Seismic hazard assessment in Central Asia: combining site effects investigations and probabilistic seismic hazard

vorgelegt von
Master of Science
Shahid Ullah
aus Malakand, Pakistan

Von der Fakultät VI- Planen Bauen Umwelt
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
-Dr.-Ing.-

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. Yuri Petryna
Berichter: Prof. Dr.-Ing. Stavros Savidis
Berichter: Prof. Dr.-Ing. Paulo Bazzurro
Berichter: Prof. Dr. Stefano Parolai

Tag der wissenschaftlichen Aussprache: 19. Februar 2016

Berlin 2016

AUTHOR'S DECLARATION

I hereby declare that I have produced this thesis without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This thesis has not previously been presented in identical or similar form to any other German or foreign examination board.

Shahid Ullah

Potsdam,
08.06.2015

ABSTRACT

Central Asia is one of the world's most seismically active regions, with the highest level of seismic hazard. Usually seismic hazard is estimated considering ground motion at rock site, but since the ground motion could vary significantly over short distances due to local surficial geology, it is important to consider locally estimated site effects in seismic hazard assessment. Under the GSHAP project carried out in 1992-1999, seismic hazard was calculated at a global scale, including Central Asia. However, an update of this assessment is required in order to consider updated and more recent datasets. Therefore, purpose of this study is to: 1) assess the updated probabilistic seismic hazard at the regional level in Central Asia, and 2) consider the hazard assessment at the local level, including empirically-estimated site effects. As part of the GEM (Global Earthquake Model) initiative, this study is carried out within the EMCA (Earthquake Model Central Asia) project, which aims to calculate an updated cross border harmonized seismic hazard study at the regional level in Central Asia.

In this study, the seismic hazard is calculated for Central Asia using an updated earthquake catalogue with respect to the Soviet times and the GSHAP project. The earthquake catalogue has been assembled to cover until 2009 from different sources, containing both instrumental and historical events, and is homogenized to surface wave magnitude MLH from different magnitude scales. Shallow seismicity (< 50 km) is considered for the calculation of seismic hazard assessment in the region. Different seismic source models are used for the calculation of seismic hazard. These include the area source model and smoothed seismicity models. In smoothed seismicity models, the approaches of Frankel (1995) and Woo (1996) approach are used. In particular, along with the Gaussian kernel function with fixed correlation distance (smoothing bandwidth) approach of Frankel (1995), the adaptive kernel function proposed by Stock and Smith (2002) is also implemented inside the Frankel (1995) approach. The seismic hazard is calculated in terms of macroseismic intensity (MSK-64), intended to be used for the seismic risk maps of the region, using the open source software platform OpenQuake.

Most of the large cities in Central Asia lie on thick sediments, which influence the level of ground motion. Also, due to the current trend of urbanization, there is an urgent need to

address the site effects in an urban level seismic hazard assessment. For this purpose, the empirical site effects are evaluated by considering both earthquake and seismic noise recordings in terms of spectral ratios and from the array analysis in terms of shear wave velocity. In this study, using clustering and correlation analysis, the spatial resolution of ground motion variability is improved upon in terms of standard spectral ratios, using earthquakes recorded at a few selected sites for a relatively short amount of time, and seismic noise data collected over a denser grid. This method is applied to Bishkek, Kyrgyzstan, where a K-means clustering algorithm is used to identify three clusters of site response type based on their similarity of standard spectral ratios. The cluster's site responses are then adopted for sites where only single station noise measurements are carried out based on the results of correlation analysis.

Here a first attempt is made to take into account the influence of the shallow geological structure on the seismic hazard for Bishkek, Kyrgyzstan, by using a proxy of Vs30 that has been estimated from in-situ seismic noise array analyses, and considering response spectral ratios calculated by analysing a series of earthquake recordings of a temporary seismic network. To highlight the spatial variability of the observed ground motion, the obtained results are compared with those estimated assuming a homogeneous Vs30 value over the whole urban area, corresponding to rock site condition. The seismic hazard is evaluated in terms of peak ground acceleration (PGA) and spectral acceleration (SA) at different periods (frequencies).

The maximum hazard observed in the regional model reaches an intensity of around 8 in southern Tien Shan for a mean return period of 475 years. The maximum hazard estimated for some of the cities in the region, namely Bishkek, Dushanbe, Tashkent and Almaty, is between 7 and 8 (7-8), 8.0, 7.0 and 8.0 macroseismic intensity, respectively, for 475 years mean return period, using different approaches. Comparing these results, the current study shows that the hazard is generally higher by an order of 2 intensity units compared with that from the GSHAP project.

The maximum hazard observed for rock site condition at the urban level for Bishkek is 0.45 g at a period of 0.1 s with a maximum PGA of 0.21 g, for a 475 years mean return period. When site effects are included through the Vs30 proxy in the seismic hazard calculation, the largest spectral acceleration of 0.64 g is obtained for a period of 0.1 s. In terms of PGA, in this case

the largest estimated value reaches 0.31 g in the northern part of the city. When the variability of ground motion is accounted for through response spectrum ratios, the largest spectral acceleration reaches a value of 1.13 g at a period of 0.5 s. In general, considering site effects in the seismic hazard assessment of Bishkek leads to an increase in the estimated seismic hazard in the north of the city, which is thus identified as the most hazardous part within the study area and which is in fact further away from the faults and seismic sources. This study represents an update of the seismic hazard at regional and local scale.

ZUSAMMENFASSUNG

Zentralasien ist eine der seismisch aktivsten Regionen weltweit, was mit einer sehr hohen seismischen Gefährdung einhergeht. Um die Erdbebengefährdung zu quantifizieren, wird gewöhnlich die Bodenbewegung auf Festgestein zugrunde gelegt; da jedoch die Stärke und Dauer der Bodenbewegung im Falle eines Erdbebens aufgrund lokaler Standorteffekte räumlich stark variieren kann, ist es wichtig, diese lokalen Standorteffekte in die Gefährdungsanalyse miteinzubeziehen. Im Rahmen des GSHAP-Projekts in den Jahren 1992 bis 1999 wurde die Erdbebengefährdung auf globaler Ebene berechnet, einschließlich Zentralasiens. Zum gegenwärtigen Zeitpunkt ist jedoch eine erneute Evaluierung dieser Ergebnisse erforderlich, welche die neu verfügbaren Datensätze und die neu entwickelten Methoden berücksichtigt. Das Ziel dieser Studie besteht daher darin, 1) eine aktualisierte probabilistische seismische Gefährdungsanalyse auf regionaler Ebene in Zentralasien vorzunehmen und 2) Gefährdungsanalysen auch auf lokaler Ebene unter Berücksichtigung der vorherrschenden, empirisch bestimmten Standorteffekte durchzuführen. Dies steht im Einklang mit den Zielen von EMCA (Earthquake Model Central Asia), einem Regionalprojekt der GEM (Global Earthquake Model)-Initiative, deren Schwerpunkt auf aktualisierten und grenzüberschreitenden Erdbebengefährdungsstudien in Zentralasien liegt.

Als Grundlage für die Berechnung der seismischen Gefährdung dient ein im Vergleich zu den im vergangenen Jahrhundert durchgeführten Arbeiten aktualisierter Erdbebenkatalog. Basierend auf verschiedenen Quellen umfasst dieser Erdbebenkatalog sowohl historische als auch instrumentell aufgezeichnete Beben einschließlich des Jahres 2009; alle Ereignisse wurden auf die Oberflächenwellenmagnitudenskala harmonisiert. Obwohl der Katalog eine Reihe von tiefen Erdbeben enthält, wurden für die Gefährdungsanalyse nur oberflächennahe Beben (< 50 km Tiefe) berücksichtigt, da nur diese Beben zu einer starken Bodenbewegung und einer damit einhergehenden Gefährdung an der Erdoberfläche führen. Für die Erdbebenherde wurden bei der Berechnung unterschiedliche Modelle zugrunde gelegt. Dies umfasst zum einen geglättete Modelle (smoothed seismicity-Verfahren von Frankel (1995) und Woo (1996)) und Zonierungsmodelle (area source models), bei denen Gebiete ähnlicher Seismizität für die Berechung zugrunde gelegt werden. Unter Verwendung der Open-Source-Software-Plattform OpenQuake erfolgt die Berechnung der seismischen Gefährdung in Bezug

auf die makroseismische Intensität (MSK-64); diese Werte können dann in einem nächsten Schritt auch als Grundlage für die Risikokartierung der Region verwendet werden.

Die maximale Gefährdung auf Grundlage der regionalen Modelle erreicht eine makroseismische Intensität von ungefähr 8 bei einer durchschnittlichen Wiederkehrperiode von 475 Jahren für den südlichen Tien Shan. Die maximalen Gefährdungen, die für einige der Städte der Region, namentlich Bishkek, Dushanbe, Tashkent und Alma-Ata (Almaty), ermittelt wurden, liegen zwischen 7 und 8, 8.0, 7.0 und 8.0 bei einer durchschnittlichen Wiederkehrperiode von 475 Jahren. Ein Vergleich mit früheren Ergebnissen zeigt, dass die ermittelten Gefährdungsniveaus in der Region um bis zu zwei Intensitätsstufen höher liegt als die während des GSHAP- Projekts ermittelten.

Obgleich sich insbesondere für die Gebiete Zentralasiens, in denen die größten und wichtigsten Städte liegen, ein hohes Gefährdungsniveau zeigt, wird die Ausgangslage noch dadurch verschärft, dass die meisten dieser Städte auf dicken Sedimenten liegen, die einen erheblichen Einfluss auf die Stärke und die Dauer der Bodenbewegung haben. Daher ist es unabdingbar, diese Standorteffekte bei der Gefährdungsbeurteilung auf lokaler Ebene miteinzubeziehen. Hierzu erfolgt in einem ersten Schritt die Quantifizierung dieser Standorteffekte auf Grundlage von Messungen des seismischen Rauschens und der Analyse von Erdbebendaten, die mittels eines temporären seismischen Netzwerks im Stadtgebiet aufgezeichnet wurden. Unter Verwendung von Cluster- und Korrelationsanalysen kann die räumliche Auflösung für das Auftreten solcher Standorteffekte signifikant verbessert werden. Hierzu werden Messungen des seismischen Rauschens, welche nur kurze Zeit in Anspruch nehmen und an einer Vielzahl von Messpunkten im ganzen Stadtgebiet durchgeführt werden können, einem bestimmten Cluster des seismischen Netzwerks zugeordnet, für den detaillierte Informationen über die Standorteffekte existieren. Dieses Verfahren wird in Bishkek, der Hauptstadt Kirgistans, angewandt, wo mittels eines K-means-Cluster-Algorithmus das Stadtgebiet in drei Gebiete aufgeteilt werden kann, für die ähnliche Standorteffekte zu erwarten sind.

Auf Grundlage dieser Ergebnisse wird weiterhin versucht, den Einfluss von oberflächennahen geologischen Strukturen auf die seismische Gefährdung auf lokaler Ebene zu berücksichtigen. Hierzu wurde die mittlere Scherwellengeschwindigkeit in den obersten 30 m des Bodens (Vs30) aus mehreren Arraymessungen, die im Stadtgebiet von Bishkek durchgeführt wurden,

berechnet. Um die räumliche Variabilität der beobachteten Bodenbewegungen zu unterstreichen, wurden die ermittelten Ergebnisse mit den Resultaten unter Annahme einer im gesamten Untersuchungsgebiet homogenen Vs30-Verteilung verglichen. Die seismische Gefährdung wird anhand der höchsten Bodenbeschleunigung (PGA) und der spektralen Beschleunigung (SA) für verschiedenen Perioden (Frequenzen) evaluiert.

In Bishkek liegt die maximale Gefährdung auf lokaler Ebene für Festgestein bei 0.45 g bei einer Periode von 0.1 s mit einer maximalen PGA von 0.21 g für eine durchschnittliche Wiederholperiode von 475 Jahren. Werden lokale Bodeneffekte mit Hilfe der mittleren Schwerwellengeschwindigkeit in den obersten 30 m in der Gefährdungsberechnung berücksichtigt, erreicht die höchste Spektralbeschleunigung einen Wert von 0.64 g bei einer Periode von 0.1 s. Für die maximale Bodenbeschleunigung werden die höchsten Wert in der Größenordnung von 0.31 g im nördlichen Teil der Stadt erreicht. Wenn die Variabilität der Bodenbewegung auf Grundlage der Antwortspektren berücksichtigt wird, erreicht die höchste Spektralbeschleunigung einen Wert von 1.13 g für eine Periode von 0.5 s. Generell führt die Berücksichtigung von lokalen Bodeneffekten in der seismischen Gefährdungsanalyse von Bischkek zu einem starken Anstieg der zu erwartenden seismische Gefährdung im Norden der Stadt. Insofern kann dieser Bereich als der am stärksten gefährdete des Untersuchungsgebietes identifiziert werden, obwohl er weiter entfernt von den Verwerfungszonen und seismischen Quellen liegt. Daher führt die Wechselwirkung zwischen einem hohen Gefährdungsniveau auf regionaler Ebene in Zusammenhang mit ungünstigen lokalen Gegebenheiten zu einer signifikanten Gefährdung weiter, dicht besiedelter Teile Zentralasiens.

ACKNOWLEDGEMENTS

First of all, I would like to thank my parents for their support and encouragement to start the PhD adventure. Their continuous moral support made it easy and comfortable to make it till the end.

I am thankful to my supervisor, Dr. habil. Stefano Parolai, for his continuous support, both scientific and moral, throughout the PhD process. Thank you for providing me with numerous opportunities to work in the field and to work with leading international experts. Thank you for believing in me and for all the efforts you made to make me a better human being and a scientist.

It would be very hard to imagine my PhD work without the assistance of Dr. Dino Bindi and Dr. Marco Pilz, who helped me at every step of my PhD. I felt so privileged to have been around you people. You are not only respected scientists, but very compassionate colleagues and friends. Thank you for helping me and all the good time we spent together.

Thank you Annamaria Saponaro, my office mate and lunch buddy, for being such a good friend and a colleague. We spent most of our PhD time sitting face to face each other in the office. Thank you, Marc Wieland for being a good friend and all the support whenever I needed. Thank you Dr. Massimiliano Pittore, for guiding and helping me during my PhD work.

I would also like to thank my friend Galina Kulikova, the only person whom I knew in Germany, before coming here. Thank you for helping me with finding accommodation at the start, and for the moral support. Thank you Susanne Köster, my section's secretary, for helping me right from the beginning with all the bureaucratic issues. Thank you Dorina Kroll for helping me whenever I needed. Thank you, Dr. Claus Milkereit for giving the useful insights during my PhD work. Thank you, Dr. Kevin Fleming for improving my technical writing. I would like to thank all my colleagues and friends in Central Asia for their support during field work. Thank you, Dr. Pagani, Dr. Weatherill, and Dr. Zuccolo in Pavia for the support and collaboration. Thank you, Dr. Danciu at the ETH Zurich for the continuous support with the OpenQuake. Thank you Michael Haas, Tobias Boxberger, Bojana Petrovic,

Alina Motschmann, Michele Pantaleo, Elena Nikolaeva, Angelo Strollo, Matteo Picozzi, Christop Bach, Katrin Kieling, Ade Anggraini, Silke Eggert, Camilla Cattania, Olga Zakharova, Mehdi Nikko, Sara Atito, Sreeram Reddy, Jacqueline Tema Salzar, Nicole Richter, Katja Müller, and all my friends and colleagues at the institute for all the good time we spent together. Thank you my young friends at the student dormitory Greibnitzsee, whom I met for a short time, but spent quality time together.

I am very thankful to my country, Pakistan, the Higher Education Commission of Pakistan, my university of Engineering and Technology, Peshawar, and my fellow Pakistanis for providing me this opportunity to study abroad.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

1 INTRODUCTION

The region of Central Asia considered in this study includes the territory of five countries, namely Kazakhstan, Kyrgyzstan, Tajikistan, Uzbekistan and Turkmenistan (Figure 1.1). It is characterized by the presence of the Tien-Shan and Pamir orogenic belts, whose tectonic regime is defined by the convergence of the Indian and Eurasian plates (Molnar and Tapponnier, 1975, 1978). This intra-continental collision region is highly seismically active and capable of generating large earthquakes, as evident by the historical seismicity. Some of the world's largest earthquake have occurred in northern Tien Shan, such as the MLH 8.3 1989 Chilik earthquake, the MLH 8.2 1911 Kemin earthquake, and the MLH 7.2 1992 Suusamyr earthquake. All major cities in the countries of Central Asia have experienced large earthquakes within the last century. This includes the MLH 7.3 1948 Ashgabad earthquake, the MLH 5.2 1966 Tashkent earthquake, the MLH 7.3, 1907 Karatag earthquake which occurred close to Dushanbe, and the MLH 7 1885 Belovodski earthquake near Bishkek.

The current study is carried out within the framework of Earthquake Model Central Asia (EMCA) project[1]. EMCA is a regional partner program of Global Earthquake Model (GEM) Foundation[2], a global public/private initiative that aims at developing open standards and tools to calculate and communicate hazard and risk worldwide. The aim of the EMCA project is the evaluation of the seismic hazard and risk in Central Asia, following an end-to-end approach where all the components, from seismic hazard to vulnerability and exposure, were re-assessed and updated.

[1] http://www.emca-gem.org/
[2] http://www.globalquakemodel.org/

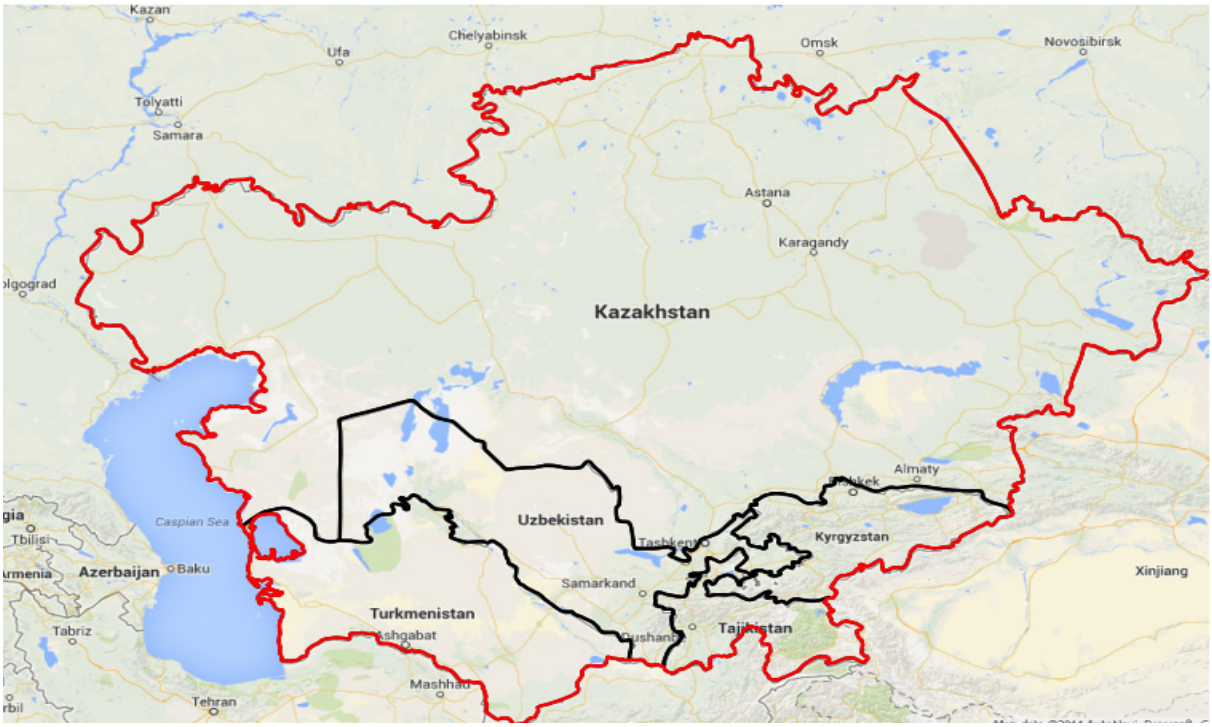


Figure 1.1 Map of the study region, which includes Kazakhstan, Kyrgyzstan, Tajikistan, Uzbekistan, and Turkmenistan.

1.1 Background

The Global Seismic Hazard Assessment Program (GSHAP, Giardini 1999) was the first project to carry out a harmonized seismic hazard assessment at the global level. This project indicated that Central Asia, which is one of the most seismically active regions in the world, is characterized by peak ground acceleration at 10% probability of exceedance in 50 years as high as 9 m/s^2. However, following the availability of new datasets and tools, an update of this hazard assessment is required at the global scale. This is being realized by the project GEM Foundation. Like EMME (Earthquake Model Middle East[3]), EMCA is a regional partnership of GEM which aims to calculate a cross-border harmonized seismic hazard and risk at the regional level. However, most of the time seismic hazard is calculated at rock sites without considering the effect of local surficial geology. This not only leads to biased results, but also in many cases to a significant underestimation of seismic hazard.

The main aim of this study is to produce updated cross-border harmonized seismic hazard maps for Central Asia, using updated, homogenized earthquake catalogues and region-specific intensity prediction equations. Most of the cities in Central Asia lie on thick

[3] http://www.emme-gem.org

sediments (Parolai et al., 2010, Pilz et al., 2013) which may greatly influence the ground motion at the surface (as explained in chapter 3). Also, considering the current trend of urbanization where people are moving towards big cities for better facilities, any urban-level seismic hazard assessment needs to properly address site effects. Considering the importance of different proxies for site effects, another aim of this study is to better quantify site effects at the local level and incorporate the different proxies of site effects into the hazard analysis at the urban level.

1.2 Organization and Outline

The second chapter of this thesis is dedicated to introducing the various concepts behind seismic hazard. The procedure to calculate the seismic hazard is described in detail for different approaches, along with their differences. Different types of seismic sources are discussed with respect to seismic hazard. The two main approaches of smoothed seismicity i.e., Frankel (1995) and the Woo (1996) are described in detail. Finally, the different types of uncertainties in seismic hazard assessment are discussed.

The third chapter describes the methods for the estimation of site effects. These include both numerical and empirical site effects estimation methods.

The fourth chapter describes the clustering analyses which are used for the classification of sites in terms of site effects in Bishkek, Kyrgyzstan. In particular the K-mean clustering algorithm is explained and discussed. The different indices are outlined for the validation of the results from the clustering analysis.

The fifth chapter explains the regional seismotectonics of Central Asia. The dominant mechanism of tectonics and seismicity is explained and the major faults responsible for large events and seismicity in the region are discussed. The history of seismic hazard studies in Central Asia is discussed, as well as an overview of existing seismic hazard studies at the regional level.

The sixth chapter deals with the seismic hazard models used for Central Asia in this study. First, the earthquake catalogue and its processing, e.g., i.e., clustering, completeness and recurrence analyses are explained in detail. Second, the seismic source models of area sources

and smoothed seismicity models are presented, which includes the smooth seismicity model of Frankel (1995) with the adaptive kernel proposed by Stock and Smith (2002).

The seventh chapter explains the results of the different source models used in chapter 6. The results are compared with each other, with respect to their underlying methodologies. The results are also quantified for the major cities in Central Asia. These results are compared with earlier studies in the region.

The eighth chapter describes the empirical site effects study carried out in Bishkek, Kyrgyzstan. The local geology of the city is described. The clustering and correlation analyses are explained for improving the spatial resolution of ground motion variability. The results of the clustering analysis are discussed along with the validation indices. These results are also discussed with respect to the geological structure of the city.

The ninth chapter describes the procedure for including the empirical site effects in the hazard analysis at the urban level for Bishkek, Kyrgyzstan. Differences arising in the results of the seismic hazard assessments are discussed for both rock site conditions and for when site effects are included. The differences in results are also discussed for different proxies of site effects.

The tenth chapter describes the conclusion of the thesis and suggestions for the future work.

2 INTRODUCTION TO SEISMIC HAZARD ASSESSMENT

Hazard is generally estimated as the mean rate of exceedance of some chosen parameter, which is often described numerically on a per-year basis (Hanks and Cornell, 2011). Hazard can be man-made or natural. Seismic hazards include all the potentially destructive effects of earthquakes, such as strong ground motion, liquefaction, landslides etc. Except for surface fault ruptures and tsunamis, all of the destructive effects of earthquakes are directly related to ground movement induced by the passage of seismic waves (Bommer, 2002). Seismic hazard analysis deals with the prediction of the severity and likelihood of occurrence of these effects at a particular site (e.g. Crowley, 2005 and the references therein). From this arises Probabilistic Seismic Hazard Assessment (PSHA), which is the calculation of the probability of exceedance of a certain level of ground shaking (e.g., 10 %) at a particular site within a certain future time of exposure (e.g., 50 years). Below is a brief description of different approaches to seismic hazard assessment (SHA).

2.1 Probabilistic Seismic Hazard Assessment (PSHA)

The classical PSHA was first introduced by Esteva (1967), Cornell (1968) and Merz and Cornell (1973). This methodology is used in the current study. This technique basically consists of four steps (Reiter 1990).

1) Identification and characterization of earthquake sources. The probability distribution of potential rupture locations within the sources is characterized. Uniform probability distributions are assigned to each source, which implies that earthquakes are equally likely to occur at any point within the source zone. These distributions are combined with the source geometry to obtain the corresponding probability distribution of source-to-site distances.
2) Characterization of the temporal distribution of earthquake recurrence. A recurrence relationship is used to characterize the seismicity of each source.
3) Estimation of the ground motion parameters at a site due to any possible size earthquake occurring at any possible point in each source zone, using attenuation relationships.

4) Calculation of hazard curves. The uncertainties in earthquake location, earthquake size and ground motion parameter are combined to obtain the probability that some ground motion parameter will be exceeded within a particular time period.

These four steps are represented graphically in figure 2.1. Different earthquake sources are defined depending on the available information. An earthquake source can be defined either in the form of a point source (smoothed seismicity approach), an area source (zonation model), or as a line source (fault model). The seismic source model describes the spatial and temporal distribution of earthquakes in a region. Both point source and zonation models are used in the current study. Due to incomplete knowledge of faults in the region, the fault model is not used in the current study.

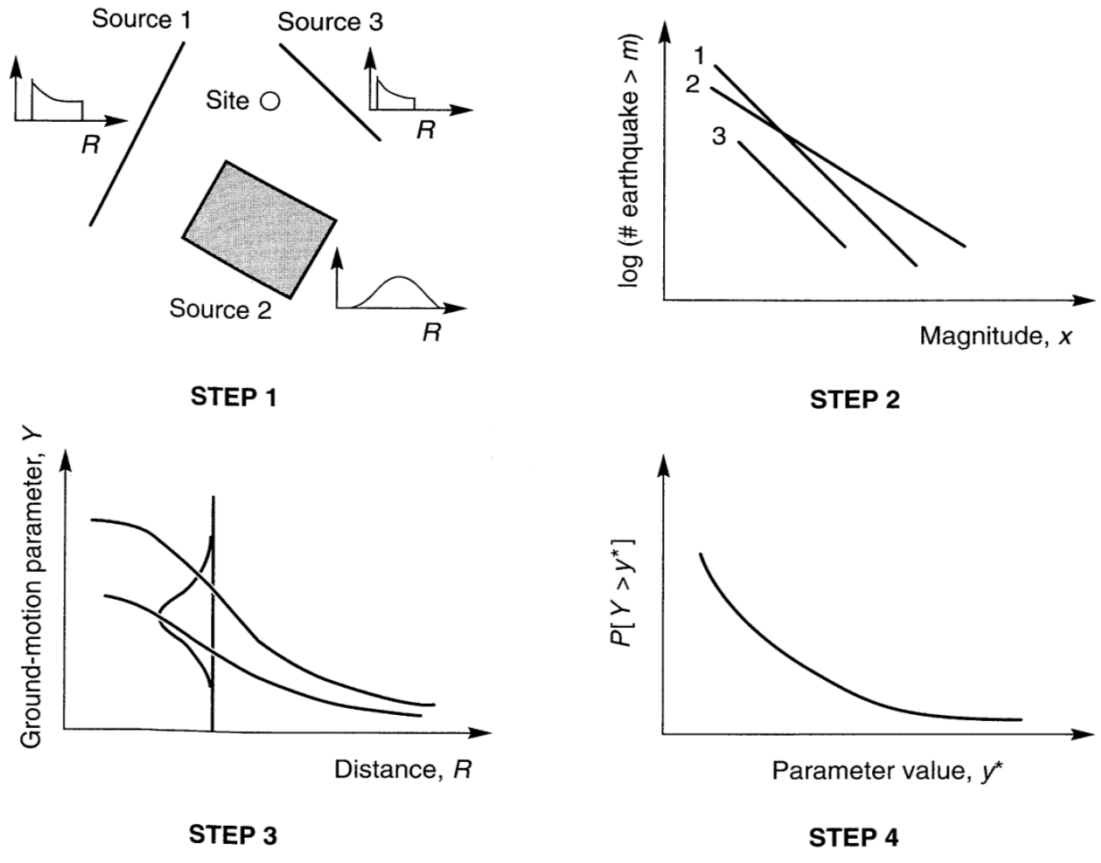


Figure 2.1 The four steps involved in the process for calculating classical PSHA (Kramer, 1996). Step 1 shows identification of seismic sources, step 2 is about the recurrence relationship for each source, step 3 is estimation of ground motion parameters for different combination of distance and magnitude, and step 4 is calculation of the final hazard curve.

Different relationships are used to describe the temporal distribution of earthquake recurrence. These relationships describe the average rate at which an earthquake of some magnitude will be exceeded. A basic assumption of PSHA is that the recurrence law obtained from past seismicity is appropriate for the prediction of future seismicity (Kramer, 1996). Gutenberg and Richter (1944) (GR), one of the first relationships describing the magnitude-frequency relationship, is given below.

$$\boldsymbol{Log(N) = a - bM} \tag{2.1}$$

where N is the mean number of events per year for a magnitude M or larger; a represents the overall activity of the seismic source (log number of events with $M >= 0$) while b represents the relative distribution of small to large earthquakes. Considering a one year period, a tells us that on average, once per year, an earthquake of magnitude (a/b) or larger occurs. A larger b value represents a region dominated by relatively smaller magnitude earthquakes, while a smaller b value represents a region with relatively higher magnitudes events. a and b values are obtained using a regression analysis of an earthquake catalogue. For the PSHA based on the Poissonian assumption, the earthquake catalogue must be declustered (i.e., the removal of foreshocks and aftershocks) and processed for completeness analysis. The applications employed in the processing of the earthquake catalogue are carried out in chapter 6. Other recurrence relationships are also used for different kind of sources and scenarios, such as bounded GR recurrence laws, characteristic earthquake recurrence laws, quadratic recurrence law, bilinear recurrence law, etc. (Kramer, 1996).

In the third step, ground motion parameters, such as peak ground acceleration (PGA), peak ground velocity (PGV), peak ground displacement (PGD), or macroseismic intensity are calculated at a site for all possible combinations of distances and magnitudes. This is carried out using Ground Motion Prediction Equation (GMPE) or Intensity Prediction Equation (IPE). Conditional probability is used to calculate ground motion. The aleatory uncertainty which is physical variability inherent to the unpredictable nature of future events, is handled by means of sigma in the GMPE or IPE. The choice of a GMPE or IPE is very important for the hazard assessment because of its significant influence on the final results.

Finally, the hazard curves are calculated for individual source zones, and combined to express the total hazard at a particular site. A hazard curve shows the mean annual rate of exceedance (MARE) of different levels of ground motion at a site. The annual rate of exceeding a

particular value, *y**, of ground motion parameter, *Y*, is calculated for one possible earthquake at one possible source location and then multiplied by the probability that that particular magnitude earthquake would occur at that particular location. The process is repeated for all possible scenario and the probabilities are summed up. The process can be expressed in its mathematical form by the total probability theorem as follows (Kramer, 1996):

$$\boldsymbol{\lambda_{y^*} = \sum_{i=1}^{N_s} v_i \int_{m_0}^{M} \int_{r_0}^{R} P[Y > y^* | m, r] \, f_{Mi}(m) f_{Ri}(r) dm \, dr} \tag{2.2}$$

where i ranges from 1 to the number of total sources, v_i is the mean rate of earthquake occurrence for magnitude m of source i, $P[Y > y^*|m,r]$ is the probability that a given earthquake of magnitude m and distance r will exceed ground motion level y^* due to aleatory variability incorporated through sigma in GMPE or IPE relationship and $f_{Mi}(m)\ and\ f_{Ri}(r)$ are the probability density functions for magnitude and distance, respectively. Please note that since the joint distribution of *m* and *r* can not be split in equation (2.2), the integral in equation (2.2) is computed numerically, by defining the range of value of magnitude *m* and distance *r* into bins (Kramer, 1996).

The MARE (λ) can be used in a Poisson model to calculate the probability of exceedance of a given ground motion level P(y), for a certain time (T) of exposure (e.g., design life of a structure). This is possible assuming that earthquakes have a Poissonian distribution in time, which means that earthquakes occur independently from each other. In its mathematical form, it is described as:

$$\boldsymbol{P(y) = 1 - e^{-\lambda_y T}} \tag{2.3}$$

Cornell and Winterstein (1986) carried out an investigative study on the applicability of Poisson and non-Poissonian models. Their findings suggest that a Poissonian model for practical seismic risk analysis is suitable, except when the seismic hazard is dominated by a single source showing strong characteristic temporal behaviour and for which the time interval since the previous significant earthquake is greater than the average inter-event time. For this reason and others related to simplicity, ease of use, and lack of sufficient data to support more sophisticated models, the Poissonian model is the most widely used in PSHA (Kramer, 1996). Since no characteristic fault sources are used in this study, the Poisson model is used here.

2.2 Deterministic Seismic Hazard Assessment (DSHA)

In general, DSHA uses discrete events and scenarios for hazard assessment. DSHA share similar steps with PSHA with some important differences. The basic four steps involved in DSHA are illustrated in figure 2.2. The first step is similar to PSHA, which involves the identification of seismic sources which are capable of generating earthquakes of a significant size and could cause considerable seismic hazard at the site. Unlike the PSHA, a controlling earthquake is selected for each source, usually referred to as the Maximum Considered Earthquake (MCE). It is the largest earthquake expected to occur within each source zone (Krinitzsky, 2002). The magnitude of MCE can be determined on the basis of geological evidence or based on historical seismicity. These scenario earthquakes are considered at a chosen distance (which is often the closest location) within each source zone with respect to the site.

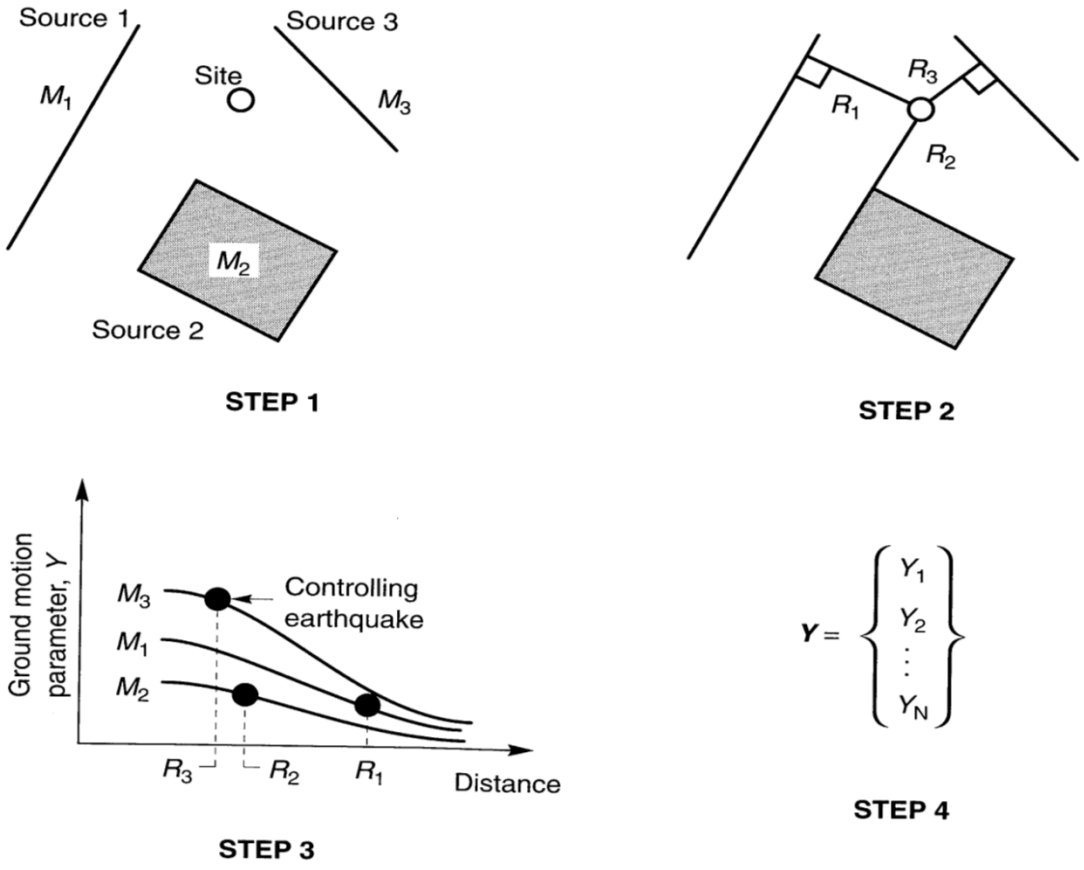


Figure 2.2 The four steps involved in the process for calculating DSHA (Kramer, 1996). The 1st step is the identification of seismic sources, the 2nd step is the calculation of a chosen distance from the site (often the closest one), the 3rd step is estimation of the ground motion parameter for the given distance and magnitude, and the 4th step is selection of the hazard from the dominant scenario.

Once the scenario earthquakes are selected for each source, the ground motion is calculated using GMPE or IPE. Due to the presence of scatter in the data, though some studies use the mean values from the predictive relationships, the current trend is to use the 84-percentile motions, which corresponds to one logarithmic standard deviation above the logarithmic mean (Krinitzsky, 2002). From this process, the ground motion is calculated at the site from each source, and the one that dominates, is considered for further application.

2.3 Differences between PSHA and DSHA

Although PSHA and DSHA have a lot of things in common as both supports to perform seismic hazard analysis, there are some substantial differences between the two approaches. DSHA generally considers the significant earthquakes at specified locations, whereas PSHA integrates the results over a wide range of possible earthquake magnitudes and source to site distances. Regarding presentation and justification of design basis ground motions (DBGM), DSHA has an important advantage of transparency in progressing from a deterministic earthquake to DBGM, whereas in PSHA this is a more complicated process handled through theorems of probabilities and a process called disaggregation, because a range of magnitudes and distances contribute to the calculated hazard (Hanks and Cornell, 2001). Another difference between the two is the handling of uncertainties. PSHA incorporates the aleatory uncertainties in the hazard analysis by means of sigma, whereas DSHA does not consider this and relies mainly on diverse expert opinion.

However, the main fundamental difference, philosophically and practically, between the two is that PSHA carries units of time while DSHA does not (Hanks and Cornell, 2001). In PSHA the hazard is defined as the mean rate of exceedance of some chosen ground motion amplitude, whereas in DSHA the hazard is defined as the ground motion at the site resulting from the main controlling earthquake.

2.4 Smooth seismicity approaches

The standard assumption in SHA is to represent the seismicity inside each area source as a homogenous Poisson process. This seismicity is further parameterized by GR magnitude frequency relationship and fixing a maximum magnitude for that source. However, most of

the time, the assumption of homogeneous seismicity within each seismic source in space and stationarity in time rarely hold. Also, the larger a source zone is, the more likely seismicity is non-homogenous and clustered in different parts of the zone. This violates the assumption of homogenous seismicity. Moreover, if the source zone is smaller, it is more likely to have temporal fluctuations in seismicity, which violates the assumption of stationarity (Woo, 1994). Hence, the smoothed seismicity approaches are used in regions having a lack of good knowledge of the relevant seismogenic structures. In smoothed seismicity approaches, the seismicity is smoothed or spread at the epicentre based on different relationships. These approaches are based on the hypothesis that the past distribution of seismicity provides information about the location of future events. Some of the approaches are explained below.

2.4.1 Frankel (1995) approach

The Frankel (1995) approach is a smoothing approach which is based on the idea that future seismicity will occur in a place where seismicity has happened in the past. Like other smoothing approaches, it mainly relies on earthquake catalogues. This method is based on recurrence relationships such as the GR relationship (equation 2.1), meaning that the b value in this equation is estimated beforehand. The seismicity is smoothed using a (fixed) Gaussian kernel to calculate the smoothed activity rate.

In this method, the area is divided into a grid of source points, e.g., 0.2x0.2 degrees. The cumulative number of events (n_i) above a threshold magnitude in each cell is counted. The cumulative number of events in each cell represents the maximum likelihood estimate of 10^a for that cell (Weichert, 1980) for earthquakes above a reference magnitude. The a is the same activity rate as used in equation 2.1. The cumulative number of events (number of events above M_{ref}) for each cell is then converted to incremental activity rates (number of events from M_{ref} to $M_{ref} + \Delta M$) using the Herman's (1977) formula, given as;

$$10^{\text{incremental a-value}} = 10^{\text{cumulative a-value}}(10^{b\Delta M} - 10^{-b\Delta M}) \quad (2.4)$$

Finally, the grid of incremental values is smoothed spatially by multiplying it by a Gaussian function with a fixed correlation distance c1 as shown below:

$$\tilde{n}_i = \frac{\sum_j n_j \exp\left(\frac{-\Delta_{ij}^2}{c_1^2}\right)}{\sum_j \exp\left(\frac{-\Delta_{ij}^2}{c_1^2}\right)} \tag{2.5}$$

where $\tilde{n}_i$ represents the smoothed activity rate for cell i, and for the total observation period (i.e., not per year), and Δ_{ij} represents the distance between cell i and j. The denominator in equation 2.5 normalizes $\tilde{n}_i$ to preserve the total number of events. In the above equation, the sum is taken over cells j within a distance of 3 x c1 of cell i as is done in the original code of Frankel (1995). In the original approach of Frankel (1995), a fixed correlation distance c1 of 50 km is used. In the original Frankel (1995) approach, this same smoothing distance is used everywhere irrespective of the density of seismicity or the sizes of magnitude at a place.

Finally, the annual rate of exceeding a given ground motion u_0 at a site is determined from summing over distance and magnitude as:

$$\lambda(u > u_0) = \sum_k \sum_l 10^{\left[\log\left(\frac{N_k}{T}\right) - b(M_l - M_{ref})\right]} P(u > u_0 | D_k, M_l) \tag{2.6}$$

where N_k represents the total of $\tilde{n}_i$ values for cells within a certain distance increment of the site, k is the index for distance and l is the index for the magnitude bin, and T is the time in years of the earthquake catalogue used to determine N_k. In the above method, the parameter b of the GR relationship can be either uniform or gridded. In case of it being uniform, the same b value is used for each grid point, whereas in the case of it being gridded, a separate b value is calculated beforehand for each grid point.

2.4.2 Woo (1996) Approach

The Woo approach (1996) is a zone free smoothing method based on the concept of fractal geometry and self-organized criticality. Differently from the Frankel (1995) approach, this method does not require the use of a recurrence relationship such as the Gutenberg-Richter equation to find the activity rates. Instead, the activity rates are determined for each magnitude interval according to a magnitude-dependent probabilistic smoothing procedure that is applied directly to the epicentres listed in the earthquake catalogue. The contribution of each event to the seismicity of the region is smeared over a distance that is magnitude dependent.

In this method, the cumulative activity rate density, λ, at a point x is computed over all the events of each magnitude interval, in which contribution of each event is inversely weighted by its effective return period.

$$\lambda(M,x) = \sum_{i=1}^{N} \left[\frac{K(M,x-xi)}{T(xi)}\right] \tag{2.7}$$

Please note the difference in the meaning of the symbols λ used in equation (2.6) for the Frankel approach and equation (2.7) above. These symbols are used to preserve the original symbols used in their respective publications. Furthermore, N is the total number of events in the catalogue, xi represents the coordinates of the epicentre being considered and T(xi) is the effective observation time period, while K(M,x) is a multi-variate probability density function in which the seismicity is smeared over distance through a magnitude dependent relationship expressed as below:

$$K(M,x) = \frac{n-1}{\pi H^2}\left[1 + (\frac{r}{H})^2\right]^{-n} \tag{2.8}$$

Here, r is the epicentral distance (km) and n is referred to as the kernel fractal scaling index. Its value lies between 1.5 and 2, corresponding to a cubic or quadratic decay of the probability density function with epicentral distance. A value of 1.5 is used in this study, as it has a negligible influence on the hazard estimates (Woo 1996, Baeuval et al., 2006, Zuccolo et al., 2013). H is a bandwidth for normalizing epicentral distances and is a function of magnitude.

$$H = ce^{dM} \tag{2.9}$$

where c and d are constants and are determined on the basis of location of events in the catalogue. These parameters are estimated following the procedure proposed by Molina et al. (2001), that is: (1) Arrange the events into different magnitude bins, e.g., 0.5 magnitude bin. (2) For each event within the same magnitude bin, the nearest epicentre distance is determined. (3) The average of all minimum distances within the same magnitude bin is determined. (4) The c and d parameters are estimated using a least square fit to the observed distance and magnitude.

Here also the activity rates are calculated at grid points, e.g., 0.2x0.2 degrees. The c and d parameters are calculated for all gridded points at which the activity rates are estimated.

Figure 2.3 shows a typical relationship for the magnitude-bandwidth for Central Asia for a magnitude bin of 0.5. This shows that for smaller magnitudes, the smoothing distance is smaller compared to the larger smoothing distance for larger magnitude events. This is consistent with seismicity observations where smaller events show more spatial clustering than larger events. This also suggests large uncertainties in the location of large earthquakes compared to the smaller events.

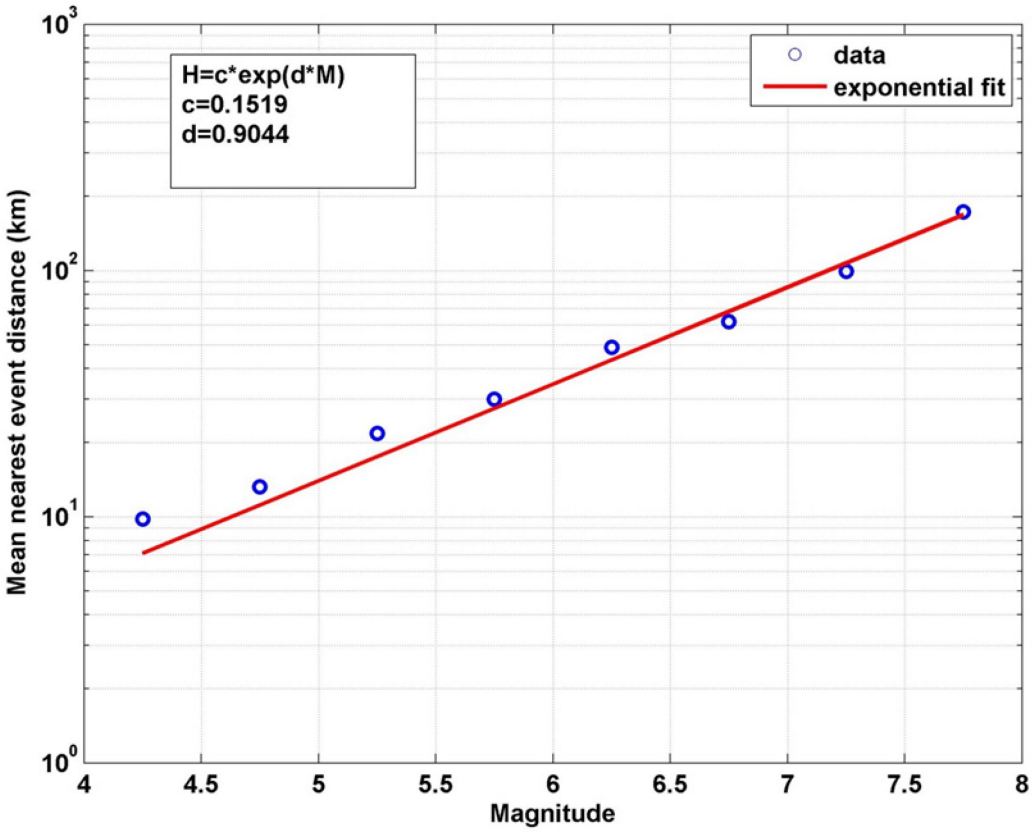


Figure 2.3 Magnitude-bandwidth relationship for estimating parameters *c* and *d* at a single site following the Woo (1996) approach.

In addition to an effective observation time period, the Woo (1996) approach also requires the uncertainties in magnitude and epicentral location of an event. Since the smoothing approaches are based on historical seismicity, which does not allow for the occurrence of events greater than the maximum observed, the uncertainty in magnitude ensures the possibility of an event greater than has been historically observed.

After the estimation of the activity rate of each class of magnitude bin for each point source, the Intensity Prediction Equation (IPE) or Ground Motion Prediction Equation (GMPE) is employed to compute the seismic hazard.

2.5 Types of uncertainties in seismic hazard assessment

Generally the uncertainties in SHA are categorized into the following two types: Aleatory and Epistemic. Aleatory uncertainty is related to the scatter in predictive relationships. It is a physical variability which is inherent in the unpredictable nature of future events and is directly incorporated into the hazard calculations. This uncertainty results from randomness in the source rupture mechanism and variability and heterogeneity of the source, the path along which the seismic rays travel and the local site conditions (SSHAC, 1997; Crowley, 2005). In other words, this uncertainty is a contribution from nature and is not reduced by gathering more data. Aleatory uncertainty is handled and quantified by means of the standard deviation (normally called sigma) in predictive relationships such as GMPE or IPE. Figure 2.4 shows the incorporation of aleatory variability (sigma) inside the predictive relationship into seismic hazard. In this figure, the conditional amplitude of a ground motion is first obtained for a particular distance and magnitude, and then the probability of exceedance of a certain level of ground motion is obtained by utilizing the sigma of the attenuation relationship.

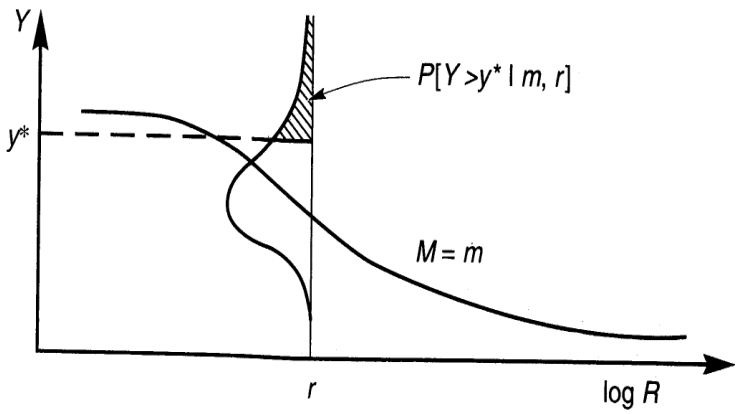


Figure 2.4 Schematic illustration of the conditional probability of exceeding a particular value of a ground motion parameter for a given magnitude and distances, implying the use of aleatory uncertainty (sigma) (Kramer, 1996).

Aleatory variability has two components, the intra-event variability, and the inter-event variability. Intra-event variability is the variability in ground motion from one location to

another (from station to station) at the same distance and with the same site classification during one earthquake. Whereas inter-event variability is the variability in ground motion at the same site from one earthquake to another with the same magnitude and rupture mechanism (Crowley, 2005). The total aleatory variability is therefore the square root of the sum of the squares of the intra- and inter-event variability.

Epistemic uncertainty on the other hand arises from a lack of knowledge, and expresses uncertainty concerning the correct value of the median. Epistemic uncertainty could be reduced by gathering more information. This uncertainty is due to the subjective judgments in the definition of various components of seismicity and hazard models. In the following are the components of the SHA process that lead inevitably to the subjective decisions and hence to creation of epistemic uncertainties (Bommer, 2002).

- Definition of seismic source zones' boundaries.
- Completeness of the earthquake catalogue, which affects the recurrence relationship in PSHA.
- Assigning the maximum magnitude Mmax to each seismic source zone.
- Selection of attenuation or predictive relationships such GMPE or IPE for a region.
- Assignment of recurrence intervals to characteristic earthquakes.

Since epistemic uncertainty arises from a lack of knowledge, it can't be quantified like aleatory uncertainty in hazard assessment. These uncertainties are included in the hazard assessment using different approaches (e.g. Ordaz and Arroyo, 2016). In this study, the epistemic uncertainty is considered in the SHA in the form of a logic tree, which makes use of the simultaneous consideration of different options for the various input parameters. Figure 2.5 shows a sample logic tree incorporating epistemic uncertainty in an attenuation model, recurrence relationships and the maximum magnitude. The weight factors to each branch are assigned based on expert judgment. These weight factors are interpreted as the relative likelihood of that model being correct. The sum of the weights in each branching level is equal to 1.0. At the tip of each branch, a hazard estimate is obtained. This estimated hazard is then multiplied with a relative weight that is obtained by multiplying the weights of the branches in that particular calculation. The final hazard is then obtained from all the branches of the logic tree.

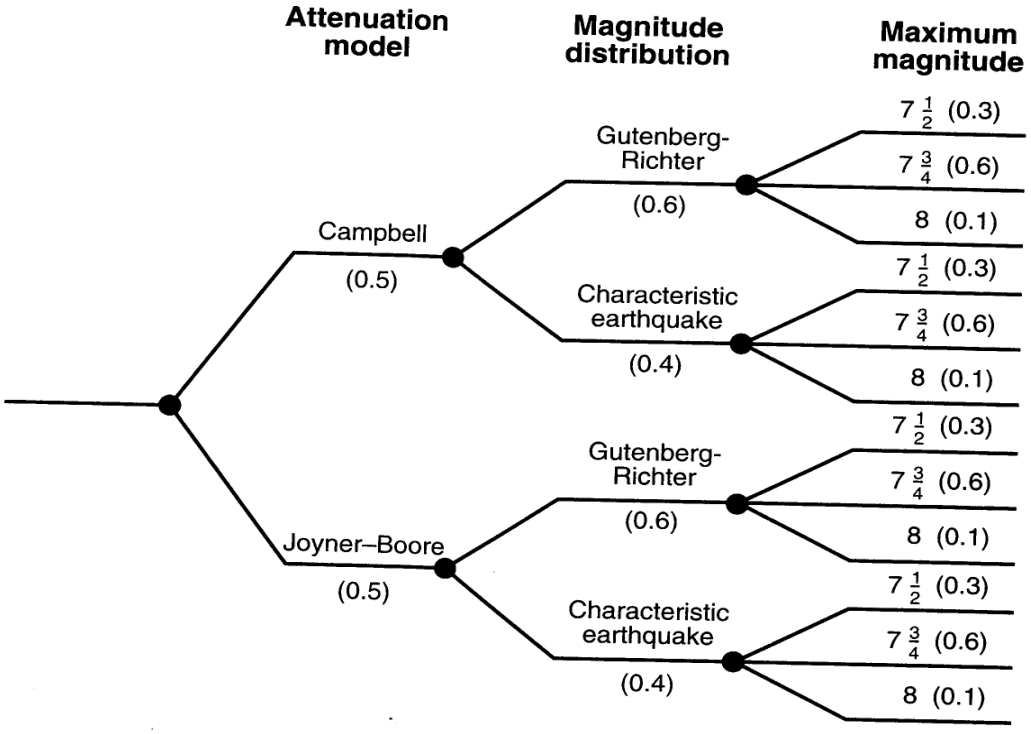


Figure 2.5 Sample illustration of a logic tree for the incorporation of epistemic uncertainty (Kramer, 1996).

3 METHODS FOR THE ESTIMATION OF SITE EFFECTS

It has long been known that earthquake motion on loose soil is greater than on rock. Daniel Drake (1815) in his book on the 1811-1812 New Madrid sequence writes “The convulsion was greater along the Mississippi, as well as along the Ohio, than in the uplands. The strata in both valleys are loose. The more tenacious layers of clay and loam spread over the adjoining hills suffered but little derangement.”

The nature of the local soil cover plays a very important role in the characteristics of surface motion and the damage caused by an earthquake. During earthquake-induced ground motion, soil layers act like a filter. They can change the frequency content of the input ground motion, while amplifying and de-amplifying the input bedrock motion at the surface depending upon frequency content and acceleration of the input motion and the dynamic properties of the soil deposits. Surficial soil deposits can also increase the duration of the ground motion due to the trapping of the seismic waves in the low velocity layer.

The ground motion recorded at a point depends upon several factors. The most important of these are: 1) the radiation strength of seismic source, 2) the attenuation of the seismic waves on their way from the source to the site of interest, and 3) the effect of near-surface geology. Figure 3.1 shows a representation of when an earthquake occurs, with the generated waves at the source propagating through the earth. When two nearby sites are located on two different materials, the ground motion recorded at the two sites can be very different. These differences in the ground motion could be very important considering the fact that the same building would react differently to different ground excitations, depending on the amplitude, duration and frequency content of ground motion.

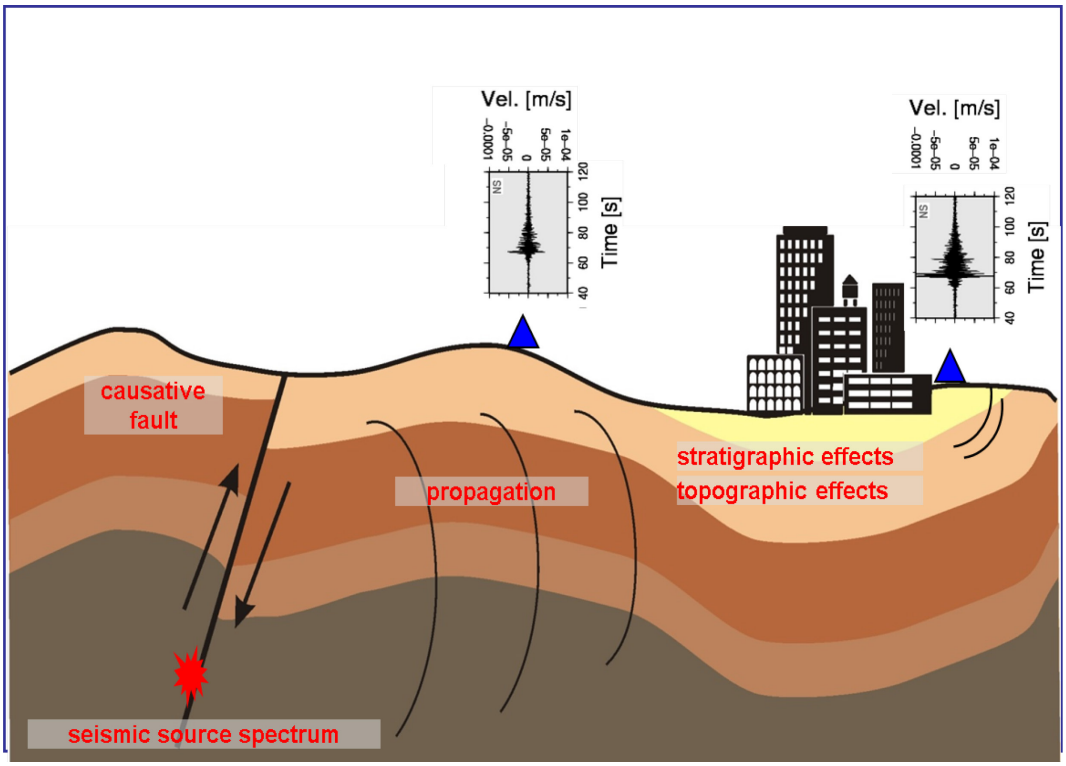


Figure 3.1 Representation of surface ground motion arising from a seismic event as a function of source, path and the site conditions (Parolai 2012).

3.1 Numerical methods

Different factors contribute to the modification of ground motion recorded at the Earth's surface, such as impedance contrast between the soft sedimentary cover and the bedrock, surface topography, heterogeneity of the earth medium, conversion of body waves to surface waves etc. (e.g., Safak, 2001). Figure 3.2 graphically shows all these effects. In general, the impedance of the soil contributes the most to site effects. Impedance is the product of seismic wave velocity and density of the soil, shown mathematically as follow:

$$Z = \rho * \upsilon \tag{3.1}$$

where ρ, and υ are the density and seismic wave velocity of the medium, respectively. In general, seismic wave velocity and density both increase with increasing depth into the crust. According to ray theory, the amplitude of motion along a ray tube is inversely proportional to the square root of seismic impedance (Aki and Richards, 2002). Hence, all seismic waves will amplify due to impedance effects as they travel to the earth's surface.

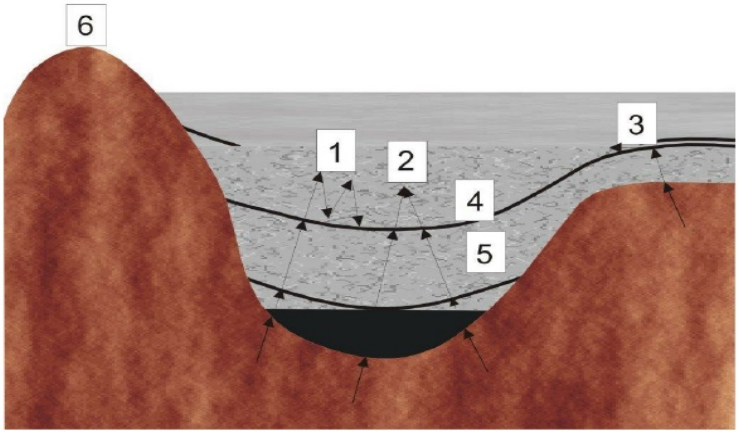


Figure 3.2 The different features that modify earthquake-induced surface ground motion: 1) Resonance due to impedance contrasts; 2) Focusing due to subsurface topography; 3) Body waves converted to surface waves; 4) Water content; 5) Heterogeneity of the earth medium; 6) Surface topography (Safak, 2001).

Many methods have been developed over the past few decades to assess the site-effects / amplification factors of soil deposits. These include both theoretical and empirical methods. Theoretical methods have been devised for 1-D, 2-D and 3-D problems considering both linear and non-linear soil response. The square root impedance (SRI) method, which is also known as the quarter wave length method, quantifies the linear site amplification based on impedance contrast (Wiggins, 1964, Joyner et al., 1981). Mathematically, SRI is represented as follows:

$$A = \sqrt{\frac{z_2}{z_1}} = \sqrt{\frac{v_2 \rho_2}{v_1 \rho_1}} \tag{3.2}$$

where z_2 and z_1 are the impedances of the bedrock and the overlaying sedimentary soft layer, respectively. To account for the free surface effect, the amplitude in equation 3.2 is multiplied by a factor of 2. Amplification due to impedance effects is frequency dependent, because longer period (low frequency) waves average the effects over a deeper (where Vs is higher) section of the crust. Frequency of impedance effects is often estimated as that corresponding to ¼ of a wavelength, e.g., in a 2 sec/km material, 1 sec (1 Hz) waves sample the top 125 m, while the 0.1 sec (10 Hz) waves are sampling the top 12 m. Since shear wave velocity is lower near the surface, the impedance amplification increases with increasing frequency.

The impedance contrast shows how strongly the waves at particular frequencies are multi-reflected within the soft surface layer, hence trapping the seismic waves' energy. Under the

assumption of vertical propagation of Shear-Horizontal (SH) waves, the modulus of 1-D site amplification H(f) is described by the Full Resonant (FR) response as:

$$|H(f)| = \frac{T}{\sqrt{1+|R|^2+2|R|\cos(4\pi f\tau)}} \tag{3.3}$$

where τ is the one way travel time in the soft sedimentary layer. T and R are the transmission and reflection coefficients for the two layers, respectively, expressed as:

$$T = \frac{2Z_2}{Z_1+Z_2} \tag{3.4}$$

$$R = \frac{Z_2-Z_1}{Z_2+Z_1} \tag{3.5}$$

The maximum of |H(f)| is obtained when $\cos(4\pi f\tau)$ is equal to -1, i.e., when $f = 1/(4\tau)$. This value of f is defined as the fundamental resonance frequency f_0 . Using the value of travel time, the famous relationship for fundamental frequency with thickness and shear wave velocity is obtained, i.e.,

$$f_0 = \frac{v_1}{4H_1} \tag{3.6}$$

where $H_1$is the thickness of the soft sedimentary layer. The SRI and the FR methods are implemented in FORTRAN language by David M. Boore in his programs site_amp and nrattle, respectively (http://www.daveboore.com/software_online.html, last accessed 16th February 2015). There are some limitations to the SRI method as pointed out by Boore (2013). The SRI method underestimates the peak response of models with large impedance contrasts, but the amplifications for those models is often close to or equal to the root mean square of the theoretical FR response of the higher modes.

A number of commercial and educational software tools are also available for calculating numerical site response of layered medium. SHAKE91 (Schnabel et al., 2012), whose educational version is Edu-Shake, and the Excel sheet-based program EERA2000 (Bardet et al., 2000) are the most widely used for linear and equivalent linear 1D analysis in the frequency domain. For time-domain non-linear analysis, the software tool DEEPSOIL (Hashash et al., 2011) and LS-DYNA (LSTC, 2009) among others are commonly used.

3.2 Empirical Methods

For generic soil, the NEHRP (National Earthquake Hazards Reduction Program) classification is used, based on Vs30. Vs30 is the average shear wave velocity over the upper 30 m. Typical engineering rock is one having an average shear wave velocity of 760 m/sec and engineering soil having an average shear wave velocity of 310 m/sec.

Table 3-1 Definition of NEHRP Site Classes (BSSC, 1994). These represent the average shear-wave velocity in the upper 30 m layer of the Earth's crust.

Site Class	Range of Shear Velocities*
A	Greater than 1500 m/sec
B	760 m/sec to 1500 m/sec
C	360 m/sec to 760 m/sec
D	180 m/sec to 360 m/sec
E	Less than 180 m/sec

In general, empirical methods are divided into reference and non-reference site methods. These methods are based on earthquake and seismic noise recordings, and are explained below in detail.

3.2.1 Reference Site Techniques

In reference site techniques, the ground motion record at one, or more, reference site is compared with other sites in the frequency domain. The reference site ground motion is either bedrock motion, rock outcropping motion, bedrock outcropping motion, or any site referred to as a rock site, where the amplification factor is ideally one. Figure 3.3 shows the different types of rock sites and a site located on sedimentary soil. Generally, S-wave or P-wave signal windows are extracted from the seismograms starting before the phase arrival and ending when a certain amount of energy is reached or after a pre-fixed amount of time. Tapering is applied at both ends of the time series data in order to bring the end points to a mean level. In general, a 5% cosine function is used for the tapering. Tapering is required to suppress

spectral leakage before calculating the Fourier Transform (FT). The time-series data is also corrected for the instrument response. The data is then processed for the base-line correction by removing the mean from the data, and finally de-trended. After processing the data, the FT is calculated. The Fourier spectrum is smoothed by using different algorithms to remove the temporal spikes from the spectrum. The analyses are performed considering a signal to noise ratio (SNR) larger than a certain threshold. The threshold is normally taken to be at least 3. For the SNR, the FT of the seismic noise is calculated for as large window as that of signal window. The analysis is carried out either for one horizontal component (N-S or E-W) or for both. In case of both components, the vectorial sum or the root mean square of the two horizontal components is used after the analysis in frequency domain (Parolai, 2012).

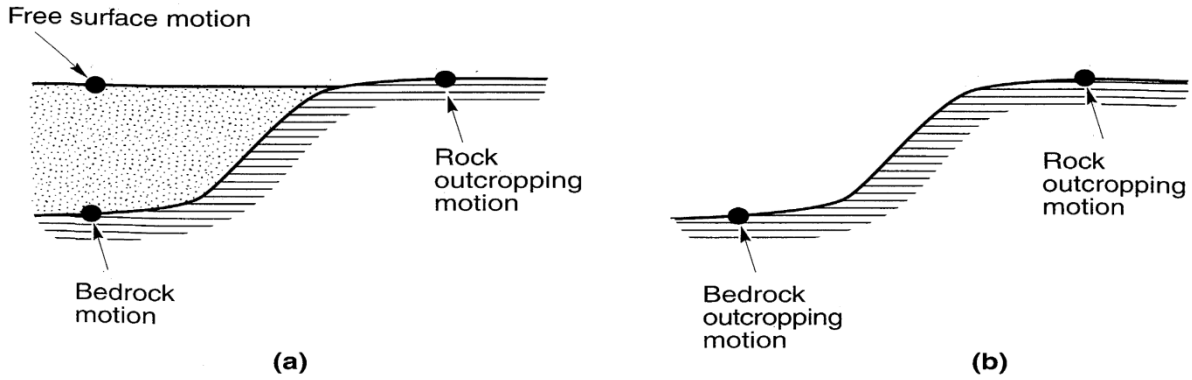


Figure 3.3: Ground response nomenclature :(a) soil overlying bedrock ;(b) no soil overlying bedrock (Steven L. Kramer, Geotechnical Earthquake Engineering).

3.2.1.1 Standard Spectral Ratio (SSR) Method

This method was first proposed by Borcherdt (1970), and now is widely used for site response analysis based on earthquake recordings. In this method, the same component of ground motion record at a site is compared with the ground motion of a reference site for the same earthquake in the frequency domain. The ratio is the site response or amplification factors of the site in the frequency domain. The reference site has to be carefully selected otherwise it will affect the site response estimated at other sites. Ideally, the reference site, which should be located on a rock site, should have a flat response. It is assumed that the ground motion of the reference site contains the same source and propagation effects (path effects) as the ground motion of the other sites. Therefore the differences observed between the sites are

explained as being due to local site effects. The source and path effects are considered to be the same for both stations if the epicentral distance of the earthquake is large enough with respect to the interstation distance. In numerical form, the method can be written as:

$$\frac{U_{ij}(f,r)}{U_{ik}(f,r)} = \frac{S_i(f)\, Z_i(f)\, A_{ij}(f,r)}{S_i(f)\, Z_k(f)\, A_{ik}(f,r)} \tag{3.7}$$

where $U_{ij}(f, r)$ is the spectral amplitude of the ground motion observed at a recording site j for an event i, $S_i(f)$ is the source function, $Z_i(f)$ is the site response, and $A_{ij}(f, r)$ is the attenuation function accounting for both linear and non-linear attenuation, with f and r representing the frequency and the hypocentral distance, respectively and the index k represents the reference site. If the reference site displays no site effects, then $Z_k(f)$ =1, and considering that $A_{ij}(f, r)$= $A_{ik}(f, r)$ and $S_i(f)$ =$S_k(f)$, thus the spectral ratio directly provides the site response at the non-reference station. The limitation of this method is that the reference station might have its own site response, hence biasing the estimated site response of the target site. Furthermore, the reference station might be located too far away from the target site, therefore the assumption of similar wave paths for the two stations may not be valid.

3.2.2 Non-reference Site Technique

One of the most challenging parts of reference site techniques is finding the appropriate reference site or rock site. As the site response is estimated with reference to a rock site, if the reference site itself has its own site response as discussed by Steidl et al. (1996), then the other target sites' estimated site responses can be significantly affected. Non-reference site techniques therefore allow the avoidance of the requirement of having a reference station, but also have some limitations. These techniques are used with both seismic noise and earthquake recordings.

3.2.2.1 Earthquake Horizontal to Vertical ratio (EHV)

This method was introduced by Langston (1979) as a means of mapping the crustal and upper-mantle structure of the Earth using tele-seismic recordings. It is based on the assumption that the vertical component of the earthquake record is not affected by local structure and that it is similar to the horizontal component of ground motion at the bedrock.

Therefore, the vertical component is de-convolved from the horizontal component of ground motion. This de-convolution is carried out in the frequency domain by dividing the Fourier Amplitude Spectrum (FAS) of the horizontal component of ground motion by its vertical component. Prior to this de-convolution procedure, the data is processed in accordance with the procedure described earlier in section 3.2.1.1. Mathematically, this method can be written as:

$$A(f) = \frac{H(f)}{V(f)} = \frac{Smoothed\ \ FAS\ of\ \ Horizontal\ record}{Smoothed\ \ FAS\ of\ \ Vertical\ record} \tag{3.8}$$

This method provides an overall frequency dependence of site response and a good estimate of the fundamental frequency of the site, although, as reported in Parolai and Richwalski (2004), the H/V from earthquake records using the S waves provides a lower level of amplification compared to SSR, especially for frequencies higher than the fundamental one. Parolai and Richwalski (2004) provided a theoretical explanation of why H/V fails in correctly estimating the level of amplification at frequencies higher than the fundamental one. The explanation is that the energy transfers to the vertical component due to S-P conversions at the bottom of the soft layer. Furthermore, as Parolai et al. (2004) note, reliable estimates of site amplification can only be derived from the H/V ratio if the sources are well distributed all around the station at different distances, as the H/V ratio depends on the incidence angle of the seismic waves (LERMO and Chavez-Garcia, 1993).

3.2.2.2 Noise Horizontal to Vertical ratio (NHV)

The Noise Horizontal to vertical ratio (NHV), also referred to as the Nakamura method, was first introduced by Nogoshi and Igarashi (1970). However, Nakamura (1989) revised it and made it popular for estimating site response. Although, according to the Nakamura explanation, the NHV is directly related to the S-waves' site response, most of the seismological community considers the predominant role of surface waves (e.g., Bard, 1999).

In this method, the ratio of the smoothed FAS of the horizontal to vertical component is calculated in the same manner as in EHV, but for seismic noise. To obtain stable and reliable results, the window length of the acquired seismic noise has to be long enough. As a rule of thumb, the total window length is selected to be long enough so as to include at least 10 cycles of the lowest frequency of interest (Parolai et al. 2001, Bard and SESAME WP02 team

2005). Hence, a 50 second window length can be considered long enough to undertake the analyses down to 0.2 Hz.

Studies dedicated to the comparison of different site response estimation techniques (e.g., Field and Jacob 1995, Bonilla et al. 1997, Parolai et al. 2004) generally show that the NHV method is capable of revealing the fundamental resonance frequency of a site (usually, although it fails to provide information about higher harmonics), but it often shows different amplitude levels than the results obtained using earthquake site response methods. Due to its simplicity and low operational budget, this method is used in many studies, especially for the resonance frequency map of cities. Parolai et al. (2010) for example used it to gain the resonance frequency map of Bishkek (Kyrgyzstan). In the present study, in combination with SSR, this method is used for improving the spatial resolution of site response in Bishkek in chapter 8.

3.2.2.3 Array analysis

Surface waves are dispersive, meaning their velocities are frequency dependent. This dispersion occurs because of the velocity variation with depth. Array analysis utilizes the seismic noise recorded simultaneously at multiple stations close to each other to estimate a dispersion curve, which describes the phase velocity versus frequency. This phase velocity is inverted using suitable algorithms (generally the genetic algorithm) to estimate shear wave velocity versus depth. The phase velocity of a wave is the rate at which the phase of the wave propagates in space. Phase of a wave is either the initial angle of a sinusoidal function at its origin or the fraction of the wave cycle that has elapsed relative to the origin. The phase velocity of surface waves from seismic noise can be extracted using different methods. The Extended Spatial Autocorrelation method (ESAC) is used here. Aki (1957, 1965) was the first to show that phase velocities in sedimentary layers can be determined using the statistical analysis of seismic noise. According to Aki, seismic noise represents the sum of plane waves propagating without attenuation in a horizontal plane in different directions with different powers, but with the same phase velocity for a given frequency. He also assumed that waves with different frequencies and different directions are statistically independent. The spatial correlation function is described as:

$$\phi(r,\lambda) = < \mathrm{u(x,y,t)(x + r\cos(\lambda), y + r\sin(\lambda), t)} > \quad (3.9)$$

where $u(x, y, t)$ is the observed velocity at point (x,y) at time t, r is the inter-station distance, λ is the azimuth and <> indicates the total average. An azimuthal average of this correlation function is given as:

$$\phi(r) = \frac{1}{\pi}\int_0^{\pi} \phi(r,\lambda)d\lambda \tag{3.10}$$

For the vertical component, the power spectrum $\phi(\omega)$(which is the Fourier Transform of the correlation function) can be related to $\phi(r)$ via the zero[th] order Hankel transform as:

$$\phi(r) = \frac{1}{\pi}\int_0^{\infty} \phi(\omega)J_0\left(\frac{\omega}{c(\omega)}r\right)d\omega \tag{3.11}$$

where ω is the angular frequency, $c(\omega)$ is the frequency-dependent phase velocity, and J_0 is the zero order Bessel function. The space correlation function for one angular frequency ω_0, normalized to the power spectrum, will be of the form:

$$\phi(r,\omega_0) = J_0\left(\frac{\omega_0}{c(\omega_0)}r\right) \tag{3.12}$$

Fitting the azimuthally average spatial correlation function obtained from the measured data to the Bessel function, the phase velocity $c(\omega_0)$ can be calculated. A fixed value of interstation distance r is used in the spatial autocorrelation method (SPAC). However, Ohori et al. (2002) and Okada (2003) showed that, since $c(\omega)$ is a function of frequency, better results are achieved by fitting the spatial-correlation function at each frequency to a Bessel function, which depends on the inter-station distances (extended spatial autocorrelation, ESAC). For every pair of stations, the function $\phi(\omega)$ can be calculated in the frequency domain by means of (Malagnini et al., 1993; Ohori et al., 2002; Okada, 2003):

$$\phi(\omega) = \frac{\frac{1}{M}\sum_{m=1}^{M} Real(mS_{jn}(\omega))}{\sqrt{\frac{1}{M^2}\sum_{m=1}^{M} mS_{jj}(\omega)\sum_{m=1}^{M} mS_{nn}(\omega)}} \tag{3.13}$$

where mS_{jn} is the cross-spectrum for the mth segment of data, between the *j*th and *n*th station, M is the total number of used segments, and the power spectra of the mth segment at station j and station n are $mS_{jj}(\omega)$ and $mS_{nn}(\omega)$, respectively.

The space correlation values for every frequency are plotted as a function of distance. An iterative grid search procedure is performed using equation (3.12) in order to find the value of phase velocity $c(\omega_0)$ that gives the best fit to the data. The iteration is performed over a large range of phase velocities (e.g., between 100 and 3000 m/s) in small steps (e.g., 1 m/s). The

best fit is achieved by minimizing the root mean square (RMS) of the differences between the values calculated with equation (3.12) and (3.13). Data points, whose correlation coefficients are higher than two standard deviations from the value obtained with the minimum misfit velocity, are removed before the next iteration of the grid search procedure. Figure 3.4 shows the spatial correlation coefficient for different frequencies obtained from the observed dataset, and the best fitting Bessel function obtained using the ESAC methodology.

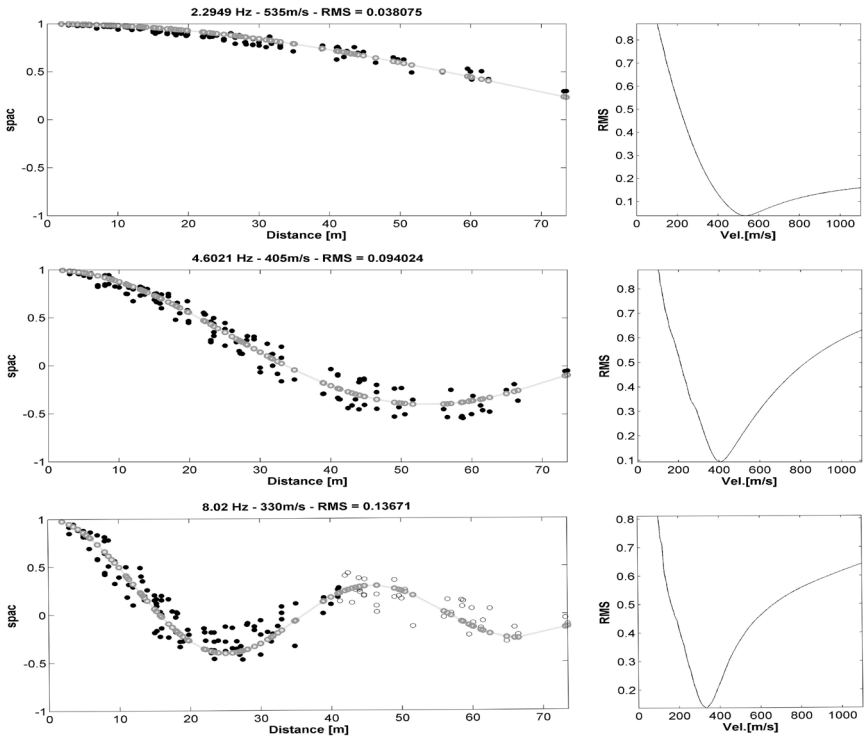


Figure 3.4 Left: Measured space-correlation function values for different frequencies (black circles) and the best-fitting Bessel function (grey circles). Right: The respective RMS error versus phase velocity curves.

Finally, the dispersion curve is obtained by plotting the phase velocity (the velocity corresponding to the minimum RMS in Figure 3.4) versus the frequency. At frequencies higher than a certain threshold, the phase velocity might increase linearly. This effect is due to spatial aliasing which limits the upper bound of the usable frequency band. It depends on the S-wave velocity structure and the minimum inter-station distance (Picozzi, 2005). For this reason, the minimum phase velocity is given by the relationship (Parolai, 2012):

$$Min\ Phase\ Velocity_{aliasing} = 2 * f * r_{minimum} \tag{3.14}$$

where $r_{minimum}$is the minimum inter-station distance in the array set up, and f is the frequency. The factor 2 might also increase from 2 to 4 depending on the array set up. At low frequencies, the RMS error function clearly indicates the lower boundary for acceptable phase velocities, but might not be able to constrain the higher ones. The frequency above which phase differences cannot be resolved any more depends on the maximum inter-station distance and the S-wave velocity structure. The maximum phase velocity is hence given by the relationship (Parolai, 2012):

$$Max\ Phase\ Velocity_{aliasing} = 3 * f * r_{maximum} \tag{3.15}$$

Figure 3.5 shows the dispersion curve estimated from the ESAC method for a site in Karakol, Kyrgyzstan. This dispersion curve shows the normal behaviour of decreasing phase velocity with increasing frequency.

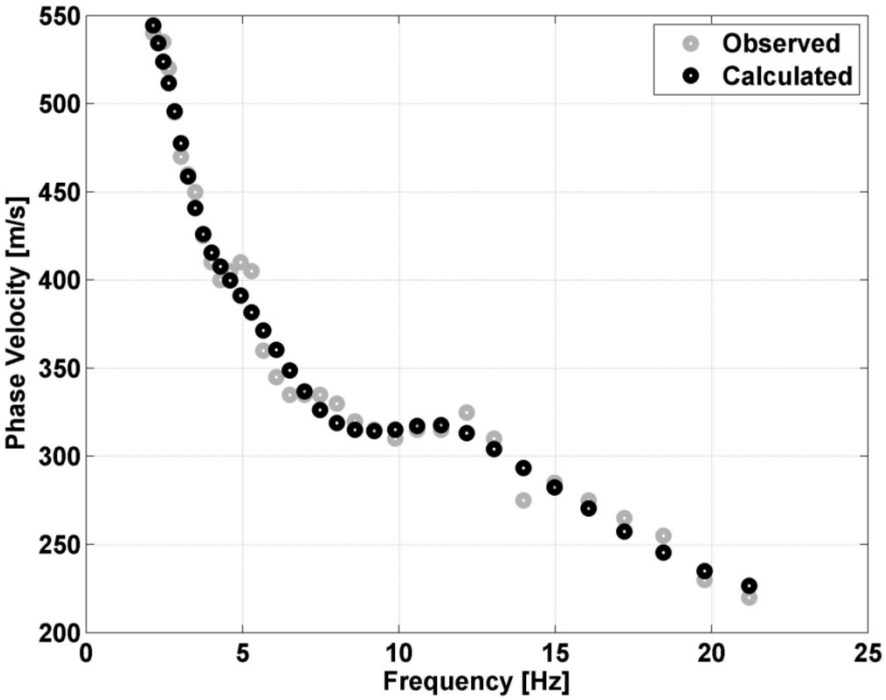


Figure 3.5 Phase velocity dispersion curves by ESAC analysis for a site in Karakol, Kyrgyzstan. The grey circles indicate the observations from the analysis, while the black circles indicate the reconstructed or calculated dispersion from the results of inversion analysis.

The dispersion curve of Rayleigh waves mainly depend non-linearly on the S-wave velocity structure, but also on the density and thickness of each layer, and the P-wave velocity structure (Yamanaka and Ishida, 1996). The dispersion curve is inverted for shear wave velocity versus depth using different algorithms. These include linear inversion (e.g.,

Tokimatsu et al., 1991) and non-linear inversion (genetic algorithm e.g., Goldberg 1989, Yamanaka and Ishida 1996) methods. In linearized methods, an initial model is assumed and an objective function is defined. Hence, the final model calculated by a linearized inversion inherently depends on an assumed initial model because of the existence of local optimal solutions. When an appropriate initial model may be generated using a priori information about the subsurface structure, linearized inversions can find an optimal solution that is the global minimum of a misfit function (Parolai et al., 2006).

In this study, the non-linear method of modified genetic algorithm (Yamanaka and Ishida, 1996) is used for the inversion of the dispersion curves. The genetic algorithm is a non-linear optimization method, which simultaneously searches locally and globally for optimal solutions by using several models. In this method, a search area is defined for the shear wave velocity, P-wave velocity and thickness of the layers. Then the algorithm uses a probabilistic process to produce an initial population of models from the defined search area (parameter space). Those models that then satisfy some particular criteria (fitness function) are selected and used as parents for the generation of children (a new population of models). This generation stage happens by means of crossover and mutation operations, which are genetic processes. The crossover process allows the merging of chromosome structure relevant to the parents' chromosome structure, by exchange and combination of each parent's structure (Gallagher and Sambridge, 1994). Hence, the children have a combination of the parents' characteristics and might also develop different features. Mutation is a random process that allows the introduction of features to children other than those from the parents.

The fitness function (f_j) is defined for each model j. The fitness function which is the inverse of the misfit function (ϕ_j) is based on the average of the differences between the observed and the theoretical phase velocities of dispersion curve, defined as:

$$\phi_j = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(d_0)_i-(d_1)_i}{\sigma(d_0)_i}\right]^2 \tag{3.16}$$

where N is the number of data, and $\sigma(d_0)_i$ is the standard deviation associated with the observed data $(d_0)_i$. Then, the fitness function is provided as:

$$f_j = \frac{1}{\phi_j} \tag{3.17}$$

Hence, the models having higher fitness (or lower misfit) values have a higher probability of proceeding to the next model generation. However, due to this considering probability, a model with a higher fitness value might not be selected for the next generation. Hence, along with crossover and mutation, two more genetic operations are used to increase the convergence. They are elite selection and dynamic mutation. Elite selection ensures that the best model appears in the next generation, replacing the worst model in the current one. Then to avoid a premature convergence of the solution into a local minimum, the dynamic mutation operation is used to increase the variety in the population.

Figure 3.6 shows the results of inversion analysis for the dispersion curve in Figure 3.5. The grey lines indicate all the tested models for different combinations of input parameters within the available parameters space. The black lines indicate all the best models within the 10% range of the misfit, and the red line indicates the best fit model. In this case, the shear wave velocity in the shallow layer is about 200 m/s, whereas in the deep layer it is about 500 m/s.

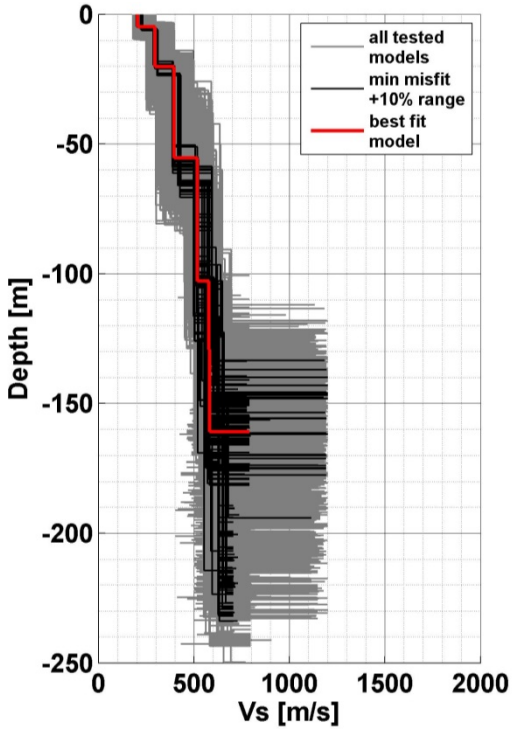


Figure 3.6 S-wave velocity model obtained by inverting the dispersion curve shown in Figure 3.5. Grey lines represent all tested models, black represent the minimum misfit model within a 10% range, and red represents the best fit model.

4 Introduction to clustering analysis

Clustering analysis is used in this study to identify groups of similar site response spectral ratios, using earthquake and seismic noise data (chapter 8). The results of clustering analysis with correlation analysis are further used for improving the spatial resolution of ground motion variability at the urban scale in Bishkek, Kyrgyzstan. Cluster analysis is an important process in data mining. It is used to organize objects into different groups based on the similarity of various selected parameters. Clustering is about partitioning a given data set into groups. Clustering analysis is done in order to obtain a simple understanding of a given data set by grouping together those elements into different groups (clusters) with similar attributes. The data points in each cluster (group) are more similar to each other than points in different clusters. Clustering analysis is used in different fields and is found under different names, for example, unsupervised learning in the field of pattern recognition, numerical taxonomy in the field of biological sciences, typology in social sciences and partition in graph theory (Theodoridis and Koutroubas, 1999, Halkidi et al., 2001). At present, various algorithms for clustering analysis are used in different fields for specific kinds of data and application.

Figure 4.1 shows the basic steps involved in the development of a clustering process. These steps are summarized below, after Fayyad et al. (1996).

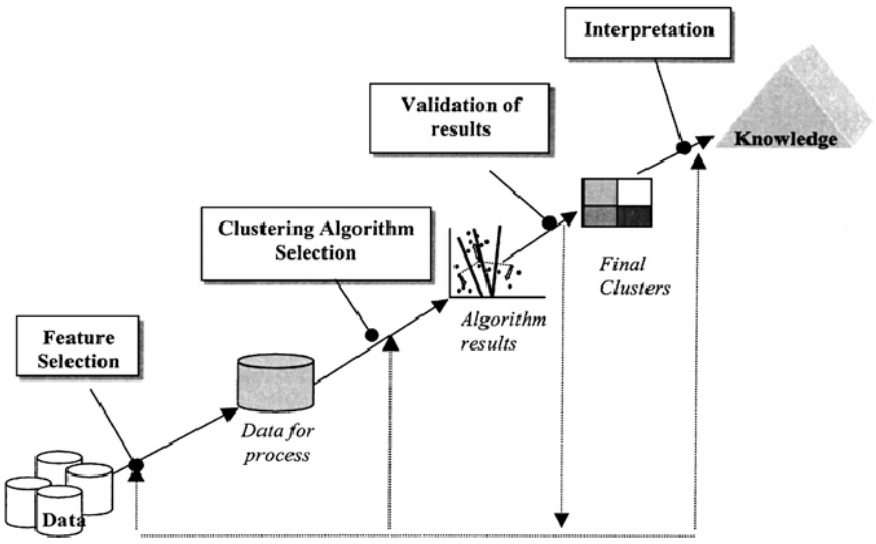


Figure 4.1 Steps involved in the clustering process (Halkidi et al., 2001).

1) First, the features of the data set are identified in which the clustering is performed. These features are carefully selected in order to encode as much information as possible regarding the purpose and interest of the task. Selection of the features is very crucial and plays an important role in the final outcome of the analysis. Pre-processing of the data is sometimes necessary prior to the clustering analysis. The spectral ratios of site effects from SSR and EHV (chapter 3) are used as features for clustering in this study (chapter 8).
2) After the feature selection and pre-processing of the data set, an appropriate clustering algorithm is selected. Most of the time, the selection of clustering algorithm depends on the type of data set at hand. A proximity measure and a clustering criterion mainly characterize a clustering algorithm as well as its efficiency to define a clustering scheme that fits the data set.
    i. Proximity measures quantify the similarity of two data points i.e., feature vectors. It is ensured that all selected features contribute equally to the computation of proximity measure.
    ii. Clustering criterion can be expressed via a cost function, which takes into account the type of clusters that are expected to occur in the data set. Thus, the clustering criterion may be defined as a "good", which leads to a partitioning that fits well the data set.

The K-mean clustering algorithm is used to classify the stations based on spectral ratios in this study.

3) The results of clustering analysis are verified using appropriate criteria and techniques. Since clustering algorithm defines clusters (groups) that are not known a priori, the final partition of data requires some kind of evaluation in most applications. The CH and Silhouette indices are used to validate the clustering results as described in section 4.2.
4) In many cases, experts' interpretation of the final result is required to confirm a conclusion of the clustering results. This is done by integrating results from other experimental evidence and by using experts' opinion.

Clustering analysis is used in a number of applications in many fields. It is used for data reduction, hypothesis generation, hypothesis testing, prediction based on groups, web mining,

spatial data analysis etc. (Theodoridis and Koutroubas, 1999). In this study, clustering analysis is used for establishing the hypothesis, that sites grouping together based on the similarities in terms of SSR also group together based on the similarities in terms of noise horizontal to vertical spectral ratios (HVSR), irrespective of the relative difference in amplification between SSR and HVSR.

4.1 K-mean Clustering Algorithm

The K-means algorithm was introduced by J.B. MacQueen in 1967. It is one of the most widely used unsupervised learning algorithms that resolve the clustering problems. It allows that one of the data belong to only one cluster, therefore this algorithm is a definite clustering algorithm. The K-means algorithm is used in this study to cluster the site response spectral ratios in terms of Standard Spectral Ratios and Horizontal to Vertical Spectral Ratios at each frequency, independently from each other. It was chosen for the clustering analysis in this study because of its relevance to the problem in hand.

The following are the main steps involved in the K-means clustering algorithm: (1) the algorithm requires that the number of clusters (k) be defined a priori. k centroids are then defined randomly to represent each cluster at each frequency. (2) The dissimilarity between an object (in this case the value of the spectral ratios at each frequency) and the centroid of each cluster is calculated. The squared Euclidean distance parameter is used here to determine this difference. (3) Based on the shortest distance, an object is assigned to the cluster whose centroid is nearest to the object. (4) When all objects are assigned in this way, a new series of k centroids for each frequency are re-calculated based on the already formed clusters. The objects are again assigned to the new centroids based on the shortest distance. The process continues iteratively until there is no further update of the centroids. The aim of this algorithm is to minimize the sum of the squared error over the whole dataset given as:

$$E = \sum_{i=1}^{k} \sum_{l=1}^{n} \sum_{j=1}^{m} \left\| x_{lj,i} - c_i \right\|^2 \tag{4.1}$$

where m is the number of objects (in this study, the number of frequencies examined in each spectra) in a data set, n is the number of observations (or number of spectra) in each cluster, k the total number of clusters, c_i is centroid of the cluster i, and $x_{lj,i}$ is the value of the j object (spectral ratio) in cluster number i of observation number l. The random selection of the

centroid of each cluster affects the results of the clustering analysis such that it might not converge to the global minimum of the chosen function. Therefore, the analysis is repeated 10 times, using different randomly chosen starting centroids and the resulting solution that provides the minimum of the sum of the squared error is chosen as the best one. The process continues until the centres of the clusters stop changing.

4.2 Validation of Clustering Algorithm

Clustering validation and evaluation of clustering results is one of the important parts of clustering analysis. Its purpose is to find the partitioning that best fits the underlying data. Cluster validation includes both quantitative and objective analysis. In most algorithms, 2D-data sets are used for experimental evaluations in order to visually verify the validity of the results. Hence, visualization of the data set is very important for the verification of clustering results. The validation algorithms play an important role in case of large multidimensional data sets, in which case the perception of clusters using visualization tools is difficult for the human eye. Clustering validation can be categorized into two classes, external clustering validation, and internal clustering validation. The main difference is whether or not external information is used for clustering validation. This external information is not present in the data. An example of external clustering validation measure is entropy, which evaluates the purity of clusters based on the given class labels (Wu et al., 2009). Internal clustering validation relies on information only present in the data set, without any regard to external information. External validation measures are mainly used for selecting an optimal clustering algorithm on a specific data set, because they know the number of clusters in advance. On the other hand, internal clustering validation measures are used both for choosing best clustering algorithms as well as the optimal cluster number (Liu et al., 2010). Since in practice, external information such as class labels is often not available in many application scenarios, internal validation measures are the only option for cluster validation. As a guide for the selection of the optimum number of clusters, the Calinski and Harabasz index (CH Index) (Calinski and Harabasz 1974) and Silhouette Index (Rousseeuw, 1987) are evaluated in this study.

4.2.1 *Calinski and Harabasz (CH) Index*

The CH Index is a variance ratio criterion that provides insight into the structure of points. In this method, the dispersion of a group of n-points is measured by the sum of the squared (SS)

distances of the points from the centroid of their clusters. It is a measure of the between-groups SS over the within-groups SS, mathematically given as:

$$CH\ \ Index = \frac{BGSS(k)}{(k-1)} \Big/ \frac{WGSS(k)}{(n-k)} \tag{4.2}$$

where BGSS is the between-groups sum of the squared distance, WGSS is the within-group sum of the squared distance, k (2, 3, 4…) is the number of clusters and "n" is the number of points (in this study, the number of spectral ratios of the stations). For the best clustering, the value of CH Index is maximized.

4.2.2 *Silhouette Index*

The Silhouette index validates the clustering performance based on the pairwise difference of between and within cluster distances. The mathematical form of the Silhouette index is given as:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{4.3}$$

where $a(i)$ is the average dissimilarity (distance) of object i to all other objects in the same cluster, and $b(i)$ is the minimum of the average dissimilarity of object i to all other objects in the closest neighbouring cluster. The value of $S(i)$ ranges from -1 to +1 with +1 representing a well-clustered sample, whereas -1 shows the sample is misfit. This is a graphical method for selecting the optimum number of clusters which shows the width of silhouette for each sample, the average silhouette width for each cluster and the overall average silhouette width for the complete data set. The overall average silhouette width is the average of the $S(i)$ for all objects in the whole dataset.

5 OVERVIEW OF EXISTING HAZARD STUDIES AND REGIONAL SEISMOTECTONICS OF THE STUDY AREA

Central Asia is one of the world's most active regions with the largest seismic hazard. However, at the same time, a significant portion of it is also considered as a stable continental region, which includes most of Kazakhstan. This chapter explains in some detail the tectonic development of this region, and previous seismic hazard studies at the regional level in Central Asia.

5.1 Regional seismotectonics of Central Asia

The deformation in Central Asia is the result of the India-Eurasia collision, with the Indian Plate migrating northwards. This collision between the two continents is considered to be one of the major tectonic events of Cenozoic time. According to Palaeomagnetic studies in Tibet, this collision has resulted in about 1900± 850 km of crustal shortening within continental Eurasia (Achache et al., 1984). Moreover, the comparison of Palaeomagnetic data and reconstruction based on magnetic anomalies show an additional 700± 300 km of convergence between India and southern Tibet. Hence, this collision between India and Eurasia has resulted in a total north-south shortening of about 2600± 900 km. Different models are proposed for the mechanism of this shortening, including continuous deformation, multiple continental under-thrusting, eastward propagating extrusion and mosaic tectonics (Patriat and Achache, 1984). Convergent plate movement between India and Eurasia is evident in particular by active tectonics beneath the Himalayas, in Tibet and in south-central Asia. According to Patriat and Achache (1984), up to 52 million years (Myr) ago the Indian plate started drifting northward with a mean velocity of 15 to 20 cm/yr. Then, between 52 and 36 Myr ago, the motion of Indian plate became rather irregular, showing changes in direction, with the mean northward velocity reduced to less than 10 cm/yr. Finally, from 36 Myr ago to the present, the convergence movement of Indian plate remained stable northward with respect to Eurasia with a constant rate of less than 5 cm/yr.

A significant amount of this post-collisional lithospheric convergence is accommodated by the convergence of Pamir and Tien Shan, which are two of the major Cenozoic mountain belts in Central Asia. The central and eastern Tien Shan forms an elongated deforming region

between two generally stable crustal elements: the Kazakh platform to the north and the Tarim basin to the south. Over a distance of ~1000–1500 km north of the Indo-Eurasian plate boundary, the central Tien Shan presently absorbs nearly one-half of the total relative plate convergence of ~45 mm/year (Abdrakhmatov et al., 1996; Holt et al., 2000). This makes the Pamir thrust system one of the regions with the highest strain rates in the India-Eurasia collision zone. The northward advance of Pamir is adjusted laterally by a series of strike-slip systems. The left lateral Darvaz-Karakul fault separates the Pamir from the Tajik-Afghan basin, where the shear is currently amounting to about 10mm/yr as approximated by GPS measurement (Ischuk et al., 2013). Focal mechanisms from moderate and large earthquakes primarily show thrust and reverse faulting with P axes oriented approximately north-south, consistent with the geodetically measured maximum shortening direction and the overall direction of Indo-Eurasian plate convergence (Tapponnier and Molnar, 1979; Nelson et al., 1987; Ghose et al., 1998). Figures 5.1 and 5.2 show the topographic maps of the study region, along with the major cities and locations discussed in this work (Figure 5.1), and the active faults and tectonic regions (Figure 5.2).

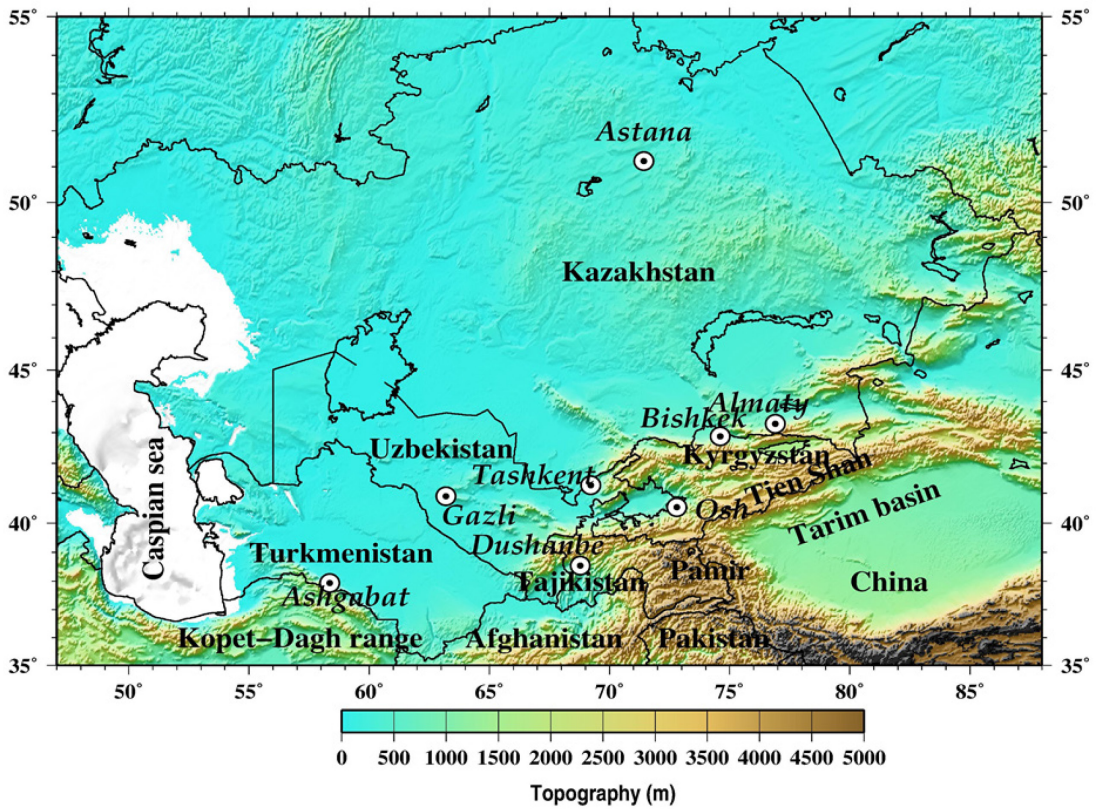


Figure 5.1 Topographic map of the study area, with the major cities and locations discussed in the text.

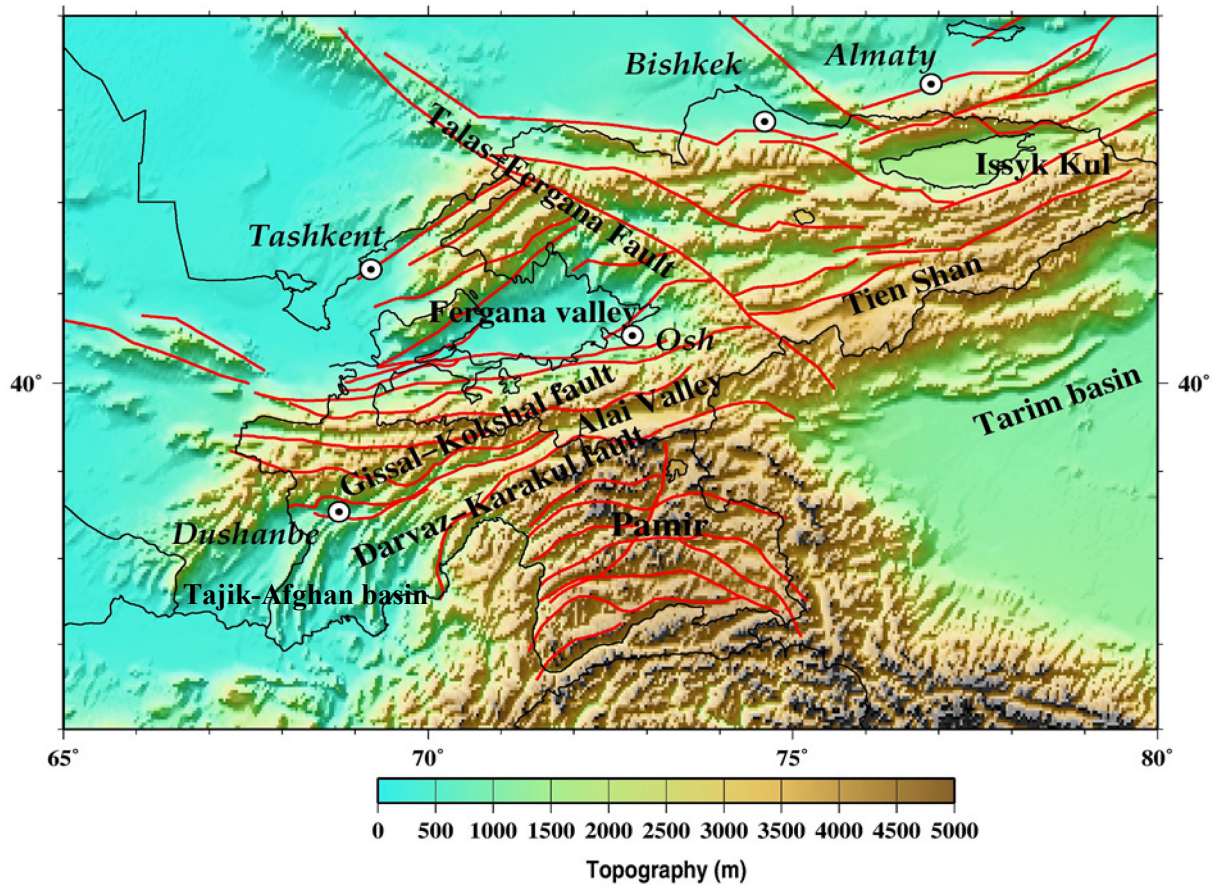


Figure 5.2 Topographic map of the study area, with the major faults outlined as red lines (Trifonov, 1996) and the major tectonic regions.

The deformation in Central Asia is mainly accommodated by reverse faults and intra-mountain basins located far from the boundaries of the major plates (Zubovich et al., 2010), with large fault systems located along the main deformation zones. Amongst these, it is worth mentioning the Pamir thrust located along the Pamir-Tien Shan convergence zone (Schurr et al., 2014); the right-lateral Talas-Fergana strike-slip fault separating the northeastern and southwestern sectors of Tien Shan, and several fault systems mapped in the northern Tien Shan, such as the Issyk-Ata and Chon-Kemin faults, which partially accommodate the convergence of the Tarim basin towards the Kazakh platform. Such convergence is absorbed over a region more than 200 km wide and, although not uniformly distributed, no single predominant fault absorbs the majority of this convergence (Zubovich et al., 2010). The Pamir has been displaced northwards by about 600 km with respect to the Tarim basin (Burtman and Molnar, 1993) and is characterized by east–west trending thrust faults and major strike-slip faults flanking the range (Lukk et al., 1995).

The Turan and south Kazak domains (TSK) cover a large area of Central Asia, from the Caspian Sea in the west, to the Tien Shan range in the east, and from the Kopet Dag range in the south, to the Aral Sea in the north. Since the Mesozoic, several continental blocks have accreted to this southern margin of Eurasia, during the successive closures of Paleotethys and Neotethys. Although most of the TSK is nowadays covered by Neogene to Quaternary continental deposits, several discrete Paleozoic basins occur within the south Kazak domain and basins have developed continuously since the Permian over the TSK. These basins provide an almost continuous record of tectonic events on the southwestern margin of Asia since the Mesozoic. However, the timing and boundary conditions controlling basin development are still poorly known. Many paleogeographic reconstructions have inferred back-arc extension in southwestern Asia during the Late Permian to Mesozoic times (Tapponnier et al., 1981; Zonenshain et al., 1990; Dercourt et al., 1993), but few data constrain this setting. Some discrete compressional events have also occurred during the Mesozoic in southwestern Asia, but their intensity and role in basin development are not well documented (e.g., Otto 1997).

5.2 Regional seismic hazard studies in Central Asia

The seismicity in Central Asia has been studied since the eighteenth century by Russian and Central Asian scientists (e.g., Kondorskaya and Shebalin, 1982; Rautian and Leith, 2002; Tatevossian 2004). The first summary of earthquakes in Central Asia was reported in the catalogue of Mushketov and Orlov (1893), based on macroseismic observations. The occurrence of large earthquakes in the area encouraged the development of projects for seismic monitoring. Examples are the installation of a seismic network in Central Asia promoted by the Russian Imperial Geographical Society following the 1887 Verny (Kazakhstan) earthquake, which caused a great deal of destruction in the city of Almaty, formerly known as Verny (Mikhailova and Kurskeev 1995), and the seismological instrument area coverage improvement after the 1948 Ashgabad (Turkmenistan) earthquake. Changes in the sensitivity and dynamic range of instruments performed in 1956 and the 1970's led to improved accuracies in location and magnitude estimation (Rautian and Leith, 2002), providing more complete data sets for seismological studies and seismic hazard assessment.

Several early studies aiming at seismic hazard assessment were carried out in Central Asia. Early seismic zoning maps of the former USSR and adjacent territories were compiled by

Mushketov (1933), in which the expected surface shaking is expressed in terms of isoseismic lines for different intensities. In 1937 Gorshkov, who was involved in Mushketov's research, compiled his own version of a seismic zoning map for Central Asia. He specified five zones of "geological similarity" (respectively for intensities V to IX) extending from areas of former seismic events. In 1957 a new seismic zoning map of the territory of the USSR, including Central Asia, was compiled under the supervision of S. V. Medvedev. The new principle of seismic zoning proposed by Gubin (1960) played a critical role in the development of the methodology for the compilation of seismic hazard maps. The next seismic zoning map of the USSR territory was compiled in 1968 by a large team of researchers under the leadership of Medvedev. Successive seismic zoning maps of the USSR were released in 1978. These maps were also incorporated into national engineering codes. In particular, the general seismic zoning (GSZ) map of 1978 (Bune and Gorshkov, 1980) included both boundaries of shaking intensities with zones for intensities from VI to IX on the MSK64 scale (Medvedev et al., 1964) as well as zones of most probable locations of severe earthquakes, differentiated by the maximum expected magnitude, ranging from MLH (Karnik, 1968) 6.1 to 8.1. The probability of occurrence of an event was also included, considering return periods of 100, 1000 and 10,000 years after the analysis of the recurrence times based on the method of Ryznichenko (1966) and historical catalogues. Figure 5.3 shows the simplified version of the 1978 GSZ for Central Asia. For most of the large cites, it indicates a maximum expected level of IX shaking. For Tashkent, the maximum expected shaking is MSK VIII. The exposure period over which this level of shaking is expected varies from location to location, as described in the original paper.

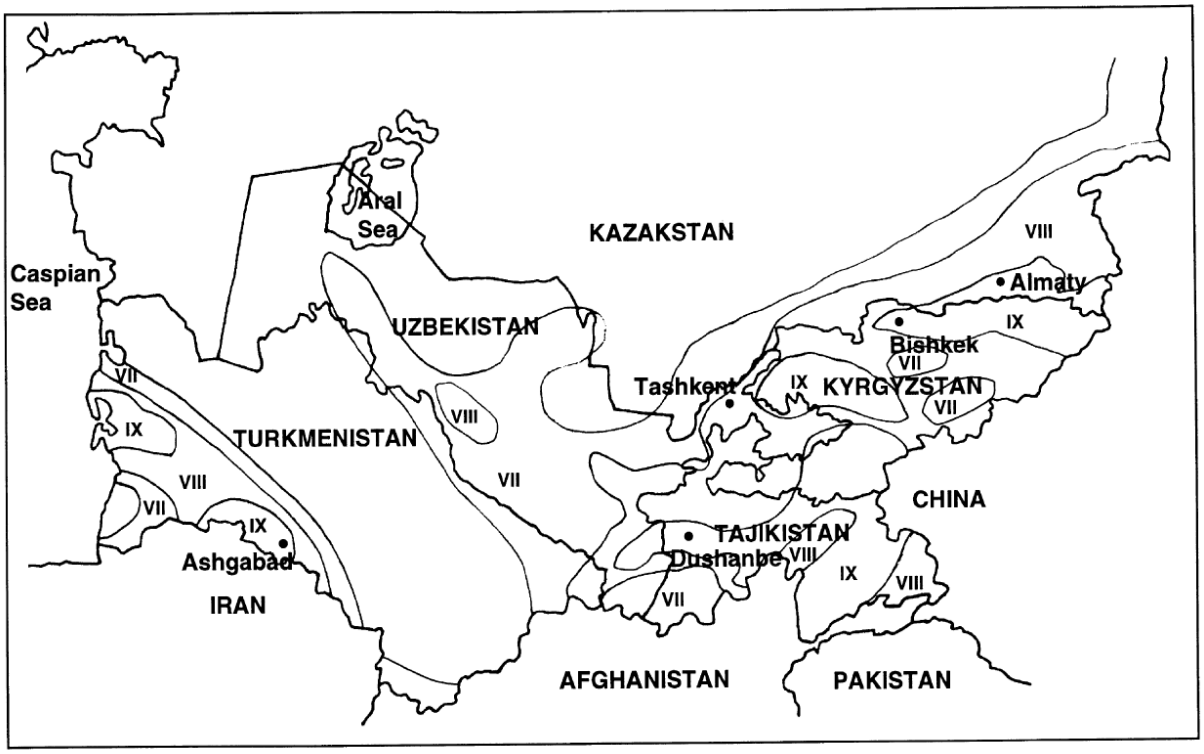


Figure 5.3 Simplified version of the 1978 official Soviet seismic hazard map showing the maximum expected seismic intensity (MSK) in Central Asia (King et al. 1996).

In the former Soviet Union, a great deal of effort was carried out in updating the seismic zoning at different spatial scales, such as detailed seismic zoning (DSZ), valid for the regional level, and seismic micro-zoning (SMZ) maps, valid at the urban scale. Such maps were characterized not only by their different spatial scales, but also by the way local seismotectonic, ground and other natural conditions were taken into account. DSZ was intended for the study of earthquake generating structures that posed seismic hazard to a certain location. For the SMZ constructed for most of the main cities in Central Asia, the influence of local site amplifications was also accounted for by introducing intensity increments related to local geological conditions. By the end of 1990, SMZ maps showing macro-seismic intensity, also considering the effect of soil and topography in a few areas, had been compiled for ten large cities in the USSR (McGuire, 2004) including all five capitals of the Central Asian countries.

Between 1991 and 1997, after the collapse of the Soviet Union, a new general seismic zoning scheme (named GSZ-97) for Northern Eurasia was developed (Ulomov, 1999) and was included as a contribution to GSHAP (Giardini, 1999). The hazard map was originally produced for macroseismic intensity and later converted to peak ground acceleration using the

Aptikaev and Shebalin (1988) relationship. GSZ-97 was the last seismic zoning that considered the territory of the former Soviet Union as a whole. Figure 5.4 shows the GSZ-97 map (m/s^2) for Northern Eurasia for 10% probability of exceedance in 50 years. The expected hazard exceeds 6 m/s^2 in most of Tajikistan and Kyrgyzstan. Similarly, for the cities of Almaty and Ashgabad, located in Kazakhstan and Turkmenistan, respectively, the expected hazard is also more than 6 m/s^2. Starting from the 1990s, seismic hazard, in terms of macroseismic intensities, was re-assessed for many capitals of the now independent countries in Central Asia (for a review, see Nurmagambetov et al., 1999). As an example, DSZ maps in terms of both macroseismic intensity and peak ground acceleration were produced for the suburbs of Almaty (formerly known as Alma-Ata)(Kyurskeyev 1993), and the probabilistic values of seismic hazard in Almaty were assessed in terms of peak accelerations and intensity (Mikhailova 1996). For Tashkent (Uzbekistan) and Bishkek (Kyrgyzstan), new probabilistic assessments have been computed by Erdik et al. (2005). On a larger scale, hazard assessments have been performed for Kyrgyzstan (Abdrakhmatov et al, 2003) and Uzbekistan, considering the Cornell (1968) approach.

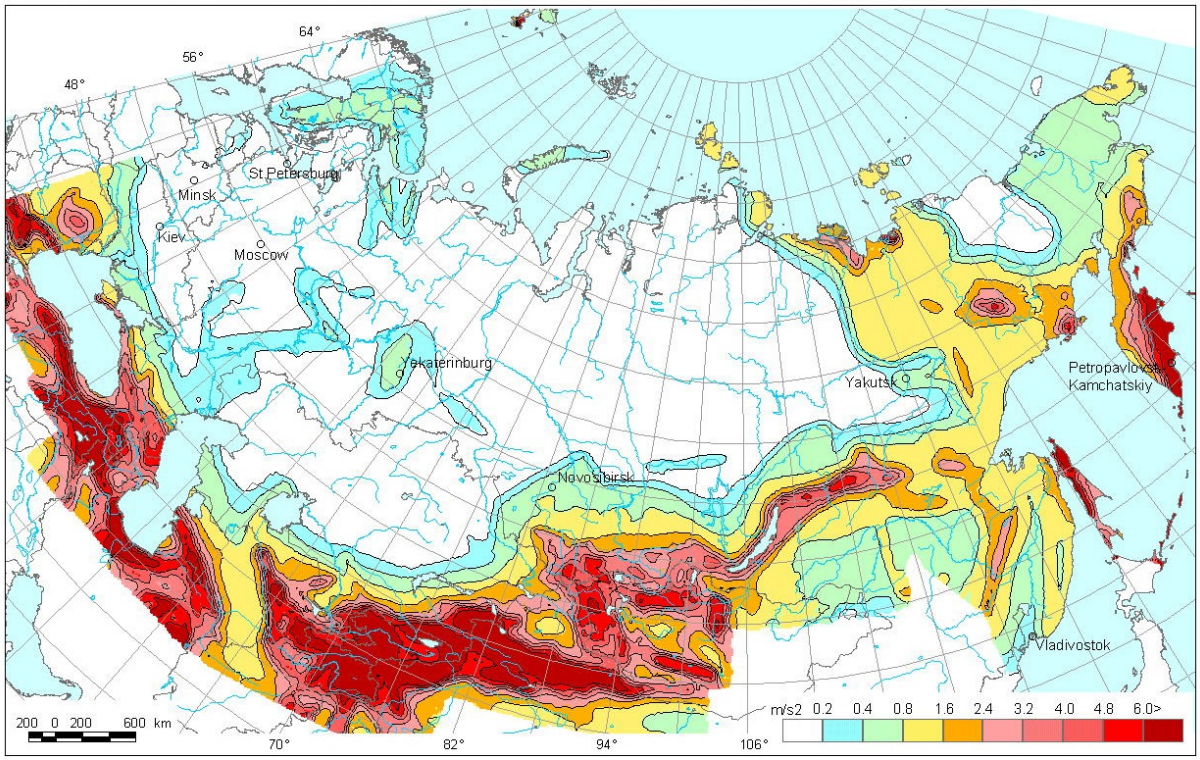


Figure 5.4 Peak ground acceleration (m/s2) map of Northern Eurasia for 10% probability of exceedance in 50 years (Ulomov, 1999).

Within the framework of the EMCA project, new hazard studies have been performed for Central Asia in term of macroseismic intensity. Starting from the distribution of intensity data for earthquakes that occurred in this region since the end of the 19th century, the seismic hazard has been assessed following the so-called site-approach (Bindi et al., 2012; D'Amico and Albarello, 2008). Based on this method, the seismic hazard assessment is based on the seismic histories available at different locations, without requiring any a-priori assumption about source zonation. The hazard is computed site by site and is expressed as the probability that the considered site will be shaken with an intensity greater than or equal to the chosen threshold over the assumed exposure time. Figure 5.5 shows the calculated hazard for Central Asia using the site-approach for 10% probability of exceedance in 50 years. It can be seen that Kyrgyzstan and Tajikistan show intensities greater than 7. In some active areas, the calculated hazard exceeds an intensity of 9.

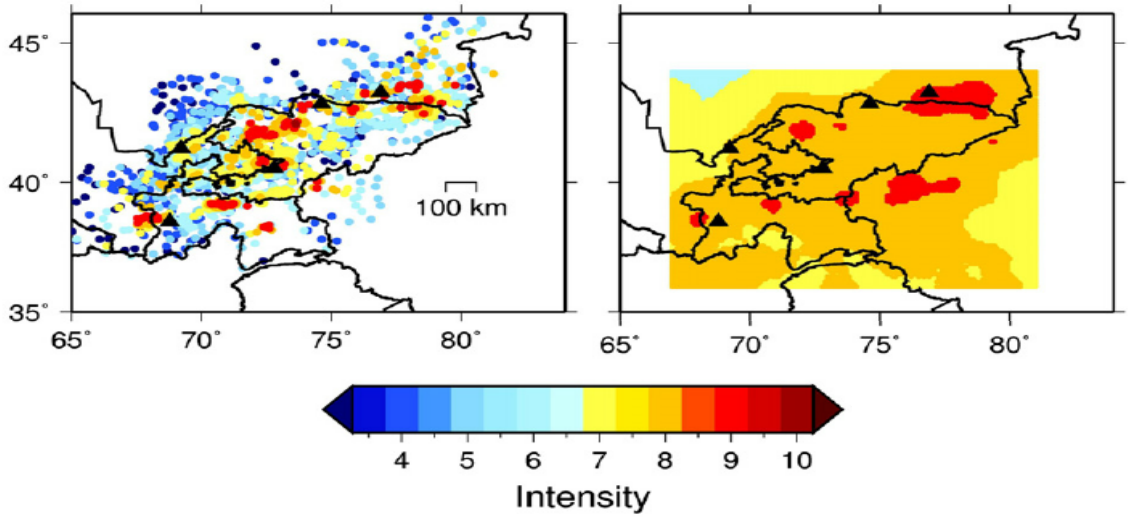


Figure 5.5 Result of PSHA for Central Asia using the site approach (Bindi et al., 2012). Left side: Seismic hazard in terms of MSK-64 considering only felt histories. Right side: Seismic hazard estimated over a grid of 0.1 degrees, considering both felt histories and virtual intensities (i.e., computed by applying an IPE to a seismic catalogue). The triangles in right hand figure show the locations of important cities.

6 SEISMIC HAZARD MODEL FOR CENTRAL ASIA

This chapter describes the seismic source models used for Central Asia in this study. Different approaches of smooth seismicity as well as the area source model are used as input for the seismic hazard calculations. For the seismic hazard calculations, first the earthquake catalogue undergoes declustering analysis, completeness analysis, and magnitude-frequency analysis. This is followed by the description of different source models.

6.1 Processing the earthquake catalogue

An earthquake catalogue is one of the basic components of any PSHA study (chapter 2). The EMCA catalogue used in this study was compiled at the Institute of Geophysical Research (IGR) of Kazakhstan. It started from the experience of the CASRI project (Central Asia Seismic Risk Initiative) (Mikhailova, 2009). The CASRI project was implemented between 2006-2009, with the participation of Kyrgyzstan, Kazakhstan, Uzbekistan and Tajikistan. One of the main products of this project was a unified catalogue including earthquakes from 2000-2005 with MLH magnitudes greater than 4.5. The EMCA catalogue is extended both spatially, by including territory of Turkmenistan (who were not involved in CASRI), and temporally, by updating the catalogue to 2009. For the compilation of the catalogue after 1990, different sources were used, including international and regional data centres and scientific publications (Mikhailova et al., 2015). Original waveform data was also analysed in case of controversial issues such as multiple earthquake locations, identification of false events, differentiation between earthquake and explosion etc.

The earthquake catalogue for the study area includes about 33600 earthquakes which have occurred in Central Asia and neighbouring areas over the time period between1800 and 2009. About 80 historical earthquakes, which occurred before 1800 C.E until 2000 B.C.E, are also included in the catalogue (Mikhailova et al., 2015). The catalogue contains events with magnitudes ranging from 1.5 to 8.3. The catalogue for the study region is significantly rich in large magnitude events and has 7805 events equal to or above MLH 4.0. MLH magnitude is based on the horizontal components of surface waves and is widely used in practice in the former USSR countries (Karnik, 1968). Figure 6.1 shows the un-processed catalogue for the study region. Different colours and shapes represent different magnitude sizes of the events. The catalogue is compiled by collecting information from several sources, such as regional

bulletins and catalogues being used to compile the catalogue from 1991 to 2009. Information about earthquakes occurring before 1990 is mainly taken from Kondorskaya and Ulomov (1995).

Instrumental observation of earthquakes in Central Asia started in 1901 with the installation of the first seismic station in Tashkent. For earthquakes before the instrumental period, the magnitudes were measured by the analysis of macroseismic data determined by iso-seismal maps. These macroseismic magnitudes are calibrated by values of surface waves magnitude MLH. After the installation of the first seismic station, several other seismic stations were installed in large cities, i.e., Samarkand (1925), Almaty (1927), Bishkek (1932), Chimkent (1932), and Semipalatinsk (1934). From time to time, this network was expanded, which also helped in improving the accuracy of the network as well as allowing smaller minimum magnitude detection. A coordination centre was established in Moscow, Russia, to join the efforts of all seismic services of the USSR's territory. Since 1964 until 1990, the same methodology and approaches were used for the seismic observations, data processing, catalogue compilation and creation of seismic zoning maps in a coordinated manner for all the USSR countries. However, the collapse of the Soviet Union in 1990 affected very much scientific practice in Central Asia. This resulted in a decrease in support for seismic networks, leading to a sharp decrease in the number of operating seismic stations, an interruption in the unified approaches among the countries, as well as in data exchange and the management system of the cross-border networks. This led to the problems associated with the efforts in restoring homogeneity to the seismic databases since 1991.

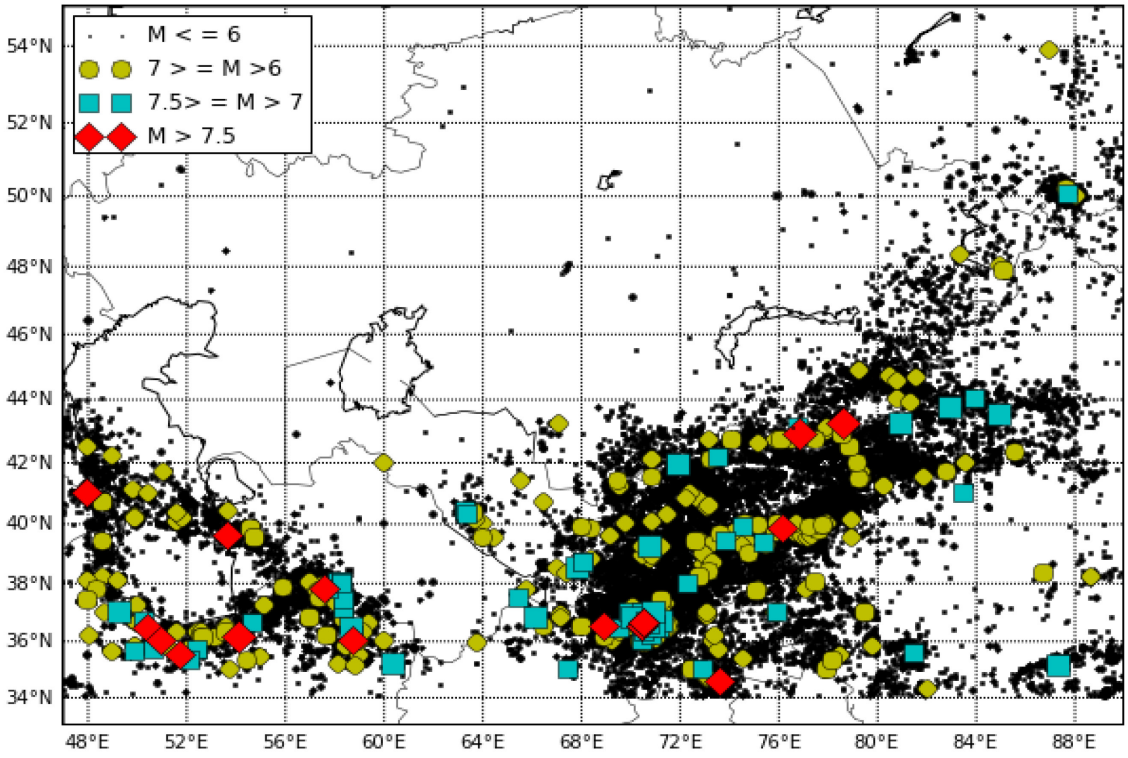


Figure 6.1 Un-declustered earthquake catalogue for the study region, having historical and instrument seismicity up to 2009. Different shapes and colours represent different magnitude scale as given in upper left corner.

The earthquakes magnitudes in Central Asia are characterized using different scales, the most common being the body-wave magnitude mb, the surface-wave magnitude MLH (Karnik, 1968) and the energy K-class (Rautian et al., 2007, and the references within). The reason for the use of different magnitudes is due to the applications of different scales, wave types, the instrumental features of global and regional networks, during different periods of time. For the joint application of catalogue data, different magnitudes are converted to the MLH scale, which is close to Ms values. This magnitude was chosen because, first, many large events were measured with this magnitude in the instrumental period, and second, the macroseismic magnitude of earthquakes that occurred before the instrumental period are based on it. Also, MLH magnitudes are used for characteristic seismic generating zones in the current maps of seismic zoning of Central Asia countries. Below are the empirical regression equations used to convert the energy class K and body wave magnitude mb to MLH (Mikhailova et al., 2015; Natalia Mikhailova, and personal communication to the author):

$$\mathrm{MLH} = 0.47\mathrm{K} - 1.15 \tag{6.1}$$

$$\mathrm{MLH} = 1.34\mathrm{mb} - 1.89 \tag{6.2}$$

The magnitude MLH is rounded to the first decimal unit before processing the catalogue further for analysis. The final EMCA catalogue contains 60 earthquakes with magnitudes greater than 7, as shown in Figure 6.1. The depth-magnitude distribution of the un-declustered catalogue is shown in Figure 6.2 with a minimum magnitude of 4.0. Therefore, magnitude MLH 4.0 is selected as the minimum magnitude here for further analysis as events below this do not cause any significant damage to structures. Most of the events in figure 6.2 are shallow and located down to 50 km depth.

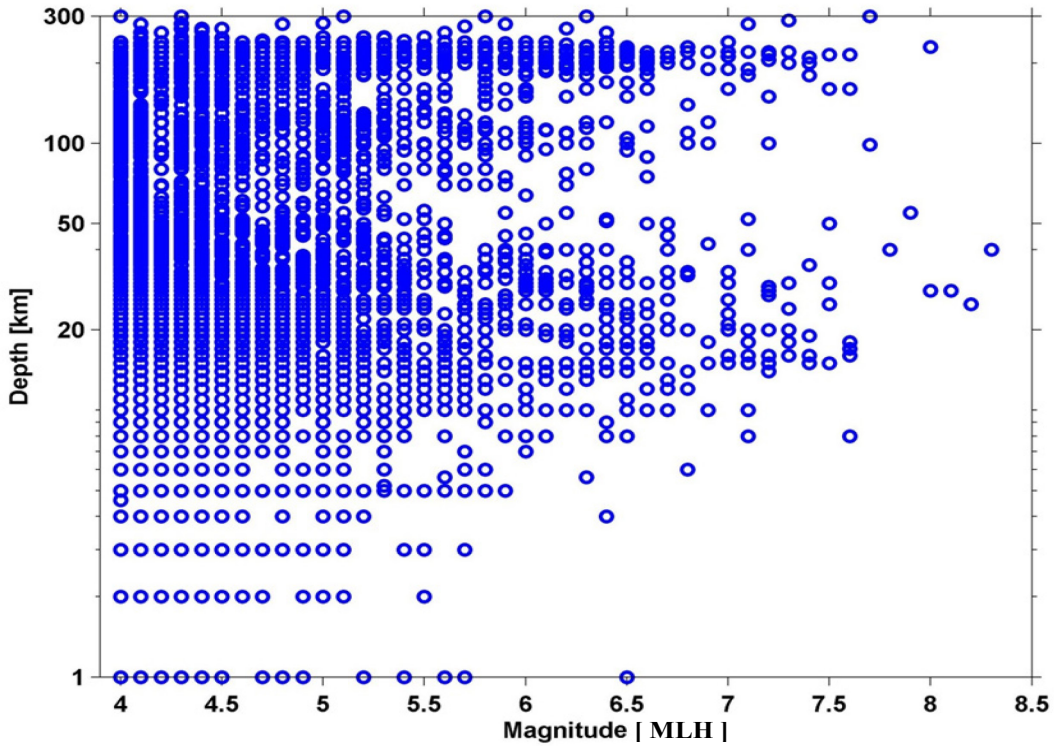


Figure 6.2 Magnitude-depth distribution of the un-declustered earthquake catalogue.

The deep seismicity (> 50 km) is mainly related to the Pamir and Hindu-Kush regions. Figure 6.3 shows the distribution of shallow seismicity down to 50 km depth (represented by black dots) and deep seismicity below 50 km (represented with red dots) in the region. Some occasional deep seismicity is also seen in the southern parts of Kyrgyzstan. Only earthquakes with focal depths less than 50 km are considered in the current study.

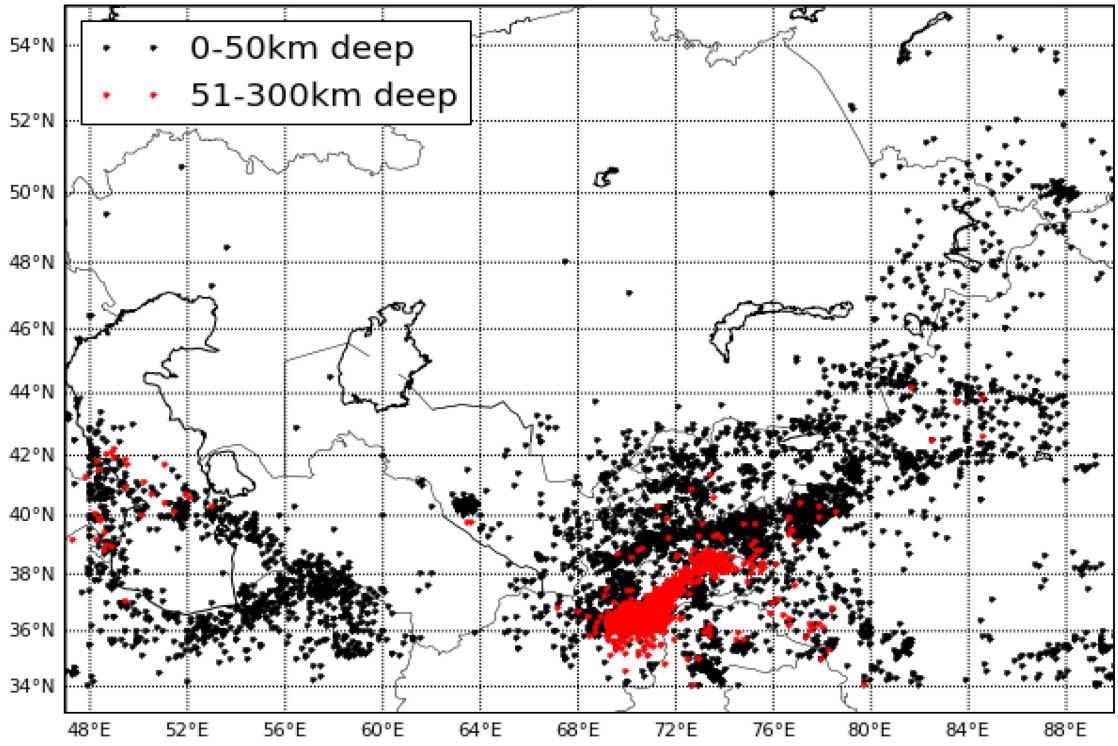


Figure 6.3 Seismicity depth distribution in Central Asia. The black dots represent seismicity down to 50 km depth, while the red dots represent the seismicity between 51 and 300 km depth.

Figure 6.4 shows a 3D histogram of the magnitude-depth distribution of seismicity. It is clear from Figure 6.4 that most of the deep events which are not considered further for hazard analysis are low magnitude events. However, although there are also some large magnitudes events (about magnitude 7) with focal depths greater than 150 km, the very deep events are not considered to be very important from a seismic hazard point of view, at least in this study, because they do not contribute significantly to the final seismic hazard.

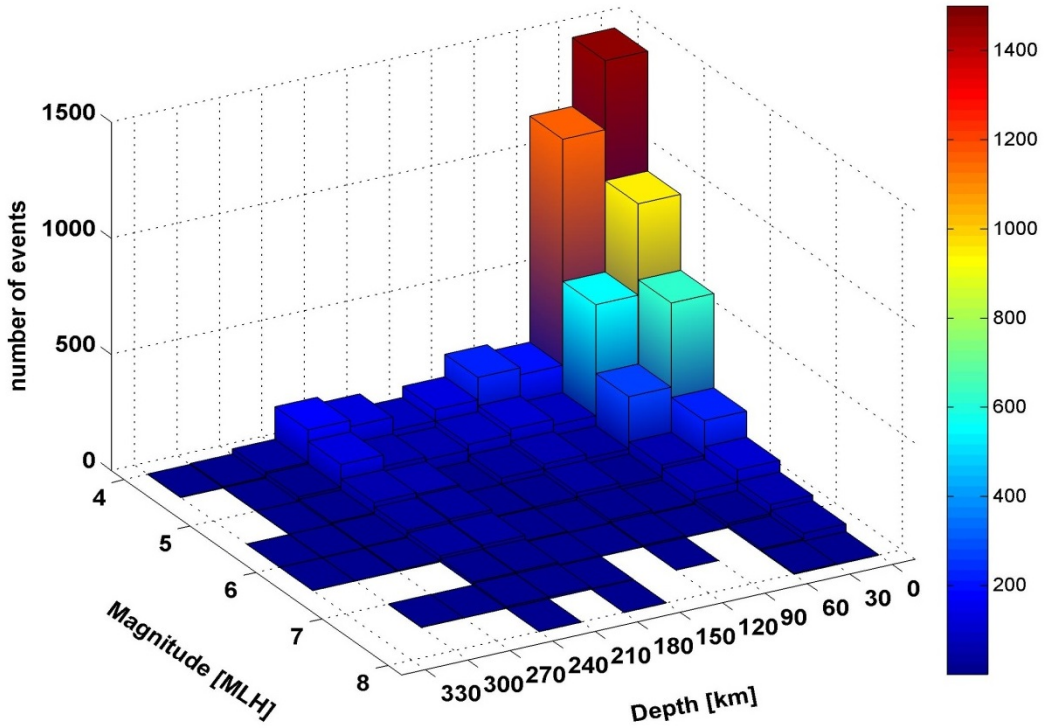


Figure 6.4 Magnitude-depth distribution showing number of events within magnitude-distance bins.

6.1.1 Declustering of the earthquake catalogue

PSHA based on the Cornell (1968) approach assumes a Poissonian process of earthquakes. This means that seismic events are considered temporally independent from each other. Hence, the earthquake catalog has to be de-clustered from dependent events before using in hazard calculations. Two de-clustering algorithms are considered in this study, namely the Gardner and Knopoff (1974) method (hereinafter referred to as GK), and the algorithm used in AFTERAN program (Musson, 1999) (hereinafter referred to as AF).

The GK method identifies the foreshocks and aftershocks within a given time and distance windows, which are fixed accordingly to the magnitude of the event. The events in a catalog are sorted in order of descending magnitude, and the dependent events (foreshocks and aftershocks) are identified if they are within the temporal and spatial windows of the largest event. The algorithm can therefore identify foreshocks and aftershocks by considering the windows forwards and backwards in time from the main shock. In the present study, the following magnitude scaling for the time and distance windows are considered: 1) the original windows as proposed by GK (1974); 2) the modified GK windows as proposed by Gruenthal and as reported in Van Stiphout et al. (2012); 3) the windows proposed by Uhrhammer (1986). These time and distance windows are summarized below:

GK

$$distance\ (km) = 10^{0.1238M+0.983}$$

$$\boldsymbol{Time\ (decimal\ days)} = \begin{cases} \boldsymbol{10^{0.032M+2.7389}\ If\ M\ \geq 6.5} \\ \boldsymbol{10^{0.5409M-0.547}\ Otherwise} \end{cases} \quad (6.3)$$

Gruenthal

$$distance\ (km) = e^{1.77+\ (0.037+1.02M)^2}$$

$$\boldsymbol{Time\ (decimal\ days)} = \begin{cases} \boldsymbol{e^{-3.95+(062+17.32M)^2}\ If\ M\ \geq 6.5} \\ \boldsymbol{10^{2.8+0.024M}\quad Otherwise} \end{cases} \quad (6.4)$$

Uhrhammer

$$distance\ (km) = e^{-1.024+0.804M}$$

$$\boldsymbol{Time\ (decimal\ days)} = \boldsymbol{e^{-2.87+1.235M}} \quad (6.5)$$

A comparison of these window sizes with magnitude are represented graphically for time and distance in figure 6.5. From the this figure, it is clear that the Uhrhammer window gives the shortest time and distance window for lower magnitude events, followed by GK (1974) and Gruenthal, whereas for large events, the Uhrhammer window gives the widest time and distance windows. The GK and Gruenthal relationships give approximately similar time and distance windows for large events.

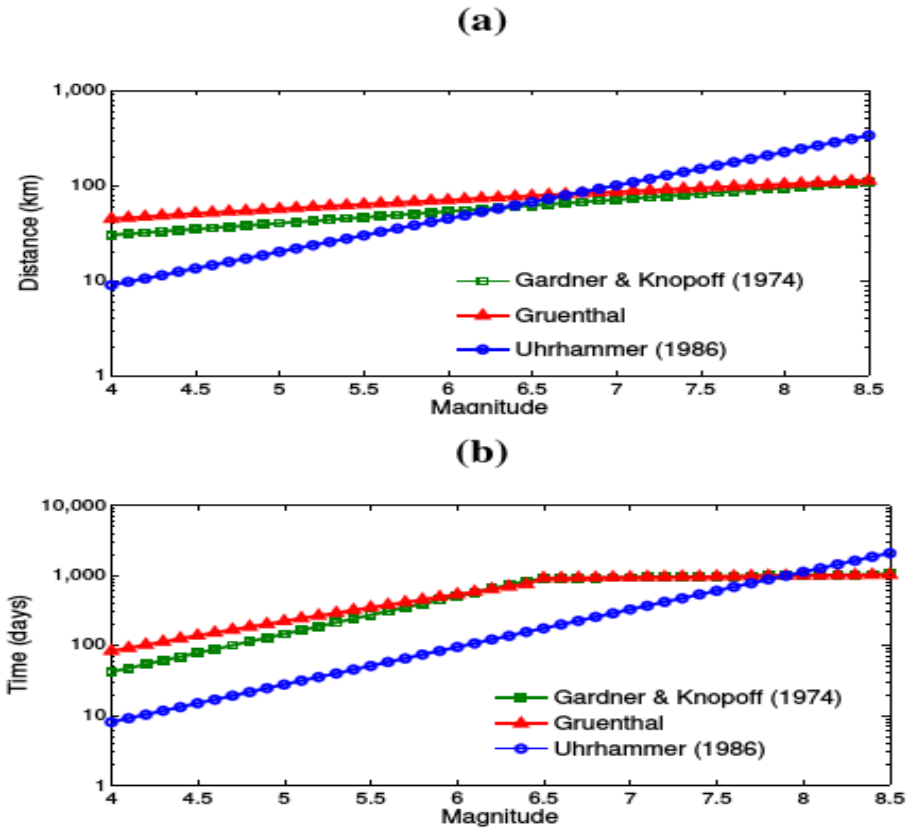


Figure 6.5 Scaling of time and distance windows with magnitude for different relationships. (a) Distance-magnitude relationships, and (b) time-magnitude relationships (Weatherill, 2014).

The AF approach is a modification of the GK approach, where a moving time window rather than a fixed time window is used. The events are first sorted into their magnitude-descending order. Then, events within fixed distance windows are identified using a moving time window of "T" days. The dependent events are declared when they occur both within the distance window and the T days' time-window. The time window is then moved to the next event, and the process is repeated.

The GK and AF algorithms are used as implemented in the OpenQuake supplementary software Hazard Modeller's Toolkit (Weatherill, 2014). The two algorithms are used with the above mentioned three windows. The GK algorithm is used for different ratios of foreshock and aftershock time and the AF algorithm is used for different lengths of the time windows. The results are summarized in Figure 6.6. From this figure, it is clear that the Gruenthal window identifies more events as being dependent, while the Uhrhammer window identifies fewer. Note that the GK algorithm identifies more dependent events not only in the instrumental part of the catalogue, but also those events with magnitudes higher than 7.0 in the historical part of the catalogue, which are supposed to be free of dependent events. In order to retain the maximum number of events and also ensure a Poissonian process, the AF algorithm with a GK distance window and 20 days' time-window is used to de-cluster the catalogue. Figure 6.7 shows the histogram of dependent events using the AF algorithm with GK distance window. As is clear, most events are in the lower magnitudes and in the most recent instrumental part of the catalogue, with about 400 events for magnitude 4. Smaller numbers of events are also identified in the historical part, but are not significant in number. Figure 6.8 shows the final earthquake catalogue for Central Asia after the de-clustering analysis. It shows about 77 events with magnitudes larger than 7 and 11 events larger than magnitude 7.5 down to a depth of 50 km.

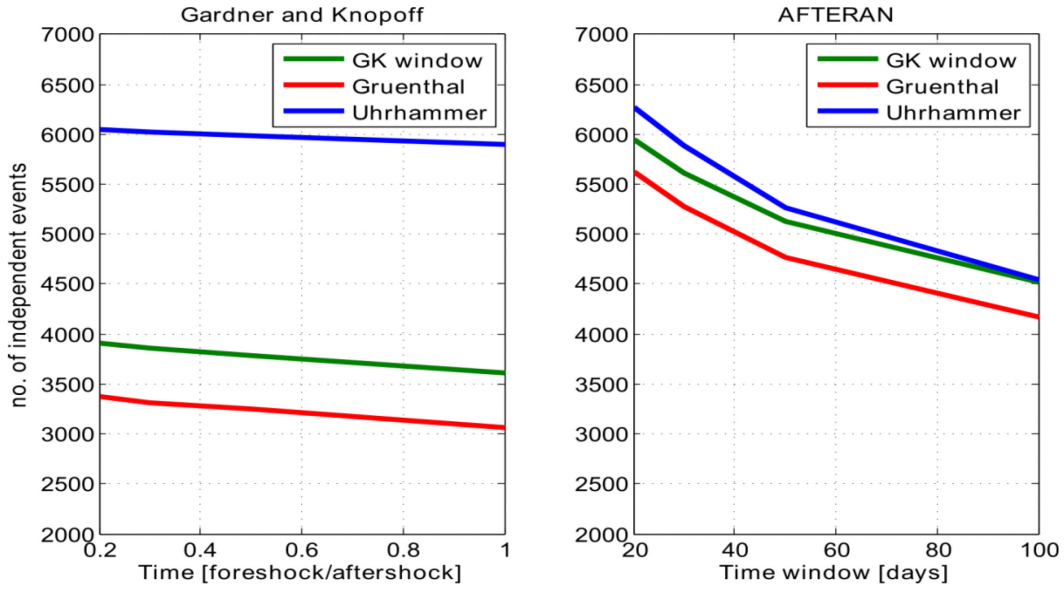


Figure 6.6 Comparison of declustering algorithms and different time and distance windows. Left side: Declustering using the GK algorithm for varying ratios of foreshocks/aftershocks. Right side: Declustering using the AF algorithm for different lengths of time window.

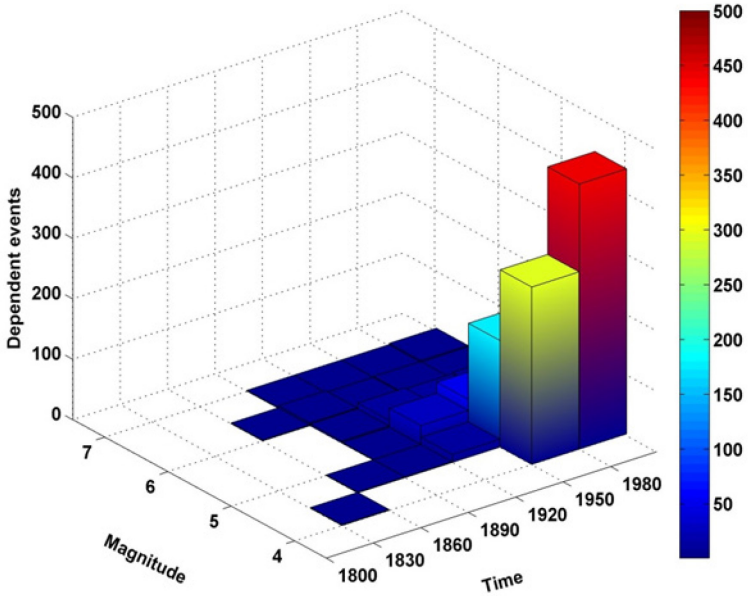


Figure 6.7 Characteristics of events identified as being dependent using the AFTERAN algorithm with the GK distance window (see Figure 6.5 (right)).

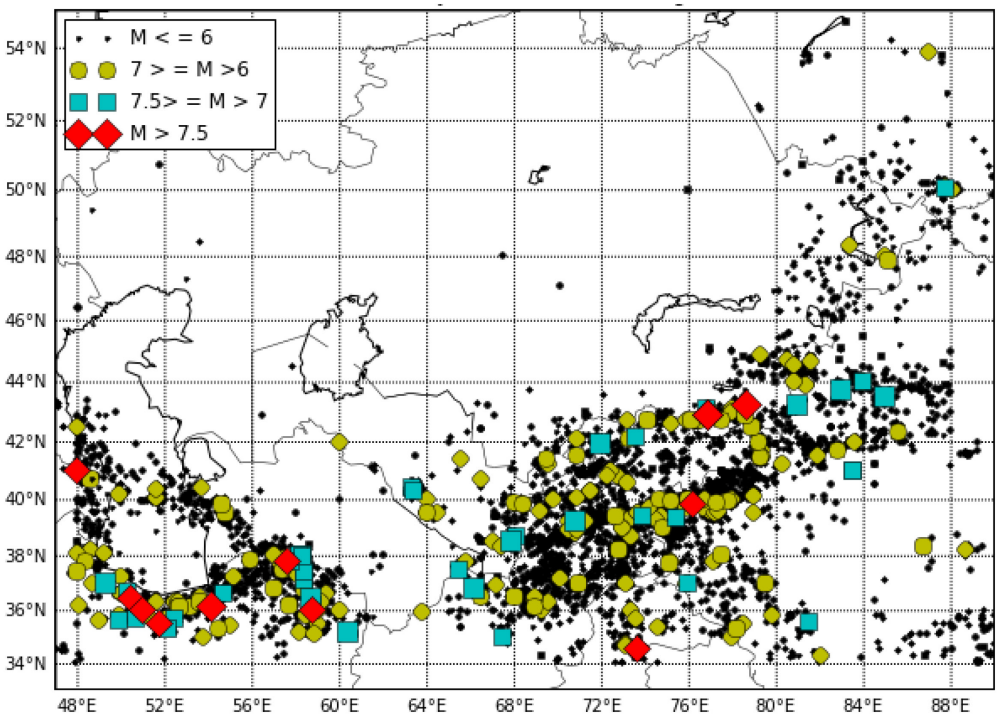


Figure 6.8 Final de-clustered earthquake catalogue for Central Asia for above magnitude 4.0 and a maximum depth of 50 km.

6.1.2 Earthquake catalogue completeness analysis

It is necessary to estimate the completeness of different magnitudes for the recurrence rates (chapter 2). Without a completeness analysis of the available samples of earthquake data, it is difficult to obtain fits for the recurrence analysis (Stepp, 1973). The magnitude of completeness (Mc) is defined as the lowest magnitude at which 100% of the earthquakes in a space-time volume are detected. The completeness of the earthquake catalogue is a function of magnitude and space (epicentre). The larger the magnitude of an event, the larger is its probability to be detected. Similarly, the detection probability of events also depends on the coverage of the seismic network. In this study, the method of Stepp (1973) is used, as also recommended by the GEM Modellers Toolkit, to estimate the time-completeness of different magnitude bins, considering the declustered catalogue. Stepp's method is an early analytical approach which is based on estimators of the mean rate of recurrence of earthquakes within given magnitude and time ranges. The completeness magnitude is identified when the observed rates of earthquakes above Mc start to deviate from the expected rate. If k_1, k_2, k_3 .... k_n are the numbers of events per unit time interval, assuming a Poissonian earthquake

sequence, the unbiased estimate of the mean rate of events per unit time interval of a given sample is given as:

$$\lambda = \frac{1}{n}\sum_{i=1}^{n} k_i \tag{6.6}$$

The variance for this is given by:

$$\sigma_\lambda^2 = \lambda/n \tag{6.7}$$

where n is the number of unit time intervals. For a unit time interval of 1 year, the standard deviation of the estimate of the mean is given by:

$$\sigma_\lambda = \sqrt{\lambda}\,/\sqrt{T} \tag{6.8}$$

where T is the sample length. Since the Poissonian assumption implies a stationary process, σ_λ behaves as $1/\sqrt{T}$ in the sub interval of the sample in which the mean rate of occurrence of a magnitude class is constant. The completeness period of a magnitude bin is inferred graphically from the analysis when σ_λ starts to deviate from $1/\sqrt{T}$. For the completeness analyses, the catalogue is divided into 0.5 magnitude interval starting from magnitude 4.0. The analyses are carried out using the GEM Hazard Modellers Toolkit. Figure 6.9 shows the results of the completeness analysis using the Stepp (1973) method for the regional declustered earthquake catalogue. The standard deviation (equation 6.8) is plotted for each magnitude bin vs the sample length T. The sample length is up to the total observation period of the catalogue, which extends until 2000 B.C.E. The black squares indicate the location in time where the standard deviation starts to depart from $1/\sqrt{T}$, hence indicating the completeness of that magnitude bin. For example, for magnitude bin 4.0-4.5, the deviation starts at a sample length of 53 years from 2009, hence representing a completeness period from 1956 for this magnitude bin. The results of this analysis for the regional earthquake catalogue are given in table 6.1.

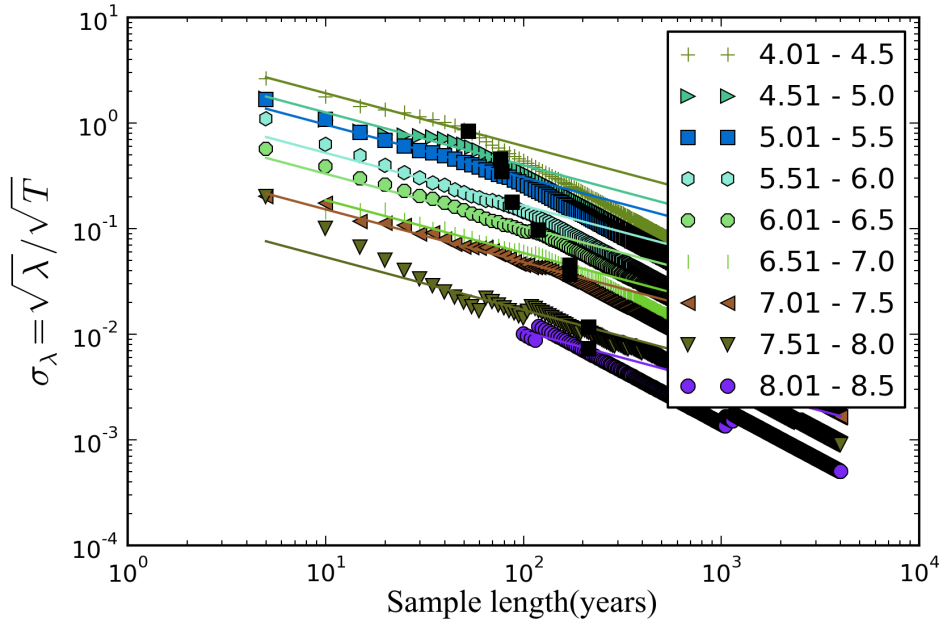


Figure 6.9 Completeness analysis using the Stepp (1973) method for the regional declustered earthquake catalogue. Different symbols in then legends represent the magnitude bins.

Figure 6.10 shows graphically the results of the completeness analysis for the regional catalogue extracted from Figure 6.9. From Figure 6.10, according to the Stepp method, the regional catalogue is completed from about 1956 above magnitude 4. This is when the monitoring network in central Asia was upgraded. Due to frequent moderate to strong events, events above magnitude 6 are complete from 1890.

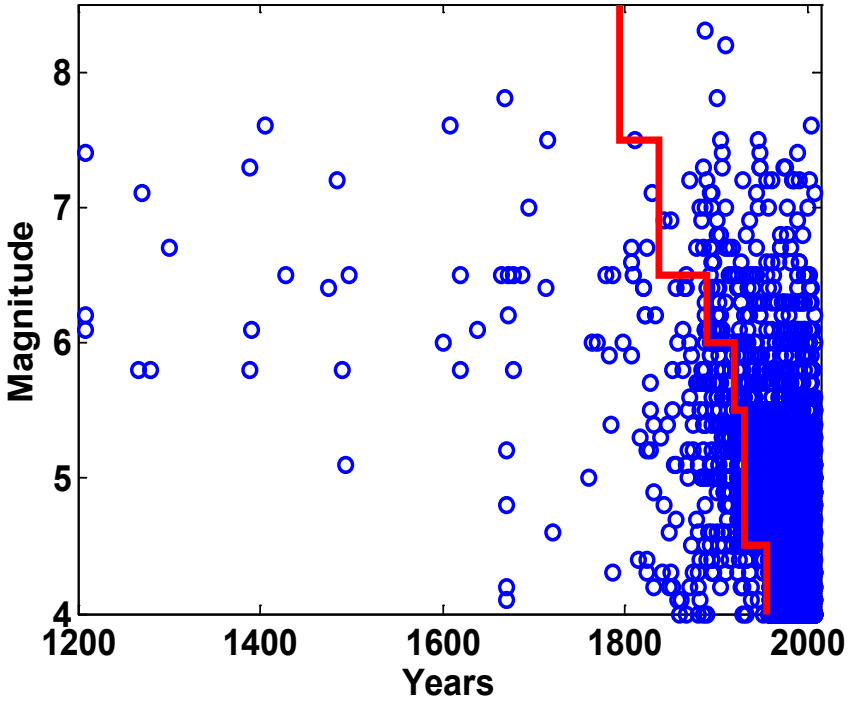


Figure 6.10 Completeness analysis for the whole declustered catalogue using Stepp's Method (Stepp, 1973). The red line shows the completeness period corresponding to each magnitude bin.

6.1.3 Earthquake recurrence analysis

The time-completeness intervals for the different magnitude bins are used to estimate the recurrence relation (Gutenberg and Richter, 1944) for the declustered catalogue. In this study, the Weichert (1980) method is used to estimate the b value (equation 2.1). This method estimates the maximum likelihood of b values for grouped magnitudes bins and allows for unequal periods of observation. According to this method, the likelihood function L of ß ($ß = b\ln(10)$) for n_i events in magnitude class m_i is given as:

$$L(\beta|n_i, m_i, t_i) = \frac{N!}{\prod_i n_i!}\prod_i Pi^{n_i} \tag{6.9}$$

where the term Pi is given as:

$$Pi = \frac{t_i \exp(-ßm_i)}{\sum_j t_j \exp(-ßm_j)} \tag{6.10}$$

An extremum of $\ln(L)$ is obtained for

$$\frac{\sum_i t_i m_i \exp(-ßm_i)}{\sum_j t_j \exp(-ßm_j)} = \frac{\sum n_i m_i}{N} = \bar{m} \tag{6.11}$$

This is solved for ß by using an iterative scheme. Figure 6.11 shows the recurrence relationship established for the regional catalogue, along with the observed values obtained from the catalogue. The relationship shows a good fit for smaller to moderate magnitude events, but it slightly overestimates the frequency of magnitudes larger than 7.

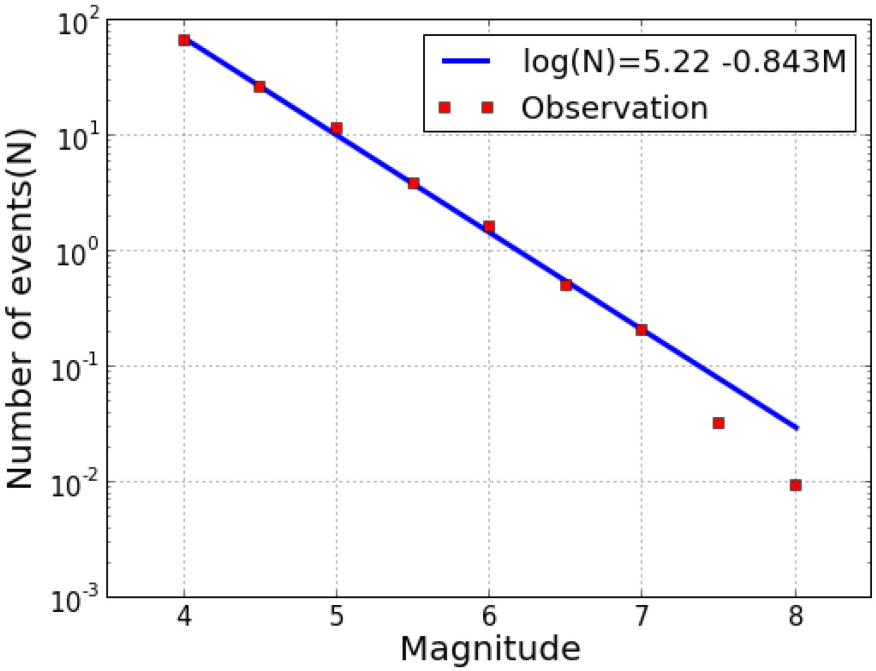


Figure 6.11 Magnitude-frequency relationship for the whole declustered catalogue based on the Weichert method, considering the regional completeness for different magnitude bins. The blue line shows the recurrence relationship, while the red squares shows the observations from earthquake catalogue.

6.2 Seismic source models for Central Asia

Two basic approaches for defining a suitable seismic source model are used in this study for Central Asia. These include the smooth seismicity approach and the area source model (chapter 2). In the smooth seismicity source model, the Frankel (1995) and Woo (1996) approaches are used. In the Frankel (1995) approach, which is based on fixed kernel distances, a new approach is developed in this study which uses adaptive kernel distances. These source models are described below.

6.2.1 *Smoothed seismicity model based on the Frankel (1995) approach with fixed kernel*

The smoothed seismicity approach of Frankel (1995) is applied considering a regular grid spacing of 0.2x0.2 degrees. Since this method requires the use of b-values (equation 2.1), both

uniform b-values and gridded b-values (non-uniform) are considered. Gridded b-values (or varying b-values) are considered because the seismicity changes from region to region, depending upon the tectonic regime. The uniform b-value is taken as a regional b-value estimated from the whole earthquake catalogue (Figure 6.11). The gridded b-values are taken from the super zones of the area source model (section 6.2.4). After calculating the cumulative activity rates in each grid point, the seismicity is smoothed using a fixed Gaussian kernel. In this study, after testing different values of smoothing distances, a fixed distance of 30 km is used for the Gaussian kernel. After the smoothed activity rates are calculated for each grid point, the seismic hazard is calculated using the OpenQuake software.

6.2.2 *Smoothed seismicity model based on the Frankel (1995) approach with adaptive kernel*

In the Frankel (1995) approach, a fixed Gaussian kernel is applied at each grid point to smooth the seismicity. This smoothing process takes into account the uncertainty in the location of the seismicity. However, earthquakes follow spatial clustering (Kagan and Jackson, 1991; Stock and Smith, 2002). This means less uncertainty in the location of future seismicity where there is higher density of events compared to a region having more scattered seismicity. Therefore, the adaptive Gaussian kernel for smoothed seismicity is also implemented in this study, using the method of Stock and Smith (2002). The adaptive kernel is based on the idea that, instead of global uniform values, different smoothing correlation distances (bandwidths) are used, depending on the density local seismicity. The adaptive kernel is estimated in three steps. First, the pilot estimator $g_1(i)$ is calculated for each cell i summed over each earthquake, as given below.

$$\boldsymbol{g_1(i) = \frac{1}{N}\sum_{e=1}^{N} exp\left(-\frac{(x_i - x_e)^2}{c_1^2}\right)} \tag{6.12}$$

where N is the total number of earthquakes and c_1 is the global smoothing kernel parameter (equation 2.4) and x_e is the location of each earthquake. The pilot estimator is used to determine the local bandwidth c_2, as given below.

$$\boldsymbol{c_2(i) = \sqrt{\frac{\mu}{g_1(i)}}} \tag{6.13}$$

where μ is the global mean of earthquake activity per area during the observation period and is calculated as:

$$\boldsymbol{\mu = \frac{Total\ no.of\ earthquakes}{total\ area\ (km^2)}} \tag{6.14}$$

The local bandwidth c_2 is used in equation (2.5) of Frankel (1995) approach as given below:

$$\boldsymbol{\tilde{n}_i = \frac{\sum_j \frac{n_j exp\left(-\Delta_{ij}^2 / c_1^2 c_2(j)^2\right)}{c_2(j)^2}}{\sum_j \frac{exp\left(-\Delta_{ij}^2 / c_1^2 c_2(j)^2\right)}{c_2(j)^2}}} \tag{6.15}$$

Equation (6.15) has the ability to adapt the degree of smoothing based on the density of seismicity in the region of interest. Similarly to equation (2.4), here the sum is also taken over cells j within a distance of 3 $C_1.C_2(i)$ of cell i. Equation (6.13) implies that the higher the seismicity at a place, the lower will be local bandwidth. Furthermore, multiplying it's value with the global correlation distance in equation (6.15) will result in a net lower smoothing distance, and vice versa.

Figure 6.12 represents the local bandwidth C_2 with global bandwidth $C_1 = 30$ km for the study area. Areas of higher seismicity have lower values of local bandwidth as is shown by dark blue colour. Lower seismicity areas have larger values of local bandwidth, and are shown by lighter to dark red.

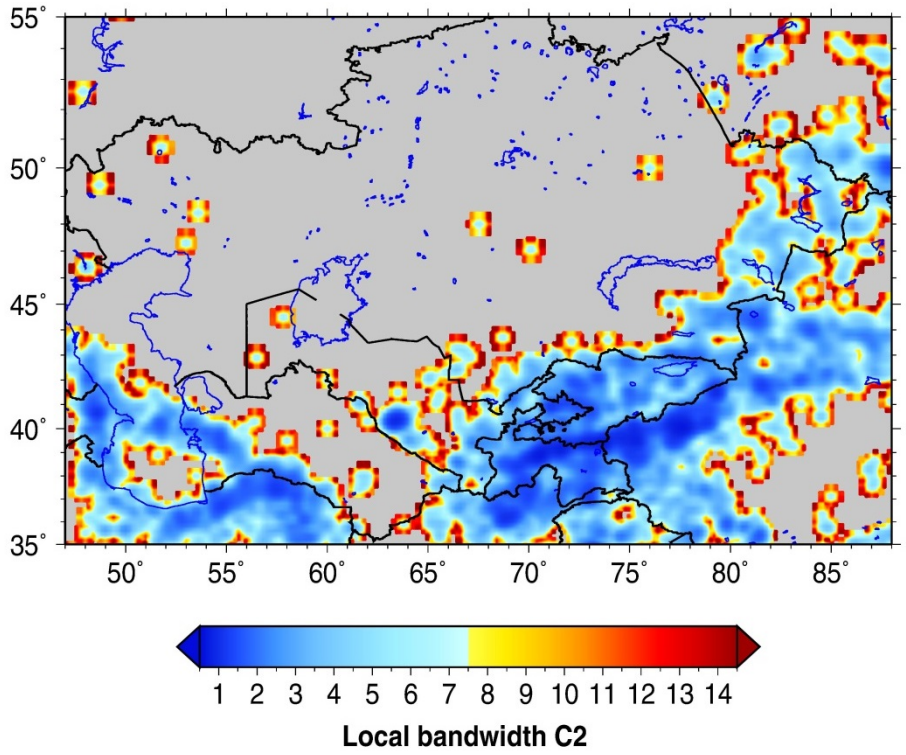


Figure 6.12 Local bandwidth C2 (equation 6.13) with C1= 30km for the study area.

6.2.3 *Smoothed seismicity model based on Woo (1996) approach*

In the Woo (1996) approach, the activity rates are calculated for the same grid size of 0.2x0.2 degrees, according to equation (2.6). The activity rates are estimated for magnitude bin size of 0.5. The seismicity is smoothed using a multi-variate probability density function (equation 2.7). The parameters c and d used in the smoothing bandwidth H (equation 2.8) are also calculated at each grid point as described in section (2.4.2). For this, the events are considered at a maximum distance of 250 km from the grid point. An exponential cure is fit to the data set to find the parameters c and d.

The Woo approach also accounts for the uncertainty in magnitude and location of events. An uncertainty of 0.5 and 0.2 for magnitude is used for earthquakes that occurred prior to or later than 1960, respectively, which corresponds to about the completeness time of magnitude 4. Regarding the uncertainty in epicentral location, the values indicated in the catalogue are used, when available. The uncertainty in epicentral location is not available for some events after 1991, and hence an uncertainty of 5 km was used. After calculating the activity rates for each grid point source, the Intensity Prediction Equation (IPE) is applied to calculate the

hazard. In the original code of Woo, the hazard is calculated at only a single point. The code has been modified in this study to account for multiple point hazard estimation.

6.2.4 *Area source model*

The seismic faults for Central Asia are not well studied. Therefore, due to the lack of detailed information about the local faults, the area sources for the study region are defined by mainly considering the pattern of crustal seismicity down to 50 km depth. Although tectonic and geological information, such as the position and strike distribution of known faults, have also been taken into account when available. The boundaries of the area sources along with the declustered seismicity distribution are shown in Figure 6.13. Large area sources (see, for example, those numbered in black, 1, 2, 5, 45 and 52 in Figure 6.13) are defined where the seismicity is scarce and there are no tectonic or geological features that would justify a further subdivision. Smaller area sources (e.g., numbers 36 and 53 in Figure 6.13, around the Talas-Fergana fault) have been designed where the seismicity can be assigned to known fault zones. Amongst these, it is worth mentioning the sources numbered 9, 57, and 58 (Figure 6.13), including the seismicity generated mainly by the Darvaz and Gissal Kokshal faults (Chapter 5, Figure 5.2). These area sources include the largest number of earthquakes (457 events), the largest one with a magnitude of 7.4. The area source 17, which includes the western Tarim basin and southern Tien Shan, is dominated by the seismicity of the Atushi-Keping fault. It is the second area source in terms of number of earthquakes with the largest event having a magnitude 6.7. Area source 13 covers the Tajik depression zone (234 events, 6.7 being the magnitude of the largest event). Of particular interest are the area sources 46 and 47,covering the Kopet-Dagh range (Turkmen-Khorasan Mountain Range), where the largest number of events greater than magnitude 7 (a total of 15) have occurred, including the 1948 MLH 7.3 Ashgabad earthquake (area source 47).

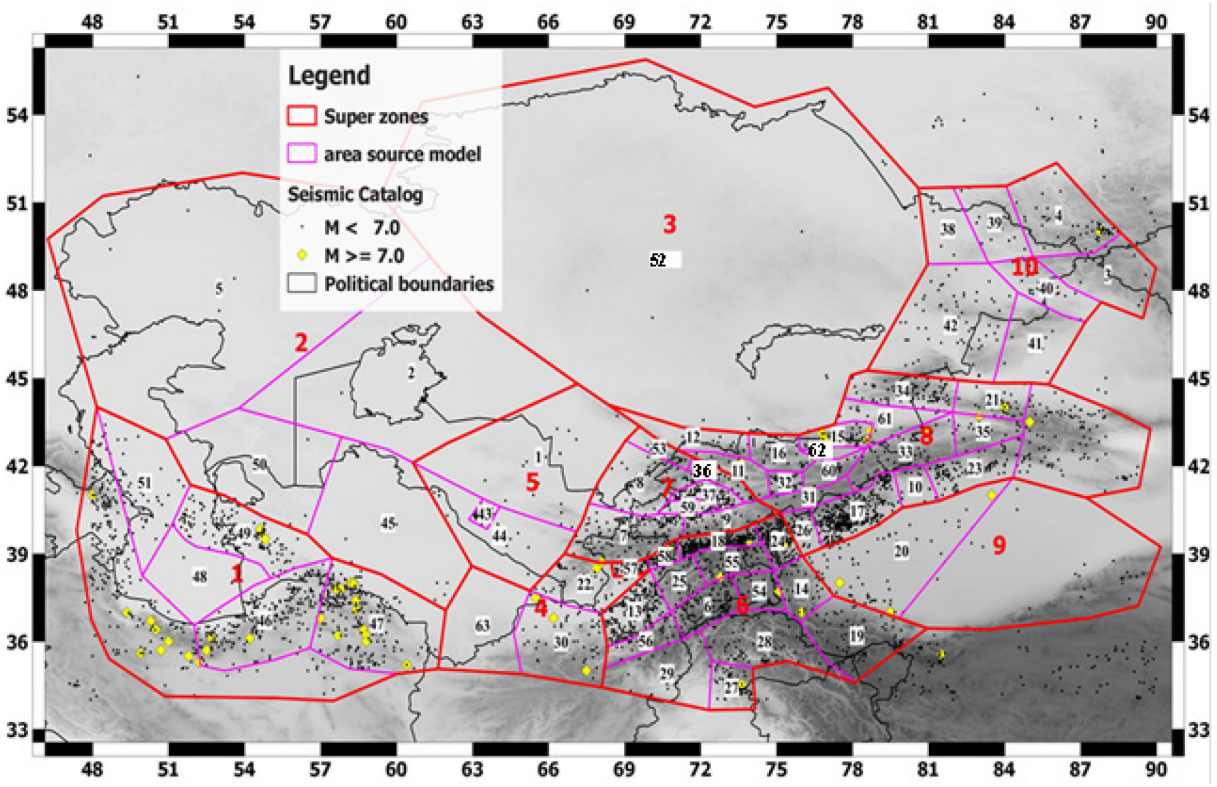


Figure 6.13 Seismic sources for the study area. The red polygons represent the super zones, the magenta colour represents the area sources. The numbers with red and black colour represent the numbering of super zones and area sources, respectively. The black dots represent the declustered earthquake catalogue MLH < 7.0, whereas the yellow diamond shapes represent events MLH >= 7.0.

The area source number 15, covering the northern Tien Shan, includes the largest known events that have affected the Central Asia region, namely the1889 M 8.3 Chilik earthquake, which is one of the largest historic intra-plate reverse faulting events, and the intra-continental 1911 M 8.2, Chon-Kemin earthquake.

Finally, area source number 43, covering part of the territory of Uzbekistan (Turan block) also includes the Gazli gas field. This area source has a peculiar concentration of seismicity within short time periods, with earthquakes larger than magnitude 7. In particular, three events with M>= 7.0 have occurred over an eight year time interval (1976-1984). These three events, which occurred in an intra-plate region, where relatively low seismic activity was known before, are considered to be possibly induced by gas extraction (Simpson and Leith, 1985). Therefore this separate source is adapted for this region.

In order to obtain a robust estimation of the necessary parameters for PSHA derived by the statistical analysis of the seismicity, due to the scarcity of data in some of the areas covered

by the model, super zones (red polygons in Figure 6.13) are introduced. These super zones are defined by combining area sources based on similarities in their tectonic regime, and taking into account local expert's judgments. Figure 6.13 shows the super zones that have been used to estimate: (1) the completeness time of the earthquake catalogue, (2) the depth distribution of seismicity, (3) the tectonic regime through focal mechanisms analysis, (4) the maximum magnitude and (5) the b values via the GR relationship (equation 2.1).

Figure 6.14 shows the completeness and the GR recurrence parameters for super zones 4 and 7. The completeness analysis (Figure 6.14 (a) and (c)), which is carried out using Stepp's method, shows completeness since 1958 and 1955 for earthquakes above magnitude 4.0 for super zones 4 and 7, respectively. Figure 6.14 (b) and (d) show the recurrence analyses for super zones 4 and 7, respectively, using the Weichert (1984) method. Super zone 4, which is dominated by large magnitude events compared to lower magnitude events, has a lower b value of 0.6, whereas super zone 7, which has a greater number of lower magnitude events compared to higher magnitude events, has a higher b value of 0.83. The plots of completeness and recurrence analyses for all the super zones are given in the appendix.

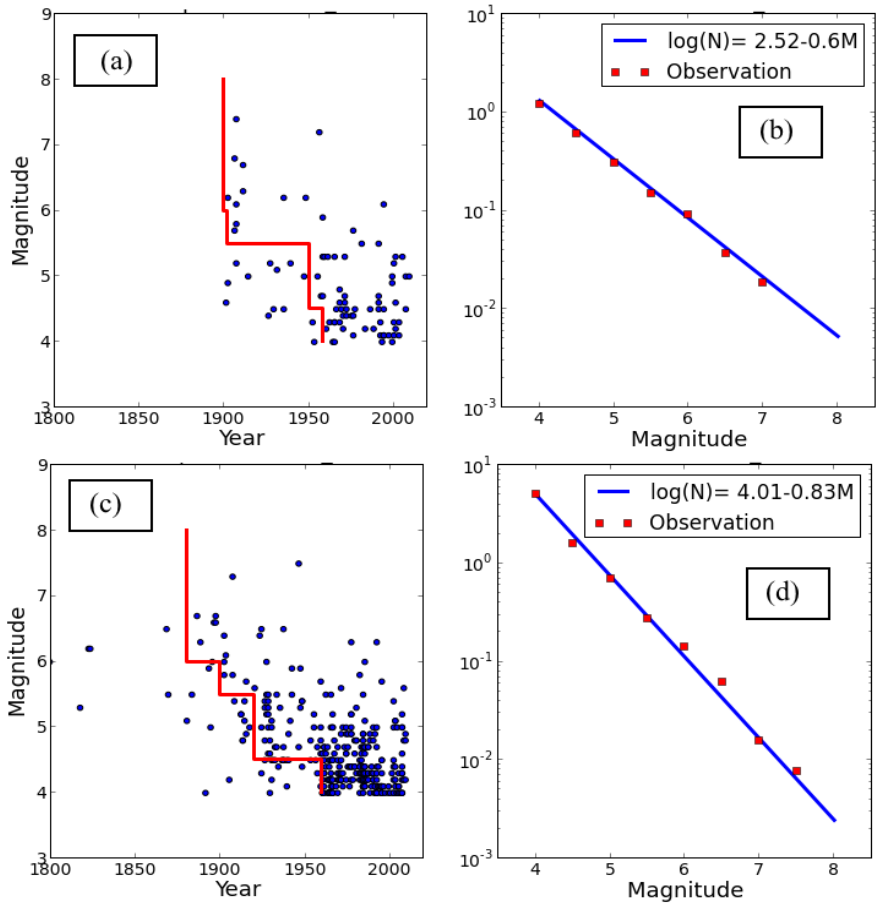


Figure 6.14 Completeness analyses and recurrence parameters estimation for super zones. (a) and (b) shows the completeness analysis and GR parameters for super zone 4. (c) and (d) shows the completeness analysis and GR parameters for super zone 7.

Table 6.1 summarizes the results of the completeness analysis for the super zones defined in Figure 6.13. The magnitude represents the central part of the bins. Super-zones 2 and 3, which are stable continental regions, do not contain enough events for statistical analysis, and hence these super-zones are assigned the completeness of the regional catalogue.

Table 6-1 Completeness analyses for the declustered catalogue above magnitude 4.0 for the super zones indicated in Figure 6.13.

Magnitude bins / Super zone	**4.25**	**4.75**	**5.25**	**5.75**	**6.25**	**6.75**	**7.25**	**7.75**	**8.25**
Whole catalogue	1956	1932	1931	1920	1890	1837	1837	1795	1795
1 Iranian block	1960	1928	1922	1885	1885	1820	1820	1820	1820
2 Kazakh block	--	--	--	--	--	--	--	--	--
3 Kazakh block	--	--	--	--	--	--	--	--	--

4 Tajik-Afghan block	1958	1950	1902	1900	1900	1900	1900	1900	1900
5 Turan block	1975	1975	1930	1930	1930	1930	1930	1930	1930
6 Pamir-Hindukush block	1962	1950	1940	1921	1921	1921	1921	1921	1921
7 Fergana block	1955	1920	1920	1900	1880	1880	1880	1880	1880
8 Tien Shan	1955	1930	1920	1890	1890	1870	1870	1870	1870
9 Tarim Basin	1970	1970	1970	1970	1945	1930	1930	1930	1930
10 Eastern Kazakh block	1958	1956	1956	1886	1866	1866	1866	1866	1866

The time-completeness interval for the different magnitude bins for each super zone is used to estimate the b-value (equation 2.1) GR parameter. Since super zones 2 and 3 belong to the stable Kazakh block, they do not have enough seismicity to determine this parameter. Hence, in this case, the b-values for these two zones are assumed to be equal to 1 (which is the average b value considered worldwide).

The earthquake catalogue for focal mechanism is extracted from the Harvard Global Centroid Moment Tensor Catalog (Ekström and Nettles, 2013). For the focal mechanism classification, the Boore et al. (1997) convention is used. This means that an event is considered to be strike-slip if the absolute value of the rake angle is <=30 or >=150 degrees, normal if the rake angle is <-30 or >-150 and reverse (thrust) if the rake angle is >30 or <150 degrees. Figure 6.15 shows the focal mechanism distribution for the study area, where blue represents reverse faulting, green strike-slip and red normal faulting. As can be seen, the study area is dominated by reverse faulting. There are some occasional normal events in the Pamir and Hindu Kush regions, as well as along the Caspian Sea in super zone 1. However, strike-slip events are found to be scattered throughout the region. The distribution of source mechanisms and their weights are estimated for the super zones. These distributions, along with the weights, are shown in Table 6.2.

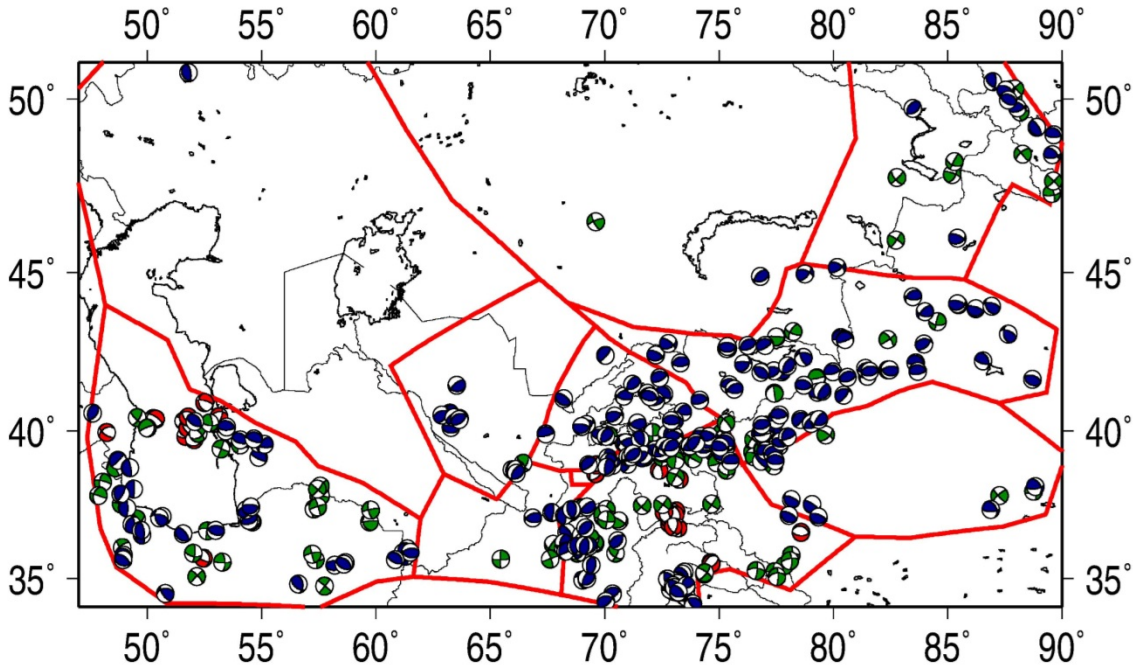


Figure 6.15 Earthquake focal mechanism map extracted from Harvard Global Centroid Moment Tensor Catalog. The blue, green and red colours represent reverse, strike-slip and normal faulting mechanisms, respectively.

For area sources, the maximum magnitude is usually taken from the historical seismicity, but due to some uncertainties in the magnitudes of the largest events, the opinions of the local experts are also included in assigning the maximum magnitude to each super zone. The results of the recurrence analysis for the super zones, along with the maximum magnitudes, are summarized in Table 6.2. For example, a maximum magnitude of 8.3 is assigned to the Tien Shan super zone. This zone shows a history of greater seismicity, including the Chilik and Kemin earthquakes. Super zones 2 and 3 are each assigned a maximum magnitude of 6, after Mooney et al. (2012), which concludes after analyses and observation of modern datasets that at least an event of magnitude 6 can occur anywhere in the world.

Table 6-2 Characteristics of the super-zones (see Figure 6.13).

S.No	Name	b	Mmax	Hypocentral depth distribution						Focal Mechanism Distribution					
				D1*	W1*	D2	W2	D3	W3	RA1*	W1*	RA2	W2	RA3	W3
1	Iranian Block	0.88	7.3	10	0.5	35	0.25	50	0.25	10	0.4	-100	0.2	100	0.4
2	Kazakh Block	1	6	15	1	--	--	--	--	70	1	--	--	--	--
3	Kazakh Block	1	6	15	1	--	--	--	--	160	0.5	120	0.5	--	--
4	Tajik-Afghan Block	0.6	7.3	8	0.3	20	0.4	35	0.3	160	0.5	60	0.5	--	--
5	Turan Block	1.06	7.3	15	1	--	--	--	--	152	0.1	100	0.9	--	--
6	Pamir-Hindukush -Karakoram	0.88	7.3	6	0.5	35	0.5	--	--	10	0.4	-100	0.1	100	0.5
7	Fergana Block	0.83	7.5	15	0.7	35	0.3	--	--	10	0.1	100	0.9		
8	Tien Shan	0.84	8.3	15	0.4	35	0.6	--	--	10	0.2	-80	0.1	100	0.7
9	Tarim Basin	0.70	6.5	33	1	--	--	--	--	10	0.5	100	0.5	--	--
10	Eastern Kazakh Block	0.72	6.5	15	1	--	--	--	--	10	0.5	100	0.5	--	--

D1* =Hypocentral depth, RA* =Rake angle ,W1*=Weight

For processing the GR parameters (a and b values) for the area sources, the completeness analysis results estimated for the super zones are assigned to the respective smaller area sources. If the individual area source has at least 20 events, the GR parameters are then estimated for the area source. Otherwise, the b value is adopted from the respective super zone to which the smaller area source belongs, and the a value is estimated based on the Weichert (1980) method. This ensures the stability in the b value as well as the variation of activity rate for different sources. The plots for recurrence analyses for all the area sources are given in the appendix. Figure 6.16 shows the number of events within each area source. The light colour shows the sources with low seismicity, i.e., less than 20 events. The numbers inside each area source shows the number of events for that source. A total of 21 source areas have less than 20 events.

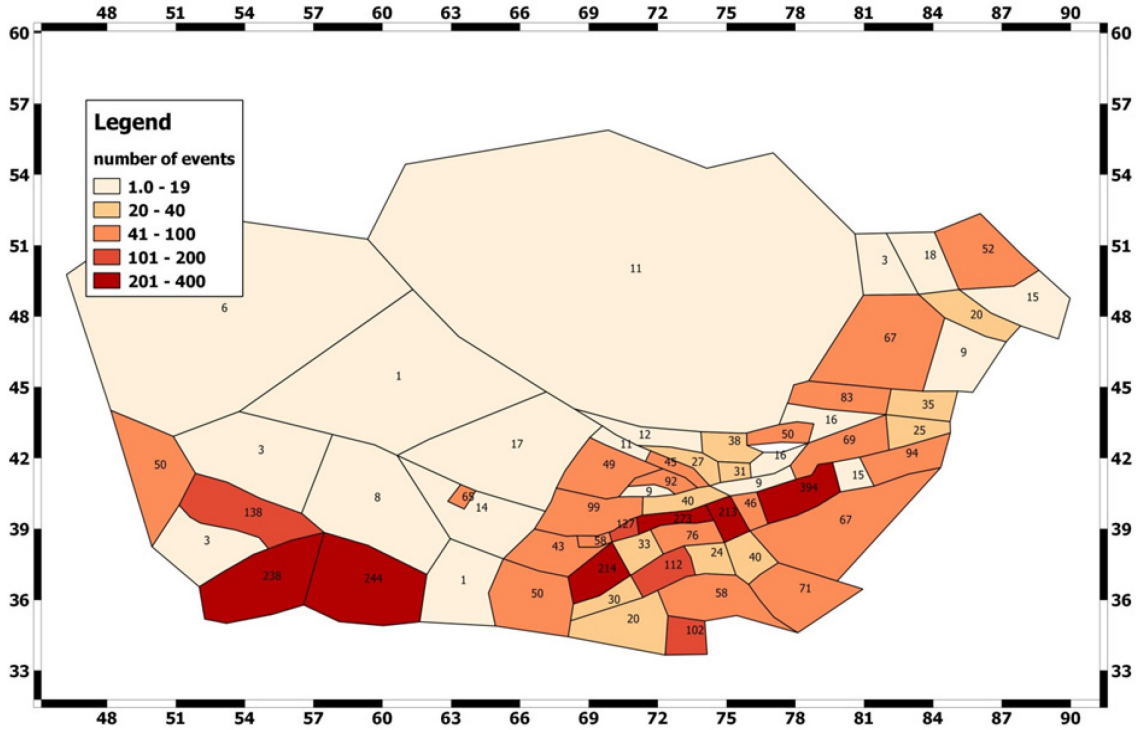


Figure 6.16 Seismicity inside each area source. The numbers in each area source represent the number of events. The variation in colour represents the seismicity according to the legend.

Figure 6.17 shows the b values distribution of GR recurrence relationship for the area sources. The light colours represent the regions with lower b values, whereas the darker colour shows regions with higher b values. Area sources having light colours (lower b values) are dominated by higher magnitude seismicity compared to regions with darker colours. Figure 6.18 shows the a values' distribution for the area sources. Lighter colour represents the lower values and darker colour shows the higher a-values. The a values range from 1 to 4.69.

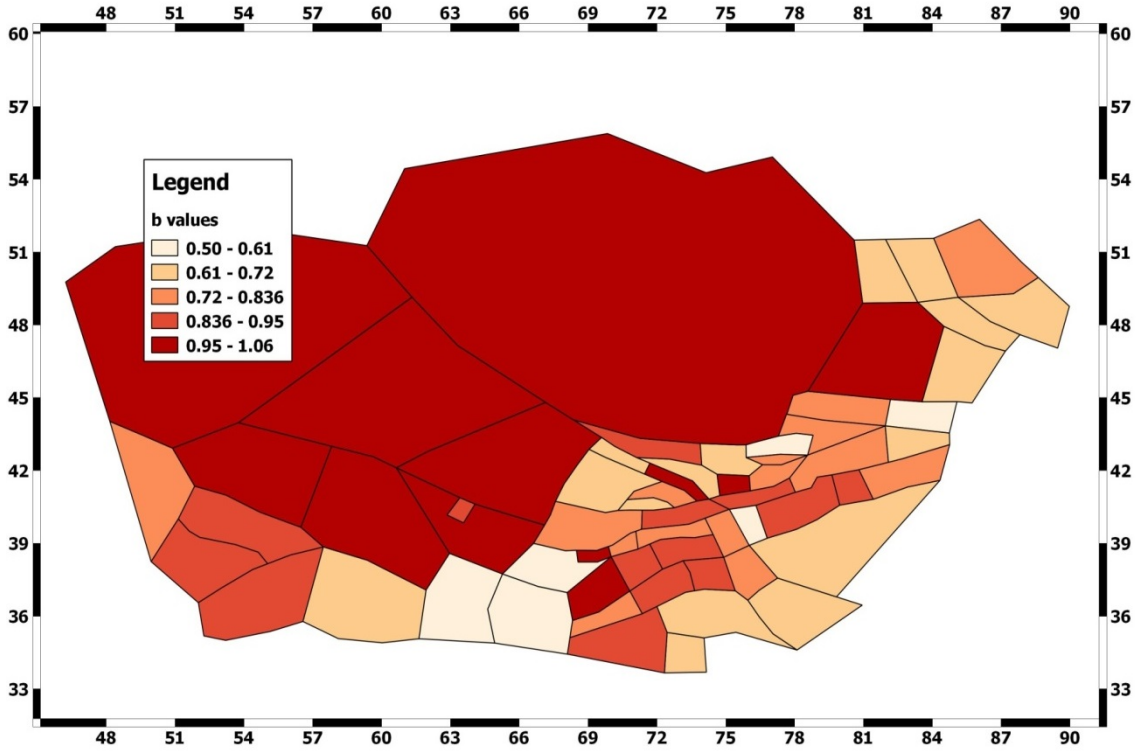


Figure 6.17 b-value distribution (of GR relationship) for the area sources model.

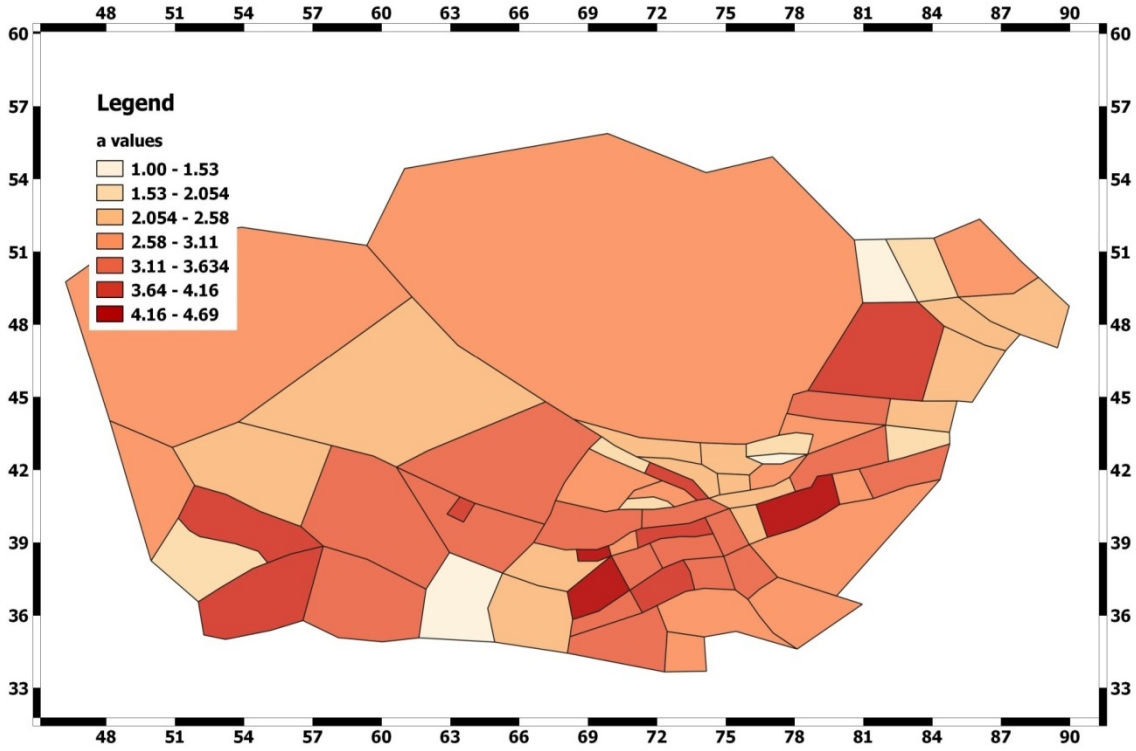


Figure 6.18 a-value distribution (from the GR relationship) for thr area source model.

For hazard calculations, each area source is assigned the maximum magnitude of their respective super zone (Table 6.2). The different parameters for the area sources are

summarized in Table 6.3. Since the IPE used in this study is considering the hypocentral depths of the events, in Table 6.3 the hypocentral depth distribution is estimated from the seismicity inside each super zone. The depth distribution is considered for maximum up to three values. Based on the number of events, the weights are assigned to each distribution. These depth distributions, along with corresponding weights, are further assigned to the area sources within the same super zones. Maximum magnitude uncertainty is considered by using a weighting scheme for the three values in a logic tree approach i.e., Mmax + 0.5, Mmax and Mmax – 0.5 of 0.2, 0.6 and 0.2, respectively.

Table 6-3 Area source parameters. (continues on next page)

Source. No	Super Zone	Mmin	Mmax	a	b	Hypocentral depth distribution					
						D1*	W1*	D2*	W2*	D3*	W3*
1	5	4	7.3	3.4	1.06	15	1	--	--	--	--
2	2	4	6	2.3	1.00	15	1	--	--	--	--
3	10	4	6.5	2.3	0.72	15	1	--	--	--	--
4	10	4	6.5	2.8	0.74	15	1	--	--	--	--
5	2	4	6	2.9	1.00	15	1	--	--	--	--
6	6	4	7.3	3.9	0.87	6	0.5	35	0.5	--	--
7	7	4	7.5	3.1	0.76	15	1	--	--	--	--
8	7	4	7.5	2.6	0.70	15	1	--	--	--	--
9	7	4	7.5	3.4	0.89	15	0.4	35	0.6	--	--
10	8	4	8.3	2.8	0.84	33	1	--	--	--	--
11	8	4	8.3	2.3	0.68	15	0.4	35	0.6	--	--
12	8	4	8.3	2.5	0.84	15	1	--	--	--	--
13	6	4	7.3	4.7	0.98	6	0.5	35	0.5	--	--
14	6	4	7.3	3.1	0.81	8	0.3	20	0.4	35	0.3
15	8	4	8.3	1.7	0.50	15	1	--	--	--	--
16	8	4	8.3	2.4	0.70	15	1	--	--	--	--
17	8	4	8.3	4.3	0.87	15	0.4	35	0.6	--	--
18	6	4	7.3	3.9	0.80	6	0.7	35	0.3	--	--
19	6	4	7.3	3.1	0.72	8	0.3	20	0.4	35	0.3
20	9	4	6.5	2.9	0.71	8	0.3	20	0.4	35	0.3
21	8	4	8.3	2.2	0.60	15	0.4	35	0.6	--	--
22	4	4	7.3	2.2	0.61	15	1	--	--	--	--
23	8	4	8.3	3.6	0.83	33	1	--	--	--	--
24	6	4	7.3	3.6	0.77	6	0.7	35	0.3	--	--
25	6	4	7.3	3.3	0.85	8	0.3	20	0.4	35	0.3
26	8	4	8.3	2.2	0.60	6	0.7	35	0.3	--	--
27	6	4	7.3	3.0	0.68	6	0.5	35	0.5	--	--
Source. No	Super Zone	Mmin	Mmax	a	b	Hypocentral depth distribution					
						D1*	W1*	D2*	W2*	D3*	W3*
29	6	4	7.3	3.1	0.84	6	0.5	35	0.5	--	--

30	4	4	7.3	2.3	0.60	15	1	--	--	--	--
31	8	4	8.3	2.5	0.84	15	0.4	35	0.6	--	--
32	8	4	8.3	3.1	0.90	15	0.4	35	0.6	--	--
33	8	4	8.3	3.4	0.80	15	0.4	35	0.6	--	--
34	8	4	8.3	3.4	0.82	15	0.4	35	0.6	--	--
35	8	4	8.3	2.2	0.70	15	0.4	35	0.6	--	--
36	7	4	7.5	3.6	0.95	6	0.7	35	0.3	--	--
37	7	4	7.5	3.0	0.73	6	0.7	35	0.3	--	--
38	10	4	6.5	1.2	0.72	15	1	--	--	--	--
39	10	4	6.5	2.0	0.72	15	1	--	--	--	--
40	10	4	6.5	2.5	0.72	15	1	--	--	--	--
41	10	4	6.5	2.1	0.72	15	0.4	35	0.6	--	--
42	10	4	6.5	4.1	1.03	15	0.4	35	0.6	--	--
43	5	4	7.3	3.9	0.92	15	1	--	--	--	--
44	5	4	7.3	3.5	1.06	15	1	--	--	--	--
45	2	4	6	3.1	1.00	15	1	--	--	--	--
46	1	4	7.3	4.1	0.89	10	0.5	35	0.25	50	0.25
47	1	4	7.3	3.4	0.72	10	0.5	35	0.25	50	0.25
48	1	4	7.3	1.8	0.88	10	0.5	35	0.25	50	0.25
49	1	4	7.3	3.8	0.88	10	0.5	35	0.25	50	0.25
50	2	4	6	2.3	1.00	15	1	--	--	--	--
51	1	4	7.3	3.1	0.79	15	1	--	--	--	--
52	3	4	6	3.0	1.00	15	1	--	--	--	--
53	7	4	7.5	2.0	0.72	15	1	--	--	--	--
54	6	4	7.3	3.3	0.90	8	0.3	20	0.4	35	0.3
55	6	4	7.3	3.5	0.84	8	0.3	20	0.4	35	0.3
56	6	4	7.3	3.2	0.79	6	0.5	35	0.5	--	--
57	6	4	7.3	4.3	1.00	8	0.3	20	0.4	35	0.3
58	6	4	7.3	3.2	0.77	8	0.3	20	0.4	35	0.3
59	7	4	7.5	2.0	0.72	6	0.7	35	0.3	--	--
60	8	4	8.3	3.0	0.83	15	0.4	35	0.6	--	--
61	8	4	8.3	3.0	0.83	15	0.4	35	0.6	--	--
62	8	4	8.3	1.0	0.83	15	0.4	35	0.6	--	--
63	4	4	7.3	1.0	0.60	15	1	--	--	--	--

D1*=Hyopcentral depth, W1*=Weight

6.3 Intensity Prediction Equation

The seismic building code is based on macroseismic intensity in Central Asia (Ulomov, 1999). Therefore, an intensity prediction equation (IPE) is used in this study to find the intensity magnitude-distance relationships in terms of MSK-64 (Chapter 2). The choice of an attenuation model is very important for the hazard assessment because of its great influence on the final results. A modified version of the Bindi et al. (2011) IPE for Central Asia is used in this work. This modification is made to consider depth dependency of events in the IPE.

The same dataset of Bindi et al. (2011) is used for this modification. In its mathematical form, the IPE derived in this study can be written as:

$$I = 1.007 * M - 2.004 * \log10(h) + 3.298 - 2.692 * 0.5 * \log10\left(\frac{R^2}{h^2} + 1\right) - 0.000423 * \left(\sqrt{R^2 + h^2} - h\right) \quad (6.16)$$

where the intensity I is expressed in the MSK-64 scale, R is the epicentral distance and h is the hypocentral depth. The total standard deviation (which accounts for the dispersion in the dataset) of equation (6.16) is 0.69. As intensity is a normal random variable compared to any strong ground motion parameter in GMPE, which are lognormal, it is made sure in OpenQuake software to deal with intensity as normal random variable. The IPE is represented graphically by Figure (6.19) for different magnitudes and hypocentral depths. From this figure, it is clear that for epicentral distances less than the hypocentral depths, low hypocentral depth events give higher intensities compared to deeper hypocentral events. For larger epicentral distances compared to the hypocentral depths, deeper events give slightly higher intensity values compared to shallow events.

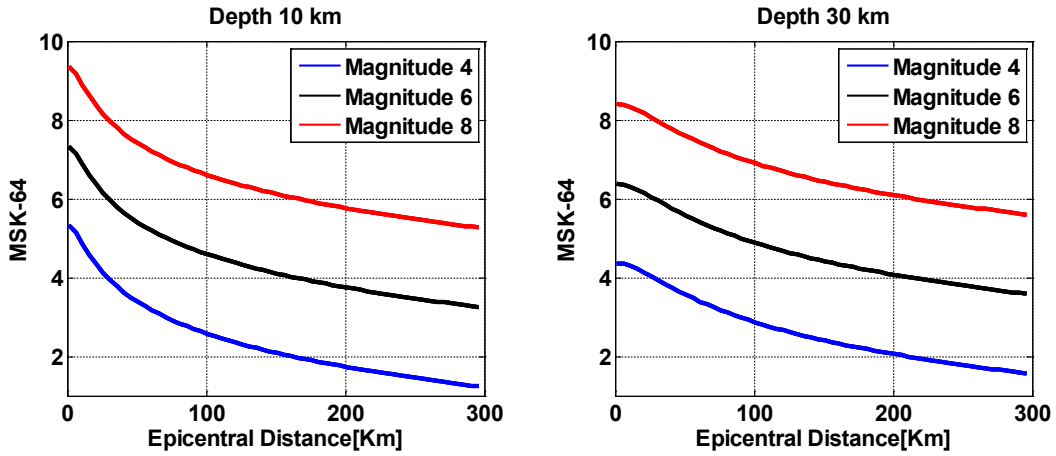


Figure 6.19 Graphical representation of the IPE used in this study for different magnitudes versus epicentral distances. The left side presents the relationship for a 10 km hypocentral depth event while the right side presents the relationship for an event with a 30 km hypocentral depth.

7 Results and Discussion of Seismic Hazard at Regional Scale

The results of seismic hazard at regional scale are summarized below for the different approaches.

7.1 Smooth Seismicity Results

Results from the smoothed seismicity approaches are discussed in detail in the next section. The fixed kernel and adaptive kernel approaches are discussed together, whereas the results of the Woo approach are presented separately.

7.1.1 *Results from Frankel Approach*

Figure 7.1 shows the incremental activity rates calculated using the Frankel (1995) approach, both for fixed and adaptive kernel bandwidths with regional (uniform) and gridded (non-uniform) *b* values. Gridded *b* values are considered from the super zones of the area source model (Table 6.2). Areas characterized by high seismicity show high activity rates, such as the Alai valley (running east-west across most of south-west Kyrgyzstan), the Fergana valley across the eastern part of Uzbekistan, Kyrgyzstan and Tajikistan, the boundary between the north-west Tarim basin in north-west China and the southern Tien Shan, and the Kopet-Dagh range (Turkmen-Khorasan Mountain Range) on the border between Turkmenistan and Iran (Figure 5.1 and 5.2). Clear differences occur between a fixed and an adaptive bandwidth for uniform *b* values (Figures 7.1(a) and (b)), especially in low and high seismicity areas. Since for a fixed bandwidth the same level of smoothing is applied everywhere, the seismicity is further smeared from the areas with a high rate of seismicity. On the other hand, using an adaptive bandwidth, the degree of smoothing depends on the seismicity, i.e., in this case, the level of smoothing is lower in areas with a high level of seismicity (e.g., the Alai valley), whereas there is stronger smoothing in areas of lower seismicity (e.g., the Tajik-Afghan block, Uzbekistan). Similar observations can be made in the case of gridded *b* values (Figures 7.1(c) and (d)). As can be seen, the activity rates are largely influenced by the *b* value (e.g., the Gazli field in central Uzbekistan as well as for western Tajikistan).

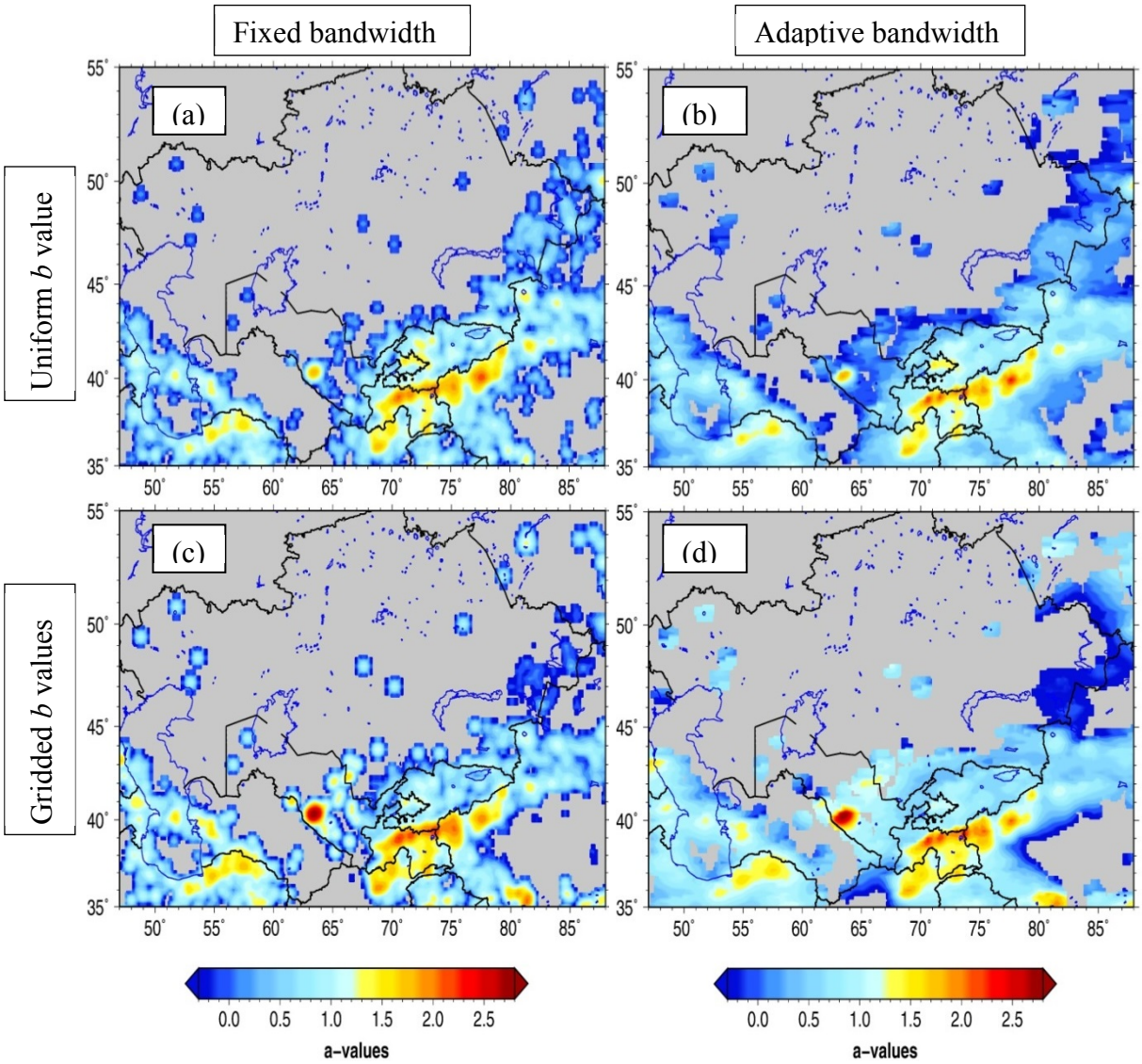


Figure 7.1 Incremental activity rates calculated using the Frankel approach. (a) Fixed bandwidth, regional b value, (b) adaptive bandwidth, regional b value, (c) fixed bandwidth, gridded b value, and (d) adaptive bandwidth gridded b value.

In Figure 7.2, the probabilistic seismic hazard for Central Asia in terms of macro-seismic intensities for 10% probability of exceedance in 50 years based on the approach of Frankel (1995) is given. Using uniform *b* value for fixed and adaptive bandwidth, the highest level of hazard is observed in the Alai valley, the Fergana valley and in the Gazli gas fields, reaching an intensity level of up to 7-8 (Figures 7.2 (a) and (b)). On the other hand, using gridded *b* values (Figures 7.2 (c) and (d)), the level of hazard is highest in the Tajik-Uzbekistan border region, reaching a level of 8-9. This high level of hazard is due to the presence of a few large magnitude events (e.g., the 1907 M=7.4 Qaratog earthquake), thus emphasizing low *b* values.

For an adaptive bandwidth, the level of hazard is smoother in the aforementioned areas and less spread around the Fergana valley, where very large magnitude events have never been recorded. The major differences between gridded *b* value with adaptive kernel and uniform *b* values and fixed kernel can also be seen in the southern part of Turkmenistan. The gridded *b* value with adaptive kernel gives much larger hazard where the *b* value is playing a major role.

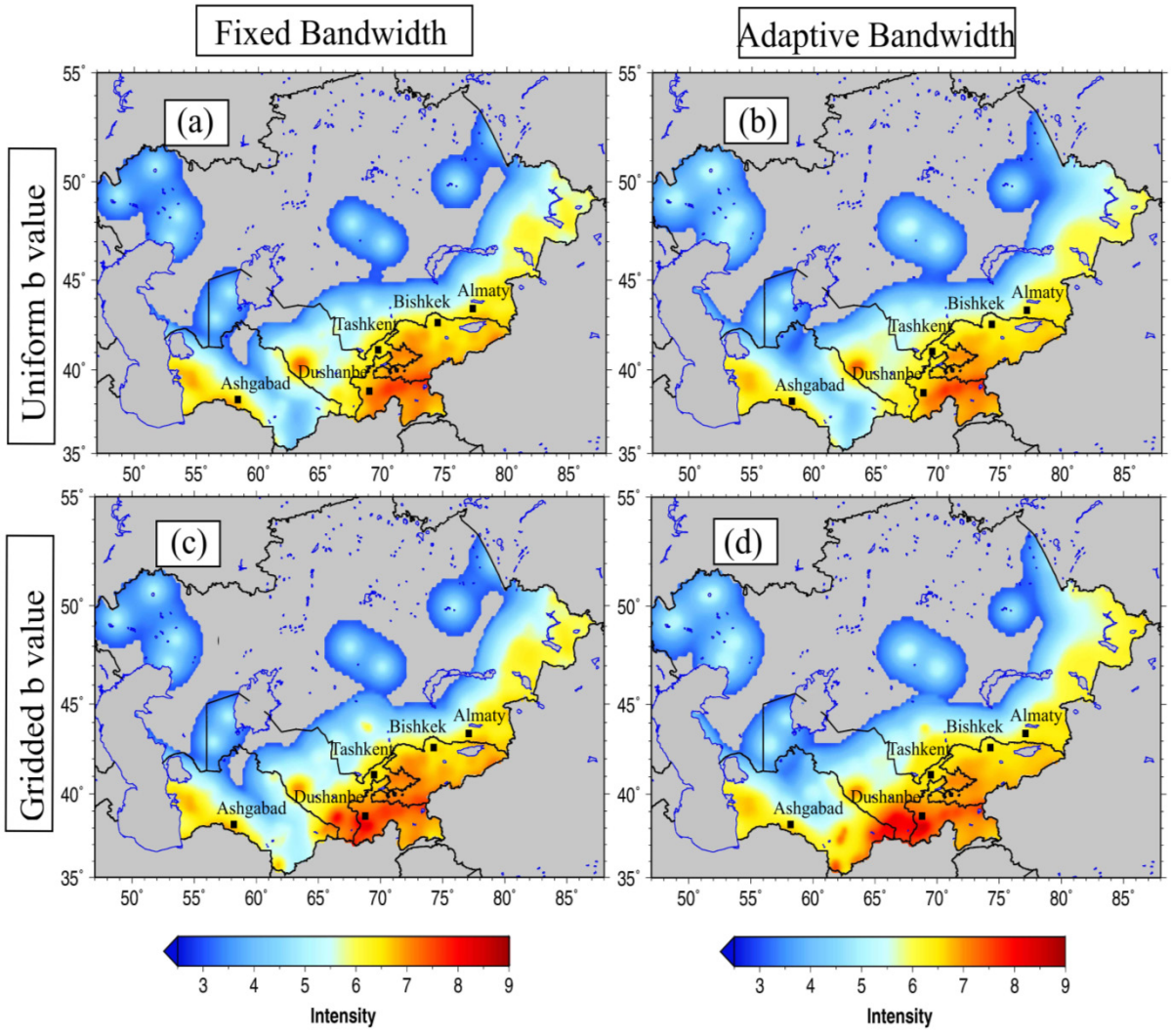


Figure 7.2 Probabilistic seismic hazard estimated for Central Asia in terms of macro-seismic intensities for 10% probability of exceedance in 50 years from the Frankel (1995) approach. (a) Fixed bandwidth, regional b value (b) adaptive bandwidth, regional b value, (c) fixed bandwidth, gridded b value, and (d) adaptive bandwidth, gridded b value.

7.1.2 *Results from Woo Approach*

Figure 7.3 shows the level of seismic hazard assessed using magnitude-dependent bandwidths for the smoothing procedure of Woo (1996). Compared to the results shown in Figure 7.2, areas of higher hazard levels are much more smeared, e.g., in the northern Tajikistan near the Issyk Kul Lake and adjacent to Almaty (Kazakhstan). This effect is also seen in Turkmenistan near Ashgabad. This is due to a fixed smoothing distance which depends on the event´s magnitude. These regions are dominated by strong magnitude events such as the Kemin earthquake (M 8.2, 1911) and Chilik earthquake (M 8.3, 1889) near Almaty, and the Ashgabad earthquake of M 7.3 in 1948 in the Kopet-Dagh region of Turkmenistan. Hence, high levels of intensity are assessed for the Issyk-Kul region in eastern Kyrgyzstan, the Tien Shan and Pamir mountain ranges, for large parts around the Gazli gas field in central Uzbekistan, and around Ashgabad, Turkmenistan, where the highest magnitude events affect the hazard estimates over large distances.

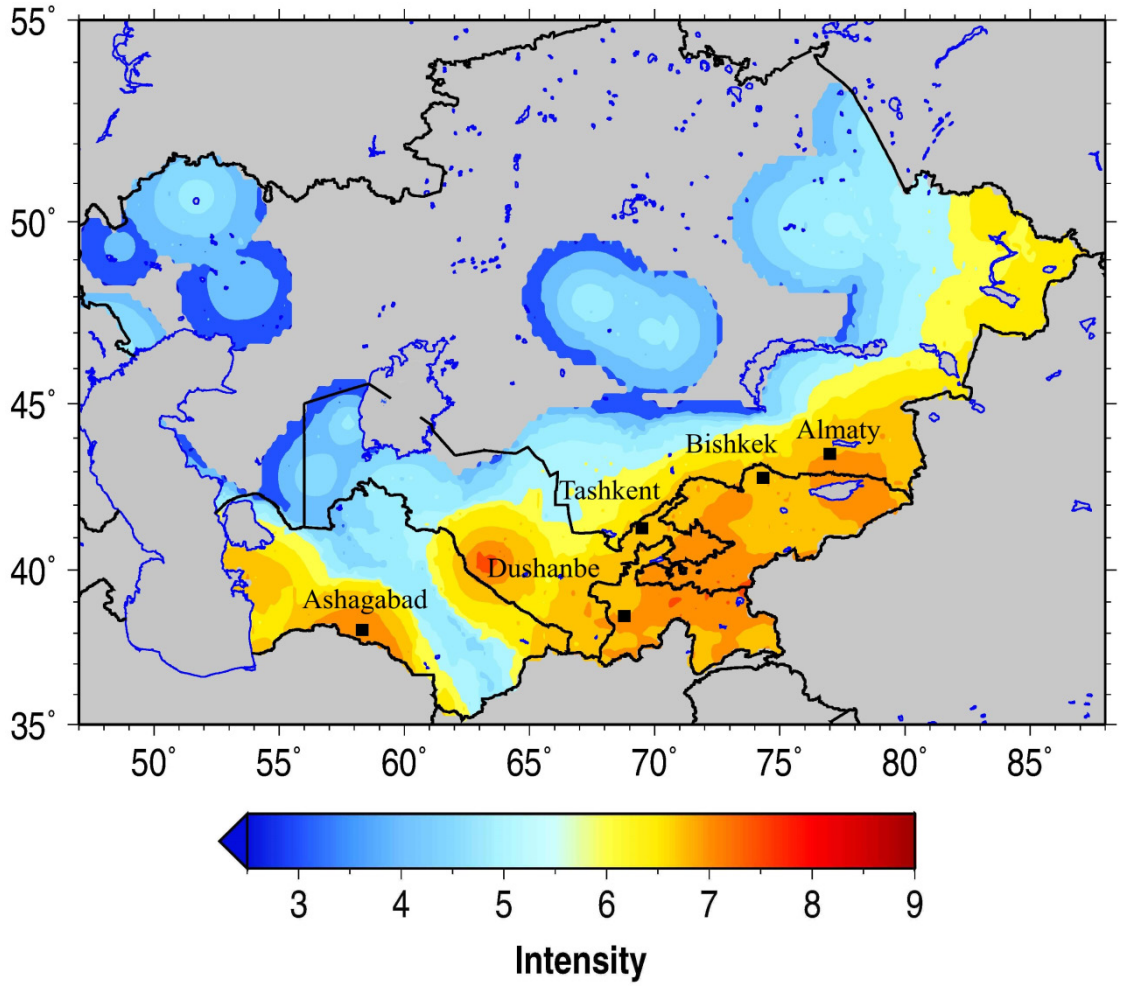


Figure 7.3 Probabilistic seismic hazard estimated for Central Asia in terms of macro-seismic intensities for 10% probability of exceedance in 50 years using the Woo (1996) approach.

7.2 Results from Area sources

Figure 7.4(a) shows the macroseismic intensity distribution obtained by considering the area source model for a probability of 10% to be exceeded, i.e., a return period of 475 years. In this case, the influence of considering the seismic sources on the estimated level of hazard can be clearly identified. As expected, areas with a high level of seismicity and with lower b-values show a significantly higher level of hazard (e.g., the Kyrgyz-Kazakh border region north of Issyk-Kul, the Alai valley around the Kyrgyz-Tajik border, the northern part of the Fergana valley). Furthermore, note that for large areas of the stable region of central Kazakhstan, intensity values of around 3 have been derived, fixing a background level corresponding to areas not, or only slightly, affected by seismicity.

Using a 2% probability of exceedance in 50 years, which corresponds to a return period of 2475 years, the maximum level of hazard reaches intensities up to 9 in the Kyrgyz-Kazakh border region (Figure 7.4 (b)). All of Kyrgyzstan and Tajikistan, as well as the eastern parts of Uzbekistan, are characterized by intensity levels higher than 7. For the stable region of central Kazakhstan, a macroseismic intensity of up to 4 is estimated.

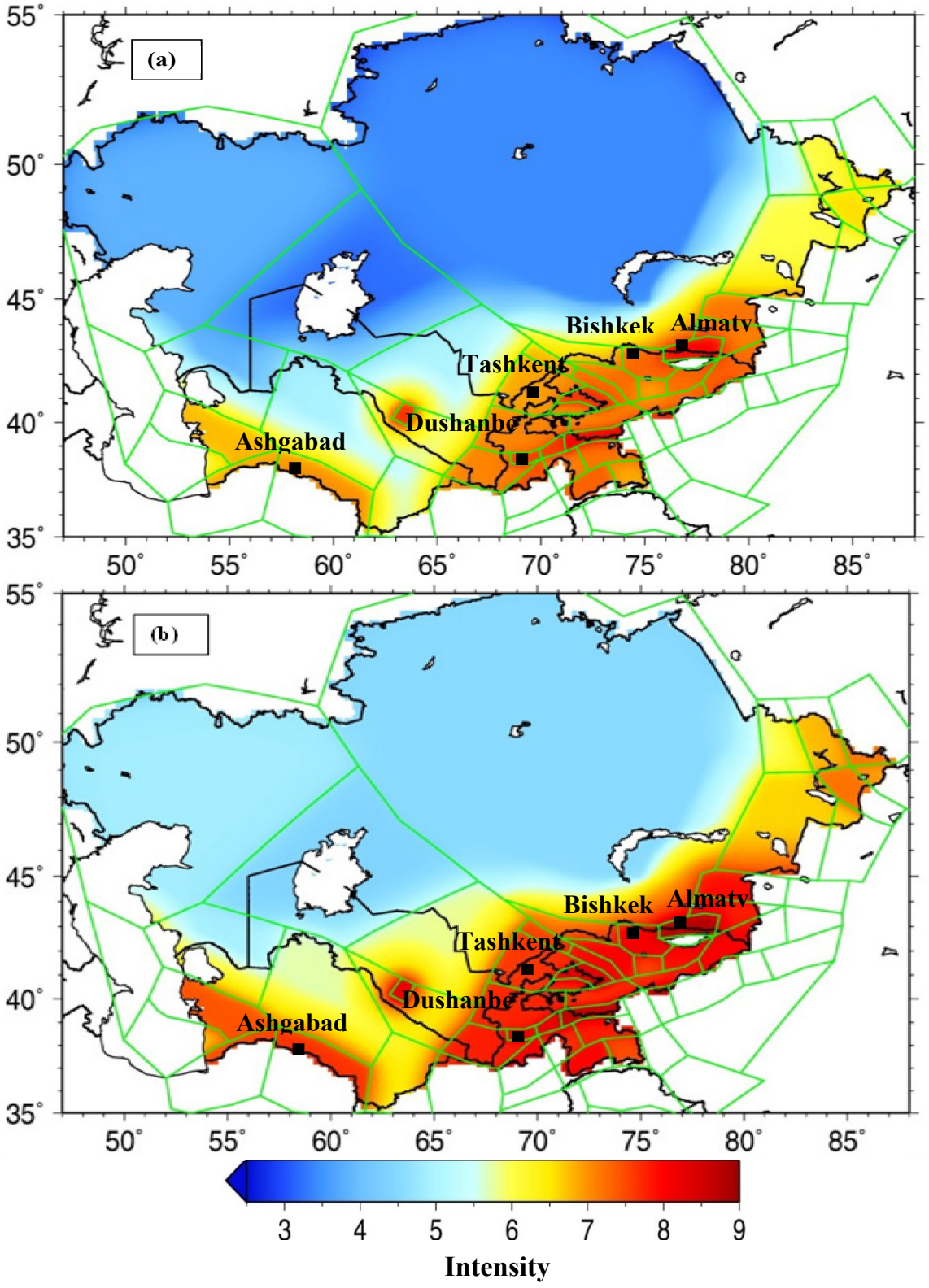


Figure 7.4 Probabilistic seismic hazard in terms of macro-seismic intensities for a (a) 10% probability and (b) 2% probability of exceedance in 50 years using the area source model. The green lines represent the area sources.

7.3 Hazard curves for selected cities

To emphasize the level of hazard for the most important cities in Central Asia, Figure 7.5 compares the level of hazard for a 10% and a 2% probability of exceedance in 50 years, derived using the smoothed seismicity approaches and area source models. In the case of the Frankel approach, the results are shown for gridded b values and an adaptive kernel.

For the Kyrgyz capital, Bishkek, the Frankel (1995) approach provides the lowest level of hazard with an intensity of 6-7 having 10% probability of exceedance in 50 years. Although the curves for the area source model and the Woo (1996) approach provide similar hazard values, only for Bishkek does the Woo (1996) approach provide the highest intensity level.

Similar conditions hold for the city of Almaty, for which the method of Frankel (1995) provides the lowest level of hazard, with intensities of around 6 and 7 for return periods of 475 and 2475 years, respectively. However, the three curves are much more spread apart. The highest hazard levels are assessed when considering the area source model, leading to an intensity level of 8 for 10% probability of exceedance in 50 years.

For Dushanbe, the agreement between the area sources curve and the curve of the Frankel (1995) approach is noticeable. Whereas the area source model shows a high probability of exceedance for low intensity values, the curve decreases more steeply for higher intensity values, approaching the curve calculated by the Woo (1996) approach. For all intensity levels, the Woo (1996) method provides the lowest level of hazard, both for low and high probabilities of exceedance.

For the Uzbek capital Tashkent, all three curves are rather congruent. All in all, the intensity levels are lower here than for the other cities, with values of around 7 for a 10% probability of exceedance in 50 years and only around 7-8 for a 2% probability of exceedance compared to, for example, 8 in Almaty.

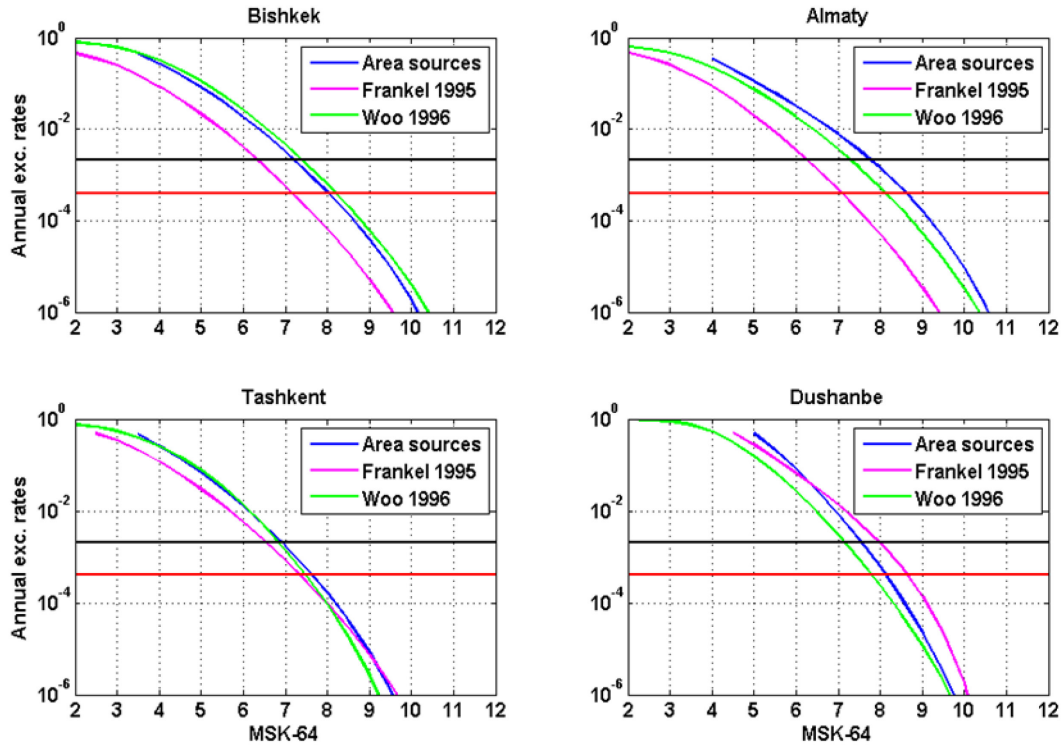


Figure 7.5 Hazard curves for the major cities of Central Asia using the smoothed seismicity approaches and area source model. In the case of the Frankel (1995) approach, the results are shown for gridded b values and adaptive kernels. The black and red lines correspond to a 10% and 2% probability of exceedance in 50 years, respectively.

7.4 Discussion and comparison with earlier studies at regional level

In principle, due to the shortness of the earthquake catalogue, and the lack of geological and tectonic knowledge, neither of the considered approaches can be considered as providing more realistic results than the other. They are therefore simply different representations of several, but well established approaches, requiring different degrees of expert judgment and/or trust in the completeness of the available information, with their own drawbacks and advantages. However, the presented results confirm that the design and geometry of the seismogenic zones for the area source model (Figure 6.13) has a major influence on the distribution of the hazard, which is strongly concentrated inside the boundaries of the seismogenic zones (see Figure 7.4). As reflected by historical seismicity, such a spatial clustering might also be correlated to a marked temporal clustering. Such observations have already been revealed by statistical analyses both at the regional (Faenza et al., 2003) and local scales (Gómez and Pacheco, 2004). In any case, the biases in most cases cannot be easily corrected by simply altering some boundaries, meaning it is clear that a different

seismotectonic zonation could be proposed and checked by using different approaches (e.g., Musson, 2006). However, one should be aware that any reduction of the sizes of the seismogenic zones also reduces the amount of data available for the statistical parameterization of seismicity in each zone. Consequently, this under sampling may adversely affect the hazard estimate. Whereas Algermissen et al. (1982) suggested a lower limit of 40 events in each zone to avoid this problem, the work by Bender (1983) provided some estimates that, with less than 25 events in each zone, the relative error on b values could exceed 25%. Therefore, to be on the safe side, for our study we adopted a regional super zone b value if the number of events is below the threshold of 20 events. As pointed out by Wiemer et al. (2009), this addresses a long-standing need in hazard assessment, and it stabilizes the resulting model by avoiding large fluctuations in b values which is evident, e.g., in super zone 8 (see Figure 6.13 and Table 6.3 and the resulting Figure 7.4).

In zone-free approaches, the predicted areas of high intensities are subject to the b value calculation scheme. The gridded seismicity model emphasizes the expectation that future large, damaging earthquakes are more likely to occur at sites near the epicenters of previous moderate earthquakes, i.e., in areas of lower b values corresponding to the southern part of this study area where large magnitude seismicity is located. On the other hand, the use of a uniform b value for fixed and adaptive bandwidths smooth's out this effect and significantly lowers the levels of hazard in these areas. The highest hazard levels can be found in zones subjected to a large number of past earthquakes, as reported in the earthquake catalogue (e.g., the Tajik-Kyrgyz border region, south-western Tajikistan), thus revealing a different spatial distribution and pattern of hazard estimates. Very little influence of large magnitude events that occurred in the northern part of the study area can be identified on the final hazard estimates (Figure 7.3).

When comparing the results, the zoning and smoothing approaches can be seen to yield similar hazard estimates for low to moderate seismicity regions. This means that the results are directly related to the seismicity models used, but their comparison can lead to very different scenarios, depending on the site being studied. Such disaggregation results strictly depend on the seismicity distribution, that is, on the proximity to sources where the seismicity is mainly concentrated. Distant, larger magnitude events dominate the hazard in areas characterized by low-seismic activity. Conversely, nearby seismicity is often the major contributor to the hazard at sites located in high-seismicity regions or in zones where the

seismic activity is characterized by weak-to-moderate, but quite frequent, events. In line with the findings of Molina et al. (2001) and Beauval et al. (2006), the fixed kernel method might yield compatible results when compared to the classical zonation approach, whereas for high seismicity regions, using a fixed bandwidth yields lower hazard values when compared to those from the zonation method.

This is particularly obvious for the most important cities in Central Asia (Figure 7.5) for which – with the exception of Dushanbe – the area source model provides a higher level of hazard. Whereas the zoning method requires homogeneous source zones and a frequency-magnitude distribution for each source zone, Woo (1996) proposed to use maps of smoothed epicentre locations, which are smoothed according to the fractal distribution of earthquakes in space (e.g., Kagan and Jackson, 2000). Although originally the method of Woo (1996) did not allow for the occurrence of magnitudes greater than the maximum magnitude observed, we have accounted for earthquakes larger than historical ones by considering an uncertainty of 0.5 in magnitude for the historical events of the catalogue (before 1960) and 0.2 in magnitude for more recent events, increasing the level of hazard.

Although the results of the latest study on the region, GSHAP (Giardini et al., 1999), who published a homogenized seismic hazard map in terms of peak ground acceleration at a 10% probability of exceedance in 50 years, have been a major step towards a cross-border approach for the entire region, the conducted PSHA, employing a rather homogeneous distribution in terms of PGA for large parts of Central Asia, might be assumed to be outdated for several reasons (Figure 7.6). This includes the fact that much progress has been made in the development of region-specific ground motion prediction equations (e.g., Bindi et al., 2011), the treatment of uncertainties within a PSHA has been improved, and new model ideas have been developed. Figure 7.6 shows the seismic hazard calculated in the GSHAP project in terms of 10% probability of exceedance in 50 years, converted from PGA to macroseismic intensity, using the same relationship (equation 7.1, Apitkaev and Shebalin (1988)) that was used for conversion from intensity to PGA in Ulomov (1999), for a comparison with the results in this study.

$$\boldsymbol{log_{10} pga(m/s^2) = 0.333\, I_{MSK} - 2.222} \qquad (7.1)$$

Comparing Figure 7.6 with the current study figure 7.4(a) shows the hazard is generally higher by 2 intensity degrees when compared to the GSHAP project results.

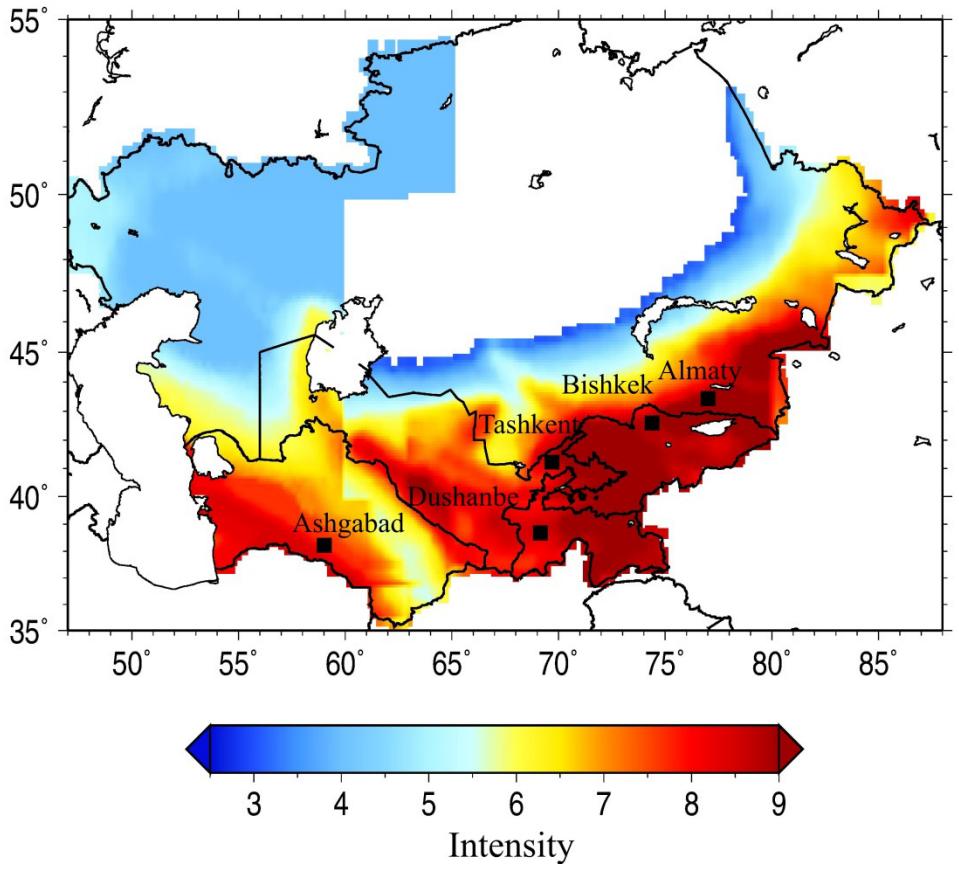


Figure 7.6 Results of PSHA for Central Asia in terms of macro-seismic intensities for 10% probability of exceedance in 50 years from GSHAP project.

Figure 7.7 shows the hazard curves (from Bindi et al. 2012) for different probabilities of exceedance in 50 years for some of the important cities in Central Asia using the site approach. Considering the results from the site approach shows a higher level of seismic hazard compared to the values from this study. Figure 7.7 shows seismic hazard of intensities 9, 9, 9 and 10 for Bishkek, Dushanbe, Tashkent and Almaty, respectively, for 10% probability of exceedance in 50 years. These estimates of hazard from the site approach considered both the observed and virtual intensity data. The maximum hazard estimates for the above cities from this study correspond to 7-8, 8, 7, and 8 for Bishkek, Dushanbe, Tashkent, and Almaty, respectively.

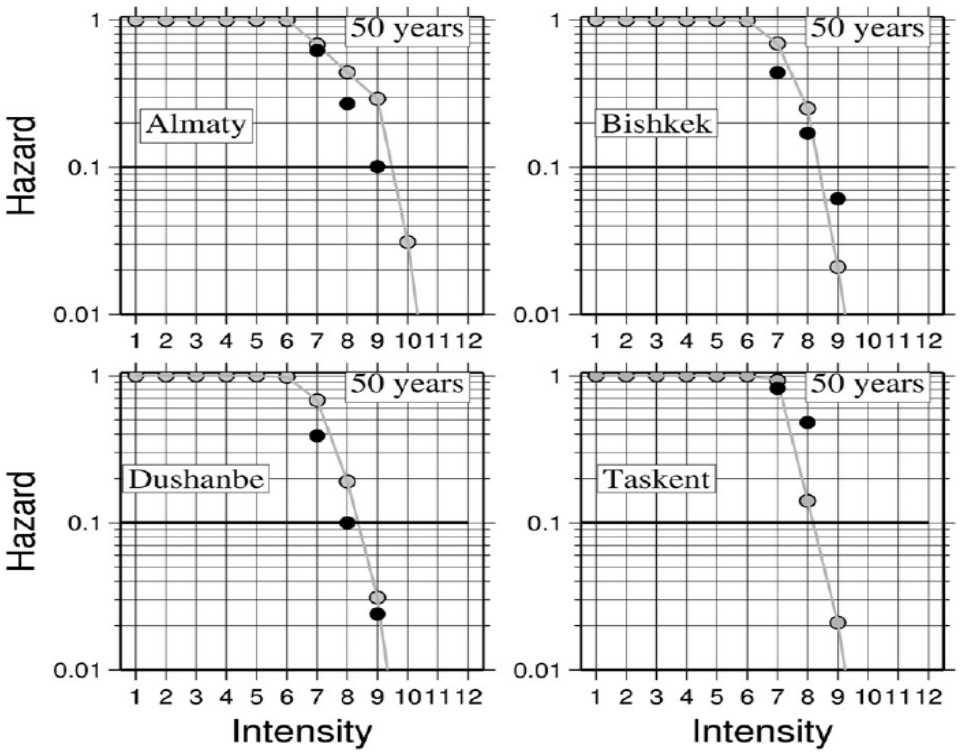


Figure 7.7 Hazard curves (grey symbols and lines) computed for major cities in Central Asia using the site approach. Black circles show the estimates of Negmatullaev et al. (1999) (Bindi et al. 2012).

While the main activity centres around the Fergana and Alai valleys remain dominant in the hazard assessments, the maps presented here are more diverse across the region. Greater discrimination is found, especially along the border between Turkmenistan and Uzbekistan, central Kyrgyzstan and eastern Tajikistan. In particular, such large regions as Central Asia are characterized by rather different seismic regimes and regionally varying features and degrees of catalogue completeness, all of which need to be taken into account for a proper seismic hazard assessment on a regional scale.

On the other hand, the results presented for the most important cities, i.e., the results for the local scale, have to be seen as part of a regional large-scale study. For site-specific investigations of engineering interest, local conditions, especially concerning the local geological conditions, the regionalization, and more detailed attenuation relationships, need to be considered. In the end, none of the models can be seen as the final result since differences still arise when based on the same data set. A zone-free approach, which does not consider expert judgment as much, might be seen as the most objective model. However, it is doubtful if the "objectivity" of such techniques that avoid regionalization can compensate for the deliberate omission of distinct geological and seismological knowledge. The weight to be

assigned at the results obtained with the two methods should be compatible with the quality and extension of historical catalogues, with the observed instrumental seismicity, and with the geological and tectonic information available. Since the final goal is to maximize the reliability of hazard estimates by making the best use of all accessible data, this capability might be best exploited to reduce uncertainty in computing seismic hazard in Central Asia, with their incomplete catalogues and where seismogenic structures are generally not easily detectable.

8 IMPROVING THE SPATIAL RESOLUTION OF GROUND MOTION VARIABILITY IN BISHKEK, KYRGYZSTAN

In the previous chapter the seismic hazard is estimated on a regional level, which does not consider site effects explicitly. However, as explained in chapter 3, site effects play an important role in seismic hazard and risk assessment, and in defining the optimal engineering design for civil structures. Nonetheless, due to the increasing trend of urbanization, target areas are often too vast to be covered by standard approaches. In this chapter, a new method is proposed to improve the spatial resolution of ground motion variability in the PSHA. The approach uses earthquakes recorded at a few selected sites for a relatively short amount of time, and seismic noise data collected over a denser grid, exploiting Standard Spectral Ratios (SSR). The method is applied to Bishkek, Kyrgyzstan.

8.1 Introduction

Bishkek, which only developed as a city during the twentieth century, is the capital and main cultural, industrial and financial centre of the Kyrgyz Republic, with a population greater than 900,000 inhabitants. The country has witnessed several moderate to large earthquakes in the past. The most important include the August 3, 1885 M 7 Belovodski and January 3, 1911 M 8.2 Kemin events (Kalmetieva et al. 2009). These earthquakes generated in Bishkek macroseismic intensities as large as VII-VIII and VI MSK, respectively (Januzakov et al. 2003). Being located in a highly seismically active region and experiencing increasing levels of urbanization, Bishkek is faced with a high seismic risk. In fact, according to Erdik et al. (2005), based on night-time occupancy and considering a 2% probability of exceedance in 50 years, the city could potentially experience an event that may cause 34,000 deaths with 90,000 injured persons who would need to be treated at hospitals. Furthermore, recent studies into possible worst case scenarios (Bindi et al., 2011) showed that for a magnitude 7.5 earthquake at the Issyk-Ata fault, located ~10 km from the city, over 48000 buildings are expected to be damaged with more than 22000 collapsing, leading to 16624 casualties and 93447 injuries. Bindi et al. (2011) also outlined the importance of site amplification in determining the distribution of damages within the city. However, the authors could only use site amplification functions from 18 sites within the urban area, where seismic stations were

deployed during a seismological experiment, while higher spatial resolution scenarios (including a high spatial resolution of building variability) require a much larger number of sites to gain adequate knowledge of the variability of ground motion. In this chapter, the resolution of ground motion variability is improved by combing the results of earthquake data and seismic noise recordings.

As described in chapter 3, seismic noise recordings have recently become an important tool for site effects estimation (e.g., Bard, 1999; Parolai et al., 2001; D'amico et al., 2004), using both single stations and arrays to retrieve information about the fundamental resonance frequency of a site and the S-wave structure of the uppermost layers. Picozzi et al. (2009) and Parolai et al. (2010), for example, used single station noise measurements to infer resonance frequency maps of Istanbul, Turkey, and Bishkek, Kyrgyzstan, respectively. Parolai et al. (2005) and Boxberger et al. (2011) used inversions of surface waves from seismic noise array measurements to characterize the shallow geology in Cologne (Germany) and Bevagna (Italy), respectively, which in turn can be used to estimate the site response.

In several studies of urban site effects (Parolai et al., 2004; Pilz et al., 2009), temporary arrays of a few tens of stations have been used to record weak motion events for assessing the site response. Combining such results with those derived by seismic noise analysis (mainly Horizontal to Vertical Spectral Ratios, HVSRs), it is possible to obtain a better picture of ground motion variability over an investigated area. However, a quantitative analysis that allows a link between the site response estimated from earthquake recordings and the HVSR was not carried out and the possibility of exploiting this correlation as a tool to improve the knowledge about the spatial variability of ground motion was not investigated.

Different methods of clustering analysis have been recently applied, either to seismic noise or to earthquake HVSR, to identify areas presenting similar HVSR characteristic and likely site amplification. Bragato et al. (2007), following the approach of Rodriguez and Midorikawa (2002), clustered homogeneous areas based on the similarity of seismic noise HVSRs, following a Bayesian approach. Tselentis et al. (2011) used the Self-Organizing Map (SOM) technique for cluster analysis, based on the similarity of earthquake HVSR. In both studies, clustering was identified, but since neither the HVSR of the noise, nor of earthquakes represents the true site response (Field and Jacob, 1995; Bindi et al., 2000; Bard, 2004; Parolai et al., 2004), it is not possible to assign to a site an appropriate amplification function

that could then be used when calculating ground motion and risk scenarios. Recently, Strollo et al. (2012) used correlation analysis between HVSRs of Single station Seismic Noise measurement Points (SSNP) and seismic noise of a few stations for which spectral intensity ratios (SIRs) had been calculated using earthquake records to spatially extend the SIRs to the large number of SSNPs.

8.2 Geology of Bishkek

Bishkek is located in the north-western part of the Tien Shan Mountains. The region is known for its high level of seismicity, with the occurrence of large earthquakes within the context of a continent to continent collision (Bullen et al, 2001; Kalmetieva et al., 2009). This seismicity is associated with a complex faulting system resulting from the collision of the Indian sub-continent with Eurasia. The city of Bishkek lies over the central part of the Chu Basin, one of the largest depressions of the northern Tien Shan. The basin extends about 50 km north-south and 150 km east-west (Bullen et al., 2001). The basin is bounded by the Kyrgyz Range to the south and the Chu-Ili Mountains in the north. Much of the Kyrgyz Range is made up of Ordovician granite and fine grained Devonian sedimentary rocks. It is cut by many short, irregular imbricate thrust faults that are laterally truncated by strike-slip faults. The Paleozoic bedrock is thrust northward over ~ 4 km of younger Cenozoic molasses at the margin of the Chu basin. This deformation front is stepped northward into the Chu basin which exposes a thick succession of Neogene strata along the Issyk-Ata fault, located ~10 km north of the basin's boundary. The Paleozoic basement depth below the urban area of Bishkek generally decreases from the north (~1 km) to the south (~3 km). The following seven Tertiary formations overlay the Paleozoic bedrock (from bottom to top): (1) Kokturpak (siltstones), (2) Kokomeren (sandstones), (3) Sera Firma (claystones), (4) Dzhel´dysu (mudstones) and (5) Saryagach (sandstones and pebbly conglomerate), all belonging to the Paleogene, and (6) Chu (mudstones and sandstones) and, (7) Sharpyldak (coarse pebble conglomerate), belonging to the Neogene (Bullen et al. 2001). Figure 8.1 shows the geological map of the Chu basin and its surroundings (Bullen et al., 2001).

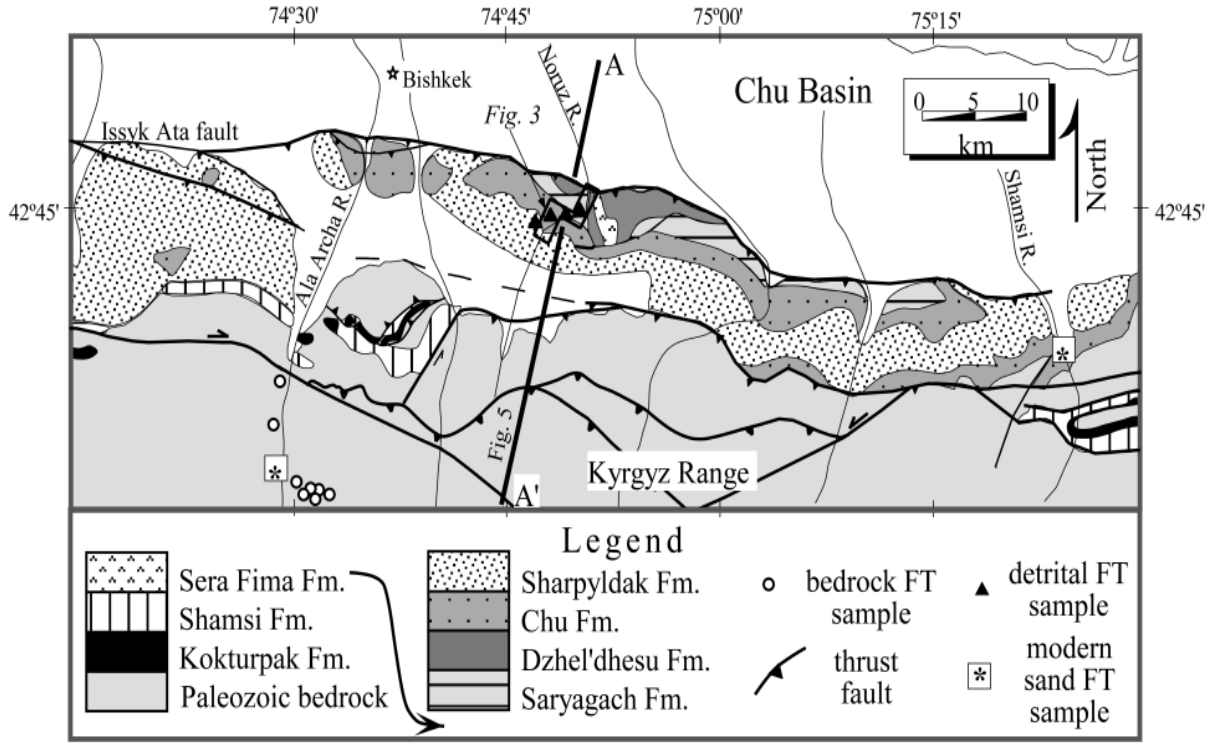


Figure 8.1 Geological map of the Chu basin and adjacent Kyrgyz Range (Bullen et al. 2001).

Quaternary sediments overlay the Tertiary deposits with a thickness of about 200-300 m. Figure 8.2 shows the presence of different types of Quaternary cover with some tertiary outcrops (i.e., Paleogene and Neogene). The geology of the study area in the south consists of alluvial material of rubble and gravel forming 15-40 m thick outcrops. The southern half of the Bishkek area, which consist of alluvial gravels, rubble and sandy materials with a total thickness of about 25-50 m, constitute the shallowest geological layers. The northern part of Bishkek consists of Quaternary sediments made up of rubbly-bench gravel with clay-sand lenses and break stone bench gravel outcrop. In the north, the alluvial material becomes finer with gravels alternating to silt layers. In the northernmost part of the city, eolic sediments (loess) with thickness reaching tens of meters are found and are occasionally cut by recent alluvial deposits (Parolai et al., 2014). Figure 8.3 shows the available large scale NS geological cross section of the study area, modified after Baeva (1999). The shape of the Tertiary and Paleozoic basement can be seen in this figure, with the Tertiary basement getting thinner in the north.

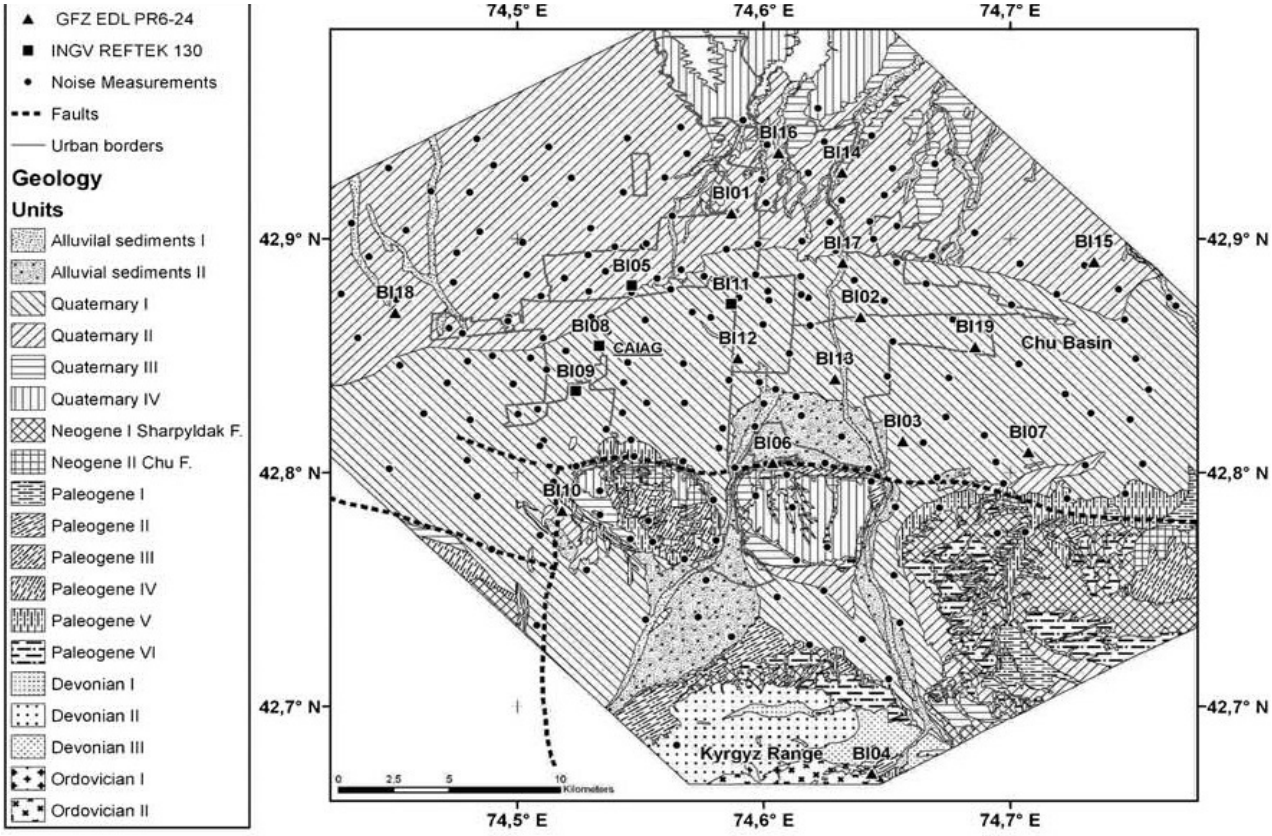


Figure 8.2 Geological map of Bishkek, Kyrgyzstan. Different symbols represent stations of the seismic network; Squares and triangles shows the temporary stations, while the black dots show the locations of the single station noise measurements as explained in the "Dataset and analysis" section.

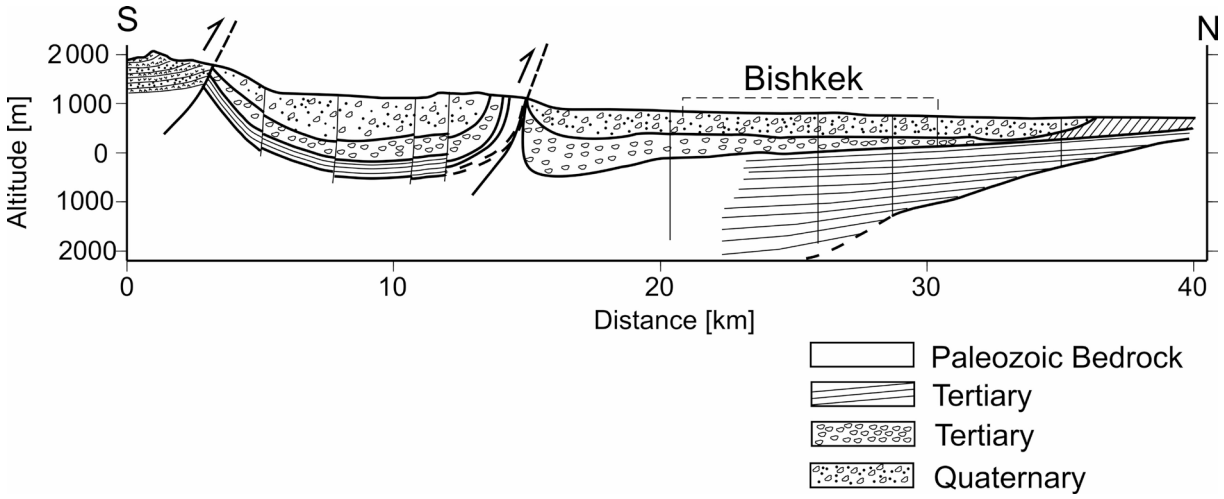


Figure 8.3 Geological cross-section of the Bishkek basin (redrawn after Baeva, 1999).

8.3 Data set and analyses

A temporary seismological network of 19 stations (Fig. 8.2), which operated in continuous mode, was installed in Bishkek from 16 August 2008 until the middle of November 2008 (Parolai et al., 2010). 15 stations used Earth Data Logger (EDL) 24-bit acquisition systems with 13 of them employing L4C-3D sensors having a 1 Hz corner frequency, one Güralp CMG-ESPC 60 and one IO-3D sensor with a corner frequency of 4.5 Hz. The other 4 stations used Reftek-130 digitizers with Le3D-5_s sensors. At all stations, data was recorded at 100 samples per second. The network recorded 56 events; including 50 crustal earthquakes with magnitudes between ML 1.6 and M_w 6.6, occurring at distances between 35 and 1527 km from station BI08 (see Fig.8.2). The dataset includes 5 deep earthquakes with magnitudes between mb 5.1 and M_w 6.8, which occurred in the Pamir-Hindu Kush region, and an event of magnitude M_w 7.3 which occurred at a distance of 5685 km in the Okhotsk Sea (Russia) at a depth of 492 km.

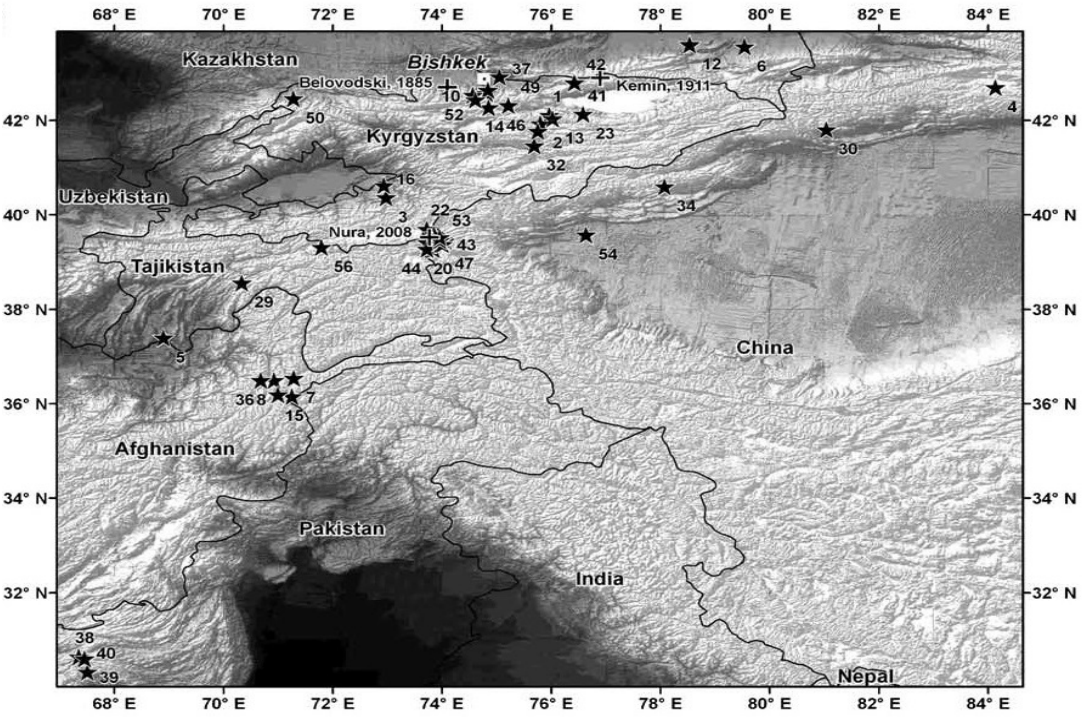


Figure 8.4 Epicentres (stars) of the earthquakes used for site response analysis. Crosses indicate the locations of historical earthquakes (Parolai et al., 2010).

The site effects are estimated at the stations of the temporary seismic network from earthquakes in term of SSR and earthquake H/V spectral ratios (HVSR) (chapter 3) (Parolai et al. 2010). For the calculation of noise HVSR at the location of the each station of temporary

network, the seismic noise before each earthquake (pre-event seismic noise) is used, employing the same length as that of the S-wave window for the calculation of SSR. Along with the network of 19 stations, single station seismic noise was also recorded in the urban area covering the whole city at nearly 200 points. The seismic noise was recorded using an EDL 24-bit digital acquisition system, equipped with a 1 Hz Mark L4C-3D sensor at 100 samples per second. Seismic noise was recorded for about 30 minutes at each point. The data was split into 60 s window lengths, and analysed for single station noise HVSR (Parolai et al., 2010).

8.4 Clustering Analysis

The clustering analysis (chapter 4) is performed over the results of the site response analysis described in the previous section. 15 out of 19 of the temporary seismic network stations are used in this study, since some stations had not recorded the events observed by the reference station BI04, from which SSRs and pre-event noise HVSR were calculated (see table 8.1). The analyses were carried out over a frequency range from 0.1-2 Hz for SSRs using logarithmically equal-spaced bins. The lower bound of the selected frequency range was chosen to allow for each site's fundamental frequency to be covered, and to be within the range where the used instruments work properly. The fundamental frequencies are expected to be low, owing to the sedimentary cover thickness in the area. The upper bound of the considered frequency range is selected by acknowledging that at higher frequencies, the HVSR of the earthquake data for the reference site (BI04) is showing site amplification (see Parolai et al. 2010). Although this band restriction toward higher frequencies might limit the ability to discriminate the shallow geological structural variations, it is shown later that the changes in the thin and shallow Quaternary layers can still be successfully captured by using the chosen frequency band. As is explained in the Discussion section, this might indicate that the shallow geological variations are linked to the deeper ones.

Table 8-1 Details of the temporary seismic network stations used for the clustering analysis.

Agency	Station Name	Latitude (N)	Longitude(E)	Digitizer	Sensor
GFZ	BI01	42.9109	74.5869	EDL	MARK-L4C-3D
GFZ	BI02	42.8665	74.6392	EDL	MARK-L4C-3D
GFZ	BI03	42.8133	74.6564	EDL	MARK-L4C-3D
GFZ	BI04	42.6712	74.6438	EDL	MARK-L4C-3D
GFZ	BI06	42.8040	74.6036	EDL	MARK-L4C-3D
GFZ	BI07	42.8087	74.7074	EDL	MARK-L4C-3D
INGV	BI08	42.8543	74.5332	Reftek-130	LENNARTZ-LE3D-5s
INGV	BI09	42.8350	74.5238	Reftek-130	LENNARTZ-LE3D-5s
GFZ	BI10	42.7836	74.5182	EDL	MARK-L4C-3D
INGV	BI11	42.8722	74.5868	Reftek-130	LENNARTZ-LE3D-5s
GFZ	BI13	42.8400	74.6289	EDL	MARK-L4C-3D
GFZ	BI14	42.9282	74.6317	EDL	MARK-L4C-3D
GFZ	BI15	42.8901	74.7341	EDL	MARK-L4C-3D
GFZ	BI16	42.9368	74.6061	EDL	IO-3D 4.5Hz
GFZ	BI18	42.8683	74.4505	EDL	MARK-L4C-3D
GFZ	BI19	42.8537	74.6858	EDL	MARK-L4C-3D

The K-means clustering algorithm, as described in chapter 4, is used to subdivide the dataset into groups based on their similarity of spectral ratios. This means that there is no a priori geographical location information. The K-means clustering algorithm is applied separately to the SSRs and the seismic noise HVSR of examined 15 stations of the temporary seismic network. The analyses were performed by varying the number of clusters (from 2 to 7). As a guide for the selection of the optimum number of clusters, the Calinski and Harabasz index (CH Index) (Calinski and Harabasz, 1974) and Silhouette index (Rousseeuw, 1987) are evaluated for both the SSRs and seismic noise HVSR.

Figure 8.5 gives the CH Index (a, b) and Silhouette index (c, d) for various numbers of clusters for the SSRs (a, c) and seismic noise HVSRs (b, d). Both validation methods indicate that 2 is the optimal number of clusters. Cluster 1 is composed of the four stations in the north (BI01, BI14, BI15 and BI18) while cluster 2 is composed of the rest of the stations. However, in order to allow a smoother spatial variation of ground motion, we have preferred to choose 3 clusters that, anyway, provide the second best CH and Silhouette index results.

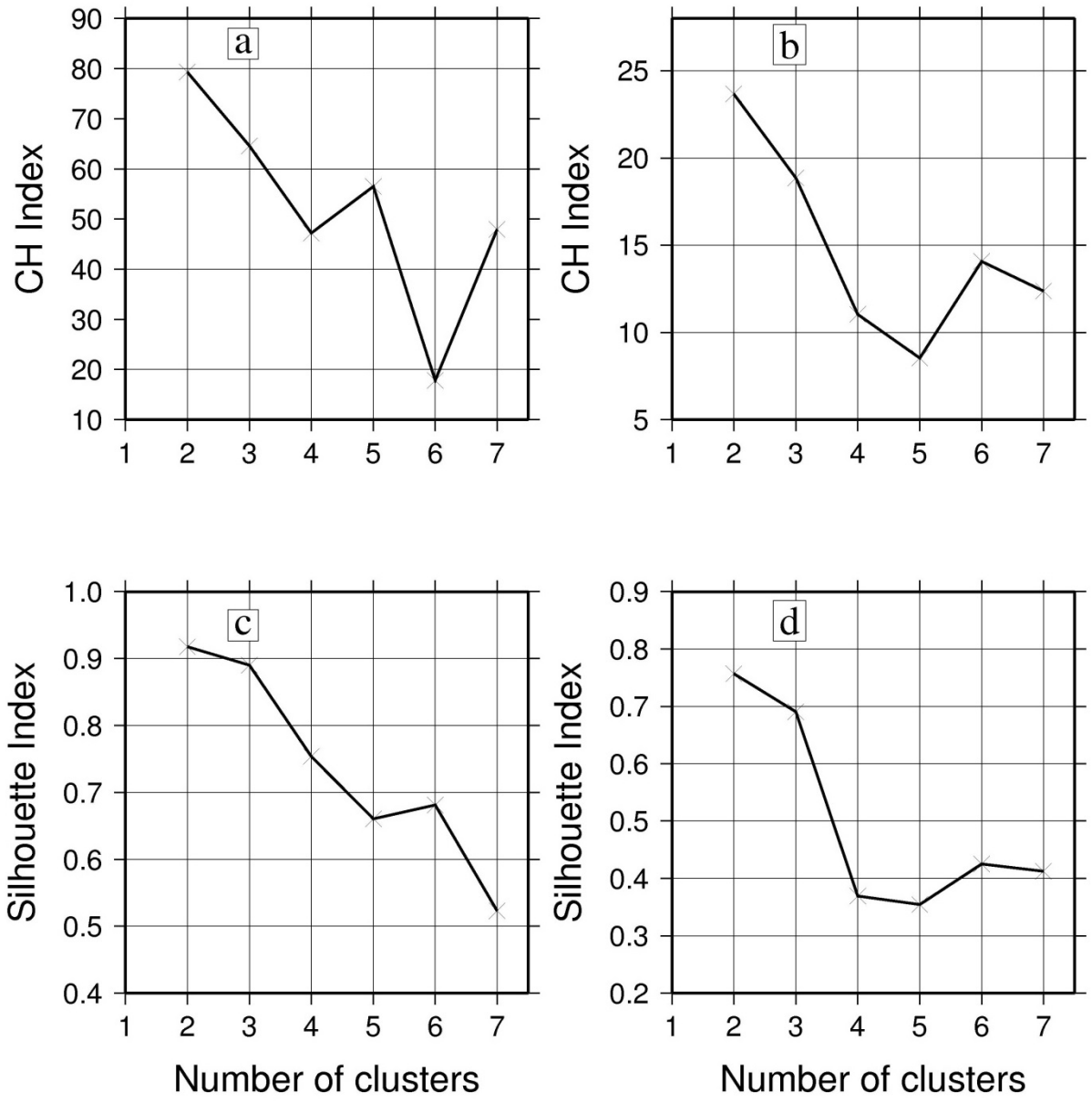


Figure 8.5 Calinski and Harabasz (CH) and Silhouette Indices vs the number of clusters. (a) and (b) shows the CH Index for the SSRs and seismic noise HVSRs clustering, respectively. (c) and (d) shows the Silhouette Index for the SSRs and seismic noise HVSRs clustering, respectively.

8.5 Clustering results

The three selected clusters (for both SSRs and seismic noise HVSRs) are depicted in Figures 8.6 and 8.7, respectively. They show a very similar spatial distribution, independent of whether seismic noise or earthquake data are used, although at each station, the level and shape of amplification from SSR and HVSR might look quite different.

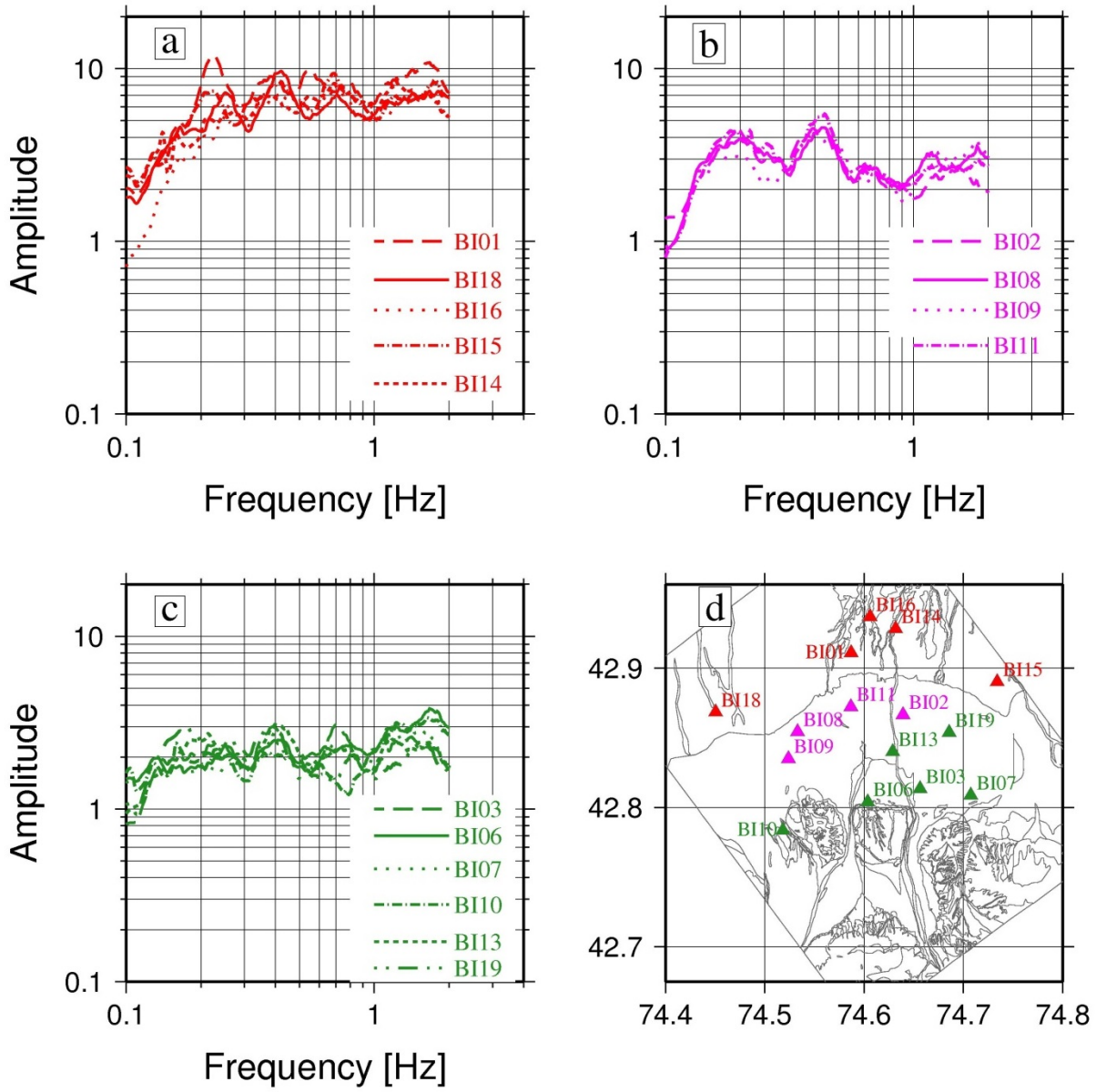


Figure 8.6 Results of the clustering analysis using the K-means algorithm on the SSR. (a), (b), and (c) show the resulting spectra subdivided into different clusters. (d) Shows the spatial distribution of the stations and their respective clusters over geological map of the area.

Figure 8.6 (a) shows the resulting SSRs for the stations clustered in the north of the city. It reveals amplifications generally between 5-9 over the examined range of frequencies, with fairly constant values after around 0.2 Hz. Figure 8.6 (b) is the same for the stations in the centre of the city, and shows a clear amplification of up to 5 at lower frequencies, in particular at 0.2 Hz and 0.4 Hz. For frequencies higher than 0.6 Hz, the amplification remains between 2 and 3. Figure 8.6 (c) presents the SSRs for the stations in the southern-most and south-east of the city, showing a lower amplification compared to the first two clusters, with a fairly constant amplification of between 1.5 and 3 over the examined frequency range, with

occasional peaks at 0.2, 0.4 and 1.5Hz. Figure 8.6 (d) shows the distribution of the different clusters superimposed on a geological map of the city (see Figure 8.2 for details and the geological units).

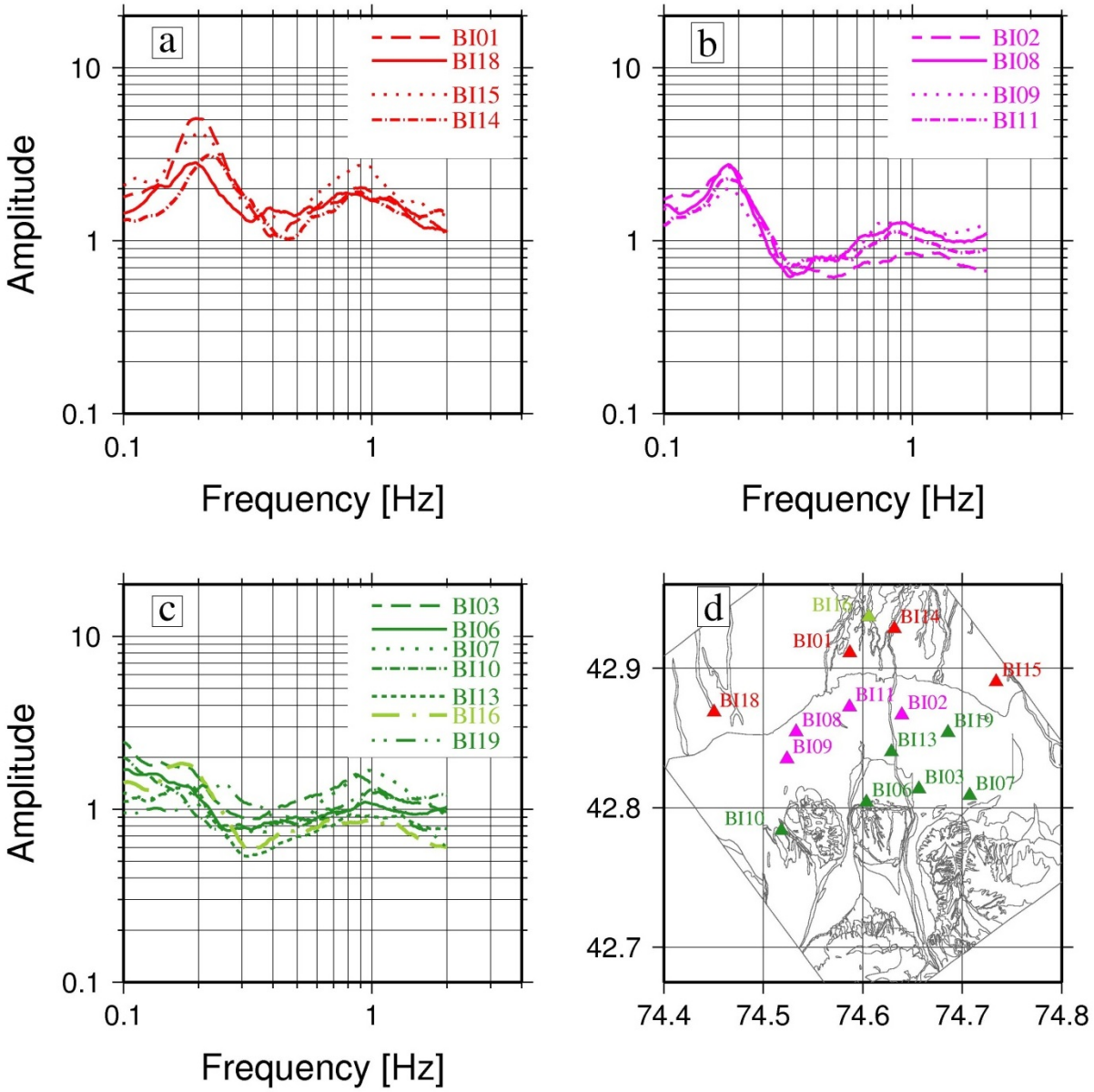


Figure 8.7 Results of the clustering analysis using the K-means algorithm on the seismic noise HVSRs of the temporary network stations. (a), (b), and (c) show the resulting spectra subdivided into different clusters. (d) Shows the spatial distribution of the stations and their respective clusters over a geological map of the area. The anomalous behaviour of BI16 is discussed in the text.

Figure 8.7 is the same as for Figure 8.6, but for the seismic noise HVSR. In HVSR, the position and amplitude of peak plays an important role in the classification of the geological structure (Parolai et al., 2005; Picozzi et al., 2009). Figure 8.7 (a) clusters the stations with

higher amplitudes at around 0.2 Hz and a second prominent peak at around 0.9 Hz. Figure 8.7 (b) clusters the stations with clear peaks at less than 0.2Hz, with a second peak, although not as prominent as in the first cluster. The amplitude of the peaks is also lower than in the first cluster, which possibly results from the deeper bedrock with a lower impedance contrast. Figure 8.7 (c) shows the cluster of stations having a lower flat first peak at around 0.2 Hz, where the amplitude being lower than that for the first two clusters. The trend of lower resonance frequencies towards the southern part of the city is in agreement with the thickening of the sedimentary cover to Paleozoic bedrock, as proposed by Bullen et al. (2001). Figure 8.7 (d) shows the distribution of stations over a geological map, clustered together based on the noise HVSR (see Figure 8.2 for details and the geological units).

It is interesting that sites clustered together by similar SSR are also clustered based on their similarity of seismic noise HVSR. This means that, irrespective of the difference between the SSRs and seismic noise HVSR for a station, the variability in the noise HVSR is consistent with that of the SSRs. This analogy provides the opportunity for SSR to be extended to SSNPs where earthquake data is not available. In general, stations having higher amplitude values of SSR in the northern part of the city also show higher peaks in the seismic noise HVSR, while stations in the southern part that have a lower amplitude in the SSRs also show lower amplitudes in the seismic noise HVSR. Considering the seismic noise HVSR clustering, only station BI16 shows anomalous behaviour (highlighted in a different colour in Figure 8.7a, d) that agrees better with cluster 3 than with the nearby stations in cluster 1 because of the low HVSR peak at lower frequencies. A possible reason for this deviation is that the station at this site was equipped with an EDL 24-bit acquisition system with 4.5 Hz sensor. According to Strollo et al. (2008), the HVSR of the noise peak at lower frequencies obtained by such a sensor is always biased due to the strong filtering effect of the geophone, leading to lower spectral ratio amplitudes at frequencies less than 0.5 Hz. Figure 8.8 shows the power spectral density comparison for the self-noise of EDL 24-bit acquisition system equipped with 1 Hz and 4.5 Hz sensors and the seismic noise recorded at a random place in the city (SSNM no. 45). The 4.5 Hz sensor shows a higher level of self-noise compared to 1 Hz sensor for the same digitizer and gain level. Also from this figure, it is clear that below 0.8 Hz, the self noise of the 4.5Hz sensor becomes larger than the noise signal recorded with 1 Hz sensor. Hence, station BI16 is not considered in further analysis.

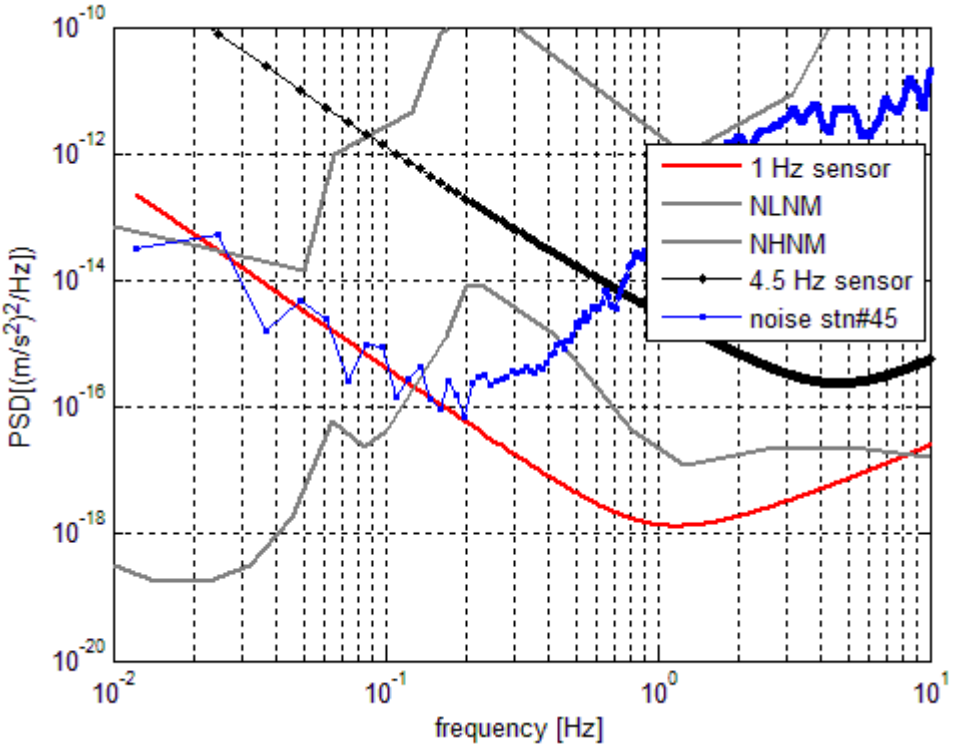


Figure 8.8 Power Spectral Density (PSD) for the self-noise of instruments. The red continuous line shows PSD for self-noise of L4C-3D I Hz sensor with EDL 24-bit digitizer and Gain 10. The black line is for IO-3D 4.5 Hz sensor with EDL 24-bit digitizer and Gain 10. The blue line shows the PSD for the vertical component of a single station noise measurement point. The grey lines show the PSD of the NLNM and NHNM of Peterson (1993).

8.6 Correlation analysis

Since sites having similar SSR also share similar noise HVSR, this analogy is used to take advantage of the HVSR of the single station noise measurement points (SSNP) to improve the spatial resolution of the microzonation (intended here as assessing the site response variation over short distances to be used later in the hazard assessment) in the city. The following correlation scheme to relate the seismic noise HVSR of 177 SSNP out of the 200 (this was because some time series were discarded, as a preliminary analysis of the ratios suggested serious installation problems) with the 14 seismic noise HVSR at the location of temporary seismic network stations (for which SSR are available).

First of all, the Pearson cross-correlation coefficients (Davis, 1986) are calculated between 177 SSNP and the 14 stations using seismic noise HVSR. The Pearson correlation coefficient is the covariance of the two variables divided by the product of their standard deviations, given as,

$$\rho_{jk} = \frac{\left(\frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j).(x_{ik} - \bar{x}_k)}{(n-1)} \right)}{\left(\sqrt{\frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{(n-1)}} \right) . \left(\sqrt{\frac{\sum_{i=1}^{n} (x_{ik} - \bar{x}_k)^2}{(n-1)}} \right)} \tag{8.1}$$

where ρ represents the correlation coefficient, j is an index that goes from 1 to the number of SSNP (177), k is an index that goes from 1 to the number of stations of temporary seismic network (14), i is an index of the frequency (0.1-2Hz), x is the seismic noise HVSR value at frequency i, and $\bar{x}$ is the mean value of x. The value of ρ ranges from -1 to +1.

To quantify the degree of similarity, the Pearson correlation coefficients are multiplied by the degree of fit, which is given by,

$$f_{jk} = \frac{1}{\sqrt{\frac{\sum_{i=1}^{n} (x_{ij} - x_{ik})^2}{n}}} \tag{8.2}$$

This is basically the standard error between the seismic noise HVSR of the SSNP and the stations of the temporary seismic network. Equation (8.2) is used to take into account the difference in amplitude between the noise HVSR of the SSNP and the stations of the temporary seismic network. The greater the difference, the lower the value of equation (8.2), while multiplying equation (8.1) by equation (8.2) will suppress the correlation coefficient and vice versa.

$$CC = \rho.f \tag{8.3}$$

The Matrix CC in equation (8.3) has dimensions of [177 x 14], where 177 represents the number of SSNP and 14 represents the stations of temporary seismic network. Based on the coefficient values of matrix CC, the selection and classification of SSNP with respect to the stations of temporary seismic network are carried out in two steps. In the first step, only those

SSNP are selected which have good similarity of seismic noise HVSR with the stations of the temporary seismic network. This selection is based on the coefficients of CC matrix, which is kept to a threshold value of at least 0.5. Changing the value of this threshold affects only the total number of SSNPs to be considered for further analysis (but not their assignation to a certain cluster). This leads to the selection of about 81% of the 177 SSNP having a CC coefficient of at least 0.5 with any station of the temporary seismic network.

In the second step, the SSNP that are selected in the first step are assigned to a particular cluster, which the stations of the temporary seismic network are already making in terms of seismic noise HVSR (Figure 8.7). A SSNP is assigned to a specific cluster of seismic noise HVSR of the temporary seismic network based on the condition that the maximum CC coefficient between the SSNP and the temporary stations of the cluster is greater than 15% of the maximum CC coefficient when it is compared to the temporary stations of the second best cluster. This way, the SSNP are assigned to particular clusters of seismic noise HVSR of the temporary seismic network. The analyses are run with different threshold levels, and it is noted that while it affects the number of SSNP assigned to a particular cluster/group, mainly in the transition zone (Figure 8.7(b)), the overall distribution of SSNP remains the same. About 79% of the classified SSNP (i.e., 79% of the classified 81%) are finally assigned through this process to different clusters. Noise measurement points that could not satisfy the second criterion were assigned to a specific cluster based on the nearest distance and geology, provided their CC coefficients were within 15% of the maximum. This leads to a total of 143 out of 177 noise measurement points being assigned to the three clusters in this way. Figure 8.9 shows examples of higher to lower CC coefficients values for the seismic noise HVSR of station BI18 compared to the SSNP in the first cluster. Note that larger the value of the CC, the greater is the resemblance to the seismic noise HVSR of the station of temporary network (NHVSR). Figure 8.10 shows the distribution of SSNP and the permanent stations over the geological map of the area. As can be seen from this figure, most of the SSNP are assigned to the clusters in the south and in the north, with few in the transition zone. Some SSNP in the north-west of the city are assigned with the cluster in the south (green colour symbols in Figure 8.10), which also represent the fact that the basin is becoming thicker here, as shown by the resonance frequency map by Parolai et al. (2010).

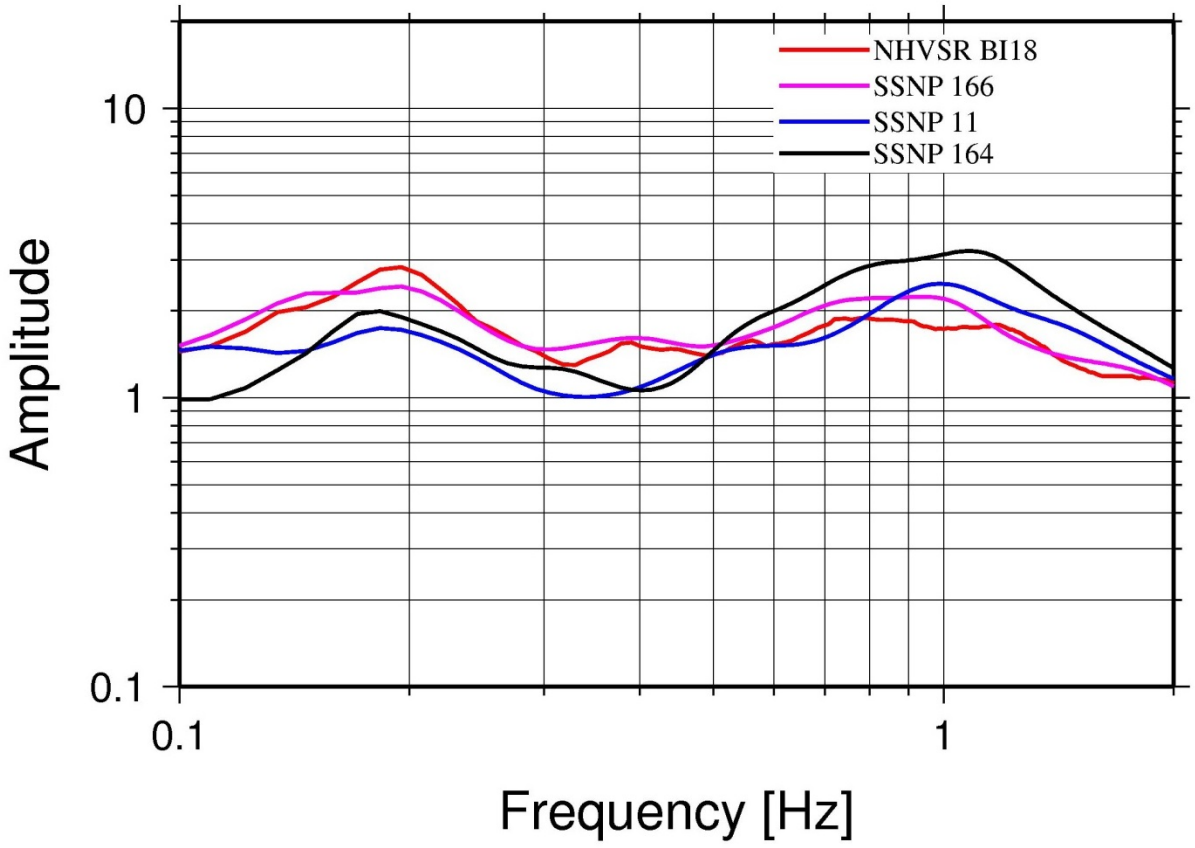


Figure 8.9 Comparison of seismic noise HVSR for SSNP with that from a temporary network station (BI18). The red line shows the noise HVSR of station BI18, magenta shows HVSR of SSNP #166 having a CC coefficient of 3.7, blue is for SSNP 11 having a medium CC coefficient of 2.4 and the black line is for SSNP 164 having a lower CC coefficient of 1.05. The location of these SSNPs are shown in Figure 8.10.

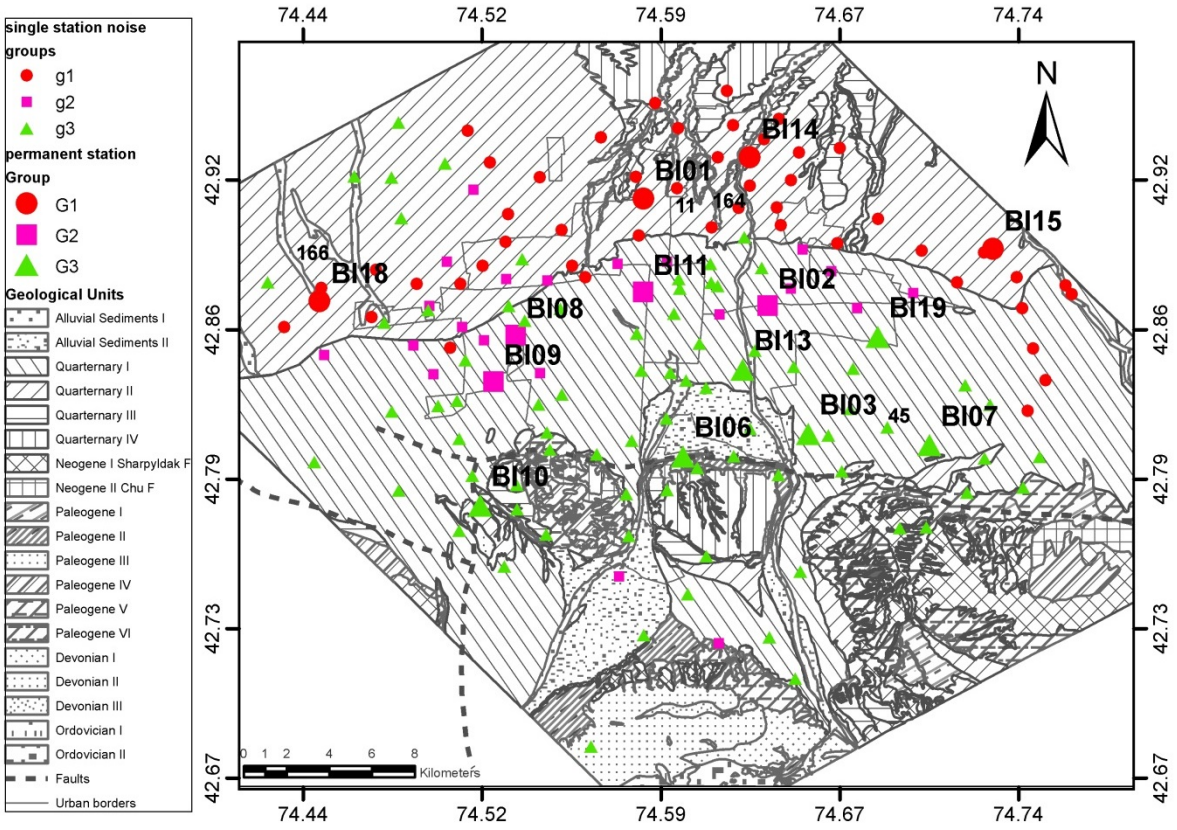


Figure 8.10 The location of the Single station Seismic Noise measurement Points (SSNP) and the temporary network used in this study, superimposed over the geological map of Bishkek, Kyrgyzstan. The bigger size shapes indicate the permanent stations, while the smaller sizes represent the SSNP. Different colours represent the different clusters or groups.

After the SSNP are assigned to different clusters, they are also associated with the site amplification ratios. The site response associated with each SSNP is the logarithmic average of the SSR of each cluster to which the SSNP was assigned after the correlation analysis. Fig. 8.11 shows the SSR for each cluster. Different dashed coloured lines represent different clusters, while the bold continuous line in each cluster represents the logarithmic average of that cluster. The grey area shows the frequency range where analyses are not considered because of the spectra ratios being affected by the site response at the reference site (see Parolai et al., 2010).

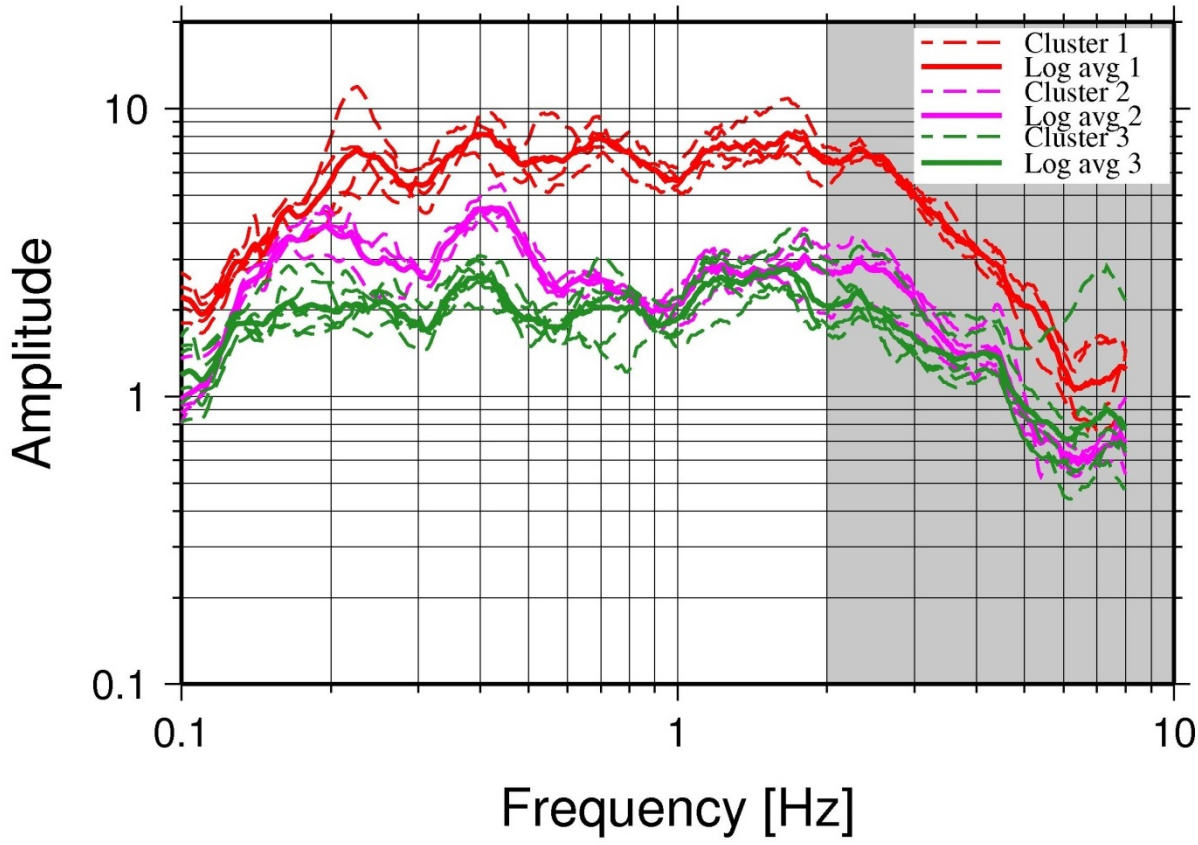


Figure 8.11 Logarithm average of SSRs of each temporary station. Different colours indicate the different clusters, while bold colours in each cluster show the logarithmic average of that cluster. The grey area represents the frequency range where analyses are not considered.

In order to better depict the spatial variation of amplification in the study area, the site response amplitude in terms of SSRs is interpolated using natural neighbours' spatial interpolation. For each temporary station, its own logarithmic average site response is retained as estimated from the earthquake analysis (Parolai et al., 2010). Fig. 8.12 shows the site response estimated for different frequencies over the study area. The figure shows that the largest amplification mainly affects the northern part of the urban area of Bishkek. In particular, the whole frequency band from 0.2 to 2 Hz shows amplification values larger than 5, with maximum values reaching nearly 8 at around 0.4 Hz. The sudden increase of the amplification values corresponding to the geological contact between the different Quaternary materials is striking. The effect of the different Quaternary sediments, responsible for the change in the site response amplitudes between the north and south of the city, can be clearly seen around latitude 42.85 in Figure 8.10.

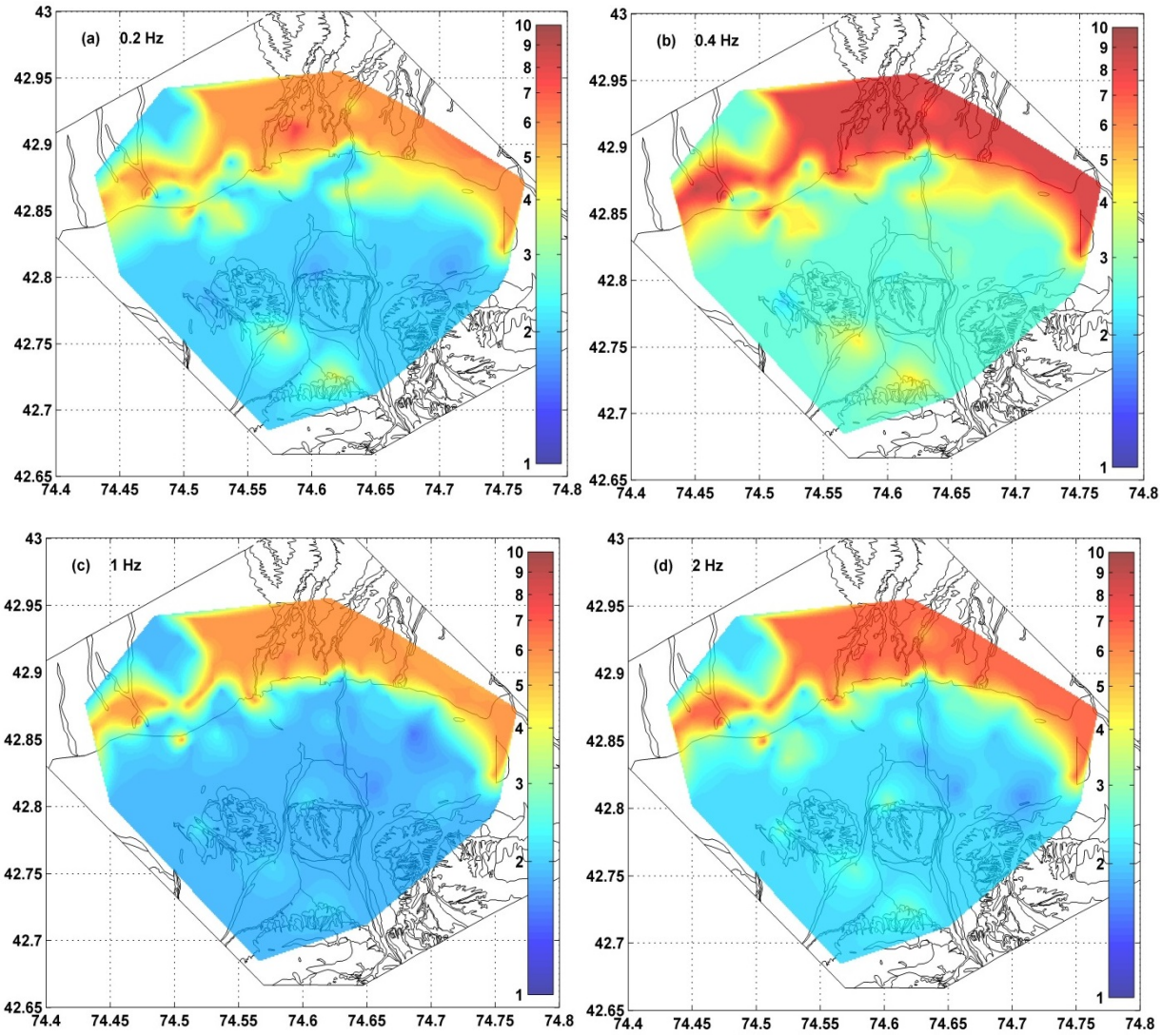


Figure 8.12 Spatial distribution of SSRs at selected frequencies (0.2, 0.4, 1 and 2Hz) using Natural Neighbours Interpolation.

8.7 Discussion

In this study, the spatial resolution of ground motion variability in terms of SSR has been improved upon in Bishkek, Kyrgyzstan, using earthquake and seismic noise data. The appropriateness of seismic noise as a proxy for extrapolating the earthquake site response is evaluated with the help of cluster analysis. Using the K-means algorithm, clusters of sites that display a similar response have been identified based on SSR and seismic noise HVSR. This indicates that, irrespective of the shape of the SSR and of the seismic noise HVSR, sites showing similar SSR also share similar noise HVSR. Using this analogy, the site response in

terms of SSR has been linked with the SSNP recorded in the city for which earthquake recordings were not available. This is accomplished by using a correlation analysis of the seismic noise HVSR between the 14 stations of a temporary network and 177 SSNP.

The three clusters identified in the study area, which cover the northern part of the city, a central transition zone and the southern part, follow well the geological structure of the basin. The depth distribution of the Paleozoic bedrock in the urban area of Bishkek is such that iso-lines are aligned east-west (Bullen et al., 2001, their Fig. 4). The clustering patterns identified in this work follow the same depth distribution, as the spectral ratios reflect the deeper geological structure.

Note that although the clustering analysis performed on the SSR in this study only considers frequencies up to 2 Hz (due to the amplification at the employed reference station), this procedure may also be considered to be valid at higher frequencies since all stations are equally affected by the site response at the reference site. The obtained results allow an efficient microzonation with a higher spatial resolution using seismological parameters alone, and the proposed approach could be used in general as a tool wherever a “good” rock reference site is available. However, in terms of its application for seismic hazard and risk scenarios, in the case at hand, a broader band site response is necessary to correctly evaluate the absolute value of ground motion in the different parts of the city. Suitable correction factors will thus have to be taken into account when considering the effect of the site amplification at the reference site.

9 PSHA FOR BISHKEK INCLUDING EMPIRICALLY ESTIMATED SITE EFFECTS

As shown in the previous chapters, variability in the surface geology potentially leads to the modification of earthquake-induced ground motion over short distances. Although this effect is of major importance when seismic hazard is assessed at the urban level, it is very often not appropriately accounted for. In this chapter, the influence of the variability of shallow geological structure on the seismic hazard assessment for Bishkek, Kyrgyzstan, is calculated. The influence of geological structure on the seismic hazard assessment is considered using different proxies for the site effects. This includes the shear wave velocity and the response spectral ratios.

9.1 Introduction

Since Cornell (1968), several modifications have been proposed to the method to account for specific effects that can affect the final assessment of seismic hazard. Amongst these, particular attention has been dedicated towards the importance of the local variability of ground motion due to the shallow geological structure. Campbell (1997) proposed considering the depth of the basement rock when modelling long-period site response, whereas Boore et al. (1997), among others, suggested the use of Vs30 (the average S-wave velocity over the upper-most 30 meters, based on the travel time from the surface to a depth of 30 m) as a proxy to account for site effects in seismic hazard assessment. More recently, Campbell and Bozorgnia (2008) and Boore et al. (2008) proposed to estimate 1D site effects based on Vs30 and PGA values through relationships that account for non-linear soil behaviour. Campbell and Bozorgnia (2008) additionally attempted to account for 2D-3D effects by modelling the basin response. Note that the above mentioned studies only broadly average the site effects in ground motion prediction relationships. Petersen et al. (1997) accounted for site effects in the PSHA of three southern California counties by using ground motion prediction equations representative of three generic site conditions (alluvium, soft-rock and rock-site) that were valid at the regional level, as deduced from geological maps. Steidl (2000) suggested adopting corrections to rock-site-ground motion prediction equations to be used for PSHA at the regional level by estimating regional amplification factors. These factors are determined by averaging ratios between observed and predicted PGAs and 5% damped SA at 0.3, 1.0 and 3.0

s periods. These amplification factors were therefore dependent on the site-class and the amplitude of ground motion.

Only a few studies have focused on improving probabilistic seismic hazard assessment at the local level. Bazzurro and Cornell (2004) suggested integrating the 1D site response estimated by numerical simulations, while also accounting for uncertainties in the subsoil structure, directly into the PSHA. Erdik et al. (2005) evaluated the seismic hazard for Bishkek accounting for a first-order approximation of the potential site amplification effects, quantifying the building vulnerability and, finally, evaluating the urban earthquake risk. Parolai et al. (2007) used numerical simulations to estimate the site effects in terms of response spectrum ratio for the Cologne (Germany) area. These site effects, along with their uncertainties, were included in hazard analyses by modifying the ground motion prediction equations with respect to a rock site.

As described in the previous chapter, the geology of Bishkek is such that any seismic hazard assessment for the town, which could be used as the basis of a rigorous earthquake risk assessment, must take into account the effect of the shallow geological structure on the spatial variability of earthquake ground motion. Here, the spatial ground motion variability is examined by considering both the shear wave velocity structure below the city, and the response spectra ratio derived by analysing the earthquake recordings collected by an urban seismological network installed in Bishkek for a few months (Parolai et al., 2010). In the first case a standard approach based on the use of Vs30 as a proxy for site effects, while in the second, an approach similar to that of Parolai et al. (2007) is used. Due to the particular sedimentary cover existing in Bishkek (sediments up to several kilometres thick with high shear-wave velocities, especially in the south, of more than 600 m/s) the first approach is expected to be more appropriate for a high-frequency-related parameter such as PGA. The approach based on the calculation of the response spectra ratio is expected to provide a more accurate description of the ground motion variability for the long to intermediate spectral periods, given that the shorter ones are affected by site effects at the reference rock-station. The second approach does not account for nonlinearity, but this can be justified by the low frequencies considered.

9.2 Earthquake catalogue and the source model

The earthquake catalogue used here is the same as that employed for the seismic hazard assessment at the regional level (chapter 6). Similarly to the regional hazard model described in chapter 6, crustal events (hypocentral depths down to 50 km) are used in this study. The same declustering process is applied as that for the regional earthquake catalogue. In order to carry out the seismic hazard assessment for Bishkek, two superzones (defined on the basis of seismo-tectonic and geological data) are selected from the seismic zonation model developed in chapter 6. Figure 9.1 shows the distribution of the events located within these superzones after their selection from the de-clustered Central Asia catalogue. This earthquake catalogue in this region is rich in moderate to large magnitude events, and in particular it includes 12 events with magnitudes greater than 7. While most of the seismicity is concentrated in the south of Kyrgyzstan, in the Tien Shan range, the larger magnitude events (Mw >7) are located to the north of the analysed area. They includes the MLH 8.3, 1889 Chilik and MLH 8.2 1911 Chon-Kemin earthquakes (Bindi et al., 2013; Bindi et al., 2014). The latter is the strongest known event in the Tien-Shan region for which instrumental recordings are also available.

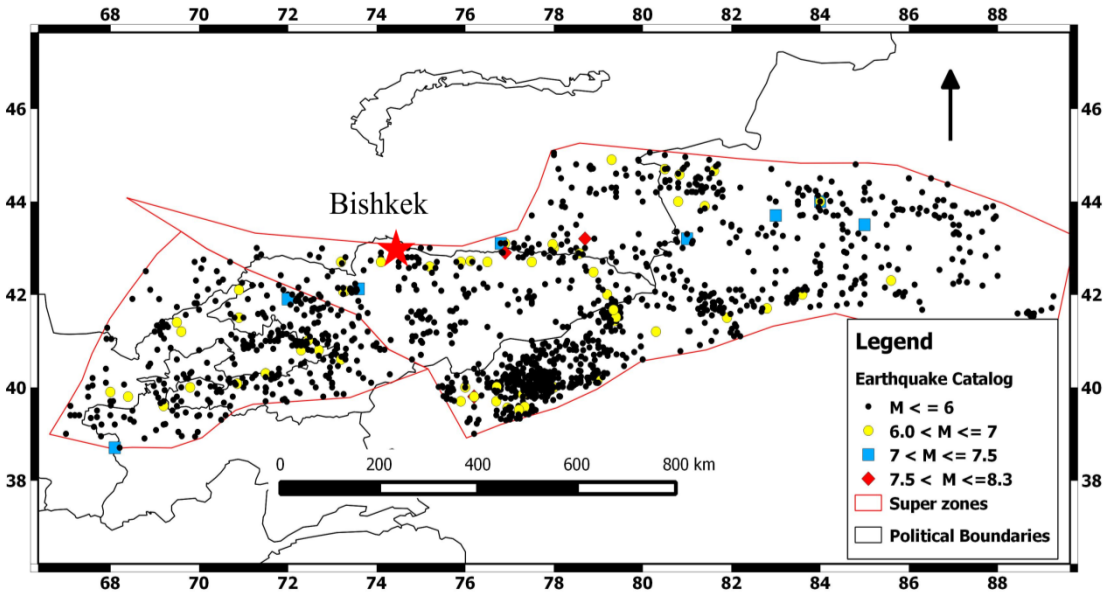


Figure 9.1 De-clustered homogenized catalogue for Central Asia. Different colours and symbol sizes represent different magnitudes ranges as shown in the legend. The location of Bishkek is shown with a star.

It is worth noting that the catalogue is dominated by earthquakes with reverse (thrust) faulting mechanisms (as indicated by the global CMT catalogue in chapter 6), while a smaller number of events appear to display strike slip and normal faulting mechanisms. In particular, in superzone 7 (Figure 9.2), 90% of the events show a reverse faulting mechanism and only 10% are strike slip. In superzone 8 (Figure 9.2), 70% of the earthquakes are reverse fault, 20 % strike slip and 10 % normal faulting events.

The area sources around Bishkek (Figure 9.2) are selected from the regional area source model for Central Asia as defined in Chapter 6. The sources are selected in order to cover an area within a radius of 200 km from Bishkek.

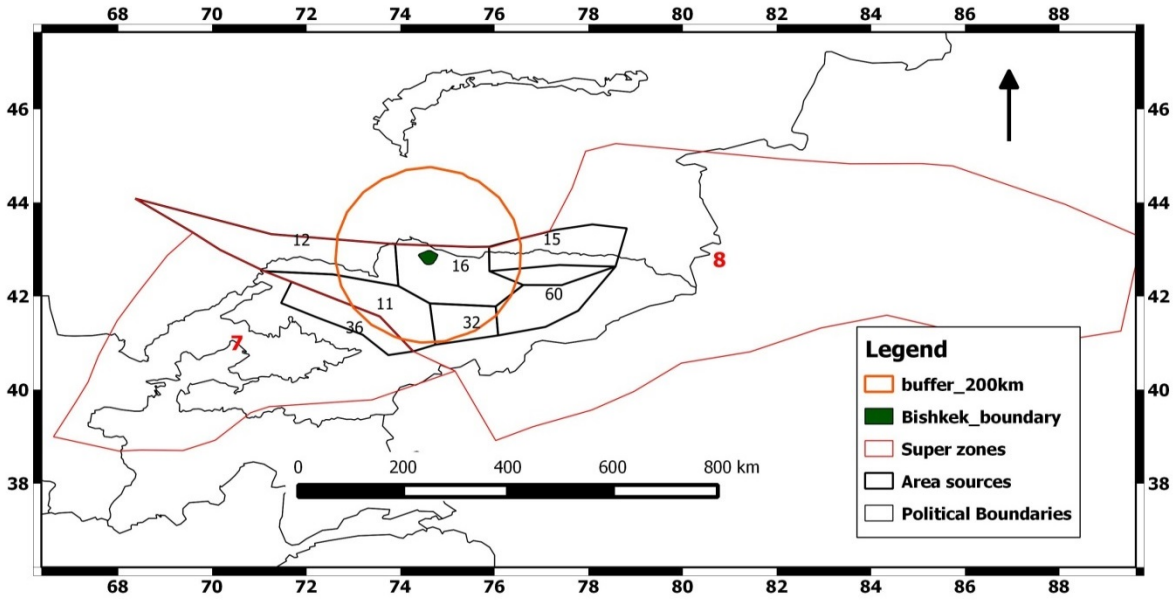


Figure 9.2 Area source models for Bishkek. The areas sources are from the EMCA seismic zonation model, from which the superzones have been defined. The numbers in red indicate the numbering of superzones, while the black numbers indicate the area sources.

Since the IPE used for the regional hazard assessment does not consider explicitly the site effects, a ground motion prediction equation (see section 9.4) is used to estimate the hazard which uses moment magnitude (Mw) as a magnitude scale. For this reason, the magnitude scale of the earthquake catalog is converted to Mw from MLH using regression analysis. A total least square (orthogonal) a quadratic regression analysis is used to estimate the following relationship between the Mw and MLH:

$$Mw = 0.14\,MLH^2 - 0.82\,MLH + 5.97 \qquad (9.1)$$

The orthogonal regression is used here which considers the variability or uncertainty in both variables (in this case, both in Mw and in MLH). Figure 9.3 shows the results of the regression analyses between MLH and Mw based on quadratic total least square regression. The figure also shows the relationships obtained using a least square analysis, and for comparison, the exponential relationship derived by using the global ISC catalog (Storchak et al., 2013). Although the three relationships give almost the same results in the lower magnitude range (below MLH 6.5), only the newly estimated relationship based on the total least square gives the best fit over the larger magnitude ranges.

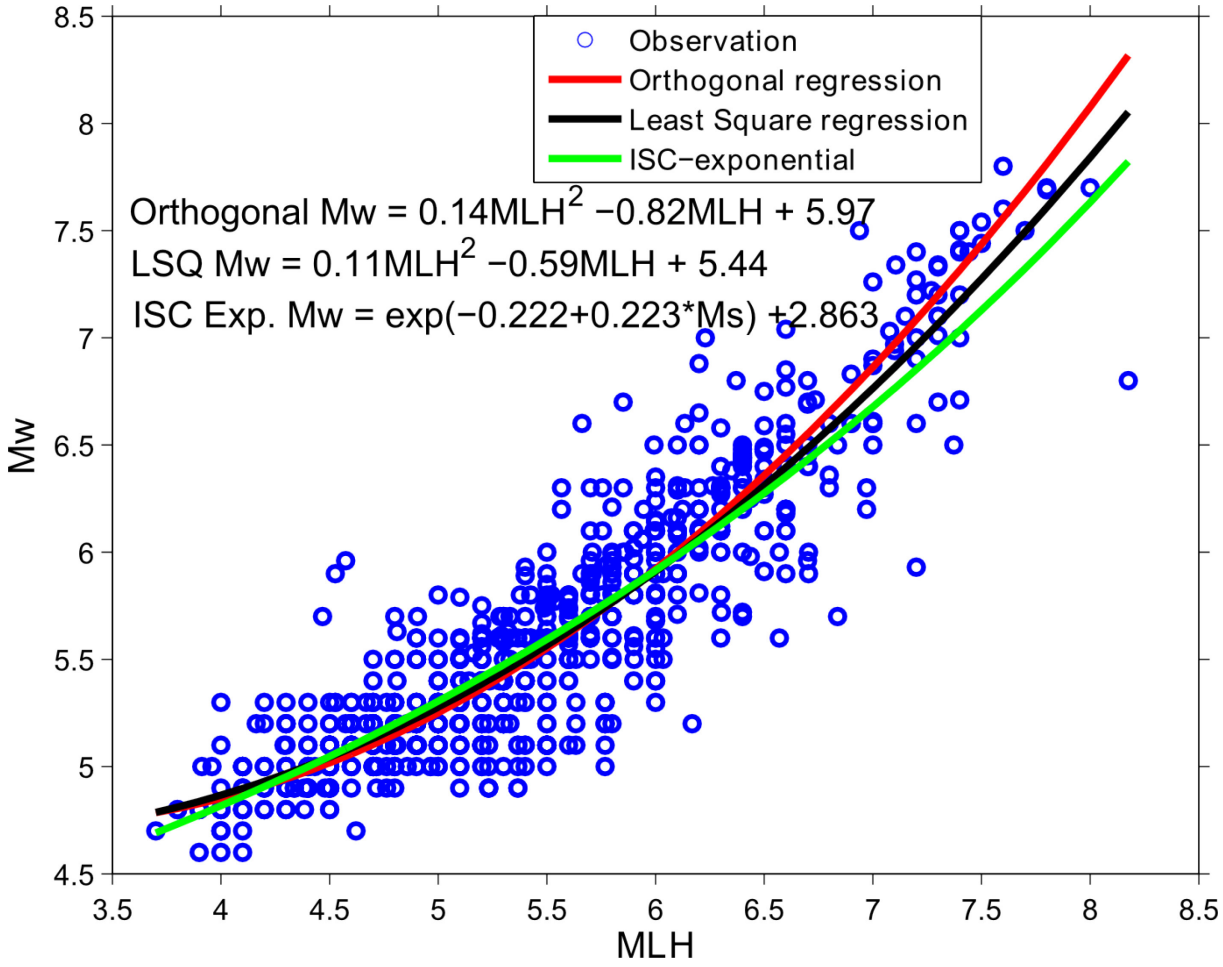


Figure 9.3 Regression analysis between MLH and Mw magnitudes. The blue circles represent the observations used, the black curve represents the least square quadratic regression, the red curve shows the orthogonal regression analysis.

After the magnitude conversion, the seismicity in the sources is processed in the same manner as described in chapter 6. This includes the completeness analyses and GR parameters estimation. The result of the completeness analysis for the shallow catalog combined in the

two superzones, shown in table 9.1, indicates a completeness of 1956 for magnitude 4.7 and above. The recurrence analyses are estimated for the two superzones based on the completeness analysis of table 9.1. It is noted from Figures 9.4 and 9.5 that the obtained a and b values are very similar, being close to 7 and 1.3, respectively.

Table 9-1 Magnitude classes and completeness analysis using the Stepp (1973) method.

Magnitude (Mw)	**4.7**	**5**	**5.6**	**6.2**	**6.8**	**7.4**	**8**
Completeness	1956	1942	1894	1860	1860	1860	1860

The maximum magnitude is assigned to each area source based on the value estimated for the superzone they belong to. In this case, magnitude values of 7.5 and 8.3 are considered for zones 7 and 8 (Figures 9.1 and 9.2), based on the largest known earthquake that occurred in the area. Different parameters for each of the considered area sources are summarized in Table 9.2.

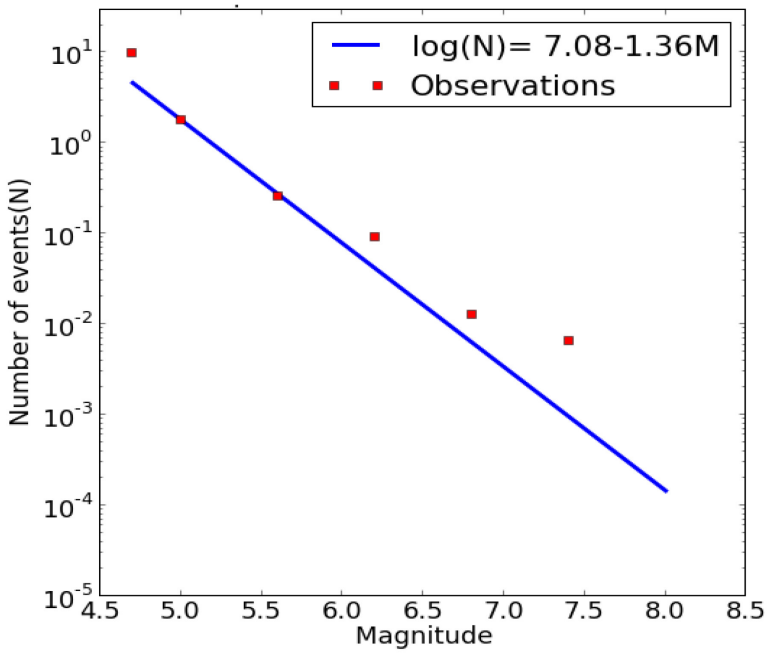


Figure 9.3 Recurrence plot for superzone (SZ) 7 (see Figure 9.2).

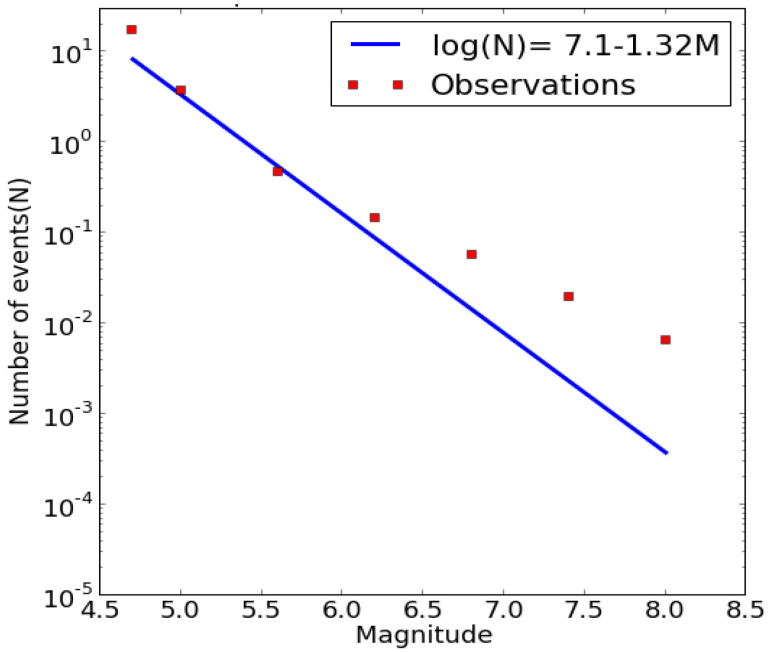


Figure 9.4 Recurrence plot for superzone 8 (see Figure 9.2).

Table 9-2 Summary of the area sources parameters for the study region (see Figure 9.2).

Area source	Super zone	No. of events	a-value	b-value	Mmin	Mmax
11	8	30	6.09	1.42	4.7	8.3
12	8	21	4.089	1.03	4.7	8.3
15	8	91	4.41	0.96	4.7	8.3
16	8	81	5.92	1.31	4.7	8.3
32	8	43	5.28	1.21	4.7	8.3
36	7	98	6.85	1.5	4.7	7.5
60	8	46	6.47	1.46	4.7	8.3

9.3 Site effects

In the previous chapter, the spatial resolution of site effects in Bishkek is improved, following the clustering and correlation analyses applied to the SSR and noise HVSR. In particular, areas are identified that share similar site responses in terms of Fourier spectra ratios. Using this information, the influence of the site response in determining the hazard level is studied

considering both a proxy for site effects (Vs30) and by modifying the ground motion prediction equations by considering the standard response spectra ratio of empirical data. The calculation of these site effects proxies are explained below.

9.3.1 *Vs30 approach*

The Vs30 values used in this study are obtained from in-situ measurements of S-wave velocity derived by the analysis of microarray of noise data (Parolai et al., 2005, Boxberger et al., 2011). Array measurements are carried out ad-hoc, considering the results of chapter 8, at sites representative of each area sharing similar site response, following the results of cluster and correlation analysis. Figure 9.6 shows the set-up and results of an array recording site in Bishkek near station BI08 (Parolai et al., 2010). This is located near the Central Asian Institute for Applied Geosciences (CAIAG) in Bishkek. The shear wave velocity (Figure 9.6 (c)) is about 600 m/s in the shallowest layers down to 25 m depth, with a time-averaged Vs30 of 650 m/s.

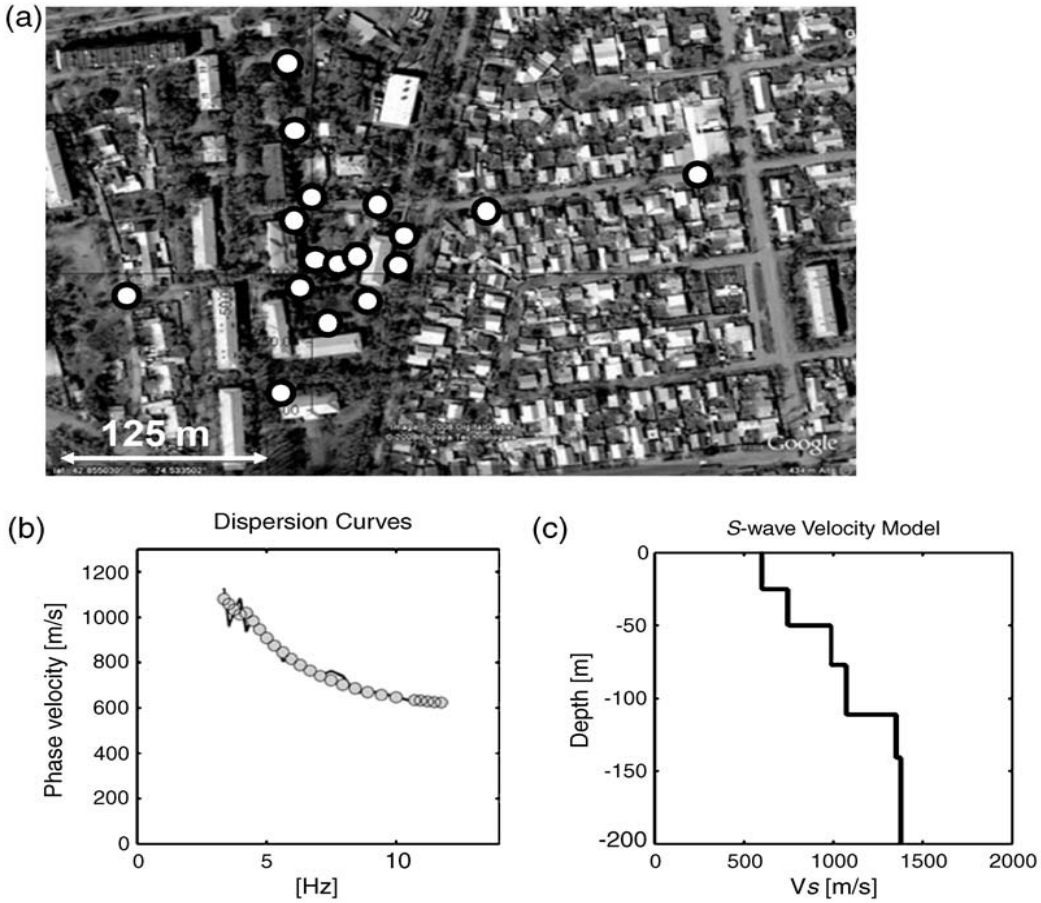


Figure 9.5 Array measurement in Bishkek (in courtyard of CAIAG) close to station BI08 (Fig 9.7). (a) Array geometry (white circles), (b) observed (black line) and reconstructed (grey circles) dispersion

curves, (c) S-wave velocity model obtained by inverting the dispersion curve in (b) (Parolai et al. 2010).

Figure 9.7 shows the results of the array analyses for the shear wave velocity versus depth at different locations in Bishkek. The red colour shows the shear wave velocity for the array analysis carried out near the Issyk Ata fault (Pilz et al., 2013). It has the time averaged Vs30 of 862 m/s. The black line shows the results of array analysis carried out in the stadium near station (BI 11), with the time average Vs30 of 650 m/s. The blue line shows the result of an array analysis in the north near station BI 01, with the lowest time averaged Vs30 of 213 m/s.

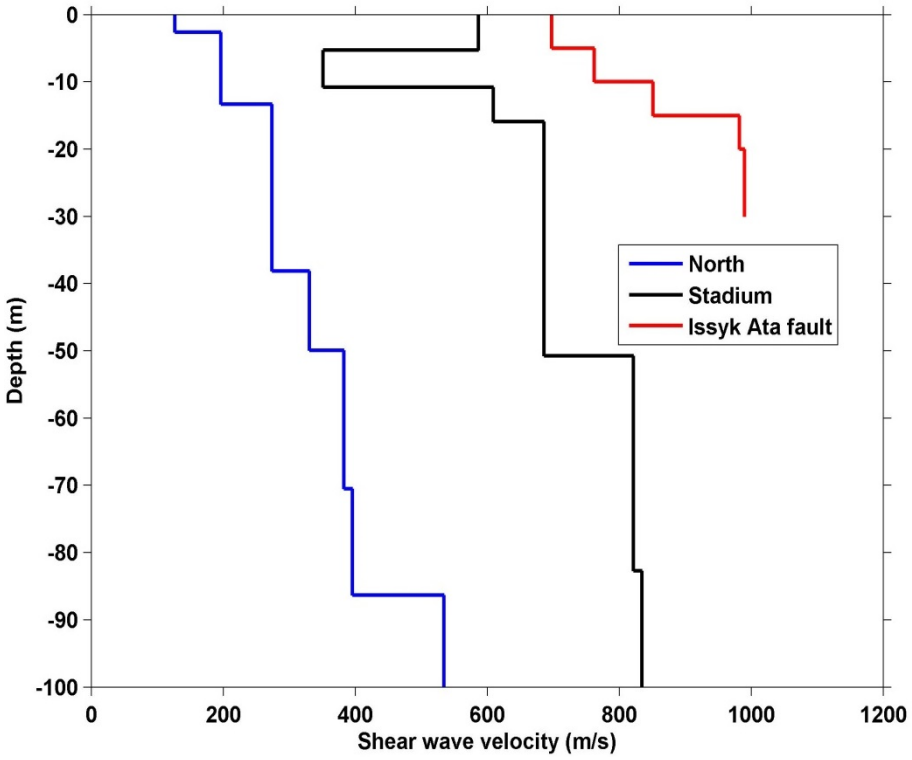


Figure 9.6 Results of array analyses in Bishkek at different locations (for their location, refer to Figure 9.8 and the text).

Figure 9.8 shows the location of the array sites together with the geological map of the area. In general, Vs30 decreases from south to north, consistent with a change in the characteristics of the shallow geological material (Parolai et al., 2014). In the northern part of Bishkek, where the Quaternary sediments mainly consist of silt or loess, the shear wave velocity (Vs30) is low, with values around 200 m/s. In the central transition zone, where the sedimentary cover is mainly composed of pebbles and gravels, the Vs30 is of the order of 650 m/s. while in the southern part of the study area, where the shallow surface material consist of boulders,

pebbles, and gravel, values as high as ~850 m/s are found. Note that at the station used as the reference for the following site response analysis, Vs30 is estimated to be about 900 m/s.

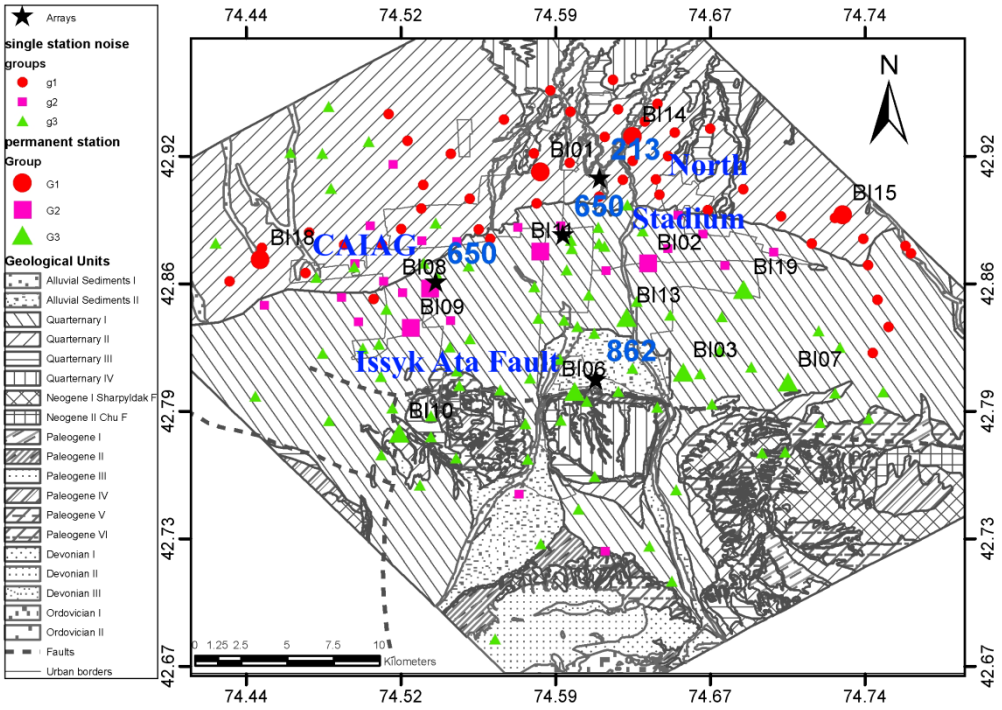


Figure 9.7 Array locations (black stars) with Vs30 values (blue colour). The different colours of the stations (single station noise measurements and permanent stations) represent different clusters in terms of the similarity of site effects from SSR.

9.3.2 *Response spectrum ratios*

As an alternative approach, the site response to be incorporated into the PSHA is also estimated in terms of response spectrum ratios with respect to the reference site (Parolai et al., 2007; Parolai et al. 2009). First, the response spectra (5% damping) are calculated for the earthquake recordings of each station of the temporary seismological network used in chapter 8. Before calculating the response spectra, the earthquake recordings are de-trended and deconvolved for instrumental response. Second, their ratios with respect to the response spectrum at the reference station are computed. The average of all the response spectra ratios of the stations in each cluster is then found. Station BI04, located in the Kyrgyz range south of Bishkek, is chosen as a reference site. It is the same station used as the reference in chapter 8. However, as noted in chapter 8, this station, although being the only one located on an apparently rock site, is affected by amplification of ground motion at frequencies higher than 2-3 Hz (0.3s - 0.5s) and with a maximum at nearly 5 Hz. For this reason, the following

analysis is carried out only considering the amplitude of the response spectra ratio (as coefficients to be used in the seismic hazard assessment) over the frequency range 0.2 Hz to 2 Hz (0.5 s to 5 s). Figures 9.9, 9.10 and 9.11 show the average response spectra ratios for all stations. These figures highlight the spatial variability of the response spectra ratios over the study area. For example, in Figure 9.9, the stations located in the north of Bishkek, where the average Vs30 is as low as 213m/s, show the largest amplifications. In general, the highest values of amplification are occurring over a broad frequency band.

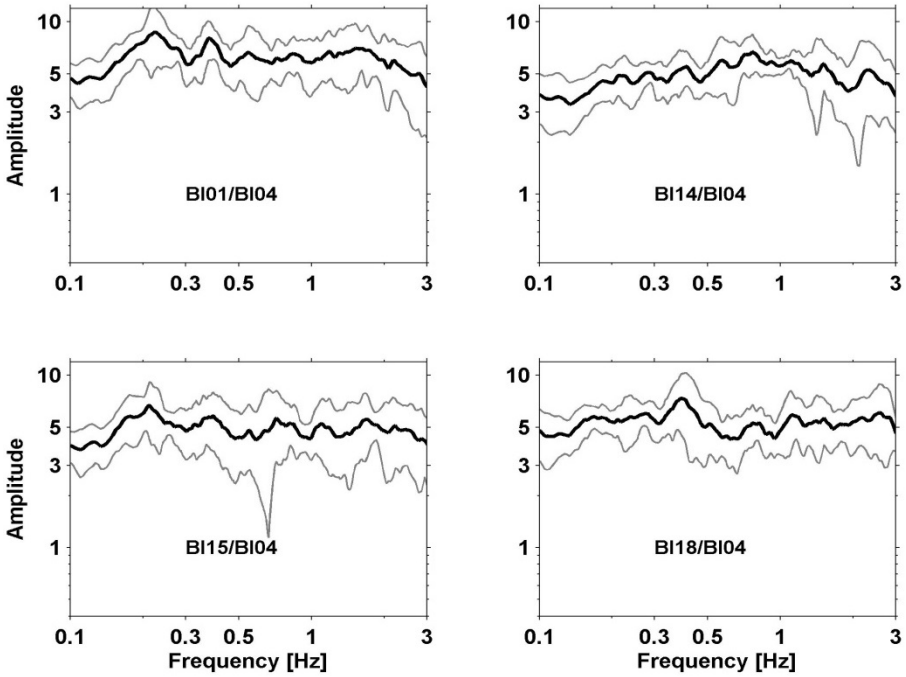


Figure 9.8 Acceleration response spectrum ratio of each station w.r.t BI04. The grey lines represent the plus/minus one standard deviation.

Figure 9.10 depicts the results obtained for the stations located in the middle of city where the average Vs30 is 650 m/s. These stations show lower average amplifications, mainly affecting frequencies lower than 0.5 Hz and higher than 1 Hz. Figure 9.11 shows the results obtained for the stations installed in the southern-most and south-east sectors of the city, where the average Vs30 is 862m/s. These response spectra ratios show a comparatively lower amplification over a broad frequency band.

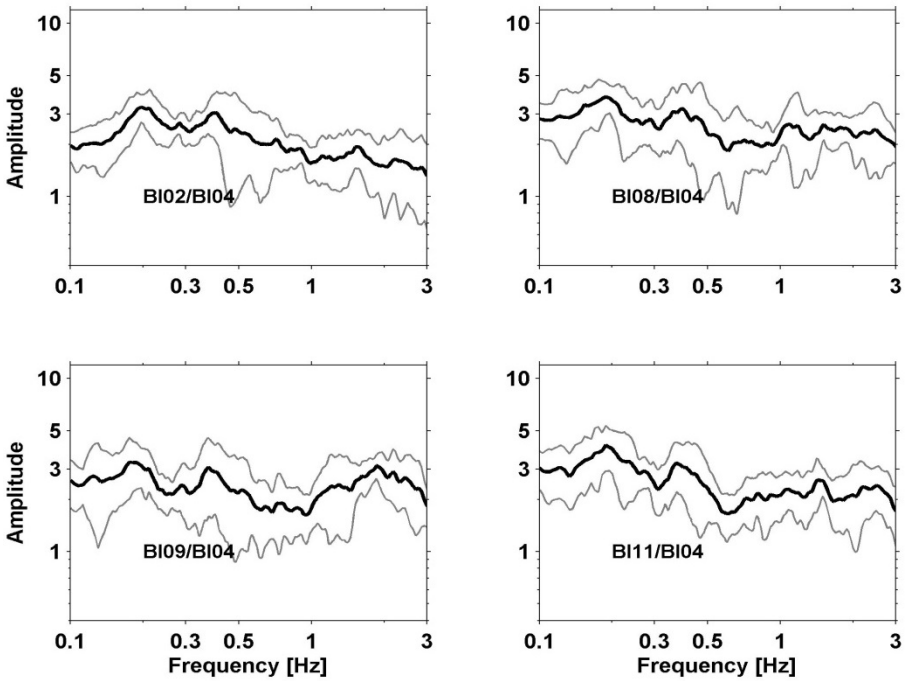


Figure 9.9 same as Figure 9.9 except being for locations in the middle of the city (see Figure 9.8).

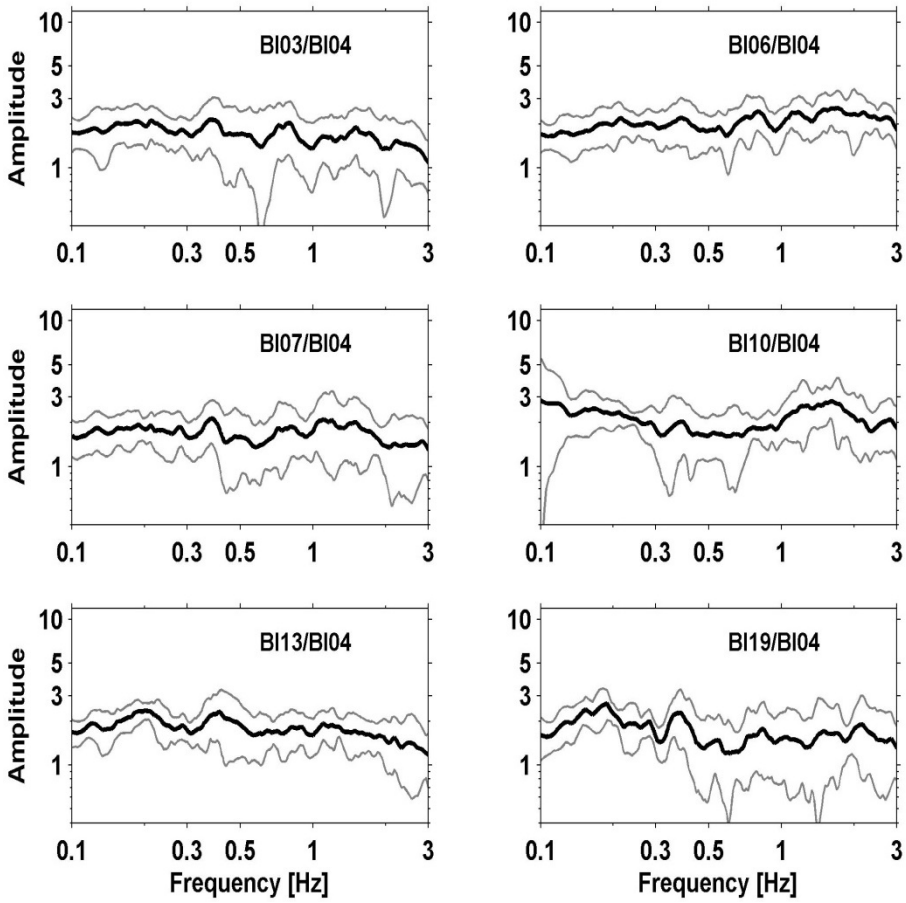


Figure 9.10 Same as Figure 9. 9, except being for the southern and south-east sectors (see Figure 9.8).

9.4 PSHA including site effects

In this study, site effects are accounted for in the PSHA by considering them directly in the ground motion prediction equation (GMPE) via the Vs30 values estimated at each site. This leads to the GMPE, which are valid for a rock site, to include the response spectra amplification factors (Bazzurro et al. 2004; Parolai et al., 2007). The Boore and Atkinson (2008) GMPE is used in this study, which is also recommended by GEM for active shallow crust conditions. The general form of the Boore and Atkinson (2008) GMPE is given as:

$$\ln(Y) = F_M(M) + F_D(R_{JB}, M) + F_S(V_{S30}, R_{JB}, M) + \varepsilon\delta_T \quad (9.2)$$

where the terms F_M, F_D, and F_S represent the magnitude scaling, the distance function, and site amplification, respectively. Y is the ground motion parameter, RJB is Joyner-Boore distance metric, M is the magnitude, δ_T is the period-dependent standard deviation and ε is the fractional number of standard deviations. In particular, the site coefficient is represented by considering a linear and non-linear amplification term:

$$F_S = F_{LIN} + F_{NL} \quad (9.3)$$

The non-linear term F_{NL} depends on the input PGA. The linear term depends on the V_{S30} through the relation:

$$F_{LIN} = b_{LIN}\ln(\frac{V_{S30}}{V_{ref}}) \quad (9.4)$$

where b_{LIN} is a period-dependent coefficient, and V_{ref} is the reference velocity (760 m/s). In the PSHA calculations, V_{S30} is used at each site, derived from the array analyses. Sites where measurements are not available have been assigned a certain Vs30 value following the clustering and correlation analysis of chapter 8. When the PSHA is carried out for the rock site, the V_{S30} of the local reference site (900 m/s) is used.

When the site effects are introduced considering the coefficients estimated by the response spectral ratios, the term F_S of the Boore and Atkinson (2008) GMPE is described as:

$$F_S = F_{LIN} + F_{NL} + \ln(T_i) \quad (9.5)$$

where $\ln(T_i)$ is the period-dependent average response spectrum ratio for each site. In order to scale the coefficients to a common reference site (consistent with that used for the response

spectra ratio calculation) in equation (9.5), the V_{S30} used in the linear site specific term is fixed to 900 m/s, independent of the location.

Consistent with the PSHA carried out at the regional level in chapter 7, the ground motion aleatory variability is truncated at 3 sigma. The PSHA for the urban area of Bishkek is carried out in terms of PGA and spectral acceleration (SA) for different periods.

9.5 Results

The PSHA assessment for the area of Bishkek is now shown for the cases when considering an hypothetic homogeneous rock site condition, and for when site effects are introduced.

Figure 9.12 shows the maximum and minimum limits of the SA with 10% probability of exceedance in 50 years when considering uniform rock site conditions for Bishkek. The values are very close and there is only a slight increase in the expected ground motion from north to south (Figure 9.13), very likely related to the dominance of the shaking produced by the larger number of earthquakes sources in the south. The maximum hazard observed for this model is represented by a SA of 0.45 g at 0.1s (10Hz) and a PGA reaching 0.21 g in the southern outskirts of the city.

Figure 9.14 shows the same maximum and minimum hazard as Figure 9.12 with a 10% probability of exceedance in 50 years estimated using the model including the variability of Vs30. The maximum hazard values are 0.64 g for the SA at 0.1s and 0.31 g for the PGA. Note that different from what was obtained when rock-site conditions are considered; the largest values are found in the northernmost area of the city (Figure 9.15). This is the area where, due to the low S-wave velocities, the highest amplification of ground motion is estimated (see Figure 9.9).

Similar trends in the results are observed when site effects are accounted for through the response spectra ratio coefficients. That is, the minimum and maximum SA curves show (Figure 9.16) larger values than those calculated for a rock site condition (Figure 9.12) and the area affected by the largest hazard becomes the northernmost one (Figure 9.17). However, in this case, the maximum hazard calculated for SA at 0.5s reaches values up to 1.13g, much larger than that obtained by considering Vs30 only. It should be noted that, owing to the

amplification at the reference station above 2 Hz, calculation for shorter periods below 5s (above 2 Hz) are not carried out.

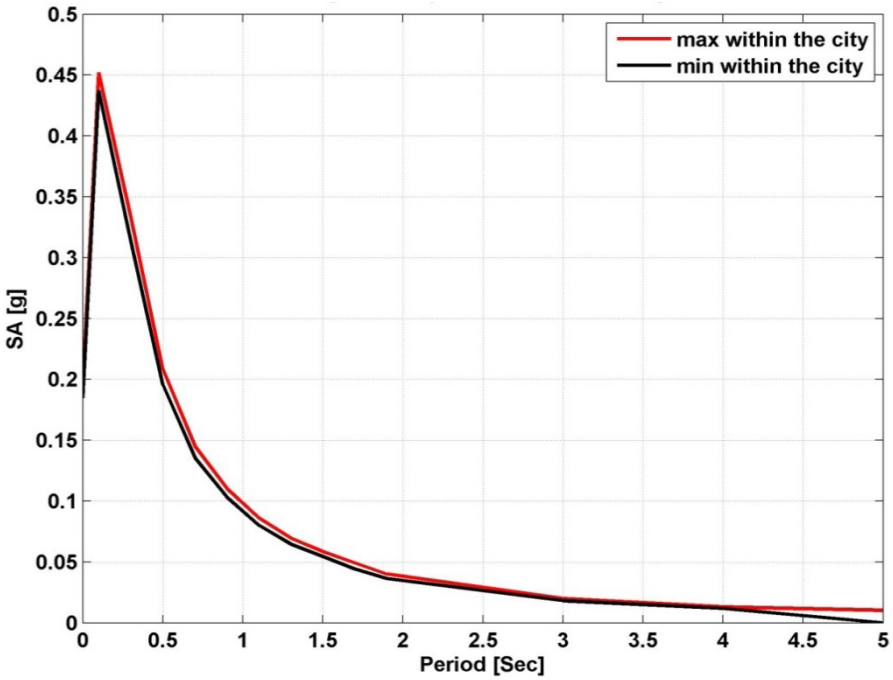


Figure 9.11 Estimated seismic hazard variation within the city for rock-site conditions (Vs30=900m/s), for 10% probability of exceedance in 50 years.

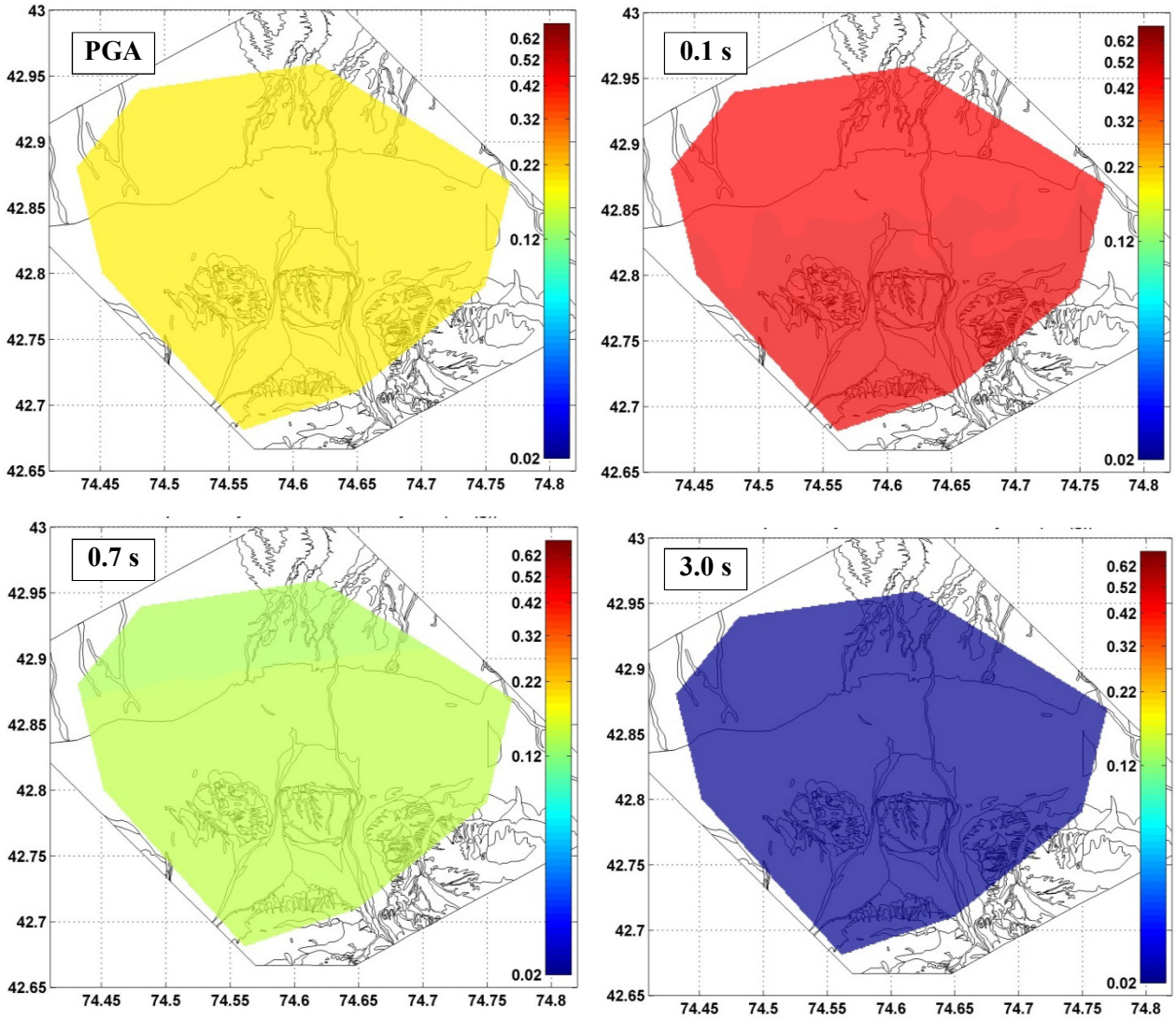


Figure 9.12 Seismic hazard in terms of 10% probability of exceedance in 50 years for rock-site conditions. The hazard is shown for PGA and spectral acceleration at different selected period in terms of g. For a detailed description of geological map of the city in the background, see figure 9.7.

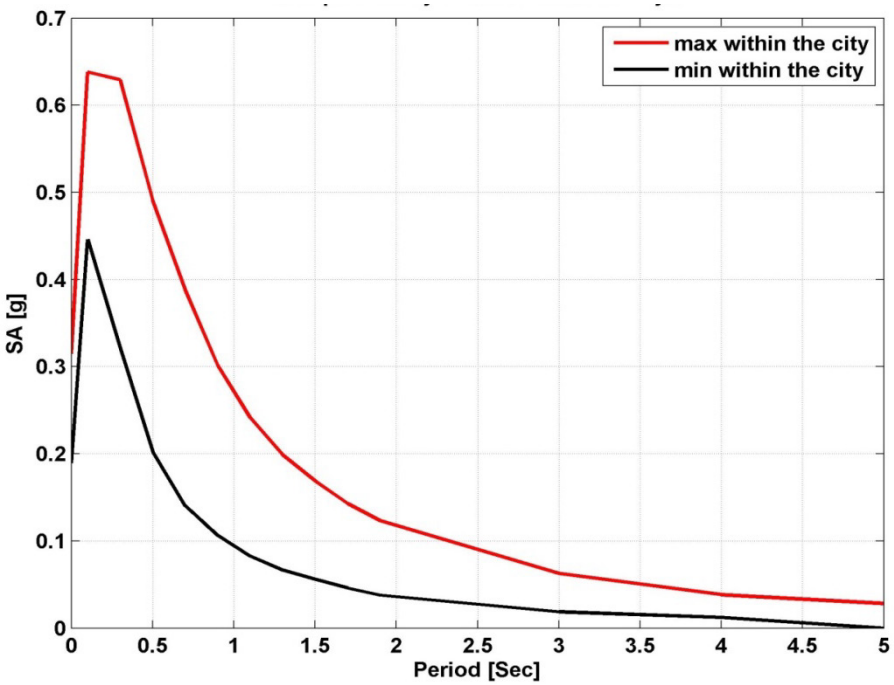


Figure 9.13 Estimated seismic hazard variations within the city considering site effects based on Vs30 variations, for a 10% probability of exceedance in 50 years.

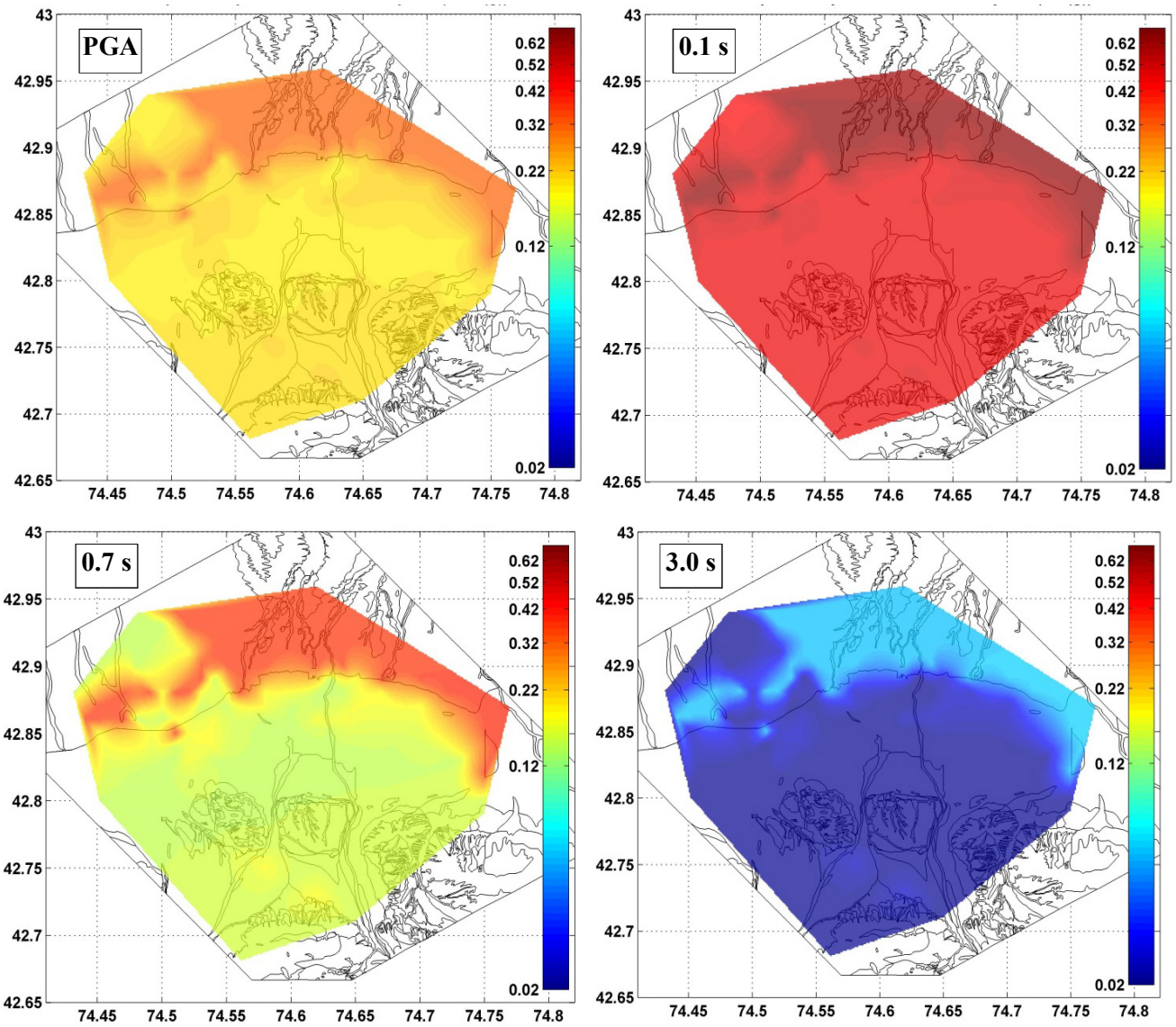


Figure 9.14 Estimated hazard for 10% probability of exceedance in 50 years considering the distribution of Vs30 values. The level of hazard is shown for PGA and spectral acceleration at different selected periods in terms of g.

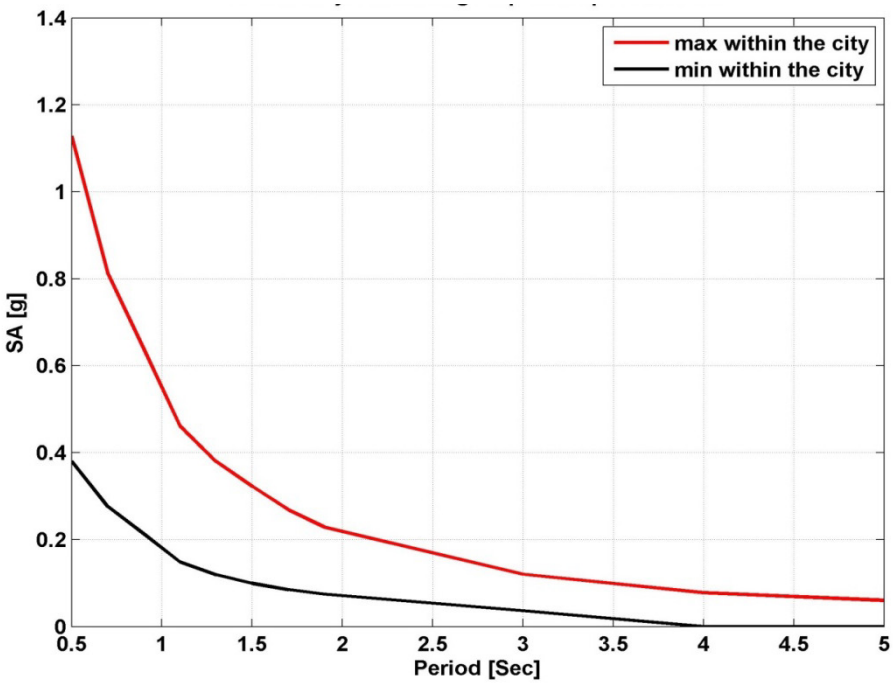


Figure 9.15 Estimated seismic hazard variations within the city considering site effects estimated from response spectrum ratios, for 10% probability of exceedance in 50 years.

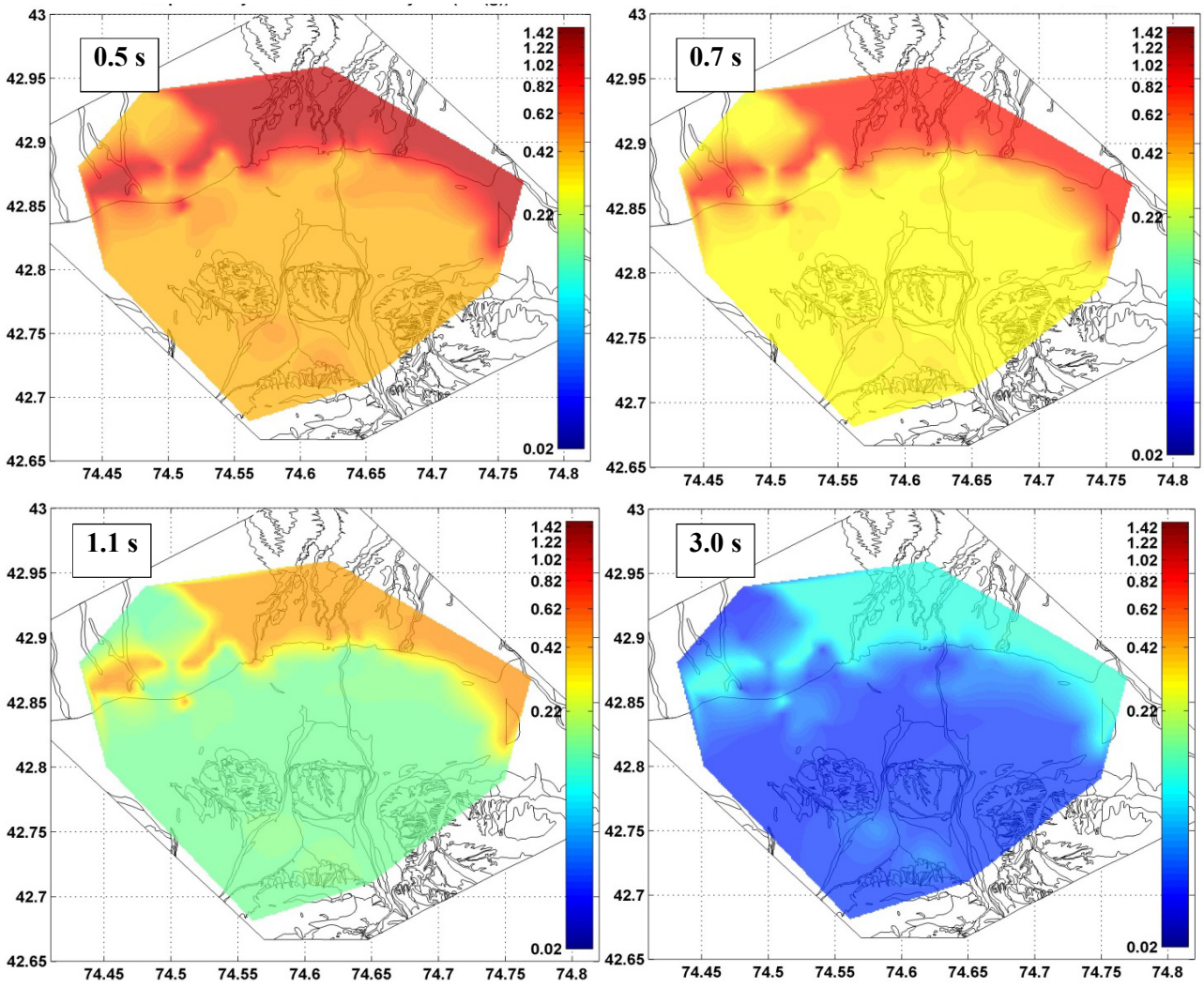


Figure 9.16 Estimated hazard for 10% probability of exceedance in 50 years considering site effects in terms of response spectrum ratios for different spectral periods.

9.6 Discussion

In this study, starting from the results of previous site effects studies, the PSHA analysis is carried out for Bishkek that accounts for the modification of ground motion over short distances. Both, the Vs30 values used in the GMPE and the response spectra amplification factors introduced into the GMPE are derived from empirical data.

The inclusion of site response into the hazard analysis shows important variations in the hazard assessment for Bishkek. The major effect in including the site response (or a proxy for it) into the PSHA is the re-prioritization of the urban territory in terms of its hazard level. In

fact, when a generic rock site condition is used, the most hazardous areas are the southernmost ones with a slight decrease of hazard towards the north. This result reflects the distribution of the seismicity and the majority of the seismogenic structures with respect to the town. Most of the active sources of seismicity are located in the east, south and south-west of Bishkek, whereas in the north lies the stable regions of Kazakhstan. On the contrary, introducing into the PSHA model parameters that describe the effect of the soil response not only masks this effect, but completely inverts the hazard trend in the urban area. In this case, the largest hazard is observed in the northern part of the city. This is very important observation considering the seismicity and seismogenic structure of the region.

The two methods used for site effects, that is the Vs30 estimated from array analyses and the response spectral ratio estimated from earthquake data, show the importance and suitability of using different proxies for site effects in seismic hazard assessment. However, the level of hazard estimated by the two proposed procedures disagrees in terms of absolute values, especially for periods less than 1 second. The differences in the estimated hazard values are also quite noticeable for longer period. This might be due to the following reasons:

1) Vs30 is a parameter representative mainly of the very shallow structure that is expected to influence especially the high-frequency part of the ground motion (and therefore the PGA). It might therefore not be the most appropriated parameter in a large basin where the site response at intermediate and long periods might be affected by the shape of the basin itself and where the site response can affect a wide frequency band (Pilz et al., 2009). The ground motion could therefore be underestimated for the medium to long periods (although unlikely to be affected by strong nonlinear behaviour).
2) On the other hand, the response spectral ratio is estimated using weak motion data. Hence, the analyses are considered in the linear range. While the approach that is followed here can be appropriate for areas with moderate seismicity, it might lead to an overestimation of the ground motion since it is not accounting for the soil's nonlinear behaviour. While this might be a relatively minor problem in the southern parts of the city where the soil mainly consists of boulders and pebbles grading slowly to gravel toward the centre of the urban area, it can be a problem for the northernmost part of Bishkek where silty deposit and loess prevail.

Although this study shows the importance of accounting for soil response in the quantitative PSHA of Bishkek, further studies are necessary in order to calibrate the procedure for including realistic site effects. The acquisition of in-situ geotechnical parameters, the analysis of borehole data and the combination of empirical data with site responses estimated from non-linear analysis will allow an improvement in the seismic hazard estimation at the urban level, while taking advantage of the track indicated by this study.

10 CONCLUSIONS

An updated earthquake catalogue up to 2009 was compiled for Central Asia within the EMCA Project. The different magnitudes are homogenized to surface wave magnitude MLH using empirical regression equations. This updated earthquake catalogue is used to calculate an updated trans-national probabilistic seismic hazard assessment for Central Asia. The seismic hazard is calculated in this study based on the direct analysis of local seismic histories provided by the smoothed seismicity approach, and the standard Cornell (1968) approach based on areal sources. The region specific intensity prediction equation is used to estimate the regional seismic hazard is in terms of macroseismic intensity (MSK-64) using the OpenQuake software. Although the methods reveal clear differences between each other in terms of absolute hazard levels, the locations of the highest levels of hazard are rather congruent. These regions cover the Alai valley, the Fergana valley, and the region north of Issyk-Kul, and are characterized by intensity levels of more than 8 for the area source model. The maximum hazard observed in the region shows an intensity level of 9 in the Kyrgyz-Kazakh border region for a return period of 2475 years. The comparison between the different approaches shows that rank differences are mainly related to the seismotectonic zoning. The maximum hazard extracted from the regional hazard model for some of the cities in the region, Bishkek, Dushanbe, Tashkent and Almaty, is between 7 and 8 (7-8), 8.0, 7.0 and 8.0 macroseismic intensity, respectively, for 475 years mean return period, using different approaches. Comparing the results of seismic hazard from this study with the earlier studies for the rock site condition, earlier studies give higher values of seismic hazard in general for the whole territory of Central Asia.

Empirically estimated site effects are included for the urban level seismic hazard assessment of Bishkek, Kyrgyzstan, considering as a proxy for site effects the Vs30 parameter, and the response spectral ratios. Vs30 is calculated from array analysis using seismic noise recordings. The site effects in terms of response spectral ratios are estimated after clustering and correlation analyses. Using K-means clustering algorithm separately over standard spectral ratios and noise horizontal to vertical spectral ratio over stations of a temporary seismic network, similar clusters are identified. This implies that stations showing similar site response in terms of standard spectral ratios also show similar seismic noise horizontal to vertical ratios. Correlation analysis is then used to establish a similarity between the stations

of temporary seismic network and a dense network of single station noise measurement points. Based on the results of correlation analysis, the cluster's site responses are then adopted for single station noise measurement points, which improves the spatial resolution of ground motion variability. The response spectrum ratios of each cluster, calculated from 5% damped response spectra using earthquake records, are also assigned to each single station noise measurement points after the results of correlation analysis.

The seismic hazard at the urban level is evaluated in terms of peak ground acceleration (PGA) and spectral acceleration (SA) at different periods (frequencies), using ground motion prediction equations. Introducing site effects into the seismic hazard assessment of Bishkek changes the orientation of the estimated hazard trend with respect to rock site conditions, which assumed homogeneous Vs30 values over the whole urban area. For the rock site condition, the southern part of Bishkek has the highest hazard, with a slight decrease of hazard towards northern part of the city. The maximum hazard observed for Bishkek considering rock site conditions is 0.21 g PGA, with 0.45 g spectral acceleration at 0.1 s in the southern part of the city for a return period of 475 years. However, including site effects into the hazard analysis, inverts the trend of hazard in the city. Using the Vs30 values, the maximum hazard is observed at 0.1 s, having 0.64 g spectral acceleration with 0.31 g PGA in the northern part of the city for a 475 years return period. For the response spectral ratios, the maximum hazard is calculated for spectral acceleration at 0.5 s reaching up to 1.13 g for a 475 years return period. Using response spectral ratios and Vs30 as proxies for site effect, shows not only the importance of including site effects in hazard, but also shows the importance of selecting the proper proxy. In particular, the Vs30 might not be the most appropriate proxy for large basins, where the effect of intermediate to long-period waves become more important, which might lead to the underestimation of ground motion. On the other hand, considering response spectral ratios as a proxy for site effects estimated from weak motion data might not be a proper choice for sites having large seismicity.

For the future, to undertake seismic hazard assessments of engineering interest, further studies are needed, including the consideration of zones of diffuse and concentrated seismicity and a comprehensive uncertainty analysis of epistemic and aleatory variables, such as alternative seismic source models, different earthquake catalogues and uncertainty in the estimated recurrence rates and vibratory ground motion values. Additional study is required to map and

characterize the local faults in the region. The effect of these faults should be included in the seismic hazard assessment as separate source model through the logic tree approach.

For the inclusion of site effects, additional data should be collected to compare the effects of different proxies over a wide range of frequencies. In particular, care must be taken to account for non-linear site effects, which otherwise might lead to overestimation of hazard estimates.

11 REFERENCES

Abdrakhmatov, K. Y., Aldazhanov, S. A., Hager, B. H., Hamburger, M. W., Herring, T. A., Kalabaev, K. B., Makarov, V. I., Molnar, P., Panasyuk, S. V., Prilepin, M. T., Reilinger, R. E., Sadybakasov, I. S., Souter, B. J., Trapeznikov, Y. A., Tsurkov, V. Y., and Zubovich, A. V., 1996, Relatively recent construction of the Tien Shan inferred from GPS measurements of present-day crustal deformation rates: Nature, v. 384, p. 450-453.

Abdrakhmatov, K., H.-B. Havenith, D. Delvaux, D. Jongmans and P. Trefois 2003. Probabilistic PGA and Arias Intensity maps of Kyrgyzstan (Central Asia), Journal of Seismology 7, 203–220

Agrawal, G, Gunopulos, R. (1998) Automatic subspace clustering of high dimensional data for data mining applications (SIGMOD'98).

Algermissen, S.T., Perkins, D.M., Thenhaus, P.C. , Hanson, S.L. , Bender B.L. (1982) Probabilistic estimates of maximum acceleration and velocity in rock in the contiguous United States U.S. Geol. Surv., Open-File Rept. (1982), pp. 80–1033

Allen, T.I., Wald, D.J. (2009). On the use of high-resolution topographic data as a proxy for seismic site conditions (VS30). Bulletin of the Seismological Society of America, Vol. 99, (2A):935–943

Aptikaev, F.F. and N.V. Shebalin (1988). Specification of correlation between level of macroseismic effect and dynamic parameters of ground movements, researches on seismic danger, Question of Engineering seismology, 29, 98-107, (in Russian)

Baeva, N.B. (1999) Geological profile, Government Agency of geology and mineralogy of Kyrgyz Republik. Kyrgyz-Hydro-geological expedition.

Bard PY (1999) Microtremor measurements: a tool for site effect estimation? In Proceedings of the Second International Symposium on the Effects of Surface Geology on Seismic Motion, Yokohama, Japan, December 1998, Vol. I–III, 1251–1279

Bard PY (2004): The SESAME project (2004) An overview and main results, In Proceedings of the 13th World Conference on Earthquake Engineering, Vancouver. Paper: 2207.

Bard, P. Y., SESAME WP02 team (2005): Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations – measurements, processing and interpretations, European Comission – Research General Directorate Project No.EVG1-CT-2000-00026 SESAME, D23.12

Bardet, J.P., Ichii, K., Lin C. H. (2000) EERA A computer program for equivalent-linear earthquake site response analyses of layered soil deposits. Technical report, University of Southern California.

Bazzurro, P., Cornell, C. A. (2004) Ground-Motion Amplification in Nonlinear Soil Sites with Uncertain Properties. Bulletin of the Seismological Society of America, Vol. 94, No. 6, pp. 2090–2109, December 2004

Beauval, C., Hainzl, S. and Scherbaum, F., 2006.Probabilistic seismic hazard estimation in low-seismicity regions considering non-Poissonian seismic occurrence, Geophys. J. Int., 164, 543-550.

Beauval C, Scotti O, Bonilla F (2006). The role of seismicity models in probabilistic seismic hazard estimation: comparison of a zoning and a smoothing approach. Geophys J. Int. 165:584–595

Bender, B. (1983): Maximum likelihood estimation of b-values for magnitude grouped data, Bull. Seismol. Soc.Am., 73, 831-851

Bezdeck, J.C., Ehrlich, R., and Full, W. (1984). FCM: Fuzzy C-Means Algorithm. Computers and Geoscience, 10(2–3), 191–203.

Bindi D, AA Gómez Capera, S Parolai, K Abdrakhmatov, MStucchi, J. Zschau (2013). Location and magnitudes of earthquakes in Central Asia from seismic intensity data: model calibration and validation. Geophys J Int.192, 710-724 doi:10.1093/gji/ggs039

Bindi D, Abdrakhmatov K, Parolai S, Mucciarelli M, Grunthal G, Ischuk A, Mikhailova N, Zschau J (2012) Seismic hazard assessment in Central Asia: Outcomes from a site approach .Soil dynamics and earthquake engineering. 37:84-91

Bindi D, Mayfield M, Parolai S, Tyagunov S, Begaliev UT, Abdrakhmatov K, Moldobekov B, Zschau J (2011) Towards an improved seismic risk scenario for Bishkek, Kyrgyz Republic. Soil Dyn Earthquake Eng 31:521-525

Bindi D, Parolai S, Spallarossa D, Catteneo M (2000) Site effects by H/V ratio: comparison of two different procedures. J Earthquake Eng 4:97-113

Bindi D., Parolai, S., Gómez-Capera, A., Locati, M.,Kalmetyeva, Z., Mikhailova,N. (2014).Locations and magnitudes of earthquakes in Central Asia from seismic intensity data, J Seismol 18, 1-21 DOI 10.1007/s10950-013-9392-1

Bindi, D, Parolai, S, Oth, A, Abdrakhmatov, K, Muraliev, A, and J. Zschau (2011). Intensity Prediction Equations for Central Asia, Geophysical Journal International 187(1):327–37. doi:10.1111/j.1365-246X.2011.05142.x.

Bonilla, L. F., Steidl, J. H., Lindley, G. T., Tumarkin, A. G., and Archuleta, R. J. (1997). Site amplification in the San Ferdinando Valley, California: variability of site effect estimation using S-wave, coda, and H/V methods. *Bull. Seism. Soc. Am.*, **87,** 710–730.

Boore, D. M., Atkinson, G. M., (2008) Ground motion prediction equations for the average horizontal component of PGA, PGV, and 5% damped PSA at spectral periods between 0.01 s and 10.0 s. Earthquake Spectra, Vol. 24, No. 1, 99-138.

Boore, D.M. (2013) The uses and limitations of the square-root-impedance method for computing site amplification. Bull. Seismol. Soc. Am. 2356-2368.

Boore, D.M., Joyner, W. B., and Fumal, T. E. (1997). Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: a summary of recent work. Seism.Res. Lett. 68, 129–153.

Borcherdt, R.D. (1970). Effects of local geology on ground motion near San Francisco Bay. Bull. Seismol. Soc. Am. 60:29-61.

Boxberger, T., Picozzi, M., Parolai, S. (2011). Shallow geology characterization using Rayleigh and Love wave dispersion curves derived by seismic noise array measurements. - Journal of Applied Geophysics,75, 2, p. 345-354.

Bragato PL, Laurenzano G, Barnaba C (2007) Automatic Zonation of Urban Areas Based on the Similarity of H/V Spectral Ratios. Bull Seismol Soc Am 97:1404-1412

Bullen, M. E., Burbank, D.W.,Garver, J. J., and Abdrakhmatov, K. Y. (2001).Late Cenozoic tectonic evolution of the northwestern Tien Shan: New age estimates for the initiation of mountain building, Geol. Soc. Am. Bull. 113, 1544–1559

Bune V.I. and G.P. Gorshkov, (1980) Seismic zonation of USSR, Bune V.I. and G.P. Gorshkov editors, Moscow: Nauka, 307 p. (in Russian)

Burtman, V.S. and Molnar, P. (1993) Geological and Geophysical Evidence for Deep Subduction of Continental Crust Beneath the Pamir. 76 p. Special Paper 281; Geological Society of America.Boulder.

Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. Communication in Statistics- Theory and Methods, 3:1-27

Campbell, K. W., Bozorgnia, Y. (2008). NGA ground motion model for Geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s. Earthquake Spectra. Vol. 24, No. 1, 139-171

Campbell, K.W. (1997).Empirical near-surface attenuation relationships for horizontal and vertical components of peak ground acceleration, peak ground velocity and pseudo-absolute acceleration response spectra. Seism. Res. Lett. 68(1):154–179.

Capon, J. (1969). High-resolution frequency-wavenumber spectral analysis. Proc. IEEE., 57, 1408-1419.

Cornell, C.A. (1968). Engineering seismic risk analysis. Bull. Seism. Soc. Am. 58, 1583-1606

Crowley, H. (2005). An investigative study on the modelling of earthquake hazard for loss assessment (PhD thesis).

D'Amico, V., and Albarello ,D. (2008). SASHA: A Computer Program to Assess Seismic Hazard from Intensity Data, Seismological Research Letters 79, 5, 663-671

D'Amico V, Picozzi M, Albarello D, Naso G , Tropenscovino S (2004) Quick estimates of soft sediment thicknesses from ambient noise horizontal to vertical spectral ratios: a case study in southern italy. J Earthquake Eng 8:895-90

Daniel Drake (1815) Natural and Statistical View or Picture of Cincinnati and the Miami Country, Illustrated by maps. With an appendix containing observations of the late earthquakes, the Aurora Borealis, and the South-west wind. Looker and Willace, Cincinnati, 1815.

Davis JC (1986) Statistics and Data Analysis in Geology-2nd edition. John Wiley and Sons, Toronto

Dercourt, J., Ricou, L.E., and Vrielynck, B., Eds., Paris: Gauthier-Villars,1993. Atlas Tethys Palaeoenvironmental Maps

Ekström, G., Nettles, M. (2013) Global CMT web page, http://www.gloabalcmt.org/. Accessed 17 April, 2014

Erdik M, Rashidov T, Safak E, Turdukulov A (2005) Assessment of seismic risk in Tashkent, Uzbekistan and Bishkek, Kyrgyz Republic. Soil Dyn Earthquake Eng 25:473-486

Ester, M., Kriegel, H-P., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceeding of 2nd Int. Conf.On Knowledge Discovery and Data Mining, Portland (pp. 226–23).

Esteva, L. (1967). Criteria for the construction of spectra for seismic design, presented at Third Panamerican Symposium on Structures, Caracas, Venezuela

Faenza, L., Marzocchi, W., Boschi, E. (2003) A nonparametric hazard model to characterize the spatio-temporal occurrence of large earthquakes; an application to the Italian catalogue. Geophys. J. Int., 155 (2003), pp. 521–531

Fayyad, M.U., Piatesky-Shapiro, G., Smuth P., Uthurusamy, R. (1996). Advances in Knowledge Discovery and Data Mining. AAAI Press.

Field, E. H., and Jacob, K. H. (1995). A comparison and test of various site response estimation techniques, including three that are non-reference- site dependent. *Bull. Seism. Soc. Am.*, **86,** 991–1005.

Frankel, A.(1995) Mapping seismic hazard in the central and eastern united states. Seismological Research Letters 66(4):8-21

Gallagher, K., and M. Sambridge (1994). Genetic algorithms: A powerful tool for large-scale non-linear optimization problems. Comput. Geosci., 20(7/8), 1229–1236.

Gardner, J. K. and L. Knopoff (1974). Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? Bulletin of the Seismological Society of America 64:1363 –1367

Ghose, S., M. W. Hamburger, and C. J. Ammon (1998). Source parameters of moderate-sized earthquakes in the Tien Shan, central Asia from regional moment tensor inversion, Geophys. Res. Lett. 25: 3,181–3,184

Giardini D. (1999). The global Seismic Hazard Assessment Program (GSHAP)—1992/ 1999. Annali di Geofisica 42(6), 957–974

Gómez , J. B. and Pacheco, A. F., 2004, The Minimalist Model of characteristic earthquakes as a useful tool for description of the recurrence of large earthquakes, Bull. Seismol. Soc. Am., 94, 1960-1967.

Gorshkov, G.P. 1937. The seismic map of the USSR, Bol'shoysovietskiy Atlas Mira, Moscow 1:93. (in Russian)

Gubin, I.E, (1960) Reularities in seismic manifestations on the territory of Tajikistan. Geology and Seismicity. M., Publ. House USSR Acad. of Sci (1960), p. 461 (in Russ.)

Guha, S, Rastogi, R., and Shim K. (1999). ROCK: A Robust Clustering Algorithm for Categorical Attributes. In Proceedings of the IEEE Conference on Data Engineering.

Guha, S., Rastogi, R., and Shim K. (1998). CURE: An Efficient Clustering Algorithm for Large Databases. In Proceedings of the ACM SIGMOD Conference.

Gutenberg, B., Richeter, C.F. (1944) Frequency of earthquakes in California. Bull. Seism. Soc. Am. 185-188

Halkidi M, Batistakis Y, Vazirgiannis M (2001) On Clustering Validation Techniques. J Intell Inf Syst 17:107-145

Hashash,Y.M.A., Groholski, D.R., Phillips,C.A., Park, D.,Musgrove,M., 2011. DEEPSOIL 5.0 User Manual and Tutorial.

Herrmann, R.B. (1977) Recurrence relations, Earthquake notes 48, 47-49

Hinneburg, A. and Keim, D. (1998). An Efficient Approach to Clustering in Large Multimedia Databases with Noise. In Proceedings of KDD Conference.

Holt, W. E., Chamot, R. N., Le-Pichon, X., Haines, A. J., Shen, T. B., and Ren, J., 2000, Velocity field in Asia inferred from Quaternary fault slip rates and Global Positioning System observations: Journal of Geophysical Research, v. 105, p. 19,185-19,209.

Horike M, Zhao B, Kawase H (2001) Comparison of site response characteristics inferred from microtremor and earthquake shear waves. Bull Seismol Soc Am 91:1526-1536

Ischuk, A., Bendick, R., Rybin, A., Molnar, P., Khan, S.F., Kuzikov, S., Mohadjer, S., Saydullaev, U., Ilyasova, Z., Schelochkov, G., and Zubovich, A.V. (2013), Kinematics of the Pamir and Hindu Kush regions from GPS geodesy, J. Geophys. Res. Solid Earth, 118, 2408–241.

Jain, A.K., Murty, M.N., and Flyn, P.J. (1999). Data Clustering: A Review. ACM Computing Surveys, 31(3), 264–323.

Januzakov KJ, Omuraliev M, Omuralieva A, Ilyasov BI, Grebennikova VV (2003) Strong earthquakes of the Tien Shan (within the Kyrgyzstan territory and adjacent regions of the countries of Central Asia), Bishkek: Ilim, 216 pp, ISBN 5-8355-1335-6.

Joyner, W. B., R. E. Warrick, and T. E. Fumal (1981). The effect of Quaternary alluvium on strong ground motion in the Coyote Lake, California, earthquake of 1979, Bull. Seismol. Soc. Am. 71, 1333– 1349.

Kagan, Y.Y, Jackson, D.D. (1991) Long-term earthquake clustering. Geophysical Journal International. 104 (1): 117-133.

Kagan, Y.Y., and Jackson, D.D., (2002). Probabilistic forecasting of earthquakes, Geophys. J. Int. 143, 438-453.

Kalmetieva ZA, Mikolaichuk AV, Moldobekov BD, Meleshko AV, Jantaev MM, Zubovich AV, (2009) Atlas of Earthquakes in Kyrgyzstan, Central Asian Institute for Applied Geosciences, Bishkek, ISBN 978–9967-25–829-7

Karnik, V. (1968). Seismicity of the European Area. Part 1. Catalogue of earthquakes (1901-1955). Publishing house of the Czechoslovak Academy of Science.

Kaufman, L., Rousseeuw, P. J., (1987). Clustering by Means of Medoids, Statistical Data Analysis Based on The L1–Norm and Related Methods, edited by Y. Dodge, North-Holland, 405–416, (1987).

King, Stephanie A (Editor), Khalturin, Vitaly I (Editor), and Tucker, Brian E (Editor) 1996. Seismic Hazard and Building Vulnerability in Post-Soviet Central Asian Republics.

Kondorskaya N.V. and Shebalin N.V. (eds), 1982. New Catalogue of strong earthquakes in the USSR from Ancient Times through 1975.2nd edition, Boulder, Colorado, 608 pp.

Kondorskaya, N. V., Ulomov, V.I., Specialized earthquake catalogue for seismic zoning of Northern Eurasia // Main achievements of the Joint Institute of Physics of the Earth named after Schmidt O. Yu. for 1992-1996. Moscow. JIPE, 1996.Vol. 1. P. 108-109.

Konno K, Ohmachi T (1998) Ground-Motion Characteristics Estimated from Spectral Ratio between Horizontal and Vertical Components of Microtremor. Bull Seismol Soc Am 88: 228-241

Kyurskeyev A., Nurmagambetov, A., Sydykov, A., Mikhailova, N.N., and Shatsilov, V.I. (1993).Detailed seismic zoning of Almaty industrial area, NovostinaukiKazakhstana. Alma-Ata, Issue 1.

Lachet, C., Hatzfeld, D., Bard, P.-Y., Theodulidis, N., Papaioannou, C., and Savvaidis, A. (1996). Site effects and microzonation in the city of Thessaloniki (Greece): comparison of different approaches. *Bull. Seism. Soc. Am.*, **86,** no. 6, 1692–1703.

Lacoss, R. T., Kelly, E.J., and Toksöz, M.N. (1969). Estimation of seismic noise structure using array. Geophysics, 29, 21-38.

Lermo J, Chavez-Garcia FJ (1994) Are microtremors useful in site response evaluation?. Bull Seismol Soc Am 84:1350-1364

Lermo, J. F., and Chavez-Garcia, F. J. (1993). Site effect evaluation using spectral ratios with only one station. Bull. Seism Soc. Am. 83, 1574-1594.

Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J. (2010) Understanding of internal clustering validation measures. IEEE International conference on data mining.

LSTC, 2009. LS DYNA Keyword User's Manual – Release 971 R4. Livermore Software Technology Corporation, Livermore, California

Lukk, A.A., Yunga, S.L., Shevchenko, V.I. and Hamburger, M.W. (1995). Earthquake focal mechanisms, deformation state, and seismotectonics of the Pamir, Tien Shan region, Central Asia. Journal of Geophysical Research 100: doi: 10.1029/95JB02158. issn: 0148-0227

MacQueen JB (1967) Some Methods for classification and Analysis of Multivariate Observations, Proceeding of 5th Berkley Symposium on Mathematical Statistics and Probability, Berkley, University of California Press 1:281-297

McGuire, R.K., (1976) FORTRAN computer program for seismic risk analysis, U.S. Geological Survey Open-File Report 76-67.

McGuire, R.K., 2004. Seismic Hazard and Risk Analysis. EERI Publication No.MNO-10.221 pp. Earthquake Engineering Research Institute, Oakland, CA.

Medvedev, S., W. Sponheuer, and V. Kárník (1964). NeueseismischeSkala Intensity scale of earthquakes, 7.Tagung der Europäischen Seismologischen Kommission vom 24.9. bis 30.9.1962. In: Jena, Veröff. Institut für Bodendynamik und Erdbebenforschung in Jena, vol 77. Deutsche Akademie der Wissenschaften zu Berlin, pp 69-7

Merz, H. A., and C. A. Cornell (1973). Seismic risk based on a quadratic magnitude-frequency law, Bull. Seism. Soc. Am. 63, 1999-2006

Mikhailova N.N (1996). Seismic hazard in quantity characteristics of strong ground motions (on the example of Almaty). Doctoral dissertation on physics and mathematics.

Mikhailova, N. N., and Kurskeev, A. K. 1995. Present status of the network for seismic observation in Kazakhstan. Journal of earthquake prediction research. V. 4, N 4, p. 497-506

Mikhailova, N. N., Mukambayev, A., Aristova, I. L., Kulikova, G., Ullah, S., Pilz, M., Bindi, D.(2015). Central Asia Earthquake catalogue from ancient time to 2009. Annals of geophysics,

Mikhailova, N.N. (2009) Seismic Risk Assessment in Central Asia: Final Project Activity Report on the work performed from 02.01.2006 to 04.30.2009 / Institute of Geophysical Researches NNC RK

Molina, S., Lindholm, C.D., Bungum, H.: Probabilistic seismic hazard analysis: zoning free versus zoning methodology. Bolletino di GeofisicaTeoricaApplicata42 (1-2), 19-39, 2001.

Molnar, P., and Tapponnier, P. 1975. Cenozoic tectonic of Asia: Effects of a continental collision. Science, vol. 189 (4201).

Molnar, P., and Tapponnier, P. 1978. Active tectonics of Tibet. J. Geophys. Res., 83, 5361-5375

Mooney, D.W., Ritsema, J., Hwang, Y.K (2012) Crustal seismicity and the earthquake catalog maximum moment magnitude (Mcmax) in stable continental regions (SCRs): Correlation with the seismic velocity of the lithosphere. Earth and Planetary Science Letters. 357-358 (2012) 78-83.

Mushketov I.V and Orlov A.P., 1893.Catalogue of earthquakes of Russian Empire (KatalogzemletryaseniyRossiyskoiImperii).Notes of Russian Geographic Soc., St. Peterburg, 26, 582 pp. (in Russian).

Mushketov, D.I. (1933). Opyt seysmicheskogo rayonirovanija S.S.S.R. Tr. Seismol.Inst. Akad. Nauk S.S.S.R., 33, 1-17

Musson, R. M. W. (2005). Intensity attenuation in the UK, J. Seism. 9, 73-86.

Musson, R.M.W. (1999). Probabilistic seismic hazard maps for the north Balkan region. Annali di Geofisica 42(2), 1109-1124

Nelson, M. R., R. McCaffrey, and P. Molnar (1987). Source parameters for 11 earthquakes in the Tien Shan, central Asia, determined by P and SH waveform inversion, J. Geophys. Res. 92: 12,629–12,648.

Nurmagambetov A., N. Mikhailova, and W. Iwan (1999). Seismic hazard of the Central Asia region, in Seismic hazard and building vulnerability in post-Soviet Central Asian republics, S.A. King, V. I. Khalturinand B. E. Tucker (eds), Kluwer Academic Publishers, Nederlands, 1-43.

Ordaz M., Arroyo D. (2016) On uncertainties in PSHA. Earthquake Spectra. DOI: 10.1193/052015EQS075M

Otto, S.C., 1997. Mesozoic-Cenozoic history of deformation and petroleum systems in sedimentary basins of Central Asia: implications of collisions on the Eurasian margin. Pet. Geosci. 3, 327-341.

Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., Henshaw, P., Butler, L., Nastasi M., ,Panzeri, L., Simionato, M., Vigano, D. (2014) OpenQuake Engine: An Open Hazard (and Risk) Software for the Global Earthquake Model. Seis. Res. Let. v. 85 no. 3 p. 692-702

Parolai S, Bormann P, Milkereit C (2001) Assessment of the Natural Frequency of the Sedimentary Cover in the Cologne Area (Germany) using Noise Measurements. J Earthq Eng 5:541-564

Parolai S, Orunbaev S, Bindi D, Strollo A, Usupaev S, Picozzi M, Di Giacomo D, Augliera P, D'Alema E, Milkereit C, Moldobekov B, Zschau J (2010) Site Effects Assessment in Bishkek (Kyrgyzstan) Using Earthquake and Noise Recording Data. Bull Seismol Soc Am 100:3068-3082

Parolai S, Picozzi M, Richwalski SM, Milkereit C (2005) Joint inversion of phase velocity dispersion and H/V ratio curves from seismic noise recordings using a genetic algorithm, considering higher modes. Geophys Res Lett 32, 1, L01303.

Parolai S, Richwalski S, Milkereit C, Bormann P (2004) Assessment of the stability of H/V spectral ratios from ambient noise and comparison with earthquake data in the Cologne area (Germany). Tectonophysics 390:57-73

Parolai, S. (2012) Chapter 14, Investigation of site response in urban areas by using earthquake data and seismic noise, DOI: 10.2312/GFZ.NMSOP-2_ch14

Parolai, S., and Richwalski, S. M. (2004). The importance of converted waves in comparing H/V and RSM site responses. Bull. Seism. Soc. Am., 94, 304–313.

Parolai, S., Cara, F., Bindi, D.,Pacor, F. (2009) Empirical site-specific response-spectra correction factors for the Gubbio basin (central Italy),Soil Dynamics and Earthquake Engineering 29 (2009) 546– 552

Parolai, S., Grünthal, G., Wahlström, R. (2007) Site specific response spectra from the combination of microzonation with probabilistic seismic hazard assessment – An example for the Cologne (Germany) Area. Soil Dynamics and Earthquake Engineering (27):49-59

Parolai, S., Moldobekov, B., Mikhailova, N., Pilz, M., Ullah, S., (2014) Report on the shallow geological investigation in Bishkek (Kyrgyzstan). Scientific Technical Report 14/12, DeutchesGeoforschungZentrum (GFZ) Potsdam.

Parolai, S., Orunbaev, S., Bindi, D., Strollo, A., Usupaev, S., Picozzi, M., Di Giacomo, D., Augliera, P., D'Alema, E., Milkereit, C., Moldobekov, B., Zschau, J., (2010) Site effects assessment in Bishkek (Kyrgyzstan) using earthquake and noise recording data. Bull. Seism. Soc. Am. 100:3068–3082

Parolai, S., P. Bormann, C. Milkereit (2001): Assessment of the natural frequency of sedimentary cover in the Cologne area (Germany) using noise measurements, J. Earthq. Eng. 5, 541-564

Parolai, S., Picozzi, M., Richwalski, S. M., Milkereit, C. (2005): Joint inversion of phase velocity dispersion and H/V ratio curves from seismic noise recordings using a genetic algorithm, considering higher modes. - Geophysical Research Letters, 32, 1, L01303.

Parolai, S., Richwalski, S. M., Milkereit, C. (2006) S-wave velocity profiles for earthquake engineering purposes for the Cologne Area (Germany). Bull. Seism. Soc. Am. 4, 65-94

Patriat, P. and Achache, J., 1984. India-Eurasia collision chronology has implications for crustal shortening and driving mechanisms of plates. Nature, 31 1, 615-621.

Petersen, M.D., Bryant, W.A., Cramer, C.H., Reichle, M.S., Real, C.R. (1997) Seismic ground-motion hazard mappinig incorporating site effects for Los Angeles, Orange, and Ventura counties, California: A geographical information system application. Bull. Seis. Soc. Am. 87(1)249:255

Picozzi M, Strollo A, Parolai S, Durukal E, Özel O, Karabulut S, Zschau J, Erdik M (2009) Site characterization by seismic noise in Istanbul, Turkey. Soil Dyn Earthquake Eng 29:469-482

Picozzi, M. (2005) Joint inversion of phase velocity dispersion and H/V ratio curves from seismic noise recordings. PhD thesis, Universita degli Studi di Siena, Italy.

Pilz M, Parolai S, Leyton F, Campos J, Zschau J (2009) A Comparison of site response techniques using earthquake data and ambient seismic noise analysis in the large urban areas of Santiago de Chile. Geophys J Int 178:713-728

Pilz, M., Parolai, S., Bindi, D. (2013). Three-dimensional passive imaging of complex seismic fault systems: evidence of surface traces of the Issyk-Ata fault (Kyrgyzstan). Geophysical Journal International. 194(3):1955-1965

Pilz, M., Parolai, S., Picozzi, M., Zschau, J. (2011). Evaluation of proxies for seismic site conditions in large urban areas: the example of Santiago de Chile. - Physics and Chemistry of the Earth, 36, 16, 1259-1266.

Rautian, T., Khalturin, V., Fujita, K., Mackey, K., and Kendall, A. (2007) Origins and methodology of the Russian energy K-class system and its relationship to magnitude scales: Seis. Res. Lett.,78, 579-590.

Rautian, T., Leith, W., 2002.Composite Regional Catalogs of Earthquakes in the Former Soviet Union. U.S. Geological Survey Open File Report02-500.

Reiter, L.: 1990, Earthquake Hazard Analysis, Columbia University Press, New York

Riznichenko, J. V. 1959. On quantitative determination and mapping of seismic activity. Annals of Geophysics, 12(2).

Riznichenko, Y. (1966). Calculation of ground shaking of the surface points caused by earthquakes in adjustment area, Physics of the Earth, 5, 2-32.

Rodriguez VH, Midorikawa S (2002) Applicability of the H/V spectral ratio of microtremors in assessing site effects on seismic motion. Earthquake Eng Struct Dyn 31(2):261–279

Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53-65

Sadigh, K., Chang, C.Y., Egan, J.A., Makdisi, F., Youngs, R.R., (1997) Attenuation relationships for shallow crustal earthquakes based on California strong motion data. Seism. Res. Lett. 68,180-189

Safak, E. (2001). Local site effects and dynamic soil behaviour. Soil Dynamics and Earthquake Engineering, Elsevier Science Ltd., Vol. 21, pp.453-458.

Schnabel, P.B., Lysmer, J., Seed, H.B., 2012. Computer Program SHAKE: A Computer Program for Earthquake Response Analysis of Horizontally Layered Sites. University of California, Berkeley, California.

Schurr, B., Ratschbacher, L., Sippl, C., Gloaguen, R., Yuan, X., Mechie, J. (2014) Seismotectonics of the Pamir. - Tectonics, 33, 8, p. 1501-1518

Senior Seismic Hazard Analysis Committee (SSHAC) (1997). Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts NUREG/CR-6372, U. S. Nuclear Regulatory Commission.

Sheikholeslami, C., Chatterjee, S., and Zhang, A. (1998). WaveCluster: A-MultiResolution Clustering Approach for Very Large Spatial Database. In Proceedings of 24th VLDB Conference, New York, USA.

Simpson, D. W., and Leith, W. (1985) The 1976 and 1984 Gazli, USSR, earthquakes –Were they induced?. Bull. Seis. Soc. Am. 75(5):1465-1468

Steidl, J.H. (2000) Site response in southern California for Probabilistic Seismic Hazard Analysis. Bull. Seis. Soc. Am. 90(6B), 149-169

Stepp, J.C. (1973) Analysis of completeness of the earthquake sample in the Puget Sound area. In: Harding ST (ed) Seismic zoning. NOAATech Report ERL 267-ESL30, Boulder

Stock, C., Smith, E.G.C.,(2002) Adaptive kernel estimation and continuous probability representation of historical earthquake catalogs. Bull. Seism. Soc. Am. 92(3):904-912

Storchak, D.A., D. Di Giacomo, I. Bondár, E. R. Engdahl, J. Harris, W.H.K. Lee, A. Villaseñor and P. Bormann, (2013). Public Release of the ISC-GEM Global Instrumental Earthquake Catalogue (1900-2009). Seism. Res. Lett., 84, 5, 810-815, doi: 10.1785/0220130034.

Strollo A, Parolai S, Bindi D, Chiuzzi L, Pagliuca R, Mucciarelli M, Zschau J (2012) Microzonation of Potenza (Southern Italy) in terms of spectral intensity ratio using joint analysis of earthquakes and ambient noise. Bull Earthquake Eng. doi:10.1007/s10518-011-9256-4

Strollo A, Parolai S, Jäckel KH, Marzorati S, Bindi D (2008) Suitability of Short-Period Sensors for Retrieving Reliable H/V Peaks for Frequencies Less Than 1 Hz. Bull Seismol Soc Am 98:671-681

Tapponnier, P. and P. Molnar (1979).Active faulting and Cenozoic tectonics of the Tien Shan, Mongolia, and Baykal regions, J. Geophys. Res. 84: 3425–3459.

Tapponnier, P., Mattauer, M., Proust, F., and Cassaigneau, C. (1981) MessozoicOphiolites, Sutures and large scale tectonic movements in Afghanistan. Earth and Planetary Science Letters, 52 (1981) 355-371.

Tatevossian, R. 2004. History of earthquake studies in Russia. Annals of Geophysics, vol 47, N. 2/3.

Theodoridis, S. and Koutroubas, K. (1999). Pattern Recognition. Academic Press.

Trifonov VG (1996) World map of active faults in Eurasia: principles, methods and results. J Earthq Prediction Res 5, 326–347

Tselentis GA, Paraskevopoulos P (2011) On the Use of Kohonen Neural Networks for Site Effects Assessments by means of H/V Weak-Motion Spectral Ratio: Application in Rio-Antirrio (Greece). Bull Seismol Soc Am 101:579-595

Uhrhammer, R. (1986). Characteristics of Northern and Central California Seismicity. Earthquake Notes 57, 21.

Ulomov, V.I., Shyumilina, L.S. A set of maps of general seismic zoning of the Russian Federation territory. GSZ – 97. Scale 1:8000000 // Explanatory note and list of cities and settlements located at seismically active regions. The map on 4 sheets. Edited by Strakhov V.N. and Ulomov V.I.). Moscow. 2000. 57 p.

Ulomov, V.I., The GSHAP Region 7 working group, (1999). Seismic hazard of Northern Eurasia. Annali di Geofisica 42, 1023–1038

Van Stiphout, T., Zhuang, J., Marsan, D. (2012), Seismicity declustering, Community online resource for statistical seismicity analysis, doi:10.5078/corssa-52382934. Available at http://www.corssa.org

Wald, D. J., and Allen, T. I. (2007). Topographic slope as a proxy for seismic site conditions and amplification, Bull. Seismol. Soc. Am. 97, 1379–1395; doi: 10.1785/0120060267.

Wang, W., Yang, J., and Muntz, R. (1997). STING: A Ststistical Information Grid Approach to Spatial Data Mining. In Proceedings of 23rd VLDB Conference.

Weatherill, G. A. (2014) OpenQuake Hazard Modeller's Toolkit - User Guide. Global Earthquake Model (GEM). Technical Report

Weichert, D. H. (1980). Estimation of earthquake recurrence parameters for unequal observation periods for different magnitudes. Bull. Seism. Soc. Am. 70, 1337-1356

Wessel P, Smith WHF (1991) Free software helps map and display data. Eos Trans AGU 72:441-461

Wiggins, J. H., Jr. (1964). Effect of site conditions on earthquake intensity, Proc. Am Soc. Civil Eng. J. Structural Div. 90, no. ST2, 279–313.

Wimer, S, Giardini, D., Fäh, D., Deichmann, N. and Sellami, S. (2009), Probabilistic seismic hazard assessment of Switzerland: best estimate and uncertainties, J Seismol, Vol. 13: 449-478.

Woo, G. (1996) Kernel estimation methods for seismic hazard area source modeling. Bull. Seism. Soc. Am. 86(2):353-362

Wu, J., Xiong, H., and Chen J. (2009) "Adapting the right measures for k-means clustering," in KDD, pp. 877–886.

Yamanaka, H., and H. Ishida (1996). Application of Generic algorithms to an inversion of surface-wave dispersion data. Bull. Seism. Soc. Am., 86, 436-444.

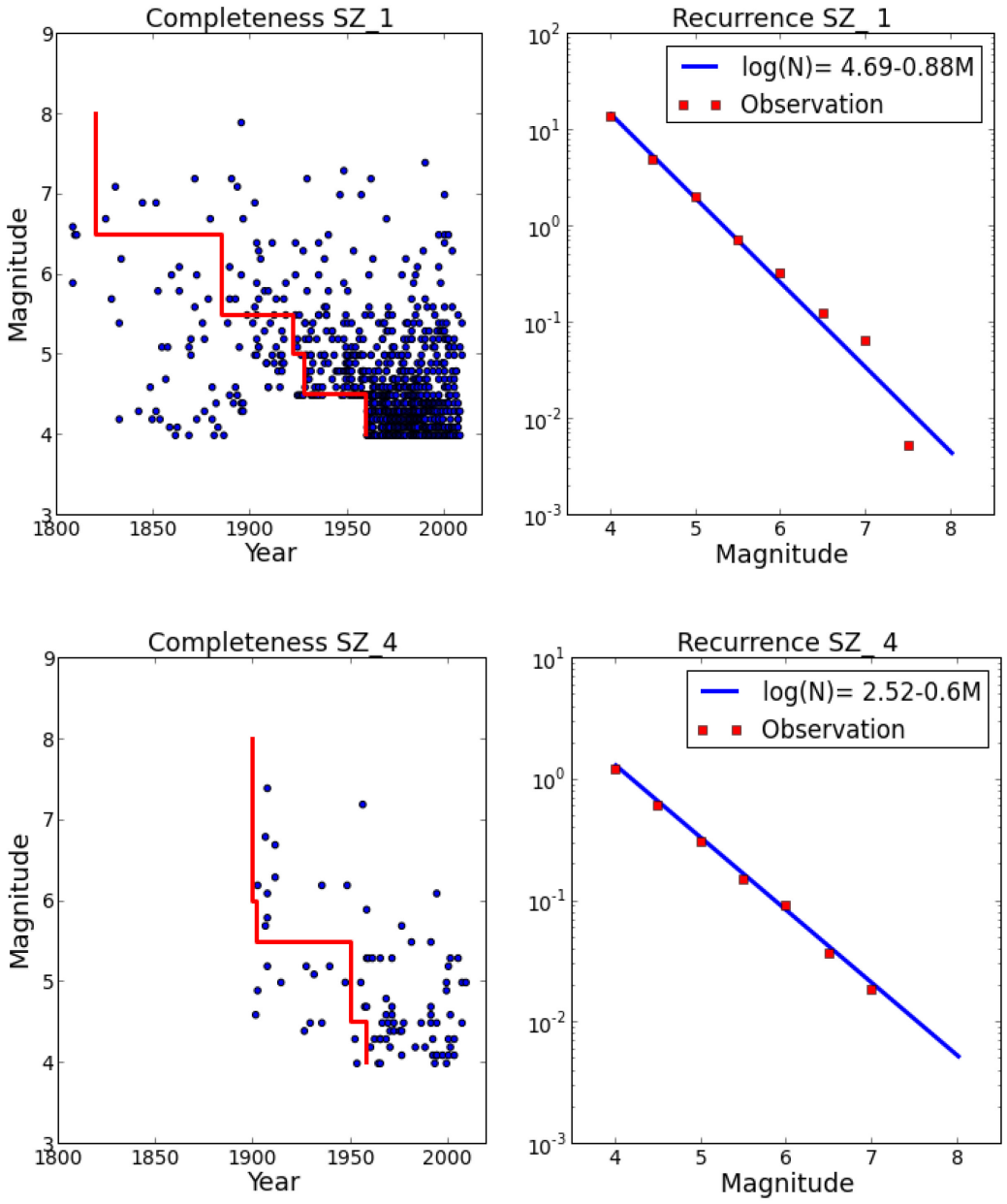
Zhang, T., Ramakrishnman, R., and Linvy, M. (1996). BIRCH: An Efficient Method for Very Large Databases. ACM SIGMOD, Montreal, Canada.

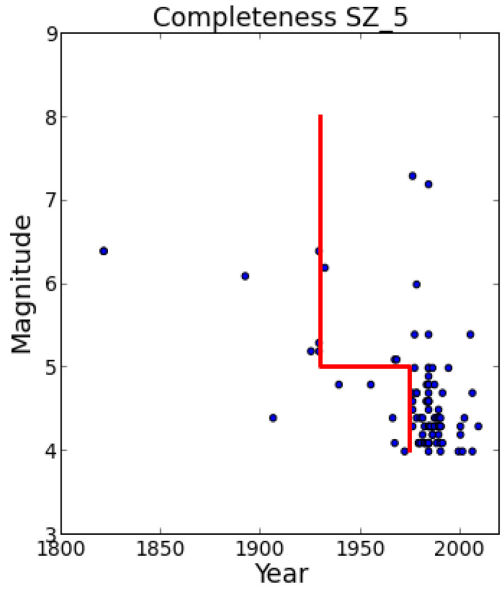
Zonenshain, L., Kuzmin, M.I.,andNatapov, L.M., 1990. Geology of the U.S.S.R.: A Plate Tectonic Synthesis. American Geophysical Union, Geodynamics Series, 21: 1-242

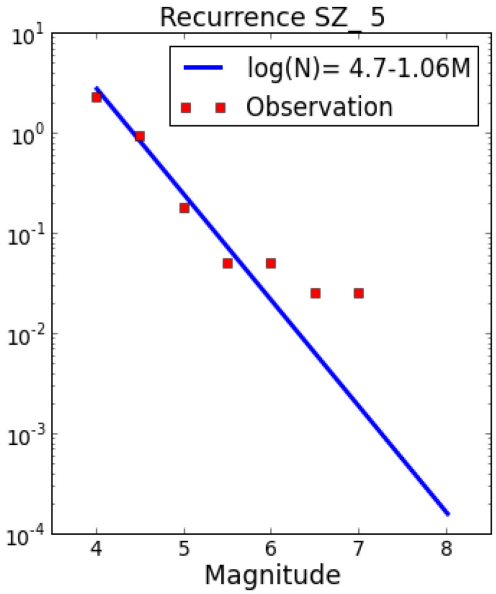
Zubovich, A.V., Wang, X., Scherba, Y.G., Schelochkov, G.G., Reilinger, R., Reigber, C., Mosienko, O.I., Molnar, P., Michajljow, W., Makarov, V.I., Li, J., Kuzikov, S.I., Herring, T.A., Hamburger, M.W., Hager, B.H., Dang, Y., Bragin, V.D. and Beisenbaev, R.T. (2010). GPS velocity field for the Tien Shan and surrounding regions. Tectonics 29: doi: 10.1029/2010TC002772. issn: 0278-7407.

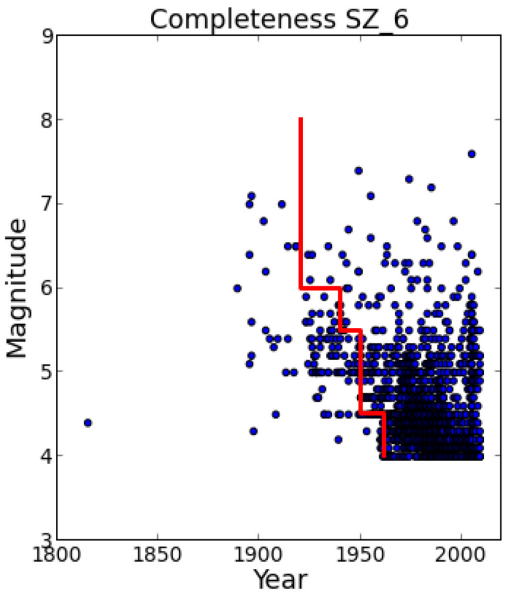
APPENDIX A: Completeness and recurrence plots for super zones

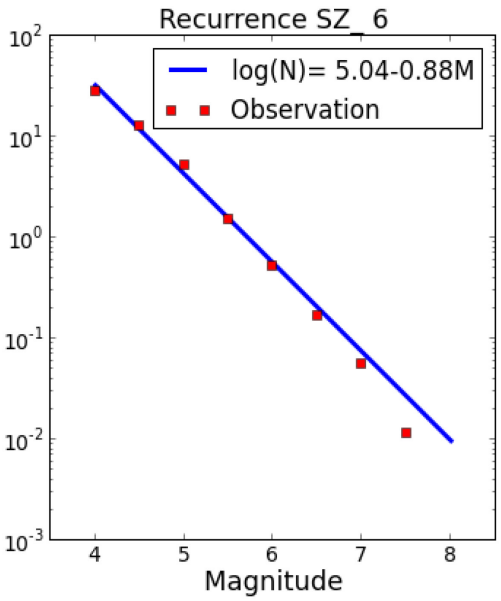
Recurrence and completeness analyses plots for super zones defined in Figure 6.13, Table 6.1 and Table 6.2. Figures are provided only for the active zones. The left side figures show the completeness analysis using Stepp1973 method. The blue dots represent the events, while the red line shows the completeness of magnitudes in time. The Right side figures show Gutenberg-Richter plots. The red squares show the observations from earthquake catalogue, while the blue line shows the recurrence relationship obtained by using the Weichert (1980) method. The number of each super zone is given on top of each figure.

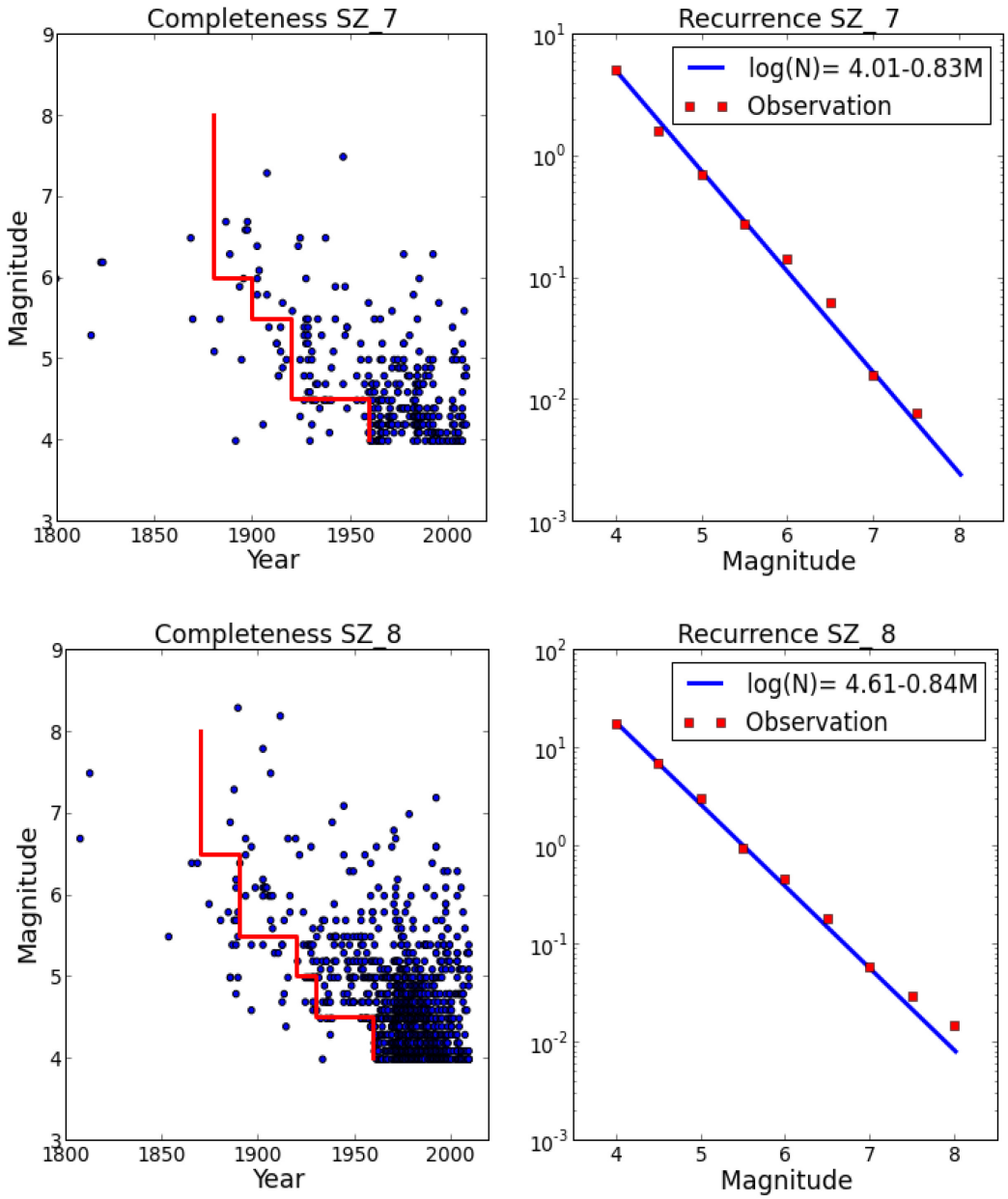


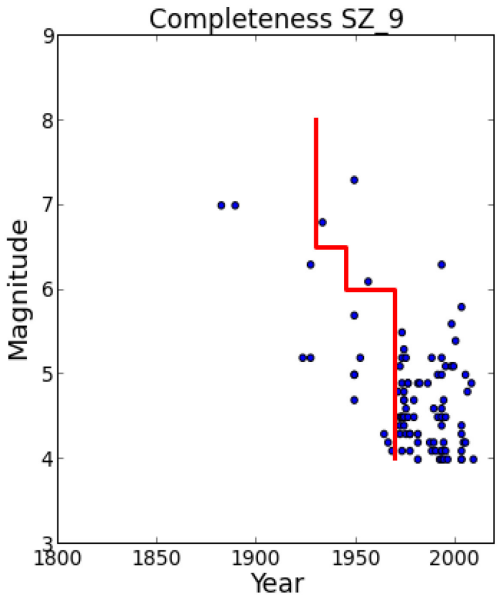

Completeness SZ_5
Magnitude
Year
9
8
7
6
5
4
3
1800
1850
1900
1950
2000

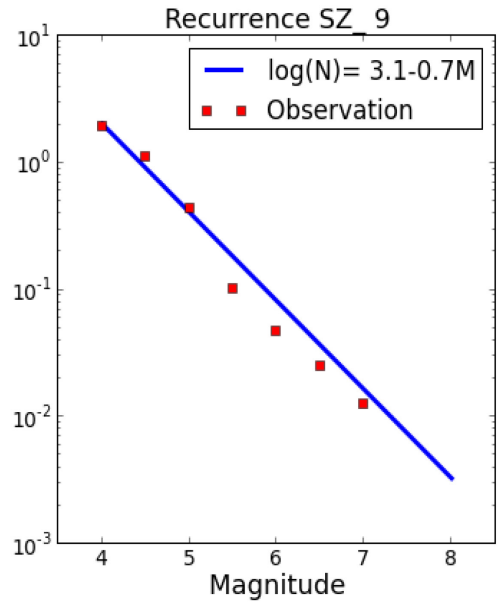

Recurrence SZ_ 5
log(N)= 4.7-1.06M
Observation
Magnitude
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
4
5
6
7
8

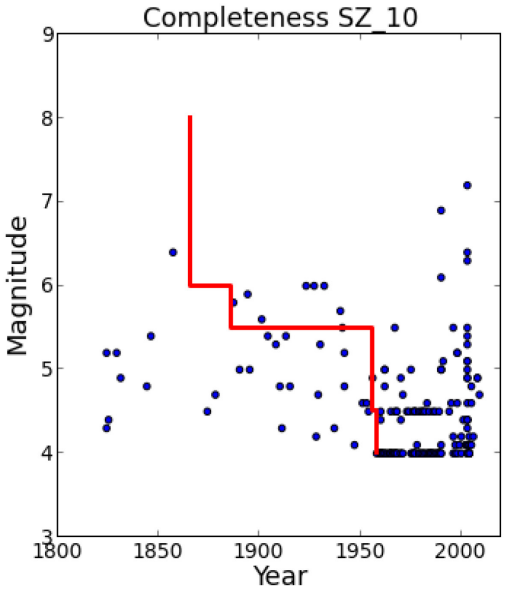

Completeness SZ_6
Magnitude
Year
9
8
7
6
5
4
3
1800
1850
1900
1950
2000

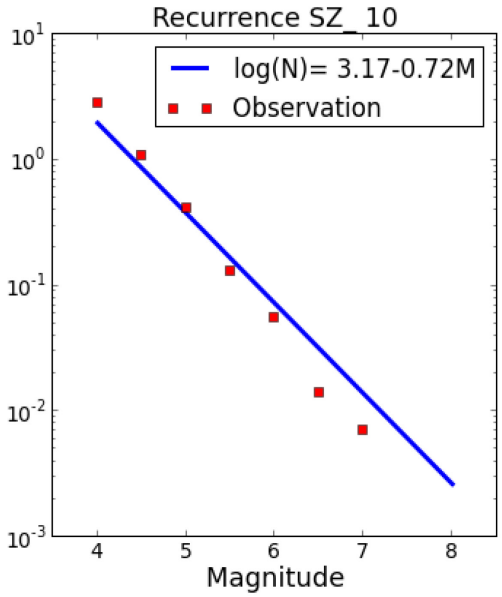

Recurrence SZ_ 6
log(N)= 5.04-0.88M
Observation
Magnitude
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
4
5
6
7
8


Completeness SZ_7
Magnitude
Year
Recurrence SZ_ 7
log(N)= 4.01-0.83M
Observation
Magnitude
Completeness SZ_8
Magnitude
Year
Recurrence SZ_ 8
log(N)= 4.61-0.84M
Observation
Magnitude


Completeness SZ_9
Magnitude
Year
9
8
7
6
5
4
3
1800
1850
1900
1950
2000

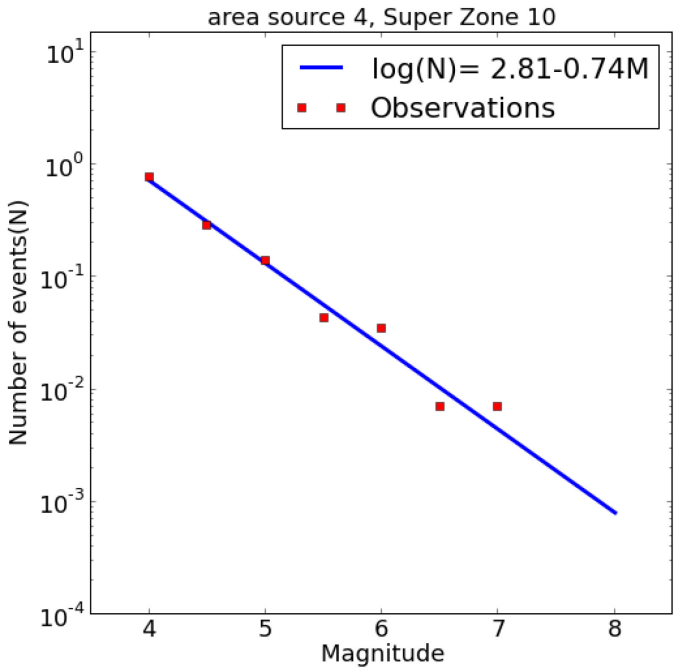

Recurrence SZ_ 9
log(N)= 3.1-0.7M
Observation
Magnitude
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
4
5
6
7
8

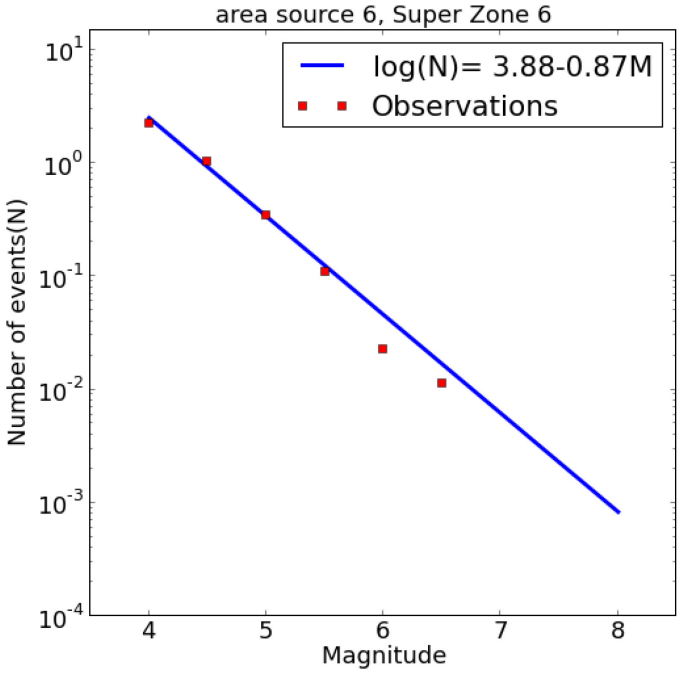

Completeness SZ_10
Magnitude
Year
9
8
7
6
5
4
3
1800
1850
1900
1950
2000

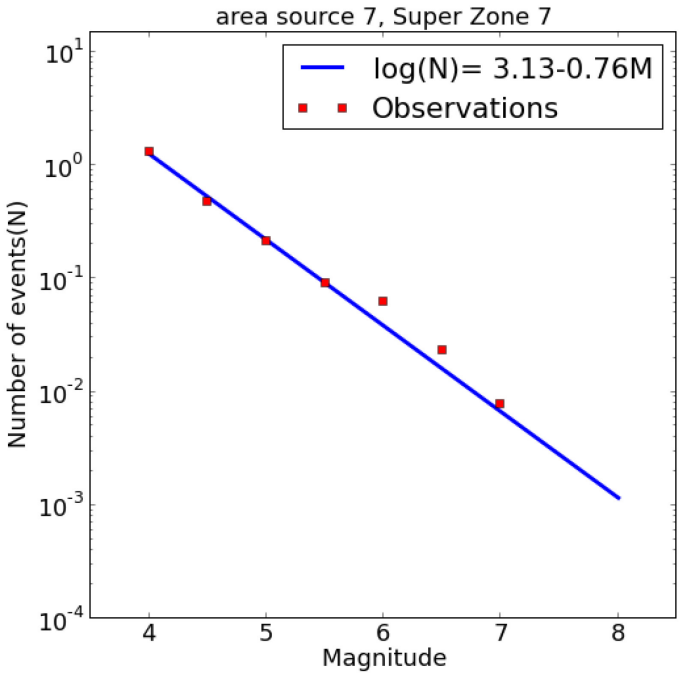

Recurrence SZ_ 10
log(N)= 3.17-0.72M
Observation
Magnitude
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
4
5
6
7
8

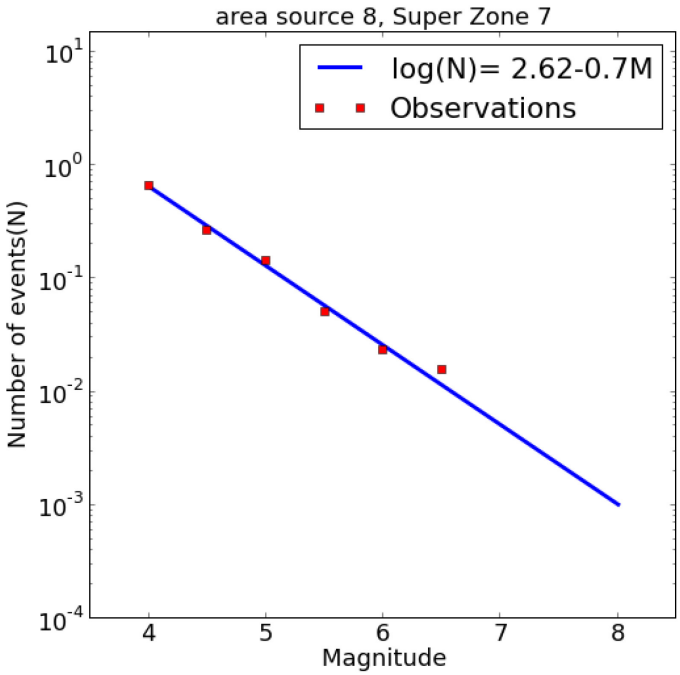
APPENDIX B: recurrence plots for area sources

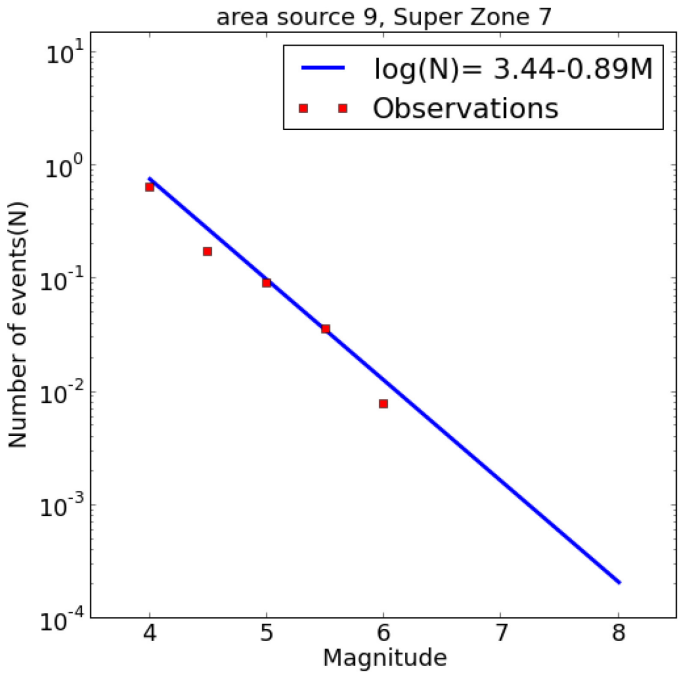
Recurrence analyses plots for area sources, defined in Figure 6.13 and Table 6.3. Figures are shown only for those sources having at least 20 events inside. The red squares show the observations from earthquake catalogue, while the blue line shows the recurrence relationship obtained by using the Weichert (1980) method. The number of each area source and the super zone to which it belongs, is given on top of each figure.

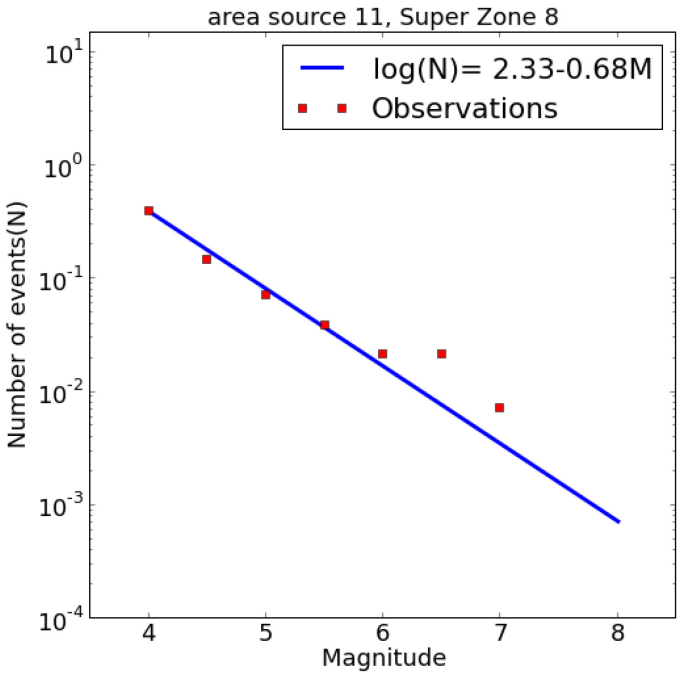


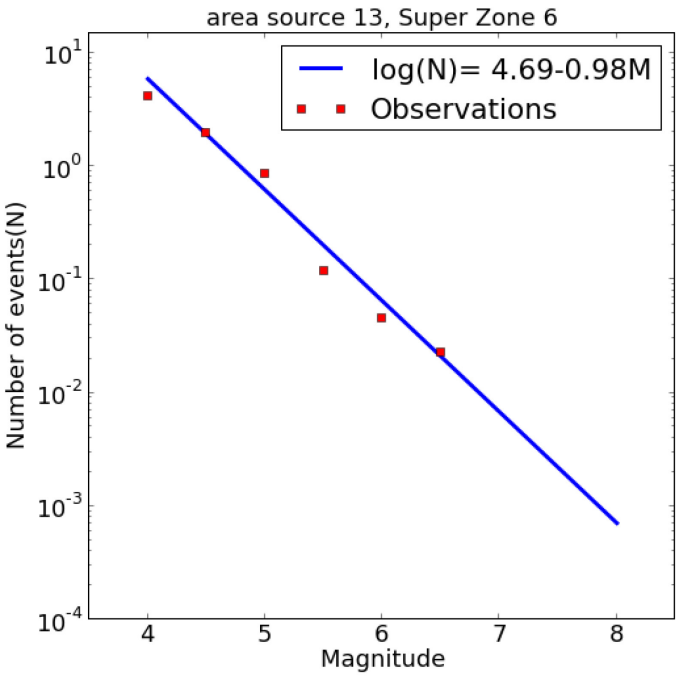


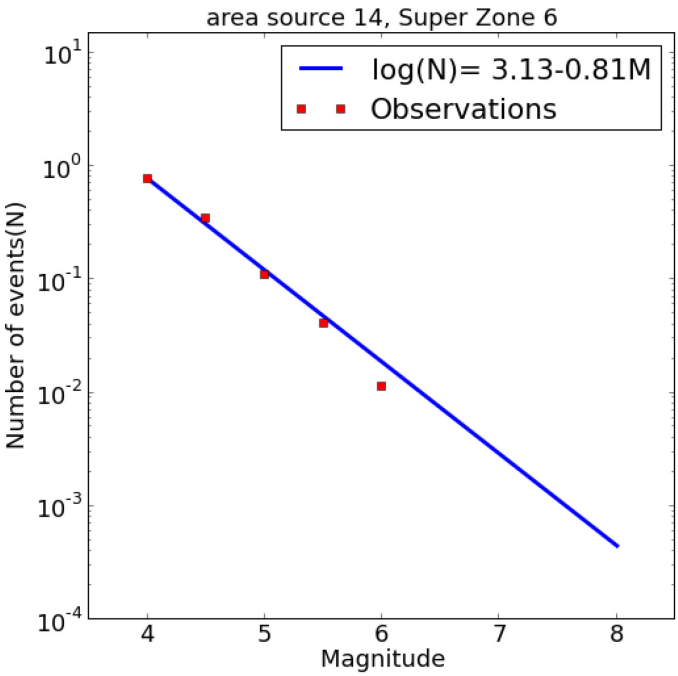


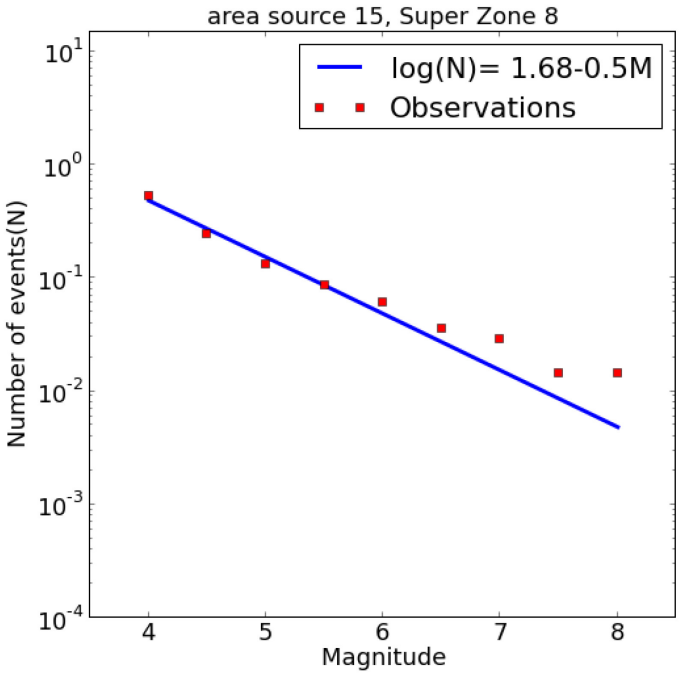


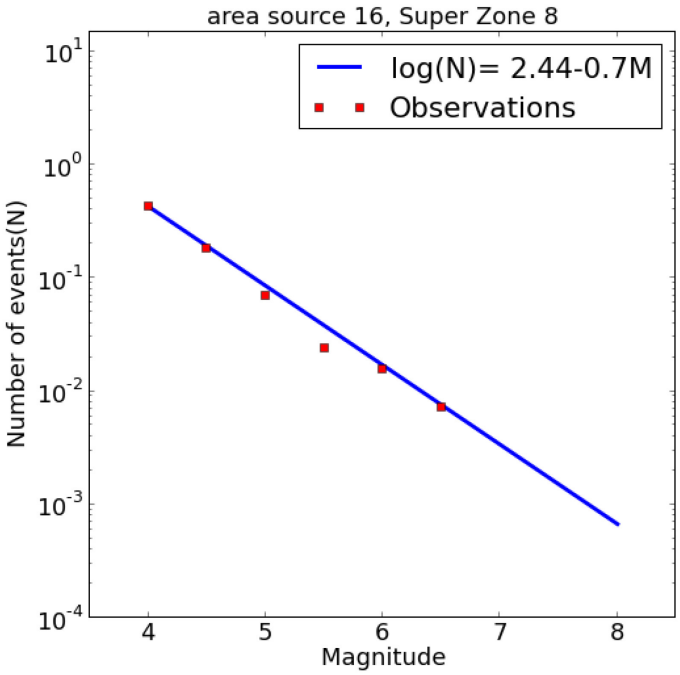

area source 9, Super Zone 7
log(N)= 3.44-0.89M
Observations
Number of events(N)
Magnitude

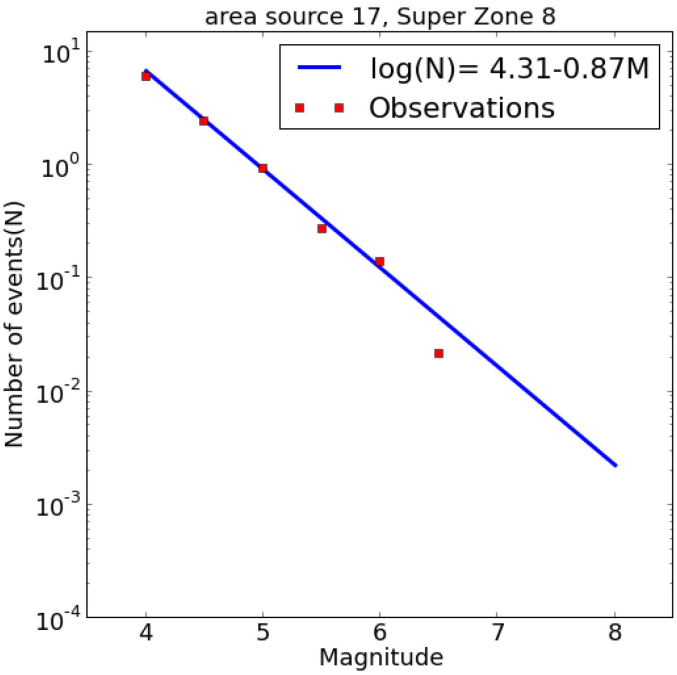

area source 11, Super Zone 8
log(N)= 2.33-0.68M
Observations
Number of events(N)
Magnitude

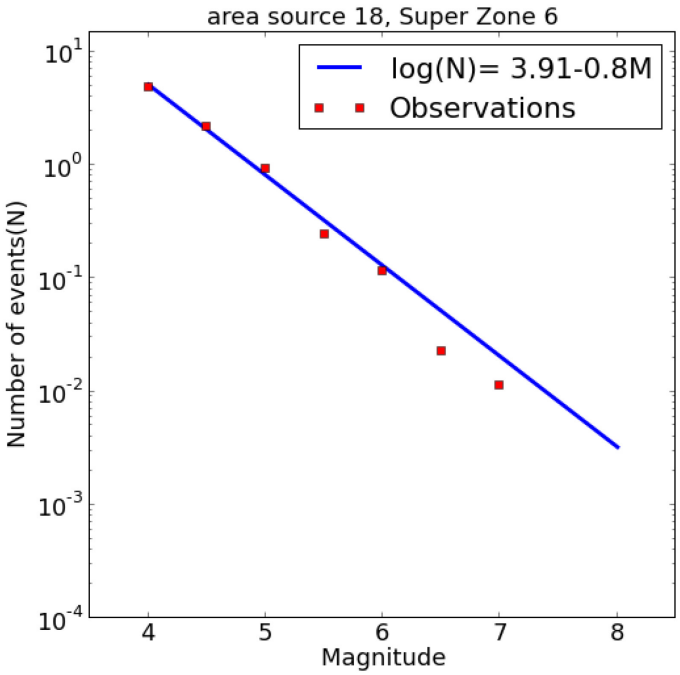

area source 13, Super Zone 6
log(N)= 4.69-0.98M
Observations
Number of events(N)
Magnitude

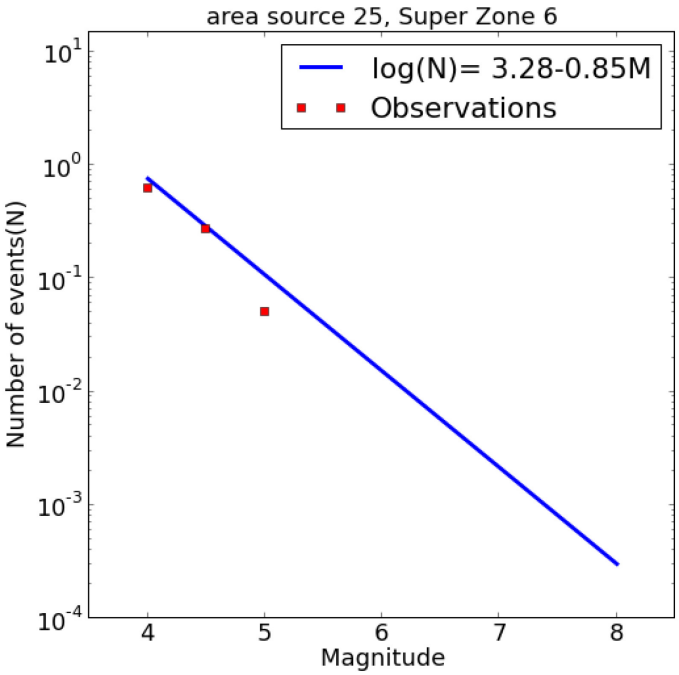

area source 14, Super Zone 6
log(N)= 3.13-0.81M
Observations
Number of events(N)
Magnitude

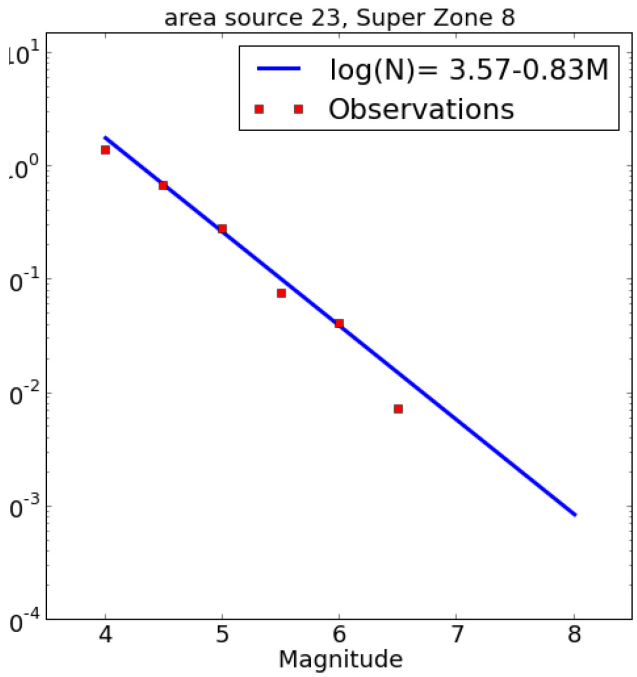

area source 15, Super Zone 8
log(N)= 1.68-0.5M
Observations
Number of events(N)
Magnitude

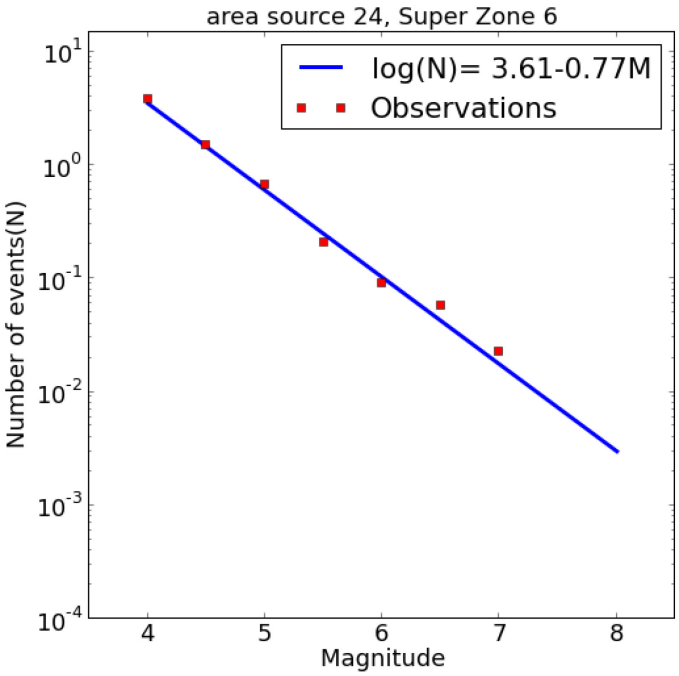

area source 16, Super Zone 8
log(N)= 2.44-0.7M
Observations
Number of events(N)
Magnitude

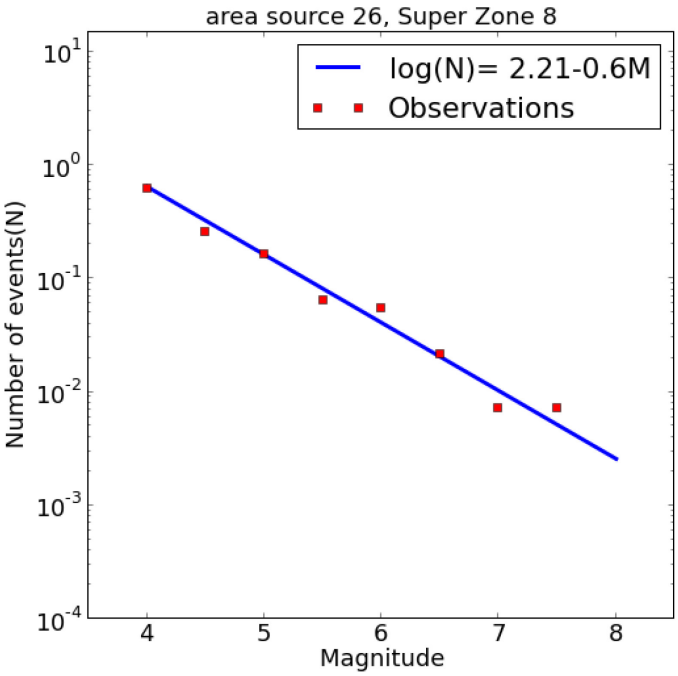

area source 17, Super Zone 8
log(N)= 4.31-0.87M
Observations
Number of events(N)
Magnitude

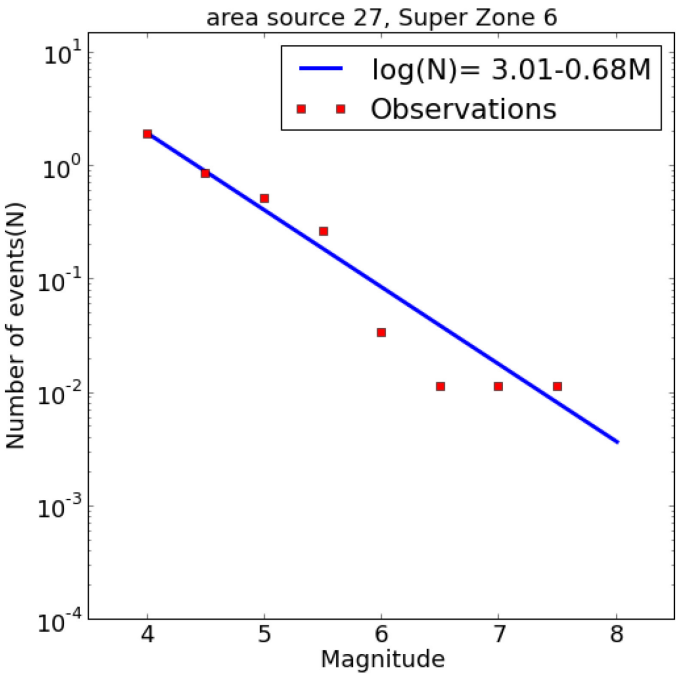

area source 18, Super Zone 6
log(N)= 3.91-0.8M
Observations
Number of events(N)
Magnitude

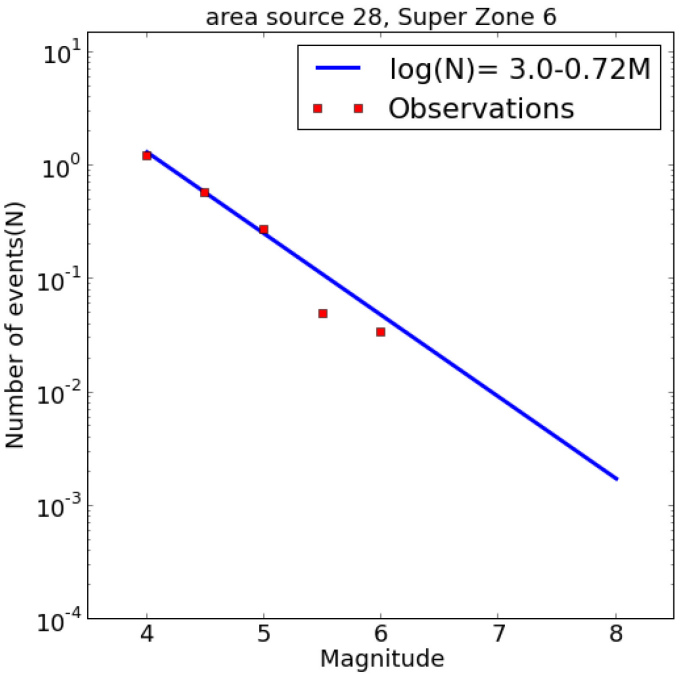

area source 25, Super Zone 6
log(N)= 3.28-0.85M
Observations
Number of events(N)
Magnitude

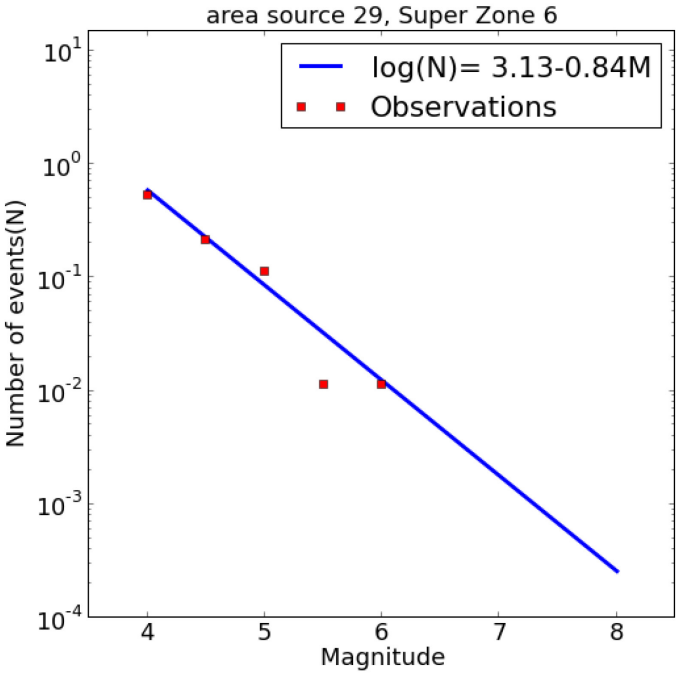

area source 23, Super Zone 8
log(N)= 3.57-0.83M
Observations
Magnitude

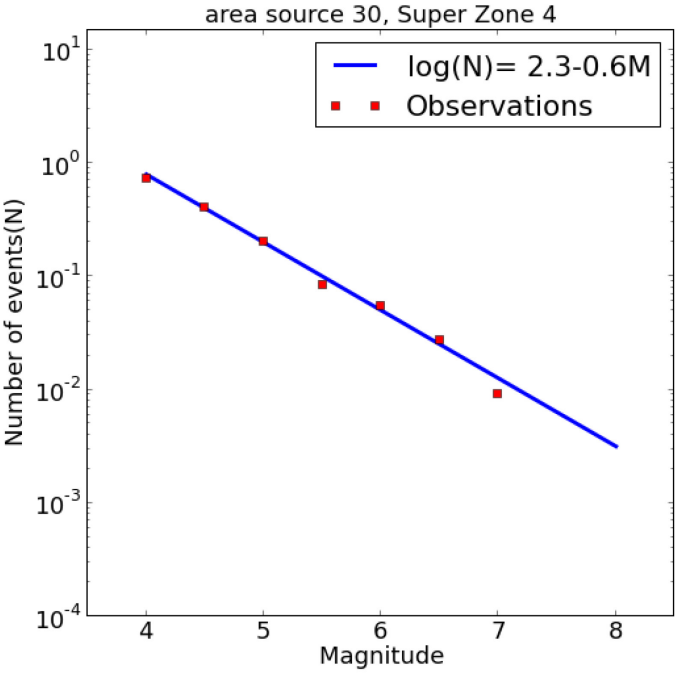

area source 24, Super Zone 6
log(N)= 3.61-0.77M
Observations
Number of events(N)
Magnitude

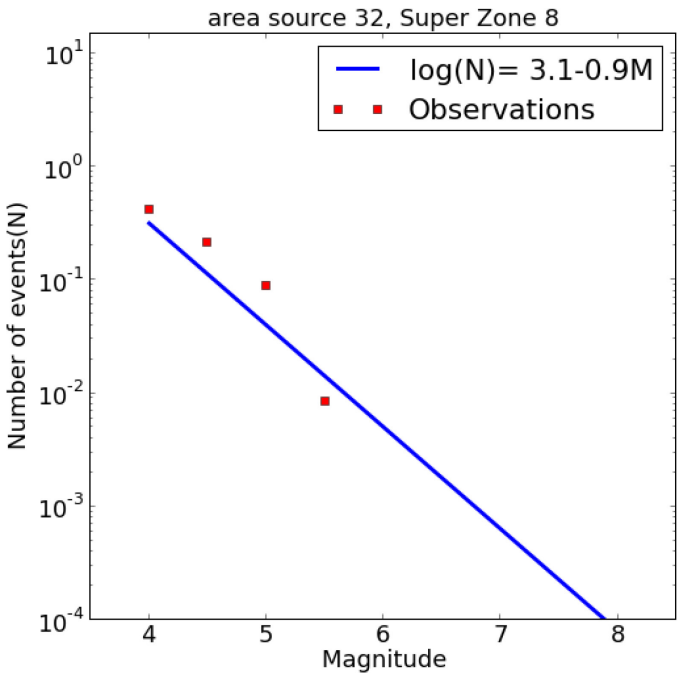

area source 26, Super Zone 8
log(N)= 2.21-0.6M
Observations
Number of events(N)
Magnitude

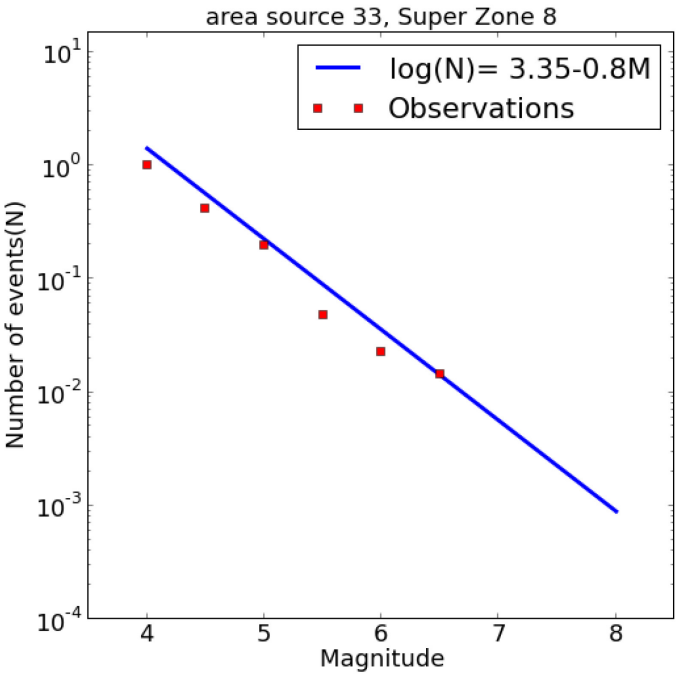

area source 27, Super Zone 6
log(N)= 3.01-0.68M
Observations
Number of events(N)
Magnitude

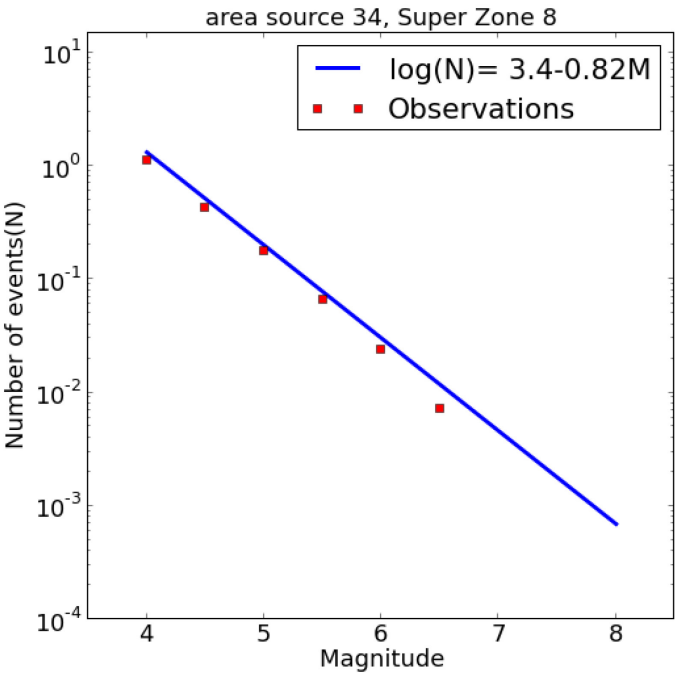

area source 28, Super Zone 6
log(N)= 3.0-0.72M
Observations
Number of events(N)
Magnitude

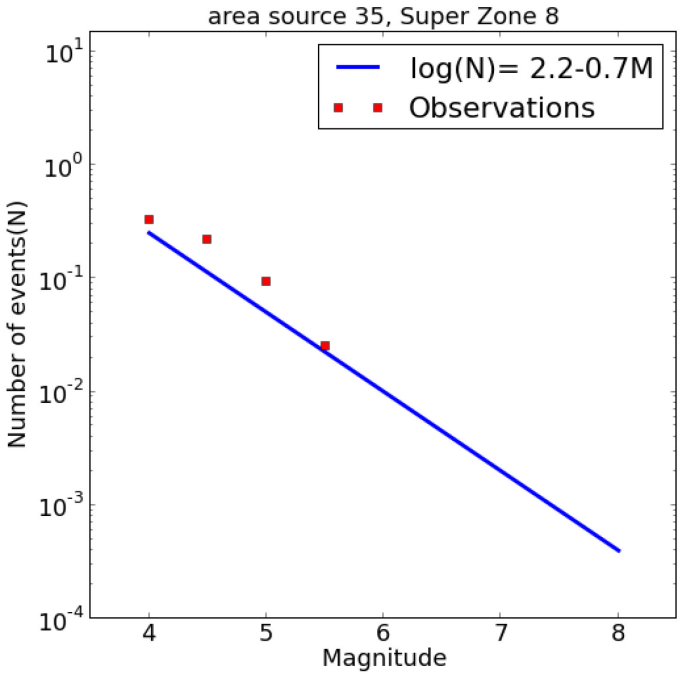

area source 29, Super Zone 6
log(N)= 3.13-0.84M
Observations
Number of events(N)
Magnitude

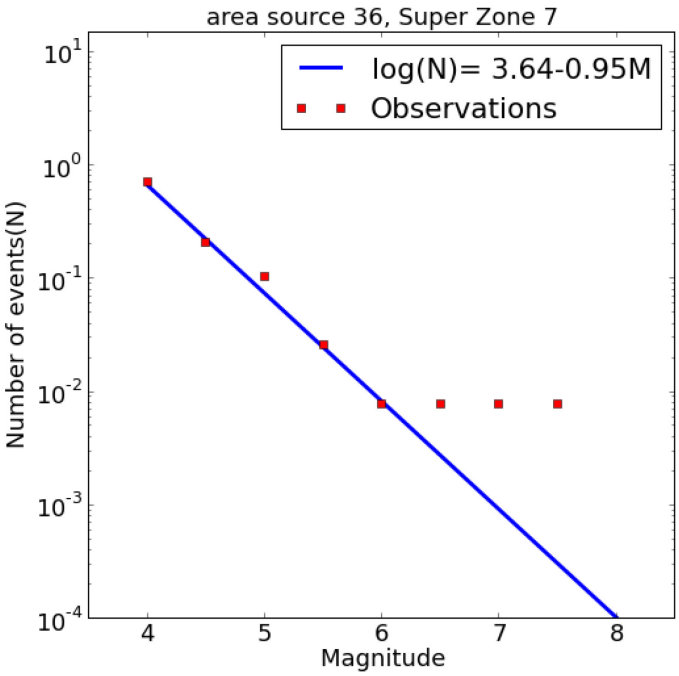

area source 30, Super Zone 4
log(N)= 2.3-0.6M
Observations
Number of events(N)
Magnitude

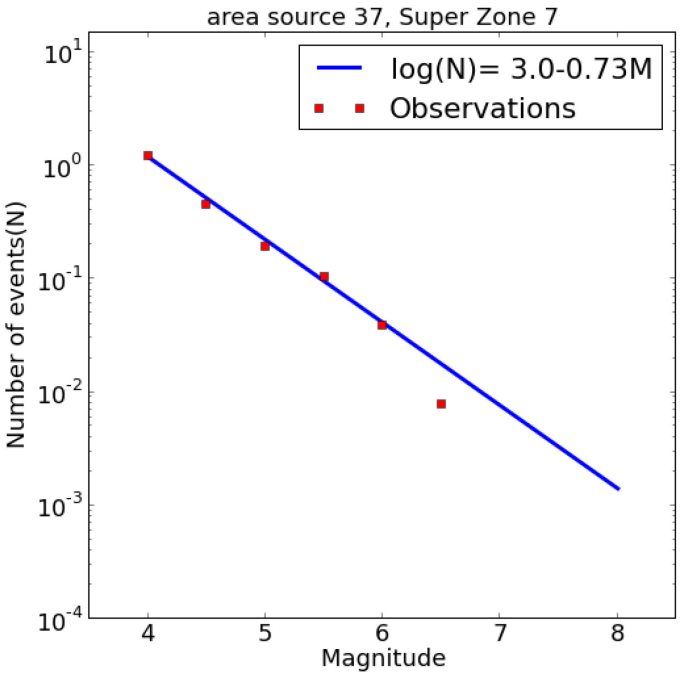

area source 32, Super Zone 8
log(N)= 3.1-0.9M
Observations
Number of events(N)
Magnitude

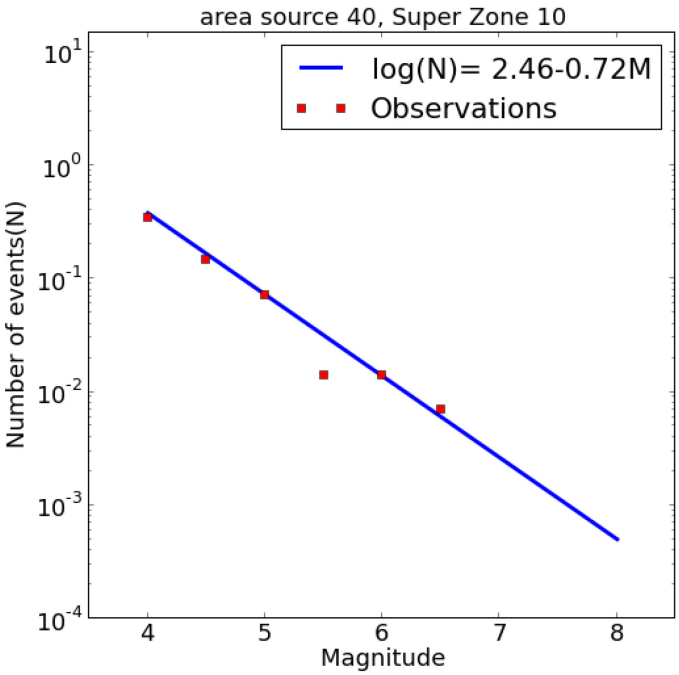

area source 33, Super Zone 8
log(N)= 3.35-0.8M
Observations
Number of events(N)
Magnitude

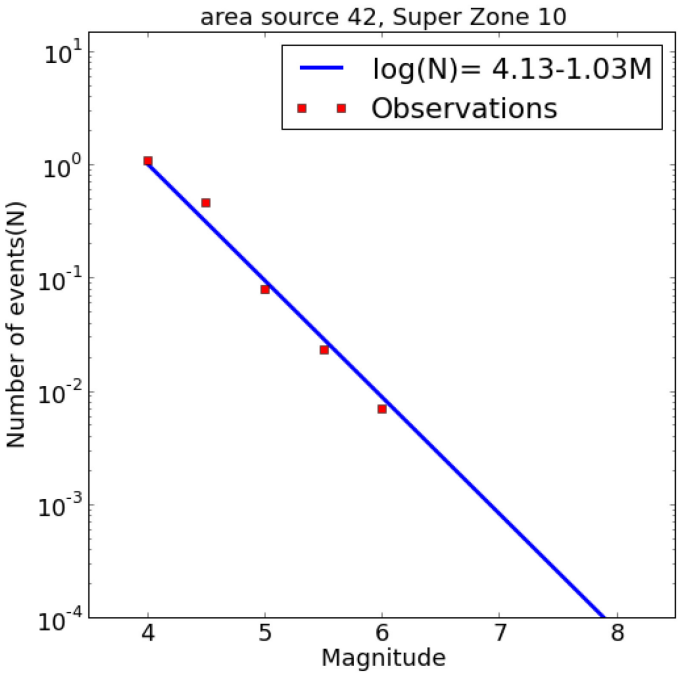

area source 34, Super Zone 8
log(N)= 3.4-0.82M
Observations
Number of events(N)
Magnitude

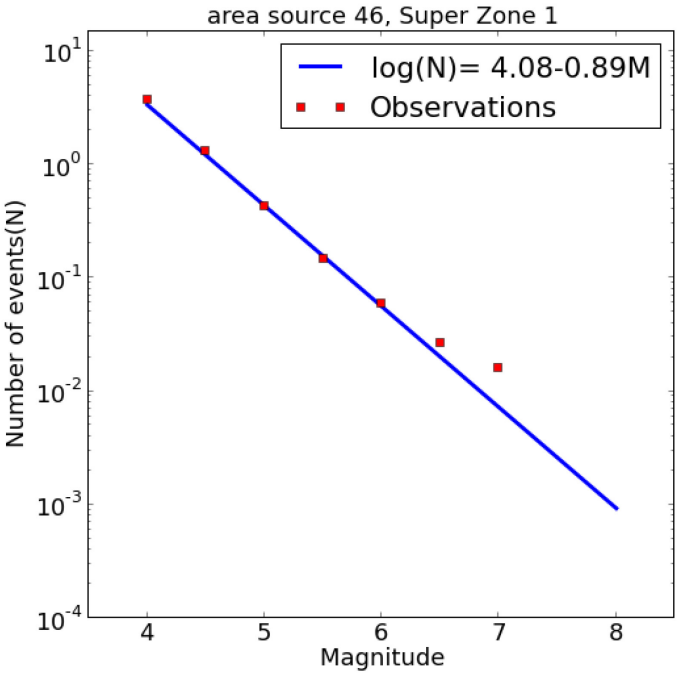

area source 35, Super Zone 8
log(N)= 2.2-0.7M
Observations
Number of events(N)
Magnitude

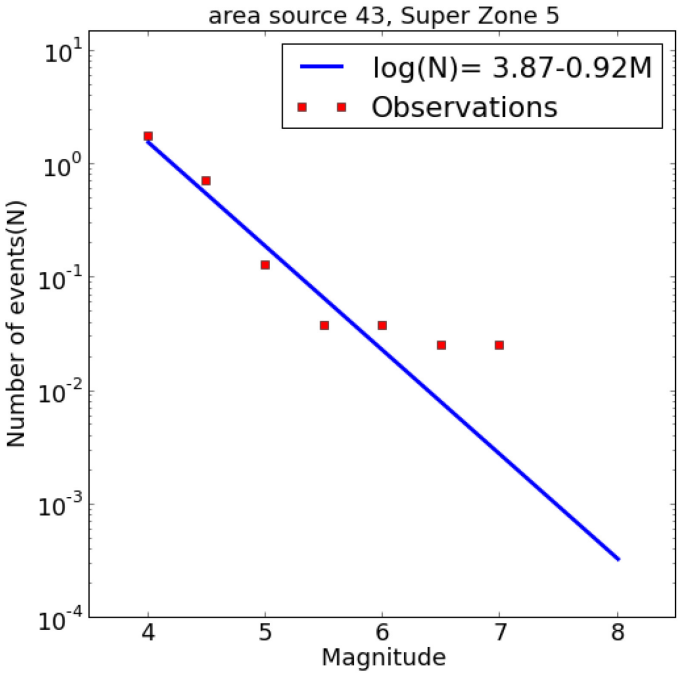

area source 36, Super Zone 7
log(N)= 3.64-0.95M
Observations
Number of events(N)
Magnitude

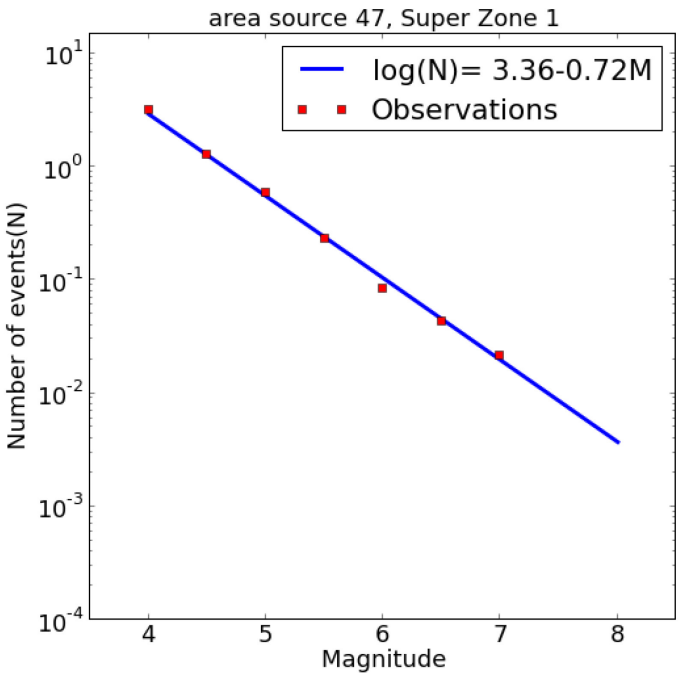

area source 37, Super Zone 7
log(N)= 3.0-0.73M
Observations
Number of events(N)
Magnitude

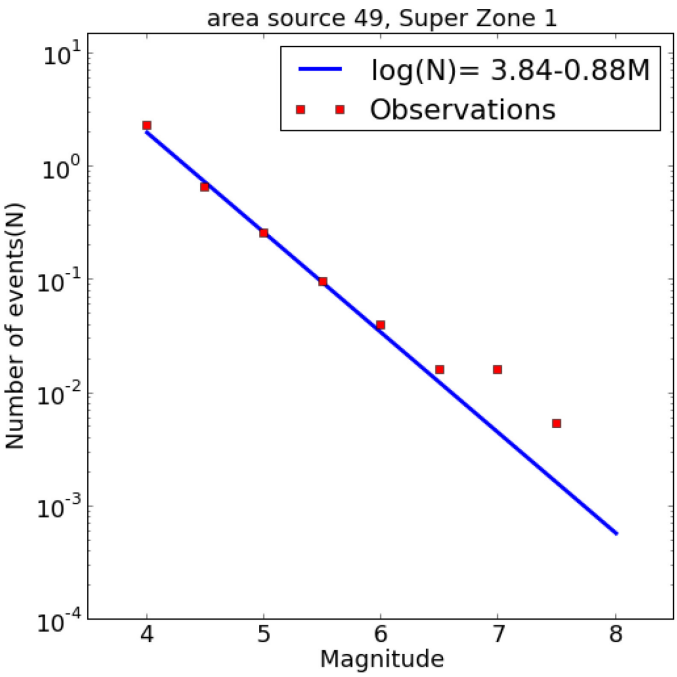

area source 40, Super Zone 10
log(N)= 2.46-0.72M
Observations
Number of events(N)
Magnitude

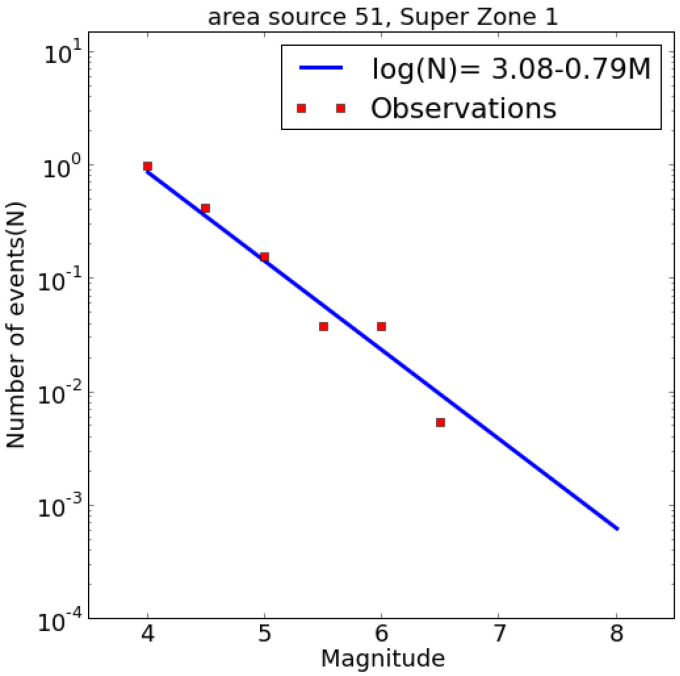

area source 42, Super Zone 10
log(N)= 4.13-1.03M
Observations
Number of events(N)
Magnitude

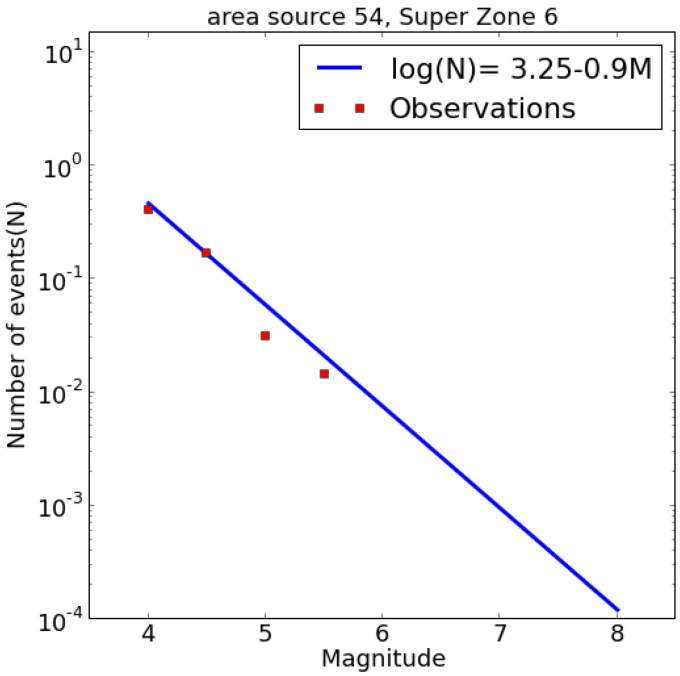

area source 46, Super Zone 1
log(N)= 4.08-0.89M
Observations
Number of events(N)
Magnitude

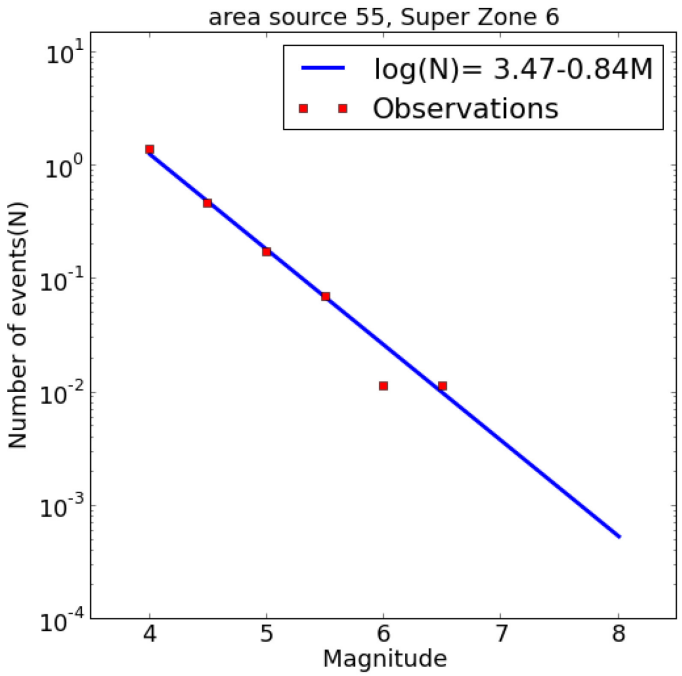

area source 43, Super Zone 5
log(N)= 3.87-0.92M
Observations
Number of events(N)
Magnitude

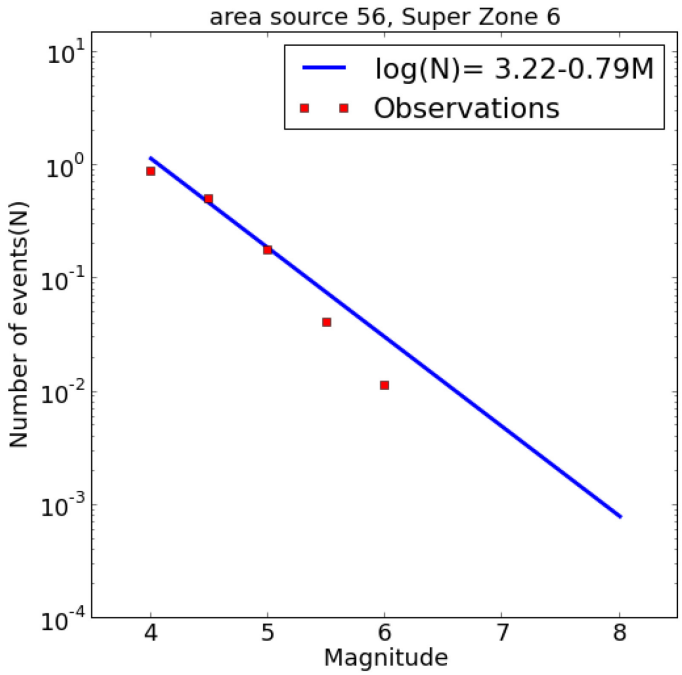

area source 47, Super Zone 1
log(N)= 3.36-0.72M
Observations
Number of events(N)
Magnitude

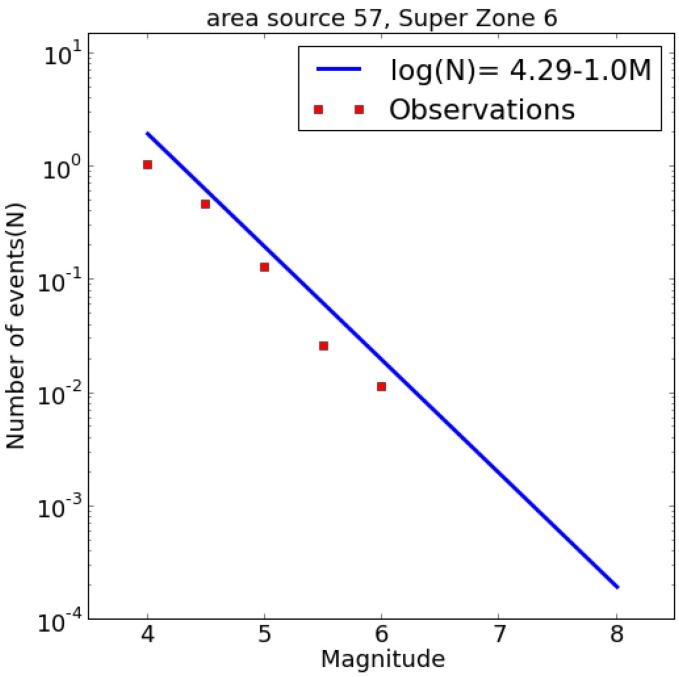

area source 49, Super Zone 1
log(N)= 3.84-0.88M
Observations
Number of events(N)
Magnitude

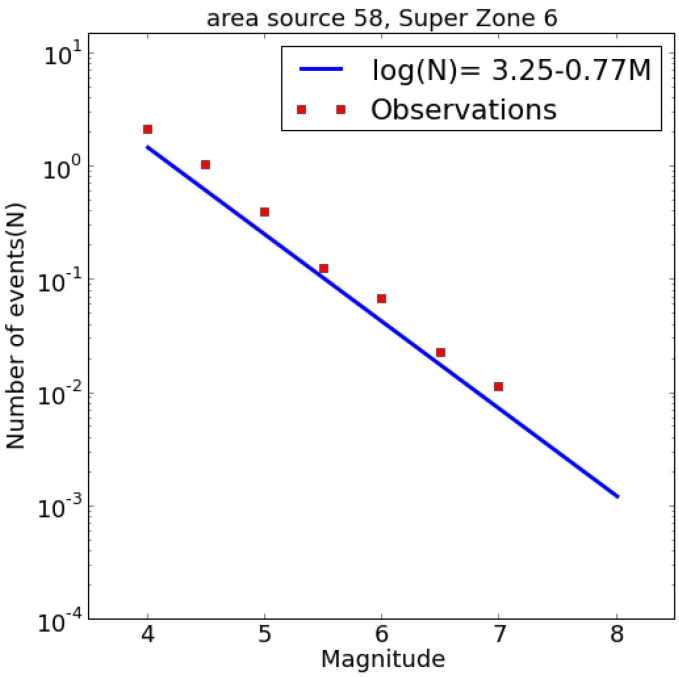

area source 51, Super Zone 1
log(N)= 3.08-0.79M
Observations
Number of events(N)
Magnitude


area source 54, Super Zone 6
log(N)= 3.25-0.9M
Observations
Number of events(N)
Magnitude


area source 55, Super Zone 6
log(N)= 3.47-0.84M
Observations
Number of events(N)
Magnitude


area source 56, Super Zone 6
log(N)= 3.22-0.79M
Observations
Number of events(N)
Magnitude


area source 57, Super Zone 6
log(N)= 4.29-1.0M
Observations
Number of events(N)
Magnitude
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
4
5
6
7
8


area source 58, Super Zone 6
log(N)= 3.25-0.77M
Observations
Number of events(N)
Magnitude
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
4
5
6
7
8