M.Sc. Engg. Thesis

Predicting Earthquakes through Digging Satellite Data

by

Sarfaraz Newaz (1014052018P)

Submitted to

Department of Computer Science & Engineering
(In partial fulfilment of the requirements for the degree of Master of Science in Computer Science & Engineering)

Department of Computer Science & Engineering
Bangladesh University of Engineering & Technology (BUET)
Dhaka 1000

September 13, 2018
Dedicated to my loving parents

Author’s Contact

Sarfaraz Newaz
Lecturer,
Computer Science and Engineering Discipline
Khulna University (KU), Khulna.
Email: snewaz@cseku.ac.bd, DreamZViewerS@gmail.com
The thesis titled “Predicting Earthquakes through Digging Satellite Data”, submitted by Sarfaraz Newaz, Roll No. 1014052018P, Session October 2014, to the Department of Computer Science & Engineering, Bangladesh University of Engineering & Technology, has been accepted as satisfactory in partial fulfillment of the requirements for the degree of Master of Science in Computer Science & Engineering and approved as to its style and contents. Examination held on September 13, 2018.

Board of Examiners

1. 
Dr. A. B. M. Alim Al Islam
Associate Professor
Department of Computer Science & Engineering
Bangladesh University of Engineering & Technology, Dhaka.

2. 
Prof. Dr. Md. Mostofa Akbar
Head and Professor
Department of Computer Science & Engineering
Bangladesh University of Engineering & Technology, Dhaka.

3. 
Prof. Dr. M. Sohel Rahman
Professor
Department of Computer Science & Engineering
Bangladesh University of Engineering & Technology, Dhaka.

4. 
Dr. Atif Hasan Rahman
Assistant Professor
Department of Computer Science & Engineering
Bangladesh University of Engineering & Technology, Dhaka.

5. 
Prof. Dr. Mafizur Rahman
Professor
Department of Civil Engineering
Bangladesh University of Engineering & Technology, Dhaka.
Candidate's Declaration

This is hereby declared that the work titled "Predicting Earthquakes through Digging Satellite Data", is the outcome of research carried out by me under the supervision of Dr. A. B. M. Alim Al Islam, in the Department of Computer Science & Engineering, Bangladesh University of Engineering & Technology, Dhaka 1000. It is also declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

Sarfaraz Newaz
Candidate
Acknowledgement

Foremost, I express my heart-felt gratitude to my supervisor, Dr. A. B. M. Alim Al Islam, for his constant supervision of this work. He helped me a lot in every aspect of this work and guided me with proper directions whenever I sought one. His patient hearing of my ideas, critical analysis of my observations, and detecting flaws (and amending thereby) in my thinking and writing have made this thesis a success.

I would also want to thank the members of my thesis committee: Prof Dr. M. Sohel Rahman, Dr. Atif Hasan Rahman, and specially the external member Prof. Dr. Mafizur Rahman, for their encouragements, insightful comments, and valuable suggestions.

I am also thankful to Saddam Hossain (PhD Candidate, CSE-BUET), Dr. Md. Shamsuzzoha Bayzid (Assistant Professor, CSE-BUET), Shahidul Islam (PhD Candidate, CSE-BUET), Mahdy Rahman Chowdhury (Lecturer, ECE-NSU), Mashrura Tasnim (PhD Candidate, University of Alberta, Canada), Mahmudul Hasan (Undergrad student, CSE-BUET), and Tarik Reza Toha (MSc student, CSE-BUET). I sought help from them a number of occasions regarding simulation set-up and performance evaluation of this thesis. In addition, I am grateful to Md. Shihabul Islam (PhD Candidate, CSE-BUET), Jannatun Noor (MSc student, CSE-BUET), Subir Saha (MSc student, CSE-BUET), and Madhusudan Basak (Lecturer, CSE-BUET) for their help and valuable suggestions regarding the writing and presentation of this thesis.

Last but not the least, I remain ever grateful to my beloved parents, who always exists as sources of inspiration behind every success of mine.
Abstract

Earthquake is, perhaps, still considered as the foremost unpredictable natural disaster experienced to date. Even though, predicting an earthquake remains an elusive goal to scientists over the last several decades, we are still to achieve a good prediction mechanism for earthquakes. To this extent, in this study, we propose a new mechanism for predicting earthquakes. Our mechanism enables prediction around four months prior to happening of an earthquake. To do so, we analyze open satellite data and find gravity gradient anomaly as a prominent pre-seismic activity. To the best of our knowledge, we are first to consider and demonstrate gravity gradient anomaly as a pre-seismic activity for predicting earthquakes. Accordingly, using this anomaly, we formulate necessary mathematical models and illustrate how we can predict earthquakes having magnitude greater than or equal to 6.0 in moment magnitude scale. We analyze efficiency of our proposed mechanism in predicting several earthquakes (magnitude \( \geq 6.0 \)) occurred in different parts of the world. Our analysis confirms that we can predict more than 90% of the earthquakes using our proposed method while having a very low false alarm rate.
# Contents

**Acknowledgement**

**Abstract**

1 **Introduction**
   1.1 Motivations of Our Work ........................................... 2
   1.2 Summary of Existing Studies ........................................ 2
   1.3 Limitations of the Existing Studies ................................. 5
   1.4 Our Contributions ................................................... 7
   1.5 Outline of Our Thesis ................................................ 8

2 **Background**
   2.1 Structure of Earth .................................................... 9
      2.1.1 Earth’s Crust ..................................................... 10
      2.1.2 Earth’s Mantle ................................................... 10
      2.1.3 Earth’s Lithosphere .............................................. 11
      2.1.4 Earth’s Asthenosphere ........................................... 11
   2.2 Tectonic Plates ....................................................... 11
   2.3 Plate Tectonics Theory ............................................... 13
      2.3.1 Convergent Boundary ............................................. 13
      2.3.2 Divergent Boundary ............................................... 14
      2.3.3 Transform Boundary ............................................... 15
   2.4 Faults ................................................................. 15
   2.5 Causes of Earthquakes ............................................... 16
2.6 Seismic Effects due to Earthquakes

2.6.1 Primary Waves

2.6.2 Secondary Waves

2.6.3 Surface Waves

3 Related Work

3.1 Precursor based Study

3.2 Strain Calculation based Study

3.3 Characteristic based Earthquakes Study

3.4 VAN Mechanism

3.5 The M8 Algorithm

3.6 Radon Emission based Study

3.7 Gravity Gradient based Study

3.8 P-waves and S-waves based Study

3.9 Artificial Intelligence based Study

4 Data Collection and Analysis

4.1 Data Source

4.1.1 Database of Previous Earthquakes

4.1.2 Data Source for Earth’s Gravitational Field

4.2 GRACE Data Format

4.3 Data Preprocessing

4.3.1 Calculation of Gravitational Field of Earth

4.3.2 Calculation of First Derivatives of the Gravitational Field of Earth

4.3.3 Calculation of Gravity Gradient Tensor (GGT)

5 Preliminary Observation

5.1 Startup

5.2 Analysis

5.3 Primary Decision

6 Methodology of Our Study

6.1 Boundary Region
6.2 Reducing Tidal and Atmospheric Effect .......................... 43
6.3 Learning from the Most Stable Part of Earth’s Nature .................. 43
6.4 Calculation of minSTD Avg and absGGT .............................. 44
6.5 Calculation of STD Multiplier ........................................ 45
6.6 GGT sum .............................................................. 45
6.7 Calculation of weightedGGT sum and cumulatedWGGT sum ............ 46
6.8 Use of Dynamic Programming ......................................... 46
6.9 Calculation of STD Multiplier using Multi Objective Evolution .......... 47
   6.9.1 Learning Based on True Positive Cases .......................... 47
   6.9.2 Learning by Avoiding False Positive Cases ....................... 48
6.10 Trade-off between Our Objectives and Calculation of a Final Value of STD Multiplier 48
6.11 Accuracy Validation .................................................. 49

7 Prediction Results and Comparative Analysis ............................. 50
  7.1 Study in the Area of Japan and Its Results ......................... 51
  7.2 Study in the Area of Chile and Its Results ......................... 52
  7.3 Study in the Area of Sumatra-Andaman and Its Results ............. 52
  7.4 Study in the Area of China and Its Results ........................ 53
  7.5 Study in the Area of India-Nepal and Its Results ................ 53
  7.6 Results at a Glance .................................................. 54
  7.7 Different Aspects of Our Proposed Mechanism ....................... 54
     7.7.1 Fixed Training and Validation Period for All Regions .......... 54
     7.7.2 Sensitivity of Division between Training and Validation Datasets over Prediction Accuracy ........................................ 61
     7.7.3 Applicability of Our Proposed Mechanism over A Region having Rare Earthquakes 66

8 Our Findings in A Nutshell ................................................. 68

9 Future Work .................................................................. 71

10 Conclusion ................................................................. 73
List of Figures

1.1 Catastrophic effects of a few recent historical earthquakes .......................... 3
1.2 Trade-offs between predicting and alarming system against no alarm system [1] .... 4

2.1 Structure of Earth ......................................................................................... 10
2.2 Tectonic plate boundaries and their moving directions [2] .............................. 12
2.3 Plate movement due to convection and different types of boundaries among tectonic plates .................................................................................................. 13
2.4 Different types of convergence and subduction zone .................................... 14
2.5 Plate movement due to convection, divergent and transform boundaries among tectonic plates ...................................................................................... 15
2.6 Different types of faults [3] .......................................................................... 16
2.7 Effect of earthquake ..................................................................................... 17
2.8 Several components of earthquake ............................................................. 18

5.1 Point A has the latitude (ϕ) of 4°N and longitude (λ) of 92°E. Point B has the latitude of 4°N and longitude of 102°E. Latitudes and longitudes for point C are 4°S and 102°E, and for D are 4°S and 92°E respectively. These four points together confine a region of our initial analysis. ................................................................. 34

5.2 M7.0+ earthquakes within the period and region of our study are shown in this figure. From Table 5.1, we can see that, there are 14 EQs within our period of study in this region. However, in this figure, we can see 13 EQs. Here, we have ignored an M7.4 EQ, which appears in the first row of the table, as we do not have enough prior data (GGT time series data of two years before occurring the EQ) for this EQ in our study. 35
5.3 Anomalies detected in time series of GGT components before three different sample earthquakes out of the 14 earthquakes as pointed in Table 5.1. Fig. 5.3a represents an EQ that took place off the west coast of Northern Sumatra in Indian Ocean in 2004. Location of this EQ was φ=3.295°N and λ=95.982°E (Sl. 2 in Table 5.1). Fig. 5.3b represents another EQ that took place near Southern Sumatra Indonesia in 2009. Location of this EQ was at φ=0.72°S and λ=99.867°E (Sl. 9 in Table 5.1). Fig. 5.3c represents another EQ that took place at φ=0.802°N and λ=92.463°E off the west coast of Northern Sumatra in 2012 (Sl. 14 in Table 5.1). The green vertical lines in all the time series shown in Fig. 5.3a, 5.3b, and 5.3c indicate the point of EQ happening. We can see, five out of six GGT components in Fig. 5.3a, three out of six components in Fig. 5.3b, and two out of six components in Fig. 5.3c show significant anomalies before two or three months of corresponding earthquakes. These anomalies are annotated by red circles.

5.4 GGT component time series for an M8.2 EQ (Sl. 14 in Table 5.1), which is the same EQ shown in Fig. 5.3c. This figure suggests that, other than just prior to the M8.2 EQ anomalies, there exist other anomalies. These anomalies are annotated by green circles.

5.5 This figure represents time series of GGT components for an M8.2 earthquake (Sl. 14 in Table 5.1), which is again the same EQ of Fig. 5.3c. However, this time, we plot other quakes occurred in nearby locations. After plotting other quakes, we can find that, the anomalies marked by green circles are just before other nearby quakes. Two anomalies marked by the green circles over $V_{xx}$ and $V_{yz}$ appear before the M7.6 EQ (Sl. 9 in Table 5.1), which took place 846.8 Km away from the original M8.2 EQ. Two more anomalies are marked by green circles over $V_{xz}$ and $V_{zz}$ appear before the M7.8 EQ (Sl. 10 in Table 5.1), which happened 543.03 Km away from the M8.2 EQ. Therefore, these anomalies are also indications of earthquakes, which took place at nearby points.

6.1 Block diagram of our proposed model to detect anomaly.
6.2 Each point under consideration for earthquake prediction is imagined as an origin (long, lat). We then consider the GGT components of all the points up to ±4° from the origin to calculate the GGT anomaly at origin. This figure is for calculation in a 1° × 1° grid. All the points up to ±4° from the origin including the origin is treated as the Boundary Region.

7.1 Progression of the M8.8 Chile-Malue earthquake (in the year 2010) from the four months period before the earthquake happening up to the happening

7.2 Progression of the M9.0 Tohoku-Oki earthquake (in the year 2011) from the four months period before the earthquake happening up to the happening

7.3 Progression of the M8.6 and M8.2 Indian Ocean earthquake (in the year 2012) from the four months period before the earthquake happening up to the happening

7.4 Progression of the M6.8 Myanmar earthquake (in the year 2012) from the four months period before the earthquake happening up to the happening

7.5 Progression of the M6.7 Imphal, India earthquake (in the year 2016) from the four months period before the earthquake happening up to the happening

7.6 Number of earthquakes vs true prediction and false prediction rate
List of Tables

4.1 Sample data file obtained from USGS .................................................. 32
4.2 Sample data file obtained from GRACE ............................................... 32
4.3 GRACE data format of the sample data shown in Table 4.2 .................... 33

5.1 All M7.0+ earthquakes information in Sumatra-Andaman region, gathered from USGS earthquake database .................................................. 36

7.1 Results obtained from our proposed mechanism for different regions of the world using different data sources .................................................. 60
7.2 Results obtained from our proposed mechanism for different regions of the world using different data sources with fixed training and validation periods .................. 61
7.3 Results obtained from our proposed mechanism for Sumatra-Andaman region considering successive two-year intervals for training and validation datasets using different data sources .................................................. 64
7.4 Results obtained from our proposed mechanism for Sumatra-Andaman region with gradually increasing time periods of the training dataset by two years using different data sources .................................................. 66
7.5 Results obtained from our proposed mechanism for the region of Iran .......... 67

8.1 Comparison among our proposed mechanism and other existing studies .... 69
8.2 Specific earthquake and region-based comparison among our proposed mechanism and other existing studies .................................................. 70
Chapter 1

Introduction

Earthquake (EQ or quake), one of the most horrible natural disasters, generally refers to a shaking of surface of the Earth. Seismic waves created by sudden release of energy in the Earth’s lithosphere result in earthquakes. They usually occur without warning and do not allow much time for people to react. Earthquakes are capable to inflict vast devastation to a large number of buildings and constructions at the blink of an eye. Therefore, earthquakes can cause serious injuries along with loss of lives and destroy different infrastructures leading to great economic losses. In this regard, prediction of earthquakes has always been a highly-emphasized study theme all over the world. Even though the prediction of earthquakes is obviously critical to the safety of mankind, it has been proven to be a very challenging issue in seismology [4]. This happens as no prediction is feasible without precise realization of the predicted phenomenon, which is always considered to be extremely difficult in case of earthquakes.

Perchance, one of the simplest adaptive formulations found in the literature, was carried out on Earthquake Prediction in 1976 [5]. Based on this study, for prediction of an earthquake, the geographical zone, the time period in which it would occur, and the magnitude range are considered to eventually allot a level of confidence for each prediction. Thus, this study presents that earthquake prediction is a branch of the science of seismology that predicts a range of time, location, and magnitude of a future earthquake. Many other studies in the literature also present the task of prediction of earthquakes in a similar way [6].

The task of predicting earthquakes has been considered to be a challenging goal to scientists from long ago. From the early 1970, researchers are trying to predict earthquakes. For example, in 1962,
CHAPTER 1. INTRODUCTION

Japanese research community was sure that they would be able to predict earthquakes within ten years since then [7]. A lot of research studies have been conducted since that time. However, till now, prediction of earthquakes has not seen the light as per expectation. Scientists have found excessive false alarms from the existing prediction mechanisms, as many of the earthquakes predicted by the existing mechanisms did not take places [8, 9]. Additionally, some prediction mechanisms are yet to have any scientific basis [10, 11, 12]. As the research studies are yet to achieve desired success in predicting earthquakes, the aftermaths due to earthquakes still continue to exhibit notable devastation even to date.

1.1 Motivations of Our Work

Unlike many other natural disasters, earthquake still remains mostly unpredictable. Therefore, most of the earthquakes happen without prior alarms and losses due to the earthquakes become catastrophic. As an evidence, since 1900, 2.3 million people have died in 2,233 earthquakes [13]. Fig. 1.1 shows some of the catastrophic effects due to a few recent historical large earthquakes.

To be more specific, in 2004, the Indian Ocean earthquake unleashed Tsunamis that killed a total of 227,300 people and made a damage of more than $10 billion [13]. Besides, according to a CEDIM report, in 2011, earthquakes and their consequences such as Tsunamis, landslides, and ground settlements caused a death of 17,900 people and a damage of $365 billion [20].

There is a trade-off in predicting and raising alarms for earthquakes. Fig. 1.2 shows the trade-off. As shown in the figure, predicting earthquakes and generating alarms are of high importance, however, limiting false alarms are of no less significance. An effective and widely-accepted prediction algorithm for earthquakes addressing the trade-off between accuracy through true alarms and inaccuracy through false alarms by suitably adjusting the bar (as shown in the figure) is yet to come into play. If such an algorithm could be devised, that would be a great relief to mankind.

1.2 Summary of Existing Studies

Earthquake has been studied over the last several decades. There exit several methods using which scientists are trying to predict earthquakes. Most of the existing studies are based on earthquake precursors, which refer to anomalous phenomena that might give effective warning of impending
earthquakes. Initially, foreshocks were considered as a precursor of an upcoming earthquake. Furthermore, there have been around 400 reports of possible precursors in scientific literature, of roughly twenty different types. We summarize some prominent precursors found in the literature below.

- Animal behaviour: For centuries, anomalous animal behaviour have been observed before several earthquakes. In these cases, animals display unusual behaviour some tens of seconds prior to an earthquake. Later, it has been found that, the animals actually respond to P-waves of the earthquakes. P-waves travel through the ground about twice as fast as the S-waves travel and the S-waves cause most severe shaking. We will discuss about P-waves and S-waves later in this
Characteristic earthquakes: The notion of characteristic earthquakes was the basis of the Parkfield prediction [12]. Parkfield was rocked by similar M6.0 earthquakes in 1857, 1881, 1901, 1922, 1934, and 1966. These suggested a pattern of breaks every 21.9 years, with a standard deviation of ±4.3 years [6, 21, 22]. Statistically, these events led to a prediction of an earthquake around 1988, or before 1993 at the latest (at the 95% confidence interval) [22]. The mechanism of such a statistical calculation tells us that the prediction is derived entirely from the trend of previous historical earthquakes.

Electromagnetic anomalies: Observations of electromagnetic disturbances before an earthquake was claimed first for the Great Lisbon earthquake in 1755 [23]. Later, in 2011, Thomas et al., reviewed and found the “most convincing” electromagnetic precursors to be Ultra Low Frequency (ULF: 0.001–10 Hz) magnetic anomalies. This type of anomaly was also recorded before the 1989 Loma Prieta earthquake [10].

VAN seismic electric signals: The most touted claim of an electromagnetic precursor was the VAN mechanism, which is also the most criticized one. This mechanism was proposed by three researchers of Physics namely Panayiotis Varotsos, Kessar Alexopoulos, and Konstantine Nomicos (VAN) from the University of Athens. In a paper in 1981, they claimed that by measuring geo-electric voltages, they could predict earthquakes of magnitude greater than 2.8
within all over Greece up to seven hours prior to their happenings. They named the geo-electric signal as “Seismic Electric Signals” (SES) [24].

In addition to the above-mentioned precursor-based studies, there are other forms of prediction studies that emerged in recent times. We summarize the studies below.

• Magnitude based study: Earthquakes of a certain magnitude are studied in different research work. For example, Shahrisvand et al., studied three great earthquakes having magnitudes greater than M8.5 [25]. Uphoff et al., Montagner et al., and Linage et al., have studied earthquakes of magnitude M9.0 [26, 27, 28]. These studies are solely based on specific magnitudes of the earthquakes.

• Region based study: This type of studies have been done for particular regions. Here, different regions are studied in isolation for earthquake prediction by few researchers [29, 30, 31]. For example, Asencio-Cortés et al., conducted their research around 200Km of Tokyo [29], Wang et al., has a research area of China [30], Molchan et al., studied the region of Italy [31], etc.

Nonetheless, there are many other existing studies that will be discussed in next chapter.

1.3 Limitations of the Existing Studies

Each of the existing types of studies face some limitations. These limitations present barriers to find a successful prediction algorithm for upcoming earthquakes in general. We summarize the limitations below.

• Precursor-based prediction: To the best of our knowledge, no generalized precursor has been found to date to be reliable for the purposes of earthquake prediction [12]. All the known precursors get confined to a set of specific earthquakes, as they produce high false alarm rates in other cases. For example, several earthquakes occurred without any foreshock or any other precursor [10, 32, 33, 34, 35, 36, 37].

  – Animal behaviour: Animals usually respond to P-waves. They predict not the earthquake itself that has already happened but only the imminent arrival of the more destructive S-waves ahead. P-waves and S-waves are discussed in Chapter 2.
CHAPTER 1. INTRODUCTION

The animal behavior is not necessarily related only to earthquakes, as they often get related to other natural phenomena such as rainfall, cyclone, etc [38]. Moreover, it is not always easy to extract the animal behavior in a uniform manner, as the behavior can be deviated as per the conditions of the animals (for example their food taking, disease, etc.) [39].

- Characteristic earthquakes: In the Parkfield case, the predicted earthquake did not occur until 2004, a decade later from the predicted time [40, 41, 42]. Researcher tried to find the cause with the Parkfield seismic data. It was found that, several M4.0 earthquakes took places at predicted time, which reduced the stresses on the of the Parkfield segment from north-west portion. This might be the cause to skip generating the predicted M6.0 earthquake [43]. This makes the Characteristic earthquakes prediction mechanism questionable. Moreover, such a prediction is yet to achieve its generality in different parts of the world.

- Electromagnetic anomalies: It is now believed that Electromagnetic anomalies observation prior to an earthquake was a system malfunction. For example, Parkfield earthquake of 2004 was closely monitored and any evidence of precursory electromagnetic signals of any type for this earthquake was not found. Eventually International Commission on Earthquake Forecasting for Civil Protection (ICEF) considered the lack of any evidence of useful precursors (including electromagnetic anomalies) before an earthquake as the failure of finding a precursory mechanism for earthquake prediction [10].

- VAN seismic electric signals: The VAN seismic electric signals has been refuted by the The International Commission on Earthquake Forecasting for Civil Protection (ICEF). The ICEF concluded that VAN mechanism could not be validated as a prediction mechanism of earthquake. This happened as the SES mechanism claimed by VAN do not support by property of Physics [10].

- Magnitude based study: The earthquakes having magnitudes greater than M8.5 are large enough and do not take places regularly. Besides, M6.0+ earthquakes are big enough to cause a huge loss in terms of human lives and economy. However, existing studies do not support the M6.0+ earthquakes in their study domain. Moreover, the existing magnitude based studies mostly remain confined to some specific cases being far from a generalized prediction mechanism.

- Region based study: Each of Asencio-Cortés et al., Wang et al., and Molchan et al., studied for only one single region [29, 30, 31]. The findings of these studies are yet to be applicable to other
parts of the world, as the crustal structures of different parts of the Earth exhibit substantial
differences [44]. Therefore, a single region based study does not work for another region.

1.4 Our Contributions

Our study is motivated by the limitations of state-of-the-art research and challenges involved in
developing a pragmatic prediction system. In our study, we propose a new mechanism for predicting
earthquakes that can be applied in different parts of the world in a generalized manner. Using our
mechanism, we can predict an earthquake around four months prior to its happening. In road to
devising our mechanism, we first analyze open satellite data and find gravity gradient anomaly as a
prominent pre-seismic activity. To the best of our knowledge, our study is the first to consider and
demonstrate gravity gradient anomaly as a pre-seismic activity for predicting earthquakes.

Accordingly, based on the gravity gradient anomaly, we perform formulation of necessary math-
ematical models and illustrate how we can predict earthquakes having magnitude greater than or
equal to 6.0 in moment magnitude scale using our models. We evaluate performance of our proposed
mechanism in predicting several earthquakes (magnitude $\geq 6.0$) occurred in different parts of the
world. Our evaluation results confirm that we can have predictions of predict more than 90% of the
earthquakes utilizing our proposed mechanism while experiencing a very low false alarm rate.

To summarize, we make the following set of contributions in this study -

- We propose a new mechanism for predicting earthquakes that can be applied in different parts
  of the world in a generalized manner. Here, we adopt gravity gradient anomaly as a prominent
  pre-seismic activity.

- We process public satellite data to demonstrate efficacy of gravity gradient anomaly to be a
  pre-seismic activity for different earthquakes.

- Subsequently, we formulate necessary mathematical models based on gravity gradient anomaly
  to perform the task of prediction of earthquakes.

- We apply the models to several earthquakes all over the world to demonstrate its high accuracy
  in predicting the earthquakes and its low false alarm rates.

It is worth mentioning that, since our model uses public satellite data that is available for all over
the world, our proposed mechanism will not be a region-specific one rather it will work for all over the world.

1.5 Outline of Our Thesis

This is how the rest of this book is organized. In Chapter 2, we light on some elementary terms related to earthquakes, background, and the causes of earthquakes. The following chapter elaborates the existing studies. Next, in Chapter 4, we present how to collect the data related to earthquakes and Earth’s gravitational field. Consequently, we also have shown the collected data format in this chapter. Then, in Chapter 5, we present our preliminary observation which leads us to propose this mechanism. Next to that, in Chapter 6, we present our methodology driven from our preliminary observation. Then, in Chapter 7, we present our findings and results. On the basis of results found from this chapter, we made a discussion in the next chapter. After the discussion, we present our future plan and finally we present the conclusion.
Chapter 2

Background

An earthquake is the sudden release of energy in the Earth’s lithosphere that creates seismic waves. Shaking, rolling, or sudden shock of the Earth’s surface is the result from an earthquake. Earthquakes can be so small in size or so weak that they cannot be felt. Again, earthquake can be so violent that it can toss people around and destroy a whole city. It can be violent enough to destroy major man made structures like building and kill thousands of people at a blink of eye. In addition, earthquakes can be felt over a huge area although they usually last very little time, usually less than a minute. In general sense, earthquake is used to describe any seismic event that can be either natural or caused by humans and that generates seismic waves. However, still there has not any successful mechanism been found for predicting an earthquake.

To describe the details of an earthquake, at first, we need to be familiar with some elementary terms. These terms are described briefly in this chapter.

2.1 Structure of Earth

The internal structure of the Earth is layered consists of a silicon made outer solid crust, a highly viscous asthenosphere and mantle, a liquid outer core which is much less viscous than the mantle, and a solid and hot inner core. That is, structure of Earth consists of different layers. In next subsections, we are going to describe some of those layers.
2.1.1 Earth’s Crust

According to geology, the Crust is the outermost solid shell of the Earth. The crust is composed of a great variety of rocks and made the solid surface of the Earth. The crust of Earth occupies less than 1% of the volume of the Earth. Earth’s crust are of two types. They are:

**Oceanic Crust**

The uppermost layer of the oceanic portion of the Earth is the oceanic crust. It has a thickness of 5 km to 10 km. Most of the case, it is made of basalt, diabase, and gabbro.

**Continental Crust**

The continental crust is the crust from the continental part of the Earth. The thickness of continental crust is from 30 km to 50 km. It is made of slightly less dense rocks than those of the oceanic crust.

From Fig. 2.1a we can see both types of crusts.

2.1.2 Earth’s Mantle

The mantle is a layer inside Earth. The interior of Earth is chemically divided into several layers. The mantle is a layer between the crust and the outer core. Earth’s mantle is made of silicate with an average thickness of 2,886 kilometers. The mantle makes up about 84% of Earth’s volume.
2.1.3 Earth’s Lithosphere

The lithosphere is the rigid, outermost shell of the Earth that is defined by its rigid mechanical properties. Earth’s lithosphere consists of the crust and the uppermost mantle, which is responsible for the hard and rigid outer layer of the Earth.

2.1.4 Earth’s Asthenosphere

The asthenosphere is the highly viscous, mechanically weak, and malleable region of the upper mantle of the Earth which can deform. Just below the lithosphere, Earth’s asthenosphere situated at depths between approximately 80 and 200 km below the surface.

Fig. 2.1b shows Earth’s crust, mantle, Lithosphere, and Asthenosphere. From the same figure we can see other cores that are out of scope of this study.

2.2 Tectonic Plates

Tectonic plates are the pieces of Earth’s crust and uppermost mantle, which is referred together as the lithosphere. In other words, the lithosphere is subdivided into tectonic plates. The tectonic plates are the large, thin, and the relatively rigid plates on the outer surface of the Earth. Tectonic plates move with respect to one another. The plates are around 100 km thick and consists of two types: Oceanic crust plate and Continental crust plate.

There are three types of tectonic plates:

• Major (or primary) plates: A major plate is any plate with an area greater than 10 million km$^2$.

There are seven major plates:

- Pacific Plate 103,300,000 km$^2$
- North American Plate 75,900,000 km$^2$
- Eurasian Plate 67,800,000 km$^2$
- African Plate 61,300,000 km$^2$
- Antarctic Plate 60,900,000 km$^2$
- Indo-Australian Plate 58,000,000 km$^2$
- South American Plate 43,600,000 km$^2$
• Minor (or secondary) plates: A minor plate is any plate with an area less than 10 million km$^2$, however, greater than 1 million km$^2$. Well known minor plates are:
  
  – Nazca Plate  5,500,000 km$^2$
  – Philippine Sea Plate  5,500,000 km$^2$
  – Arabian Plate  5,000,000 km$^2$
  – Caribbean Plate  3,300,000 km$^2$
  – Cocos Plate  2,900,000 km$^2$
  – Caroline Plate  1,700,000 km$^2$
  – Scotia Plate  1,600,000 km$^2$
  – Burma Plate  1,100,000 km$^2$
  – New Hebrides Plate  1,100,000 km$^2$

• Microplates (or tertiary plates): A microplate is any plate which has an area less than 1 million km$^2$. In a major plate map, microplates are often grouped with an adjacent major plate.

Figure 2.2: Tectonic plate boundaries and their moving directions [2]
CHAPTER 2. BACKGROUND

(b) Different types of boundaries among tectonic plates [2]

2.3 Plate Tectonics Theory

The plate tectonics theory explains the features and movement of Earth’s surface from the deepest ocean trench to the tallest mountain for both in the present and the past. According to this theory, Earth’s lithosphere has greater strength than the underlying asthenosphere. Therefore, the density variations in the mantle result in convection and this convection causes the tectonic plates to move. Fig. 2.3a shows the movement of plates due to convection in mantle. The plates are all moving in different directions and at different speeds [48]. Fig. 2.2 shows different types of plates and the direction of their movements. Where the plates meet, their relative motion determines the type of the boundary. Fig. 2.3b present different types of boundaries.

2.3.1 Convergent Boundary

An actively deforming region where more than one tectonic plates or fragments of the lithosphere move toward one another and collide, that is a convergent boundary. Convergent boundaries are also known as a destructive plate boundary as crust are destroyed here. There exist three types of convergent boundary. Fig. 2.4a shows a boundary of continental-continental convergence, whereas Fig. 2.4b shows an oceanic-oceanic convergence boundary, and Fig. 2.4c shows continental-oceanic convergence boundary. When two plates move towards one another, they form a subduction zone.
Subduction is the process in which one tectonic plate moves under another tectonic plate and sinks into the mantle as the plates converge. Fig 2.4d shows an example of subduction process and subduction zone. Rates of subduction are typically centimetres per year, with the average rate of convergence being approximately two to eight centimetres per year. Regions where this process occurs are known as subduction zones. Generally, large earthquakes take place at convergent boundaries.

(a) Continental-continental convergence [49]  
(b) Oceanic-oceanic convergence [49]  
(c) Continental-oceanic convergence [49]  
(d) Subduction zone [50]

Figure 2.4: Different types of convergence and subduction zone

2.3.2 Divergent Boundary

The plate boundary that are moving away from each other, called divergent boundary. Divergent Boundary is also known as a constructive boundary or an extensional boundary as new crust is created here. Most of the time, active divergent plate boundaries take place between oceanic plates. They exist as mid-oceanic ridges. This type of boundary is shown in Fig. 2.5a. Divergent boundaries are normally situated at deep sea, therefore, earthquakes take place there may not have affects to people. Moreover, unlike to convergent boundaries, earthquakes in divergent boundaries are not too much large.
2.3.3 Transform Boundary

A transform boundary (or conservative boundary) is where two of the tectonic plates slide alongside each other. At transform boundaries no new crust is created, therefore, the name conservative boundary comes. No subduction (or sandwiching) of plates occur here. Generally, earthquakes at transform boundaries are not as large as convergent boundary.

As an example, we can consider Linear valleys, where rock has been ground down to leave gaps. River beds can be another example that have been broken into two because the two halves have shifted in opposite directions. Fig. 2.5b shows a transform boundary. San Andreas in South America is one of the largest transform boundary in the world.

Figure 2.5: Plate movement due to convection, divergent and transform boundaries among tectonic plates

2.4 Faults

Due to the relative motion of the plates, fracture occur at the plate boundaries. These fractures are faults. A fault is an area of stress in the Earth where broken rocks slide past each other, causing a crack in the Earth’s surface. There are four major types of faults:

- Dip-slip normal fault
- Dip-slip reverse fault
- Strike-slip fault
CHAPTER 2. BACKGROUND

- Oblique-slip fault

Fig. 2.6 shows images of different types of faults.

![Fig. 2.6: Different types of faults](image)

(a) Dip slip normal fault  
(b) Dip slip reverse fault  
(c) Strike slip fault  
(d) Oblique slip fault

2.5 Causes of Earthquakes

Earthquakes can happen in any of these situations:

- When two plates collide head to head, they push each other up and mountains formed in this way. Himalayas and other great mountain ranges were created long ago in this way.

- At subduction zone, one plate dives below another plate. The diving plate is crushed and melted in mantle. In this process often magma (molten rock) rises up to the surface and creates volcano.

- When two plates slide past against each other, they create a transform fault. As an example, San Andreas fault is one of the largest transform fault.

According to the plate tectonic theory we discussed earlier in this chapter, the tectonic plates are always moving. However, at plate boundary, the surface between two plates are not too much smooth to move or slide apart. Most of the time the adjacent plates are locked at the boundary due to friction and roughness of the rock. The irregularities along the sides of plates, which are called faults, increase
the frictional resistance. The fault surface could smoothly slide past against each other if there are no such irregularities. Most fault surfaces do have such irregularities and leads to a form of stick-slip behavior. The lock situation can happen in any of the plate boundaries. The plates spend much of the time locked in place by the friction of the plates in spite of the powerful forces driving plates against each other. Once the fault is locked, continued motion between the plates with respect to each other leads to increasing stress. As a result, growth of stress increase stored strain energy in the volume around the fault surface. Increasing of stress continues until the stress has risen sufficiently to break through the irregularity of the fault surface. Eventually, however, the lock situation build up so much pressure that, suddenly the locked situation get unlocked. This sudden unlock event allows the plates abruptly snap forward, sliding over the locked portion of the fault, and release the stored energy [53]. The energy is released in different forms such as radiated elastic strain seismic waves, frictional heating of the fault surface, cracking of the rock at the plate boundary, etc. These different forms of the released energy altogether cause an earthquake. Due to an earthquake, the ground can shift a few meters! Fig. 2.7a shows such an real life example. Moreover, seismic waves from that sudden motion spreads in all directions. Fig. 2.7b illustrates the spreading of seismic waves due to earthquake in all direction. Note that, the point beneath the Earth’s surface where the rocks break and move is called the focus or hypocenter of the earthquake. The focus or hypocenter is the underground point of origin of an earthquake. The point which is directly above the focus, on the Earth’s surface, is the epicenter. An epicenter due to an earthquake is shown in Fig. 2.8a.

![Figure 2.7: Effect of earthquake](image)

As an example, two great plates, the Pacific and the North American, meet in California. The
Pacific plate is moving north whereas the North American plate is moving south. Therefore, they created an well known transform fault named the San Andreas fault at their boundary. The Pacific plate has slid about 200 miles north over the last 20 million years. If it keeps moving like this, San Francisco will become adjacent to Seattle in next 20 million years! On the other hand, the San Andreas fault curves around Los Angeles. It then again curves around Northern California. Because of that, the two plates cannot slide smoothly against each other. Moreover, the increment of stresses of plate movement here fractured the land and also created dozens of smaller fault lines. Seismologists have been studying California’s faults for decades. They now say that there is a 70% chance of a major earthquake at the Bay of San Francisco Bay Area before 2030. This forecast is based on years of study of the many faults in the area including the San Andreas.

2.6 Seismic Effects due to Earthquakes

As we have already discussed that, waves are generated due to an earthquake are known as seismic waves. These seismic waves are the seismic effect due to an earthquake. There are different types of seismic waves.
2.6.1 Primary Waves

Seismic waves that travel the fastest are called primary waves, or P-waves. P-waves arrive at a given point before any other types of seismic waves. P-waves travel through solids, liquids, and gases. Note that, P-waves are push-pull waves. As P-waves travel, they push rock particles into the particles ahead of them, thus compressing the particles. The rock particles then bounce back. They hit the particles behind them that are being pushed forward. As a result, the particles move back and forth in the direction the waves are moving. Fig. 2.8 shows the P-waves. Generally P-waves cause little damages and not so much destructive.

2.6.2 Secondary Waves

Seismic waves that do not travel through the Earth as fast as P-waves do are secondary waves, or S waves. Generally, S-waves arrive at a given point after P-waves do. Moreover, S-waves travel through solids, however, not through liquids and gases. Fig. 2.8 shows the S-waves. S-waves are more destructive than P-waves.

2.6.3 Surface Waves

The slowest-moving seismic waves are called surface waves, or L-waves. L-waves arrive at a given point after primary and secondary waves do. L-waves originate at the epicenter. Surface waves travel along the surface of the Earth, rather than down into the Earth. L-waves usually cause more damages than P-waves or S-waves. Fig. 2.8b shows the L-waves. L-waves are the most destructive waves.
Chapter 3

Related Work

Scientists have been studied earthquakes for several decades. Many of them have used different techniques to predict earthquakes. Here are some widely used prediction techniques found in literature:

3.1 Precursor based Study

Predicting earthquakes exploiting a precursor has been studied for a long period of time. For example, the Haicheng earthquake hit Haicheng, Liaoning in China with a magnitude of M7.3 on February 4, 1975 [57, 58, 32, 33, 34, 35, 36, 37, 10, 59, 60]. This prediction was done by considering the strong foreshock of M4.7 (as a precursor) that occurred one day before occurring the earthquake [61, 11]. However, in later times, a lot of earthquakes took places without any foreshock [32, 33, 34, 35, 36, 37, 10]. Therefore, only foreshock cannot be a reliable precursor for predicting earthquakes. There are other precursors for prediction, however, not a single one could be proved as a reliable one yet.

3.2 Strain Calculation based Study

Other types of studies on predicting earthquakes also exist. For example, in 1976, it was predicted that a large earthquake might take place at off-shore of Peru within a period of seven to fourteen years from mid November 1974 [8]. The prediction was based on a theory of earthquakes with a deduction on strains building in the subduction zone. Later in June 1978, the time window was narrowed down to October to November, 1981, with a main shock in the range of 9.2±0.2 [9]. However, this earthquake is yet to take place till now.
3.3 Characteristic based Earthquakes Study

An ambush had been carefully laid for the Parkfield earthquake [12]. It was predicted by the characteristics of the Parkfield segment of the San Andreas Fault. Here, an earthquake of about M6.0 has been observed every several decades - in 1857, 1881, 1901, 1922, 1934, and 1966 - at Parkfield [62, 22, 63]. Bakun et al., pointed out that, other than the 1934 quake in this list, the earthquakes occur every 22 years, with a standard deviation of ±4.3 years [22]. Counting from 1966, it was predicted with a 95% chance that the next earthquake would hit around 1988 or 1993 at the latest. 1993 came, and passed away making the prediction unsuccessful. Subsequently, there was an M6.0 earthquake on the Parkfield segment of the fault in 2004, however, without any forewarning or precursors [40, 41, 42].

Further research into the Parkfield seismic data revealed that, several M4.0 earthquakes had reduced the stresses on the north-west portion of the Parkfield segment, perhaps causing it to skip generating the predicted M6.0 earthquake [43].

3.4 VAN Mechanism

In 1981, the VAN group declare that, they found a relationship between earthquakes and Seismic Electric Signals (SES). In 1984, they claimed successful prediction of 18 earthquakes out of 23 from 19th January to 19th September in 1983 [24]. In 1991, they again claimed a prediction of six out of seven earthquakes with $M \geq 5.5$ from 1st April in 1987 to 10th August in 1989 [64]. In 1996, they published a summary of all predictions issued from 1st January 1987 to 15th June 1995 [65], amounting to 94 predictions [21]. Later in Greece, they claimed 10 successes out of 14 cases with a success rate of 70%, however, ending with an extremely high false alarm rate of 89% [65].

Besides, objections raised about underlying physics of the VAN mechanism. According to VAN’s report, the Seismic Electric Signals (SES) have been transmitted over several hundred kilometers of distances from an earthquake’s epicenter to a monitoring station. Then, VAN mechanism grabs that SES and reports back with its amplitude. In later times, wave propagation properties of SES through Earth’s crust was analyzed with more rigor. The analysis found that it is impossible for the SES to traverse such a distance with the reported amplitude [43]. In addition, VAN’s publications do not account for possible sources of electromagnetic interference (EMI) [10]. Further, in 2011, the International Commission on Earthquake Forecasting for Civil Protection (ICEF) concluded that the SES prediction capability claimed by the VAN mechanism could not be validated [10].
3.5 The M8 Algorithm

The M8 algorithm by Keilis-Borok predicted that an M7.5 earthquake might take place in a five-year window from 1st January in 1984 to 31st December in 1988 at Loma Prieta, California, USA [66]. The prediction missed both magnitude and time. The Loma Prieta earthquake occurred with a moment magnitude of 6.9 on 17th October in 1989 [66].

Later, the M8 algorithm successfully predicted the 2003 San Simeon and Hokkaido earthquakes [67, 68]. Then, this algorithm again predicted that an $M \geq 6.4$ earthquake is going to occur in Southern California on or before 5th September in 2004 [67]. The predicted earthquake did not occur. A very similar prediction was made for another earthquake on or before 14th August in 2005, in approximately the same area of Southern California. This prediction also failed.

3.6 Radon Emission based Study

On 27th March in 2009, Giampaolo Giuliani warned the mayor of the city L’Aquila of the Abruzzo region of central Italy that there could be an earthquake within 24 hours from then. An earthquake of a moment magnitude of 2.3 occurred [69] to support his prediction. On 29th March in 2009, he made a second prediction [10]. He warned the mayor of the town of Sulmona over telephone to expect a damaging or even catastrophic earthquake within 6 to 24 hours till then. Sulmona is about 55 kilometers southeast of L’Aquila. However, no quake ensued [70, 71, 72].

Subsequently, on 6th April 2009, L’Aquila was rocked by a magnitude of 6.3 earthquake [10]. After the L’Aquila event, Giuliani claimed that he had found alarming rises in radon levels just hours before of the quake [70, 73]. However, the civil authority found that Giuliani did not transmit a valid prediction before the earthquake, and thus, found no credibility of the claim.

Moreover, later, it has been found that all of these predictions did not have scientific relations with earthquakes and the predictions either were just coincidences or eventually failed [11, 12].

3.7 Gravity Gradient based Study

In recent years, while researching on earthquakes, gravity changes during or after earthquakes have been observed by the researchers. The researchers have observed co-seismic and post-seismic gravity changes for most of the M8.5+ quakes [28, 74, 75, 76, 77, 78, 79, 80, 81]. For example, Montagner et al.,
showed that gravity signals are captured using superconducting gravimeter during the 2011 Tohoku-Oki earthquake having a moment magnitude of 9.0 [27]. However, to the best of our knowledge, no significant research study has taken place to predict earthquakes beforehand exploiting the underlying nature of pre-seismic gravity signal or any of its processed form such as gradient. It is generally believed that the gravity signal gets generated when an earthquake has already taken place. In other words, it is treated as a co-seismic activity. Therefore, as per conventional belief, it cannot be used in predicting earthquakes beforehand and it can only be used in after analysis.

Accordingly, Linage et al., separated co-seismic and post-seismic gravity anomaly for the 2004 Sumatra-Andaman earthquake from 4.6 year long data of GRACE satellite [28]. Area of the rupture surface of corresponding earthquakes, having magnitudes ranging between 9.1 and 9.3, is around $1200 \times 200 \text{ km}^2$ spreading from north-west of Sumatra to the Andaman Islands [82, 83]. Linage et al., analyzed the co-seismic and post-seismic gravity anomaly at this ruptured area [28]. Subsequently, Uphoff et al., performed supercomputer simulation to have a better idea about the rapture and tsunami that took place due to this mega-thrust earthquake of Indian Ocean [26].

Besides, Chen et al., analyzed GRACE data and detected co-seismic and post-seismic deformation from the M9.3 Sumatra-Andaman earthquake occurred in 2004 [75]. Additionally, Johanson et al., calculated co-seismic and post-seismic slip of the 2004 Parkfield earthquake from space-geodetic data [84]. Heki et al., detected co-seismic gravity changes of the M8.8 earthquake in 2010 Central Chile [85] and the 2011 Tohoku-Oki earthquake [86] using GRACE satellite gravimetry. Wang et al., also detected co-seismic and post-seismic deformations of the 2011 Tohoku-Oki earthquake using GRACE gravimetry [87]. Further, Shahrisvand et al., tried to detect (not predict) pre-seismic activities from GRACE data and studied pre-seismic gravity gradient anomaly in case of M8.8 2010 Chile-Maule earthquake, M9.0 2011 Tohoku-Oki earthquake, and M8.6 2012 Indian Ocean earthquake [25]. Note that, all of these three earthquakes are of magnitudes greater than M8.5 representing extreme earthquakes. Nevertheless, they could not find any gravity anomaly for earthquakes having lower magnitudes. Thus, to the best of our knowledge, no existing study has observed any pre-seismic gravity gradient anomaly for less than or equal M8.5 earthquakes. Moreover, no existing study has successfully adopted the pre-seismic gravity gradient anomaly for predicting earthquakes in general.
3.8 P-waves and S-waves based Study

Damaging S-waves travel at about 4 km/s whereas speed of P-waves are 7-8km/s [88]. When an earthquake takes place, then both waves are generated. P-waves travel twice faster than the S-waves. As an example, a city with a distance of 130 km at any direction from the epicenter takes 18 second and 32 seconds to reach the P-waves and S-waves respectively. Thus an alarm based on P-waves can give 16 seconds time to people before the earthquake can reach to them. On the other hand, if the city is around 370 KM distance from the epicenter, then the P-waves would have reached in about 53 seconds and S-waves would have reached in about 90 seconds. Therefore, if an alarm is generated using P-waves, then people can be warned 37 seconds earlier before the earthquake can catch them.

The Japan Meteorological Agency (JMA) has two Earthquake Early Warning (EEW) systems. One system is for the general public and another is for the National Meteorological and Hydrological Services. These systems are based on P-waves. When any P-waves are detected from at least two of the 4200 seismometers placed throughout Japan, the JMA automatically analyzes the P-waves. Then they predicts the magnitude of the earthquake and the rough area of the earthquake’s epicenter [89]. They also calculate how much area would be affected by the earthquake. If the magnitude is large enough to affect people, then through TV, radio, and mobile sms, the JMA notify people that a strong earthquake is expected.

It is worth to mention that, the systems can only give warnings before seconds or one or two minutes of the powerful S-waves hit and shaking due to earthquake gets dangerous. However, this small time before earthquake can mean the difference between life and death. It can be just enough time to take cover, stop a running car to the side of the road, not going into an elevator, or stop medical surgery.

Currently, P-waves are used for forewarning the earthquakes and tsunamis. However, P-waves are slow waves (7-8km/s) in comparison to gravity signal (speed of light) [27]. Therefore, it is yet to be widely adopted as a reliable precursor by the researchers. If we can replace the P-waves by the gravity signals, then we can forewarn an earthquake more early than P-waves based system before the S-waves reach.
3.9 Artificial Intelligence based Study

In a recent study, Asencio-Cortes et al., used artificial neural network (ANN) to predict earthquakes for only regions nearby Tokyo and 200 Km of its surroundings [29]. They have used several seismical indicators in their study. However, the area under their study is highly confined to a specific region and the study is yet to be generalized for other regions. Besides, Wang et al., conducted a study on the region nearby China and showed an improved result. Their mechanism of study exploited Long Short Term Memory (LSTM) of Deep Learning [30]. They used earthquakes greater than M4.5 from 1966 to 2016 as the input of their LSTM network, which are gathered from United States Geological Survey (USGS) earthquake database. They defined a time slot of one month. This study considered 600 data items and all earthquakes greater than M4.5 in their input dataset over a time span of 50 years (1966-2016). However, we have found from the USGS database that a total of 5508 earthquakes took place in this time frame over the considered region that are greater than M4.5. That indicates occurring an average of more than 9 earthquakes having magnitude greater than or equal to M4.5 in every month. Therefore, it is difficult or even near to impossible to correctly define or evaluate the prediction of earthquake in the time granularity of each month, as almost every month experiences earthquake occurring in such a case. Moreover, this study does not investigate application of its mechanism in any other region of the world other than China.

There exist other recent studies that are trying to predict earthquakes using different methods such as machine learning, bigdata, neural network, etc [29, 30]. Some of them are using statistical data, some are using historical earthquakes data, and some are using different seismic indicators as the input to their techniques. However, to the best of our knowledge, all of them are trying to predict earthquakes of a single specific region, however, not focusing on finding a general mechanism that can be applicable to predict earthquakes over different parts of the world [29, 30].
Chapter 4

Data Collection and Analysis

To do our job, we need the data of previous earthquakes. In addition, since we plan to study the prediction of earthquakes using Earth’s gravity field, we need the data of the gravitational field of the Earth. In this chapter we will discuss how we managed the data from the data sources. We will also discuss the preprocessing of the data.

4.1 Data Source

For the data of previous earthquakes and for the data of Earth’s gravitational field, we used two different data sources.

4.1.1 Database of Previous Earthquakes

United States Geological Survey (USGS) provides the data of previous earthquakes. The USGS Earthquake Hazards Program is a part of the National Earthquake Hazards Reduction Program (NEHRP). They monitor and report about earthquakes. They also assess earthquake impacts and hazards. Moreover, they research the causes and effects of earthquakes. Around the western United States, the USGS maintains a number of fault and volcano monitoring sites. There are tiltmeters, creepmeters, and strainmeters in these monitoring sites. The monitoring sites also include other environmental parameters such as temperature and barometric pressure. The data are collected and monitored to help understand why, when, and how fault slips, volcanic activities, and large earthquakes occur. Through the installed instruments, USGS measure the crustal deformation of the Earth before, during, and after the events in real-time. The goal is to better understand these unavoidable natural
processes. Another goal is to use these data to study the earthquakes and the volcanic hazards [90].

USGS contain information about every single earthquake over the world. From their site, anyone can search the earthquake database by giving a range of magnitude, a date range, and an area of geographic location over the Earth as parameters [91]. The output of the search result can be exported in CSV format in sorted order. The output contain various seismic parameters of each earthquake within the search range. Among them, time, latitude, longitude, magnitude, and place are considered in our study. These parameters are self explanatory, therefore, we don’t need to search for the explanation of the format.

We search the USGS earthquake database using the magnitude range greater than M6.0, date range start form April 2002 to August 2016, and geographic location of five different study regions [91]. We choose this time range, because the satellite data for Earth’s gravity field is available from April 2002. Next, we save one output file for each of the different regions and divide each of the datasets as training and testing dataset. We will discuss about the different regions as well as the division into training and testing dataset in Chapter 7. Table 4.1 shows a sample data file obtained from USGS. We can see headers at the top of every column in the sample USGS data which are self explanatory.

4.1.2 Data Source for Earth’s Gravitational Field

The Gravity Recovery and Climate Experiment (GRACE) is the first space mission which can measure the variation Earth’s gravity field precisely. Therefore, data retrieved from GRACE can be used as a tool for detecting changes in the Earth’s gravity field. National Aeronautics and Space Administration (NASA) and German Aerospace Center (DLR) both jointly launched the GRACE mission on March 2002. GRACE consists of two similar spacecrafts. They both are orbiting the Earth at an altitude of \( \sim 450 \text{ km} \). Both spacecrafts are separated by \( \sim 220 \text{ km} \) and linked with K-band micro-wave ranging (KBR) satellite-to-satellite tracking (SST) system [92, 25].

GRACE satellite provides monthly spherical harmonic coefficients of Earth’s gravity field [93, 94]. It has different releases of data. Three different organizations process the GRACE data and release them. They are

- The University of Texas, Center for Space Research (UTCSR) Austin
- German Research Centre for Geosciences (GeoForschungsZentrum, GFZ)
CHAPTER 4. DATA COLLECTION AND ANALYSIS

- Jet Propulsion Laboratory (JPL) that is owned by NASA and managed by the nearby California Institute of Technology (Caltech) for NASA.

Moreover, these organizations provide GAA, GAB, GAC, GAD, and GSM data [93, 94]. Among them, in this study, we adopt both RL 05 UTCSR GSM 96 data and RL 05 JPL GSM 90 monthly data separately which are composed of fully normalized spherical harmonic coefficients [95]. Here, 96 and 90 are the degree and order of the spherical harmonic of the data respectively. Zou et al., showed that calculation using $60 \leq \text{degree} \leq 100$ provide sensible values, whereas, $\text{degree} < 60$ provide average values and $\text{degree} > 100$ provide high frequency values that may contain noises [96]. Data from GRACE of monthly field solution covering from April 2002 to August 2016 are used. We use our adopted GRACE data for calculating gravity gradients. However, gravity gradients can be measured only for specific points. Therefore, we consider the Earth as $1^\circ \times 1^\circ$ grid [25].

The Earth’s oblateness values ($C_{20}$) has been replaced by those of the Satellite Laser Ranging (SLR) because of its accuracy [92]. We find latest SLR data from NASA [97]. For the study period, we computed the full gravitational gradient tensor (second derivative of gravitational potential) in spherical coordinates in $1^\circ \times 1^\circ$ spherical regular grid. To better demonstrate the local mass anomaly, the local north-east-down frame (NED) at each point with spherical coordinate is introduced: x axis is directed to the north, the y-axis to the east, and the z-axis to downward. Based on the principal of converting between coordinate systems, the full gravitational gradient tensor in this local NED frame is obtained [98].

4.2 GRACE Data Format

The data from GRACE we found processed by different organizations stated at previous section are in compressed format. We first decompress it. Then we need to know the data format of the data available in those files. Table 4.2 shows a sample data file found from GRACE. We also find the data format. Table 4.3 shows the GRACE data format with the sample data from Table 4.2.

4.3 Data Preprocessing

Data preprocessing is the key thing to be done for any calculation. We get the processed satellite data from different organizations. Therefore, before performing any desired calculation, we need to
preprocess the data and generate output as per our required format which can be then used as input in our next step calculation. At the next subsections, we discuss the preprocessing of the processed satellite data found from different organizations.

4.3.1 Calculation of Gravitational Field of Earth

In this study, we first calculate Gravity Gradient Tensor (GGT). A GGT consists of six components. GGT cannot be computed directly [99] and has to be computed in terms of the six components. The components are $V_{xx}$, $V_{xy}$, $V_{xz}$, $V_{yy}$, $V_{yz}$, and $V_{zz}$. To calculate the components, we first need to calculate the gravitational field of Earth.

The gravitational fields of the Earth can be computed by the following equation:

$$ V(r, \theta, \lambda) = \frac{GM}{r} \times \left( \sum_{n=2}^{\infty} \sum_{m=0}^{n} \left( \frac{a_e}{r} \right)^n \times (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) \times P_{nm}(\cos \theta) \right) $$

(4.1)

Here, $C_{nm}$ and $S_{nm}$ are the spherical harmonic (SH) coefficients [98], which describe the mass distribution within the Earth; $a_e$ is the equatorial radius; $r$, $\theta$, and $\lambda$ are the radius, co-latitude, and longitude respectively; $P_{nm}$ is the associated Legendre function; $GM$ is the universal gravitational constant multiplied by the mass of the Earth. Eventually, equatorial radius $a_e$ is replaced by $R$ as in this case, the sphere is the Earth.

4.3.2 Calculation of First Derivatives of the Gravitational Field of Earth

Based on Eq. 4.1, the first derivatives of $V$ with respect to $r$, $\theta$, and $\lambda$ can be derived as follows [98]:

$$ V_r(r, \theta, \lambda) = -\frac{GM}{R^2} \left( \sum_{n=2}^{\infty} (n+1) \left( \frac{R}{r} \right)^{n+2} \sum_{m=0}^{n} (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) P'_{nm}(\cos \theta) \right) $$

(4.2)

$$ V_{\theta}(r, \theta, \lambda) = -\frac{GM}{R} \left( \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} \sum_{m=0}^{n} (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) P_{nm}(\cos \theta) \sin \theta \right) $$

(4.3)

$$ V_{\lambda}(r, \theta, \lambda) = \frac{GM}{R} \left( \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} \sum_{m=0}^{n} m(-C_{nm} \sin m\lambda + S_{nm} \cos m\lambda) P_{nm}(\cos \theta) \right) $$

(4.4)
Accordingly, the second derivatives of $V$ with respect to $r$, $\theta$, and $\lambda$ are as follows [98]:

$$V_{rr}(r, \theta, \lambda) = \frac{GM}{R^3} \left( \sum_{n=2}^{\infty} (n+1)(n+2) \left( \frac{R}{r} \right)^{n+3} \sum_{m=0}^{n} (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda)P_{nm}(\cos \theta) \right)$$  \hspace{1cm} (4.5)

$$V_{r\theta}(r, \theta, \lambda) = \frac{GM}{R^2} \left( \sum_{n=2}^{\infty} (n+1) \left( \frac{R}{r} \right)^{n+2} \sum_{m=0}^{n} (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda)P'_{nm}(\cos \theta) \sin \theta \right)$$ \hspace{1cm} (4.6)

$$V_{r\lambda}(r, \theta, \lambda) = \frac{GM}{R} \left( \sum_{n=2}^{\infty} (n+1) \left( \frac{R}{r} \right)^{n+2} \sum_{m=0}^{n} m(C_{nm} \sin m\lambda - S_{nm} \cos m\lambda)P_{nm}(\cos \theta) \right)$$ \hspace{1cm} (4.7)

$$V_{\theta\theta}(r, \theta, \lambda) = \frac{GM}{R} \left( \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} \sum_{m=0}^{n} (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) \right) \left( P''_{nm}(\cos \theta) \sin^2 \theta - P'_{nm}(\cos \theta) \cos \theta \right)$$ \hspace{1cm} (4.8)

$$V_{\theta\lambda}(r, \theta, \lambda) = \frac{GM}{R} \left( \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} \sum_{m=0}^{n} m(C_{nm} \sin m\lambda - S_{nm} \cos m\lambda)P'_{nm}(\cos \theta) \sin \theta \right)$$ \hspace{1cm} (4.9)

$$V_{\lambda\lambda}(r, \theta, \lambda) = -\frac{GM}{R} \left( \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} \sum_{m=0}^{n} m^2(C_{nm} \cos m\lambda + S_{nm} \sin m\lambda)P_{nm}(\cos \theta) \right)$$ \hspace{1cm} (4.10)

### 4.3.3 Calculation of Gravity Gradient Tensor (GGT)

The gravitational gradients (full tensor) in the local NorthEastDown (NED) frame can be further derived as [98]:

$$V_{xx}(r, \theta, \lambda) = \frac{1}{r}V_r(r, \theta, \lambda) + \frac{1}{r^2}V_{\theta\theta}(r, \theta, \lambda)$$  \hspace{1cm} (4.11)
\[ V_{zy}(r, \theta, \lambda) = V_{yz}(r, \theta, \lambda) = \frac{1}{r^2 \sin \theta} (-\cot \theta V_{\lambda}(r, \theta, \lambda) + V_{\theta \lambda}(r, \theta, \lambda)) \] (4.12)

\[ V_{xz}(r, \theta, \lambda) = V_{zx}(r, \theta, \lambda) = \frac{1}{r} V_{r \theta}(r, \theta, \lambda) - \frac{1}{r^2} V_{\theta}(r, \theta, \lambda) \] (4.13)

\[ V_{yy}(r, \theta, \lambda) = \frac{1}{r} V_{r}(r, \theta, \lambda) + \frac{1}{r^2} \cot \theta V_{\theta}(r, \theta, \lambda) + \frac{1}{r^2 \sin^2 \theta} V_{\lambda \lambda}(r, \theta, \lambda) \] (4.14)

\[ V_{yz}(r, \theta, \lambda) = V_{zy}(r, \theta, \lambda) = \frac{1}{r \sin \theta} \left( V_{r \lambda}(r, \theta, \lambda) - \frac{1}{r} V_{\lambda}(r, \theta, \lambda) \right) \] (4.15)

\[ V_{zz}(r, \theta, \lambda) = V_{rr}(r, \theta, \lambda) \] (4.16)

We already have seen from Table 4.2 and Table 4.3 that, GRACE satellite provides monthly spherical harmonic coefficients of the Earth, i.e., \( C_{nm} \) and \( S_{nm} \). We calculate the GGT components from those satellite data using these equations. Note that, GGT components can be calculated only for any specific point and calculating GGT components of a single point is time consuming. On the other hand, in this study, we need to use the GGT components of the same points in multiple times. However, time complexity of calculating GGT components multiple times for the same points are exponential. To overcome these problems, we first divide the Earth into a \( 1^\circ \times 1^\circ \) grid [25] and plan to use dynamic programming. In extent to do that, we calculate the GGT components of each point of this \( 1^\circ \times 1^\circ \) grid and store them as they are the solution of our sub-problems. Here, a single sub-problem of our dynamic programming is to calculate the GGT components of a single point. We will use the solution of these sub-problems with the solution of other points at a later time. We use solution of sub-problems and apply dynamic programming to solve the bigger problem during our calculation. We will discuss the use of dynamic programming in our study in Chapter 6.
### Table 4.1: Sample data file obtained from USGS

<table>
<thead>
<tr>
<th>time</th>
<th>latitude</th>
<th>longitude</th>
<th>depth</th>
<th>mag</th>
<th>magType</th>
<th>nst</th>
<th>gap</th>
<th>id</th>
<th>updated</th>
<th>place</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003-06-03T23:58:02.740Z</td>
<td>-17.36</td>
<td>-72.807</td>
<td>33</td>
<td>6</td>
<td>mwb</td>
<td>303</td>
<td>42</td>
<td>usp000bz00</td>
<td>2015-05-13T18:53:46.000Z</td>
<td>near the coast of southern Peru</td>
</tr>
</tbody>
</table>

### Table 4.2: Sample data file obtained from GRACE

<table>
<thead>
<tr>
<th>1</th>
<th>FIRST</th>
<th>GSM-2,2016221-2016247,0027,UTCSR,0096,0005</th>
<th>SHM</th>
<th>UT-CSR</th>
<th>20161128</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>CMMNT</td>
<td>/scratch/00079/byaa705/grav/RL05,16-08/c/iter/L2/GSM-2,2016221-2016247,0027,UTCSR,0096,0005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>CMMNT</td>
<td>Insert</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>EARTH</td>
<td>0.39860004415E+15 0.6378136300E+07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>SHM</td>
<td>96 96 1.00 fully normalized inclusive permanent tide</td>
<td>86 86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>CMMNT</td>
<td>Reported standard deviations are formal (not calibrated)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>GRCOF2</td>
<td>0 0 0.10000000000000000E+01 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 nnnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>GRCOF2</td>
<td>1 0 0.00000000000000000E+00 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 nnnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>GRCOF2</td>
<td>2 0 -0.84169638146E-03 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 ynnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>GRCOF2</td>
<td>3 0 0.957037312922E+00 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 ynnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>GRCOF2</td>
<td>4 0 0.539912297847E-03 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 ynnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>GRCOF2</td>
<td>5 0 0.68668495003E+00 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 ynnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>GRCOF2</td>
<td>6 0 -0.149987593157E-06 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 ynnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>GRCOF2</td>
<td>7 0 0.904414435031E-07 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 ynnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>GRCOF2</td>
<td>8 0 0.494819535001E-07 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 ynnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>GRCOF2</td>
<td>9 0 0.280633870023E-07 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 ynnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>GRCOF2</td>
<td>10 0 0.53375138297E-07 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 ynnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>GRCOF2</td>
<td>11 0 -0.507579103733E-07 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 ynnn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>GRCOF2</td>
<td>12 0 0.36477949570E-07 0.00000000000000000E+00 0.0000E+00 0.0000E+00 20160808.0000 20160904.0000 ynnn</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.3: GRACE data format of the sample data shown in Table 4.2

<table>
<thead>
<tr>
<th>Line 1</th>
<th>FIRST</th>
<th>GSM-2 2016221-2016247,UTCSR_0096_0005</th>
<th>SHM</th>
<th>UT-CSR</th>
<th>20161128</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>Keyword (Fixed, Mandatory, Must be first record of the file)</td>
<td>PID-2,YYYYDOY-YYYYDOY_dddd_ssss_mmmm_rrrr</td>
<td>SHM</td>
<td>SPHERICAL HARMONIC MODEL</td>
<td>Generating institute</td>
</tr>
<tr>
<td>Line 2</td>
<td>CMMNT</td>
<td>/scratch/00079/byaa705/grav/RL05_16-08/c/iter/L2/GSM-2_2016221-2016247_0027,UTCSR_0096_0005</td>
<td>Comment</td>
<td>This is the comment</td>
<td></td>
</tr>
<tr>
<td>Line 3</td>
<td>CMMNT</td>
<td>Insert</td>
<td>Comment</td>
<td>This is the comment</td>
<td></td>
</tr>
<tr>
<td>Line 4</td>
<td>EARTH</td>
<td>0.3986004415E+15</td>
<td>0.6378136300E+07</td>
<td>0.3986004415E+15</td>
<td>0.6378136300E+07</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Earth related line</td>
<td>GM: gravitational constant times mass of Earth, unit: ( m^3 s^{-2} )</td>
<td>R: mean equator radius, unit: m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Line 5</td>
<td>SHM</td>
<td>96</td>
<td>96</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>SHM Related line</td>
<td>maximum degree of model</td>
<td>scale factor applied to given std. dev. (if SCALE=0 or ' ', no std. dev. given)</td>
<td>'fully normalized' or 'unnormalized....'</td>
<td>'inclusive permanent tide' or 'exclusive permanent tide' or 'not applicable.....'</td>
</tr>
<tr>
<td>Line 6</td>
<td>CMMNT</td>
<td>Reported standard deviations are formal (not calibrated)</td>
<td>Comment</td>
<td>This is the comment</td>
<td></td>
</tr>
<tr>
<td>Line 7</td>
<td>GRCOF2</td>
<td>0</td>
<td>0</td>
<td>0.100000000000E+01</td>
<td>0.000000000000E+00</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>GRACE Coefficient related line</td>
<td>degree ( l )</td>
<td>order ( m )</td>
<td>( C_m ) coefficient</td>
<td>( S_m ) coefficient</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>C_m adjusted: ( y ): yes, ( n ): no, S_m adjusted: ( y ): yes, ( n ): no, stochastic a priori inf. for ( C_m ): ( y ): yes, ( n ): no, stochastic a priori inf. for ( S_m ): ( y ): yes, ( n ): no</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5

Preliminary Observation

Figure 5.1: Point A has the latitude ($\phi$) of 4°N and longitude ($\lambda$) of 92°E. Point B has the latitude of 4°N and longitude of 102°E. Latitudes and longitudes for point C are 4°S and 102°E, and for D are 4°S and 92°E respectively. These four points together confine a region of our initial analysis.

Going beyond the approaches of the existing studies [25], we plan to analyze whether there arise any anomaly before earthquakes (can be of smaller magnitudes) using the GRACE satellite data. The existing studies generally remain confined to only epicenters of a few earthquakes under their
Figure 5.2: M7.0+ earthquakes within the period and region of our study are shown in this figure. From Table 5.1, we can see that, there are 14 EQs within our period of study in this region. However, in this figure, we can see 13 EQs. Here, we have ignored an M7.4 EQ, which appears in the first row of the table, as we do not have enough prior data (GGT time series data of two years before occurring the EQ) for this EQ in our study.

consideration for such an analysis. Unlike to them, we plan to focus not only to the epicenters of few earthquakes, however, also on several different regions of the Earth in parallel.

5.1 Startup

At an initial stage, we try to find out anomaly in Gravity Gradient Tensor (GGT) [25] before earthquakes in the region of Sumatra-Andaman area, more specifically for the region having \( \phi = 4^\circ N \) to \( \phi = 4^\circ S \), and \( \lambda = 92^\circ E \) to \( \lambda = 102^\circ E \) (Fig. 5.1) during the time range from April 2002 to August 2016. Reason behind choosing this time frame is - GRACE satellite data is available for only this period of time. Primarily, we gather all the information of the quakes having magnitude greater than or equal to M7.0 that took places within the defined region and time frame. As a result, we got only 14 quakes from the USGS earthquake database as presented in Table 5.1 and Fig. 5.2. We plot the
Table 5.1: All M7.0+ earthquakes information in Sumatra-Andaman region, gathered from USGS earthquake database

<table>
<thead>
<tr>
<th>Serial</th>
<th>Time</th>
<th>Latitude ($\phi$)</th>
<th>Longitude ($\lambda$)</th>
<th>Magnitude</th>
<th>Place</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>02/11/2002</td>
<td>2.824</td>
<td>96.085</td>
<td>7.4</td>
<td>Simeulue Indonesia</td>
</tr>
<tr>
<td>2</td>
<td>26/12/2004</td>
<td>3.295</td>
<td>95.982</td>
<td>9.1</td>
<td>Off the west coast of Northern Sumatra</td>
</tr>
<tr>
<td>3</td>
<td>28/03/2005</td>
<td>2.085</td>
<td>97.108</td>
<td>8.6</td>
<td>Northern Sumatra Indonesia</td>
</tr>
<tr>
<td>4</td>
<td>12/09/2007</td>
<td>-2.625</td>
<td>100.841</td>
<td>7.9</td>
<td>Kepulauan Mentawai region Indonesia</td>
</tr>
<tr>
<td>5</td>
<td>12/09/2007</td>
<td>-4.438</td>
<td>101.367</td>
<td>8.4</td>
<td>Southern Sumatra Indonesia</td>
</tr>
<tr>
<td>6</td>
<td>13/09/2007</td>
<td>-2.13</td>
<td>99.627</td>
<td>7.0</td>
<td>Kepulauan Mentawai region Indonesia</td>
</tr>
<tr>
<td>7</td>
<td>20/02/2008</td>
<td>2.768</td>
<td>95.964</td>
<td>7.4</td>
<td>Simeulue Indonesia</td>
</tr>
<tr>
<td>8</td>
<td>25/02/2008</td>
<td>-2.486</td>
<td>99.972</td>
<td>7.2</td>
<td>Kepulauan Mentawai region Indonesia</td>
</tr>
<tr>
<td>9</td>
<td>30/09/2009</td>
<td>-0.72</td>
<td>99.867</td>
<td>7.6</td>
<td>Southern Sumatra Indonesia</td>
</tr>
<tr>
<td>10</td>
<td>06/04/2010</td>
<td>2.383</td>
<td>97.048</td>
<td>7.8</td>
<td>Northern Sumatra Indonesia</td>
</tr>
<tr>
<td>11</td>
<td>25/10/2010</td>
<td>-3.487</td>
<td>100.082</td>
<td>7.8</td>
<td>Kepulauan Mentawai region Indonesia</td>
</tr>
<tr>
<td>12</td>
<td>10/01/2012</td>
<td>2.433</td>
<td>93.21</td>
<td>7.2</td>
<td>Off the west coast of Northern Sumatra</td>
</tr>
<tr>
<td>13</td>
<td>11/04/2012</td>
<td>2.327</td>
<td>93.063</td>
<td>8.6</td>
<td>Off the west coast of Northern Sumatra</td>
</tr>
<tr>
<td>14</td>
<td>11/04/2012</td>
<td>0.802</td>
<td>92.463</td>
<td>8.2</td>
<td>Off the west coast of Northern Sumatra</td>
</tr>
</tbody>
</table>

Time series of each of the six components of GGT for each of the quakes. We take two years of the GGT components just prior to each quake and calculate the mean of this two years. We also calculate the standard deviation (STD) of GGT components over two years just prior to each quake. Then, we draw five lines on the time series. The lines are mean (average) line, ± STD lines, and ± 2STD lines.

5.2 Analysis

Here, we find significant anomalies in the time series before the earthquakes. Fig. 5.3 shows the anomalies just two or three months before the three different earthquakes of M9.1, M7.6, and M8.2, which took places at different times in the same region very near to each other. Note that, in each of the subfigures, we can find more anomalies other than the marked ones. To further analyze these anomalies we mark the other anomalies of Fig. 5.3c in Fig. 5.4.
(a) Anomalies detected in GGT components time series for an M9.1 earthquake

(b) Anomalies detected in GGT components time series for an M7.6 earthquake
CHAPTER 5. PRELIMINARY OBSERVATION

38

(c) Anomalies detected in GGT components time series for an M8.2 earthquake

Figure 5.3: Anomalies detected in time series of GGT components before three different sample earthquakes out of the 14 earthquakes as pointed in Table 5.1. Fig. 5.3a represents an EQ that took place off the west coast of Northern Sumatra in Indian Ocean in 2004. Location of this EQ was $\phi=3.295^\circ$N and $\lambda=95.982^\circ$E (Sl. 2 in Table 5.1). Fig. 5.3b represents another EQ that took place near Southern Sumatra Indonesia in 2009. Location of this EQ was at $\phi=0.72^\circ$S and $\lambda=99.867^\circ$E (Sl. 9 in Table 5.1). Fig. 5.3c represents another EQ that took place at $\phi=0.802^\circ$N and $\lambda=92.463^\circ$E off the west coast of Northern Sumatra in 2012 (Sl. 14 in Table 5.1). The green vertical lines in all the time series shown in Fig. 5.3a, 5.3b, and 5.3c indicate the point of EQ happening. We can see, five out of six GGT components in Fig. 5.3a, three out of six components in Fig. 5.3b, and two out of six components in Fig. 5.3c show significant anomalies before two or three months of corresponding earthquakes. These anomalies are annotated by red circles.

Fig. 5.4 points out some other anomalies other than those that appear just before the quake. Our further investigation reveals that, some other quakes took places in the time series just after the newly marked anomalies in Fig. 5.4. To demonstrate these other quakes further, in Fig. 5.5, we show these nearby quakes plotted on the same time series. All three figures (Fig. 5.3, 5.4, and 5.5) in combination demonstrates worthiness of analyzing anomaly in GGT in predicting earthquakes.
5.3 Primary Decision

From our observations, we find the following:

- Not each of the GGT components exhibit notable anomaly before each of the earthquakes, and
- Not the same GGT component always exhibits notable anomaly before all earthquakes event at the same region.

Based on these findings, we perform our future study. We present methodology of our study in the next section.

Figure 5.4: GGT component time series for an M8.2 EQ (Sl. 14 in Table 5.1), which is the same EQ shown in Fig. 5.3c. This figure suggests, other than just prior to the M8.2 EQ anomalies, there exist other anomalies. These anomalies are annotated by green circles.
Figure 5.5: This figure represents time series of GGT components for an M8.2 earthquake (Sl. 14 in Table 5.1), which is again the same EQ of Fig. 5.3c. However, this time, we plot other quakes occurred in nearby locations. After plotting other quakes, we can find that, the anomalies marked by green circles are just before other nearby quakes. Two anomalies marked by the green circles over $V_{x x}$ and $V_{y z}$ appear before the M7.6 EQ (Sl. 9 in Table 5.1), which took place 846.8 Km away from the original M8.2 EQ. Two more anomalies are marked by green circles over $V_{x z}$ and $V_{z z}$ appear before the M7.8 EQ (Sl. 10 in Table 5.1), which happened 543.03 Km away from the M8.2 EQ. Therefore, these anomalies are also indications of earthquakes, which took place at nearby points.
Chapter 6

Methodology of Our Study

According to our preliminary observation, since all GGT components do not exhibit anomaly before an earthquake, we plan to design a model to combine all GGTs and detect if any of the GGT components exhibits anomaly. Fig. 6.1 shows the block diagram detecting anomaly using our proposed system. Here, $G_{xy}$ reflect the anomaly. Since any of the six GGT components may have the anomaly, we need to combine them and detect the anomaly from the combined one. The right most box in Fig. 6.1 show the box from which we can detect the anomaly.

![Figure 6.1: Block diagram of our proposed model to detect anomaly](image)

In addition, anomalies can be detected at nearby points. As a result, nearby surrounding points might accumulate effect to a center point. Therefore, in addition with combining all GGT components
for a single point, our model incorporates GGT anomalies from the nearby surrounding points in such a way that increasing distance of the surrounding points reduces the effect of the points to the center point. To do so, while incorporating the GGT anomalies of surrounding points, we multiply them with corresponding weights, which depend on the distance between the center point and the nearby point.

Finally, our model determines a threshold value applicable to the GGT components. If any GGT component exceeds this threshold value, that indicates prediction of a potential earthquake. It is worth mentioning that the same threshold value may not work for different regions of the world. The reason behind this is, the crustal structure and strength are not same for all parts of the Earth [44]. Moreover, tectonic plates move with different speeds and directions at different locations of the Earth [48]. Therefore, the same threshold would not work for everywhere on the Earth. To address this issue, we propose to divide the Earth in some small blocks, and then find appropriate thresholds for each blocks. Subsequently, we can apply those specific thresholds for their corresponding blocks of the Earth to predict earthquakes of the blocks. Accordingly, threshold values in our proposed mechanism will be region-specific. Thus, for different regions, the output threshold values will be different.

Our proposed mechanism learn from previous earthquakes of a block and calculate the threshold for the same block based on that learning. Afterwards, the mechanism test the threshold on some test data. To do this, we divide our dataset into two parts, training dataset and validation dataset. The division process of the data into training and validation set is described in Chapter 7.

6.1 Boundary Region

After calculating the GGT components at the data processing step, we apply our proposed mechanism. While applying the mechanism to a particular point in the study region, at first, we imagine the point as the center of a coordinate system. Then we consider all the points from $-4^\circ$ to $+4^\circ$ from the center point with a granularity of $1^\circ$ in both latitude and longitude directions. We call this area as the “Boundary Region” for the center point. This is shown in Fig 6.2. From this figure we can see that, Boundary Region for a center point consists a total of 81 points including the center point. Here, we take the points of Boundary Region in our calculation considering that, the points of the Boundary Region exhibit impact upon the center point. Mathematically, we define the Boundary Region as follows:
CHAPTER 6. METHODOLOGY OF OUR STUDY

Figure 6.2: Each point under consideration for earthquake prediction is imagined as an origin (long, lat). We then consider the GGT components of all the points up to ±4° from the origin to calculate the GGT anomaly at origin. This figure is for calculation in a 1° × 1° grid. All the points up to ±4° from the origin including the origin is treated as the Boundary Region.

\[
\text{Boundary Region} = \forall_{i=\phi+4^\circ}^{\phi-4^\circ} \forall_{i=\lambda+4^\circ}^{\lambda-4^\circ} \text{ Point}(i,j) \tag{6.1}
\]

6.2 Reducing Tidal and Atmospheric Effect

Now, if we want to predict earthquake for a particular point in a particular month, then we first need to take two years average of each GGT components at each points of Boundary Region for that particular point and up to that particular month. The two years are considered just prior from that particular month. The two years average of these GGT components helps us to reduce the impact of tidal effects, atmospheric effects, and noise [25].

6.3 Learning from the Most Stable Part of Earth’s Nature

Note that, we actually do not take the average of GGT components for exactly two years. In many cases, we observed that, after a great earthquake, there is a huge gravity change. If we consider these changes in the calculation of finding average of GGT components, then the result will contain unexpected noise. However, we just want to know the gravity anomaly from the regular gravity distribution which is the stable part of Earth’s nature. Therefore, we apply filtering to reduce the effect of noise induced by the huge gravity change due to some other great earthquakes.
To do the filtering, for a single GGT component and for a specific point, according to last section, at first we take 24 months (two years) into consideration just prior to the selected month. Now, keeping the starting month of this 24 months fixed, we decrease the end month of these 24 months by one down to the middle of these 24 months. We calculate the standard deviation (STD) of the GGT component for each of these reduced periods. On the other hand, we do the same by keeping the end month fixed and increasing the starting month of these 24 months by one up to the middle of these 24 months. Then we take the minimum STD from this set of calculated STDs. We also take into consideration the associated starting and ending month of this minimum STD. While calculating the STD, we set the limit of increment of the starting month or that of the decrement of the ending month up to the middle of the 24 months, because, we want to ensure at least one year average. Instead of taking fixed two years average, we take average of the GGT component for the time period denoted by starting and ending month. We do the same for all the six GGT components for all the points and for all the months in our study region and period and get the average.

\[
minSTD = \min \left( \forall \text{Month}_{\text{cur}} - 12 \leq \text{Month}_{\text{end}} \leq \text{Month}_{\text{cur}} - 1 (STD(GGT)) \right) \\
\cup \forall \text{Month}_{\text{cur}} - 24 \leq \text{Month}_{\text{start}} \leq \text{Month}_{\text{cur}} - 12 (STD(GGT)) \right) \quad (6.2)
\]

### 6.4 Calculation of minSTDAvg and absGGT

Subsequently, for each single GGT component, for a particular month, and for a specific point, we calculate difference between the GGT component (over all possible duration prior to that particular month) and the average of GGT over the period from \(\text{Month}_{\text{start}}\) to \(\text{Month}_{\text{end}}\), as specified above. After that, we take the STD and absolute values of the calculated differences for corresponding start and end period. Subsequently, we calculate average of the absolute values of the differences for corresponding start and end period, and denote this newly calculated average as minSTDAvg. Additionally, we call the absolute values of GGT components as absGGT.
6.5 Calculation of STDMultiplier

While experimenting with each individual GGT component, we observe that all of the six GGT components do not reflect anomaly before an earthquake. Even if we consider GGT components of the epicenter of an earthquake, this statement holds true. Rather, only some of the GGT components show anomaly before an earthquake. Moreover, the same set of components does not reflect anomaly for different EQs even in the same region. Nonetheless, different sets of components reflect the anomaly at different times. Therefore, to grab the anomaly, we need to consider all six GGT components together.

To grab the anomaly considering all six GGT components together, we plan to individually compare value of each GGT component against a certain threshold value. We represent this threshold value as a multiple of the corresponding STD (the STD, which we have already calculated from its absGGT). The multiplier is named as STDMultiplier and mathematically as follows:

\[
threshold = STD \times STDMultiplier
\]

6.6 GGTsum

When the value of a GGT component becomes greater than corresponding threshold times of STD from its minSTDAvg, it serves as a potential indicator for an impending earthquake. Here, since we need to consider all the GGT components, we sum up the absGGTs above corresponding threshold for each of the six GGT components. In this way, we get a summed up value of all the six threshold_filtered_absGGTs for a point in the Boundary Region. We call it as GGTsum and calculate it as follows:

\[
GGTsum = \sum_{i=1}^{6} threshold\_filtered\_absGGT_i
\]

Since, the points of the Boundary Region might exhibit impact on the center point, we sum up GGTsums of all the points of the Boundary Region. To do so, we calculate GGTsums for all 81 points of the Boundary Region for a particular center point (a point in our study area).
6.7 Calculation of weightedGGTsum and cumulatedWGGTsum

While summing up the GGTsums of different points, the distance between a point and the center may vary for different points in the Boundary Region. It is obvious that, distant points should exhibit less impacts than a nearer point from the center. To realize this lessening impact, we multiply each threshold-filtered absGGT by a weight while calculating the GGTsums. The weight depends on distance between the point of GGT calculation and the center. In this way, instead of summing up all the threshold_filtered_absGGTs directly for a particular point, we sum up all the weighted-threshold_filtered_absGGTs for that point. Accordingly, we call it weightedGGTsum. Further, we sum up the weightedGGTsums of all 81 points in the Boundary Region to deduce a cumulatedWGGTsum. Here, we use the following set of equations:

\[
weight = \begin{cases} 
1, & \text{if the point is the center point} \\
\frac{0.75}{\max(\phi \text{ distance from center}, \lambda \text{ distance from center})}, & \text{otherwise}
\end{cases}
\]  

(6.5)

\[
\text{weightedGGTsum} = \sum_{i=1}^{i=6} weight \times \text{threshold_filtered_absGGT}_i
\]  

(6.6)

\[
\text{cumulatedWGGTsum} = \sum_{i \in \text{Points in the Boundary Region}} \text{weightedGGTsum}_i
\]  

(6.7)

We do the same calculations for all months under our study period. By doing this, we get a time series values of cumulatedWGGTsum pertinent to the center point. We calculate the time series values for all the points of the region under our study.

6.8 Use of Dynamic Programming

While calculating the minSTD, we need the GGT components of last 24 months for each point in each month. As we have already discussed in Chapter 4 that we saved the GGT components for each point for every month, we used those saved values to speed up the process and not to run redundant calculations that may lead to exponential computation time.

Besides, while calculating cumulatedWGGTsum, the same values of these variables are required
for multiple times. This happens as a point may be in different Boundary region as a boundary region consists of 81 points. Therefore, calculation of the values of minSTD, minSTDAvg, absGGT, and weightedGGTsum for the same point again may incur redundant computation, which in turn may lead to exponential time complexity. To solve this issue, we store values for each point and for each month. Accordingly, we use dynamic programming over the stored values. Thus, whenever we need any value of minSTD, minSTDAvg, absGGT, or weightedGGTsum, we can find it already calculated and we just read the value from the stored ones.

6.9 Calculation of STDMultiplier using Multi Objective Evolution

Our next task is to find out the STDMultiplier in such a way that we can get an anomaly before an earthquake. Here, we have two different objectives. The first objective is to predict as many earthquakes as possible. On the other hand, the second objective is to generate as small number of alarm alarms as possible. We realize accumulation of both the objectives one by one.

6.9.1 Learning Based on True Positive Cases

To achieve our first objective, we start with an initial value of the STDMultiplier as 1. Accordingly, we check if we can predict EQs in the training dataset with 100% accuracy with this value of STDMultiplier. If so, then we multiply the STDMultiplier by 2, get the next evolution, and again check the same. We multiply the STDMultiplier again by 2 in case of achieving 100% accuracy, get the next evolution, and check the same. We do this exponential search up to having an accuracy of 100%. During this exponential search based on true positive cases, we ignore the false positive cases.

Now, when the accuracy falls below 100% in our exponential search, we take the corresponding STDMultiplier as the upper limit. In addition, we divide this STDMultiplier by 2, get the previous evolution, and take this value as the lower limit. Next, we perform a binary search over the evolution range defined by the upper and lower limits, and find out the specific STDMultiplier value for which exact 100% true prediction can be achieved. We take decimal precision up to two digits for calculating the STDMultiplier in this process of binary search. We consider this STDMultiplier value as STDMultiplierForTruePrediction.
6.9.2 Learning by Avoiding False Positive Cases

To achieve our second objective, we take the false positive cases into consideration after performing the above-mentioned tasks for getting the highest accuracy in prediction. At first, we start from the value of last calculated STDMultiplier (STDMultiplierForTruePrediction) and check if the training dataset gives us zero percent of false positive case. If not, then we multiply the STDMultiplier by 2, get the next evolution, and check again the same. In this way, we run the exponential search again to get zero percent false positive case. However, to get a suitable value of STDMultiplier for zero percent of false positive case may not be possible in all cases. Because, while minimizing the false positive cases by increasing the STDMultiplier as a power of “2”, we will lose the true prediction percentage in parallel. To solve this trade off, we set a minimum threshold of true prediction percentage as 75% while performing the exponential search for false positive cases. In other words, whatever the false positive cases there could be, we do not allow a true prediction percentage below 75%.

By setting the minimum threshold for true prediction percentage at 75%, we conduct the exponential search on the value of STDMultiplier up to the point where false positive cases get to zero or true positive cases get below 75%. After getting the STDMultiplier for which this condition meets, we again perform a binary search in the similar manner to pin point the best value STDMultiplier for which the condition holds. We consider this value as $STDMultiplierForFalsePrediction$.

6.10 Trade-off between Our Objectives and Calculation of a Final Value of STDMultiplier

We now consider 100 points between $STDMultiplierForTruePrediction$ and $STDMultiplierForFalsePrediction$ with equidistant interval. Next, we calculate the true and false positive cases for each of these 100 points in the interval. Finally, we find the value of the STDMultiplier among these 100 points for which difference between corresponding true prediction and false prediction becomes the maximum. We consider the value of this STDMultiplier as the $finalSTDMultiplier$.

$$finalSTDMultiplier = \max_{i \in \text{Values in the equidistant points}} (true\ prediction_i - false\ prediction_i) \quad (6.8)$$
6.11 Accuracy Validation

As the finishing step, we test the validation dataset using the finalSTDMultiplier and find a results on true prediction and false prediction. This results portray efficacy of our proposed approach. The results of validation along with the validation dataset is described in Chapter 7.
Chapter 7

Prediction Results and Comparative Analysis

For demonstrating generality of our proposed mechanism, we take five different regions around the globe in consideration for our analysis. The regions are chosen based on different frequencies of earthquakes happenings over there. Three of them have been taken as highly-frequent earthquake prone areas. One has been taken as medium-frequent earthquake prone area. The last one has been taken as less-frequent earthquake prone area. For all regions, period of our study spans from April 2002 to August 2016, as GRACE satellite data is available for only this period of time.

- Japan, Chile, and Sumatra-Andaman areas have been chosen as highly-frequent earthquake prone areas.
- China has been chosen as the medium-frequent earthquake prone area
- India-Nepal area has been chosen as the less-frequent earthquake prone area

Note that, we take different numbers of months as our training periods for the different regions, considering the event of experiencing an earthquake after a substantial amount of time. This substantial amount of time having no earthquake refers to a quiet period. Such quiet periods often segregate between two different eras of earthquake happening [100, 101]. Our training periods span from 2002 up to starting of the quiet periods, and our validation periods span the time onward. Thus, the quiet periods denote demarcations between the training and validation periods in our study. As the quiet periods vary over the different regions, our training and validation periods also vary in term of the
numbers of months over different regions. Nonetheless, later in this study, we will investigate impacts of varying training periods over the accuracy of prediction results.

7.1 Study in the Area of Japan and Its Results

For Japan, our study region is $\phi = 33^\circ N$ to $44^\circ N$ and $\lambda = 131^\circ E$ to $145^\circ E$. We take all the earthquakes having magnitude greater than or equal to M6.0 for this area from USGS earthquake database. As a result, we get a total of 167 earthquakes within our study period. For the purpose of our study, we divide data on these EQs into two sets - a training set and a validation set. We differentiate these sets by corresponding time periods. For training dataset, we take the training period of a total 99 months spanning from 2002 to 2010 for both UTCSR and JPL data. During this training period, a total of 57 EQs occurred. For validation dataset, we take 54 and 60 months spanning from 2011 to 2016 for UTCSR and JPL data respectively. A total of 110 EQs occurred within this time period. Note that, we take the first 99 months as our training period, as there was an earthquake after a substantial amount of time following the 99th month. We run our proposed mechanism on the training dataset using GGTs calculated from UTCSR data and calculated the final STDMultiplier as 4.40631866 (for detail, see methods). Now, when we run our proposed mechanism on validation dataset with the calculated STDMultiplier, we get 26 different months for which, the combined GGT anomaly model exceeds the EQ possibility threshold. 102 EQs out of 110 from the validation dataset took place within four months prior to these 26 months. That results in 93% of true prediction. Besides, while using the UTCSR data for the Japan region, we have not found any false positive cases resulting in a 0% of false positive cases.

Subsequently, we run the mechanism on JPL data with same training dataset and calculate the 3.07517609 as the STDMultiplier. By running our proposed mechanism on validation dataset using this STDMultiplier, we get 29 different months that exceed their GGT anomaly from the EQ possibility threshold. 106 EQs out of 110 took place within four months period prior to the 29 months resulting in 96% of true prediction. Here, we get three months more that exceed their threshold of anomaly, however, no EQ took place within four months prior from these months. Therefore, these three months indicate false alarms resulting in 3% of false positive cases. Note that the reported false positive case is calculated with respect to the total number of earthquakes occurred in the region during the time period under consideration.
7.2 Study in the Area of Chile and Its Results

Next, we run our proposed mechanism on the region of Chile, specifically, from $\phi = 17^\circ S$ to $58^\circ S$ and $\lambda = 66^\circ W$ to $81^\circ W$. For this region, we get a total of 93 earthquakes within the period of our study from the USGS earthquake database. Here, we take first 87 months from 2002 to 2009 as the training period and got 39 earthquakes within the training dataset. Next, 67 and 73 months from 2010 to 2015 for UTCSR and JPL data respectively contribute 54 earthquakes in the validation dataset. Note that, we take the first 87 months as our training period, as there was an earthquake after a substantial amount of time following the 87th month. We calculate the STDMultiplier as $5.7890625$ for UTCSR data. By running our proposed mechanism on the validation data using this STDMultipliers, we see that GGT anomaly exceeds over the EQ possibility threshold for 33 different months. 48 EQs out of 54 took place within four months period prior to these 24 months among those 33 months resulting in 88% true prediction. Other nine months exhibit false alarms here with a false positive of 16%. Besides, by running our proposed mechanism with newly calculated STDMultiplier of $5.81314987$ for the JPL data, we get exceeding of GGT anomaly ove EQ possibility threshold in 31 different months. 46 EQs out of 54 took place within four months prior to 16 months among those 31 months resulting in 85% true prediction. Other 15 months contribute to false alarms resulting in 25% false positive cases.

7.3 Study in the Area of Sumatra-Andaman and Its Results

For Sumatra-Andaman area, we chose the study region from $\phi = 4^\circ N$ to $4^\circ S$ and $\lambda = 92^\circ E$ to $102^\circ E$. We get 78 EQs in this area within the time specified. Here, we set 88 months from 2002 to 2009 as the training period for both UTCSR and JPL data. As the validation period, we set 65 and 71 months from 2010 to 2015 for UTCSR and JPL respectively. Thus, we get 59 EQs in the training dataset and 19 EQs in the validation dataset. Note that, we take the first 88 months as our training period, as there was an earthquake after a substantial amount of time following the 88th month. For UTCSR, we calculate the STDMultiplier as $4.32947144$, and for JPL, we calculate it as $3.1484375$. While using UTCSR data, we get GGT anomaly exceeding the EQ possibility threshold in 29 months. 24 months out of the 29 have 15 earthquakes within four months period prior to them. These 15 EQs out of 19 results in 79% true positive case. Besides, five months exhibit anomaly where no EQ took place within four months period indicating 25% false positive case ($5 FP / (5 FP + 15 True EQ) = 25\%$). In
case of JPL data, 38 months exhibit the anomaly, where 29 months cover 17 EQs within four months period prior to them. Out of 19 EQs, true predictions of 17 EQs result in 89% true positive case. However, anomaly in nine months without any EQ within four months period prior to them results in 35% false positive cases.

7.4 Study in the Area of China and Its Results

In case of China, we take the region from \( \phi = 23^\circ N \) to \( 45^\circ N \) and \( \lambda = 75^\circ E \) to \( 119^\circ E \). From USGS earthquake database, we get 45 M6.0+ earthquakes in this region within the specified time period. From 2002 to 2011, we take a total of 112 months as the training period and get 28 earthquakes in the training dataset for both UTCSR and JPL data. Subsequently, 41 and 47 months contribute as validation periods for UTCSR and JPL respectively. Note that, we take the first 112 months as our training period, as there was an earthquake after a substantial amount of time following the 112th month. We calculate STDMultiplier as 5.97513809 for UTCSR and as 5.17301788 for the JPL data. A total of 19 different months show anomalies within the validation period for UTCSR data. 15 out of the 17 EQs with 88% true positive case took place within the four months period from one of the anomaly indicating 17 months out of the 19. Only two months exhibit anomaly, however, no EQ occurred within four months prior to those two months resulting in 12% false positive case. For JPL data, the true and false positive cases remain as it is that we have found for UTCSR data.

7.5 Study in the Area of India-Nepal and Its Results

In case of India-Nepal region, we take \( \phi = 18^\circ N \) to \( 30^\circ N \) and \( \lambda = 85^\circ E \) to \( 96^\circ E \) as the region under our study. From USGS earthquake database, we find only 16 EQs in this region within the time period previously specified. Since, the total number of EQ is small here, we set 137 months from 2002 to 2014 as our training period to get 11 EQs in the training dataset. Other 16 and 22 months from 2015 to 2016 contribute to validation period and we get five EQs in the validation dataset during this period. Note that, we take the first 137 months as our training period, as there was an earthquake after a substantial amount of time following the 137th month. The STDMultipliers are calculated here as 3.25698532 and 5.68906021 from UTCSR and JPL data respectively. With the multipliers, we can find the GGT anomaly exceeding EQ possibility threshold in 11 different months for UTCSR data. All five EQs in the validation dataset are within four months period prior to one of the 11 months
resulting in 100% true positive case. Note that, as an anomaly indicates a potential earthquake within a period of four months, multiple anomalies can indicate one single EQ by iteratively appearing within the period. The iterative appearances of anomalies result in 11 anomalies for the five EQs in this case. Additionally, we do not find any false positive here resulting in 0% false positive case. For JPL data, we get anomaly in 12 months. All five EQs in the validation dataset are within four months period prior to these 12 months with 100% true positive case. On the other hand, similar to the previous case, we do not find any false positive here too resulting in 0% false positive case.

7.6 Results at a Glance

Table 7.1 summarizes all the results elaborated in previous sections for different regions. Besides, Fig 7.1 - 7.5 demonstrate GGT anomalies for few earthquakes having different magnitudes and different regions of occurring. Here, in some cases, the anomaly gets disappeared. However, this happens only in a limited number of cases though having anomalies in adjacent months. Thus, the anomalies in adjacent months remain capable enough to produce advance alarm through predicting the earthquake.

7.7 Different Aspects of Our Proposed Mechanism

We preform several studies for investigating different aspects of our mechanisms. One aspect is to take a fixed training and validation period for all the experimental region. Another aspect is to vary the training and validation time periods, and analyze sensitivity of division between the training and validation datasets on the prediction results. One more aspect is to explore applicability of our proposed mechanism over a region that rarely experiences earthquakes.

7.7.1 Fixed Training and Validation Period for All Regions

We take a fixed period of 2003-2012 as the training period and 2013-2016 as the validation period for the region of China, Japan, India with Nepal, and Chile. The only exception is that we take 2003-2010 as the training period and 2011-2016 as validation period for the Sumatra-Andaman region. Reason behind varying the period for Sumatra-Andaman is, we get only five earthquakes in the validation dataset if we would take the same time period (2013-2016) similar to other regions as its validation period.
(a) GGT anomaly started before 4 months of the M8.8 Chile-Maule earthquake.

(b) GGT anomaly continued before 3 months of the M8.8 Chile-Maule earthquake.

(c) GGT anomaly continued before 2 months of the M8.8 Chile-Maule earthquake.

(d) GGT anomaly continued before 1 month of the M8.8 Chile-Maule earthquake.

(e) At the month of the M8.8 Chile-Maule earthquake, black circles indicate earthquakes. Here, we can see more than one earthquake. Those are the aftershocks of the M8.8 Chile-Maule earthquake.

Figure 7.1: Progression of the M8.8 Chile-Maule earthquake (in the year 2010) from the four months period before the earthquake happening up to the happening...
(a) GGT anomaly started before 4 months of the M9.0 Tohoku-Oki earthquake.

(b) GGT anomaly disappeared before 3 months of the M9.0 Tohoku-Oki earthquake.

(c) GGT anomaly taken place again before 2 months of the M9.0 Tohoku-Oki earthquake. This time extent of the anomaly appears to be greater.

(d) GGT anomaly continued before 1 month of the M9.0 Tohoku-Oki earthquake. Again, extent of the anomaly appears to be much greater.

(e) At the month of the M9.0 Tohoku-Oki earthquake, black circles indicate earthquakes. Here, we can see more than one earthquake. Those are the aftershocks of the M9.0 Tohoku-Oki earthquake.

Figure 7.2: Progression of the M9.0 Tohoku-Oki earthquake (in the year 2011) from the four months period before the earthquake happening up to the happening
(a) GGT anomaly started before 4 months of the M8.6 and M8.2 Indian Ocean earthquakes.

(b) GGT anomaly continued before 3 months of the M8.6 and M8.2 Indian Ocean earthquakes.

(c) GGT anomaly continued before 2 months of the M8.6 and M8.2 Indian Ocean earthquakes.

(d) GGT anomaly disappeared before 1 month of the M8.6 and M8.2 Indian Ocean earthquakes.

(e) At the month of the M8.6 and M8.2 Indian Ocean earthquakes, black circles indicate earthquakes.

Figure 7.3: Progression of the M8.6 and M8.2 Indian Ocean earthquake (in the year 2012) from the four months period before the earthquake happening up to the happening
(a) GGT anomaly started before 4 months of the M6.8 Myanmar earthquake.

(b) GGT anomaly continued before 3 months of the M6.8 Myanmar earthquake.

(c) GGT anomaly continued before 2 months of the M6.8 Myanmar earthquake.

(d) GGT anomaly disappeared before 1 month of the M6.8 Myanmar earthquake.

(e) At the month of the M6.8 Myanmar earthquake, black circle indicates the earthquake.

Figure 7.4: Progression of the M6.8 Myanmar earthquake (in the year 2012) from the four months period before the earthquake happening up to the happening
(a) GGT anomaly started before 4 months of the M6.7 Imphal, India earthquake.
(b) GGT anomaly continued before 3 months of the M6.7 Imphal, India earthquake.
(c) GGT anomaly continued before 2 months of the M6.7 Imphal, India earthquake.
(d) GGT anomaly continued before 1 month of the M6.7 Imphal, India earthquake.
(e) At the month of the M6.7 Imphal, India earthquake, black circle indicates the earthquake.

Figure 7.5: Progression of the M6.7 Imphal, India earthquake (in the year 2016) from the four months period before the earthquake happening up to the happening
Table 7.1: Results obtained from our proposed mechanism for different regions of the world using different data sources

<table>
<thead>
<tr>
<th>Experimental region</th>
<th>Data source</th>
<th>Training dataset (# of EQs)</th>
<th>Training period (years)</th>
<th>Validation dataset (# of EQs)</th>
<th>Validation period</th>
<th>Threshold value calculated</th>
<th># of EQs predicted (Training)</th>
<th>True positive % (Training)</th>
<th># of EQs predicted (Validation)</th>
<th>True positive % (Validation)</th>
<th># of false alarms (Validation)</th>
<th>False positive % (Validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>China and Nepal</td>
<td>UTCSR GSM 96</td>
<td>28</td>
<td>2002-2011</td>
<td>17</td>
<td>2012-2016</td>
<td>5.975138092</td>
<td>27</td>
<td>96</td>
<td>15</td>
<td>2</td>
<td>88</td>
<td>12</td>
</tr>
<tr>
<td>China and Nepal</td>
<td>JPL GSM 95</td>
<td>28</td>
<td>2002-2011</td>
<td>17</td>
<td>2012-2016</td>
<td>5.173017883</td>
<td>27</td>
<td>96</td>
<td>15</td>
<td>2</td>
<td>88</td>
<td>12</td>
</tr>
<tr>
<td>Japan</td>
<td>UTCSR GSM 96</td>
<td>55</td>
<td>2002-2010</td>
<td>110</td>
<td>2011-2016</td>
<td>4.406318665</td>
<td>51</td>
<td>92</td>
<td>102</td>
<td>0</td>
<td>93</td>
<td>0</td>
</tr>
<tr>
<td>Japan</td>
<td>JPL GSM 95</td>
<td>55</td>
<td>2002-2010</td>
<td>110</td>
<td>2011-2016</td>
<td>3.075176086</td>
<td>53</td>
<td>96</td>
<td>106</td>
<td>3</td>
<td>96</td>
<td>3</td>
</tr>
<tr>
<td>India and Nepal</td>
<td>UTCSR GSM 96</td>
<td>11</td>
<td>2002-2014</td>
<td>5</td>
<td>2015-2016</td>
<td>3.256985321</td>
<td>11</td>
<td>100</td>
<td>5</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>India and Nepal</td>
<td>JPL GSM 95</td>
<td>11</td>
<td>2002-2014</td>
<td>5</td>
<td>2015-2016</td>
<td>5.689060211</td>
<td>10</td>
<td>90</td>
<td>5</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Chile</td>
<td>UTCSR GSM 96</td>
<td>39</td>
<td>2002-2009</td>
<td>54</td>
<td>2010-2015</td>
<td>5.789062500</td>
<td>39</td>
<td>100</td>
<td>48</td>
<td>9</td>
<td>89</td>
<td>16</td>
</tr>
<tr>
<td>Chile</td>
<td>JPL GSM 95</td>
<td>39</td>
<td>2002-2009</td>
<td>54</td>
<td>2010-2015</td>
<td>5.813149872</td>
<td>34</td>
<td>87</td>
<td>46</td>
<td>15</td>
<td>85</td>
<td>25</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For China with part of Nepal region, we get 31 and 14 earthquakes in training and validation datasets respectively. We achieve 100% and 85% true positive rate with a false positive rate of only 6% and 7% for JPL and UTCSR data respectively.

For Japan, the training dataset consists of 149 earthquakes whereas the validation dataset consists of 17 earthquakes. Here, JPL data gives us 88% true prediction with 16% false positive cases. However, UTCSR data gives only 47% true prediction in this case with a 0% false cases.

For India and Nepal region, we get only 6 earthquakes in training dataset and 10 earthquakes in validation dataset. Here, JPL gives us 100% true prediction with 0% false cases. However, again, UTCSR performs poor this time. UTCSR gives only 20% true positive cases with 0% false positive cases. This could be due to less number of earthquakes in the training dataset.

We get 86 and 45 earthquakes in the region of Chile as training and validation datasets respectively. Here, JPL gives 93% true positive with only 4% false positive cases. UTCSR gives 57% true positive with 0% false positive case.
Finally, for the region of Sumatra-Andaman, we get 64 and 14 earthquakes respectively in training and validation datasets. We get 85% and 71% true positive cases with 40% and 28% false positive cases for JPL and UTCSR data respectively.

Table 7.2 shows the results of this experiment.

Table 7.2: Results obtained from our proposed mechanism for different regions of the world using different data sources with fixed training and validation periods

<table>
<thead>
<tr>
<th>Experimental region</th>
<th>Data source</th>
<th>Training dataset (# of EQs)</th>
<th>Training period (years)</th>
<th>Validation dataset (# of EQs)</th>
<th>Validation period (years)</th>
<th>Threshold value calculated</th>
<th>True positive % (Training)</th>
<th>True positive % (Validation)</th>
<th>False positive % (Validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>China + Nepal</td>
<td>JPL GSM 90</td>
<td>31</td>
<td>2003-2012</td>
<td>14</td>
<td>2013-2016</td>
<td>5.1884</td>
<td>90</td>
<td>100</td>
<td>6</td>
</tr>
<tr>
<td>Japan</td>
<td>JPL GSM 90</td>
<td>149</td>
<td>2003-2012</td>
<td>17</td>
<td>2013-2016</td>
<td>2.7338</td>
<td>98</td>
<td>88</td>
<td>16</td>
</tr>
<tr>
<td>India + Nepal</td>
<td>JPL GSM 90</td>
<td>6</td>
<td>2003-2012</td>
<td>10</td>
<td>2013-2016</td>
<td>2.6172</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Chile</td>
<td>JPL GSM 90</td>
<td>86</td>
<td>2003-2012</td>
<td>45</td>
<td>2013-2016</td>
<td>2.6953</td>
<td>100</td>
<td>93</td>
<td>4</td>
</tr>
<tr>
<td>Sumatra-Andaman</td>
<td>JPL GSM 90</td>
<td>64</td>
<td>2003-2010</td>
<td>14</td>
<td>2011-2016</td>
<td>3.1484</td>
<td>100</td>
<td>85</td>
<td>40</td>
</tr>
<tr>
<td>China + Nepal</td>
<td>UTCSR GSM 96</td>
<td>31</td>
<td>2003-2012</td>
<td>14</td>
<td>2013-2016</td>
<td>5.9702</td>
<td>96</td>
<td>85</td>
<td>7</td>
</tr>
<tr>
<td>Japan</td>
<td>UTCSR GSM 96</td>
<td>149</td>
<td>2003-2012</td>
<td>17</td>
<td>2013-2016</td>
<td>5.1009</td>
<td>92</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>India + Nepal</td>
<td>UTCSR GSM 96</td>
<td>6</td>
<td>2003-2012</td>
<td>10</td>
<td>2013-2016</td>
<td>5.0547</td>
<td>83</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>Chile</td>
<td>UTCSR GSM 96</td>
<td>86</td>
<td>2003-2012</td>
<td>45</td>
<td>2013-2016</td>
<td>5.7804</td>
<td>93</td>
<td>57</td>
<td>0</td>
</tr>
<tr>
<td>Sumatra-Andaman</td>
<td>UTCSR GSM 96</td>
<td>64</td>
<td>2003-2010</td>
<td>14</td>
<td>2011-2016</td>
<td>4.3295</td>
<td>95</td>
<td>71</td>
<td>28</td>
</tr>
</tbody>
</table>

7.7.2 Sensitivity of Division between Training and Validation Datasets over Prediction Accuracy

We explore the aspect of sensitivity of division between training and validation datasets over prediction accuracy through varying time intervals pertinent to training and validation datasets in the region of Sumatra-Andaman. Here, we mainly focus on two types of time variations. In the first case, we take a
fixed time of two years for both training and validation datasets. We vary this two-year intervals over
the time period from 2003 to 2016 keeping the training and validation datasets in successive years.
Thus, this study also explores what we would obtain in case we consider only the recent two-year data
in training rather that considering a longer period of history as we have done in our earlier analysis.

In the later study, we start with a small training dataset of only two years with a validation dataset
of 12 years. Then, we gradually increase the training time by two years and decrease the validation
time by the same duration. This study explores what we would get while having longer period of
history in our consideration.

**Study with Consecutive Training and Validation Datasets Each having A Duration of
Two Years**

We perform this study over the region of Sumatra-Andaman. Here, we take earthquakes of two years of
different disjoint time slices over the time period from 2003 to 2014 as the training dataset. Besides,
we take earthquakes from successive two years after the training period as the validation dataset.
Thus, the first time slices under study cover earthquakes over the period of 2003-2004 as training
dataset and earthquakes over the period of 2005-2006 as the validation dataset. In this study, the
number of earthquakes in the training dataset is only five whereas the number of earthquakes in the
validation dataset is 26. Since the training dataset is too small in terms of the number of earthquakes
within it, we get only 30\% and 11\% success rate for UTCSR and JPL data respectively with a 0\%
false positive case.

The next time slices cover earthquakes over the period of 2005-2006 as the training dataset and
earthquakes over the period of 2007-2008 as the validation dataset. Here, the number of earthquakes
in the training and validation datasets are 26 and 23 respectively. This time, we can predict 100\% of
earthquakes in the validation dataset using both UTCSR and JPL data. Besides, false positives are
4\% and 0\% for UTCSR and JPL respectively.

Next, we take earthquakes over the period of 2007-2008 as training dataset and that over 2009-
2010 as validation dataset. This time the training dataset consists of 23 earthquakes and validation
dataset consists of ten earthquakes. We can predict 100\% of these earthquakes using both UTCSR
and JPL data. We also notice that, we do not experience any false positive cases in case of JPL data,
however, UTCSR data exhibits 9\% false alarms.

Later, we consider earthquakes over the time period of 2009-2010 as training dataset and that
over 2011-2012 as validation dataset. This time, the numbers of earthquakes in both training and validation datasets are only ten and nine respectively. Since, the training dataset contains a small number of earthquakes here, we get only 55% and 22% prediction accuracy using UTCSR and JPL data respectively with 0% false alarms with both the types of data.

In the next study, earthquakes from the time period over 2011-2012 are considered in training dataset whereas that over 2013-2014 are considered in validation dataset. Nine and three earthquakes are found in the training and validation datasets respectively in this case. As the training dataset again contains a very small number of earthquakes, we can predict only 33% of earthquakes with both the types of data. Moreover, we experience 50% false alarms with UTCSR data. Fortunately, JPL data does not raise any false alarm in this case.

Finally, we perform our last study of this kind by taking earthquakes over the period of 2013-2014 in training dataset and that over 2015-2016 in validation dataset. The numbers of earthquakes found in training and validation datasets are three and two respectively in this case. Though, we can predict 100% of these earthquakes in spite of having very small number of earthquakes in the training dataset, we experience 50% and 75% false alarms with UTCSR and JPL data respectively.

Table 7.3 shows the results of these experiments. The results demonstrate that we can achieve considerable accuracy only if there exist a good number of earthquakes in the training dataset. Therefore, considering a couple of years even if in the latest history may result in degraded prediction accuracy when the latest history does not contain a good number of earthquakes.

**Study with Gradually Increasing Training Periods**

In this study, we gradually increase the training period by two years for the region of Sumatra-Andaman. Here, we take rest of the periods in the validation dataset.

Similar to the previous investigation, we first take earthquakes from the two years period over 2003-2004 as training dataset. However, unlike to previous investigation, we take earthquakes from the rest 12 years period over 2005-2016 as the validation dataset. Here, we get five earthquakes in the training dataset and 73 earthquakes in the validation dataset. Since, the training dataset contains a very small number of earthquakes, we achieve only 9% and 46% true prediction having 46% and 11% false alarms using JPL and UTCSR data respectively.

Our next study of this kind increases the training period by two years while decreasing the validation period also by two years. Thus, we increase the period from 2003-2004 to 2003-2006 for training
Table 7.3: Results obtained from our proposed mechanism for Sumatra-Andaman region considering successive two-year intervals for training and validation datasets using different data sources

<table>
<thead>
<tr>
<th>Experimental region</th>
<th>Data source</th>
<th>Training dataset (# of EQs)</th>
<th>Training period (years)</th>
<th>Validation dataset (# of EQs)</th>
<th>Validation period (years)</th>
<th>Threshold value calculated</th>
<th># of EQs predicted (Training)</th>
<th>True positive % (Training)</th>
<th># of EQs predicted (Validation)</th>
<th>True positive % (Validation)</th>
<th># of false alarms (Validation)</th>
<th>False positive % (Validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>5</td>
<td>2003-2004</td>
<td>26</td>
<td>2005-2006</td>
<td>684.0541134</td>
<td>5</td>
<td>100</td>
<td>3</td>
<td>0</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>UTCSR GSM96</td>
<td>5</td>
<td>2003-2004</td>
<td>26</td>
<td>2005-2006</td>
<td>4.739166107</td>
<td>4</td>
<td>80</td>
<td>8</td>
<td>0</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>26</td>
<td>2005-2006</td>
<td>23</td>
<td>2007-2008</td>
<td>3.1484375</td>
<td>26</td>
<td>100</td>
<td>23</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>UTCSR GSM96</td>
<td>26</td>
<td>2005-2006</td>
<td>23</td>
<td>2007-2008</td>
<td>3.3823125</td>
<td>26</td>
<td>100</td>
<td>23</td>
<td>1</td>
<td>100</td>
<td>4</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>23</td>
<td>2007-2008</td>
<td>10</td>
<td>2009-2010</td>
<td>4.788200673</td>
<td>23</td>
<td>100</td>
<td>10</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>UTCSR GSM96</td>
<td>23</td>
<td>2007-2008</td>
<td>10</td>
<td>2009-2010</td>
<td>4.3828125</td>
<td>23</td>
<td>100</td>
<td>10</td>
<td>1</td>
<td>100</td>
<td>9</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>10</td>
<td>2009-2010</td>
<td>9</td>
<td>2011-2012</td>
<td>6.299404144</td>
<td>9</td>
<td>90</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>UTCSR GSM96</td>
<td>10</td>
<td>2009-2010</td>
<td>9</td>
<td>2011-2012</td>
<td>6.840444565</td>
<td>9</td>
<td>90</td>
<td>5</td>
<td>0</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>9</td>
<td>2011-2012</td>
<td>3</td>
<td>2013-2014</td>
<td>3.94493542</td>
<td>8</td>
<td>88</td>
<td>1</td>
<td>0</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>UTCSR GSM96</td>
<td>9</td>
<td>2011-2012</td>
<td>3</td>
<td>2013-2014</td>
<td>4.0859375</td>
<td>9</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>33</td>
<td>50</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>3</td>
<td>2013-2014</td>
<td>2</td>
<td>2015-2016</td>
<td>1.6484375</td>
<td>3</td>
<td>100</td>
<td>2</td>
<td>6</td>
<td>100</td>
<td>75</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>3</td>
<td>2013-2014</td>
<td>2</td>
<td>2015-2016</td>
<td>3.1015625</td>
<td>3</td>
<td>100</td>
<td>2</td>
<td>2</td>
<td>100</td>
<td>50</td>
</tr>
</tbody>
</table>

dataset set while we decrease the period from 2005-2016 to 2007-2016 for validation dataset. We get 31 earthquakes in the training dataset over this four years of training period and 47 earthquakes in validation dataset over this ten years of validation period. We achieve 95% of true prediction in this case for both the types of data. However, we get 25% and 21% false alarms using JPL and UTCSR data respectively.

Next, we again increase our period for training dataset by two years, i.e., from 2003-2006 to 2003-2008, and decrease the period for validation dataset by two years, i.e., from 2007-2016 to 2009-2016. For these training and validation periods, we get 54 and 24 earthquakes respectively. We achieve 91% true prediction from this study. We also get 29% false alarms from JPL data and 24% false alarms from UTCSR data.

Later, we consider earthquakes in the training period over 2003-2010 by increasing the period again by two years and earthquakes in the validation period over 2011-2016 by decreasing the period.
by two years. For these training and validation periods, we get 64 and 14 earthquakes respectively. Here, JPL data give 85% true prediction with 40% false alarms. On the other hand, UTCSR gives 71% true prediction with 28% false alarms.

The next study considers earthquakes over the period of 2003-2012 in the training dataset and that over 2013-2016 in the validation dataset. The numbers of earthquakes in training and validation datasets are 73 and five respectively. We get 80% true prediction with 63% false prediction from JPL data. We also get 40% true prediction with 50% false alarms from UTCSR data.

Finally, we perform our last study of this kind by taking earthquakes over the period of 2003-2014 in the training dataset and that over 2015-2016 in the validation dataset. We get 76 earthquakes in the training dataset and only two earthquakes in the validation dataset. This time, we get 100% true prediction using JPL data with 77% false alarms. Besides, we could not predict any earthquake from the two of them (0% true prediction) while we get a single false alarm using UTCSR data.

Table 7.4 shows these results. The results demonstrate that there exists mostly a marginally decreasing trend in true prediction with an increase in the training period and there exists mostly an increasing trend in false prediction with an increase in the training period. Overall, we can achieve the best performances both in term of true prediction and false alarms when the periods of training and validation dataset are closely balanced. This is also reflected in Fig 7.6.
Table 7.4: Results obtained from our proposed mechanism for Sumatra-Andaman region with gradually increasing time periods of the training dataset by two years using different data sources

<table>
<thead>
<tr>
<th>Experimental region</th>
<th>Data source</th>
<th>Training dataset (# of EQs)</th>
<th>Training period (years)</th>
<th>Validation dataset (# of EQs)</th>
<th>Validation period (years)</th>
<th>Threshold value calculated</th>
<th># of EQs predicted (Training)</th>
<th>True positive % (Training)</th>
<th># of EQs predicted (Validation)</th>
<th># of false alarms (Validation)</th>
<th>True positive % (Validation)</th>
<th>False positive % (Validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>5</td>
<td>2003-2004</td>
<td>73</td>
<td>2003-2016</td>
<td>684.0541134</td>
<td>5</td>
<td>100</td>
<td>7</td>
<td>6</td>
<td>9</td>
<td>46</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>UTCSR GSM96</td>
<td>5</td>
<td>2003-2004</td>
<td>73</td>
<td>2005-2016</td>
<td>4.739166107</td>
<td>4</td>
<td>80</td>
<td>46</td>
<td>6</td>
<td>63</td>
<td>11</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>31</td>
<td>2003-2006</td>
<td>47</td>
<td>2007-2016</td>
<td>3.1474375</td>
<td>31</td>
<td>100</td>
<td>45</td>
<td>15</td>
<td>95</td>
<td>25</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>UTCSR GSM96</td>
<td>31</td>
<td>2003-2006</td>
<td>47</td>
<td>2005-2016</td>
<td>3.3828125</td>
<td>31</td>
<td>100</td>
<td>45</td>
<td>12</td>
<td>95</td>
<td>21</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>54</td>
<td>2003-2008</td>
<td>24</td>
<td>2009-2016</td>
<td>3.1484375</td>
<td>54</td>
<td>100</td>
<td>22</td>
<td>9</td>
<td>91</td>
<td>29</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>UTCSR GSM96</td>
<td>54</td>
<td>2003-2008</td>
<td>24</td>
<td>2009-2016</td>
<td>3.3828125</td>
<td>54</td>
<td>100</td>
<td>22</td>
<td>7</td>
<td>91</td>
<td>24</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>64</td>
<td>2003-2010</td>
<td>14</td>
<td>2011-2016</td>
<td>3.1484375</td>
<td>64</td>
<td>100</td>
<td>12</td>
<td>8</td>
<td>85</td>
<td>40</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>UTCSR GSM96</td>
<td>64</td>
<td>2003-2010</td>
<td>14</td>
<td>2011-2016</td>
<td>4.329471436</td>
<td>61</td>
<td>95</td>
<td>10</td>
<td>4</td>
<td>71</td>
<td>28</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>73</td>
<td>2003-2012</td>
<td>5</td>
<td>2013-2016</td>
<td>3.019041901</td>
<td>72</td>
<td>98</td>
<td>4</td>
<td>7</td>
<td>80</td>
<td>63</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>UTCSR GSM96</td>
<td>73</td>
<td>2003-2012</td>
<td>5</td>
<td>2013-2016</td>
<td>4.329471436</td>
<td>69</td>
<td>94</td>
<td>2</td>
<td>2</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>JPL GSM90</td>
<td>76</td>
<td>2003-2014</td>
<td>2</td>
<td>2015-2016</td>
<td>3.033704529</td>
<td>74</td>
<td>97</td>
<td>2</td>
<td>7</td>
<td>100</td>
<td>77</td>
</tr>
<tr>
<td>Sumatra Andaman</td>
<td>UTCSR GSM96</td>
<td>76</td>
<td>2003-2014</td>
<td>2</td>
<td>2015-2016</td>
<td>4.328315277</td>
<td>71</td>
<td>93</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

7.7.3 Applicability of Our Proposed Mechanism over A Region having Rare Earthquakes

To analyze applicability of our mechanism in case of rarely-happening earthquakes, we choose a region where earthquakes happen very rarely. Iran is such a place, as here a strong earthquake (M6.0+) takes place in every two years on an average. For Iran, our study region covers from $\phi = 29^\circ N$ to $43^\circ N$ and from $\lambda = 42^\circ E$ to $52^\circ E$. Over the time period from 2002 to 2016, here, we get only eight M6.0+ earthquakes. This number is too small considering other cases considered earlier in this study.

First, we consider earthquakes over the period of 2002-2011 in training dataset and that over 2012-2016 in validation dataset. Thus, we get four earthquakes in both training and validation datasets.

We perform our proposed mechanism using this training and validation dataset and found 100% true prediction with a false alarm rate of 69% in case of using UTCSR data. On the other hand, JPL data gives us 50% true prediction with 88% false alarms. The results are poor substantially below what
we have got earlier, as we do not have enough earthquakes in our training dataset.

To avoid the poor performance, we explore a different approach. Here, we do not divide the dataset in training and validation datasets. Rather, we take the whole dataset as the validation dataset. Since, in this case, we do not have any training dataset, we can not calculate STDMultiplier directly. We calculate the average of STDMultiplier from all the previously experimented region and validate the dataset using this value of STDMultiplier for both UTCSR and JPL data. This time, we get 87% true prediction with 84% false alarms for UTCSR data, and 62% true prediction with 90% false alarms for JPL data. This approach also gives us poor results similar to the earlier case.

We again try with yet another approach. Here, we take all the earthquakes in the validation dataset and set the maximum STDMultiplier value found from our other previously experimented regions as the STDMultiplier while running our proposed mechanism in this case. This time, we get 37% true prediction with 75% false alarms for UTCSR data. We also get 62% true prediction with 84% false alarms for JPL data.

We conclude that, we do not have enough earthquakes to learn the appropriate STDMultiplier value for this region where earthquake takes place very rarely. Therefore, we can predict the earthquakes in this region using our proposed mechanism, however, the rate of false alarms becomes too high in this region.

Table 7.5: Results obtained from our proposed mechanism for the region of Iran

<table>
<thead>
<tr>
<th>Experimental region</th>
<th>Data source</th>
<th>Training dataset (# of EQs)</th>
<th>Training period (years)</th>
<th>Vali-dation dataset (# of EQs)</th>
<th>Validation period (years)</th>
<th>Threshold value calculated</th>
<th># of EQs predicted (Training)</th>
<th>True positive % (Training)</th>
<th># of false alarms (Validation)</th>
<th>False positive % (Validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iran</td>
<td>JPL GSM90</td>
<td>4</td>
<td>2002-2011</td>
<td>4</td>
<td>2012-2016</td>
<td>4.297625122</td>
<td>4</td>
<td>100</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Iran</td>
<td>UTCSR GSM96</td>
<td>4</td>
<td>2002-2011</td>
<td>4</td>
<td>2012-2016</td>
<td>4.3671875</td>
<td>4</td>
<td>100</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Iran</td>
<td>JPL GSM90</td>
<td>N/A</td>
<td>2002-2016</td>
<td>N/A</td>
<td>N/A</td>
<td>4.66558176</td>
<td>N/A</td>
<td>N/A</td>
<td>5</td>
<td>45</td>
</tr>
<tr>
<td>Iran</td>
<td>UTCSR GSM96</td>
<td>N/A</td>
<td>2002-2017</td>
<td>N/A</td>
<td>N/A</td>
<td>4.66558176</td>
<td>N/A</td>
<td>N/A</td>
<td>7</td>
<td>39</td>
</tr>
<tr>
<td>Iran</td>
<td>JPL GSM90</td>
<td>N/A</td>
<td>2002-2018</td>
<td>N/A</td>
<td>N/A</td>
<td>5.97513809</td>
<td>N/A</td>
<td>N/A</td>
<td>5</td>
<td>32</td>
</tr>
<tr>
<td>Iran</td>
<td>UTCSR GSM96</td>
<td>N/A</td>
<td>2002-2019</td>
<td>N/A</td>
<td>N/A</td>
<td>5.97513809</td>
<td>N/A</td>
<td>N/A</td>
<td>3</td>
<td>9</td>
</tr>
</tbody>
</table>
Chapter 8

Our Findings in A Nutshell

We compare our proposed mechanism with several other existing studies and find that our mechanism outperforms all those studies every one in terms of both matrices - true positive and false positive. Table 8.1 shows the comparison. On the other hand, Table 8.2 shows specific earthquake and region based comparison among ours and other existing methods.

It is worth mentioning that our endover in this study is to predict M6.0+ earthquakes all over the world. In extent to achieve that,

- We propose a new earthquake prediction mechanism based on gravity gradient tensor using satellite data.
- We test our proposed mechanism for five different regions of the Earth and successfully predict the earthquakes having magnitudes greater than or equal to M6.0.
- We apply our proposed mechanism over the data of all the five regions obtained from two different data sources.
- We get a successful prediction rate (true positive) of 91% on an average having only 13% false alarms (false positive) on an average.
- Our false alarm percentage is very low in most of the regions except in one region (Sumatra-Andaman). We analyze the false cases in the exceptional region and find that the false alarms are generated at the boundary of the region under study.
<table>
<thead>
<tr>
<th>Author, year</th>
<th>Mechanism</th>
<th>Prediction window</th>
<th>Experimental data source</th>
<th>Experimental data gathered from</th>
<th>Experimental areas</th>
<th>Tested in multiple regions</th>
<th>Studied EQs</th>
<th>True positive</th>
<th>False positive</th>
<th>EQ range</th>
<th>Data range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newaz et al., 2017 [28]</td>
<td>Our proposed mechanism</td>
<td>4 months</td>
<td>GRACE satellite data and seismic data</td>
<td>UTCSR, JPL, and USGS</td>
<td>Japan, Chile, Sumatra-Andaman, India-Nepal, and China</td>
<td>Yes</td>
<td>399 EQs</td>
<td>91%</td>
<td>13%</td>
<td>M6.0 to above</td>
<td>14 years</td>
</tr>
<tr>
<td>Asencio-Cortes et al., 2017 [29]</td>
<td>EQP-ANN</td>
<td>0.25 months</td>
<td>Seismic data</td>
<td>NSGS</td>
<td>200 Km around Tokyo</td>
<td>No</td>
<td>400 EQs</td>
<td>70%</td>
<td>Too high</td>
<td>M5.0 to above</td>
<td>2 years</td>
</tr>
<tr>
<td>Wang et al., 2017 [30]</td>
<td>LSTM</td>
<td>1 month</td>
<td>Seismic data</td>
<td>USGS</td>
<td>China</td>
<td>No</td>
<td>5508 EQs</td>
<td>77%</td>
<td>23%</td>
<td>M4.5 to above</td>
<td>50 years</td>
</tr>
<tr>
<td>Molchan et al., 2017 [31]</td>
<td>Probabilistic Parimutuel Gambling (PG) with statistical data</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Italy</td>
<td>No</td>
<td>10 EQs</td>
<td>70%</td>
<td>N/A</td>
<td>M5.4 to M5.6</td>
<td>18 years</td>
</tr>
<tr>
<td>Uphoff et al., 2017 [26]</td>
<td>ADER-DG</td>
<td>N/A</td>
<td>Open source meshing software Gmsh</td>
<td>Zenodo</td>
<td>Sumatra-Andaman</td>
<td>N/A</td>
<td>Only Sumatra-Andaman earthquake in 2004</td>
<td>N/A</td>
<td>N/A</td>
<td>M9.0</td>
<td>N/A</td>
</tr>
<tr>
<td>Montagner et al., 2016 [27]</td>
<td>Statistical data analysis from Super Conducing Gravimeter</td>
<td>Few minutes</td>
<td>Mizusawa VERA observatory and five F-net stations closest to the observatory</td>
<td>National Astronomical Observatory of Japan</td>
<td>Japan</td>
<td>N/A</td>
<td>Only Tohoku-Oki earthquake in 2011</td>
<td>N/A</td>
<td>N/A</td>
<td>M9.0</td>
<td>2 months</td>
</tr>
<tr>
<td>Shahrisvand et al., 2014 [25]</td>
<td>NARX neural network</td>
<td>1.25 Months</td>
<td>GRACE satellite data</td>
<td>UTSR</td>
<td>Chile, Japan, and India</td>
<td>N/A (Not multiple area, Just three specific EQ)</td>
<td>Only Chile Maule earthquake in 2010, Tohoku-Oki earthquake in 2011, and Indian Ocean earthquake in 2012</td>
<td>N/A</td>
<td>M8.5 to above</td>
<td>9 years</td>
<td></td>
</tr>
<tr>
<td>Linage et al., 2009 [28]</td>
<td>Normal-modes summation</td>
<td>No window, co-seismic and post-seismic study</td>
<td>GRACE satellite data</td>
<td>UTSR</td>
<td>Sumatra-Andaman</td>
<td>N/A</td>
<td>Only Sumatra-Andaman earthquake in 2004</td>
<td>N/A</td>
<td>N/A</td>
<td>M9.0</td>
<td>4.6 years</td>
</tr>
</tbody>
</table>
Table 8.2: Specific earthquake and region-based comparison among our proposed mechanism and other existing studies

<table>
<thead>
<tr>
<th>Incident / Metric</th>
<th>Montagner et al., 2016</th>
<th>Shahrisvand et al., 2014</th>
<th>Newaz et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chile Maule earthquake (2010, Chile)</td>
<td>Predicted</td>
<td>Predicted</td>
<td>Predicted</td>
</tr>
<tr>
<td>Tohoku-Oki earthquake (2011, Japan)</td>
<td>Not Studied</td>
<td>Predicted</td>
<td>Predicted</td>
</tr>
<tr>
<td>Indian Ocean earthquake (2012, Sumatra-Andaman)</td>
<td>Not Studied</td>
<td>Predicted</td>
<td>Predicted</td>
</tr>
<tr>
<td>True positive for M6.0+ earthquake in the region of Chile</td>
<td>Not Studied</td>
<td>Not Studied</td>
<td>89%</td>
</tr>
<tr>
<td>True positive for M6.0+ earthquake in the region of Japan</td>
<td>Not Studied</td>
<td>Not Studied</td>
<td>96%</td>
</tr>
<tr>
<td>True positive for M6.0+ earthquake in the region of Sumatra-Andaman</td>
<td>Not Studied</td>
<td>Not Studied</td>
<td>89%</td>
</tr>
<tr>
<td>False positive for M6.0+ earthquake in the region of Chile</td>
<td>Not Studied</td>
<td>Not Studied</td>
<td>16%</td>
</tr>
<tr>
<td>False positive for M6.0+ earthquake in the region of Japan</td>
<td>Not Studied</td>
<td>Not Studied</td>
<td>3%</td>
</tr>
<tr>
<td>False positive for M6.0+ earthquake in the region of Sumatra-Andaman</td>
<td>Not Studied</td>
<td>Not Studied</td>
<td>25%</td>
</tr>
</tbody>
</table>

- These false alarms may be indicators of other earthquakes that took place just outside of our region of study.
- We envision that we could avoid such false alarms if we can have prediction of all over the world all together, or at least cover a much larger region.
- Moreover, these alarms might be a signal of M6.0+, however, the energy could have been released due to the tectonic plates stress with a few EQs less than M6.0. As our study is for M6.0+ EQs, we could not detect such small-magnitude EQs.
Chapter 9

Future Work

In this study, we present a new approach of predicting earthquakes based on satellite data. We point out avenues of potential future work of the study below.

- Up to now, we have applied our proposed mechanism mostly to five different regions of the world. However, since our proposed mechanism possesses a generalized flavour, we plan to divide the whole world into small blocks and apply our mechanism to all of the blocks in future.

- In our study, we have found disappearances of our GGT-based precursors within four months period prior to earthquake happenings in a few cases (for example Fig. 7.2, 7.3, and 7.4). This could have been happened due to different environmental factors, which we have left for further investigation in future.

- We plan to combine our mechanism with existing other methods, for example M8 algorithm [102], VAN mechanism [24], etc., to explore whether the combined mechanisms exhibit more accuracy or not. Besides, in recent times, some researchers have claimed that they found electromagnetic variation near the epicenter during earthquakes [103]. Combining GGT anomaly with the electromagnetic anomaly can be another research avenue we want to explore in future.

- We also plan to use different types of AI mechanisms with our proposed one to explore whether such integration improves the prediction performance.

- We plan to define some policies [104, 105, 106, 107] and frameworks after an alarm being raised by our proposed method. Such policies, after getting an alarm of a looming earthquake within the
next four months, can be used by government organs, NGOs, and general public. For example, the policy could be related to general people awareness about the upcoming disaster, sudden evacuation planning during an earthquake, emergency rescue planning after the earthquake, adoption of methods to absorb minimal loss, implementation of the plannings in root level, evaluation of the feedback from general people about the alarms, and reorientation of the model to the region after the earthquake. Investigating these aspects are left for our future work.
Chapter 10

Conclusion

Various mechanism have been proposed over the past several decades to predict earthquakes. However, still an effective algorithm is yet to come into play. Therefore, in this study, we propose a new mechanism for predicting earthquakes based on open satellite data using gravity gradient tensor components. Unlike others, our mechanism is a generalized mechanism and can be applied to any part of the world.

In this study, first, we thoroughly investigate the existing research studies in literature and point out their limitations. Two common limitations are - 1) Existing mechanisms are yet to present a generalized mechanism of predicting earthquakes in any part of the world, and 2) They are prone to high false alarm rates. These limitations eventually lead us to design a novel mechanism for predicting earthquakes. Our mechanism can be applied to any part of the world and have a good success rate along with a low false alarm rate. We evaluate performance of our proposed mechanism and compare it with that of other recent studies. Comparison results demonstrate significant performance improvement in terms of various performance metrics using our proposed mechanism over state-of-the-art approaches. Specially, the performance improvement in terms of successful prediction (true positive) is quite substantial compared to that of other approaches.

In our proposed mechanism, we use gravity gradient tensor components, which we calculated from the open satellite data. We perform several pre-processing as per our customized requirements on the publicly-available satellite data provided by two different organizations namely UTCSR-Austin and JPL-NASA. The pre-processing can be used in future research studies pertinent to calculating operational parameter of our mechanism for different other parts of the world.
We plan to extend our work using electromagnetic anomaly in future. Combining gravity gradient
tensor anomaly with electromagnetic anomaly can be another research direction in road to enhancing
accuracy of the prediction further. Nonetheless, we want to explore other such combinations in future.
Bibliography


BIBLIOGRAPHY


[73] N. Squires and G. Rayne, “Italian earthquake: expert’s warnings were dismissed as scaremongering,” The Telegraph, Apr 2009.


