A GRID-BASED SEISMIC HAZARD ANALYSIS APPLICATION

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

BY

ÇELEBİ KOCAİR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

SEPTEMBER 2010

Approval of the thesis:

A GRID-BASED SEISMIC HAZARD ANALYSIS APPLICATION

submitted by **ÇELEBİ KOCAİR** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** _____

Dr. Cevat Şener
Supervisor, **Computer Engineering Department, METU** _____

Prof. Dr. Ayşen Akkaya
Co-supervisor, **Statistics Department, METU** _____

Examining Committee Members:

Asst. Prof. Dr. Tolga Can
Computer Engineering Department, METU _____

Dr. Cevat Şener
Computer Engineering Department, METU _____

Prof. Dr. Ayşen Akkaya
Statistics Department, METU _____

Asst. Prof. Dr. Sinan Kalkan
Computer Engineering Department, METU _____

Ali Yıldız, M.Sc.
ASELSAN _____

Date: _____

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

Name, Last Name: ÇELEBİ KOCAİR

Signature :

ABSTRACT

A GRID-BASED SEISMIC HAZARD ANALYSIS APPLICATION

Kocair, Çelebi

M.S., Department of Computer Engineering

Supervisor : Dr. Cevat Şener

Co-Supervisor : Prof. Dr. Ayşen Akkaya

September 2010, 55 pages

The results of seismic hazard analysis (SHA) play a crucial role in assessing seismic risks and mitigating seismic hazards. SHA calculations generally involve magnitude and distance distribution models, and ground motion prediction models as components. Many alternatives have been proposed for these component models. SHA calculations may be demanding in terms of processing power depending on the models and analysis parameters involved, and especially the size of the site for which the analysis is to be performed. In this thesis, we develop a grid-based SHA application which provides the necessary computational power and enables the investigation of the effects of applying different models. Our application not only includes various already implemented component models but also allows integration of newly developed ones.

Keywords: Grid, Grid Computing, Seismic Hazard Analysis

ÖZ

GRID TABANLI SİSMİK TEHLİKE ÇÖZÜMLEME UYGULAMASI

Kocair, Çelebi
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi : Dr. Cevat Şener
Ortak Tez Yöneticisi : Prof. Dr. Ayşen Akkaya
Eylül 2010, 55 sayfa

Sismik tehlike çözümlemesinin (STÇ) sonuçları sismik risklerin değerlendirilmesinde ve sismik tehlikenin azaltılmasında büyük önem taşır. STÇ hesaplamaları bileşen olarak genellikle büyüklük ve uzaklık dağılımı modelleri ile yer hareketi tahmin modellerini içerir. Bu bileşen modelleri için bir çok seçenek önerilmiştir. STÇ hesaplamaları; içerdiği modellere, çözümleme parametrelerine ve özellikle çözümlemenin yapılacağı bölgenin büyüklüğüne bağlı olarak yüksek işlem gücü gerektirebilir. Bu tezde, gerekli berim gücünü sağlayan ve farklı modellerin uygulanmasının etkilerinin incelenmesine olanak sunan grid tabanlı STÇ uygulaması geliştirilmiştir. Uygulamamız kendi içinde çeşitli bileşen modelleri bulundurmakla kalmayıp yeni geliştirilen modellerin de eklenmesine izin vermektedir.

Anahtar Kelimeler: Grid, Grid Hesaplama, Sismik Tehlike Çözümlemesi

To my family

ACKNOWLEDGMENTS

I would like to thank my supervisor Dr. Cevat Şener for his constant attention and sincere mentoring, and my co-supervisor Prof. Dr. Ayşen Akkaya for her continuous enthusiasm and persistent support.

I would like to thank Asst. Prof. Dr. Tolga Can for never denying me his time and efforts when I was in need. I would also like to thank Asst. Prof. Dr. Sinan Kalkan and Ali Yıldız for their valuable comments and suggestions.

My parents deserve much gratitude, because they taught me to pursue my decisions with faith and determination. Although being physically apart from each other, my brother Çağdaş Kocair managed to support me throughout this laborious period. And, I am heartily thankful to Dilan Kalmış for her encouragement and moral support, among many other things. Without them, this thesis and most of the things that are worth living in my life would not be realized.

Many thanks are owed to Ümit Sivri for helping me with part of the GUI implementation.

Lastly, I would like to present my regards and best wishes to all of those who supported me in any respect from the preliminaries to the completion of this thesis.

I had been partially funded by Turkcell Akademi and TBD (Türkiye Bilişim Derneği) under ”gnçtrkcll Yüksek Lisans Burs Programı” during my graduate studies. This study was further supported within the Seismology Virtual Organization of the FP7 SEE-GRID-SCI project, funded by the European Commission under the contract RI-211338.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Seismic hazards are very important aspects of public safety and need to be analyzed with respective consideration. Seismic Risk Assessment (SRA) is the study of quantifying the probabilities of occurrence of losses due to seismic hazards. In order to provide these probabilities, SRA requires information on earthquake-related phenomena as input. Seismic Hazard Analysis (SHA) tries to provide this information by describing ground shaking, ground failure, fault rupture etc. that have potential to cause harm and the associated occurrence frequencies. That is to say, SHA tries to quantify probabilities of occurrences of future earthquakes and the damages and losses they can evoke. The outputs of SHA can be used, via SRA, for assessing public safety and hazards mitigation, establishing appropriate insurance rates, for improving earthquake-resistant design and construction or emergency plans with the help of zoning maps. All these processes require a reliable seismic hazard assessment, and this requirement makes SHA a very complex and time-consuming study. In order to provide good estimations and realistic outcomes, the natural uncertainties connected to earthquakes have to be integrated into SHA, increasing the assessment duration intolerably. Also, data storage problems may arise because the amount of the required data can be particularly large for some instances. Since SHA is very important for the aforementioned issues, it is essential that these problems are solved. The fields of application increase as well as the rate, required customizations and deadlines; and current SHA applications face difficulties answering the demands. It is only rational to try and find a less time-consuming and more generic way of providing the outcomes of SHA than the current solutions.

The purpose of this thesis is to provide a solution for dealing with the difficulties that are encountered in SHA studies. Since these difficulties are mostly computational it is reasonable to search for a more powerful computing approach to attack this problem, which yields the

idea of using grid computing for that purpose. The virtually unlimited resources, both in terms of storage capability and computational power, provided by grid computing seem to be the only option for SHA. In this thesis, an attempt to utilize the grid computing resources for constructing a powerful solution for SHA is realized.

In the scope of this work, after briefly studying the theoretical background information regarding earthquake-related phenomena, numerous scientific studies related to the methodologies used in SHA are investigated. Afterwards, currently offered software solutions for SHA are examined and their strengths and weaknesses are attempted to be determined. Using the obtained background information, a Grid-based approach is proposed for the solution of aforementioned problems encountered in SHA studies. By means of following the proposed approach, an application that uses both computing and storage resources provided by the Grid infrastructure is implemented.

The rest of the thesis is organized as follows: In the next chapter, a brief introduction to SHA concepts and some information regarding the necessary calculations will be given. Chapter 3 will introduce grid computing. In Chapter 4, a brief survey on available software for SHA will be presented. Chapter 5 will describe the structure of the developed application. In Chapter 6 some details about the implementation of the application will be explained followed by a description of the tests performed. The last chapter will conclude the thesis by briefly summarizing the work done and pointing out some future directions.

CHAPTER 2

BACKGROUND ON SEISMIC HAZARD ANALYSIS

Seismic hazard analysis (SHA) methods can be classified into two main categories, namely deterministic and probabilistic approaches. In the following sections; first, the fundamental concepts will be explained, and then the probabilistic and deterministic approaches to SHA will be described in detail.

2.1 Fundamental Concepts

In both deterministic and probabilistic SHA approaches, analyses are performed at a *site region*. This site region may be a point or a rectangular region representing the location of interest, such as a construction area or an existing building. A city, a country, a continent, or even the whole world may be used as the site of interest in large-scale analyses.

Seismic sources are essential components in SHA. A seismic source can be defined as a region that has almost invariant features in terms of seismicity [6]. For SHA, it is necessary to characterize all seismic sources near the site of interest. This characterization mainly involves identifying the locations and the geometries of the sources. Seismic sources are commonly categorized geometrically as follows:

- *Line sources* correspond to actual faults, and are represented with straight lines or, more generally, with a series of line segments. Since faults have three-dimensional planar structures, line sources actually represents a map-view of the fault plane [49].
- *Area sources* are used for describing the regions where many small faults are co-located or the previous earthquake activity cannot be associated with well-known fault struc-

tures, or representing the faults which cannot be represented as line sources [2]. Area sources are depicted as polygons with arbitrary boundaries.

- *Point sources* are used for modeling past seismic activity concentrated on a small area far from the site of interest, possibly originating from volcanic or geothermal activities [2].

In addition to the location and the geometry of a seismic source; seismicity parameters, namely *maximum earthquake magnitude* and *earthquake recurrence*, specific to that source must be determined as a part of the seismic source characterization process [6].

As the name implies, maximum earthquake magnitude is the largest possible magnitude of an earthquake that a specific seismic source may produce. Two approaches are commonly used for determining maximum earthquake magnitude. The first approach makes use of historical earthquake evidence. Among the earthquakes corresponding to the source, the one with the largest magnitude is selected, and its magnitude with some increment (generally 0.5 magnitude units) is used as the maximum earthquake magnitude. The second approach is used when the underlying fault structure of the seismic source is known. The maximum earthquake magnitude is determined through empirical regression between earthquake magnitude and a geometric feature of the fault such as its total length or rupture length. For example, the following relations between earthquake magnitude and surface rupture length are proposed by Wells and Coppersmith [52]:

$$\begin{aligned} M_w &= 5.16 + 1.12 \log L, \quad \text{(strike-slip)} \\ M_w &= 5.00 + 1.22 \log L, \quad \text{(reverse)} \\ M_w &= 4.86 + 1.32 \log L, \quad \text{(normal)} \end{aligned} \tag{2.1}$$

where M_w is moment magnitude and L is the fault rupture length in kilometers. Here, it should be noted that when the fault rupture length cannot be determined empirically, generally one-half of the total fault length is used as the rupture length [34].

Earthquake recurrence of a seismic source is defined as the frequencies of earthquakes with distinct magnitudes generated by that source [6]. A recurrence relation between earthquake magnitudes and the number of earthquake occurrences is used to describe the earthquake recurrence. The most widely used earthquake recurrence relation is proposed by Gutenberg

and Richter [26] and is given as follows:

$$\log N(m) = a - bm, \tag{2.2}$$

where m is the Richter magnitude, $N(m)$ is the number of earthquakes with magnitudes greater than m, and a and b are constants. The constants a and b are determined through regression of historical earthquake data, and the most commonly used methods for this purpose are least squares estimation and maximum likelihood estimation.

Since estimating the ground motions expected to occur on the site of interest is the main focus of SHA, *ground motion attenuation relations* (or *ground motion prediction models*) are crucial for SHA. Ground motions are quantified using intensity measures. Peak ground acceleration (PGA), peak ground velocity (PGV), and spectral acceleration (SA) are the most commonly used ground motion intensity measures. The ground motion attenuation relations aim to estimate the ground motion intensity on the site region in terms of a selected intensity measure. They predict the probability distribution of ground motion intensity as a function of one or more predictor variables such as magnitude, distance, fault type and local site conditions [3]. Baker [3] notes that ground motion prediction models are usually developed using statistical regression on previously observed ground motion intensity values, and gives the general form of ground motion prediction models as follows:

$$\ln Y = \overline{\ln Y}(M, R, \theta) + \sigma_{\ln Y}(M, R, \theta) \cdot \varepsilon, \tag{2.3}$$

where $\ln Y$ is the natural logarithm of the chosen ground motion intensity measure (Y) and it is modeled as a random variable. $\overline{\ln Y}$ and $\sigma_{\ln Y}$ are the predicted mean and the standard deviation of this random variable, respectively. Both the predicted mean and the standard deviation are given as functions of earthquake magnitude (M), distance (R) and other predictor variables (θ). Finally, ε is a standard normal random variable which represents the observed variability in $\ln Y$.

Many ground motion attenuation relationships are proposed for being used in SHA studies. Some of them are developed for specific regions and some are generic, i.e. they can be used for any region. As an example, the following generic ground motion attenuation relationship is proposed by Cornell et al. [10]:

$$\begin{aligned} \overline{\ln Y} &= -0.152 + 0.859M - 1.803 \ln (R + 25), \\ \sigma_{\ln Y} &= 0.57, \end{aligned} \tag{2.4}$$

where Y is PGA in units of g. Here, it should be noted that the relation depends only on magnitude and distance, and also the standard deviation is constant for all magnitude and distance values.

2.2 Probabilistic Seismic Hazard Analysis

The main idea behind the probabilistic approach to SHA is to provide a way of assembling all the uncertainties while assessing the seismic hazard. The uncertainties in the location, time, and magnitude of future earthquakes are considered in probabilistic seismic hazard analysis (PSHA) studies [32]. Furthermore, uncertainties related to ground motions are also considered by means of the ground motion prediction models, as mentioned in the previous section. By combining the probability distributions corresponding to these aforementioned uncertainties, PSHA estimates the probability of observing a ground motion with an intensity greater than a particular level at the site of interest in the future.

Baker [3] describes PSHA as the following five-step process:

1. The seismic sources which can generate considerable ground motions on the site region are determined.

2. The probability distribution of source-to-site distances is characterized for each selected source.

3. Different earthquake magnitudes that each selected source can produce is characterized as a probability distribution.

4. An appropriate ground motion prediction model is selected in order to quantify variation of ground motion intensity.

5. The distributions are combined using the total probability theorem in order to evaluate the exceeding probabilities.

In the following subsections the details of these steps are explained.

2.2.1 Selecting the Seismic Sources

Among all the seismic sources characterized as described in Section 2.1, the ones which can produce damaging ground motions on the site of interest are determined as the first step of PSHA. Since earthquakes at large distances will not produce significant ground motions at the site of interest even if they have large magnitudes, sources that are far from the site will not contribute to the ground motion at the site [6]. Hence, the distance between the site and the source is generally used as the selection criterion, and commonly the sources within a 150-300 kilometers radius of the site region are selected.

2.2.2 Characterizing the Distribution of Source-to-site Distances

After relevant sources are identified; the distributions of source-to-site distances, which rationalize the uncertainties in earthquake locations, are characterized for the selected sources. In general, a probability density function (PDF), generally denoted as $f_R(r)$, is derived to represent the distance distribution.

For point sources, the corresponding PDF is given as:

$$f_R(r) = \begin{cases} 1 & \text{if } r = r_0, \\ 0 & \text{otherwise,} \end{cases} \tag{2.5}$$

where r_0 is the distance between the site and the point source.

For line and area sources, potential earthquake locations are generally assumed to be uniformly distributed throughout the source [32]. Although this assumption is not necessary, i.e. non-uniform distributions can also be used; it simplifies the characterization of source-to-site distance distributions.

One way to characterize the distance distribution for a line or an area source is to derive it analytically by making use of the geometry of the source. Baker [3] provides examples of this derivation for both line and area sources. Another commonly used approach is to split the seismic source into smaller equal-sized elements, i.e. line segments for line sources and rectangles for area sources; and approximate the distance distribution numerically using the distances between the center of each element and the site [32]. One other method for characterizing the distribution of source-to-site distances again uses the idea of dividing the

source as in the previous approach. As opposed to the previous approach though; this method considers each element as a point source, which is located on the center of the element, and uses the PDF in equation 2.5.

2.2.3 Characterizing the Distribution of Earthquake Magnitudes

After the distribution of source-to-site distances are characterized, the next step is to determine a probability distribution of earthquake magnitudes that each selected source may produce. As in the case of source-to-site distance distributions; a probability density function, generally denoted as $f_M(m)$, represents the distribution of earthquake magnitudes.

The earthquake recurrence described in Section 2.1 constitutes the basis for deriving the magnitude distribution of a seismic source. Generally, the PDF for the distribution of earthquake magnitudes depends on the constants (especially the so-called *b-value*) determined through statistical regression of the corresponding recurrence relation.

Three most commonly used models for assessing the distribution of earthquake magnitudes are described next.

Gutenberg-Richter model

This model is derived directly using Gutenberg-Richter recurrence relation described in Section 2.1. Cumulative distribution function (CDF) for the magnitudes of earthquakes is defined as the following ratio [3]:

$$\begin{aligned} F_M(m) &= P(M \leq m \mid m > m_{min}) \\ &= \frac{\text{Rate of earthquakes with} \quad m_{min} < M \leq m}{\text{Rate of earthquakes with} \quad m_{min} < M}, \qquad m > m_{min} \end{aligned} \tag{2.6}$$

where m_{min} is some minimum magnitude, generally taken as 4 or 4.5. The earthquakes with magnitudes smaller than m_{min} are ignored in SHA calculations due to their lack of producing strong ground motions.

Without loss of generality, Gutenberg-Richter recurrence relation given in Equation 2.2 can be rewritten for the rate of earthquake occurrences as follows:

$$\log \lambda_m = a - bm,$$

where λ_m denotes the rate of earthquakes with magnitudes greater than m. Substituting this form of the relation in Equation 2.6, the CDF is obtained as:

$$F_M(m) = \frac{\lambda_{m_{min}} - \lambda_m}{\lambda_{m_{min}}} = 1 - 10^{-b(m-m_{min})}. \qquad m > m_{min}$$

Taking the derivative of the above CDF, the PDF for the distribution of earthquake magnitudes is obtained as follows:

$$f_M(m) = -b\ln(10)10^{-b(m-m_{min})}. \qquad m > m_{min} \tag{2.7}$$

Bounded Gutenberg-Richter model

The Gutenberg-Richter model described above estimates the distribution of earthquake magnitudes without an upper bound. However, this unboundedness is not coherent with the real situation. Hence, bounding the earthquake magnitudes with a constraint on maximum possible magnitude value, the following PDF for the distribution is obtained:

$$f_M(m) = \frac{b\ln(10)10^{-b(m-m_{min})}}{1 - 10^{-b(m_{max}-m_{min})}}, \qquad m_{min} < m < m_{max} \tag{2.8}$$

where m_{max} is the maximum earthquake magnitude, which is determined while identifying the seismic source characteristics as described in Section 2.1.

Characteristic earthquake model

Schwartz and Coppersmith [39] proposed that earthquakes with magnitudes approximately equal to the maximum magnitude are frequently produced by individual faults and fault segments. Such earthquakes are called characteristic earthquakes, and their magnitudes vary in a range of one-half magnitude units.

Using the characteristic earthquake model, Youngs and Coppersmith [54] derived a distribution for earthquake magnitudes, with the following PDF:

$$f_M(m) = \begin{cases} k\beta e^{-\beta(m-m_0)} & m_0 \le m < m_1 - 0.5, \\ k\beta e^{-\beta(m_1-3/2-m_0)} & m_1 - 0.5 \le m \le m_1, \end{cases} \tag{2.9}$$

where m_0 is the minimum earthquake magnitude and m_1 is the characteristic earthquake magnitude, instead of which the maximum earthquake can be also used. β and k are are constants

defined as follows:

$$\beta = b \ln 10,$$
$$k = \left[1 - e^{-\beta(m_1 - 0.5 - m_0)} + \beta e^{-\beta(m_1 - 3/2 - m_0)} 0.5\right]^{-1}.$$

2.2.4 Selecting the Attenuation Relation

The last step before combining all the distributions for obtaining exceeding probabilities is to determine the ground motion attenuation relation to be used in the analysis. Although generally only one attenuation relation is used for the whole SHA study, it is actually possible to choose different attenuation relations for each selected source.

While choosing the attenuation relation, its suitability for the site region and the seismic sources should be inspected. As already mentioned in Section 2.1, some ground motion prediction models are region-specific; and hence they can be used only in that particular region (or maybe in a region similar to that particular region in terms of both site and source characteristics). Furthermore, many attenuation relations necessitate some constraints for the site and/or the seismic source. For example; some relations may be developed for only sites of specific soil type, hence the underlying soil type (rock, soil, stiff soil, etc.) of the site region should be known before choosing to use such attenuation relations. As another example; an attenuation relation may change formulation according to the fault type (normal, strike-slip, reverse) of the seismic source. Hence, that attenuation relation cannot be used for a seismic source when the fault type of the corresponding fault cannot be determined, or when the source does not actually represent a fault. Another condition that ground motion prediction models commonly require is about the source-to-site distances. Many attenuation relations require that the site and the source are not further than a particular distance.

2.2.5 Calculating the Probabilities

The final step of PSHA is to evaluate the exceeding probabilities by combining the modeled uncertainties in the previous steps. Firstly, the selected attenuation relation is used for calculating the probability of exceeding a particular ground motion intensity level when the magnitude and the distance are given. Since natural logarithm of the ground motion intensity is observed to be normally distributed, the exceeding probability for a particular level of

ground motion intensity can be calculated as follows:

$$P(Y > y \mid m, r) = 1 - \Phi\left(\frac{\ln y - \overline{\ln Y}}{\sigma_{\ln Y}}\right), \tag{2.10}$$

where Φ is the standard normal cumulative distribution function.

Equation 2.10 evaluates the exceeding probability when the magnitude and the distance are known. However; as already mentioned future earthquake locations and magnitudes are uncertain, and they are modeled as probability distributions. These distributions and the conditional probability formulated in Equation 2.10 are combined using the total probability theorem as follows [3]:

$$P(Y > y) = \int_{m_{min}}^{m_{max}} \int_{0}^{r_{max}} P(Y > y \mid m, r)\ f_M(m)\ f_R(r)\ dr\ dm. \tag{2.11}$$

In order to obtain the probability that ground motion intensity exceeds a particular level on the site region, the formula above uses integration to sum up the conditional exceeding probabilities of all possible magnitude and distance values.

Analysis studies are generally interested in determining the frequency of earthquake occurrences. By applying a simple modification to Equation 2.11, the rate of observing ground motions with intensity levels greater than a particular level can be calculated as follows [3]:

$$\lambda(Y > y) = \lambda(M > m_{min}) \int_{m_{min}}^{m_{max}} \int_{0}^{r_{max}} P(Y > y \mid m, r)\ f_M(m)\ f_R(r)\ dr\ dm, \tag{2.12}$$

where $\lambda(Y > y)$ is the rate of ground motions with intensities greater than y, and $\lambda(M > m_{min})$ is the rate of earthquakes, produced by the seismic source, with magnitudes greater than m_{min}. $\lambda(M > m_{min})$ is determined by using the historical earthquake catalog data.

Here, it should be noted that Equation 2.12 computes the exceeding rate of ground motion intensity that a single seismic source causes. Considering all relevant seismic sources, the total rate of exceeding a particular ground motion intensity level at the site is the sum of the exceeding rates computed individually for each seismic source. Hence, the total exceeding rate can be formulated as follows [3]:

$$\lambda(Y > y) = \sum_{i=1}^{n_{sources}} \lambda(M_i > m_{min}) \int_{m_{min}}^{m_{max}} \int_{0}^{r_{max}} P(Y > y \mid m, r)\ f_{M_i}(m)\ f_{R_i}(r)\ dr\ dm. \tag{2.13}$$

Although numerical integration methods can handle the integrals in Equation 2.13, generally these integrals are converted to summations by means of splitting the magnitude and distance

ranges into small intervals. Baker [3] gives the formulation, when possible magnitude and distance ranges are divided into n_M and n_R intervals respectively, as follows:

$$\lambda(Y > y) = \sum_{i=1}^{n_{sources}} \lambda(M_i > m_{min}) \sum_{j=1}^{n_M} \sum_{k=1}^{n_R} P(Y > y \mid m_j, r_k)\, P(M_i = m_j)\, P(R_i = r_k), \quad (2.14)$$

where $P(M_i = m_j)$ and $P(R_i = r_k)$ are probabilities of particular magnitude and distance values m_j and r_k respectively. These probabilities are calculated as:

$$P(M_i = m_j) = F_{M_i}(m_{j+1}) - F_{M_i}(m_j),$$
$$P(R_i = r_k) = F_{R_i}(r_{k+1}) - F_{R_i}(r_k),$$

where F_{M_i} and F_{R_i} are CDFs of the magnitude and distance distributions respectively. Since calculation of these probabilities uses CDFs; in order to use the summation formula in Equation 2.14, magnitude and distance distributions are required to be characterized using their CDFs instead of their PDFs.

2.3 Deterministic Seismic Hazard Analysis

The deterministic seismic hazard analysis (DSHA) approach involves a scenario-based methodology. The worst-case earthquake scenario, i.e. the earthquake which will generate the largest ground motion intensity level at this site, is assumed and the ground motion that the predicted scenario will produce at the site region is analyzed. The DSHA methodology can be defined as the following four-step process based on the description provided by Reiter [37]:

1. The seismic sources that are capable of producing effective ground motions at the site region are identified.
2. The distances between the site and the selected sources are characterized.
3. The controlling earthquake, i.e. the earthquake that will generate the largest ground motion intensity, is determined.
4. The seismic hazard at the site region is evaluated using the controlling earthquake.

The first step actually is the same as the one in PSHA, i.e. the seismic sources within a certain proximity of the site are selected. On the other hand; in the second step only the closest distances between the site region and the selected sources are considered as opposed to the

case in PSHA methodology, in which source-to-site distances are characterized as probability distributions. Again in contrast to the case in PSHA, no magnitude distributions are characterized in DSHA. The maximum possible earthquake magnitude is used for each source for determining the controlling earthquake. After selecting a suitable attenuation relation, for each source the closest distance and the maximum magnitude is used to evaluate the ground motion that is expected to be produced at the site. The largest ground motion intensity value obtained among those evaluations is selected as the ground motion produced by the controlling earthquake, and it is treated as the output of DSHA study.

CHAPTER 3

BACKGROUND ON GRID COMPUTING

In the scope of computing; the term *Grid*, in its most general form, is used to describe an infrastructure that combines geographically distributed computer systems for providing high-throughput computing capabilities. Foster et al. [16] introduced the concept of *virtual organizations* (VO) by defining the grid computing as an approach for solving large-scale problems by means of a collaborative sharing of various computational resources dynamically among multiple institutions and organizations.

Grid computing provides not only transparent and reliable access to additional, possibly under-utilized, computing and storage resources for individual users, but also a collaborative research infrastructure for scientific communities.

In the rest of this chapter; first, the general architecture of Grid systems is described, and then some efforts to classify types of Grid systems are examined.

3.1 Grid Architecture

A protocol architecture, as can be examined in Figure 3.1, for the Grid is proposed by Foster et al. [16]. The proposed architecture follows a layered approach similar to the Internet protocol architecture. The top-most three layers correspond to Application layer in the Internet protocol architecture, whereas the Fabric layer is analogous to the Link layer. The Connectivity layer, on the other hand, relates to the combination of the Transport and Internet layers in the Internet protocol architecture. In the following parts these layers are described briefly.

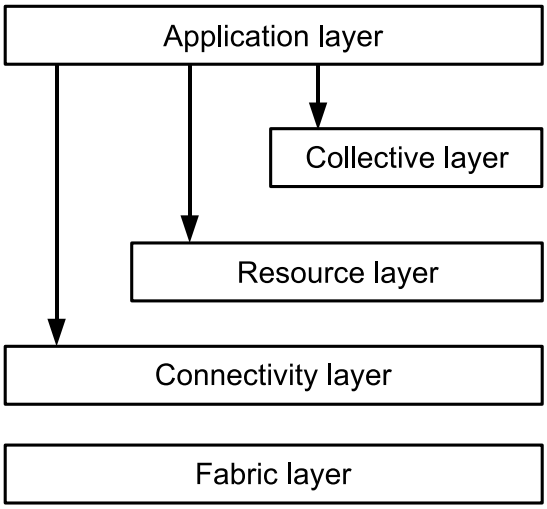

Figure 3.1: The Grid protocol architecture, adapted from [16]

3.1.1 Fabric Layer

The Fabric layer involves the physical and logical resources included in the Grid infrastructure. Hardware resources such as cluster computers and storage arrays constitute physical resources, whereas a file system implementation or a data catalog may be considered as examples of logical resources. This layer mainly necessitates state discovery and resource-specific management functionalities for the resources involved. More advanced operations are not generally requested, because such a demand will complicate the deployment of the resources to the infrastructure.

3.1.2 Connectivity Layer

As the name implies, the Connectivity layer provides the necessary communication mechanisms among Grid resources. Moreover, this layer also involves authentication protocols in order to ensure trust among the users and the resources. Connectivity layer implementations mostly utilize existing protocols defined in TCP/IP stack.

3.1.3 Resource Layer

The Resource layer provides information and management protocols which access the relevant functions in the Fabric layer, passing through communication and authentication mechanisms of the Connectivity layer, for discovering and managing single resources. Information protocols are analogous to state discovery functions in the Fabric layer, i.e. they are used for monitoring the state and structure of the resources. Management protocols provide secure instantiation and management of various operations performed on the resources.

3.1.4 Collective Layer

The Collective layer provides the necessary functionality, in terms of protocols and services, for organizing the interactions among collections of available Grid resources. Resource allocation, job scheduling, data replication, and collective resource monitoring services can be given as examples to the services that this layer involves.

3.1.5 Application Layer

The Application layer in the Grid protocol architecture involves the user applications. The applications in this top-most layer are developed by making use of many protocols and services implemented within the scope of the other layers in the architecture, as depicted in Figure 3.1.

3.2 Grid Classification

Although, mostly a major classification based on the main focus of the system, namely computational and data Grids, is agreed upon; there exists no standard classification of Grid systems. In this section some classification efforts are described.

Krauter et al. [33] proposed a classification for Grid systems, seen in Figure 3.2, which adds *service Grids* to the major classification mentioned above. Service Grids describe the systems providing large-scale services that cannot be provided by single machines. These type of Grids are further divided into three sub-categories: *On-demand*, *collaborative* and *multimedia* Grids. The first sub-category represents the systems that are capable of aggregating resources

on a dynamic basis for the services they provide. Collaborative Grids enable interactions among people and applications involved in the same or similar virtual study groups. As the name implies, the last sub-category involves an environment for real-time multimedia applications. Moreover, the proposed taxonomy also divides *computational Grids* into two categories, namely *distributed supercomputing* and *high throughput*. The former sub-category involves the systems that are capable of executing single jobs in parallel, whereas the latter describes the systems favoring stream-type jobs involving parameter studies.

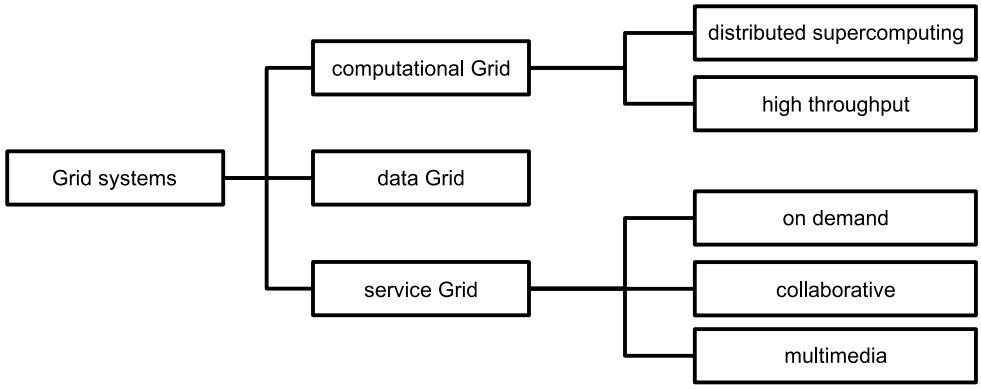

Figure 3.2: A Grid systems taxonomy, adapted from [33]

Another categorization based on system topology, provided by Ferreira et al. [13], is as follows:

- *Intragrids* consist of one or more computer clusters, which are connected through an internal private network, operated by single organizations, and they provide only a small number of Grid services.

- *Extragrids* involve more than one intragrid connected through a wide area network, operated by multiple organizations, and provide a more dynamic environment than intragrids for partner integration purposes.

- *Intergrids* constitute a more general form of extragrids, since they provide an infrastructure for a collaborative community involving many organizations and multiple business partners.

3.3 SEE-GRID Infrastructure

The regional Grid infrastructure in the South Eastern European (SEE) region is mostly referred as the SEE-GRID. This infrastructure includes the national Grids of most of the countries in the region. After the infrastructure was constructed within the scope of the SEE-GRID (South-Eastern European Grid-enabled eInfrastructure Development) project [46], the SEE-GRID-2 [41] and SEE-GRID-SCI (SEE-GRID eInfrastructure for regional eScience) [43] projects not only expanded and improved the infrastructure but also strengthened the communication and collaboration among the scientific communities in the SEE region. A snapshot view of SEE-GRID-SCI eInfrastructure, which expands the original SEE-GRID infrastructure, can be seen in Figure 3.3 [44], provided by Scientific Computing Laboratory (SCL) at the Institute of Physics Belgrade.

Figure 3.3: SEE-GRID-SCI eInfrastructure as of 15 June 2010, taken from [44]

3.3.1 Overview of the Infrastructure

The SEE-GRID infrastructure is described using the following three sub-groups [42]:

- *Operational infrastructure* represents the core Grid infrastructure. It consists of the Grid sites and the necessary services deployed on them for providing the major func-

tionality of the Grid. The required tools and services in this infrastructure are provided by gLite middleware [19].

- *Operational support infrastructure* assists the Grid administrators in maintaining the operation of the Grid infrastructure. It mainly includes different tools which collect and serve monitoring data, such as the availability of the deployed services or the statuses of the resources in different Grid sites.

- *User support infrastructure* provides some services for aiding end-users and maintaining the communication among administrators. It mainly contains mailing lists and technical forums.

3.3.2 Details of Operational Infrastructure

As already mentioned in Subsection 3.3.1, the operational infrastructure contains Grid sites and services implementing the main Grid functionality. Grid sites typically consist of the following components:

- *Computing Element (CE)*: A CE represents the computing resources that are provided by a Grid site. It includes a Grid Gate (GG), a Local Resource Management System (LRMS), and a set of Worker Nodes (WNs) [8]. The WNs are the computers where the Grid jobs are actually executed. The GG initiates job executions on WNs by using the LRMS.

- *Storage Element (SE)*: An SE represents the storage resources that are provided by a Grid site. An SE includes a Storage Resource Manager (SRM) for managing the available storage resources such as large disk arrays or tape-based storage systems [8]. Different SRM implementations are available for managing different types of storage resources.

- *User Interface (UI)*: A UI provides the necessary environment for the users to use the Grid. The users perform Grid-related operations, such as authentication, job management, and file management, through using the related tools provided by a UI where they have personal accounts [8].

The core services deployed in the scope of the operational infrastructure are as follows:

- *Virtual Organization Membership Service (VOMS)* is an authorization service that manages and serves the information about the users in a VO [42]. The information provided by this service includes the roles, groups and capabilities of the users.

- *Berkeley Database Information Index (BDII)* is a Grid service that periodically checks and serves the information about computing and storage resources in the Grid infrastructure [8].

- *Workload Management System (WMS)* manages the Grid jobs. For job submission, the attributes related to the job are defined using Job Description Language (JDL) and according to those attributes the WMS determines the most suitable CE for the execution of the job [8]. Other job management operations such as status checking and output retrieval are also handled by the WMS.

- *Resource Broker (RB)* is another service for job management operations. It determines the appropriate resources for a submitted Grid job, schedules the job, and monitors it [42]. Although RB service is deployed in the SEE-GRID infrastructure; it has been abandoned in favor of the WMS, since the latter is more robust and provides more functionality [40].

- *Relational Grid Monitoring Architecture (R-GMA)* is used for gathering accounting information [42]. Both system-level and user-level accounting data are collected and published by this service [8].

- *LCG File Catalog (LFC)* manages the mappings between actual Grid files and their logical names [8]. It provides a hierarchical namespace structure and integrated Grid authentication mechanisms, and also supports access control lists [42].

- *MyProxy Service* provides storage and retrieval mechanisms for user credentials. Actual Grid authentication system involves *proxy certificates* which hold the user credentials, and these proxies expire after a predefined time. MyProxy service provides an automatic proxy renewal mechanism, enabling the execution of jobs that require long running times [8].

- *File Transfer Service (FTS)* is a service that controls the file transfer operations among the SEs in the Grid, and it is mostly used for large-scale data transfers [8].

- *ARDA Metadata Catalog (AMGA)* is a service that provides an interface for database access. It supports integrated Grid authentication mechanisms and enables Grid applications to use different types of databases [8].

CHAPTER 4

RELATED WORK

As described in Chapter 2, SHA requires a great deal of numerical computations even in the case of a single site location and a single seismic source. Since actual SHA studies involve large site regions, or thousands of site locations, and many seismic sources, the use of computers is inevitable for SHA studies. In addition, many SHA studies investigate the effects of applying different SHA models in the analyses; and hence, they definitely require automated procedures for SHA computations. Furthermore, due to the dynamic nature of seismic events; repetition of the analyses is occasionally required, which again introduces the necessity of automation. Despite some computer programs written specifically for some SHA studies, there also exist software projects and products to be used as generic solutions. In this chapter, some widely used software for SHA are described.

SEISRISK III [5] is the final revision of Fortran programs used by the United States Geological Survey (USGS) for seismic hazard mapping prior to 1996. Although SEISRISK III is not used by USGS for producing seismic hazard maps since 1996; it is still being used, mostly with some modifications by individual efforts, for producing seismic hazard maps around the world and for teaching purposes. SEISRISK III mainly computes the ground motion exceedance probabilities, and it requires that all the data including information regarding seismic sources and tables for attenuation relations are provided as inputs prepared by many other programs. Hence, it cannot be considered as a complete SHA solution.

USGS further provides the software [1] used for producing the 2008 Update of the United States National Seismic Hazard Maps [36]. Although the most recent developments in SHA methodology are utilized in computations for producing the maps, the software involves region specific attributes since it is developed for seismic hazard mapping of the U.S. The soft-

ware uses different Fortran programs for different parts of SHA computations, and C codes for some input/output operations.

EZ-FRISK [12], developed by Risk Engineering Inc., is a widely used commercial product for SHA. EZ-FRISK supports both probabilistic and deterministic approaches, and it also provides capabilities other than SHA, namely spectral matching and site response analysis. It comes with a comprehensive database of attenuation relations, and a regional seismic source database. It is also possible to purchase additional seismic source databases for almost the entire world. Furthermore, EZ-FRISK also allows user-defined seismic sources and attenuation relations through its graphical user interface (GUI).

FRISK88M [17], which is another commercial product by Risk Engineering Inc., provides advanced PSHA capabilities. It uses multiple weighted input parameters designated to represent both randomness and uncertainty, and follows a logic tree approach depending on those weighted parameters through PSHA computations. Although FRISK88M lacks a GUI and requires the seismic sources and the attenuation relations are specified in text input files, there are a few tools developed for input pre-processing and output post-processing.

OpenSHA [15] is a project, conducted jointly by Southern California Earthquake Center (SCEC) and USGS, for developing a framework for SHA. The goal of the project is described as to build a ”community modeling environment” for supporting interdisciplinary research in SHA. Several standalone applications are already implemented within the project. However, the main purpose of OpenSHA is to build a framework for SHA, where any SHA model can be plugged in and used in the analyses. In order to achieve this modular structure, an object-oriented approach is employed. The applications and the framework are implemented in Java programming language not only since Java is an object-oriented language, but also to provide platform-independence and to enable GUI and web-based access. Furthermore, SHA codes written in other programming languages may also be used by means of implementing wrappers. Although the source code has not been released yet; as its name implies, OpenSHA will be open source.

The major deficiency of the solutions mentioned above is the running time of the analysis computations. In case of large-scale analysis studies or when the analysis involves complex models that require intensive numerical calculations, SHA computations may require a few days to be completed on a single processor. Recent versions of EZ-FRISK uses multiple pro-

cessor cores to overcome this issue, and it is claimed to provide 40 to 60 percent decrease in execution time. However, multi-core processing does not suffice for many SHA studies. For example, analyses that involve logic tree computations may require intolerably long running times even when executed on a quad-core processor. Therefore; considering the computational power required, the use of grid computing is vital for actual SHA computations. OpenSHA project includes an application, developed by Field et al. [14], utilizing grid computing for hazard map calculations. However, the details regarding how SHA computations are distributed among available grid resources are not clearly explained and also the utilized grid is a relatively small one.

CHAPTER 5

A GRID-BASED APPROACH TO SHA

The main goal of our application is to perform SHA computations by utilizing grid computing resources. While providing alternative SHA models, it also enables the use of new models for assessing seismic hazards. In this chapter, the reasons for using grid computing will be explained first, and then the structure of the application will be described.

5.1 Use of Grid Computing

Depending on the complexity of the analysis models involved, the size of the analysis site, or the number of neighboring seismic sources; SHA calculations may require a great amount of processing power. The main idea behind our application is to utilize grid resources in order to provide that necessary processing power for SHA computations. This use of grid resources not only helps reduce the time required for analysis studies but also provides more precise results since discrete numerical calculations can be performed with finer granularity. Furthermore, this grid-based approach not only shortens the time required for evaluating SHA studies that incorporate logic tree methodology but also enables the use of a greater number of branches in such studies; and hence it helps to quantify and evaluate uncertainties associated with seismic hazards more precisely.

Another intent of our application is to benefit from the storage resources provided by grid infrastructure. Grid storage resources are used for storing the input data required for SHA calculations, such as information regarding seismic sources and site conditions. Depending on the analysis models used, the number of seismic sources available, or the variety of available site-related attributes; this input data may demand a great amount of storage. This grid-based

storage approach also allows future SHA models to use any type of data as input by only inserting the related information into this already available data repository.

5.2 Application Description

Our application consists of four main components as depicted in Figure 5.1. The ENGINE constitutes the core of our application, since utilization of grid resources for performing all SHA calculations is realized by this component. The DATA component is responsible for providing input data stored in grid storage resources when requested by the ENGINE. The data and the functions related to SHA models, i.e. distance distribution models and ground motion attenuation models, are provided by the MODEL component. The GUI component eases the usage of the application by means of a web-based interface.

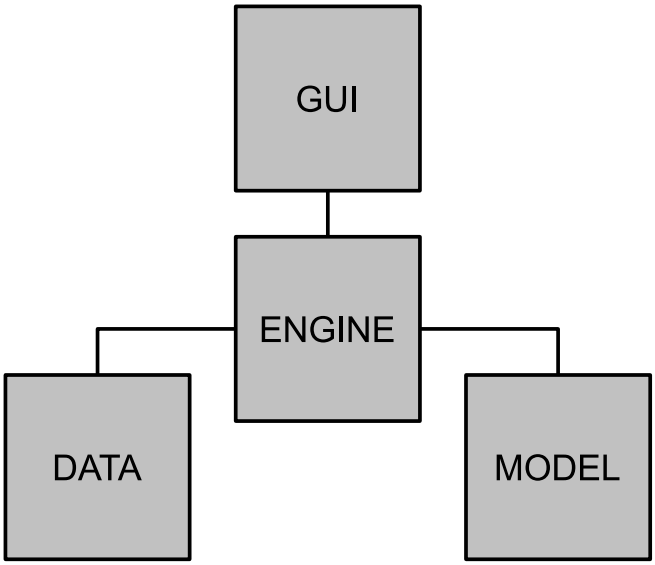

Figure 5.1: Components of the application

In the rest of this chapter, the details of these components and the interactions among them are described.

5.2.1 The DATA Component

The major task of the DATA component is to supply the required input data for the analysis models. The structure of the component is shown in Figure 5.2. The data provided by the DATA component can be examined in two main categories according to the method of access. The first category of data is accessed using an application service already deployed on the grid infrastructure. On the other hand, the second data category is accessed directly from grid storage elements.

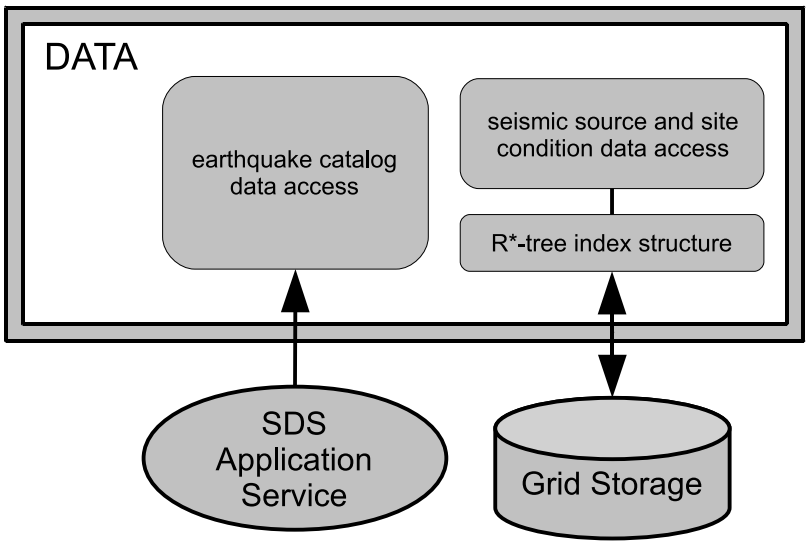

Figure 5.2: Structure of DATA component

The first data category involves earthquake information. As described in Chapter 2, information about the previous earthquakes is necessary for determining the maximum earthquake magnitudes and characterizing the distributions of earthquake magnitudes. Therefore, SHA calculations require access to a catalog of historical earthquake data. In the scope of our application, this access is provided through Seismic Data Server Application Service (SDSAS) [35]. The purpose of SDSAS is to serve seismic data provided by national seismology institutions of the countries in the South Eastern European (SEE) region. In addition to historical earthquake catalog data, station information and seismic waveform data are also available through the high-level interface, which is based on C++ iterators, provided by the application service. As this high-level iterator interface provides custom querying capabilities; instead of gathering all earthquakes, the DATA component is able to retrieve only the information for

earthquakes that are relevant to the current SHA study. The relevant earthquakes are determined based on their locations and magnitudes. Firstly, the earthquakes that are outside a rectangular region covering the site and selected seismic sources are eliminated. Secondly; since only the earthquakes with magnitudes greater than a particular minimum earthquake magnitude are considered in the analyses, the earthquakes with small magnitudes are eliminated. Furthermore; since the earthquake catalog involves various magnitude scales, the magnitudes of retrieved earthquakes are converted to moment magnitude, which is the mostly used magnitude scale in SHA models, using the following empirical relations derived by Yenier et al. [53]:

$$M_w = 0.571M_s + 2.484, \quad 3.0 \leq M_s < 5.5$$

$$M_w = 0.817M_s + 1.176, \quad 5.5 \leq M_s \leq 7.7$$

$$M_w = 0.953M_L + 0.422, \quad 3.9 \leq M_L \leq 6.8$$

$$M_w = 0.764M_d + 1.379, \quad 3.7 \leq M_d \leq 6.0$$

$$M_w = 1.104m_b + 0.194, \quad 3.5 \leq m_b \leq 6.3$$

where M_w is moment magnitude, M_s is surface wave magnitude, M_L is local magnitude, M_d is duration magnitude, and m_b is body wave magnitude.

The second data category involves all input data except earthquake information. This information includes data related to seismic sources and site conditions. The DATA component maintains this data as a repository on grid storage elements, and provides access to that repository as requested by the ENGINE. Since the repository involves spatial data, it requires a spatial index structure for efficient access. For indexing the seismic data repository in our application, R*-tree spatial index structure [4] is chosen. The R*-tree is a variant of the R-tree spatial index structure [27], which can be considered as a multi-dimensional generalization of the B^+-tree structure. The R-tree structure organizes the spatial objects by splitting the space into overlapping minimum bounding rectangles (MBR) that enclose the objects. While splitting the space, R-tree tries to optimize the areas of MBRs, whereas the R*-tree structure additionally uses other optimization criteria on overlaps and margins in order to improve performance of both point and rectangle queries. The DATA component maintains an R*-tree structure for the seismic data repository of our application by means of adapting SaIL spatial index library [28], which provides an efficient R*-tree implementation, for being used on grid storage elements.

Currently, 39 seismic source zones supplied by Demircioğlu et al. [11] are inserted into the seismic data repository. In addition to the coordinates of the polygonal regions representing the supplied area sources, seismic source information present in the repository also includes fault length and fault type information for the sources corresponding to actual fault lines or segments. The fault lengths are assigned by summing up the lengths, computed using provided coordinates, of fault lines and segments inside the seismic source zones. The types of corresponding faults are determined by manually inspecting and matching the provided fault lines with the ones compiled by Kayabalı and Akın [31], and using the types of the faults they provide.

One site condition attribute commonly required by models used in SHA calculations is the average shear-velocity from the surface to 30 meters depth (${V_S}^{30}$). USGS manages a global ${V_S}^{30}$ map server [21] that provides ${V_S}^{30}$ values for the whole world based on a method, proposed by Wald and Allen [51], correlating topographic slope with shear velocity. The map server provides predefined maps together with accompanying numerical data, and also allows custom maps to be constructed. Combining the predefined maps provided for İstanbul, Turkey, and Southern Europe; ${V_S}^{30}$ values for 18,729,288 different locations are inserted into the seismic data repository for being accessed as required by SHA models.

Some models used in SHA calculations require soil profile characteristics of the site, instead of using average shear velocity values. For such kind of models; the mapping between soil profile and shear velocity, defined by National Earthquake Hazards Reduction Program (NEHRP) [7], shown in Table 5.1 is used to determine the type of soil profile on the site region.

Table 5.1: NEHRP site classification

Site Class	Soil Profile	${V_S}^{30}$ Condition
A	Hard rock	${V_S}^{30} > 1500$ m/s
B	Rock	760 m/s $< {V_S}^{30} \leq 1500$ m/s
C	Very dense soil and soft rock	360 m/s $< {V_S}^{30} \leq 760$ m/s
D	Stiff soil	180 m/s $\leq {V_S}^{30} \leq 360$ m/s
E	Soil	${V_S}^{30} < 180$ m/s

5.2.2 The MODEL Component

As explained in Chapter 2, SHA involves various alternative analysis models for different parts of the calculations. The MODEL component maintains such alternative models. Implementations of magnitude distribution models and ground motion prediction models reside in this component as depicted in Figure 5.3.

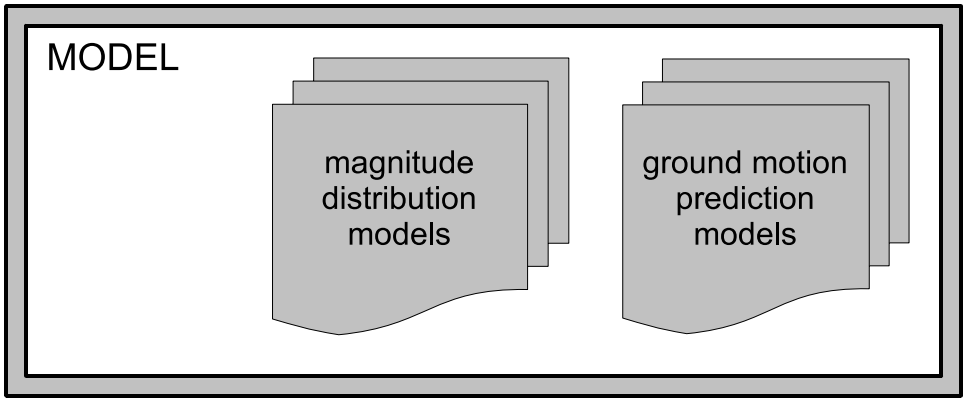

Figure 5.3: Structure of MODEL component

All magnitude distribution models described in Subsection 2.2.3, namely Gutenberg-Richter, bounded Gutenberg-Richter, and characteristic earthquake models, are already implemented. In addition to the ground motion attenuation relation by Cornell et al. [10] formulated in Equation 2.4, two other attenuation relationships proposed by Campbell [9] and Sadigh et al. [38] are also present as ground motion prediction models.

Furthermore, the MODEL component also allows adding new magnitude distribution and ground motion prediction models. The component provides abstract base classes that serve as interfaces for new models. To add a new magnitude distribution model a function representing the PDF of the distribution and two initialization functions must be implemented. One of the initialization function is for setting constant values included in the model that depend on study parameters that are invariant with respect to the site and sources, and the second initialization function is for setting the parameters of the model that depend on the site and source related attributes. The implementation of new ground motion prediction models involve defining two functions describing the model, namely the predicted mean and standard deviation functions corresponding to the random variable representing the ground motion intensity as described

in Section 2.1, together with again two initialization function similar to the ones in the case of magnitude distribution models.

5.2.3 The GUI Component

The GUI of our application is a web-based interface that interacts with the users. It is developed as a portlet for the P-GRADE Grid Portal [30], which provides a web-based environment for managing grid jobs on various grid platforms. The portlet developed for our application provides a graphical interface, seen in Figure 5.4, for setting the parameters involved in SHA studies.

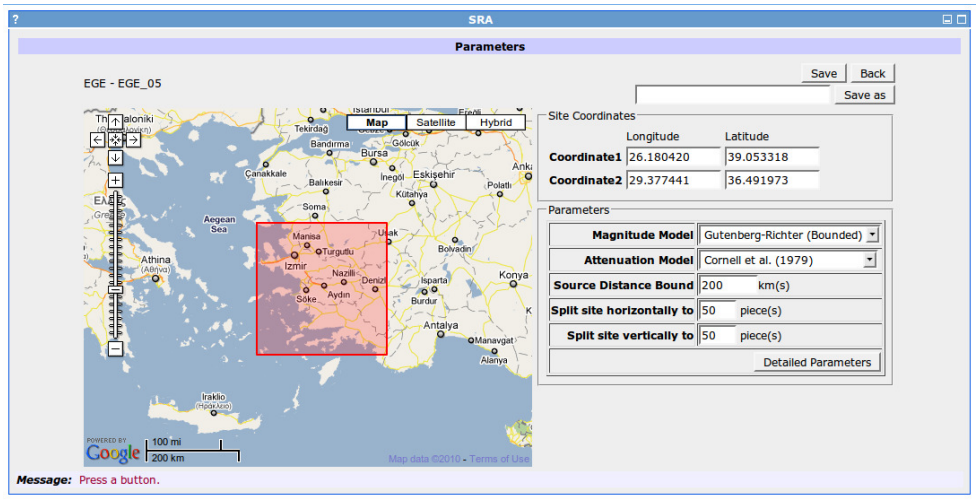

Figure 5.4: Parameter settings via web-based GUI

Setting the parameters for the study involves selecting the rectangular site region, choosing among available magnitude distribution and ground motion prediction models, and setting other study-specific parameters. The site region can be defined either by means of manually entering the coordinates of its two opposite corners or by selecting the rectangle using the embedded Google-powered map [24]. The selection among available ground motion prediction and magnitude distribution models, setting the *source distance bound*, which determines the sources to be considered in the calculations, and defining the number of subparts that the site region will be divided into can be performed via the graphical interface. Furthermore, the GUI also allows setting more advanced analysis parameters, such as the minimum earthquake

magnitude to be considered in the calculations, via a text area provided in *detailed parameters* page.

The GUI component organizes analysis studies of the users in the form of *projects*, each of which may contain one or more *parameter sets*. This organization allows the user to manage analysis studies in a systematic way. For instance, the effects of different parameters and models for a particular site may be investigated by creating different parameter sets for that particular site of interest in the scope of a single project. Furthermore, since each parameter set corresponds to single grid job that can be managed individually, by means of the interface seen in Figure 5.5, many analysis studies may be conducted simultaneously.

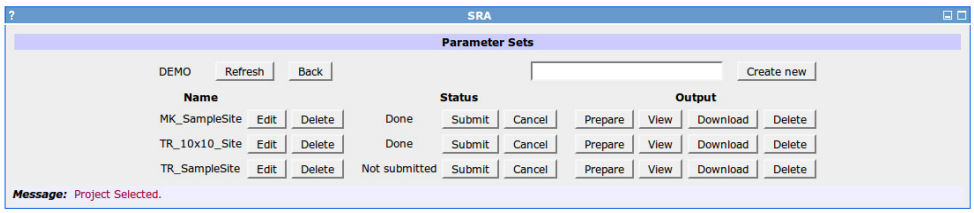

Figure 5.5: Management of different parameter sets in a project

Moreover, the GUI also provides the chance to directly visualize the analysis results by means of drawing resulting output graphs in the web browser. The user may directly download the results as an archive file, or decide whether or not to download after examining the output graphs.

5.2.4 The ENGINE Component

The ENGINE is the backbone of our application. All of the calculations related to the analysis are performed within the scope of this component. Furthermore, utilization of grid computing resources is also the responsibility of the ENGINE. The general structure of the component is depicted in Figure 5.6.

The *pre-processing module* prepares the necessary setup for grid job submission. Either a predefined number of grid jobs are prepared for the analysis study, or the parameters supplied by the user via the GUI and current availability of grid computing resources are examined in

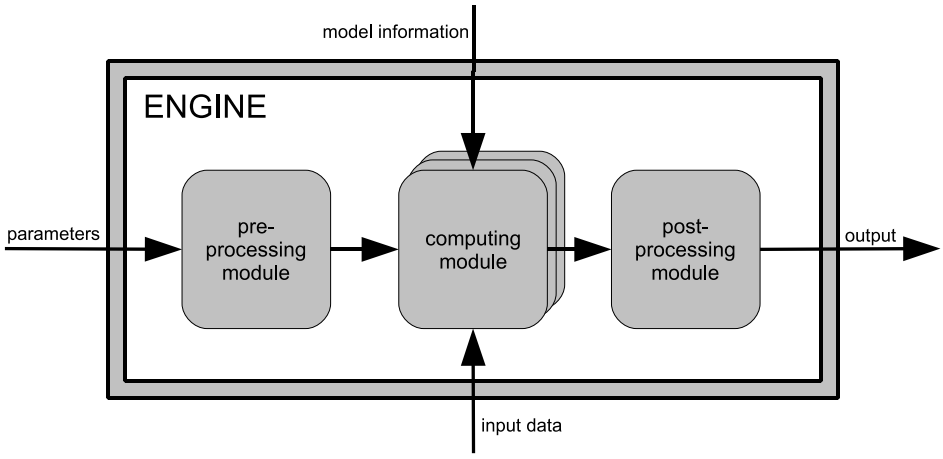

Figure 5.6: Structure of ENGINE component

order to determine the number of grid jobs that will be submitted for the analysis. When the site region is divided into sub-regions according to the parameters, each grid job handles the calculations for equal (or nearly equal) number of those sub-regions.

The implementations of core SHA calculations reside in the *computing module*. This part of the ENGINE can be thought as representing the executable and accompanying files transferred to computing elements (CE) on the grid. Hence, this module involves multiple instances as depicted in Figure 5.6. The following operations are performed on the CEs by the computing module:

1. The information supplied by pre-processing module, in the form of a parameter file and command line arguments, is processed in order to determine the analysis parameters, and what types of outputs will be generated for which part of the site region.

2. The necessary input data is obtained through the DATA component. This input data includes the seismic sources near the site region, corresponding earthquake information, and information regarding the site conditions.

3. The information regarding the selected ground motion prediction and magnitude distribution models are obtained through the MODEL component.

4. Preliminary calculations are performed for characterizing the chosen seismic sources using the earthquake catalog data obtained. As described in Chapter 2, these calcula-

tions involve determining earthquake recurrence constants, namely a and b values, the maximum magnitude, and rate of earthquakes with magnitudes greater than the minimum value.

5. Finally; core SHA calculations are performed using the gathered input data and the selected models, and the resulting data for requested output types are stored in binary format for being processed by post-processing module.

As already mentioned, the site region is divided into sub-regions in the scope of our application. Similarly; while performing SHA calculations the seismic sources are also split into smaller equal-sized cells, as illustrated in Figure 5.7, which are then considered as point sources.

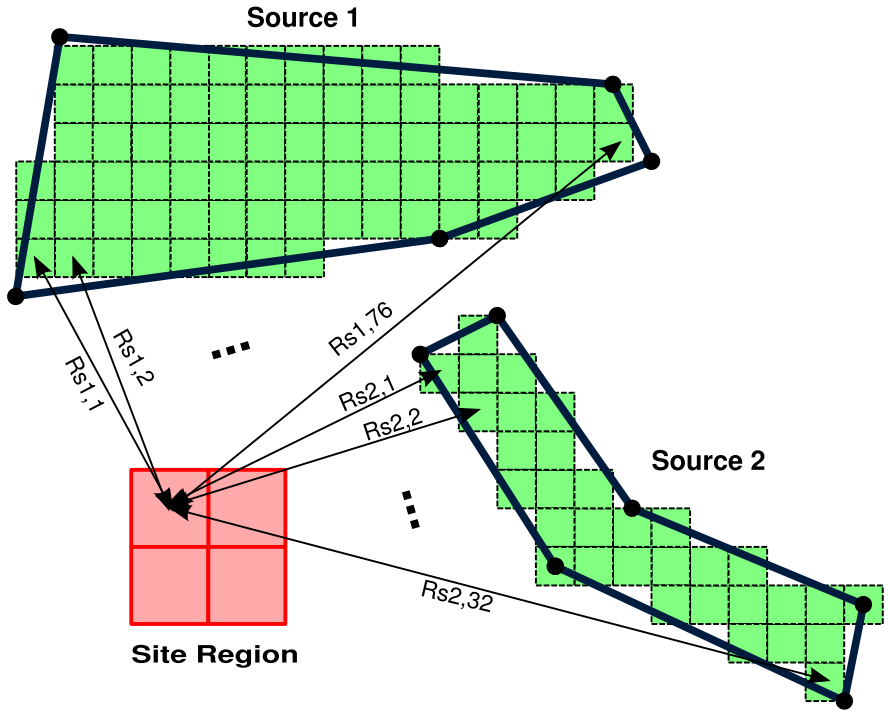

Figure 5.7: Gridded view of the site and seismic sources

This gridded seismic source approach eliminates the computational burden of deriving the source-to-site distance distributions for arbitrarily shaped seismic sources and allows the use of the simple source-to-site distance PDF, formulated in Equation 2.5, for point sources. Furthermore, the use of this PDF for point sources enables the calculation of total exceeding rates

by the following simpler formula, eliminating the double integral involved in Equation 2.13.

$$\lambda(Y > y) = \sum_{i=1}^{n_{sources}} \lambda(M_i > m_{min}) \int_{m_{min}}^{m_{max}} P(Y > y \mid m, R_i)\ f_{M_i}(m)\ dm. \tag{5.1}$$

It should be noted here that the above formulation is valid only when all the sources involved in the analysis are point sources. In the scope of our application, the sources in the above equation correspond to the cells which the actual seismic sources are divided into. Moreover; the rates of observing earthquakes with magnitudes greater than the minimum earthquake magnitude, i.e. $\lambda(M_i > m_{min})$, for these cell-type point sources directly relate to the rates for their corresponding sources. Considering the randomness of earthquake occurrences, a uniform distribution of these rates throughout the source zone is assumed; and the relation between them is formulated as follows:

$$\lambda(M_i > m_{min}) = \frac{1}{N_{cells}}\ \lambda(M > m_{min}),$$

where $\lambda(M > m_{min})$ is the rate of earthquakes with magnitudes greater than m_{min} for the corresponding seismic source, and N_{cells} is the number of cells that the source is divided into.

Our application is capable of producing the following types of output for PSHA:

- *Annual rate of exceedance*: The calculated values indicate how frequently events with ground motion intensities exceeding a given level occur per year in the site of interest. Annual exceeding rates are calculated using Equation 5.1 for different levels of ground motion intensities. Since the equation calculates the exceedance values for a single location, the rates calculated for sub-regions should be combined in order to determine the annual rates of exceedance for the whole site region. Again a uniformness assumption, similar to the one mentioned above for seismic source, is incorporated; and the total rate of exceedance for the site region is calculated using the following formula:

  $$\lambda(Y > y) = \frac{1}{N_{cells}}\ \sum_{i=1}^{N_{cells}} \lambda(Y_i > y), \tag{5.2}$$

  where $\lambda(Y_i > y)$ is the annual rate of exceeding intensity level y for the i^{th} sub-region, and N_{cells} is the total number of sub-regions that the site is split into.

- *Probability over years*: This output type describes the probabilities of observing earthquake events which are expected to produce ground motions, with intensities exceeding

a certain level, at the site region over the forthcoming years. Calculation of such probabilities require information about the inter-event times of earthquakes, and these times are most commonly modeled using Poisson distribution [3]. Hence, the probability of observing at least one event within a specific time period is calculated by the following formula:

$$P(\text{at least one event with } Y > y \text{ in time } t) = 1 - e^{-\lambda(Y>y)\,t}. \tag{5.3}$$

Our application calculates the probabilities over different year periods for a particular ground motion intensity level by utilizing the above formula. These probability values are calculated using the exceedance rate for the whole site region.

- *Magnitude deaggregation*: This type of output is used for investigating the probabilities of earthquake scenarios with different magnitudes, given the probable occurrence of a ground motion with intensity exceeding a certain level. Baker [3] presents the following formula for magnitude deaggregation in terms of rates:

$$P(M = m \mid Y > y) = \frac{\lambda(Y > y\ ,\ M = m)}{\lambda(Y > y)},$$

where $\lambda(Y > y\ ,\ M = m)$ represents the rate of occurrences for earthquake events, with magnitude m, causing ground motions with $Y > y$; and it is calculated, by a simple modification to Equation 5.1, as follows:

$$\lambda(Y > y\ ,\ M = m) = \sum_{i=1}^{n_{sources}} \lambda(M_i > m_{min})\ P(Y > y \mid m, R_i)\ P(M_i = m). \tag{5.4}$$

For a given specific ground motion intensity level, our application calculates the deaggregation values for different magnitudes.

- *Probability map*: In order to help a better visualization of seismic hazard at the site region, our application produces a probability map. This map is constructed for a specific ground motion level using the probabilities calculated particularly for each sub-region of the site.

For DSHA studies, our application calculates the ground motion intensity value for the worst-case earthquake scenario at the site region as described in Chapter 2. For each sub-region in the site of interest, the closest distances to every selected source are determined; and then using those distances and maximum earthquake magnitudes specific to each source, the chosen attenuation relation is evaluated. Among the calculated ground motion intensity values,

the maximum one is determined for each sub-region in the site; and finally among the sub-regions the one with the maximum intensity value calculated is reported. Furthermore, annual rate of exceedance and probability values are also provided for that deterministically calculated ground motion intensity value.

As the computing module finishes the calculations, binary output produced are transfered back from the grid CEs. Afterwards, the *post-processing module* combines all the outputs generated by the submitted grid jobs, and produces the actual SHA results. For combining the outputs related to probability map, only concatenation of produced values into a single file is adequate. Whereas, for other types of outputs, some simple calculations are necessary. For instance; to produce the actual annual rate of exceedance results, the produced values are summed up considering the number of sub-regions handled by each grid job.

Furthermore; for helping easier visual investigation of the analysis results, the post-processing module also prepares three graphs for annual rate of exceedance, probability over years, and magnitude deaggregation outputs. The task of drawing the probability map mentioned above is also handled by this module.

CHAPTER 6

IMPLEMENTATION AND TESTING

In this chapter; first, the details regarding the implementation are given, and then the tests performed are described.

6.1 Implementation Details

Our application is developed mainly using C++ programming language. The DATA and MODEL components are implemented completely using C++ codes. Moreover; in the ENGINE component, the whole computing module and the output combining part of the post-processing module are also implemented using C++. Other parts of the ENGINE component are implemented as numerous shell scripts. Java programming language and JavaServer Pages (JSP) [29] technology are used for implementing the GUI, since it is developed as a portlet for the P-GRADE Grid Portal as already mentioned.

6.1.1 Implementation of the DATA component

As described in Subsection 5.2.1; the DATA component uses the C++ iterators provided by SDSAS for accessing earthquake catalog data, and the R*-tree spatial index structure implementation provided by SaIL for storing and accessing seismic source and site information. By default, SaIL allows memory- and disk-based usage for the R*-tree indexes. However, it also provides an interface, namely *IStorageManager*, for implementing user-defined storage management for both index and data entries. In order to adapt the R*-tree implementation provided by SaIL for being used in the Grid infrastructure; a new storage manager, which is able to access the files stored on the Grid storage elements (SE), is implemented. The

implemented storage manager uses the Grid File Access Library (GFAL) [18] functions for performing file operations on the SEs.

6.1.2 Implementation of the MODEL component

The MODEL component, as explained in Subsection 5.2.2, provides interfaces for magnitude distribution and ground motion attenuation models. Abstract C++ base classes with virtual member functions are used for the purpose of defining these interfaces. Furthermore, implementations for alternative models also reside in this component in the form of C++ classes which inherit from the provided abstract classes.

6.1.3 Implementation of the GUI component

As already mentioned, the implementation of the GUI component involves the usage of Java programming language and JSP technology. In particular, JSP technology is used for constructing the visual parts shown on the tabbed interface of the P-GRADE Portal. For example, direct visualization of analysis results in terms of graphs is provided using JSP codes. Java codes, on the other hand, provide the required back-end functionality. For instance, Java class methods are used for managing a particular directory tree structure that provides the project-based organization described in Subsection 5.2.3 for analysis studies.

6.1.4 Implementation of the ENGINE component

In the ENGINE component, the whole computing module and a part of the post-processing module are implemented using C++ programming language, as previously stated. Other parts of this component are implemented as shell scripts written for the GNU Bourne Again Shell (BASH) [48].

The pre-processing module includes scripts for Grid job management. These scripts handle job related operations by means of invoking the relevant commands provided by gLite Workload Management System (WMS) [20].

The computing module uses GNU Scientific Library (GSL) [25] for performing numerical calculations. The integration operation in the calculation of total rate of exceedance (Equa-

tion 5.1), the evaluation of standard normal cumulative distribution function in the calculation of ground motion intensity exceeding probability (Equation 2.10), and the least squares estimation of earthquake recurrence for seismic sources (described in Section 2.1) are performed using relevant functions provided by GSL.

Since analysis calculations involve the distances between site and source points, the calculation of the distance between two points is necessary. For this purpose; first, the following formula (historically known as the *haversine formula*) is used for determining the angular difference between them [45]:

$$\Delta\widehat{\sigma} = 2\arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 \sin^2\left(\frac{\Delta\lambda}{2}\right)}\right),$$

where $\Delta\widehat{\sigma}$ is the spherical angular difference between the two points, ϕ_1 and ϕ_2 represent the latitudes of the two points, and $\Delta\phi$ and $\Delta\lambda$ are the angular differences between their latitudes and longitudes respectively. Afterwards, the distance is calculated as $r\Delta\widehat{\sigma}$, where r is the average radius of the earth, which is taken as 6371.01 kilometers.

Furthermore, in order to characterize the properties of seismic sources the relevant earthquakes should be determined. This process requires testing whether an earthquake location is inside or outside the polygonal region representing a seismic source. For this purpose, a simple inside-outside test for polygons which involves the odd-even rule is implemented.

As mentioned in Subsection 5.2.4, the post-processing module prepares three output graphs and a probability map. Similar to the case for the pre-processing module, shell scripts are used for automating these tasks. Gnuplot [22] graphing utility is used for plotting the resulting graphs. An example annual rate of exceedance graph can be seen in Figure 6.1.

For the case of the probability map, The Generic Mapping Tools (GMT) [47] collection is used. Contouring tool *pscontour* included in GMT is used for producing contours of the probability values obtained throughout the site region. Using GMT, the probability map is produced both as an image file and as a KMZ archive that can be further examined using Google Earth [23]. In Figure 6.2 an example probability map can be seen.

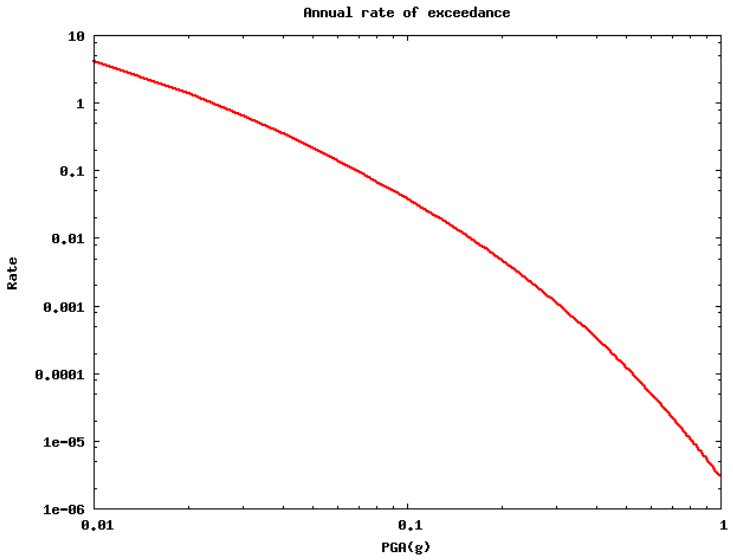

Figure 6.1: An example annual rate of exceedance graph

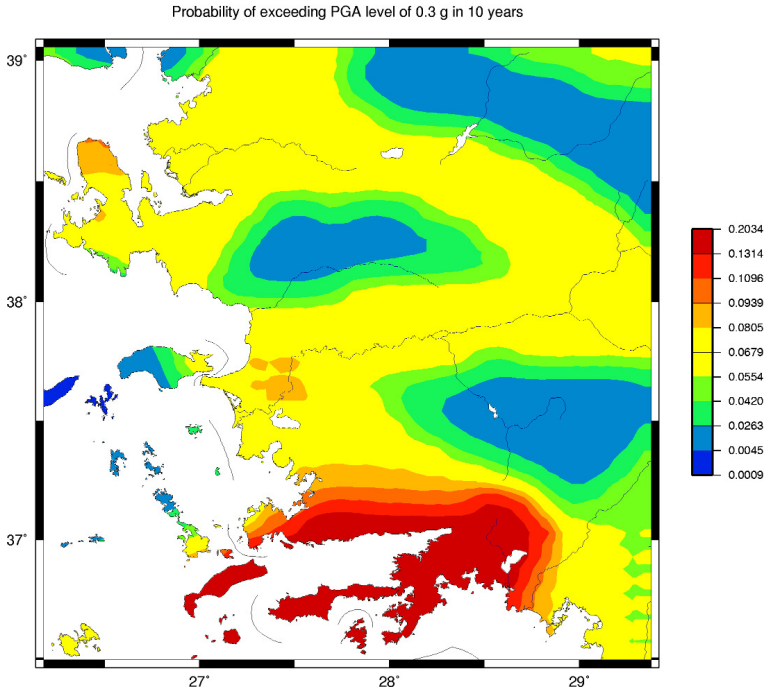

Figure 6.2: An example probability map

6.2 Testing Details

In the following part, after explaining the details regarding validation of the developed application, the performance and scalability of the application are examined. Finally, a comparison of the application with the other SHA software packages is presented.

6.2.1 Validation

For the purpose of validating the calculations performed by our application, a verification study for PSHA software [50] is chosen as the reference. The study evaluates the calculations performed by many different software packages, including EZ-FRISK, FRISK88M, OpenSHA, and the Fortran programs from USGS mentioned in Chapter 4. Two sets of test cases are used for the numerical verification of the software. The first set is designed for testing the basic characteristics of the programs, such as the strategy used for modeling fault planes and area sources, the usage of recurrence models, and the integration of the standard deviation in attenuation relationship calculations. The second test case set is used for testing more advanced concepts, such as multiple seismic sources, deaggregation, and logic tree based calculation.

From the first test set, a single test case (case 10) is selected for numerical evaluation of our application. This choice is made based on the fact that this case is the only one which tests area source related calculations in the first set. Although the second test set includes a case which also involves area sources, the cases in that second set cannot be used for verification since their solutions are not provided.

The selected test case includes a circular area source and four site locations as depicted in Figure 6.3. The coordinates for the source and the sites are provided, and it is also given that the area source has a fixed depth of 5 kilometers. For the source, the b-value is given as 0.9, and the value 0.0395 is provided as the annual rate of observing earthquakes with magnitudes greater than 5. Furthermore, a grid spacing of 1 kilometer is given for the source. The truncated exponential model (the bounded Gutenberg-Richter model in our terminology), with bounds $m_{max} = 6.5$ and $m_{min} = 5.0$, is selected as the magnitude distribution model to be used. The attenuation relation by Sadigh et al. [38] is chosen, but its standard deviation is taken as 0. The site locations are assumed to have rock-type soil profile.

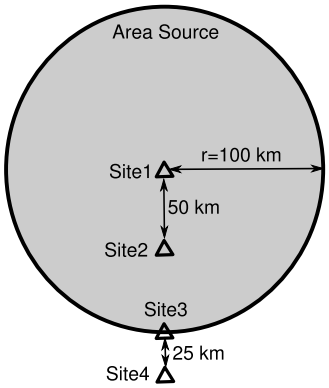

Figure 6.3: The area source and site locations for the test case, adapted from [50]

After fixing the provided study parameters, and modifying the standard deviation function of the selected attenuation relationship; the given coordinates for the source and site locations are supplied as inputs to our application. Here, it should be noted that a grid spacing of 0.009 degrees, which approximates the grid spacing of 1 kilometer for locations nearby the given source, is used since our application divides the seismic sources into cells using a degree-based approach. Once the testing environment is set, the annual probabilities of exceeding different PGA levels are calculated for each site. The results are provided in Table 6.1 together with the mean values for the software tested in the original study.

In Table 6.1, the relative errors between the mean and calculated probabilities are also provided. As observed from the table, our calculations deviate from the mean values by at most 2.91 percent, and the average 0.93 percent deviation demonstrates the validity of our implementation.

Table 6.1: The results of validation tests

	Site 1			Site 2		
PGA (g)	Mean	Calculated	Error	Mean	Calculated	Error
0.001	3.87E-02	3.87E-02	0%	3.87E-02	3.87E-02	0%
0.01	2.19E-02	2.18E-02	0.46%	1.82E-02	1.81E-02	0.55%
0.05	2.97E-03	2.96E-03	0.34%	2.96E-03	2.94E-03	0.68%
0.1	9.22E-04	9.18E-04	0.43%	9.21E-04	9.12E-04	0.98%
0.15	3.59E-04	3.59E-04	0%	3.59E-04	3.57E-04	0.56%
0.2	1.31E-04	1.32E-04	0.76%	1.31E-04	1.31E-04	0%
0.25	4.76E-05	4.71E-05	1.05%	4.76E-05	4.68E-05	1.68%
0.3	1.72E-05	1.68E-05	2.33%	1.72E-05	1.67E-05	2.91%
0.35	5.38E-06	5.35E-06	0.56%	5.37E-06	5.33E-06	0.74%
0.4	1.18E-06	1.20E-06	1.69%	1.18E-06	1.20E-06	1.69%

	Site 3			Site 4		
PGA (g)	Mean	Calculated	Error	Mean	Calculated	Error
0.001	3.87E-02	3.87E-02	0%	3.82E-02	3.83E-02	0.26%
0.01	9.29E-03	9.32E-03	0.32%	5.31E-03	5.33E-03	0.38%
0.05	1.37E-03	1.39E-03	1.46%	1.24E-04	1.25E-04	0.81%
0.1	4.37E-04	4.41E-04	0.92%	1.67E-06	1.63E-06	2.4%
0.15	1.74E-04	1.76E-04	1.15%			
0.2	6.42E-05	6.47E-05	0.78%			
0.25	2.31E-05	2.27E-05	1.73%			
0.3	8.32E-06	8.45E-06	1.56%			
0.35	2.65E-06	2.66E-06	0.38%			
0.4	5.96E-07	5.84E-07	2.01%			

6.2.2 Performance and Scalability

As described in Subsection 5.2.4, all the produced SHA outputs involve the rates of exceeding particular intensity levels. Furthermore, as can be observed from Equation 5.1, calculation of these rates mainly requires the numerical integration of the related functions provided by the selected ground motion prediction and magnitude distribution models for the analysis study. Hence, the complexity of the used models directly affects the running time of SHA computations. For demonstrating the effects of using different models, the calculation times are measured for a sample analysis study. The study involves a relatively small soil site, which is divided into 100 sub-regions, and the seismic sources within a 250 kilometers radius of the site region. The results are summarized in Table 6.2.

Table 6.2: The calculation times for a sample SHA study using different models

Magnitude Model	Attenuation Model	Calculation Time (s)
Bounded Gutenberg-Richter	Cornell et al.	329
Bounded Gutenberg-Richter	Campbell	600
Bounded Gutenberg-Richter	Sadigh et al.	1692
Gutenberg-Richter	Cornell et al.	244
Gutenberg-Richter	Campbell	588
Gutenberg-Richter	Sadigh et al.	1453
Characteristic	Cornell et al.	4324
Characteristic	Campbell	3131
Characteristic	Sadigh et al.	3491

As can be observed from Table 6.2, the selected models directly influence the running times. The Gutenberg-Richter model requires less time for calculations compared to its bounded version. The characteristic model causes longer calculation times, since it involves more complex calculations than the other magnitude distribution models. For the case of ground motion attenuation models, the simple model proposed by Cornell et al. is the fastest one, except when used together with the characteristic model. This behavior is due to the numerical integration involved in the calculations. The attenuation models proposed by Campbell and Sadigh et al. together with the characteristic magnitude model constitute functions that can be integrated in fewer iterations compared to the attenuation model by Cornell et al.

Equations 5.1 and 5.2 further indicate that the numerical integration is performed for all pairs of source and site cells while calculating the rates. Hence, both the number of cells that the

site region is divided into and the number of cell-type point sources that represent the seismic sources affect the time required for computations directly. The following formulation can be used to describe the running time requirement for calculating the rate of exceeding a specific ground motion intensity level, t_{rate}, in terms of the time required for numerical integration, $t_{integration}$:

$$t_{rate} = N_{site\_cells} \cdot N_{source\_cells} \cdot t_{integration}, \tag{6.1}$$

where N_{site_cells} is the number of sub-regions that the site is split into and N_{source_cells} is the total number of cell-type point sources that the seismic sources are divided into.

Furthermore, the running time of the calculations depends on the types of the outputs that will be produced. For instance, the annual rate of exceedance output involves the calculation of exceeding rates for different intensity levels. On the other hand, each of the other output types requires the rate to be calculated only once for a specific ground motion intensity value. Hence, the total running time required for calculations, $t_{calculation}$, can be formulated as follows, by using the formula for the rate calculations presented in Equation 6.1:

$$\begin{aligned} t_{calculation} &= N_{IM\_levels} \cdot t_{rate} \\ &= N_{IM\_levels} \cdot N_{site\_cells} \cdot N_{source\_cells} \cdot t_{integration}, \end{aligned} \tag{6.2}$$

where N_{IM_levels} is the number of different intensity levels involved in the requested outputs. By default, our application calculates the annual rate of exceedance values for 100 different intensity levels and one exceeding rate for each of the other output types. Hence, the default value of N_{IM_levels} is 104. Here, it should be noted that the output types other than the annual rate of exceedance also involve different types of calculations. However, the time required for those calculations is negligible since it is dominated by the time required for numerical integration. For instance, the magnitude deaggregation output type requires evaluating the involved SHA models as seen in Equation 5.4. However, since the numerical integration involves many such evaluations, the time required for those evaluations can be neglected. Similarly, the probability over years output type and the probability mapping process require the calculation of probabilities which involves exponentiation, as seen in Equation 5.3. However, the time required for exponentiation can also be neglected in the presence of numerical integration.

Apart from the running time of the calculations, the total analysis time further depends on the time required for obtaining the input data. The time required to complete this data gathering

process mainly depends on the size of the site region. As the size of the site region increases, more time is required to obtain the site related data from the Grid SEs since the size of the data also increases. In addition, the number of seismic sources, which are selected as described in Subsection 2.2.1, affects the time required for obtaining input data. As the number of selected seismic sources increases, the size of source related data, i.e. source information and earthquake catalog data, also increases and more time is required for obtaining the data.

Since the necessary calculations are shared equally (or almost equally) among the Grid jobs constructed for the analysis, as described in Subsection 5.2.4, our application is expected to scale out optimally in terms of calculation time. However, the aforementioned input data gathering process is repeated in each Grid job. Thus, the expected total running time of an analysis on the Grid, $t_{expected}$, can be formulated as follows:

$$t_{expected} = t_{data} + \frac{1}{N_{jobs}} \cdot t_{calculation}, \tag{6.3}$$

where t_{data} represents the time required to gather the input data, N_{jobs} is the number of Grid jobs constructed for the analysis, and $t_{calculation}$ represents the time required to complete the calculations on a single processor.

For testing the scalability of our application, a sample SHA study is executed using different numbers of Grid jobs. This study involves a relatively large site region, which is split into 2,500 sub-regions, and the seismic sources within a 300 kilometers radius of the site. Using the bounded Gutenberg-Richter magnitude model and the attenuation model of Cornell et al., the analysis calculations are performed by constructing 1, 2, 4, 5, 10, 20, 25, and 50 Grid jobs. The obtained time measurements are illustrated as a graph in Figure 6.4. This graph shows the times spent both for the calculations and for obtaining the input data. Here, it should be noted that the provided time values are the maximum ones obtained among the Grid jobs for a particular analysis instance. Furthermore, the total running times, i.e. the longest total times observed among the corresponding Grid jobs, are also provided in Figure 6.4.

The measurements depicted in Figure 6.4 indicate that our application scales out almost optimally in terms of calculation times for the cases with 2, 4, 5, and 10 Grid jobs. For larger numbers of Grid jobs, although shorter calculation times are observed, the measured values differ from the expected calculation times. This behavior can be explained by the heterogeneity of the Grid infrastructure. Since the hardware configurations of the WNs vary among

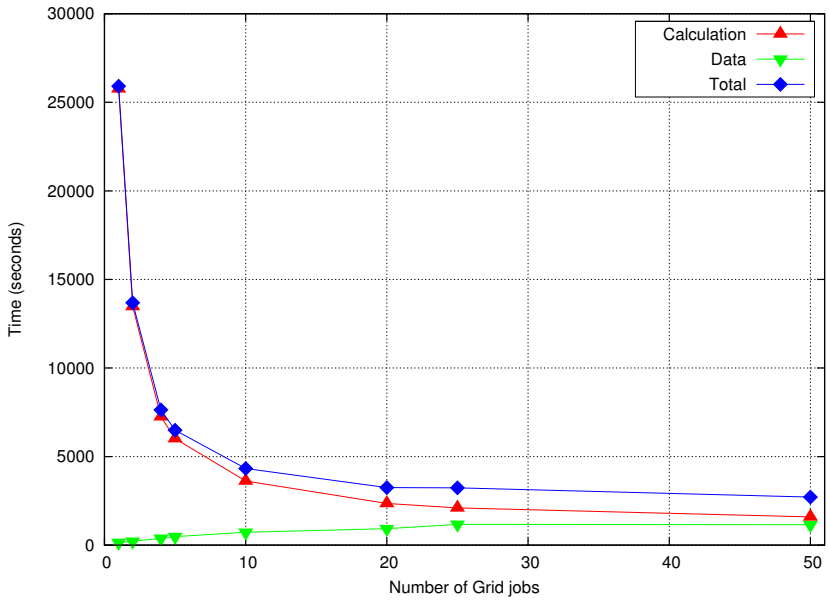

Figure 6.4: Time measurements for different numbers of Grid jobs

different Grid sites, the times required to perform the same calculations may also vary among them. As the number of Grid jobs constructed for an analysis study increases, the chance that the jobs are executed on different Grid sites also increases, probably causing longer running times.

Another observation that can be made from Figure 6.4 is about the data gathering times. Although all the measurements involve the same site region and the same seismic sources, the time spent for obtaining the input data varies among different Grid job configurations. The reason behind this variation can be explained by the loads due to the simultaneous data access requests on the Grid SEs and SDSAS. Since the calculation times will dominate the data gathering times, especially in large-scale analyses; this situation does not interfere with the scalability of our application.

Finally, using the calculation time measurement of the single-job case in Figure 6.4, the formula for total calculation time derived in Equation 6.2 can be verified. For this purpose, the time requirement of numerical integration should be determined first. Since the bounded Gutenberg-Richter magnitude model and the attenuation model by Cornell et al. are used in the analysis, the related time measurement in Table 6.2 can be used for obtaining that necessary time value. Since a total of 1,688 source cells and 100 site cells are constructed in

the analysis study for comparing the effects of using different models, the time required for numerical integration can be determined as 0.000018741 seconds. On the other hand, the SHA study for testing the scalability of our application involves 2,500 site cells, as already mentioned, and a total of 5,145 source cells. Hence, by using Equation 6.2, the expected calculation time on a single processor can be calculated as 25,070 seconds conforming to the obtained value of 25,762 seconds for the single-job case.

6.2.3 Discussion

In this part, the developed application is compared with the other software packages described in Chapter 4. This comparison is based on the aspects of extensibility and parallel execution capability.

Regarding extensibility, our application and OpenSHA are the most successful solutions. Both software provide interfaces for implementing new SHA models and allow those new models to be plugged into the analyses. On the other hand, EZ-FRISK provides some extensibility by allowing user-defined attenuation relations. However, this feature does not provide adequate freedom in developing new models. Since FRISK88M and SEISRISK III both require the attenuation models to be supplied in some pre-defined formats, they do not have the capability to be extended. The hazard mapping software provided by USGS also does not offer any ways to be extended apart from modifying the source codes.

Although the excessive processing power requirement in SHA calculations necessitates parallel execution, only EZ-FRISK, OpenSHA, and our application provide this capability. Although EZ-FRISK offers multi-core processing, a single machine is not adequate considering the high amount of processing power required especially for large-scale analyses. OpenSHA utilizes a small Grid of 100 workstations for performing hazard mapping calculations. However, even such a Grid may not be adequate for performing calculations in large-scale SHA studies. On the other hand, our application utilizes both the computing and storage resources available in SEE-GRID infrastructure, described in Section 3.3, and constitutes a powerful solution for parallel execution of SHA computations.

CHAPTER 7

CONCLUSION

Due to the dynamic and uncertain nature of seismic hazards, SHA studies require continuous updating. However, the uncertain nature of earthquake-related phenomena also causes the models involved in SHA to become extremely complex in terms of the calculations required. This complexity leads to the fact that SHA calculations demand high amounts of computational resources.

In this thesis, the power of grid computing is utilized for the purpose of supplying SHA studies with the necessary computational resources. The Grid-based SHA application developed not only provides a powerful infrastructure for existing SHA models but also offers a chance to quickly evaluate and validate possible new models. The developed application shortens the time required to complete analysis calculations by means of allowing parallel execution. Furthermore, our Grid-based application also provides a way to utilize the available grid resources for efficiently accessing to SHA-related spatial data.

As the results of the tests performed indicate, the implemented Grid-based application is able to perform SHA calculations correctly and within acceptable durations. Hence, the developed application enables large-scale seismic analyses to be completed in reasonable time, providing the opportunity to update SHA results on a continuous basis. Although only a subset of the available SHA models are currently implemented, the application provides the necessary flexibility for implementing and using any other model. This flexibility further helps the development of new and arbitrarily complex SHA models by means of providing a powerful framework that can be used for testing and validating purposes throughout the development process.

One possible extension to this work may be about utilizing the storage resources available in the Grid infrastructure. By utilizing virtually unlimited grid storage resources, it is possible to store the results of the analyses. Those stored results later may be used not only for helping facilitate the collaboration among different researchers but also for preventing repetition of previously performed analysis studies.

REFERENCES

[1] 2008 NSHM Software. http://earthquake.usgs.gov/hazards/products/conterminous/2008/software/. Last visited on 23 August 2010.

[2] J. G. Anderson and M. D. Trifunac. Uniform risk functionals for characterization of strong earthquake ground motion. *Bulletin of the Seismological Society of America*, 68(1):205, 1978.

[3] J. W. Baker. An introduction to probabilistic seismic hazard analysis (PSHA). White paper, version 1.3, 2008.

[4] N. Beckmann, H. P. Kriegel, R. Schneider, and B. Seeger. The R*-tree: an efficient and robust access method for points and rectangles. In *Proceedings of the 1990 ACM SIGMOD international conference on Management of data*, pages 322–331. ACM Press New York, NY, USA, 1990.

[5] B. Bender and D. M. Perkins. SEISRISK III: a computer program for seismic hazard estimation. *U.S. Geological Survey Bulletin 1772*, 1987.

[6] R. H. Budnitz, G. Apostolakis, D. M. Boore, L. S. Cluff, K. J. Coppersmith, C. A. Cornell, and P. A. Morris. Recommendations for probabilistic seismic hazard analysis: Guidance on uncertainty and use of experts. NUREG/CR-6372, Volume 1, U.S. Nuclear Regulatory Commission, 1997.

[7] Building Seismic Safety Council (BSSC). NEHRP recommended provisions for seismic regulations for new buildings and other structures. 2003 Edition, FEMA-450, Federal Emergency Management Agency, Washington, DC, 2003.

[8] S. Burke, S. Campana, E. Lanciotti, P. M. Lorenzo, V. Miccio, C. Nater, R. Santinelli, and A. Sciabà. *gLite 3.1 User Guide*. Enabling Grids for E-science (EGEE), 2009.

[9] K. W. Campbell. Empirical near-source attenuation relationships for horizontal and vertical components of peak ground acceleration, peak ground velocity, and pseudo-absolute acceleration response spectra. *Seismological Research Letters*, 68:154–179, 1997.

[10] C. A. Cornell, H. Banon, and A. F. Shakal. Seismic motion and response prediction alternatives. *Earthquake Engineering & Structural Dynamics*, 7(4):295–315, 1979.

[11] M. Demircioğlu, K. Sesetyan, E. Durukal, and M. Erdik. Assessment of earthquake hazard in Turkey. In *4th International Conference on Earthquake Geotechnical Engineering (ICEGE), Thessaloniki - Greece*, June 2007.

[12] EZ-FRISK - Software for Earthquake Ground Motion Estimation. http://www.ez-frisk.com/. Last visited on 23 August 2010.

[13] L. Ferreira, V. Berstis, J. Armstrong, M. Kendzierski, A. Neukoetter, M. Takagi, R. Bing-Wo, A. Amir, R. Murakawa, O. Hernandez, J. Magowan, and N. Bieberstein. Introduction to Grid computing with Globus. IBM Redbook SG24-6895-01, IBM Corp., International Technical Support Organization, 2003.

[14] E. H. Field, V. Gupta, N. Gupta, P. Maechling, and T.H. Jordan. Hazard map calculations using grid computing. *Seismological Research Letters*, 76(5):565, 2005.

[15] E. H. Field, T. H. Jordan, and C. A. Cornell. OpenSHA: A community-modeling environment for seismic hazard research. *Seismological Research Letters*, 74(4):406–419, 2003.

[16] I. Foster, C. Kesselman, and S. Tuecke. The anatomy of the Grid: Enabling scalable virtual organizations. *International Journal of High Performance Computing Applications*, 15(3):200, 2001.

[17] FRISK88M - Risk Engineering-Software. http://riskeng.com/SoftwareHTML/software_frisk.html. Last visited on 23 August 2010.

[18] GFAL man pages. http://grid-deployment.web.cern.ch/grid-deployment/gis/GFAL/GFALindex.html. Last visited on 11 September 2010.

[19] gLite - Lightweight Middleware for Grid Computing. http://glite.web.cern.ch/glite/. Last visited on 27 September 2010.

[20] gLite WMS: Workload Management System. http://web.infn.it/gLiteWMS/. Last visited on 11 September 2010.

[21] Global Vs30 Map Server, USGS. http://earthquake.usgs.gov/hazards/apps/vs30/. Last visited on 04 September 2010.

[22] Gnuplot homepage. http://www.gnuplot.info/. Last visited on 11 September 2010.

[23] Google Earth. http://www.google.com/earth/index.html. Last visited on 11 September 2010.

[24] Google Maps API Family. http://code.google.com/apis/maps/. Last visited on 05 September 2010.

[25] GSL - GNU Scientific Library. http://www.gnu.org/software/gsl/. Last visited on 11 September 2010.

[26] B. Gutenberg and C. F. Richter. Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, 34(4):185, 1944.

[27] A. Guttman. R-trees: a dynamic index structure for spatial searching. In *Proceedings of the 1984 ACM SIGMOD international conference on Management of data*, pages 47–57. ACM, 1984.

[28] M. Hadjieleftheriou, E. Hoel, and V. J. Tsotras. SaIL: A spatial index library for efficient application integration. *Geoinformatica*, 9(4):367–389, 2005.

[29] JavaServer Pages Technology. http://java.sun.com/products/jsp/. Last visited on 11 September 2010.

[30] P. Kacsuk and G. Sipos. Multi-grid, multi-user workflows in the P-GRADE Grid Portal. *Journal of Grid Computing*, 3:221–238, 2005.

[31] K. Kayabalı and M. Akın. Seismic hazard map of Turkey using the deterministic approach. *Engineering Geology*, 69(1-2):127–137, 2003.

[32] S. L. Kramer. *Geotechnical Earthquake Engineering*. Prentice Hall, 1996.

[33] K. Krauter, R. Buyya, and M. Maheswaran. A taxonomy and survey of grid resource management systems for distributed computing. *Software: Practice and Experience*, 32(2):135–164, 2002.

[34] R. K. Mark. Application of linear statistical models of earthquake magnitude versus fault length in estimating maximum expectable earthquakes. *Geology*, 5(8):464, 1977.

[35] C. Özturan, B. Bektaş, and M. Yılmazer. Seismic data server application service and web interface. In *SEE-GRID-SCI User Forum 2009*, December 2009.

[36] M. D. Petersen, A. D. Frankel, S. C. Harmsen, C. S. Mueller, K. M. Haller, R. L. Wheeler, R. L. Wesson, Y. Zeng, O. S. Boyd, D. M. Perkins, N. Luco, E. H. Field, C. J. Wills, and K. S. Rukstales. Documentation for the 2008 update of the United States national seismic hazard maps. Open-File Report 2008-1128, US Geological Survey, 2008.

[37] L. Reiter. *Earthquake Hazard Analysis: Issues and Insights*. Columbia University Press, New York, 1990.

[38] K. Sadigh, C. Y. Chang, J. A. Egan, F. Makdisi, and R. R. Youngs. Attenuation relationships for shallow crustal earthquakes based on California strong motion data. *Seismological Research Letters*, 68:180–189, 1997.

[39] D. P. Schwartz and K. J. Coppersmith. Fault behavior and characteristic earthquakes: examples from the Wasatch and San Andreas fault zones. *Journal of Geophysical Research*, 89(B7):5681–5698, 1984.

[40] SEE-GRID-2 consortium. Infrastructure overview and assessment. SEE-GRID-Deliverable-3.4, FP6 Research Infrastructures - SEE-GRID-2, 2008.

[41] SEE-GRID-2 Project. http://www.see-grid.eu/. Last visited on 26 September 2010.

[42] SEE-GRID consortium. SEE-GRID infrastructure evaluation results. SEE-GRID-Deliverable-4.3, FP6 Research Infrastructures - SEE-GRID, 2006.

[43] SEE-GRID eInfrastructure for regional eScience. http://www.see-grid-sci.eu/. Last visited on 26 September 2010.

[44] SEE-GRID Infrastructure Development. https://http.ipb.ac.rs/documents/seegrid_infrastructure_development/. Last visited on 05 October 2010.

[45] R. W. Sinnott. Virtues of the Haversine. *Sky and Telescope*, 68(2):159, 1984.

[46] South-Eastern European Grid-enabled eInfrastructure Development. http://www.see-grid.org/. Last visited on 26 September 2010.

[47] The Generic Mapping Tools. http://gmt.soest.hawaii.edu/. Last visited on 11 September 2010.

[48] The GNU Bourne-Again SHell. http://tiswww.case.edu/php/chet/bash/bashtop.html. Last visited on 11 September 2010.

[49] P. C. Thenhaus and K. W. Campbell. Seismic hazard analysis, in Earthquake Engineering Handbook. *W. F. Chen and C. Scawthorn (Editors), CRC Press, Boca*, 2003.

[50] P. Thomas, I. Wong, and N. Abrahamson. Verification of probabilistic seismic hazard analysis computer programs. PEER Report 2010/106, Pacific Earthquake Engineering Research Center, College of Engineering, University of California, Berkeley, 2010.

[51] D. J. Wald and T. I. Allen. Topographic slope as a proxy for seismic site conditions and amplification. *Bulletin of the Seismological Society of America*, 97(5):1379, 2007.

[52] D. L. Wells and K. J. Coppersmith. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America*, 84(4):974–1002, 1994.

[53] E. Yenier, Ö. Erdoğan, and S. Akkar. Empirical relationships for magnitude and source-to-site distance conversions using recently compiled Turkish strong-ground motion database. In *The 14th World Conference on Earthquake Engineering*, October 2008.

[54] R. R. Youngs and K. J. Coppersmith. Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates. *Bulletin of the Seismological Society of America*, 75(4):939, 1985.