Topographic Effects in Strong Ground Motion

Manisha Rai

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Civil Engineering

Adrian Rodriguez-Marek, Chair
Martin C. Chapman
Russell A. Green
Joseph Wartman

August 10, 2015
Blacksburg, Virginia

Keywords: Topographic effects, Terrain classification, Seismic hazard, Ground motion prediction equation, NGA-West2 dataset

Topographic Effects in Strong Ground Motion

Manisha Rai

ABSTRACT

Ground motions from earthquakes are known to be affected by earth's surface topography. Topographic effects are a result of several physical phenomena such as the focusing or defocusing of seismic waves reflected from a topographic feature and the interference between direct and diffracted seismic waves. This typically causes an amplification of ground motion on convex features such as hills and ridges and a de-amplification on concave features such as valleys and canyons. Topographic effects are known to be frequency dependent and the spectral accelerations can sometimes reach high values causing significant damages to the structures located on the feature. Topographically correlated damage pattern have been observed in several earthquakes and topographic amplifications have also been observed in several recorded ground motions. This phenomenon has also been extensively studied through numerical analyses. Even though different studies agree on the nature of topographic effects, quantifying these effects have been challenging. The current literature has no consensus on how to predict topographic effects at a site. With population centers growing around regions of high seismicity and prominent topographic relief, such as California, and Japan, the quantitative estimation of the effects have become very important. In this dissertation, we address this shortcoming by developing empirical models that predict topographic effects at a site. These models are developed through an extensive empirical study of recorded ground motions from two large strong-motion datasets namely the California small to medium magnitude earthquake dataset and the global NGA-West2 datasets, and propose topographic modification factors that quantify expected amplification or deamplification at a site.

To develop these models, we required a parameterization of topography. We developed two types of topographic parameters at each recording stations. The first type of parameter is developed using the elevation data around the stations, and comprise of parameters

such as smoothed slope, smoothed curvature, and relative elevation. The second type of parameter is developed using a series of simplistic 2D numerical analysis. These numerical analyses compute an estimate of expected 2D topographic amplification of a simple wave at a site in several different directions. These 2D amplifications are used to develop a family of parameters at each site. We study the trends in the ground motion model residuals with respect to these topographic parameters to determine if the parameters can capture topographic effects in the recorded data. We use statistical tests to determine if the trends are significant, and perform mixed effects regression on the residuals to develop functional forms that can be used to predict topographic effect at a site. Finally, we compare the two types of parameters, and their topographic predictive power.

To my family :

my dearest husband Shrey

our loving parents

my caring brother Anurag

Acknowledgments

I want to thank my advisor, Prof. Adrian Rodriguez-Marek. Thank you for providing me with the opportunity to work with you on such an exciting project. I am really grateful to you for the guidance you have given me all these years. You have been the kindest, most supportive, and most patient advisor I could have asked for. You have been a friend and a family to me in Blacksburg. I also want to thank your lovely wife, Tina, and your adorable kids Naira and Alex who have always been so affectionate and kind to me.

Thanks to my committee members, Prof. Green, Prof. Chapman, and Prof. Wartman for their invaluable guidance and suggestions during my proposal and throughout my PhD, which helped shape my thesis. Thanks Prof. Green for your feedback on running the finite difference analyses. Thanks Prof. Chapman, for giving me feedback on the thesis, and for suggesting appropriate changes. Your comments have really helped me improve the quality of the thesis.

I would like to thank my co-authors on the journal papers: Dr. Alan Yong, who provided the shear wave velocity data for the ground motion stations in the dataset, and Dr. Brian Chiou who provided the ground motion residuals for the analyses. They both reviewed the papers thoroughly, which certainly improved the quality of the work. I want to thank Prof. Assimaki, who reviewed parts of our numerical analysis, and provided invaluable insights.

I thank all the professors in the department that I have taken classes or have had

a chance to interact with. You have greatly inspired me with your immense knowledge, expertise, and your dedication towards the profession. I thank all our geotechnical staff members in Patton hall, for helping me with all the administrative and technical issues. You all have been very kind and supportive throughout my stay at Blacksburg. Special thanks to Lindy Cranwell, Bonnie O. Franklin, Kira Showalter, Leigh Anne, and Sandy Simpkins for their help throughout the program. Thanks to Alec Flint, and Milko for their continuous technical support.

The National Science Foundation and the California Strong Motion Instrumentation program sponsored my research, and I greatly appreciate their support.

Thanks to all my friends and fellow geotechnical students for their friendship, help and support. I want to especially thank Ashly Cabas for being such a great friend. Thanks for the long tutoring sessions on Soil behavior. I will forever owe my qualifying exam victory to you. Thanks to Alex Reeb, and Jian Song for always addressing all my FLAC related question, and Sam Lasley and Haitham Dawood for all the technical discussions. Thanks to Abeera Batool for the help during my early years in the program and for being such a wonderful and kind friend. Thanks to Maryam, Melis, Rajat, and Sayantani for their friendship.

I thank my friends and ex-roommates, Sony, Harshada for being the most amazing roommates ever. You made my stay in Blacksburg so much more wonderful. I still remember our long conversations over tea, and those home cooked meals with both of you. I want to thank my friends Atashi, Meeta, Tina and Polomi for their friendship and kindness. I also thank my friends Rajan, Sudhir, Ankur and Abhinav for keeping us a good company.

I thank my husband Shrey, for his love, friendship, encouragement, and support. I am so blessed to have you in my life. You are an inspiration to work hard, pursue excellence, and being humble. You have taught me that kindness is the most important quality in a human being.

I thank my parents, Atul Kumar Rai, and Beena Rai, for all that I am today. They

have showered me with so much love and affection. Words cannot express my gratitude towards them. They have always given me the freedom to pursue any career I liked, and helped and supported me in all my life choices. I thank my parents-in-law, Sudhir Kumar Shahi and Rashmi Shahi, for being second parents to me. They have loved me and supported me like their daughter. I want to thank my brother, Anurag Rai, and my bhabhi Samata Rai for always being there for me. Thank you for calling me and checking up on me Bhaiya. With you around, I always know there is someone looking out for me.

Finally, I want to thank my lovely cats golu baby and goli. These fluffy balls of fur continue to brighten my day.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Objectives 3

1.3 Organization and key contributions 4

2 Literature Review **6**

2.1 Topographic effects 6

2.1.1 Observational and Experimental studies 6

2.1.2 Analytical and numerical studies 9

2.2 Summary of the literature review 14

2.3 Ground motion prediction equations 15

2.3.1 Introduction 15

2.3.2 Regression techniques and residual components 16

2.3.3 Site-effects in GMPEs and building codes 18

2.4 Concluding Remarks 19

3 An empirical model to predict topographic effects in strong ground motion using California small to medium magnitude earthquake database **20**

3.1 Abstract . . . 20

3.2 Introduction . . . 21

3.3 Ground motion data . . . 24

3.4 Methodology . . . 25

3.4.1 Topographic parameterization . . . 25

3.4.2 Residual analysis . . . 31

3.5 Regression/Model development . . . 39

3.5.1 Proposed model . . . 41

3.5.2 Model diagnostics . . . 42

3.6 Conclusion . . . 42

3.7 Acknowledgments . . . 48

4 Empirical terrain-based topographic modification factors for use in ground motion prediction **49**

4.1 Abstract . . . 49

4.2 Introduction . . . 50

4.3 Ground motion data . . . 51

4.4 Methodology . . . 52

4.4.1 Topographic Parameterization . . . 52

4.4.2 Residual analysis . . . 56

4.4.3 Significance test . . . 59
4.4.4 Other factors . . . 68
4.4.5 Regional differences . . . 69
4.5 Model development . . . 72
4.5.1 Model diagnostic . . . 76
4.6 Comparison with proposed topographic modification factors in Chapter 3 . . . 78
4.7 Summary and conclusion . . . 79
4.8 Acknowledgments . . . 81

5 Topographic proxies from 2-D numerical analyses 82

5.1 Abstract . . . 82
5.2 Introduction . . . 83
5.3 Topographic Parametrization . . . 85
5.3.1 Generating cross-sectional profiles . . . 86
5.3.2 Generate Finite Difference meshes . . . 87
5.3.3 Parameter computation . . . 92
5.3.4 Residual Analysis . . . 95
5.4 Summary and Conclusion . . . 102
5.5 Acknowledgment . . . 105

6 Conclusion 106

6.1 Empirical findings and contributions . . . 106

6.2 Limitations . . . 109
6.2.1 Data Limitations . . . 109
6.2.2 Modeling limitation . . . 111
6.3 Future work . . . 111
6.3.1 Studying other ground motion datasets . . . 111
6.3.2 Ground-motion directionality . . . 112
6.3.3 Topographic effects on vertical component of ground-motion . . . 112
6.3.4 PSHA . . . 112
6.4 Concluding remarks . . . 113

A Sensitivity analysis **114**

B Spatial clusters in site residuals **117**

Bibliography **122**

List of Figures

3.1 Locations of ground motion recording stations (green dots) in the SMME database used for this study are shown. Only stations with three or more recordings were selected resulting in a total of 459 stations. 23

3.2 Magnitudes versus closest distance for records in the SMM database for stations with three or more recording. 24

3.3 Measured V_{S30} values (Louie, 2005, 2007; Louie et al., 2011; Ancheta et al., 2014; Yong et al., 2013) are plotted against the values in the Chiou et al. (2010) database for the 208 stations where new measurements are available. . 26

3.4 Local curvature (C) and smoothed curvatures at a scale of 240 m, 360 m and 720 m plotted against H_{1250} for the stations used in the study 29

3.5 A comparison of a) elevation raster, b) H_{1250} and c) C_{360}. Note the similarity between the H_{1250} and c) C_{360} 30

3.6 Site residuals ($\delta S2S_s$) from Equation 3.7 are plotted against H_{1250} Mean site residuals for the three classes of H_d: *high*, *low*, and *intermediate*, using a threshold of 15 m are shown as straight lines. 35

3.7 Mean $\delta S2S_s$ for each site class (*high*/*intermediate*/*low*) along with a +/- 1 standard deviation error is shown for scales from 500 m to 3000 m. A threshold of 15 m is used for classifications. 36

3.8 ϕ_{S2S} for each site class (*high*, *low*, and *intermediate*) is shown for all the scales from 500 m to 3000 m. A threshold of 15 m is used for classifications. 37

3.9 P-values from variance analysis of $\delta S2S_s$ are presented in a graphical format for a) PGA, b) T = 0.3 s and c) 1 s. P-values of 0.05 and less indicate that the means site residuals for the three topographic class *high*, *low*, and *intermediate* are significant for that scale and threshold pair. This analysis is not possible on a scale of 250 m and thresholds of 20 m, 25 m as for this combination of scales and threshold, not all the three classes were populated by stations (hence the white regions in the plot). 38

3.10 The multi-linear functional form shown in Equation 3.12 is plotted with H_{1250} for a) T = 0.3 s and b) T =1 s. The transition from *low* to *intermediate* occurs from x = -40 m to -15 m and from *intermediate* to *high* from x = 15 m to 40 m. No correction is proposed for the PGA. 43

3.11 $\delta S2S_s$ terms from Equation 3.11 are plotted with H_{1250}. A moving average of the$\delta S2S_s$ terms computed using local regression (loess) is also shown. Most of the residual trend from Figure 3.6 is removed suggesting that Equation 3.12 accounts for most of the topographic effect in the ground-motion intensity. . 44

3.12 $\delta S2S_s$ terms from Equation 3.11 are plotted with V_{S30} values at the station. A moving average of the $\delta S2S_s$ terms computed using local regression (loess) is also shown. Absence of any trend indicates that using the (Chiou et al., 2010) functional form with the new V_{S30} values doesnt produce any bias in V_{S30} scaling. 45

3.13 $\delta S2S_s$ from Equation 3.11 plotted with H_{1250} at the station for different V_{S30} bins. A moving average of the $\delta S2S_s$ terms computed using local regression (loess) is also shown. Lack of any significant trend suggests that there is no coupling between V_{S30} and topography. 46

4.1 Locations of earthquake hypocenters and ground motion recording stations for the data used in this study. Only stations with three or more recordings are considered. This filtering of data resulted in stations within a) California and b) Japan 52

4.2 Magnitude distance distribution for the data used in the study. 53

4.3 The top row consists of smoothed elevation raster, with a) no smoothing, b) smoothing using a scale of 360 m, and c) smoothing using a scale of 720 m. The middle row shows corresponding smoothed slopes and the bottom row shows corresponding smoothed curvatures. Both smoothed slopes and curvatures are computed using corresponding smoothed elevations. 55

4.4 The variation of relative elevation with scale is shown here. a) is the elevation raster, b) is H_{500}, c) is H_{1500}, and d) is H_{3000}. Note that at smaller scales, finer features are visible. As the scale is increased, broader features become more prominent and finer details are lost. 57

4.5 Scatterplot of C_{360} and H_{1500}. The two parameters have a high linear correlation. 58

4.6 Variation of site residuals ($\delta S2S_s$) from Equation 4.1 with smoothed slope at a scale of 180 m (S_{180}). A moving average of the $\delta S2S_s$ terms computed using local regression (loess) is also shown. There is no visible trend between the two variables indicating that slope does not systematically affect ground motions. 60

4.7 Variation of site residuals ($\delta S2S_s$) from Equation 4.1 with relative elevation at s scale of 1500 m (H_{1500}). A moving average of the $\delta S2S_s$ terms computed using local regression (loess) is also shown. 61

4.8 A terrain is classified using a threshold (t) values of 0.5 (top) and 1 (bottom) for scales of a) 500 m, b) 1500 m and c) 3000 m. 62

4.9 Mean site residuals ($\delta S\bar{2}S_s$) for the *high*, *intermediate* and *low* class along with the +/- 1 standard deviation error bars. 64

4.10 Correlation coefficients values between the site residuals and the relative elevation parameters computed at scales of 500 m, 1500 m, and 3000 m 66

4.11 p-value from ANOVA and Tukey's pairwise t-test at different period. The stations are classified using a scale of 500 m, 1500 m, and 3000 m and a threshold t of 0.5, 1. 67

4.12 Variation of site-to-site variability or the $\phi S2S_s$ values for *high*, *low*, and *intermediate* sites with period. The stations are classified using a scale d of 1500 m, and a threshold t of 0.5 69

4.13 Variation of the site terms ($\delta S2S_s$) with V_{S30} at the ground motion stations for the three topographic classes. The classification is done using a scale of 1500 m and a threshold corresponding to $t = 0.5$. Moving average value of site terms for a topographic class is also shown along with the 95 percent confidence interval. The average value for site terms within each class does not seem to change significantly with the V_{S30} values, suggesting that the two are not correlated. 70

4.14 Histogram of Closest distance, Magnitude and $Z_{1.0}$ value for the three topographic classes. The classification is done using a scale (D) of 1500 m and a threshold (t) of 0.5. 71

4.15 Variation of site residuals ($\delta S2S_s$) from Equation 4.1 (prior to the removal of topographic effects) at a period of 0.5 s as a function of H_{1500} for Japan, Northern California (NCal), and Southern California (SCal). Its is evident that most of the residual trends observed in the total data is coming from SCal stations. Also, for Japan, majority of the stations have H_{1500} value less than zero. 71

4.16 The resulting bootstrapped models from Equation 4.3 for a period of 0.5 s along with the histograms of coefficients c_{low}, c_{int} and c_{high}. The histograms show the mean value of the coefficient in a red dotted line. 73

4.17 Proposed modification factors, c_{low} (H_{1500} < -20 m) and c_{high} (H_{1500} > 20 m) for topography. Linear interpolation should be used to estimate the modification factors for absolute H_{1500} values of 17 m to 20 m. 74

4.18 Scatterplot of site terms ($\delta S2S_s$) from Equation 4.2 with H_{1500} at the station. Lack of a significant trend suggests that the functional form in Equation 4.3 has accounted for most of the bias in the data. 76

4.19 Scatterplot of site residuals ($\delta S2S_s$) from Equation 4.2 with V_{S30} at the station. Lack of any significant trend with respect to V_{S30} indicates that the functional form used in Equation 4.3 is not introducing any systematic bias with respect to V_{S30}. 77

4.20 A comparison of the topographic modification factors c_{high}, c_{low} obtained from using different functional forms and datasets. 79

5.1 Detailed flowchart of the methodology used for topographic parameters from finite difference analysis. 84

5.2 3D terrain around a site (left) and cross-sectional profiles across the station in 6 different directions (right) are shown for the Tamalpais peak B station (37.9231, -122.5983). The recording station is located at [0, 0] m. 87

5.3 Schematic illustration of the finite difference model used for the analysis at one of the recording stations. The model uses a realistic topographic cross-section profile at the top. The station is located at the surface, equidistant from both the lateral edges. The height of stations from the base is same for all stations. Free-field boundary condition is applied to the lateral boundaries and quiet boundary conditions are applied at the base. 90

5.4 Acceleration and velocity time history for input motion with 10 cycles, and a period of 1 s. 91

5.5 A comparison of the normalized PGA values obtained from our analysis, with the Bouckovalas and Papadimitriou (2005) for the case of 30 degree slope, height 50 m, damping less than 5 % and an input acceleration of $H/\lambda = 0.2$. 93

5.6 Normalized PGA values for the six cross-sections of the Tamalpais peak B station are shown for an input wave of period 0.5 s, and 1 s. The amplification and deamplification patterns seem to emerge at distances proportional to the wavelength of input motion. Also note that the normalized PGA values are greater than one for most of the azimuths, indicating an average amplification at the station. 94

5.7 Scatterplot of the five $lnAmp$ parameters computed using wave of period T = 1s, and a constant shear wave velocity of 500 m/s for each station. The corresponding values of correlation coefficients, and the parameter density plot is also shown. 97

5.8 Coefficient of determination (R^2) value from the loess regressions on the intra-event residuals from the Chiou and Youngs (2014) model using the five $lnAmp$ parameters from Approach 1. 98

5.9 Coefficient of determination (R^2) value from the loess regressions on the intra-event residuals from the Chiou and Youngs (2014) model using the five $lnAmp$ parameters from Approach 2. 99

5.10 Scatterplot of the $lnAmp_{avg}$ obtained from the analysis using Approach 1, and Approach 2, and the relative elevation parameter H_d for d values of 250 m - 3000 m. The relative elevation parameter was computed at each site in Chapter 4. The values are shown for an input wave of period 0.5 s and 1 s. . 100

5.11 Correlation coefficient values between $lnAmp_{avg}$ from Approach 1, and H_d for d values of 250 m - 3000. The three lines correspond to input wave of period 0.5 s, 1 s and 2s. The peaks are shown with solid symbols. Note that the respective peaks occur at a scale equal to the wavelength of the input motion. 101

5.12 Scatterplot of loess predictions using the geometry-based H_{1000} parameter, and the predictions from loess regression using the $lnAmp_{avg}$ parameter for the 2 s wave, at periods of 0.3s, 0.5s, 1.0 s, and 3.0 s. The solid black line shown the x = y line. 103

5.13 Comparison of the coefficient of determination (R^2) values from the loess regressions on the intra-event residuals using the H_{1000} parameter, and the $lnAmp_{avg}$ parameters using 2s input wave. 104

A.1 Variation of mean $\delta S2S$ for the three topographic classes as a function of N, where N is the minimum number of recordings at a site for it to be included in the study. +/- standard deviation error bars are also shown. 115

A.2 Variation of ϕ_{s2s} for the three topographic classes as a function of N, where N is the minimum number of recordings at a site for it to be included in the study. +/- standard deviation error bars are also shown. 116

B.1 The spatial variation of the site terms ($\delta S2S_s$) from Equation 4.3 are shows for the stations in the three topographic classes for California and Japan. The points are colored based on the site residual values at that station for T = 0.5 s. The standard deviation values used in the plots correspond to the standard deviation value of the site residuals within that class. The standard deviation of site terms and the total number of stations in each class are shown in the figure. 119

B.2 High-high (HH), high low (HL), low-high (LL) and low-low (LL) clusters for stations in the three topographic classes in California and Japan. HH corresponds to stations which have a high site residual and are surrounded by similar stations. These clusters are also referred to as hot spots. LL corresponds to stations that have a low site residual is surrounded by stations with similarly low values of site residuals. Such clusters are called cold spots. HL and LH are stations that have high value of site terms and are surrounded by stations with low values of site terms or vice versa. These points are called as outliers. In Japan, there were too few stations in the High class to get a reliable clustering analysis result and therefore the results have not been shown. The points only mark the locations of the stations. 121

List of Tables

3.1 y_N values for the models at periods of 0.3 s and 1 s. 40

3.2 Models used to capture topographic effects 41

3.3 y_N values for the models at periods of 0.3 s and 1 s. 42

4.1 Period dependent coefficients for f in Equation 4.3 75

5.1 Parameter values assumed for the analysis in the two approaches. 92

5.2 List of parameters used as a proxy for topographic effects. 95

B.1 The global statistics from the Global Morans I tool on stations in the three topographic classes for California and Japan. 118

Chapter 1

Introduction

1.1 Motivation

Earth's surface topography is known to affect earthquake ground motions. Typically, amplification of ground motion occurs on hills and ridges and de-amplification occurs in canyons and valleys, with respect to a similar site on a flat terrain (e.g., Boore, 1972; Davis and West, 1973; Spudich et al., 1996). There have been cases where unusually high peak ground accelerations (PGAs) were observed in the vicinity of a topographic feature. An example of such an amplification is the recording at the Tarzana site from the 1994 Northridge earthquake where a PGA of 1.8 g was observed, whereas the neighboring regions recorded PGAs of less than 1g (e.g., Shakal et al., 1994; Bouchon and Barker, 1996; Spudich et al., 1996). Numerical studies confirmed that these amplifications were caused due to the topography at the site (Spudich et al., 1996); however, they could not explain the large magnitudes of the observed amplifications. Another very well-known case of topographic amplification is the Pacoima Dam abutment recording from the San Fernando earthquake which recorded a PGA of 1.25 g (Trifunac and Hudson, 1971). Boore (1973) used finite difference analysis to predict the topographic effects at the Pacoima site and concluded that topography would be expected to amplify the recorded ground motion by 50 percent at high frequencies.

There are several other cases where the evidence of topographic amplification was observed in the form of uneven damage distribution in the neighboring regions. In most of these cases, buildings that were located on the hilltops suffered much more damage than buildings that were located on the base or a flat region. Examples of such earthquake are the Lambesc earthquake (France 1909), the San Fernando earthquake (1971), the Friuli earthquake (Italy 1976), the Irpinia earthquake (Italy 1980), the Chile earthquake (1985), the Whittier Narrows earthquake (1987), The Egion earthquake (Greece 1995) and the Athens (Greece 1999) earthquake, the Eje-Cafetero earthquake (Colombia 1998), and the Bingöl earthquake (Turkey 2003) (Assimaki et al., 2005).

A number of researchers have studied the effects of topography on ground motions using numerical and analytical schemes (e.g., Davis and West, 1973; Rogers et al., 1974; Griffiths and Bollinger, 1979; Tucker et al., 1984; Geli et al., 1988; Bouchon and Barker, 1996; Assimaki and Gazetas, 2004; Meunier et al., 2008; Assimaki and Jeong, 2013). These studies provide valuable insights into the physics behind the observed effects, and show that a number of factors influence the magnitude of amplifications observed at the surface and the frequencies at which these effects are observed. These factors include: the type of the incident wave, wavelength, the angle of incidence, slope angles, and soil stratigraphy (e.g., Assimaki and Gazetas, 2004; Hartzell et al., 1994; Assimaki and Gazetas, 2004; Assimaki et al., 2005; Assimaki and Jeong, 2013), among others.

Numerical analyses however have certain limitations. Most of the numerical studies are performed using an idealized 2D geometry with simple soil layering and properties. As a consequence, they are often found to under-predict the amplifications observed in the field (Geli et al., 1988). A few studies have used more realistic 3D models of a terrain to compute amplifications due to surface topography (e.g., Lee et al., 2009; Chaljub et al., 2010; Maufroy et al., 2012) . Though, such studies are able to predict the effects observed in the field more accurately, they are usually more computationally expensive to perform than the 2D studies (Chaljub et al., 2010). Also, most of these studies do not verify their findings with the recorded data, and therefore their results are not particularly useful for predictions.

There is no consensus in the scientific community on how to quantify topographic effects and account for these in the predictions of ground motion intensity. Consequently, the design of buildings located on or near a topographic feature, such as a hill or a ridge, can potentially be un-conservative. Not only are the estimations of spectral acceleration on ridges possibly un-conservative, but there could also be a higher uncertainty (standard deviation) associated with their prediction. This increase in the median prediction and uncertainty will lead to a higher hazard at the site located near a hill or a ridge. With cities developing around regions of significant topographic relief, this problem has become increasingly important.

In this dissertation, we address this issue by developing a comprehensive framework to parameterize topography and propose models to account for topographic effects in ground motion prediction equations. The models are based on an extensive empirical analyses of recorded ground motions and use simple topographic parameters for predictions. The simplicity of the parameters will make them widely applicable to most engineering problems.

1.2 Objectives

The main objectives of this study are:

- To develop a framework to parameterize earth's surface topography at a site.
- Study ground motion model residuals to test if topographic effects are observable in the recorded data.
- Conduct statistical tests to select parameters that are good proxies for topographic effects in the ground motions, and find periods where the effects of topography on ground motions are significant.
- Develop and propose models to predict topographic effects in the ground motions.

1.3 Organization and key contributions

This dissertation is divided into 6 chapters and 2 appendices. Chapters 3, 4 and 5 are manuscripts that have either been submitted to journals or are in the process of submission at the time of the submission of the dissertation. **Chapter 2** consists of a review of related literature and includes important findings from these studies.

In **Chapter 3**, we study the ground motion residuals from a subset of the Chiou et al. (2010) Small to Medium Magnitude Earthquake (SMME) dataset. The SMME dataset consists of ground motion recordings from earthquakes in California with magnitudes less than 5.5. We first propose a topographic parameterization scheme that quantifies topography at a site. The scheme uses the elevation data in a neighborhood of the site for this purpose. We classify the terrain underneath the ground motion recording stations in the dataset into 3 broad classes based on the values of the computed topographic parameter at the station. We compute the average site residuals for each of the three classes of topography, and use statistical tests to determine if the site residuals in the three classes are statistically different from one another. Using mixed effects regression, and the topographic parameters, we fit several different models to remove the trends from the data. We compare these models based on their overall fit to the data and finally select one model. The model is proposed as a modification to the Chiou et al. (2010) model and can be used to predict topographic effects at a generic site. As all the recordings in the dataset were from earthquakes in California, the applicability of the model is limited to California.

In **Chapter 4**, we extend the analysis from Chapter 3 to a subset of the global NGA-West2 dataset from Chiou and Youngs (2014). The NGA-West2 dataset is a much larger dataset consisting of ground motion recordings from 7 geographic regions around the world; though the subset we use in the study consists of recording from California and Japan only. The dataset consists of spectral acceleration (S_a) values at 105 periods from 0.01s to 10 s from earthquakes of magnitude 3 and higher. In addition to the topographic parameters from Chapter 3, we also study smoothed slopes at the station as a potential topographic

parameter. Similar to the approach used in Chapter 3, we first classify the terrain at the location of the stations in three broad classes, based on the value of the topographic parameter, and a threshold. We then perform statistical tests to determine if the mean site residuals in the three classes are significantly different from one another. We propose a model for topographic modification, and also propose estimates of epistemic uncertainties in the model using bootstrapping. These epistemic uncertainties can be used in the logic-tree implementation of probabilistic seismic hazard analysis (PSHA).

In **Chapter 5**, we explore an alternative parameterization scheme for topography that is computed using results of simplistic 2D numerical analyses. For these analyses, we use the same subset of the NGA-West2 dataset as in Chapter 4. Using the elevation data at the location of the ground motion recording station, we create 2D cross-sectional profiles of the ground surface at the ground motion stations in six different directions. We generate a finite difference mesh in FLAC software for each station-azimuth pair and distort their top surface to fit the shape of the corresponding cross-sections. Using simple soil properties, boundary conditions, and a harmonic input, we perform dynamic analyses on the resulting meshes and record the time histories at the top surface of each mesh. Using these synthetic time histories, we compute the PGAs and normalize them with PGA from a 1D analysis with no topography. Using the 6 normalized PGA values at station from 6 directions, we compute a family of parameters for each ground motion recording station. We study the intra-event residuals from the Chiou and Youngs (2014) ground motion model for trends with respect to these parameters and compare them to those observed in Chapter 4.

Chapter 6 presents the summary, important findings, and conclusions from this work. **Appendix A and B** contain additional details from **Chapter 4**. **Appendix A** consists of the same analysis presented in Chapter 4, but with a higher cut-off threshold for the number of recordings per station. This analysis was done to check how the ground motion variabilities are affected by the choice of the subset of the ground motion dataset. **Appendix B** presents the results of a cluster analysis that was performed to test if there was any spatial correlation between the high residuals or the low residuals, within each topographic class.

Chapter 2

Literature Review

This chapter reviews publications on topographic effects and ground motion prediction equations (GMPEs). In the first part, we review past publications on experimental, numerical, and analytical studies on topographic effects. We conclude this section with a brief summary of our findings from the review. In the second part we review research on GMPEs, where we first provide a brief background on the topic, leading to an introduction of the NGA-West2 dataset. We then cover some of the existing regression techniques that are used to fit these predictive equations, and also briefly discuss different residual components in the regression. Lastly, we talk about site-effects and how they are accounted for in the predictions.

2.1 Topographic effects

2.1.1 Observational and Experimental studies

Boore (1972), in his discussion on the observed evidence of topographic effects, revisited some cases where the induced vertical ground accelerations locally exceeded gravity and resulted in regions of churned, overturned ground and boulders. These events included the Indian earthquake of 1897 (Richter, 1958), the Cedar Mountain earthquake of 1932

(Gianella and Callaghan, 1934), the Hebgen Lake earthquake of 1959 (Hadley, 1964), among others. They quoted Hadley (1964) who noted that these regions of churned earth were found mostly on ridges, crests, or other topographic eminences. These observations led them to identify a dependence of ground motion amplifications on topography. Since these early mentions of topographic effects, observational evidence of topographic amplifications has been found in numerous other cases (e.g., Celebi, 1987, 1991; Hartzell et al., 1994; Hatzfeld et al., 1995; Gazetas et al., 2002). These studies showed that the structures that were located on ridges suffered much more damage than those located on the base. An example of a case with a distinctive observational evidence of topographic effects was the M 7.8 Central Chile earthquake where the four- to five-storied buildings located on the ridges of the Canal Beagle subdivision suffered extensive damage (Celebi, 1987). Celebi (1991) studied the aftershock data from the earthquake recorded on a dense array temporarily established at Canal Beagle and found that the spectral ratios of the observed ground motions were amplified on the ridges of Canal Beagle for a frequency of 2 to 4 Hz and 8 Hz. The 2 to 4 Hz frequency range was well within the fundamental frequency of typical four- to five-storied buildings that suffered heavy damage during the main event.

Similar damage patterns were reported in the field during the Lambesc earthquake (France 1909), the San Fernando earthquake (1971), the Friuli earthquake (Italy 1976), the Irpinia earthquake (Italy 1980), the Chile earthquake (1985), the Whittier Narrows earthquake (1987), the Egion earthquake (Greece 1995) and the Athens (Greece 1999) earthquake, the Eje-Cafetero earthquake (Colombia 1998) and the Bingöl earthquake (Turkey 2003) (Assimaki et al., 2005). An exception to this pattern was observed during the Whittier Narrows earthquake of California in 1987, where most of the damage was concentrated along the slopes of the hill instead of the hill tops (Kawase and Aki, 1990). Kawase and Aki (1990) suggested that that this anomaly was probably a result of the combined effects of topography and critically incident SV waves.

In addition to the observational evidence of topographic effects showing non-uniform damage in the vicinity of a topographic feature, there is also ample evidence of topographic

amplifications in the recorded ground motions from earthquakes. Two notable cases are the Pacoima dam abutment recording from the San Fernando earthquake of 1971 that recorded a maximum horizontal peak ground acceleration (PGA) of 1.25 g (Trifunac and Hudson, 1971) and an extremely high PGA of 1.78 g observed on a relatively gently sloped site in Tarzana, California during the M 6.7 Northridge earthquake of 1994 (Shakal et al., 1994). Both these PGAs were one of the highest recorded at the time. A number of other studies show that ground motions were typically amplified on the hill tops and ridges, with respect to the motions at the base and that these effects were frequency dependent (e.g., Davis and West, 1973; Griffiths and Bollinger, 1979; Tucker et al., 1984; Celebi, 1987). The spectral periods at which amplifications occurred were typically proportional to the size of the feature on which the recordings were made (e.g., Davis and West, 1973; Tucker et al., 1984; Celebi, 1991). Davis and West (1973) deployed several seismographs at the base and crest of Kagel Mountain and Josephine peak in California and recorded aftershocks from the San Fernando earthquake of 1971. They also deployed seismographs at the base, crest, and slope of Butler Mountain near Nevada and recorded the signals from cavity collapse in a Nevada Test Site detonation. They computed the crest to base ratios of the 5 percent damped pseudo-velocity spectra and found that the magnitude and periods of amplification varied with the size of the mountains. In general, frequency-dependent amplifications were observed at the crest of all the three mountains. The smallest mountain amplified the motion in a narrow range of period (peaking at 0.3 to 0.5 s), the medium size mountain amplified motions over a slightly broader range (peaking at 0.4 to 0.5 s), and the largest mountain showed less amplifications, but over a broader range of periods. A reversal of this general trend of observing highest amplifications on the ridges and hill tops was observed by Griffiths and Bollinger (1979) who noted de-amplifications on one of the five ridges that had a less pronounced peak than the other. They believed that this was either a result of an irregular, 3D configuration of this particular ridge compared to the relatively 2D slopes on the other ridges, or a result of a combined effect of multiple ridge systems.

A few other factors have been shown to influence the magnitude of amplifications in the

recorded ground motions such as the wave type, where observed amplifications were higher for horizontally polarized signals than for vertically polarized signals (Tucker et al., 1984), and the direction of the feature, where a directional resonance in the top to base spectral ratios were observed in the direction transverse to the elongation of the hill (Spudich et al., 1996).

2.1.2 Analytical and numerical studies

An earthquake produces different types of body (P, S) and surface waves (Rayleigh, Love). The body waves that are generated when a fault ruptures interact with the surficial layers and the surface of the earth generating surface waves (Kramer, 1996). When a topographic feature is present, the seismic waves experience focusing or defocusing and scattering. The focusing, defocusing, and the interference of the direct and diffracted wave causes wave amplifications and deamplifications at the surface (Bard, 1999). This problem has been studied analytically and numerically by several researchers in the past. With the advent of advanced numerical techniques, researchers have been able to identify a number of factors that affect the response of a topographic feature to an incident wave. These factors include: the type of incident waves, the angle of incidence, the dimensionless frequency (fl/V_s, where f is the frequency of input motion, l is the characteristic length of the feature, and V_s is the shear-wave velocity), slope angle, smoothed curvatures, the direction of elongation of the feature, shear wave velocity structure, and the material damping, among others. Some of these numerical studies are reviewed below.

To study topographic effects, researchers have used methods such as finite difference (e.g., Boore, 1972, 1973; Boore et al., 1981; Bouckovalas and Papadimitriou, 2005), finite elements (e.g., Smith, 1975; Assimaki and Gazetas, 2004; Assimaki et al., 2005; Tripe et al., 2013; Assimaki and Jeong, 2013), boundary element methods (e.g., Sánchez-Sesma et al., 1982; Nguyen and Gatmiri, 2007) and discrete wave-number method (Bard, 1982; Durand et al., 1999). Some of these studies are parametric studies where the effect of changing one

or more model parameters on topographic effects is studied, while others are case studies that try to model past cases of earthquakes where topographically correlated amplifications were observed. A majority of case studies consider 2D models and often under-predict the observed magnitude of amplifications in the recorded ground motions (Geli et al., 1988; Bouchon and Barker, 1996). Some believe that this discrepancy in the predicted and observed amplification is a result of the numerical models lacking the complexity of the real-world scenarios (Geli et al., 1988; Bouchon and Barker, 1996), while others suggest that the choice of reference site to compute amplifications can lead to either overestimation or underestimation of the amplification factors on top of the hills (Tucker et al., 1984; Geli et al., 1988). Nevertheless, most of these numerical studies provide good qualitative estimates of the observed amplification. According to a review article by Geli et al. (1988), most of the numerical studies predicted amplifications on top of the ridges over a broad range of frequencies, and a de-amplification of the low frequency waves (for wavelengths larger than mountain width) at mountain base, with an alternation of amplification and de-amplification at higher frequencies. On mountain slopes, the phase and amplitude varied rapidly causing differential motion. In the ensuing, we look at the effects of some of the factors that have been shown to influence the ground motions in numerical studies.

Incident wave-type

A few 2D numerical studies have compared the effects of different incident wave types (SH, P, SV) on the predicted amplification. Typically, for simple 2D slopes, amplification of the SV waves was found to be higher than the SH waves (Ashford and Sitar, 1997). In case of incident SV waves, generally a zone of amplification existed near the slope due to interaction of the primary SV and the diffracted Rayleigh waves (e.g., Boore et al., 1981; Bouchon, 1973; Wong, 1982). In addition to the nature of the input waves, its polarization also affects the response of the feature, e.g. Bouchon and Barker (1996) studied the response of a three dimensional topography to incident shear waves polarized in different directions

using the boundary integral equation method. They observed that the hill produced higher amplification of ground motions when the input waves were polarized transverse to the direction of elongation of the hill compared to when it was polarized parallel to the direction of the elongation of the hill. They also observed a rotation of the polarization of particle motions from the base to the top of the hill, with particle motions being more transverse to the hill at the top than the base.

Angle of incidence

Some researchers have investigated the effects of incidence angle on the topographic amplifications for cases of simple 2D topographies (Boore, 1973; Wong and Jennings, 1975; Ashford and Sitar, 1997) and suggest that it has a significant effect on the amplification of waves. Ashford and Sitar (1997) studied the response of a stepped half-space subjected to an inclined shear wave from 0 to 30 degree using generalized consistent transmitting boundaries and found that the horizontal motion at the crest of the slope is amplified for waves traveling into the slope and de-amplified for waves traveling away from the slope, in comparison to the free-field motion behind the crest of the slope. They measured the incidence angles clockwise from the vertical axis, and referred to the waves with positive incidence angle as traveling away from the slope, and those with negative angles as traveling into the slope. They also suggested that these amplifications can be twice the amplifications observed for a vertically propagating wave. A similar study on a more realistic model using inclined incident shear waves showed that even though topographic amplifications are greater for inclined waves, the magnitude of horizontal and vertical acceleration at the crest is usually greatest for the case of vertical incidence. They suggest that it is therefore sometimes acceptable to use vertical incidence for site response analysis and force-based slope stability analysis because it simplifies the problem and does not affect the magnitude of maximum accelerations. Boore (1973) used a simple finite difference model to study the topographic effects observed on Pacoima Dam site during the 1971 San Fernando earthquake in California. As the topography at the

site of the Pacoima dam abutment recording station was complicated, they considered two site geometries to adequately represent the site: A ridge model with two angles of incidences at 0 and 45 degrees and a canyon model with vertical incident waves. They observed that the ridge model with vertical incidence produced the greatest amplification for frequencies between 3 Hz and 10 Hz. At shorter frequencies, the effect of the canyon became relatively more important and produced attenuation of wave motion.

Site geometry and wavelength of input motion

Past numerical studies have shown that topographic effects are more pronounced for input wavelengths that are comparable to the dimensions of the feature (e.g., Geli et al., 1988; Ashford et al., 1997; Ashford and Sitar, 1997). In addition, the slope angle has been shown to influence the amplifications observed at the crest in 2D slopes (e.g., Ashford et al., 1997; Bouckovalas and Papadimitriou, 2005; Tripe et al., 2013; Maufroy et al., 2015). In particular, Ashford et al. (1997) studied the response of steep slopes using generalized consistent transmitting boundaries and found that the topographic effects are most visible for slopes greater than 60 degrees and decrease with slope angles. Also, for a stepped layer over a half space, the response is dominated by the natural frequency of the site behind the crest. Bouckovalas and Papadimitriou (2005) performed a parametric study using a simple 2D finite difference slope model and studied the influence of slope inclination, normalized height of the slope (H/λ, where H is the height of the slope and λ is the wavelength), number of significant excitation cycles and critical hysteretic damping on the amplifications. They found that both slope angle and the normalized height of the slope have a significant effect on horizontal and vertical ground motions. They also found that the only effect of damping is on the distance to the free-field, while the number of significant cycles of excitation has only a minor effect overall. More recently, Maufroy et al. (2015) performed numerical analysis using a realistic 3D model of a terrain, and found that the average amplifications at a site were highly correlated with curvatures that were smoothed using a window proportional in

size to the wavelength of input motion.

Soil stratigraphy

Soil stratigraphy plays an important role in the surface response to input motion and can cause both amplification and de-amplifications of the wave (e.g., Assimaki and Gazetas, 2004; Assimaki et al., 2005; Assimaki and Jeong, 2013). When topography is present at the surface, the observed response is a result of a combined effects of the topography and the surface layering effects (Geli et al., 1988). Assimaki and Gazetas (2004) studied different configurations of layered soil profile in a 2D slope and found that the topographic amplification of the motion is more pronounced for a layered medium in a certain frequency range due to multiple reflections and transmissions of the vertically incident S-waves through the horizontal layers, causing an additional amplification at the crest. Assimaki and Jeong (2013) performed site-specific numerical simulations to understand the severe damage caused during the 2010 M 7 Haiti earthquakes and found that the observed amplifications were most likely caused by a combined effects of soil layering and topographic amplifications.

As mentioned earlier, most of these results were based on the analysis of 2D numerical models. A natural limitation of such studies is that the models they used to predict the amplifications in the field are highly simplified representation of the real world scenario. In reality, almost all the topographic features are continuous three-dimensional features. To address this problem, some researchers considered more realistic 3D models to estimate topographic effects at the site (e.g., Lee et al., 2009; Chaljub et al., 2010; Maufroy et al., 2012, 2015). These studies show that topographic effects are strongly dependent on the source frequency content, where the high frequency waves interact with small-scale topographies and cause amplified peaks. Most of these studies also agree with the other 2D studies that observed amplification along the ridges and a deamplification along the depressions; although 3D studies were better than the 2D studies at predicting the magnitude of the observed effects (e.g., Maufroy et al., 2015; Chaljub et al., 2010). However, 3D numerical analyses

are numerically expensive to perform Chaljub et al. (2010), and hence may not always be the best choice for developing predictive equations. Maufroy et al. (2015) proposed a model that predicted average amplification at a site based on the frequency-scaled curvature of the ground surface at the site. However, given the model was based purely on numerical analysis, it has a limited use for predictions without an extensive validation using recorded ground motions. Another limitation of the study was that it only analyzed motions in a small region, with uniform sub-surface conditions across every station and so it is difficult to determine if the results are acceptable outside the test area.

2.2 Summary of the literature review

An extensive literature review on topographic effects has been done in the past (Geli et al., 1988; Assimaki et al., 2005). In this section, we consolidate the findings from these studies and from our current review:

- In the time domain, grounds motions are amplified on convex features such as hill tops and ridges and de-amplified on concave features such as valleys and canyons. In the frequency domain, a broadband amplification of ground motions occurs on the top of the hill and a de-amplification occurs at the base with alternating amplification and deamplification at higher frequencies. On the slopes, the amplitude and phase vary rapidly causing a differential motion.

- There is a qualitative agreement between theory and observation regarding the existence of ground motion amplification; however, the amplitude of such effects is mostly under-predicted in numerical analyses. This under-prediction is believed to be a result of the numerical models not being complex enough. Some believe that this under-prediction is a result of an inappropriate selection of the reference station in the field for computing amplifications.

- Topographic effects are prominent for wavelengths comparable to size of the irregularity. For wavelengths larger than the dimensions of the feature, amplifications are negligible.

- The magnitude of crest to base amplification ratios depends on the ratio of mountain height to mountain half width and is found to be proportional to this ratio. The amplifications also depend on the angle of incidence of the incoming wave, more so at the near critical incidence.

- The amplifications are also affected by incident wave type. The amplifications are typically larger for horizontal components than for vertical component of the motion. The amplifications were however not found to be affected by the damping ratio of the soil and the duration of the motion.

- In numerical studies using more realistic topographic, amplifications were found to be strongly correlated to frequency-scaled curvatures.

- Topographic amplifications are affected by soil stratigraphy. In particular, amplifications are magnified in the presence of soft, shallow soil layers.

2.3 Ground motion prediction equations

2.3.1 Introduction

Ground-motion prediction equations (GMPEs) are used to predict the distribution of ground-motion intensity at a given site due to an earthquake, using predictor variables like the earthquake magnitude, the source-to-site distance, and other source-, site-, and path- characteristics. A number of GMPEs have been developed in the past. Gunberg and Green (2011) compiled a list of 45 GMPEs and provided MATLAB scripts and spread sheets for the models. Douglas (2014) provided the most updated, and comprehensive summary of

all the empirical GMPEs published between 1964 and 2014. The report consisted of 365 empirical GMPEs for the prediction of peak ground acceleration (PGA) and 232 empirical models for the prediction of elastic spectral coordinates.

In the earlier years of their development, GMPEs were simple and often required only two parameters, the magnitude and the distance, for predictions. With a multi-fold increase in the size of the recorded ground motion database, and a better understanding of different phenomena affecting ground motions, the prediction equations became more complex with time. In 2003, the Pacific Earthquake Engineering Research Center (PEER) started a large collaborative research program, now known as the NGA-West project, to develop next-generation attenuation relationships for shallow-crustal earthquakes in active tectonic regions (Bozorgnia et al., 2014). The project was completed in 2008, with the development of a large ground-motion database, and a set of GMPEs. This was followed by the NGA-West2 project that expanded the original NGA-West dataset to include other small-, medium-, and large-magnitude earthquakes, and addressed some of the shortcomings of the NGA-West project. The NGA-West2 database consists of one of the largest sets of recorded ground-motions and associated metadata, available globally. It spans seven regions around the world, with spectral acceleration values available at 105 periods between 0.01 s and 10 s. A large number of recordings in this dataset make it ideal for empirical studies.

2.3.2 Regression techniques and residual components

GMPEs are fit using regression on the recorded ground motion data. Earlier prediction equations used simple linear regression, where the PGA or any other intensity measure was expressed as a linear function of the predictor variables. Joyner and Boore (1981) introduced a two-stage regression method to decouple magnitude and distance dependence. A drawback of these approaches was that it put more importance to an earthquake that had more recordings. To effectively handle the problem of weighing, Brillinger and Preisler (1984) for the first time used the random-effects model which incorporated an explicit earthquake-to-

earthquake component of variance and a record-to-record component. Their approach used a maximum likelihood method to estimate the parameters in the prediction. Maximum-likelihood approach is a parameter estimation technique that maximizes the likelihood of observing the given data. Abrahamson and Youngs (1992) provided an alternative algorithm for using the random-effects model in regression analysis that used an iterative approach to maximize a likelihood function. Their method was more stable than the maximum likelihood algorithm used by Brillinger and Preisler (1984). Joyner and Boore (1993) also introduced a new method for implementing the Brillinger and Preisler (1984) regression.

Most of the recent ground motion models were fitted using the mixed effects regression approach. A mixed-effects model consists of fixed-effects and random-effects. The fixed-effects are of primary interest and are assumed to be repeatable. The random-effects are sampled from a random distribution. A typical GMPE can be written as:

$$lnS_a(T) = f(M, R, \theta) + \delta B_e + \delta W_{es} \tag{2.1}$$

where $lnS_a(T)$ is the natural logarithm of the observed spectral acceleration at a recording station for period T; $f(M, R, \theta)$ is the prediction from a ground motion model and is modeled as a fixed-effect. M is magnitude of the earthquake, R is the source-to-site distance and θ is an array of other predictor variables and model coefficients; δB_e is the between-event residual, also known as the inter-event residual; and δW_{es} is the within-event residual, or intra-event residual. Both δB_e and δW_{es} are assumed to be independent random effects, modeled as normally distributed random variables with zero mean and a standard deviation of τ and ϕ , respectively. The δW_{es} term could be further partitioned to:

$$\delta W_{es} = \delta S2S_s + \delta WS_{es} \tag{2.2}$$

where $\delta S2S_s$ is the site term or the average error in prediction at a site, after removing the effects of earthquake, and the δWS_{es} is the site-and-event corrected residual. Again, both

these effects are random effects, and are modeled as independent normal random variables with zero means and a standard deviations of ϕ_{S2S} and ϕ_{ss}, respectively. ϕ_{S2S} is known as the site-to-site variability, and ϕ_{ss} is known as the within-site variability or the single-station sigma. The notations used for different components of the error term are taken from Atik et al. (2010). While most of the current ground motion models use a residual decomposition approach similar to those shown in Equations 2.1 and 2.2, Stafford (2014) showed that there are shortcomings with the use of such linear additive random effects, and that by changing the variance structure for random effects, the results of the regression can change. They recommended the use of a more complex treatment of mixed-effects models.

2.3.3 Site-effects in GMPEs and building codes

The effects of shallow surface deposits are captured by recent GMPEs using simplified parameterizations of the soil profile. The most common parameter used for this purpose is V_{s30} or the time-averaged shear-wave velocity in the top 30 m of the soil. Other parameter used in recent GMPEs along with V_{s30} is the depth to a stiff layer ($Z_{1.0}$). However, some researchers argue against the validity of such a parameter as a proxy for site effects (Mucciarelli and Gallipoli, 2006). Topographic effects are also site-effects, but they remain unaccounted for in the predictions due to a lack of proper quantification. The only codes that have included the topographic effects are the French building code (Bard et al., 1995), followed by the European building codes (Eurocode, 1994). These codes recommended factors to modify the design spectra; however these factors were frequency-independent and were not validated with the recorded data. A few researchers have proposed topographic amplification factors based on numerical studies (e.g., Assimaki et al., 2005; Bouckovalas and Papadimitriou, 2005; Maufroy et al., 2015). However, like the European codes, these factors have also not been validated with recorded data.

2.4 Concluding Remarks

In light of our current review, we conclude that even though there is a multitude of evidence suggesting that topographic amplifications exist in the field and that they are prominent for frequencies of engineering interest, we do not yet have a proper quantitative tool to account for these effects in the predictions. The magnitudes of predicted amplifications from a numerical analysis are typically lower than those observed in the recorded ground motions. 3D numerical analyses have managed to bridge some of the gap between the observed and predicted amplifications. A few studies have proposed factors for topographic amplifications; however they are based purely on numerical simulations, and are not verified with recorded data. To address this problem, we study large database of recorded ground motions and propose empirical models that can be used to predict topographic effects at a site..

Chapter 3

An empirical model to predict topographic effects in strong ground motion using California small to medium magnitude earthquake database

Adapted from Rai, M., A. Rodriguez-Marek and A. Yong (2015). An empirical model to predict topographic effects in strong ground motion using California small to medium magnitude earthquake database. *Earthquake Spectra (accepted for publication)*

3.1 Abstract

We develop a model to predict the effects of topography on earthquake ground motions using a database of small to medium magnitude earthquake from California. The proposed model

relies on a parameter called relative elevation that quantifies topography using the elevation of a site relative to its surroundings. We also investigate an alternative parameterization of topography called smoothed curvature. We study the bias in the residuals from the Chiou et al. (2010) ground motion model with respect to these parameters and fit a model to remedy those biases. We then compare these models by assessing their goodness of fit to the data. The proposed model for topographic effects is intended as a correction to the Chiou et al. (2010) small to medium magnitude earthquake prediction model.

3.2 Introduction

Surface topographic features are known to affect earthquake ground motions (Bouchon, 1973; Boore et al., 1981; Bard, 1982). Typically, these motions tend to amplify near hills and ridges and de-amplify near valleys (e.g., Boore, 1972; Davis and West, 1973; Spudich et al., 1996). Numerous case studies from past earthquakes have shown that buildings located close to hills or ridges suffer more damage than those located at the base (Trifunac and Hudson, 1971; Celebi, 1987; Geli et al., 1988; Kawase and Aki, 1990; Hatzfeld et al., 1995; Bouchon and Barker, 1996). Examples of topography-induced amplifications can be seen in the Tarzana hill station recording from the 1994 Northridge earthquake, which recorded peak ground accelerations (PGA) close to 1.8 g, (Shakal et al., 1994; Bouchon and Barker, 1996; Spudich et al., 1996) and in the Pacoima Dam station recording with a PGA of 1.25 g during the 1971 San Fernando earthquake (Trifunac and Hudson, 1971).

Previous studies based on theoretical and numerical models indicate that topography influences the characteristics of both the intensity and frequency in ground motion (Davis and West, 1973; Rogers et al., 1974; Griffiths and Bollinger, 1979; Tucker et al., 1984; Geli et al., 1988; Bouchon and Barker, 1996; Assimaki and Gazetas, 2004; Meunier et al., 2008; Assimaki and Jeong, 2013). Topographic amplification usually occurs in a frequency band that depends on relief dimensions (e.g., Boore, 1972, 1973; Davis and West, 1973; Durand

et al., 1999). Most theoretical and numerical models however tend to under predict the observed amplifications in the field because they are mostly based on simplifying assumptions such as isolated 2D geometries, simple soil layering, and/or an elastic soil regime Geli et al. (1988). By contrast, there are many other factors that influence the magnitude of amplifications, such as the nature and direction of incident waves, slope inclination, soil stratigraphy, material damping, and sub-surface shear-wave velocity structure (e.g., Assimaki and Gazetas, 2004; Hartzell et al., 2014; Burjánek et al., 2014), all of which can be challenging to estimate accurately.

Despite the aforementioned studies, the current state of knowledge for quantifying topographic effects remains nascent. No framework currently exists to include these effects for engineering purposes. As a direct consequence, predictive equations and design codes may be associated with a greater uncertainty or a larger standard deviation and may underestimate the intensity of actual ground shaking beneath structures located on hills and ridges.

In this paper we present a ground motion model that accounts for the effects of topography on spectral accelerations, based on records from the California small-to-medium magnitude (SMM) earthquake database published by Chiou et al. (2010). To quantify topography, we use two geomorphometric parameters namely smoothed curvature and relative elevation. Using an updated compilation of the shear-wave velocity data (Louie, 2005, 2007; Louie et al., 2011; Ancheta et al., 2014; Yong et al., 2013), we perform residual analyses to estimate misfits or biases in the SMM records with respect to these topographic parameters. A highlight of this study is that it inherently separates the effects of shear-wave velocity (as parameterized by the average shear-wave velocity in the upper 30 m, or V_{s30}) from the effects of topography, unlike site-specific studies that cannot isolate the effects of topography from those due to changes in shear-wave velocity.

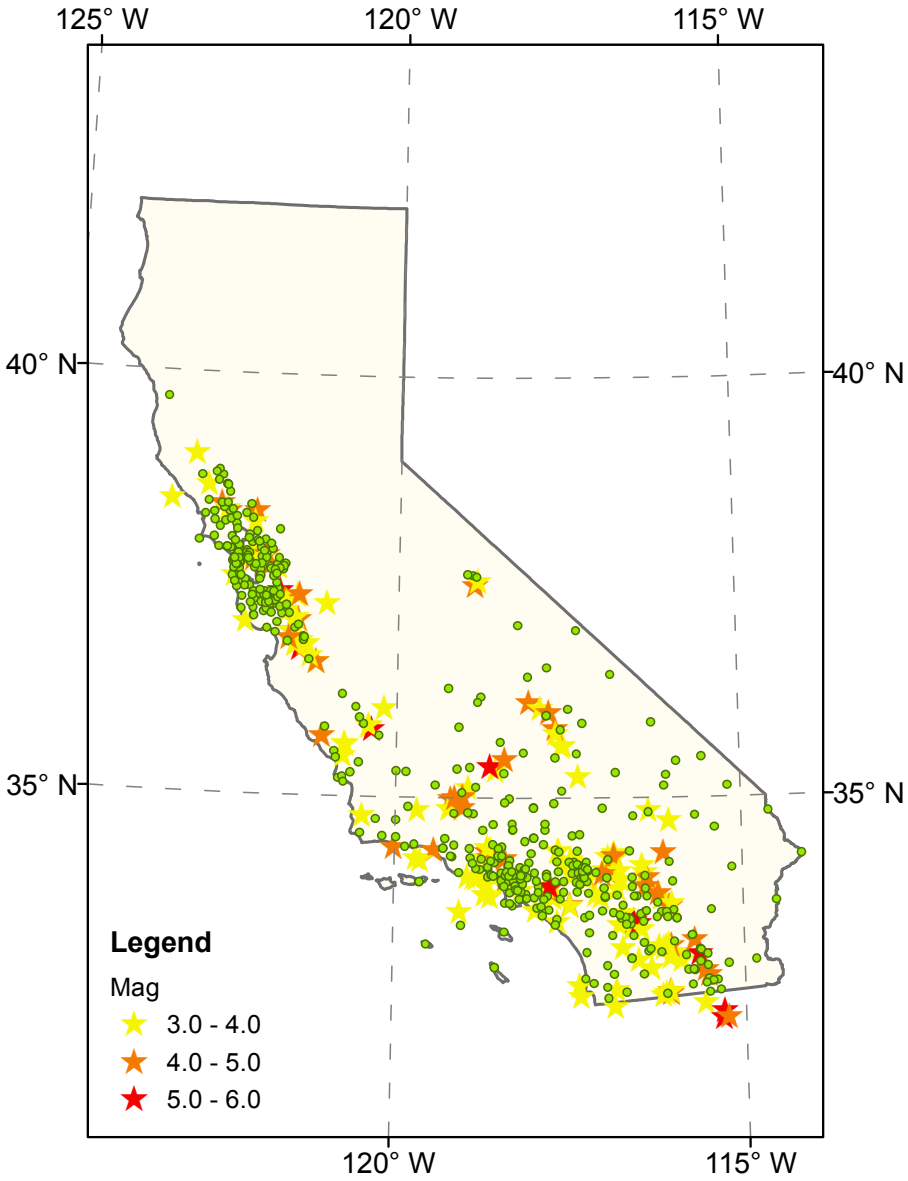

Figure 3.1: Locations of ground motion recording stations (green dots) in the SMME database used for this study are shown. Only stations with three or more recordings were selected resulting in a total of 459 stations.

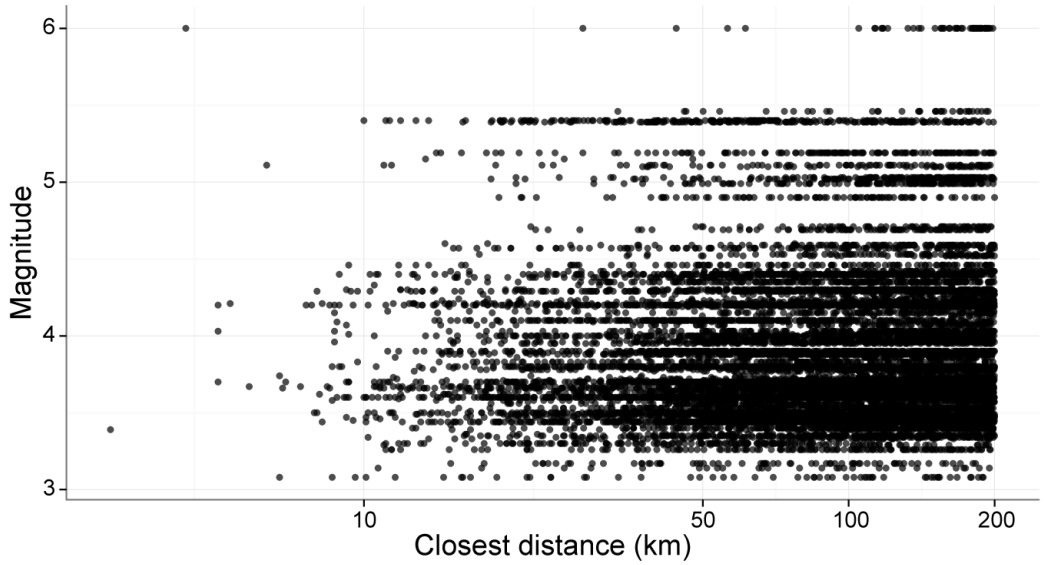

Figure 3.2: Magnitudes versus closest distance for records in the SMM database for stations with three or more recording.

3.3 Ground motion data

The small-to-medium magnitude earthquake database contains recordings of 188 earthquakes in California with magnitudes ranging from 3 to 6, recorded at 802 unique stations (Chiou et al., 2010). Multiple recordings at each station are important to this study because these observations allow robust estimates of the systematic residuals at each site ($\delta S2S_s$), following the notation in Atik et al. (2010). Here, we only utilize stations that recorded at least three events, reducing the number of useable stations to 459 (Figure 3.1). We consider recordings for peak ground acceleration (PGA), and 5%-damped pseudo spectral acceleration (PSA) for periods of 0.3 s and 1 s and utilize the ground motion prediction equation (GMPE) of Chiou et al. (2010). The magnitude-distance distribution of the recordings used in the study is shown in Figure 3.2.

The SMM database includes V_{S30}, the time-averaged shear-wave velocity in the top 30 m of soil beneath the recording station, and $Z_{1.0}$, the depth to the soil layer with 1 km/s shear-wave velocity. These parameters are used for constraining site amplification.

The current version of the database contains many V_{S30} values that are inferred based on surface geology as a proxy since actual V_{S30} measurements were not available. Recently, new measurements of V_{S30} became available for various stations in California (Louie, 2005, 2007; Louie et al., 2011; Ancheta et al., 2014; Yong et al., 2013). The V_{S30} values obtained by Louie (2005, 2007); Louie et al. (2011) were based on the standalone refraction microtremor ($ReMi^{TM}$) technique, whereas Yong et al. (2013) used multiple combinations of active and passive, body and surface wave techniques. The V_{S30} values compiled by Ancheta et al. (2014) included values based on a mixture of down-hole and other geophysical methods. Amongst the three data sources, results from Yong et al. (2013) are believed to provide the most robust measurements of V_{S30}; hence, when V_{S30} values are available from multiple sources, the first preference is to apply results from Yong et al. (2013), followed by Ancheta et al. (2014) and then Louie (2005, 2007); Louie et al. (2011). The previous V_{S30} assignments by Chiou et al. (2010) are substituted with V_{S30} from the highest-quality measurement that is closest to the station, thus, we only consider measurement sites within a distance of 300 m from a station (Yong et al., 2013). The search distance of 300 m is chosen to verify that the measurement site was indeed close enough to the station, such that the character of the underlying geology remains representative of the station. Figure 3.3 compares the measured V_{S30} values with the V_{S30} from the original Chiou et al. (2010) database. Still, not all stations have measured V_{S30}. Thus, following Chiou et al. (2010), we retain their proxy-based V_{S30} for stations where no measurements are available.

3.4 Methodology

3.4.1 Topographic parameterization

Some of the most common geometric parameters for terrain classification include elevation, slope, curvature, and surface roughness (Goldgof et al., 1990; Irvin et al., 1997; Burrough et al., 2000; MacMillan et al., 2000; Iwahashi and Pike, 2007). Apart from these parameters,

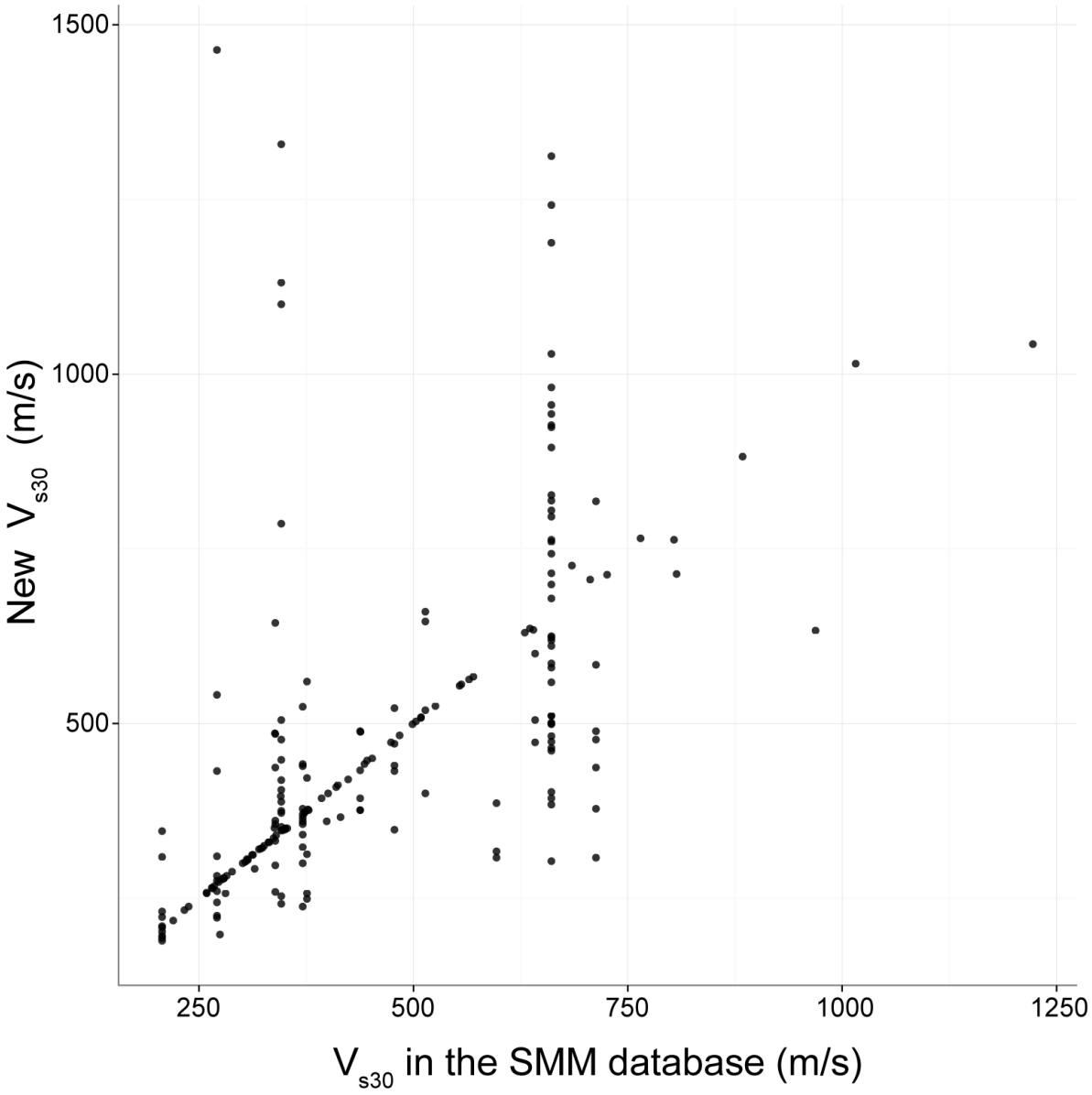

Figure 3.3: Measured V_{S30} values (Louie, 2005, 2007; Louie et al., 2011; Ancheta et al., 2014; Yong et al., 2013) are plotted against the values in the Chiou et al. (2010) database for the 208 stations where new measurements are available.

scale is also an important component of land element classification (Schmidt and Hewitt, 2004). Most terrain classification methods use combinations of these parameters to classify a terrain into a finite number of terrain types or classes. One such classification scheme was employed by Iwahashi and Pike (2007) to iteratively classify landforms into 8, 12, or 16 terrain types using parameters such as slope, surface texture, and local convexity. Their algorithm divided a region into terrain classes by iteratively combining the binarized images derived from the three previously listed variables. The dividing threshold for each variable was its mean value. As part of an earlier study, we did not find correlation between the terrain classes in the Iwahashi and Pike (2007) based classification map and site residuals from the SMM dataset (Rai et al., 2012). In this study, we explore two other parameters derived from the 30 m resolution elevation data in the form of a 1-arc-s Digital Elevation Model (DEM) of California downloaded from the USGS seamless server (`http://viewer.nationalmap.gov/viewer/`, accessed on 11/17/2013). The elevation data is typically a raster data, which consists of rows and columns of grid cells, with each cell storing the elevation value at the center of the grid. These rasters are used to compute the parameters for each station in ArcGIS (ESRI, 2011).

The first parameter we consider is relative elevation (H). Also referred to as topographic position index, this parameter has been used by other researchers, particularly in watershed studies (Guisan et al., 1999; Jones et al., 2000). The parameter H is computed for each grid cell of the raster by taking the difference between the elevation raster and the mean elevation raster. The mean elevation raster is defined as a raster that contains, for each cell, the mean value of the elevations within a specified neighborhood of that cell. Mean elevations are computed using the focal statistics tool in ArcGIS. In this study, we use a circular neighborhood to computed mean elevation. The diameter of the circle used for calculating the mean elevation is referred to as the scale (d) and the relative elevation calculated at a scale d is referred to as H_d. A positive H_d value at a cell location indicates that the cell has a higher elevation than the mean elevation in the neighborhood, a negative H_d value indicates that cell has a lower elevation than the mean elevation in the neighborhood, and an H_d value

of zero for the cell indicates that the cell is located in a region of constant slope. For this reason, H_d is effective in highlighting features such as ridges ($H_d > 0$), valleys ($H_d < 0$), and plains/uniform slopes ($H_d = 0$). Like elevation, H_d is inherently continuous and varies smoothly from one cell to another over a terrain. Continuous values of H_d can be separated into three discrete zones or classes using a positive and a negative threshold parameter: *high*, where H_d is greater than the positive threshold setting; *intermediate*, where H_d is between the positive and the negative thresholds; and, *low*, where H_d is less than the negative threshold. Note that this classification depends on the values selected for both scale, which is used to compute mean elevation, and the threshold parameter used to set class boundaries. These selections affect the classifications of a region in a terrain. Thus, we consider various scales and threshold values to determine the effects of these two parameters on the classification of the terrain. More recently, Burjánek et al. (2014) used this parameter to empirically study the effects of surface topography on ground motions. Their study was based on the analysis of recorded ground motions at sites located in Switzerland and Japan, and showed that the strong amplifications observed at sites with significant topography was mainly controlled by the subsurface velocity structure, and not so much by the shape of the topography. This analysis was based on sites in Switzerland and Japan and may be representative of those regions.

The second parameter we consider is smoothed curvature (C). We define smoothed curvature as the curvature of the mean elevation raster. To compute the smoothed curvature raster, we use the curvature tool in ArcGIS with the mean elevation raster as an input. Recent studies using 3D numerical simulations have shown that smoothed measures of surface curvature strongly correlate with the median topographic amplifications (Maufroy et al., 2015). The C parameter depends on the size of the neighborhood used to compute the mean elevations. Following the convention for H_d, we designate smoothed curvature as C_d where d is the diameter of the smoothing window used for computing mean elevations.

For each cell corresponding to the site of a station, we compute the C_d values at scales between 120 m to 1200 m at intervals of 60 m and H_d values at scales of 250 m to 3000 m at

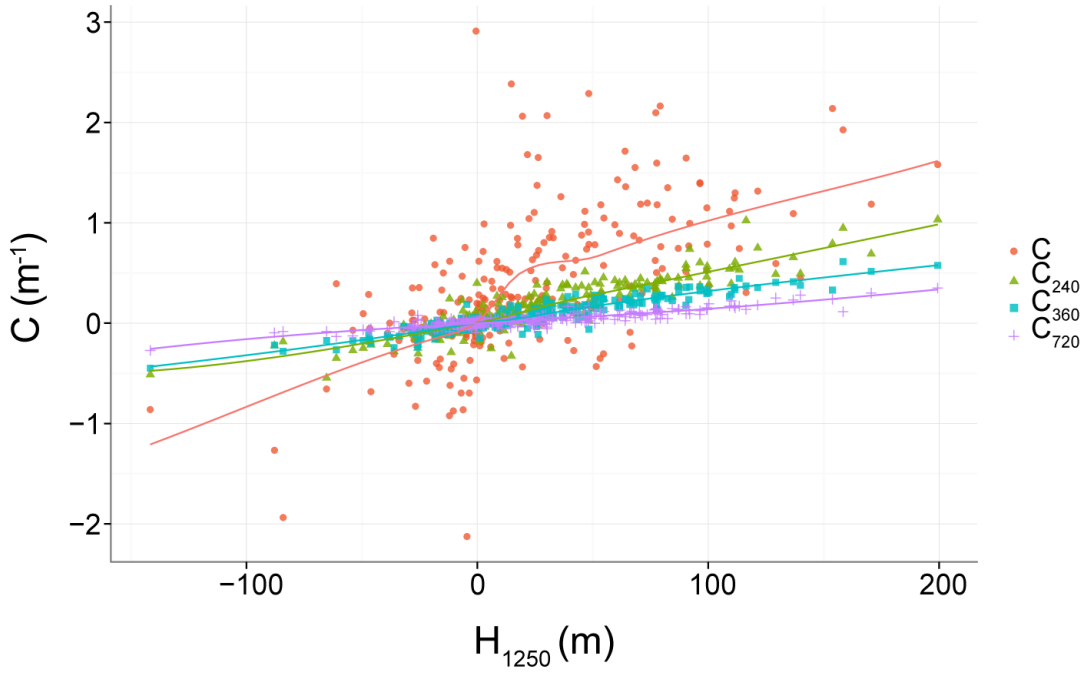

Figure 3.4: Local curvature (C) and smoothed curvatures at a scale of 240 m, 360 m and 720 m plotted against H_{1250} for the stations used in the study

intervals of 250 m. Although these parameters are quite different by definition, we observe a fairly high correlation between the two parameters. Figure 3.4 shows C_d values at each station plotted against H_{1250}. The figure indicates that the smoothed curvature at given scales are linearly correlated to relative elevation at 1250 m. The strongest correlation of C_d with H_{1250} is for C_{360}, where correlation coefficient $\rho = 0.94$. Figure 3.5 shows the H_{1250} raster and C_{360} raster calculated from the elevation raster. Note the strong resemblance between the H_{1250} raster and the C_{360} raster. This degree of similarity suggests that the two variables are very similar at some scale(s) (Figure 3.4); thus, they express similar information about the terrain. Although H_d and C_d are highly correlated, we explore both parameters in our analysis.

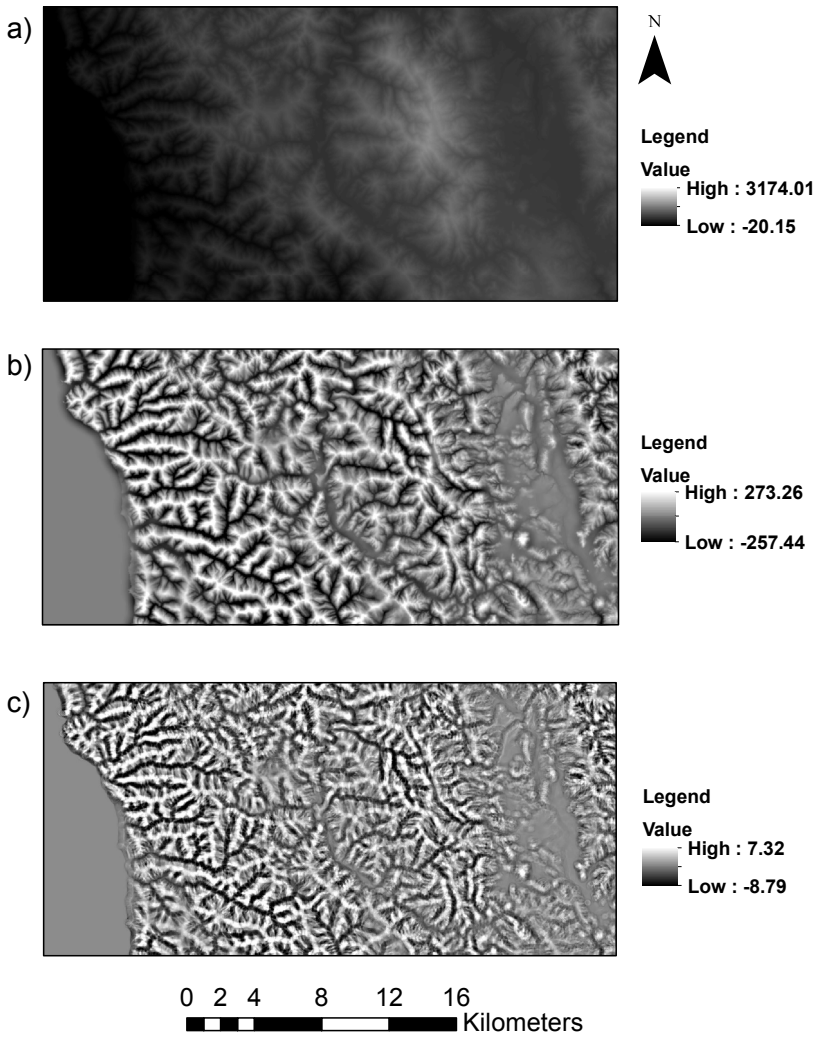

Figure 3.5: A comparison of a) elevation raster, b) H_{1250} and c) C_{360}. Note the similarity between the H_{1250} and c) C_{360}

3.4.2 Residual analysis

Ground motion at a site is typically predicted using a ground motion prediction equation (GMPE) (Kramer, 1996). For the purpose of this study, spectral acceleration (S_a) is used as the measure of intensity of the ground motion. A typical ground motion prediction equation is modeled as:

$$lnS_a(T) = f(M, R, \theta) + \delta B_e + \delta W_{es} \tag{3.1}$$

where $lnS_a(T)$ is the natural logarithm of the observed spectral acceleration at a site for period T; $f(M, R,)$ is the median prediction from a ground motion model where M is magnitude of the earthquake, R is the source-to-site distance and θ is an array of other predictor variables and model coefficients; δB_e is the between-event residual, also known as the inter-event residual; and δW_{es} is the within-event residual, or intra-event residual. δB_e is the average residual of the GMPE prediction from a given earthquake, e, which is common at all sites recording that event. δW_{es} is the shift of an observation at a station, s, from the median prediction for an earthquake, e, over all stations at which the event was recorded. Both δB_e and δW_{es} are modeled as random normal variables with zero means and standard deviations of τ and ϕ, respectively. The δW_{es} term can be further partitioned into two components:

$$\delta W_{es} = \delta S2S_s + \delta WS_{es} \tag{3.2}$$

where $\delta S2S_s$ is the site residual and δWS_{es} is the site-and-event corrected residual. The term $\delta S2S_s$ is the repeatable site residual that is common for each ground-motion recorded at a site. Both terms are modeled as random normal variables with zero means and standard deviations of ϕ_{S2S} and ϕ_{ss}, respectively. The ϕ_{S2S} term is defined as the site-to-site variability and ϕ_{SS} is known as the event-corrected single-station variability. As topographic effects are

site-specific, the bias in residuals from topography is expected to be present in the site residual.

In this study, we use the Chiou et al. (2010) model as our reference model to study topographic effects. We elected to use the Chiou et al. (2010) functional form over developing a new GMPE because we wish to leverage the predictive capabilities of an established model. Furthermore, we do not seek to create a new GMPE, but to understand topographic effects. The general form of the Chiou et al. (2010) ground motion model is as follows:

$$ln(y) = ln(y_{ref}) + \phi_1 f_1(V_{s30}) + \phi_2 f_2(V_{s30}, y_{ref}, \delta B_e; \phi_3, \phi_4) + f_3(Z_{1.0}; \phi_5, \phi_6, \phi_7, \phi_8) + \delta B_e + \delta W_{es} \tag{3.3}$$

where f_1, f_2, and f_3 are given by:

$$f_1 = min(ln(\frac{V_{s30}}{1130}), 0) \tag{3.4}$$

$$f_2 = (e^{\phi_3(min(V_{s30},1130)-360)} - e^{\phi_3(1130-360)}) ln(\frac{y_{ref} e^{\delta B_e} + \phi_4}{\phi_4}) \tag{3.5}$$

$$f_3 = \phi_5(1 - \frac{1}{cosh(\phi_6 max(0, Z_{1.0} - \phi_7))}) + \frac{\phi_8}{cosh(0.15 max(0, Z_{1.0} - \phi_7))} \tag{3.6}$$

and y is the observed ground motion intensity at a site; y_{ref} is the median predicted ground motion at the reference site; $\phi_1 - \phi_8$ are the ground motion model coefficients. Note that if the ground motion model is refitted using the updated V_{S30} values, we can expect the values of some of the coefficients in the model to change; specifically the ones associated with functions that contain the V_{S30} parameter, that are functions f_1 and f_2 . This may also result in a change in the distribution of the residual components δB_e and δW_{es} at some sites. As we are interested in the site terms (that are computed from the δW_{es} term) for the study

of topographic effects, it is important that we refit the SMM model for the updated V_{S30} values and re-compute the residual components. Our preliminary analyses indicate that the changes in V_{S30} values at some of the sites did not result in significant changes in the δB_e term. This is likely because the updated V_{S30} values did not cause a systematic bias in the site terms. Moreover, the $ln(y_{ref})$ and the $f_3(Z_{1.0})$ terms do not depend on the V_{S30} value. We thus use the same functional form and coefficients for these terms in the new model. The only terms that change with the V_{S30} update are f_1 and f_2 and thus we refit coefficients ϕ_1 and ϕ_2 in the new model. The coefficients ϕ_3 and ϕ_4 have higher order effects and make the regression problem non-linear. Furthermore, some of the changes in model predictions, due to changes in the values of ϕ_3 and ϕ_4, can be absorbed by varying the values of ϕ_1 and ϕ_2. Thus, for simplicity, we fix the coefficients ϕ_3 and ϕ_4 to values estimated by Chiou and Youngs (2008) while performing the regression. Note that these values did not change in the updated results provided in Chiou et al. (2010). Subtracting lny_{ref}, δB_e, and $f_3(Z_{1.0})$ from both sides of Equation 3.3 and substituting δW_{es} from Equation 3.2 gives us:

$$\delta W^* = \phi_1 f_1(V_{s30}) + \phi_2 f_2(V_{s30}, y_{ref}, \delta B_e; \phi_3, \phi_4) + \delta S2S_s + \delta WS_{es} \tag{3.7}$$

where

$$\delta W^* = ln(y) - ln(y_{ref}) - \delta B_e - f_3(Z_{1.0}) \tag{3.8}$$

After evaluating δW^*_{es} for each recording, we recomputed coefficients ϕ_1, ϕ_2, and $\delta S2S_s$, δWS_{es} using the updated V_{S30} values. Figure 3.6 shows a plot of $\delta S2S_s$ from Equation 3.7 with H_{1250} for PGA and PSA at T = 0.3 s and T = 1 s. Note that for periods of 0.3 s and 1 s, the average site term is positive for higher values of H_{1250}, negative for lower values of H_{1250}, and has a value close to zero for intermediate values of H_{1250}. Similar trends are observed with relative elevations at all other scales. Figure 3.7 shows the variation of mean site residuals for each topographic class with scale. There are clear biases in the mean site residuals for some scales and periods but only for the *high* and *low* topographic classes.

These trends possibly suggest amplifications and de-amplifications in ground motions within the respective classes. The apparent topographic biases are greater for *low* sites for smaller scales, especially at a spectral period of 1 s, however these values are poorly constrained, as indicated by a higher standard error.

Figure 3.8 shows a plot of sample standard deviation of the site terms (ϕ_{S2S}) or the site-to-site variability for each of the topographic classes at different scales. A higher variance in the site terms (i.e., larger ϕ_{S2S}) can be expected for *high* and *low* sites compared to *intermediate* sites in the period range where topography significantly affects ground motions. A probable reason for such an observation could be that our proposed classification scheme groups features together and ignores complicated aspects of ridge and/or valley geometry, such as the direction and dimensions of the ridge, steepness of the slope, and other such details. Omitting these factors that contribute to the overall site response can increase the uncertainty in the response of these features to an earthquake. *Intermediate* sites, on the other hand, are often flat terrains with relatively smaller slopes and thus the sites classified as *intermediate* are relatively less likely to show the same variability in their site response. In this study, we observe that ϕ_{S2S} is higher for *low* and/or *high* sites than the *intermediate* sites, but only at some scales and some periods (see, for example, Figure for T = 0.3 s). Topographic effects are also known to be frequency dependent, where the frequency depends on the dimensions of the topographic feature (e.g., Geli et al., 1988), which can explain why this trend is not observed across all scales and periods.

Even though the trends in site residuals are visually evident, it is important to determine whether the difference in the mean site terms is statistically significant. To test this hypothesis, we apply the analysis of variance (ANOVA) method (Rice, 1995) to determine if the mean $\delta S2S_s$ are significantly different for the three topographic classes. Given that the classification of a station as *high*, *low*, or *intermediate* depends on the choice of scale and threshold, the results of ANOVA also depend on the scale and threshold chosen. We thus repeated this analysis for all other combinations of scales and thresholds. The p-values from ANOVA are presented in graphical format in Figure 3.9. Low p-values indicate that there

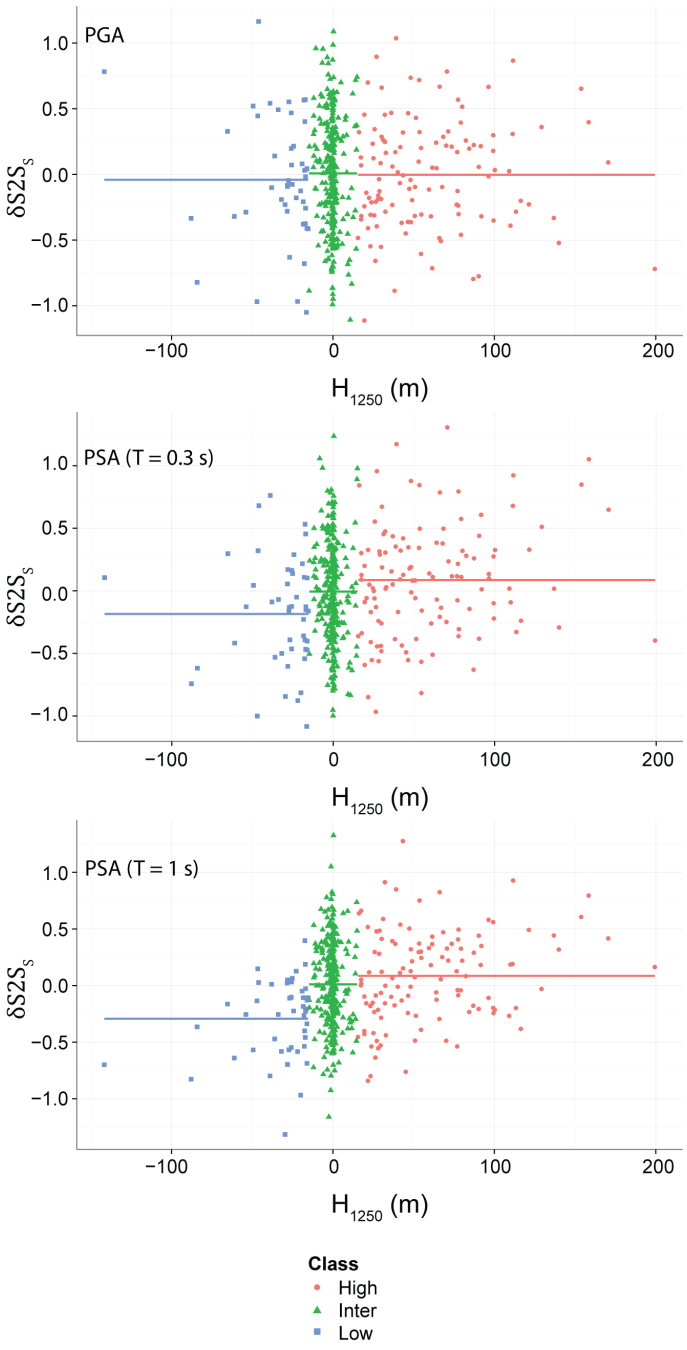

Figure 3.6: Site residuals ($\delta S2S_s$) from Equation 3.7 are plotted against H_{1250} Mean site residuals for the three classes of H_d: *high*, *low*, and *intermediate*, using a threshold of 15 m are shown as straight lines.

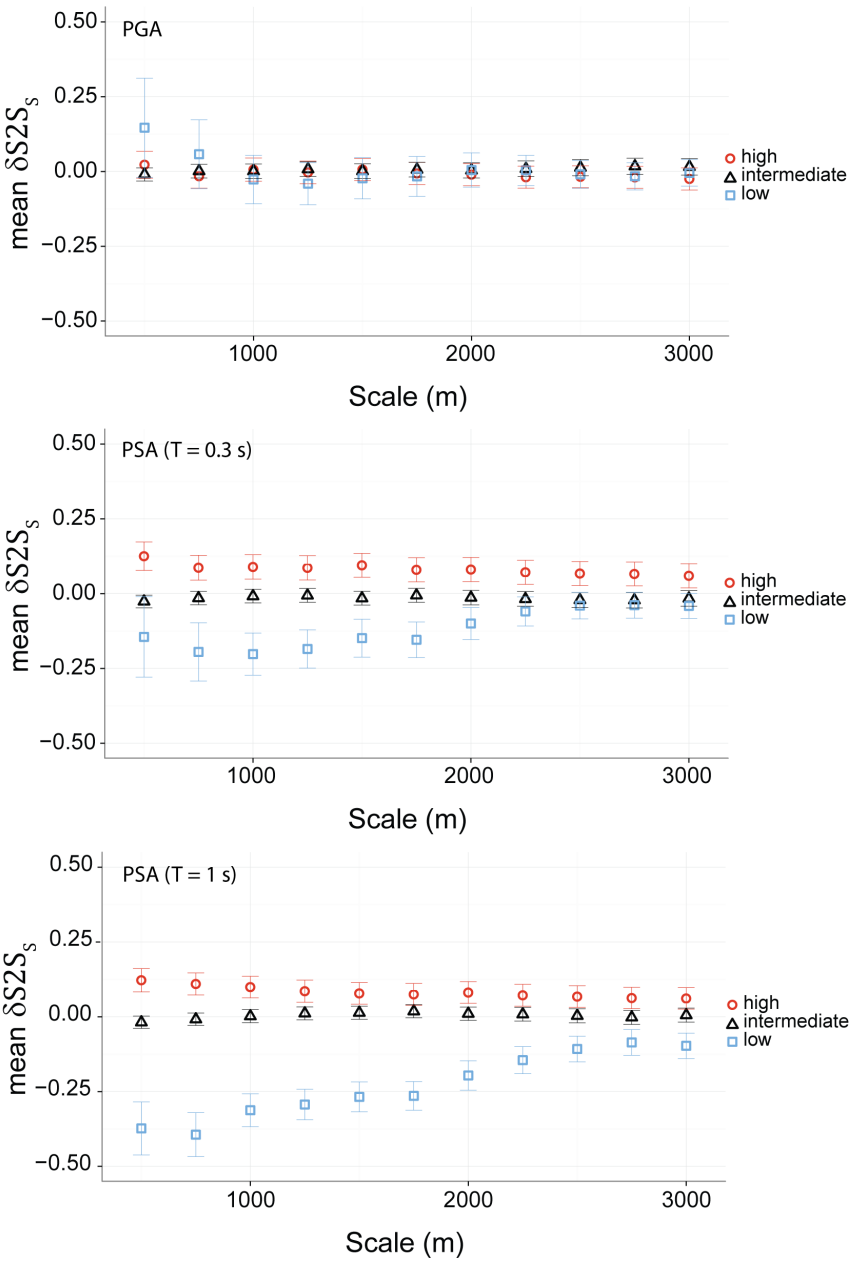

Figure 3.7: Mean $\delta S2S_s$ for each site class (*high/intermediate/low*) along with a +/- 1 standard deviation error is shown for scales from 500 m to 3000 m. A threshold of 15 m is used for classifications.

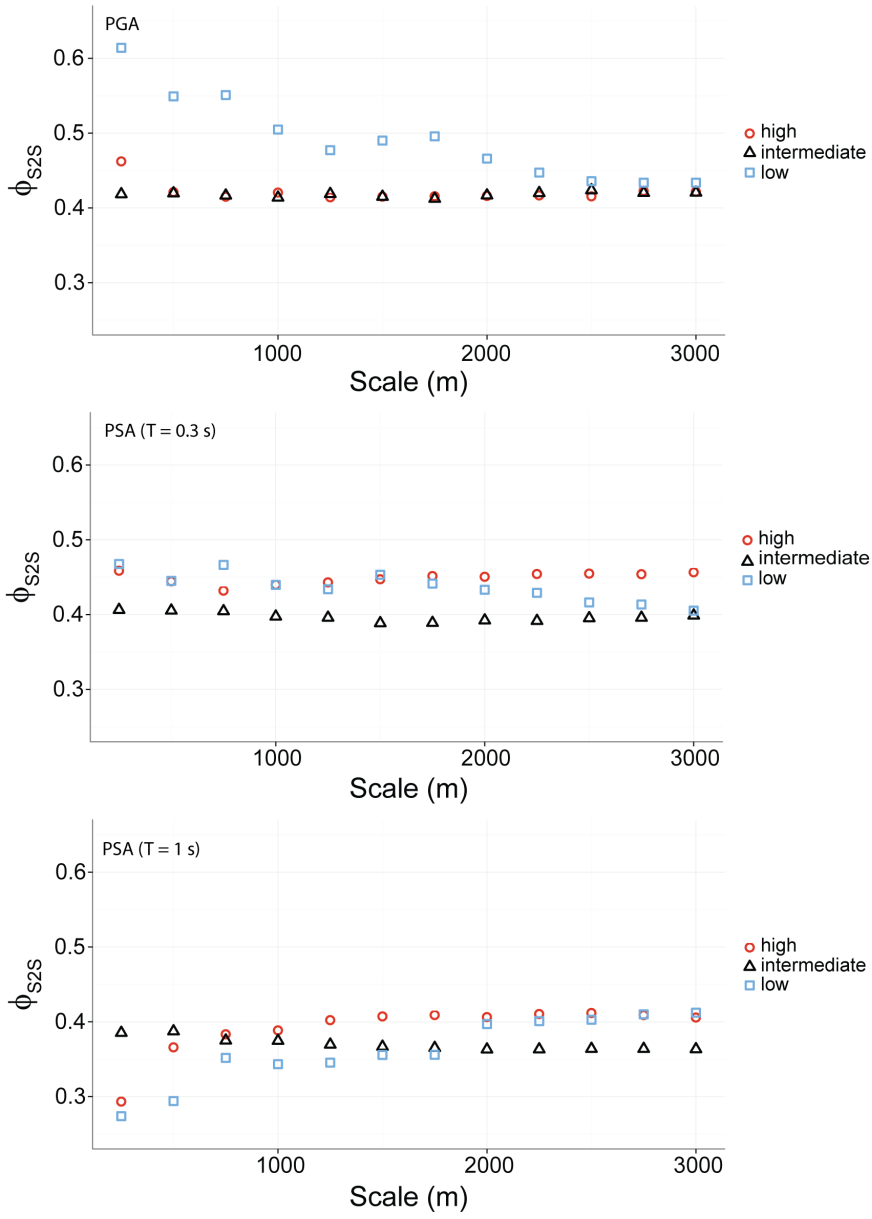

Figure 3.8: ϕ_{S2S} for each site class (*high*, *low*, and *intermediate*) is shown for all the scales from 500 m to 3000 m. A threshold of 15 m is used for classifications.

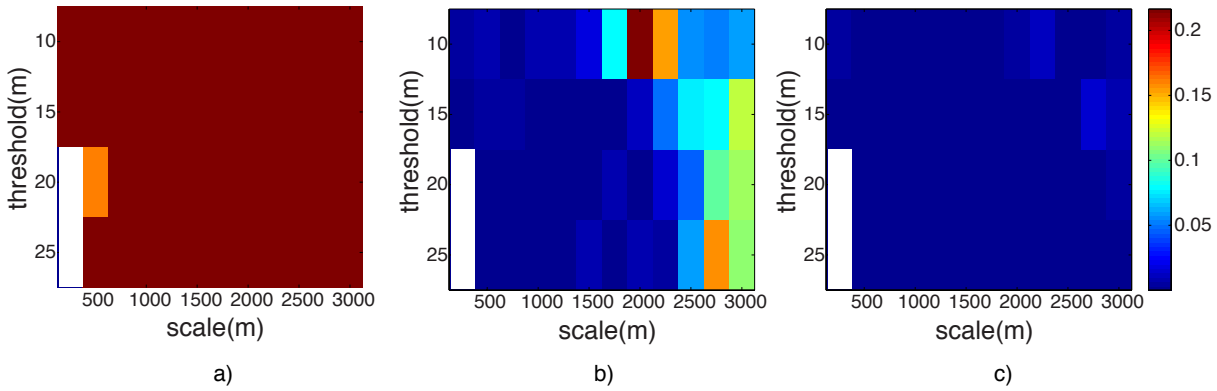

Figure 3.9: P-values from variance analysis of $\delta S2S_s$ are presented in a graphical format for a) PGA, b) T = 0.3 s and c) 1 s. P-values of 0.05 and less indicate that the means site residuals for the three topographic class *high*, *low*, and *intermediate* are significant for that scale and threshold pair. This analysis is not possible on a scale of 250 m and thresholds of 20 m, 25 m as for this combination of scales and threshold, not all the three classes were populated by stations (hence the white regions in the plot).

is a higher chance that the mean residual for all the three topographic classes are not equal. Note that for PGA, there is no statistically significant differences between classes at almost all scales and thresholds, but for T = 0.3 s and for T = 1 s, there are statistically significant differences in the mean at various scales and thresholds. The results from the ANOVA test can be used to select a scale and a threshold that ensures a high probability that the mean residuals of the three topographic classes would be significantly different. It can be seen in the figure that various combinations of scales and thresholds have p-values of less than 0.05 for T = 0.3 s and for T = 1 s, and one could potentially choose any of those pairs. We also observed that the correlation coefficients between site residuals at T = 0.3 s and the relative elevation parameter are larger for smaller scales (d = 250 m to 500 m) and at T = 1 s, the correlation is larger for larger scales (d = 1250 to 1500). However, the differences in terms of a predictive model are small, and therefore, for simplicity we chose to use a single scale for both spectral periods. We select 1250 m as our scale and 15 m for threshold because these parameters account for at least 46 stations in each category, which allows for a stable estimate of the average site terms within each class.

3.5 Regression/Model development

In the previous section we demonstrated that spectral acceleration residuals, at some periods, show a bias that is remedied by the topographic parameter. To model this dependency, a generic topographic term is added to Equation 3.7. The basic form of the resulting regression equation is given as:

$$\delta W^* = f_{topo} + \phi_1 f_1(V_{s30}) + \phi_2 f_2(V_{s30}, y_{ref}, \delta B_e; \phi_3, \phi_4) + \delta S2S_s + \delta WS_{es} \tag{3.9}$$

where ϕ_1 and ϕ_2 are the coefficients for functions f_1 and f_2, V_{S30} is the updated value of V_{S30} , and f_{topo} is a function of a topographic parameter. The topographic parameters used here are H_{1250} and C_{360}. Even though these variables are highly correlated, we consider both of them to see which of these parameters fit the data better when we use the same functional form for both of them. To model topographic effects, we consider two functional forms: linear and the multi-linear. The linear functional form varies linearly with the topographic parameter. The multi-linear functional form has three constant levels of amplification for three topographic classes: *high*, *low*, and *intermediate*, with two transition zones connecting them in order to keep the predictions continuous across the class boundaries. To classify stations as *high*, *low*, and *intermediate*, we used a threshold of 15 m with the parameter H_{1250} and a threshold of 0.04 with the parameter C_{360}. The threshold for H_{1250} is obtained through ANOVA (see previous section). The C_{360} threshold is approximated using the relation between H_{1250} and C_{360} (Figure 3.4). As we are fitting two functional forms using two parameters, we have a total of 4 models. The descriptions of these models are given in Table 3.1. Note that model 1 in the table corresponds to Equation 3.7 that does not contain a term for topography, and that has been refitted using the updated V_{s30} values. We have included this model in the discussion so that it can be used as a base model to compare the performance of other models.

Table 3.1: y_N values for the models at periods of 0.3 s and 1 s.

Model No.	Model Type
1	No Topo Term
2	Linear H_{1250}
3	Multi-linear H_{1250}
4	Linear C_{360}
5	Multi-linear C_{360}

In this regard, we define a variable y_N (index N corresponds to model N from Table 3.1 such that:

$$y_N = \sqrt{(\phi_{S2S1})^2 - (\phi_{S2SN})^2} \tag{3.10}$$

where ϕ_{S2SN} is the ϕ_{S2S} value for model N. y_N , by definition, quantifies the reduction in the site to site variability when topography is accounted for in the model. We use Model 1 as a reference for this computation as it does not contain any topographic term. Any change in the site to site variability is then attributable to the introduction of a topographic term in the model. A higher y_N value indicates a greater reduction in the site to site variability and hence a model with a higher y_N should be preferred over other models. The y_N values are computed using the results from the regression and are presented in Table 3.2.

We find that all of the models lead to similar y_N values at both periods. It is difficult to discern which of these models fit the data better based just on the y_N values. However, we feel that models 3 and 5 may extrapolate better for the parameter values outside the range of values used to fit the model, as the model predicts constant values beyond a given threshold. Between models 3 and 5, we choose to focus the discussion on Model 3 because it is based on the H_{1250} parameter which is relatively easier to compute.

Table 3.2: Models used to capture topographic effects

Model No.	$y_N(0.3s)$	$y_N(1s)$
2	0.0822	0.1070
3	0.0790	0.1160
4	0.0838	0.1137
5	0.0740	0.1109

3.5.1 Proposed model

Our proposed model is given by:

$$ln(y) = f_{topo} + \phi_1 f_1 + \phi_2 f_2 + f_3(Z_{1.0}) + \delta B_e + \delta S2S_s + \delta WS_{es} \tag{3.11}$$

where

$$f_{topo} = \begin{bmatrix} c_{low} & H_{1250} < -40 \\ c_{int} & -15 < H_{1500} < 15 \\ c_{high} & H_{1500} > 40 \end{bmatrix} \tag{3.12}$$

with linear transition zones between H_{1250} of -40 m to -15 m, and between 15 m to 40 m. The values of the three coefficients shown in equation 3.12 are provided in Table 3.3. Figure 3.10 shows the function f_{topo} plotted against H_{1250} for T = 0.3 s and T = 1 s. For sites with low H_{1250}, our model predicts strong negative bias, which is controlled by very few data points (see Figure 3.6); hence, these values must be used with caution.

The c_{int} values in Table 3.3 appear to indicate a negative bias for the *intermediate* sites. However, most of this bias emerges from refitting coefficients ϕ_1 and ϕ_2 (Equation 3.9). Note that the original values of ϕ_1 were -0.6444, -0.799, and ϕ_2 were -0.2405, -0.0699 for spectral periods of 0.3 s, and 1 s, respectively. We have no reason to believe that the *intermediate* sites should be biased with respect to the original GMPE, given that a majority of the

Table 3.3: y_N values for the models at periods of 0.3 s and 1 s.

T (s)	c_{low}	c_{int}	c_{high}	ϕ_1	ϕ_2	ϕ_{s2s}	ϕ_{ss}
0.3	-0.2921	-0.1346	0.0654	-0.7398	-0.4333	0.4301	0.5130
1	-0.5911	-0.1211	0.0564	-0.9236	-0.3511	0.3874	0.4463

stations are in this class. What is more relevant for predictive purposes is the differences between the topographic classes, which can be computed from the difference between the c_{low} and c_{high} coefficients, and the c_{int} coefficients in Table 3.3. These differences are equivalent to amplification/deamplification values at *high* and *low* sites of 1.22 and 0.85 for T = 0.3, and 1.19 and 0.63 for T = 1 s, respectively.

3.5.2 Model diagnostics

Figure 3.11 shows $\delta S2S_s$ from the proposed model (Equation 3.11) plotted with H_{1250}. The absence of any trend suggests that the proposed model for topography properly accounts for its effects. Figure 3.12 shows the site terms ($\delta S2S_s$) from the proposed model plotted with the updated V_{S30} values. Note the absence of any significant trend, which indicates that the introduction of a topographic term did not introduce bias in the V_{S30} scaling term, suggesting that both effects are uncoupled. To further check for any unaccounted interaction between topography and V_{S30} , plots of $\delta S2S_s$ against H_{1250} for different V_{S30} bins are made (Figure 3.13). The plots lack any significant trend indicating that the model described by Equation 3.12 adequately accounts for both the effects of topography and V_{S30} together.

3.6 Conclusion

We examine the effects of topography on earthquake ground motion using the small to medium magnitude earthquake ground motion database from Chiou et al. (2010) and geo-

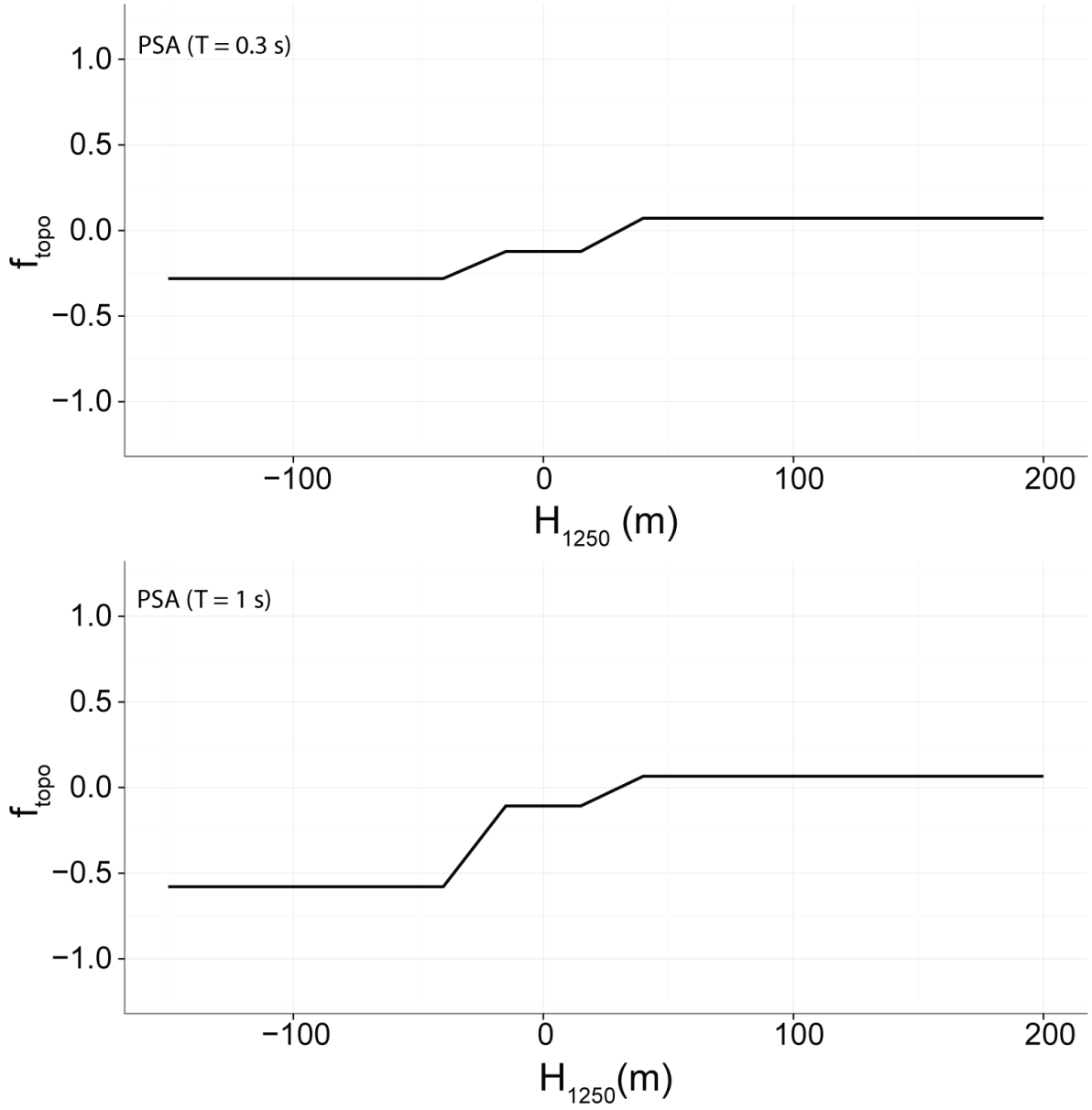

Figure 3.10: The multi-linear functional form shown in Equation 3.12 is plotted with H_{1250} for a) T = 0.3 s and b) T =1 s. The transition from *low* to *intermediate* occurs from x = -40 m to -15 m and from *intermediate* to *high* from x = 15 m to 40 m. No correction is proposed for the PGA.

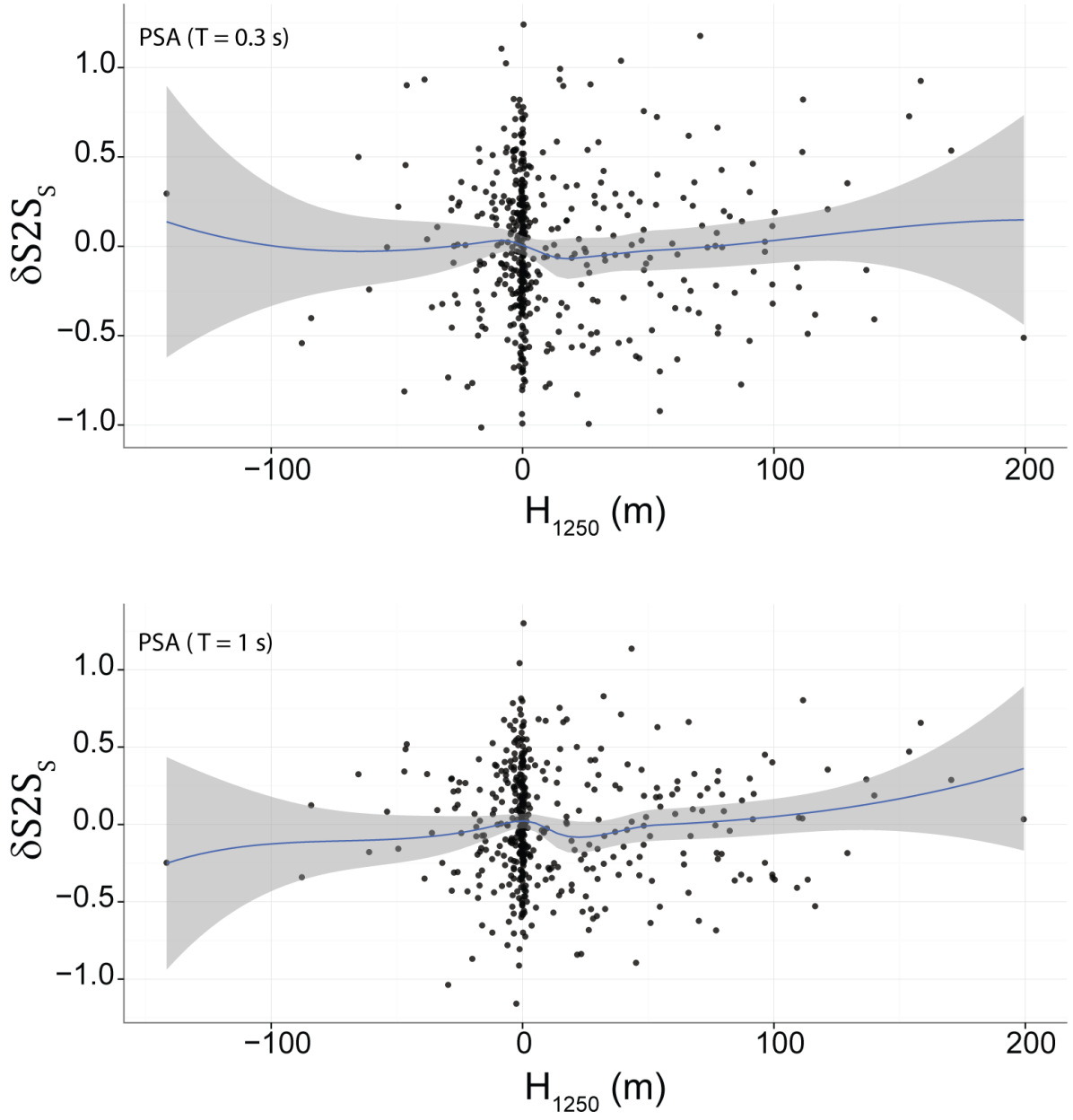

Figure 3.11: $\delta S2S_s$ terms from Equation 3.11 are plotted with H_{1250}. A moving average of the$\delta S2S_s$ terms computed using local regression (loess) is also shown. Most of the residual trend from Figure 3.6 is removed suggesting that Equation 3.12 accounts for most of the topographic effect in the ground-motion intensity.

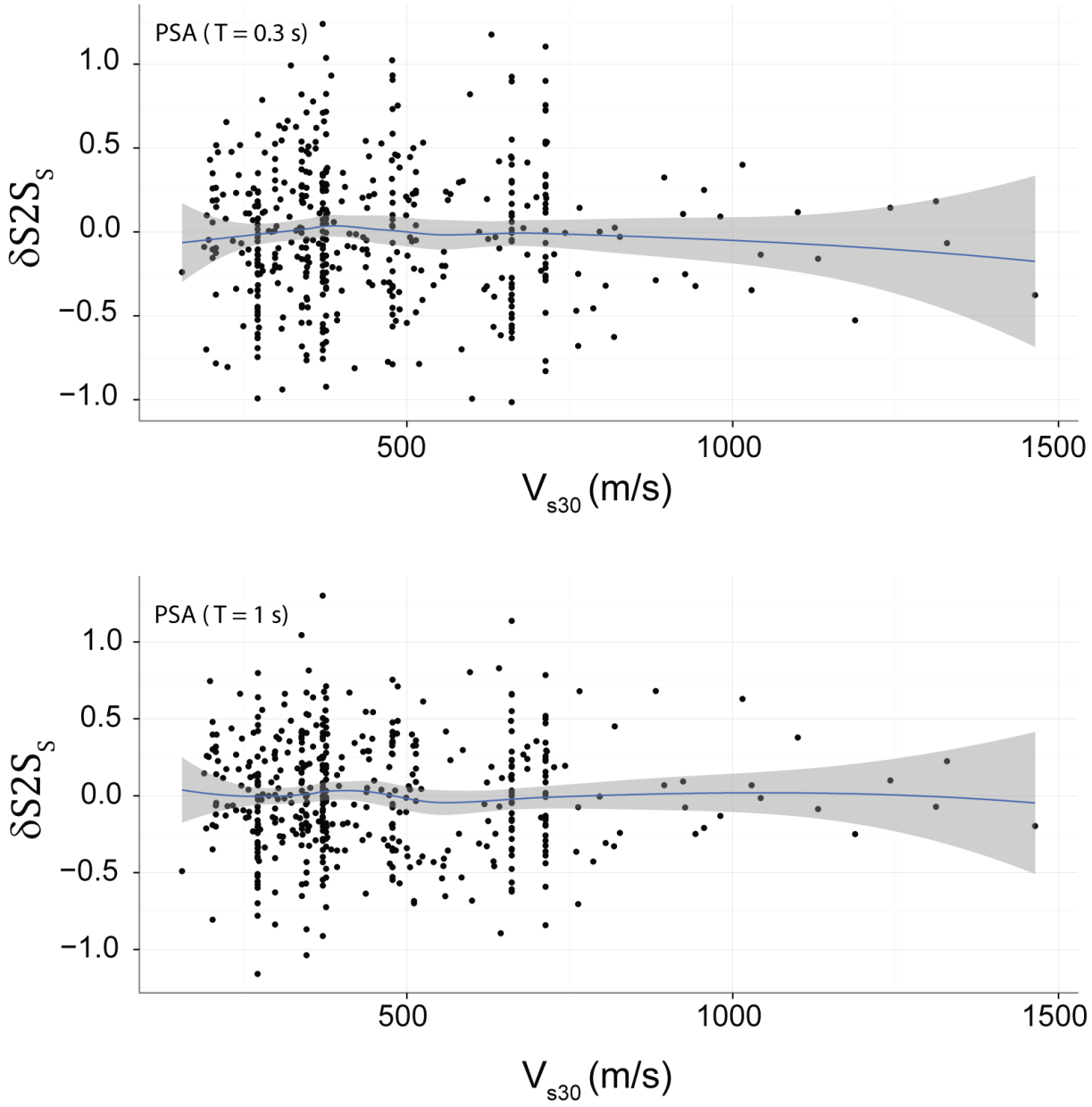

Figure 3.12: $\delta S2S_s$ terms from Equation 3.11 are plotted with V_{S30} values at the station. A moving average of the $\delta S2S_s$ terms computed using local regression (loess) is also shown. Absence of any trend indicates that using the (Chiou et al., 2010) functional form with the new V_{S30} values doesnt produce any bias in V_{S30} scaling.

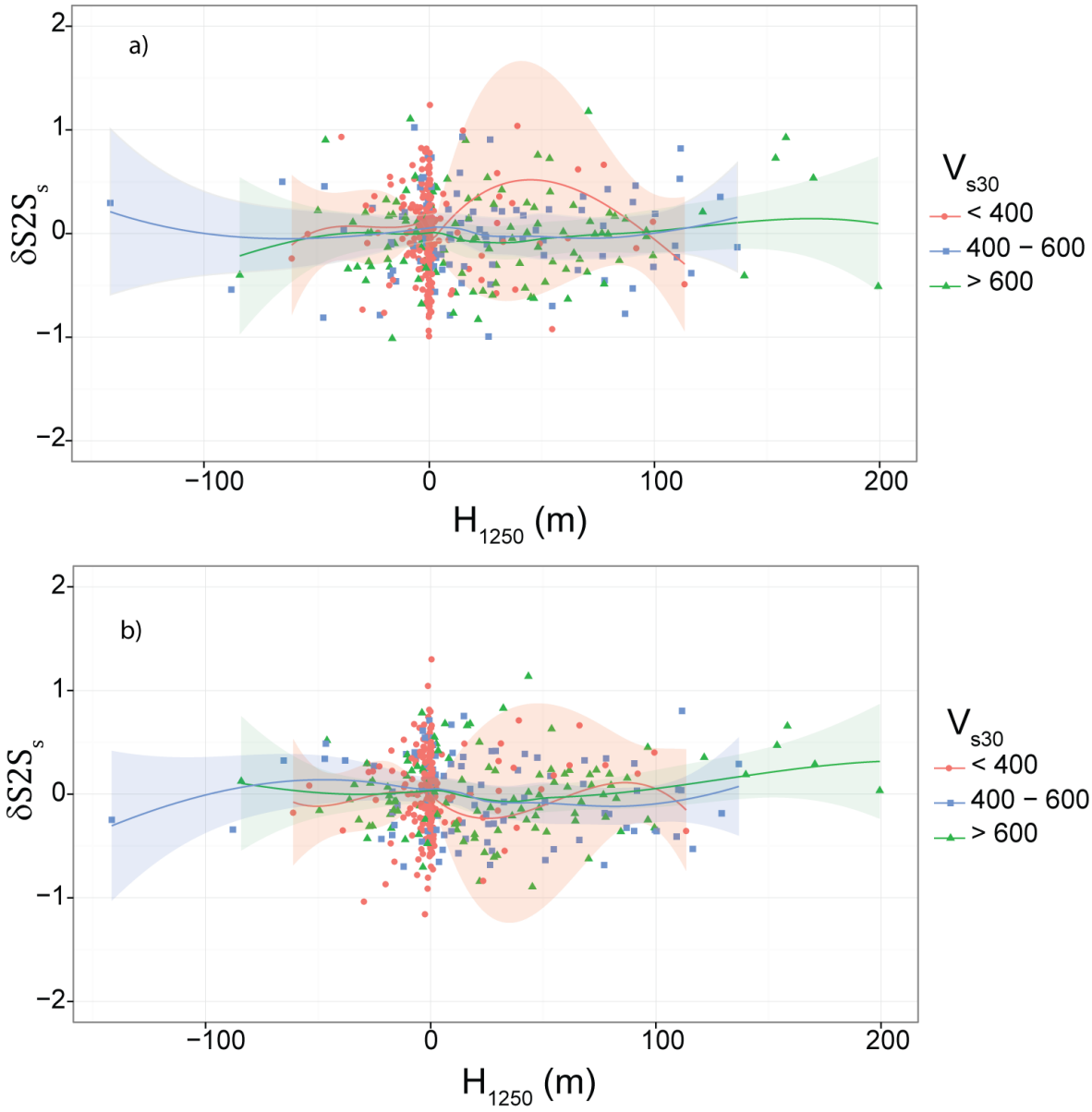

Figure 3.13: $\delta S2S_s$ from Equation 3.11 plotted with H_{1250} at the station for different V_{S30} bins. A moving average of the $\delta S2S_s$ terms computed using local regression (loess) is also shown. Lack of any significant trend suggests that there is no coupling between V_{S30} and topography.

metric parameters derived from elevation data for California. The two topographic parameters considered are relative elevation and smoothed curvature. An analysis of site residuals from the ground motion model with respect to these topographic parameters shows that both parameters are effective in capturing the bias in the ground motion residuals. In an effort to include these topographic terms in the existing ground motion model, a linear mixed effects regression is performed using two types of functional forms using each of the topographic parameters. We propose a multi-linear functional form using relative elevation to correct the Chiou et al. (2010) SMM model for PSA at periods of 0.3 s and 1 s. We do not propose modifications for the PGA model because we did not observe any residual trends with the topographic parameters considered. As the data for the study comes from California, further study would be required to test the applicability of this model to regions outside California.

This model does not purport to address all the aspects of topographic effects as the proposed corrections only address the average effects of topography. There could still be some effects not accounted for by the model. More elaborate numerical analysis might be a better choice for site-specific analysis. However, the numerical analysis results typically depend on a large number of variables (e.g. input motion, frequency content, angle of incidence, etc.). When these variables are properly accounted for, the resulting variability in the topographic amplification factors tend to be large. Therefore, site-specific numerical analyses should be conducted with extreme care and should aim at capturing potential uncertainties. We would also like to highlight that this paper is not an attempt to understand the physical basis of topographic effects, as various other publications have already addressed this aspect. Our goal was to extract the average effects of topography from the given database of recorded ground motions, and we achieved it through a simple parameterization scheme. This approach has its own obvious limitations. However, the observed biases in the dataset indicate that first-order predictions of the biases can be made with this simple parameterization. This effort does not negate further efforts to obtain better predictions using more elaborate physical models. Finally, we note that this model has to be used with caution in hazard analysis because it is based only on two spectral periods.

3.7 Acknowledgments

This research was partially supported by the National Science Foundation under award number CMII 1132373 and the California Strong Motion Instrumentation Program (CSMIP) under Standard Agreement No. 1012-955. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors would like to acknowledge Dr. Brian Chiou who provided the residuals for the Chiou et al. (2010) database. The constructive comments from the Editor, the Editorial Board Member and three anonymous reviewers have also greatly improved this manuscript.

Chapter 4

Empirical terrain-based topographic modification factors for use in ground motion prediction

Adapted with little modifications from

Rai, M., A. Rodriguez-Marek and B. S. Chiou (2015). Empirical terrain-based topographic modification factors for use in ground motion prediction. *Eartquake Spectra (under review)*

4.1 Abstract

We perform an empirical analysis to study the effects of surface topography on ground motions using the global NGA-West2 dataset and the Chiou and Youngs (2014) model as the reference ground-motion model. We quantify topography at a site using three geomorphometric parameters: smoothed curvature, smoothed slope, and relative elevation. We examine the relationship between the site residuals, which represent the average modeling error of the reference GMPE at a station, and the three topographic parameters. Our analysis shows

that there is a systematic trend in the site residuals with respect to the relative elevation parameter and the smoothed curvatures and that these trends are statistically significant. To account for the observed trends, we propose a modification to the Chiou and Youngs (2014) model to account for topographic effects on both the median and the standard deviation.

4.2 Introduction

This paper presents a model to predict the effects of earth's surface topography on earthquake ground motion. The model is based on an empirical analysis of the intra-event residuals from the Chiou and Youngs (2014) ground motion prediction equation (GMPE) and proposes smoothed period-dependent factors that can be used with GMPEs to predict topographic effects at any site in the future. This work is motivated by and builds on the study presented in Chapter 3 where we observed a systematic bias in the site residuals from a small-to- moderate magnitude earthquake (SMME) GMPE (Chiou et al., 2010) with respect to parameters derived using the surface elevations near the recording stations. In Chapter 3 we proposed a topographic modification to the Chiou et al. (2010) model; however, that model had certain limitations. First, the Chiou et al. (2010) model was available for only two spectral periods, and so the dependence of scale or the dimension of the feature on spectral periods could not be effectively evaluated. Second, the model cannot be used for hazard analyses that require predictions over a full range of spectral periods. The current work addresses both these shortcomings and extends that study to the larger NGA-West2 database.

We parameterize topography at the recording stations using three parameters; relative elevation, smoothed curvature, and smoothed slope. All the parameters are derived using the elevation data near the recording stations. We use the intra-event residuals from the Chiou and Youngs (2014) ground motion model to compute site terms at each recording station, and test the significance of the trends in the site residuals with respect to the above three topographic parameters. Other potential factors that may cause such trends

are evaluated and subsequently eliminated. These include potential sampling bias in the data, improper modeling of site terms, or other unaccounted for spatial phenomena. We also investigate the relationship between the topographic dimension and the periods they affect, as numerous studies have indicated a link between the two (e.g., Geli et al., 1988; Sánchez-Sesma and Campillo, 1991; Durand et al., 1999; Maufroy et al., 2015). We compute associated uncertainties in the proposed modification factors using bootstrapping. These uncertainties can be used to model epistemic uncertainties in a logic tree implementation of the probabilistic seismic hazard analysis.

4.3 Ground motion data

We use the subset of the NGA-West2 database used by Chiou and Youngs (2014). This subset consists of ground motion recordings from 300 earthquakes of magnitude 3 and higher, at 3208 stations located in the regions of California, Alaska, Japan, Taiwan, China, Turkey, Italy, Iran, and New Zealand. We use the intra-event residuals from the Chiou and Youngs (2014) GMPE to study topographic effects. As topographic effects are site effects, the residual component of interest for this study are the site residuals, which represent the average error in prediction at a site, after removing the effects of earthquakes. To obtain site residuals, the intra-event residuals (δW_{es}) are partitioned as follows:

$$\delta W_{es} = \delta S2S_s + \delta WS_{es} \tag{4.1}$$

where $\delta S2S_s$ are the site residuals, δWS_{es} are the site-and-event corrected residuals, and the subscripts e and s represent a given event and a given site, respectively. Each of the components of Equation 4.1 are assumed to be zero mean random variables with standard deviations ϕ, ϕ_{S2S}, and ϕ_{SS} for δW_{es}, $\delta S2S_s$ and δWS_{es}, respectively. For additional discussion on the partition of residuals from GMPEs see Atik et al. (2010) or Rodriguez-Marek et al. (2011). To get stable estimates of site residuals at each station, we only include stations

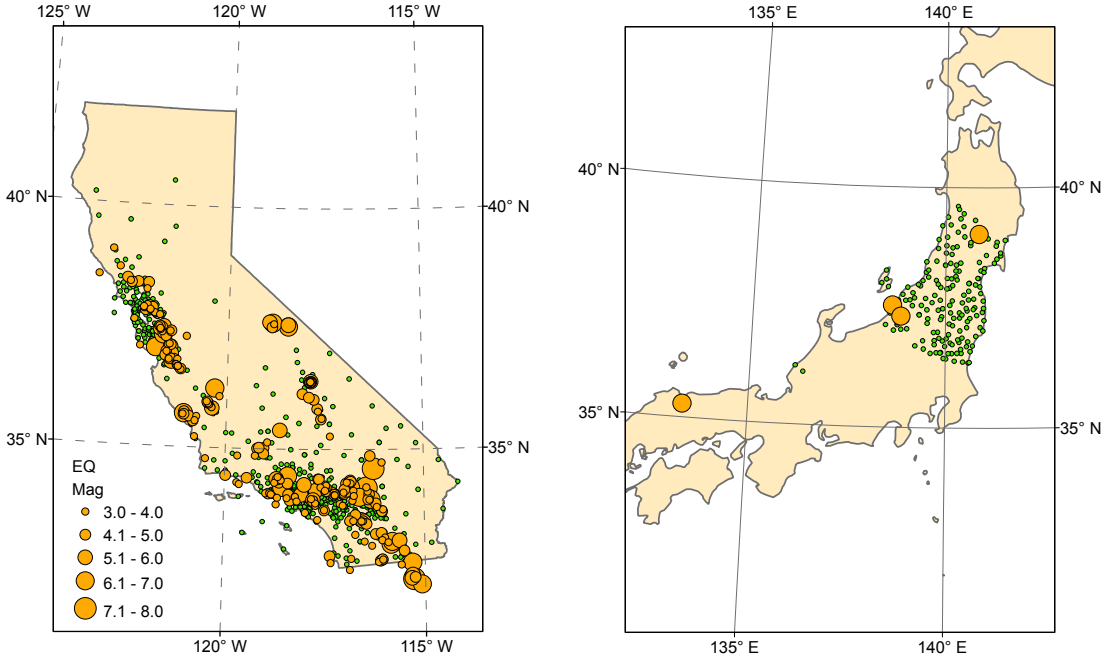

Figure 4.1: Locations of earthquake hypocenters and ground motion recording stations for the data used in this study. Only stations with three or more recordings are considered. This filtering of data resulted in stations within a) California and b) Japan .

with three or more recording on them. This constraint results in a total of 9,195 ground motions at 798 stations, located in California and Japan (Figure 4.1). The magnitude and distance distribution of the selected records are shown in Figure 4.2. The dataset consists of ground motion residuals at 105 spectral periods from 0.01 s to 10 s.

4.4 Methodology

4.4.1 Topographic Parameterization

Capturing topographic effects in a form that can be included in a GMPE requires the definition of a set of parameters that quantify topography. Geomorphometric parameters such as

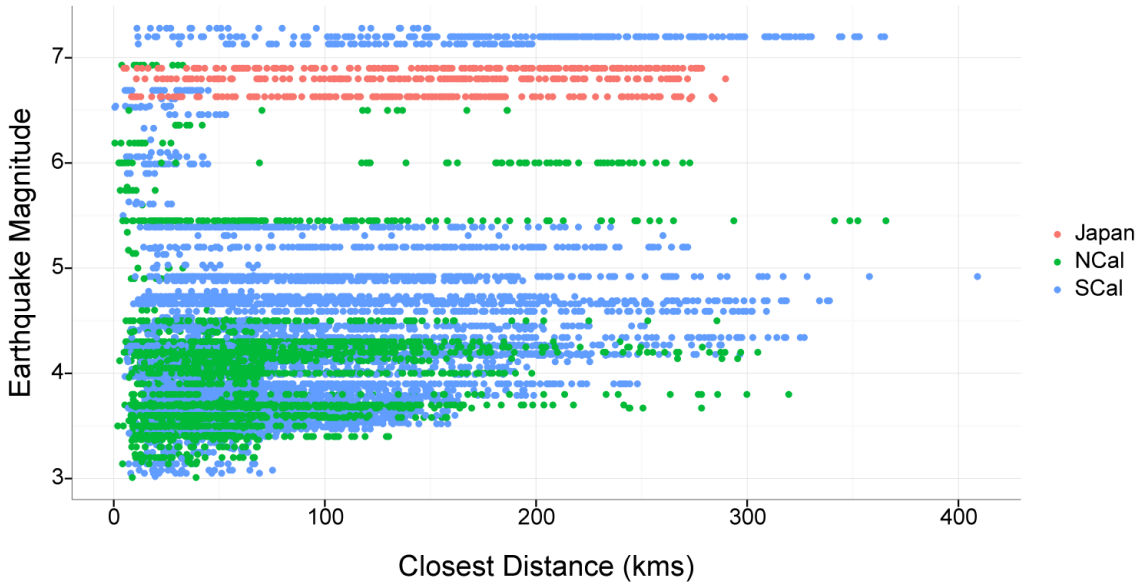

Figure 4.2: Magnitude distance distribution for the data used in the study.

slope, curvature, surface roughness, relative elevation are some important descriptors of land surface geometry have often been used for land element classification (e.g., Goldgof et al., 1990; Irvin et al., 1997; Burrough et al., 2000; Jones et al., 2000; MacMillan et al., 2000). Out of the above parameters, slope is often the most studied parameter in the context of topographic effects (e.g., Nguyen and Gatmiri, 2007; Ashford et al., 1997; Bouckovalas and Papadimitriou, 2005). These studies suggested that the amplifications at the crest of a 2D slope typically increased with slope angles. More recently, Maufroy et al. (2015) used numerical simulations to show that, for a realistic 3D topography, the smoothed measures of surface curvature are strongly correlated with median topographic amplifications. In Chapter 3, we found that both the smoothed curvature (C) and the relative elevation (H) parameter are correlated with biases in ground motion residuals. Based on these findings, we select the following three parameters for our study: slope (S), curvature (C) and relative elevation (H).

Slope is defined as the first spatial derivative of elevation and quantifies the steepness of the earths surface. Curvature is the second spatial derivative of a surface and quantifies

the convexity/concavity of the surface. Curvature values are positive for convex feature such as a hill or a ridge, and are negative for concave feature such as a valley or a canyon. A curvature of zero indicates the presence of a region where the slopes do not change and so it can be either a side of slope or a plain. Slopes and curvatures of a surface on earth can be computed using the built-in tools in ArcGIS (ESRI, 2011) on the elevation data. These tools however use a conventional 3 by 3 matrix of neighboring elevations from a digital elevation model (DEM) to compute these values. DEM is a raster data format where elevation values are arranged in a rectangular grid, or cells. Due to the default choice of neighborhood size in these computations, the outputs from these tools provide curvature and slope values that are local to a cell. As we are interested in using these parameters as a proxy for topography at a ground motion recording station, using such a local estimate of the parameter at the station will disregard the overall dimension or the scale of the feature the station is sitting on. To overcome this issue, we consider simple modifications of both these parameters: smoothed slopes and smoothed curvatures. To compute these parameters we first smooth out the DEM using different scales (window size) and then perform slope and curvature computations on the smoothed DEM. The window we use for smoothing is a circle centered at the grid cell at which the parameter values are to be computed. We refer to the diameter of this circle as the scale parameter (d). Smoothing fades away smaller irregularities (relative to the window size) from the DEM, and retains larger features. As a result, when we compute curvature and slopes of the smoothed DEM, we get the slopes and curvature values corresponding to those of the larger features. In this study, we use three scales for slope and curvature computations: $d = 180$ m, 360 m and 720 m. The effect of smoothing on slope and curvature parameter is shown in Figure 4.3. Note that at larger values of scale, the elevation raster loses its sharpness due to an increased level of smoothing. Similarly, the corresponding slope and curvature rasters show broader features.

The third parameter we use is relative elevation (H_d), also referred to in the literature as topographic position index. It is the difference between the elevation at a point on the surface and the mean elevation in the neighborhood of the point. This parameter has been

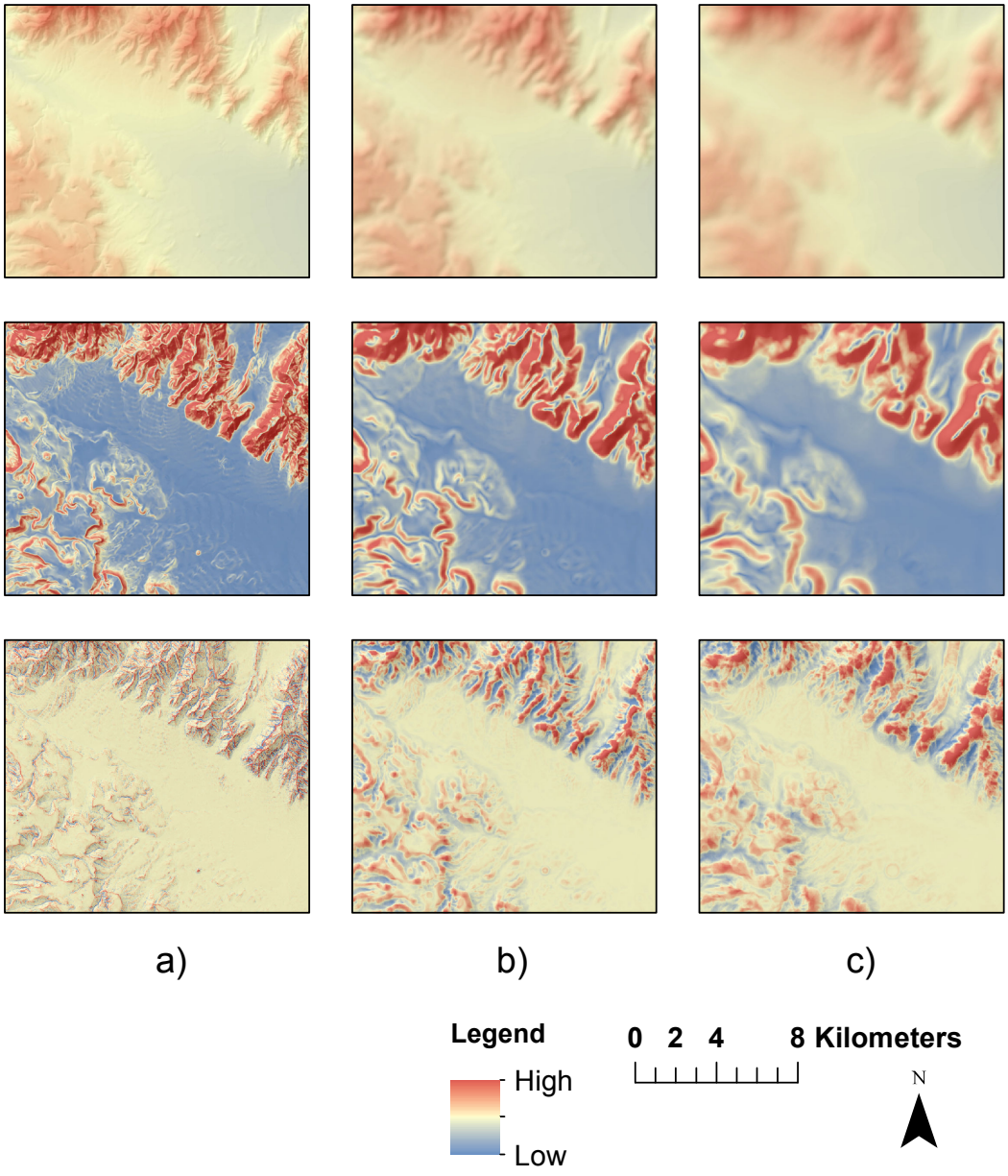

Figure 4.3: The top row consists of smoothed elevation raster, with a) no smoothing, b) smoothing using a scale of 360 m, and c) smoothing using a scale of 720 m. The middle row shows corresponding smoothed slopes and the bottom row shows corresponding smoothed curvatures. Both smoothed slopes and curvatures are computed using corresponding smoothed elevations.

used in the past to delineate ridges, slopes, and valleys in watershed study (e.g., Guisan et al., 1999; Jones et al., 2000). The details on the computation of this parameter can be found in Chapter 3. In this study, we compute H_d values at three scales (d): 500 m, 1500 m, and 3000 m. The effect of scale d on H_d is shown in Figure 4.4. We can see in the figure that as the value of d increases, the finer details are lost, and wider features are extracted.

All of the above topographic parameters are computed in ArcGIS using the elevation data around the stations. The elevation data for California stations is the National Elevation Dataset downloaded from the National map viewer service maintained by the USGS (`http://viewer.nationalmap.gov/viewer/`, last visited on 10/15/2014) and has a horizontal resolution of 1 arc-s or approximately 30 m. The elevation data for the Japanese stations is the ASTER GDEM v2 data downloaded from the ASTER GDEM website (`http://asterweb.jpl.nasa.gov/gdem.asp`, accessed on 10/15/2014), which is a joint product by Ministry of Economy, Trade and Industry of Japan and the United States National Aeronautics and Space Administration. The horizontal resolution of this data is also 1 arc-s.

In Chapter 3, we found that relative elevations and smoothed curvatures were highly correlated. A similar relationship was observed for this dataset too (Figure 4.5). This implies that the use of both parameters simultaneously does not provide any additional information about the topography, especially in an empirical study such as this. Therefore we decided to only use slope and relative elevation for further analysis. We select relative elevation over curvature as it is relatively easier to compute.

4.4.2 Residual analysis

In this section, we examine the site residuals for any trends with respect to the topographic parameters. Upfront, we did not observe any trends in the site residuals with respect to the smoothed slopes at any scales or periods (Figure 4.6). This may seem contrary to the findings from previous studies where slope angles were shown to affect the ground motion intensities

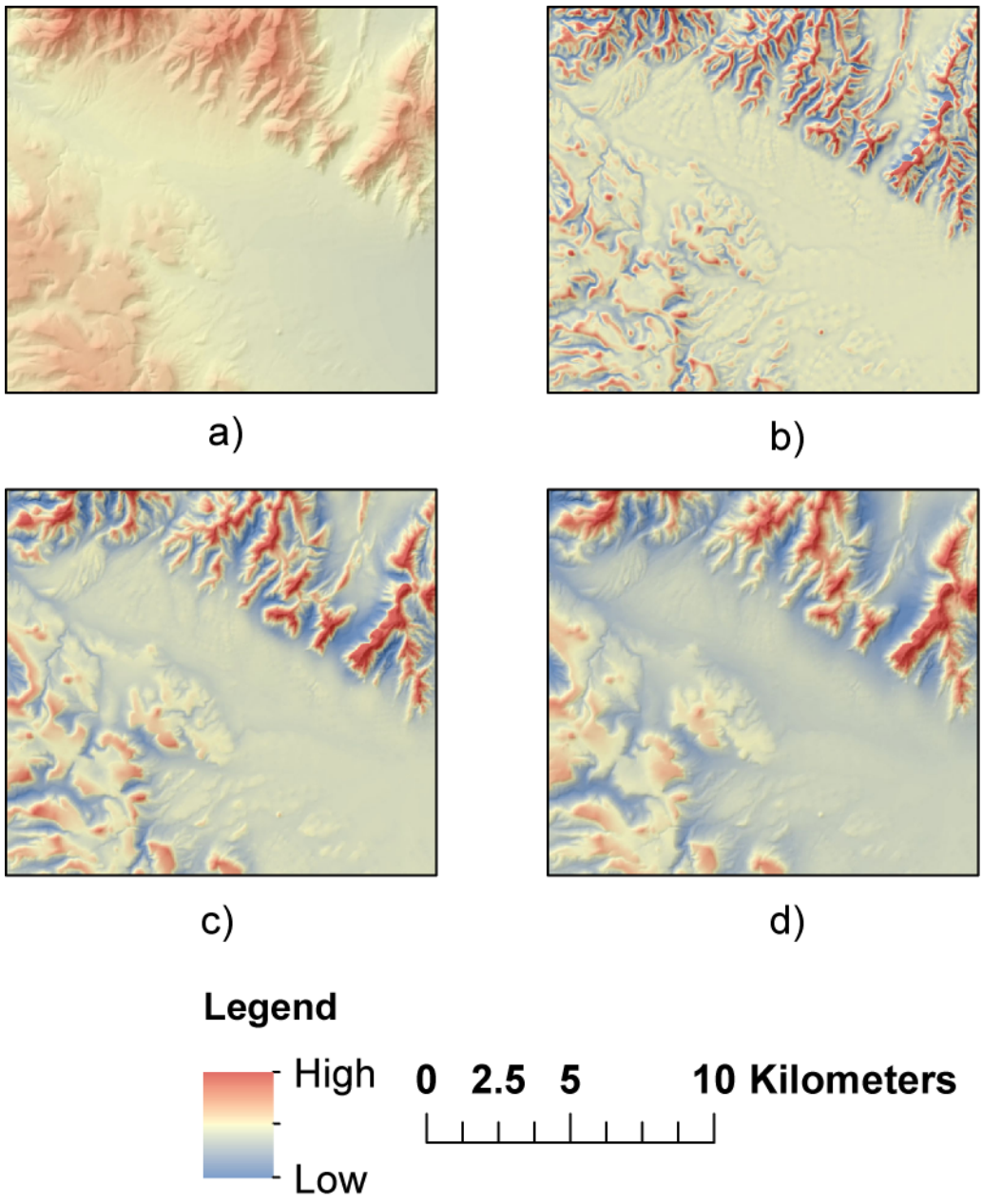

Figure 4.4: The variation of relative elevation with scale is shown here. a) is the elevation raster, b) is H_{500}, c) is H_{1500}, and d) is H_{3000}. Note that at smaller scales, finer features are visible. As the scale is increased, broader features become more prominent and finer details are lost.

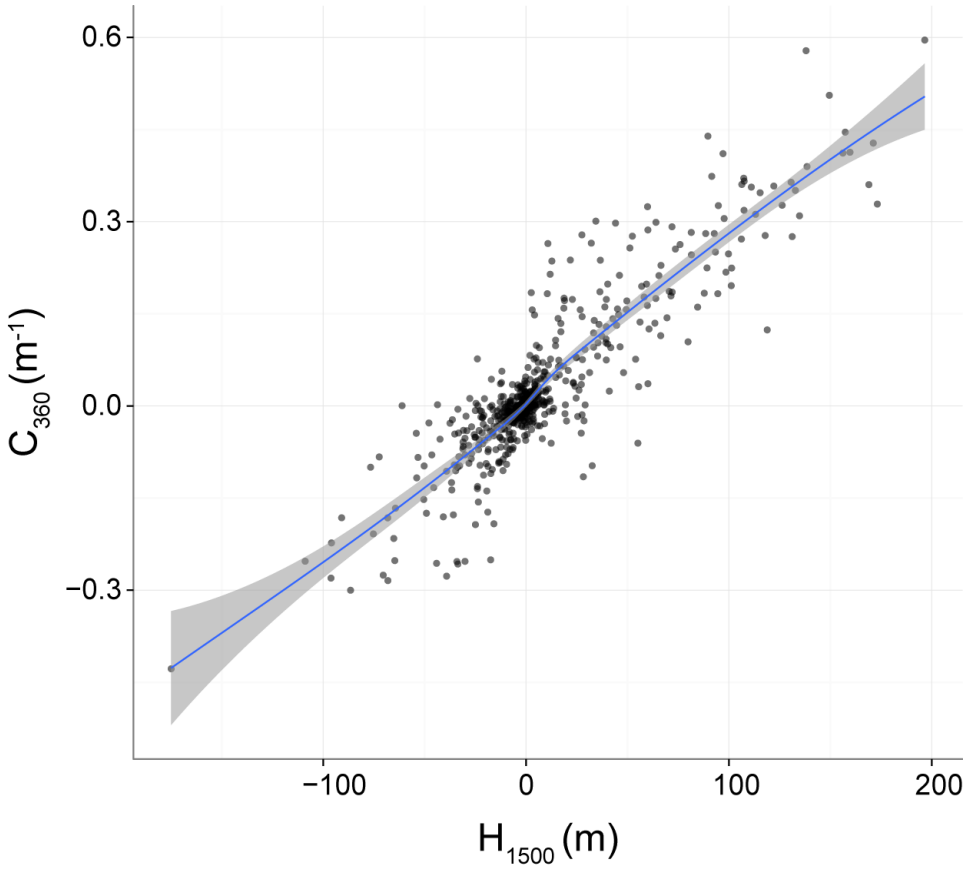

Figure 4.5: Scatterplot of C_{360} and H_{1500}. The two parameters have a high linear correlation.

at the crest and troughs of 2D slopes (e.g., Boore, 1972; Bouckovalas and Papadimitriou, 2005), with steeper slopes causing higher amplifications. However, it is important to realize that for a 2D slope of a fixed height, a change in the slope gradient also implies a change of smoothed curvature at the crest of the slope, with steeper slopes resulting in higher smoothed curvatures; therefore it is likely that the observed amplification at the crest in these studies were indeed a result of higher curvatures. Because of its lack of correlation with site residuals, we do not further pursue smoothed slope as a proxy of topography for quantifying topographic effects on ground motions. On the other hand, the site residuals show a clear positive trend with parameter H_{1500} at periods of 0.3 s, 1 s, and 3 s (Figure 4.7). In the next section, we test whether the residual trends with respect to H_d are statistically significant and evaluate other potential factors that could cause similar trends to show up in the data.

4.4.3 Significance test

We test the statistical significance of the residual trends by first dividing the stations into three classes based on the value of the topographic parameter H_d and then testing if the mean site residuals in each of the three classes are statistically different from each other. To do this, we denote the group of stations with $H_d > t\sigma_{H_d}$ as $High$, the group with $H_d < -t\sigma_{H_d}$ as Low, and the group with $t\sigma_{H_d} < H_d < t\sigma_{H_d}$ as $Intermediate$. Here, t is a constant threshold and is used to set the class boundaries. σ_{H_d} is the standard deviation of H_d values at the stations in the dataset used in this study. We use t values of 0.5 and 1, to see the effect of changing threshold on the classification and on the mean site residuals within each class. Figure 4.8 shows an example of the terrain shown in Figures 4.3 and 4.4 classified using this approach. Note that this classification depends on the value of d (used to compute mean elevation), and t (used to set the class boundaries). As we have selected 3 values of d and 2 values of t, there are a total of 6 d-t combinations. Each of these d-t combinations leads to a different classification (see Figure 4.8).

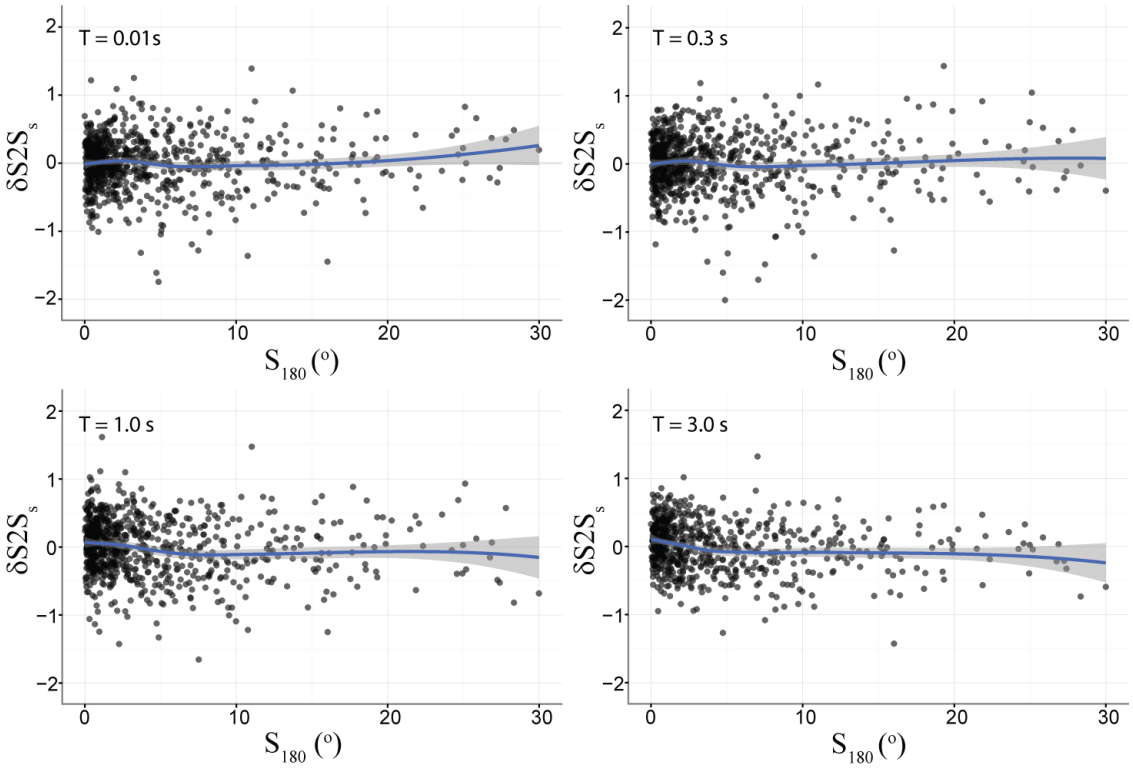

Figure 4.6: Variation of site residuals ($\delta S2S_s$) from Equation 4.1 with smoothed slope at a scale of 180 m (S_{180}). A moving average of the $\delta S2S_s$ terms computed using local regression (loess) is also shown. There is no visible trend between the two variables indicating that slope does not systematically affect ground motions.

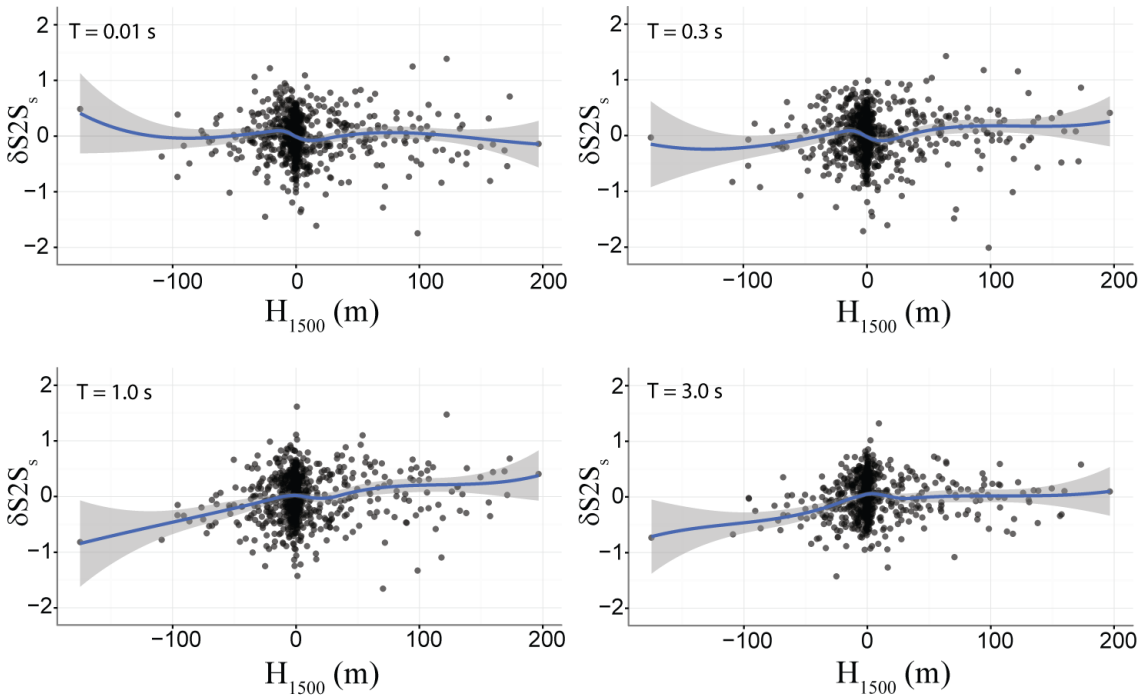

Figure 4.7: Variation of site residuals ($\delta S2S_s$) from Equation 4.1 with relative elevation at s scale of 1500 m (H_{1500}). A moving average of the $\delta S2S_s$ terms computed using local regression (loess) is also shown.

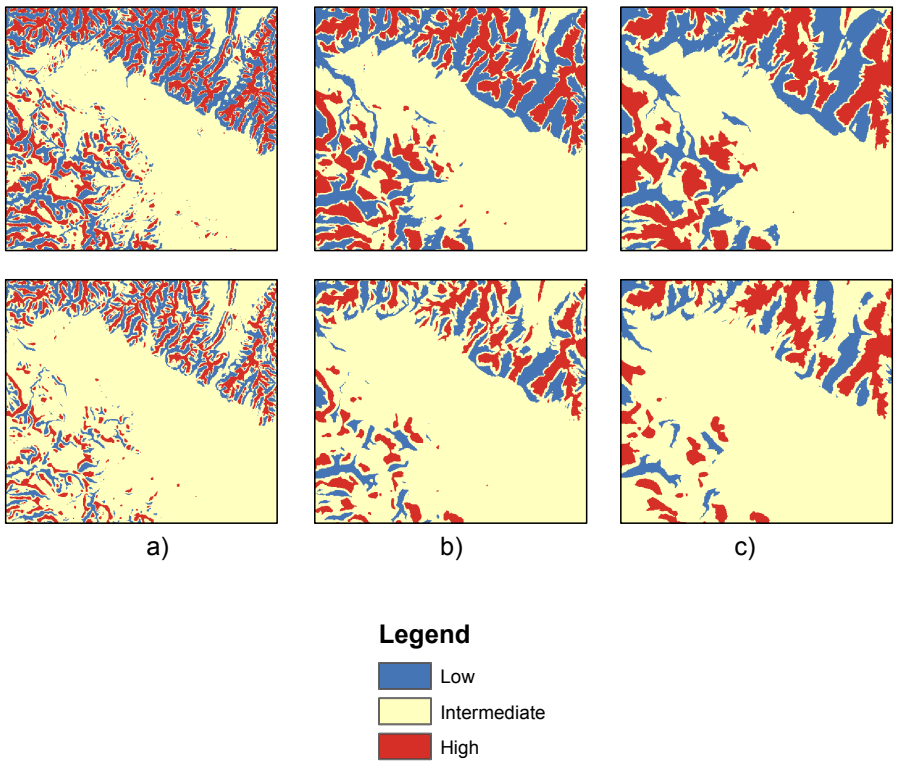

Figure 4.8: A terrain is classified using a threshold (t) values of 0.5 (top) and 1 (bottom) for scales of a) 500 m, b) 1500 m and c) 3000 m.

After determining the topographic class for each ground-motion station for a particular combination of d and t, we use a mixed-effects regression along with Equation 4.1 on stations from a single class to compute class-specific ϕ_{S2S} and ϕ_{SS}, as well as the mean site residual ($\delta\bar{S2S_s}$), which is the average of the site residuals in that class. Regressions were conducted using the R package lme4 (Bates and Maechler, 2010). We repeat this process for stations in the other two classes and also for other combinations of d and t. The resulting $\delta\bar{S2S_s}$ values for the three classes are shown as a function of the spectral period for different combination of d and t (Figure 4.9). We can see that there is an intermediate period range (T = 0.2 to T = 1 s) for which the $\delta\bar{S2S_s}$ value for the *high* class becomes greater than the $\delta\bar{S2S_s}$ value for the other two classes. Higher $\delta\bar{S2S_s}$ value implies that the recorded ground motions on the stations in that class were on average higher than predicted by the GMPE. The $\delta\bar{S2S_s}$ value for the *low* class typically reduces as period increases and becomes the lowest of the three classes for periods longer than 0.5 s. This means that at longer periods, the recorded motions in the *low* class are on average lower than the median predictions. As a majority of the stations in the dataset are classified as *intermediate* (557 classified as *intermediate*, 103 as *low*, and 138 as *high*), the ($\delta\bar{S2S_s}$) value for the *intermediate* class can be expected to approximate the total $\delta\bar{S2S_s}$ for the data. Thus, the $\delta\bar{S2S_s}$ value for the stations in the *intermediate* class is closer to zero compared to the other two classes. As the ground motion are presented in the log scale, the average amplification and de-amplification for each class can be computed by taking the exponent of the observed $\delta\bar{S2S_s}$ for that class.

We test the statistical significance of observed differences in the mean site residuals for the three topographic classes using the analysis of variance (ANOVA) method and the pair-wise Tukey's HSD test (Miller, 2011). The null hypothesis is that the means of the three sets are equal. These tests return p-values; a p-value of less than 0.05 implies that the null hypothesis can be rejected with a 95 percent significance level. Figure 4.11 shows the p-values from the ANOVA and Tukey's test for the 6 *d*-*t* pairs. A common trend in all of the 6 plots is that once the p-value from pairwise Tukey's test for the *intermediate-low* pair becomes lower than 5 percent it continues to be lower than 5 percent. This means that

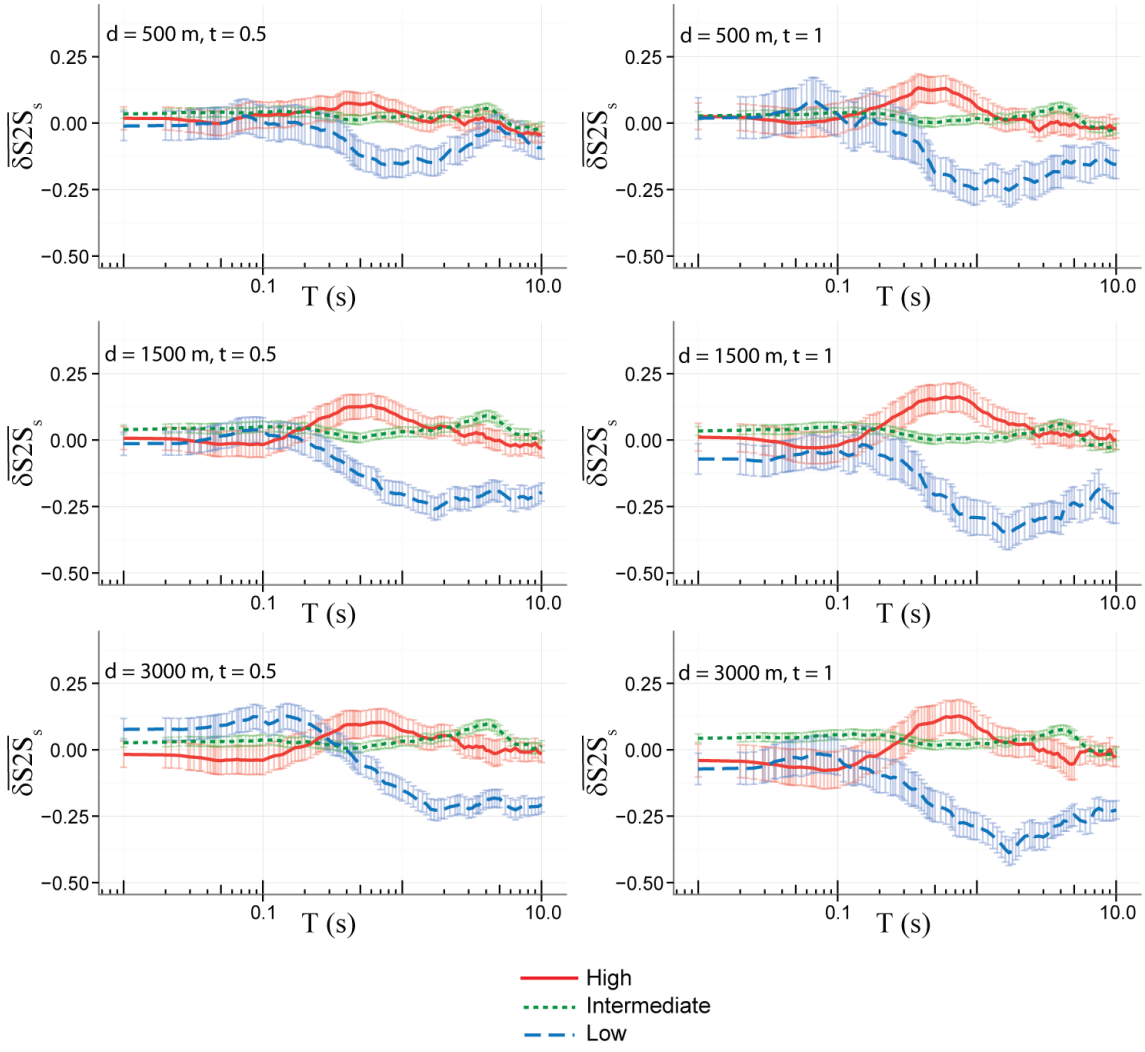

Figure 4.9: Mean site residuals ($\delta S\bar{2}S_s$) for the *high*, *intermediate* and *low* class along with the +/- 1 standard deviation error bars.

the difference between the mean site residuals of the *intermediate* and *low* class becomes significant at a certain period and then remains to be significant at all subsequently longer periods. Another important observation is that the p-value for the *intermediate* − *high* pair is less than 5 percent for only three scale and threshold combinations: $d = 500$ m, $t = 1.0$ and $d = 1500$ m, $t = 0.5$ or 1.0 and only for a narrow period range. This may mean that the features that seem to affect ground motions for the given periods are extracted only at certain scales. This also means that for a certain scale, the topographic effects are visible only for a certain period range. Previous numerical studies have shown that the effects of topography at a given period are significant when the topographic features have dimensions comparable to the associated wavelengths (e.g. Geli et al., 1988; Sánchez-Sesma and Campillo, 1991; Durand et al., 1999; Maufroy et al., 2015). To investigate the exact relationship between spectral periods that are affected by topography, and the scales where the effects are most prominent in the recorded data, we compute correlation coefficients between $\delta S2S_s$ (T), and H_d for different values of scales (d). These correlation coefficients will help us identify scales at which the site terms at a given period are most correlated with the relative elevation parameter. The values of these coefficients are shown in Figure 4.10. The cells in the figure are colored based on the numerical value contained in each cell. Note that the highest value of the coefficient in the figure is less than 0.2, suggesting that these correlations are very weak. These low values can be attributed to the high scatter in the site residuals. The relative values of these coefficients can tell us if there is any improvement in correlations moving from one scale to another, and from one period to another. For smaller periods (T < 0.2), the coefficients are either weakly negative or weakly positive, indicating that the parameter is not able to capture topographic effects at these periods, at any of the considered scales. At a scale of 500 m, there are relatively higher positive correlations ($\rho > 0.1$) at about 0.4 s. The correlation reaches a maximum at about 0.6 s, and diminishes after about 1 s. At a scale of 1500 m, the correlation is prominent at 0.4 s, reaches peak at 0.75, and diminishes after 2 s. For 3000 m, the correlation coefficient becomes significant at 0.5 s, reached maximum at 0.75 s, and declines at about 3 s. These results suggest that

Spectral Period (s)	Scale (m)		
	500	1500	3000
0.01	0.0094	-0.0346	-0.0636
0.05	-0.0242	-0.0642	-0.0935
0.1	-0.013	-0.0663	-0.0975
0.15	0.0078	-0.0465	-0.0708
0.2	0.0423	-0.0117	-0.0379
0.25	0.0767	0.0301	0.0016
0.3	0.0976	0.0582	0.0299
0.4	0.1445	0.1088	0.0787
0.5	0.1596	0.1331	0.1077
0.6	0.1627	0.1461	0.1248
0.75	0.1541	0.159	0.1504
1	0.1232	0.134	0.1308
1.5	0.0911	0.1259	0.1396
2	0.0818	0.1156	0.1411
3	0.0514	0.0942	0.1193
4	0.0335	0.0726	0.0911
5	0.0392	0.0691	0.086
7.5	0.0359	0.0842	0.1087
10	0.0779	0.1206	0.1419

Figure 4.10: Correlation coefficients values between the site residuals and the relative elevation parameters computed at scales of 500 m, 1500 m, and 3000 m

with an increase in scale, topographic effects are observable at larger periods. A possible explanation is that at smaller scales, the parameter extracts smaller features that influence smaller periods. At larger scales, larger features are extracted that affect higher periods. However, from Figure 4.9, we know that these higher correlation coefficients at a scale of 500 m and 3000 m must be controlled by stations in the *low* class, given there are no significant biases in the site residuals for *high* and *intermediate* class at any periods for these scales. The only scale that is able to capture the amplifications in the *high* class is 1500 m. For this reason, we select H_{1500} as the parameter to include in the model. Also, out of the two t values, we select a t value of 0.5 as a lower value of threshold would result in more stations in the *high* and *low* classes (Figure 4.7) and allow for a model that is well constrained. The t of 0.5 $\sigma_{H_{1500}}$ corresponds to a threshold of 17 m.

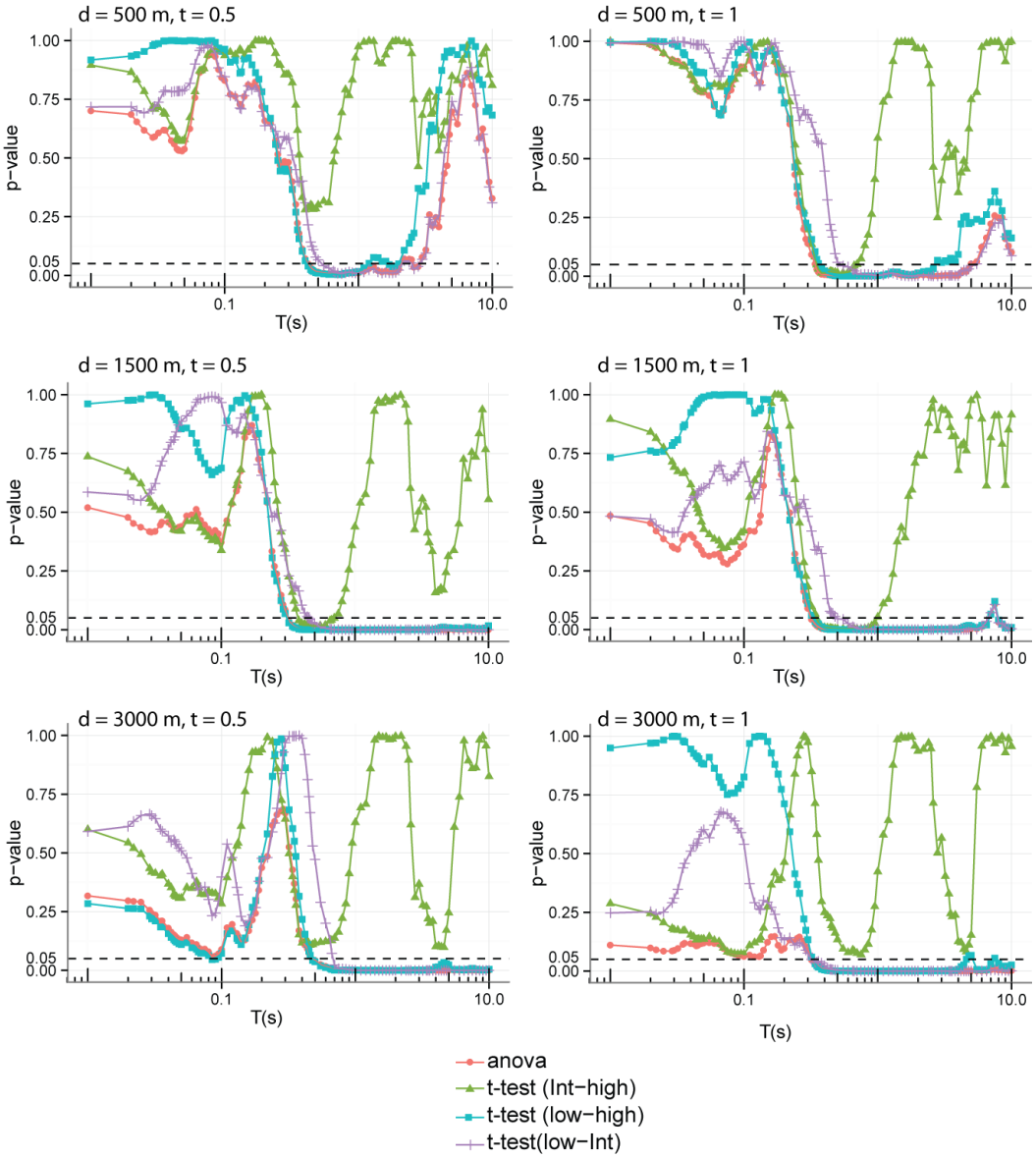

Figure 4.11: p-value from ANOVA and Tukey's pairwise t-test at different period. The stations are classified using a scale of 500 m, 1500 m, and 3000 m and a threshold t of 0.5, 1.

The site-to-site variability (ϕ_{S2S}), and the within-site variability (ϕ_{SS}) for the three topographic classes is shown in Figure 4.12. We expect to see a higher variability in the site residuals of *high* and *low* classes, compared to the *intermediate* class, in the period range where topography affects the ground motions. The reason behind such an expectation is that our classification scheme would group different kinds of ridges into the *high* class irrespective of other aspects of ridge geometry such as the direction of elongation of the ridge and the dimensions of the ridge. These factors can increase variability in the response of the *high* class as a whole to the ground motions. A similar argument will hold for the *low* sites. On the other hand, the majority of the station classified as *intermediate* in our dataset have smaller slopes and are thus mostly in plains. A flat region is less likely to have the same amount of variability in their response to an earthquake. We see this expected behavior, only up to a period of about 1 s. It is likely that after 1s, the number of observations in the dataset drops, making it difficult to constrain ϕ_{S2S} values. Some small differences were also observed in ϕ_{SS} values, with slightly higher ϕ_{SS} values for the *low* stations at short periods and small values of the scale parameter. The ϕ_{S2S} values were however not found to be stable when we selected fewer stations by increasing the limit on minimum number of recordings per site (see Appendix A). We therefore use the same site-to-site variability and single-station variability for all three classes in our model.

4.4.4 Other factors

The significance tests confirm that the differences in mean site residuals of the three topographic classes are statistically significant at some periods. Such trends may also result from other factors such as an improper modeling of site-effects (modeled using V_{s30}) or deep basin effects (modeled using $Z_{1.0}$), in which case our topographic parameter would be serving as a proxy for these effects. To test the first concern, we plot the site residuals with V_{s30} values for the three topographic classes (Figure 4.13). The moving average of site residuals for *high*, *low*, and *intermediate* class does not indicate a significant change in mean site residuals for

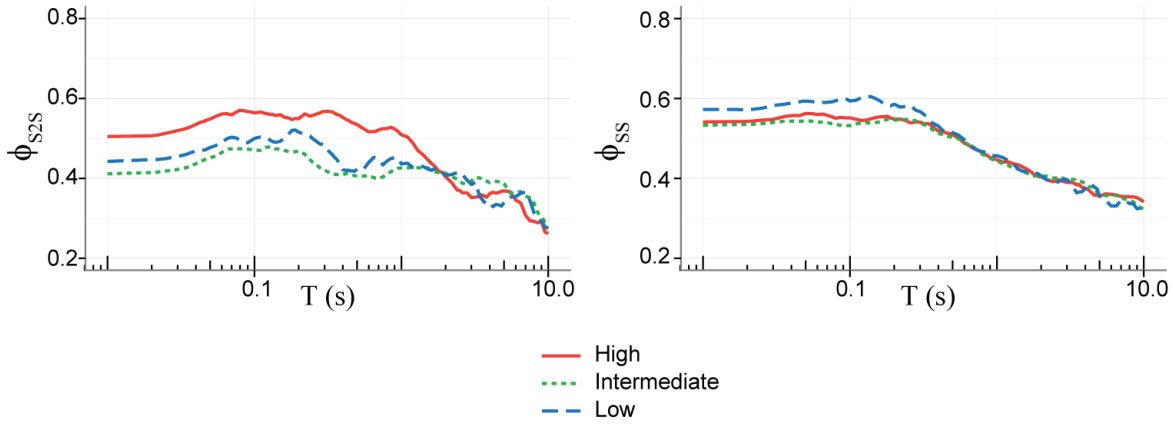

Figure 4.12: Variation of site-to-site variability or the $\phi S2S_s$ values for *high*, *low*, and *intermediate* sites with period. The stations are classified using a scale d of 1500 m, and a threshold t of 0.5

each class with V_{S30} suggesting that the observed biases in the site residuals with respect to H_{1500} parameter are more or less independent of V_{S30} values. We also analyze the histograms of magnitude, distance, and $Z_{1.0}$ (Figure 4.14) and they show a similar distribution of the parameter value across the three topographic classes, suggesting that sampling biases do not affect the topographic classes.

4.4.5 Regional differences

In this section, we check if there is a regional variation in the observed trends. We plot the site residuals with H_{1500} for the three regions of Japan, North California and South California for a period of 0.5 s (Figure 4.15). We observe that a majority of stations in Japan have an H_{1500} value of zero or lower, meaning that only a few of the stations are in the *high* class. Similarly, for Northern California, most the stations have a H_{1500} value of zero or higher, meaning that only a few stations are in the *low* class. Southern California is more balanced in terms of its distribution of stations in the *high* and *low* classes. We can also see that most of the trend in the site residuals is contributed from the stations in southern California. However, more data would be necessary to properly constrain regional differences.

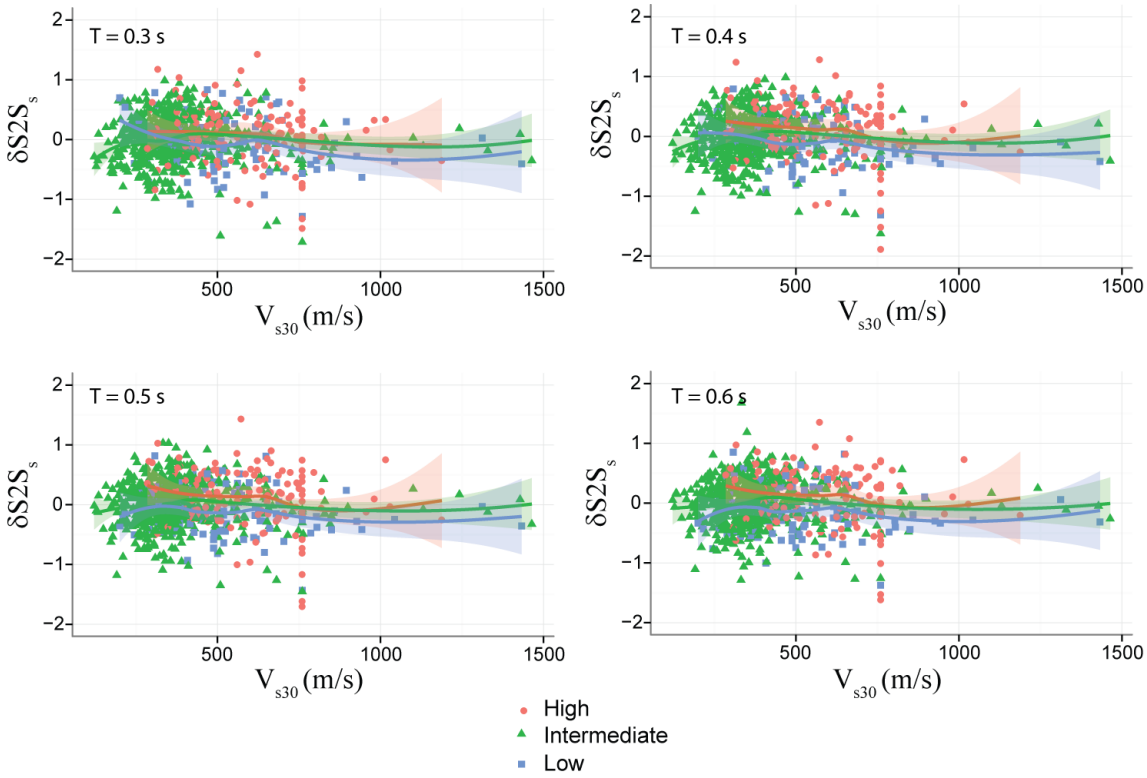

Figure 4.13: Variation of the site terms ($\delta S2S_s$) with V_{S30} at the ground motion stations for the three topographic classes. The classification is done using a scale of 1500 m and a threshold corresponding to $t = 0.5$. Moving average value of site terms for a topographic class is also shown along with the 95 percent confidence interval. The average value for site terms within each class does not seem to change significantly with the V_{S30} values, suggesting that the two are not correlated.

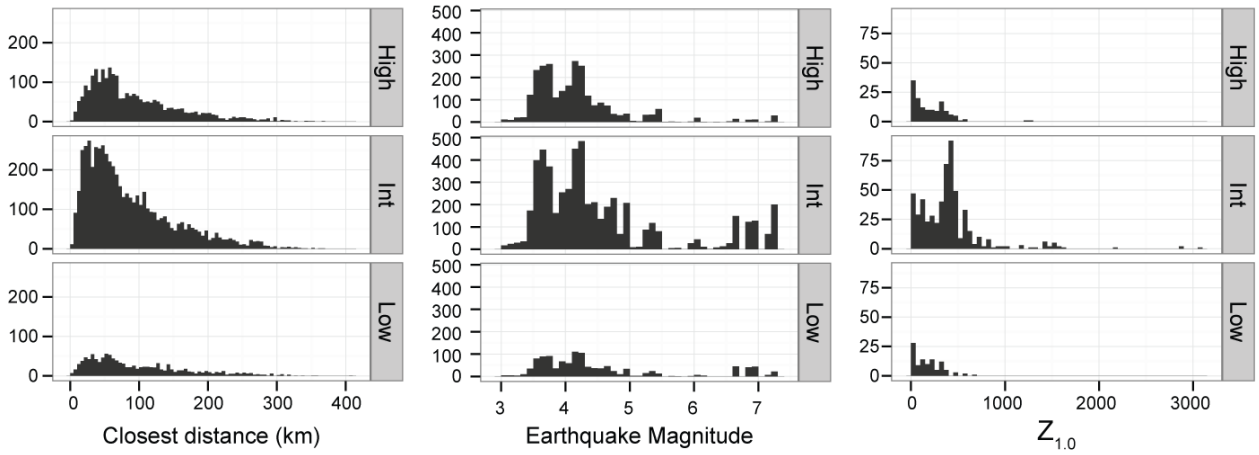

Figure 4.14: Histogram of Closest distance, Magnitude and $Z_{1.0}$ value for the three topographic classes. The classification is done using a scale (D) of 1500 m and a threshold (t) of 0.5.

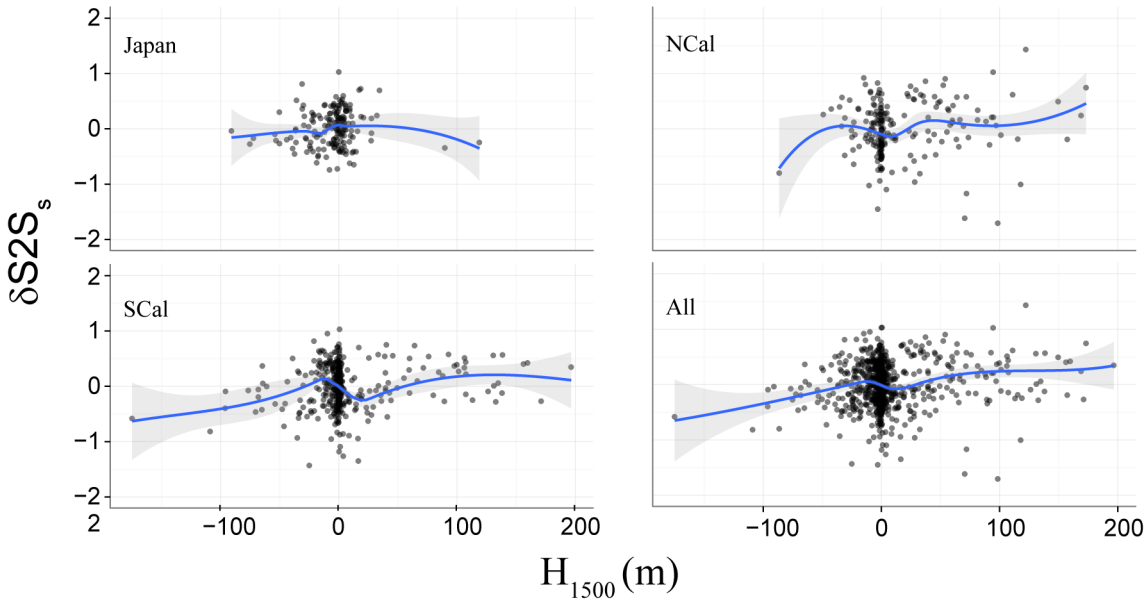

Figure 4.15: Variation of site residuals ($\delta S2S_s$) from Equation 4.1 (prior to the removal of topographic effects) at a period of 0.5 s as a function of H_{1500} for Japan, Northern California (NCal), and Southern California (SCal). Its is evident that most of the residual trends observed in the total data is coming from SCal stations. Also, for Japan, majority of the stations have H_{1500} value less than zero.

4.5 Model development

To account for the bias in residuals with respect to topography, we add a term containing our topographic parameter to the right side of Equation 4.1, as follows:

$$\delta W_{es} = f(H_{1500}) + \delta S2S_s + \delta W_{es} \tag{4.2}$$

where δW_{es} is the intra-event residual from the Chiou and Youngs (2014) ground motion model, f is a function of the topographic parameter H_{1500}, $\delta S2S_s$ is the site residual after accounting for topographic effects, and δW_{es} is the site and event corrected residual. Following Chapter 3, we select a multi-linear functional form for f that has constant levels of spectral acceleration value for each topographic class with a linear transition from one class to the next class. The choice of this functional form is based on the fact that we observed distinct behavior within each class, as demonstrated by the different levels of mean site residuals (Figure 4.9), and that this behavior is also in agreement with previous studies. A multi-linear form proposes average levels of amplifications expected at sites within each class (e.g., *high*, *low*, or *intermediate*), and therefore it is much more robust in predictions than fitting say a linear model, that would predict increasing values of amplification for higher values of the topographic parameter, even though we do not have physical evidence to support such a model. The functional form we selected for accounting topographic effect is given by:

$$f_{topo} = \begin{bmatrix} c_{low} & H_{1500} < -20 \\ c_{int} & -17 < H_{1500} < 17 \\ c_{high} & H_{1500} > 20 \end{bmatrix} \tag{4.3}$$

with linear transition zones for intermediate values of H_{1500} (e.g., -20 to -17 or 17 to 20 m). In the model, we have kept the transition zone from *low*/*high* class to *intermediate* class very steep such that the coefficient values can be more realistically constrained, which was an issue in the model presented in Chapter 3.

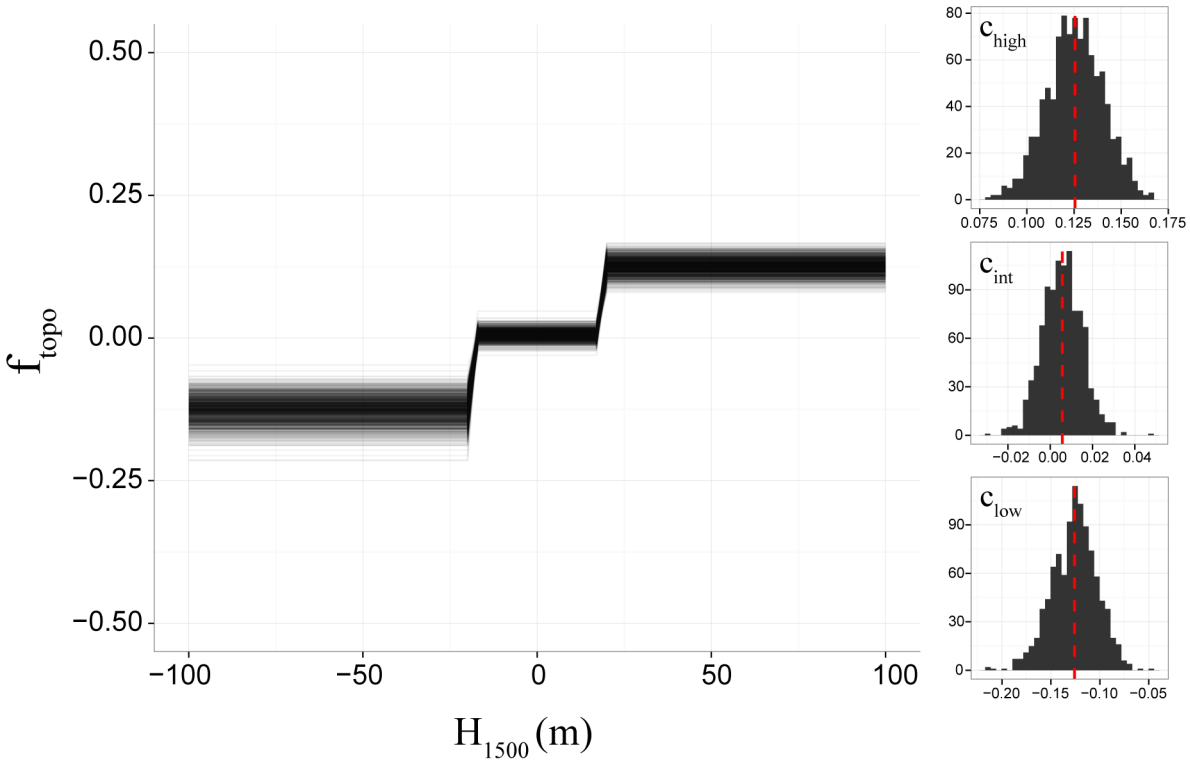

Figure 4.16: The resulting bootstrapped models from Equation 4.3 for a period of 0.5 s along with the histograms of coefficients c_{low}, c_{int} and c_{high}. The histograms show the mean value of the coefficient in a red dotted line.

The median value of the model in Equation 4.3 has an epistemic uncertainty associated with it. We capture this uncertainty using *bootstrapping* with replacement (e.g., Efron and Tibshirani, 1993) and fit Equation 4.3 to the data for 1000 random samples (1000 samples produced stable estimates for the parameter distributions). For each sample, different values of model coefficients for each period are obtained. The models obtained from bootstrapping are shown in Figure 4.16 for a period of 0.5 s. We compute the mean and the standard deviations of the three coefficients of the regression from the 1000 realizations. These standard deviations are a measure of the epistemic uncertainty in the estimated values of these coefficients. The values of cross-correlation between the bootstrapped coefficients c_{low}, c_{int}, c_{high} are very small (< 0.05 for most cases), indicating that the coefficients are independently distributed. This implies that these uncertainities can be estimated independently.

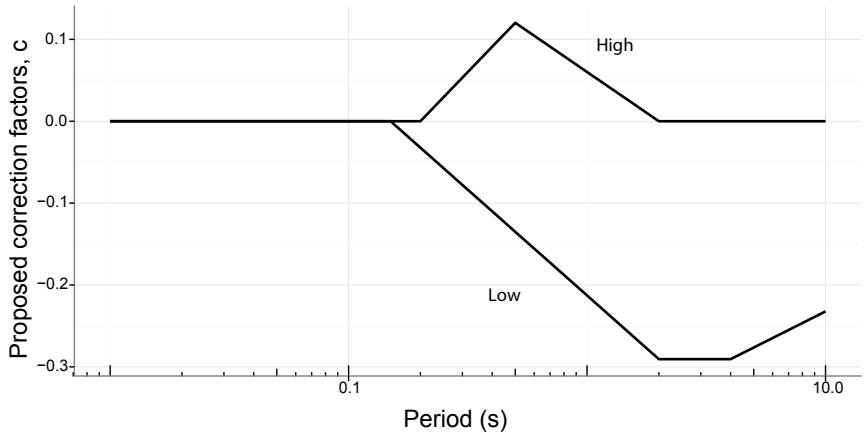

Figure 4.17: Proposed modification factors, c_{low} (H_{1500} < -20 m) and c_{high} (H_{1500} > 20 m) for topography. Linear interpolation should be used to estimate the modification factors for absolute H_{1500} values of 17 m to 20 m.

The value of the coefficient c_{int} obtained from regression was fairly close to zero at all periods, and there is no reason to assume that they are different than zero. Because, we want to preserve the difference between different levels, we subtract c_{int} from c_{high}, and c_{low} and smooth these values, ensuring that the c_{high}, c_{low} values gradually reach zero for periods where they were not significantly different than c_{int}. The updated coefficient values of coefficients c_{low}, c_{high}, and ϕ_{s2s} obtained through the mixed-effects regression on the full data set, and the epistemic uncertainties (σ) associated with each coefficient, obtained from bootstrapping are reported in Table 4.1. These values can be added directly to the GMPE to estimate topographic effects at a site. We note that the *high* sites in the dataset showed a maximum amplification at a period of 0.5 s, and the *low* sites showed maximum deamplification between periods of 2 - 4 s. We could not establish the exact reason for this behavior; however we think that the majority of sites classified as *high* and *low* might be experiencing some sort of topographic resonance at these periods. Figure 4.17 shows the variation of the proposed modification factors with period.

Table 4.1: Period dependent coefficients for f in Equation 4.3

Spectral Period(s)	c_{low}	$\sigma(c_{low})$	c_{high}	$\sigma(c_{high})$	ϕ_{S2S}	ϕ_{SS}
0.01	0	-	0	-	-	-
0.05	0	-	0	-	-	-
0.10	0	-	0	-	-	-
0.15	0	-	0	-	-	-
0.2	-0.0323	0.0263	0	-	0.4894	0.5518
0.25	-0.0573	0.0248	0.0293	0.0167	0.4704	0.5497
0.3	-0.0778	0.0255	0.0532	0.0175	0.4580	0.5428
0.4	-0.1100	0.0254	0.0910	0.0162	0.4396	0.5165
0.5	-0.1351	0.0226	0.1202	0.0158	0.4346	0.5060
0.75	-0.1805	0.0220	0.0851	0.0155	0.4335	0.4680
1	-0.2128	0.0219	0.0601	0.0142	0.4450	0.4460
1.5	-0.2583	0.0195	0.0250	0.0134	0.4309	0.4192
2	-0.2906	0.0192	0	-	0.4110	0.4054
3	-0.2906	0.0207	0	-	0.3854	0.3948
4	-0.2906	0.0213	0	-	0.3776	0.3830
5	-0.2764	0.0199	0	-	0.3772	0.3602
7.5	-0.2506	0.0236	0	-	0.3406	0.3483
10	-0.2323	0.0263	0	-	0.2802	0.3268

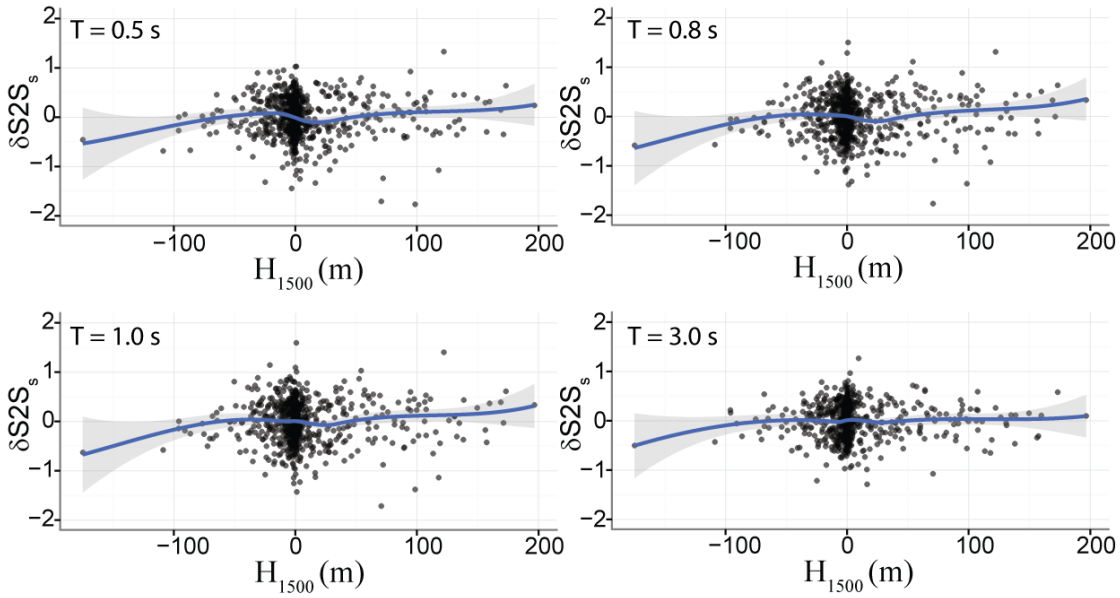

Figure 4.18: Scatterplot of site terms ($\delta S2S_s$) from Equation 4.2 with H_{1500} at the station. Lack of a significant trend suggests that the functional form in Equation 4.3 has accounted for most of the bias in the data.

4.5.1 Model diagnostic

The site residuals from Equation 4.2 (which include the effects of topography) are plotted with H_{1500} and V_{S30} for periods of 0.5 s, 0.8 s, 1 s, and 3 s (Figure 4.18 and 4.19). Lack of any trend with H_{1500} indicates that the functional form has adequately accounted for the effects associated with H_{1500}. The apparent bias at very low H_{1500} values ($H_{1500} < -150$) is due to one point alone; however, note that the confidence intervals still cover the zero line, so that apparent trend is not statistically significant. No trend with respect to V_{S30} indicates that the modification for topographic effects did not introduce any biases with V_{S30}, and hence there is no interaction between V_{S30} and H_{1500} in our dataset.

We attributed the trends associated with H_{1500} to the topography, as this parameter is derived using the topography at the site. However, the possibility that the observed trends result from other site-effects, such as basin effects, cannot be completely eliminated. This may happen if, for example, a majority of stations in the *high* class were also located on a

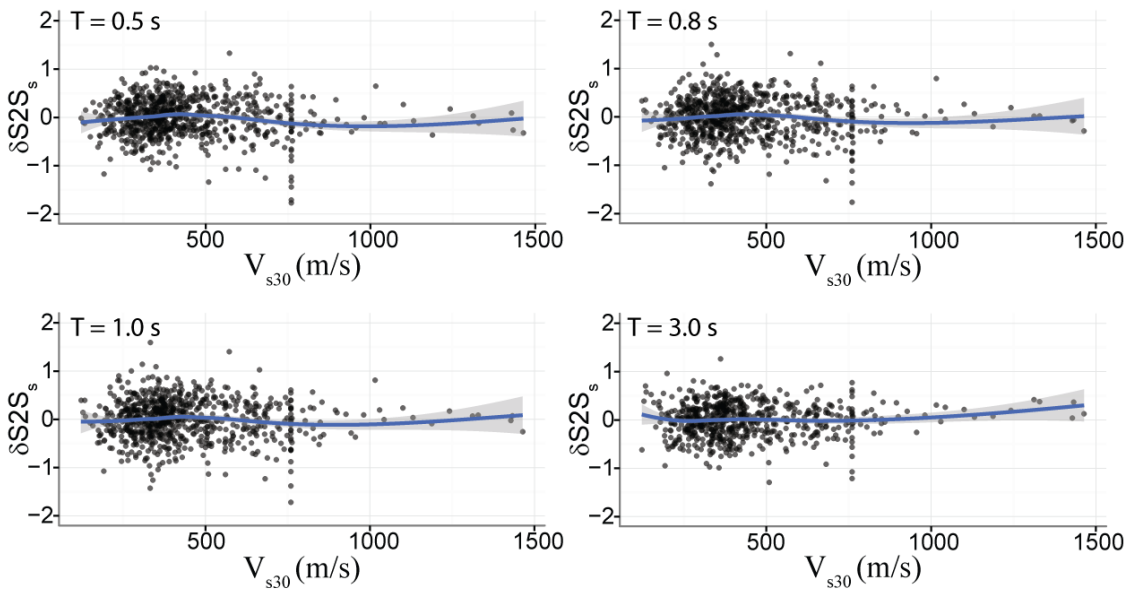

Figure 4.19: Scatterplot of site residuals ($\delta S2S_s$) from Equation 4.2 with V_{S30} at the station. Lack of any significant trend with respect to V_{S30} indicates that the functional form used in Equation 4.3 is not introducing any systematic bias with respect to V_{S30}.

basin, where ground motions are typically amplified. In such a case, one cannot say with certainty that the average amplifications at those stations are only a result of topographic effects. However, in such a case, we can expect to see clusters of sites with higher site residuals in the data where basin effects are prominent. Doing a spatial clustering analysis of the site residuals will highlight any such potential spatial phenomenon. So, given the site residuals at each station in the dataset, we want to find out whether there are any significant spatial clusters within each topographic class: *high*, *intermediate*, and *low*. The details of this analysis are included in Appendix B. In summary, we found that there were no significant clusters of stations with high site residuals or clusters of stations with low site residuals within the *high* and *low* topographic classes. This means that the trends we observed in the site residuals earlier were not caused due to any other site-effects. The only place where we see significant clustering is in the *intermediate* class of stations in both California and Japan, however that doesnt affect conclusions related to topographic effects. Note that no significant biases in the site residuals in that class were observed.

4.6 Comparison with proposed topographic modification factors in Chapter 3

There are three major differences between the model presented in this chapter and the model presented in Chapter 3. The first difference is in the choice of scale parameter to compute relative elevations that are used in the predictions. In Chapter 3 we used a scale of 1250 m, whereas in the current study we use a scale of 1500 m. The second difference is in the choice of thresholds and transition zones in the fitted multi-linear function. In Chapter 3 model, the transition zones were between absolute H_{1250} values of 15 m, and 40 m, whereas in this chapter we selected a transition zone between H_{1500} value of 0.5 $\sigma_{H_{1500}}$ (17 m) and 20 m. The third difference is the dataset used to fit the models; In Chapter 3 we used a small to medium magnitude earthquake dataset, with recording from only within California, whereas in this study we use a much larger dataset of recorded ground motion including recordings from higher magnitude earthquakes. The differences in the functional form, the parameter value used to fit the function form, and the dataset affects the resultant coefficient values in the model. To do a fair comparison of the topographic modification factors from Chapter 3 and the current work, we fit two additional functional forms using the current data. In the first model, we first compute H_{1250} for the stations in this study and then fit a multi-linear function of H_{1250} that is identical to the one we proposed in Chapter 3. In the second model, we fit a multi-linear function of H_{1250} using transition zones as defined in this chapter. For both these models, we compute and plot the $c_{high} - c_{int}$, and $c_{low} - c_{int}$, coefficients, along with the predictions from Chapter 3, and unsmoothed modification factors computed in this chapter (Figure 4.20). We can see in the figure that the difference in the modification factor values when we use H_{1250} parameters, instead of the H_{1500} parameter, are very small. However, the difference between our unsmoothed factors and the factors obtained if the functional form from Chapter 3 is used with the same data is comparatively larger. This difference is a result of using different thresholds in the fitted functional form. Finally, there are differences in the coefficients when the same functional form is fitted with two different

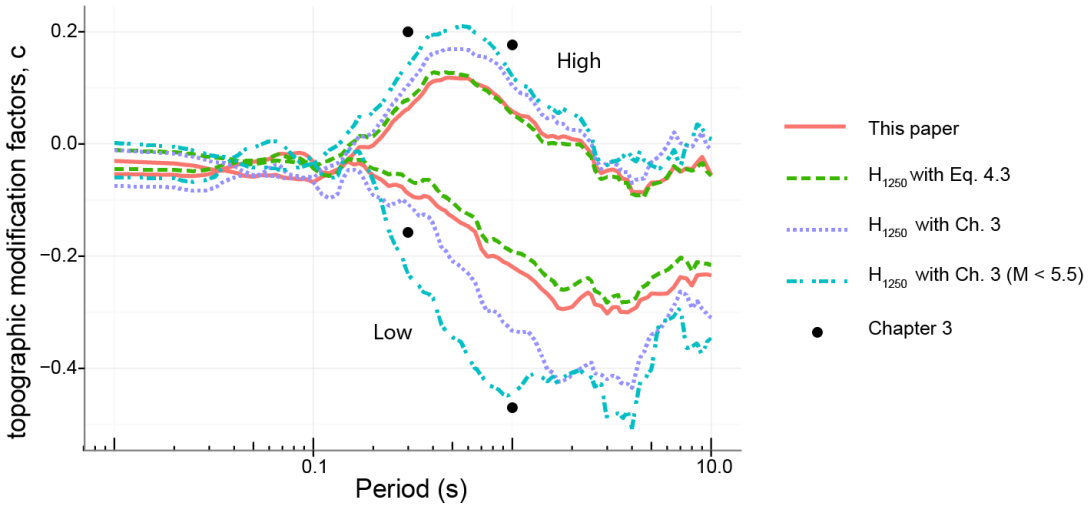

Figure 4.20: A comparison of the topographic modification factors c_{high}, c_{low} obtained from using different functional forms and datasets.

datasets. To test if these differences are caused due to using recording from higher magnitude earthquake, we refitted the functional form presented in Chapter 3, using the current dataset, but only including recordings from earthquake of magnitude 5.5 and lower. We find that in the resultant model, the biases for the *low* and *high* sites increased for *intermediate* periods (0.2 s to 1 s) compared to other models.This increase was greater for *low* sites, than the *high* sites, suggesting that earthquake magnitude could potentially influence topographic effects. Further studies would be necessary to understand this phenomenon better. Nevertheless, the proposed modification factors are providing conservative estimates of expected amplifications and deamplification compared to all the other models discussed here.

4.7 Summary and conclusion

We used the NGA-West2 database to empirically study the effects of topography on earthquake ground motions. Topography was quantified using geometric parameters that were

computed from the elevation data around the station. The three topographic parameters studied were smoothed slope, smoothed curvature, and relative elevation. Two of the parameters, smoothed curvature and relative elevation, were found to be linearly correlated. Of these two parameters, we only used relative elevation parameter for the regression analysis, as it a relatively simpler parameter to compute.

We used the intra-event residuals from the Chiou and Youngs (2014) model to compute the average event-corrected residuals at each station. We then examined these site residuals with respect to the smoothed slope and the relative elevation to see if there are systematic trends in the data with respect to these parameters. We found that the site residuals did not carry any significant trend with respect to slope. We observed a statistically significant trend in the site residuals with the relative elevation parameter for a band of periods and only at some scales. We studied the effect of changing scale on affected periods, and found that at higher scales, higher periods are affected. The trends in the site residuals with respect to relative elevation were found to be stronger in Southern California than in Northern California, and were very weak in the Japanese data.

We proposed a modification to the Chiou and Youngs (2014) model to account for the observed biases in the ground motion residuals. The proposed modification predicts an average maximum amplification of about 13 % at *high* sites at 0.5 s , and a maximum deamplification of about 25 % at *low* sites between 2 - 4 s. To account for epistemic uncertainties in the modification, we used the method of bootstrapping and proposed standard deviations for the coefficients obtained from the regression. These values can be incorporated in a logic tree when performing a seismic hazard computation. We also performed a cluster analysis of the site residuals from the resultant model to test if there are any other site effects that are causing the observed biases in the data in the *high* and *low* class. Our analysis showed that there is no significant evidence of any spatial effects in the two topographic classes.

We compared the proposed model to the proposed model from Chapter 3 that was fitted using the Chiou et al. (2010) small to medium magnitude earthquake dataset. We

found that there are differences in the two predictions. These differences are a combined effect of three factors, namely the choice of the scale, thresholds in the fitted functional form, and the differences in the dataset. We found that the earthquake magnitude appears to have an effect on the modification factors in the final model, with larger factors when only smaller earthquakes are included in the regression. Additional studies will be required to understand if there is a dependence of topography on earthquake magnitude.

4.8 Acknowledgments

This research was partially supported by the National Science Foundation under award number CMII 1132373 and the California Strong Motion Instrumentation Program (CSMIP) under Standard Agreement No. 1012-955. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Chapter 5

Topographic proxies from 2-D numerical analyses

Adapted from Rai, M., Rodriguez-Marek, A. (2015). Topographic proxies from 2-D numerical analyses. *To be submitted*

5.1 Abstract

We use simplistic 2D numerical analyses at ground motion stations from the NGA-West2 dataset to develop parameters to predict the effect of surface topography on response spectra. The simplistic numerical analyses use sinusoidal input motions, uniform soil profiles, elastic soil properties, and absorbing boundary conditions. We define several parameters using the natural logarithm of 2D amplifications in peak ground acceleration of a sinusoidal input motion in different orientations. The natural log of 2D amplifications when averaged over multiple orientations was found to have the most predictive power among the studied parameters. We also explore the relationship between the topographic parameters developed in this study, and the topographic parameters developed at the same sites in Chapter 4.

5.2 Introduction

The presence of topographic irregularities has been shown to significantly affect the intensity and frequency of strong ground motion during earthquakes (Bouchon, 1973; Boore et al., 1981; Bard, 1982). Numerous studies have documented the widespread structural damage on hills and ridges (Trifunac and Hudson, 1971; Celebi, 1987; Geli et al., 1988; Kawase and Aki, 1990; Hatzfeld et al., 1995; Bouchon and Barker, 1996), and several studies have documented actual amplifications of ground motions (Hartzell et al., 1994; Lee, 1994; Shakal et al., 1994; Bouchon and Barker, 1996; Spudich et al., 1996). Past analytical and numerical studies have shown that the ground motions are amplified on hills and ridges, and deamplified on valleys and canyons (Davis and West, 1973; Rogers et al., 1974; Griffiths and Bollinger, 1979; Tucker et al., 1984; Geli et al., 1988; Bouchon and Barker, 1996; Assimaki and Gazetas, 2004; Meunier et al., 2008; Assimaki and Jeong, 2013; Maufroy et al., 2015), although a quantitative estimation of the effect still remains difficult (Geli et al., 1988). Some of this difficulty arises from the challenges of accurately modeling some factors that also influence the response of topography to a ground motion. These effects include, the angle of incidence, type of input wave, soil properties, damping ratio, shear wave velocity profile (Geli et al., 1988; Assimaki et al., 2005) and the choice of reference station. Due to a lack of proper quantification, these effects are not yet accounted for in ground motion prediction equations (GMPEs).

Since topographic effects are systematically observed in the ground motions, they should also be predictable. Prediction of topographic effects can only be achieved by studying the dependence of the effects at a site on a family of parameters. Several numerical studies have investigated the effects of simple site geometry on ground motions by using parameters such as the 2D slope gradient, height, and width. Although these studies have helped us understand the effects of different factors on earthquake ground motions, without a proper verification with a large set of empirical data, these results cannot be used for predictions. This issue was addressed in Chapters 3 and 4, where we studied the biases in the site residuals

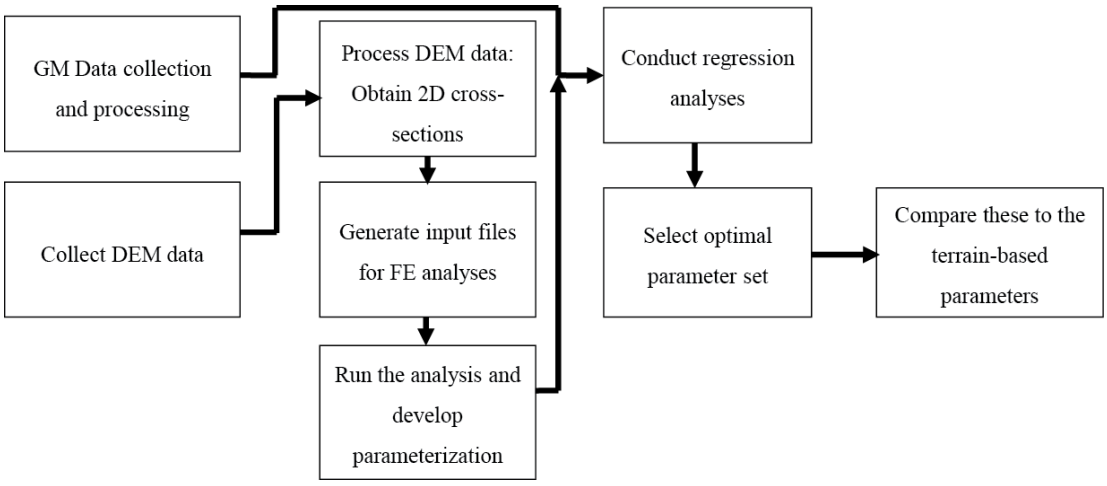

Figure 5.1: Detailed flowchart of the methodology used for topographic parameters from finite difference analysis.

from the GMPE predictions with respect to several topographic parameters derived using elevation data around the stations. We found significant biases in the site residuals with respect to a parameter based on the relative elevation of the station. Site residuals are the average error in prediction at a site, after removing the effects of earthquake. The relative elevation parameter (H_d) is the difference between elevation at a site and the mean elevation in a neighborhood around a site and it quantifies how high or low a point on the surface is compared to its surrounding. Although, the parameters in Chapters 3 and 4 were shown to predict topographic effects, better parameters can still be developed to improve the modeling of these biases. The goal of the study is to develop new parameters to predict topographic effects at a site, and compare them to the elevation-based parameters used in Chapter 4. The proposed parameters are based on the prediction of amplification/deamplification from simplistic 2D numerical analysis at the site; these parameters are used to predict the biases in GMPEs. Figure 5.1 shows a flowchart of the outline of our proposed methodology.

5.3 Topographic Parametrization

We compute amplification of an input wave at a site from a 2D analyses using multiple cross-sections, and use these amplifications to develop a family of topographic proxies for the site. To compute these amplifications, we perform simplistic numerical analyses for each station in the dataset by using planar 2D meshes with simplistic V_s profiles, elastic soil properties, transmitting boundary conditions, and sinusoidal input motions. The top surface of the mesh is fit to the shape of the cross-sectional profile at the site of the station. A cosine tapered sinusoidal velocity is used as an input at the base of the mesh, and the resultant time history is recorded at the top surface of the mesh. Using the recorded motion, we compute the peak ground acceleration (PGA) at the station. These analyses are not meant to accurately evaluate the expected topographic amplification at a site. We only intend to compute an approximate estimate of amplification due to topography at a site and use these amplifications as an input parameter in a regression analysis to predict average amplification at a site. This concept is somewhat similar to using the V_{s30} parameter to quantify site amplification. V_{s30} is a measure of travel time of a vertically propagating shear wave over the last 30 m of a profile, assuming perfectly horizontal layering, and is used extensively as a proxy for site response in the GMPEs.

The proposed parameterization scheme is divided into 3 steps. The first step is to obtain the cross-sectional profiles at the site of each recording station in the dataset. For each station, we compute profiles in six orientations. The second step is to set up the mesh for the analyses. In this step, we fit the top surface of the mesh to the shape of the corresponding cross-section in that orientation. For each 2D analysis, we also perform a 1D analysis, where we use a flat horizontal surface at the top of the mesh. We take the ratio of PGAs from the 2D and 1D analyses to compute amplification due to topography alone. The third and the last step is to use the amplifications obtained in the previous step to develop a family of topographic parameters at every station. The details of individual steps are given below.

5.3.1 Generating cross-sectional profiles

We obtain 2D cross-sectional profiles of the ground surface at the site of each recording station in six equally spaced orientations, each 30 degrees apart. Multiple 2D sections at the site capture the details of terrain geometry in different directions, and enable us to obtain the seismic 2D response of the feature in different directions. For this study, we use the subset of the NGA-West2 ground motion dataset used by Chiou and Youngs (2014) and filter it to only include stations that have 3 or more recordings on them. We use this filter to ensure that the corresponding site residuals are properly constrained. Properly estimating site residuals is important here, as we are studying topographic effects, and these effects are site-specific. The filtered dataset consists of 798 recording stations in California and Japan. The locations of these stations, earthquake hypocenter, and the Magnitude-distance distribution can be found in Chapter 4. Since we use six cross-sections for every recording station, there are a total of 4788 cross-sections.

To obtain the cross-sections at each recording station, we use the ground surface elevation data in the neighborhood of the recording stations. The elevation data is typically available in a raster format that defines a space as an array of equally sized cells, arranged in rows and columns, with each cell containing the elevation data. To obtain the cross-sectional profiles at the stations in the NGA-West2 dataset, we use the 30 m resolution National Elevation Dataset (NED) (`http://viewer.nationalmap.gov/viewer/`, last visited on 10/15/2014) for California and a 30 m ASTER GDEM v2 data for Japan (`http://asterweb.jpl.nasa.gov/gdem.asp`, accessed on 10/15/2014). Both these datasets provide the elevations for a bare earth model that is free from any vegetation or man-made structures. The horizontal units for distance in the downloaded raster are in degrees of latitude and longitude. We project these rasters to their corresponding Universal transverse Mercator (UTM) projection to convert these distances from degrees to meters. We obtain 2D cross-sectional profiles at the station by interpolating the elevations from the corresponding projected raster along a straight line centered at each station and striking at six azimuth

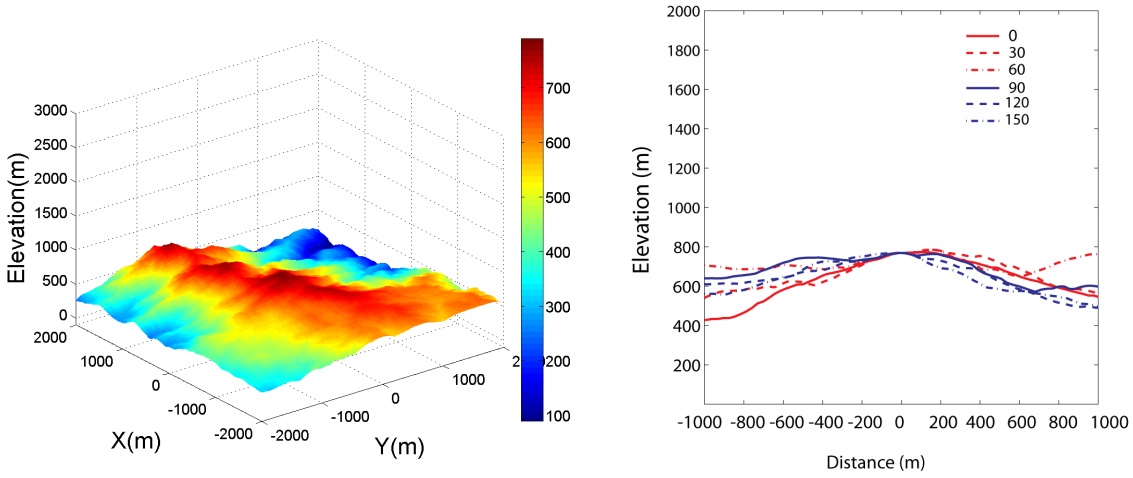

Figure 5.2: 3D terrain around a site (left) and cross-sectional profiles across the station in 6 different directions (right) are shown for the Tamalpais peak B station (37.9231, -122.5983). The recording station is located at [0, 0] m.

angles between 0 to 150 degrees, each 30 degrees apart. Figure 5.2 shows an example of the resultant cross-sections for the Tamalpais peak station. We subtract the elevation at the recording station from the entire profile. As these cross-sections are later used to generate meshes for analysis, this step ensures that in the resulting meshes, every station is at the same height from the base of the mesh, such that the only difference between two sites is its surface topography, and we are solely accounting for topographic effects. The process of projecting the rasters and creating cross-sections in multiple directions for each station is automated using python scripts that use the ArcPy package to control ArcGIS (ESRI, 2011).

5.3.2 Generate Finite Difference meshes

A 2000 m by 1000 m deep rectangular finite difference mesh is generated in FLAC 5.0 (Itasca Consulting Group, 2011). The base and the top of the grid are set at an elevation of -1000 m and 0 m, respectively. We use two approaches to generate meshes for the analysis. In the first approach, we assign a constant shear-wave velocity profile (V_s) of 500 m/s at every station.

In the second approach, we assign a V_s value equal to the shear wave velocity in the top 30 m of the site (V_{s30}). Past numerical studies have shown that shallow soil layers can magnify the effects of surface topography (Assimaki and Gazetas, 2004; Assimaki et al., 2005; Assimaki and Jeong, 2013), and so the first approach looks at the effects of surface topography on ground motions without any interaction from the soil, whereas the second approach partially addresses the effect of different shear wave velocities on topographic effects.

Approach 1: A constant Vs profile of 500 m/s at every station

We assume a grid size equal to a tenth of the wavelength of the input motion in the x-direction and one-twentieth of the wavelength in the y-direction. The grid sizes in numerical analyses are constrained by the shortest wavelength in the input motion, and a grid size equal to a tenth of the wavelength is typically recommended. Even though an aspect ratio of 1:1 for the grid is preferred, we assumed a finer grid size in the y-direction to compensate for the fact that the grid may need to be stretched to fit the shape of the cross-sectional profile. The mesh is generated for a rectangular section, and then the top surface is distorted to fit the shape of the cross-sectional profile at the station. Note that the process of reshaping the mesh to fit the cross-section at the site does not induce any distortion at the site of the station, because the height of the station stays the same in the process. As a result, the aspect ratio in the proximity of the stations remains at the initial 2:1 ratio. This ratio is well within the acceptable range of 5:1 that provides reasonable numerical accuracy. We assign the following material properties to the soil: mass density $\rho = 2\ Mg/m^3$, Poissons ratio $= 1/3$, $V_s = 500$ m/s (Bouckovalas and Papadimitriou, 2005; Tripe et al., 2013) and a target damping ratio of 5 % using Rayleigh damping. Though Rayleigh damping is a frequency dependent damping and soil damping is not, an appropriate choice of the central frequency can ensure that the damping is constant over the range of frequencies of interest. We use the central frequency for damping equal to the frequency of the input motion.

Approach 2: A constant shear-wave velocity profile equal to V_{s30} at the station

In this approach, we assume a constant grid size of 10 m in both x and y directions for all the stations used in the study. The grid size is sufficiently small to propagate a wave of period 0.5 s with a V_{s30} of 200 m/s. As 97 % of the stations in the selected dataset have a V_{s30} value greater than 200 m/s, this grid size is appropriate for our analysis given we will only use input waves of period 0.5 s and higher. The mesh is distorted as in Approach 1. We assign the same material properties to the mesh as in the first approach but use a damping ratio of 2 %. A lower damping ratio is assumed such that the site resonances are not excessively dampened.

The base of the mesh is assumed to be an absorbing boundary and to model it in FLAC, quiet boundaries are applied at the base of each model in both x and y directions. To model the quite boundaries, FLAC uses the viscous boundary developed by Lysmer and Kuhlemeyer (1969). This scheme uses dashpots that are attached independently to the boundary in the normal and shear direction. When a quiet boundary is applied at the base of the mesh, the accelerations time histories can no longer be applied directly and shear stresses are applied instead. The shear stresses to be applied are computed using the following equation:

$$\sigma_s = -2\rho \, C_s \, v_s \tag{5.1}$$

where σ_s is the applied shear stress in the horizontal plane, ρ is mass density of the material, C_s is the s-wave velocity in the medium, and v_s is the input shear particle velocity. To minimize reflections from the lateral boundaries, free-field boundaries are applied to both the lateral boundaries. Free-field boundary conditions in FLAC simulate the effect of an infinite elastic medium surrounding the model. A schematic illustration of the mesh and the applied boundary conditions are shown in Figure 5.3. We use harmonic velocity waves with 10 - 20 total cycles as an input. We taper the motion at both the ends using a cosine taper to ensure a gradual build-up and decay of acceleration and velocity. The acceleration and

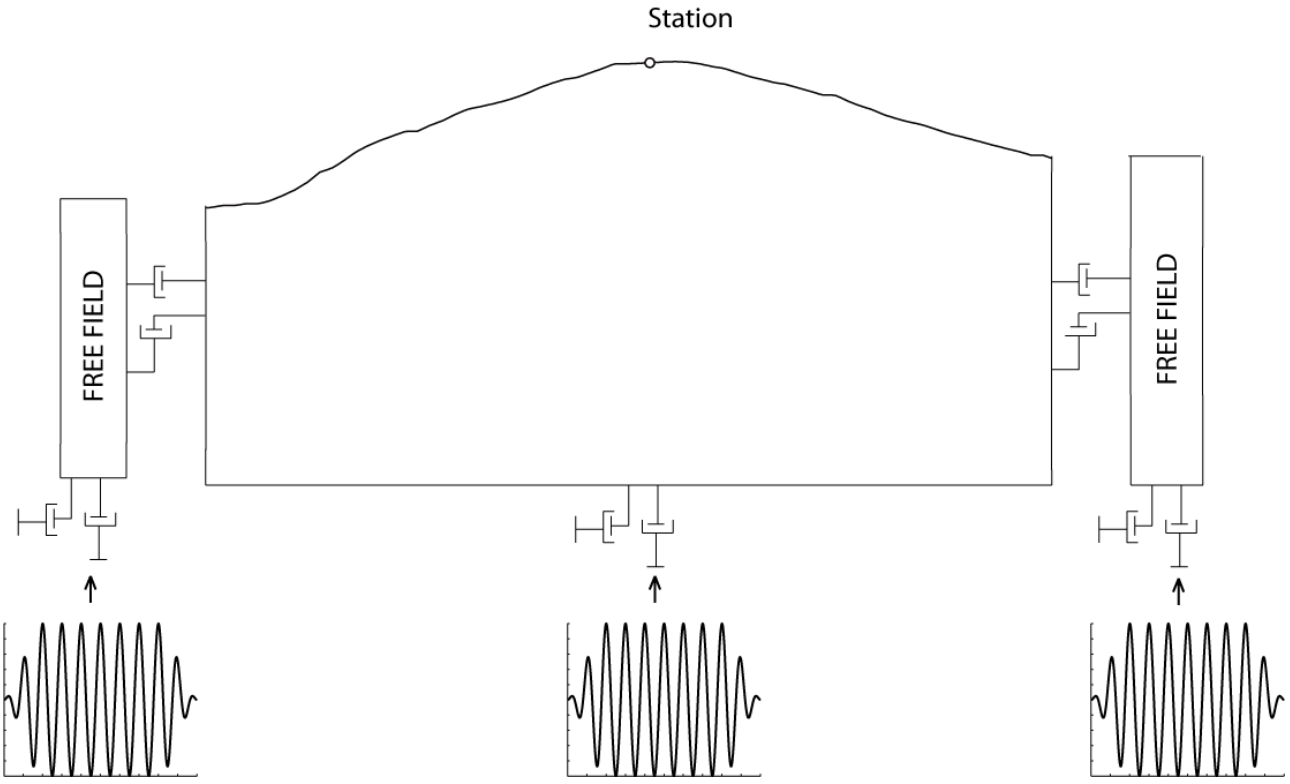

Figure 5.3: Schematic illustration of the finite difference model used for the analysis at one of the recording stations. The model uses a realistic topographic cross-section profile at the top. The station is located at the surface, equidistant from both the lateral edges. The height of stations from the base is same for all stations. Free-field boundary condition is applied to the lateral boundaries and quiet boundary conditions are applied at the base.

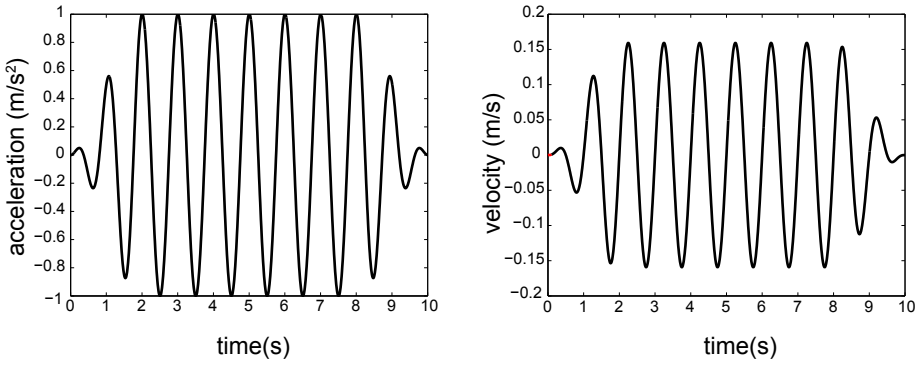

Figure 5.4: Acceleration and velocity time history for input motion with 10 cycles, and a period of 1 s.

velocity time histories are shown in Figure 5.4. The input motion is baseline corrected by default.

After the first set of runs, we discovered that the ground motion reaches steady state within the first 5 or 6 cycles, and so subsequent analyses were performed using 10 total cycles of input motion. Table 5.1 summarizes the modeling details for each of the approaches. To study the influence of the frequency of input motion on the resultant amplifications, we use the following three periods of input motions; 0.5 s, 1 s, and, 2 s. We selected these three periods because they are well within the periods of interest for typical structures, and they also lead to reasonable run times in our analyses. For the analysis with the 2.0 s input wave, we double the mesh length and width in both directions, so that the sites are at similar wavelengths-distance away from the boundary, as in the case of 1.0 s input motion. For this reason, we did not perform the 2.0 s analyses for Approach 2, as it became very computationally expensive.

We verify the numerical accuracy of our methodology with the case of a simple 2D slope from Bouckovalas and Papadimitriou (2005) subjected to an input sinusoidal motion. They used a 50 m high slope with an angle of 30 degree. We selected this study because the authors had also used FLAC software for their analyses, and so the differences from using different software are minimized. For this verification study, we first generate a mesh using

Table 5.1: Parameter values assumed for the analysis in the two approaches.

Parameters	Approach 1	Approach 2
V_s (m/s)	500	V_{s30}
Grid size (x, y) (m)	$\lambda/10$	10
Damping (%)	5	2
Number of cycles	10, 20	10
Density, ρ (Mg/m^3)	2	2
Poissons Ratio, ν	1/3	1/3

properties listed in Table 1 for Approach 1, and input a sinusoidal acceleration at the base using Equation 5.1. We also perform two 1D analyses, one with a column height equal to the height at the crest, and other with a height equal to the height of the base of the slope. Using the recorded time histories at the surface of the mesh, we compute the steady state PGA from the 2D analysis, and normalize it with the steady-state PGAs from the two 1D analyses. Figure 5.5 shows a comparison of the results from Bouckovalas and Papadimitriou (2005), and the results from our analysis. The results agree well with each other at all distances from the station.

5.3.3 Parameter computation

We compute the steady-state Normalized PGAs at every station in six directions. We use the term $NormalizedPGA$ for the ratio of the PGA from the 2D analysis to the PGA from the corresponding 1D analysis with a flat top surface. Figure 5.6 shows the variation of normalized PGA at the surface of Tamalpais peak station computed using Approach 1 (Figure 5.2). A pattern of amplification and de-amplification emerges along the surface, and the distances over which these variations occur seem to be proportional to the wavelength of input motion. This observation is consistent with observations made by Bouckovalas and Papadimitriou (2005). Note that the amplifications do not reach one as we approach the edge of the mesh, indicating that the free-field distances have not been attained yet.

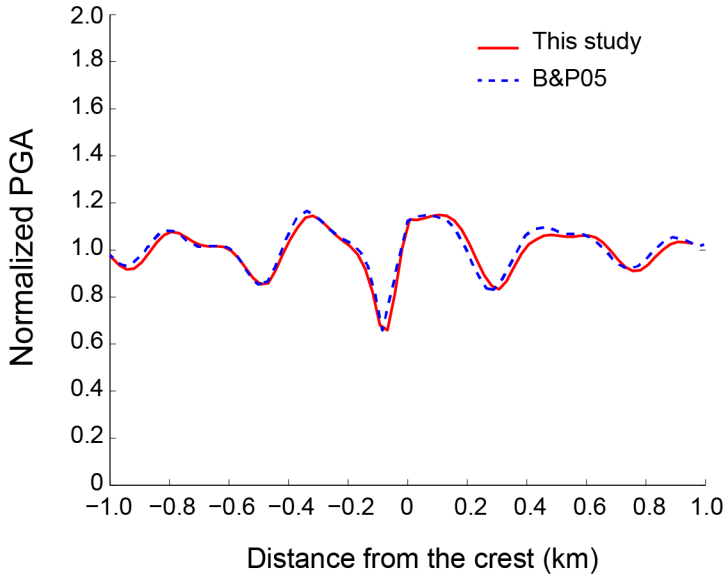

Figure 5.5: A comparison of the normalized PGA values obtained from our analysis, with the Bouckovalas and Papadimitriou (2005) for the case of 30 degree slope, height 50 m, damping less than 5 % and an input acceleration of $H/\lambda = 0.2$.

However, this should not be a concern for this study, as our objective is not to predict an exact amplification at the site but to simply obtain topographic proxies that can be used for a regression exercise to predict topographic effects in ground motions.

Using these Normalized PGA values, we compute the sets of parameters listed in Table 5.2. The term *lnAmp* in the table is used to define the natural logarithm of the Normalized PGAs. We use the natural logarithm of amplification as a parameter because the ground motion residuals are also in the units of natural logarithm of spectral accelerations, and we are using these parameters to model the trends in these residuals. To compute $lnAmp_{par}$ and $lnAmp_{perp}$, we compute the angle of the line joining the station to the earthquake hypocenter. Using this angle, we select one of the six azimuths where we obtained the cross-sections and select the one that is closest to the computed angle, and assign the corresponding *lnAmp* value in that direction to $lnAmp_{par}$. We repeat this process to obtain the $lnAmp_{perp}$, this time selecting the azimuth closest to the angle perpendicular to the computed angle. The parameterization results in a total of 15 *lnAmp* parameter at each site from Approach 1, and

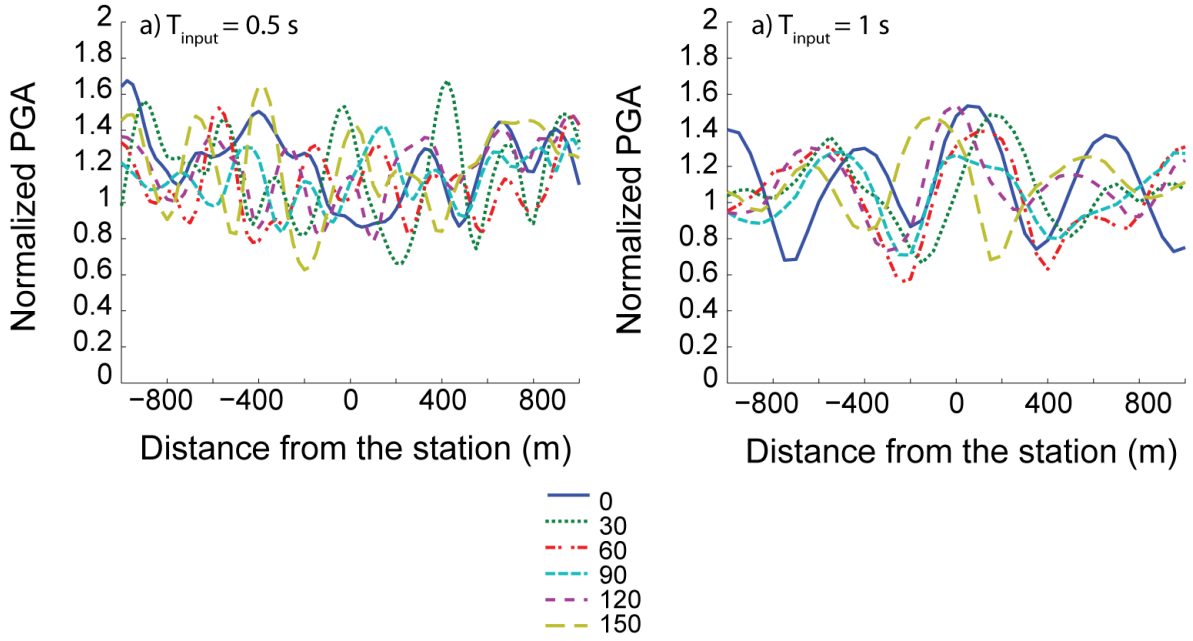

Figure 5.6: Normalized PGA values for the six cross-sections of the Tamalpais peak B station are shown for an input wave of period 0.5 s, and 1 s. The amplification and deamplification patterns seem to emerge at distances proportional to the wavelength of input motion. Also note that the normalized PGA values are greater than one for most of the azimuths, indicating an average amplification at the station.

Table 5.2: List of parameters used as a proxy for topographic effects.

Parameter	Description
$lnAmp_{avg}$	Avg. of the 6 *lnAmp* values
$lnAmp_{min}$	Min of the 6 *lnAmp* values
$lnAmp_{max}$	Max of the 6 *lnAmp* values
$lnAmp_{par}$	*lnAmp* value in the direction of incoming waves
$lnAmp_{perp}$	*lnAmp* value perpendicular to the incoming wave

a total of 10 *lnAmp* parameter from Approach 2. Using these parameters, we now study the trends in the intra-event residuals from the Chiou and Youngs (2014) GMPE to determine if one or more of these parameters can predict biases in the residuals.

5.3.4 Residual Analysis

Figure 5.7 shows the pair-wise correlation between *lnAmp* parameters from Approach 1, for wave period of 1 s. The figure shows that most of these correlations are moderate with only three pair of parameters showing higher correlations (R >0.8). Similar relationship between parameters was observed for input wave periods of 0.5 s, and 2.0 s. As the correlations are not very high, the parameters most likely have different predictive powers. We use all of these parameters to fit models to explain the bias in GMPE predictions. Residual analyses of these models are performed to find the parameter with most predictive power. We fit smoothed models to the intra-event residuals from the Chiou and Youngs (2014) GMPE by performing loess regression (Cleveland et al., 1992) as a function of each of the *lnAmp* parameters. For this regression, we use an α value of 1. α controls the degree of smoothing in the loess regression, with a value of 1 resulting in maximum smoothed function. We use this setting to avoid over-fitting the data. The coefficient of determination, or the R^2 values from these regressions are shown in Figure 5.8. Note that these R^2 values are of the order of 0.01, which seem to be rather small. These models are only accounting for a site-specific effect, thus we are only reducing a part of the total variance i.e. the site-specific variance.

Other components of variability are still present in the residuals, even after removing the site-specific biases. Therefore the overall reduction in the variability of the intra-event residual after the regression is small. Even though these R^2 values are small, they still provide information about the relative predictive abilities of individual $lnAmp$ parameters. We can see in Figure 5.8 that out of the 5 $lnAmp$ models, the $lnAmp_{avg}$ model has the highest R^2 value overall. Also, within the $lnAmp_{avg}$ parameter, at shorter periods (T <0.65), the T = 1 s waves results in highest R^2, and for larger periods, the T = 2 s waves gives highest R^2. This shows that the amplification experienced by higher period wave better predicts the spectral amplifications at higher periods, and vice versa.

We repeat this analysis for the two periods in Approach 2 (Figure 5.9). Again, we find that the $lnAmp_{avg}$ regression leads to the higher R^2 values, compared to the other 4 parameters. We also note that for periods less than 1.2 s, the T = 0.5 s wave results in higher R^2, whereas for periods higher than 1.2 s, the T = 1.0 s wave gives higher R^2. This trend is similar to the trends observed for approach 1, where the lower period wave was better modeling amplifications for lower periods and vice versa. Comparing the R^2 values from $lnAmp_{avg}$ regression from the 2 approaches, we find that the T = 1.0 s from Approach 1 performs relatively better than the T = 1 s wave from Approach 2. We therefore use the $lnAmp_{avg}$ parameter from Approach 1 for further analysis.

We compare the $lnAmp_{avg}$ from Approach 1 and the H_d parameter computed in Chapter 4 to evaluate their correlation. Figure 5.10 shows the scatterplot of the $lnAmp_{avg}$ values for wave period of 0.5 s and 1 s with the H_d for d = 500 m. The figure shows a strong linear correlation between the two variables at both the periods. We compute the correlation coefficient between $lnAmp_{avg}$ and the corresponding H_d for d ranging from 250 m to 3000 m (Figure 5.11). The overall high coefficients of correlation (R >0.8) indicate that the $lnAmp_{avg}$ is highly correlated with H_d at most scales (d). The figure also shows that the correlations between $lnAmp_{avg}$ parameter and H_d parameter are maximum at a scale, d, of: 250 m for the 0.5 s wave, 500 m for the 1.0 s wave, and 1000 m for the 2 s wave. In the analyses, we assumed a shear-wave velocity of 500 m/s for all the mesh, and so the

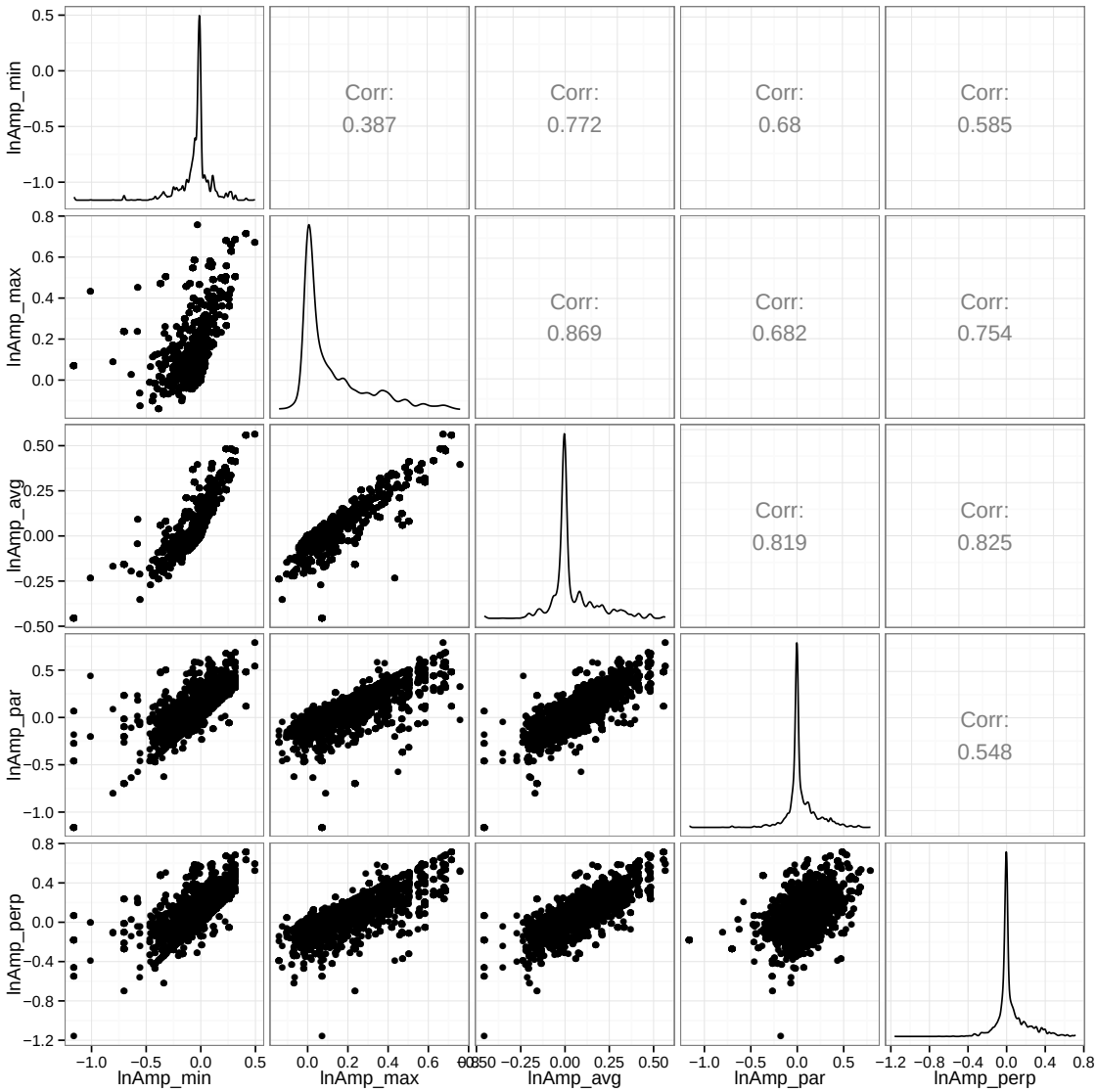

Figure 5.7: Scatterplot of the five *lnAmp* parameters computed using wave of period T = 1s, and a constant shear wave velocity of 500 m/s for each station. The corresponding values of correlation coefficients, and the parameter density plot is also shown.

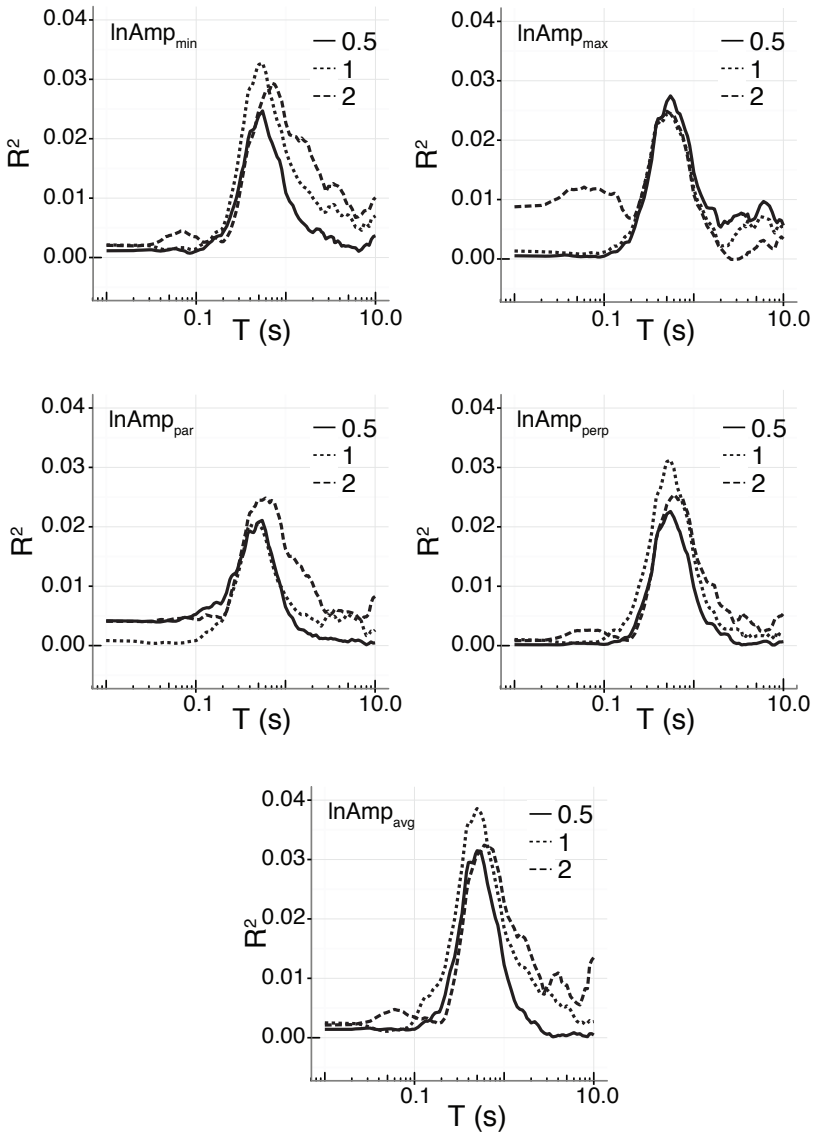

Figure 5.8: Coefficient of determination (R^2) value from the loess regressions on the intra-event residuals from the Chiou and Youngs (2014) model using the five *lnAmp* parameters from Approach 1.

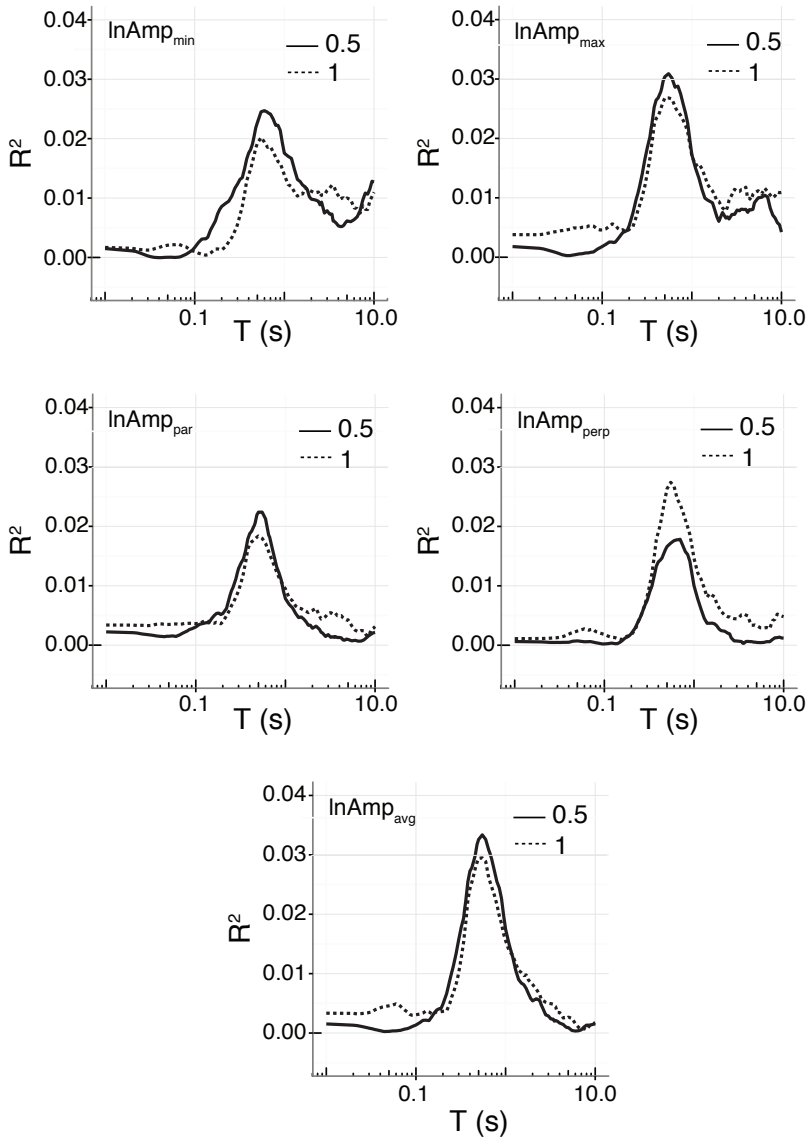

Figure 5.9: Coefficient of determination (R^2) value from the loess regressions on the intra-event residuals from the Chiou and Youngs (2014) model using the five *lnAmp* parameters from Approach 2.

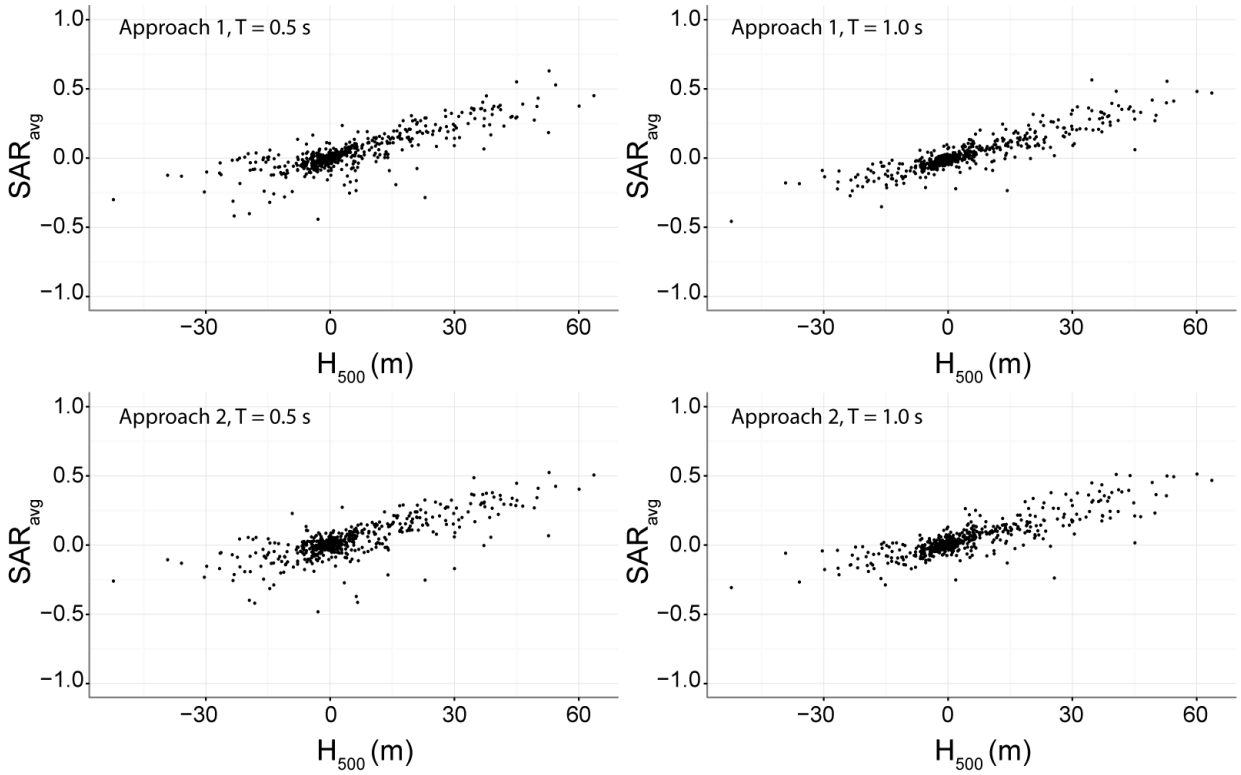

Figure 5.10: Scatterplot of the $lnAmp_{avg}$ obtained from the analysis using Approach 1, and Approach 2, and the relative elevation parameter H_d for d values of 250 m - 3000 m. The relative elevation parameter was computed at each site in Chapter 4. The values are shown for an input wave of period 0.5 s and 1 s.

corresponding wavelengths ($\lambda = V_s T$) for the three waves are 250 m for the 0.5 s wave, 500 m for the 1.0 s wave, and 1000 m for the 2 s wave. Note that these values are same as the scales or the H_d parameter at which correlations are maximum for the three cases. This is an important result as it shows that the relative elevation parameter at a given scale is effectively modeling the average elastic 2D amplification of a wave of wavelength equal to the scale. A similar observation was made by Maufroy et al. (2015) who noted that the amplifications at a site were highly correlated with the smoothed curvatures. However, this result also contrasts the findings from Burjánek et al. (2014) who found that the amplifications at sites with pronounced topography are mostly controlled by sub-surface shear wave velocities and not so much by the shape of the topographic feature.

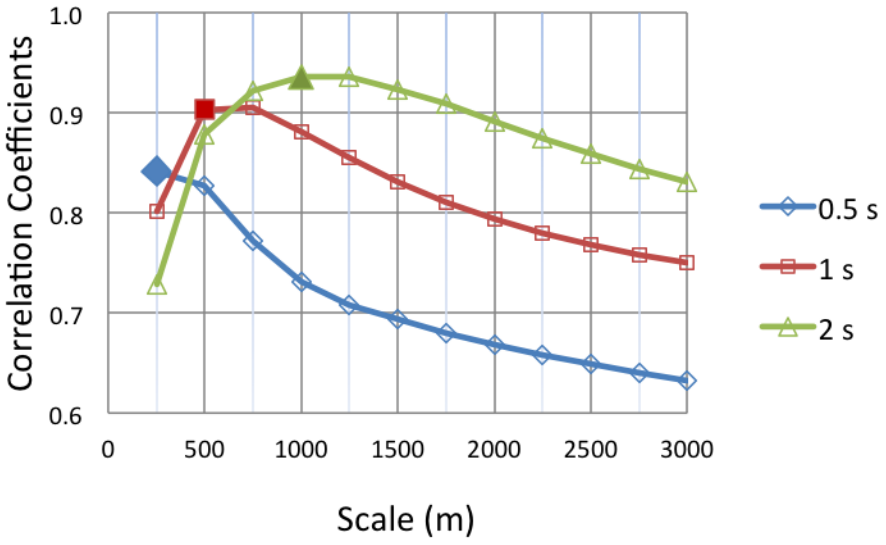

Figure 5.11: Correlation coefficient values between $lnAmp_{avg}$ from Approach 1, and H_d for d values of 250 m - 3000. The three lines correspond to input wave of period 0.5 s, 1 s and 2s. The peaks are shown with solid symbols. Note that the respective peaks occur at a scale equal to the wavelength of the input motion.

The high correlation between H_d parameters, and the $lnAmp_{avg}$ parameters shows that the two parameters have similar information and they both capture topographic information. Thus they should have very similar predictive power. Figure 5.12 shows a comparison of the predictions from the two loess models, one fitted using the H_{1000} parameter and the other fitted using the $lnAmp_{avg}$ parameter for an input wave of period 2 s. The comparison is shown for periods of 0.3 s, 0.5 s, 1 s, and 3 s. The predictions from the two models are similar at all periods and the points are scattered around the x = y line, suggesting that both parameters express similar information about topographic amplification at the sites. The two predictions match more closely in the first quadrant, where both predictions are positive, and less so in the 3rd quadrant, where both predictions are negative. To determine which of the two models performs better at reducing topographic biases in the residuals, we compare the R^2 values from the two fitted models (Figure 5.13). The figure shows that the loess model fitted using the H_{1000} parameter resulted in higher R^2 value at most periods. This suggests that the H_{1000} parameter is able to model the trends in the ground motion

residuals comparatively better than the $lnAmp_{avg}$ parameter computed for 2 s wave. Since computing the $lnAmp_{avg}$ parameters are expensive, and we are not gaining any additional reduction in the standard deviation in the fitted model, we recommend using the model presented in Chapter 4 for predicting topographic effects at a site.

5.4 Summary and Conclusion

We developed a family of parameters at the recording stations from the NGA-West2 dataset to predict topographic effects on response spectra. We used simplistic 2D numerical analyses to compute approximate estimates of topographic amplifications at the sites in multiple orientations, and used the natural logarithm of these amplifications to develop five different parameters at each site. We used two different approaches for modeling the meshes. In the first approach we used a constant shear wave velocity of 500 m/s at each site, and in the second approach we used a shear wave velocity equal to the shear wave velocity in the upper 30 m of the soil at the site of the recording station. The second approach is used to partially address the effect of different shear wave velocity on topographic effects. We compared the predictive powers of these parameters to find which one most effectively models the errors in the ground motion prediction from the GMPE. Our analysis showed that the natural logarithm of 2D amplification at the sites from the first approach, when average over multiple orientations, most effectively predicts the biases in the intra-event residuals from the GMPE. The amplifications from 1 s wave was found to more effectively model shorter period (T <0.65) residuals, and the amplifications from 2 s wave was found to more effectively model longer period residuals.

We compared the mean of the natural log of 2D amplification from multiple orientation or the $lnAmp_{avg}$ values at sites with the relative elevation parameter or H_d at the same sites from Chapter 4, and found that the two are highly correlated at most values of the scale parameter d. We found that for a given wavelength of input motion, the correlations between

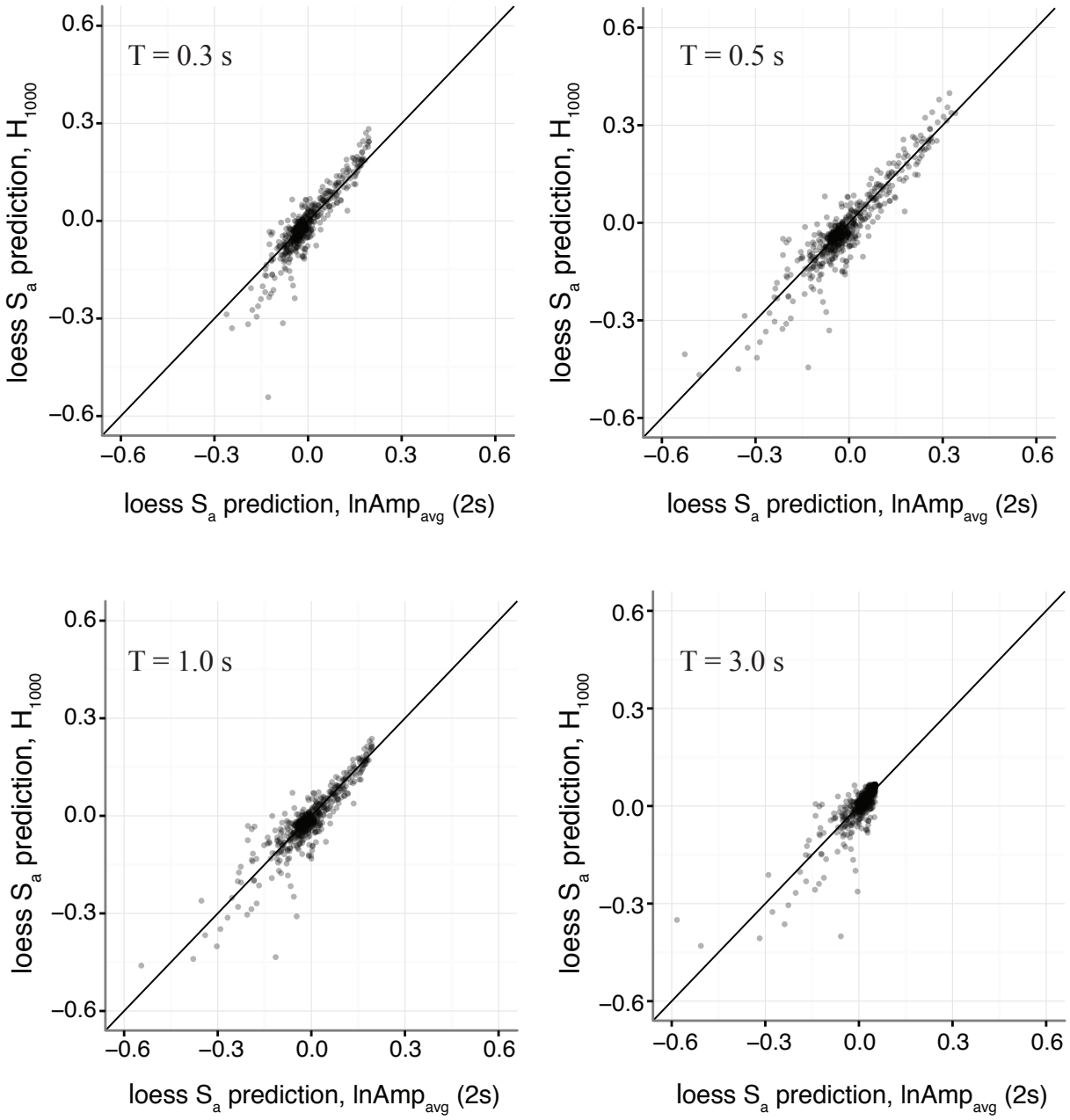

Figure 5.12: Scatterplot of loess predictions using the geometry-based H_{1000} parameter, and the predictions from loess regression using the $lnAmp_{avg}$ parameter for the 2 s wave, at periods of 0.3s, 0.5s, 1.0 s, and 3.0 s. The solid black line shown the x = y line.

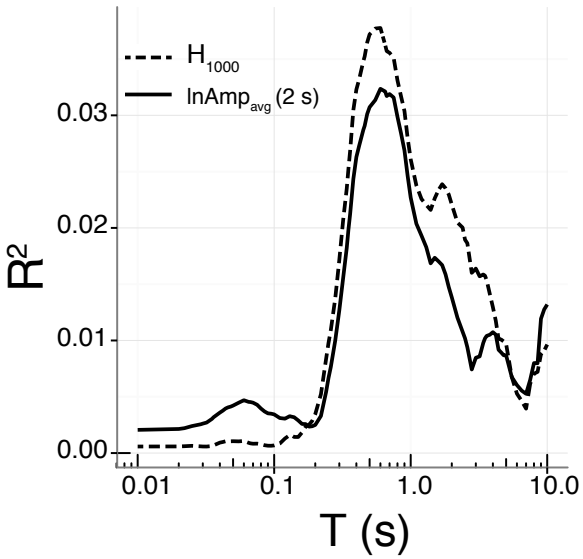

Figure 5.13: Comparison of the coefficient of determination (R^2) values from the loess regressions on the intra-event residuals using the H_{1000} parameter, and the $lnAmp_{avg}$ parameters using 2s input wave.

resulting $lnAmp_{avg}$ values and the H_d parameter reaches a maximum at a d value equal to the wavelength of input motion. This result shows that there is a physical link to why the relative elevation parameter was able to model the topographic bias in the data, and the reason is that it is effectively modeling the elastic 2D amplification at the sites. Due to the high correlations between the $lnAmp_{avg}$ parameter and the H_d parameter used in Chapter 4, the two parameters carry very similar information, and also have very similar predictive power. The coefficient of determination from the loess model was comparatively larger when we use the relative elevation parameter instead of the $lnAmp_{avg}$ parameter, and so we suggest using the model from Chapter 4 for predicting topographic effects. The proposed approach for computing topographic effects, that is of using simplistic numerical models to obtain parameters that are later used in regression analyses, can also be used for capturing other effects of site amplification.

5.5 Acknowledgment

This research was partially supported by the National Science Foundation under award number CMII 1132373 and the California Strong Motion Instrumentation Program (CSMIP) under Standard Agreement No. 1012-955. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We are grateful to Prof. Dominiki Assimaki for her guidance on the project.

Chapter 6

Conclusion

This dissertation focused on developing models to predict topographic effects in ground motions. The following sections summarize briefly the important findings and limitations, and suggest future work related to this dissertation.

6.1 Empirical findings and contributions

Topographic features affect earthquake ground-motions. Yet, the current literature on this subject is inconclusive on how to account for these effects in the predictions of ground-motion intensity. Including these effects into ground-motion predictions will improve the accuracy of seismic hazard computations at sites with significant topographic relief, which in turn will reduce earthquake risk, and improve the resiliency of communities located in areas with topographic relief.

To include the effect of topography on ground-motion, we first need to parameterize this effect. Previous studies have identified that parameters such as curvature and slope of the surface controls the response of the feature at a site to the ground-motion. However, such findings are typically, based on either the numerical analysis of a region of a limited extent,

or on the numerical analyses of simplified geometries. Additionally, the response at the sites in such studies is not always verified with recorded data. As a result, these results cannot be used for predictions. We addressed the issue of topographic parameterization at a site by developing two types of site parameters. The first set of parameters was computed using the site geometry. These parameters were developed using the elevation data around the site, and comprised of the smoothed curvatures, smoothed slopes, and relative elevations, at several scales. Chapter 3 and 4 of this dissertation cover the details of these geometry-based topographic parameters. The second set of parameters was computed using simplistic 2D numerical analysis at the site. To compute these parameters, we computed 2D amplification due to surface topography at a site from multiple orientations, and used these amplification values to develop five different parameters. The details of this parameterization can be found in Chapter 5 of the dissertation.

To quantify the effect of surface topography on ground-motions, we performed residual analysis to determine if there were any significant trends in the ground-motion model residuals with respect to any of the above parameters. For the residual analysis, we used two different ground-motion datasets. In Chapter 3, we studied the California small-to-medium magnitude earthquake dataset that consisted of spectral accelerations at two periods. In Chapter 4, we extended the study to the much larger NGA-West2 dataset, with spectral accelerations available at 105 periods between 0.01 s and 10.0 s, which covers most periods of engineering interest. Since the ground-motion recordings were available at multiple periods, the NGA-West2 dataset provided us an opportunity to examine the effects of changing the scale that was used to compute geometric parameters, on periods where topographic effects were prominent in the data. Changing the scale in the parameter computation affected the types of geometries extracted. For example, we found that at smaller scales, smaller and narrower topographic features were extracted, whereas at larger scales, bigger and wider features were extracted. It was important to check the effects of scales on periods because past studies have shown that the size of the feature has an effect on the wave periods that are affected by topography, and we wanted to check that dependence in our data. We found,

in our analysis of the NGA-West2 dataset, that with an increase in the scale the spectral periods where topographic effects are observable increases slightly. However, this dependence was not found to be strong, and therefore, we selected to use the same scale for classification, at all periods.

In Chapters 3 and 4, we proposed two separate models to account for topographic effects in the median prediction. In each of these models, we proposed two correction factors at each period, corresponding to the high and low classes. The sites were classified as high or low, based on the value of the topographic parameter, and a fixed threshold. The proposed corrections were positive for high sites, and negative for low sites, showing that the ground-motion residuals were biased with respect to the considered topographic parameter. These biases were in agreement with the expected behavior, given that the high class mostly corresponded to sites on the top of the hills, or on the ridges, where ground motions are known to amplify, and the low corresponded to valleys adjacent to a hill, where ground motions typically de-amplify. In Chapter 4, we also proposed an estimate of the variability in the proposed correction factors for the high and low classes so that they can be incorporated in the probabilistic seismic hazard analysis (PSHA).

The model presented in Chapter 4 was an improvement over the model presented in Chapter 3, as the NGA-West2 dataset that was used to fit the model in Chapter 4 consisted of spectral accelerations for 105 periods, compared to only 2 periods in the SMM dataset used in Chapter 3. In Chapter 4 we evaluated regional differences in the trends in site residuals with respect to the parameters. However, due to lack of sufficient data points in each class for every region, we could not establish the regional differences. As a result, we proposed a single model for all regions.

In Chapter 5, we extended the analysis presented in Chapters 3 and 4, and used the second type of parameter (based on the Finite difference analyses) to quantify topography at the sites. We computed 2D amplifications due to surface topography at each of the sites in the NGA-West2 dataset in six different orientations and used them to compute five

different parameters at each site. Using residual analysis, we compared these parameters for their predictive power, and found that the log of the amplifications, when averaged over multiple orientations, most effectively predicted the biases in the site residuals from ground-motion models. We found that the amplification of a lower period wave was able to more effectively predict biases at shorter periods, and amplifications of the higher period wave better predicted biases at higher periods. Although, we observed in Chapter 5 that the above parameter were highly correlated with the previously studied geometry-based parameters at the same sites from Chapter 4, and hence the value of the Finite difference based parameter is diminished considering that the geometry based parameters have a stronger predictive power. The correlations between the numerical analysis based parameters and the geometry-based parameters were found to peak when the geometry-based parameters were computed at scales equal to the wavelength of input motion used to compute the numerical analysis based parameters. This is an important result, as it shows that the geometry-based topographic parameters used in Chapters 3 and 4, were essentially modeling elastic amplifications due to surface topography. The proposed approach, of using the output of simplified numerical models as input to regression analyses, is a novel approach that could have applications in other areas of site effects prediction.

6.2 Limitations

6.2.1 Data Limitations

Following are the limitations of this study due to the constraints of the data:

- One of the major limitations of our study is that it is empirical in nature. As a result, the quality of these models depends on the quality of the data used to fit them. Also, the applicability of the model is limited by the range of the data used to fit the model, and extrapolation can be an issue.

- For the residual analysis, we only used recording stations with three or more measurements on them, in order to effectively constrain the site terms. Constraining the site terms was important as topographic effects are site-specific, and we wanted to ensure that we are correctly estimating the site-terms at every site. Although, the NGA-West2 dataset is a large global dataset, this filtering eliminated many stations from the analysis, and only retained stations in California and Japan. Consequently, our analysis in Chapter 4 is limited to these two regions, and is only applicable to these regions. Similarly, in Chapter 3 we studied the Small-to-medium magnitude earthquake dataset, with recordings from only within California, and so the same limitations applied there.

- There were a fewer numbers of ground-motion stations in a region of significant topographic relief, as compared to stations on relatively flatter terrains. Therefore, the variability in the estimated median value of the topographic biases was higher for stations with significant topographic relief. These values could have been more realistically constrained if a database with a more uniform distribution of recording station was available. Moreover, most of the sites in Japan were classified as low, and due to a non-uniform distribution of recording stations with respect to topography, we could not effectively quantify the regional variation in topographic effects.

- For computing geometry-based topographic parameters at the Japanese station, we used the ASTER GDEM v2 elevation data. Although this dataset is the most updated, and the only global digital elevation model data available at a high resolution, it is known to be less accurate than the national elevation dataset that was available for California. As a result, the parameters computed from the elevation data might contain some amount of error. The amount of this error may control the parameter values at these sites, and in turn affect the regression results. With the availability of a higher resolution, and more accurate data, the parameter values can be estimated more accurately, and any potential errors could be minimized.

6.2.2 Modeling limitation

The models presented in Chapter 3 and 4 were developed using the ground-motion model residuals. We chose to study the residuals from the ground-motion model and propose corrections to an existing ground-motion model, rather than re-fitting a full model because fitting a ground-motion model is a more challenging task. In fitting a full model, one would need to account for any possible interactions between different effects, and ensure that the models extrapolate well outside the range of the empirical data used for fitting. This generally includes using theoretical constraints, and supplementing empirical data with numerical simulation. The models we presented in Chapters 3 and 4 are pure empirical fits to the ground-motion residuals, and only a small amount of smoothing was applied to the coefficients from the mixed-effects regression. Though, our models are a big first step towards quantifying topographic effects, and subsequently including these effects in future seismic risk assessment, additional research might be necessary before these models can be used universally. We would also like to stress that the proposed corrections are the average effects over a wide range of topographic feature with similar values of the parameter, and so they might not accurately predict the effects at a specific site. At sites where more accurate predictions of site effects are required, conducting a site-specific analysis might be a better choice.

6.3 Future work

6.3.1 Studying other ground motion datasets

In this dissertation, we studied the effects of topography on ground-motion by only using recorded measurements from stations in California, and Japan. An extension of this study would be to look at ground-motion datasets from other regions to determine how these results extrapolate in regions outside California and Japan.

6.3.2 Ground-motion directionality

Previous studies have shown that the ground-motions can be polarized in the presence of a topographic feature, resulting in increased ground motion intensity in one orientation (Spudich et al., 1996; Burjánek et al., 2014). Since our primary goal was to include the effects of surface topography in ground-motion prediction equations, we only studied residuals of the spectral accelerations in the median orientation, which does not account for ground-motion directionality. An interesting study would be to perform a polarization analysis of ground-motion to determine the amount of polarization caused by these topographic features, and to try to quantify them.

6.3.3 Topographic effects on vertical component of ground-motion

Topography also affects earthquake ground motions in the vertical direction, and may cause high vertical ground-motions. These amplifications can cause heavy infrastructure damage in the surrounding regions. Studying a large database of ground motions in the vertical direction can help us to better understand and quantify these effects.

6.3.4 PSHA

Our model proposes corrections to the median ground motion predictions for sites with significant topographic relief. These models would impact the hazard estimate at a site in such areas. It would be meaningful to see how our predictions impact the overall hazard estimate at the sites around regions of topographic relief.

6.4 Concluding remarks

This dissertation proposed empirical models to predict the effects of surface topography on earthquake ground motions. These models are proposed as corrections to the current GMPE. We propose two period-dependent correction factors for each spectral period, along with their standard deviations. These correction factors are positive for sites with a larger value of the topographic parameter suggesting amplifications at the site, and negative for sites with a larger negative value of the topographic parameter suggesting de-amplifications at the sites. Although, care has been taken to ensure that the models extrapolate well outside the range of the data, it should be noted that these models are pure empirical fits to the ground-motion model residuals, and so more research would be necessary before these results can be applied to other sites outside the regions studied here. Also, the proposed correction pertains to the average effects of topographic features that satisfy a certain criteria, and so if more accurate predictions of the effects are required, a site-specific analysis might be more suitable.

Appendix A

Sensitivity analysis

To determine the effects of changing the threshold of minimum number of recordings per station for selecting the dataset for analysis (N), we re-run our regressions using three other, higher values of N.The resulting mean site residuals, $\delta S2S$, and the corresponding site-to-site variabilty (ϕ_{s2s}) for the three topographic classes are shown below. The recording stations are classified using a scale (d) of 1500 m, and a threshold (t) of 0.5.

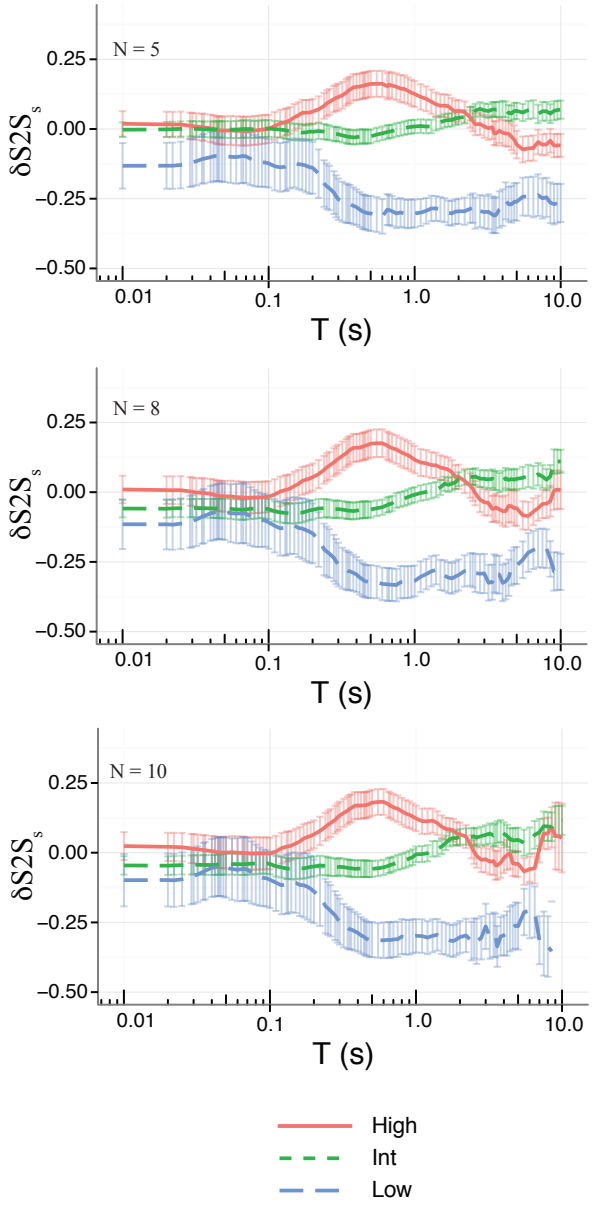

Figure A.1: Variation of mean $\delta S2S$ for the three topographic classes as a function of N, where N is the minimum number of recordings at a site for it to be included in the study. +/- standard deviation error bars are also shown.

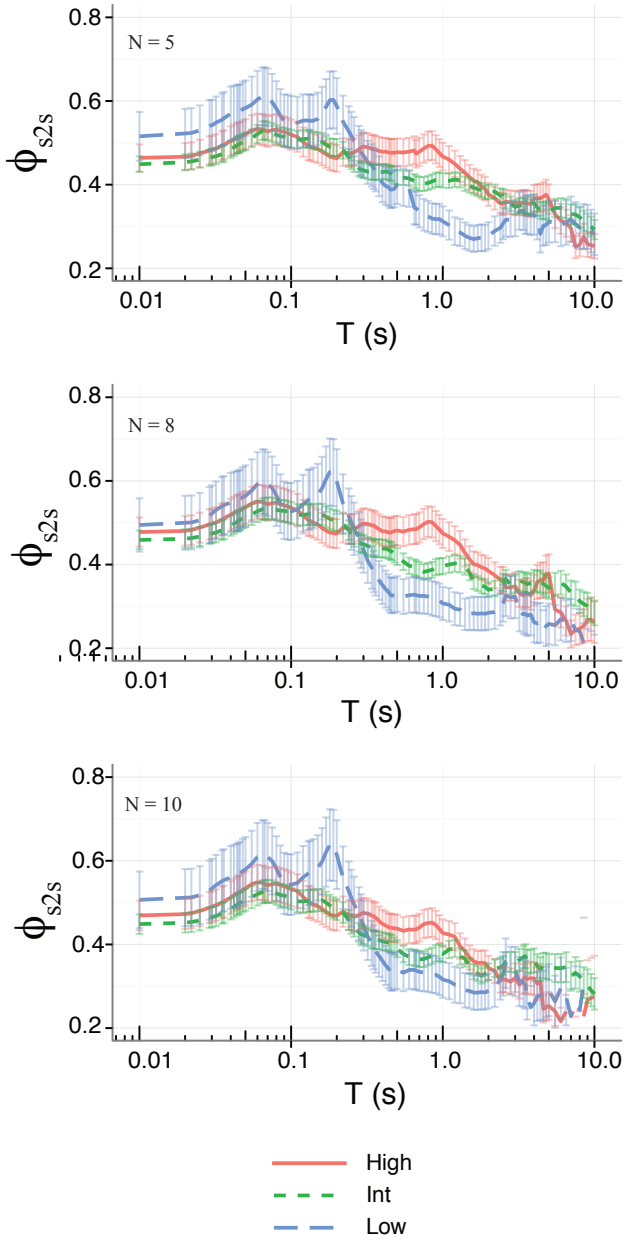

Figure A.2: Variation of ϕ_{s2s} for the three topographic classes as a function of N, where N is the minimum number of recordings at a site for it to be included in the study. +/- standard deviation error bars are also shown.

Appendix B

Spatial clusters in site residuals

To do this analysis, the $\delta S2S_s$ values from Equation 4.3 are used. The cluster analysis process comprises of 3 steps:

1. Finding out if there is a significant spatial clustering in the data. This is done by computing a global statistics which quantifies the presence of spatial clusters across the entire dataset. There is a tool that does this analysis in ArcGIS is called the Global Morans I tool. The tool returns a p-value and a z-value indicating the significance and intensity of clustering in the data, respectively. A positive z-value indicates that the data is clustered and a negative value indicated the data is dispersed.

2. If there is statistically significant clustering, the next step is to determine the scale or distance at which the clustering is most prominent. The scales can be chosen in two ways:

   (a) Based on the theoretical understanding of the phenomenon.

   (b) Using an iterative process to determine the scale or distance at which the clustering occurs in the dataset. This can be done using the incremental Global Morans I tool in ArcGIS.

Table B.1: The global statistics from the Global Morans I tool on stations in the three topographic classes for California and Japan.

	CA			Japan		
	high	int	low	high	int	low
Moran's Index	0.173	0.147	0.257	-0.064	0.137	0.850
Variance	0.070	0.001	0.197	0.021	0.004	0.209
z-score	0.683	4.150	0.616	0.252	2.228	1.913
p-value	0.495	0.000	0.538	0.801	0.026	0.056
Dist. threshold(km)	127.1	81.9	149.3	164.2	31.0	49.2

3. Apply local statistics to identify hot and cold spots. Hot spots are areas with features of similarly high values and cold spots are features with similarly low values. This step can be done using Local Morans I tool in ArcGIS.

Figure B.1 shows the spatial variation of $\delta S2S_s$ (T = 0.5 s) values within the three classes. A spectral period of 0.5 s is chosen as the difference in mean site terms was found to be significant at this period. The standard deviation values used to color the points are standard deviation of site values within that class. As we are interested in finding out if there are clusters of stations with similarly high or low residuals within each topographic class, the stations are subset based on the topographic class and for each of the subset perform step 1. The Global Morans I tool requires that we specify a conceptualization of spatial relationship which specifies what kind of spatial relationships among features are used, and a threshold distance which is a cutoff distance beyond which features are ignored in the analysis. If this value is not provided, it is derived automatically to ensure that every feature has at-least one neighbor. We used an inverse distance spatial relationship. With inverse distance conceptualization, all features impact/influence all other features, but the impact decreased with the distance between the features. We do not use a threshold distance at this step, so it is computed automatically. The important statistics from this analysis is presented in Table B.1.

Each of the analysis outputs a p and z value that quantifies the significance and inten-

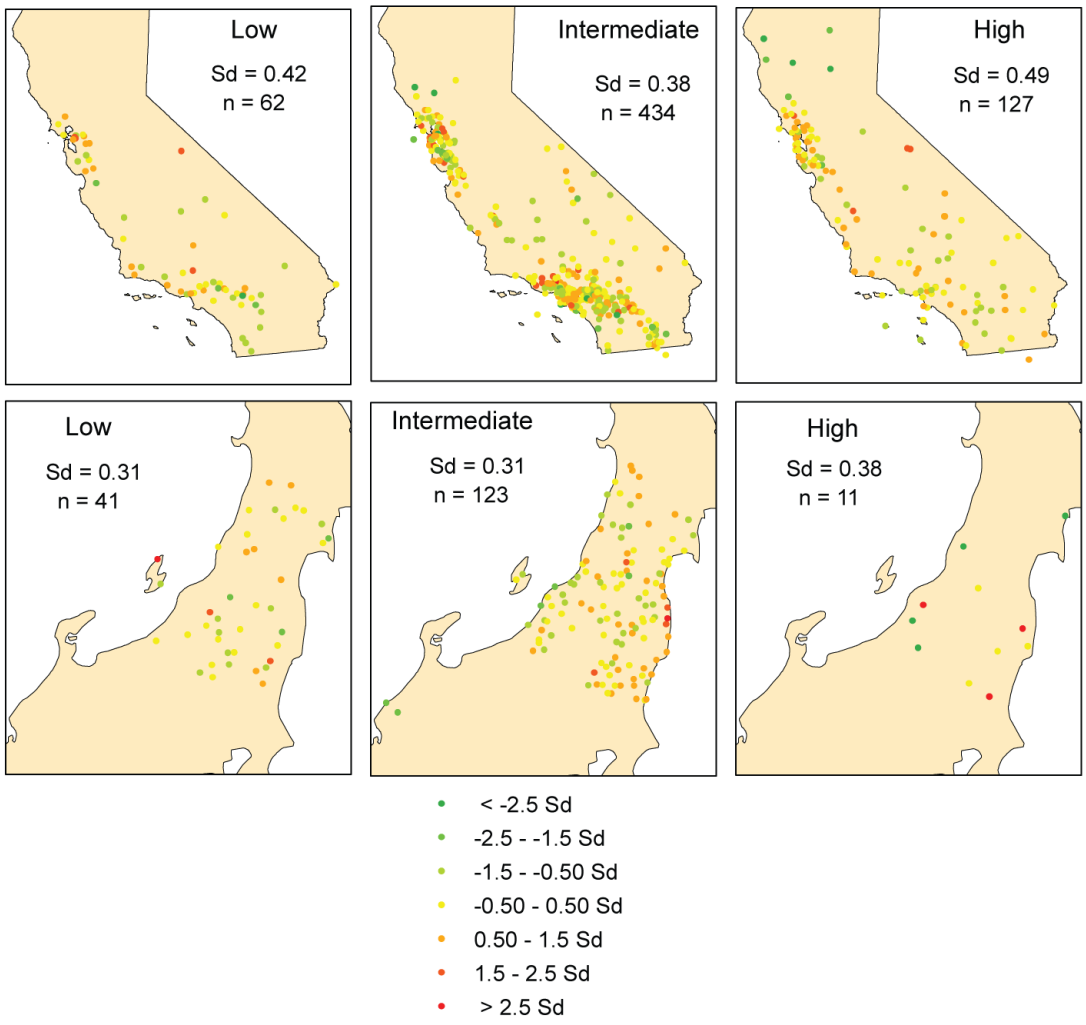

Figure B.1: The spatial variation of the site terms ($\delta S2S_s$) from Equation 4.3 are shows for the stations in the three topographic classes for California and Japan. The points are colored based on the site residual values at that station for T = 0.5 s. The standard deviation values used in the plots correspond to the standard deviation value of the site residuals within that class. The standard deviation of site terms and the total number of stations in each class are shown in the figure.

sity of clustering, respectively. Positive values of z indicate clustering, and negative values indicate dispersion. If a statistical significance is assumed at 95 %, we can see from the table that clustering is observed only for Japan and California for the stations in the intermediate class.

For the two cases where clustering is found to be significant, we compute the threshold distance or scale at which the clustering is most prominent. To do this, we use the incremental spatial autocorrelation tool, which measures spatial autocorrelation for a series of distances and reports the distances where most significant clustering is observed. Using this approach, the threshold distances are found to be 90 km for California and 50 km for Japan, for stations in the intermediate class.

Next, we perform a cluster and outlier analysis to determine the hot and cold spots in the dataset. Again, we use an inverse distance spatial relationship with a threshold distance obtained in the previous step. We can see in Figure B.2 that there is a hot spot in the southern part of California, denoted by black points and a cold spot around the coast in the northern part of California, denoted by blue spots. Similarly, we see a zone of hot spots in Japan along the eastern coastline, and a small cluster of cold spot on the western coast. This indicates that there is a spatial phenomenon at play at these locations that are amplifying and de-amplifying ground motions in the vicinity. With more data, it may be possible to understand the underlying phenomenon better. However, given that we did not observe any significant spatial clusters in the stations within the high and low topographic class, we can reject the hypothesis that there is any other spatial phenomenon influencing the ground motions in these two classes.

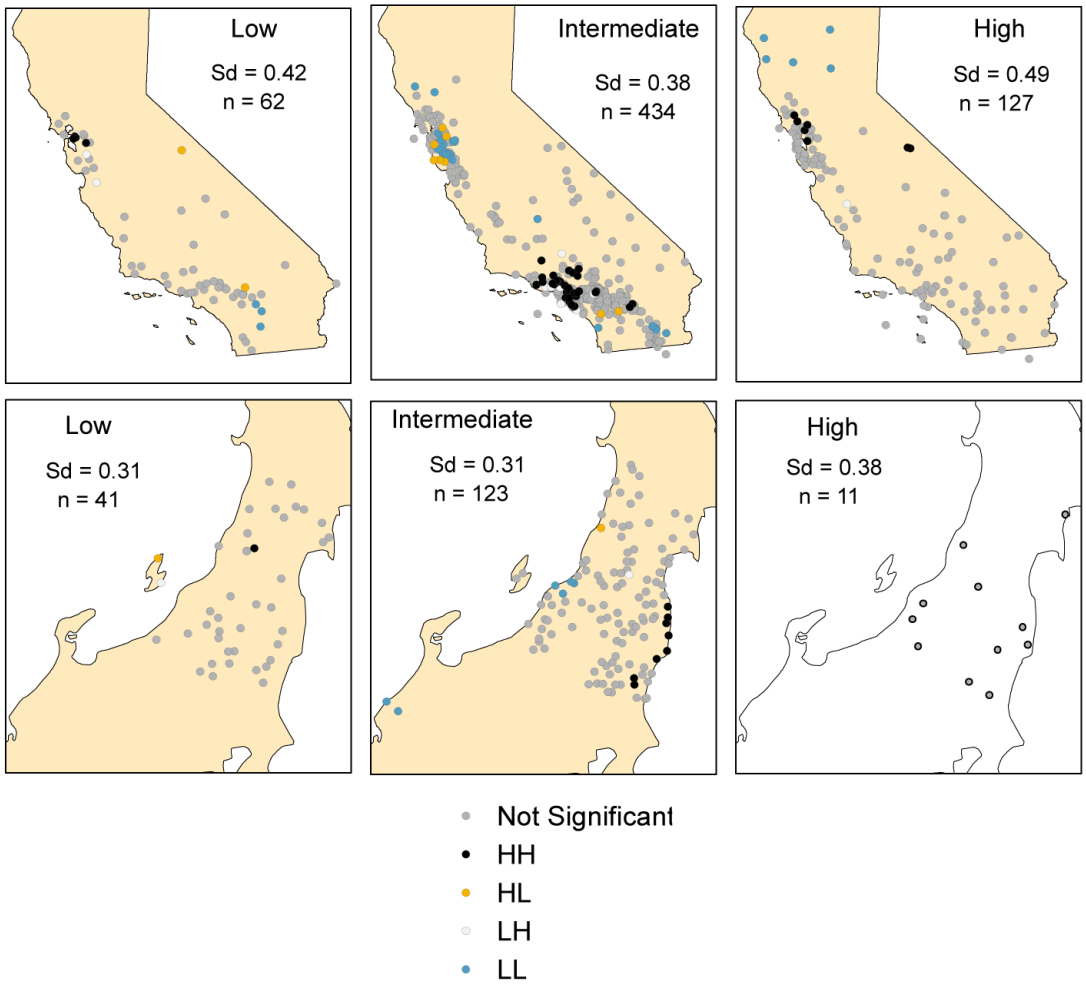

Figure B.2: High-high (HH), high low (HL), low-high (LL) and low-low (LL) clusters for stations in the three topographic classes in California and Japan. HH corresponds to stations which have a high site residual and are surrounded by similar stations. These clusters are also referred to as hot spots. LL corresponds to stations that have a low site residual is surrounded by stations with similarly low values of site residuals. Such clusters are called cold spots. HL and LH are stations that have high value of site terms and are surrounded by stations with low values of site terms or vice versa. These points are called as outliers. In Japan, there were too few stations in the High class to get a reliable clustering analysis result and therefore the results have not been shown. The points only mark the locations of the stations.

Bibliography

Abrahamson, N. A. and R. R. Youngs (1992). SHORT NOTES. *Bulletin of the Seismological Society of America 82*(1), 505–510.

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, T. Kishida, and J. L. Donahue (2014). Nga-west2 database. *Earthquake Spectra 30*(3), 989–1005.

Ashford, S. A. and N. Sitar (1997). Analysis of topographic amplification of inclined shear waves in a steep coastal bluff. *Bulletin of the seismological society of America 87*(3), 692–700.

Ashford, S. A., N. Sitar, J. Lysmer, and N. Deng (1997). Topographic effects on the seismic response of steep slopes. *Bulletin of the Seismological Society of America 87*(3), 701–709.

Assimaki, D. and G. Gazetas (2004). Soil and topographic amplification on canyon banks and the 1999 Athens earthquake. *Journal of earthquake engineering 8*(01), 1–43.

Assimaki, D., G. Gazetas, and K. E. (2005, June). Effects of Local Soil Conditions on the Topographic Aggravation of Seismic Motion: Parametric Investigation and Recorded Field Evidence from the 1999 Athens Earthquake. *Bulletin of the Seismological Society of America 95*(3), 1059–1089.

Assimaki, D. and S. Jeong (2013, September). Ground-Motion Observations at Hotel Mon-

tana during the M 7.0 2010 Haiti Earthquake: Topography or Soil Amplification? *Bulletin of the Seismological Society of America 103*(5), 2577–2590.

Assimaki, D., E. Kausel, and G. Gazetas (2005, November). Soil-Dependent Topographic Effects: A Case Study from the 1999 Athens Earthquake. *Earthquake Spectra 21*(4), 929–966.

Atik, L. A., N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn (2010, August). The Variability of Ground-Motion Prediction Models and Its Components. *Seismological Research Letters 81*(5), 794–801.

Bard, P. (1999). Local effects on strong ground motion: physical basis and estimation methods in view of microzoning studies. *Proceedings of the Advances Study Course Seismotectonic and Microzonation techniques in Earthquake Engineering: Integrated Training in Earthquake risk reduction practices*, 75–124.

Bard, P., C. Czitrom, J. Durville, P. Godefroy, J. Meneroud, P. Mouroux, and A. Pecker (1995). Guidelines for seismic microzonation studies. *Delegation of major risks of the French Ministry of the environment-direction for prevention, pollution and risks*.

Bard, P.-Y. (1982, December). Diffracted waves and displacement field over two-dimensional elevated topographies. *Geophysical Journal International 71*(3), 731–760.

Bates, D. and M. Maechler (2010). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-33.

Boore, D. M. (1972). A note on the effect of simple topography on seismic SH waves. *Bulletin of the Seismological Society of America 62*(1), 275–284.

Boore, D. M. (1973). The effect of simple topography on seismic waves: implications for the accelerations recorded at Pacoima Dam, San Fernando Valley, California. *Bulletin of the Seismological Society of America 63*(5), 1603–1609.

Boore, D. M., S. C. Harmsen, and S. T. Harding (1981). Wave scattering from a step change in surface topography. *Bulletin of the Seismological Society of America 71*(1), 117–125.

Bouchon, M. (1973). Effect of topography on surface motion. *Bulletin of the Seismological Society of America 63*(2), 615–632.

Bouchon, M. and J. S. Barker (1996). Seismic response of a hill: the example of Tarzana, California. *Bulletin of the Seismological Society of America 86*(1A), 66–72.

Bouckovalas, G. D. and A. G. Papadimitriou (2005, August). Numerical evaluation of slope topography effects on seismic ground motion. *Soil Dynamics and Earthquake Engineering 25*(7-10), 547–558.

Bozorgnia, Y., N. A. Abrahamson, L. A. Atik, T. D. Ancheta, G. M. Atkinson, J. W. Baker, A. Baltay, D. M. Boore, K. W. Campbell, B. S.-J. Chiou, and others (2014). NGA-West2 research project. *Earthquake Spectra 30*(3), 973–987.

Brillinger, D. R. and H. K. Preisler (1984). An exploratory analysis of the Joyner-Boore attenuation data. *Bulletin of the Seismological Society of America 74*(4), 1441–1450.

Burjánek, J., B. Edwards, and D. Fäh (2014). Empirical evidence of local seismic effects at sites with pronounced topography: A systematic approach. *Geophysical Journal International*, ggu014.

Burrough, P., P. van Gaans, and R. MacMillan (2000, July). High-resolution landform classification using fuzzy k-means. *Fuzzy Sets and Systems 113*(1), 37–52.

Celebi, M. (1987). Topographical and geological amplifications determined from strong-motion and aftershock records of the 3 March 1985 Chile earthquake. *Bulletin of the Seismological Society of America 77*(4), 1147–1167.

Celebi, M. (1991). Topographical and geological amplification: case studies and engineering implications. *Structural Safety 10*(1), 199–217.

Chaljub, E., P. Moczo, S. Tsuno, P.-Y. Bard, J. Kristek, M. Kser, M. Stupazzini, and M. Kristekova (2010, August). Quantitative Comparison of Four Numerical Predictions of 3d Ground Motion in the Grenoble Valley, France. *Bulletin of the Seismological Society of America 100*(4), 1427–1455.

Chiou, B., R. Youngs, N. Abrahamson, and K. Addo (2010, November). Ground-Motion Attenuation Model for Small-To-Moderate Shallow Crustal Earthquakes in California and Its Implications on Regionalization of Ground-Motion Prediction Models. *Earthquake Spectra 26*(4), 907–926.

Chiou, B.-J. and R. R. Youngs (2008, February). An NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra. *Earthquake Spectra 24*(1), 173–215.

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.

Cleveland, W. S., E. Grosse, and W. M. Shyu (1992). Local regression models. *Statistical models in S*, 309–376.

Davis, L. L. and L. R. West (1973). Observed effects of topography on ground motion. *Bulletin of the Seismological Society of America 63*(1), 283–298.

Douglas, J. (2014). Fifty years of ground-motion models. In *Second European Conference on Earthquake Engineering and Seismology (2ECEES): a joint event of the 15th European Conference on Earthquake engineering & 34th General Assembly of the European Seismological Commission*, Istanbul, Turkey.

Durand, S., S. Gaffet, and J. Virieux (1999). Seismic diffracted waves from topography using 3-D discrete wavenumber-boundary integral equation simulation. *Geophysics 64*(2), 572–578.

Efron, B. and R. J. Tibshirani (1993). *An introduction to the bootstrap*. Monographs on statistics and applied probabilities; 57. New York: Chapman & Hall.

ESRI (2011). Arcgis desktop: Release 10. redlands, ca: Environmental systems research institute.

Eurocode (1994). No. 8: Design provisions for earthquake resistance of structures. *Part 2*, 1998–2.

Gazetas, G., P. V. Kallou, and P. N. Psarropoulos (2002). Topography and soil effects in the MS 5.9 Parnitha (Athens) earthquake: the case of Admes. *Natural Hazards 27*(1-2), 133–169.

Geli, L., P.-Y. Bard, and B. Jullien (1988). The effect of topography on earthquake ground motion: a review and new results. *Bulletin of the Seismological Society of America 78*(1), 42–63.

Gianella, V. P. and E. Callaghan (1934, January). The Earthquake of December 20, 1932, at Cedar Mountain, Nevada and Its Bearing on the Genesis of Basin Range Structure. *The Journal of Geology 42*(1), 1–22.

Goldgof, D. B., T. S. Huang, and H. Lee (1990). Terrain analysis from curvature profiles. *International Journal of Imaging Systems and Technology 2*(3), 169–182.

Griffiths, D. W. and G. A. Bollinger (1979). The effect of Appalachian Mountain topography on seismic waves. *Bulletin of the Seismological Society of America 69*(4), 1081–1105.

Guisan, A., S. B. Weiss, and A. D. Weiss (1999, July). GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology 143*(1), 107–122.

Gunberg, K. A. and R. Green (2011). Cgpr report no. 66, ground motion predictive (attenuation) equations. Technical report, Virginia Tech.

Hadley, J. B. (1964). *Landslides and related phenomena accompanying the Hebgen Lake earthquake of August 17, 1959*. Geological Survey Professional Paper 435.

Hartzell, S., M. Meremonte, L. Ramirez-Guzman, and D. McNamara (2014, February). Ground Motion in the Presence of Complex Topography: Earthquake and Ambient Noise Sources. *Bulletin of the Seismological Society of America 104*(1), 451–466.

Hartzell, S. H., D. L. Carver, and K. W. King (1994). Initial investigation of site and topographic effects at Robinwood Ridge, California. *Bulletin of the Seismological Society of America 84*(5), 1336–1349.

Hatzfeld, D., J. Nord, A. Paul, R. Guiguet, P. Briole, J.-C. Ruegg, R. Cattin, R. Armijo, B. Meyer, A. Hubert, P. Bernard, K. Makropoulos, V. Karakostas, C. Papaioannou, D. Papanastassiou, and G. Veis (1995, November). The Kozani-Grevena (Greece) Earthquake of May 13, 1995, Ms = 6.6. Preliminary Results of a Field Multidisciplinary Survey. *Seismological Research Letters 66*(6), 61–70.

Irvin, B. J., S. J. Ventura, and B. K. Slater (1997). Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. *Geoderma 77*(2), 137–154.

Itasca Consulting Group, I. (2011). Flac fast lagrangian analysis of continua, ver. 5.0. minneapolis: Itasca.

Iwahashi, J. and R. J. Pike (2007, May). Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology 86*(3-4), 409–440.

Jones, K. B., D. T. Heggem, T. G. Wade, A. C. Neale, D. W. Ebert, M. S. Nash, M. H. Mehaffey, I. A. Goodman, K. A. Hermann, A. R. Selle, D. Bolgrien, S. Augustine, J. Pedersen, C. J. Lin, J. M. Viger, D. Chiang, Y. Zhong, J. Baker, and R. D. V. Remortel (2000, January). Assessing Landscape Condition Relative to Water Resources in the

Western United States: A Strategic Approach. In S. S. Sandhu, B. D. Melzian, E. R. Long, W. G. Whitford, and B. T. Walton (Eds.), *Monitoring Ecological Condition in the Western United States*, pp. 227–245. Springer Netherlands.

Joyner, W. B. and D. M. Boore (1981). Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 imperial valley, california, earthquake. *Bulletin of the Seismological Society of America 71*(6), 2011–2038.

Joyner, W. B. and D. M. Boore (1993). Methods for regression analysis of strong-motion data. *Bulletin of the Seismological Society of America 83*(2), 469–487.

Kawase, H. and K. Aki (1990). Topography effect at the critical SV-wave incidence: possible explanation of damage pattern by the Whittier Narrows, California, earthquake of 1 October 1987. *Bulletin of the Seismological Society of America 80*(1), 1–22.

Kramer, S. L. (1996, January). *Geotechnical Earthquake Engineering* (1 edition ed.). Upper Saddle River, N.J: Prentice Hall.

Lee, S.-J., D. Komatitsch, B.-S. Huang, and J. Tromp (2009, feb). Effects of topography on seismic-wave propagation: An example from northern taiwan. *Bulletin of the Seismological Society of America 99*(1), 314–325.

Lee, W. H. K. (1994). *Digital seismograms of selected aftershocks of the Northridge earthquake recorded by a dense seismic array on February 11, 1994 at Cedar Hill Nursery in Tarzana, California*. US Department of the Interior, US Geological Survey.

Louie, J. N. (2005). Improving Next-Generation Attenuation models with shear-velocity measurements at all TriNet and strong-motion stations in LA. pp. 43 pp.

Louie, J. N. (2007). Shear-wave velocity map for California: collaborative research with CGS, and UNR. pp. 98 pp.

Louie, J. N., T. Mayo, S. Hauksson, Z. Reynolds, and E. Hall-Patton (2011). Shear-velocity measurements at CISN stations along the southern San Andreas Fault. pp. 98 pp.

Lysmer, J. and R. L. Kuhlemeyer (1969). Finite dynamic model for infinite media. *Journal of the Engineering Mechanics Division 95*(4), 859–878.

MacMillan, R., W. Pettapiece, S. Nolan, and T. Goddard (2000, July). A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. *Fuzzy Sets and Systems 113*(1), 81–109.

Maufroy, E., V. M. Cruz-Atienza, F. Cotton, and S. Gaffet (2015). Frequency-scaled curvature as a proxy for topographic site-effect amplification and ground-motion variability. *Bulletin of the Seismological Society of America 105*(1), 354–367.

Maufroy, E., V. M. Cruz-Atienza, and S. Gaffet (2012). A robust method for assessing 3-D topographic site effects: A case study at the LSBB Underground Laboratory, France. *Earthquake Spectra 28*(3), 1097–1115.

Meunier, P., N. Hovius, and J. A. Haines (2008, November). Topographic site effects and the location of earthquake induced landslides. *Earth and Planetary Science Letters 275*(3-4), 221–232.

Miller, R. G. J. (2011, October). *Simultaneous Statistical Inference* (2nd ed. 1981. Softcover reprint of the original 2nd ed. 1981 edition ed.). New York, NY: Springer.

Mucciarelli, M. and M. R. Gallipoli (2006). Comparison between vs30 and other estimates of site amplification in italy. In *First European Conference on Earthquake Engineering and Seismology*.

Nguyen, K.-V. and B. Gatmiri (2007, February). Evaluation of seismic ground motion induced by topographic irregularity. *Soil Dynamics and Earthquake Engineering 27*(2), 183–188.

Rai, M., A. Rodriguez Marek, and A. Yong (2012). Topographic effects in strong ground motion. In *Proocedings 12 world conference on earthquake engineering, Lisbon*.

Rice, J. A. (1995). *Mathematical statistics and data analysis*, Volume 2nd. Belmont, CA: Duxbury Press.

Richter, C. F. (1958). Elementary Seismology. W. H. Freeman and Co., San Francisco.

Rodriguez-Marek, A., G. A. Montalva, F. Cotton, and F. Bonilla (2011). Analysis of single-station standard deviation using the kik-net data. *Bulletin of the Seismological Society of America 101*(3), 1242–1258.

Rogers, A. M., L. J. Katz, and T. J. Bennett (1974). Topographic effects on ground motion for incident P waves: A model study. *Bulletin of the Seismological Society of America 64*(2), 437–456.

Sánchez-Sesma, F. J. and M. Campillo (1991). Diffraction of P, SV, and Rayleigh waves by topographic features: A boundary integral formulation. *Bulletin of the Seismological Society of America 81*(6), 2234–2253.

Sánchez-Sesma, F. J., I. Herrera, and J. Avilés (1982, April). A boundary method for elastic wave diffraction: Application to scattering of SH waves by surface irregularities. *Bulletin of the Seismological Society of America 72*(2), 473–490.

Schmidt, J. and A. Hewitt (2004, August). Fuzzy land element classification from DTMs based on geometry and terrain position. *Geoderma 121*(3-4), 243–256.

Shakal, A. F., M. Huang, R. Darragh, T. Cao, R. Sherburne, P. Malhotra, C. Cramer, R. Sydnor, V. Graizer, G. Maldonado, C. Peterson, and J. Wampole (1994). CSMIP strong-motion records from the Northridge, California earthquake of January 17 1994. *Report OSMS 94-07, Calif. Div. Mines Geol., Sacramento, California.*.

Smith, W. D. (1975). The application of finite element analysis to body wave propagation problems. *Geophysical Journal of the Royal Astronomical Society 42*(2), 747–768.

Spudich, P., M. Hellweg, and W. H. K. Lee (1996). Directional topographic site response at Tarzana observed in aftershocks of the 1994 Northridge, California, earthquake: implications for mainshock motions. *Bulletin of the Seismological Society of America 86*(1B), S193–S208.

Stafford, P. J. (2014). Crossed and nested mixed-effects approaches for enhanced model development and removal of the ergodic assumption in empirical ground-motion models. *Bulletin of the Seismological Society of America 104*(2), 702–719.

Trifunac, M. D. and D. E. Hudson (1971). Analysis of the Pacoima dam accelerogram-San Fernando, California, earthquake of 1971. *Bulletin of the Seismological Society of America 61*(5), 1393–1411.

Tripe, R., S. Kontoe, and T. Wong (2013, July). Slope topography effects on ground motion in the presence of deep soil layers. *Soil Dynamics and Earthquake Engineering 50*, 72–84.

Tucker, B. E., J. L. King, D. Hatzfeld, and I. L. Nersesov (1984). Observations of hard-rock site effects. *Bulletin of the Seismological Society of America 74*(1), 121–136.

Wong, H. L. (1982, August). Effect of surface topography on the diffraction of P, SV, and Rayleigh waves. *Bulletin of the Seismological Society of America 72*(4), 1167–1183.

Wong, H. L. and P. C. Jennings (1975). Effects of canyon topography on strong ground motion. *Bulletin of the Seismological Society of America 65*(5), 1239–1257.

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. Technical report.