GROUND MOTION TO INTENSITY CONVERSION EQUATIONS (GMICEs) FOR TÜRKİYE: EVALUATION OF REGIONAL DIFFERENCES WITH PARAMETRIC AND NON-PARAMETRIC REGRESSION METHODS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

KUBİLAY ALBAYRAK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
EARTHQUAKE STUDIES

JANUARY 2023

Approval of the thesis:

GROUND MOTION TO INTENSITY CONVERSION EQUATIONS (GMICEs) FOR TÜRKİYE: EVALUATION OF REGIONAL DIFFERENCES WITH PARAMETRIC AND NON-PARAMETRIC REGRESSION METHODS

submitted by **KUBİLAY ALBAYRAK** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Earthquake Studies, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Ayşegül Askan Gündoğan
Head of the Department, **Earthquake Studies** _____

Prof. Dr. Ayşegül Askan Gündoğan
Supervisor, **Earthquake Studies, METU** _____

Assist. Prof. Dr. Fatma Yerlikaya Özkurt
Co-Supervisor, **Industrial Engineering, Atılım University** _____

Examining Committee Members:

Prof. Dr. Altuğ Erberik
Civil Engineering, METU _____

Prof. Dr. Ayşegül Askan Gündoğan
Earthquake Studies, METU _____

Assist. Prof. Dr. Fatma Yerlikaya Özkurt
Industrial Engineering, Atılım University _____

Assoc. Prof. Dr. Onur Pekcan
Civil Engineering, METU _____

Assoc. Prof. Dr. Mustafa Koçkar
Civil Engineering, Hacettepe University _____

Date: 26.01.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name: Kubilay Albayrak

Signature:

ABSTRACT

GROUND MOTION TO INTENSITY CONVERSION EQUATIONS (GMICEs) FOR TÜRKİYE: EVALUATION OF REGIONAL DIFFERENCES WITH PARAMETRIC AND NON-PARAMETRIC REGRESSION METHODS

Albayrak, Kubilay
Master of Science, Earthquake Studies
Supervisor: Prof. Dr. Ayşegül Askan Gündoğan
Co-Supervisor: Assist. Prof. Dr. Fatma Yerlikaya Özkurt

January 2023, 97 pages

Earthquakes cause damage to the built environment with the ground motions generated due to seismic energy release. Effects of earthquakes can be measured quantitatively through instrumental measures such as peak ground acceleration (PGA) and peak ground velocity (PGV) or qualitatively by macroseismic (felt) intensity levels. It is important to study the correlations between macroseismic intensity and instrumental ground motion parameters. Such relationships for Türkiye exist but they mostly have employed relatively limited datasets. In this study, three sets of data from Türkiye are employed: The first one is from the Aegean-Mediterranean Region, the second is from tectonic regions in Türkiye with dominantly strike slip mechanisms, and the third is the combination of these two datasets. These datasets are gathered to correlate the ground motion parameters with felt intensity levels, and to study potential regional differences. The entire dataset is composed of 69 earthquakes of which instrumental ground motion data and intensity

data are available. Initially, the relationships between Modified Mercalli Intensity (MMI) and log (PGA) as well as log (PGV) are studied with linear regression method using 3140 data pairs of MMI and PGA&PGV. Next, Principal Component Analysis (PCA) is performed for 2187 data points composed of magnitude (Mw), PGA (cm/s^2), PGV (cm/s), peak ground displacement (PGD) (cm), epicentral distance (km), significant duration (D_5_95), Arias intensity (m/s), focal depth (km), average 30-meter shear wave velocity (V_s30) (m/s), and the number of responses to select the parameters which most influence MMI levels. Based on the results of PCA, multiple linear regression is then performed with explanatory variable couples of PGA and epicentral distance as well as PGV and epicentral distance where MMI is the response variable. Finally, to study potential non-linearities in the data, the multivariate adaptive regression splines (MARS) method is used via piecewise linear functions. Not only the relationships are derived but also regional differences are captured with the analyses performed in this study. The presented equations can be used for ShakeMap applications and disaster management considerations in the future.

Keywords: Felt-Intensity, Parametric and Non-Parametric Regression methods, Türkiye, Regional Evaluation, Modified Mercalli Intensity

ÖZ

TÜRKİYE İÇİN YER HAREKETİNDEN ŞİDDETE DÖNÜŞÜM (YHŞD) DENKLEMLERİ: BÖLGESEL FARKLARIN PARAMETRİK VE PARAMETRİK OLMAYAN REGRESYON YÖNTEMLERİYLE DEĞERLENDİRİLMESİ

Albayrak, Kubilay
Yüksek Lisans, Deprem Çalışmaları
Tez Yöneticisi: Prof. Dr. Ayşegül Askan Gündoğan
Ortak Tez Yöneticisi: Dr. Öğr. Üyesi Fatma Yerlikaya Özkurt

Ocak 2023, 97 sayfa

Depremler, sismik enerji salınımı nedeniyle oluşan yer hareketleri ile çevreye zarar vermektedir. Depremlerin etkileri, niceliksel olarak maksimum yer ivmesi (MYİ) ve maksimum yer hızı (MYH) gibi aletsel ölçümlerle veya niteliksel olarak makrosismik (hissedilen) şiddet seviyeleriyle ölçülebilir. Makrosismik şiddet ile aletsel yer hareketi parametreleri arasındaki korelasyonları incelemek önemlidir. Belirtilen ilişkiler Türkiye için literatürde mevcuttur, ancak bu çalışmalarda nispeten sınırlı veri setleri kullanılmıştır. Bu çalışmada ise Türkiye'den üç veri seti kullanılmıştır: Birincisi Ege-Akdeniz Bölgesi'nden, ikincisi Türkiye'de doğrultu atım mekanizmalarının baskın olduğu tektonik bölgelerden ve üçüncüsü bu iki veri setinin birleşiminden oluşmaktadır. Bu veri kümeleri, yer hareketi parametrelerini hissedilen şiddet seviyeleriyle ilişkilendirmek ve potansiyel bölgesel farklılıkları tanımlayabilmek için oluşturulmuştur. Tüm veri seti, aletsel yer hareketi verileri ve şiddet verileri mevcut olan 69 depremden oluşmaktadır. İlk olarak, değiştirilmiş

Mercalli şiddeti (DMŞ) ile log (MYİ) ve log (MYH) arasındaki ilişkiler, her biri 3140 veri çiftinden oluşan DMŞ ve MYİ ve DMŞ ve MYH parametreleri kullanılarak doğrusal regresyon yöntemi ile incelenmiştir. Ardından, depremin büyüklüğü (Mw), MYİ (cm/s2), MYH (cm/s), maksimum yer değiştirme (MYD) (cm), dış merkez mesafesi (km), belirgin süre (D_5_95), Arias yoğunluğu (m/s), odak derinliği (km), ortalama 30 metrelik kayma dalgası hızı (Vs30) (m/s) parametrelerinden oluşan 2187 veri noktası dikkate alınarak MMI seviyesini en çok etkileyen parametreleri tanımlayabilmek için temel bileşen analizi (TBA) yapılmıştır. TBA'nın sonuçlarına dayanarak, MMI seviyesinin tanımlanabilmesi için açıklayıcı değişken olarak tanımlanan MYİ-dış merkez mesafesi ve MMI ile MYH-dış merkez mesafesi çiftleri ile çoklu doğrusal regresyon gerçekleştirilmiştir. Son olarak, verilerdeki potansiyel doğrusal olmama durumlarını incelemek için, parçalı doğrusal fonksiyonlar aracılığıyla çok değişkenli uyarlanabilir regresyon eğrileri yöntemi kullanılmıştır. Bu çalışmada yapılan analizlerle sadece şiddet ve yer hareketi parametreleri arasındaki ilişkiler türetilmemiş; aynı zamanda bölgesel farklılıklar da yakalanmıştır. Sunulan denklemler gelecekte ShakeMap uygulamaları ve afet yönetimi kararlarında kullanılabilecek niteliktedir.

Anahtar Kelimeler: Hissedilen şiddet, Parametrik ve Parametrik olmayan Regresyon Yöntemleri, Bölgesel Değerlendirme, Değiştirilmiş Mercalli Şiddeti (MMI)

To my family, especially to my grandfather

And to my advisor Prof. Dr. Ayşegül Askan Gündoğan

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor Prof. Dr. Ayşegül Askan Gündoğan, whose profound knowledge, support, and encouragement gave me precious insight into not only my future academic life but also the rest of my life. This thesis would not have been possible without her guidance and foresight. I am such a very lucky engineer to have an adviser like her.

I owe my sincere gratitude to Assist. Prof. Dr. Fatma Yerlikaya Özkurt for her invaluable guidance and supervision. Her contributions to my dissertation process were not only about knowledge but also about her kind attitude.

I owe a debt of gratitude to my grandfather Nazmi Albayrak; my father, Orhan Albayra; my mother, Cennet Albayra; my brother Çağatay Albayrak, and my sister Ezgi Albayrak for their unconditional support. They are always proud of me, and I am always proud of them. I know that love cannot disappear even if I lose my grandfather. I will always try to be just like what my grandfather told me to be.

I would like to thank Servet Özcan, Müge Akın, and Mutluhan Akın for their belief and support. I will always be grateful to you.

I also would like to thank Okan Koçkaya for acting like a brother for his support and guidance in my life.

Finally, I would like to thank the Scientific and Technological Research Council of Türkiye (TÜBİTAK) for its support.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

PGA	: Peak Ground Acceleration
PGV	: Peak Ground Velocity
PGD	: Peak Ground Displacement
Mw	: Moment Magnitude
Ia	: Arias Intensity
V_s30	: The time-averaged shear-wave velocity to a depth of 30 m
D_5_95	: Significant Duration
Sa	: 5% damped Spectral Acceleration
PSa	: 5% damped Pseudo Spectral Acceleration
PSv	: 5% damped Pseudo Spectral Velocity
SI	: Spectral Intensity
PGM	: Peak Ground Motion
CDI	: Community Decimal Intensity
MMI	: Modified Mercalli Intensity
EMS-98	: The European Macroseismic Scale
JMA	: The Japanese Meteorological Agency Scale
MCS	: The Mercalli-Cancani-Sieberg Intensity Scale
PCA	: Principal Component Analysis
MARS	: Multivariate Adaptive Regression Splines

GMICEs : Ground Motion to Intensity Conversion Equations

AFAD : Disaster and Emergency Management Presidency of Türkiye

USGS : The United States Geological Survey

DYFI : “Did you feel it?”

VIF : Variance Inflation Factor

AIC : Akaike’s Information Criterion

CHAPTER 1

INTRODUCTION

An earthquake causes sudden release of seismic energy resulting in seismic waves travelling through the earth. Effects of earthquakes can be measured quantitatively through instrumental measures such as peak ground acceleration (PGA), and peak ground velocity (PGV) or quantitatively by macroseismic (felt) intensity levels. Before the advance of digital seismometers, felt intensity was the sole expression of ground shaking levels and effects on buildings. Nowadays, it is important to study the correlations between macroseismic intensity and instrumental ground motion parameters since such correlations are critical inputs for disaster management and risk reduction purposes.

Macroseismic intensity levels which express the level of shaking can be defined in the field via reconnaissance surveys after earthquakes by experts or through online questionnaires such as “Did You Feel It?” system by USGS. (Wald et al. 1999-b, Boatwright and Phillips, 2012). There are alternative intensity scales to Modified Mercalli Intensity (MMI) in the literature, but the most common intensity scale is MMI followed by the European Macroseismic Scale (EMS-98), and the Japanese Meteorological Agency Scale (JMA).

Following the early studies of Giuseppe Mercalli, MMI was modified in 1931 by Wood and Neumann (Wood and Neumann, 1931). Since the original scale was not compatible with the ground motion-related intensity measurements, especially for higher intensity levels, the full-length scale was later modified by Stover and Coffman in 1993 (Stover and Coffman, 1993). Roman numerals are used to define the intensity levels, which is described in a range from undetectable shaking to catastrophic destruction. The intensity levels through MMI are in a range between I

and XII where X+ corresponds to total destruction. The detailed definitions of intensity levels are shown in Appendix A.

Although the responses for intensity level measurements are naturally subjective as they rely on human reporting, correlations between ground motion parameters with intensity levels are used to intervene quickly after an earthquake in order to define the areas of potential damage. ShakeMap applications are now in use to visually assess the spatial distributions of intensity levels for earthquakes all over the world. Such maps require well-defined correlations of the intensity levels with instrumental ground motion parameters. These correlations should be based on regional datasets since both intensity and ground motion parameters carry local characteristics. Such relationships for Türkiye exist (e.g.: Arıoğlu et al. 2001; Bilal and Askan, 2014) but these studies have mostly employed relatively limited datasets. In this study, novel ground motion to intensity conversion equations are derived using three sets of data from Türkiye along with parametric and non-parametric regression methods.

1.1 Objective and Scope of the Thesis

This thesis aims to study the relationships between ground motion parameters and macro seismic or felt intensity for earthquakes which occurred in Türkiye within the instrumental era. To fulfill this objective, in this study, three sets of data from Türkiye are employed: First one is from the Aegean-Mediterranean Region, second is from tectonic regions with dominantly strike-slip mechanisms and third is the combination of these two datasets. These datasets are gathered to correlate the ground motion parameters with felt intensity levels and study potential regional differences. The entire dataset is composed of 69 earthquakes of which instrumental ground motion data and intensity data are available. Alternative methods used in this thesis include simple linear regression, PCA, multiple linear regressions and MARS approaches.

The main objective of this study is to develop regression models to estimate MMI as a function of measured or computed ground motion parameters. These equations are named as ground motion to intensity conversion equations (GMICEs). In this study, the ground motion dataset is obtained from the Turkish National Strong Ground Motion Database operated by AFAD (www.tadas.afad.gov.tr) while the intensity dataset is gathered from USGS-DYFI. Intensity data available from field surveys of past earthquakes are not used in this study for sustaining the homogeneity of the USGS-DYFI database.

In Chapter 2, the literature survey is represented where the previous studies are reviewed in detail.

In Chapter 3, the available dataset that is used for this study is described. The databases of instrumental ground motion parameters and intensity are defined.

In Chapter 4, the methodologies used in this thesis are defined in detail. Linear regression, PCA, multiple linear regression and MARS models are discussed in detail.

In Chapter 5, the numerical results of this thesis are presented. The resulting GMICEs are compared with previous studies, and the regional differences are discussed in detail.

In Chapter 6, the summary and conclusions of this thesis are presented. The recommendations for further studies are also listed.

CHAPTER 2

LITERATURE REVIEW

2.1 General

Predictive relationships between macroseismic intensity scales and instrumental parameters are common worldwide. However, the most common correlations are based on MMI-PGA and MMI-PGV pairs. In this section, the previous studies are discussed briefly.

2.2 Review the Most Commonly Used Past Relationship between Felt Intensity and Instrumental Ground Motion Parameters

2.2.1 Trifunac and Bready (1975)

Trifunac and Brady (1975) carried out a correlation analysis between Modified Mercalli Intensity (MMI) and PGA, PGV, and PGD. The dataset in this study comprises 57 earthquakes and 187 strong motion accelerograms recorded in the Western United States. The analyses in that study are valid for the MMI values between V and VIII. Three local geological conditions of soft, intermediate, and hard are used to correlate MMI with peak ground motion values. PGA is used for stiffer soil conditions where PGV is employed for softer soil conditions.

2.2.2 Murphy and O'Brien (1977)

Murphy and O'Brien (1977) conducted a correlation analysis between Modified Mercalli Intensity and PGA. The dataset in that study comprises 875 data points measured from nearly 1500 strong motion accelerograms worldwide so the resulting

equations are available for use all over the world. The PGA values lower than 10 cm/s^2 PGA are not included in that study since the authors believe that this elimination reduces uncertainty of the dataset. The results show that PGA is not directly proportional to the intensity level and PGV is suggested to be used in the future for intensity correlations.

2.2.3 Wald et al. (1999-a)

Wald et al. (1999-a) conducted a regression analyses between Modified Mercalli Intensity and PGA & PGV. The dataset is composed of data from eight California earthquakes, which are the 1971 (M 6.7) San Fernando, the 1979 (M6.6) Imperial Valley, the 1986 (M 5.9) North Palm Strings, the 1987 (M 5.9) Whittier Narrows, the 1989 (M 6.9) Loma Prieta, the 1991 (M 5.8) Sierra Madre, the 1992 (M 7.3) Landers, and the 1994 (M 6.7) Northridge Earthquake due to their well-recorded regional strong-motion records with relatively higher intensity observations. The MMI-PGA relationship in that study is valid between MMI levels of V and VIII, and while the MMI-PGV relationship is valid between MMI levels of V and IX. It is suggested by the authors that the MMI-PGA correlation equation should be used for lower MMI values whereas MMI-PGV correlation equation for higher MMI values.

2.2.4 Atkinson and Sonley (2000)

Atkinson and Sonley (2000) conducted a correlation analysis between Modified Mercalli Intensity and 5% damped Pseudo Spectral Acceleration. The study dataset comprises 29 California earthquakes with a moment magnitude range from 4.9 to 7.4 and distance range of 1 to 300 km. The authors conclude that magnitude is also a dependent parameter of MMI at low frequencies of PSA while distance becomes important at higher frequencies.

2.2.5 Arıoğlu et al. (2001)

Arıoğlu et al. (2001) conducted a correlation analysis between Modified Mercalli Intensity and PGA. The study database is composed of 14 strong ground motion records of the 17 August 1999 Kocaeli Earthquake. Although this study is valuable being the first local ground motion to intensity conversion equation for Türkiye, the results are not considered to be sufficiently robust and general due to the limited strong ground motion dataset.

2.2.6 Boatwright et al. (2001)

Boatwright et al. (2001) conducted a correlation analysis between Modified Mercalli Intensity with PGA, PGV, and PSv ordinates at 14 different period values ranging between 0.1 to 7.5 seconds. The database comprises data from 66 free-field strong ground motion stations that recorded the 1994 Northridge Earthquake (Mw 6.7). The authors conclude that PGV and average PSv are better correlated with MMI than PGA. Additionally, the MMI-PGV correlation provides lower uncertainty than the average MMI-PSv correlation.

2.2.7 Karim and Yamazaki (2002)

Karim and Yamazaki (2002) conducted a correlation analysis between Japan Meteorological Agency (JMA) seismic intensity with PGA, PGV, and SI. The database is composed of two datasets. The first dataset includes 13 major earthquakes, which occurred in Japan, the United States, and Taiwan, with 879 records obtained at sites located on non-liquefied soils. Second database includes 7 seven major earthquakes, which occurred in Japan and the United States, with 17 records obtained at sites located on liquefied soils. A two-stage linear regression analysis is used in this study. The correlation coefficient is found to be the highest for SI, and for multivariate analysis, the SI-PGA pair has a higher correlation

coefficient than the PGA-PGV pair. Additionally, PGV is concluded to be the lowest correlation parameter for JMA intensity estimations.

.

2.2.8 Wu et al. (2004)

Wu et al. (2004) conducted correlation analyses between Modified Mercalli Intensity and PGA, Sa at 1 second, and PGV. The database is composed of 30000 strong ground motion records. The authors conclude that PGA and Sa have a higher correlation with MMI than PGV. However, PGV and Sa at 1s have a better correlation with MMI than PGA, which is not observed to be stable for small earthquakes.

2.2.9 Kaka and Atkinson (2004)

Kaka and Atkinson (2004) conducted a correlation analysis of the Modified Mercalli Intensity with PGV, and Sa at 1, 5, and 10 Hz. The database comprises post-1982 earthquakes from eastern North America, with moment magnitude values ranging between 3.6 to 7.25. The study employs 232 MMI-PGV data pairs and 199 MMI-PSA pairs. For regression analyses, linear least squares regression method is used in this study. The authors conclude that PGV is the best parameter to estimate MMI values for ShakeMap applications due to the lowest uncertainty in predictions.

2.2.10 Atkinson and Kaka (2006)

Following their previous study, Atkinson and Kaka (2006) later conducted a correlation analysis between Modified Mercalli Intensity and PGA, Sa at 0.5, 1, and 3.3 Hz, and PGV. The database comprises 22 New Madrid Seismic Zone earthquakes calibrated from California ShakeMap/DYFI data. The authors conclude that the MMI-PGV correlation is the most reliable one. The uncertainty due to magnitude

dependency of PGV and PSa (at 0.5 and 1 Hz) as well as the distance dependency of PGA and PSa (3.3 Hz) are minimized by including these trends in correlation equations.

2.2.11 Atkinson and Kaka (2007)

Atkinson and Kaka (2007) conducted a correlation analysis between Modified Mercalli Intensity and PGA, PGV, and Spectral Acceleration at 0.5, 1, and 3.3 Hz. This study provides two equations for lower and higher value of 0.48 for log (PGV). The database comprises the Central United States Region and the earthquakes, calibrated from California ShakeMap/DYFI Data. The authors conclude that the refined results based on PGV are a reasonable choice for predictive equations.

2.2.12 Tselentis and Danciu (2008)

Tselentis and Danciu (2008) analyzed the correlation between Modified Mercalli Intensity with PGA, PGV, PGD, Arias Intensity, and Cumulative Absolute Velocity. The database is composed of 89 Greek earthquakes with 310 ground motion records from the earthquakes which occurred between 4 November 1973 and 7 September 1999. The characterizations of the ground motions are defined as short duration, low-energy content, and high frequency. In this study, two empirical relationships are developed. For the first relationship, the magnitude and epicentral distance are excluded for MMI prediction. For the second relationship, the entire dataset is used for the calculations. The authors conclude that Arias Intensity exhibits the lowest standard deviation in the first model while PGA shows similar performance in the second model. The local site effects, which are included in the calculation with a dummy variable, slightly affects the prediction of MMI. The magnitude and epicentral distance are the main dependent variables with PGA to estimate MMI.

2.2.13 Faenza and Michelini (2010)

Faenza and Michelini (2010) conducted a correlation analysis between Mercalli-Cancani-Sieberg (MCS) Intensity and PGA as well as MCS and PGV. The orthogonal distance regression technique is used in this study. The database is composed of 66 earthquakes between 1972 and 2004 within the range of 3.9 to 6.9-moment magnitude with 266 pairs of Intensity-PGM data. The authors conclude that the correlation analysis does not depend on earthquake magnitude and source-to-site distance.

2.2.14 Yaghmaei-Sabegh, Tsang, and Lam (2011)

Yaghmaei-Sabegh, Tsang, and Lam (2011) conducted correlation analyses between Modified Mercalli Intensity and PGA, PSa, and PGV. The dataset comprises data from 10 earthquakes between 1978 and 2003 recorded in the Iranian Plateau with a moment magnitude range of 4.4 to 7.4. The authors conclude that the MMI-PGV correlation has the lowest residual variance.

2.2.15 Alvarez et al. (2012)

Alvarez et al. (2012) conducted a correlation analysis between macroseismic intensity and PGA, PGV, moment magnitude, and epicentral distance. Alternative methods including Support Vector Regression, Multilayer Perceptrons, and Genetic Programming are employed in this study. The dataset is composed of ground motions from the 1989 Loma Prieta earthquake, the 1992 Petrolia earthquake, and the California earthquakes which occurred after 2000, with a total of 843 ground motion records from 63 earthquakes. The authors conclude that the Multilayer Perceptron method is the most effective approach for the nonlinear correlation of macroseismic intensity and ground motion parameters.

2.2.16 Bilal and Askan (2014)

Bilal and Askan (2014) conducted a correlation analysis between Modified Mercalli Intensity and PGA as well as PGV. The database comprises 14 earthquakes with a moment magnitude range from 5.7 to 7.4 forming 92 MMI-ground motion pairs. The linear least regression method is used in this study to present two sets of equations. The first set of equations is simple regression between MMI and PGA, PGV, as well as PSa at ordinates of 0.3, 1, and 2 seconds. The second set is the refined equations with addition of moment magnitude and epicentral distance to the independent variables. The authors conclude that PGA is the main ground motion parameter to predict MMI values for rigid structures while PGV correlates well with ductile structures. This study is the first systematic GMICEs for Türkiye and is still in use by international agencies to estimate MMI levels after earthquakes in Türkiye.

2.2.17 Caprio et al. (2015)

Caprio et al. (2015) conducted a correlation analysis between Modified Mercalli Intensity and peak ground motion parameters, magnitude, and hypo central distance. In this study, the correlation equations are presented for California, Greece, Italy, New Madrid region in CEUS, and worldwide as well as altogether to signify potential regional differences. The total dataset is composed of 2380 observations for these regions. The MMI-PGM correlation pairs are arranged to be within a maximum distance of 2 kilometers between the recorded ground motion stations and the responses. The dataset comprises earthquakes with a magnitude range between 2.5 and 7.3 recorded between 1965 and 2005 from worldwide recorded ground motions. The orthogonal regression technique is used in this study. The authors conclude that MMI-PGV and MMI-PGA correlation pairs have regional dependencies for magnitude and distance terms. It is also observed that the intensity scale variations also affect the regional differences. The authors suggested to use regional relationships whenever models can be calibrated regionally.

2.2.18 Du et al. (2020)

Due et al. (2020) conducted a correlation analysis between Modified Mercalli Intensity and PGA as well as PGV by novel probabilistic relationships. These relationships assume that peak ground motion values are randomly distributed for each MMI value following a normal distribution. The database is composed of 37 earthquakes that occurred in Western China between 1994 and 2017. Intensity maps are plotted with the seismic intensity values calculated by the Bayesian formula, in addition to standard deviation maps which are constructed through spatial interpolation. The authors conclude that the probabilistic approach performs better than the traditional methods in rapid estimation of earthquake intensities.

2.2.19 Ahmadzadeh et al. (2020)

Ahmadzadeh et al. (2020) conducted a correlation analysis between Modified Mercalli Intensity and PGA as well as PGV. The database is composed of 23 earthquakes with moment magnitude range from 5.1 to 7.3 recorded between 1977 and 2017 in Iran. Least squares regression method is used in this study. The authors conclude that in the first model, magnitude and distance are excluded, but the residual analysis pointed out the dependencies on these parameters. Then, refined predictive models are derived which cover these parameters in the study. The predictive equations of this study have standard errors between 0.5 to 2 intensity units.

2.2.20 Ardeleanu et al. (2020)

Ardeleanu et al. (2020) conducted a correlation analysis between Medvedev-Sponheuer-Karnik Intensity Scale and PGA, PGV, PGD, Cumulative Absolute Velocity, Arias Intensity, and Destructiveness Potential Factor. This study is valid for intensity levels between V and VIII in the Carpathian bend zone. The database is

composed of 5 events with Mw ≥ 6 that occurred in the Vrancea region during the half past century. Simple least squares regression method is used in this study. The authors conclude that the differences between this study and previous studies are due to the different data sets. Although the dataset is limited, the linear regression method is observed to be effective. Additionally, when the magnitude, distance, and geological site conditions are excluded, Arias Intensity is found to be the most stable predictor, and PGA has the most significant standard deviation.

2.2.21 Tao et al. (2020)

Tao et al. (2020) conducted a correlation analysis between macroseismic intensity and sixteen different ground motion parameters including PGA, PGV, Arias Intensity, Housner Intensity, Acceleration Spectrum Intensity, Velocity Spectrum Intensity, and others. Support vector analysis is used as the main method in this study. The dataset comprises 25 pairs of macroseismic intensity values and ground motion parameters. The support vector regression results are valid for macroseismic intensity values between VI to IX due to the high seismicity rate in the study region. The authors conclude that PGA is the most critical parameter, and PGV is the fifth important parameter among the nine parameters examined in this study. Different than other studies with linear regression methods, in this study PGA is reported to perform better than PGV in predicting macroseismic intensity. The Gaussian Kernel Support Vector Analysis provides better accuracy percentage, correlation coefficient, and performance measures of minimum average mean squared error than the linear methods. Additionally, magnitude and epicentral distance are found to be insignificant.

2.2.22 Gomez-Capera et al. (2020)

Gomez-Capera et al. (2020) conducted regression analyses between macroseismic intensity and PGA, PGV, and PSa at 0.2, 0.3, 1, and 2 seconds. The dataset comprises

67 earthquakes within a magnitude range from 4.2 to 6.8 between 1972 and 2016, resulting in 240 macroseismic intensity-ground motion correlation pairs. The MMI and ground motion data are paired within a range of 2 kilometers. The nonlinear form of the logarithmic method is used to predict the intensities levels. The study is valid for II ≤ macroseismic intensity ≤ X–XI and geometric values within a range 0.9 ≤ PGA (cm/s2) ≤ 587. The uncertainty is around 1 unit for intensity values. Based on to the residual analysis, the authors conclude that the regressions are not significantly dependent on moment magnitude and epicentral distance. The lowest uncertainty value is obtained for PGV, so it is concluded to be the best parameter to predict the intensity for their dataset

CHAPTER 3

AVAILABLE DATABASE

3.1 General

In this study, ground motion to intensity conversion equations are studied for three regions of Türkiye, which are named as “Aegean-Mediterranean region”, “Strike-Slip region in Türkiye” and “Türkiye” which is the combination of these datasets. The entire dataset is comprised of 69 earthquakes, which are presented in detail in Appendix B, with 3140 MMI-peak ground motion parameter pairs for simple linear regression and 2187 MMI-ground motion parameter data points for multiple linear regression and MARS method. Instrumental ground motion parameters, which are moment magnitude, peak ground acceleration, peak ground velocity, peak ground displacement, focal depth, epicentral distance, significant duration and Arias intensity are obtained from Disaster and Emergency Management Presidency of Türkiye (AFAD) while the corresponding intensity values The United States Geological Survey (USGS) “Did you feel it?” (DYFI) system. These institutional databases, study areas as well as the ground motion parameters used in this study will be introduced in this chapter in detail.

3.2 AFAD Database

The AFAD database is Türkiye's official national strong ground motion database, with data from approximately 750 stations (https://tadas.afad.gov.tr/). It provides information about recorded earthquakes and the corresponding strong ground motion stations which recorded the events as well as the instrumental ground motion parameters. Moment magnitude (Mw), and focal depth (km) of an earthquake as well

as epicentral distance (km), peak ground acceleration for all directions (PGA, cm/s^2), and 30-meter average shear wave velocity (V_s30, m/s) of the selected stations are readily available parameters at the AFAD database. For the other instrumental ground motion parameters of peak ground velocity (PGV, cm/s), peak ground displacement (PGD, cm), Arias intensity (m/s), and effective duration (D_5_95, second), an opensource MATLAB software code by Carlton (2005) was used. The lack of V_s30 value at some of the strong ground motion stations in Türkiye reduced the number of data points significantly from 3114 to 2171.

The faults in Türkiye as well as the earthquake distributions of the entire dataset are presented in Figure 3.1. in Figure 3.2, respectively.

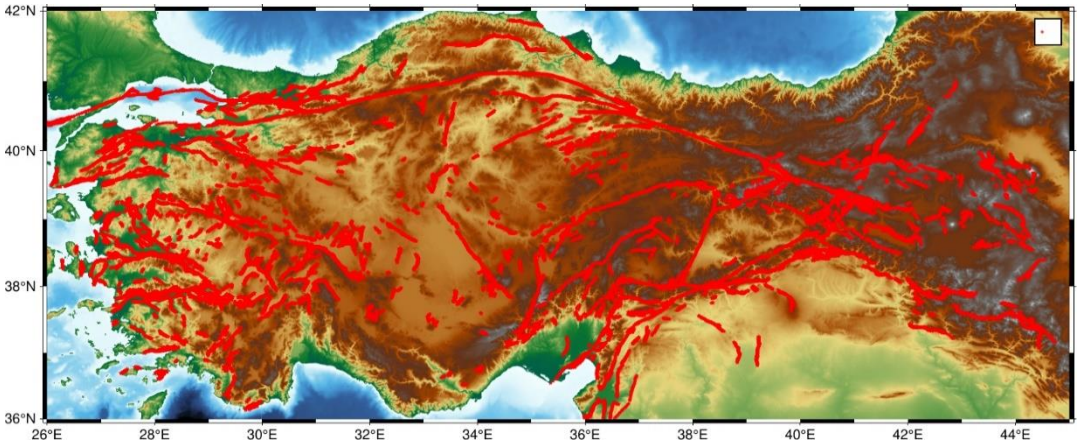


Figure 3.1 The faults in Türkiye

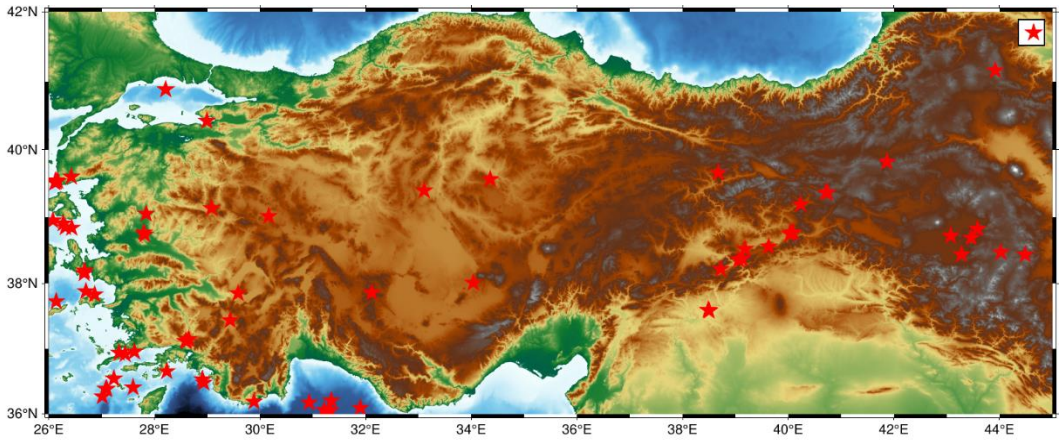


Figure 3.2 The earthquakes used in this study

3.3 USGS-DYFI Database

USGS has a system called “Did you feel it?”, which is based on an internet-based questionnaire to obtain the felt intensity levels after an earthquake from the public. The link to the DYFI questionnaire is https://earthquake.usgs.gov/earthquakes/eventpage/tellus. DYFI page provides information on the intensity map, the plot of intensity as a function of distance, the plot of responses as a function of time, and a table of DYFI responses for each earthquake. The responses to the questionnaire are evaluated based on the Community Decimal Intensities (CDI), which state the earthquake effects over an area as a single intensity level, provided by Dengler and Dewey (1998). The CDI is an aggregate of the weighted sums of the various indices of the DYFI questionnaires composed of eight questions with weights and ranges. The details are presented in Appendix C. The calculations are defined by Wald et al. (2012) as follows:

1. Each answer is turned into a numeric value from 0 to 1.
2. The averages of all answers are calculated within specified community.
3. The community weighted sum (CWS) is formed by taking the weighted sum of all averages.

4. DYFI intensity level is calculated as follows:

$$CDI=3.40 \ln(CWS)-4.38 \quad (3.1)$$

where CDI is rounded off to the first decimal place.

After estimating the regional intensities, DYFI maps are created within minutes for a magnitude 1.9 or greater earthquake. Although DYFI maps are based on direct reports of earthquake effects by the affected people, ShakeMaps are primarily based on point location measurements of the ground motion parameters with seismic intensity levels (Worden et al., 2010). Since DYFI data is compatible and consistent with ShakeMaps, DYFI intensities are used to calibrate the equations used by ShakeMap to convert ground motions into intensity levels (Boatwright and Philips, 2012).

In this study, the selected earthquakes from the AFAD strong motion database are matched with the USGS database to exclude the ones that are not recorded by AFAD stations to prevent instrumental ground motion parameter differences. Thus the USGS database is only used for intensity levels. For data pairs, the instrumental ground motion parameters are paired with the number of responses as well as intensity levels within ± 5 kilometers distance between ground motion station and the DYFI response point. An example of MMI levels and the number of responses from DYFI system of Elazığ Sivrice Earthquake, 2020 is presented at Appendix D.

3.4 Study Areas

In this study, three regions, which are Aegean-Mediterranean, Strike-Slip, and Türkiye as the entire region, are studied to define probable regional differences as well as the relationships between felt intensity and instrumental ground motion parameters. The description of Strike-Slip region is based on dominant focal mechanism of earthquakes. The Aegean-Mediterranean region is defined according

to density of earthquakes and the geographical region. Mostly normal and reverse mechanisms are observed in this region.

The descriptive statistics for available datasets according to defined methods are presented in Appendix E. Figure 3.3, show the boundaries of the regions in this study while Table 3.1 and Table 3.2 list the city names in the regions.

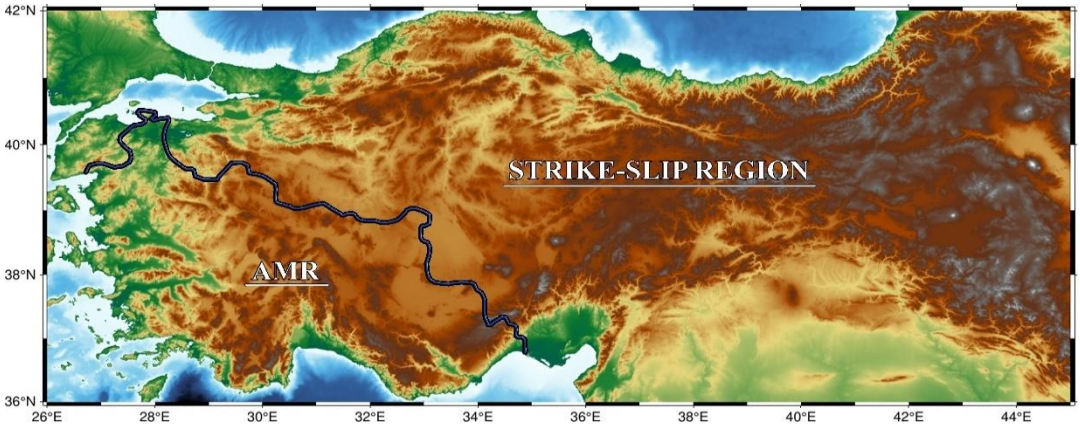


Figure 3.3 Boundaries of regions used in this study

Table 3.1 Cities that are used in the Aegean-Mediterranean Region database

Balıkesir	Manisa	Kütahya	Afyon	Izmir
Aydın	Denizli	Burdur	Muğla	Antalya
Konya	Mersin	Uşak	Isparta	

Table 3.2 Cities that are used in Strike-Slip Region database

Adana	Adıyaman	Ağrı	Ankara	Bilecik
Bingöl	Bursa	Çanakkale	Çorum	Diyarbakır
Edirne	Elazığ	Erzincan	Erzurum	Eskişehir
Gaziantep	Giresun	Hatay	Istanbul	Kayseri
Kırklareli	Kocaeli	Malatya	Kahramanmaraş	Mardin
Ordu	Rize	Sakarya	Samsun	Sivas
Tekirdağ	Tokat	Trabzon	Tunceli	Şanlıurfa
Van	Yozgat	Batman	Yalova	Osmaniye
Düzce				

3.5 Ground Motion Parameters Used in This Study

The instrumental ground motion parameters that are used in this study are Moment Magnitude (Mw), Epicentral Distance (km), Focal Depth (km), Peak Ground Acceleration (PGA, cm/s^2), 30-meter average shear wave velocity (V_s30, m/s), Peak Ground Velocity (PGV, cm/s), Peak Ground Displacement (PGD), Arias Intensity, and Effective Duration (D_5_95).

3.5.1 Moment Magnitude (Mw)

The moment magnitude is a quantitative measure of an earthquake magnitude developed by Kanamori and Hanks (1979). The calculation of moment magnitude is based on the seismic moment (M_0), which considers the fault geometry, material rigidity at the fault level, and fault displacement for calculating the energy release caused by an earthquake. Therefore, moment magnitude is indeed the only magnitude scale to measure earthquake magnitudes most reliably. The moment magnitude is defined in terms of the seismic moment as follows:

$$M_W = \frac{2}{3} \log M_0 - 10.7 \quad (3.2)$$

$$M_0 = D\, A\, \mu \quad (3.3)$$

where M_0 is the seismic moment, D is the average fault displacement, A is the total area of the fault surface, and μ is the average rigidity.

3.5.2 Epicentral Distance and Focal Depth

Epicentral distance is defined as the distance on the ground surface between the site and the focus of an earthquake, which is the point that an earthquake occurs on the seismic source. Focal depth is the closest distance between the focus and the ground

surface. Figure 3.4 shows the visual descriptions of epicentral distance and focal depth.

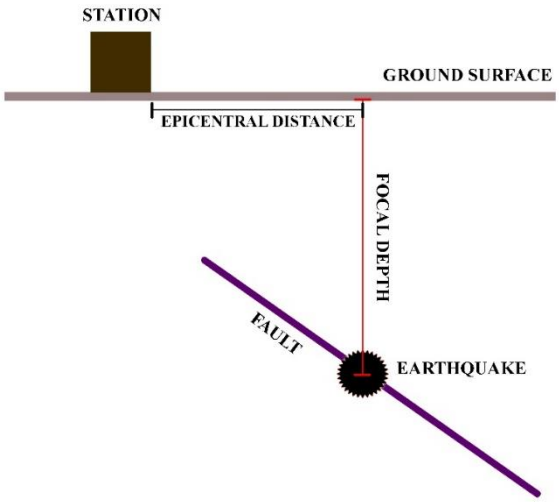


Figure 3.4 Visual description of epicentral distance and focal depth

3.5.3 Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV), and Peak Ground Displacement (PGD)

Peak ground acceleration, peak ground velocity, and peak ground displacement are defined as the maximum acceleration, velocity, and displacement of the ground due to strong ground motion of an earthquake, respectively. Since the earthquake shaking occurs in three directions, these parameters are expressed in three directions, which are North-South, East-West, and Up-Down. In this study, the pre-processed strong ground motion records of AFAD are used to compute PGA, PGV, and PGD in horizontal directions. Then, the parameters are defined as a single PGA, PGV, and PGD value for each strong motion record by taking geometric means of the corresponding horizontal components.

3.5.4 30-meter Average Shear Wave Velocity (V_s30)

V_S30 is defined as the time-averaged shear-wave velocity to a depth of 30 m. As of now, it is the globally accepted site proxy which provides unambiguous definitions of site classes for building codes and site coefficients for site dependent response spectra.

3.5.5 Arias Intensity (I_a)

Arias Intensity is an indicator for the energy content of ground motions including both the duration and the amplitude of the whole ground motion time history (Arias, 1970). The principal assumption of Arias Intensity is that the amount of damage experienced by a structure is proportional to the energy dissipated by the structure per unit weight during the overall duration of the earthquake-induced motion (Arias 1970).

The Arias Intensity is defined as follows:

$$I_a = \frac{\pi}{2g}\int [a(t)]^2 dt \qquad (3.4)$$

where a(t) is the acceleration in m/sec^2, g is the acceleration of gravity, and I_a is the Arias intensity in m/sec.

3.5.6 Significant Duration (D_5_95)

Significant duration is the time between 5% and 95% of Arias Intensity (Trifunac and Brady,1975). It measures the damage potential during an earthquake by Husid plots, which show the buildup of the energy of an accelerogram with time and the time interval for 5 to 95 percent energy buildup. Local site conditions, distance from the station and the source, and fault characteristics are the main parameters that affect

the significant duration. Figure 3.5 shows the computation of Significant Duration from Husid Graph.

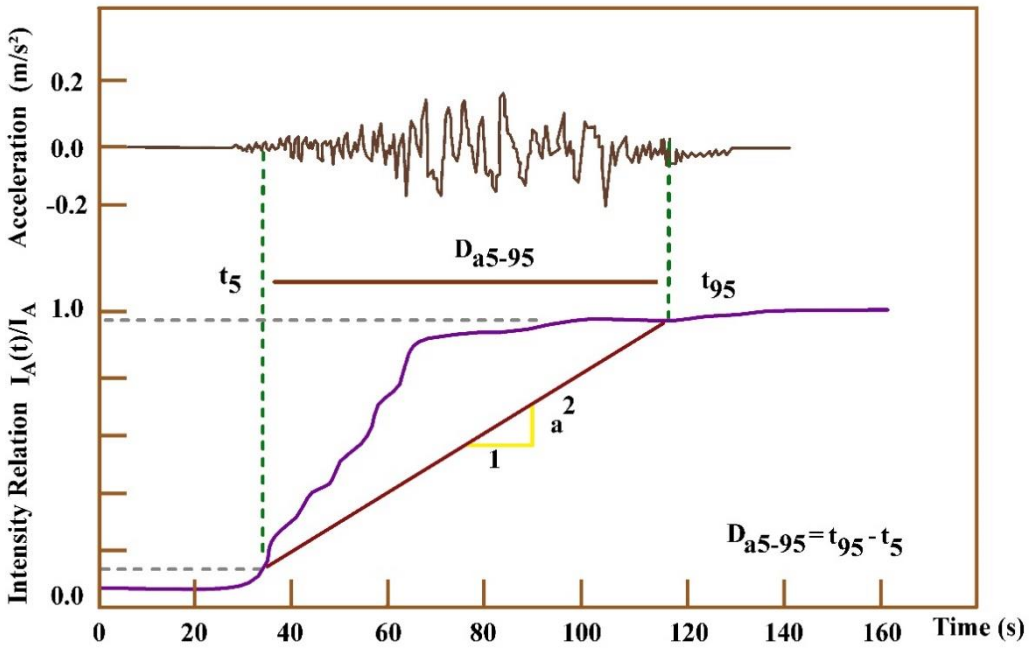


Figure 3.5 The computation of the significant duration from Husid Graph (modified from Husid,1969)

CHAPTER 4

METHODOLOGY

4.1 General

In this thesis, simple linear regression, multiple linear regression, principal component analysis, and multivariate adaptive regression splines are used as the main statistical approaches. The objectives are twofold: First one is to select the most influential parameters on intensity levels and second is to define the conversion relationships using the selected parameters between each pair of datasets in terms of linear considerations.

4.2 Linear Regression

Linear regression, which is a parametric regression method, is used to estimate the relationships between a dependent variable (also called response or outcome variable) and one or more independent variables (also called predictors or explanatory variables) in terms of their linear combinations. The regression analysis has a descriptive purpose and a predictive purpose. The descriptive purpose is to derive a relationship between the predictor and the response variables. The predictive purpose is to estimate the response variable based on the value of predictors. Regression analysis aims to fit a model to an input data by minimizing the differences between the actual response variable and the estimated response variable. The response variable is estimated by the best fit of a line. The error is defined in terms of least squares, where the squared difference between the actual and modeled values are minimized. Least square minimization is used both for simple and multiple linear regression models.

This study uses simple linear regression to obtain relationships between log (PGA)-MMI and log (PGV)-MMI pairs. The dataset is composed of 3140 data pairs for each correlation. For multiple linear regression, MMI is obtained by the predictor variables of PGA&epicentral distance and PGV&epicentral distance pairs. The dataset is composed of 2187 seismic parameter pairs for each correlation.

4.2.1 Methodology

Multiple linear regression involves the prediction of the response variable (Y) based on two or more predictor variables ($X_1, X_2, \ldots, X_p$). The model form is defined as follows:

$$Y_i = B_0 + B_1 X_{1^i} + B_2 X_{2i} + \cdots + \epsilon_i \qquad (4.1)$$

where

Y_i: the i^{th} $(i = 1, 2, \ldots, n)$ response value

B_0:the intercept of the least-squares regression line on the response axis

B_j: the regression slope for the j^{th} $(j = 1, 2, \ldots, p)$ independent variable

X_{ji}: the i^{th} value of the j^{th} independent variable

ϵ_{i}: the i^{th} random error value

Based on a normal distribution of the residuals, the linear form of the simple linear regression is rearranged as follows:

$$Y = Normal(b_1 X + b_0, \epsilon) \qquad (4.2)$$

where b_1 is the slope of the least-squares regression line, b_0 is the intercept and ϵ is the standard deviation of the variation of Y.

The coefficient of variation (R^2), which determines how well the dataset is fitted through the regression line, is the proportion of variance in the response variable (Y) that is predicted by the independent variable. Since R^2 is the proportion of variance, it takes values between 0 and 1, where 1 is the theoretical best fit of data to regression line and 0 means the data does not fit properly. The coefficient of variation is defined as follows:

$$R^2 = 1 - \frac{SSE}{TSS} \quad (4.3)$$

TSS: the total sum of squares and defined as follows:

$$TSS = \sum_{i=1}^{k} (Y_i - \bar{Y})^2 \quad (4.4)$$

$\bar{Y}$: the mean of the response variable

SSE: the sum of squares of errors and defined as follows:

$$SSE = \sum_{i=1}^{k} \left(Y_i - \widehat{Y}_i\right)^2 \quad (4.5)$$

$\widehat{Y}_i$: the i^{th} $(i = 1, 2, \dots, N)$ estimated value of the response variable and it is defined as follows:

$$\widehat{Y}_i = m\, X_i + n \quad (4.6)$$

The standard error of the estimated response parameter (*Y)*, σ, is defined in Equation 4.7.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} (Y_i - \hat{Y})^2}{(k-2)}} \quad (4.7)$$

The standard deviation of the residuals or error terms is assumed to be equivalent to the standard error of the estimated response parameter. This is only valid for normal distribution of variables. Additionally, the correlation coefficient or Pearson's R is equal to the square root of the coefficient of variation.

4.3 Principle Component Analysis (PCA)

Principle component analysis (PCA) is a technique where high dimensional data is represented in a lower dimensional form while preserving the maximum amount of information. The methodology is based on a linear transformation of an initial dataset from n dimensional space to another space that has the same number of dimensions, which is composed of principal components that are completely uncorrelated and orthogonal or perpendicular to each other (Wang, 2008).

Principal components are obtained by the maximization of the variances that is assumed to contain the largest information of the initial dataset. Hence, the first principal component has the largest variance and the consecutive principal components have lower variances. On one hand they do not have any real meanings since they are constructed as linear combinations of the initial variables, on the other hand the contributions of variables for each principal component state the most influential parameter.

In this study, PCA is performed to reduce the complexity and the computational cost of multiple linear regression by defining the most influential variables that are completely uncorrelated to each other.

4.3.1 Methodology

PCA is a matrix operation-based computing method, therefore the dataset is formed in a matrix, where rows and columns are composed of the number of samples and variables, respectively. A sample matrix composed of p columns and n rows is defined as follows:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \quad (4.8)$$

where, x_{ij} is the predictor data, i is the index for the variables i= 1, 2…., n, and j is the index for the numerical value of the variables for the defined response variable and j=1, 2…., p.

PCA is a sensitive analysis. Thus, the initial dataset has to be analyzed on similar measurement scales. Therefore, to prevent the scaling effects especially for lower variance principal components, standardization of initial dataset is the most important step before analysis. Additionally, if all the variables used in PCA are measured on the same scale, the centering of the data, which is subtracting the mean from each variable, is used instead of standardization.

Standardization is defined as subtracting the mean values from each variable and dividing the result by the standard deviation. The mathematical formula is defined as follows:

$$X_{std} = \begin{bmatrix} (x_{11} - \overline{x_1})/\sigma_1 & \cdots & (x_{1p} - \overline{x_p})/\sigma_p \\ \vdots & \ddots & \vdots \\ (x_{n1} - \overline{x_1})/\sigma_1 & \cdots & (x_{np} - \overline{x_p})/\sigma_p \end{bmatrix} \quad (4.9)$$

where $\boldsymbol{X}_{std}$ is the standardized value matrix, x_{ij} is the observed value, $\overline{x_j}$ is the mean of jth variable, and σ_j is the standard deviation of the jth variable. The standardized data has a mean of zero and a standard deviation of one for each variable. The $\overline{x_j}$ is computed as follows:

$$\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \forall\, j \quad (4.10)$$

Additionally, the square of the standard deviation is defined as variance, and after the standardization of the variables, each variable has a variance of one. The variance is defined as follows:

$$var_j = {\sigma_j}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})^2, \forall\, j \quad (4.11)$$

The relationship between parameters is defined by the correlation matrix, which is obtained by the matrix operations. The correlation matrix from standardized dataset is defined as follows:

$$X_{cor} = \frac{1}{n-1} X_{std}^T X_{std} = \begin{bmatrix} 1 & \cdots & c_{1p} \\ \vdots & \ddots & \vdots \\ c_{p1} & \cdots & 1 \end{bmatrix} \quad (4.12)$$

where, c_{ij} is the correlation coefficient between x_i and x_j parameters for $(i = 1, 2, \dots, p)$ and $(j = 1, 2, \dots, p)$.. The correlation coefficient is defined as follows:

$$c_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} = \frac{\sum_{i=1}^{n}[(x_{ij} - \overline{x_j})(x_{in} - \overline{x_n})]}{\sqrt{\sum_{i=1}^{n}(x_{ij} - \overline{x_j})^2}\sqrt{\sum_{i=1}^{n}(x_{im} - \overline{x_m})^2}} \quad (4.13)$$

where, σ_{ij} is the population covariance, σ_i and σ_j are the standard deviations of the variables. The diagonal elements of the correlation matrix are 1 due to standardization and it is defined as the *p* x *p* symmetric positive matrix.

Since PCA is the linear combination of the variables, the directions of variables and variance differences are obtained through eigenvector and eigenvalue analyses, respectively. While eigenvectors are composed of directions and magnitudes for each defined variable, PCA uses only one eigenvector as the combination of variables in terms of contribution coefficients. For the variance difference as stated before, eigenvalues determine variances for each principal component, where the largest variance signifies the first principal component as well as the others are sequenced likewise.

To calculate the eigenvalues and eigenvectors as well as elimination of redundant elements, the correlation matrix has to be diagonalized as follows:

$$X_D = K^T X_{cor} K = diag(\lambda_1 \dots \lambda_p) \quad (4.14)$$

where, *K* is the rotation matrix and X_{cor} is the correlation matrix.

The uncorrelated variables are defined as vectors in the rotated space as follows:

$$yi = K^T x_i,\ xi = K y_i, \forall\, j \qquad (4.15)$$

The eigenvalues and eigenvectors are computed as follows:

$$(X_D - \lambda_j I) v_j = 0\ for\ j = 1,2 \ldots, p \qquad (4.16)$$

where, I is the identity matrix, λ_j is the eigenvalue and v_j is the eigenvector defined for each variable. The solution of the characteristic equation defines the principal components of the correlation matrix when ordered from the highest to the lowest eigenvalues as follows:

$$X_D v_j = \lambda_j v_j\ for\ j = 1,2 \ldots, p \qquad (4.17)$$

$$\det (X_D - \lambda I) = 0 \qquad (4.18)$$

where, p is the number of eigenvalues.

After the calculations of eigenvalues and eigenvectors, principal components can be arranged according to feature extraction and feature selection. The former considers each initial variable equally and the latter is based on the selection of most important variables. In this study, feature extraction is performed with classical methods, which use eigenvalues as variance indicators. The first classic method is Kaiser Rule (Kaiser, 1960), which retains all the principal components that have eigenvalues equal or higher than 1. Since the highest eigenvalue gives the highest variation, the lower eigenvalues can be ignored by this approach (Bohm and Zech, 2010). The other method is the percentage-based variance comparison with respect to the total variance. The sum of eigenvalues of variables is defined as the total variance where the percentage of each component's variance is computed as follows:

$$\mathrm{PC_n} = [\lambda_j / (\textstyle\sum_{j=1}^{p} \lambda_j)]^*100 \qquad (4.19)$$

where, $\mathrm{PC_n}$ is the n^{th} principal component.

4.4 Multivariate Adaptive Regression Splines

The previous regression methods are based on parametric statistics. Parametric statistics do not handle the non-linearities between independent variables. Thus, in this study, multivariate adaptive regression splines (MARS), one of the available non-parametric approaches, is employed to handle potential non-linearities between independent variables using piecewise linear functions (Friedman, 1991). Although the response variable is highly biased as in this study, MARS can offer quick predictions that have significantly low variance and low bias. It has only one main drawback regarding the data checks, which are not calculated directly. However, given its simplicity and effective interpretation of the results, MARS is preferred as the non-parametric approach for regression in this study.

In this study, MARS is performed for each study region individually to obtain the regional differences. Since MARS algorithm selects the most influential variables automatically, the regional differences are expressed not only in terms of correlation coefficients, but also in terms of the included predictor variables in the prediction equations.

4.4.1 Methodology

MARS is based on a linear model that uses linear and/or non-linear functions of independent variables to explain a given data structure. For a well-defined data structure, MARS algorithm enables to use step functions and polynomial regression functions to obtain non-linearities and intersection points between independent variables.

In addition, the best prediction of the response variable can be achieved by increasing the degree of polynomial regression functions as seen in Figure 4.1, the optimal degree of polynomial used in the calculations is limited to 4, where the

multicollinearity between independent variables increases beyond this threshold value.

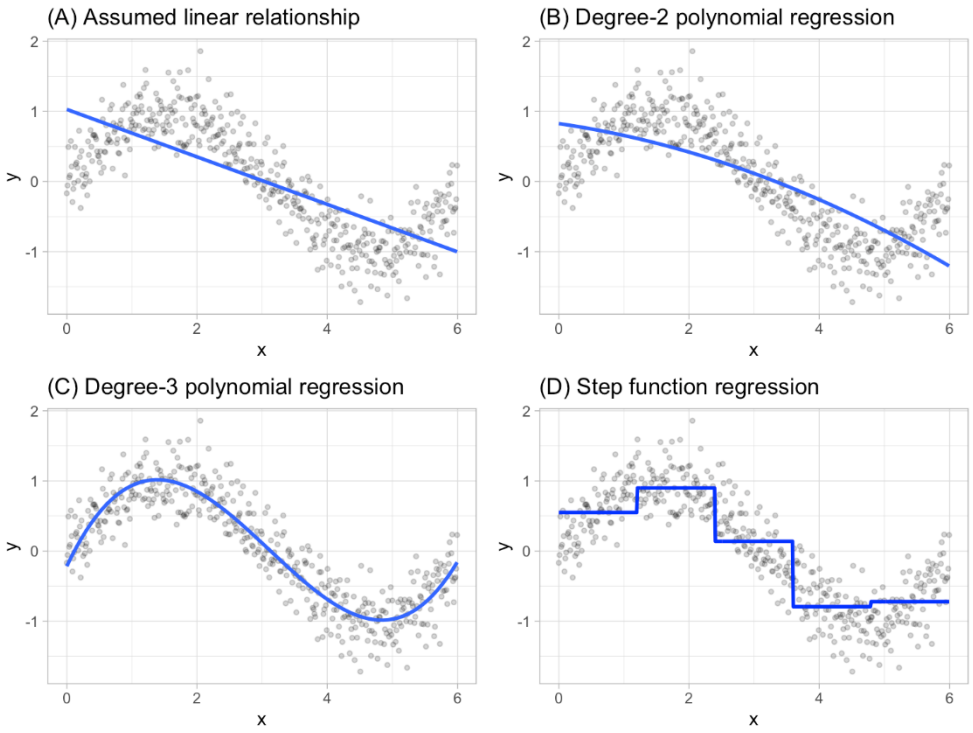


Figure 4.1 Alternative approaches for linear and non-linear patterns in given data (adopted from https://bradleyboehmke.github.io/HOML/)

The linear formulation of MARS is defined as follows:

$$Y_i = B_0 + B_1 X_i + B_2 X^2{}_i + \cdots + B_d X^d{}_i + \epsilon_i \qquad (4.20)$$

The other alternative to polynomials is to use step functions, which break the linear functions into pieces to define the locally operated constant piecewise linear functions or the basis functions (Weber et al, 2012). The algorithm is based on turning continuous variable into an arranged categorical variable. The formulation is defined as follows:

$$Y_i = B_0 + B_1 C_1(x_i) + B_2 C_2(x_i) + B_3 C_3(x_i) + \cdots + B_d C_d(x_i) + \epsilon_i \quad (4.21)$$

where, $C_1(x_i)$ represents x_i values ranging from $c_1 \leq x_i < c_2$, $C_2(x_i)$ represents x_i values ranging from $c_2 \leq x_i < c_{3\ldots}$, $C_d(x_i)$ represents x_i values ranging from $c_d \leq x_i < c_{d+1}$.

The cut points, also called knots, define the non-linear relationships between variables. These points determine the relationships through the linear regression model, also called a hinge function, with candidate piece(s) as shown in Figure 4.2 for different number of knots. The increase in the number of knots also increases handling of non-linearities between variables hence smaller residuals are obtained accordingly. In addition, the increase in the accuracy of the response variable, the generalized data could be a better representative of the initial dataset since hinge functions do not require data preparation. MARS algorithm offers to determine the number of knots if they are not believed to be the best representative of data structure by the cross-validation analysis of the independent variables.

The general MARS model is as follows:

$$Y = \beta_0 + \sum_{m=1}^{M} \beta_m H_m(x^m) + \epsilon, \quad (4.20)$$

where Y is the response variable, ϵ is an error term which is assumed to have zero mean and a finite variance. Here, β_m are the unknown coefficients for the mth basis function ($m = 1,2,\ldots,M$) and for the constant 1 ($m = 0$). The functions H_m ($m = 1,2,\ldots,M$) are hinge (basis) functions and they can be in a form of main or interaction. For a observed data pair (x_i, y_i) ($i = 1,2,\ldots n$), the form of the mth basis function for the multiple independent variables is as follows:

$$H_m(x^m) := \prod_{J=1}^{K_m} \left[s_{\kappa_j^m} \times \left(x_{\kappa_j^m} - \tau_{\kappa_j^m} \right) \right]_+, \quad (4.21)$$

where $[q]_{\{+\}} := max\{0, q\}$, K_m is the number of truncated linear functions multiplied in the mth basis function, $x_{\kappa}^{m}{}_{j}$ is the input variable corresponding to the jth truncated linear function in the mth basis function, $\tau_{\kappa}^{m}{}_{j}$ is the knot value corresponding to the variable $x_{\kappa}^{m}{}_{j}$, and $s_{\kappa}^{m}{}_{j}$ is the selected sign $+1$ or -1.

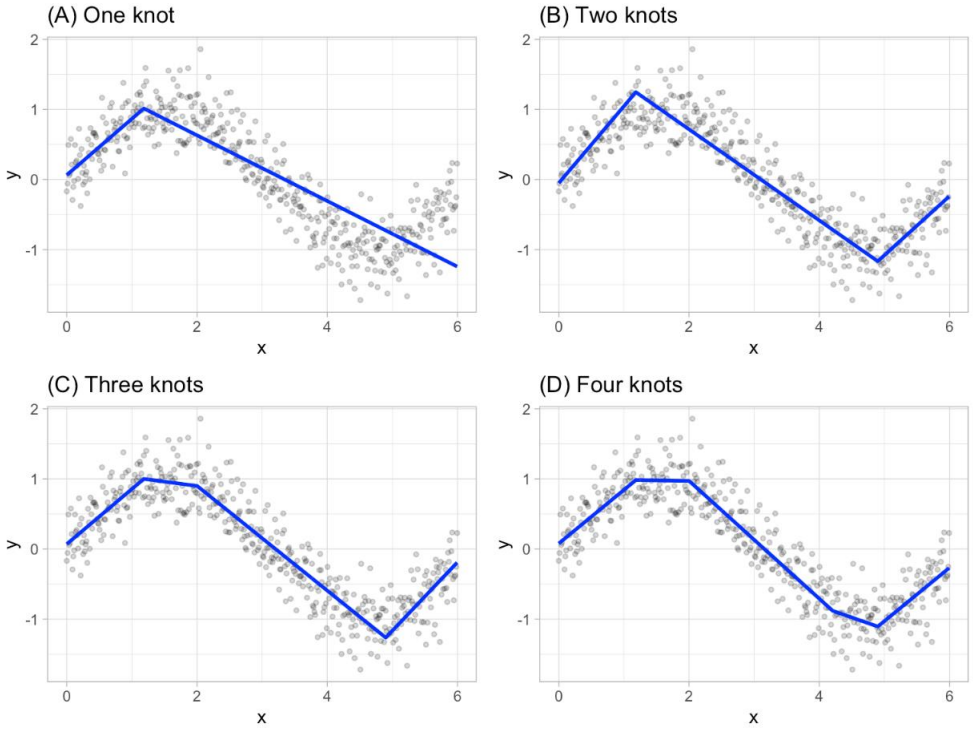


Figure 4.2 Examples of fitted regression splines for different number of knots (adopted from https://bradleyboehmke.github.io/HOML/)

A basic MARS model is obtained by "earth package" in R software (Torsten, 2022) where the dataset of each study region is analyzed by all possible knots to obtain the optimal number based on the estimated change in R^2 of less than 0.001 (Hastie and Lumley, 2019). Although the change in the coefficient of determination is very low, Generalized Cross-Validation, which generates the estimated leave-one-out cross-validation error metric, is performed to regularize the trade-off between model complexity and goodness-of-fit (Golub, Heath, and Wahba,1979).

CHAPTER 5

RESULTS AND DISCUSSION

5.1 General

In this study, initially, the relationships between MMI and log (PGA) as well as log (PGV) are studied with linear regression method using 3114 data pairs of MMI and PGA&PGV. Next, PCA is performed for 2171 data points composed of Mw, PGA, PGV, PGD, epicentral distance, D_5_95, Arias intensity, focal depth, V_s30, and the number of responses to select the parameters which mostly influence MMI levels. Based on the results of PCA, multiple linear regression is then performed with explanatory variable couples of PGA and epicentral distance as well as PGV and epicentral distance where MMI is the response variable. Finally, to study the potential non-linearities in the databases, MARS method is used via piecewise linear functions. The results are compared in terms of statistical coefficients (e.g.: coefficient of determination). GraphPad Prism 9 software (30 days trial version) is used for linear calculations. The estimated linear relationships between felt intensity and ground motion parameters are within the 95% confidence interval limit. Additionally, the residuals are plotted. For MARS computations, R software is used. With the use of these methods, not only predictive relationships are derived but also regional differences are captured.

5.2 Implementation of Linear Regression Method

Linear regression is performed for PGA-MMI and PGV-MMI data pairs with the following formulations:

$$MMI_{est} = B_1 + B_2 * \log(PGA) \quad (5.1)$$

$$MMI_{est} = B_3 + B_4 * \log(PGV) \quad (5.2)$$

where, B_i values are the regression coefficients.

5.2.1 Database 1: Türkiye

The statistical parameters for the mean representative PGA and PGV values are presented at Table 5.1.

Table 5.1 Statistical parameters of the mean representative PGA and PGV values for each MMI level for the Türkiye database

MMI	log(PGA)	log(PGV)
1	-0.048	-0.816
2	0.436	-0.483
3	0.492	-0.386
4	0.728	-0.200
5	1.085	0.115
6	1.053	0.110
7	1.645	0.575
8	1.233	0.426
9	2.239	1.400

The estimated MMI in terms of PGA for the Türkiye database is computed to be as follows:

$$MMI_{est} = 1.290 + 3.766 * \log(PGA) \quad (5.3)$$

Figure 5.1 shows the best fit line plot and the residual plot for MMI-log(PGA) correlation of the Türkiye database. The R^2 value is 0.8897. The P-value is 0.0001. These numerical values indicate moderate to strong correlation.

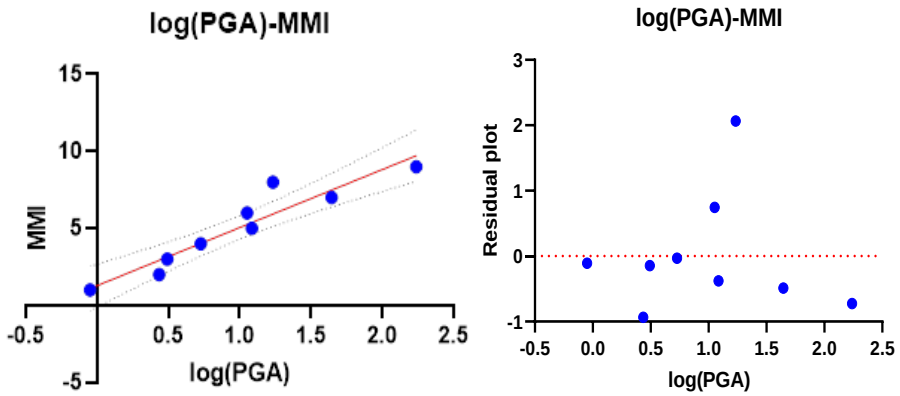


Figure 5.1 The best fit line with the mean representative dataset plot and the residual plot of the Türkiye database for MMI-log(PGA) correlation

The estimated MMI in terms of PGV for the Türkiye database is computed to be as follows:

$$MMI_{est} = 4.687+3.919*\log(PGV) \quad (5.4)$$

Figure 5.2 shows the best fit line plot and the residual plot of MMI-log (PGV) correlation of the Türkiye database. R^2 is 0.9036. The P-value is <0.0001. These numerical values indicate moderate to strong correlation.

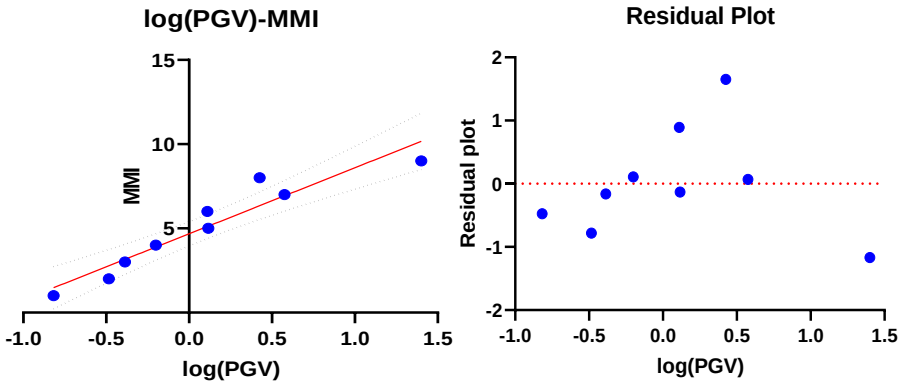


Figure 5.2 The best fit line with the mean representative dataset plot and the residual plot of the Türkiye database for MMI-log(PGV) correlation

5.2.2 Database 2: Aegean-Mediterranean Region

The statistical parameters for the mean representative PGA and PGV values are presented at Table 5.2.

Table 5.2 Statistical parameters of the mean representative PGA and PGV values for each MMI level for the AMR database

MMI	log(PGA)	log(PGV)
1	0.132	-0.414
2	0.509	-0.424
3	0.576	-0.347
4	0.752	-0.220
5	1.144	0.171
6	1.327	0.225
7	1.693	0.597
8	1.462	0.525

The estimated MMI in terms of PGA for the AMR database is computed to be as follows:

$$MMI_{est} = 0.331 + 4.390 * \log(PGA) \quad (5.5)$$

Figure 5.3 shows the best fit line plot and the residual plot of MMI-log(PGA) correlation of the AMR database. R^2 is 0.9339. The P-value is <0.0001. These numerical values indicate a strong correlation.

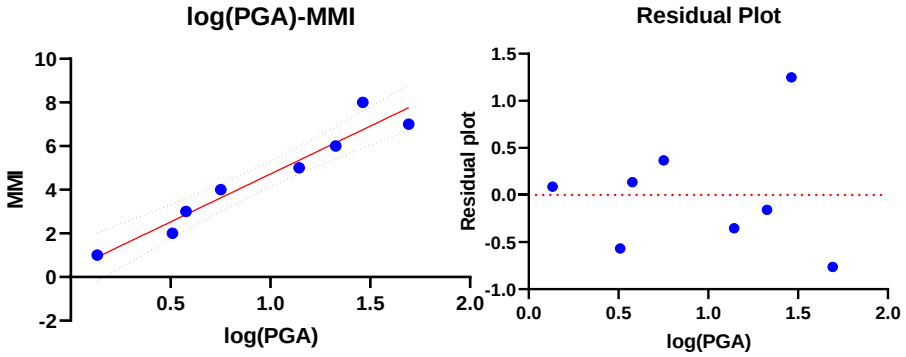


Figure 5.3 The best fit line with the mean representative dataset plot and the residual plot for of the AMR database for MMI-log(PGA) correlation

The estimated MMI in terms of PGV for the AMR dataset is computed to be as follows:

$$MMI_{est} = 4.420 + 5.597 * \log(PGV) \quad (5.6)$$

Figure 5.4 shows the best fit line plot and the residual plot of MMI-log(PGV) correlation of AMR database. R^2 is 0. 9193. The P-value is 0.0002. These numerical values indicate a strong correlation.

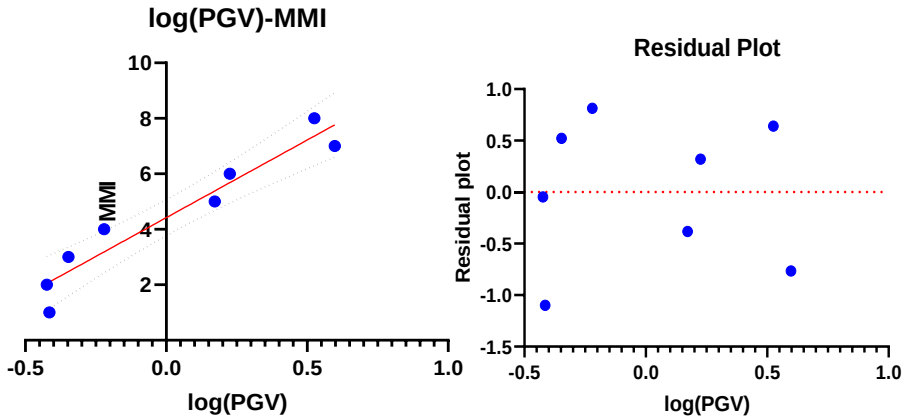


Figure 5.4 The best fit line with the mean representative dataset plot and the residual plot of the AMR database for MMI-log(PGV) correlation

5.2.3 Database 3: Strike-Slip Region

The statistical parameters for the mean representative PGA and PGV values are presented at Table 5.3.

Table 5.3 Statistical parameters of the mean representative PGA and PGV values for each MMI level for the Strike-Slip Region database

MMI	log(PGA)	log(PGV)
1	-0.077685185	-0.88214134
2	0.373885704	-0.534663255
3	0.412192236	-0.424261758

Table 5.3 (cont'd)

4	0.705540196	-0.181384963
5	1.010860097	0.048815176
6	0.885175145	0.037353096
7	1.478207052	0.495517408
8	1.142279865	0.386734624
9	2.23926059	0.386734624

The estimated MMI in terms of PGA for the Strike-Slip database is computed to be as follows:

$$MMI_{est} = 1.6 + 3.745*\log(PGA) \quad (5.7)$$

Figure 5.5 shows the best fit line plot and the residual plot of MMI-log(PGA) correlation of the Strike-Slip database. R^2 is 0.8667. The P-value is 0.0003. These values indicate a moderate to strong correlation.

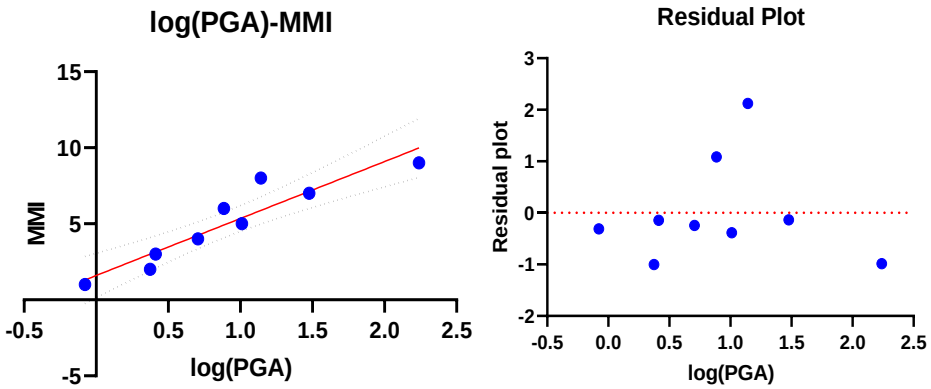


Figure 5.5 The best fit line with the mean representative dataset plot and the residual plot of the Strike-Slip database for MMI-log(PGA) correlation

The estimated MMI in terms of PGV for Strike-Slip database is computed to be as follows:

$$MMI_{est} = 4.852 + 3.850*\log(PGV) \quad (5.8)$$

Figure 5.6 shows the best fit line plot and the residual plot of MMI-log(PGA) correlation of the Strike-Slip database. R^2 is 0.8954. The P-value is 0.0001. These values indicate a moderate to strong correlation.

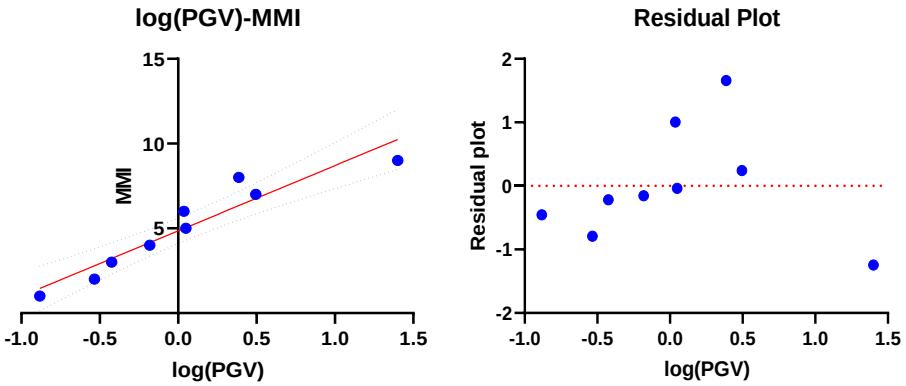


Figure 5.6 The best fit line with the mean representative dataset plot and the residual plot of the Strike-Slip database for MMI-log(PGV) correlation

5.2.4 Comparison of Correlations in Alternative Study Areas

The correlation equations for different study areas in this thesis are shown in Table 5.4. Next, Figure 5.7 compares the MMI-PGA and MMI-PGV correlations for different regions, respectively.

Table 5.4 PGA-based correlation equations of MMI for different study areas in this thesis

Study Areas/Databases	MMI-PGA Correlation	R^2
1: Türkiye	MMI=1.290+3.766* log(PGA)	0.8897
2: AMR	MMI=0.331+4.390* log(PGA)	0.9339
3: Strike-Slip	MMI=1.600+3.745* log(PGA)	0.8667

Table 5.5 PGV-based correlation equations of MMI for different study areas in this thesis

Study Areas/Databases	MMI-PGV Correlation	R^2
1: Türkiye	MMI=4.687+3.919* log(PGA)	0.9036
2: AMR	MMI=4.420+5.597* log(PGA)	0.9193
3: Strike-Slip	MMI=4.852+3.850* log(PGA)	0.8954

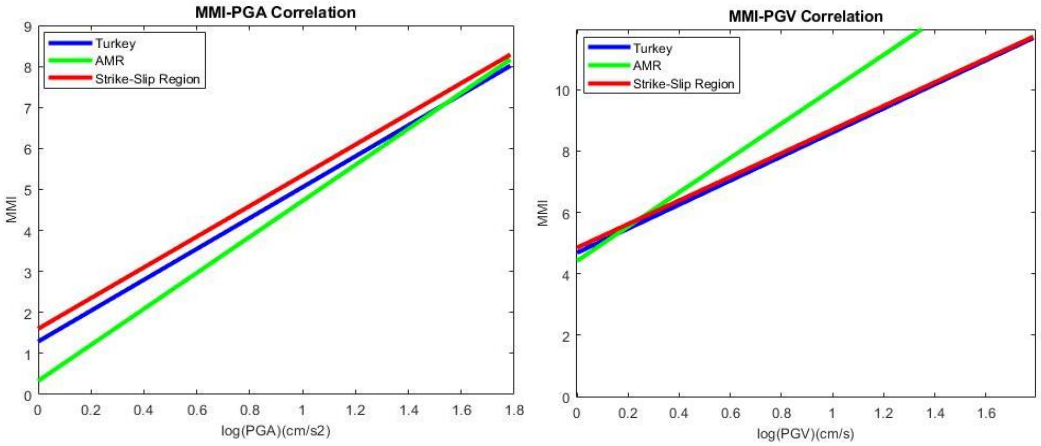


Figure 5.7 Comparison of MMI-log(PGA) and MMI-log(PGV) correlations for the study areas in this thesis

It is observed that not only the coefficient of determination of Türkiye dataset but also those of the other two dataset's for MMI-log(PGV) correlation are higher than the coefficient of determination values for the corresponding MMI-log(PGA) correlations. PGV is a better indicator for more ductile reinforced concrete structures while PGA is better correlated to damage in rigid masonry structures (Erberik, 2008 a,b). Given the majority of the structural type in Türkiye, PGV based ground motion to intensity conversion equations are strongly recommended for Türkiye. However, in regions with less ductile reinforced concrete structures and in regions with more rigid structures, PGA-based equations are suggested.

It is observed that entire Türkiye and Strike-Slip regions, whose dataset is composed of dominantly strike-slip earthquakes, exhibit compatibility for MMI-PGA and MMI-PGV relationships. However, resulting equations for the AMR region underestimate MMI values by approximately 1 unit when MMI-PGA correlation is

used and 2 units when MMI-PGV correlation is used. The differences are mostly attributed to the differences in focal mechanisms of earthquakes, building stock, and regional seismic parameters.

The highest coefficient of correlation is obtained for MMI-log(PGA) relationship in AMR, although the contribution of AMR dataset to the entire region dataset is not significant for both correlations as presented in Figure 5.7. This shows that it is necessary to examine AMR individually. In this thesis, ground motion to intensity conversion equations of AMR are successfully obtained according to coefficient of determination values higher than 90%. Additionally, this also emphasizes that the regional distinction is defined appropriately. For future events in Aegean and Mediterranean region, use of this particular correlation is recommended.

5.2.5 Comparison of the Correlations in This Study with the Most Commonly Used MMI-Ground Motion Correlation Studies

The resulting ground motion to intensity conversion (MMI-PGA and MMI-PGV) equations of Türkiye are compared with the most commonly used MMI-PGA and MMI-PGV correlations worldwide. Since the previous studies are not partially region based, the comparison is performed between the equation for the entire Türkiye dataset to the correlations in prior studies. The correlation equations selected for the comparison of MMI-PGM relationships are presented in Table 5.6. Figure 5.8 compares these equations with those obtained for MMI-PGA in this thesis.

Table 5.6 Equations for MMI-PGA correlations selected for comparisons against this study

Previous Studies	Equation
Wald et al. (1999-a)	MMI=-1.66+3.660*log(PGA)
Arıoğlu et al. (2001)	MMI=-1.078+1.748*ln(PGA)
Tselentis and Danciu (2008)	MMI=-0.946+3.563*log(PGA)
Faenza and Michelini (2010)	MMI=1.680+2.580*log(PGA)

Table 5.6 (cont'd)

Bilal and Askan (2014)	MMI=0.132+3.884*log(PGA)
This Study	MMI=1.290+3.766*log(PGA)

It is observed in Figure 5.8 that, for a given acceleation level, MMI-PGA correlation

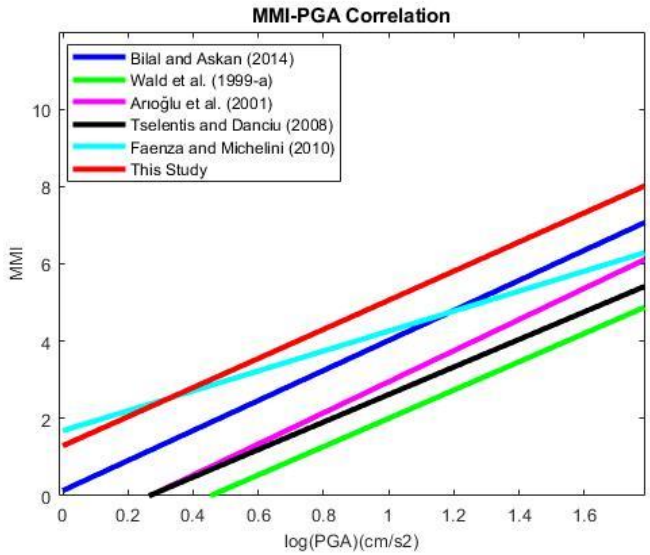


Figure 5.8 Comparison of this study with the most commonly used studies for MMI-PGA correlation

equation for California by Wald et al. (1999-a), with a dominant strike-slip focal mechanism; MMI-PGA correlation equation for Greece by Tselentis and Danciu (2008), and MMI-PGA correlation equation for Italy by Faenza and Michelini (2010) underestimate the felt-intensity values from the corresponding equation derived in this thesis. In addition to the natural subjectivity of MMI values, the differences may arise from region-specific ground motion characteristics, building response as well as subjective human responses. Even for California and Türkiye which have similar tectonic structures, differences in the conversion equations are visible which points to the differences in building characteristics and behavior. This observation confirms the need for regional correlations.

For Türkiye, Arıoğlu et al. (2001) underestimate MMI levels by approximately 2 units, and Bilal and Askan (2014) underestimates MMI levels by 1 unit. The difference between these studies and the relationship derived in this thesis is precisely the volume of the new dataset. The entire dataset of this study is composed of 69 earthquakes although Arıoğlu et al. (2001) dataset is composed of only 1 earthquake, and Bilal and Askan's (2014) dataset is composed of 14 different earthquakes. The correlations in this study better exhibit the MMI levels from large events when compared to that by Bilal and Askan (2014).

Table 5.7 Equations for MMI-PGV correlation selected for comparisons against this study

Previous Study	Equation
Wald et al. (1999-a)	MMI=2.350+3.470*log (PGV)
Atkinson and Kaka (2004)	MMI=3.960+1.790*log (PGV)
Tselentis and Danciu (2008)	MMI=3.300+3.358*log (PGV)
Faenza and Michelini (2010)	MMI=5.110+2.350*log (PGV)
Bilal and Askan (2014)	MMI=0.319+5.021*log (PGV)
This Study	MMI=4.687+3.919*log (PGV)

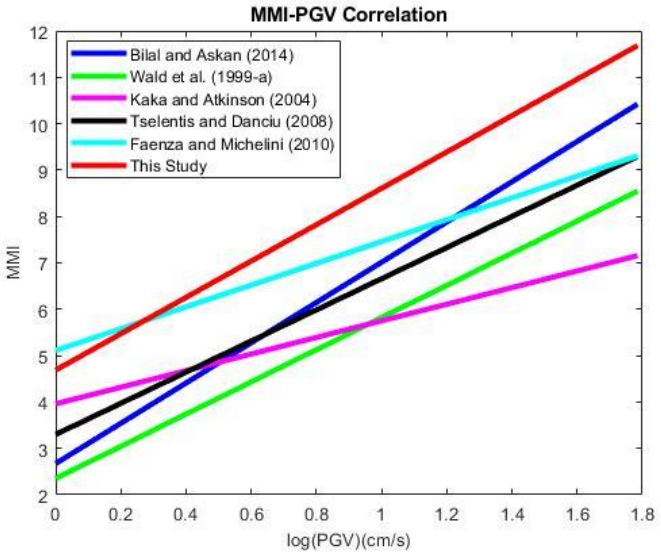


Figure 5.9 Comparison of this study with the most commonly used studies for MMI-PGV correlation

Faenza and Michelini (2010) overestimate the MMI levels in this study for values up to MMI= V and underestimate for the rest of the scale. The remaining equations from other studies underestimate MMI levels. The results show that the ground motion to intensity conversion equations for MMI-PGV correlations certainly need to be constructed locally.

Bilal and Askan (2014) underestimate the MMI levels computed from PGV by approximately 2 units for Türkiye. The main difference is the increased number of ground motion stations which help exhibit regional differences.

5.3 Implementation of Principal Component Analysis Method

Principal component analysis is performed for the entire database of Türkiye, AMR, and Strike-Slip databases to define the most influential variables on MMI levels. Next, the correlation matrix is used to retain only uncorrelated variables. Since the entire Türkiye database has almost the same correlation coefficients in the correlation matrix, the selected explanatory variables of this dataset is used for further calculations. In this study, PGA and PGV are examined individually since these parameters are the main explanatory variables in ground motion to intensity conversion equations.

The correlation matrix of the entire Türkiye database is presented in Figure 5.10. The AMR database and the Strike-Slip database correlation matrices are presented in Appendix F and Appendix G, respectively.

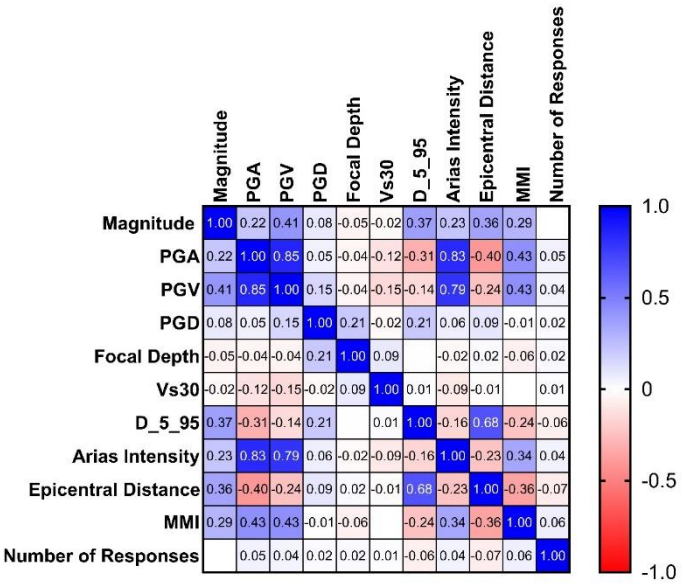


Figure 5.10 The correlation matrix of the Türkiye database

According to the correlation matrix of the Türkiye database, PGA&PGV, PGA&Arias intensity, PGV&Arias intensity, epicentral distance&D_5_95 are highly correlated with each other. PGV&magnitude, magnitude&D_5_95 are moderately correlated with each other.

For MMI-based relationships of variables, P value summary of the Türkiye database is used of which the details are presented at Appendix H. PGA, PGV, magnitude, Arias intensity, D_5_95, and epicentral distance are defined as the most important parameters but, as stated earlier, there are correlations among these variables. To prevent multicollinearity, it is necessary to retain only the uncorrelated or very weakly correlated variables. Based on the correlation values in Figure 5.10, the uncorrelated variables are selected as PGA&epicentral distance and PGV&epicentral distance.

5.3.1 PGA-based Principal Components

The PGA-based eigenvalues of the principal components of the Türkiye database are presented in Figure 5.11.

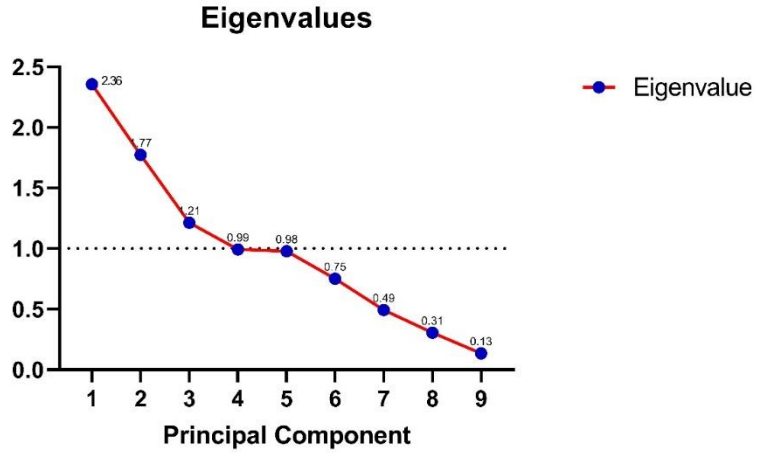


Figure 5.11 PGA-based eigenvalues of principal components for the Türkiye database

The PGA-based proportion of variance of principal components for the Türkiye database is presented in Figure 5.12.

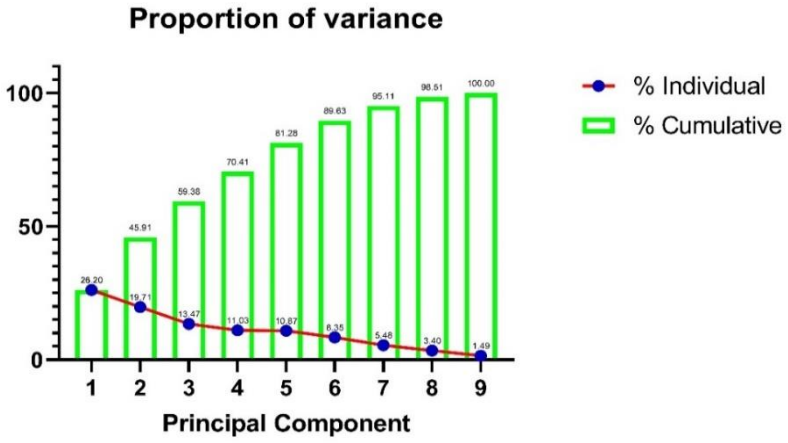


Figure 5.12 The proportion of variance of PGA-based principal component analysis of the Türkiye database

PGA-based eigenvalue analyses defined 9 eigenvalues with their proportions of variances but according to Kaiser rule, the first principal components are the most significant. Since the first principal component has the highest variance and the highest eigenvalue, the most influential variable of the first principal component is the most contributing variable. As presented in Table 5.8, PGA is the top contributor variable. Next , epicentral distance, D_5_95, and Arias intensity are ordered in terms of decreasing contributions.

Table 5.8 The contribution of variables for PGA-based principal component analysis of the Türkiye database

Variables	PC1	PC2	PC3
Magnitude	0.011	0.344	0.021
PGA	0.280	0.142	1.966E-05
PGD	0.005	0.068	0.350
Focal Depth	0.001	2.082E-05	0.528
Vs30	0.005	0.017	0.061
D_5_95	0.225	0.138	0.001
Arias Intensity	0.207	0.197	1.717E-05
Epicentral Distance	0.257	0.090	0.008
Number of Responses	0.006	1.687E-05	0.029

The loadings of variables for PGA-based principal component analysis is presented in Figure 5.13.

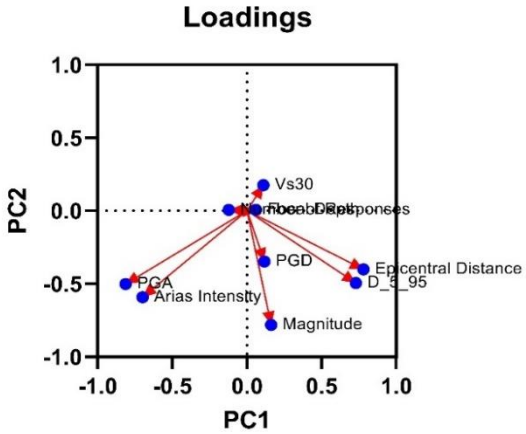


Figure 5.13 The loadings of variables between the first principal components for PGA-based principal component analyses

The loadings in Figure 5.13, which are used to define the correlations between parameters visually, show that V_s30, focal depth, and number of responses do not have significant affects on PGA-based principal components. However, the correlations between PGA & Arias intensity, and epicentral distance & D_5_95 are obtained just as in the correlation matrix.

5.3.2 PGV-based Principal Components

The eigenvalues of principal components are presented at Figure 5.14.

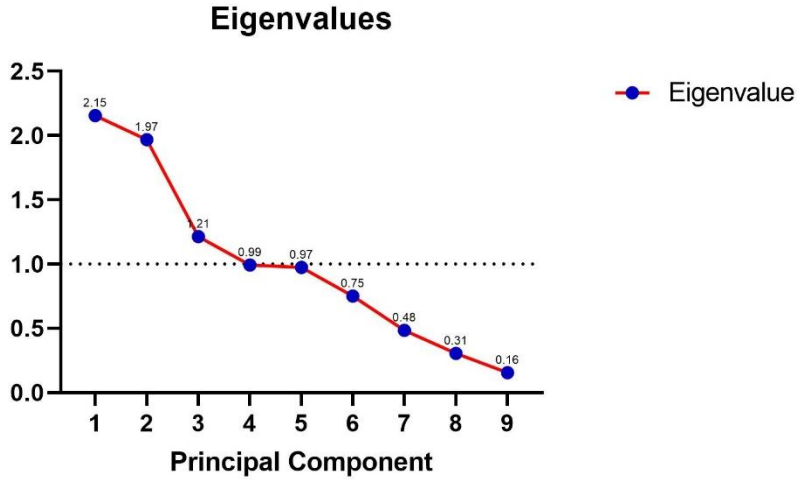


Figure 5.14 PGV-based eigenvalues of principal components for the Türkiye database

The proportion of variance of PGV-based principal component analysis is presented in Figure 5.15.

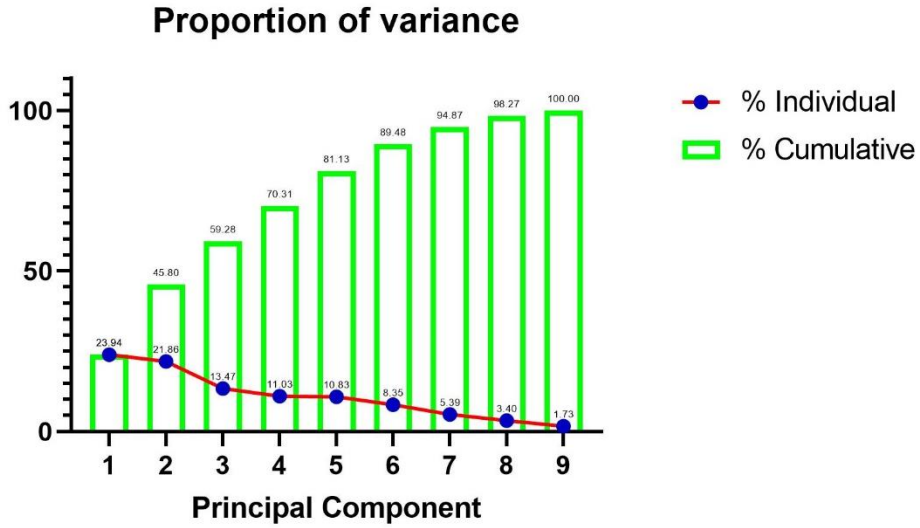


Figure 5.15 The proportion of variance of PGV based principal component analysis of the Türkiye database

PGV-based eigenvalue analyses defined 9 eigenvalues with their proportions of variance. According to Kaiser rule, as presented in Table 5.9, PGV is the most contributing variable with a slightly higher contribution than Arias intensity, and epicentral distance, and D_5_95, and are ranked in order of decreasing contributions.This is an interesting result since the correlation coefficients between MMI-PGA and PGA-Arias intensity are higher than the correlation coefficients between MMI-PGV and PGV-Arias intensity.

Table 5.9 The contribution of variables for PGV-based principal component analysis

Variables	PC1	PC2	PC3
Magnitude	0.001	0.329	0.021
PGV	0.296	0.136	1.77E-05
PGD	0.001	0.069	0.349
Focal Depth	0.002	1.38E-05	0.528
Vs30	0.012	0.010	0.062
D_5_95	0.172	0.209	0.001
Arias Intensity	0.293	0.082	1.29E-05

Table 5.10 (cont'd)

Epicentral Distance	0.214	0.161	0.008
Number of Responses	0.007	0.001	0.030

The loadings of variables for PGV-based principal component analysis is presented at Figure 5.16.

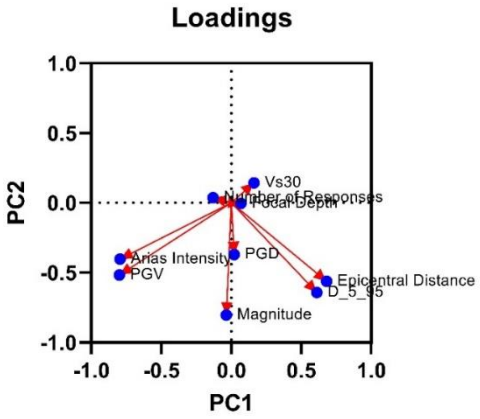


Figure 5.16 The loadings of variables between the first principal components of PGV-based principal component analyses

The loadings in Figure 5.16 shows the same relationships between variables for PGV-based principal component analyses as in the case of PGA.

5.4 Implementation of Multiple Linear Regression Method

Multiple linear regression is performed for each study database defined in this thesis by using the most influential variables that are identified by principal component analysis. These variable couples are PGA&epicentral distance and PGV&epicentral distance. The formulations of the equations are defined as follows:

$$MMI_{est} = B_1 + B_2 * (\text{PGA}) + B_3 * (\text{Epicentral Distance}) \quad (5.9)$$

$$MMI_{est} = B_4 + B_5 * (\text{PGV}) + B_6 * (\text{Epicentral Distance}) \quad (5.10)$$

where, B_i values are the regression coefficients.

5.4.1 Database 1: Türkiye

The estimated MMI in terms of PGA and epicentral distance for the Türkiye database is computed to be as follows:

$$MMI_{est} = 3.774 + (0.02401)\,(\text{PGA}) - (0.003085)\,(\text{Epicentral Distance}) \quad (5.11)$$

The regression coefficients and the standard errors are presented in Table 5.10.

Table 5.10 Regression coefficients and standard errors of Equation 5.11

Parameter estimates	Variable	Estimate	Standard error
B_1	Intercept	3.774	0.05979
B_2	PGA	0.02401	0.001447
B_3	Epicentral Distance	-0.003085	0.0002869

The multicollineartiy of the parameters is defined by Variance Inflation Factors (VIF), which are presented in Table 5.11.

Table 5.11 VIFs of Equation 5.11

Multicollinearity	Variable	VIF
B_1	Intercept	
B_2	PGA	1.193
B_3	Epicentral Distance	1.193

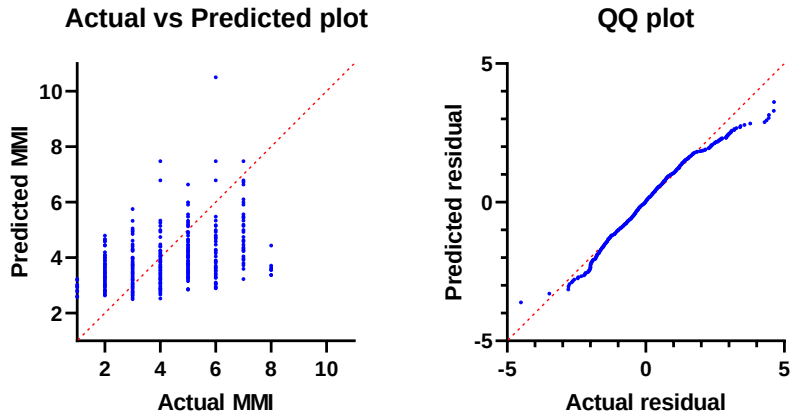


Figure 5.17 The estimated and observed MMI levels and the Q-Q plot for MMI-PGA-epicentral distance correlation of the Türkiye database

The estimated MMI in terms of PGV and epicentral distance for Türkiye dataset is computed as follows:

MMI_{est} =3.834+(0.2659)*(PGV)-(0.003778)*(Epicentral Distance) (5.12)

The regression coefficients and the standard errors are defined in Table 5.12 while VIFs are presented in Table 5.13.

Table 5.12 Regression coefficients and standard errors of Equation 5.12

Parameter estimates	Variable	Estimate	Standard error
B_4	Intercept	3.834	0.0550
B_5	PGV	0.266	0.0140
B_6	Epicentral Distance	-0.004	0.0002

Table 5.13 VIFs of Equation 5.12

Multicollinearity	Variable	VIF
B_4	Intercept	
B_5	PGV	1.061
B_6	Epicentral Distance	1.061

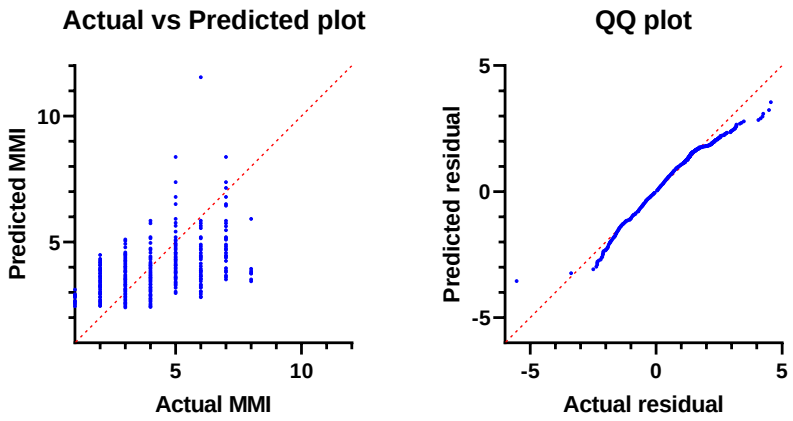


Figure 5.18 The estimated and observed MMI levels and the Q-Q plot for MMI-PGV-epicentral distance correlation of the Türkiye database

For the correlations of PGA-epicentral distance with MMI and PGV-epicentral distance with MMI, there is no multicollinearity between variables since VIF is lower than the limiting value of 4. Since P value is <0.0001 for both of these regressions, these regressions are statistically significant. The Q-Q plots show that there are saturations at the two ends, so the fitted database into the regression model is obtained moderately. MMI values are overestimated up to MMI=III, and underestimated for higher levels compared to the best-fit of estimation resulting in lower correlation coefficients, which are presented in Table 5.14.

Table 5.14 The goodness of fit parameters of multiple linear regression analysis of the Türkiye database

Goodness of Fit	PGA-Epicentral Distance	PGV-Epicentral Distance
Degrees of Freedom	2168	2168
Multiple R	0.477	0.506
R squared	0.227	0.256
Adjusted R squared	0.227	0.255
Sum of Squares	2308	2223
RMSE	1.031	1.012
AICc	141.200	59.290

According to goodness of fit parameters, PGV-epicentral distance based GMICE is a better fit for the entire Türkiye database.

Since Bilal and Askan (2014) derived ground motion to intensity equations based on moment magnitude, epicentral distance and PGA&PGV, the resulting equations in this thesis are not compared with the equations from that study.

5.4.2 Database 2: Aegean-Mediterranean Region

The estimated MMI in terms of PGA and epicentral distance for the AMR database is computed to be as follows:

$$MMI_{est} = 3.575 + (0.02797)*(\text{PGA}) - (0.002561)*(\text{Epicentral Distance}) \quad (5.13)$$

The regression coefficients and the standard errors are defined at Table 5.15 and VIFs are presented in Table 5.16.

Table 5.15 Regression coefficients and standard errors of Equation 5.13

Parameter estimates	Variable	Estimate	Standard error
B_7	Intercept	3.575	0.0790
B_8	PGA	0.028	0.0020
B_9	Epicentral Distance	-0.003	0.0004

Table 5.16 VIFs of Equation 5.13

Multicollinearity	Variable	VIF
B_7	Intercept	
B_8	PGA	1.165
B_9	Epicentral Distance	1.165

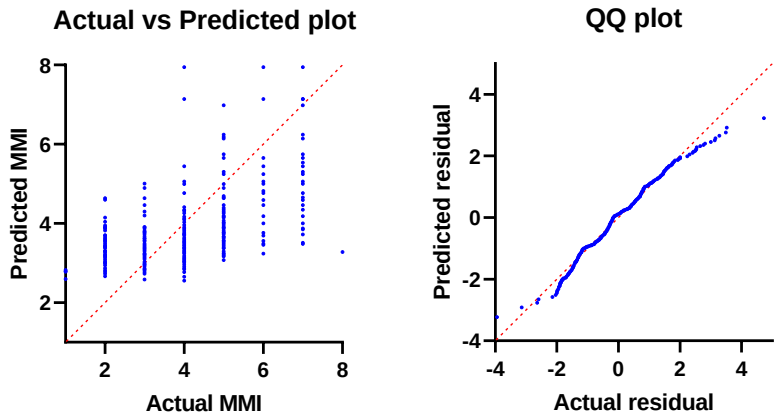


Figure 5.19 The estimated and observed MMI levels and the Q-Q plot for MMI-PGA-epicentral distance correlation of the AMR database

The estimated MMI in terms of PGV and epicentral distance for the AMR database is computed to be as follows:

$$MMI_{est} = 3.688 + (0.2899)*(\text{PGV}) - (0.003539)*(\text{Epicentral Distance}) \quad (5.14)$$

The regression coefficients and the standard errors are defined at Table 5.17 and VIFs are presented in Table 5.18.

Table 5.17 Regression coefficients and standard errors of Equation 5.14

Parameter estimates	Variable	Estimate	Standard error
B_{10}	Intercept	3.688	0.0750
B_{11}	PGV	0.289	0.0170
B_{12}	Epicentral Distance	-0.004	0.0004

Table 5.18 VIFs of Equation 5.14

Multicollinearity	Variable	VIF
B_{10}	Intercept	
B_{11}	PGV	1.070
B_{12}	Epicentral Distance	1.070

The estimated MMI and the observed MMI levels are plotted at Figure 5.20.

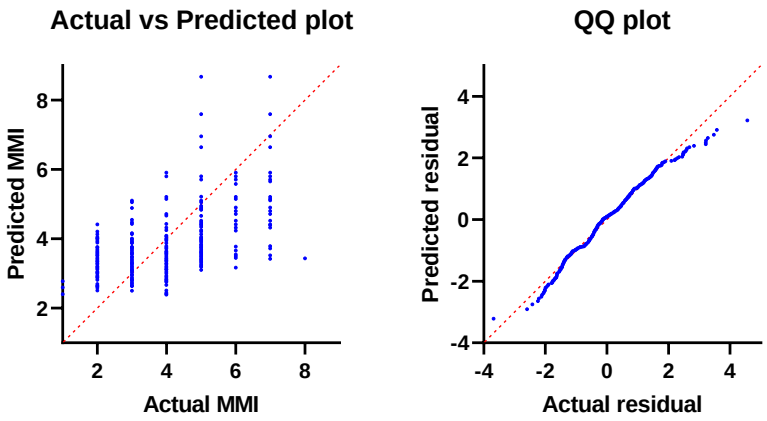


Figure 5.20 The estimated and observed MMI levels and the Q-Q plot for MMI-PGV-epicentral distance correlation of the AMR database

For the correlation of PGA-epicentral distance with MMI and PGV-epicentral distance with MMI, there is no multicollinearity between variables since VIF is lower than 4. The Q-Q plots show that the assumption of normal distribution of variables is moderately significant due to the skewness at the beginning and the end of the graph. MMI levels are overestimated up to MMI=III, and underestimated for higher levels compared to the perfect estimation because of low correlation coefficients. Additionally, P values lower than 0.0001 for both of these regressions provides statistical significance. The correlation coefficients are shown in Table 5.19.

Table 5.19 The goodness of fit parameters of multiple linear regression analysis of the AMR database

Goodness of Fit	PGA-Epicentral Distance	PGV-Epicentral Distance
Degrees of Freedom	1107	1107
Multiple R	0.533	0.537
R squared	0.284	0.288
Adjusted R squared	0.283	0.287
Sum of Squares	1049	1043

Table 5.19 (cont'd)

RMSE	0.973	0.969
AICc	-54.510	-61.390

According to goodness of fit parameters, PGV-epicentral distance based GMICE is a better fit for the AMR database.

5.4.3 Database 3: Strike-Slip Region

The estimated MMI in terms of PGA and epicentral distance for the Strike-Slip database is computed to be as follows:

$$MMI_{est} = 4.181 + (0.01578)*(\text{PGA}) - (0.004374)*(\text{Epicentral Distance}) \quad (5.15)$$

The regression coefficients and the standard errors are defined at Table 5.20 and VIFs are presented in Table 5.21.

Table 5.20 Regression coefficients and standard errors of Equation 5.15

Parameter estimates	Variable	Estimate	Standard error
B_{13}	Intercept	4.181	0.0990
B_{14}	PGA	0.016	0.0030
B_{15}	Epicentral Distance	-0.004	0.0004

Table 5.21 VIFs of Equation 5.15

Multicollinearity	Variable	VIF
B_{13}	Intercept	
B_{14}	PGA	1.251
B_{15}	Epicentral Distance	1.251

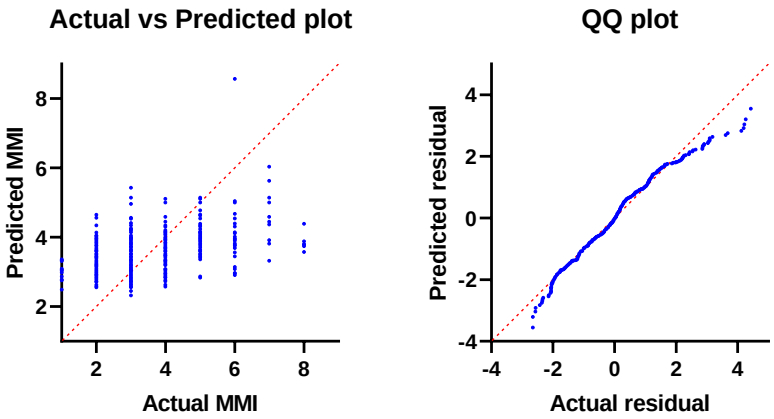


Figure 5.21 The estimated and observed MMI levels and the Q-Q plot for MMI-PGA-epicentral distance correlation of the Strike-Slip database

The estimated MMI in terms of PGV and epicentral distance for the Strike-Slip database is computed to be as follows:

$$MMI_{est} = 4.129 + (0.2273)*(\text{PGV}) - (0.004624)*(\text{Epicentral Distance}) \quad (5.16)$$

The regression coefficients and the standard errors are defined at Table 5.22 and VIFs are presented in Table 5.23.

Table 5.22 Regression coefficients and standard errors of Equation 5.16

Parameter estimates	Variable	Estimate	Standard error
B_{16}	Intercept	4.129	0.0880
B_{17}	PGV	0.227	0.0230
B_{18}	Epicentral Distance	-0.005	0.0004

Table 5.23 VIFs of Equation 5.16

Multicollinearity	Variable	VIF
B_{16}	Intercept	
B_{17}	PGV	1.058
B_{18}	Epicentral Distance	1.058

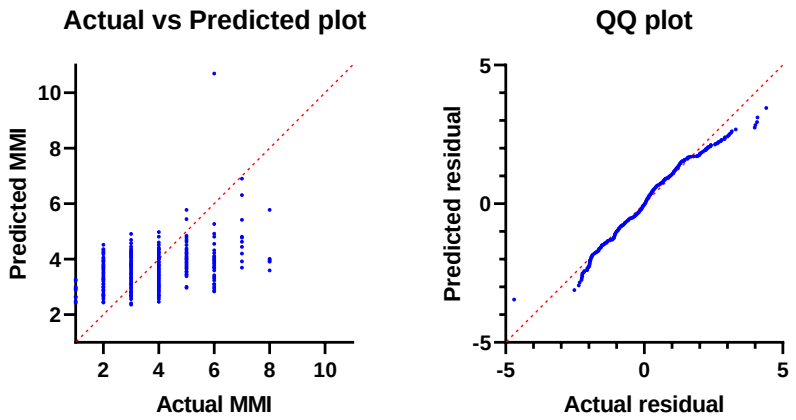


Figure 5.22 The estimated and observed MMI levels and the Q-Q plot for MMI-PGV-epicentral distance correlation of the Strike-Slip database

For the correlation of PGA-epicentral distance with MMI and PGV-epicentral distance with MMI, there is no multicollinearity between variables since VIF is lower than 4. The Q-Q plots show that the assumption of normal distribution of variables is moderately significant due to the skewness at the two ends. MMI levels are overestimated up to MMI=III and underestimated for higher levels compared to the perfect estimation because of low correlation coefficients, which are presented in Table 5.24.

Table 5.24 The goodness of fit parameters of multiple linear regression analysis of the Strike-Slip database

Goodness of Fit	PGA-Epicentral Distance	PGV-Epicentral Distance
Degrees of Freedom	1058	1058
Multiple R	0.441	0.489
R squared	0.195	0.239
Adjusted R squared	0.193	0.238
Sum of Squares	1223	1155
RMSE	1.074	1.044
AICc	158.700	97.930

According to goodness of fit parameters, PGV-epicentral distance based GMICE is a better fit for the Strike-Slip database.

5.5 Implementation of Multivariate Adaptive Regression Splines Method

Multivariate adaptive regression splines method is used for defining the non-linear relationships between ground motion parameters and felt intensity levels for each study region defined in this study. In this thesis, the best estimation function in the R software is used to define the hinge functions as well as intercept value through the databases.

5.5.1 Database 1: Türkiye

Entire database of Türkiye database is composed of 2171 data points. After the best estimation of felt intensity through ground motion parameters, 6 ground motion parameters are selected as the predictor variables, which are magnitude, epicentral distance, Arias intensity, D_5_95, focal depth, and number of responses, and 13 terms, which are composed of 12 hinge functions as well as the intercept value, are defined as the best model of felt intensity estimation through selected ground motion variables.

The estimated MMI in terms of selected variable threshold values of hinge functions and coefficients for the entire Türkiye database is defined in Equation 5.17.

$$MMIest = \begin{cases} 13.955598945, & Intercept \\ 1.647982134, & 6.5 \geq Magnitude \\ -0.880186683, & Magnitude \geq 6.5 \\ -0.030721920, & 93.6855 \geq Epicentral\ Distance \\ -0.038201614, & 65.9992 \geq Epicentral\ Distance \\ 0.063755330, & 85.6238 \geq Epicentral\ Distance \\ 0.017941812, & 19.44 \geq Focal\ Depth \\ 0.034690745, & \text{Focal Depth} > 19.44 \\ -0.003590779, & 28.0179 \geq D\_5\_95 \\ -15.294538078, & Arias\ Intensity \geq 0.064895 \\ -0.172547439, & 2 \geq Number\ of\ Responses \\ 0.170705315, & 47 \geq Number\ of\ Responses \\ -0.182530239, & Number\ of\ Responses > 47 \end{cases} \quad (5.17)$$

The model summary of MARS for the entire database of Türkiye database is presented at Figure 5.23.

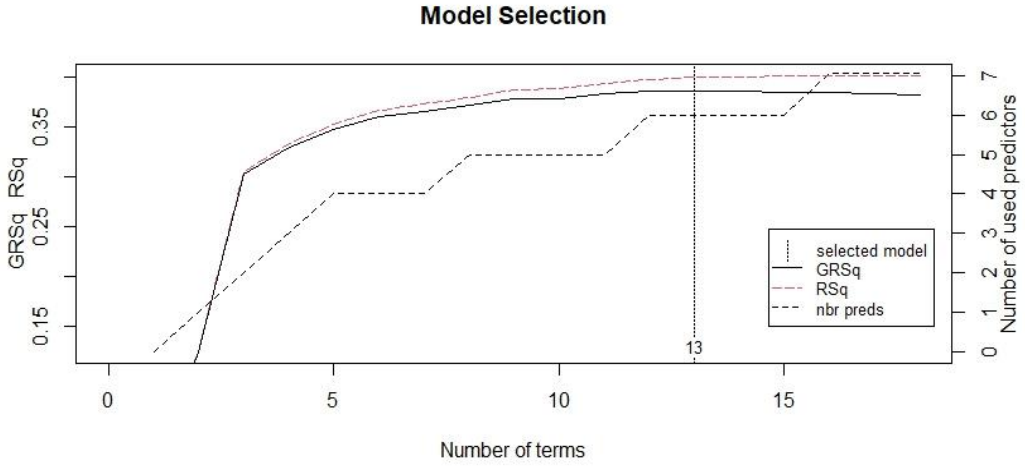


Figure 5.23 Model summary of MARS for the entire Türkiye database

The goodness of fit parameters for the best model of MARS for the entire database of Türkiye is presented at Table 5.25.

Table 5.25 The correlation coefficients between MMI and selected predictor variables for the entire database of Türkiye

Goodness of fit	Value
RMSE	0.900
R^2	0.413
MAE	0.688
RMSESD	0.025
R^2SD	0.033
MAESD	0.022

The resulting equation of best MARS model for the entire Türkiye database shows that magnitude is selected as the main explanatory variable rather than PGA and PGV. This model is still valid due to the correlation between PGA&PGV and magnitude. Although the increase in complexity of the model reduces the coefficient of determination, MARS model has a higher R^2 value when compared to multiple linear regression for the same database.

GRSq is generalized coefficient of determination and RSq is the coefficient of determination of the best model of MARS. As presented in Figure 5.23, the increased number of variables reduce the coefficient of determination for GRSq. Since we use the best estimate function in R, RSq is used. According to R^2 of this model, the prediction of MMI is of moderate power.

5.5.2 Database 2: Aegean-Mediterranean Region

The AMR database is composed of 1111 data points. After the best estimation of felt intensity through ground motion parameters, seven ground motion parameters are selected as the predictor variables, which are PGA, PGD, magnitude, epicentral distance, D_5_95, focal depth, and number of responses. 16 terms, which are composed of 15 hinge functions as well as the intercept value, are defined as the best model of felt intensity estimation through selected ground motion variables.

The estimated MMI in terms of selected variable threshold values of hinge functions and coefficients for the AMR database is defined in Equation 5.18.

$$MMIest = \begin{cases} -22.911524682, & Intercept \\ -0.008626564, & PGA \geq 46.998 \\ 16.159749441, & 0.030289 \geq PGD \\ -16.179232565, & 1.78988 \geq PGD \\ 15.917248393, & \text{PGD} > 1.78988 \\ 9.556664445, & 6.5 \geq Magnitude \\ -12.056791705, & 6.6 \geq Magnitude \\ -0.917524992, & \text{Magnitude} > 6.5 \\ 0.008014414, & 102.849 \geq Epicentral\ Distance \\ -0.012357381, & 45.4583 \geq Epicentral\ Distance \\ 0.021261746, & 14.9 \geq Focal\ Depth \\ 0.086767620, & Focal\ Depth > 14.9 \\ 0.013238413, & 113.878 \geq D\_5\_95 \\ 0.007216839, & D\_5\_95 > 113.878 \\ -0.003918730, & 45 \geq Number\ of\ Responses \\ -0.022255002, & Number\ of\ Responses > 45 \end{cases} \quad (5.18)$$

Figure 5.24 shows the model summary of MARS for the AMR database.

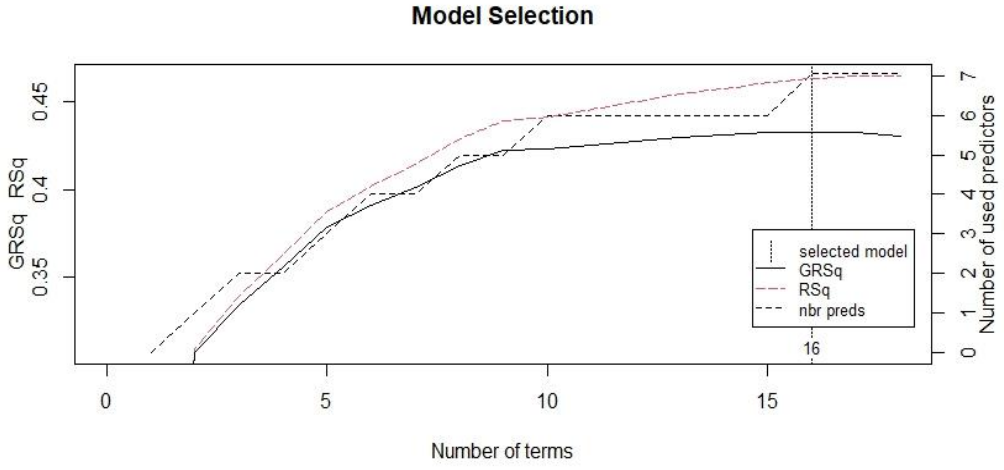


Figure 5.24 Model summary of MARS for the AMR database

The goodness of fit parameters for the best model of MARS for the AMR database is presented at Table 5.26.

Table 5.26 The correlation coefficients between MMI and selected predictor variables for the AMR database

Goodness of fit	Value
RMSE	0.848
R^2	0.458
MAE	0.642
RMSESD	0.079
R^2SD	0.068
MAESD	0.059

The resulting equation of best MARS model for the AMR database shows that PGA and PGD are both selected as the explanatory variables. Although PGD is not a very stable ground motion parameter and a very limited number of structures are prone to PGD, it is also defined as an additional variable. The coefficient of determination is still higher for the best model of MARS for the AMR database than multiple linear regression. According to R^2 of this model, the prediction of MMI is of moderate power.

5.5.3 Database 3: Strike-Slip Region

The Strike-Slip region database is composed of 1060 data points. After the best estimation of felt intensity through ground motion parameters, 6 ground motion parameters are selected as the predictor variables, which are PGV, magnitude, epicentral distance, focal depth, D_5_95, and number of responses. 14 terms, which are composed of 13 hinge functions as well as the intercept value, are defined as the best model of felt intensity estimation through selected ground motion variables.

The estimated MMI in terms of selected variable threshold values of hinge functions and coefficients for the Strike-Slip database is defined in Equation 5.19.

$$MMIest = \begin{cases} -21.31270794, & Intercept \\ 0.08556835, & 3.114 \geq PGV \\ -3.15695807, & 5.8 \geq Magnitude \\ 1.96748389, & Magnitude > 5.8 \\ 4.03798728, & 6.6 \geq Magnitude \\ 3.40218257, & 5.2 \geq Magnitude \\ -2.47561783, & 15.96 \geq Focal\ Depth \\ 2.27452904, & 5.8 \geq Focal\ Depth \\ 0.30225069, & 22.48 \geq Focal\ Depth \\ 2.19126126, & Focal\ Depth > 15.96 \\ -0.00939018, & 241.257 \geq Epicentral\ Distance \\ 0.00823639, & Epicentral\ Distance > 241.257 \\ -0.01943522, & D\_5\_95 > 29.5509 \\ -0.28982486, & Number\ of\ Responses > 2 \end{cases} \quad (5.19)$$

Figure 5.25 shows the model summary of MARS for the Strike-Slip database

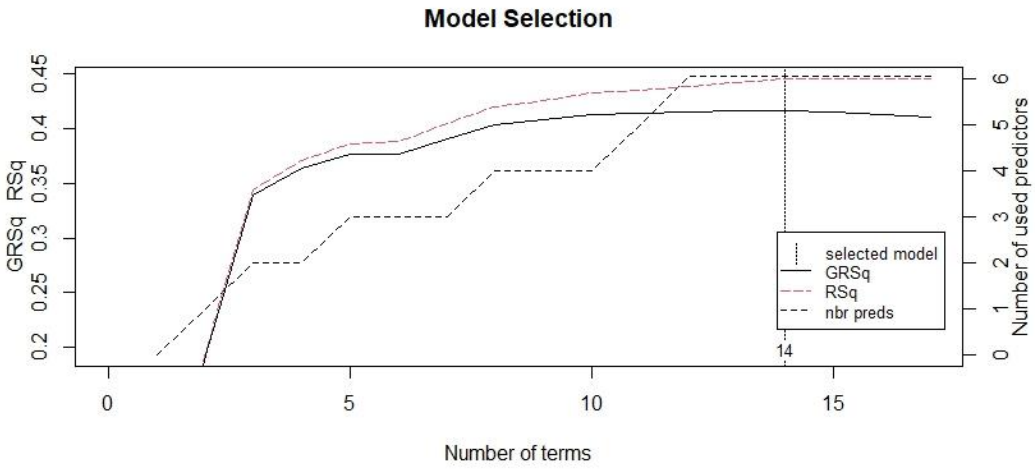


Figure 5.25 Model summary of MARS for the Strike-Slip database

The goodness of fit parameters for the best model of MARS for the Strike-Slip database is presented at Table 5.27.

Table 5.27 The correlation coefficients between MMI and selected predictor variables for the Strike-Slip database

Goodness of fit	Value
RMSE	0.915
R^2	0.424
MAE	0.697
RMSESD	0.065
R^2SD	0.094
MAESD	0.044

The resulting equation of best MARS model for the Strike-Slip database shows that PGV is selected as the main explanatory variable rather than PGA. Although PGV indicates the regional differences specifically, this result is not compatible with the entire Türkiye database, which has the same focal mechanism for selected earthquakes. The statistical parameters show that this correlation model has a higher R^2 value when compared to multiple linear regression for the same database. According to R^2 of this model, the prediction of MMI is of moderate power.

CHAPTER 6

SUMMARY AND CONCLUSIONS

6.1 Summary and Conclusions

This thesis derives ground motion to intensity conversion equations for entire Türkiye, Aegean-Mediterranean Region, and Strike-Slip dominant regions. Initially, simple linear regression method is used to obtain the equations for MMI-log(PGA) and MMI-log(PGV) pairs. Next, principal component analysis is performed to define the most influential additional variables for PGA and PGV-based ground motion to intensity conversion equations. After defining the most influential and uncorrelated variables from PCA, multiple linear regression method is performed for these variables. Finally, multivariate adaptive regression splines method is performed to define any potential non-linearities between ground motion variables in the correlation models through piecewise linear functions.

The main conclusions based on the numerical results in this thesis are listed as follows:

- Türkiye and Strike-Slip regions exhibit compatibility with each other for PGA and PGV correlations. This is mostly because the majority of the data in Türkiye come from regions with dominant focal mechanism of strike slip. However, the models based on AMR dataset is observed to be different than these two regions.
- The highest correlation coefficient is obtained for the simple conversion equation of MMI-log(PGA) of AMR dataset. The contribution of AMR data is not significant on the resulting equations of the entire dataset when compared to the Strike-Slip region’s resulting equations. This observation

along with the previous conclusion underlines regional differences and indicates that AMR is required to be examined individually.

- The highest correlation coefficients are obtained for simple linear regressions for all the regions than those from more complex models. This is mostly because when the number of explanatory variables increase, the predictive power of regression equations decreases.
- The MMI-log(PGV) correlations based on linear regression method perform better in all regions. Thus, when compared to PGA, PGV seems to be a better indicator for ground motion to intensity conversion equations for Türkiye. This conclusion is different than the findings of Bilal and Askan (2014) where PGA was found to be a better indicator of damage. The differences in the findings may arise from the fact that this study encompasses a larger dataset including MMI values in a wider range while Bilal and Askan (2014) mostly dealt with large earthquakes causing more severe damage in non-ductile structures where PGA becomes critical.
- Wald et al. (1999-a) MMI-PGA correlation equation for California, Tselentis and Danciu (2008) MMI-PGA correlation equation for Greece and FaTenza and Michelini (2010) MMI-PGA correlation equation for Italy underestimate the felt-intensity levels in Türkiye. These differences are mostly due to the fact that both the design considerations and construction styles of the buildings as well as the key ground motion parameters exhibit regional characteristics. Thus, use of region-specific datasets become critical in GMIC models.
- There are minor differences between the findings of this study and previous MMI models in Türkiye. The difference between these previous studies and this thesis is mainly due to the enlarged dataset used herein. Arıoğlu et al. (2001) ‘s dataset is composed of only one earthquake, and Bilal and Askan’s (2014) dataset is composed of 14 different earthquakes whereas the entire dataset in this thesis is composed of data from 69 earthquakes.

- According to principal component analysis, PGA-epicentral distance and PGV-epicentral distance are found to be the most influential variables for ground motion to intensity conversion equations. It is interesting to note that Vs30 is not found to be correlated with MMI levels and thus it is not included as an explanatory variable. It is possible to comment that effect of local site conditions is already included in the PGA values, thus another explicit explanatory variable is not needed statistically.
- The best MARS model for the Strike-Slip dataset selected PGV as the main predictor variable. This finding, in contrast to the findings of Bilal and Askan (2014), is compatible with simple linear regression results of this study. Finally, based on the findings from alternative methods in this study, MMI-PGA correlations are suggested for regions with rigid masonry buildings and MMI-PGV correlations are recommended for regions where the ductile reinforced concrete buildings are the major building type.
- The best MARS model for the AMR dataset selected PGA and PGD as predictor variables. Although PGD is not a very stable ground motion parameter and a very limited number of structures are prone to PGD, it is still statistically selected as an additional predictor variable since PGA is defined in the model.
- Since MARS uses piecewise linear functions to model ground motion variables with MMI, the most critical variables of PGA and PGV are not forced to be defined in the resulting equations to make them statistically significant. The best MARS model for the entire Türkiye dataset resulting equation is an example of this concern since it selects neither PGA nor PGV. Thus, magnitude is selected as the main explanatory variable, which is also defined in terms of PGA and PGV.
- According to the statistical parameters, MARS has higher correlation coefficients than multiple linear regression. Although both methods are based

on a linear modeling of variables, MARS with its piecewise functions provides a better representative of multiple linear regression.

- The resulting equations of this thesis are based on a wide variety of variables with DYFI system based MMI levels, and derived equations can be used for ShakeMap applications and disaster management considerations in Türkiye.
- The ShakeMap applications in Türkiye is employed by AFAD-RED but the equations used in constructing these maps are based on ground motion models and GMICEs. Since these steps combine two standard deviations, it is strongly recommended to use resulting equations in this thesis not to increase the uncertainty in ShakeMap applications.
- Finally, simple correlations have practical value and are of critical importance immediately after the earthquakes. More complex models however may provide refined information whenever the explanatory variables are available possibly after a long time following the earthquakes.

6.2 Recommendations

Recommendations for future studies can be summarized as follows:

- The database in this thesis is gathered between 2005-2022 for magnitude 4 and above earthquakes. As the databases keep expanding, updated GMIC models are necessary.
- Building stock as well as the construction types, number of stories, and building age can be included in the regression equations for complex GMICEs to identify the damage levels of specified regions according to these parameters.
- The Strike-Slip region has lower coefficient of determination value when compared to the AMR for all methods. This shows that it can be divided into regions to obtain higher coefficient of determination values.

- The methods, which can give more accurate results with relatively small number of datasets can be used to obtain province-based GMICEs. So, in future efforts, use of other approaches including machine learning algorithms, unsupervised learning tools etc. are suggested.
- In this study, train and test datasets are not defined. In future studies, such classifications are recommended to train and test the prediction models.
- Verification of the proposed models should be performed via ShakeMaps in the future earthquakes in Türkiye.

REFERENCES

Ahmadzadeh, S., Doloei, G. J., & Zafarani, H. (2020). "Ground Motion to Intensity Conversion Equations for Iran. Pure and Applied Geophysics", 177(11), 5435–5449. https://doi.org/10.1007/s00024-020-02586-x

Alvarez, D. A., Hurtado, J. E., & Bedoya-ruíz, D. A. (2012). "Prediction of modified Mercalli intensity from PGA , PGV , moment magnitude , and epicentral distance using several nonlinear statistical algorithms",489–511. https://doi.org/10.1007/s10950-012-9291-x

Ardeleanu, L., Neagoe, C., & Ionescu, C. (2020). "Empirical relationships between macroseimic intensity and instrumental ground motion parameters for the intermediate-depth earthquakes of Vrancea region, Romania", Natural Hazards, 103(2), 2021–2043. https://doi.org/10.1007/s11069-020-04070-0

Arıoğlu, E., Arıoğlu, B., Girgin, C., (2001). "Doğu Marmara Depreminin İvme Değerleri Açısından Değerlendirilmesi", Beton Prefabrikasyon, 57-58, 5-15.

Atkinson, G. M. and Kaka, S. L. I. (2007). "Relationships between felt intensity and instrumental ground motion in the Central United States and California", Bulletin of the Seismological Society of America, 97(2), pp. 497–510.

Atkinson, G. M., and Kaka, S. I. (2006). "Relationships Between Felt Intensity and Instrumental Ground Motion for New Madrid ShakeMaps", (March), pp. 1–27.

Atkinson, G. M., and Sonley, E. (2000). "Empirical relationships between Modified Mercalli Intensity and response spectra", Bulletin of the Seismological Society of America, 90(2), pp. 537–544.

Bilal, M. and Askan, A. (2014). “Relationships between felt intensity and recorded ground-motion parameters for Türkiye”, Bulletin of the Seismological Society of America, 104(1), pp. 484–496.

Boatwright, J., Thywissen, K. and Seekins, L. C. (2001). “Correlation of ground motion and intensity for the 17 January 1994 Northridge, California, earthquake”, Bulletin of the Seismological Society of America, 91(4), pp. 739–752.

Boatwright, J., and E. Phillips (2012). “Exploiting the demographics of “Did You Feel It?” responses to estimate the felt area of moderate earthquakes.” 84th Annual Meeting of the Eastern Section of the Seismological Society of America, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 28–30 October 2012.

Bohm, G. and Zech, G. (2010). “Introduction to statistics and data analysis for physicists”, Statsref.Com.

Caprio, M., Tarigan, B., Worden, C. B., Wiemer, S., & Wald, D. J. (2015). “Ground Motion to Intensity Conversion Equations (GMICEs): A Global Relationship and Evaluation of Regional Dependency Ground Motion to Intensity Conversion Equations (GMICEs): A Global Relationship and Evaluation of Regional Dependency”. August 2016, 0–15. https://doi.org/10.1785/0120140286

Dengler, L. A., and J. W. Dewey (1998). “An Intensity Survey of Households Affected by the Northridge, California, Earthquake of 17 January, 1994”, Bulletin of the Seismological Society of America, 88, 441-462.

Du, K., Ding, B., Bai, W., Sun, J., & Bai, J. (2020). “Quantifying Uncertainties in Ground Motion-Macroseismic Intensity Conversion Equations. A Probabilistic Relationship for Western China”, Journal of Earthquake Engineering, 2469(May). https://doi.org/10.1080/13632469.2020.1750509

Erberik, M. A. (2008a). “Generation of fragility curves for Turkish masonry buildings considering in-plane failure modes”, Earthq. Eng. Struct.Dyn. 37, 387–405.

Erberik, M. A. (2008b). “Fragility-based assessment of typical mid-rise and low-rise RC buildings in Türkiye”, Eng. Struct. 30, 1360–1374.

Faenza, L. and Michelini, A. (2010). “Regression analysis of MCS intensity and ground motion parameters in Italy and its application in ShakeMap”, Geophysical Journal International, 180(3), pp. 1138–1152.

Friedman, Jerome H. 1991. “Multivariate Adaptive Regression Splines.” The Annals of Statistics. JSTOR, 1–67.

Golub, Gene H, Michael Heath, and Grace Wahba (1979). “Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter.” Technometrics 21 (2). Taylor & Francis Group: 215–23.

Gomez-Capera, A. A., D’Amico, M., Lanzano, G., Locati, M., & Santulin, M. (2020). “Relationships between ground motion parameters and macroseismic intensity for Italy”, Bulletin of Earthquake Engineering, 18(11), 5143–5164. https://doi.org/10.1007/s10518-020-00905-0

Hanks, Thomas C.; Kanamori, Hiroo (May 10, 1979), "A Moment magnitude scale" (PDF), Journal of Geophysical Research, 84 (B5): 2348–50, Bibcode:1979JGR....84.2348H, doi:10.1029/JB084iB05p02348, archived from the original on August 21, 2010.

Kaka, S. L. I., and Atkinson, G. M. (2004). “Relationships between instrumental ground-motion parameters and modified Mercalli intensity in eastern North America”, *Bulletin of the Seismological Society of America*, 94(5), pp. 1728–1736.

Karim, Kazi R. and Yamazaki, F (2002). “Correlation of JMA instrumental seismic intensity with strong motion parameters”, Earthquake Engineering and Structural Dynamics, 31(5), 1191-1212. https://doi.org/10.1002/eqe.158

Mukherjee, S., & Gupta, V. K. (2002). “Correlation of JMA instrumental seismic intensity with strong motion parameters”, Earthquake Engineering and Structural Dynamics, 31(5), 1191–1212. https://doi.org/10.1002/eqe.158

Murphy, J. R., and O’Brien, L. J. (1977). “The correlation of peak ground acceleration amplitude with seismic intensity and other physical parameters”, Bull. Seismol. Soc. Am. 67, 877–915

Stover, C. W., and Coffman, J. L. (1993). “Seismicity of the United States, 1568-1989 (Revised)”, U. S. Geological Survey Professional Paper 1527, 418 p.

Tao, D., Ma, Q., Li, S., Xie, Z., Lin, D., & Li, S. (2020). “Support vector regression for the relationships between ground motion parameters and macroseismic intensity in the Sichuan-Yunnan Region”, Applied Sciences (Switzerland), 10(9). https://doi.org/10.3390/app10093086.

Trevor Hastie, Stephen Milborrow. “Derived from mda:mars by, and Rob Tibshirani. Uses Alan Miller’s Fortran utilities with Thomas Lumley’s leaps wrapper. 2019. *Earth:* Multivariate Adaptive Regression Splines”, https://CRAN.R-project.org/package=earth.

Torsten Hothorn (2022). “CRAN Task View: Machine Learning & Statistical Learning. Version 2022-03-07”, URL https://CRAN.R-project.org/view=MachineLearning.

Trifunac, M. D. and Brady, A. G. (1975). “On the correlation of seismic intensity scales with the peaks of recorded strong ground motion”, Bulletin of the Seismological Society of America.

Tselentis, G. A. and Danciu, L. (2008). “Empirical relationships between modified Mercalli intensity and engineering ground-motion parameters in Greece”, Bulletin of the Seismological Society of America, 98(4), pp. 1863–1875.

Wald, D. J., V. Quitoriano, T.H. Heaton, and H. Kanamori (1999-a). “Relationships between peak ground accerelation, peak ground velocity, and modified Mercalli intensity in California”, Earthq. Spectra 15, no. 3, 557-564

Wald, DJ., Quitoriano, V., Heaton, TH., Kanamori, H., Scrivner, CW., Worden, BC. (1999-b). “TriNet “ShakeMaps”: Rapid generation of peak ground-motion and intensity maps for earthquakes in southern California”, Earthquake Spectra 15(3): 537–556

Wang, Z.Y., Chen, L.E. (2008). “Application of principal component analysis in influencing factor weighting”, Mathematics in Practic,38 (4): 26-29.

Weber, G. W., Batmaz, I., Köksal, G., Taylan, P., & Yerlikaya-Özkurt, F. (2012). “CMARS: A new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization”, Inverse Problems in Science and Engineering, 20(3), 371–400. https://doi.org/10.1080/17415977.2011.624770

Wood, H. O., and Neumann, Frank (1931). “Modified Mercalli Intensity Scale of 1931”, Seismological Society of America Bulletin, v. 21, no. 4, p. 277-28.

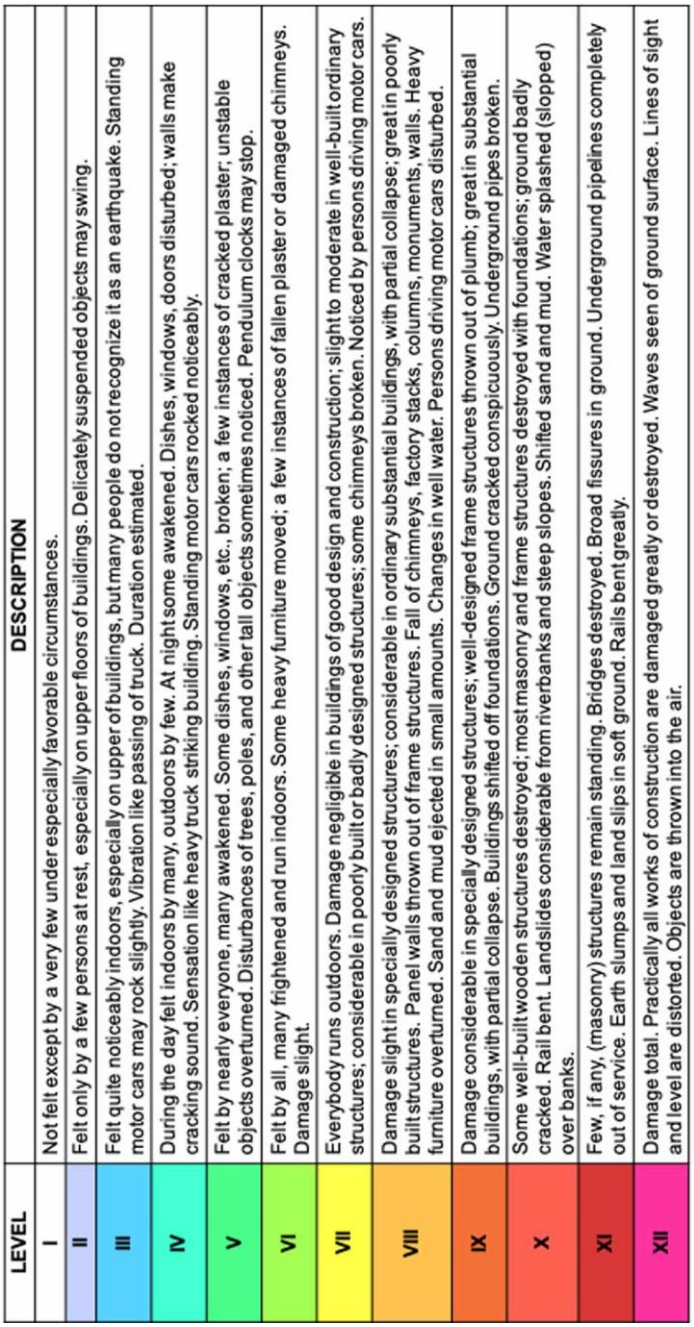
Worden, C. B., D. J. Wald, K. Lin, G. Cua, and D. Garcia (2010). “A revised ground-motion and intensity interpolation scheme for ShakeMap”, Bulletin of the Seismological Society of America, 100, 3083-3096.

Wu, Y. M., Hsiao, N. C. and Teng, T. L. (2004). ”Relationships between strong ground motion peak values and seismic loss during 1999 Chi-Chi, Taiwan Earthquake”, Natural Hazards, 32(3), pp. 357–373.

Yaghmaei-Sabegh, S., Tsang, H. H. and Lam, N. T. K. (2011). “Conversion Between Peak Ground Motion Parameters and Modified Mercalli Intensity Values”, Journal of Earthquake Engineering, 15(7), pp. 1138–1155.

APPENDICES

A. Modified Mercalli Intensity Scale



LEVEL	DESCRIPTION
I	Not felt except by a very few under especially favorable circumstances.
II	Felt only by a few persons at rest, especially on upper floors of buildings. Delicately suspended objects may swing.
III	Felt quite noticeably indoors, especially on upper of buildings, but many people do not recognize it as an earthquake. Standing motor cars may rock slightly. Vibration like passing of truck. Duration estimated.
IV	During the day felt indoors by many, outdoors by few. At night some awakened. Dishes, windows, doors disturbed; walls make cracking sound. Sensation like heavy truck striking building. Standing motor cars rocked noticeably.
V	Felt by nearly everyone, many awakened. Some dishes, windows, etc., broken; a few instances of cracked plaster; unstable objects overturned. Disturbances of trees, poles, and other tall objects sometimes noticed. Pendulum clocks may stop.
VI	Felt by all, many frightened and run indoors. Some heavy furniture moved; a few instances of fallen plaster or damaged chimneys. Damage slight.
VII	Everybody runs outdoors. Damage negligible in buildings of good design and construction; 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	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. Persons driving motor cars disturbed.
IX	Damage considerable in specially designed structures; well-designed frame structures thrown out of plumb; great in substantial buildings, with partial collapse. Buildings shifted off foundations. Ground cracked conspicuously. Underground pipes broken.
X	Some well-built wooden structures destroyed; most masonry and frame structures destroyed with foundations; ground badly cracked. Rail bent. Landslides considerable from riverbanks and steep slopes. Shifted sand and mud. Water splashed (slopped) over banks.
XI	Few, if any, (masonry) structures remain standing. Bridges destroyed. Broad fissures in ground. Underground pipelines completely out of service. Earth slumps and land slips in soft ground. Rails bent greatly.
XII	Damage total. Practically all works of construction are damaged greatly or destroyed. Waves seen of ground surface. Lines of sight and level are distorted. Objects are thrown into the air.

Figure A.1 Modified Mercalli Intensity Scale

B. The Earthquakes used in this study

Table B.1 The earthquakes used in this study

Number	Earthquake ID	Date	Latitude	Longitude	Magnitude
1	252972	17.10.2005 05:45	38.1921	26.677	5.5
2	263786	20.10.2005 21:40	38.1535	26.6708	5.8
3	266472	24.10.2006 14:00	40.4221	28.9937	5.2
4	62424	26.12.2007 23:47	39.396	33.1073	5.6
5	100074	8.03.2010 02:32	38.7665	40.0712	5.1
6	100169	8.03.2010 11:12	38.7452	40.0342	5
7	101233	24.03.2010 14:11	38.7713	40.0935	5.1
8	128573	19.05.2011 20:15	39.1328	29.082	5.8
9	133752	23.06.2011 07:34	38.5562	39.6307	5.4
10	139913	22.09.2011 03:22	39.6597	38.6777	5.6
11	141933	23.10.2011 10:41	38.689	43.4657	7
12	142682	25.10.2011 14:55	38.823	43.5857	5.4
13	146118	8.11.2011 22:05	38.7192	43.0778	5.4
14	146290	9.11.2011 19:23	38.4382	43.2825	5.6
15	167145	10.06.2012 12:44	36.5302	28.9073	6
16	168752	25.06.2012 13:05	36.4792	28.9333	5
17	210047	28.12.2013 15:21	36.048	31.332	5.9
18	272073	4.09.2014 21:00	36.172	30.9301	5.2

Table B.1 (cont'd)

Number	Earthquake ID	Date	Latitude	Longitude	Magnitude
19	283239	6.12.2014 01:45	38.904	26.2741	5.1
20	313035	6.10.2015 21:27	36.1846	29.8853	5.2
21	322860	10.01.2016 17:40	39.564	34.358	5
22	350630	27.09.2016 20:57	36.405	27.5966	5.2
23	360180	6.02.2017 03:51	39.5423	26.1318	5.3
24	360268	6.02.2017 10:58	39.5275	26.1373	5.3
25	360450	7.02.2017 02:24	39.514	26.1161	5.2
26	361551	12.02.2017 13:48	39.5336	26.17	5.3
27	363883	2.03.2017 11:07	37.5955	38.4866	5.5
28	368412	13.04.2017 16:22	37.1533	28.647	5.1
29	373447	27.05.2017 15:53	38.7358	27.8156	5.1
30	375576	12.06.2017 12:28	38.8486	26.313	6.2
31	376890	17.06.2017 19:50	38.8381	26.436	5.3
32	381491	20.07.2017 22:31	36.9198	27.4435	6.5
33	381868	21.07.2017 17:09	36.941	27.332	5
34	385714	8.08.2017 07:42	36.9576	27.6236	5.1
35	396691	22.11.2017 20:22	37.1206	28.5921	5
36	396950	24.11.2017 21:49	37.1146	28.6045	5.1

Table B.1 (cont'd)

Number	Earthquake ID	Date	Latitude	Longitude	Magnitude
37	411017	24.04.2018 00:34	37.5836	38.5036	5.1
38	420513	12.09.2018 06:21	36.0535	31.2135	5.2
39	431610	20.02.2019 18:23	39.6011	26.4261	5
40	433515	20.03.2019 06:34	37.4401	29.4335	5.5
41	444581	8.08.2019 11:25	37.851	29.584	6
42	447923	26.09.2019 10:59	40.8818	28.214	5.8
43	457038	22.01.2020 19:22	39.0488	27.8443	5.4
44	457758	24.01.2020 17:55	38.3593	39.063	6.8
45	458439	25.01.2020 16:30	38.374	39.131	5.1
46	466527	23.02.2020 05:52	38.436	44.489	5.9
47	475667	14.06.2020 14:24	39.365	40.714	5.7
48	475841	15.06.2020 06:51	39.3678	40.7435	5.6
49	476430	25.06.2020 10:03	38.472	44.0285	5.4
50	476470	26.06.2020 07:21	38.7676	27.8018	5.5
51	476668	28.06.2020 17:43	36.6563	28.2336	5.2
52	478393	4.08.2020 09:37	38.2193	38.7243	5.2
53	480704	20.09.2020 19:08	38.011	34.037	5.1
54	483762	30.10.2020 11:51	37.879	26.703	6.6

Table B.1 (cont'd)

Number	Earthquake ID	Date	Latitude	Longitude	Magnitude
55	483846	30.10.2020 15:14	37.8331	26.869	5.1
56	490172	5.12.2020 12:44	36.0878	31.8998	5.2
57	491958	27.12.2020 06:37	38.5218	39.1813	5.3
58	494497	1.02.2021 08:35	38.9483	26.0788	5.1
59	499838	13.04.2021 20:28	36.5425	27.2331	5.1
60	505093	21.06.2021 22:14	36.3838	27.0975	5.3
61	505564	25.06.2021 18:28	39.192	40.2348	5.2
62	507881	1.08.2021 04:31	36.3843	27.0805	5.5
63	508454	3.08.2021 12:38	36.267	27.0148	5.2
64	510421	31.08.2021 11:04	39.0133	30.1641	5
65	515594	8.11.2021 17:43	37.8618	32.1165	5.1
66	516236	19.11.2021 12:40	39.8208	41.868	5.1
67	516819	30.11.2021 04:00	37.7275	26.1403	5.1
68	518758	5.01.2022 03:21	36.2016	31.3505	5.1
69	520777	13.02.2022 18:25	41.1583	43.9213	5.3

C. USGS DYFI Questionnaire

Table C.1 USGS DYFI questionnaire and weights

Weight	Range	Question
5x	0-1	Did you feel it?
1x	0-5	How would you describe the shaking?
1x	0-5	How did you react?
2x	0-1	Was it difficult to stand or walk?
5x	0-1	Did objects rattle, topple over, or fall off shelves?
2x	0-1	Did pictures move of get knocked askew?
3x	0-1	Did furniture slide, topple, or become displaced?
5x	0-3	Was there damage to the building?

D. An example of MMI Levels and the Number of Responses from the DYFI system of Elazığ Sivrice Earthquake 2020

Location	MMI	Responses	Distance ↑	Latitude	Longitude
Silivri Elazığ Turkey	VIII	2	11 km	38.450°N	39.310°E
Elazığ Elazığ Turkey	VI	14	25 km	38.680°N	39.230°E
Malatya Malatya Turkey	IV	2	82 km	38.360°N	38.310°E
Yeşilyurt Malatya Turkey	III	1	88 km	38.300°N	38.250°E
Diyarbakır Diyarbakır Turkey	IV	4	106 km	37.920°N	40.230°E
Fethiye Muğla Turkey	III	1	917 km	36.620°N	29.110°E

Figure D.1 An example of MMI levels and the number of responses from DYFI system of Elazığ Sivrice Earthquake, 2020

E. Descriptive Statistics of the Available Datasets

The descriptive statistics for linear regression methods and multivariate adaptive regression splines method are defined in the tables below.

Table E.1 Descriptive statistics of Türkiye dataset for linear regression analysis method

	PGA	PGV	MMI
Number of values	3114	3114	3114
Minimum	0.01066	0.001756	1
Median	4.443	0.5647	3
Maximum	661.8	29.17	9
Range	661.8	29.17	8
Mean	9.571	1.061	3.493
Std. Deviation	20.03	1.63	1.196
Std. Error of Mean	0.3588	0.02922	0.02142

Table E.2 Descriptive statistics of Eagan-Mediterranean Region dataset for linear regression analysis method

	PGA	PGV	MMI
Number of values	1515	1515	1515
Minimum	0.1793	0.02643	1
Median	5.402	0.609	3
Maximum	661.8	18.1	8
Range	661.6	18.08	7
Mean	11.75	1.172	3.565
Std. Deviation	24.95	1.763	1.194
Std. Error of Mean	0.6409	0.0453	0.03068

Table E.3 Descriptive statistics of Strike-Slip Region dataset for linear regression analysis method

	PGA	PGV	MMI
Number of values	1599	1599	1599
Minimum	0.01066	0.001756	1
Median	3.788	0.5429	3
Maximum	282.1	29.17	9
Range	282	29.17	8
Mean	7.512	0.9559	3.423
Std. Deviation	13.55	1.487	1.194
Std. Error of Mean	0.3387	0.03718	0.02986

Table E.4 Descriptive statistics of Türkiye dataset for MARS method

	Number of values	Minimum	Maximum	Range	Mean	Median	Std. Deviation	Std. Error of Mean
Magnitude	2171	5	7	2	5.842	5.8	0.5602	0.01202
PGA	2171	0.01804	282.1	282	9.355	4.386	16.72	0.3589
PGV	2171	0.006497	29.17	29.17	1.064	0.5574	1.618	0.03472
PGD	2171	0.001318	24.84	24.83	0.495	0.1581	1.276	0.02738
Focal Depth	2171	3.97	77.22	73.25	15.77	11.92	10.25	0.2201
Vs30	2171	131	1862	1731	466.4	390	279.3	5.994
D_5_95	2171	2.28	212.1	209.9	47.6	40	29.73	0.638
Arias Intensity	2171	2.45E-08	0.3433	0.3433	0.005466	0.0004772	0.02391	0.0005131
Epicentral Distance	2171	13.73	435.1	421.4	169.7	154.4	84.34	1.81
MMI	2171	1	8	7	3.475	3	1.173	0.02518
Number of Responses	2171	1	483	482	8.248	2	30.99	0.6651

Table E.5 Descriptive statistics of Aeagan-Mediterranean Region for MARS method

	Number of values	Minimum	Maximum	Range	Mean	Median	Std. Deviation	Std. Error of Mean
Magnitude	1111	5	6.8	1.8	5.797	5.8	0.5438	0.01632
PGA	1111	0.1793	160.7	160.5	10.84	4.964	18.73	0.5623
PGV	1111	0.02643	18.1	18.08	1.133	0.5714	1.753	0.05262
PGD	1111	0.00538	24.84	24.83	0.4884	0.1639	1.246	0.0374
Focal Depth	1111	3.97	77.22	73.25	18.55	14.9	12.26	0.368
Vs30	1111	131	1323	1192	473.7	401.5	263.7	7.914
D_5_95	1111	6.345	212.1	205.8	41.24	33.52	26.79	0.804
Arias Intensity	1111	7.27E-07	0.3433	0.3433	0.008046	0.0005207	0.03154	0.0009466
Epicentral Distance	1111	16.93	408.6	391.6	141.6	133.9	68.89	2.068
MMI	1111	1	8	7	3.515	3	1.149	0.0345
Number of Responses	1111	1	483	482	8.174	2	26.03	0.7813

Table E.6 Descriptive statistics of Strike-Slip region dataset for MARS method

	Number of values	Minimum	Maximum	Range	Mean	Median	Std. Deviation	Std. Error of Mean
Magnitude	1060	5	7	2	5.888	5.8	0.5738	0.01762
PGA	1060	0.01804	282.1	282	7.803	4.108	14.17	0.4351
PGV	1060	0.006497	29.17	29.17	0.9923	0.5465	1.461	0.04487
PGD	1060	0.001318	13	13	0.5022	0.1491	1.307	0.04014
Focal Depth	1060	5.01	43.6	38.59	12.84	10.35	6.431	0.1975
Vs30	1060	181	1862	1681	457.4	377	291.6	8.956
D_5_95	1060	2.28	192.2	189.9	54.28	48.66	31.17	0.9575
Arias Intensity	1060	2.45E-08	0.2176	0.2176	0.002768	0.000434	0.01074	0.0003299
Epicentral Distance	1060	13.73	435.1	421.4	199.1	196.8	88.96	2.732
MMI	1060	1	8	7	3.433	3	1.197	0.03678
Number of Responses	1060	1	483	482	8.331	2	35.47	1.089

F. The Correlation Matrix of the AMR Database

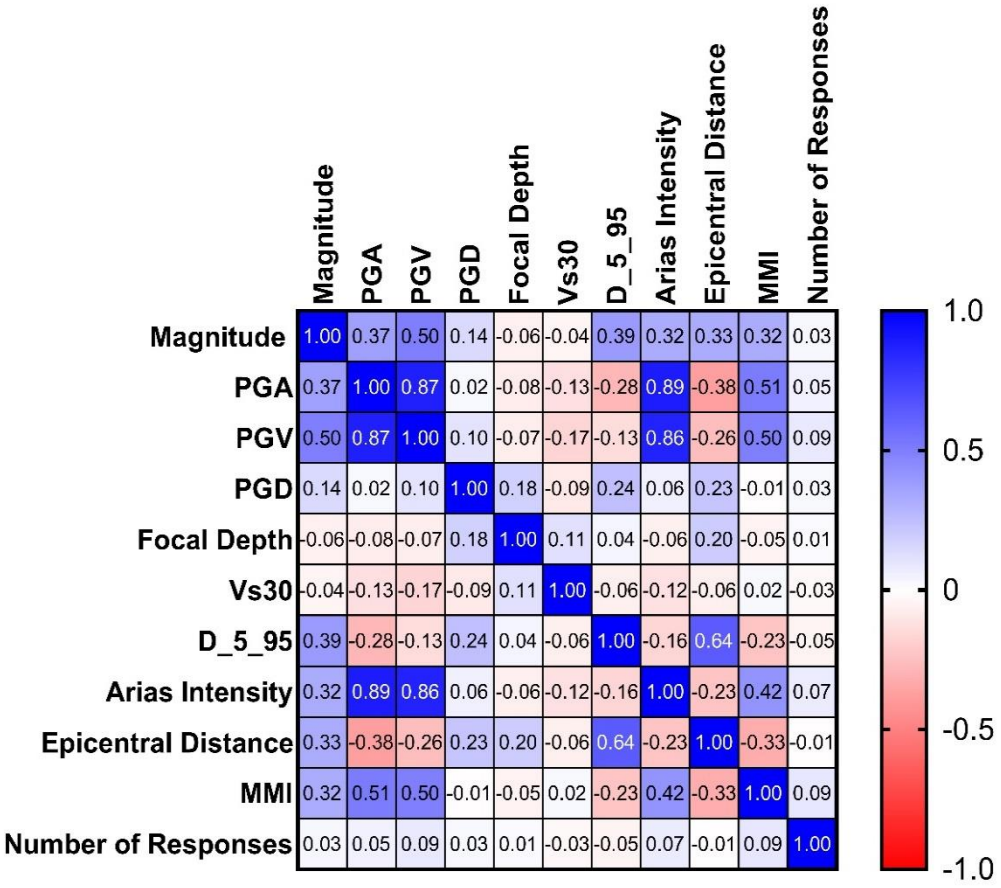


Figure F.1 The correlation matrix of the AMR database

G. The Correlation Matrix of the Strike-Slip Database

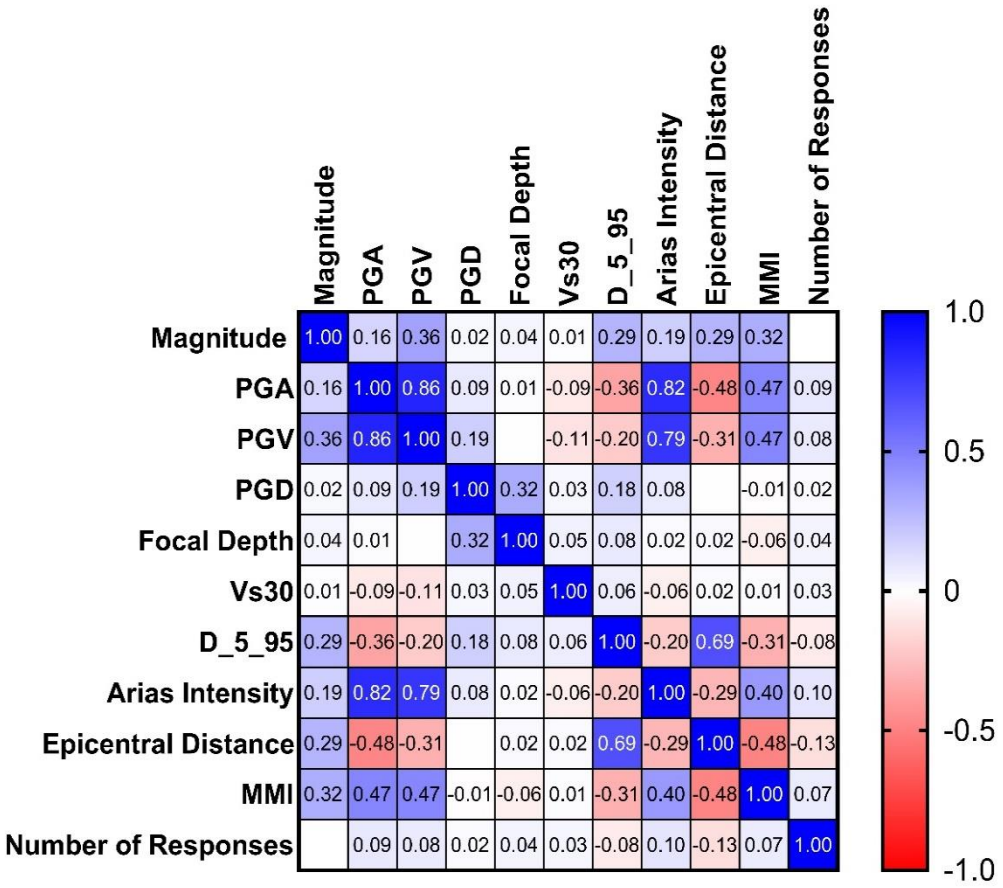


Figure G.1 The correlation matrix of the Strike-Slip database

H. P value summary of the Türkiye database correlation matrix for MMI-based relationships of variables

Table H.1 Descriptive statistics of Strike-Slip region dataset for MARS method

	P value summary	P (two-tailed)	Pearson R
MMI vs Magnitude	****	<0.0001	0.29
MMI vs PGA	****	<0.0001	0.43
MMI vs PGV	****	<0.0001	0.43
MMI vs PGD	ns	0.3515	-0.01
MMI vs Arias Intensity	****	<0.0001	0.34
MMI vs D_5_95	****	<0.0001	-0.24
MMI vs Focal Depth	**	0.0087	-0.06
MMI vs Epicentral Distance	****	<0.0001	-0.36
MMI vsVs30	ns	0.9501	-
MMI vs Number of Responses	**	0.0039	0.06