}\ln\left(\frac{\gamma+\gamma_{ref}}{\gamma_{ref}}\right)}{\gamma^2/(\gamma+\gamma_{ref})} - 2\right] \ \text{(Unit: \%)} \tag{A16}$$

and ξ_{min} (or "D_{min}", in some other literature), is calculated as

$$\xi_{min} = \left[\phi_6 + \phi_7 \cdot \text{PI} \cdot (\text{OCR})^{\phi_8}\right]\left(p_0'\right)^{\phi_9} \cdot (1 + \phi_{10}\ln f) \tag{A17}$$

where $\phi_6 = 0.8005$, $\phi_7 = 0.0129$, $\phi_8 = -0.1069$, $\phi_9 = -0.2889$, $\phi_{10} = 0.2919$ are calibrated by Darendeli, and f is the earthquake frequency, taking a nominal value of 1.0 Hz.

Chapter 4

Improved site factors for the Western United States

This chapter presents SAG19, a set of improved site factors[1] calculated from 1D site response simulations. Compared to other site factors, SAG19 uses methods that are an improvement over the state of the art. SAG19 takes V_{S30}, z_1, and PGA as input parameters, and has a reference site condition of 1,046 m/s. SAG19 is the first site factor product to provide phase-shift factors, which describe the time lag between the output and input motions.

Section 4.2 documents the details of the simulation methods of SAG19. Section 4.1.3 compares the difference between the methodologies used in SAG19 and other site factors, and Section 4.4 compares the results of SAG19 with others.

Contents of this chapter

[1] See Nomenclatures (on page xix) for definition of "site factors".

4.1 Introduction

4.1.1 Site response in seismic hazard assessments

Characterizing site response on a regional scale (city- to state-wide) is an integral part to seismic hazard assessments, whose results are usually in the form of the following products: seismic hazard maps—such as the U.S. Geological Survey National Seismic Hazard Maps (NSHM) (*Frankel et al.*, 1996, 2002; *Petersen et al.*, 2008, 2014), open-source seismic hazard software—such as the Global Earthquake Model[2], or proprietary catastrophe models.

These products play an important role in many policy-making processes in both the public and private sectors. Examples of such processes include the U.S. building code of earthquake resistant structural design (*Building Seismic Safety Council*, 2015), the California Earthquake Authority Earthquake Insurance Premium Calculator[3], disaster management and mitigation strategies (such as Hazus[4]), planning and seismic safety applications (such as the California Seismic Hazards Mapping Acts[5]), and many site-specific engineering analyses by industries and governments (such as those applied by the US Department of Defense and the US Bureau of Reclamation) (*Petersen et al.*, 2014).

The specific strategies, with which the above-mentioned seismic hazard assessment products incorporate site amplification, evolve with time.

For example, earlier versions of the NSHM ground motion models incorporated site effects through a scaling parameter that was based on site classification (for example, *Boore et al.*, 1993), or coarser qualitative distinctions such as the binary soil versus rock classification (*Campbell*, 2003). *Boore et al.* (1997) introduced a parameterization that included explicitly the average shear-wave velocity in the upper 30 m, widely known as V_{S30}, while *Abrahamson and Silva* (1997) introduced additional terms to further account for nonlinear site response in ground motion models.

The latest NSHM revision (i.e., the 2014 version) has incorporated ground motion models from the following five NGA-West2 GMPEs[6]: ASK14 (*Abrahamson et al.*, 2014), BSSA14 (*Boore et al.*, 2014), CB14 (*Campbell and Bozorgnia*, 2014), CY14 (*Chiou and Youngs*, 2014), and I14 (*Idriss*, 2014), for the Western United States (WUS); and modified versions

[2] https://www.globalquakemodel.org/

[3] https://www.earthquakeauthority.com/California-Earthquake-Insurance-Policies/Earthquake-Insurance-Premium-Calculator

[4] http://www.fema.gov/hazus/

[5] https://www.conservation.ca.gov/cgs/shma

[6] NGA-West2 GMPEs: see Nomenclatures (on page xix) for definition.

of some NGA-West2 GMPEs, among others, for the Central and Eastern United States (CEUS). These five GMPEs employ different methodologies in determining the site factors:

- ASK14 partially incorporates the numerical simulations performed by *Kamai et al.* (2014) to constrain its site amplification terms.
- CB14 partially incorporates the numerical simulations by *Walling et al.* (2008), which used the same simulation scheme as *Kamai et al.* (2014) but was based on a smaller soil profile database, to constrain its site amplification terms.
- The site amplification in BSSA14 is constrained using the work by *Seyhan and Stewart* (2014), which is a semi-empirical approach combining the simulation of *Kamai et al.* (2014) and empirical data analysis.
- CY14 derived their nonlinear site amplification terms purely based on data.
- I14 only includes linear site amplification terms, and is thus applicable only to stiffer sites ($V_{S30} \geqslant 450$ m/s).

4.1.2 Site response simulations for generating site factors

Because the amount of recordings of strong ground shaking on soft soils is very scarce, using numerically simulated site amplification factors to augment the recorded earthquake data, as performed in *Walling et al.* (2008) and *Kamai et al.* (2014) (henceforth referred to as KAS14), is currently a good approach.

Another recent study by *Hashash et al.* (2018) (henceforth referred to as HIH18) also attempted to develop site amplification factors for WUS.

HIH18 and KAS14 share similarities in some of their procedures, while having differences in others. A comparison between KAS14, HIH18, and SAG19 is summarized in Table 4.1. And detailed explanations of the differences are presented next, in Section 4.1.3.

4.1.3 Differences in the methods/procedures between KAS14, HIH18, and SAG19

This subsection summarizes the differences in the methods and procedures used in the three site factor models (KAS14, HIH18, and SAG19).

Procedure \ Model	KAS14	HIH18	SAG19
Base V_S profiles	From NGA-West2 database	Mostly from NGA-West2 database	Generated using the SVM (from Chapter 2)
Randomized V_S profiles	Uses randomization scheme by *Toro* (1995)*	Uses a modified randomization scheme based on *Toro* (1995)*	Uses an improved randomization scheme based on *Toro* (1995)
Small-strain damping ($D_{\min}$)	Unclear	*Campbell* (2009)	*Archuleta and Liu* (2004)
Reference V_{S30}	1,180 m/s	Unclear*	1,046 m/s
Stress-strain relations of soils	Does not consider soil strength*	The GQ/H model by *Groholski et al.* (2016) to incorporate soil strength	The HH model (from Chapter 3) to incorporate soil strength
Input ground motions	Response spectra generated on point sources*	Recorded motions from NGA-West2 database*	Simulated motions
Site response simulation method	RVT equivalent linear method*	Nonlinear method	Nonlinear method (for amplitude), equivalent linear method (for phase)
Form of amplification factor	Response spectra*	Fourier spectra	Both Fourier and response spectra
Phase factor	Not provided*	Not provided*	Provided

Table 4.1: A brief comparison of the methods/procedures of three site factor models for WUS: KAS14 (*Kamai et al.*, 2014), HIH18 (*Hashash et al.*, 2018), and SAG19 (this chapter). The asterisks (*) indicate that the methods/procedures have some disadvantage that could lead to worse results.

4.1.3.1 Base V_S profiles

The base V_S profiles are 1D profiles from which randomized profiles (more realistic) are generated. Three models all use base V_S profiles directly or indirectly from measured V_S profiles in WUS, hence are quite similar.

4.1.3.2 Randomized V_S profiles

KAS14 uses the randomization scheme proposed by *Toro* (1995), and HIH18 modifies the Toro scheme to eliminate velocity reversal. Both KAS14 and HIH18 could lead to the randomized profiles having very different V_{S30} and z_1 values than the base profiles. (The V_{S30} and z_1 of the base profiles are the target values.) The randomization scheme in SAG19 addresses this issue (see Section 4.2.2).

4.1.3.3 Small-strain damping ($D_{\min}$)

The small-strain damping (a.k.a., $D_{\min}$) used in KAS14 is not documented. HIH18 adopts the Q_S-V_S relationship proposed by *Campbell* (2009): $Q_S = 7.17 + 0.0276V_S$ (unit of V_S: m/s). SAG19 adopts the relationship proposed by *Archuleta and Liu* (2004):

$$Q_S = \begin{cases} 0.06V_S, & V_S \leqslant 1000 \text{ m/s} \\ 0.14V_S, & 1000 < V_S < 2000 \text{ m/s} \\ 0.16V_S, & V_S \geqslant 2000 \text{ m/s} \end{cases} \tag{4.1}$$

where V_S should use the unit of m/s. $D_{\min}$ is calculated as $D_{\min} = 1/(2Q_S)$. The actual $D_{\min}$ values of the two relationships—for typical V_S of soils—are both between 0.5% and 6%, and are quite similar.

4.1.3.4 Reference site condition

The reference site condition (or "reference V_{S30}") of HIH18 is nominally 760 m/s, but evidence in Section 4.4.2.3 shows that HIH18 uses higher values as its actual reference site condition for stiffer sites (V_{S30} > 500 m/s), and the actual values are unclear. This makes it difficult for the potential users to use the HIH18 site factors for stiffer sites.

On the other hand, the reference site conditions of SAG19 is 1,046 m/s, and SAG19 does not have the same issue.

4.1.3.5 Stress-strain relations of soils

KAS14 uses the *Electric Power Research Institute (EPRI)* (1993) model and the "peninsular range" model (defined in *Walling et al.*, 2008) as the stress-strain relations, which do not take into account shear strength of soils. On the other hand, SAG19 uses the HH model (from Chapter 3) that incorporates shear strength of soils and captures the small- and large-strain soil shearing behaviors well, and thus offers much better site response predictions for medium-to-strong motions. HIH18 uses a stress-strain model (named GQ/H, proposed by *Groholski et al.*, 2016) that can capture soil strength, but contrary to the HH model, GQ/H has not demonstrated the ability to either fit laboratory test data (for larger strains) or reconstruct realistic soil behaviors with only V_S information.

4.1.3.6 Input ground motions

KAS14 uses simulated response spectra from a point source (generated by the computer code RASCALS written by *Silva and Lee*, 1987), rather than earthquake time histories, as

input ground motions, which is not very realistic. HIH18 selects 145 recorded time histories from the NGA-West2 database from sites with V_{S30} > 760 m/s. This means that not all the selected motions are recorded strictly on the reference site condition (nominally 760 m/s for HIH18). On the other hand, SAG19 uses simulated ground motions on sites with reference V_{S30} = 1,046 m/s (more in Section 4.2.4), which can mitigate the potential issues in KAS14 and HIH18.

4.1.3.7 Site response simulations

KAS14 uses the random-vibration-theory-based equivalent linear method to simulate site response. The equivalent linear method (originally proposed by *Seed and Idriss*, 1970) is an approximated solution to the nonlinear wave equation. Several recent studies, such as *Kaklamanos et al.* (2013, 2015) and *Shi and Asimaki* (2017) (i.e., Chapter 3), have shown that the equivalent linear method under-predicts site amplification of stronger input motions (PGA higher than about 0.05g), especially at higher frequencies (higher than 5 Hz).

4.1.3.8 The form of the amplification factors

KAS14 provide the simulation results in the form of amplification between output and input response spectra, the disadvantage of which will be shown in Section 4.3.3.1. Both HIH18 and SAG19 provides the amplification factors as Fourier-based ratios.

4.1.3.9 The inclusion of phase factors

Only SAG19 provides phase factors, which reflect the time lag between the output and input ground motions.

To date, all regional-scale site-factor-based approaches have omitted phase factors. This may be acceptable for single-site analyses where the timing of the ground motion is of secondary importance. But for spatially distributed infrastructure, such as bridges, railways, and pipelines, different wave arrival times at different parts of the structure can cause complex internal seismic responses (such as rocking). Therefore the omission of arrival times could result in an underestimation of actual structural damages.

4.2 Site response simulations for developing site factors

This section documents the detailed procedures to run site response simulations for developing the SAG19 site factors.

4.2.1 One-dimensional base V_S profiles

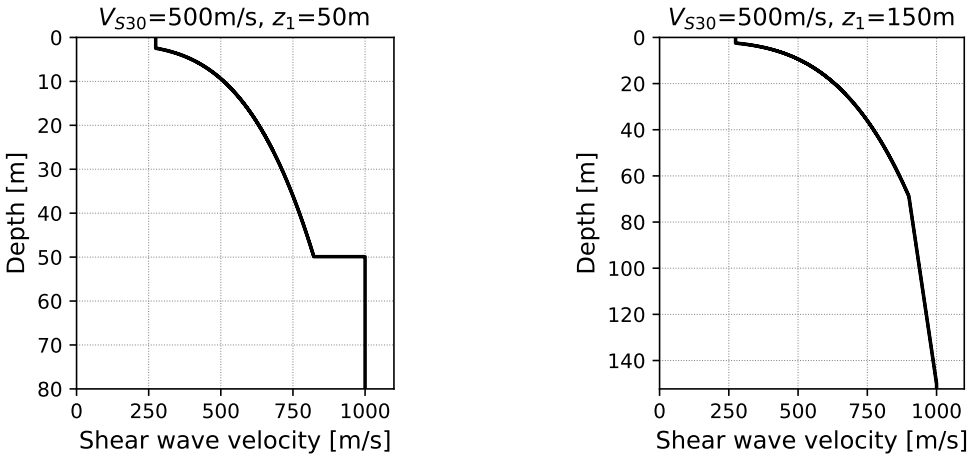
Since V_{S30} is the only available parameter to represent the stiffness of a soil site, we need to "reconstruct" realistic 1D soil profiles from a single V_{S30} value. Here we use the Sediment Velocity Model (SVM) proposed in *Shi and Asimaki* (2018) (Chapter 2) for this task. We pick the following target V_{S30} values: 175, 200, 250, 300, 350, ..., 900, and 950 m/s.

The only available parameter to represent the basin depth is z_1, and we choose the following target z_1 values: 8, 16, 24, 36, 75, 150, 300, 450, 600, and 900 m. Not all combinations of V_{S30} and z_1 are realistic, so we only pick the combinations shown in Table 4.2, which is mostly consistent with KAS14.

V_{S30} [m/s] \ z_1 [m]	8	16	24	36	75	150	300	450	600	900
175	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
200	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
250	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
300	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
350	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
400	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
450	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
500	✓	✓	✓	✓	✓	✓	✓	✓	✓	
550	✓	✓	✓	✓	✓	✓	✓	✓	✓	
600	✓	✓	✓	✓	✓	✓	✓	✓		
650	✓	✓	✓	✓	✓	✓	✓			
700	✓	✓	✓	✓	✓	✓				
750	✓	✓	✓	✓	✓	✓				
800	✓	✓	✓	✓	✓					
850	✓	✓	✓	✓						
900	✓	✓								
950	✓	✓								

Table 4.2: Summary of site response cases considered in this project. The table cells with tick symbols (✓) are the V_{S30}-z_1 combinations that we pick. There are 128 combinations in total.

We denote the analytical (i.e., smooth) soil profiles generated by the SVM as the "base



(a) $V_{S,\mathrm{SVM}} \leqslant 1000$ m/s at z_1 (b) $V_{S,\mathrm{SVM}} > 1000$ m/s at z_1

Figure 4.1: Two examples showing how the SVM handles different combinations of V_{S30} and z_1

profile". If $V_{S,\mathrm{SVM}}$ (the V_S value of the base profile) is lower than 1,000 m/s at the depth of z_1, the base profile below z_1 is truncated and replaced with a homogeneous 1000 m/s "half space", thus creating a velocity contrast at z_1. If, on the other hand, $V_{S,\mathrm{SVM}} > 1000$ m/s at z_1 (which is possible due to errors in the target z_1 value or any potential inaccuracies of the SVM), our SVM code finds $z_{0.9}$ (depth where $V_{S,\mathrm{SVM}} = 900$ m/s), and replaces the velocity values between $z_{0.9}$ and z_1 with a linear interpolation. As a result, all our base V_S profiles terminates at 1,000 m/s and at the depth of z_1. Figure 4.1 demonstrates these two cases.

4.2.2 Randomization of V_S profiles

The use of V_{S30} and z_1 to represent local site conditions is an inevitable compromise in situations where very little information is known about a site, and this has been adopted as the "standard practice" of the seismic hazard analysis field. Therefore, by using only V_{S30} and z_1 information, the base profiles do not reflect the inherent variability of the real-world soil stratification, which could lead to the undesirable consequence that the simulation results are overly representative of smooth soil profiles and not so representative of less smooth ones.

To address this issue, we produce 40 randomized V_S profiles per base profile, using the procedures below.

4.2.2.1 Soil layer randomization

As indicated in Equations (2.7), (2.8), and Figure 2.12 (on page 28), soil layer thickness tends to be thicker for deeper soil layers, and so does the uncertainty bound for thickness. With Equations (2.7) and (2.8), we generate a series of layer thicknesses. *Toro* (1995) proposed a thickness-depth relationship that is in general consistent with ours.

4.2.2.2 V_S uncertainty bounds

At each soil layer (defined by the randomized layer thicknesses above), we generate randomized V_S from a distribution characterized by the base V_S and the uncertainty bound. This process is mostly consistent with the schemes in *Toro* (1995), with only one change: the uncertainty bound for V_S.

Toro (1995) proposed the following uncertainty bound for V_S, denoted as $\sigma_{\ln V_S}$:

$$\sigma_{\ln V_S} = \begin{cases} 0.37, & V_{S30} < 180 \text{ m/s} \\ 0.31, & 180 \leqslant V_{S30} < 360 \text{ m/s} \\ 0.27, & 360 \leqslant V_{S30} < 760 \text{ m/s} \\ 0.36, & 760 \leqslant V_{S30} < 1500 \text{ m/s} \end{cases} \tag{4.2}$$

Note that their proposed relationship is a function of V_{S30} (not V_S), and thus is a constant for all soil layers of specific soil profile.

Instead of Equation (4.2) above, we use Equation (2.10) (on page 29) to describe the V_S uncertainty bound. Equation (2.10) is repeated here for convenience:

$$\sigma_{\ln V_S} = \left(-7.769 \times 10^{-10}\right) V_S^3 + \left(1.597 \times 10^{-6}\right) V_S^2 - \left(8.7240 \times 10^{-4}\right) V_S + 0.4233$$

The curve-fitting figure for this equation is shown in Figure 2.13 (on page 29).

The range of $\sigma_{\ln V_S}$ in Equation (2.10) is consistent with Toro's empirical values in Equation (4.2), but Figure 2.13 reveals that $\sigma_{\ln V_S}$ does not vary monotonically with V_S, which is not reflected in Equation (4.2). Therefore, using Equation (2.10) allows us to better model the actual variability of V_S.

KAS14 uses Equation (4.2) without modifications. On the other hand, HIH18 uses a constant $\sigma_{\ln V_S}$ of 0.2 for all the soil layers regardless of their V_S and their V_{S30}. Furthermore, HIH18 makes the following modifications based on *Toro* (1995):

1. HIH18 disables soil layer randomization.

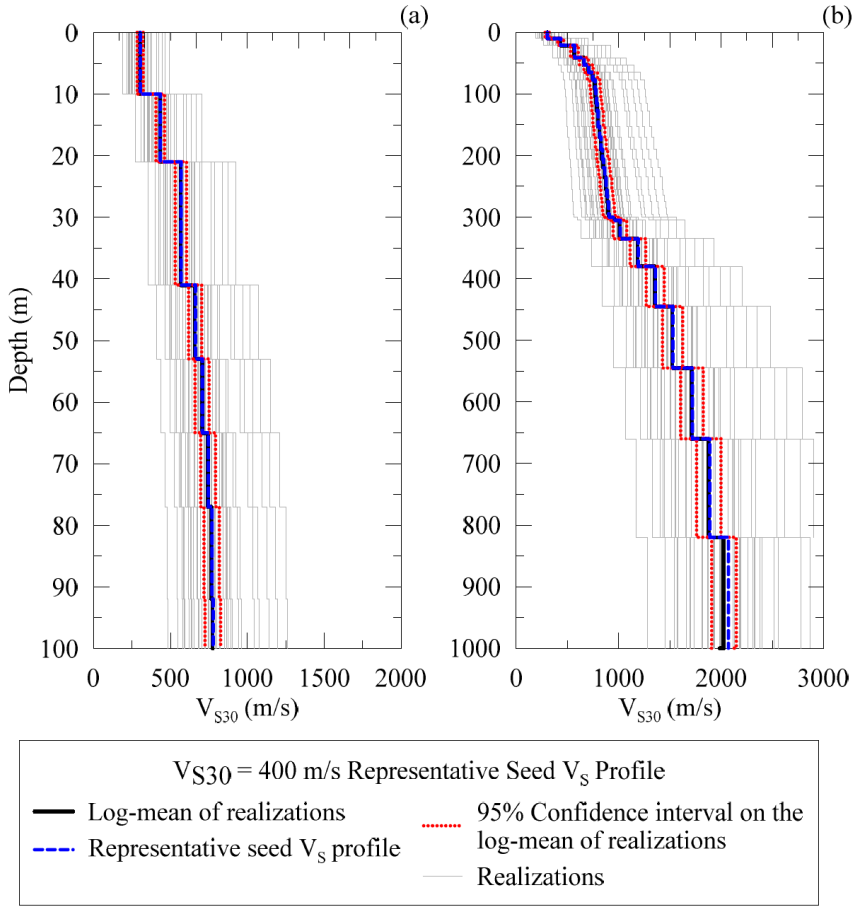


Figure 4.2: Randomized V_S profiles produced using the method in HIH18. Taken from the HIH18 report (*Hashash et al.*, 2018). Subplot (b) shows the whole profile from 0 m to 1000 m, and subplot (a) is the zoomed-in version of (b).

2. HIH18 assigns a single random seed value for all the soil layers within a randomized profile, in order to prevent velocity reversals[7].

The randomized V_S profiles used in HIH18 are demonstrated in Figure 4.2. We can see that a randomized profile is either stiffer or softer than the base profile—for all the layers, and all the randomized profiles are as "smooth" as the base profile. These randomized profiles are likely not fully representative of the real-world variations, thus introducing biases in the results.

[7]Velocity reversal: the phenomenon that the deeper soil layers have lower V_S than the shallower layers.

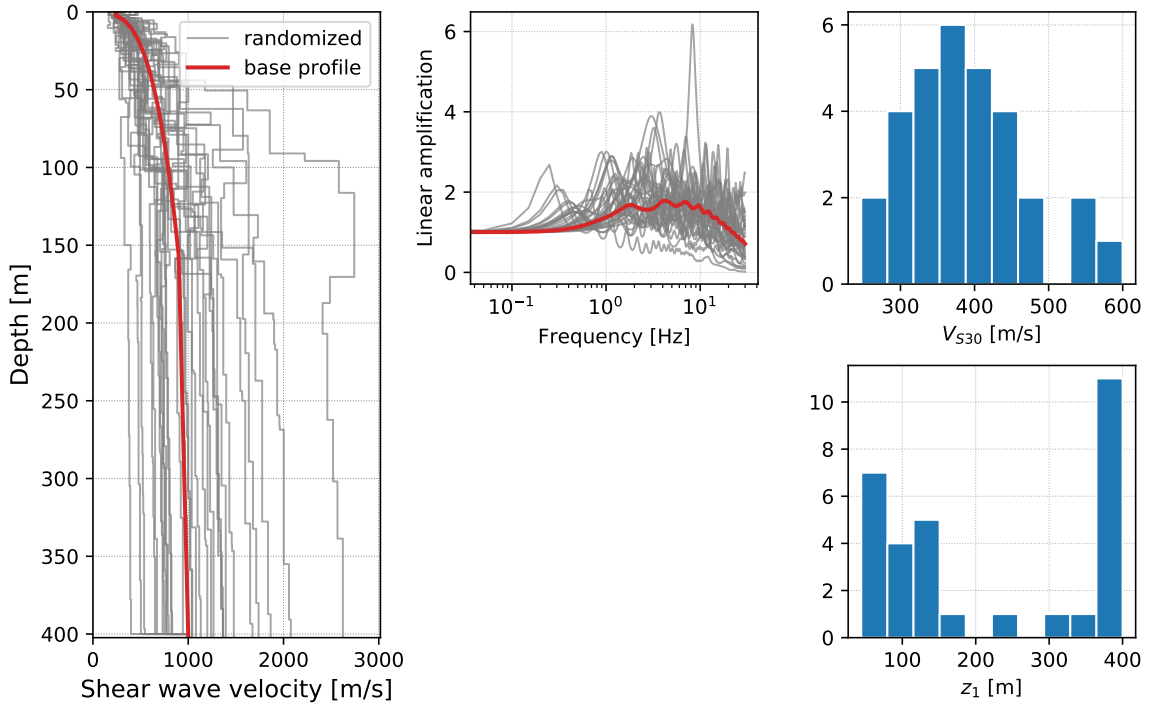


Figure 4.3: A demonstration of the undesirable situation of the randomized profiles having very different V_{S30} and z_1 from the "target V_{S30}" (400 m/s) and "target z_1" (400 m).

4.2.2.3 Target V_{S30} and target z_1

From Figure 4.2, we can also observe that the randomized V_S profiles may have very different V_{S30} and z_1 values than the base profile, which is an issue that KAS14 also shares.

Figure 4.3 illustrates this point by showing a base profile of V_{S30} = 400 m/s and z_1 = 400 m, 30 randomized profiles from this base profile, their respective linear amplifications, and the distribution of their V_{S30} and z_1. From the figure, we see that the randomized V_{S30} values range from ~250 m/s to ~600 m/s, and there are about 1/3 randomized profiles having $z_1 < 200$ m.

This indicates that the amplification factors calculated from some of the randomized profiles may actually correspond to a much softer/stiffer site condition with a much shallower bedrock depth, and such amplification factors will be incorrectly attributed to V_{S30} = 400 m/s and z_1 = 400 m. Additionally, the randomized V_S of the last layer is oftentimes much higher than 1,000 m/s, which no longer corresponds to the target reference V_{S30} (1,046 m/s).

To address this issue, we apply the following criteria, which a randomized profile needs to satisfy to be included for simulations:

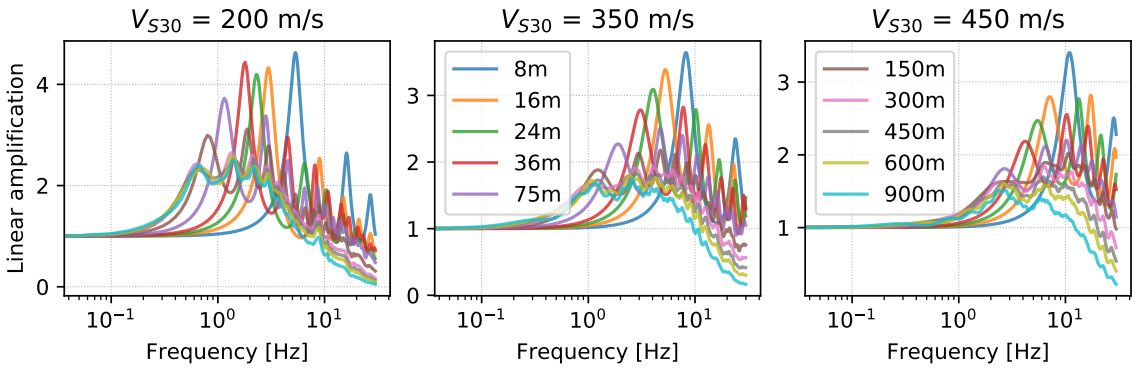


Figure 4.4: Demonstration of the diminishing effect of z_1 to site amplification. The three subplots shows the linear site amplification from SVM-generated profiles of the same V_{S30} but different z_1. For the same V_{S30}: when z_1 is low, increasing z_1 affects linear site amplification a lot, and as z_1 gets higher, its effect diminishes.

1. The absolute difference between the randomized and target V_{S30} is < 25 m/s
2. The relative difference (between the randomized profile and the base profile) of the last soil layer's V_S is $< 5\%$
3. The relative difference of the randomized and target z_1 is $< 20\%$

The rationale behind choosing these three bounds are as follows:

1. The V_{S30} spacing for our simulations are 50 m/s, so the randomized V_{S30} acceptance range is ± 25 m/s
2. The site amplification can be quite sensitive to the impedance contrast between the last soil layer and the bedrock, so the acceptance range is chosen to be only 5%
3. The effect of z_1 on the site amplification diminishes with depth (i.e., more prominent for lower z_1 and less for higher z_1), as shown in Figure 4.4, so the acceptance range is more lenient, at 20%.

Figure 4.5 demonstrates the randomized profiles that satisfy these criteria for the same example base profile (V_{S30} = 400 m/s, z_1 = 400 m). We can see that the randomized V_{S30} and z_1 are within a close range from the target values, and the randomized linear amplifications are closer to the base amplification too.

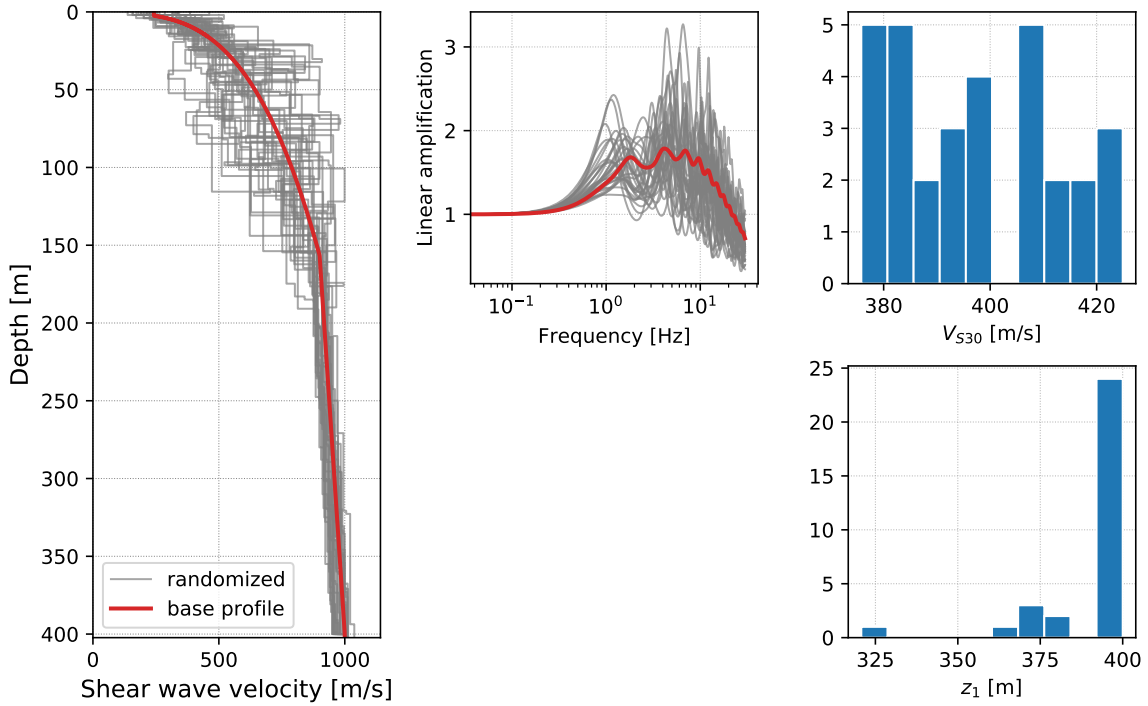


Figure 4.5: A demonstration of randomized profiles that satisfy the criteria (i.e., having V_{S30}, z_1, and V_S of last soil layer close to the target values).

4.2.3 Stress-strain relations of soils

We use the hybrid hyperbolic (HH) model proposed in *Shi and Asimaki* (2017) (Chapter 3) as the stress-strain relation. To generate modulus reduction and damping curves for each layer of the randomized V_S profile, we follow the "HHC procedure" (in Section 3.6) with the following minor adjustments:

- If $\tau_{\mathrm{MKZ}}|_{\gamma=\gamma_t} \geqslant 0.85\tau_f$ (meaning that the MKZ stress curve and the estimated τ_f are not quite consistent), or the last step in Section 3.6.1 does not find a satisfactory d value, relax the lower bound for searching γ_t, such as to 0.005% or 0.001%.
- If μ is too low ($\leqslant 0.04$), adjust μ according to the uncertainty bound of $\log_{10}(\gamma_{M=2})$: 0.236 (reported in *Vardanega and Bolton*, 2011) and the "3σ rule". (Note: μ and $\gamma_{M=2}$ are inversely proportional to each other, as seen in Equations 3.10 and 3.11.)

These adjustments are for rare edge cases where some soil layers have large V_S and small τ_f (shear strength), and they are meant to prevent the stress curves from decreasing with strain. These edge cases happen most likely due to the uncertainties in empirically estimating τ_f, or the uncertainties in Darendeli's formulas for the MKZ parameters. Among all the soil

layers of all the randomized V_S profiles that we generate for this chapter, only 6% of them require the adjustments above.

4.2.4 Input ground motions

We need a wide variety of ground motions (either recorded or simulated) as input to run the site response simulations, and the qualified input ground motions for this project need to satisfy the following criteria:

1. They need to be in the time domain. Response spectra cannot be used as input.
2. They need to be recorded (or simulated to occur) on rock sites whose V_S is very close to 1,000 m/s.
3. Their PGA need to fall within ±10% of these levels: $0.01g$, $0.05g$, $0.1g$, $0.2g$, $0.3g$, $0.4g$, $0.5g$, $0.75g$, $1.0g$, $1.25g$, and $1.5g$. (These PGA levels are therefore the input PGA, or rock-outcrop PGA.)
4. They need to be actual ground motions recorded in WUS, or simulated motions for WUS.

The NGA-West2 database (*Ancheta et al.*, 2014) has many recorded ground motions from around the world. However, we find less than 10 recordings that satisfy all four criteria above. (And it is worth noting that the WUS records in that database are almost exclusively in California.) Therefore, we need to use simulated ground motions as our input. We use the Broadband Platform (BBP) software, version 17.3.0, developed by the Southern California Earthquake Center (*Maechling et al.*, 2014) to generate the motions. The BBP takes a kinematic earthquake source description (magnitude and rupture surface geometry) as input, simulates the rupture process, and propagates the rupture through a simplified 1D crustal velocity model to get the shaking time histories on the ground surface.

The basic parameters that we use in the BBP are summarized below:

- We use a historical event, the 1994 Northridge Earthquake (M_w 6.7), and several scenario events (strike-slip and reverse, with magnitudes ranging from M_w 5.5 to M_w 7.2).
- The maximum stress drop for the strongest event (M_w 7.2) is 100 bars (10 MPa).
- The closest site-to-epicentral distance is 0.5 km, and the farthest distance is 160 km.

- We use the "LA basin" 1D crustal velocity profile in the BBP (the other two velocity profiles for California are very similar to the LA basin profile, and the BBP does not have a crustal velocity profile for the Pacific Northwest).

We perform the ground motion simulations with the parameters above, without applying BBP's built-in site effects adjustments, and from thousands of output ground motions, we pick 440 motions that satisfy the third criterion above (the PGA requirements), with each PGA level having 40 ground motions.

Subsequently, because the crustal velocity profile of the LA basin used in the BBP has four topmost layers (totaling 20 m thick) with V_S below 1,000 m/s, we need to adjust the simulated motions to remove the site amplification introduced by those four layers, so that the resultant ground motions are truly motions on rock outcrop. To this end, for each ground motion, we perform a 1D linear deconvolution, which back-propagates the time series downwards through those four layers. Essentially, this deconvolution step is equivalent to running the BBP propagation through an adjusted 1D crustal V_S profile whose topmost layer is 1,000 m/s. Since this adjusted 1D crustal V_S profile has a V_{S30} of 1,046 m/s, the reference site condition of SAG19 is a V_{S30} of 1,046 m/s, which is quite close to three of the NGA-West2 GMPEs: 1,180 m/s of ASK14 (*Abrahamson et al.*, 2014), 1,100 m/s of CB14 (*Campbell and Bozorgnia*, 2014), and 1,130 m/s of CY14 (*Chiou and Youngs*, 2014).

4.2.5 Numerical site response solver

We have 128 base V_S profiles corresponding to 128 V_{S30}-z_1 combinations (Table 4.2), and each base profile has 40 randomized profiles, hence $128 \times 40 = 5,120$ profiles in total. For each randomized profile, we propagate 440 input ground motions, hence $5,120 \times 440 = 2,252,800$ simulations in total.

We use two methods for each of the 2,252,800 simulations: the nonlinear method and the equivalent linear method, which correspond to "NL_{HH}" and "EQ_{HH}" in Chapter 3 respectively.

4.3 Simulation results

This section presents the simulation results, and highlights how different aspects of the simulations affect the results.

4.3.1 Formulas

Denote the input ground motion in the time domain as $a_{\text{in}}(t)$, and the output ground motion as $a_{\text{out}}(t)$.

Take the Fourier transform (FFT) of the motions, and their frequency-wise ratio is the Fourier spectra ratio, denoted as $\text{FSR}(f)$:

$$\text{FSR}(f) = \frac{\text{FFT}\left(a_{\text{out}}(t)\right)}{\text{FFT}\left(a_{\text{in}}(t)\right)} \tag{4.3}$$

where "(f)" means that FSR is in the frequency domain. FSR is an array of complex numbers, and its amplitude and phase are the amplification factor (AF) and the phase factor (PhF) respectively:

$$\text{AF}(f) = \left|\text{FSR}(f)\right|, \ \text{PhF}(f) = \arg\left(\text{FSR}(f)\right) \tag{4.4}$$

Amplification factors describe how much each frequency of the ground motion is amplified by the soil layers, and phase factors describe how much each frequency is delayed by the soil layers.

Calculate the response spectra (RS) of the input and output motions, using the method introduced in Section 5.2 of *Chopra* (2001), and their frequency-wise ratio[8] is the response spectral ratio, denoted as $\text{RSR}(f)$:

$$\text{RSR}(f) = \frac{\text{RS}\left(a_{\text{out}}(t)\right)}{\text{RS}\left(a_{\text{in}}(t)\right)} \tag{4.5}$$

$\text{RSR}(f)$ is an array of real values, and it describes the amplification of different frequencies, not the delay.

Section 4.3.3 presents the amplification factor results, and Section 4.3.4 presents the phase factor results.

4.3.2 Aggregating all the simulation results

We aggregate the 2,252,800 simulation results to calculate the site factors (AF, PhF, and RSR) for the 1,408 unique V_{S30}-z_1-PGA combinations[9]. At each V_{S30}-z_1-PGA combination, there are $40 \times 40 = 1,600$ unique output-input ground motion pairs[10]. Then we calculate

[8] Response spectra are more often presented as functions of periods (T), not frequencies (f). But since $T = 1/f$, we present response spectra as functions of f in this chapter.

[9] There are 128 unique V_{S30}-z_1 combinations, each having 11 input PGA levels, hence there are $128 \times 11 = 1,408$ total V_{S30}-z_1-PGA combinations.

[10] Because there are 40 randomized V_S profiles per unique V_{S30}-z_1 combination, and 40 input ground motions per input PGA level.

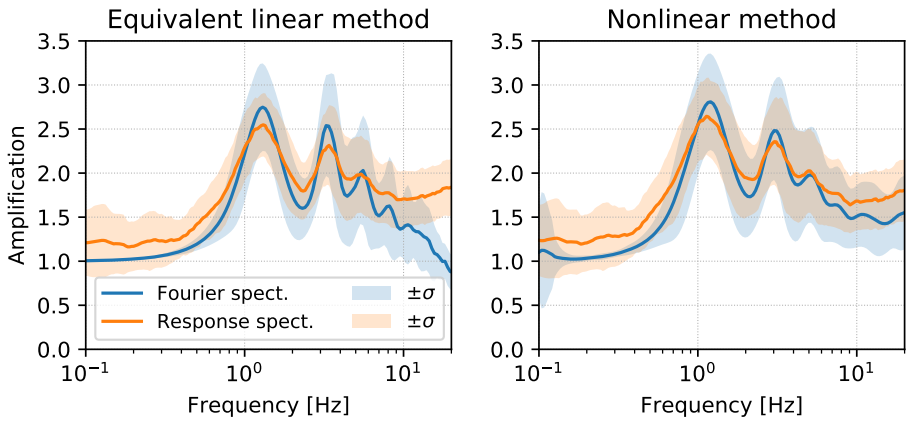


Figure 4.6: A comparison of the response spectral amplification factors and Fourier-based amplification factors, for both the equivalent linear and nonlinear methods. This example uses $V_{S30} = 250$ m/s, $z_1 = 75$ m/s, and PGA $= 0.01g$.

the mean and the (unbiased) standard deviation of these 1,600 output-input ratios (i.e., AF, PhF, and RSR), and denote them as the amplification/phase factors of a certain V_{S30}-z_1-PGA combination.

Note that we performed both NL_{HH} and EQ_{HH} simulations, so there are two versions of site factors for each V_{S30}-z_1-PGA combination.

4.3.3 Amplification factors

This subsection discusses how amplification factors are affected by different aspects, such as the simulation method, V_{S30}, z_1, or PGA.

4.3.3.1 Response spectral versus Fourier spectral amplification factors

The NGA GMPEs have presented the site factors as RSR(f). And when adjusting time-domain signals to incorporate site effects, seismologists and engineers have often been directly using RSR in place of FSR, often overlooking the adverse implications of such a choice.

Figure 4.6 demonstrates the difference between RSR and FSR. We can see the peak amplifications of RSR are lower than those of FSR, and the overall shape of RSR are more "spread out" across frequencies. Especially, in higher frequencies (above 10 Hz), RS shows an trend of increasing with frequency. The example in the figure uses $V_{S30} = 250$ m/s, $z_1 = 75$ m/s, and PGA $= 0.01g$; similar patterns also occur for other V_{S30}-z_1-PGA combinations.

This phenomenon stems from the fact that response spectra only indirectly reflect the

frequency contents of a ground motion through a series of single-degree-of-freedom oscillators, and hence the response spectral value at a specific frequency reflects not only energies in that frequency, but those in other frequencies. Therefore, using RSR instead of FSR to adjust time-domain signals can introduce undesirable—and more importantly, unquantifiable—errors. Moreover, RSR does not have information about the phase delay of different frequencies, which means that the output and input motions are synchronous.

Bora et al. (2016) focused on this particular issue and have documented the same phenomenon. They also proposed procedures to adjust RSR to FSR, which, alas, required additional information (such as stress drop) not available for the users of site factors (who often only know V_{S30}, z_1, and PGA).

4.3.3.2 Equivalent linear versus nonlinear amplification factors

The validation study in Section 3.4 (on page 50) points out the disadvantage of the equivalent linear method for site response analysis: it can "over-damp" the higher frequencies of a ground motions, sometimes to an extent of complete elimination of high frequencies.

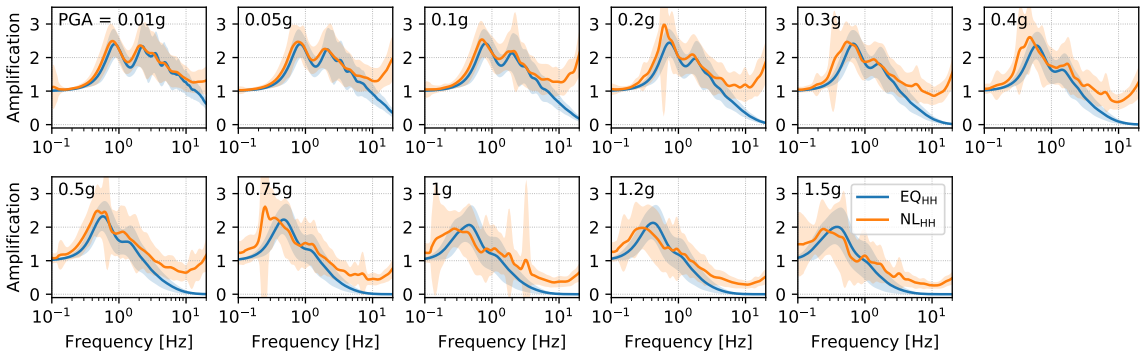
Figure 4.7 (page 86) illustrates this difference by showing an example of $V_{S30} = 250$ m/s and $z_1 = 150$ m/s (other V_{S30}-z_1 combinations also exhibit the same pattern). For |FSR|, EQ_{HH} produces lower amplification than NL_{HH} above around 5 Hz, and almost zero amplification for stronger input motions. And for RSR, the difference between EQ_{HH} and NL_{HH} is quite small.

This indicates that KAS14's use of RSR as the amplification factor unintentionally mitigates the disadvantages of their use of the equivalent linear method. Nevertheless, the unquantifiable errors introduced by RSR (see Section 4.3.3.1) prompts us to discourage its use, and hence to also discourage using the equivalent linear method. Therefore, the recommended amplification factors of SAG19 are |FSR|, and produced by NL_{HH}.

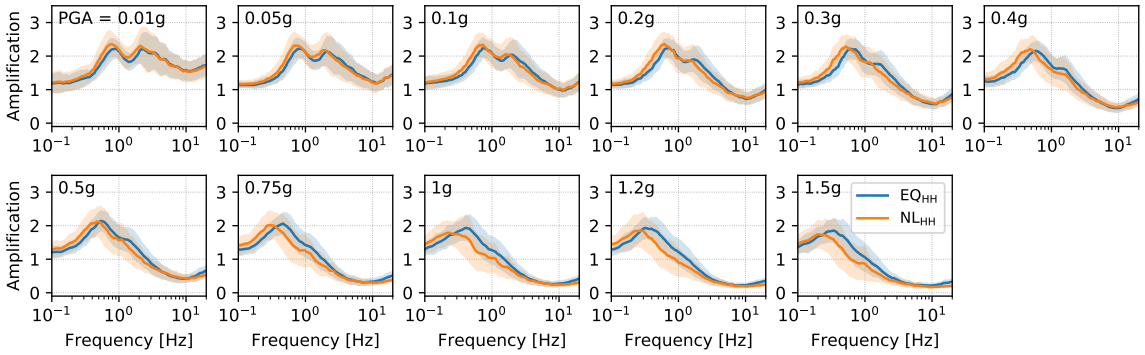
4.3.3.3 Trend of amplification with respect to input PGA

Figure 4.8 shows the trend of the amplification with respect to input PGA (i.e., PGA on rock outcrop) for six representative V_{S30}-z_1 combinations. There are two major trends associated with the increase of PGA: (1) the shift in the fundamental frequency and (2) the change in the overall shape of the amplitude.

The fundamental frequency, f_0, is defined as the lowest frequency at which the amplification peaks. From all subplots, we observe that f_0 decreases with increasing PGA. This is because when PGA increases, the nonlinearity level in the soils increases, effectively making the



(a) Fourier spectra amplification



(b) Response spectral amplification

Figure 4.7: A comparison of amplification factors produced by the equivalent linear method (EQ_{HH}) and the nonlinear method (NL_{HH}). This example uses $V_{S30} = 250$ m/s, $z_1 = 150$ m/s. The input PGA (i.e., rock-outcrop PGA) are annotated in each subplot. (a) The Fourier spectra amplification (i.e., |FSR|); (b) the response spectral amplification (i.e., RSR).

soils softer (i.e., having lower V_S), hence exhibiting lower f_0. This phenomenon is also confirmed in seismic records, such as in the work by *Wu et al.* (2009); *Wu and Peng* (2011).

On the other hand, the change in the overall shape of the amplitude with respect to PGA is not as monotonic as the change of f_0. With increasing PGA, here are the trends:

- For softer sites (e.g., $V_{S30} = 200$ m/s in Figure 4.8), the overall shape shifts leftwards and downwards.
- For medium-stiffness sites (e.g., $V_{S30} = 500$ m/s in Figure 4.8), the overall shape shifts leftwards, with amplitudes unchanged in lower frequencies and slightly decreasing in higher frequencies.
- For stiff sites (e.g., $V_{S30} = 800$ m/s in Figure 4.8), the shape shifts leftwards, with amplitudes increasing in lower frequencies and slightly decreasing in higher frequen-

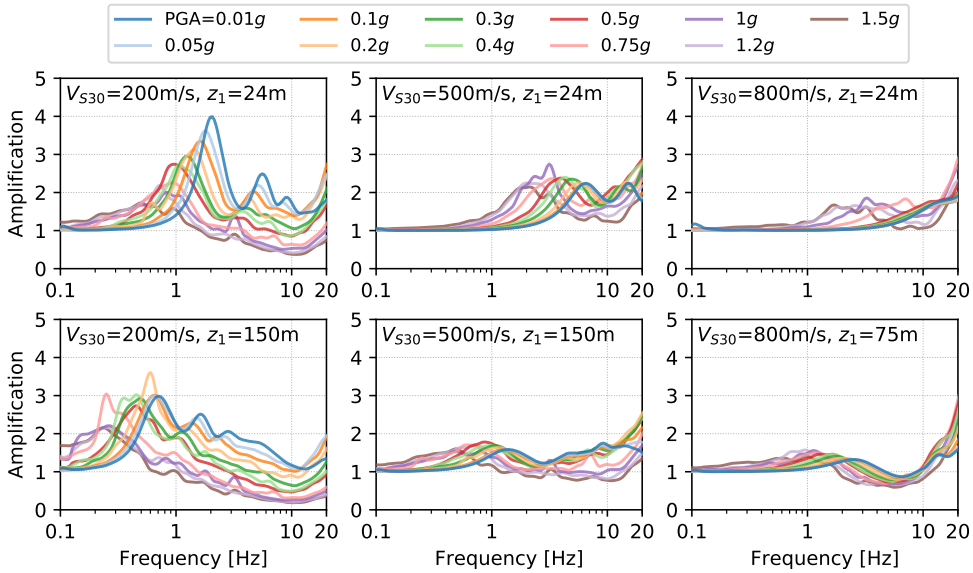


Figure 4.8: The trend of the amplification factors with respect to input PGA, for six representative V_{S30}-z_1 combinations. These factors are Fourier-based, and produced by NL$_{HH}$.

cies.

4.3.3.4 Trend of amplification with respect to V_{S30}

Figure 4.9 shows the trend of the amplification with respect to V_{S30}, for eight representative z_1-PGA combinations.

The fundamental frequency (f_0) and the whole shape of the amplification shift rightwards with the increase of V_{S30}. The overall amplification level shifts downwards for all other z_1-PGA combinations except (16 m, 0.75g) and (16 m, 1.5g), which indicates that when soil nonlinearity increases (i.e., input PGA increases) on shallow sites (low z_1), there would be higher site amplification. This is because even though soil nonlinearity increases damping (which dampens the motion), it also significantly increases the velocity contrast between soils and the bedrock, so the overall site amplification still increases. This does not seem to be the case for deeper sites.

4.3.3.5 Trend of amplification with respect to z_1

Figure 4.10 shows the trend of the amplification with respect to z_1, for 12 representative V_{S30}-PGA combinations.

The overall shape of the amplification, as well as f_0, shifts to the left as z_1 increases from the 8 m to about 300 m. Beyond 300 m, the amplification becomes more or less unchanged

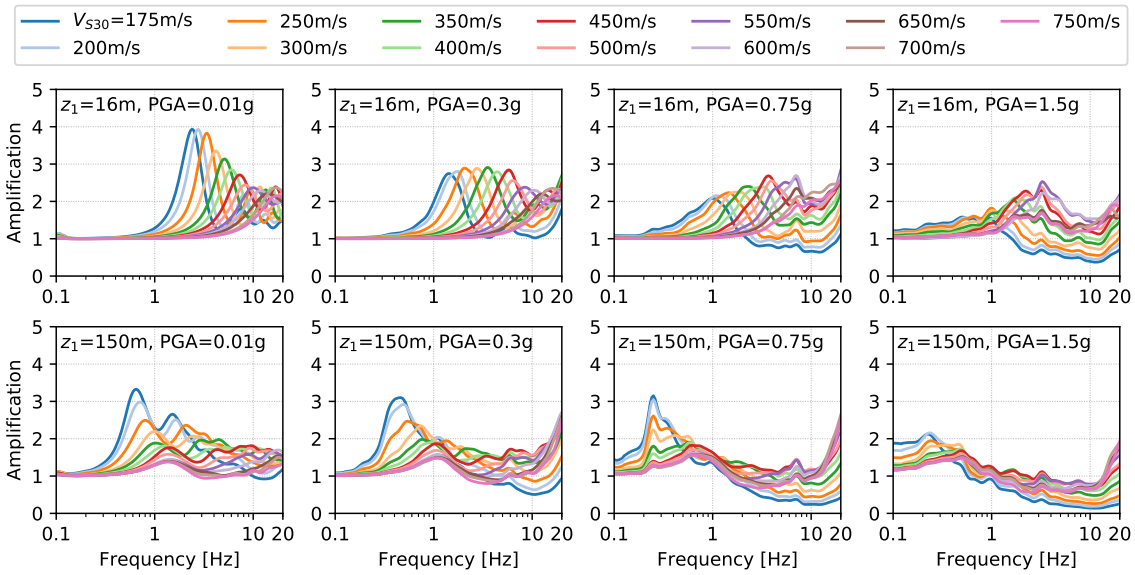


Figure 4.9: The trend of the amplification factors with respect to V_{S30}, for eight representative z_1-PGA combinations. These factors are Fourier-based, and produced by NL_{HH}.

for different V_{S30} and PGA levels. Note that this observation is naturally consistent with Figure 4.4 (on page 80), which is only the linear case (i.e., small input PGA). This indicates that basin depth, z_1, can only significantly influence site amplification when it is shallower than around 300 m.

4.3.4 Phase factors

This subsection discusses the characteristics of phase factors, the practicalities of calculating phase factors, and how phase factors can be affected by aspects such as V_{S30}, z_1, and PGA.

4.3.4.1 Unwrapping phase factors

The phase factors, PhF(f), are always bounded between $[-\pi, \pi]$, so in order to aggregate the site response simulations results (by averaging different randomized V_S profiles of the same V_{S30}-z_1 combination as well as different input motions of the same PGA), we need to unwrap the phase angles[11].

However, the phase factors calculated from NL_{HH} contain too much noise (due to the lack of energy at many discrete frequency values) to be correctly unwrapped. This is because the unwrapping algorithm traverses along a given phase signal, and adjusts the signal by 2π if

[11] For example, if we average two unwrapped phase angles: 0.95π and 1.01π, we correctly get 0.98π. But if these two angles are wrapped, they become 0.95π and -0.99π respectively, and their average becomes -0.02π, which is incorrect.

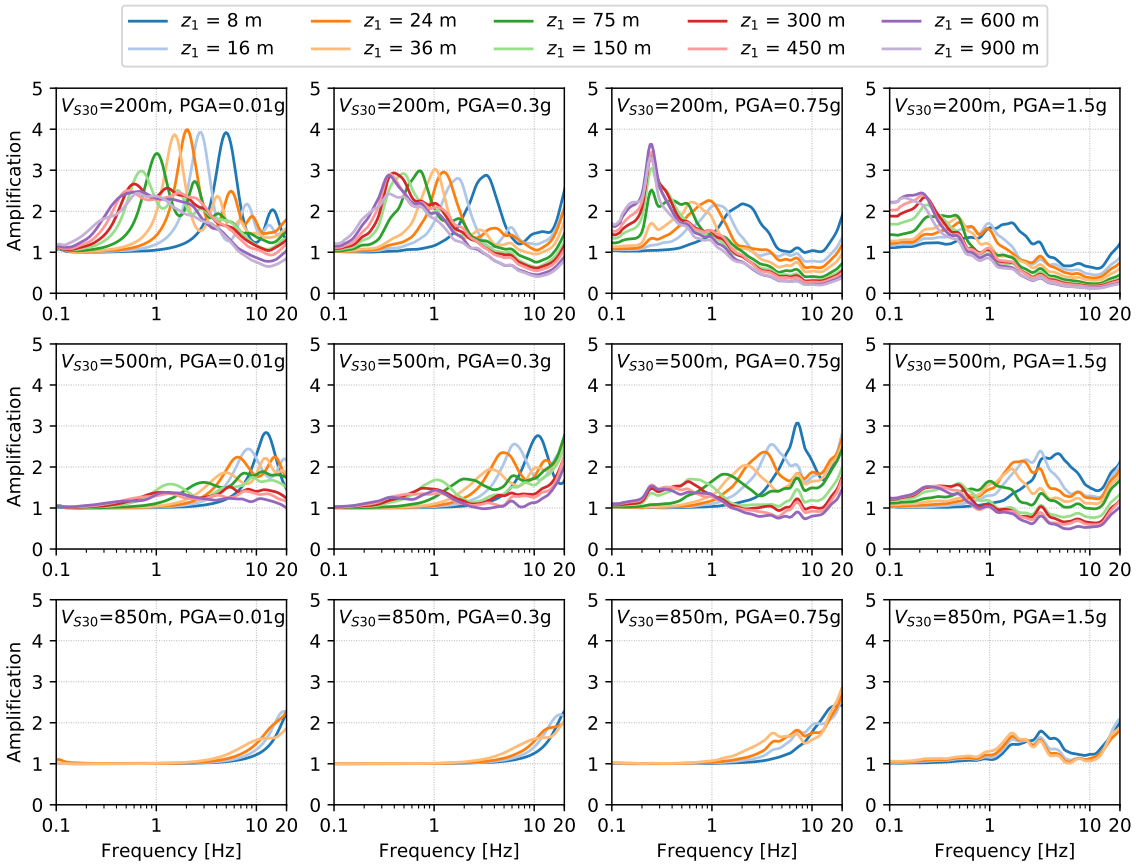


Figure 4.10: The trend of the amplification factors with respect to z_1, for 12 representative V_{S30}-PGA combinations. These factors are Fourier-based, and produced by NL_{HH}.

it detects a "valid jump" (i.e., a discontinuity greater than 2π) between two adjacent signal points. Therefore, signals with excessive noises would have their "valid jumps" disguised behind noise, making the unwrapping algorithm unable to properly unwrap it.

Figure 4.11 (on page 91) shows the comparison between the phase factors calculated from EQ_{HH} and NL_{HH} (for the same V_{S30}, z_1, and PGA). We can see that the NL_{HH} results are often too noisy to be correctly unwrapped[12], except for linear site response on stiff and shallow sites, such as Figure 4.11(a).

Therefore, we choose to adopt the phase factors calculated from EQ_{HH} in SAG19. Since the equivalent linear method can prescribe overly low V_S and overly high damping for the higher frequencies, we expect to introduce some errors. However, this has become an inevitable compromise.

[12]Correctly unwrapped phase factors need to be negative for all frequencies, because the phase values indicate the phase shifts between the output and the input. Since wave propagation always takes positive amount of time, the phase shifts are always negative.

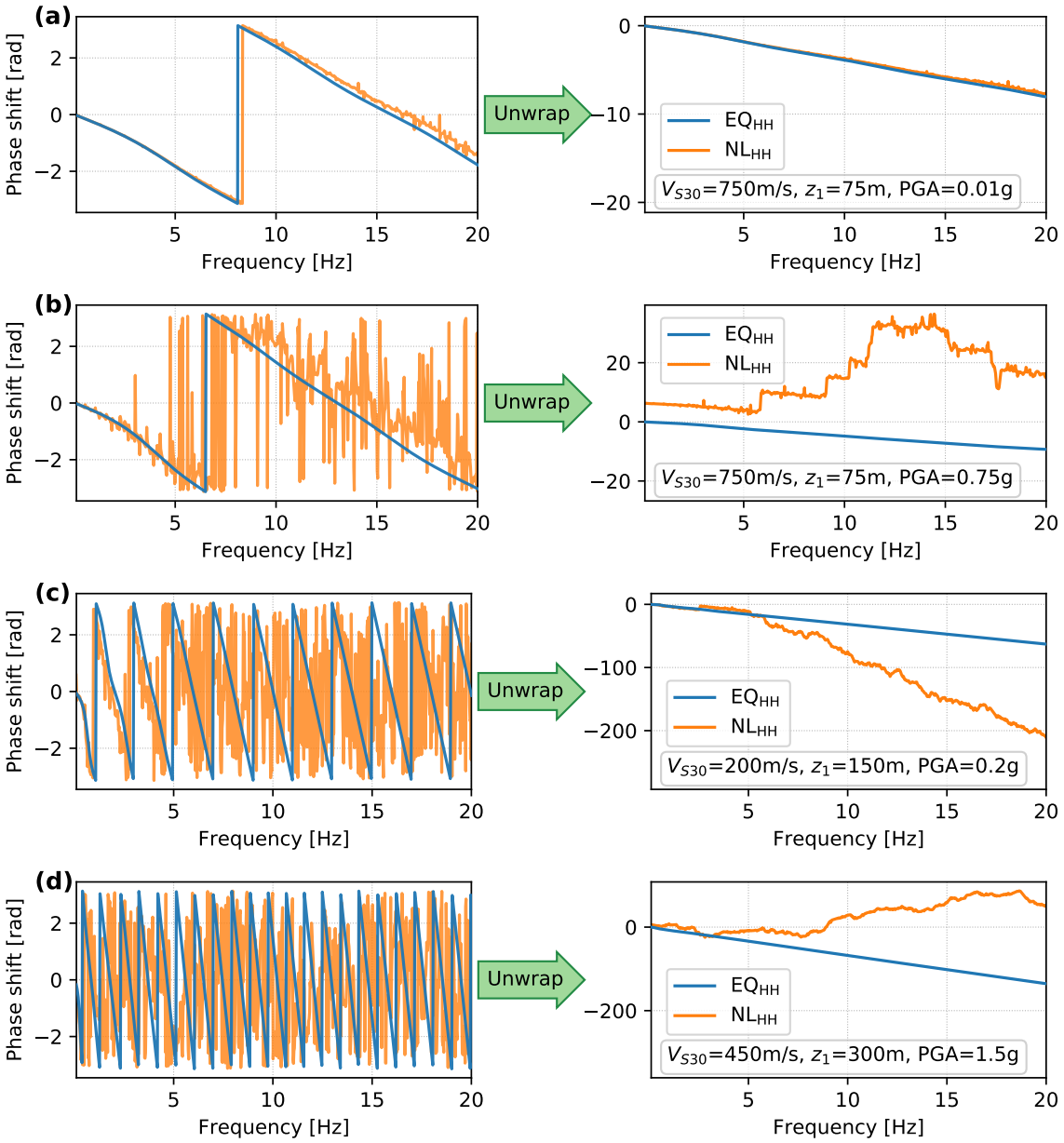


Figure 4.11: A comparison between the phase factors by EQ$_{HH}$ and NL$_{HH}$. The NL$_{HH}$ results contain too much noises to be correctly unwrapped.

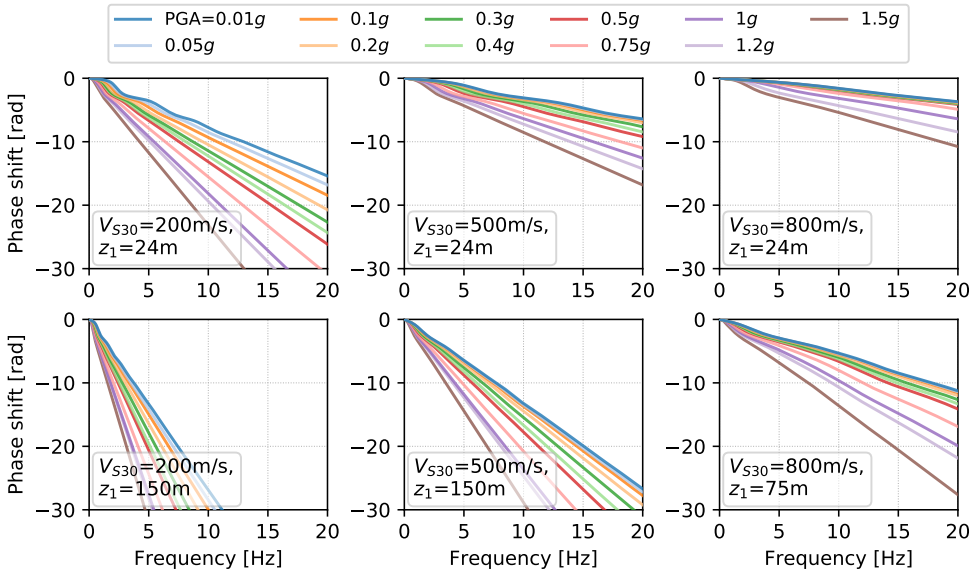


Figure 4.12: The trend of the phase-shift factors with respect to input PGA, for six representative V_{S30}-z_1 combinations. These factors are Fourier-based, and produced by EQ$_{\text{HH}}$.

The amplification factors calculated from NL$_{\text{HH}}$ also contain a lot of noise, but unlike phase, the amplitude does not need to be unwrapped, therefore much of the noise cancels each other out in the averaging process, making the NL$_{\text{HH}}$ amplification factors smooth enough.

4.3.4.2 Trend of phase with respect to input PGA

Figure 4.12 (on page 92) shows the trend of the phase with respect to input PGA (i.e., PGA on rock outcrop) for six representative V_{S30}-z_1 combinations. We can see that higher PGA causes higher soil nonlinearity, hence reduced V_S, which increases the time lag (steeper overall slope in the figure).

4.3.4.3 Trend of phase with respect to V_{S30}

Figure 4.13 (on page 92) shows the trend of the phase with respect to V_{S30}, for eight representative z_1-PGA combinations. With the increase of V_{S30}, the overall velocities in the soils increase, which reduces the time lag.

4.3.4.4 Trend of phase with respect to z_1

Figure 4.14 shows the trend of the phase with respect to z_1, for 12 representative V_{S30}-PGA combinations. Increasing z_1 leads to increasing travel time, hence more time lag between the output and the input motions.

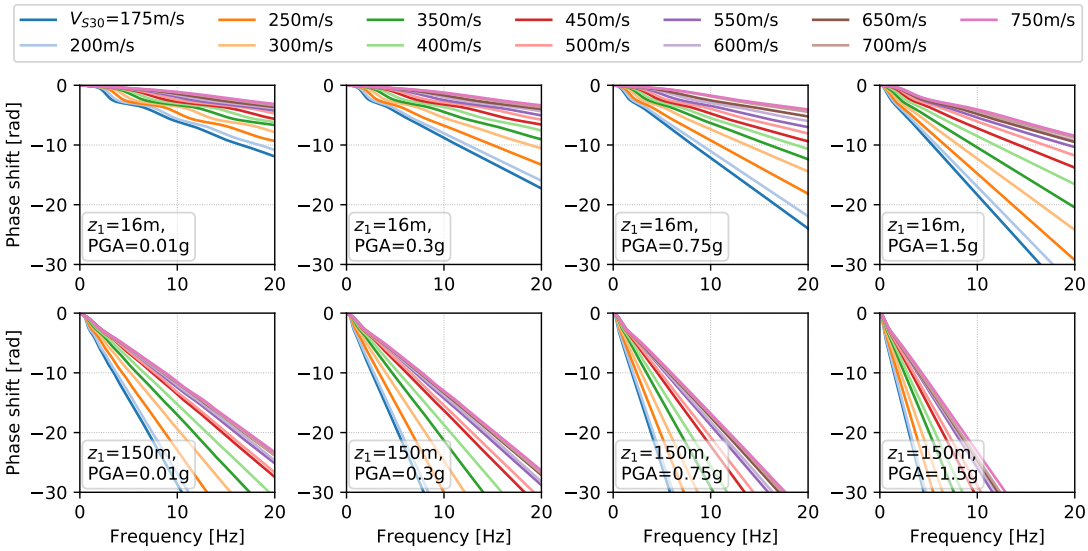


Figure 4.13: The trend of the phase-shift factors with respect to V_{S30}, for eight representative z_1-PGA combinations. These factors are Fourier-based, and produced by EQ$_{HH}$.

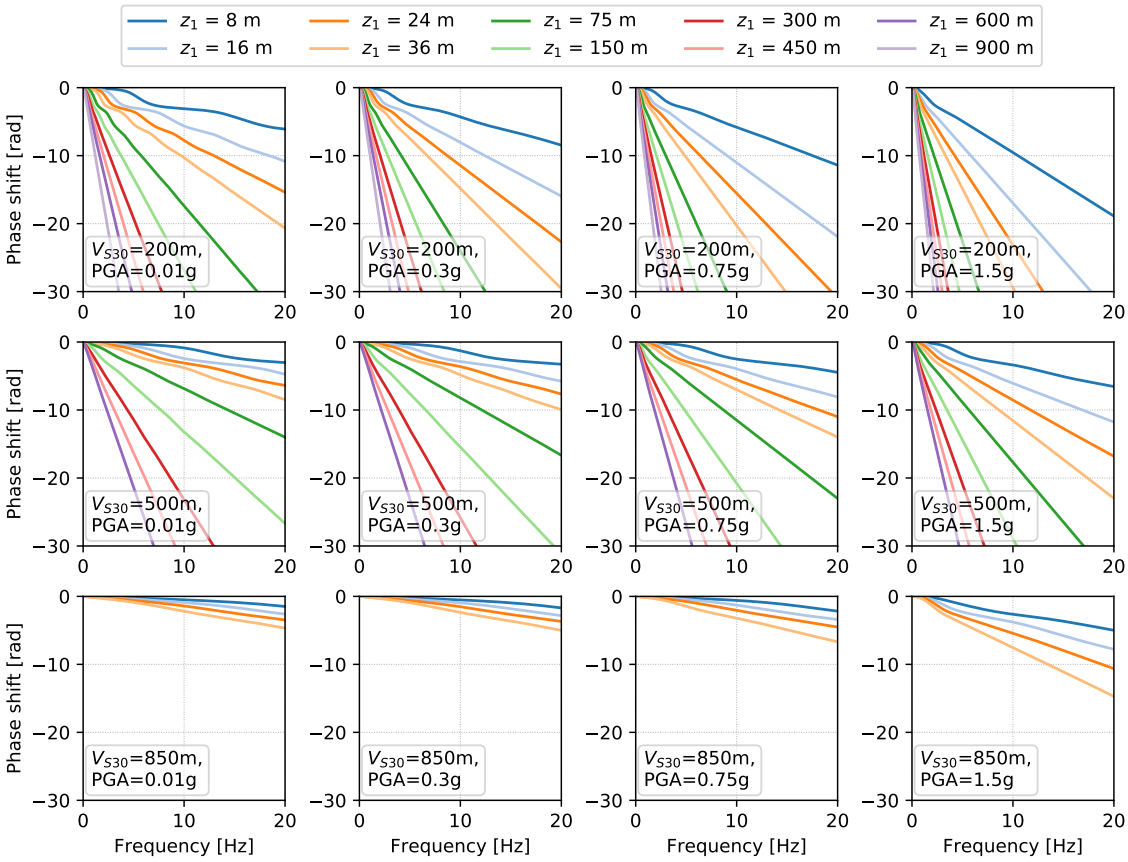


Figure 4.14: The trend of the phase-shift factors with respect to z_1, for 12 representative V_{S30}-PGA combinations. These factors are Fourier-based, and produced by EQ$_{HH}$.

4.4 Comparison of amplification factors of KAS14, HIH18, and SAG19

This section compares the amplification factors of KAS14, HIH18, and SAG19. Phase factors are not compared because neither KAS14 or HIH18 produces them.

The KAS14 results come from the electronic supplements (Microsoft Excel files) that come with *Kamai et al.* (2013), and the HIH18 results come from digitizing the PDF file of *Hashash et al.* (2018) using Adobe Acrobat.

4.4.1 Reference site conditions

Reference site conditions are quantified by reference V_{S30} (henceforth denoted as V_{ref}), which is the threshold beyond which a certain site factor model (e.g., SAG19 or HIH18) assumes that there is no site effects. For example, if a site factor model's V_{ref} is 760 m/s, then for sites with $V_{S30} \geqslant 760$ m/s, this site factor model reports all 1's (i.e., no amplification). This means that a site factor model whose V_{ref} is on the lower end (such as 760 m/s) has a limited capability in predicting site amplification—especially nonlinear amplification.

4.4.1.1 Adjusting different reference V_{S30}

The V_{ref} are different between KAS14 (1,180 m/s), HIH18 (nominally, 760 m/s), and SAG19 (1,046 m/s). Different NGA-West2 GMPEs also uses different V_{ref}: ASK14 (which partially adopts the simulation results of KAS14) uses 1,180 m/s, CB14 uses 1,100 m/s, CY14 uses 1,130 m/s, and BSSA14 uses 760 m/s.

Therefore, to compare different site factor models, we first need to adjust them according to their respective V_{ref}. Here we use a procedure recommended by Professor Adrian Rodriguez-Marek of Virginia Tech (personal communication, December, 2018) and Professor Jonathan P. Stewart of UCLA (personal communication, January, 2019).

Suppose we want to adjust site amplification factors produced by a site factor model whose $V_{\text{ref}} = V_{\text{raw}}$ (e.g., 760 m/s) into a target reference $V_{S30} = V_{\text{target}}$ (e.g., 1,046 m/s). The procedure is as follows:

1. From the four NGA-West2 GMPEs mentioned above, find the one whose V_{ref} is closest to V_{raw}, then:
    a) Calculate the predicted ground motion (as a function of period or frequency) using V_{raw} and a reasonable range of magnitude, epicentral distance, ... Denote the predicted ground motion as GM_{raw}.

b) Calculate the predicted ground motion (as a function of period or frequency) using V_{target} and a reasonable range of magnitude, epicentral distance, ... Denote the predicted ground motion as $\text{GM}_{\text{target}}$.

c) Calculate the ratio: $R = \text{GM}_{\text{raw}}/\text{GM}_{\text{target}}$. Different magnitudes and epicentral distances can result in slightly different R, so take the mean value as necessary, but if both V_{raw} and V_{target} are large, the variation in R is usually very small. Note that R is a function of frequency (or period).

2. Using the same GMPE as Step 1,

a) Calculate the predicted PGA using V_{raw} and a reasonable range of magnitude, epicentral distance, ... Denote the PGA as PGA_{raw}.

b) Calculate the predicted PGA using V_{target} and a reasonable range of magnitude, epicentral distance, ... Denote the PGA as $\text{PGA}_{\text{target}}$.

c) Calculate the ratio: $\beta_{\text{PGA}} = \text{PGA}_{\text{target}}/\text{PGA}_{\text{raw}}$. Note that β_{PGA} is a single number.

3. In the "raw" site factor model ($V_{\text{ref}} = V_{\text{raw}}$), pick a V_{target}-z_1-PGA combination, and calculate the site amplification factor using $\text{PGA}/\beta_{\text{PGA}}$ as input PGA. Denote the resultant site amplification factor as AF_{raw}.

4. $R \cdot \text{AF}_{\text{raw}}$ should be comparable with the amplification factor queried from the "target" site factor model using the V_{target}-z_1-PGA combination.

With the procedure above, we adjust the KAS14 and HIH18 amplification factors to be comparable with the SAG19 amplification factors. We use the V_{S30}-z_1 combinations from in KAS14 that are also within the valid range of HIH18 and SAG19 (i.e., V_{S30} = 190, 270, 400, 560, 760, 850, and 900 m/s; z_1 = 50, 120, 250, 500, 1000, and 2000 ft), and some representative PGA levels: 0.01, 0.05, 0.1, 0.2, 0.5, 0.75, 1.0, and $1.25g$.

4.4.2 Analyses of the differences

Figures 4.15 to 4.24 (pages 98 to 107) show the comparison of amplification factors of KAS14, HIH18, and SAG19.

This subsection briefly summarizes the differences observed in those figures.

4.4.2.1 Linear site amplification

We use the subplot with PGA = $0.01g$ to compare the linear site amplification from KAS14, HIH18, and SAG19.

For softer sites ($V_{S30} \leqslant 400$ m/s), when $z_1 \geqslant 120$ ft, the HIH18 and SAG19 linear results are quite similar. But when $z_1 = 50$ ft, the HIH18 linear amplifications appear to have a lower f_0 and higher amplification than SAG19. This indicates that site amplifications are rather sensitive to z_1 when z_1 is small. It is worth noting that at these shallow sites, the f_0 of linear amplifications by KAS14 are very similar to SAG19.

Overall, the linear site amplifications by KAS14 are higher than those by HIH18 and SAG19 by a non-negligible amount—for $V_{S30} \leqslant 560$ m/s and all z_1. Especially in some cases—such as (V_{S30}, z_1) =(190 m/s, 500 ft), (190 m/s, 1000 ft), (190 m/s, 2000 ft), (270 m/s, 500 ft), (270 m/s, 1000 ft), and (270 m/s, 2000 ft)—the linear amplification of KAS14 is almost 2.0 at $f = 0.1$ Hz, which clearly is beyond the threshold of random error[13]. Therefore, we believe that the KAS14 linear site factors should be used with caution.

Since linear site amplification in theory depends only on the site condition (V_S and small-strain damping profiles), and the base V_S profiles used in KAS14, HIH18, and SAG19 are quite similar, we narrow down the causes for KAS14's different results to the following three possibilities:

1. The small-strain damping ($D_{\min}$) used in KAS14 is too small. As mentioned in Section 4.1.3.3, KAS14 did not document how they picked $D_{\min}$. (The $D_{\min}$ values used in HIH18 and SAG19 both come from literature and are similar.)

2. The V_S profile randomization scheme by *Toro* (1995) used in KAS14. However, Figure 4.3 (on page 79) shows that even with Toro's original randomization scheme, the average linear amplification is still not as high as in KAS14. So this possibility can be eliminated.

3. The random-vibration-theory (RVT) equivalent linear method used in KAS14. RVT differs from the classical equivalent linear method (i.e., the EQ_{HH} method used in SAG19) in that RVT modifies response spectra, rather than time-domain signals. And we think this is the most likely cause.

[13] One-dimensional wave propagation theories indicate that the amplification should be 1.0 for very low frequencies, regardless of small-strain damping level.

4.4.2.2 Nonlinear site effects

Nonlinear site effects manifest when the input ground motion increases. So in Figures 4.15 to 4.24, the differences between the subplots with PGA $> 0.01g$ and the subplot with PGA $= 0.01g$ are the nonlinear site effects.

From the figures, we observe different levels of f_0 reduction among KAS14, HIH18, and SAG19, for the same amount of PGA increase (f_0 reduction associated with soil nonlinearity is discussed earlier in Section 4.3.3.3). The f_0 reduction in SAG19 the largest, followed by KAS14. But in HIH18, we see almost no f_0 reduction—compared to the linear case (PGA $= 0.01g$)—even when PGA is as high as $1.25g$.

We attribute this observation in HIH18's results to the fact that HIH18 uses the GQ/H model (*Groholski et al.*, 2016) as the nonlinear stress-strain model. The GQ/H model has not demonstrated the ability to fit its parameters to laboratory data well or to reconstruct realistic soil behaviors from V_S alone. On the other hand, the HH model used in SAG19 has demonstrated the ability to fit its parameters satisfactorily to laboratory data (Figure 3.4 on page 41), as well as reconstruct realistic stress-strain curves from V_S (Figure 3.5 on page 42). And as a result, the site response analyses based on the HH model obtained satisfactory prediction performance, even for some extremely strong ground motions (see Figure 3.15 on page 67).

4.4.2.3 Site amplification for stiffer sites (with larger V_{S30})

A closer look at the stiffer sites ($V_{S30} \geqslant 760$ m/s, Figures 4.23 and 4.24) reveals that the HIH18 site amplifications are still larger than 1.0 (before V_{ref} adjustments). By definition, there should not be any amplification on sites with $V_{S30} \geqslant V_{\text{ref}}$, because the nominal V_{ref} reported in HIH18 is 760 m/s.

This is most likely due to the fact that HIH18 did not truncate its V_S profiles at 760 m/s. The V_S profiles used in the simulation should end at a depth of z_1 and with an ending V_S of 760 m/s. If this is not the case, then the actual V_{ref} is the actual ending V_S of the profile. Therefore, HIH18's actual V_{ref} is no longer the nominal value (760 m/s), and it is even possible that HIH18 has different actual V_{ref} for different input V_{S30}. Therefore, for applications where the V_{ref} of a site factor model needs to be known exactly, we recommend using SAG19.

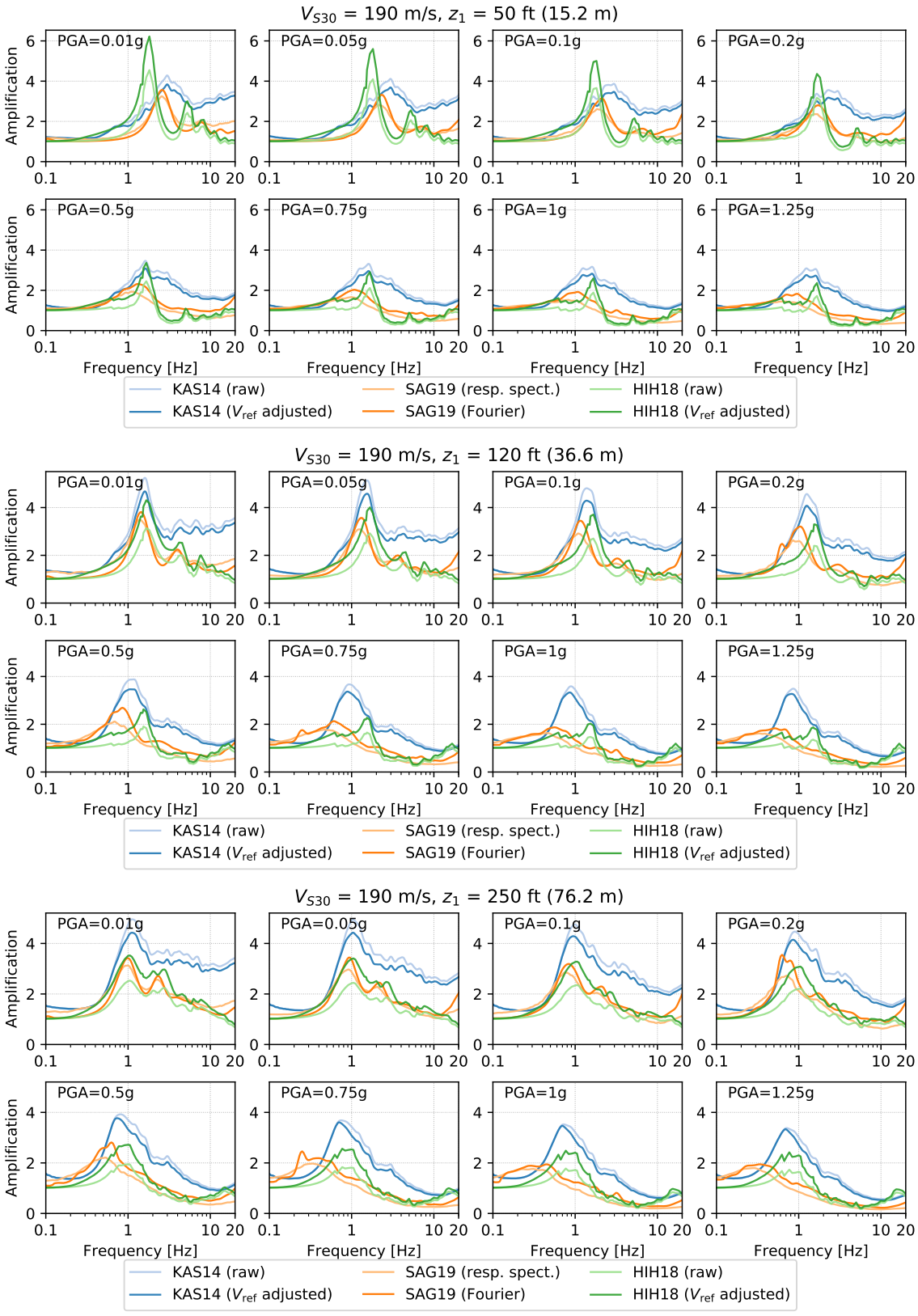


Figure 4.15: Comparison of site amplification factors between KAS14, HIH18, and SAG19, for three V_{S30}-z_1 combination: (190 m/s, 50 ft), (190 m/s, 120 ft), and (190 m/s, 250 ft). Each subplot corresponds to a PGA level.

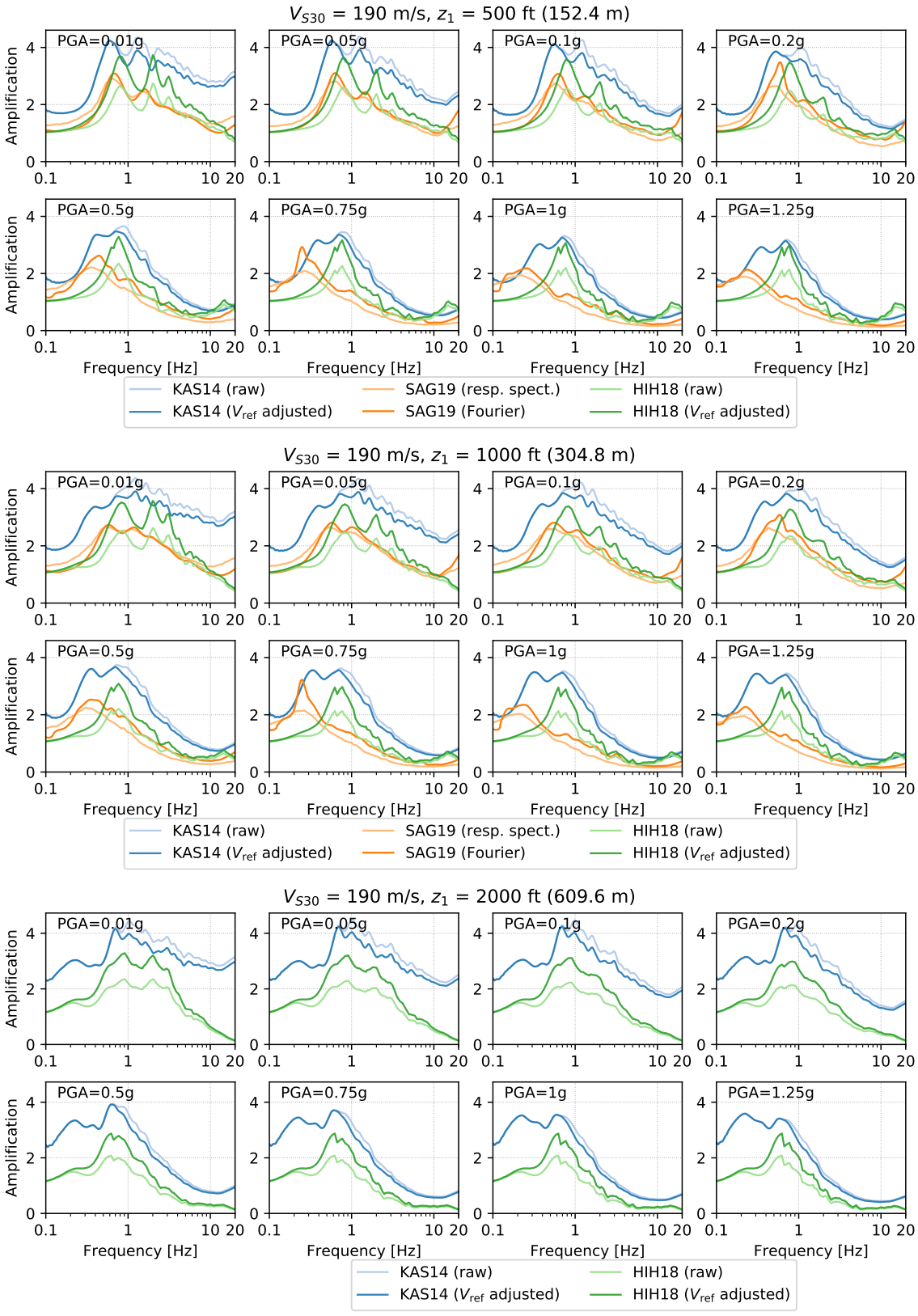


Figure 4.16: Comparison of site amplification factors between KAS14, HIH18, and SAG19, for three V_{S30}-z_1 combination: (190 m/s, 500 ft), (190 m/s, 1000 ft), and (190 m/s, 2000 ft). Each subplot corresponds to a PGA level.

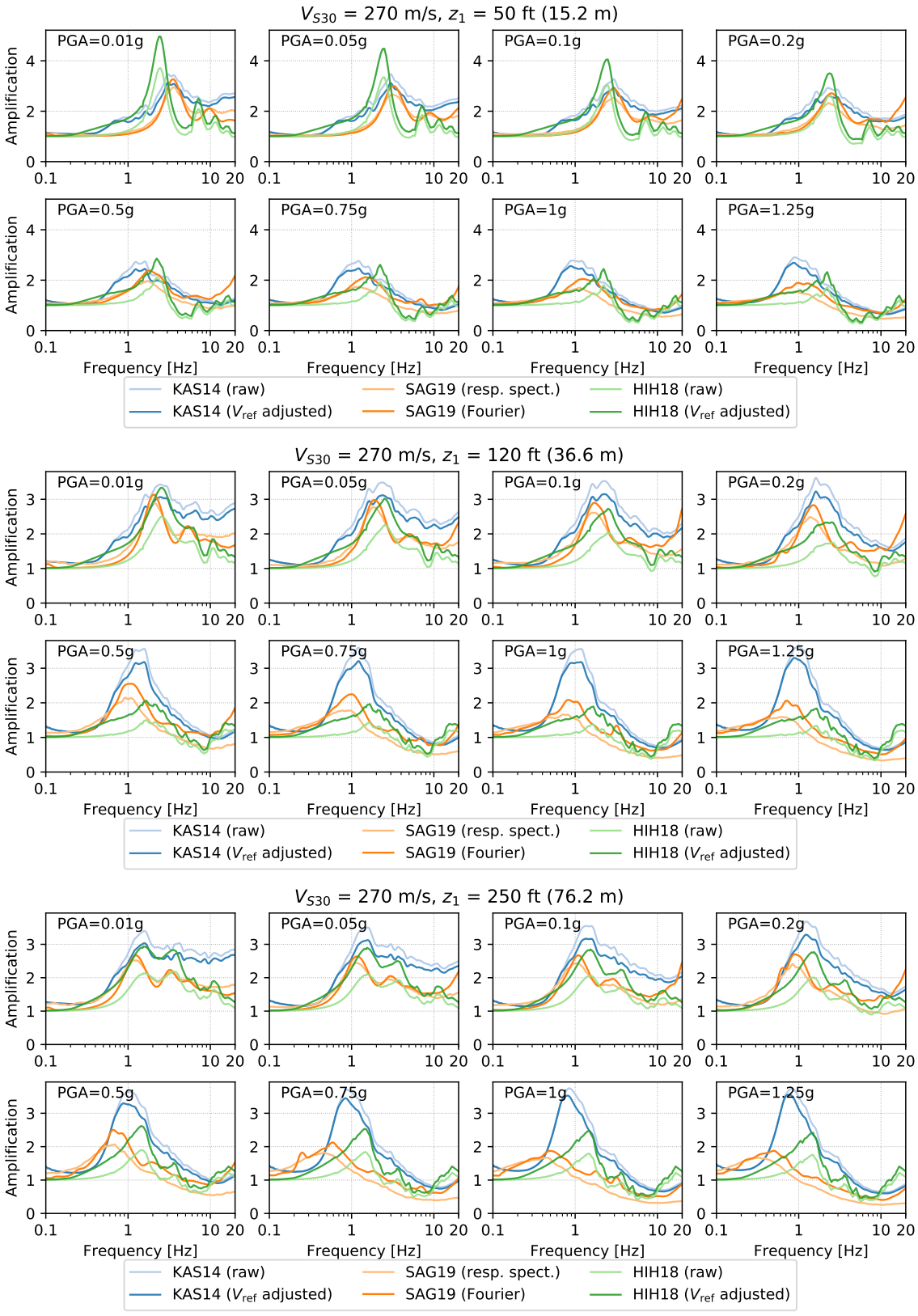


Figure 4.17: Comparison of site amplification factors between KAS14, HIH18, and SAG19, for three V_{S30}-z_1 combination: (270 m/s, 50 ft), (270 m/s, 120 ft), and (270 m/s, 250 ft). Each subplot corresponds to a PGA level.

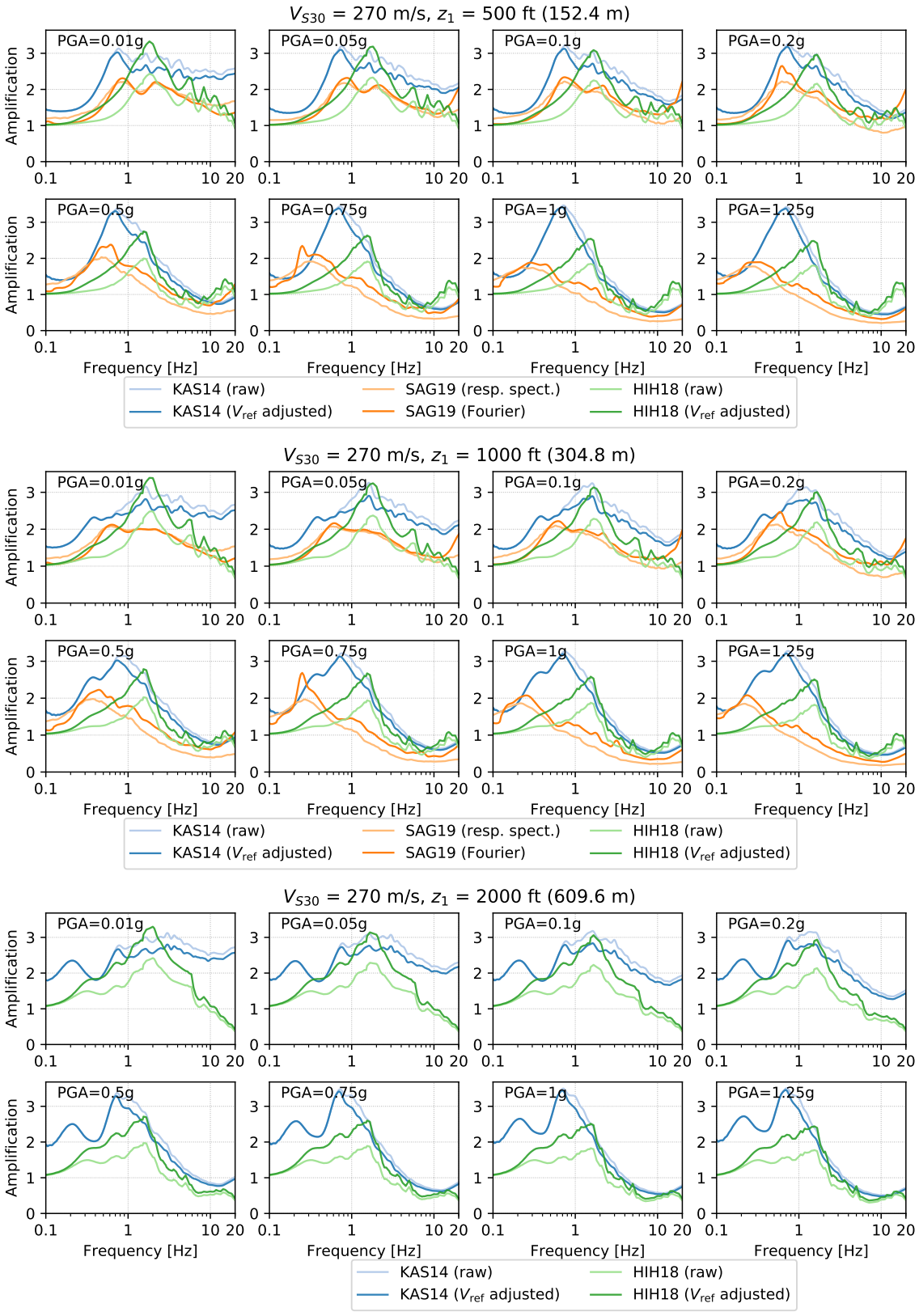


Figure 4.18: Comparison of site amplification factors between KAS14, HIH18, and SAG19, for three V_{S30}-z_1 combination: (270 m/s, 500 ft), (270 m/s, 1000 ft), and (270 m/s, 2000 ft). Each subplot corresponds to a PGA level.

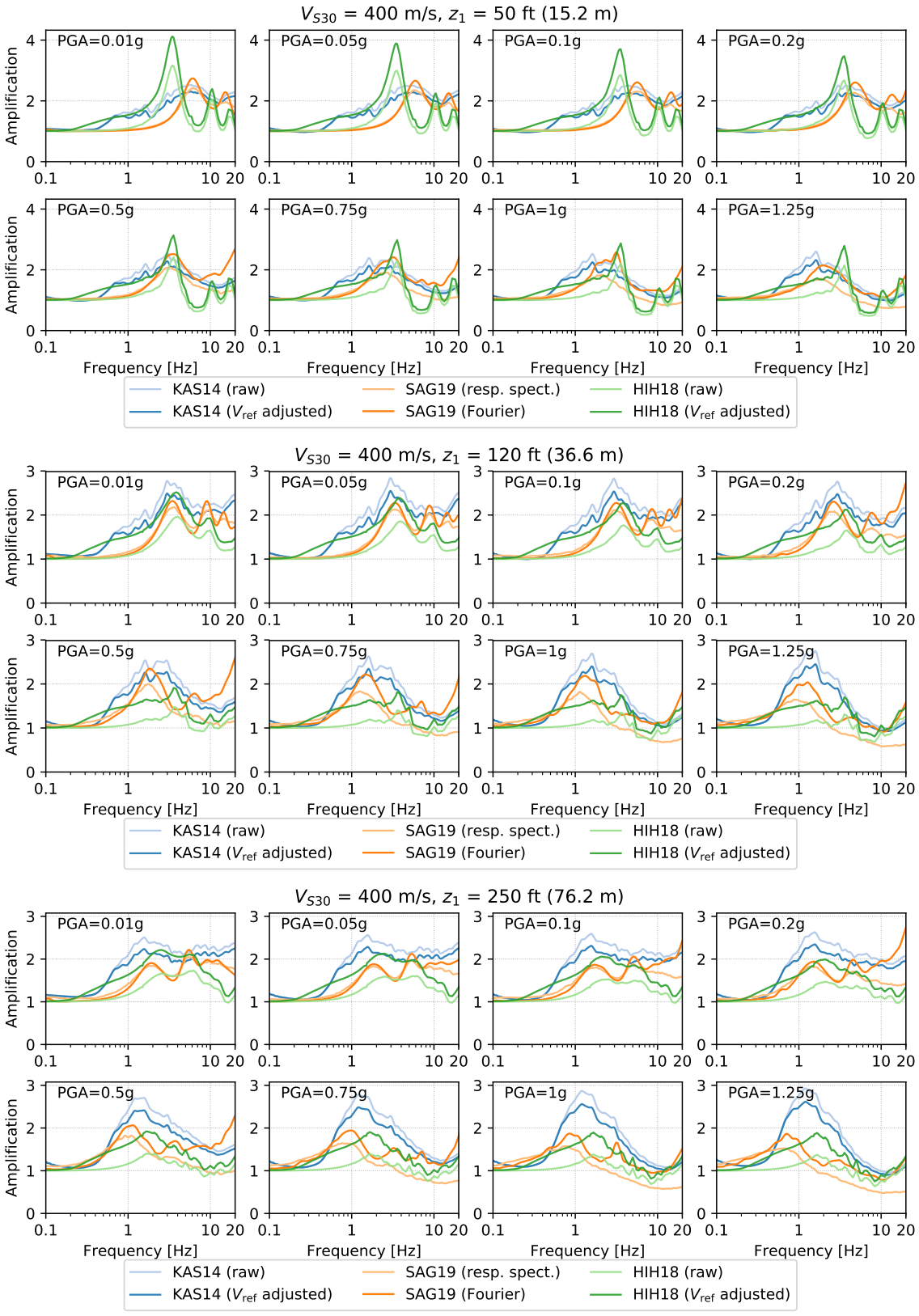


Figure 4.19: Comparison of site amplification factors between KAS14, HIH18, and SAG19, for three V_{S30}-z_1 combination: (400 m/s, 50 ft), (400 m/s, 120 ft), and (400 m/s, 250 ft). Each subplot corresponds to a PGA level.

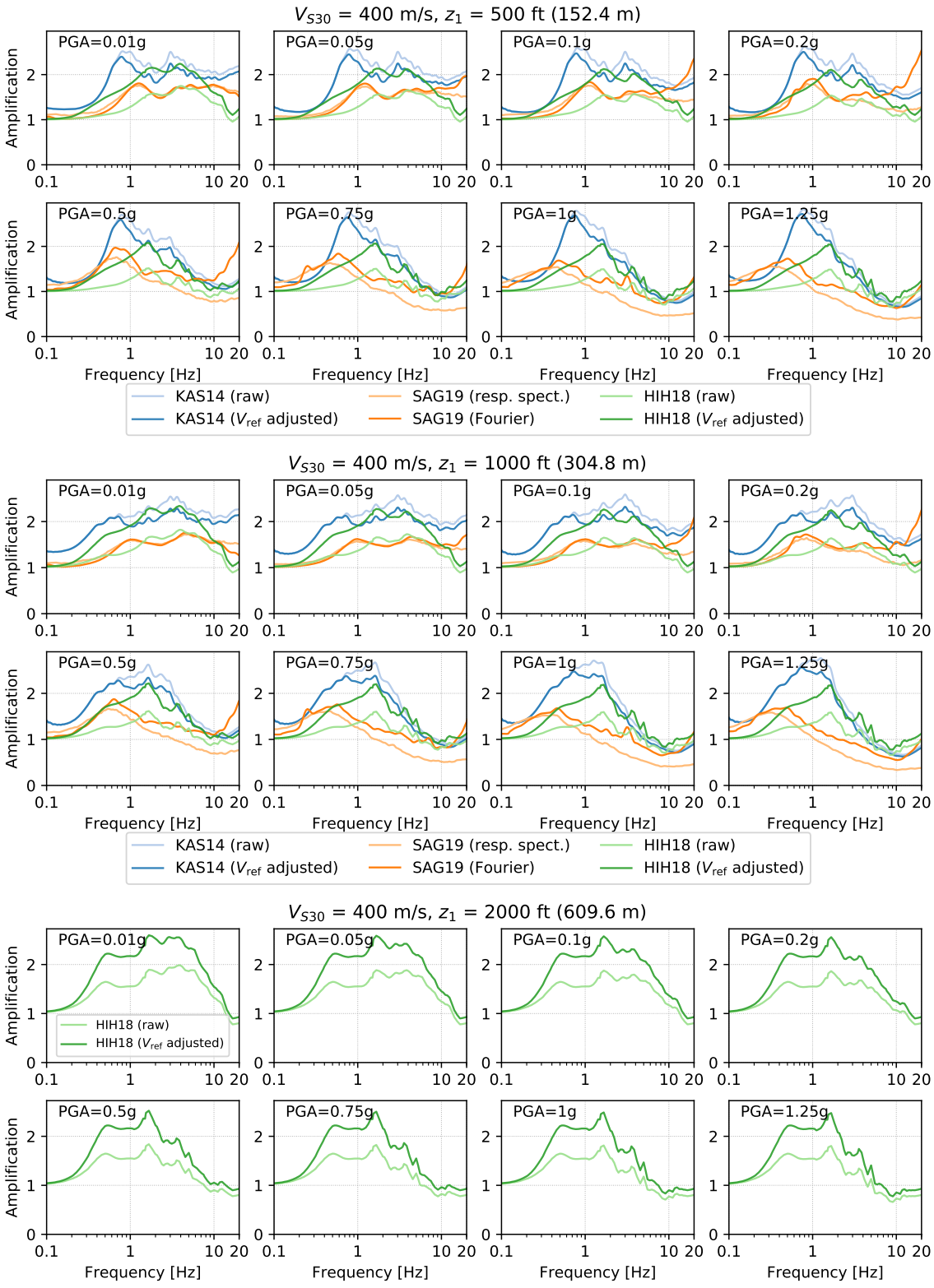


Figure 4.20: Comparison of site amplification factors between KAS14, HIH18, and SAG19, for three V_{S30}-z_1 combination: (400 m/s, 500 ft), (400 m/s, 1000 ft), and (400 m/s, 2000 ft). Each subplot corresponds to a PGA level.

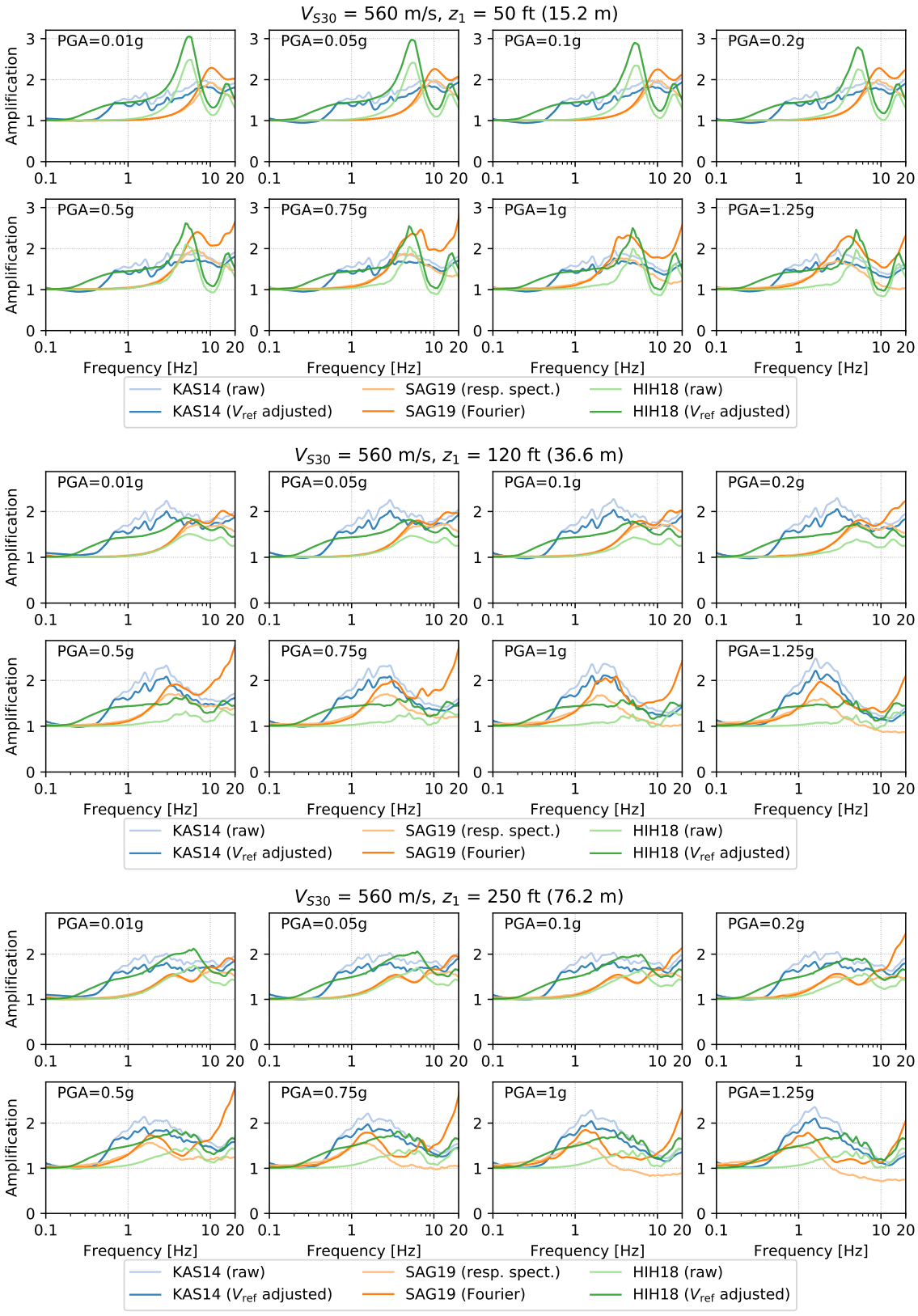


Figure 4.21: Comparison of site amplification factors between KAS14, HIH18, and SAG19, for three V_{S30}-z_1 combination: (560 m/s, 50 ft), (560 m/s, 120 ft), and (560 m/s, 250 ft). Each subplot corresponds to a PGA level.

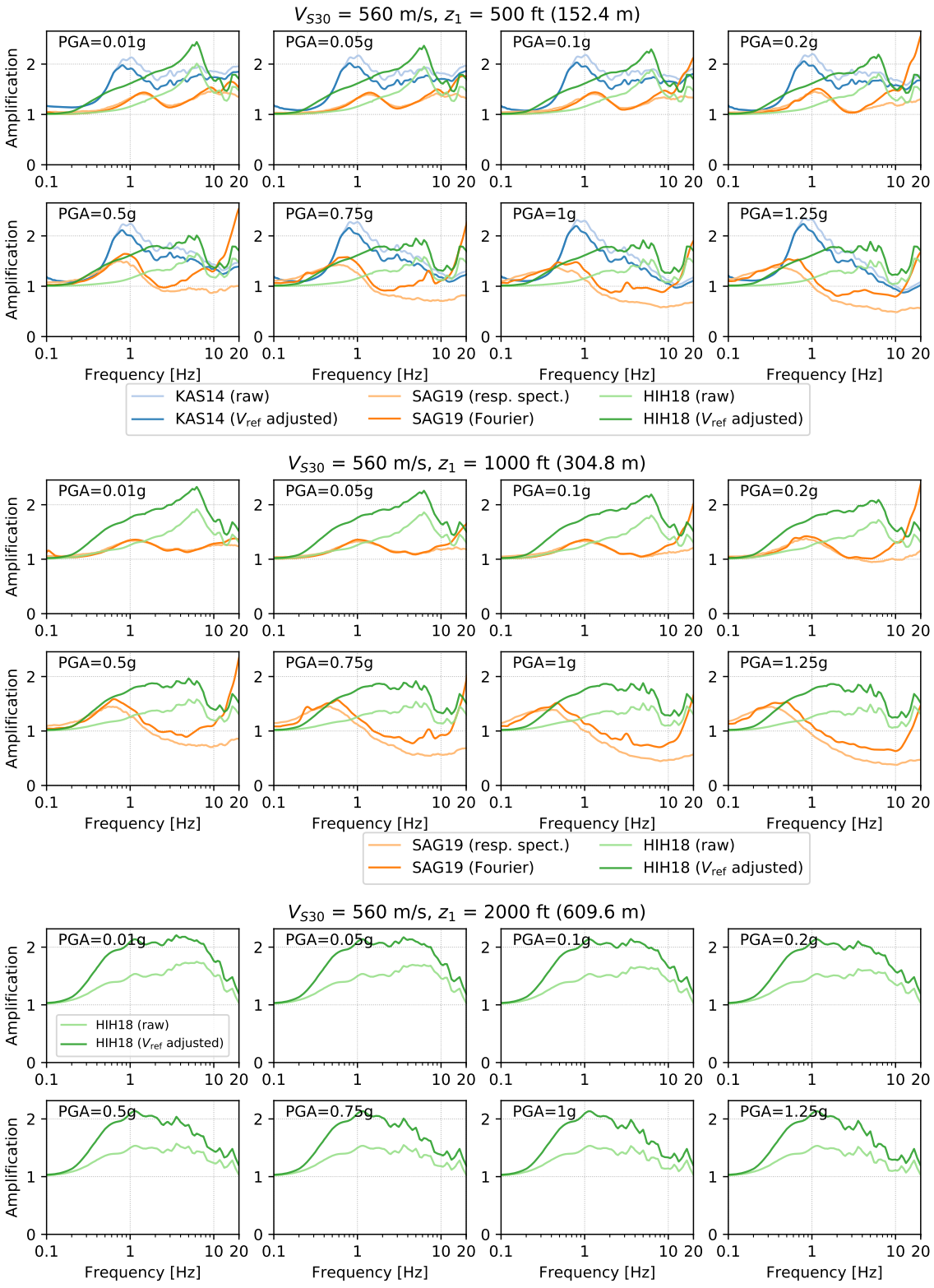


Figure 4.22: Comparison of site amplification factors between KAS14, HIH18, and SAG19, for three V_{S30}-z_1 combination: (560 m/s, 500 ft), (560 m/s, 1000 ft), and (560 m/s, 2000 ft). Each subplot corresponds to a PGA level.

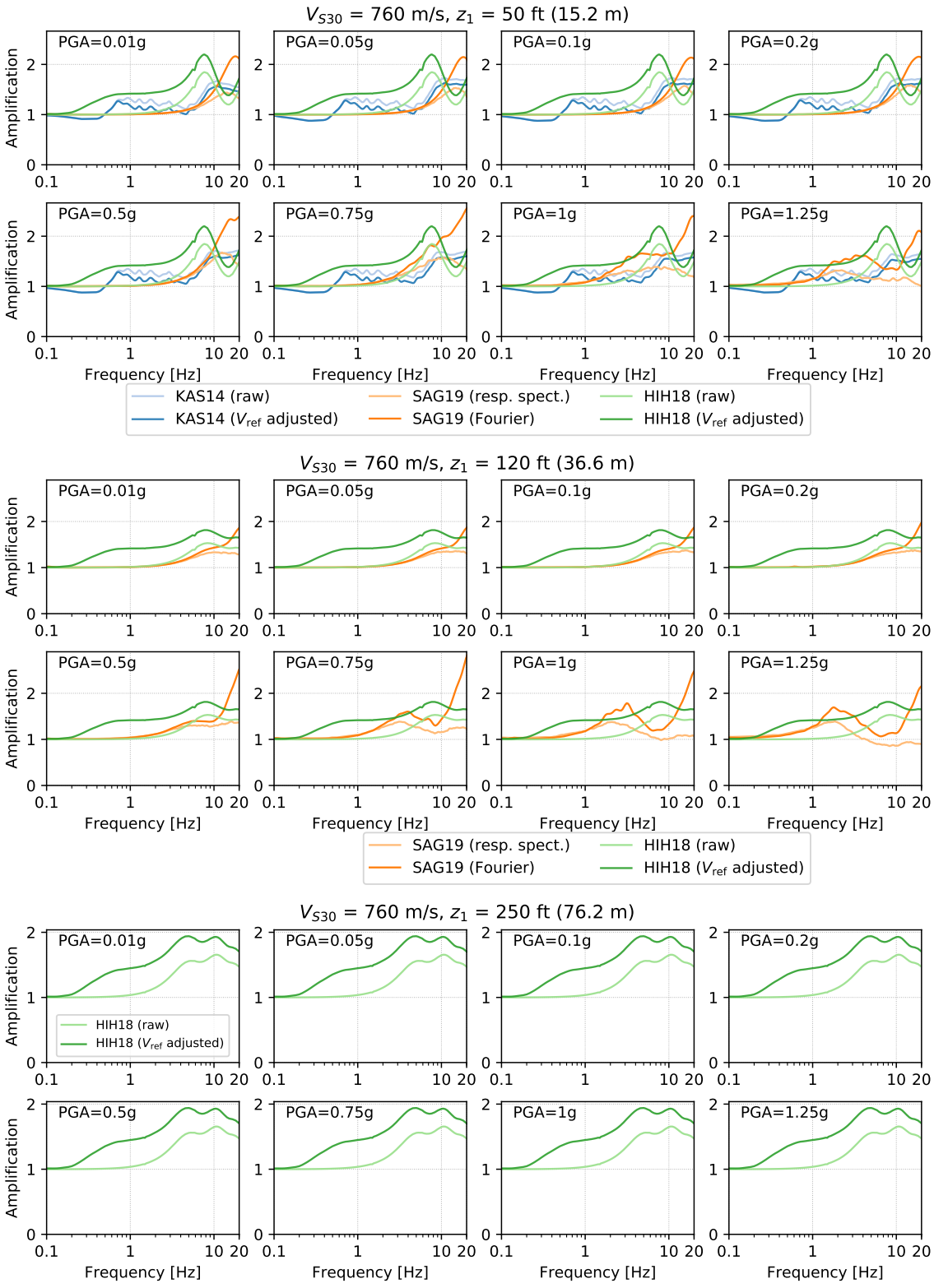


Figure 4.23: Comparison of site amplification factors between KAS14, HIH18, and SAG19, for three V_{S30}-z_1 combination: (760 m/s, 50 ft), (760 m/s, 120 ft), and (760 m/s, 250 ft). Each subplot corresponds to a PGA level.

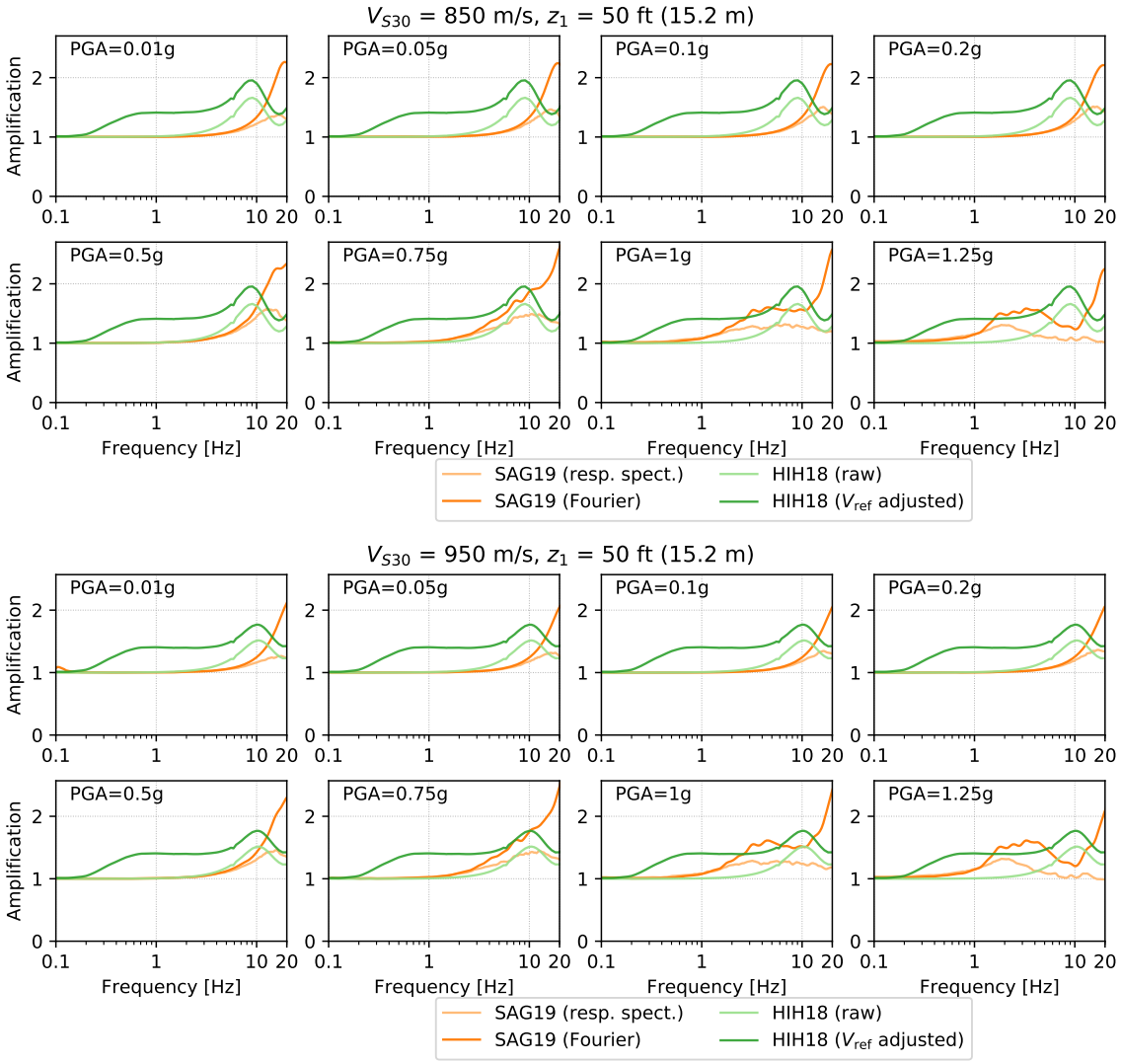


Figure 4.24: Comparison of site amplification factors between KAS14, HIH18, and SAG19, for two V_{S30}-z_1 combination: (850 m/s, 50 ft), (950 m/s, 50 ft). Each subplot corresponds to a PGA level.

4.5 Using SAG19 factors in ground motion simulations

This section presents a demonstration of using the SAG19 site factors to adjust simulated ground motions on rock outcrop. We run the Broadband Platform (BBP), version 17.3.0, developed by the Southern California Earthquake Center to simulate the 1994 M_w 6.73 Northridge earthquake. The BBP takes kinematic earthquake source parameters, and we use the following parameters for the Northridge event, which are the suggested values in the BBP and slightly different from what was used in *Graves and Pitarka* (2010):

- Moment magnitude (M_w): 6.73
- Fault length: 20 km; down-dip width: 24.9 km
- Strike: 122°; dip: 40°; rake: 105°
- Depth (from the ground surface to the top of the fault plane): 5.0 km
- Coordinates of the top-center point of the fault plane: 34.344°N, 118.515°W
- Hypocenter location: 6.0 km (along strike direction from the top-center point) and 19.4 km (along dip direction from the top-center point)

The BBP uses a hybrid approach to simulate ground motions: below 1 Hz, it uses a deterministic 3D wave-propagation approach, assuming that the crustal velocity structure is purely 1D (layered media), and above 1 Hz, it uses a semi-stochastic approach (explained in *Graves and Pitarka*, 2010). The sites for which we simulate the motions are the same as those in *Graves and Pitarka* (2010).

We use the V_{S30} value (from the NGA-West2 database) and z_1 value (queried from CVM-S4.26.M01) of each site to generate SVM profiles[14], and use the same procedure in Section 4.2.3 (on page 80) to generate stress-strain curves and damping curves for soils.

The current version of the BBP (17.3.0) has an option to turn on/off the site effect adjustments. If the users turn on the adjustments, it uses a combination of BSSA14 (*Boore et al.*, 2014) and CB14 (*Campbell and Bozorgnia*, 2014) site amplification factors (no phase) to add site response to the simulated rock motions. We run the BBP simulation for the same sites twice: the first time with BBP's site effect adjustments turned on[15], and the second time with its site effect adjustments turned off. We then de-convolve the output motions

[14] For this demonstration study, we use the base profile only. No randomization.

[15] We provide the same V_{S30} for each site (i.e., from the NGA-West2 database) to the BBP

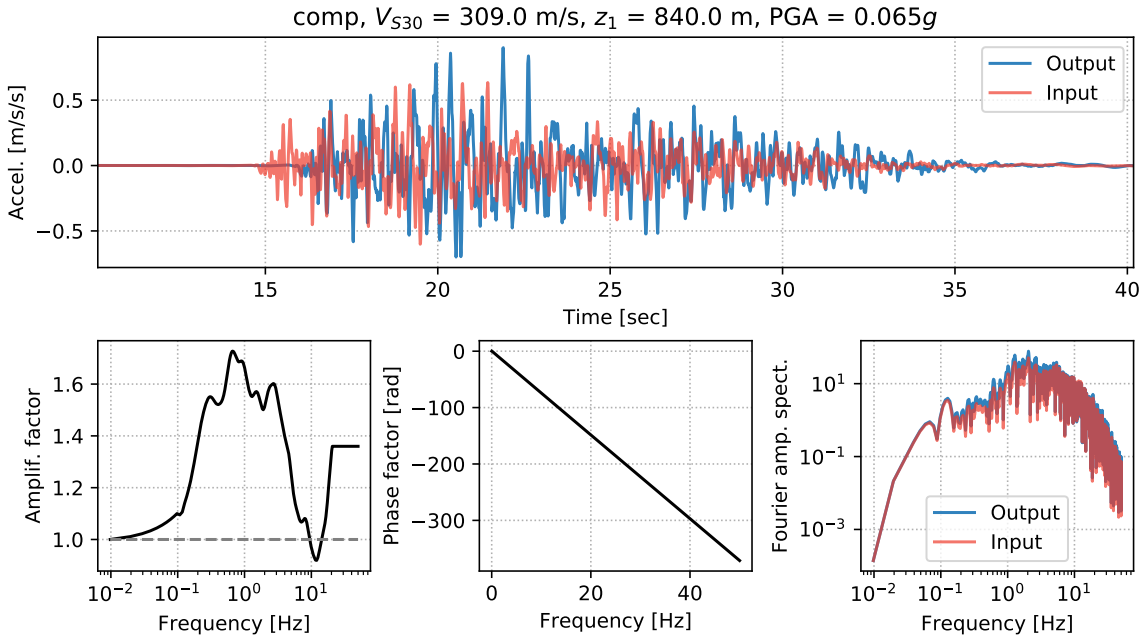


Figure 4.25: Demonstration of SAG19 adjustment to the rock-outcrop motion simulated by the BBP on site "comp" (in East Compton, CA). This is a deep site (z_1 = 840 m), so the time lag introduced by the phase factor is about 2 seconds.

from the second time, following the identical deconvolution procedure in Section 4.2.4 (on page 81), to obtain the true rock-outcrop motions.

We then adjust these BBP-generated rock-outcrop motions using these three methods:

1. The linear 1D wave-propagation method, using the base SVM profiles
2. The nonlinear 1D wave-propagation method (i.e., NL_{HH}) using the base SVM profiles and the stress-strain and damping curves mentioned above
3. Using SAG19 site factors (both amplitude and phase): V_{S30} and z_1 for each site are available (described above), and input PGAs are calculated from the rock-outcrop motions. Examples of two of the sites are shown in Figures 4.25 and 4.26 to demonstrate how amplification and phase affect the input motions.

After excluding a few rock sites ($V_{S30} > 1000$ m/s), we have 116 soil sites left.

We compare the adjusted ground motions with actual recordings, and calculate the goodness-of-fit scores (as defined in Section 3.3.3 on page 48). We show the overall goodness-of-fit for different frequency bands in Figure 4.27. Note that a positive goodness-of-fit means over-prediction, and a negative goodness-of-fit means under-prediction.

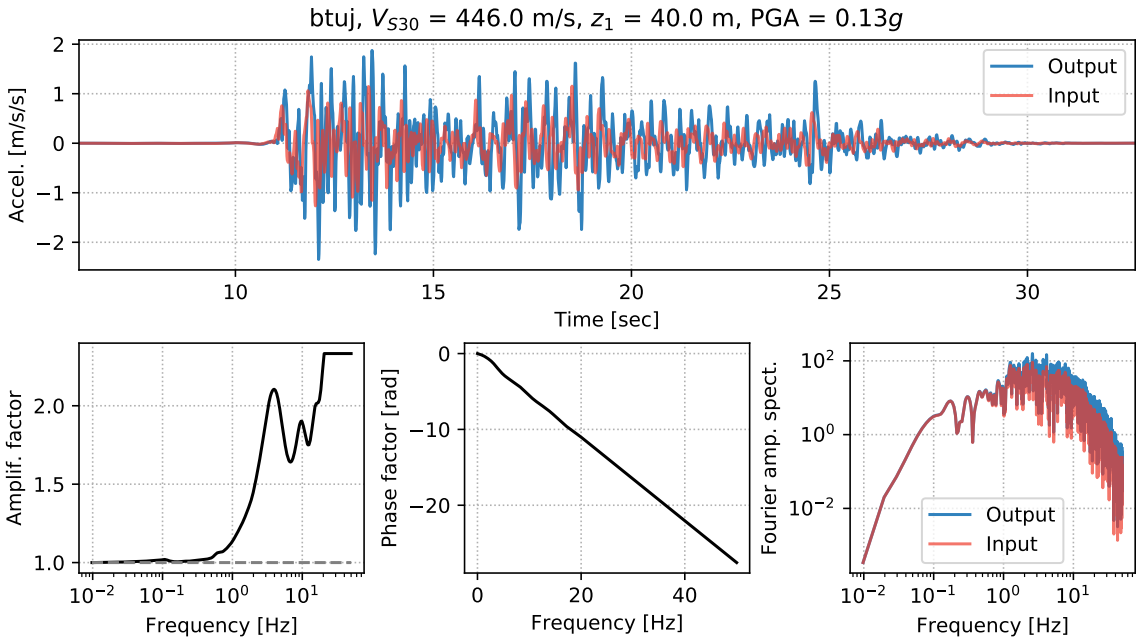


Figure 4.26: Demonstration of SAG19 adjustment to the rock-outcrop motion simulated by the BBP on site "btuj" (in Tujunga, CA). This is a shallow site ($z_1 = 40$ m), so the time lag introduced by the phase factor is very small.

From Figure 4.27, as expected, we see that the input motions always have the lowest overall intensity (compared to the recordings). Also, the NL_{HH} and SAG19 scores are always lower or on the same level as the linear scores. This is to be expected.

An unexpected observation is that for [0.2, 1] Hz, [1, 5] Hz, and [0.2, 20] Hz, the "BBP 17.3.0" scores are even higher than the linear scores. (The linear site response is almost always an overestimation of the true nonlinear site response, as shown in Figures 3.14 on page 66). Since the SVM model have been demonstrated to offer credible predictions (in Section 2.3 on page 20), the linear site response in this section are credibly represented. Therefore, we conclude that the BSSA14 and/or CB14 site factors are likely and overestimation. Unlike the validation study in Section 3.3, where the input ground motions are accurate (because they are recordings), the input motions in this validation exercise may be inaccurate. Therefore, even though the "BBP 17.3.0" scores are usually the closest to 0, it does not mean that the BSSA14 and/or CB14 site factors can correctly capture site response.

4.6 Conclusions

This chapter presents the SAG19 site factor model, which has both amplification and phase components. Compared to two other recent models (KAS14 by *Kamai et al.*, 2014 and

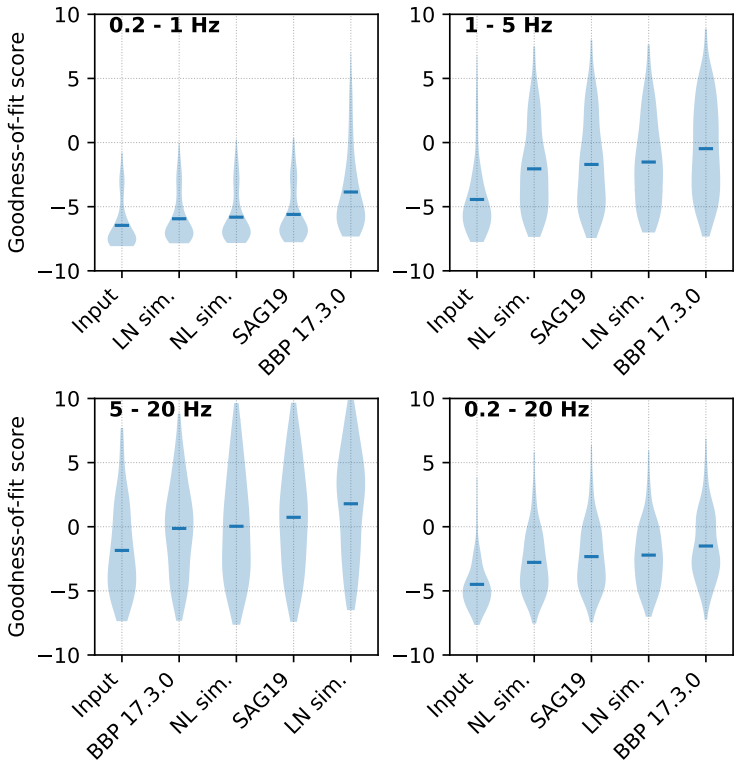


Figure 4.27: The overall goodness-of-fit scores for the Northridge event simulations, compared against actual recordings. The shapes of the shaded areas represent the distributions: the wider (horizontally) the shade at a certain goodness-of-fit level, the more sites with that goodness-of-fit score. And the horizontal bars represent the mean goodness-of-fit scores. "LN sim.": linear simulations. "NL sim.": nonlinear simulations (i.e., NL_{HH}). "SAG19": adjusted using SAG19 site factors. "Input": simulated rock-outcrop motions from the BBP. "BBP 17.3.0": simulated rock-outcrop motions from the BBP (with site effect adjustments turned on).

HIH18 by *Hashash et al.*, 2018), SAG19 as the following main advantages:

- It uses an improved V_S profile randomization scheme that captures natural variability of soil profiles well.
- It uses a nonlinear stress-strain model that has been validated to perform very well for predicting ground motions from weak to extremely strong.
- Its reference site condition is 1,046 m/s, enabling nonlinear site effects for stiffer soil sites (V_{S30} between 760 m/s and 1,046 m/s) to be modeled.
- It uses the time-domain nonlinear method to simulate ground motions, which provides better site effects prediction quality over the prevalent equivalent linear method.

- It provides both amplification factors and phase factors, the latter of which allows the timing of wave arrivals to be adequately modeled, hence can improve the seismic hazard analyses of spatially distributed infrastructure (bridges, pipelines, etc.).

The SAG19 site factors have the following potential uses:

- For improving the traditional ground motion prediction equations, and by extension, the seismic hazard models for California and the rest of WUS.
- For physics-based newer-generation seismic hazard models, such as CyberShake (*Graves et al.*, 2011), enabling the inclusion of nonlinear site effects.
- For deterministic 2D or 3D earthquake simulations, when the crustal material model used in the simulations lacks spatial resolution in shallow depths (a few hundred meters).
- For adding site response into hybrid (deterministic and stochastic) earthquake simulations, such as the Broadband Platform of the Southern California Earthquake Center

Chapter 5

Conclusions

Contents of this chapter

5.1 Summary of previous chapters

This thesis presents the author's research in the improvement of earthquake site response analysis. Improving site response analysis helps advance the state of the art in modeling earthquake ground motions, which can contribute to both civil engineering design and seismic hazard quantification.

Chapter 2 presents a model called the sediment velocity model (SVM). This model is calibrated with measured data, and is capable of estimating 1D shear-wave velocity (V_S) profiles of a location in California given only the V_{S30} information. This model bridges the gap between the need for detailed descriptions of near-surface linear properties of soils (i.e., V_S profiles) and the fact that V_{S30} is the only accessible information for the majority of California. We show that the predictive capability of the SVM is better than the "Geotechnical Layer model" by *Ely et al.* (2010), and also better than CVM-H and CVM-S (two 3D velocity models in the Community Velocity Models of the Southern California Earthquake Center). Apart from the model itself, we also present some empirical correlations that characterizes the stochastic nature in the data, which can be used in generating randomized 1D V_S profiles. The SVM has two limitations due to the limited coverage of available data: (1) the SVM can only estimate V_S profiles in California or regions with similar geological formation to California, but the methodology presented in the chapter is transferable to other regions; (2) estimating V_S values up to 1,000 m/s requires some extrapolation for softer sites (shown in Figure 2.4), but this can be improved by including deeper V_S measurements in the future.

Chapter 3 presents a model called the hybrid hyperbolic (HH) model. It is capable of incorporating shear strengths of soils, a feature lacking in the currently popular Modified Kondner-Zelasko (MKZ) model. Incorporating shear strengths enables us to better model nonlinear behaviors of soils during cyclic loading, thus improving the predictive capability

of nonlinear site response simulations. The HH model can also "construct" nonlinear stress-strain curves from only V_S information, making it a practical tool for regional-scale site response analysis. Compared with other similar strength-incorporated models to date, the HH model is the only one that has demonstrated the ability to satisfactorily fit or predict: (1) larger-strain ($> 1\%$) laboratory shearing test data, and (2) nonlinear site response of extreme ground motions. The HH model is only applicable to 1D soil shearing behavior, so it does not capture 2D/3D effects such as heterogeneity and dilatancy.

Chapter 4 presents a project that exemplifies the potentials of the models in Chapters 2 and 3: the development of nonlinear site factors for the Western United States. Site factors are a useful product in civil engineering design, in the development of regional seismic hazard maps, and in regional-scale ground motion simulations. With the improvements brought by the SVM and the HH model, the site factors developed in this chapter have the potential to better reflect the actual nonlinear site response. Additionally, the site factors presented here is the first product that provides phase shift, which enables the adjustment of wave arrival time, and can improve the hazard estimation for spatially distributed systems such as bridges and pipelines.

5.2 Future work

The models and results presented in this thesis open up some new avenues of related research.

For example, the difference in the predictive performance of the SVM and CVM-S4.26.M01 (in Chapter 2) prompts us to reexamine the way CVM-S currently integrates near-surface V_S measurements. Ideally, CVM-S (or other 3D crustal velocity models) can combine measured V_S profiles and the "background profiles" (predicted by the SVM) in a "natural" way, in which locations closer to the measurements reflect more "influences" of the measurements, and locations far away from any measurements can "fall back" to the SVM-generated profiles. Doing so in a "non *ad hoc*" manner requires additional data for calibration, such as obtaining V_S profiles densely (every few hundred meters) for a whole city.

Another example is the extension of the HH model to 2D and 3D. The so-called "one-dimensional-two-component" or "one-dimensional-three-component" simulations directly use 1D stress-strain models and extend them into 2D and 3D. The effects of omitting 2D/3D soil behaviors (such as dilatancy) on ground motion modeling accuracy need to be quantitatively studied.

The third potential avenue is using the nonlinear site factors to improve physics-based seis-

mic hazard products such as CyberShake (developed by *Graves et al.*, 2011). CyberShake cannot model nonlinear site response directly, because it is based on the principle of reciprocity that inherently assumes linear ground responses. Therefore, the site factors can be used as an "add-on" to adjust the results of CyberShake, improving its ability to better model seismic hazards.

Bibliography

Aagaard, B. T., T. M. Brocher, D. Dolenc, D. Dreger, R. W. Graves, S. Harmsen, S. Hartzell, S. Larsen, and M. L. Zoback (2008), Ground-motion modeling of the 1906 San Francisco earthquake, Part I: Validation using the 1989 Loma Prieta earthquake, *Bulletin of the Seismological Society of America*, *98*(2), 989–1011.

Abrahamson, N., and W. J. Silva (1997), Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seismological Research Letters*, *68*(1), 94–127.

Abrahamson, N. A., W. J. Silva, and R. Kamai (2014), Summary of the ASK14 ground motion relation for active crustal regions, *Earthquake Spectra*, *30*(3), 1025–1055.

Ancheta, T. D., R. B. Darragh, J. P. Stewart, E. Seyhan, W. J. Silva, B. S.-J. Chiou, K. E. Wooddell, R. W. Graves, A. R. Kottke, D. M. Boore, et al. (2014), NGA-West2 database, *Earthquake Spectra*, *30*(3), 989–1005.

Anderson, J. G. (2004), Quantitative measure of the goodness-of-fit of synthetic seismograms, in *Proc. 13th World Conf. on Earthquake Eng.*, Vancouver, British Columbia, Canada, Int. Assoc. Earthquake Eng.

Aoi, S., T. Kunugi, and H. Fujiwara (2004), Strong-motion seiemograph network operated by NIED: K-NET and KiK-net, *Journal of Japan Association for Earthquake Engineering*, *4*(3), 65–74.

Archuleta, R. J., and P. Liu (2004), Improved prediction method for time histories of near-field ground motions with application to southern CALIFORNIA, *Final report*, U.S. Geological Survey.

Asimaki, D., and J. Shi (2017), *SeismoSoil user manual, v1.3*, California Institute of Technology.

Assimaki, D., and W. Li (2012), Site- and ground motion-dependent nonlinear effects in seismological model predictions, *Soil Dynamics and Earthquake Engineering*, *32*, 143–151.

Assimaki, D., and J. Steidl (2007), Inverse analysis of weak and strong motion downhole array data from the Mw 7.0 Sanriku-Minami earthquake, *Soil Dynamics and Earthquake Engineering*, *27*(1), 73–92.

Assimaki, D., J. Steidl, and P.-C. Liu (2006), Attenuation and velocity structure for site response analyses via downhole seismogram inversion, *Pure and Applied Geophysics*, *163*, 81–118.

Assimaki, D., W. Li, J. M. Steidl, and J. Schmedes (2008), Quantifying nonlinearity susceptibility via site response modeling uncertainty at three sites in the Los Angeles basin, *Bulletin of the Seismological Society of America*, *98*(5), 2364–2390.

Beresnev, I., and K.-L. Wen (1996), Nonlinear soil response—a reality?, *Bulletin of the Seismological Society of America*, *86*(6), 1964–1978.

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, *Open-File Report 2003-191*, U.S. Geological Survey.

Boore, D. M., and M. W. Asten (2008), Comparisons of shear-wave slowness in the Santa Clara Valley, California, using blind interpretations of data from invasive and noninvasive methods, *Bulletin of the Seismological Society of America*, *98*(4), 1983–2003.

Boore, D. M., and L. T. Brown (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*(3), 222–229.

Boore, D. M., and E. M. Thompson (2007), On using surface-source downhole-receiver logging to determine seismic slownesses, *Soil Dynamics and Earthquake Engineering*, *27*(11), 971–985.

Boore, D. M., W. B. Joyner, and T. E. Fumal (1993), Estimation of response spectra and peak accelerations from WESTERN NORTH AMERICAN earthquakes: An interim report, *Open-file report 93-509*, US Geological Survey.

Boore, D. M., W. B. Joyner, and T. E. Fumal (1997), Equations for estimating horizontal response spectra and peak acceleration from Western North American earthquakes: A summary of recent work, *Seismological Research Letters*, *68*(1), 128–153.

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

Bora, S. S., F. Scherbaum, N. Kuehn, and P. Stafford (2016), On the relationship between Fourier and response spectra: Implications for the adjustment of empirical ground motion prediction equations (GMPEs), *Bulletin of the Seismological Society of America*, *106*(3), 1235–1253, doi:https://doi.org/10.1785/0120150129.

Brown, L. T., D. M. Boore, and K. H. Stokoe (2002), Comparison of shear-wave slowness profiles at 10 strong-motion sites from noninvasive SASW measurements and measurements made in boreholes, *Bulletin of the Seismological Society of America*, *92*(8), 3116–3133.

Budhu, M. (2011), *Soil Mechanics and Foundations, Third Edition*, John Wiley and Sons, Inc, New York, New York.

Building Seismic Safety Council (2015), NEHRP recommended seismic provisions for new buildings and other structures (FEMA P-1050), 2015 edition, *FEMA report P-1050*.

Campbell, K. W. (2003), Prediction of strong ground motion using the hybrid empirical method and its use in the development of ground-motion (attenuation) relations in Eastern North America, *Bulletin of the Seismological Society of America*, *93*(3), 1012–1033.

Campbell, K. W. (2009), Estimates of shear-wave Q and κ_0 for unconsolidated and semi-consolidated sediments in Eastern North America, *Bulletin of the Seismological Society of America*, *99*(4), 2365–2392.

Campbell, K. W., and Y. Bozorgnia (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.

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

Chopra, A. K. (2001), *Dynamics of Structures*, Prentice-Hall.

Darendeli, M. B. (2001), Development of a new family of normalized modulus reduction and material damping curves, Ph.D. thesis, University of Texas at Austin.

de Coulomb, C.-A. (1776), Essai sur une application des règles des maximis et minimis à quelques problèmes de statique relatifs à l'architecture,, *Mémoires de mathématique et de physique*, *7*, 343–387.

Drucker, D. C., and W. Prager (1952), Soil mechanics and plastic analysis or limit design, pp. 157–165.

Electric Power Research Institute (EPRI) (1993), Guidelines for determining design basis ground motions, *Method and guidelines for estimating earthquake ground motion in eastern North America*, *1*.

Ely, G. P., P. Small, T. H. Jordan, P. J. Maechling, and F. Wang (2010), A Vs30-derived near-surface seismic velocity model, in *poster at American Geophysical Union Annual Conference*, (URL: http://earth.usc.edu/ gely/vs30gtl, last accessd: March 2016).

EPRI (1993), Guidelines for determining design basis ground motions, *Report No. TR-102293*, Electric Power Research Institute, Palo Alto, California.

Field, E. H., P. A. Johnson, I. A. Beresnev, and Y. Zeng (1997), Nonlinear ground-motion amplification by sediments during the 1994 Northridge earthquake, *Nature*, *390*, 599–602.

Frankel, A. D., C. Mueller, T. Barnhard, D. Perkins, E. Leyendecker, N. Dickman, S. Hanson, and M. Hopper (1996), National seismic-hazard maps — documentation June 1996, *Open-file report 96-532*, US Geological Survey, 110 pp.

Frankel, A. D., M. D. Petersen, C. S. Mueller, K. M. Haller, R. L. Wheeler, E. Leyendecker, R. L. Wesson, S. C. Harmsen, C. H. Cramer, D. M. Perkins, et al. (2002), Documentation for the 2002 update of the national seismic hazard maps, *Open-file report 02-420*, 33.

Fu, H., C. He, B. Chen, Z. Yin, Z. Zhang, W. Zhang, T. Zhang, W. Xue, W. Liu, W. Yin, G. Yang, and X. Chen (2017), 18.9-Pflops nonlinear earthquake simulation on Sunway TaihuLight: enabling depiction of 18-Hz and 8-meter scenarios, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, Denver, Colorado, November 12-17, 2017*, p. 2, Association for Computing Machinery.

Gingery, J. R., and A. Elgamal (2013), Shear stress-strain curves based on the G/Gmax logic: A procedure for strength compatibility, in *IACGE 2013: Challenges and Recent Advances in Geotechnical and Seismic Research and Practices*, pp. 721–729.

Graves, R., T. H. Jordan, S. Callaghan, E. Deelman, E. Field, G. Juve, C. Kesselman, P. Maechling, G. Mehta, K. Milner, D. Okaya, P. Small, and K. Vahi (2011), CyberShake: A physics-based seismic hazard model for Southern California, *Pure and Applied Geophysics*, *168*(3), 367–381, doi:10.1007/s00024-010-0161-6.

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

Groholski, D. R., Y. M. Hashash, B. Kim, M. Musgrove, J. Harmon, and J. P. Stewart (2016), Simplified model for small-strain nonlinearity and strength in 1D seismic site response analysis, *Journal of Geotechnical and Geoenvironmental Engineering*, *142*(9), 04016,042.

Gutenberg, B. (1957), Effects of ground on earthquake motion, *Bulletin of the Seismological Society of America*, *47*(3), 221–250.

Hardin, B. O., and V. P. Drnevich (1972a), Shear modulus and damping in soils: measurement and parameter effects, *Journal of the Soil Mechanics and Foundations Division*, *98*(6), 603–624.

Hardin, B. O., and V. P. Drnevich (1972b), Shear modulus and damping in soils: design equations and curves, *Journal of the Soil Mechanics and Foundations Division*, *98*(7), 667–692.

Hartzell, S., L. Ramírez-Guzmán, M. Meremonte, and A. Leeds (2016), Ground motion in the presence of complex topography II: Earthquake sources and 3d simulations, *Bulletin of the Seismological Society of America*, *107*(1), 344–358.

Hartzell, S. H. (1998), Variability in nonlinear sediment response during the 1994 Northridge, California, Earthquake, *Bulletin of the Seismological Society of America*, *88*(6), 1426–1437.

Hartzell, S. H., L. F. Bonilla, and R. A. Williams (2004), Prediction of nonlinear soil effects, *Bulletin of the Seismological Society of America*, *94*(5), 1609–1629.

Hashash, Y. M. A., C. Phillips, and D. R. Groholski (2010), Recent advances in non-linear site response analysis, in *Fifth International Conference on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics*.

Hashash, Y. M. A., O. Ilhan, and J. A. Harmon (2018), Fourier amplitude spectrum site amplification modelling in Western United States, *Tech. rep.*, Southern California Earthquake Center.

Haskell, N. A. (1953), The dispersion of surface waves on multilayered media, *Bulletin of the seismological Society of America*, *43*(1), 17–34.

Hayashi, H., M. Honda, and T. Yamada (1994), Modeling of nonlinear stress strain relations of sands for dynamic response analysis, in *Earthquake Engineering, Tenth World Conference, Balkema, Rotterdam*.

Idriss, I. M. (2014), An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes, *Earthquake Spectra*, *30*(3), 1155–1177.

Jones, L. M., R. Bernknopf, D. Cox, J. Goltz, K. Hudnut, D. Mileti, S. Perry, D. Ponti, K. Porter, M. Reichle, et al. (2008), The ShakeOut scenario, *US Geological Survey Open-File Report*, *1150*, 308.

Kaklamanos, J., and B. A. Bradley (2016), Improving our understanding of 1D site response model behaviour: Physical insights for statistical deviations from 114 Kik-Net sites.

Kaklamanos, J., B. A. Bradley, E. M. Thompson, and L. G. Baise (2013), Critical parameters affecting bias and variability in site-response analyses using KiK-net downhole array data, *Bulletin of the Seismological Society of America*, *103*(3), 1733–1749.

Kaklamanos, J., L. G. Baise, E. M. Thompson, and L. Dorfmann (2015), Comparison of 1D linear, equivalent-linear, and nonlinear site response models at six KiK-net validation sites, *Soil Dynamics and Earthquake Engineering*, *69*, 207–219.

Kamai, R., N. A. Abrahamson, and W. J. Silva (2013), Nonlinear horizontal site response for the NGA-West2 project, *Tech. rep.*, Pacific Earthquake Engineering Research Center.

Kamai, R., N. A. Abrahamson, and W. J. Silva (2014), Nonlinear horizontal site amplification for constraining the NGA-West2 GMPEs, *Earthquake Spectra*, *30*(3), 1223–1240.

Kanai, K., R. Takahashi, and H. Kawasumi (1956), Seismic characteristics of ground, in *Proc. World Conference on Engineering, Earthquake Engineering Research Institute*, 31, pp. 1–16.

Kohler, M., H. Magistrale, and R. Clayton (2003), Mantle heterogeneities and the SCEC reference three-dimensional seismic velocity model version 3, *Bulletin of the Seismological Society of America*, *93*(2), 757–774.

Kondner, R. L., and J. S. Zelasko (1963), A hyperbolic stress-strain formulation of sands, in *Proc. 2nd Pan Am. Conf. on Soil Mech. and Found. Engrg., Brazilian Association of Soil Mechanics, Silo Paulo, Brazil*, pp. 289–324.

Konno, K., and T. Ohmachi (1998), Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor, *Bulletin of the Seismological Society of America*, *88*(1), 228–241.

Koutsoftas, D. C. (1978), Effect of cyclic loads on undrained strength of two marine clays, *Journal of the Geotechnical Engineering Division*, *104*(GT5), 609–620.

Kramer, S. L. (1996), *Geotechnical Earthquake Engineering*, Prentice Hall, Upper Saddle River, New Jersey.

Kristeková, M., J. Kristek, and P. Moczo (2009), Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals, *Geophysical Journal International*, *178*(2), 813–825.

Ladd, C. C. (1991), Stability evaluation during staged construction: the 22nd Terzaghi Lecture, *Journal of Geotechnical Engineering*, *117*(4), 540–615.

Ladd, C. C., and L. Edgers (1972), Consolidated-undrained direct-simple shear tests on saturated clays, *Research Report R72-82, Soils Publication 284*, Department of Civil Engineering, MIT, Cambridge, MA.

Lee, E.-J., P. Chen, T. H. Jordan, P. B. Maechling, M. A. Denolle, and G. C. Beroza (2014), Full-3-D tomography for crustal structure in Southern California based on the scattering-integral and the adjoint-wavefield methods, *Journal of Geophysical Research: Solid Earth*, *119*(8), 6421–6451.

Li, W., and D. Assimaki (2010), Site- and motion-dependent parametric uncertainty of site-response analyses in earthquake simulations, *Bulletin of the Seismological Society of America*, *100*(3), 954–968.

Liu, P., R. J. Archuleta, and S. H. Hartzell (2006), Prediction of broadband ground-motion time histories: Hybrid low/high-frequency method with correlated random source parameters, *Bulletin of the Seismological Society of America*, *96*(6), 2118–2130.

Liu, P.-C., and R. J. Archuleta (2006), Efficient modeling of Q for 3D numerical simulation of wave propagation, *Bulletin of the Seismological Society of America*, *96*(4A), 1352–1358.

Lozos, J. C., D. D. Oglesby, J. N. Brune, and K. B. Olsen (2015), Rupture and ground-motion models on the northern San Jacinto fault, incorporating realistic complexity, *Bulletin of the Seismological Society of America*, *105*(4), 1931–1946.

Maechling, P. J., F. Silva, S. Callaghan, and T. H. Jordan (2014), SCEC Broadband Platform: System architecture and software implementation, *Seismological Research Letters*, *86*(1), 27–38.

Magistrale, H., K. McLaughlin, and S. Day (1996), A geology-based 3D velocity model of the Los Angeles basin sediments, *Bulletin of the Seismological Society of America*, *86*(4), 1161–1166.

Magistrale, H., S. Day, R. W. Clayton, and R. Graves (2000), The SCEC Southern California reference three-dimensional seismic velocity model version 2, *Bulletin of the Seismological Society of America*, *90*(6B), S65–S76.

Mai, P. M., W. Imperatori, and K. B. Olsen (2010), Hybrid broadband ground-motion simulations: Combining long-period deterministic synthetics with high-frequency multiple S-to-S backscattering, *Bulletin of the Seismological Society of America*, *100*(5A), 2124–2142.

Masing, G. V. (1926), Eigenspannungen und Verfestigung beim Messing, in *Proceedings of the 2nd international congress of applied mechanics*, vol. 100, pp. 332–335.

Matasovic, J., and M. Vucetic (1993), Cyclic characterization of liquefiable sands, *Journal of Geotechnical Engineering*, *119*(11), 1805–1822.

Mayne, P., and F. Kulhawy (1982), K_0-OCR relationships in soil, *Journal of the Geotechnical Engineering Division*, *108*(GT6), 851–872.

Mayne, P., P. Robertson, and T. Lunne (1998), Clay stress history evaluated from seismic piezocone tests, in *Geotechnical Site Characterization, ISC '98*, edited by P. K. Robertson and P. W. Mayne, pp. 1113–1118, Balkema, Rotterdam, ISBN 90 5410 939 4.

Mayne, P. W., J. A. Schneider, and G. K. Martin (1999), Small- and large-strain soil properties from seismic flat dilatometer tests, in *Pre-failure Deformation Characteristics of Geomaterials*, edited by L. . L. P. Jamiolkowski, pp. 419–426, Balkema, Rotterdam, ISBN 90 5809 075 2.

McCarron, W., J. Lawrence, R. Werner, J. Germaine, and D. Cauble (1995), Cyclic direct simple shear testing of a Beaufort Sea clay, *Canadian Geotechnical Journal*, *32*(4), 584–600.

Milne, J. (1908), *Seismology*, 2nd edition ed., London.

Muravskii, G. (2005), On description of hysteretic behavior of materials, *International Journal of Solids and Structures*, *42*, 2625–2644.

National Oceanic and Atmospheric Administration (), "Sumatra, Indonesia Earthquake and Tsunami, 26 December 2004", https://www.ngdc.noaa.gov/hazard/26dec2004.html. (Last accessed: March 16, 2019).

Neumann, F. (1954), *Earthquake intensity and related ground motion*, University of Washington Press.

Olsen, K., and R. Takedatsu (2015), The SDSU broadband ground-motion generation module BBtoolbox version 1.5, *Seismological Research Letters*, *86*(1), 81–88.

Olsen, K. B., and J. E. Mayhew (2010), Goodness-of-fit criteria for broadband synthetic seismograms, with application to the 2008 Mw 5.4 Chino Hills, California, earthquake, *Seismological Research Letters*, *81*(5), 715–723.

Petersen, M. D., A. D. Frankel, S. C. Harmsen, C. S. Mueller, K. M. Haller, R. L. Wheeler, R. L. Wesson, Y. Zeng, O. S. Boyd, D. M. Perkins, et al. (2008), Documentation for the 2008 update of the United States National Seismic Hazard Maps, *Open-file report 2008-1128*, US Geological Survey, 128 pp.

Petersen, M. D., M. P. Moschetti, P. M. Powers, C. S. Mueller, K. M. Haller, A. D. Frankel, Y. Zeng, S. Rezaeian, S. C. Harmsen, O. S. Boyd, N. Field, R. Chen, K. S. Rukstales, N. Luco, R. L. Wheeler, R. A. Williams, , and A. H. Olsen (2014), Documentation for the 2014 update of the United States National Seismic Hazard Maps, *Open-file report 2014-1091*, US Geological Survey, 255 pp.

Phillips, C., and Y. M. Hashash (2009), Damping formulation for nonlinear 1D site response analyses, *Soil Dynamics and Earthquake Engineering*, *29*(7), 1143–1158.

Pyke, R. M. (1979), Nonlinear soil model for irregular cyclic loading, *Journal of the Geotechnical Engineering Division*, *105*(6), 715–726.

Ramberg, W., and W. R. Osgood (1943), Description of stress-strain curves by three parameters, *Tech. Rep. Technical Note No. 902*, National Advisory Committee For Aeronautics, Washington DC.

Ramirez-Guzman, L., R. W. Graves, K. B. Olsen, O. S. Boyd, C. Cramer, S. Hartzell, S. Ni, P. Somerville, R. A. Williams, and J. Zhong (2015), Ground-motion simulations of 1811–1812 New Madrid earthquakes, central United States, *Bulletin of the Seismological Society of America*, *105*(4), 1961–1988.

Reid, H. F. (1910), The California earthquake of April 18, 1906, *The Mechanics of the Earthquake, Carnegie Institute of Washington*, *2*(87).

Richardson, A. M., and R. V. Whitman (1963), Effect of strain-rate upon undrained shear resistance of a saturated remoulded fat clay, *Geotechnique*, *13*(4), 310–324.

Rix, G. J., G. L. Hebeler, and M. C. Orozco (2002), Near-surface V_S profiling in the New Madrid Seismic Zone using surface-wave methods, *Seismological Research Letters*, *73*(3), 380–392.

SCECpedia (2016), "Community Velocity Models", https://scec.usc.edu/scecpedia/Community_Velocity_Models, last accessed: 2017-10-20.

SCECpedia (2016), "CVM-S4.26.M01", https://scec.usc.edu/scecpedia/CVM-S4.26.M01. (Last accessed: March 19, 2019).

Seed, H. B., and I. M. Idriss (1970), Soil moduli and damping factors for dynamic response analyses, *UCB/EERC- 70/10*, Earthquake Engineering Research Center, University of California at Berkeley.

Seidalinov, G., and M. Taiebat (2014), Bounding surface SANICLAY plasticity model for cyclic clay behavior, *International Journal for Numerical and Analytical Methods in Geomechanics*, *38*, 702–724.

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

Shi, J., and D. Asimaki (2017), From stiffness to strength: Formulation and validation of a hybrid hyperbolic nonlinear soil model for site-response analyses, *Bulletin of the Seismological Society of America*, *107*(3), 1336–1355.

Shi, J., and D. Asimaki (2018), A generic velocity profile for basin sediments in California conditioned on Vs30, *Seismological Research Letters*, *89*(4), 1397–1409.

Silva, W., and K. Lee (1987), WES RASCAL code for synthesizing earthquake ground motions, state-of-the-art for assessing earthquake hazards in the United States, *Report 24*, Pacific Engineering and Analysis, El Cerrito, CA.

Stephenson, W. J., J. N. Louie, S. Pullammanappallil, R. A. Williams, and J. K. Odum (2005), Blind shear-wave velocity comparison of ReMi and MASW results with boreholes to 200 m in Santa Clara Valley: Implications for earthquake ground-motion assessment, *Bulletin of the Seismological Society of America*, *95*(6), 2506–2516.

Stewart, J. P., and A. O. Kwok (2008), Nonlinear seismic ground response analysis: Code usage protocols and verification against vertical array data, in *Proceedings of the 4th Decennial Geotechnical Earthquake Engineering and Soil Dynamics Conference, ASCE, Sacramento, CA*.

Sun, X., S. Hartzell, and S. Rezaeian (2015), Ground-motion simulation for the 23 august 2011, mineral, virginia, earthquake using physics-based and stochastic broadband methods, *Bulletin of the Seismological Society of America*.

Süss, M. P., and J. H. Shaw (2003), P wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California, *Journal of Geophysical Research: Solid Earth*, *108*(B3).

Taborda, R., and J. Bielak (2013), Ground-motion simulation and validation of the 2008 Chino Hills, California, earthquake, *Bulletin of the Seismological Society of America*, *103*(1), 131–156.

Taborda, R., S. Azizzadeh-Roodpish, N. Khoshnevis, and K. Cheng (2016), Evaluation of the Southern California seismic velocity models through simulation of recorded events, *Geophysical Journal International*, *205*(3), 1342–1364.

Thompson, E., D. J. Wald, and C. Worden (2014), A vs30 map for california with geologic and topographic constraints, *Bulletin of the Seismological Society of America*, *104*(5), 2313–2321.

Thompson, E. M., L. G. Baise, R. E. Kayen, and B. B. Guzina (2009), Impediments to predicting site response: Seismic property estimation and modeling simplifications, *Bulletin of the Seismological Society of America*, *99*(5), 2927–2949.

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

Thomson, W. T. (1950), Transmission of elastic waves through a stratified solid medium, *Journal of applied Physics*, *21*(2), 89–93.

Toro, G. (1995), Probabilistic models of site velocity profiles for generic and site-specific ground-motion amplification studies, *Tech. rep.*, Brookhaven National Laboratory, Upton, NY.

U.S. Geological Survey (), Can you predict earthquakes?, https://www.usgs.gov/faqs/can-you-predict-earthquakes (Last accessed: March 12, 2019).

Vardanega, P. J., and M. D. Bolton (2011), Strength mobilization inclays and silts, *Canadian Geotechnical Journal*, *48*, 1485–1503.

Vardanega, P. J., and M. D. Bolton (2013), Stiffness of clays and silts: normalizing shear modulus and shear strain, *Journal of Geotechnical and Geoenvironmental Engineering*, *139*(9), 1575–1589.

Vrettos, C. (1996), Simple inversion procedure for shallow seismic refraction in continuously nonhomogeneous soils, *Soil Dynamics and Earthquake Engineering*, *15*, 381–386.

Vucetic, M., and R. Dobry (1991), Effect of soil plasticity on cyclic response, *Journal of Geotechnical Engineering*, *117*(1), 89–107.

Walling, M., W. Silva, and N. Abrahamson (2008), Nonlinear site amplification factors for constraining the NGA models, *Earthquake Spectra*, *24*(1), 243–255.

Wills, C., C. Gutierrez, F. Perez, and D. Branum (2015), A next generation Vs30 map for California based on geology and topography, *Bulletin of the Seismological Society of America*.

Wills, C. J., and K. B. Clahan (2006), Developing a map of geologically defined site-condition categories for California, *Bulletin of the Seismological Society of America*, *96*(4A), 1483–1501.

Wu, C., and Z. Peng (2011), Temporal changes of site response during the 2011 Mw 9.0 off the Pacific coast of Tohoku earthquake, *Earth, planets and space*, *63*(7), 51.

Wu, C., Z. Peng, and D. Assimaki (2009), Temporal changes in site response associated with the strong ground motion of the 2004 Mw 6.6 Mid-Niigata earthquake sequences in Japan, *Bulletin of the Seismological Society of America*, *99*(6), 3487–3495.

Yamamoto, Y., and J. W. Baker (2013), Stochastic model for earthquake ground motion using wavelet packets, *Bulletin of the Seismological Society of America*, *103*(6), 3044–3056.

Yee, E., J. P. Stewart, and K. Tokimatsu (2013), Elastic and large-strain nonlinear seismic site response from analysis of vertical array recordings, *Journal of the Geotechnical and Geoenvironmental Engineering*, *139*(10), 1789–1801.

Yong, A., S. E. Hough, J. Iwahashi, and A. Braverman (2012), A terrain-based site-conditions map of California with implications for the contiguous United States, *Bulletin of the Seismological Society of America*, *102*(1), 114–128.

Yong, A., A. Martin, K. Stokoe, and J. Diehl (2013), ARRA-funded Vs30 measurements using multi-technique approach at strong-motion stations in California and Central-Eastern United States, *Open-file report 2013-1102*, U.S. Geological Survey.

Zalachoris, G. (2014), Evaluation of one-dimensional site response methodologies using borehole arrays, Ph.D. thesis, University of Texas, Austin.

Zalachoris, G., and E. M. Rathje (2015), Evaluation of one-dimensional site response techniques using borehole arrays, *Journal of Geotechnical and Geoenvironmental Engineering*, *141*(12), 04015,053.

Zhang, Z., W. Zhang, X. Chen, C. Fu, et al. (2017), Rupture dynamics and ground motion from potential earthquakes around Taiyuan, China, *Bulletin of the Seismological Society of America*, *107*(3), 1201–1212.