
Electronic Thesis and Dissertation Repository

1-12-2012 12:00 AM

Relationships between Earthquake Ground Motions and Modified Mercalli Intensity

Andrea Sweny, *The University of Western Ontario*

Supervisor: Dr. Gail Atkinson, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Geophysics

© Andrea Sweny 2012

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Geophysics and Seismology Commons](#)

Recommended Citation

Sweny, Andrea, "Relationships between Earthquake Ground Motions and Modified Mercalli Intensity" (2012). *Electronic Thesis and Dissertation Repository*. 368.
<https://ir.lib.uwo.ca/etd/368>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

RELATIONSHIPS BETWEEN EARTHQUAKE GROUND MOTIONS AND
MODIFIED MERCALLI INTENSITY

(Spine title: Relationships between Earthquake Ground Motions and MMI)

(Thesis format: Monograph)

by

Andrea Sweny

Graduate Program in Earth Sciences

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Andrea Sweny 2012

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor:

.....
Dr. G. Atkinson

Supervisory Committee:

.....
Dr. K. Tiampo

.....
Dr. G. Atkinson

Examiners:

.....
Dr. M. El Naggar

.....
Dr. K. Tiampo

.....
Dr. R. Shcherbakov

The thesis by

Andrea Sweny

entitled:

Relationships between Earthquake Ground Motions and Modified Mercalli Intensity

is accepted in partial fulfillment of the
requirements for the degree of
Master of Science

.....
Date

.....
Chair of the Thesis Examination Board

Abstract

With widespread Internet use, web-based earthquake reports are a valuable source of felt intensity data with broad geographical coverage. Paired with ground motion data, Internet surveys like the USGS Did You Feel It? (DYFI) website can be used to derive relationships between ground motion parameters and Modified Mercalli Intensity (MMI). These are particularly important for regions with sparse seismographic coverage where the quality and density of intensity data may be superior to ground motion data. MMI-ground motion relationships are also needed to infer ground motions of historical earthquakes.

Using well-documented earthquakes in across North America, we analyze correlations between felt intensity and ground motions, including peak ground acceleration (PGA), peak ground velocity (PGV), and spectral acceleration (PSA) at 0.3, 1, and 3 seconds. These relationships are explored using correlations, linear regressions, and geostatistical methods. Magnitude, distance and site effects are analyzed, and Kriging is used to explore interactions among these variables.

Keywords: Felt intensity, Modified Mercalli Intensity, Ground Motion, Earthquake, Did You Feel It

Contents

Certificate of Examination	ii
Abstract	iii
List of Figures	viii
List of Tables	xiv
1 Introduction	2
1.1 Overview	2
1.2 Seismic Intensity Scales	3
1.3 Instrumental measures of earthquake ground motion	4
1.4 Ground motion - Intensity Relationships	6
1.4.1 Bilinear Relationships	7
1.4.2 Magnitude and Distance Dependence	9
1.4.3 Choice of Ground Motion Parameter	10
1.4.4 Choice of Functional Form	10
1.5 Moment Magnitude	11

1.6	Regional Differences	11
2	Data	15
2.1	Summary of Earthquake Characteristics	16
2.2	Did You Feel It?	17
2.3	Canadian Intensity Data	19
2.4	Data Analysis Methods	21
2.4.1	Matching DYFI and Instrumental Data	22
2.4.2	Comparison of Geocoded and Zip-code aggregated Data Matching	23
2.4.3	Canadian Intensity Data Analysis	25
2.4.4	Estimating Site Condition	30
3	Geostatistics	38
3.1	Overview	38
3.2	Ordinary Kriging	39
3.3	Empirical Variogram and Variogram Modeling	40
3.4	Block Kriging and The Change-of-Support Problem	43
3.5	Application of Kriging to Seismic Data	45
3.5.1	Methods in Kriging Seismic Data	46
3.5.2	Cross-validation of Kriging results	48
4	Results	50
4.1	Summary Statistics	50

4.2	Decay of MMI and Ground Motions with Distance	56
4.3	Correlations between Ground Motion Parameters and MMI	62
4.3.1	Distance, Magnitude and V_{S30} Dependence of Correlations	65
4.4	Linear Regressions	70
4.4.1	Residuals	71
4.5	Region Dependence of Correlations	78
4.6	Kriging and Spatial Analysis	89
4.6.1	Semivariograms and Spatial Continuity	89
4.6.2	Semivariogram Model Parameters	91
4.6.3	Kriging Estimation of MMI and Ground Motion Parameters	93
4.6.4	Kriging V_{S30}	101
4.6.5	Block Kriging	102
5	Summary of Results, Conclusions and Discussion	108
5.1	Variability of MMI and Ground Motion Parameters	109
5.2	Correlation Coefficients and Regression Statistics	110
5.3	Eastern vs. Western North America	111
5.4	V_{s30} as a Proxy for Site Condition	112
5.5	Kriging for Ground Motions and Seismic Intensity	113
	Bibliography	116
	Appendix: R Functions	121

List of Figures

1.1	MMI-PGA linear regressions from previous studies.	9
1.2	Locations of felt DYFI and intensity responses for two M 5.4 earthquakes: Ocotillo, CA (2010); Mt. Carmel, IL (2008)	12
1.3	Estimates of quality factor as a function of frequency for Northeastern and Western United States; from Atkinson and Silva [2000] and Atkinson and Boore [1995]. . .	14
2.1	MMI vs distance overlaid with PGA vs distance for M 5.4 Borrego Springs, and M 7.2 Baja. Data are grouped into distance bins and the mean values of MMI and PGA are plotted as points. Error bars on the average values indicate standard deviation for the binned data.	18
2.2	MMI vs distance overlaid with PGA vs distance for M 5.2 Anza, and M 5.4 Chino Hills.	20
2.3	Yellow circle indicates location of the epicenter for the M 5.0 2010 Val des Bois, Quebec earthquake. Individual felt intensity reports for Canada, and aggregate DYFI MMI values for the U.S. are shown as small red dots. Seismograph stations are shown as blue triangles.	21
2.4	Distance-magnitude distribution of matched MMI-ground motion observations. Note the great concentration of data at M less than 6.0, and at distances 10-100 km.	23
2.5	Comparison of DYFI intensity maps - zip code vs. geocoded [USGS, 2011]	24
2.6	Comparison of MMI values matched with stations from Zip coded DYFI intensities and Geocoded DYFI intensities.	26

2.7	Comparison of matched MMI-PGV observations from geocoded and zip code-based DYFI intensities for Borrego Springs event. The four plots show geocoded MMI matched at increasing distances from ground motion observations.	27
2.8	Comparison of matched instrumental-MMI observations for PGV (top) and PGA (bottom) from geocoded (right) and zip code-based (left) DYFI intensities for all earthquakes with both types of DYFI intensities.	28
2.9	Plot of zip code residuals against geocoded residuals for all MMI vs. ground motion regressions. The 45-degree line is plotted for reference.	29
2.10	The gridding scheme used to average the Riviere du Loup intensity data. The large yellow circle shows the earthquake's epicenter. Red circles indicate individual intensity observations. Major cities are labeled. All observations within a grid cell are averaged, and assigned a location equal to the center of the cell.	31
2.11	Illustration from Chen and Scawthorn [2003]. Bedrock overlain with a soil deposit of depth H and shear wave velocity V_S	33
2.12	Locations of V_{S30} measurements (red) and stations (black) in California.	35
2.13	Histogram showing the distribution of V_{S30} values in California.	36
2.14	Averages over zip codes of Chiu et al direct V_{S30} values minus averages of topographic slope estimates from Allen and Wald [2011]	37
3.1	Six classes of commonly-used variogram models. The nugget effect captures the pure noise component of a variogram, when there is positive variance at zero separation distance. Spherical, exponential and Gaussian semivariograms are bounded, exhibiting increasing variance with separation distance up to a threshold distance called the range. Beyond the range, these transitional semivariograms level off at a value called the sill. The linear and logarithmic semivariograms are unbounded, and increase indefinitely [R Development Core Team, 2009].	42
3.2	Examples of bounded (A) and unbounded (B) variogram models [Sarma, 2009]. Bounded variograms increase up to a certain separation distance, after which they level out at the sill. Unbounded variograms continue to increase indefinitely.	42

3.3	Block Kriging: An average of an areal support (box) is computed based on a set of individual observations (block points). The box is deaggregated into a grid of regularly spaced points, Kriging is performed over those points, and averaged. Source: Wackernagel [2003]	45
3.4	Sample results from cross-validation of Vs30 Kriging. Each plot represents a separate simulation in which a sample of 100 of 802 points are removed, and Kriging is performed. The actual data points are then compared with their estimates, and a histogram of the differences are displayed.	49
4.1	Shows the number of DYFI observations versus number of ground motion observations for all events. The events with over 400 DYFI observations and over 140 ground motion observations are highlighted; these include the 2007 M 5.6 Alum Rock, 2008 M 5.4 Mount Carmel, 2009 M 4.7 Inglewood, 2005 M 4.9 Yucaipa, 2010 M 5.4 Borrego Springs and 2008 M 5.4 Chino Hills events.	51
4.2	Histogram showing the number of observations for the combined data set of all events, over different distances. On the left are distances from 0-200 km, with a scale from 0-500 observations. On the right, distances from 200-800 km, with a scale of 0-50 observations.	53
4.3	Comparing the attenuation of MMI with distance for different earthquakes, based on region and magnitude. Small dots show individual MMI observations, while large squares are the mean MMI over various distances. Error bars show standard deviations associated with these means. From the top left corner clockwise: (1) Mt. Carmel and Borrego Springs, ENA and WNA both M 5.4. (2) 2008 M 5.4 Mount Carmel and 2011 M 5.8 Virginia, both in ENA of different magnitudes. (3) M 7.2 Baja, M 5.7 Ocotillo and M 4.7 Inglewood - three earthquakes in Western North America of varying magnitudes. (4) 2011 M 9.0 Tohoku and 2010 M 7.2 Baja, both large magnitude earthquakes.	58
4.4	Attenuation of PGV with distance from the epicenter. Top: M 7.2 Baja, M 5.4 Borrego Springs and M 4.7 Inglewood earthquakes. Individual observations shown by small circles, averages over distance bins shown by squares with standard deviation bars. Bottom: M 5.4 Borrego Springs and Mount Carmel events.	60
4.5	Attenuation of PGA with distance. Top: M 7.2 Baja, M 5.4 Borrego Springs and M 4.7 Inglewood earthquakes. Bottom: M 5.4 Mount Carmel in CEUS compared with same magnitude Borrego Springs in California.	61

4.6	Violin plots for log PGV (top, cm/s) and log PGA (bottom, cm/s ²) associated with different MMI ranges. The black rectangle spans the first to third quartile, with the white circle representing the mean. The kernel density function for the set of ground motion parameter values is pasted on both sides. Note that the group with MMI > 5.5 consists of only 15 observations. All other groups contain at least 75 observations.	64
4.7	Relationship between Log ground motion parameters and Vs30 for fixed MMI ranges. For MMI 2-3 (top) and MMI 4-4.5 (bottom), log PGV, log PGA and log Average PSA is plotted against log of Vs30/760 m/s. We can see a clear trend of decreasing ground motion with increasing Vs30. This implies that lower Vs30 sites are associated with stronger ground motions for a given MMI level.	66
4.8	Graphical representation of the correlation coefficients presented in Table 4.7. . . .	68
4.9	MMI-ground motion plots with symbols distinguishing different epicentral distance ranges (<i>R</i>).	72
4.10	MMI-ground motion plots with symbols distinguishing eastern and western North America.	73
4.11	Residuals from MMI-ground motion regressions with symbols distinguishing eastern and western North America.	75
4.12	Residuals from equation (4.2), plotted against epicentral distance for the six ground motion parameters. Points represent individual residuals; square symbols are averages over distance bins with standard deviation bars.	79
4.13	Boxplot of residuals from equation (4.2) based on magnitude. Thick black line indicates mean of the magnitude bin, the lower and upper edges of the box represent the first and third quartiles, and standard errors are represented by bars.	80
4.14	Density function of residuals from MMI-Ground Motion regression (4.2), separated by V_{S30} values above and below 360 m/s.	81
4.15	Top: MMI-PGV bilinear relationship from AK07 and Worden et al. [2011] overlaid on data from this study, showing systematic differences in data used. The outlying points from ENA data appear to force the regression line down below AK07 and Worden et al. [2011]. Bottom: Comparison of residuals from Atkinson and Kaka [2007] MMI-PGV curve and residuals from the simple MMI-PGV relationship found in this study, plotted against epicentral distance. The overall shape of the residuals, and their behaviour with distance, is similar for both relationships. . . .	82

4.16	MMI vs PGA for M5.4 Mount Carmel and Borrego Springs events. Top: the raw MMI and PGA data. Bottom: raw data are detrended using the distance component c_2 from equation (4.8).	85
4.17	Histograms of residuals from regression (4.5), in which pooled Mount Carmel and Borrego Springs data are used to regress MMI on PGA. The left side shows residuals from Mount Carmel, the right side residuals from Borrego Springs. Note how the Mount Carmel residuals are skewed to the negative side, while the Borrego Springs are skewed to the positive side.	86
4.18	Top: MMI vs PSA for M5.4 Mount Carmel (right) and Borrego Springs (left) events. Note that the lower frequency PSA for Mount Carmel is 0.2 seconds, while for Borrego Springs it is 0.3 seconds. Bottom: MMI vs PGV for M5.4 Mount Carmel and Borrego Springs events.	88
4.19	Top: Empirical variogram and Gaussian model for Inglewood MMI. The model fits the data quite well and represents a clear choice, particularly at small separation distances where fit is most important. Bottom: Empirical semivariogram for Yorba Linda MMI, with two different choices of model semivariogram: Gaussian (solid line) and Spherical (dashed line). Illustrates the subjective error involved in selecting the best semivariogram model.	90
4.20	Kriging of Yorba Linda MMI based on Gaussian (left) and spherical (right) variogram models. The x-axis represents longitude, and the y-axis is latitude; the scale shown represents MMI values. The differences are modest, reflecting the only slight differences between the two models.	91
4.21	Illustration of a failure to match a variogram model to the empirical variogram from the 2003 M6.5 San Simeon event. Top: the MMI-distance plot shows a dearth of observations close to the epicenter. Bottom: The empirical variogram is erratic with almost no structure, making it impossible to fit a variogram model.	94
4.22	Trend surface of MMI from M5.7 Ocotillo earthquake. Horizontal axes are longitude and latitude, and the surface represents the best-fitting polynomial surface for the MMI observations.	95
4.23	Left: MMI-PGV data and linear relationship for 2009 M4.7 Inglewood event. Right: Histogram of residuals from linear regression.	96

4.24	Empirical variogram for M4.7 Inglewood event, PGV. Two different variogram models are shown; Gaussian (left), and Spherical (right). The spherical and Gaussian models both fit equally well, and only differ by the unknown behaviour near the origin. The spherical model was chosen.	98
4.25	Illustration of the steps involved in the Kriging of PGV for M4.7 Inglewood event. Horizontal axis is longitude, vertical axis is latitude. Epicenter is depicted as circular dot. Scale represents PGV in cm/s. Top: Kriging of residuals from the regression of PGV against epicentral distance. Bottom left: Spatial representation of the linear distance trend component of PGV. Bottom right: Complete Kriging estimate made by adding residual component and trend component.	99
4.26	Kriging of MMI (top left), PGA (top right), PGV (bottom left) and ratio of $V_{S30}/760$ m/s (bottom right) from M4.7 Inglewood earthquake. Axes represent longitude and latitude; epicenter represented by small circle.	100
4.27	Kriging of MMI (left) and PGV (right) from M5.6 Alum Rock earthquake. Epicenters are located at the points of highest MMI and PGV values (red areas).	101
4.28	Topographic map of California, from USGS digital elevation model [USGS, 2012].	102
4.29	Variogram showing spatial continuity of V_{S30} in California.	103
4.30	Top: Block Kriging of PGV (cm/s) over zip codes in California; Bottom: chloropleth map of zip-coded MMI for M4.7 Inglewood earthquake. In the chloropleth map, different colours represent different MMI values.	105
4.31	Block Kriging of PGV and chloropleth map of MMI for M5.2 Alum Rock earthquake.	106
4.32	Comparison of matched data from block Kriging (left) and by matching stations with zip codes (right) for M4.7 Inglewood earthquake (top) and Alum Rock earthquake (bottom)	107

List of Tables

1.1	The abridged Modified Mercalli Intensity (MMI) scale of 1931 from Wood and Neumann [1931]	5
1.2	Summary of previous linear MMI-PGA relationships, $MMI = c_0 + c_1 \log Y$. ENA = Eastern North America; CA = California.	8
2.1	Date, location, and magnitude of events.	16
2.2	Shear wave velocity values corresponding to NEHRP site classes.	33
4.1	Summary statistics for DYFI data. Includes the type of intensity response (either aggregated by zip code or geocoded locations from DYFI data, or as single response locations for Canadian events), the number of locations (for zip codes, these must represent an area with more than 3 responses), the mean and maximum MMI value, and the range of distances covered.	52
4.2	Summary statistics for instrumental data, including number of observations, mean PGA and PGV, and distance range of observations.	54
4.3	Description of matched DYFI-ground motion data, including number of observations, the maximum distance between station and MMI observation used for matching geocoded or locational data (Match dist.), and the MMI and distance ranges covered.	55

4.4	Weighted least-squares regression statistics for the MMI-distance relation $\overline{\text{MMI}} = c_0 + c_1 \log(\overline{R})$, where R is epicentral distance. These regressions used binned DYFI data, where the MMI values are averaged over distance bins, and weighted by the number of points in each bin. The intercept and slope are recorded, along with the standard error of the estimated coefficients. * Note: MMI is undefined above 12, but the simple linear regression extrapolates to a higher intercept for the M7.2 Baja and M9.0 Tohoku events.	59
4.5	Number of ground motion observations within MMI ranges from 2.0 to 7.99, and minimum and maximum values of ground motion parameters within these ranges.	63
4.6	Mean and standard deviation for ground motion parameters associated with different MMI ranges. PGV, PGA and average PSA over 0.3, 1.0 and 3.0 seconds.	65
4.7	Correlation coefficients for MMI and ground motion parameters, for different distance and magnitude ranges. Starred rows indicate results were not statistically significant with p-values greater than 0.05.	69
4.8	Coefficients from estimation of equation (4.2), all data. Standard error given in brackets.	71
4.9	Coefficient estimates from regressions (4.3), (4.4) and (4.5), with standard errors. (*) indicates coefficients are not significant at 0.1 level, and (**) indicates coefficient is significant only at 0.1 level.	78
4.10	Comparison of coefficient estimates from equations (4.3) and (4.4), and those from Atkinson and Kaka [2007] and Worden et al. [2011].	78
4.11	Estimates and standard errors for coefficients of the linear regressions (4.6), (4.7), (4.8) and (4.9), M5.4 events Mount Carmel and Borrego Springs. Estimates and standard errors are given.	84
4.12	Estimates and standard errors for coefficients of the linear regressions (4.10) and (4.11), M5.4 events Mount Carmel and Borrego Springs. Estimates and standard errors are given.	87
4.13	Variogram parameters for ground motions and MMI, inter-event comparison.	92

Chapter 1

Introduction

1.1 Overview

The intensity of ground shaking produced by an earthquake can be assessed using both quantitative and qualitative measures. While seismographic instruments record the velocity or acceleration of ground movement, the strength of ground shaking may also be described by its physical manifestations: human reactions to the shaking, the behaviour of furniture or other free-standing objects, damages to buildings and infrastructure, or visible surface waves, to name a few. Quantitative observations of ground motions are ideal, as these give accurate and precise measures of ground shaking. But seismometers are sparsely located in some regions, especially those with less-frequent severe earthquakes. But while extensive seismographic coverage may not exist in many regions (for instance Eastern North America), Internet-based earthquake reporting by citizens makes it easy to rapidly obtain large numbers of qualitative observations. Citizens can now easily report information about earthquake ground shaking and its physical effects, which can then be processed into a value of Modified Mercalli Intensity (MMI). This value summarizes the observed intensity of ground shaking over a given geographical area. Making use of this readily obtainable and rich data can enhance our understanding of ground motions in regions where instrumental coverage is limited, or for historical earthquakes where such instrumental data is altogether absent.

To make use of these qualitative MMI observations, we need to relate them back to quantitative measures of ground motions, as it is these measures which will ultimately be used to estimate

seismic hazard and inform the design of building codes. It is natural for ground motion measurements to be well correlated with MMI, as the latter represents the physical manifestations of the former. But how well correlated are they? Are they linearly related, or is the relationship better described by some other functional form? How do regional differences, magnitude, distance or local site effects influence these relationships? This report will examine the relationships between instrumental ground motion observations and MMI: assessing the strength of correlations, deriving predictive relationships between the variables, and looking at the spatial continuity of correlations. Data from a choice set of selected events are used - earthquakes in both eastern and western North America for which there are rich MMI data as well as good instrumental ground motion data.

These analyses are done using a variety of different methods. First, the most basic relationships between MMI and ground motion are elucidated using simple numerical associations - using matched MMI-ground motion data to compare levels of ground motion parameters for different MMI ranges. Basic correlation coefficients provide information about the strength of these relationships, and using data grouped over magnitudes and distances, the effects of these secondary variables can be observed. More precise quantitative assessments of MMI-ground motion correlations are then performed using linear regressions, and residuals from these regressions are analyzed to determine distance, magnitude and site effects. Finally, spatial analysis of ground-motion and MMI data is performed using Kriging; this provides insight into detailed MMI-ground motion interactions that may not be captured by numerical analysis alone.

1.2 Seismic Intensity Scales

Seismic intensity scales rate the severity of an earthquake using human perceptions and observations of physical effects and damage. Lower intensity levels are characterized primarily by human perceptions. Was the earthquake felt by most people, or just a few? Was it felt only by those outdoors at the time, or also those inside buildings? Were people awakened from sleep by the ground motions or disturbances? At intermediate levels, intensity scales use a combination of human reactions and physical manifestations. Criteria such as whether people experienced difficulty walking, and outward observations such as the rattling or falling of objects are used. High intensities (≥ 6) are categorized primarily based on physical damage to structures, as well as observations of surface waves and cracks in the ground.

There are three main intensity scales in use: the Medvedev-Sponheuer-Karnik (MSK) scale, the Modified Mercalli Intensity (MMI) scale, and the Japan Meteorological Agency (JMA) intensity scale [Shabestari and Yamazaki, 2001]. In North America, the MMI scale [Wood and Neumann, 1931] is used almost exclusively. Precursors to the MMI scale originated with the seismologist M. S. Rossi, who devised a 10-division scale to classify Italian earthquakes [Gorshkov and Shenkareva, 1960]. Rossi later combined his scale with that of Swiss limnologist Francois-Alphonse Forel to create the Rossi-Forel scale, in which the value I referred to the intensity at the farthest epicentral distance for which the earthquake would register on a seismograph, and X corresponded to total destruction [Reed, 2008].

In 1902 Mercalli adapted the Rossi-Forel scale to include an earthquake's effect on human behaviors - whether people were awoken from sleep, had difficulty maintaining balance, etc. This scale was further modified in 1932 to take into account effects on recent innovations like motorized vehicles and tall buildings, and the scale was expanded to include 12 categories [Cormier, 2011]. Richter revised the scale a final time in 1956 to produce the modern MMI scale that is widely used today. Descriptions of the 12 MMI levels are presented in table 1.1.

1.3 Instrumental measures of earthquake ground motion

Seismometers record time histories of ground displacement, velocity or acceleration from an earthquake at a specific site. Motions are usually recorded for three directions: one vertical and two orthogonal components. Seismograms provide information about the duration of ground shaking and its frequency and amplitude content. A number of parameters are typically used to capture the major features of seismograms. Peak ground displacement (PGD), peak ground velocity (PGV) and peak ground acceleration (PGA) are the maximum recorded values for each of these variables. Most earthquakes exhibit greater horizontal ground motions than vertical, so peak ground motion values usually correspond to the peak *horizontal* ground motion values.

While the time history of acceleration gives peak ground acceleration, velocity and displacement, it is also of interest to understand the frequency content of ground motions. Structures have natural frequencies of vibration, and their response to ground shaking depends on this frequency content. This information is captured in the response spectrum, which represents the maximum response of single degree-of-freedom linear oscillator with a specific natural frequency [Jennings, 2003].

MMI	Equivalent Magnitude	Description
I	1.0 - 2.0	Not felt except by a very few under especially favorable circumstances.
II	2.0-3.0	Felt only by a few persons at rest, especially on upper floors of buildings. Delicately suspended objects may swing.
III	3.0-4.0	Felt quite noticeably indoors, especially on upper floors of buildings. Many people do not recognize it as an earthquake. Standing motor cars may rock slightly. Vibration like passing of truck. Duration estimated.
IV	4.0-5.0	During the day felt indoors by many, outdoors by few. At night some awakened. Dishes, windows, doors disturbed; walls made cracking sound. Sensation like heavy truck striking building. Standing motor cars rocked noticeably.
V	5.0- 6.0	Felt by nearly everyone; many awakened. Some dishes, windows, etc. broken. A few instances of cracked plaster; unstable objects overturned. Disturbance of trees, poles and other tall objects sometimes noticed. Pendulum clocks may stop.
VI	5.0 - 6.0	Felt by all; many frightened and run outdoors. Some heavy furniture moved. A few instances of fallen plaster or damaged chimneys. Damage slight.
VII	6.0	Everybody runs outdoors. Damage negligible in buildings of good design; slight to moderate in well-built ordinary structures; considerable in poorly built or badly designed structures. Some chimneys broken. Noticed by persons driving motor cars
VIII	6.0-7.0	Damage slight in specially designed structures; considerable in ordinary substantial buildings with partial collapse; great in poorly built structures. Panel walls thrown out of frame structures. Fall of chimneys, factory stacks, columns, monuments, walls. Heavy furniture overturned. Sand and mud ejected in small amounts. Changes in well water. Disturbed persons driving motor cars.
IX	7.0	Considerable damage to specially built structures; buildings shifted off foundations. Ground cracked conspicuously. Underground pipes broken.
X	7.0 - 8.0	Most masonry and frame structure and their foundations destroyed. Ground badly cracked. Rails bent. Landslides considerable from river banks and steep slopes.
XI	8.0	Few, if any (masonry) structures remain standing. Bridges destroyed. Broad fissures in the ground; Earth slumps and land slips in soft ground.
XII	8.0 or greater	Total damage. Waves seen on ground surfaces. Objects thrown upward into the air.

Table 1.1: The abridged Modified Mercalli Intensity (MMI) scale of 1931 from Wood and Neumann [1931]

Given the time history of ground acceleration, this displacement due to a given level of ground motions can be found for any natural frequency and damping level, and the response spectrum is simply the maximum absolute value of this displacement over the entire time history. The *pseudo acceleration spectrum* (PSA) is used in this study, and is derived from the response spectrum by multiplying each value by the square of the natural frequency ω .

1.4 Ground motion - Intensity Relationships

As mentioned above, seismic intensity is defined using physical manifestations of ground shaking, so it is natural for measures of intensity to be well-correlated with measures of ground motion. These physical manifestations reflect not only the overall level of ground motion, but also its duration and frequency content. Empirical relationships between seismic intensity and ground motion parameters can be used to infer characteristics of ground motions when instrumental data is absent; in the reverse direction, they can be used to estimate intensities arising from hypothetical ground motions for seismic hazard analysis. One of the key applications of MMI-ground motion relationships is the generation of Shakemaps - automatically generated maps of ground shaking and intensity that are created rapidly following an earthquake [Wald et al., 1999b, 2006].

The value of accurately estimating MMI from ground motions (and vice versa) has generated a long history of ground motion-intensity relationships. While the majority of these relationships use the peak ground motion parameters PGA and PGV, some are also based on duration of ground shaking and its frequency content [Trifunac and Westermo, 1977, Wald et al., 1999a], response spectra [Atkinson and Sonley, 2000], and other quantitative measures of ground shaking. As these relationships could potentially be region-dependent, studies have been conducted using both global and regional data.

In their work developing relationships between peak ground motion parameters and Modified Mercalli intensity, Trifunac and Brady [1975] provide an overview of previous studies. These preliminary works sought to find basic linear relationships between intensity and ground motion, of the form

$$\text{MMI} = c_0 + c_1 \log(Y) \quad (1.1)$$

where Y is a ground motion parameter - PGA, PGV, or PSA at some frequency.

While these early relationships were rudimentary and based on limited data, recent studies (like Atkinson and Kaka [2007] using more extensive data sets still largely support this basic linearity. With the advent of Internet intensity reporting, the most recent ground motion-intensity relationships have been developed using MMI data derived from the USGS Did You Feel It? website [Atkinson and Wald, 2007, Atkinson and Kaka, 2007, Worden et al., 2011] (see section 2.2).

The coefficients from basic linear regressions of ground motions vs. intensity over the last 60 years are compiled in table 1.2 and plotted in figure 1.1. From this figure, it is easy to see that these linear relationships differ in both slope and intercept values. This variation among results may be due to differences in the source data used to estimate the regression coefficients. Regional differences may come into play, but discrepancies in results can also be introduced simply by the selection of different events from the same region. Differences in how events from distinct regions are sampled from the distance and magnitude distribution may influence results and compound any regional differences present in the data.

Another source of variability among ground motion vs intensity relationships is the methodology involved in generating coefficient estimates. Different ways of manipulating the data that are designed to ameliorate the uncertainty in MMI values may introduce different biases into the results. For example, the method employed here of using average MMI and ground motion values over MMI bins in regression could have different effects depending on the subsets of data used. It is as yet not clear how much of the variation in MMI vs ground motion relationships can be attributed to either data selection or methodology differences; but it is clear that significant uncertainty in these relationships persists. In the future, a comprehensive data set of matched MMI and ground motion parameter values may be helpful in isolating the effects of methodology and data selection.

1.4.1 Bilinear Relationships

While MMI varies linearly with ground motion, a marked increase in slope often occurs for higher MMI levels. Atkinson and Kaka [2007] attributes this phenomenon to the nature of the MMI scale, and the change it undergoes from low MMI to high. At low levels MMI is determined primarily by human experiences, while high MMI is based on physical damages. To account for this change in slope, Atkinson and Kaka [2007] and others use a predetermined MMI value (5) as a change point, and perform separate regressions for MMI above and below the change point.

Source	Year	Parameter	Region	Intercept c_0	Slope c_1
Kawasumi	1951	PGA		-0.732	3.89
Gutenberg & Richter	1958	PGA		1.5	3
Neumann	1954	PGA		0.13	3.25
Hershberger	1956	PGA		2.09	2.33
Trifunac & Brady	1975	PGA		-0.047	3.33
Murphy & O'Brien	1977	PGA	Global	-1	4
Atkinson & Sonley	2000	PGA		-2.826	4.348
Wu et al.	2003	PGV	Taiwan	1.89	2.14
Kaka & Atkinson	2004	PGV		3.96	1.79
Atkinson & Kaka	2007	PGV (≤ 0.48)		4.37	1.32
		PGV (> 0.48)		3.54	3.03
Tseletic & Danciu	2008	PGA	Greece	-0.946	3.563
		PGV		3.30	3.36
Dangkua & Cramer	2011	PGV	ENA	5.13	1.59
			CA	4.62	1.54
		PGA	ENA & CA	2.60	1.58
		PSA 0.3 s	ENA & CA.	2.33	1.57
		PSA 1.0 s	ENA	4.26	1.26
			CA	3.22	1.47
		PSA 2.0 s	ENA	5.00	1.27
			CA	3.80	1.68

Table 1.2: Summary of previous linear MMI-PGA relationships, $MMI = c_0 + c_1 \log Y$. ENA = Eastern North America; CA = California.

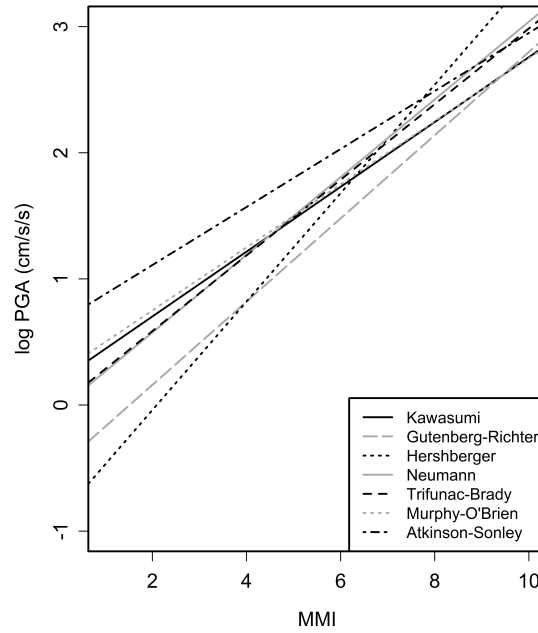


Figure 1.1: MMI-PGA linear regressions from previous studies.

1.4.2 Magnitude and Distance Dependence

Analysis of the residuals from these basic regressions reveal other variables that influence the relationship between MMI and ground motions. The principal of these are distance, magnitude, and local site conditions. Atkinson and Kaka [2007] plot residuals as a function of both magnitude and distance, showing a clear negative trend in both cases. Thus some of the unexplained variation from the simple MMI-ground motion regressions can be explained by distance and magnitude dependencies, and the following regression can be carried out to determine the extent of these parameters' influence:

$$\sigma_{\text{MMI}} = c_2 + c_3 \mathbf{M} + c_4 (\log R) \quad (1.2)$$

where \mathbf{M} is moment magnitude and R is epicentral distance.

The systematic trends found in these residuals suggest that distance and magnitude should be included in the original model. The reason this is not done is because of a common methodology employed in generating coefficient estimates for the original regression equation (Equation 1.1). In the MMI vs. ground motion relationships developed here and elsewhere (e.g. [Atkinson and

Kaka, 2007]), the initial MMI vs. ground motion regression is performed using grouped means. The matched data are first divided into bins based on either their MMI values or ground motion parameter values, and the mean values (weighted by the number of data points in each group) are used to determine the regression coefficients. This method is used to reduce the impact of outlying points, so that the resulting regression line is designed to fit the data where it is most abundant.

1.4.3 Choice of Ground Motion Parameter

Early ground motion-intensity relationships were based on peak ground acceleration, which is popular because it is easily obtained from unprocessed accelerograms. It is now understood that PGA is not necessarily well correlated with damage; time histories with the same PGA can vary greatly in frequency content, duration of strong shaking, and energy level [Cormier, 2011]. PGV has less contribution from the high-frequency component of the acceleration record, and more from the frequency range associated with most buildings. Hence PGV is often cited as being a better indicator of damage potential than PGA [Kappos, 2002]. PGD is considered the least reliable indicator of the three, due to noise at longer periods as well as errors introduced during the filtering and integration of accelerograms [Agarwal and Shrikhande, 2006].

1.4.4 Choice of Functional Form

While a linear or bilinear ground motion-intensity relationship is usually assumed, some researchers have proposed alternative functional forms. Since the data do tend to exhibit strong linearity, and the exact nature of any nonlinearity is unknown, the use of nonlinear models may be suspect.

McCann et al. [1980] argues that because the large amount of scatter in ground motion-MMI relationship can support many different functional relationships, it makes sense to use a function that offers more flexibility than a strict linear relationship; the authors instead fit a quadratic form to the data:

$$\log(\text{PGA}) = c_0 + c_1\text{MMI} + c_2\text{MMI}^2 \quad (1.3)$$

Another approach is to use nonparametric estimation to derive the appropriate relationship. Worden et al. [2011] derive joint probability density functions between MM and ground motion pa-

rameters, using them to derive reversible bilinear relationships. They also attempt to derive more precise relationships using linear combinations of PGA and PGV, to account for the different ranges over which each parameter may be better correlated with MMI. Their results validate the inclusion of magnitude and distance terms in MMI-ground motion regressions, as these terms significantly reduce the residual errors.

1.5 Moment Magnitude

Throughout this report, an earthquake's size will be described by its moment magnitude, denoted M . Moment magnitude was defined by Hanks and Kanamori [1979] as

$$M = \frac{2}{3} \log_{10} M_0 - 10.7 \quad (1.4)$$

where M_0 is the seismic moment, itself a product of the rupture area A , average displacement D of the rupture, and the shear modulus μ along the rupture [Datta, 2010]:

$$M_0 = \mu AD \quad (1.5)$$

Other magnitude scales- such as local magnitude or body-wave magnitude - are based on the amplitude of seismic waves, and tend to saturate at magnitudes above around 7.5. Because moment magnitude is based on the physical characteristics of the rupture and the total amount of energy radiated, it does not suffer from this shortcoming.

1.6 Regional Differences

In developing ground motion-intensity relationships, care is often taken to specify the region being studied. Ground motions display quite different behaviour in different geologic settings. Even a cursory look at Figure 1.2 - showing the locations of DYFI and intensity responses for two M 5.4 earthquakes - reveals fundamental differences between the way that ground motions are propagated and felt in different regions. The Mt. Carmel, IL earthquake was felt over a very large area relative to the felt area of the Ocotillo, California event. Bakun and McGarr [2002] examine differences

among continental regions in the behaviour of seismic intensity with distance from the source, and find a rapid decrease of intensity in Western North America, with a more moderate, gradual decline in intensity in ENA. This marked difference in felt area across North America results from difference in seismic wave attenuation, a combination three effects: geometric spreading, intrinsic or viscoelastic attenuation, and scattering. Geometric spreading is simply the decrease in energy

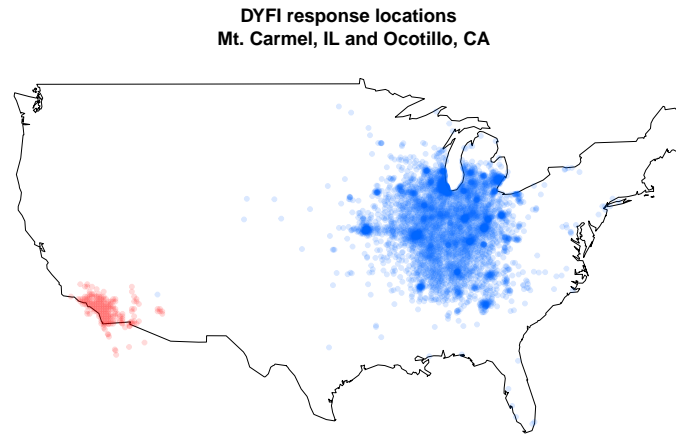


Figure 1.2: Locations of felt DYFI and intensity responses for two **M** 5.4 earthquakes: Ocotillo, CA (2010); Mt. Carmel, IL (2008)

density of a seismic wave that occurs when the wavefront expands as it travel farther from the source. For surface waves, this spreading takes place in two dimensions, like ripples emanating from a stone thrown into water. Body waves travel within Earth's interior, and spreading occurs in three dimensions. For a body wave traveling in a medium of uniform density and velocity, geometric spreading causes amplitude decay that is proportional to the inverse of its distance from the source: $\frac{1}{R}$. For surface waves, this amplitude decay is proportional to the inverse of the square root of distance from the source: $\frac{1}{\sqrt{R}}$.

A traveling wave also generates friction as it travels; this generates heat, and the energy lost in the process results in intrinsic attenuation. Finally, a wave will also encounter inhomogeneities as it travels, undergoing scattering as cracks and imperfections within Earth's crust deflect it in multiple

directions [Cormier, 2011]. The crust in Eastern North America is older and less fractured than in the west, allowing seismic waves to travel farther and retain higher amplitudes as they travel away from the source.

In addition to the differences in felt area, earthquakes in Eastern North America exhibit more energy at higher frequencies of ground motion relative to those in Western North America. This is caused by differences in anelastic attenuation - the loss of high-frequency components of seismic waves. A wave traveling from the source will undergo not only a decrease in amplitude, but also a shift from higher to lower frequencies.

The attenuation properties of a medium are characterized by the Quality factor Q , which appears to be both region- and frequency-dependent. A seismic wave will undergo anelastic attenuation at frequency f according to the exponential decay function $\exp(-\pi ft/Q(f))$, where

$$\frac{1}{Q(f)} = 2\pi \frac{\Delta E}{E} \quad (1.6)$$

and ΔE is the energy lost by the wave over one cycle, and E is the maximum energy during the cycle [Aki and Richards, 2002].

Two of the relationships between Q and frequency that have been estimated for Eastern North America (ENA) and Western North America are provided below, and plotted in figure 1.3 for frequencies up to 5 seconds. The Quality factor for Eastern North America is substantially larger than for Western North America at all frequencies. Thus amplitudes will decay more rapidly in the West relative to Eastern North America, and ground motions will be sustained over larger distances in the East for a given level of ground shaking at the epicenter.

$$\begin{array}{ll} Q(f) = 180f^{0.45} & \text{Western United States - Atkinson and Silva [2000]} \\ Q(f) = 680f^{0.36} & \text{NorthEastern United States - Atkinson and Boore [1995]} \end{array}$$

Anelastic attenuation may be accompanied by fundamental differences in earthquake characteristics between the two regions to cause the shift to higher frequencies in ENA. This higher frequency content leads to higher PGA values for ENA earthquakes (for the same magnitude and distance), since PGA tends to be carried by higher frequencies [Atkinson and Boore, 1990].

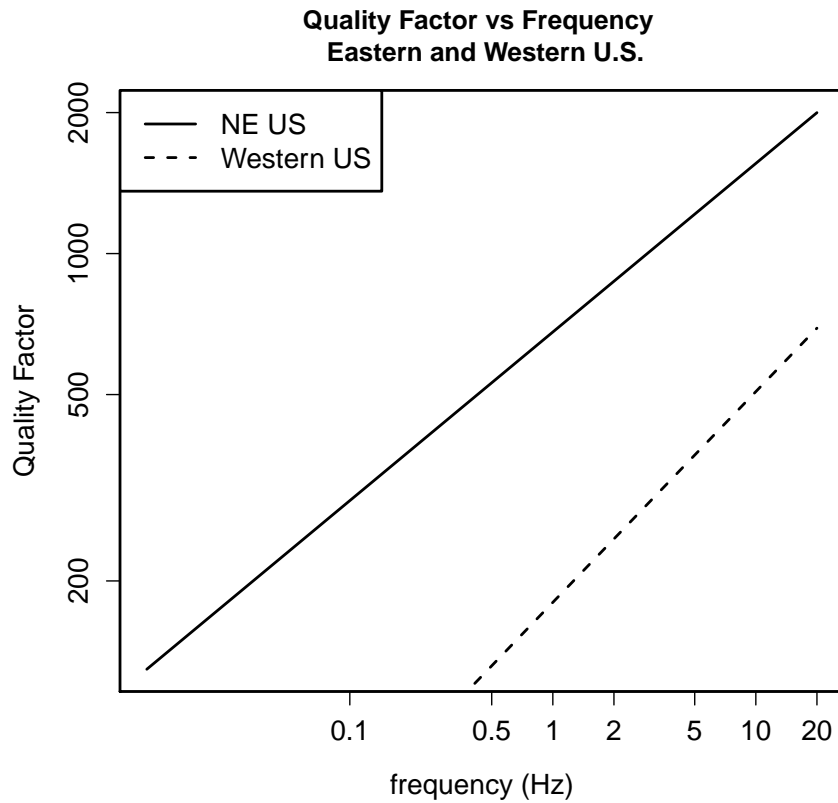


Figure 1.3: Estimates of quality factor as a function of frequency for Northeastern and Western United States; from Atkinson and Silva [2000] and Atkinson and Boore [1995].

Chapter 2

Data

To provide a general overview of the ground motion and intensity data used in this study, this section will provide summary statistics and some introductory plots. Basic characteristics of the earthquakes examined are provided, including magnitude, location and date. This section will describe the various types of data used, and the following sections will provide greater detail about how the data are handled, and what methods are used to prepare the data for analyses.

Initially, this study was to be an in-depth look at ground motions and Modified Mercalli Intensity for two moderate sized events - the **M** 5.4 Mount Carmel, IL and the **M** 4.7 Inglewood, CA earthquakes. These earthquakes were well-recorded, and large numbers of Did You Feel It? responses made provided rich MMI data. As the study progressed, other earthquakes with both instrumental and intensity data were added. The two years over which this study was conducted saw several new events of interest - the **M** 5.0 Val des Bois earthquake in eastern Canada, the **M** 9.0 Tohoku earthquake in Japan, and the **M** 5.8 Mineral earthquake in Virginia. A number of other well-recorded earthquake with abundant DYFI data were also included along the way. This section will describe the earthquakes used and their locations, along with key features of the intensity and ground motion data.

2.1 Summary of Earthquake Characteristics

Since California has the most dense instrumental coverage in North America, as well as frequent moderate-sized earthquakes, much of the data used in this study comes from California earthquakes. The 6 earthquakes with epicenters outside of California come from Quebec, Canada; Baja, Mexico; Mineral, Virginia; Mt.Carmel, IL; and Tohoku, Japan. Table 2.1 gives the name of each event, its magnitude, date, and location. Figure 2.4 shows how the data occupy the space of epicentral distance and magnitude. The majority of data lies within 100km of epicenters, and for MMI below 6.0. Observations outside of this range come from the **M7.2** Baja and **M9.0** Tohoku earthquakes.

Event Name	Magnitude	Date	Location	Longitude	Latitude
California					
Alum Rock	5.6	30 October 2007	California	-121.78	37.43
Anza 2001	5.1	31 October 2001	California	-116.51	33.51
Anza 2005	5.2	12 June 2005	California	-116.57	33.53
Big Bear City 1992	6.2	28 June 1992	California	-116.83	34.20
Big Bear City 2003	5.4	22 February 2003	California	-116.85	34.31
Borrego Springs	5.4	07 July 2010	California	-116.49	33.42
Chatsworth	4.6	09 August 2007	California	-118.62	34.30
Chino Hills	5.4	29 July 2008	California	-117.76	33.95
Inglewood	4.7	17 May 2009	California	-118.34	33.94
Lake Elsinore	4.7	02 September 2007	California	-117.48	33.73
Ocotillo	5.7	14 June 2010	California	-115.92	32.70
Parkfield	6.0	28 September 2004	California	-120.45	35.77
San Bernardino	4.5	08 January 2009	California	-117.29	34.11
San Simeon	6.5	22 December 2003	California	-121.10	35.71
Whittier Narrows	4.4	16 March 2010	California	-118.07	34.00
Yorba Linda	4.4	03 September 2002	California	-117.78	33.92
Yucaipa	4.9	16 June 2005	California	-117.01	34.06
Outside California					
Baja	7.2	04 April 2010	Mexico	-115.29	32.26
Mineral	5.8	23 August 2011	Virginia	-77.93	37.94
Mount Carmel	5.4	18 April 2008	Illinois	-87.89	38.45
Riviere du Loup	5.4	06 March 2005	Canada	-69.73	47.75
Val des Bois	5.0	23 June 2010	Canada	-75.50	45.90
Tohoku	9.0	11 March 2011	Japan	142.37	38.30

Table 2.1: Date, location, and magnitude of events.

2.2 Did You Feel It?

Since the MMI scale uses observations of both physical effects and the impact of ground motion on human activities, it is ideal to have both detailed reports of an earthquake's destructiveness as well as subjective accounts of how people perceive and experience ground shaking. Increasingly pervasive use of the Internet makes obtaining such subjective reports easy. The Did You Feel It? (DYFI) website [USGS, 2011] run by the USGS is an Internet-based reporting system, where earthquake witnesses can fill out a questionnaire about observed earthquake effects [Wald et al., 1999a]. Respondents input their geographic location, the time they felt the event, whether they were inside or in a building, and whether they were awake or asleep. Other typical questions include queries about how long the shaking lasted, whether it was difficult to stand or walk, if objects were rattled or overturned, and observations about any structural damage that may have occurred (e.g. cracks, ceiling damage, broken windows). The questionnaire is in multiple-choice format; this makes it easy to fill out and reduces the probability of inaccurate reporting. The responses for each question can be aggregated by either zip code, city, or (using geocoding of responses) by regularly sized boxes, and the average responses for the area is then translated into an MMI value. Thus each data point reflects an average effect over the entire area, in concert with the true nature of the Modified Mercalli Intensity scale.

Atkinson and Wald [2007] provide a good overview of the “Did You Feel It?” (DYFI) data and its use in engineering seismology. They highlight the fact that while individual DYFI responses may be imprecise, the sheer volume of observations makes the DYFI data surprisingly robust. This is an example of the law of large numbers concept in statistics - as the number of observations increases indefinitely, their average value will approach the true mean. Figures 2.1 and 2.2 show examples of the binned DYFI data overlaid on a plot of binned peak ground acceleration (PGA) for four California events: **M5.4** Borrego Springs, **M7.2** Baja, **M5.2** Anza, and **M5.4** Chino Hills. The data match up fairly well, especially at intermediate epicentral distances of 50-100 km.

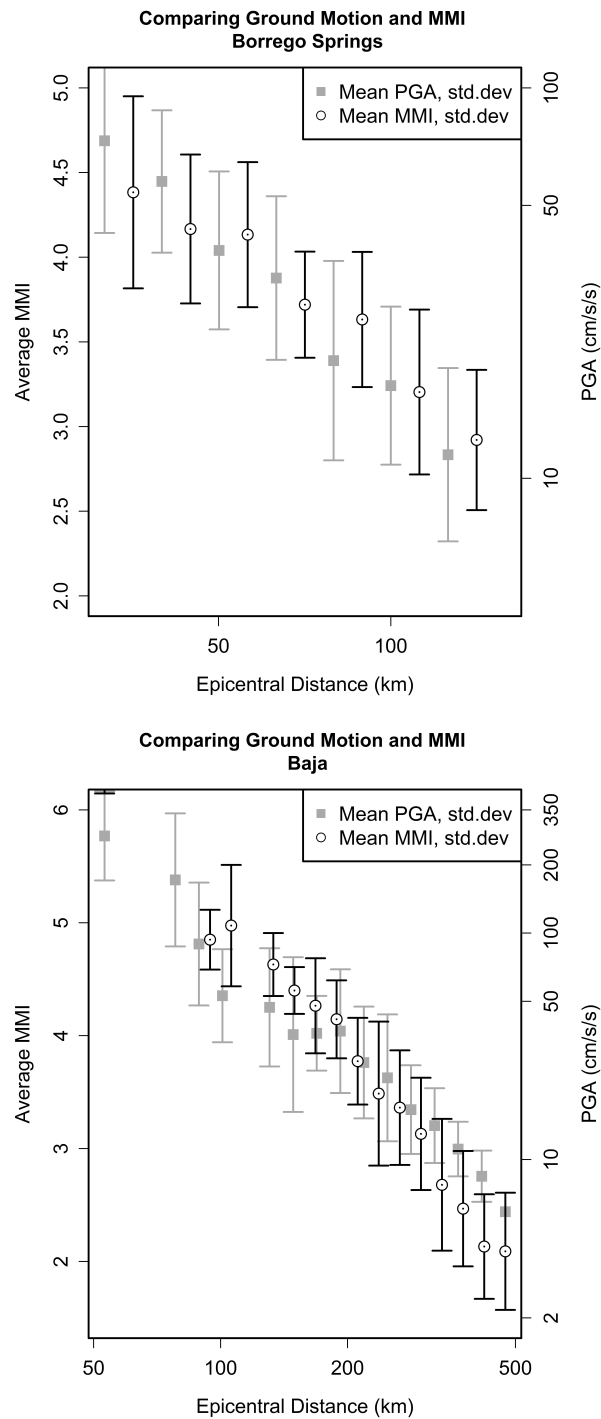


Figure 2.1: MMI vs distance overlaid with PGA vs distance for **M5.4** Borrego Springs, and **M7.2** Baja. Data are grouped into distance bins and the mean values of MMI and PGA are plotted as points. Error bars on the average values indicate standard deviation for the binned data.

2.3 Canadian Intensity Data

Listed in Table 2.1 are two earthquakes that occurred in Quebec, Canada: the **M5.4** Riviere du Loup earthquake, and the recent 2010 **M5.0** Val des Bois earthquake. The Canadian MMI data for these events were collected and aggregated differently than the USGS Did You Feel It? data. These MMI values are based on observations at a particular location, and do not represent an average of responses over an areal region. Such data need to be dealt with differently when matching stations to MMI values; this different methodology is discussed in section 2.4.3.

The locations of the Canadian intensity data as well as the DYFI MMI data are shown in Figure 2.3. The Canadian intensity data consist of direct responses from individual households or persons. Information collected from each response includes computed MMI value, city, postal code, latitude and longitude, epicentral distance, hypocentral distance, and building type and construction.

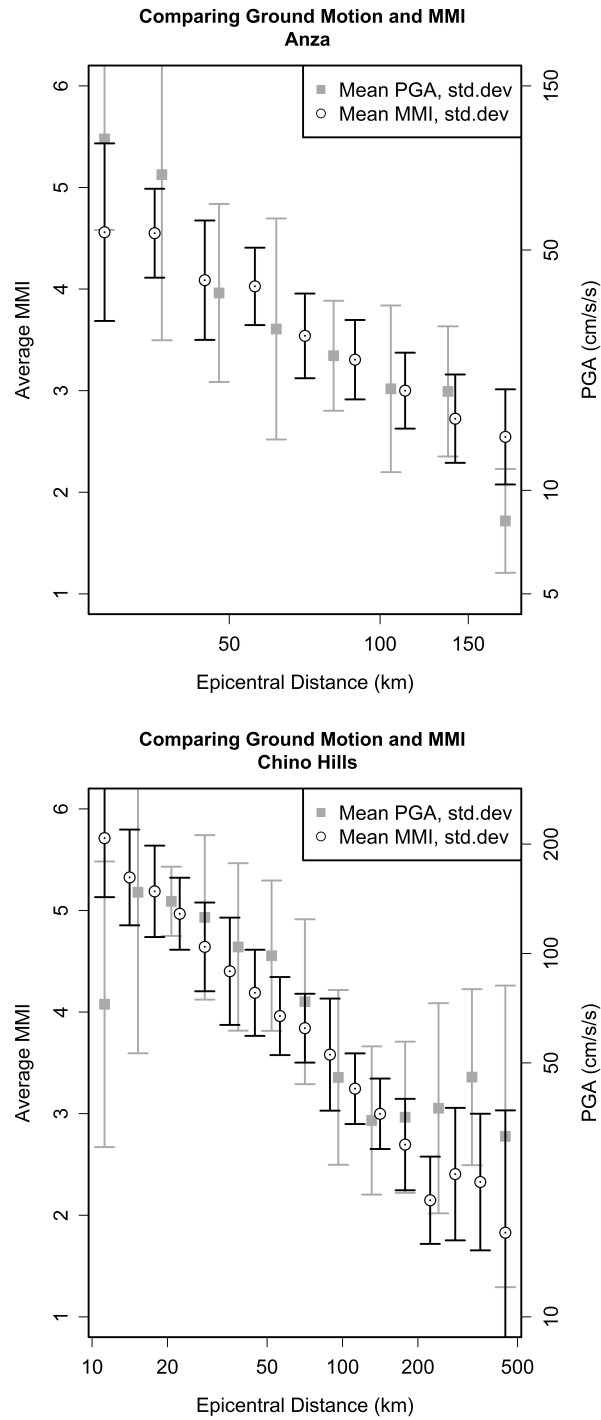


Figure 2.2: MMI vs distance overlaid with PGA vs distance for M5.2 Anza, and M5.4 Chino Hills.

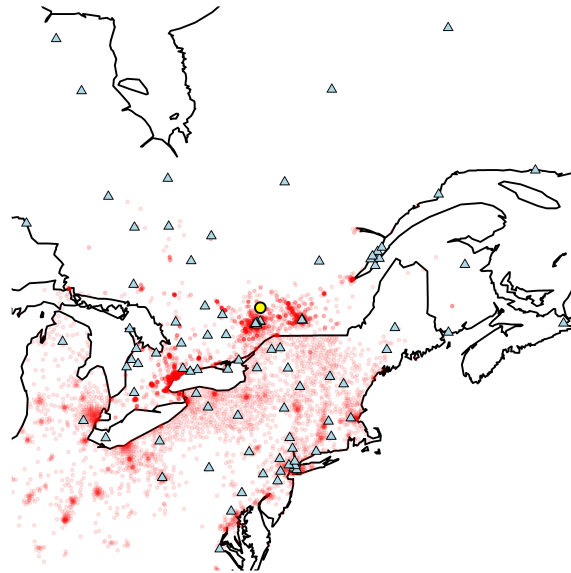


Figure 2.3: Yellow circle indicates location of the epicenter for the M5.0 2010 Val des Bois, Quebec earthquake. Individual felt intensity reports for Canada, and aggregate DYFI MMI values for the U.S. are shown as small red dots. Seismograph stations are shown as blue triangles.

2.4 Data Analysis Methods

One of the principal challenges of developing MMI-ground motion relationships is being able to appropriately match ground motion data with MMI information. Ground motion data are observed at discrete locations in space, while MMI from the DYFI algorithm data are inherently areal - an DYFI value represents an average intensity over a geographic region. This problem is exacerbated when intensity values are given as discrete observations, so that they must be aggregated and transformed into an areal average, and then matched with ground motion data. The following section describes how these various types of data are handled. The geostatistical problem of matching data types with different underlying spatial supports is also addressed in section 3.

2.4.1 Matching DYFI and Instrumental Data

In this study, the USGS DYFI data are filtered to include only zip codes or geocoded areas with at least three responses. Any single report is subject to human error, but averaging multiple reports within one region minimizes this error and greatly improves data quality. The aggregating of felt reports over geographical areas is problematic when one wants to match MMI values with local site condition information or ground motion parameters obtained from seismic stations, which are discrete points. In geostatistical terms, the MMI and instrumental data have different “supports”: the former is only meaningful over geographical areas, while the latter is defined only at a single location in space. The difficulty in trying to relate two types of data with differing supports is called the change-of-support problem (COSP), and is discussed in more detail in section 3.4.

Two methods are used to deal with this change of support problem. The first is to simply match each station with the MMI of the zip code in which it is located. This presents a difficulty for zip codes in which there are multiple stations - though they will certainly have different ground motion parameter values, they will be associated with the same MMI. To account for this, values of ground motion parameters are also averaged by area, in an attempt to achieve the same effect that aggregation has on MMI values; that is, the ground motion parameters used may sometimes represent averages over a geographical area rather than point observations.

While this approach seems like a reasonable solution to the problem, in fact we are losing information about the spatial variability of ground motions by considering averages over areas. An alternative method based on geostatistics can address the COSP presented by the areal MMI data. Kriging is a spatial interpolation technique that can use point observations to estimate averages over the same areal supports as the MMI data. This method is called *block kriging*. An overview of ordinary Kriging, block kriging and geostatistics is presented in section 3.

A summary illustration of the magnitudes and distances spanned by the matched data is given by figure 2.4. Magnitudes below 6 and distances between 10-100 km are well-represented. Above M 6.0, and at closer and farther distance ranges, the data are much more sparse.

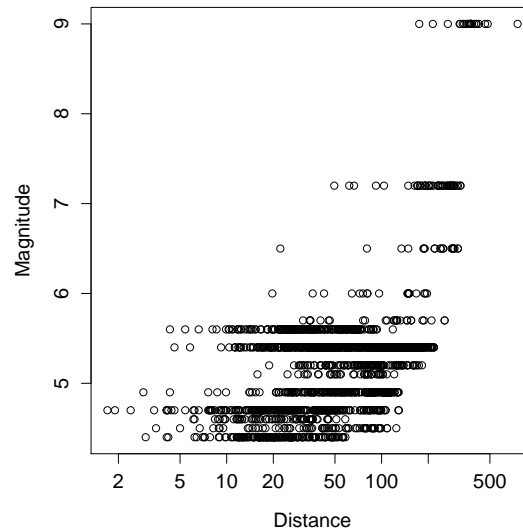


Figure 2.4: Distance-magnitude distribution of matched MMI-ground motion observations. Note the great concentration of data at M less than 6.0, and at distances 10-100 km.

2.4.2 Comparison of Geocoded and Zip-code aggregated Data Matching

MMI is an inherently areal form of data - it is defined as an average value over a region, and not as a specific point value. The Did You Feel It? website initially used zip codes as a convenient way to aggregate responses and derive MMI values. Zip codes are irregularly shaped and of variable size, ranging from a few city blocks to many square miles. The two types of aggregations are illustrated in Figure 2.5, which shows the DYFI intensities for the Borrego Springs event. The variability in zip code size is quite pronounced, particularly with respect to proximity to the Los Angeles area. Zip codes within the Los Angeles area are quite small, while those toward the eastern California border are extensive. When geocoding is used to aggregate on a regular grid, the large range of zip code sizes results in some trade-offs in data accuracy and the number of observations. Some of the larger zip codes are split into multiple intensity observations, while many of the smallest zip codes must be combined into a single intensity. This variability can be a source of error; the larger the zip code, the less precise an intensity average will be. Recently DYFI intensities have also been calculated using geocoded locations. If respondents provide their address to the DYFI website (90-95 % of them do), the exact geographical location in terms of longitude and latitude can be determined. Then these responses are grouped using a regular grid with areas of a few square kilometers. To match the geocoded DYFI intensities to instrumental stations, each

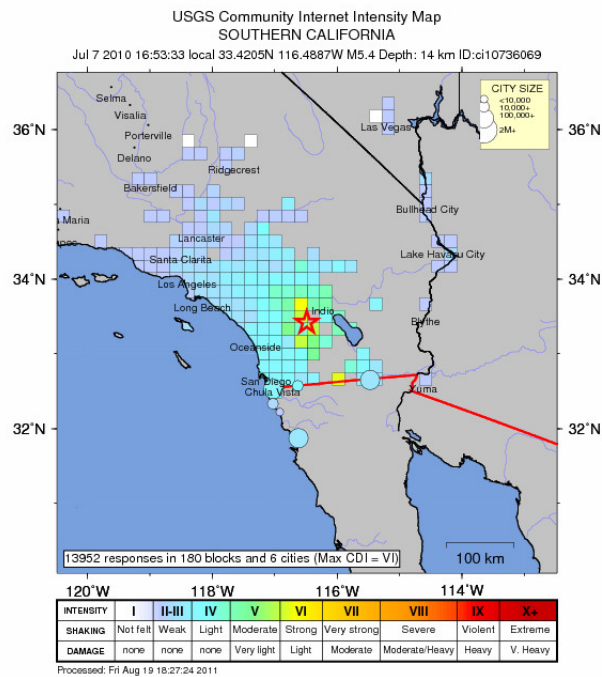
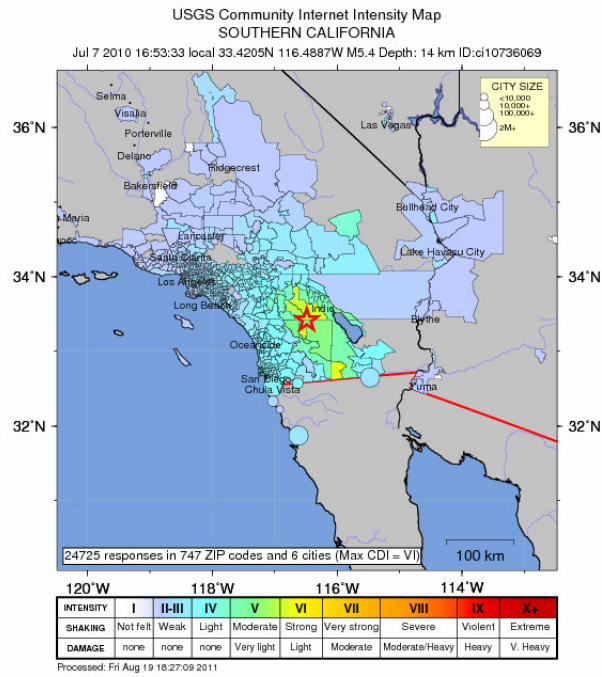


Figure 2.5: Comparison of DYFI intensity maps - zip code vs. geocoded [USGS, 2011]

DYFI intensity was assigned a “location” based on the longitude and latitude coordinates. Then all intensities within a given threshold distance from the station - typically 2 or 3 km - were averaged to give the matched MMI value. As with the zip code based DYFI data, only stations associated with more than 3 DYFI responses were retained. And just as with the Canadian intensity data, this method suffers from the error associated with averaging computed MMI values rather than averaging questionnaire responses.

How do the matched station-geocoded MMI data compare to the matched zip code based data? A number of the earthquakes studied here had both zip code based MMI and geocoded MMI values, allowing for a comparison between the two methods of aggregation. Figure 2.6 plots the two types of data against one another. While most of the points lie near the 45 degree line, there are a few outliers where the zip coded data are markedly dissimilar from the geocoded data.

Figure 2.7 shows the MMI-PGV relationship for matched data from the M 5.4 2010 Borrego Springs earthquake, using both zip codes and geocoding. The four plots show progressively larger distance tolerances for the geocoded data, up to 5 km from the station to the DYFI grid. The first thing to notice is the number of observations in each case: with the zip code data, there are 294 matched stations, while at 2 km distance there are only 26. At 3,4 and 5 km there are 59, 106 and 184 matched observations, respectively. Geocoded matches also appear to represent a subset of the zip code matches. And while the comparison using a single earthquake does not reveal any marked reduction in scatter, we can see in Figure 2.8 that when multiple events are combined, the geocoded data does indeed filter out many of the outlying points. This is to be expected, as the geocoded data should more accurately reflect “true” MMI values, allowing for closer station-MMI matches and better matched data. Figure 2.9 plots residuals from regressions using zip-code matched data against residuals from regressions using geocoded data. The 45 degree line is plotted for reference, and most of the points are clustered around this line, with a view outliers.

2.4.3 Canadian Intensity Data Analysis

To make the Canadian felt intensity reports commensurate with the aggregated DYFI data, the individual intensity responses are averaged in a similar manner. For each seismograph station, we find all intensity observations within a specified distance. This matching distance was set to 2km, 5km, 10km and 15km, to see how including MMI-ground motion matches at greater distances

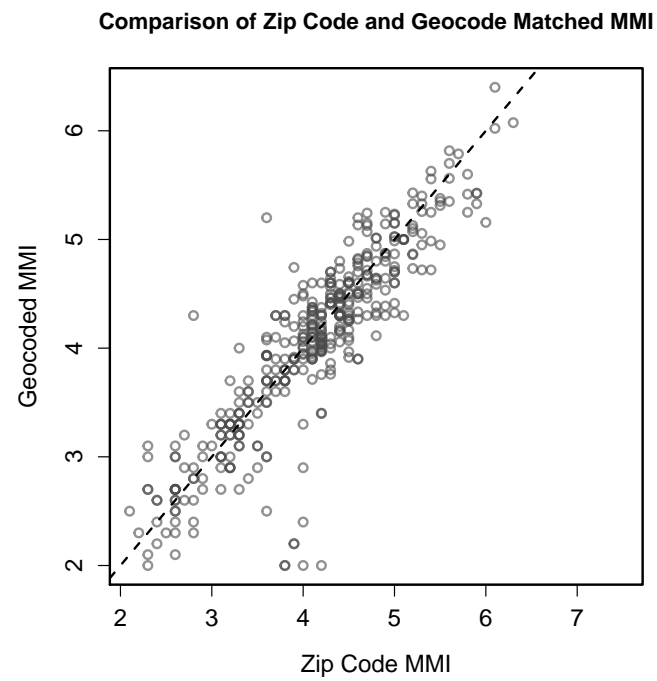


Figure 2.6: Comparison of MMI values matched with stations from Zip coded DYFI intensities and Geocoded DYFI intensities.

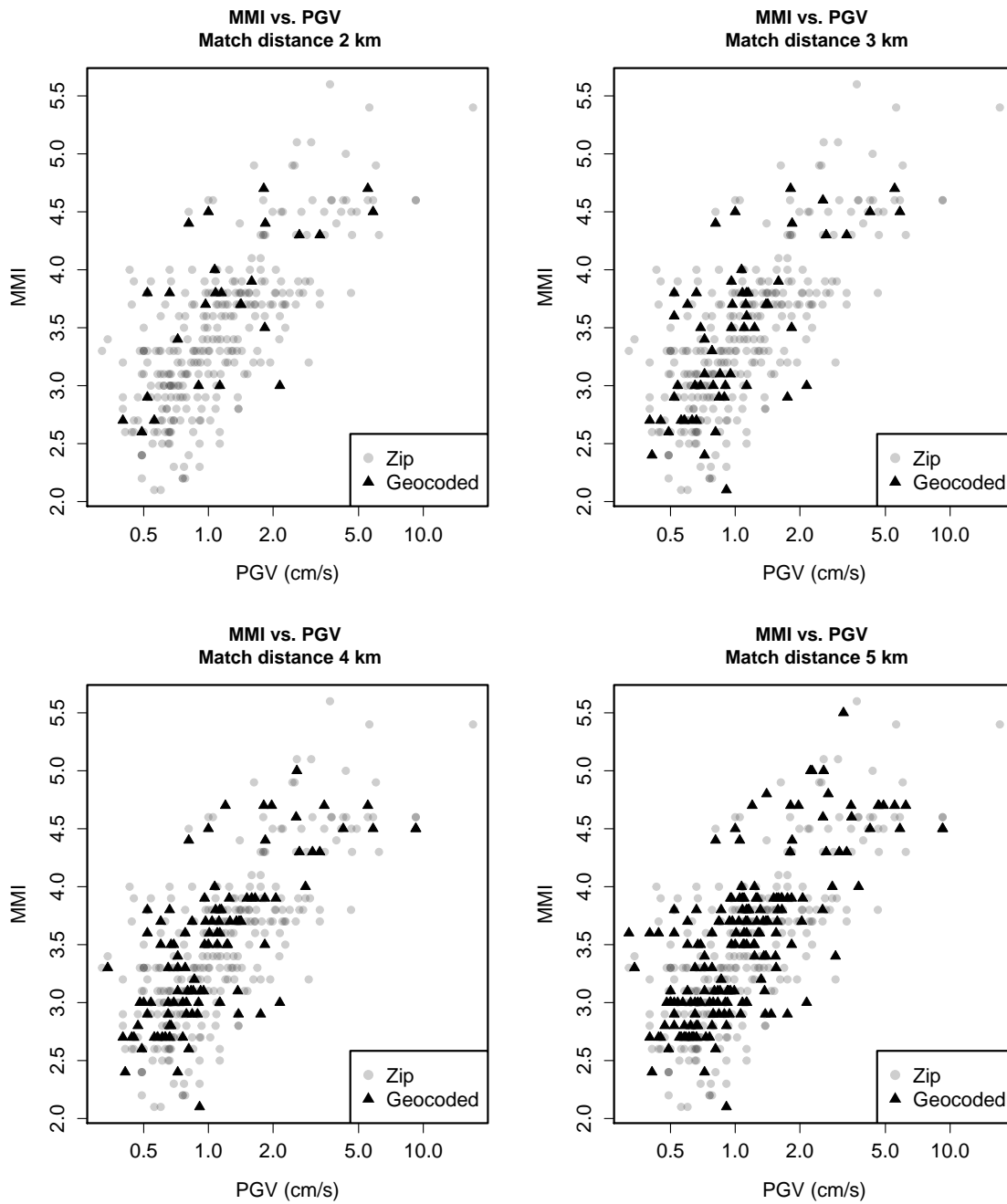


Figure 2.7: Comparison of matched MMI-PGV observations from geocoded and zip code-based DYFI intensities for Borrego Springs event. The four plots show geocoded MMI matched at increasing distances from ground motion observations.

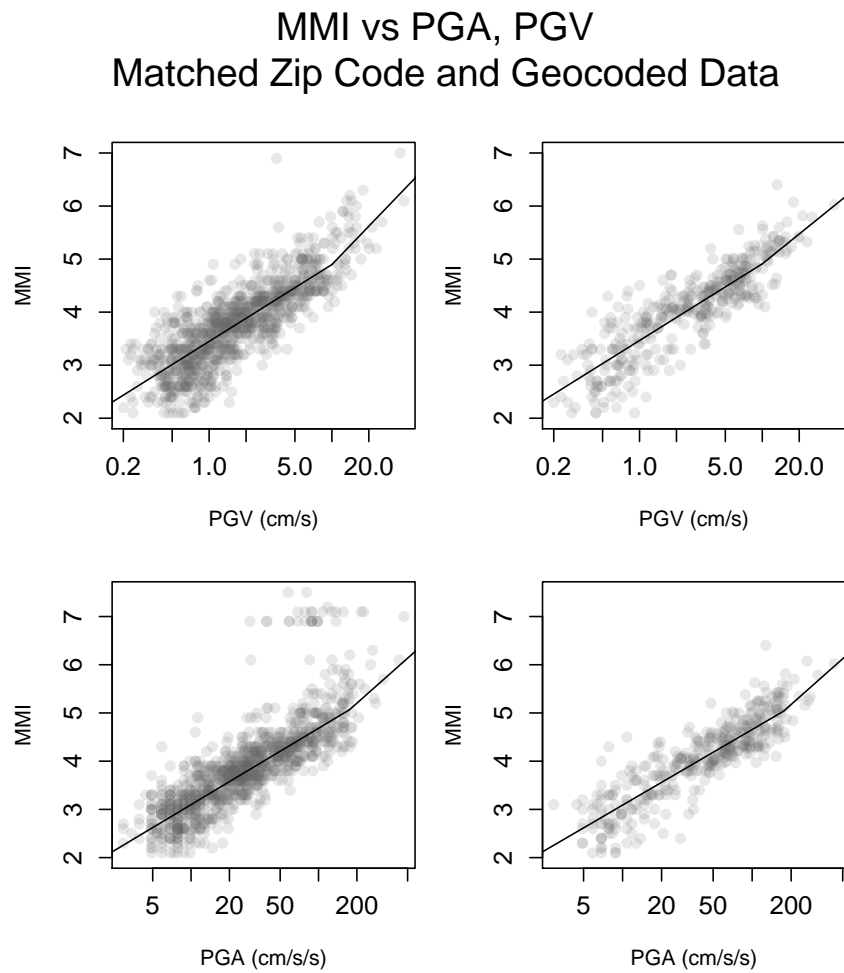


Figure 2.8: Comparison of matched instrumental-MMI observations for PGV (top) and PGA (bottom) from geocoded (right) and zip code-based (left) DYFI intensities for all earthquakes with both types of DYFI intensities.

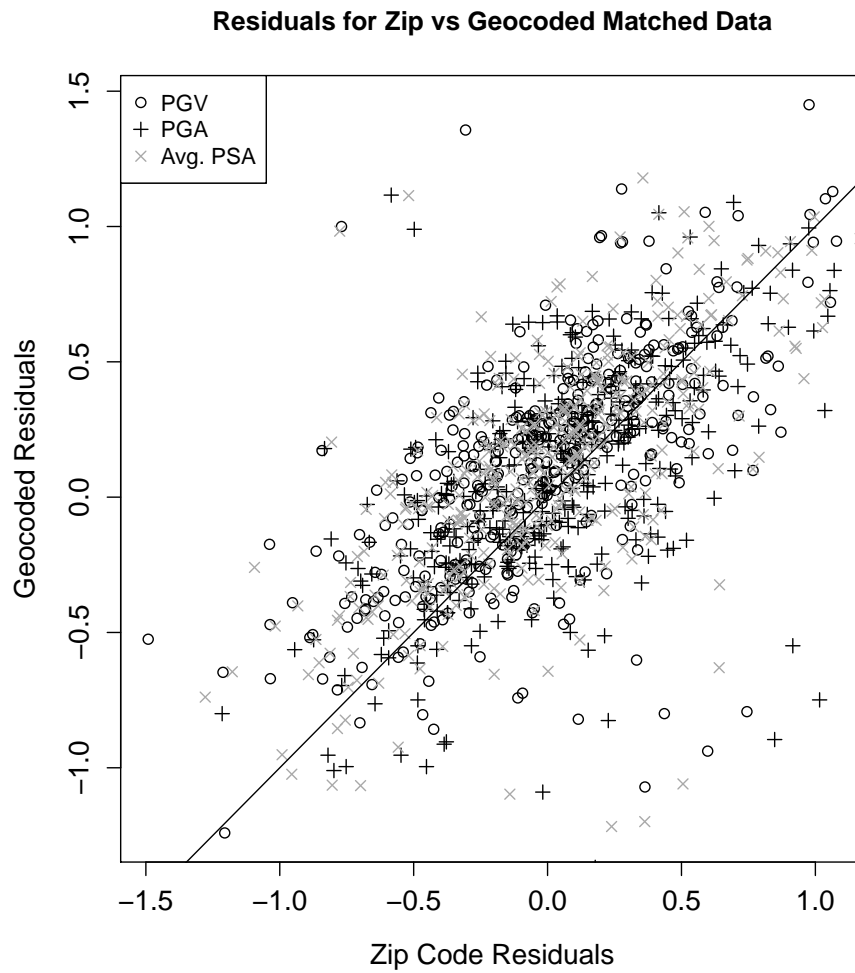


Figure 2.9: Plot of zip code residuals against geocoded residuals for all MMI vs. ground motion regressions. The 45-degree line is plotted for reference.

affected the quality of the data. As this distance threshold was increased, the variability of the matched data increased. There was no obvious indication about the distance threshold at which matching MMI data with ground motion observations is no longer very meaningful. As such, the somewhat arbitrary distance of 2km was chosen, in accordance with other studies [Worden et al., 2011].

It should be noted that this averaging method differs slightly from that used by the USGS DYFI website. There, the *responses* to individual questions are averaged *before* they are used to derive an MMI value, instead of computing the MMI value for each questionnaire and then averaging them. The above procedure is performed on the intensity data only when it needs to be matched with the instrumental data. For purposes in which the intensity data are used alone (e.g. in looking at the attenuation of MMI with distance), they are averaged over a grid. The cell size of the grid is determined using a trial-and-error procedure, in order to achieve a balance between the number of MMI averages, and the number of discrete responses for each average. If the cells are too big, there are too few MMI values for any meaningful analysis; too small, and we risk errors associated with MMI values derived from only a handful of responses.

For the M5.0 2010 Val des Bois earthquake, the chosen cell size is $.15 \times .15$ degrees; this yields 6875 cells, 476 of which contain an intensity observation and 206 of which contain 3 or more observations. For the M 2005 Riviere du Loup, the cell size is $.25 \times .25$ degrees, the number of cells encompassing observations is 164, with 95 containing at least 3 observations. The study region and gridding for both Val des Bois and Riviere du Loup is shown in Figure 2.10. Within each cell, the MMI for all observations are averaged, and the epicentral distance is taken to be the distance from the center of the cell.

2.4.4 Estimating Site Condition

As seismic waves travel through the earth, their amplitudes are influenced by the materials with which they come into contact. The soil profile at a location on Earth's surface can significantly affect the magnitude of ground motions and felt intensity experienced at the site. The energy of a seismic wave is the product of its velocity and amplitude. When the seismic wave travels from a higher velocity layer (like hard bedrock) to a slower layer (like soft topsoils), its velocity decreases. As a consequence of the conservation of energy, the seismic wave's amplitude must increase, and

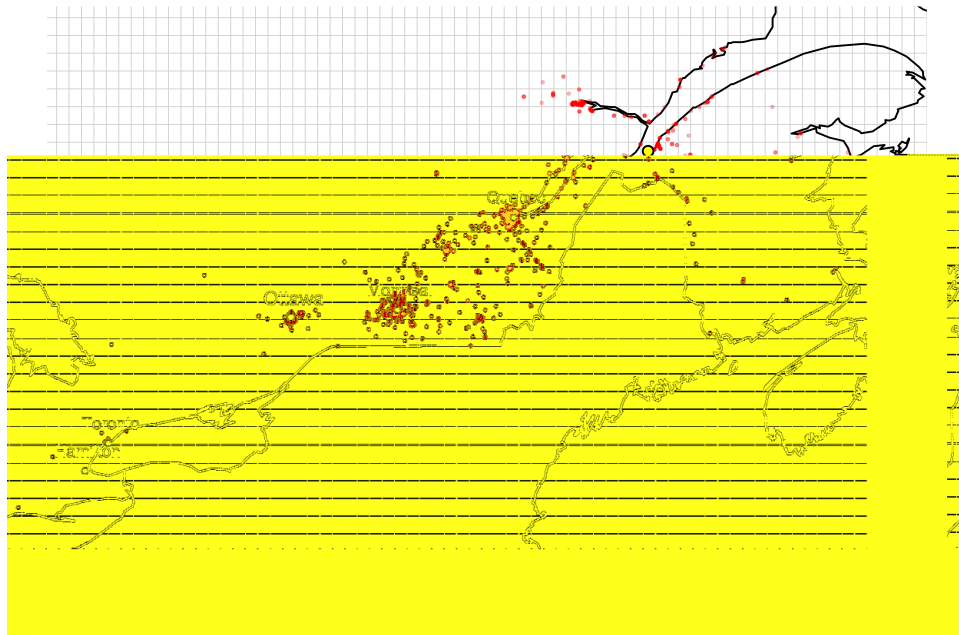


Figure 2.10: The gridding scheme used to average the Riviere du Loup intensity data. The large yellow circle shows the earthquake's epicenter. Red circles indicate individual intensity observations. Major cities are labeled. All observations within a grid cell are averaged, and assigned a location equal to the center of the cell.

ground shaking is amplified. Amplification is typically greater for longer periods of vibration than for shorter ones. A widely used proxy for seismic amplification is the average shear wave velocity to 30 meters (V_{S30}), although the strength of its correlation with amplification factors is disputed [Castellaro et al., 2008]. Shear waves are characterized by the direction in which they displace material as they travel; primary waves cause rarefaction and compression in the direction of travel, whereas shear waves displace particles perpendicular to their direction, producing shear stresses.

To illustrate, consider a simple model from Chen and Scawthorn [2003] in which a shear wave travels vertically through a soil layer with thickness H overlaying bedrock (Figure 2.11). Let the shear wave velocity of this layer be V_S . The horizontal displacement through the soil layer is then

$$u(z, t) = A \cos \frac{\omega}{V_S} z e^{i\omega t} \quad (2.1)$$

Where ω is the frequency of ground motions. Amplification is defined as ratio of maximum displacement on the surface to maximum displacement at the bottom of the soil layer. In this model, the amplification is

$$A = \frac{1}{\sqrt{\cos^2(\omega H/V_S) + (\eta\omega H/V_S)^2}} \quad (2.2)$$

where η is the damping ratio. Amplification depends on the frequency of ground motion, achieving a maximum at frequencies ω_n , $n = 0, 1, 2, \dots$ for which

$$\omega_n = \frac{V_S}{H} \left(\frac{\pi}{2} + n\pi \right) \quad (2.3)$$

The ω term in the denominator of Equation (2.2) means that amplification decreases with frequency. Thus the maximum amplification occurs at the lowest frequency, when $n = 0$ and $\omega_n = \frac{\pi V_S}{2H}$. This is called the fundamental frequency, and the period of vibration to which it corresponds is called the characteristic or fundamental site period T_0 :

$$T_0 = \frac{4H}{V_S} \quad (2.4)$$

This dependence of the fundamental period (and hence the magnitude of the amplification) on the shear wave velocity is what motivates the use of V_{S30} as a proxy for amplification at a site [Finn, 2001]. The National Earthquake Hazards Reduction Program (NEHRP) has a classification scheme for site conditions based on V_{S30} values, described in table 2.4.4. While it is convenient to use a single parameter to characterize site response, variability in soil conditions can occur over

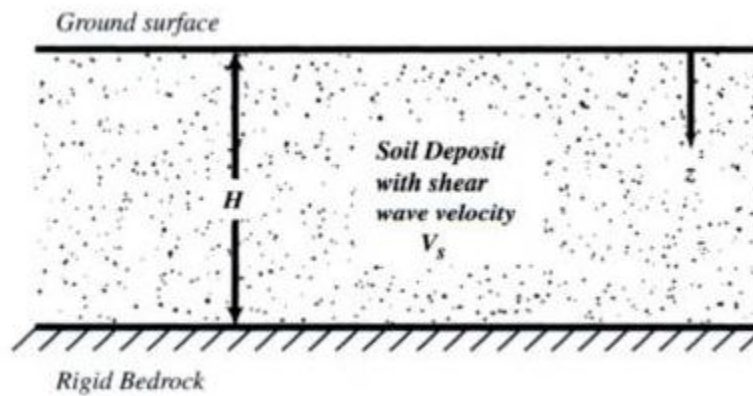


Figure 2.11: Illustration from Chen and Scawthorn [2003]. Bedrock overlain with a soil deposit of depth H and shear wave velocity V_s .

Site Class	Physical Description	Shear wave velocity (m/s)
A	Hard Rock	> 1500
B	Rock	760-1500
C	Very dense soil and soft rock	360-760
D	Stiff soil	180-360
E	Soft soil	< 160

Table 2.2: Shear wave velocity values corresponding to NEHRP site classes.

minute distance scales, which are often finer than the number of shear wave velocity measurements available. To a certain extent, the geologic features of an area can give indications of local site conditions.

Wald and Allen [2007] use topographic slope as a proxy for V_{s30} measurements, reasoning that steep slopes must be comprised of hard rock materials (lest they be quickly eroded), while soft, fine soils are ultimately deposited in flat regions. Often site response must be estimated using two sets of proxies: the correlation between either surficial geology or topographic slope is used to estimate V_{s30} , and V_{s30} in turn is used as a proxy for site response [Borcherdt, 1994]. Direct measurements of V_{s30} are ideal to estimate site response, but are not widely available for many regions, especially at the sites where ground motions are recorded.

The V_{s30} values used for California consist of 802 direct V_{s30} measurements. Figure 2.12 shows the locations of the V_{s30} measurements, as well as the locations of all ground motion measurements from California. This figure illustrates that most of the ground motion measurements appear to be close by to at least one V_{s30} measurement. Figure 2.13 plots a histogram showing how the data are distributed. The density of V_{s30} measurements made matching them with stations relatively simple; Any V_{s30} values from within 2 km of a station are averaged, and most stations have at least one measurement within this range.

With the available V_{s30} information from California it is possible to compare direct measurements with those estimated using the topographic slope method of Wald and Allen [2007]. A V_{s30} map of California was obtained using the global V_{s30} Map Server from the USGS [Allen and Wald, 2011]. These estimates were averaged over California zip codes. The direct V_{s30} measurements were similarly averaged, and then the difference between them (direct - topographic slope estimate) for each zip code was computed. The results are shown in figure 2.14.

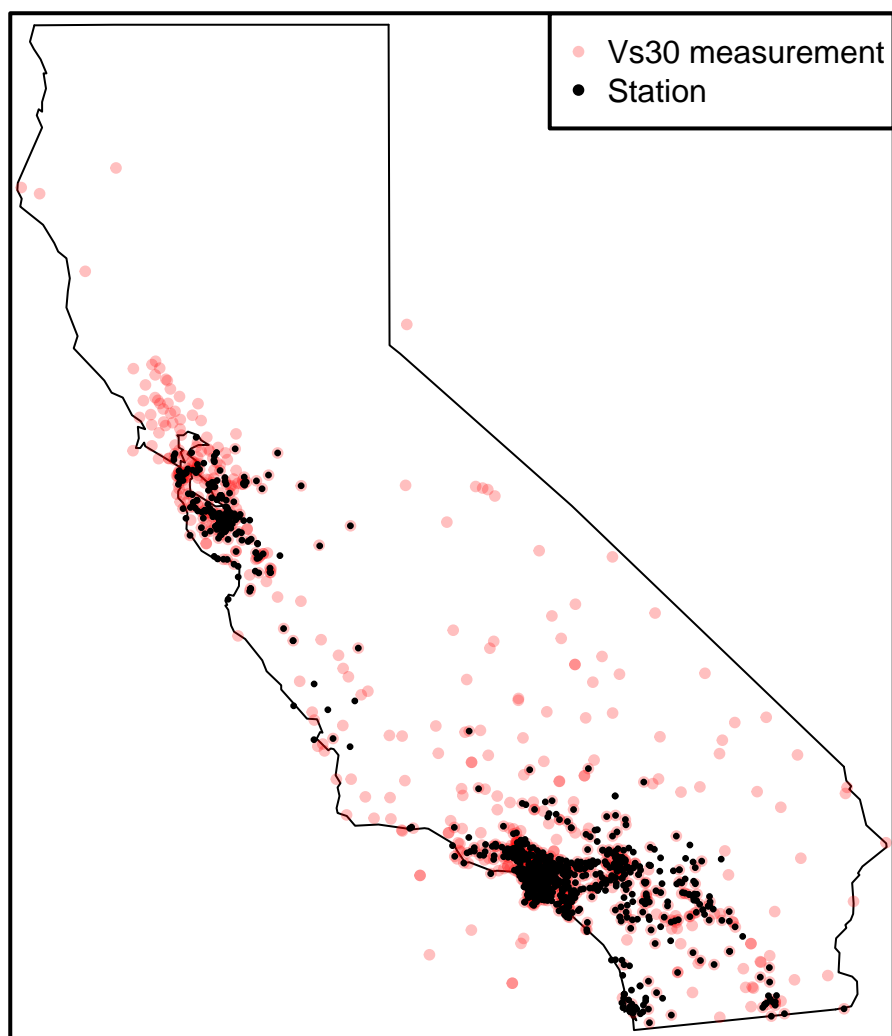


Figure 2.12: Locations of V_{S30} measurements (red) and stations (black) in California.

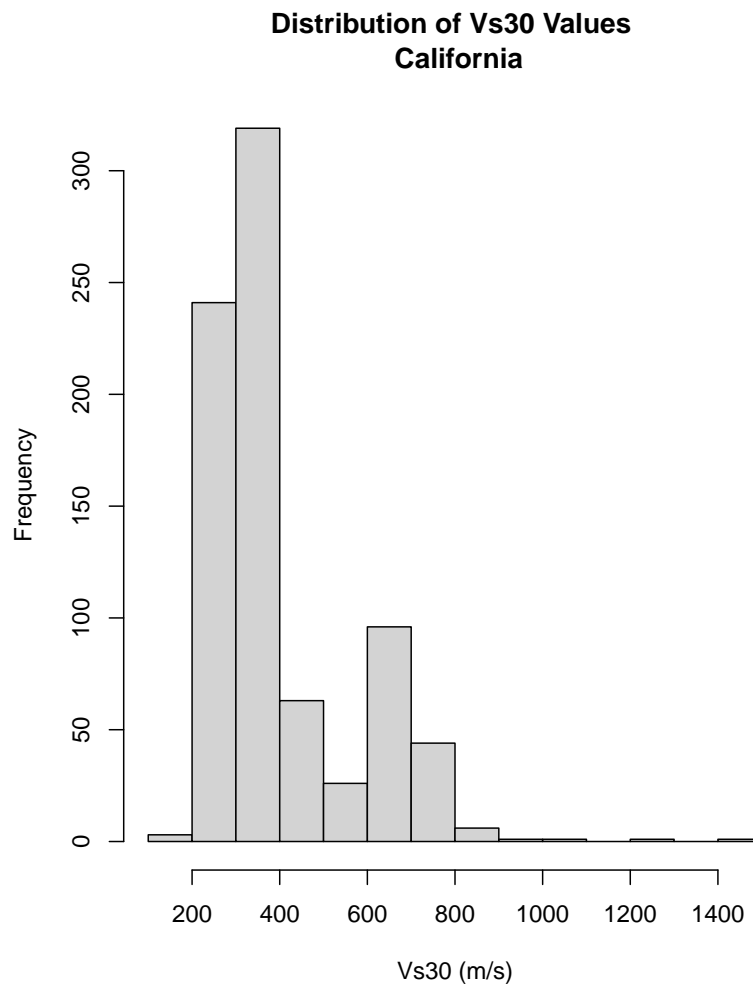


Figure 2.13: Histogram showing the distribution of V_{S30} values in California.

Figure 2.14: Averages over zip codes of Chiu et al direct V_{S30} values minus averages of topographic slope estimates from Allen and Wald [2011]

Chapter 3

Geostatistics

3.1 Overview

Geostatistics is a branch of statistics dealing with the analysis of correlated spatial data. It originated with South-African mining engineer D.Krige, who addressed the problem of predicting the gold content of unmined blocks [Gómez-Hernández, 1996]. The technique he employed, which was later termed “Kriging” in deference to Krige’s contributions to the field, is one of the fundamental elements of geostatistics. He described the method as “the *multiple regression* procedure for arriving at the *best linear unbiased predictor* or best linear *weighted moving average* predictor of the ore grade of an ore block (of any size) by assigning an optimum set of weights to all the available and relevant data inside and outside the ore block.” [Cressie, 1990] His methods were later consolidated and formalized by G.Matheron [1963], distinguishing geostatistics as a field in its own right, apart from classical statistics.

Geostatistical methods assume (and make use of) an underlying structure or continuity to spatial data, so that nearby observations are more likely to have similar values than those located farther apart. For geostatistics to be applicable to a variable or physical quantity it must exhibit gradual changes in value, without sharp or random transitions from high to low. A *regionalized variable* $z(\mathbf{x})$ is one whose value depends on its location in space \mathbf{x} , and which exhibit both spatial correlation and irregularity on a smaller scale [Chils and Delfiner, 1999].

Kriging is one of the fundamental techniques of geostatistics, used to interpolate known quantities of a variable to estimate values at unknown locations. It takes advantage of the assumption that the data Z are correlated and not random - that they are comprised of a spatial trend component μ as well as a stochastic residual component ϵ [Journel and Rossi, 1988]:

$$Z(\mathbf{x}) = \mu(\mathbf{x}) + \epsilon(\mathbf{x}) \quad (3.1)$$

where the expected value of the data $\mathbf{E}[Z(\mathbf{x})] = \mu(\mathbf{x})$ and $\mathbf{E}[\epsilon(\mathbf{x})] = 0$.

This dichotomy between structured and random aspects of spatial variables aptly describes the behavior of earthquake ground motions. While MMI and ground motion parameters exhibit very clear and predictable attenuation characteristics on a large scale, there are also significant local variations in both ground shaking intensity and its physical effects. As Carr and Glass [1985] points out, we can have two smokestacks located 50ft apart, and one may be destroyed by an earthquake, while the other is unscathed.

3.2 Ordinary Kriging

Consider a set of known data observations $Z(\mathbf{x})$. If we assume the trend component $\mu(\mathbf{x})$ is unknown, the *ordinary kriging predictor* at a point x_0 is a linear weighted average of values at a set of n known locations:

$$\sum_{i=1}^n \alpha_i Z(x_i) \quad (3.2)$$

This weighted average may include all known values, or only a subset of values within a certain threshold distance from the point of interest. The scope of the weighted average - i.e. which points are included in the estimation - depends on the covariance structure of the data. This will be discussed in the following section.

The weights α_i given above must sum to one, and are selected so as to minimize the mean-squared prediction error

$$\mathbf{E} \left(Z(x_0) - \sum_{i=1}^n \alpha_i Z(x_i) \right)^2 \quad (3.3)$$

Then solving for the best linear unbiased predictor (BLUP) is tantamount to solving for the ap-

propriate weights, α_i [Oliver, 2010]. The BLUP weights are solved in terms of the covariances between each of the samples (an $n \times n$ matrix \mathbf{C} with $(i, j)^{\text{th}}$ element $C(x_i, x_j)$), and the covariances between the samples and the point x_0 to be estimated: $\mathbf{c} = (C(x_0, x_1), C(x_0, x_2), \dots, C(x_0, x_n))'$:

$$\alpha = (\mathbf{c} + (1 - \mathbf{c}'\mathbf{C}^{-1}\mathbf{1})(\mathbf{1}'\mathbf{C}^{-1}\mathbf{1})^{-1}\mathbf{1})' \mathbf{C}^{-1} \quad (3.4)$$

This solution requires covariances between known and unknown locations, which are not provided by the data. Furthermore, the uniqueness of the BLUP requires that the covariance matrix be strictly positive definite. These two problems are addressed by using a positive definite *variogram model* to estimate these covariances.

3.3 Empirical Variogram and Variogram Modeling

The Kriging estimator depends crucially on the covariance structure of the data. Its existence and uniqueness requires that we have a model specifying the covariances between all pairs of points - both known and unknown - in the region of interest. The *empirical variogram* categorizes points based on their separation distance h (or vector \mathbf{h}), and computes the covariance for all pairs of points separated by h :

$$\gamma(h) = \frac{1}{2} E \left[(z(x) - z(x+h))^2 \right] \quad (3.5)$$

But the empirical variogram cannot be used directly in the Kriging estimation; it represents discrete estimates of covariance, whereas we need a continuous function capable of evaluating the covariance between any pair of points. A variogram model is used instead: a function which is positive definite and closely fits the empirical variogram. There are six classes of variogram models most commonly used: the nugget effect, spherical, exponential, Gaussian, linear and logarithmic. These models may be used together additively, since any linear combination of a positive definite function with nonnegative coefficients is also a positive definite function. For instance, the nugget effect model may be combined with the spherical model to capture the random noise observed in the empirical variogram when $\gamma(0) \neq 0$. Descriptions of these models are provided below, and plotted in Figure 3.1.

Variogram models have three important features: the nugget, sill and range. The nugget is the y-axis intercept. Theoretically, variograms should be increasing functions that are zero at zero

separation distance. Values at points that are nearly overlapping should be nearly identical, and as we consider pairs that are farther apart, we expect their covariances to increase. In practice, this is not always the case, and the empirical variogram may be nonzero at small values of h . When this occurs, $\gamma(0)$ is referred to as the *nugget*. In a similar vein, we would like variograms to level off at some separation distance. This represents an upper bound on the distance at which points can “influence” one another. This defines the radius of a finite neighbourhood over which to perform the weighted average estimation. If a variogram displays this asymptotic behaviour, the maximum value is referred to as the *sill*, and the distance at which this leveling off occurs is called the *range*.

Three of the variogram models - the spherical, exponential and Gaussian - have similar shape, but differ in their behaviour near the origin and they way the sill is reached. These three models are called the *transition models*, because pairs of points are correlated if they are separated by less than the range, and uncorrelated at distances beyond the range. For transition models, the neighbourhood of points used in the corresponding Kriging estimate consists of all points within a distance less than the range of the variogram model. For all others, the neighbourhood is effectively global, and all known points are used in the estimate.

There are subtle differences between the three transition models. The spherical model is nearly linear near the origin, and levels out abruptly. The Gaussian model represents the greatest spatial continuity, with a change in inflection for small separation distances. It is relatively flat near the origin, with increasing steepness at intermediate distances. The exponential model exhibits linear behaviour near the origin, and gradually levels out to approach the sill asymptotically. The linear and logarithmic models do not level out - they are unbounded and have infinite range and sill. Depictions of bounded and unbounded variogram models are given in Figure 3.2.

Nugget Effect: Describes purely random noise,

$$\gamma(h) = \begin{cases} 0, & h = 0; \\ N, & h > 0. \end{cases} \quad (3.6)$$

Spherical: Let C be the sill and a be the range.

$$\gamma(h) = \begin{cases} C \left(\frac{3}{2} \frac{h}{a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right), & h \leq a; \\ C, & h > a. \end{cases} \quad (3.7)$$

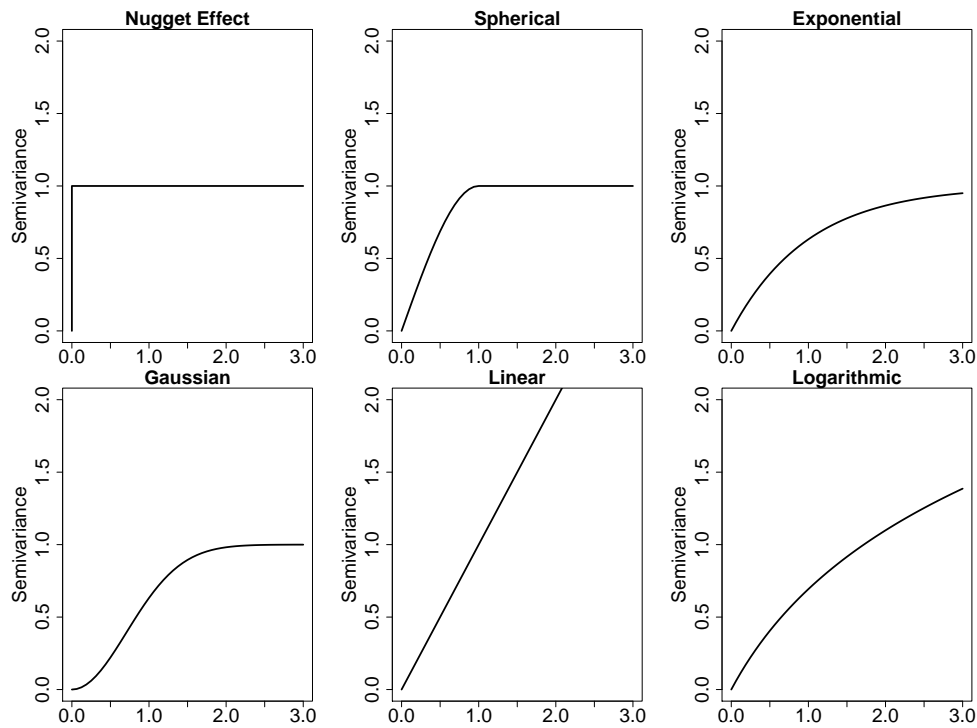


Figure 3.1: Six classes of commonly-used variogram models. The nugget effect captures the pure noise component of a variogram, when there is positive variance at zero separation distance. Spherical, exponential and Gaussian semivariograms are bounded, exhibiting increasing variance with separation distance up to a threshold distance called the range. Beyond the range, these transitional semivariograms level off at a value called the sill. The linear and logarithmic semivariograms are unbounded, and increase indefinitely [R Development Core Team, 2009].

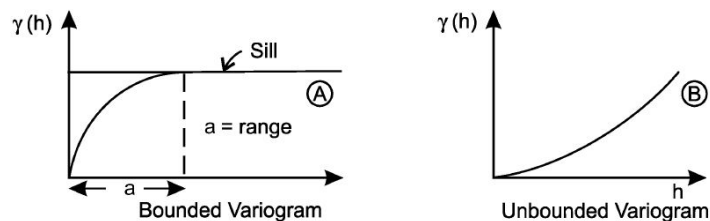


Figure 3.2: Examples of bounded (A) and unbounded (B) variogram models [Sarma, 2009]. Bounded variograms increase up to a certain separation distance, after which they level out at the sill. Unbounded variograms continue to increase indefinitely.

The tangent at the origin intercepts the sill at $\frac{2}{3}$ of the range.

Exponential:

$$\gamma(h) = C \left[1 - \exp\left(-\frac{h}{a}\right) \right] \quad (3.8)$$

Since the exponential model only approaches the sill asymptotically, the range is infinite. The *practical range* is the distance at which the variogram reaches 95 % of the sill. The tangent at the origin intersects the sill at about $\frac{1}{5}$ the range.

Gaussian:

$$\gamma(h) = C \left(1 - \exp\left[-\left(\frac{h}{a}\right)^2\right] \right) \quad (3.9)$$

Linear:

$$\gamma(h) = C \times h \quad (3.10)$$

Logarithmic:

$$\gamma(h) = C \log(h), \quad h > 0 \quad (3.11)$$

3.4 Block Kriging and The Change-of-Support Problem

The Kriging procedures described above work on discrete sets of observations, interpolating the values to create a continuous map of estimates. When Kriging is performed on ground motion data, we get an estimate of what the ground motion parameters might be at every location in the region of interest. But the DYFI data are not presented in this form; the DYFI responses are aggregated over discrete areal units, and don't form a continuous map. For purposes of matching ground motion and intensity data, we would like them to have the same type of spatial supports.

Many types of data do not make sense as point objects; health data such as cancer rates, for instance, are necessarily averaged over the irregular spatial units of counties or zip codes [Goovaerts, 2008]. Such areal spatial units are termed *blocks*. MMI is inherently aerial, and DYFI responses are unreliable as point sources, as a single intensity response is subject to human error. This is the fundamental tradeoff of aggregation - sacrificing small-scale spatial variations in the data to reduce noise associated with extreme or outlying values.

To match the DYFI data with the appropriate ground motion and site condition observations over geographical units, we must estimate the average values of the point-support ground motion observations over the zip codes or geocoded areas. This problem - when the nature of a process is inherently of one type of support, while our observations are made using another (usually larger) support - is known as the *change of support problem*, and is extensively addressed in the statistics literature, particularly with applications to ecology and biostatistics [Gelfand et al., 2001, Gotway and Young, 2002]. In this case, the change of support can be used to make the association between point ground motion observations and aerial intensity observations more intuitive.

To give the estimated ground motion data the same support (aerial) as the DYFI data, we use *block kriging*, a method for interpolating average values over an aerial region. In block kriging we have a set of point observations of a regionalized variable, and a set of areal supports over which we want to estimate its average value (see Figure 3.3). To do this, each support is deaggregated into a set of regularly spaced points. Ordinary Kriging is performed using these points, and then the average for all points within the support is the block kriging estimate. Details of the block Kriging mathematics are as follows:

Consider a variable with point support $Z(\mathbf{x})$, and let $\mathcal{A}(\mathbf{x}_0)$ be an aerial support with centroid \mathbf{x}_0 . The true average of $Z(\mathbf{x})$ over A is given by

$$Z_{\mathcal{A}}(\mathbf{x}_0) = \frac{1}{|\mathcal{A}|} \int_{A(\mathbf{x}_0)} Z(\mathbf{x}) d\mathbf{x} \quad (3.12)$$

and the unknown average value over A can be estimated using a weighted average of the known point observations, just as in point-to-point Kriging:

$$Z_{\mathcal{A}}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i) \quad (3.13)$$

With block kriging, the semivariogram uses each sample observation \mathbf{x}_i and all of the points within the area \mathcal{A} :

$$\gamma(\mathbf{x}_i, \mathcal{A}) = \frac{1}{|\mathcal{A}|} \int_{\mathbf{x} \in \mathcal{A}} \gamma(\mathbf{x}_i - \mathbf{x}) d\mathbf{x} \quad (3.14)$$

This integral can be approximated by discretizing \mathcal{A} and computing the variogram using this finite set of points \mathbf{x}_j , $j = 1, \dots, N$; the average semivariance between an observation \mathbf{x}_i and all of the

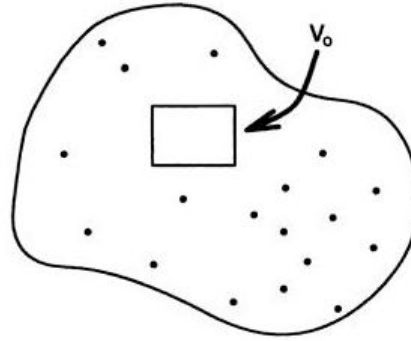


Figure 3.3: Block Kriging: An average of an areal support (box) is computed based on a set of individual observations (block points). The box is deaggregated into a grid of regularly spaced points, Kriging is performed over those points, and averaged. Source: Wackernagel [2003]

discretized points is then

$$\gamma(\mathbf{x}_i, \mathcal{A}) = \frac{1}{N} \sum_{j=1}^N \gamma(\mathbf{h}_{ij}) \quad (3.15)$$

where \mathbf{h}_{ij} is the distance between the observation \mathbf{x}_i and the discretized point in \mathcal{A} , \mathbf{x}_j [de Gruijter et al., 2006].

The above sections described the statistical underpinnings of the Kriging methods employed in this study. The following sections provide an overview of how Kriging has been applied to seismological data.

3.5 Application of Kriging to Seismic Data

While both ground motions and felt intensity exhibit consistent spatial patterns (namely attenuation with distance, although this can vary for different regions of North America), spatial variability can arise due to differences in location relative to the fault, differences in travel paths of seismic waves, and local site conditions [Zerva and Zervas, 2002]. These variations can profoundly impact the severity of ground motions and their physical manifestations. While information about the surficial geology of the site and shear wave velocity profiles can help capture these variations, their efficacy is limited by the complexities of seismic wave propagation and amplification, along with data availability. The spatial behaviour of ground motions, and in particular the correlation of

ground motions from a single earthquake at two different sites, is of special interest for insurance companies who want to limit the amount of correlated loss associated with an earthquake [Park et al., 2010].

Jayaram and Baker [2009] use variograms to assess the spatial correlation among residuals from the model

$$\ln(Y_{ij}) = \overline{\ln(Y_{ij})} + \epsilon_{ij} + \eta_j \quad (3.16)$$

where Y_{ij} is the ground motion parameter at location i from earthquake j , $\overline{Y_{ij}}$ is the median ground motion value predicted from GMPEs, ϵ_{ij} is the intra-event residual and η_j is the inter-event residual. The authors fit semivariograms to residuals from the above model using spectral acceleration from the 1994 **M6.7** Northridge and **M7.6** 1999 Chi-Chi earthquakes, looking specifically at how the range of the semivariogram changes with period. They find the exponential model to yield the best fit, especially for small separation distances h . The range (a measure of the geographic extent of correlation) is shown to decrease with increasing frequency of ground motion. At shorter periods, waves are more susceptible to minute heterogeneities as they travel from source to receiver, whereas long period waves are unaffected. Semivariogram ranges from the Chi-Chi earthquake were also significantly larger than those for Northridge, which the authors explain as increased correlation due to homogeneity of local site effects and correlation due to errors in V_{S30} measurement. The authors also calculate correlations for a few of the same earthquakes used in this study: 2003 **M5.4** Big Bear City, 2004 **M6.0** Parkfield, 2005 **M5.2** Anza, 2007 **M5.6** Alum Rock and 2008 **M5.4** Chino Hills, finding similar patterns in the period-range relationship.

Rubeis et al. [2005] apply Kriging to intensity data from an Italian earthquake, using the resulting estimates to reduce scatter in intensity attenuation curves.

3.5.1 Methods in Kriging Seismic Data

In this study, Kriging is used to provide visual details about spatial correlations among MMI, ground motions and V_{S30} . It is also used in an attempt to compute average ground motions over zip code areas, as an alternative matching procedure for MMI and ground motion data. R Development Core Team [2009], a statistical programming language, contains many user-written packages for spatial data analysis. A brief outline of the steps used in performing ordinary kriging is as follows:

Estimate the trend surface and collect residuals Before one can use ordinary kriging (OK), we have to remove any large-scale trend in the data. As already mentioned, the major feature of earthquake intensity and ground motion data is attenuation with epicentral distance. This trend was estimated two different ways. In the first attempt, the best-fitting degree-2 polynomial surface was used; however, these surfaces ultimately did a poor job of matching the data, and instead the simple linear regression of MMI against distance was used.

Create spatial grid over which to perform kriging estimates One must create a grid with the appropriate number and closeness of points necessary for the spatial analysis. This was typically a 0.1 degree square grid, although for smaller magnitude events a greater fineness of 0.05 degrees was used.

Compute and plot empirical variogram A function from the **R** package `gstat` computes the empirical semivariogram and plots it.

Select and fit variogram model Selecting the appropriate variogram model type was done visually, by overlaying best-fitting models on the empirical variogram. The package `gstat` contains a function which searches for the optimal parameters (sill, range, nugget) given an empirical variogram, a model type and a nugget effect. Best-fitting are superimposed on the empirical variogram to assess which model is appropriate.

Perform ordinary kriging With the variogram model in hand, ordinary kriging of the trend-surface residuals over the spatial grid is performed. The function `predict` from the `gstat` package computes Kriging estimates given the variogram, variogram model, and grid.

Add trend component back in The final step is to compute the trend component for all points in the grid, and add this to the Kriged residuals.

Visualization Maps of Kriging estimates are visualized by converting the grid of trend and residuals components to an image using the function `as.image.SpatialGridDataFrame`. The resulting image is overlaid on a map of California.

The final step in the Kriging procedure used here is cross-validation. Because the trend component of the data is removed prior to estimation, the known values of the parameters will not be identical to the Kriging estimates at those locations. This provides an opportunity to compare known values with estimates.

3.5.2 Cross-validation of Kriging results

Cross-validation of Kriging estimates is used to test their robustness to variations in data quality. It is performed using a jack-knifing method, in which some samples are removed prior to Kriging and compared with their estimates. For the direct V_{S30} data from California, a random sample of 100 of the 802 points are removed; the variogram estimation and Kriging procedure are performed, and the predicted values are compared to the actual values of the original 100 sample points. Histograms of the actual-predicted residuals from cross-validation of the Kriging of V_{S30} data are shown in figure 3.4. As desired, most of the residuals cluster around zero, although some outliers of very large differences occur. After introducing the motivation for this study, the fundamental concepts involved, and the data and statistical methods used, the following sections will begin to present results from this study. These results will progress from simple summary statistics and correlation coefficients, to more informative analyses using regression statistics and geostatistical methods. The latter half of this report will conclude with a discussion of these results and a conclusion.

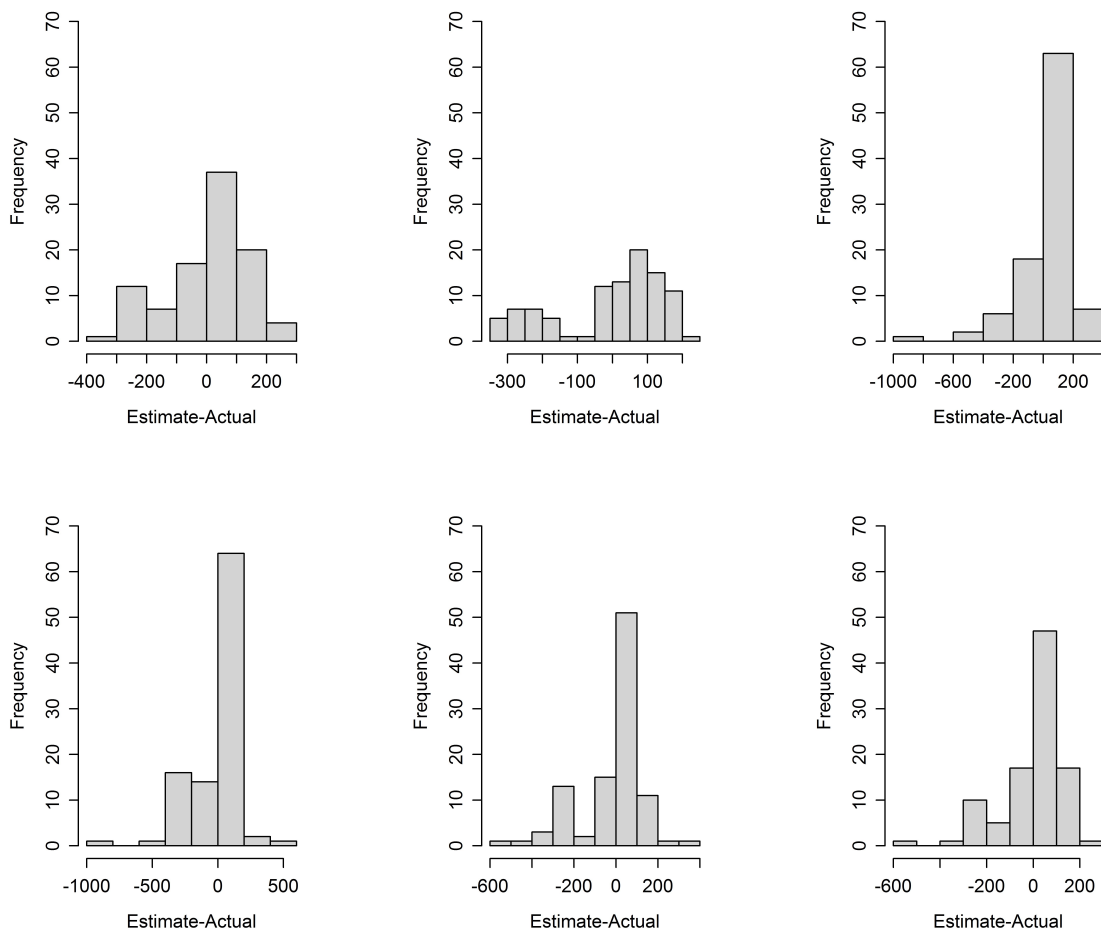


Figure 3.4: Sample results from cross-validation of Vs30 Kriging. Each plot represents a separate simulation in which a sample of 100 of 802 points are removed, and Kriging is performed. The actual data points are then compared with their estimates, and a histogram of the differences are displayed.

Chapter 4

Results

4.1 Summary Statistics

To get a sense for the scope of the data used in this study, summary statistics are provided for the three types of data sets: the DYFI responses (Table 4.1), the ground motion data (Table 4.2) and the matched ground motion-MMI data (Table 4.3). For the DYFI data, information is provided about whether geocoded or zip code locations were used, the number of observations, mean and maximum MMI values, and the distance range of observations. The number of DYFI observations varies greatly based on the size and location of the earthquake, with a minimum of 180 for the **M** 5.7 Ocotillo event to a maximum of 2410 for the **M** 5.4 Mount Carmel event.

Greater numbers of observations allow for more accurate matching of DYFI with ground motion amplitude data. To illustrate which events are best documented in terms of both DYFI and ground motion data, Figure 4.1 plots the number of both kinds of observations for each event. The events with over 400 DYFI observations and over 140 ground motion observations are labeled; these include the 2008 **M**5.4 Mount Carmel, 2007 **M**5.6 Alum Rock, 2009 **M**4.7 Inglewood, 2005 **M**4.9 Yucaipa, 2010 **M**5.4 Borrego Springs and 2008 **M**5.4 Chino Hills events. These events will be the focus of greater in-depth analysis because of the better quality of matched data. In addition, the 2005 **M**5.2 Anza, 2010 **M**5.4 Big Bear City and 2010 **M**4.4 Whittier Narrows events are included in this set of well-documented events, based on the large number of matched observations available (see Table 4.3).

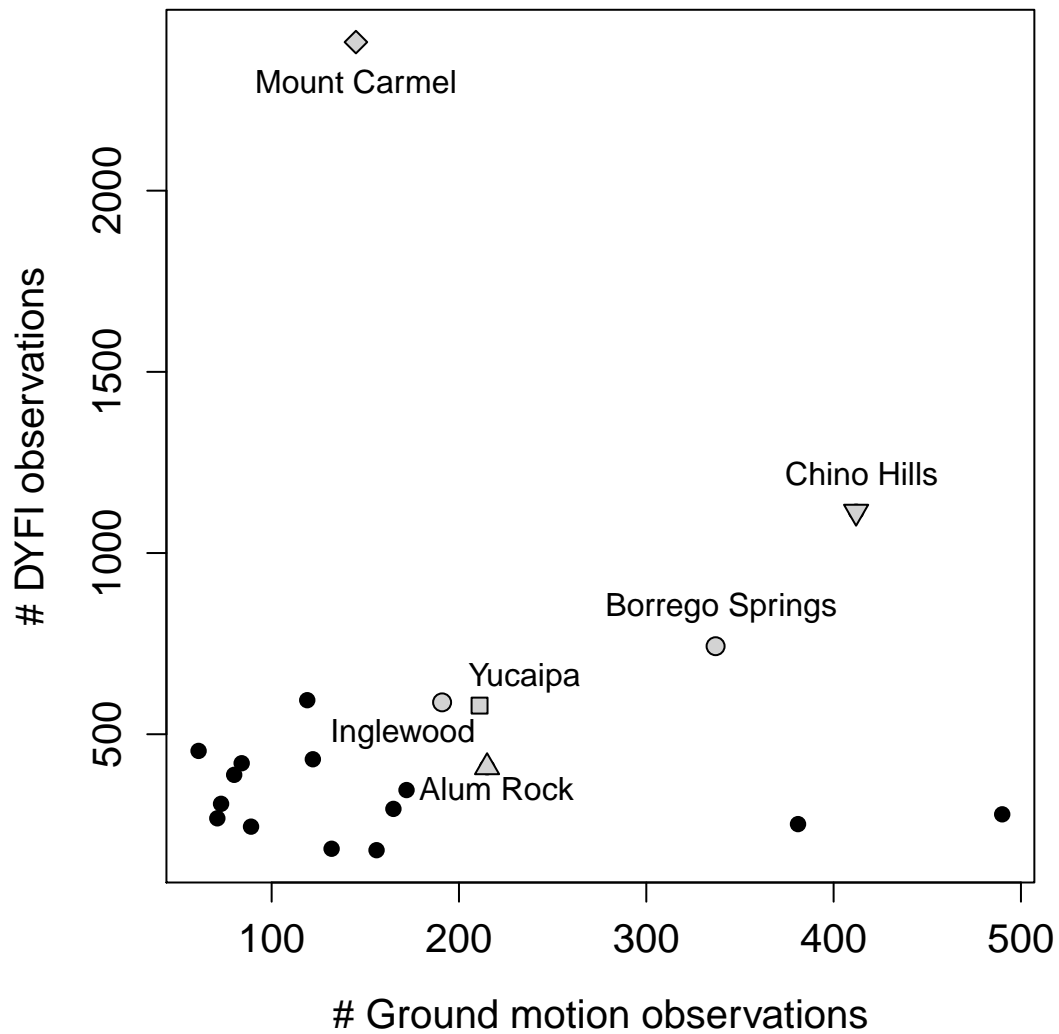


Figure 4.1: Shows the number of DYFI observations versus number of ground motion observations for all events. The events with over 400 DYFI observations and over 140 ground motion observations are highlighted; these include the 2007 **M** 5.6 Alum Rock, 2008 **M** 5.4 Mount Carmel, 2009 **M** 4.7 Inglewood, 2005 **M** 4.9 Yucaipa, 2010 **M** 5.4 Borrego Springs and 2008 **M** 5.4 Chino Hills events.

Event	Type	# obs.	Max. MMI	Mean MMI	Distances (km)
Alum Rock	Zip	409	6.3	3.3	3-282
Anza 2001	Zip	308	4.8	2.8	9-312
Anza 2005	Zip	594	5.6	2.9	10-338
Baja	Geocoded	279	7.6	4.1	17-2118
Big Bear 1992	Zip	184	8.7	4.7	5-652
Big Bear 2003	Zip	431	5.6	2.6	7-254
Borrego Springs	Zip	743	5.8	2.9	18-2388
Chino Hills	Geocoded	1114	7.3	4.5	1-172
Chatsworth	Zip	268	5.2	2.9	2-132
Inglewood	Zip	588	5.4	2.9	0-347
Lake Elsinore	Zip	454	4.4	2.6	3-134
Milpitas	Geocoded	245	4.3	2.2	4-154
Mineral	Zip	473	5.2	2.8	5-691
Mount Carmel	Zip	2410	6.9	3.1	4-849
Ocotillo	Geocoded	180	7.5	3.0	12-411
Parkfield	Geocoded	420	7.3	2.7	11-413
Riviere du Loup	Location	2164	6.0	4.4	6-1053
San Bernardino	Geocoded	346	4.6	2.6	6-187
San Simeon	Geocoded	338	8.1	3.1	10-489
Tohoku	Cities	252	7.9	4.5	103-3618
Val des Bois	Location, Zip	8479	7.0	3.0	7-2581
Whittier Narrows	Geocoded	294	4.6	2.2	4-170
Yorba Linda	Zip	388	4.6	2.8	2-144
Yucaipa	Zip	579	5.5	3.3	2-579

Table 4.1: Summary statistics for DYFI data. Includes the type of intensity response (either aggregated by zip code or geocoded locations from DYFI data, or as single response locations for Canadian events), the number of locations (for zip codes, these must represent an area with more than 3 responses), the mean and maximum MMI value, and the range of distances covered.

The instrumental ground motion data span a wide range of distances and ground motion values. Instrumental data are generally less abundant than intensity data. Nonetheless, some events are quite well-recorded, with ground motion measurements from 200-400 different stations. PGV values range from 0.01 to 62.9 cm/s, while PGA ranges from 0.13 to 1249 cm/s². The epicentral distances are depicted in Figure 4.2. The largest number of observations occur in the 25-75 km range, and generally at nearer distances of less than 200km. At distances greater than 200km, the numbers of observations fall to about 10% of those within 200km, due to increasing noise that degrades the instrumental signal.

These well-documented earthquakes allow for detailed analysis of ground motions, Modified Mercalli Intensity, and their interactions, as well as the role of contributing variables like region, distance, magnitude and site condition on these relationships. Subsequent sections will explore these relationships using correlations, regression statistics and spatial interpolation.

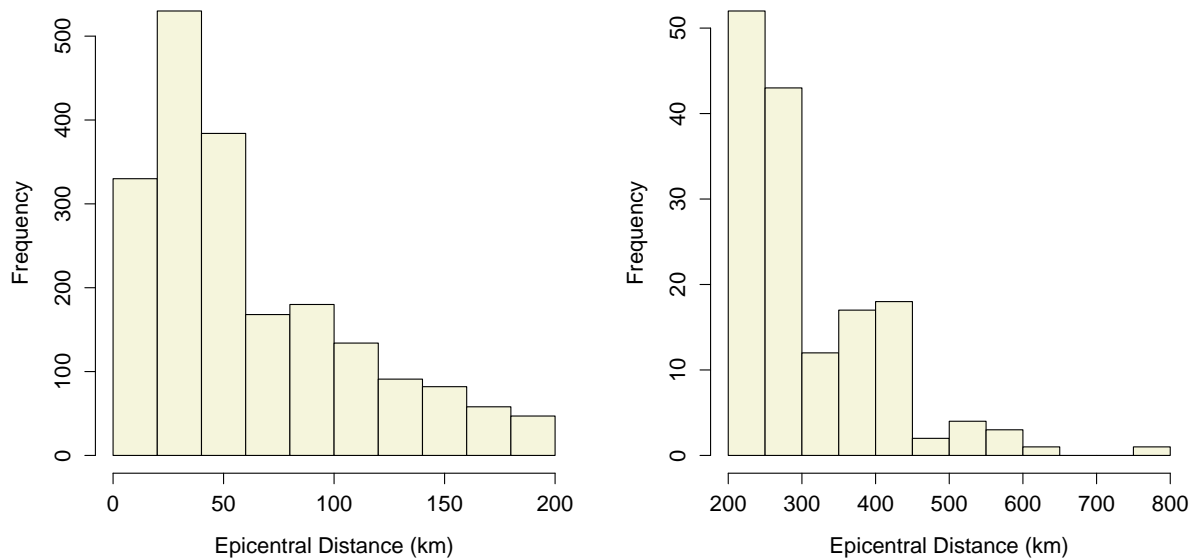


Figure 4.2: Histogram showing the number of observations for the combined data set of all events, over different distances. On the left are distances from 0-200 km, with a scale from 0-500 observations. On the right, distances from 200-800 km, with a scale of 0-50 observations.

Event	# observations	Mean PGV	Mean PGA	Distance Range (km)
Alum Rock	215	4.68	66.65	5-151
Anza 2001	73	1.14	40.69	11-133
Anza 2005	119	2.0	47.00	19-182
Baja	490	6.59	33.40	16-448
Big Bear 1992	132	NA	59.68	11-180
Big Bear 2003	122	1.11	21.59	8-156
Borrego Springs	337	1.44	22.84	18-217.6
Chino Hills	412	4.68	66.65	5-151
Chatsworth	71	1.41	32.61	6 - 58
Inglewood	191	2.52	63.52	2-129
Lake Elsinore	61	0.93	32.47	14 - 67
Milpitas	89	0.58	13.57	3-81
Mount Carmel	145	42.35	9.81	10-953
Ocotillo	156	1.29	20.81	16 - 277
Parkfield	84	16.23	255.67	1-196
Riviere du Loup	77	0.23	10.54	6-985
San Bernardino	172	1.18	31.00	2-124
San Simeon	99	3.91	52.01	13-311
Tohoku	381	13.56	159.00	127-1283
Val des Bois	94	0.09	7.17	47-1167
Whittier Narrows	165	1.44	43.25	1 - 69
Yorba Linda	80	1.18	28.60	4-59
Yucaipa	211	1.45	44.31	3-129

Table 4.2: Summary statistics for instrumental data, including number of observations, mean PGA and PGV, and distance range of observations.

Event	# observations	Match dist.(km)	MMI range	Dist range (km)
All Events	2157	2.62	44.44	2-755
Alum Rock	208	NA	2.6-6.1	4-118
Anza 2001	38	NA	2.0-4.8	16-121
Anza 2005	114	NA	2.0-5.4	19-182
Baja	48	2	2.6-7.2	50-323
Big Bear 1992	37	NA	3.6-7.9	11-168
Big Bear 2003	101	NA	2.0-4.4	17-156
Borrego Springs	319	NA	2.1-5.3	18-218
Chatsworth	70	NA	2.2-4.8	6-58
Chino Hills	237	2	2.5-6.1	5-55
Chatsworth	70	NA	2.2-4.8	6-58
Inglewood	189	NA	2.0-5.4	2-129
Lake Elsinore	60	NA	2.0-4.3	14-67
Milpitas	50	2	2.0-4.3	3-65
Mount Carmel	35	NA	2.0-6.3	10-447
Ocotillo	27	3	2.1-5.0	31-255
Parkfield	19	5	3.1-5.1	20-196
Riviere du Loup	20	3	3.5-5.5	6-531
San Bernardino	67	2	2.0-4.6	4-124
San Simeon	24	4	3.0-7.6	22-311
Tohoku	23	3	3.3-7.9	175-755
Val des Bois	16	2	2.0-4.1	47 - 601
Whittier Narrows	99	2	2.0-4.6	3-57
Yorba Linda	72	NA	2.0-4.6	4-58
Yucaipa	196	NA	2.5-5.5	3-129

Table 4.3: Description of matched DYFI-ground motion data, including number of observations, the maximum distance between station and MMI observation used for matching geocoded or locational data (Match dist.), and the MMI and distance ranges covered.

4.2 Decay of MMI and Ground Motions with Distance

The general behaviour of ground motion and MMI data can be elucidated using simple correlations, plots and regression statistics. The most prominent spatial feature of both ground motion and seismic intensity data is that of attenuation with distance - as the distance from the epicenter increases, the strength of ground shaking and its physical effects are muted. Furthermore, the frequency content of the seismic waves is altered as the earth acts as a low-pass filter, preferentially attenuating the high-frequency content of the signal. This concomitant decay of ground motions and MMI with distance was discussed briefly in section 2.2 and illustrated in Figures 2.1 and 2.2. The following section will extend this theme by providing empirical relationships to quantify the decrease in ground motion intensity and MMI with distance.

Figure 4.3 compares the behaviour of the zip-code aggregated DYFI values with distance for earthquakes of different magnitudes and in different regions. A weighted least-squares linear regression was performed on MMI values averaged over distance ranges, with weights equal to the number of observations in each distance bin; these coefficients are given in Table 4.4. The bottom right figure illustrates how the shape of the MMI-distance curve changes with magnitude for California events. The smaller **M4.7** Inglewood event shows a more gradual slope, while the larger events exhibit steeper slopes. As expected, the larger magnitude event is felt over greater distances, and at any given distance, the average MMI for the smaller event is less than that of larger events.

The top right of Figure 4.3 shows a similar comparison between the **M5.4** Mt.Carmel and **M5.8** Virginia events. Interestingly, the larger magnitude Virginia event exhibits smaller MMI at every distance than Mt. Carmel, despite its larger magnitude. A possible explanation for this result could be the unique local geology of the Mt. Carmel,IL region. It is located near the north end of the Mississippi embayment, a wedge-shaped region underlain by unconsolidated marine sediments [Arsdale, 2009]. A lower-than-average stress drop for the **M5.9** Virginia event could contribute to the observed lower MMI relative to Mount Carmel. However the stress drop for the Mount Carmel event is estimated at 100 bars [Hartzell and Mendoza, 2011], while the stress drop for the Mineral, VA event has been estimated at 300 bars or higher [Cramer et al., 2011]. But as the Mineral, VA event is studied in greater detail, the estimated stress drop may be revised.

It is also of interest to note that while the slopes of different magnitude events in California are different, the slopes of the linear MMI-distance relationships for Mount Carmel and Virginia are

virtually the same (1.5 and 1.6, respectively). The different attenuation properties of Eastern and Western North America may be causing this difference in seismic intensity decay between the two regions.

In the top left of Figure 4.3, the differences in MMI-distance behaviour between eastern and western North America are illustrated. Both the 2010 Borrego Springs and 2008 Mt. Carmel events are **M5.4**, but we can see that at any distance, the Mt. Carmel MMI is reported as higher than that of Borrego Springs. The larger felt area is also obvious in this comparison; the maximum distance of the DYFI felt reports for Borrego Springs is 392 km, while for Mt. Carmel it is 2238 km. Finally, the bottom left figure compares two large magnitude events: Baja and Tohoku. The linear regressions for both of these events extrapolate to undefined MMI values, with the Tohoku intercept being significantly larger than the Baja intercept. The slopes for these two events are commensurate (4.3 for Baja, 4.4 for Tohoku), indicating similar attenuation behaviour.

A similar comparison involving the behaviour of ground motion parameters with distance from the epicenter is shown in Figures 4.4 and 4.5. PGV from **M7.2** Baja, **M5.4** Borrego Springs and **M4.7** Inglewood earthquakes are plotted against distance. The general shape of the attenuation curves are similar, with a vertical shift in the magnitude of PGV. Interestingly, the smaller Inglewood event shows larger PGV than Borrego Springs at close distances, from 20-50 km. PGA versus distance is also plotted for two **M5.4** events, Mount Carmel IL and Borrego Springs CA. Unlike the case for MMI (in which the Borrego Springs event showed smaller MMI than Mount Carmel at all distances), the two sets of PGA values appear to follow a similar attenuation pattern. For the distance ranges in which the two data sets overlap, the Mount Carmel PGA-distance points occur within the cloud of Borrego Springs observations.

A comparison of Figure 4.3 with figures 4.4 and 4.5 illustrate some disparities in the way MMI and ground motion parameters decay with distance from the epicenter. Patterns observed in the MMI-distance plots - like a clear vertical shift with magnitude, and the region and magnitude dependence of slope - are not necessarily mirrored in the ground motion-distance plots. In figure 4.4, for instance, we can see that the behaviour of PGV with distance for the **M 5.4** Borrego Springs and the smaller **M4.7** Inglewood events is almost identical. We also see a clear regional difference in the decay of ground motions with distance when comparing **M5.4** Mount Carmel and Borrego Springs PGV, while the PGA-distance relationships for these events appear much more consistent. These differences in the behaviour of ground motions and MMI will translate into weaker MMI-ground motion correlations and less accurate MMI-ground motion equations, which

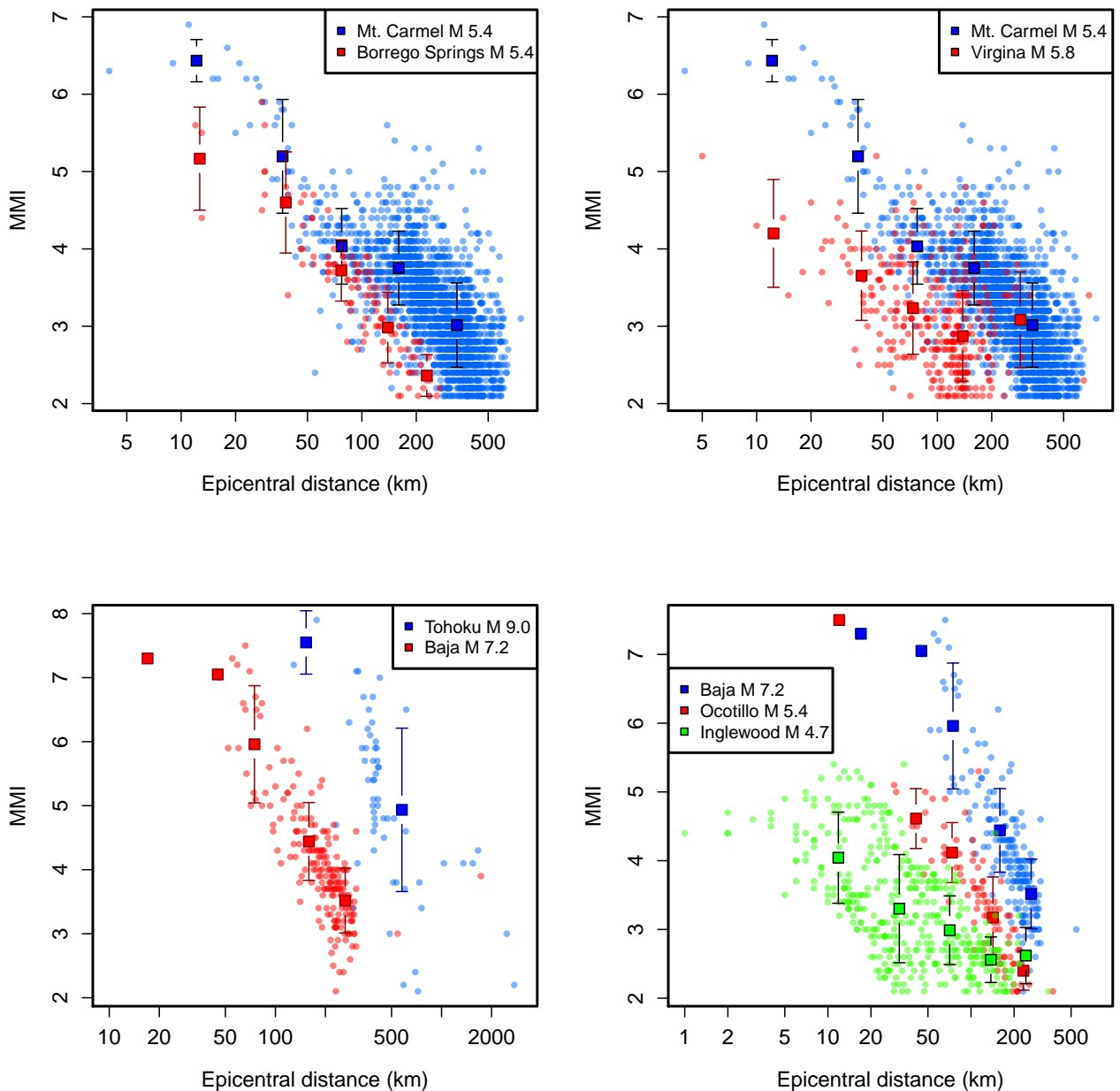


Figure 4.3: Comparing the attenuation of MMI with distance for different earthquakes, based on region and magnitude. Small dots show individual MMI observations, while large squares are the mean MMI over various distances. Error bars show standard deviations associated with these means. From the top left corner clockwise: (1) Mt. Carmel and Borrego Springs, ENA and WNA both M5.4. (2) 2008 M5.4 Mount Carmel and 2011 M5.8 Virginia, both in ENA of different magnitudes. (3) M7.2 Baja, M5.7 Ocotillo and M 4.7 Inglewood - three earthquakes in Western North America of varying magnitudes. (4) 2011 M9.0 Tohoku and 2010 M7.2 Baja, both large magnitude earthquakes.

will be addressed in the following section.

Event	Intercept (c_0)	SE	Slope (c_1)	SE
Baja	13.9*	1.1	-4.3	0.50
Borrego Springs	9.1	0.45	-2.9	0.22
Inglewood	5.4	0.22	-1.5	0.12
Mount Carmel	9.1	0.47	-2.5	0.19
Ocotillo	10.4	0.52	-3.4	0.24
Tohoku	16.8*	0.95	-4.4	0.35
Virginia	6.0	0.42	-1.6	0.20

Table 4.4: Weighted least-squares regression statistics for the MMI-distance relation $\overline{\text{MMI}} = c_0 + c_1 \log(\overline{R})$, where R is epicentral distance. These regressions used binned DYFI data, where the MMI values are averaged over distance bins, and weighted by the number of points in each bin. The intercept and slope are recorded, along with the standard error of the estimated coefficients. ***Note:** MMI is undefined above 12, but the simple linear regression extrapolates to a higher intercept for the M7.2 Baja and M9.0 Tohoku events.

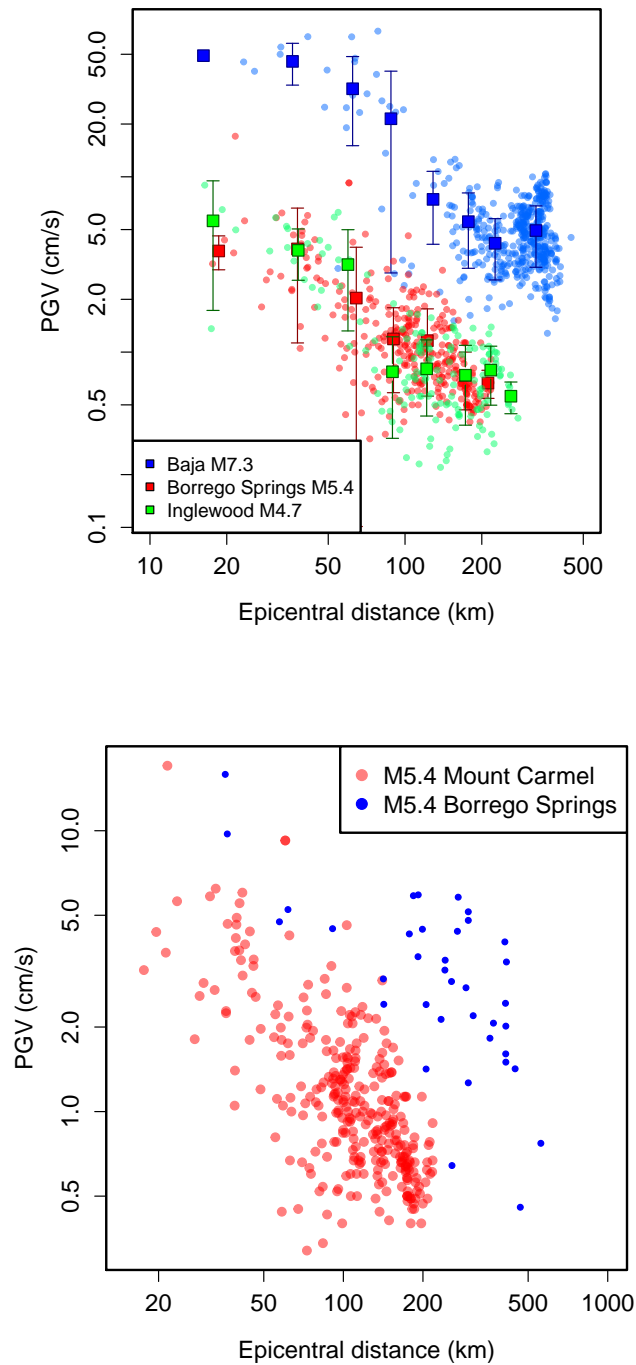


Figure 4.4: Attenuation of PGV with distance from the epicenter. Top: **M7.2** Baja, **M5.4** Borrego Springs and **M4.7** Inglewood earthquakes. Individual observations shown by small circles, averages over distance bins shown by squares with standard deviation bars. Bottom: **M 5.4** Borrego Springs and Mount Carmel events.

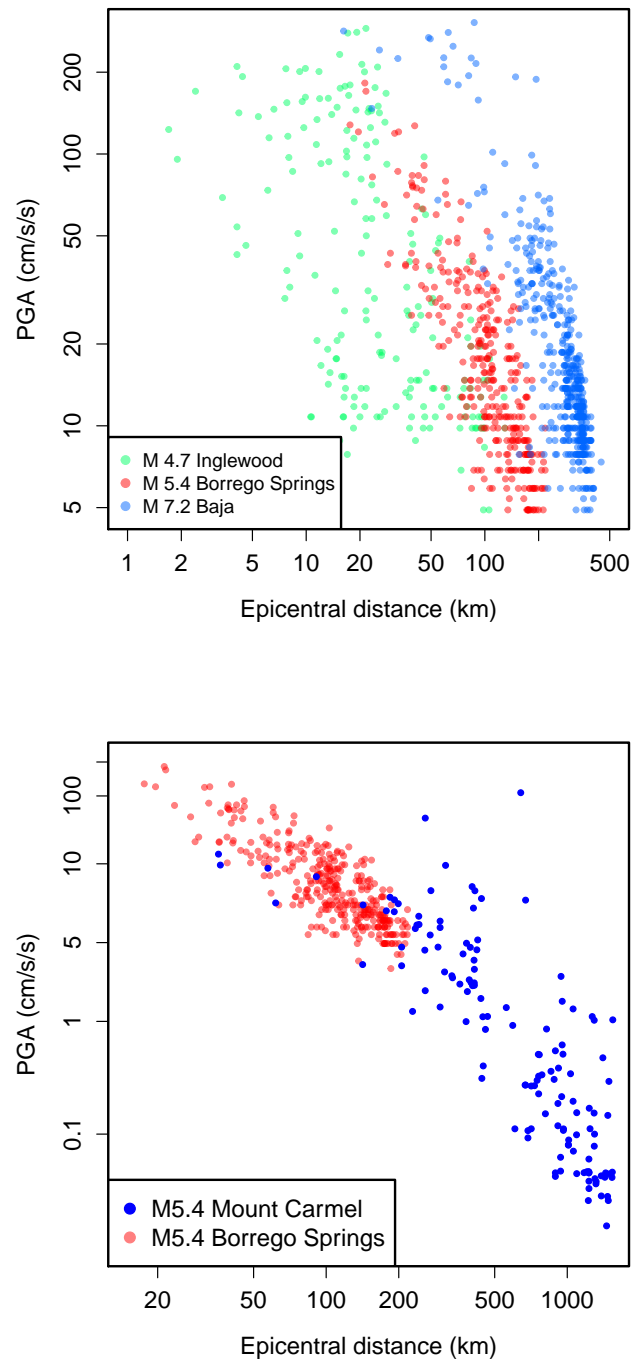


Figure 4.5: Attenuation of PGA with distance. Top: **M**7.2 Baja, **M**5.4 Borrego Springs and **M**4.7 Inglewood earthquakes. Bottom: **M**5.4 Mount Carmel in CEUS compared with same magnitude Borrego Springs in California.

4.3 Correlations between Ground Motion Parameters and MMI

The simplest way to assess correlations between ground motion parameters and MMI is to look at the values of ground motions associated with specific MMI ranges. The degree to which different MMI ranges have overlapping ground motion values will tell us how precisely MMI is pinned down by a ground motion parameter. If the same ground motion parameter value is often associated with multiple MMI values, that parameter may be of questionable value in developing MMI-ground motion relationships. By simply grouping ground motion data by MMI values, we can look at how well the minimum, maximum and mean values of the ground motion parameters track MMI, and get a crude measure of the scatter or variation within the data. For a given MMI range, we can group the ground motion values of all of the matched ground motion-MMI values with MMI within this range, and look at the minimum, maximum and mean values. Table 4.5 gives the minima and maxima of log ground motion for each MMI range, and Table 4.6 shows the mean and standard deviation for these sets.

It is immediately clear that the minimum ground motion parameter values associated with MMI ranges provide virtually no information about the MMI-ground motion relationships. For example, when tracking the minimum ground motions in each MMI range, the log PGV values for MMI from 2.50 to 5.99 show no trend, and seem to vary randomly. The minimum for the highest MMI range is actually associated with the *lowest* minimum log PGV value. PGA shows more variation in minimum values, but they're still inconsistent: the minimum for the 6.0-7.99 MMI range is also the lowest. The maxima follow the increase in MMI better, but like the minima they represent a single observation, and thus are subject to the varagies of individual variation. Nonetheless, the maxima and minima are important for evaluating the error involved in trying to predict MMI from ground motion or vice versa. Violin plots are used to illustrate this error.

A violin plot is similar to a box plot; the median and interquartile range are displayed using a white circle and black rectangle, respectively. Unlike a boxplot though, the violin plot provides a visual representation of the entire distribution. The kernel density function for the set of ground motion parameters pertaining to the different MMI ranges is pasted on both sides of the rectangle. Violin plots of the ground motion parameters associated with these MMI ranges depict differences in the range and median for different MMI ranges (Figure 4.6).

The violin plots illustrate the inconsistent nature of the maxima and minima of the MMI bins.

The medians are reliable and monotonic increasing. Also note differences in the shape of the distributions: at the lower and higher MMI ranges, the distributions are more peaked around the mean, while for intermediate (3.5-4.5) MMI ranges the distributions are flattened. These MMI values are more likely to be associated with a broad range of ground motion parameter values.

It is also of interest to observe the degree of overlap in these distributions. Uncertainty in MMI-ground motion relationships arises from an inability to attribute a single MMI range to a ground motion parameter value. It is valuable to have certain parameter values which we can assign with great confidence to a given MMI range. For example, the line drawn at log PGV equal to 0.5 crosses all of the MMI bins, meaning that an observed value for log PGV of 0.5 could be associated with a wide range of MMI values. In contrast, a line drawn at log PGV equal to 1.5 only crosses the distribution for MMI greater than 5.5. An observed log PGV of 1.5 or greater would allow for greater accuracy in the MMI match estimate than at lower PGV values.

After looking at a direct summary of the ground motion parameter values associated with MMI ranges, it is natural to ask how other variables - those that are known to influence MMI and ground motions - influence these relationships. Given the large amount of variation shown in MMI-ground motion matching, we would like to be able to use other variables to explain some of the scatter.

MMI range	Pairs	log PGV		log PGA		log PSA 0.3		log PSA 1.0		log PSA 3.0	
		min	max	min	max	min	max	min	max	min	max
2.0-2.45	120	-0.68	1.12	-0.31	2.73	-2.3	-0.27	-3	0.47	NA	1.54
2.5-2.99	284	-1.65	1.44	-0.53	2.46	NA	1.96	NA	2.09	NA	2.55
3.0-3.49	482	-1.96	1.8	-0.9	2.76	-2.52	2.01	-3	2.27	NA	1.98
3.5-3.99	555	-1.05	1.61	0.25	2.51	-2.22	1.58	NA	1.51	NA	2.15
4.0-4.49	353	-0.72	1.2	0.2	2.44	-2.1	1.89	-3	1.27	NA	2.15
4.5-4.99	201	-0.47	1.18	0.62	2.23	-1.92	1.53	-2.7	1.81	NA	2.29
5.0-5.99	126	-1.05	1.51	0.17	2.64	-2.15	2.02	-3	2.25	NA	2.08
6.0-7.99	36	-1.94	1.77	-0.59	3.10	-2.12	2.00	-2.70	2.51	NA	2.44

Table 4.5: Number of ground motion observations within MMI ranges from 2.0 to 7.99, and minimum and maximum values of ground motion parameters within these ranges.

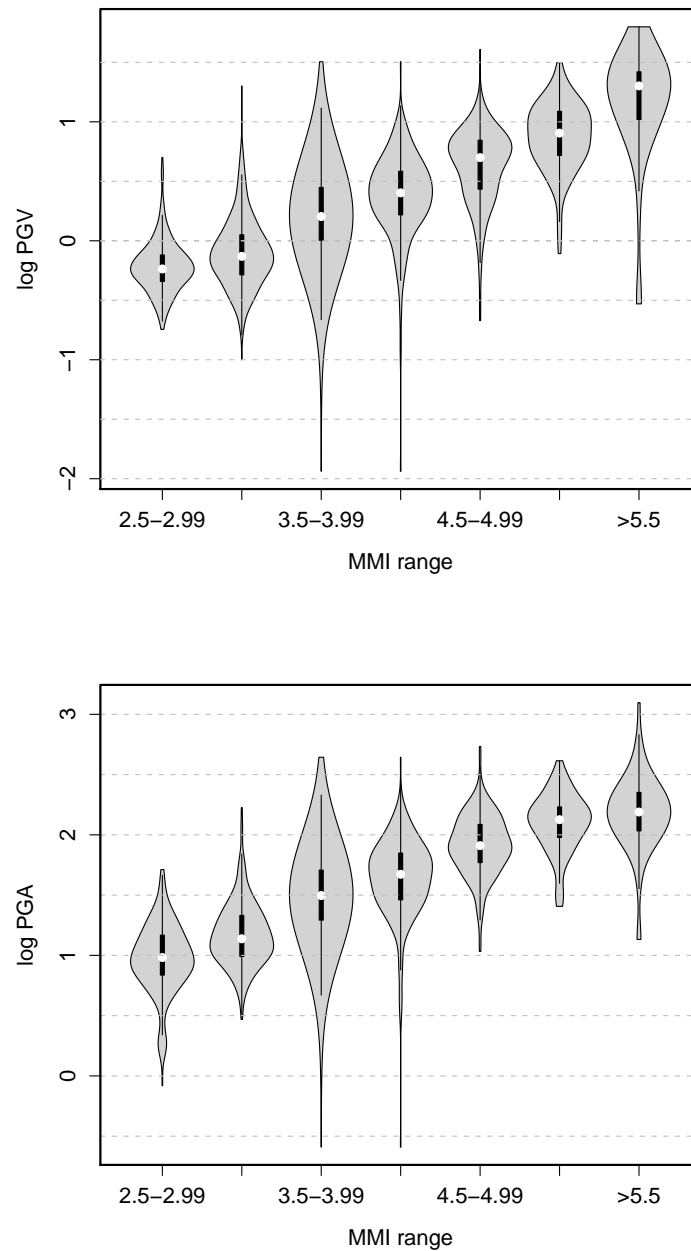


Figure 4.6: Violin plots for $\log \text{PGV}$ (top, cm/s) and $\log \text{PGA}$ (bottom, cm/s^2) associated with different MMI ranges. The black rectangle spans the first to third quartile, with the white circle representing the mean. The kernel density function for the set of ground motion parameter values is pasted on both sides. Note that the group with $\text{MMI} > 5.5$ consists of only 15 observations. All other groups contain at least 75 observations.

MMI range	log PGV	SD	log PGA	SD	log Avg. PSA	SD
2.0-2.5	0.16	0.45	1.44	0.48	1.29	0.44
2.5-3.0	0.13	0.45	0.38	0.46	NA	NA
3.0-3.5	0.11	0.46	1.36	0.49	1.25	0.47
3.5-4.0	0.11	0.43	1.39	0.45	1.26	0.45
4.0-4.5	0.14	0.40	1.40	0.43	1.33	0.38
4.5-5.0	0.20	0.43	1.44	0.42	1.32	0.39
5.0-6.0	0.18	0.52	1.42	0.45	1.30	0.49
6.0-7.9	0.25	0.60	1.50	0.51	1.35	0.54

Table 4.6: Mean and standard deviation for ground motion parameters associated with different MMI ranges. PGV, PGA and average PSA over 0.3, 1.0 and 3.0 seconds.

4.3.1 Distance, Magnitude and V_{S30} Dependence of Correlations

The basic relationships between MMI and ground motion parameters were assessed in the previous section by looking at the ground motions associated with a given MMI range. To understand how other variables may influence these correlations, we can look at how distance, magnitude and V_{S30} influence ground motion parameter values within the fixed MMI ranges used above.

The ground motion values within these ranges show a large degree of scatter - as we saw with the violin plots, the same MMI values may be associated with widely varying ground motion values. By plotting ground motions from fixed MMI ranges against V_{S30} , we'll be able to see whether any of this variation can be explained by site conditions.

Figure 4.7 shows PGV, PGA, and average PSA over 0.3, 1.0 and 3.0 seconds versus the ratio of V_{S30} to the reference site 760 m/s for the MMI ranges 2-3 and 4-4.5. The number of observations in these ranges are 328 and 267, respectively. The plots show a clear decrease in the strength of ground motions with increasing V_{S30} . For a site with relatively lower shear wave velocity, larger ground motions are associated with the same MMI as somewhat smaller ground motions at relatively higher V_{S30} sites. To rule out a correlation between MMI and V_{S30} that could generate a spurious relationship, the average MMI values for each V_{S30} bin were also calculated, and they were effectively the same. The influence of distance and magnitude on correlations between MMI and ground motion was evaluated using straightforward correlation coefficients. The correlation coefficient (r) is a measure of the strength of the linear relationship between two variables. The absolute value of r is 0 for uncorrelated variables, and 1 if there is a perfect linear relationship

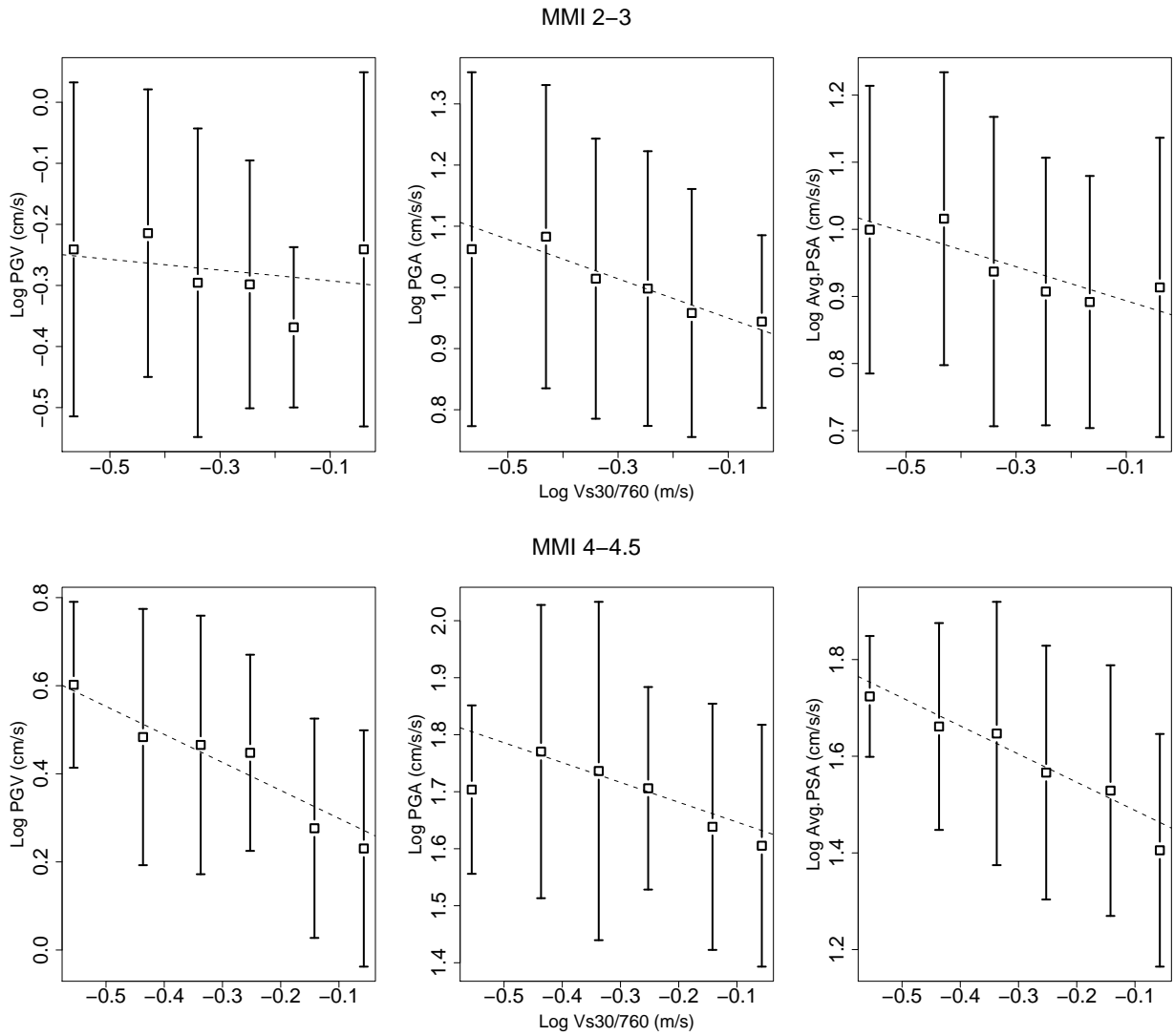


Figure 4.7: Relationship between Log ground motion parameters and Vs30 for fixed MMI ranges. For MMI 2-3 (top) and MMI 4-4.5 (bottom), log PGV, log PGA and log Average PSA is plotted against log of Vs30/760 m/s. We can see a clear trend of decreasing ground motion with increasing Vs30. This implies that lower Vs30 sites are associated with stronger ground motions for a given MMI level.

between them. The empirical correlation coefficient is

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2(y_i - \bar{y})^2}} \quad (4.1)$$

[Dalgaard, 2008] The data were divided into different sets based on magnitude (less than 5.0, 5.0-6.0 and greater than 6.0) and distance (distance bins representing near, intermediate, and far), and the correlation coefficient between MMI and log PGV, log PGA and log average PSA was calculated for each of these sets, and compiled in Table 4.7. Graphical representations of this data for magnitudes less than 5.0, and between 5.5 and 6.0, are shown in Figure 4.8. There are some predictable and some unexpected trends in the correlation coefficients. Predictably, larger magnitude events show better correlation at greater distances from the epicenter. This result is simply a corollary of the fact that higher MMI values are better correlated with ground motion parameters. At greater distances, larger magnitude events retain higher MMI. At distances from 100-120 km, correlations are the lowest (and often insignificant), with a curious increase at distances over 120 km for PGA and average PSA, but not PGV. This breakdown in correlations could be attributed to the interactions between seismic waves arriving directly from the source and those reflected off the Mohorovicic discontinuity.

While correlation coefficients give as a good sense for how strong the relationships between MMI and ground motion parameters are, linear regressions are necessary to derive predictive equations for the purpose of converting between the two. The following section estimates linear regression coefficients for several different MMI-ground motion equations, using the auxiliary variables of magnitude, distance and V_{S30} .

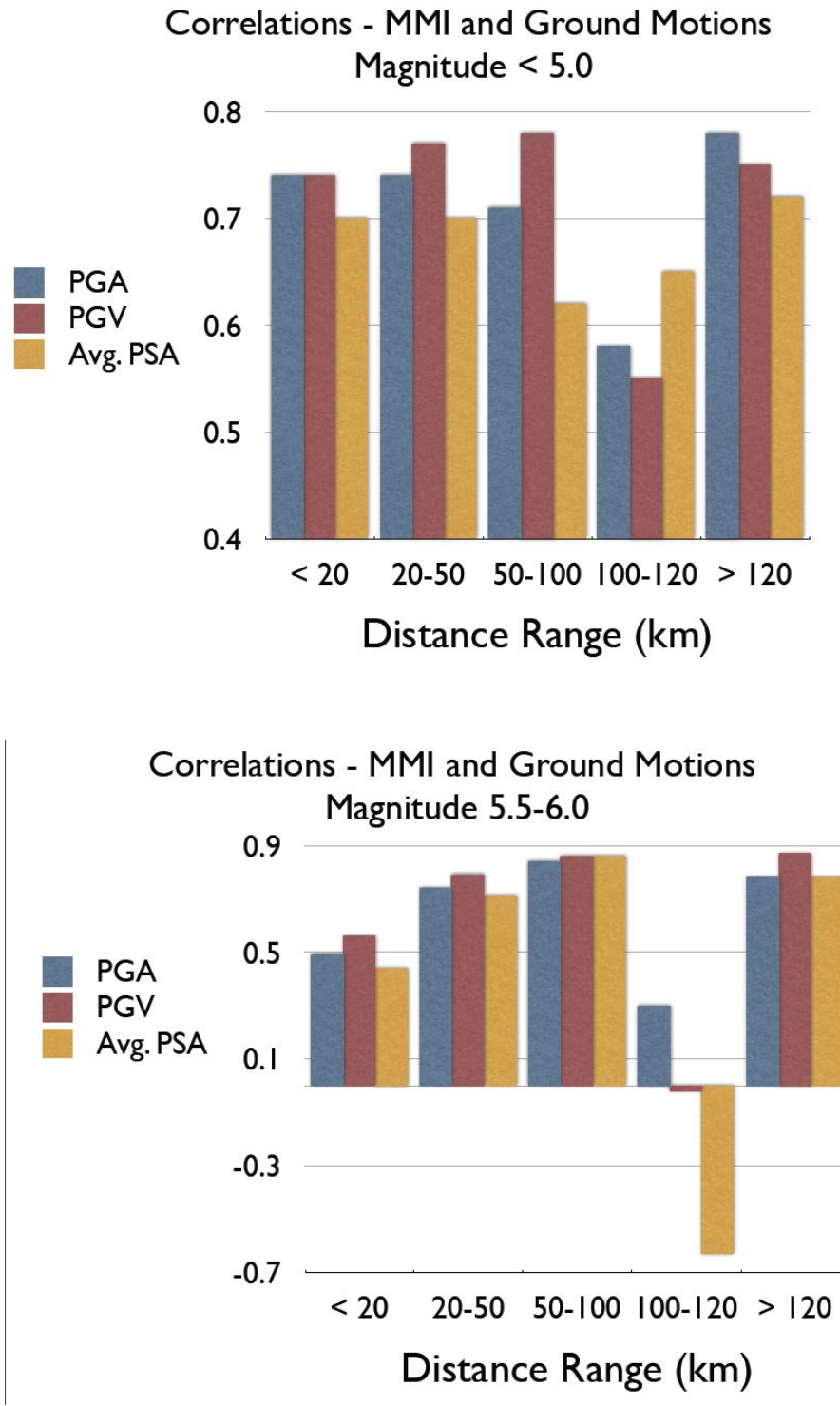


Figure 4.8: Graphical representation of the correlation coefficients presented in Table 4.7.

Magnitude	Distance	Log PGV	Log PGA	Log Avg.PSA
Less than 5	< 20	0.74	0.74	0.70
	20-50	0.74	0.77	0.70
	50-100	0.71	0.78	0.62
	100-120	0.58	0.55	0.65
	> 120	0.78	0.75	0.72
5 to 5.5	< 20	0.64	0.62	0.33
	20-50	0.65	0.68	0.56
	50-100	0.52	0.52	0.27
	100-120*	0.23	0.33	0.24
	> 120	-0.07	0.32	NA
5.5 to 6	< 20	0.49	0.56	0.44
	20-50	0.74	0.79	0.71
	50-100	0.86	0.84	0.86
	100-120*	0.13	-0.02	-0.63
	> 120	0.78	0.87	0.78

Table 4.7: Correlation coefficients for MMI and ground motion parameters, for different distance and magnitude ranges. Starred rows indicate results were not statistically significant with p-values greater than 0.05.

4.4 Linear Regressions

One of the main goals of investigating correlations between MMI and ground motions is to find predictive relationships between the two measures. As discussed in the introduction, the simplest relationship is the linear model relating MMI and ground motion:

$$\text{MMI} = c_0 + c_1 \log(Y) \quad (4.2)$$

where Y may represent PGA, PGV, PSA at 0.3, 1.0 or 3.0 seconds, or the average of PSA over these three frequencies.

The coefficients for equation (4.2) are estimated using weighted least squares. The data used in these regressions are the matched MMI and ground motion data from all events listed in Table 2.1, including the M9.0 Tohoku earthquake. These matched data were found using geocoded DYFI observations where available, otherwise they used zip code DYFI observations.

The regressions were performed by first averaging ground motions and MMI values over MMI ranges, from 2.0 up to 5.5. For example, one data point consists of the mean ground motion and MMI of all matched data with an MMI value from 3.0 to 3.49. Above MMI 5.5, the data are sparse, and only one or two binned averages were possible. As a change in the slope of MMI-ground motion relations has been observed around MMI 5.0, to incorporate higher MMI with a piecewise linear form would be to estimate a linear segment with only two or three points. For the following analysis, the data are truncated at MMI 5.5, with the understanding that a second linear segment with different slope exists for $\text{MMI} > 5.0$.

The regression is performed using these average values weighted by the number of observations making up each average. The MMI ranges used were 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, and 4.5-5.5. This averaging procedure allows for a more robust estimate that is less influenced by the large number of observations for low to medium MMI, and is used in previous studies of MMI-ground motion relationship [Kaka and Atkinson, 2004, Atkinson and Kaka, 2007]. The estimated coefficients from these regressions are presented in Table 4.8. Plots showing the individual MMI-ground motion observations distinguished by epicentral distance and region are given in Figures 4.9 and 4.10.

Ground Motion Parameter	Intercept (c_0)	Log Y (c_1)
PGV	3.38 (0.04)	2.25 (0.10)
PGA	0.52 (0.22)	2.28 (0.15)
PSA 0.3 sec	5.92 (0.44)	1.96 (0.36)
PSA 1.0 sec	6.15 (0.64)	1.45 (0.37)
PSA 3.0 sec	5.36 (1.22)	0.74 (0.54)
Avg. PSA	0.53 (0.25)	2.46 (0.18)

Table 4.8: Coefficients from estimation of equation (4.2), all data. Standard error given in brackets.

4.4.1 Residuals

While equation (4.2) does a fair job of describing the relationship between MMI and ground motion parameters, it is well understood that other variables have significant predictive influence on the MMI-ground motion correlations. Distance and magnitude are routinely included in these relationships, usually by way of a secondary linear regression using residuals from equation (4.2), e.g. Atkinson and Kaka [2007]. Figures 4.9 and 4.10 plot the matched ground motion-MMI data from North America, with different symbols illustrating differences in distance ranges and regions.

The first set of plots distinguishing distance reveal that many of the outlying points farthest from the regression line are associated with distances greater than 100km. This illustrates what was found using MMI-ground motion correlation coefficients broken down by regions - at greater distances, the correlation becomes weaker. The breakdown in correlation around 100-120 km from the epicenter appears here in the residuals. These outlying points could be an artifact of the Moho bounce, where seismic waves reflecting from the Mohorovicic discontinuity (the boundary between Earth's crust and mantle) interact with waves arriving directly from the source.

The second set of plots shows the departure of the ENA data from the California data. In particular, a group of points in the lower right quadrant of the plot represent observations from the M5.4 Mount Carmel, IL earthquake, while those ENA points above the regression line correspond to the Canadian data. These differences may be reflecting different site conditions, which are leading to a difference in the frequency content of ground motions in the central U.S. compared to southeastern Canada.

Figures 4.12 and 4.13 plot residuals from the MMI-ground motion regressions against distance and magnitude to further visualize the trends. At distances less than 200 km, the residuals from

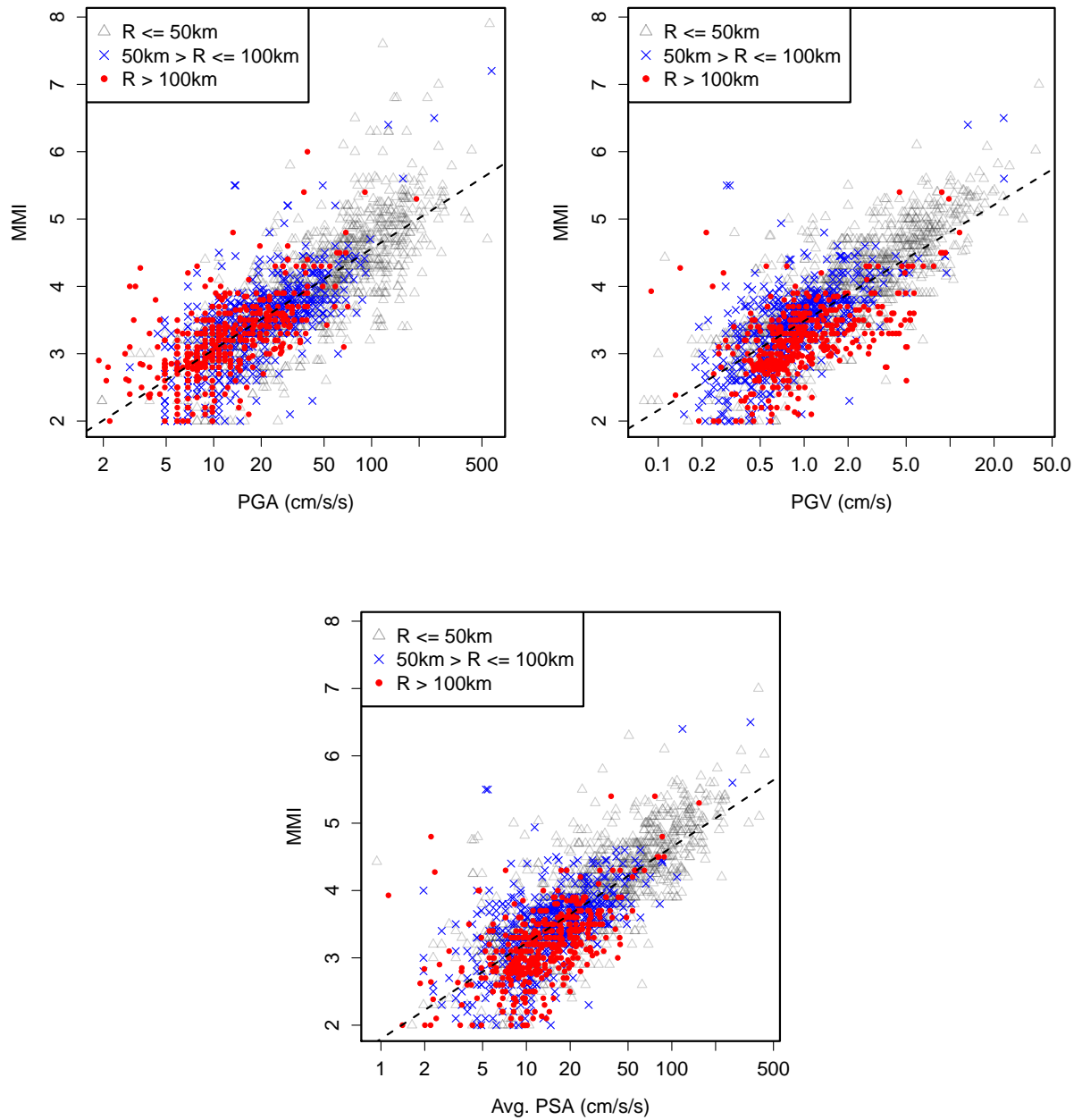


Figure 4.9: MMI-ground motion plots with symbols distinguishing different epicentral distance ranges (R).

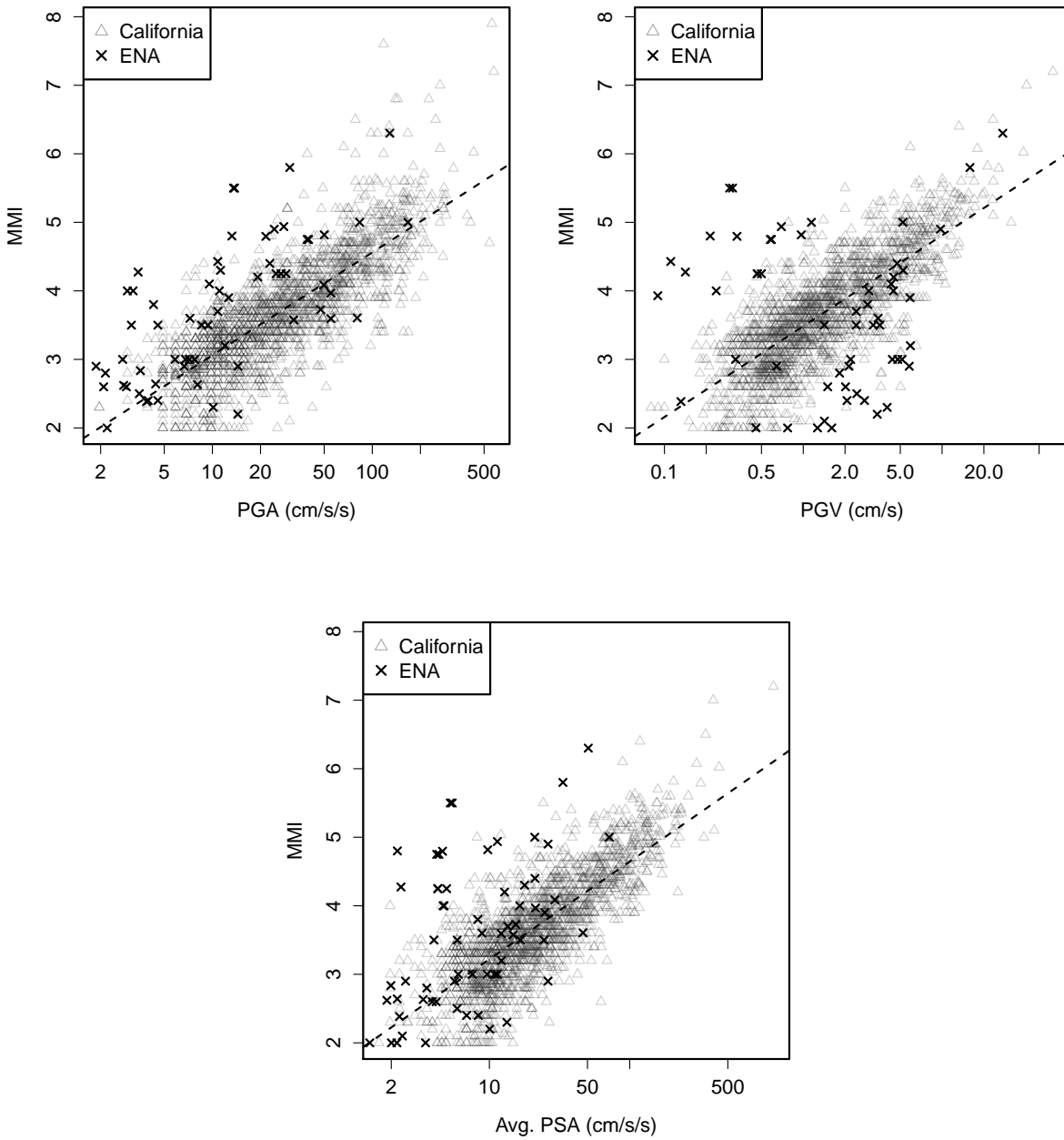


Figure 4.10: MMI-ground motion plots with symbols distinguishing eastern and western North America.

all ground motion parameters show a decreasing trend with increasing distance from the epicenter. This trend is most pronounced for PGV, and for distances less than 100 km. A slight increase of residuals around 200 km occurs for every parameter except PGV, and most notably for PSA at 3.0 seconds. Figure 4.11 again plots residuals from the MMI-ground motion regressions against epicentral distance, with different symbols distinguishing ENA and California.

The pattern of magnitude dependence is less clear. Residuals are higher for smaller magnitude events for PGV, and PSA at 1.0 and 3.0 seconds. For PGA and PSA at 0.3 second, residuals are slightly larger for magnitudes greater than 5.5 than for those below 5.5. No magnitude trend is observed for the residuals from the PSA average.

In addition to magnitude and distance dependence, we can investigate whether the residuals from equation (4.2) exhibit V_{S30} dependence as well. The variation in residuals from equation (4.2) explained by V_{S30} is less pronounced. To determine whether the residuals were influenced by V_{S30} , they were separated into two sets based on whether the associated V_{S30} value was above or below 360 m/s. The density functions of the residuals from each set are plotted in Figure 4.14. For PGV, residuals for the relatively higher V_{S30} values are marginally skewed toward the positive side, with little apparent differences in magnitude. The relatively higher residuals for PGA and average PSA, however, show a larger peak in residuals around zero, and accompanying skinnier tails for outlying residual values. The difference is small, but indicates slightly less precision in the MMI-ground motion relationship for stations on soft soil sites.

To incorporate distance, magnitude and site-dependence into equation (4.2), we perform three additional regressions using the residuals.

$$\epsilon_{\text{MMI}} = c_0 + c_1 \log(R) + c_2 \mathbf{M} \quad (4.3)$$

$$\epsilon_{\text{MMI}} = c_0 + c_1 \log(R) + c_2 \mathbf{M} + c_3 V_{S30}^{\text{ref}} \quad (4.4)$$

$$\epsilon_{\text{MMI}} = c_0 + c_1 \log(R) + c_2 \mathbf{M} + c_3 D_{VS30} \quad (4.5)$$

where ϵ_{MMI} are the residuals from equation (4.2), R is epicentral distance, \mathbf{M} is magnitude, V_{S30}^{ref} is $\frac{V_{S30}}{760\text{m/s}}$, and D_{VS30} is dummy variable with value 1 if V_{S30} is above 360 m/s, and 0 if it is below 360 m/s.

The estimated coefficients from the above equations are listed in Table 4.9. They are consistent across ground motion parameters and equations; the distance effect is negative, the magnitude

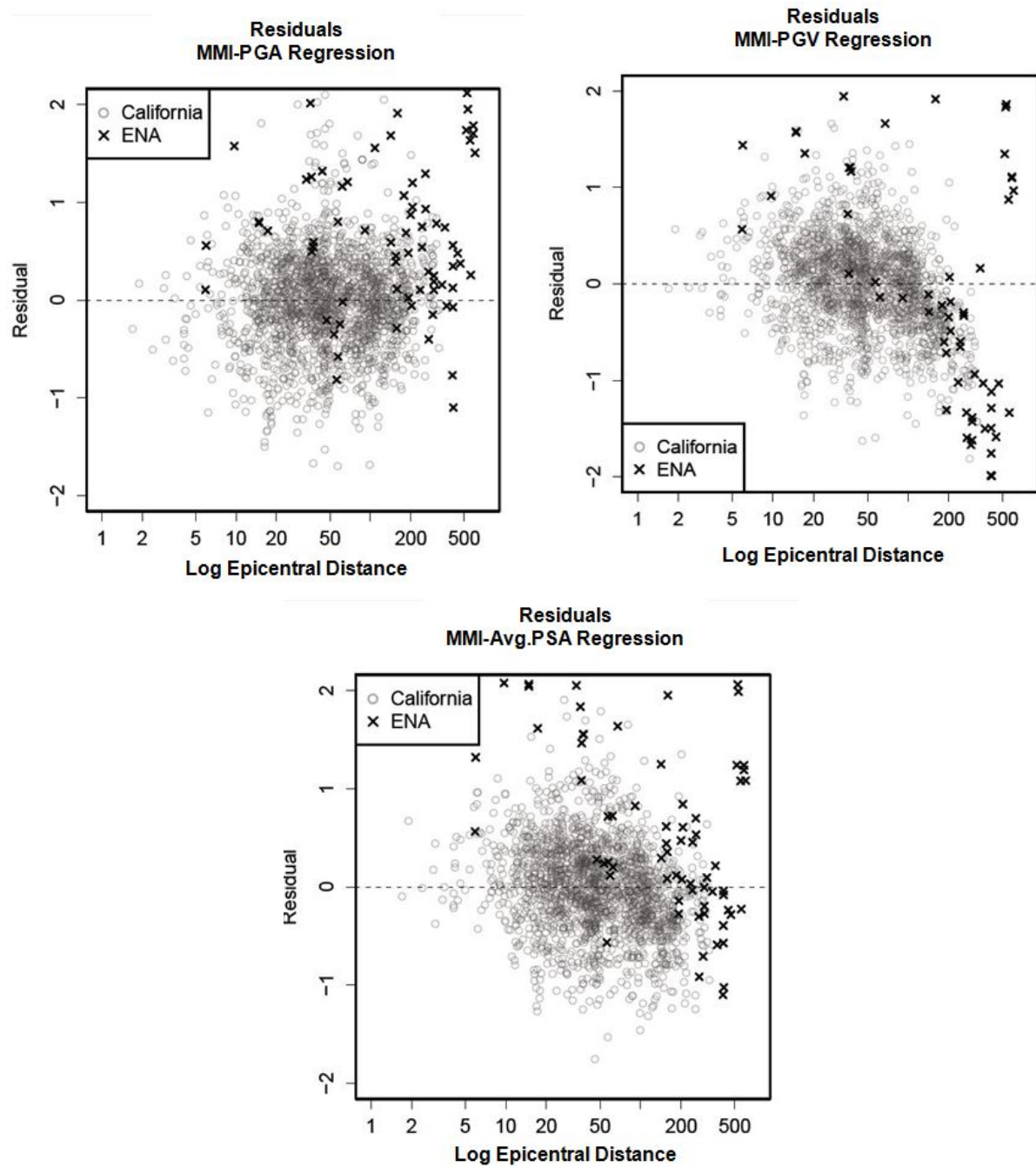


Figure 4.11: Residuals from MMI-ground motion regressions with symbols distinguishing eastern and western North America.

effect is positive, and the V_{S30} effect is negative. Differences in the magnitude and log distance coefficients among the three equations are slight; the inclusion of the V_{S30} term has only a small effect on the original regression (4.3). For PGV and PGA, the ratio of V_{S30} to the reference site 760 m/s was significant at the 0.05 level, while for average PSA, the reference V_{S30} was not significant, but the dummy variable was significant at the 0.10 level.

As a reference, the coefficients for magnitude and distance from equations (4.3) and (4.4) are presented beside the same estimates from Atkinson and Kaka [2007] (AK07) and Worden et al. [2011] in Table 4.10. Note that the Worden et al. [2011] probabilistic analysis results use derived MMI cutoff points for their bilinear functional forms, which differ from the more arbitrary cutoff of MMI equal to 5 used here. Their coefficients are found using data for which MMI is less than 4.22 for PGA, and MMI less than 4.56 for PGV. The following differences are observed, broken down by ground motion parameter (PGA or PGV) and coefficient (intercept, magnitude slope and log distance slope):

1. **Intercept for PGV:** For all three sets of results, the PGV intercepts are all positive, and of roughly comparable magnitude.
2. **Magnitude Slope for PGV:** AK07 and Worden are both negative and nearly identical. This study finds a small positive effect, for both regressions (with and without site effect term).
3. **Log Distance Slope for PGV:** AK07 finds a small positive effect, this study finds a negative effect, and Worden finds zero effect.
4. **Intercept for PGA:** For all three sets of results, the PGA intercepts are all negative, and of comparable magnitude.
5. **Magnitude Slope for PGV:** This study reveals the strongest positive effect, AK07 finds a weak positive effect, and Worden finds a small negative effect.
6. **Log Distance Slope for PGA:** AK07 and Worden results are similar, with a strong positive effect. This study finds a small negative effect.

Figure 4.15 contains a plot of MMI vs PGV for all events, with binned averages marked, with the AK07 and Worden et al. [2011] MMI-PGV bilinear relationships overlaid. The slope clearly matches that of the binned data, but the intercept of AK07 relationship is shifted, showing a systematic difference in the data used. As expected from the similar slopes, residuals from both the

linear relationship obtained from the binned data, and the bilinear relationship from AK07, they show similar shape and behaviour with distance. The AK07 residuals are simply shifted downwards, reflecting the upward translation of the AK07 MMI-PGV curve relative to the relationship found in this study.

Regression	PGV			PGA			Avg.PSA		
	(4.3)	(4.4)	(4.5)	(4.3)	(4.4)	(4.5)	(4.3)	(4.4)	(4.5)
Intercept	0.17	0.21	0.26	-1.46	-1.54	-1.57	-1.6	-1.72	-1.72
SE	0.07	0.08	0.08	0.06	0.08	0.08	0.06	0.08	0.08
Log Dist	-0.44	-0.43	-0.41	-0.16	-0.17	-0.18	-0.19	-0.18	-0.18
SE	0.03	0.03	0.03	0.02	0.03	0.03	0.02	0.03	0.03
Magnitude	0.1	0.07	0.07	0.33	0.36	0.36	0.35	0.37	0.37
SE	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.02
Vs30 Ref.		0.11			-0.09			-0.03*	
SE		0.06			0.04			0.05	
Vs30 Dummy			-0.01*			-0.03*			-0.03**
SE			0.02			0.02			0.02

Table 4.9: Coefficient estimates from regressions (4.3), (4.4) and (4.5), with standard errors. (*) indicates coefficients are not significant at 0.1 level, and (**) indicates coefficient is significant only at 0.1 level.

Source	Intercept	Mag. Slope	Log Dist Slope
	PGV		
AK07	0.47	-0.19	0.26
Worden	0.9	-0.18	0
Eq.32 (no site effect term)	0.17	0.1	-0.44
Eq. 33 (site effect term)	0.21	0.07	-0.43
	PGA		
AK07	-1.96	0.02	0.98
Worden	-0.91	-0.17	1.02
Eq.32	-1.46	0.33	-0.16
Eq. 33	-1.54	0.36	-0.17

Table 4.10: Comparison of coefficient estimates from equations (4.3) and (4.4), and those from Atkinson and Kaka [2007] and Worden et al. [2011].

4.5 Region Dependence of Correlations

Given the inherent differences in the behaviour of ground motions between eastern and western North America, it is natural to ask whether the strength of correlation between MMI and ground motion parameters is significantly different between the two regions, and whether the optimal choice of parameter is different. In eastern North America, earthquakes tend to generate high PGA values, but these are associated with short duration and high frequencies. This makes the damage

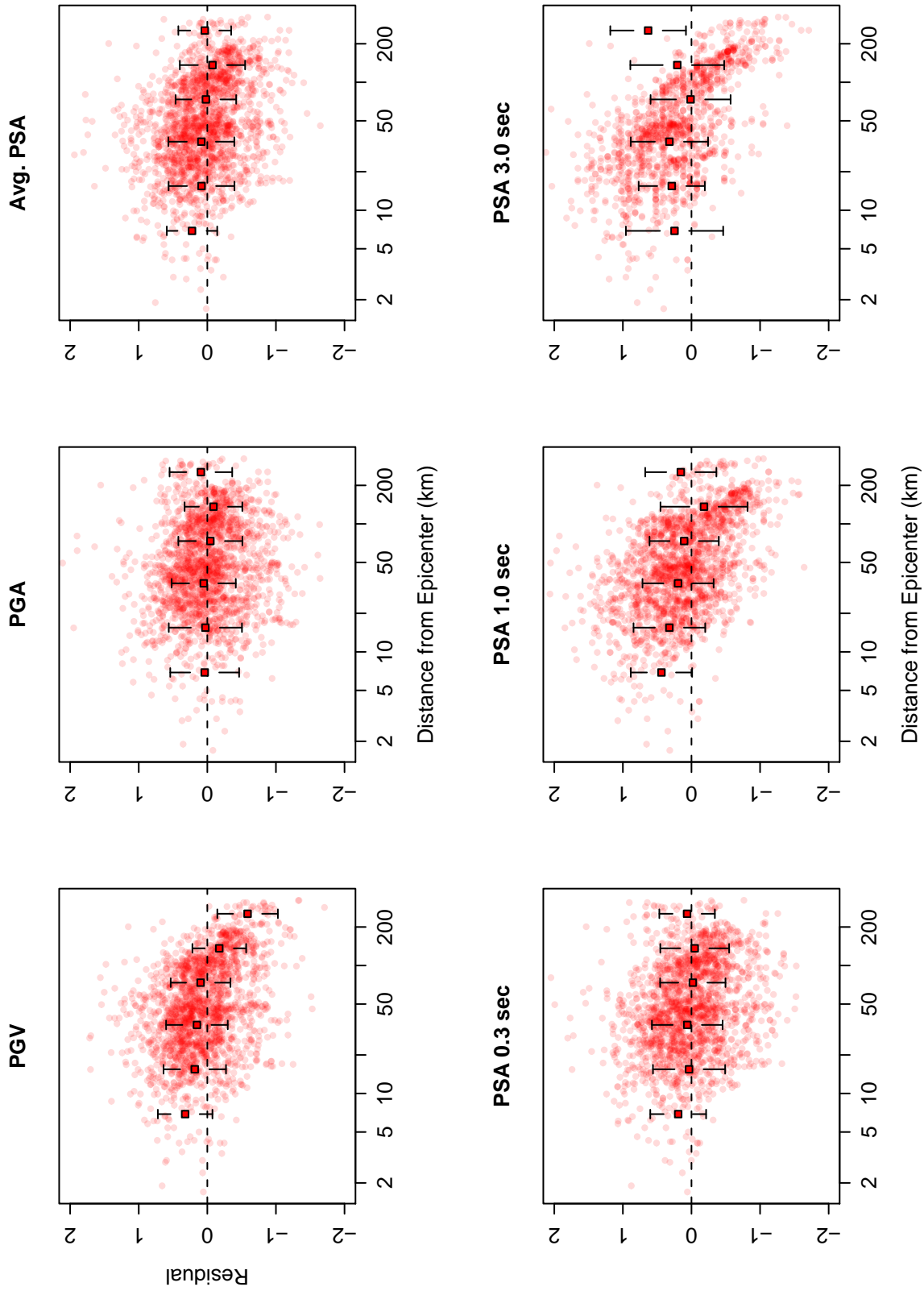


Figure 4.12: Residuals from equation (4.2), plotted against epicentral distance for the six ground motion parameters. Points represent individual residuals; square symbols are averages over distance bins with standard deviation bars.

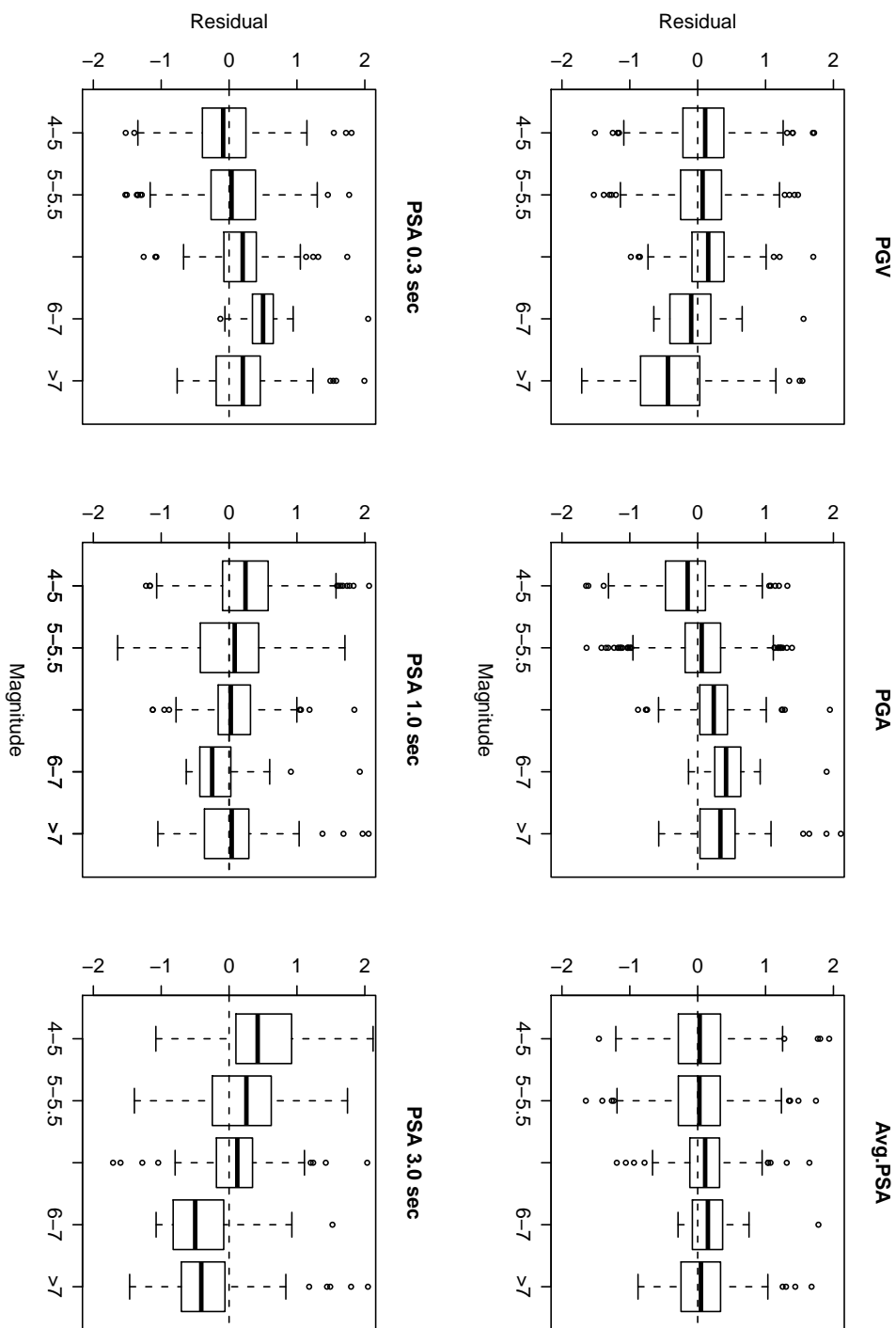


Figure 4.13: Boxplot of residuals from equation (4.2) based on magnitude. Thick black line indicates mean of the magnitude bin, the lower and upper edges of the box represent the first and third quartiles, and standard errors are represented by bars.

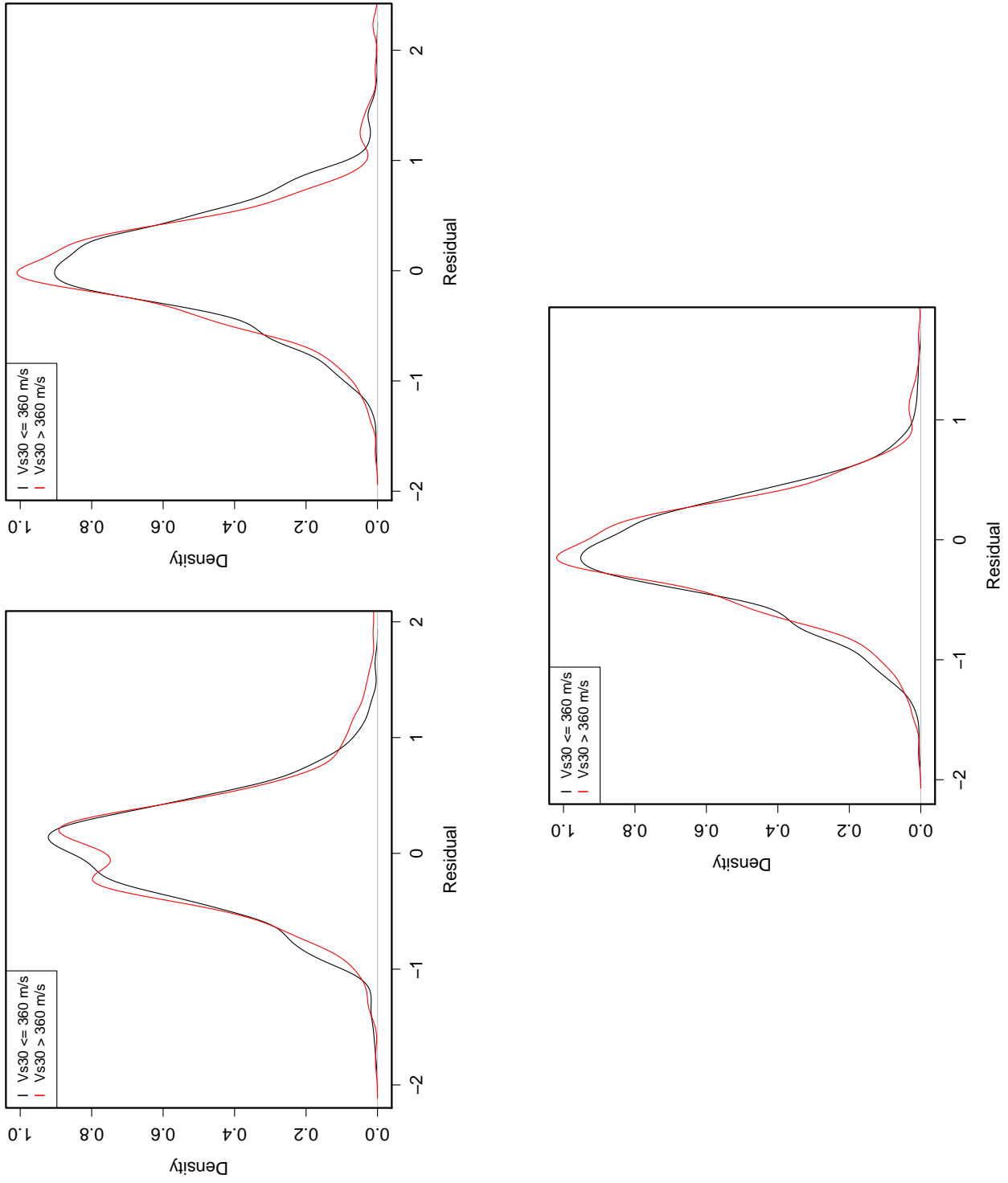


Figure 4.14: Density function of residuals from MMI-Ground Motion regression (4.2), separated by V_{s30} values above and below 360 m/s.

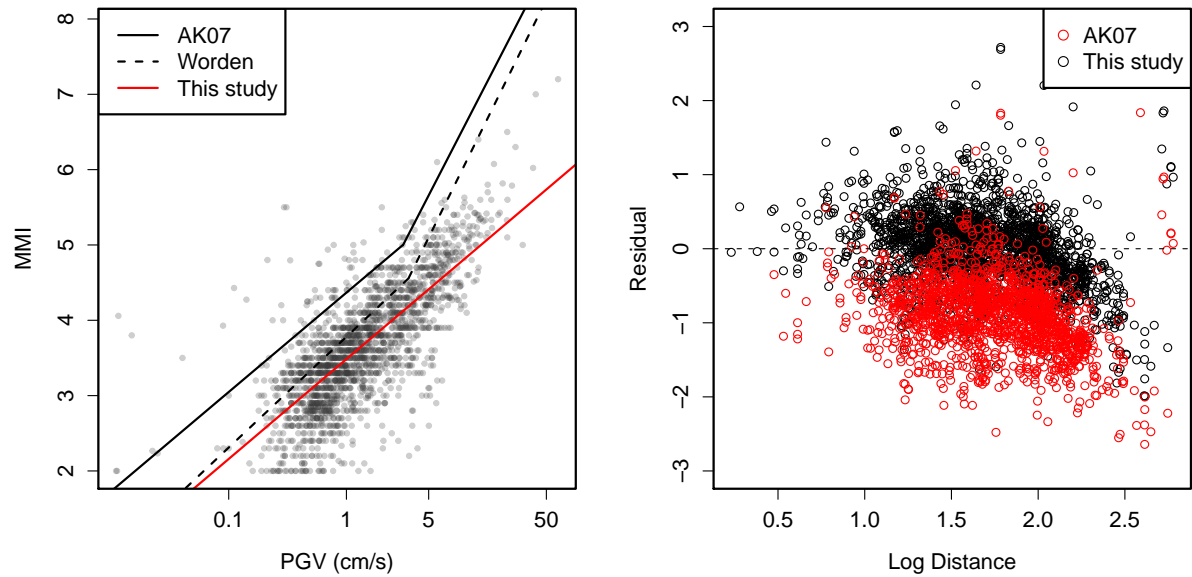


Figure 4.15: Top: MMI-PGV bilinear relationship from AK07 and Worden et al. [2011] overlaid on data from this study, showing systematic differences in data used. The outlying points from ENA data appear to force the regression line down below AK07 and Worden et al. [2011]. Bottom: Comparison of residuals from Atkinson and Kaka [2007] MMI-PGV curve and residuals from the simple MMI-PGV relationship found in this study, plotted against epicentral distance. The overall shape of the residuals, and their behaviour with distance, is similar for both relationships.

associated with these PGA values potentially far less severe than the same value of PGA, but at lower frequencies and for longer duration [Leblanc and Klimkiewicz, 1994]. Thus we may expect PGA in eastern North America to be less well-correlated with MMI. MMI values above 5.0 are correlated with structural damage, and structural damage may be less well correlated with the high PGA values in ENA.

The M5.4 Borrego Springs event possesses abundant matched MMI-ground motion observations, and is conveniently the same magnitude as the Mount Carmel event in the central/eastern United States. This affords a good opportunity to assess the strength of correlation between PGA and MMI for two same-sized events in different continental regions.

Figure 4.16 is a plot of MMI vs PGA for Mount Carmel and Borrego Springs; the coefficients for the linear regression are given in Table 4.11. While there are many more observations for Borrego Springs, it is still clear that the MMI-PGA trend for Mount Carmel lies above the same trend for Borrego Springs. Put another way, for any give level of PGA, the corresponding MMI is higher for Mount Carmel than for Borrego Springs. This runs counter to the idea that PGA in ENA may over predict damage. However the distance ranges of these observations differ greatly, with distances from 10-447 km for Mount Carmel and 18-218 km for Borrego Springs. To adjust for these differences, a second linear regression was performed which included a distance component; coefficients for this regression are included in Table 4.11 as well. Note that the coefficients for log distance are strongly negative and differential, with a greater distance impact for Borrego Springs.

Using these coefficients, the data were detrended for distance, taking the actual PGA values and subtracting $c_2 \times \log(R)$. The resulting data are shown in the right hand side of Figure 4.16. While the Mount Carmel MMI-PGA data appear to be more in concert with the Borrego Springs data now, it is still not clear that they represent the same MMI-PGA relationship. As a check, a third regression was performed using a dummy variable that assumes a value of 1 for the Mount Carmel data and 0 for the Borrego Springs data (D_{MC}). The results, in Table 4.11, show a significant positive coefficient for this dummy variable, indicating that for the same PGA values, the corresponding MMI for Mount Carmel is on average 0.6 units greater than that for Borrego Springs.

A final demonstration of the difference in MMI-PGA relationship between these two events was made using a pooled regression. Average MMI values over PGA bins were used, and the resulting residuals were separated by event. Histograms of these residuals are shown in Figure 4.17, with Mount Carmel residuals on the left and Borrego Springs residuals on the right. The Mount

Carmel residuals are clearly skewed toward negative values, while the Borrego Springs residuals are skewed to the positive side, suggesting that the MMI-PGA relationships for these two events are distinct.

To summarize, the four different regressions performed in the above analysis are listed below.

$$\text{MMI} = c_0 + c_1 \log \text{PGA} \quad (4.6)$$

$$\text{MMI} = c_0 + c_1 \log(\text{PGA}) + c_2 \log(R) \quad (4.7)$$

$$\overline{\text{MMI}} = c_0 + c_1 \log(\overline{\text{PGA}}) \quad (4.8)$$

$$\text{MMI} = c_0 + c_1 \log(\text{PGA}) + c_2 \log(R) + c_3 D_{MC} \quad (4.9)$$

where R represents epicentral distance, $\overline{\text{PGA}}$ and $\overline{\text{MMI}}$ are mean values averaged over binned PGA values, and D_{MC} is the dummy variable described above.

Event	Eqn.	c_0	SE	c_1	SE	c_2	SE	c_3	SE
Mount Carmel	(4.6)	1.74	0.16	1.79	0.13	NA	NA	NA	NA
	(4.7)	5.50	0.54	0.90	0.15	-1.27	0.18	NA	NA
Borrego Springs	(4.6)	1.70	0.06	1.42	0.05	NA	NA	NA	NA
	(4.7)	8.27	0.25	0.31	0.05	-2.56	0.10	NA	NA
Both	(4.8)	1.63	0.11	1.53	0.09	NA	NA	NA	NA
Both	(4.9)	7.87	0.24	0.37	0.05	-2.40	0.09	0.64	0.07

Table 4.11: Estimates and standard errors for coefficients of the linear regressions (4.6), (4.7), (4.8) and (4.9), M5.4 events Mount Carmel and Borrego Springs. Estimates and standard errors are given.

The results shown here are counterintuitive: if the high PGA values in ENA are associated with high frequencies, which in turn are supposed to be less well correlated with damage, then we should expect that for a given value of PGA, the matched MMI for Mount Carmel should be *lower* than for Borrego Springs. In fact, we see the opposite. This observation may be caused by site effects, which were not considered in this analysis. Mount Carmel is situated in the Mississippi embayment, and the deep sediments therein may obfuscate the expected results. To further investigate this result, the MMI-PSA relationships at two different frequencies are plotted in Figure 4.18. The lower frequency is 0.3 seconds for Borrego Springs and 0.2 seconds for Mount Carmel. It is immediately apparent that the distinction between the two frequencies is much greater for Mount Carmel. The Borrego Springs data shows a fair amount of intermingling between the two frequen-

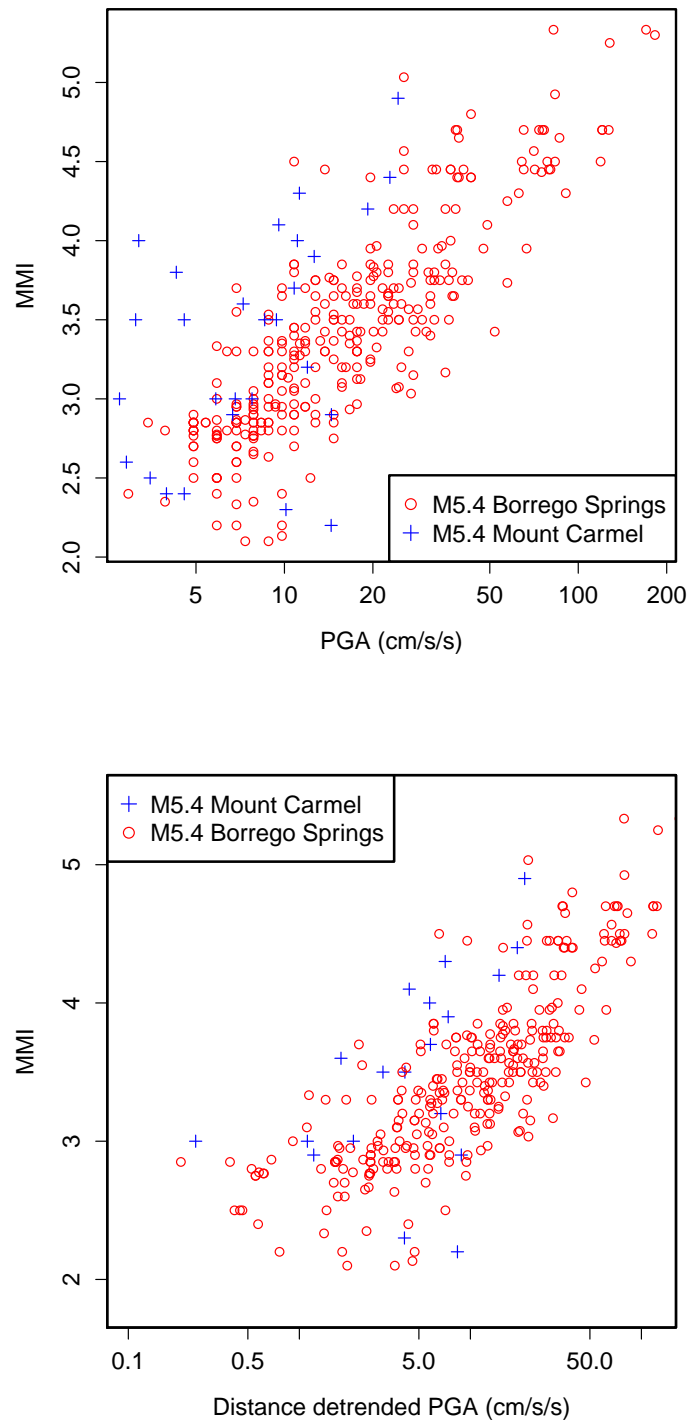


Figure 4.16: MMI vs PGA for M5.4 Mount Carmel and Borrego Springs events. Top: the raw MMI and PGA data. Bottom: raw data are detrended using the distance component c_2 from equation (4.8).

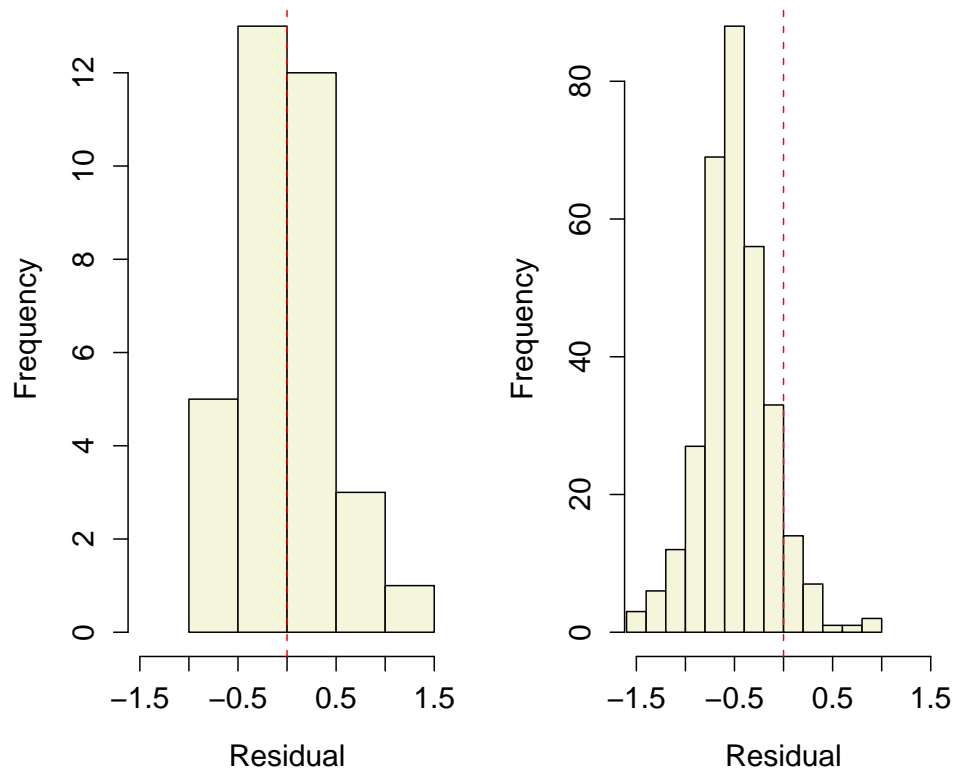


Figure 4.17: Histograms of residuals from regression (4.5), in which pooled Mount Carmel and Borrego Springs data are used to regress MMI on PGA. The left side shows residuals from Mount Carmel, the right side residuals from Borrego Springs. Note how the Mount Carmel residuals are skewed to the negative side, while the Borrego Springs are skewed to the positive side.

cies; for some PSA values, the 0.3 second and 1.0 second points overlap. This is not the case for the Mount Carmel data, which exhibit a strong difference in slope. For a given PSA value, the higher frequency PSA always produces a higher MMI value.

The above analysis for the MMI-PGA relationship was repeated for PGV. The results show a stronger disparity between the two events, as shown in Figure 4.18. Any given MMI value is associated with lower PGV for the Borrego Springs event than for the Mount Carmel event. MMI-PGV regressions were performed with and without a distance term, and the coefficient estimates are given in Table 4.17:

$$\text{MMI} = c_0 + c_1 \log \text{PGV} \quad (4.10)$$

$$\text{MMI} = c_0 + c_1 \log(\text{PGV}) + c_2 \log(R) \quad (4.11)$$

Where R is epicentral distance. In the preceding sections, magnitude, distance, site and region ef-

Event	Eqn.	c_0	SE	c_1	SE	c_2	SE
Mount Carmel	(4.10)	-0.06	0.47	2.26	0.31	NA	NA
	(4.11)	7.72	0.89	0.50	0.26	-2.23	0.24
Borrego Springs	(4.10)	3.31	0.02	1.57	0.10	NA	NA
	(4.11)	8.61	0.16	0.22	0.06	-2.59	0.08

Table 4.12: Estimates and standard errors for coefficients of the linear regressions (4.10) and (4.11), M5.4 events Mount Carmel and Borrego Springs. Estimates and standard errors are given.

fects on MMI-ground motion relationships have been elucidated using simple averaging statistics, correlation coefficients, and linear regression coefficients. The rest of this analysis uses a qualitative, visual approach to understanding the interrelationships between MMI, ground motions and the auxiliary variables that influence them.

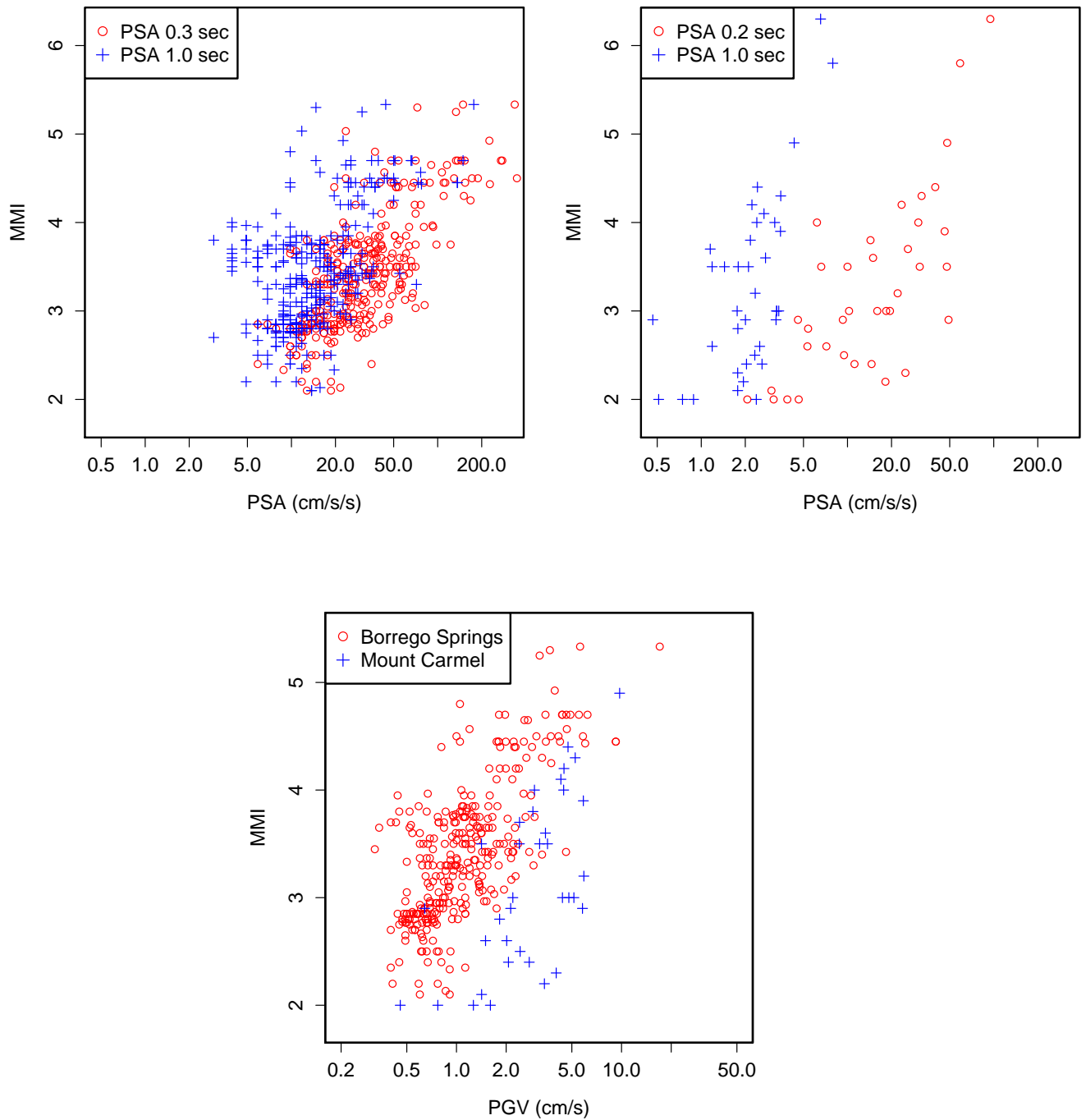


Figure 4.18: Top: MMI vs PSA for M5.4 Mount Carmel (right) and Borrego Springs (left) events. Note that the lower frequency PSA for Mount Carmel is 0.2 seconds, while for Borrego Springs it is 0.3 seconds. Bottom: MMI vs PGV for M5.4 Mount Carmel and Borrego Springs events.

4.6 Kriging and Spatial Analysis

4.6.1 Semivariograms and Spatial Continuity

Section 3 described the semivariogram and its role in describing spatial continuity of a regionalized variable. This section will look at the empirical semivariograms for both MMI and ground motions for different events, and assess features of their spatial variation using variogram differences. Recall that the key features of the semivariogram are the model's functional form, and the nugget, sill and range parameters. The nugget is the Y-intercept of the semivariogram, and ideally should be near-zero for a regionalized variable. This corresponds to the intuitive notion that there should be little variance among observations that are spatially very close to one another. The nugget effect must always be estimated or extrapolated from existing data, as it represents the variance between samples that are closer together than the smallest separation distance.

The general shape of the variogram at small separation distances greatly influence the choice of model. The Gaussian, exponential, linear and spherical models vary predominantly based on their behaviour near the origin. Often the choice between Gaussian and spherical models is somewhat subjective; while some variables have clear-cut model choices, others may be somewhat ambiguous, with more than one model fitting the variogram. The left side of Figure 4.19 shows the empirical semivariogram for the **M4.7** Inglewood and **M4.4** Yorba Linda events. The empirical semivariogram for Inglewood exhibits a distinct point of inflection as it approaches zero separation distance, making the Gaussian model the unequivocal choice. On the right side, The Yorba Linda semivariogram is more obscure, and by overlaying the best-fitting Gaussian and spherical models, we can see how difficult (if not impossible) it is to discern the best choice. The difference is subtle in this case, as evidenced by the modest changes in Kriging outcome shown in Figure 4.20, but larger differences can potentially impact the resulting Kriging estimate in a significant way. We can see from Table 4.13 that all three of the transitional model types are represented in the MMI variogram models.

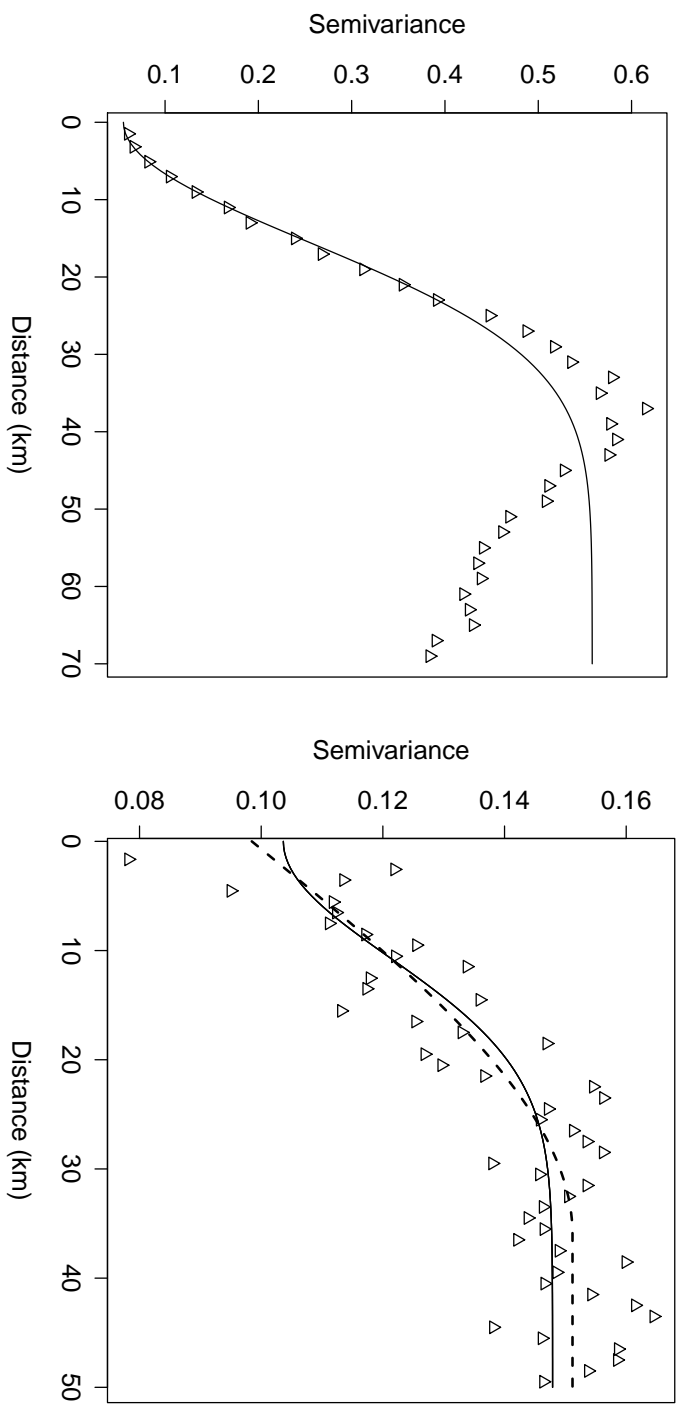


Figure 4.19: Top: Empirical variogram and Gaussian model for Inglewood MMI. The model fits the data quite well and represents a clear choice, particularly at small separation distances where fit is most important. Bottom: Empirical semivariogram for Yorba Linda MMI, with two different choices of model semivariogram: Gaussian (solid line) and Spherical (dashed line). Illustrates the subjective error involved in selecting the best semivariogram model.

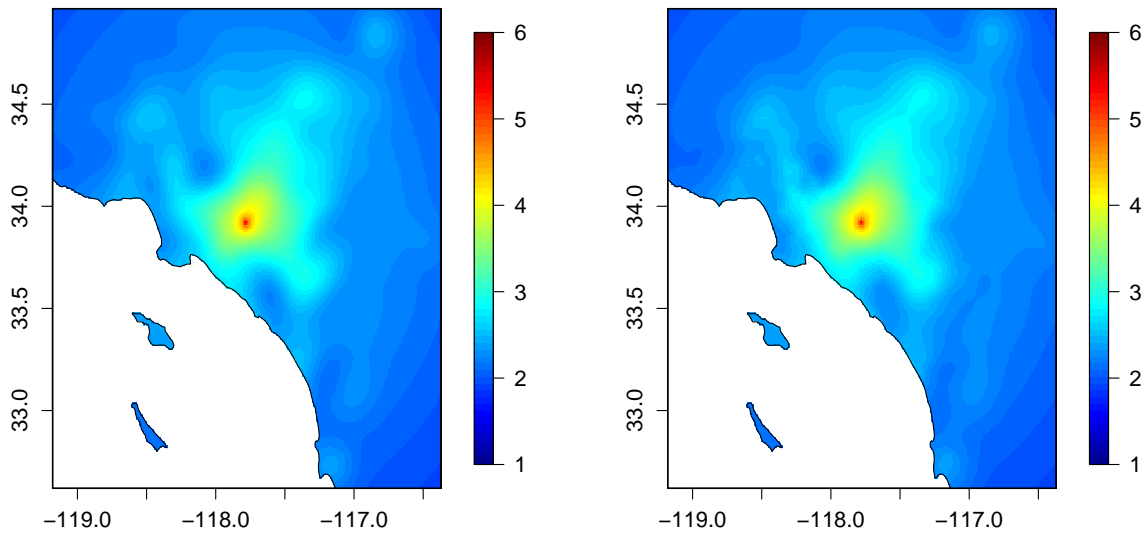


Figure 4.20: Kriging of Yorba Linda MMI based on Gaussian (left) and spherical (right) variogram models. The x-axis represents longitude, and the y-axis is latitude; the scale shown represents MMI values. The differences are modest, reflecting the only slight differences between the two models.

4.6.2 Semivariogram Model Parameters

It may be of interest to understand how the spatial continuity of ground motions and intensities change with the magnitude and region of an event. An easy way to assess inter-event spatial variability is to simply compare the best-fitting variogram model type and the three parameters of the semivariogram - the nugget effect, sill and range. Table 4.13 lists these parameters for a subset of earthquakes.

The range is a key element of the variogram and of spatial continuity. There are some expected patterns in the fitted range values; for instance, the Mount Carmel and Chino Hills events are both magnitude 5.4, but the range for Mount Carmel MMI is 103 km, while for Chino Hills it is 16 km. This reflects the differences in attenuation between eastern and western North America discussed earlier. Similarly we observe some degree of correlation between range and magnitude. The two highest magnitude events - **M6.0** Parkfield and **M7.2** Baja - both have much greater ranges than any other California event: 70 km and 99 km, respectively. This correlation is imperfect though, which could be due simply to errors in the variogram model selection process. The **M5.6** Alum

Event	Parameter	Model	Nugget	Range (km)	Sill
Ground Motion Parameters					
Alum Rock	PGV	Exponential	0.01	24.8	0.08
Anza 2005	PGV	Exponential	0.02	8.7	0.07
Borrego Springs	PGA	Spherical	0.02	56.1	0.02
Chino Hills	PGV	Exponential	0.02	21.4	0.05
	PGA	Exponential	0.03	24.6	0.03
Inglewood	PGV	Gaussian	0.04	19.4	0.20
	PGA	Gaussian	0.04	21.1	0.22
Whittier Narrows	PGV	Exponential	0.03	12.4	0.07
Modified Mercalli Intensity					
Alum Rock		Exponential	0.004	12.11	0.15
Anza 2001		Spherical	0.10	35.1	0.05
Anza 2005		Gaussian	0.09	11.9	0.04
Baja		Spherical	0.09	99.3	0.19
Chino Hills		Spherical	0.18	16.4	0.23
Inglewood		Gaussian	0.06	21.84	0.50
Mount Carmel		Spherical	0.15	103	0.08
Parkfield		Spherical	0.05	70.4	0.18
San Bernardino		Spherical	0.08	62.6	0.06
Yucaipa		Gaussian	0.08	17.0	0.03
Yorba Linda		Spherical	0.10	36.0	0.05

Table 4.13: Variogram parameters for ground motions and MMI, inter-event comparison.

Rock event has an MMI range of only 12 km, the same as the smaller M4.4 Whittier Narrows, and smaller than the M4.7 Inglewood MMI range of 22 km. Finally, we can examine correlation between the ranges of MMI, PGA and PGV for the same event. Only a few earthquakes possessed data which allowed both MMI and ground motion parameter variograms to be modeled. These included the M4.7 Inglewood, with ranges for MMI, PGV and PGA equal to 22, 27 and 26 km. These are similar, but this was not the case for the M5.6 Alum Rock; here, the range for the PGA variogram was only 4 km, for MMI it was 12 km, and for PGV it was 30 km. Again, these disparities could simply be due to moel error, or there could be some actual underlying differences in spatial continuity.

Not all events lend themselves to MMI or ground motion variogram estimation. Sometimes the data are too sparse, with not enough pairs at certain separation distances to get a meaningful semi-variance value. Other times the data are simply erratic, with semivariance changing wildly from one separation distance to the next. Figure 4.21 shows MMI observations versus distance, and the empirical semivariogram for the San Simeon event. The data are too sparse at close distances to the epicenter, and the resulting variance among points with small separation distances is highly scattered and erratic. Any of the models could be made to fit such an empirical semivariogram, and all would perform equally poorly. Essentially the San Simeon event lacks enough data at the right distances to get any clear idea of the spatial correlation. While we can be sure that this correlation exists, it is not reflected in the data.

4.6.3 Kriging Estimation of MMI and Ground Motion Parameters

The dual effects of attenuation with distance and amplification due to site conditions can interact to give MMI and ground motions irregular spatial patterns. If MMI and ground motions were governed only by distance from the epicenter, a map of these variables would show regular concentric circles of steadily decreasing values. By interpolating known data to create continuous spatial maps, we can see how much actual ground motion and intensity data depart from this pure distance effect, and relate these variations back to site condition information.

An outline of the Kriging procedure used was given in Section 3.3; to supplement that explanation, this section will begin with a more detailed walkthrough of the procedure for Kriging PGV for the 2009 M4.7 Inglewood event.

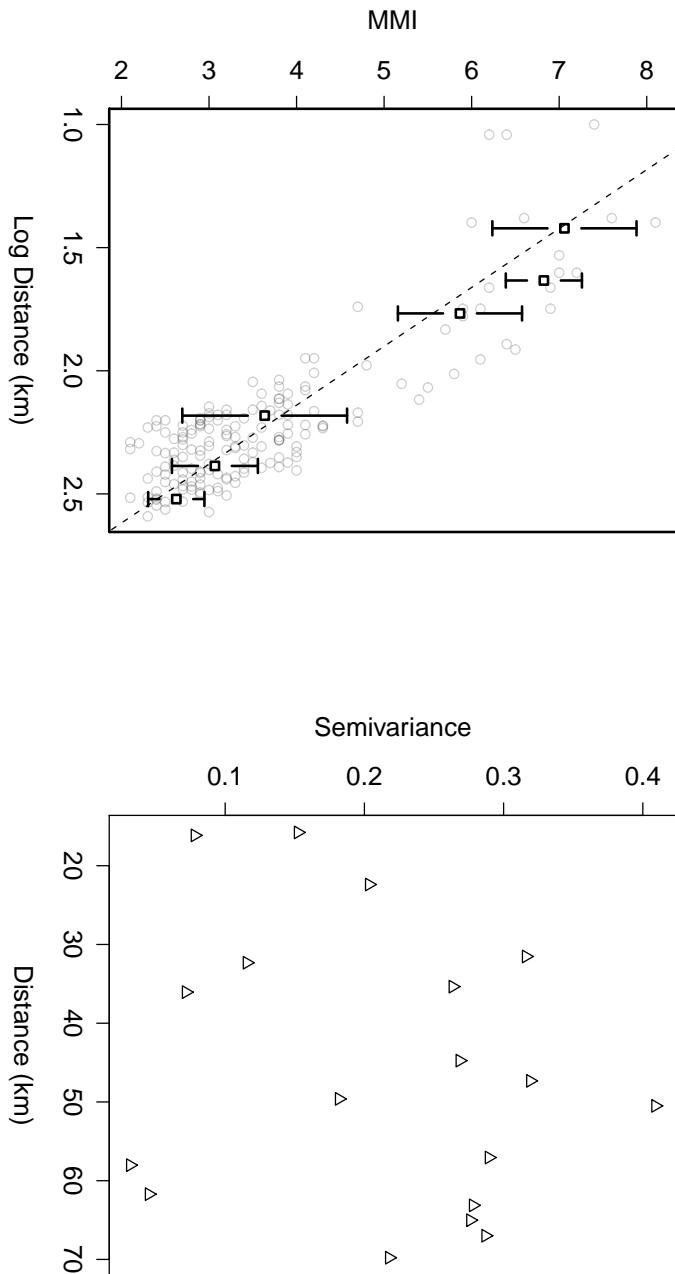


Figure 4.21: Illustration of a failure to match a variogram model to the empirical variogram from the 2003 **M6.5** San Simeon event. Top: the MMI-distance plot shows a dearth of observations close to the epicenter. Bottom: The empirical variogram is erratic with almost no structure, making it impossible to fit a variogram model.

1. The first step entails detrending the data for the large scale decrease with distance. Initially this was done using an **R** function called `surf.ls`, part of the `spatial` package that fits a trend surface by ordinary least-squares. This method yielded poor results however, as the function chooses the best-fitting polynomial of a given degree and often matched outlying data at the expense of points near the epicenter. Two examples of these unsuccessful attempts are shown in Figure 4.22. An alternative approach used weighted linear regressions of averaged MMI and ground motion data to remove the distance trend. A plot of the data with linear model is shown in Figure 4.23, along with a histogram of residuals.

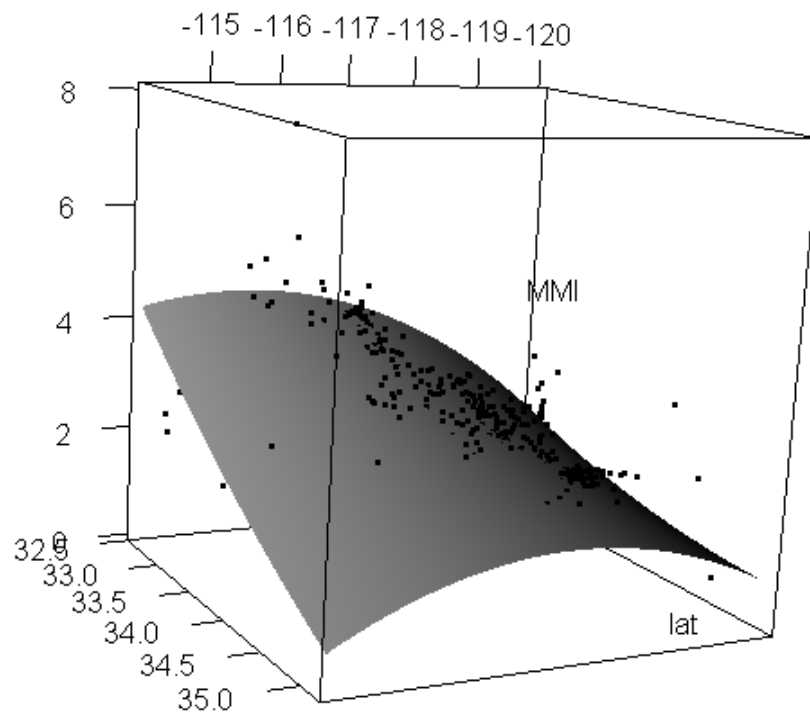


Figure 4.22: Trend surface of MMI from M5.7 Ocotillo earthquake. Horizontal axes are longitude and latitude, and the surface represents the best-fitting polynomial surface for the MMI observations.

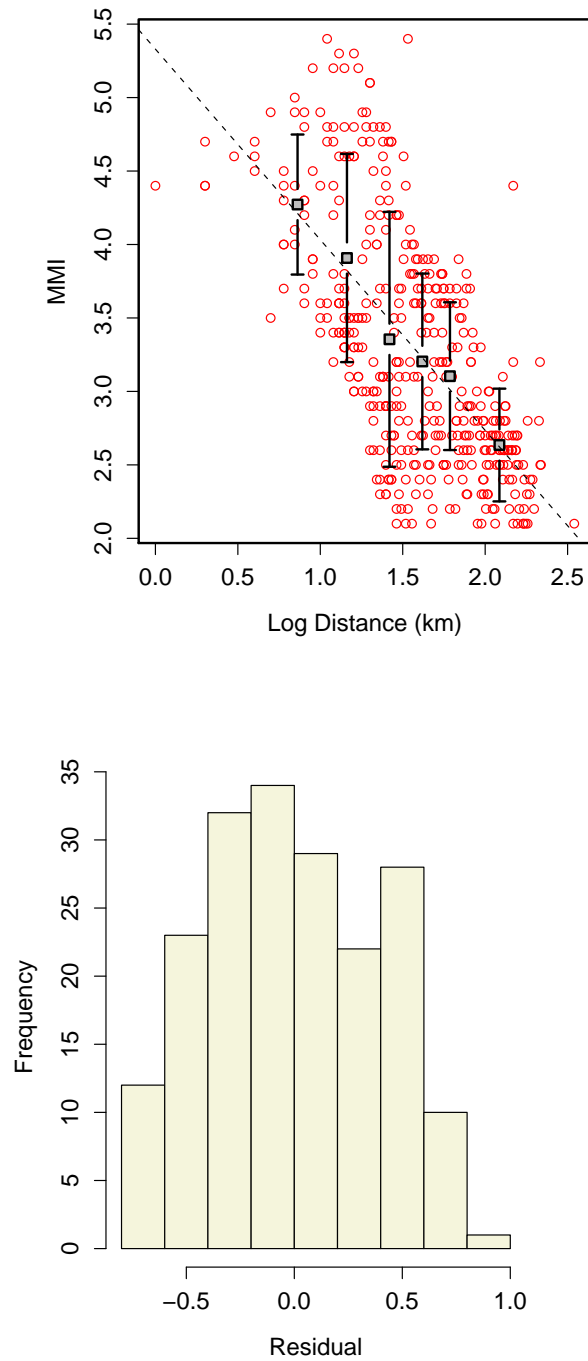


Figure 4.23: Left: MMI-PGV data and linear relationship for 2009 M4.7 Inglewood event. Right: Histogram of residuals from linear regression.

2. A `gstat` object is created, which holds the latitude and longitude coordinates of the data and the value of residual at each location. A grid is also created based on the range of coordinates; these are the locations where PGA will be estimated.
3. The empirical variogram is plotted using the `gstat` routine `variogram`. It takes the `gstat` object, a cutoff distance, and a value to determine how many separation distances are used to compute the empirical variogram.
4. The nugget effect and range are roughly estimated by visual inspection, and these are used to create variogram models using the `vgm` function. Ground motion data are best fit using exponential, spherical or gaussian models. Figure 4.24 shows the empirical variogram for Borrego Springs with the best-fitting individual of these three variogram model families.
5. The `gstat` object is updated to include the variogram model. Kriging is performed using the `predict` function, which uses the `gstat` object, the variogram model, and the grid. The residuals Kriging estimates are mapped in Figure 4.25, top left.
6. The distances between the epicenter and all grid coordinates are computed. These are then used in the linear model of PGA vs distance to find what the pure distance trend component is at each location. An image of the trend component is given in the top right of Figure 4.25. This is added to the grid of Kriged residuals to yield the final estimates.
7. The `overlay` function overlays the Kriging grid locations on a shapefile map of California to find which grid points fall within the boundary of California. Those that lie outside the boundaries are assigned NA.
8. The resulting grid is converted into an image using `as.image.SpatialGridDataFrame` and plotted; the result is shown in Figure 4.25, bottom right.
9. These steps are repeated for MMI and PGV, PGA, average PSA, and Vs30 in the region of interest. Figure 4.26 shows an example of the final result.

A second example of the final results of this Kriging procedure is presented in Figure 4.27, which shows a map of the Kriging estimates for MMI and PGV values from the M5.6 Alum Rock earthquake. The striking feature of these maps are the correlations among irregularities. Patches of relatively higher MMI appear in a ray to the south-east of the epicenter, with MMI values 0.5-1

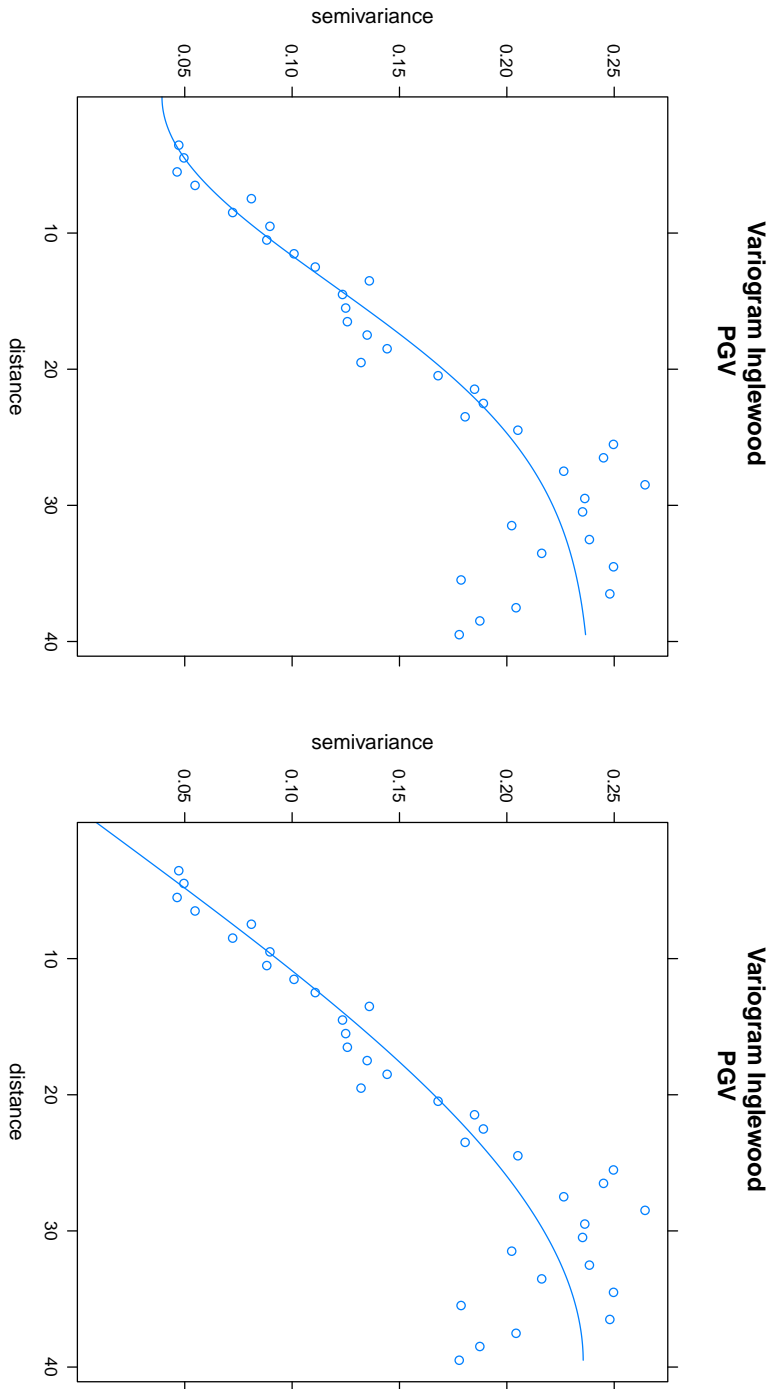


Figure 4.24: Empirical variogram for M4.7 Inglewood event, PGV. Two different variogram models are shown; Gaussian (left), and Spherical (right). The spherical and Gaussian models both fit equally well, and only differ by the unknown behaviour near the origin. The spherical model was chosen.

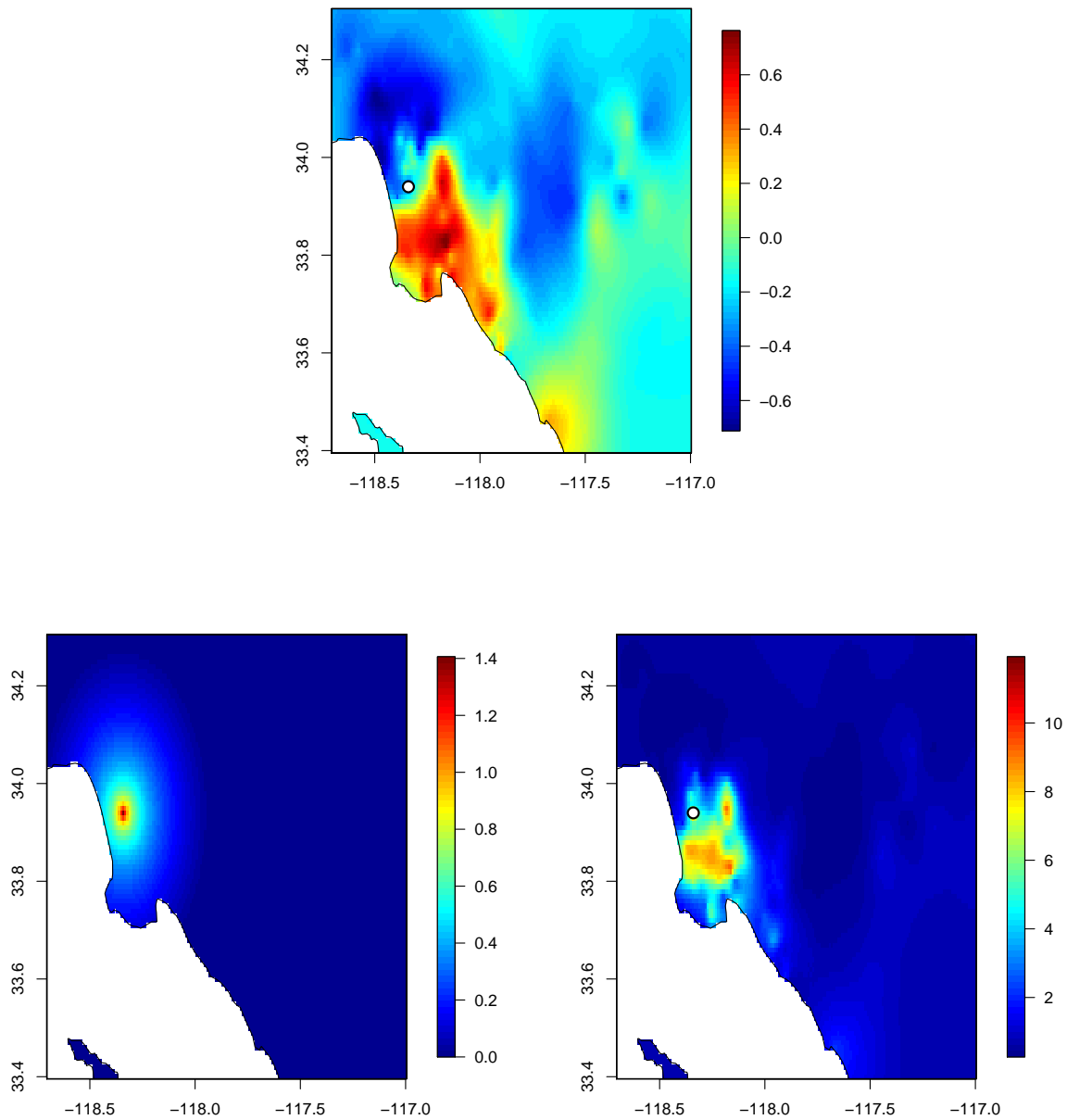


Figure 4.25: Illustration of the steps involved in the Kriging of PGV for **M4.7** Inglewood event. Horizontal axis is longitude, vertical axis is latitude. Epicenter is depicted as circular dot. Scale represents PGV in cm/s. Top: Kriging of residuals from the regression of PGV against epicentral distance. Bottom left: Spatial representation of the linear distance trend component of PGV. Bottom right: Complete Kriging estimate made by adding residual component and trend component.

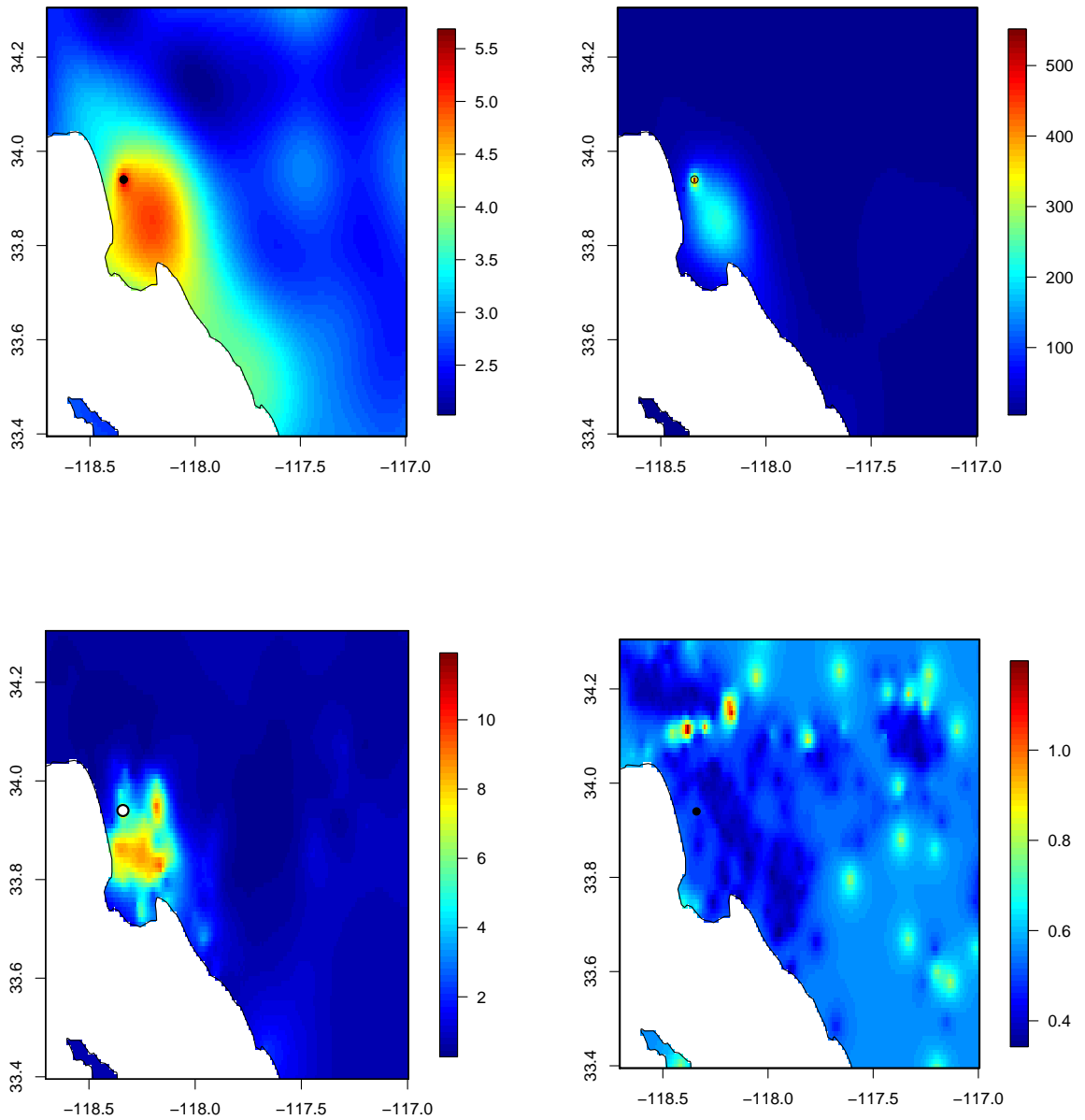


Figure 4.26: Kriging of MMI (top left), PGA (top right), PGV (bottom left) and ratio of $V_{S30}/760$ m/s (bottom right) from M4.7 Inglewood earthquake. Axes represent longitude and latitude; epicenter represented by small circle.

units above other regions at the same distance. This pattern is reflected in the PGV map, especially close to the epicenter where there appears to be an elliptical region of high PGV. Figure 4.28 shows a topographic map of California, and matching the elliptical areas of relatively higher MMI and ground motions to this map shows regions higher elevation running in approximately the same direction. It could be that an aspect of the topography in this region is amplifying the ground motions and influencing MMI. More in-depth pattern matching of this topography with the maps of Kriging estimates would be able to quantify this influence.

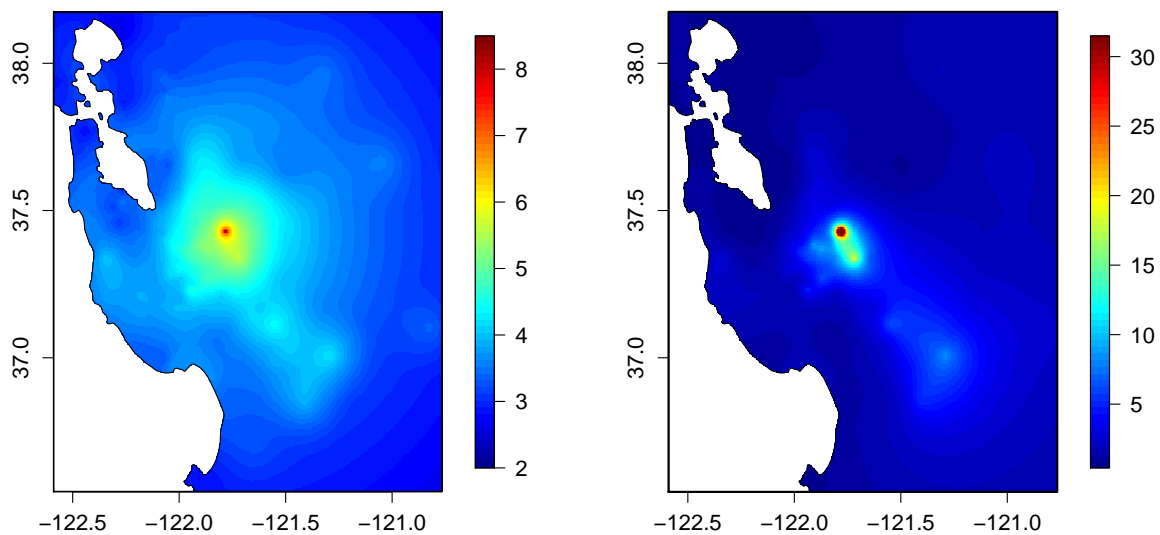


Figure 4.27: Kriging of MMI (left) and PGV (right) from M5.6 Alum Rock earthquake. Epicenters are located at the points of highest MMI and PGV values (red areas).

4.6.4 Kriging V_{S30}

While the MMI and ground motion data show a great deal of spatial continuity, the V_{S30} data vary more erratically over space. Figure 4.29 shows the empirical semivariogram for the California

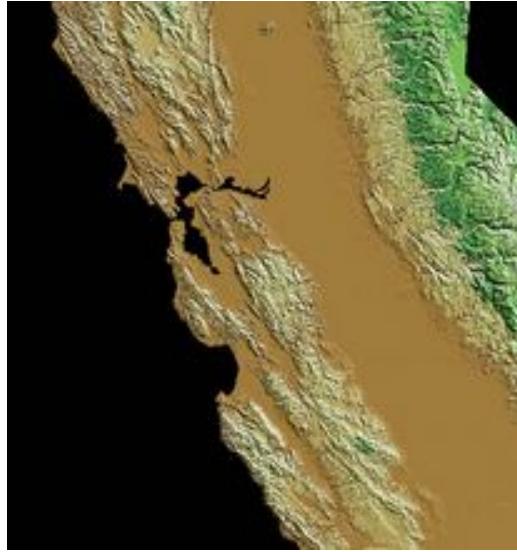


Figure 4.28: Topographic map of California, from USGS digital elevation model [USGS, 2012].

V_{S30} data. The model semivariogram is exponential, with a range of 2.14 km, a nugget of 0.002, and a sill of 0.03. The very small range indicates that V_{S30} changes over short distances, with no correlation of V_{S30} values located more than 2km apart. Given that the average distance between observations is greater than 2, the potential of Kriging for this data is quite limited. When all of California is used as support, the resulting map looks more like individual points plotted rather than a smooth, continuous map of values. On a smaller scale the results can still be used to examine V_{S30} values in regions where MMI or ground motions show anomalous behaviour.

The final use of Kriging to be discussed is block kriging, used both as a visual tool for assessing spatial correlation, as well as an alternative way to match MMI and ground motion data.

4.6.5 Block Kriging

Given that MMI data are unalterably areal in nature, an intuitively appealing way to match MMI and ground motion data is to estimate average ground motions over compatible areas, converting

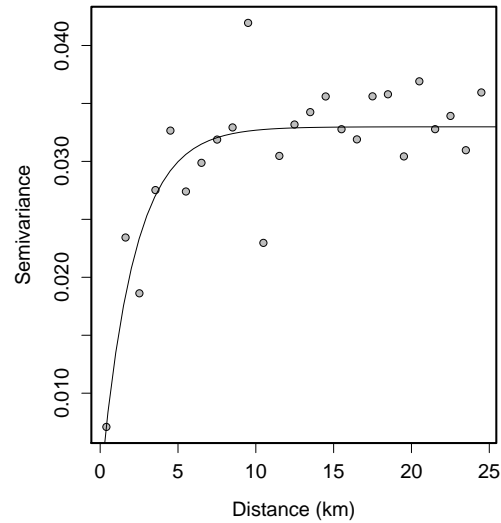


Figure 4.29: Variogram showing spatial continuity of V_{S30} in California.

ground motions from a point support to an areal support. As discussed in section 3, block Kriging differs from ordinary Kriging in a simple way. Block Kriging takes a set of areal supports (or blocks) and deaggregates them into a set of regularly spaced points. Ordinary Kriging is performed over these points, and the resulting average of all points within a block is the Kriging estimator for that area.

The method used for block Kriging involves some additional steps to estimate the trend component for the polygonal zip code areas. For ordinary Kriging, the trend at each point was calculated from the linear model of ground motion or MMI with distance. Now the trend must be an average over the zip code. So again, the trend is calculated at each point, and these points are converted to a Spatial Points object. Then these can be overlaid on the zip code shapefile map to find which points lie in each polygon. All the points within each polygon are averaged to get the average trend component for each zip code.

Block Kriging results from the M4.7 Inglewood, CA earthquake are shown in Figure 4.30. The

top of the figure shows Kriging for PGV; the bottom shows a choropleth map for MMI, in which different colours are assigned based on average values of each zip code region. An interesting feature of these maps is the ring of PGV of 0.5 cm/s inside a ring of PGV 0.25 cm/s. Similarly on the MMI block Kriging map, a ring of MMI 2 is flanked on one side by MMI 4.0 and the other side by MMI 2.5 and 3.5. There is also a number of zip codes on the east side of the map with MMI 3.5 (turquoise) that correspond to the lowest PGV values (0.15-0.25 cm/s). How well do the block-Kriged matched data stack up against the station-zip code matched data? Figure 4.32 shows the MMI-PGV plot for both types of data for M4.7 Inglewood and M5.2 Alum Rock. The block-kriging matched data are more numerous, since the limiting quantity is the number of distinct zip codes and not the number of stations, which tends to be less. For both events, the block-kringed data show anomalous clusters of points that don't seem to fit the overall MMI-PGV relationship. For the Inglewood event, the cluster occurs between MMI 2.0 and 2.5 and PGV 0.5-1.0 cm/s, with group of points associated with larger PGV than the general trend. For Alum Rock, the cluster occurs in the area of MMI 3-4 and PGV 0.5-1.0 cm/s, this time with lower PGV than the trend. It is unknown what the underlying cause for these anomalous matches are, but likely it is something systematic in the block kriging procedure used.

The Kriging results presented above offer a more qualitative approach to understand MMI-ground motion interactions than purely statistical analyses, and can provide a valuable complement to more rigorous methods. Kriging maps allow for easy visual identification of interesting spatial relationships, which may not be readily apparent from the traditional MMI-ground motion plots. Combined with maps of local site conditions, Kriging maps can illustrate some of the interrelationships between MMI, ground motion, and V_S 30.

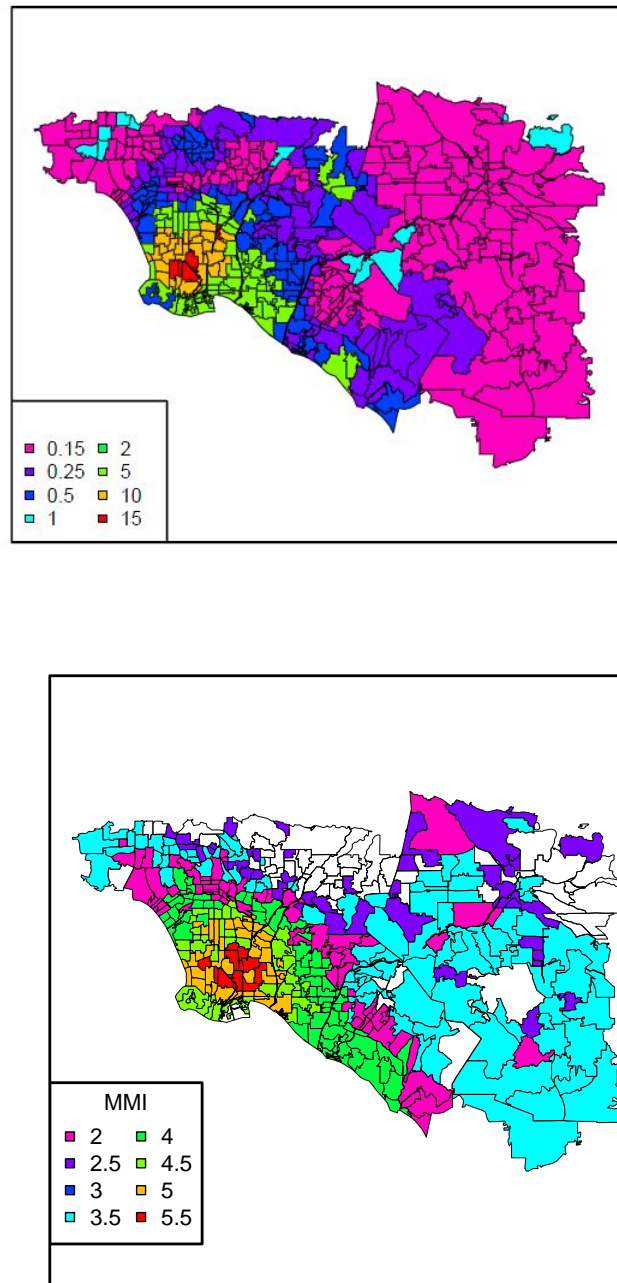


Figure 4.30: Top: Block Kriging of PGV (cm/s) over zip codes in California; Bottom: chloropleth map of zip-coded MMI for M4.7 Inglewood earthquake. In the chloropleth map, different colours represent different MMI values.

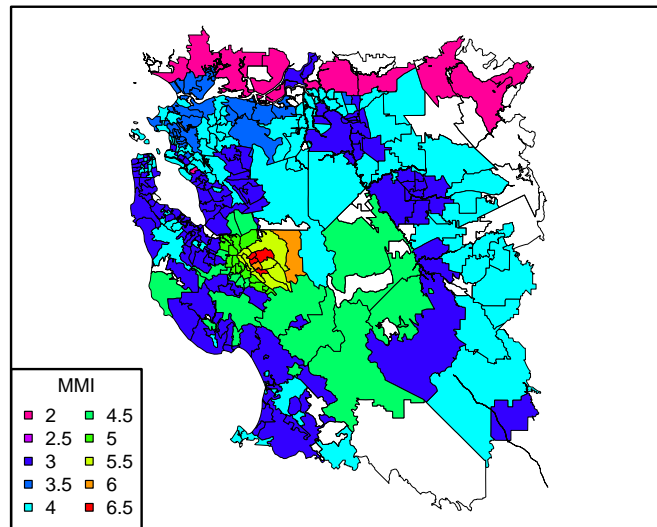
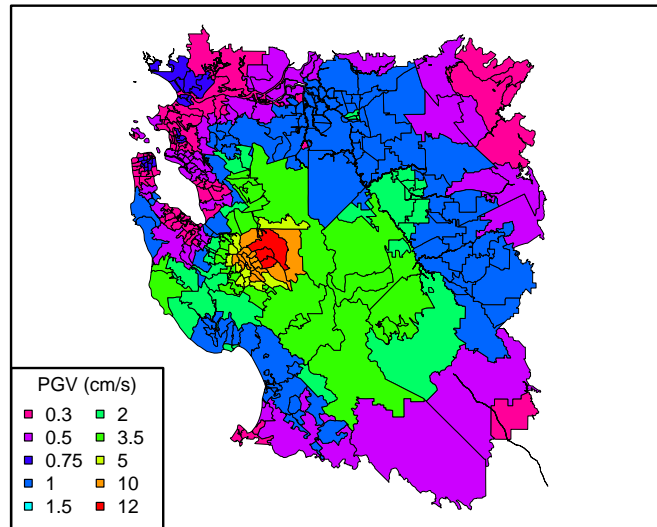


Figure 4.31: Block Kriging of PGV and chloropleth map of MMI for M5.2 Alum Rock earthquake.

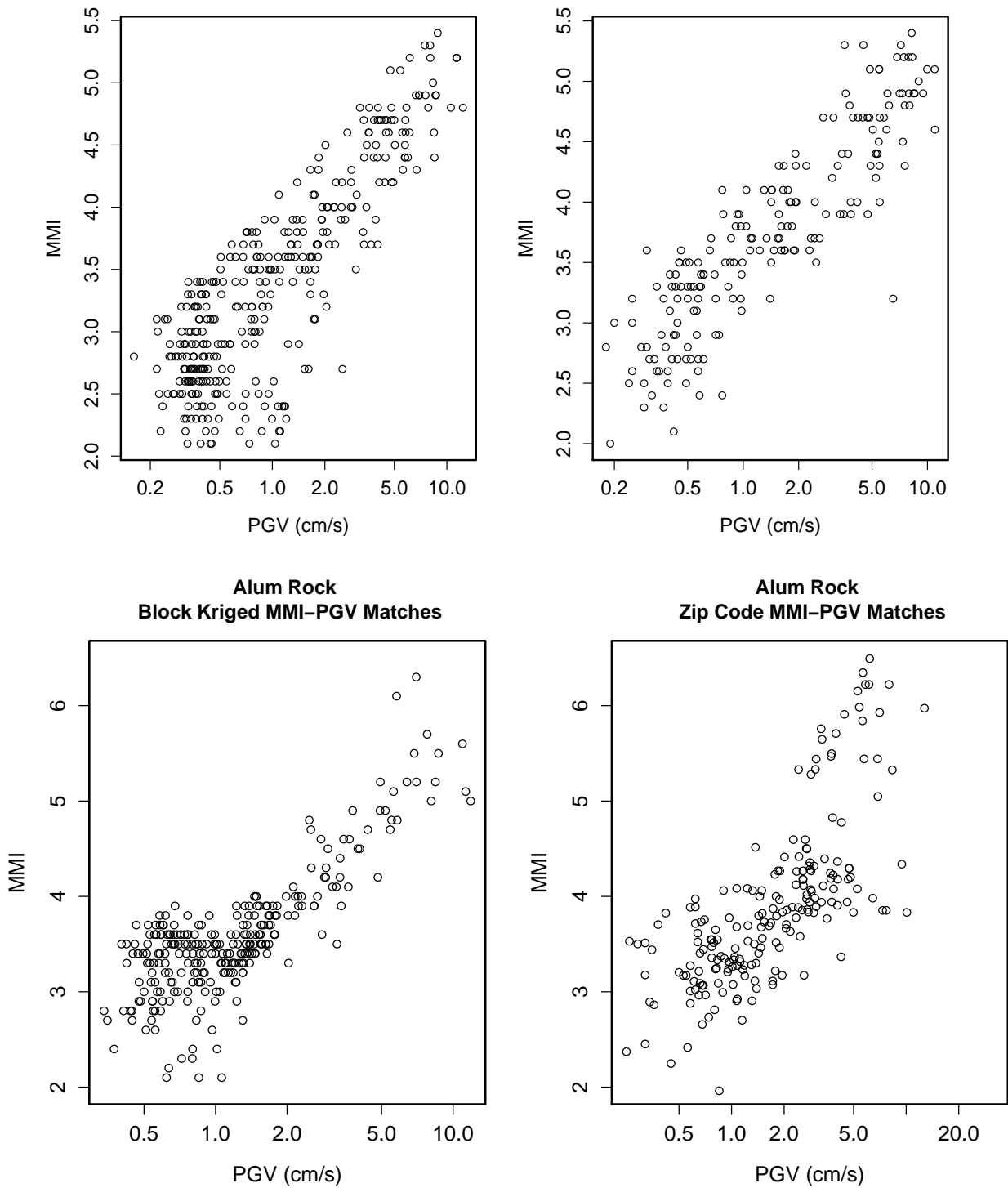


Figure 4.32: Comparison of matched data from block Kriging (left) and by matching stations with zip codes (right) for M4.7 Inglewood earthquake (top) and Alum Rock earthquake (bottom)

Chapter 5

Summary of Results, Conclusions and Discussion

This study expands upon earlier work developing relationships between Modified Mercalli Intensity and ground motion parameters. Using some well-recorded earthquakes with rich MMI data from the Did You Feel It? website, this study looks at the strength of correlation between ground motion parameters and MMI, and the effect of distance, magnitude and site conditions from a variety of different perspectives. The following sections are organized thematically, summarizing the key results presented here and discussing the results: what conclusions can be drawn, the limitations of this study, and possibilities for future research.

5.1 Variability of MMI and Ground Motion Parameters

It was shown throughout this study that there is a large degree of variability among MMI observations. Because of the way the MMI scale is defined - with lower MMI values described by human experiences and reactions, and higher MMI values corresponding to physical effects and damages to structures - it would be natural for the variance of higher and lower MMI ranges to differ. One might expect the variability of low MMI values to be greater than higher MMI. An alternative approach to the weighted regression used here would be to estimate separate variance parameters $\sigma_1, \sigma_2, \dots, \sigma_n$ for each of the n MMI ranges used.

One important aspect of MMI variability was not addressed here: the number of observations making up the spatially averaged MMI value. While the MMI vs. ground motion parameter regressions were performed on grouped averages and weighted by the number of *matched* data points, the simple model used did not recognize the quality of the individual MMI values making up the matched data. The data was filtered to contain only those MMI values representing at least three individual responses, but there was still a great amount of variability among the number of individual DYFI responses comprising the MMI observations. For example, an MMI value derived from the average of only three DYFI reports was treated exactly the same way as one derived from 200 DYFI reports. Obviously the accuracy of these two MMI values could be vastly different. Moreover, if we have two MMI values each consisting of the same number of individual DYFI responses, the quality of the MMI value could vary based on the variability among the DYFI reports themselves. If one MMI observation is the result of ten very different DYFI reports, while the other was obtained by average ten very similar DYFI reports, the latter would likely be a more reliable observation. Greater attention to this additional source of variability in the MMI vs. ground motion regressions would lead to more precise coefficient estimates.

MMI vs. ground motion regressions performed using separate variance parameters for each group of matched data points (as described above) might be able to incorporate this extra information about MMI variability. If the individual responses comprising each MMI value were known, a variance parameter could be ascribed to each MMI observations; these could then be aggregated and weighted by the total number of data points for each matched MMI-ground motion group used in the regression.

5.2 Correlation Coefficients and Regression Statistics

The simple grouping of MMI and ground motion data portrayed the basic correlations between them, highlighting the large degree of uncertainty in trying to associate MMI values with particular ground motion parameter values. Differences in the behaviour of MMI and ground motions with epicentral distance also shed light on the limitations of MMI-ground motion relationships.

Correlation coefficients were used to quantify the strength of correlations between MMI and ground motions over different distance and magnitude ranges. Weighted least squares linear regressions provide estimates of coefficients for predictive equations, allowing the conversion between MMI and ground motion parameters. These regressions included magnitude, distance and V_{S30} information to capture additional sources of variation. The coefficients are comparable, but different than those found in previous studies. The inclusion of a V_{S30} term was found to be only marginally effective, implying a need for denser observations to determine the value of including site conditions in MMI-ground motion regressions.

The differences between regression results presented here and those from previous studies (e.g. Atkinson and Kaka [2007], Worden et al. [2011]) are likely due to a combination of data selection, regional differences and variations in methodology. There are numerous sources of uncertainty in

the development of MMI vs ground motion relationships, beginning with the data itself and how it is selected, how ground motion parameters are matched with MMI values, to how the regressions are performed and the statistical techniques used to deal with the highly scattered MMI data. The potential for biases to be introduced at any of these steps highlights the sensitivity of MMI-ground motion relationships to data selection and methodology, and suggests that a comprehensive database of matched MMI and ground motion data may improve regression coefficient estimates.

5.3 Eastern vs. Western North America

One of the central goals of this study was the comparison of MMI-ground motion relationships between Eastern and Western North America, for select well-recorded events. This was done using two M5.4 earthquakes, one from each region: the 2008 Mount Carmel, IL event from the east, and the 2010 Borrego Springs event from the west. This analysis suggests an underlying difference in the behaviour of matched MMI-ground motion data between the two regions. For the CEUS Mount Carmel earthquake, MMI increases more rapidly with increasing PGA, even when differences in the distance ranges of the two events is taken into account. For a given level of PGA, the associated MMI is always larger for the Mount Carmel event than for Borrego Springs. This result runs counterintuitive to the idea that high PGA values in ENA occur at high frequencies, which tend not to affect MMI as much as large accelerations at lower frequencies. This phenomenon may be caused by marked differences in site conditions for the two events. Mount Carmel is located at the north end of the Mississippi embayment, where deep sediments may mute PGA and disguise high PGA values. Detailed information about V_{S30} , surficial geology, soil profiles, or other site condition proxies in both regions could validate this explanation.

This comparison also revealed differences in the MMI-PGV relationships between the two events.

For a given PGV value, the associated MMI for Mount Carmel was significantly lower than for Borrego Springs - an opposite effect to that observed for PGA. This phenomenon was not observed for the Canadian data, which is consistent with the explanation that unique conditions in the embayment are at work to produce this effect. These comparisons suggest that region-specific adjustments to residuals from ground motion-MMI regressions may improve predictive relationships. Though the data used here are limited to only three events from eastern North America, it is noteworthy that the two Canadian events - Riviere du Loup and Val des Bois - are associated with significant positive residuals from the overall MMI-ground motion regression, while the Mount Carmel event is associated with negative residuals.

5.4 V_{S30} as a Proxy for Site Condition

The use of V_{S30} values in MMI-ground motion equations provided insight into the effect of site conditions on MMI-ground motion correlations, as well as the value and limitations of including this information. While the regression coefficients for the reference V_{S30} terms in the PGA and PGV regressions were significant at the 0.05 level, the resulting change in the other coefficients was modest, and the MMI-ground motion regressions with only magnitude and distance terms are likely adequate. But the results suggest the possibility of generating more accurate relationships if a denser array of V_{S30} information was available.

Alternatively, the mixed outcome of including V_{S30} terms in the MMI vs. ground motion regressions might suggest that different measures of site condition should be used. Ideally, more detailed data about the entire V_{S30} profile at the ground motion recording stations would provide a much more accurate estimate of the true amplification. Even information about the depth to bedrock at the V_{S30} sites used here would clarify some of the inherently confounding aspects of using V_{S30}

data. For example, the V_{S30} value at two sites could be identical, but one site might consist of a shallow layer of extremely soft soil, while the other consists of a much deeper layer of firm soil or soft rock. The amplification at these two sites would then be vastly different, yet this would not be captured in the V_{S30} values.

Additionally there are aspects of an area's local geology that might influence MMI values but not ground motion values. While the ground motions of earthquakes used in this study are likely insufficient to cause liquefaction, for larger ground motions it might be necessary to include information about the liquefaction potential of the region being studied. With enough detailed information, a composite proxy for site amplification may be able to explain a significant amount of the variability in MMI-ground motion relationships. Such a proxy might consist of information about topography, soil profiles and depth to bedrock, shear wave velocity profiles, liquefaction potential and surficial geology. As these data become more widely available, they can serve to further improve MMI-ground motion equations.

5.5 Kriging for Ground Motions and Seismic Intensity

The use of Kriging as a complementary means of assessing the relationships between MMI and ground motion had two applications. The first was to create maps visualizing the spatial variation in ground motions, MMI and V_{S30} , in order to detect curious relationships that may not be revealed by purely statistical methods. The pattern of MMI and PGV in the Kriging maps for the M5.6 Alum Rock earthquake shows how the local geography and geology of a region can influence both seismic intensity and instrumental ground motion measurements. The application of Kriging to this type of comparison could be extended to include pattern-matching analyses using maps of topography, surficial geology, liquefaction potential, or another related variable. In

a pattern-matching analysis, the Kriging estimates of ground motion parameters and MMI could be compared individually to the value of other variables at the same locations in space, and the strength of their spatial correlation could be quantified.

The model variogram fitting provides basic information about spatial continuity of MMI and ground motions for a number of earthquakes. Some patterns within the variogram model parameters could be found, but there is a large degree of uncertainty associated with variogram model selection. It would be interesting to delve more thoroughly into the spatial variability of ground motion parameters and MMI with a more detailed study. An expanded data set and greater attention to variogram modeling might reveal interesting relationships between spatial continuity, magnitude, and region. In particular, a more rigorous method of selecting the appropriate variogram model should be employed. In this study, the best-fitting members of each variogram family were evaluated by visual inspection, and not through an optimization procedure. The extra step of quantifying the performance of each best-fitting member might improve the Kriging estimates.

The second use of Kriging employed in this study was as a new method to generate matched MIM-ground motion data. The quantity of matched observations available is typically limited by the number of seismographic stations. The novel procedure used here to match ground motions to DYFI observations using block Kriging has the potential to produce larger datasets for developing MMI-ground motion relationships. By generating ground motion estimates for every zip code, the number of matched observations becomes dependent of the number of DYFI zip code areas, which are always more numerous than seismographic stations.

The data matched using this method preserve the general trends seen in the traditionally matched data, with some anomalous features that are likely an artifact of errors introduced during the Kriging process. This could include subjective error originating from the selection of variogram models, which was performed in a somewhat ad hoc fashion. More rigorous testing and validation of

model selection could yield a better variogram model, and hence more precise Kriging estimates.

The DYFI website is increasingly using geocoding to pin down street-level observations, which can then be aggregated over a regular areal grid, rather than the irregular ZIP code areas which differ greatly in size and shape. The method used here required information about the boundaries of areal regions, which was easily available for ZIP code data. Using similar boundary data for the grid-aggregated DYFI observations would almost certainly improve the accuracy of this matching method. Since block Kriging is an average of estimates of a region, the smaller the region, the better this average will reflect the actual ground motion values across the region.

Bibliography

- P. Agarwal and M. Shrikhande. *Earthquake Resistant Design Of Structures*. Prentice-Hall, 2006.
- K. Aki and P. G. Richards. *Quantitative Seismology*. University Science Books, 2002.
- T. Allen and D. Wald. Global Vs30 Map Server. <http://earthquake.usgs.gov/hazards/apps/vs30/>, 2011.
- R. V. Arsdale. *Adventures Through Deep Time: The Central Mississippi River Valley and Its Earthquakes*. Geological Society of America, 2009.
- G. M. Atkinson and D. Boore. New ground motion relations for eastern North America. *Bulletin of the Seismological Society of America*, 85:17–30, 1995.
- G. M. Atkinson and D. M. Boore. Recent trends in Ground Motion and Spectral Response Relations for North America. *Earthquake Spectra*, 6(1), 1990.
- G. M. Atkinson and S. I. Kaka. Relationships between Felt Intensity and Instrumental Ground Motion in the Central United States and California. *Bulletin of the Seismological Society of America*, 97(2):497–510, 2007.
- G. M. Atkinson and W. Silva. Stochastic Modeling of California Ground Motions. *Bulletin of the Seismological Society of America*, 90(2):255–274, 2000.
- G. M. Atkinson and E. Sonley. Empirical Relationships between Modified Mercalli Intensity and Response Spectra. *Bulletin of the Seismological Society of America*, 90(2):537–544, 2000.
- G. M. Atkinson and D. J. Wald. “Did You Feel It?” Intensity Data: A Surprisingly Good Measure of Earthquake Ground Motion. *Seismological Research Letters*, 78(3):362–368, 2007.
- W. Bakun and A. McGarr. Differences in attenuation among the stable continental regions. *Geophysical Research Letters*, 29(23):1–4, 2002.
- R. D. Borcherdt. Estimates of Site-Dependent Response Spectra for Design (Methodology and Justification). *Earthquake Spectra*, 10(4), 1994.
- J. R. Carr and C. E. Glass. Treatment of Earthquake Ground Motion Using Regionalized Variables. *Mathematical Geology*, 17(3):221–241, 1985.

- S. Castellaro, F. Mulargia, and P. L. Rossi. Vs30: Proxy for Seismic Amplification? *Seismological Research Letters*, 79:540–543, 2008.
- W.-F. Chen and C. Scawthorn. *Earthquake Engineering Handbook*. CRC Press, 2003.
- J.-P. Chils and P. Delfiner. *Geostatistics: modeling spatial uncertainty*. John Wiley & Sons, Inc., 1999.
- V. F. Cormier. Seismic, Viscoelastic Attenuation. In H. Gupta, editor, *Encyclopedia of Solid Earth Geophysics*, page 1280. Springer, 2011.
- C. Cramer, J. R. Kutliroff, and D. T. Dangkua. The 2011 Mineral, VA M5.8 Earthquake Ground Motions and Stress Drop: An Important Contribution to the NGA East Ground Motion Database. In *83rd Annual Meeting of the Eastern Section of the Seismological Society of America*. Univ. of Memphis, October 2011.
- N. Cressie. The Origins of Kriging. *Mathematical Geology*, 22(3):239–252, 1990.
- P. Dalgaard. *Introductory Statistics with R*. Springer Science+Business Media, LLC, 2008.
- T. Datta. *Seismic Analysis of Structures*. John Wiley and Sons, 2010.
- J. de Gruijter, D. Brus, M. Bierkens, and M. Knotters. *Sampling for Natural Resource Monitoring*. Springer-Verlag Berlin, 2006.
- W. Finn. Earthquake Engineering. In R. Rowe, editor, *Geotechnical and Geoenvironmental Engineering Handbook*, page 629. Kluwer Academic Publishers, 2001.
- A. E. Gelfand, L. Zhu, and B. P. Carlin. On the change of support problem for spatio-temporal data. *Biostatistics*, 2(1):31–45, 2001.
- G. Matheron. Principles of Geostatistics. *Economic Geology*, 58:1246–1266, 1963.
- J. J. Gómez-Hernández. Geostatistics and Hydrology: An Overview. In V. C. de la Asociación Español de Sistemas de Información Geográfica, editor, *Geographical Information: form research to application through Cooperation*, pages 457–466. IOS Press, 1996.
- P. Goovaerts. Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units. *Mathematical Geosciences*, 40:101–128, 2008.
- G. Gorshkov and G. Shenkareva. On the Correlation of Seismic Scales. *U.S. Joint Publications Reseach Service*, 1960.
- C. A. Gotway and L. J. Young. Combining Incompatible Spatial Data. *Journal of the American Statistical Association*, pages 632–648, June 2002.
- T. C. Hanks and H. Kanamori. A Moment Magnitude Scale. *Journal of Geophysical Research*, 84: 2348–2350, 1979.

- S. Hartzell and C. Mendoza. Source and Site Response Study of the 2008 Mount Carmel, Illinois, Earthquake. *Bulletin of the Seismological Society of America*, 101:951–963, 2011.
- N. Jayaram and J. W. Baker. Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering and Structural Dynamics*, pages 1687–1708, 2009.
- P. C. Jennings. An Introduction to the Earthquake Response of Structures. In W. H. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger, editors, *International handbook of earthquake and engineering seismology, Part B*, pages 1097–1125. Academic Press, 2003.
- A. Journel and M. Rossi. When do we need a trend model? *Study on Statistics and Environmental Factors in Health*, 1988.
- S. I. Kaka and G. M. Atkinson. Relationships between Instrumental Ground-Motion Parameters and Modified Mercalli Intensity in Eastern North America. *Bulletin of the Seismological Society of America*, 94(5):1728–1736, 2004.
- A. J. Kappos. *Dynamic loading and design of structures*. Spon Press, 2002.
- G. Leblanc and G. Klimkiewicz. Seismological issues: history and examples of earthquake hazard assessment for Canadian nuclear generating stations. *Geological Survey of Canada Open File 2929*, pages 614–615, 1994.
- M. W. McCann, F. Sauter, and H. C. Shah. A Technical Note on PGA-Intensity Relations with Applications to Damage Estimation. *Bulletin of the Seismological Society of America*, 70(2): 631–637, 1980.
- M. Oliver. *Geostatistical Applications for Precision Agriculture*. Springer Science + Business Media, 2010.
- J. Park, P. Bazzurro, and J. Baker. Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. *10th International Conference on Application of Statistics and Probability in Civil Engineering, Tokyo Japan*, 2010.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- C. Reed. *Earth Science: decade by decade*. Facts on File, Inc, 2008.
- V. D. Rubeis, P. Tosi, C. Gasparini, and A. Solipaca. Application of Kriging Technique to Seismic Intensity Data. *Bulletin of the Seismological Society of America*, 95(2):540–548, 2005.
- D. Sarma. *Geostatistics with Applications in Earth Sciences*. Springer, 2009.
- K. T. Shabestari and F. Yamazaki. A Proposal of Instrumental Seismic Intensity Scale Compatible with MMI Evaluated from Three-Component Acceleration Records. *Earthquake Spectra*, 17: 711–723, 2001.

- M. Trifunac and A. Brady. On the Correlation of Seismic Intensity Scales with the Peaks of Recorded Strong Ground Motion. *Bulletin of the Seismological Society of America*, pages 139–162, 1975.
- M. Trifunac and B. Westermo. A Note on the Correlation of Frequency-dependent duration of strong earthquake ground motion with the Modified Mercalli Intensity and the Geologic Conditions at the Recording Stations. *Bulletin of the Seismological Society of America*, pages 917–927, 1977.
- USGS. Did you feel it? <http://earthquake.usgs.gov/earthquakes/dyfi/>, 2011.
- USGS. California Map Society. <http://californiamapsociety.org>, 2012.
- H. Wackernagel. *Multivariate geostatistics: an introduction with applications*. Springer-Verlag Berlin Heidelberg, 2003.
- D. J. Wald and T. I. Allen. Topographic Slope as a Proxy for Seismic Site Conditions and Amplification. *Bulletin of the Seismological Society of America*, 97(5):1379–1395, 2007.
- D. J. Wald, V. Quitoriano, T. H. Heaton, and H. Kanamoria. Relationships between PGA, PGV, and MMI in California. *Earthquake Spectra*, 15(3):557–564, 1999a.
- D. J. Wald, V. Quitoriano, C. Serivner, and C. Worden. TriNet "ShakeMaps": rapid generation of instrumental ground motion and intensity maps for earthquakes in southern California. *Earthquake Spectra*, 15, 1999b.
- D. J. Wald, B. C. Worden, V. Quitoriano, and K. L. Pankow. Shakemap Manual: Technical Manual, Users Guide and Software Guide. *Advanced National Seismic System*, 2006.
- H. O. Wood and F. Neumann. Modified Mercalli Intensity Scale of 1931. *Bulletin of the Seismological Society of America*, 21(4):27–283, 1931.
- C. Worden, M. Gerstenberger, D. Rhoades, and D. Wald. Probabilistic Relationships Between Ground-Motion Parameters and Modified Mercalli Intensity, 2011.
- A. Zerva and V. Zervas. Spatial variation of seismic ground motions. *Applied Mechanics Reviews*, 55(3):271–297, 2002.

Appendix: R Functions

The **R** language for statistical computing is an open-source program; as such, it possesses many user-written packages designed to perform specific tasks. The functions within many add-on packages were used extensively during this research. The following list describes these packages and the functions used.

Package: `Hmisc`

Author: Frank E Harrell Jr , with contributions from many other users

Maintainer: Charles Dupont

Description: Contains many functions useful for data analysis, high-level graphics, utility operations, functions for computing sample size and power, importing datasets, imputing missing values, advanced table making, variable clustering, character string manipulation, conversion of S objects to LaTeX code, and recoding variables.

Functions

`cut2`: Given a data vector and a set of values at which to make cuts, returns a vector giving the bin into which each element of the data vector falls. Used to create averages over ground motion, MMI and distance bins.

Package: `gstat`

Author: Edzer Pebesma and others

Maintainer: Edzer Pebesma

Description: variogram modelling; simple, ordinary and universal point or block (co)kriging, sequential Gaussian or indicator (co)simulation; variogram and variogram map plotting utility functions.

Functions:

`gstat`: Creates `gstat` objects, which hold all the necessary information for geostatistical prediction.

`fit.variogram`: fit ranges and/or sills from a simple or nested variogram model to a sample variogram.

`predict.gstat`: Simple, ordinary and universal kriging, cokriging, point or block-kriging.

`variogram`: Calculates the sample variogram from data.

`vgm`: Generates a variogram model, or adds to an existing model.

Package: `fields`

Author: Reinhard Furrer, Douglas Nychka and Stephen Sain

Maintainer: Doug Nychka

Description: For curve, surface, and function fitting with an emphasis on splines, spatial data and spatial statistics.

Functions:

`image.plot`: Combines **R** image function with some automatic placement of a legend. `rdist.earth`:

Given longitude and latitude of two points, calculates the distance between them on Earth's surface.

Package: geoR

Author: Paul J. Ribeiro Jr.

Maintainer: Paulo J. Ribeiro Jr and
Peter J. Diggle

Description: Geostatistical analysis including traditional, likelihood-based and Bayesian methods.

Functions:

`lines.variomodel`: Adds a line with a variogram model specified by the user to a current variogram plot.

Package: maptools

Author: Nicholas J. Lewin-Koh and Roger Bivand

Maintainer: Roger Bivand

Description: Set of tools for manipulating and reading geographic data, in particular ESRI shapefiles.

Functions:

`readShapeSpatial`: Reads data from a shapefile into a `SpatialDataFrame` object.

Curriculum Vitae

Name: Andrea Sweny

Post-Secondary Education and Degrees: B.Sc. Mathematics, Honours - Queen's University, Kingston
2002-2006
M.A. Economics - University of Western Ontario, London
2006-2007

Honours and Awards: Graduate Teaching Assistant of the Year, Department of Economics
2007-2008
Ontario Graduate Scholarship
2008-2011

Conferences: Seismological Society of America, 2011
Canadian Geophysical Union, 2011

Related Work Experience: Teaching Assistant
The University of Western Ontario
2006 - 2010
Catastrophe Risk Analyst
Validus Research
2011