SEISMIC WAVEFORM ANALYSIS RELEVANT TO TSUNAMI WARNING

By

AJIT KUNDU
Enrolment No: PHYS01201304018

Bhabha Atomic Research Centre, Mumbai

A thesis submitted to the
Board of Studies in Physical Sciences

In partial fulfillment of requirements
for the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

September, 2018

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by Shri Ajit Kundu entitled "Seismic Waveform Analysis Relevant to Tsunami Warning" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman – Prof. G. Rami Reddy — **Date: 21/09/2018**

Guide / Convener – Prof. S. Basu — **Date: 21/09/2018**

Examiner – Prof. Arun K. Tangirala — **Date: 21/09/2018**

Member 1- Prof. A. K. Bhattacharjee — **Date: 21/09/2018**

Member 2- Prof. S. Mukhopadhyay — **Date: 21/09/2018**

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I hereby certify that I have read this thesis prepared under my direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 21/09/2018

Place: Mumbai

Guide: Prof. S. Basu

ACKNOWLEDGEMENTS

First and foremost I would like to express my deep and sincere gratitude to my guide Prof. S. Basu (Associate Dean, HBNI) for his invaluable guidance, constant encouragement, unstinted inspiration, keen interest and good wishes. I greatly appreciate his contributions, comments and insights during the course of writing this thesis. His patience and support has helped me get through some challenging moments. For this, and much more, I remain indebted.

I sincerely express my grateful thanks to Dr. S. Mukhopadhyay (Head, SD, BARC) for his encouragement and critically going through the manuscript of this thesis. His ability to explain things simply and clearly are something I have admired right from my first visit to him, and continue to remain in awe even today.

I sincerely thank Shri Y. S. Bhadauria for his valuable suggestion and continuous support during the course of the work. I also thank Shri P. Mane for extending support and help at various stages of the thesis work.

I would like to express my sincere thanks to my doctoral committee members, Prof. G Rami Reddy and Prof. A. K. Bhattacharjee for their critical comments and suggestions during annual review presentations.

Last but not the least; I would like to offer my whole-hearted indebtedness to my family members for their moral support and motivation.

DEDICATIONS

Dedicated to

My Parents

Shri Raghunath Kundu

Smt. Gayatri Kundu

My Wife

Smt. Nilanjana Kundu

And

My Daughter

Miss. Ankita Kundu

List of Publications

Journal

1. *Prompt identification of tsunamigenic earthquakes from 3-component seismic data*.

Ajit Kundu, Y.S. Bhadauria, S. Basu, S. Mukhopadhyay
Physics of the Earth and Planetary Interiors, 2016,259, 10-17.

BARC Internal Report

1. *Detecting seismic events using short-time Fourier transform and artificial neural network.*

Ajit Kundu, Pratap Mane, Y.S. Bhadauria, S. Basu, S. Mukhopadhyay
Report No. BARC/2017/I/019

Conferences

1. *Artificial Neural Network Based Estimation of Moment Magnitude with Relevance to Earthquake Early Warning.*

Ajit Kundu, Y.S. Bhadauria, S. Basu, S. Mukhopadhyay
IEEE International Conference on Wireless Communications, Signal Processing and Networking, 22-24 March, 2017.

2. *Application of ANN and SVM for Identification of Tsunamigenic Earthquakes from 3-component Seismic Data.*

Ajit Kundu, Y.S. Bhadauria, S. Basu, S. Mukhopadhyay
IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, 19-20 May, 2017 .

Ajit Kundu

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Ajit Kundu

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Ajit Kundu

CONTENTS

3-component broad-band station contain sufficient information required for categorizing a tsunamigenic earthquakes and hence azimuthal coverage of stations may not be necessary. The direct mapping of rms amplitudes of seismic phases to earthquake category using ANN proves to be extremely useful for the regions that are not adequately instrumented for azimuthal coverage. This chapter ends by bringing out the future scope.

and of 100% of the events with absolute error less than 0.35. Subsequently an ANN based non-linear mapping of rms amplitudes of seismic phases to earthquake category has been developed. This mapping algorithm promptly identifies shallow focus (depth < 70 km) tsunamigenic earthquakes using the seismograms recorded by a single 3-component station located at regional distance. Besides location and magnitude, the root mean square amplitudes of seismic phases (i.e. P, S and LR) recorded by a single 3-component station have been considered as inputs and earthquake category as desired output to ANN. The trained ANN (say ANN–I) has been found to categorize 100% of the new earthquakes successfully as tsunamigenic or non-tsunamigenic. The proposed method has been corroborated by an alternate mapping technique of earthquake category estimation. The second method involves computation of focal parameters (using say ANN-II), estimation of water volume displaced at the source and eventually deciding category of the earthquake. The method has been found to identify 95% of the new earthquakes successfully. The performance of direct mapping of rms amplitudes of seismic phases to earthquake category using ANN-I has been checked with the competitive mapping method namely SVM. The SVM have been found to categorize 91% of the new earthquakes successfully as tsunamigenic and non-tsunamigenic. The fair agreement between the two methodologies again ensures that the direct mapping of rms amplitudes of seismic phases to earthquake category using ANN is a reliable technique. The method would prove to be extremely useful for the regions that are not adequately instrumented for azimuthal coverage.

Finally, Chapter 5 lists out major achievements of the research work emphasizing the role of machine learning techniques in identifying the tsunamigenic earthquakes. It has been observed that seismic records of a single

Chapter 4 presents an application of the proposed method. For detection of seismic event, short period array (GBA) of India has been considered and for estimating moment magnitude as well as identifying the earthquake category, a single 3-component broad band station (PALK) of Sri Lanka has been chosen. The seismic event is detected by analyzing the change in spectral characteristics of noise and event using STFT and ANN. The smoothened amplitude spectra computed using STFT at noise and event window of width 3 sec in the frequency range 0.2 Hz to 20 Hz at an interval of 0.4 Hz have been fed to an ANN which has been trained with past recorded events. The trained ANN has been found to have detection capability of 99% with an average error of ± 0.036 sec in onset pick up. The method has been tested with dataset of 148 events and the results have been compared with the conventional method of detection (STA/LTA) which has been found to detect 84% of the new events with an average error of ±1.73sec in onset pick up. Once the event is detected online, its epicentral location is estimated by using conventional method. In this method, from the coherent signals recorded by GBA the time lags along the two arms are measured. Apparent phase velocity and back-azimuth are then estimated from these time lags using standard relations. From the knowledge of apparent velocity the epicentral distance is computed using a polynomial fit to the theoretical distance - velocity data. Knowing distance and back-azimuth the source location can be computed. For estimating moment magnitude (M_w) the amplitude spectrum computed in the frequency range from 0.0098 Hz to 10 Hz from the vertical component of seismograms recorded by a single 3–component station have been used to train an ANN with M_w as desired output. The trained ANN has been found to estimate M_w of about 73% of the new events with absolute error less than 0.1

Computing the P wave displacement amplitude spectra from the same seismogram, an ANN may be used to map these spectral amplitudes with the moment magnitude of the event. Confirming the magnitude of the event greater than 6.0 its category (i.e., tsunamgenic/non-tsunamigenic (T or NT)) could be estimated. Since the amplitudes of the seismic waves recorded at a 3-component station at regional distance from a given geographical area differ from tsunamigenic (slow rupture earthquake) to non-tsunamigenic earthquakes originating from that area even if other conditions remain same the earthquake could be classified as tsunamigenic/non-tsunamigenic based on this change in amplitudes. For this task, an ANN could be trained with rms phase amplitudes along with location and magnitude as inputs and the as desired output for a large number of both tsunamigenic and non-tsunmaigenic earthquakes. Subsequently the trained ANN could be used to identify new tsunamigenic earthquake originating from the same geographical region based on the recordings from the given station. Again for a given source-station path, the seismic phase amplitudes are function of source strength (magnitude) and focal parameters. Therefore, another may be used for inverse mapping of this function to obtain the focal parameters. Knowing magnitude and focal parameters the volume of water displaced at the source could be estimated. Comparing the water volumes with the given earthquake category a tsunami-volume scale may be established which could be used to find the category of new earthquake. This method of estimation of earthquake category will theoretically justify the direct mapping method of estimating earthquake category. This mapping could also be verified using the other non-linear mapping technique such as SVM. The proposed method has been developed for poorly instrumented region of the world where multi-station data may not be available.

computed using spectral technique such as STFT (Short-Time Fourier Transform). The window length could be selected such that it accommodates the onset times at two farthest channels along each of the arms of L shaped array corresponding to maximum lag. To map the spectral amplitudes with event or noise category, a machine learning algorithm such as ANN (Artificial Neural Network) may be used. For this, an ANN has to be trained with spectral amplitudes as inputs and event/noise (i.e. binary decision) as desired output for large number of events. After training, the trained ANN could be used to find the onset of new event online. For this, smoothened spectral amplitudes will be computed in a predefined time window for all the channels of the seismic array. These spectral amplitudes would then be fed to the trained ANN to obtain decisions corresponding to all the channels. If certain number say, K out of all the decisions show event for the given window it would be considered as a seismic event, otherwise it would be classified as noise. Once an event is declared, the starting time of the window, called the preliminary onset is noted and the procedure for fine detection will be started over a data block starting from preliminary onset minus one second to preliminary onset plus 3 sec. Within this data block corresponding to each channel, amplitude spectra will be computed for the same time window which will slide in steps of one sample to find the sample number at which the ANN shows decision as event. If the decisions are found to match K number of channels, the occurrence of an event will be declared. From the time lags of the seismic event recorded by various sensors of a seismic array the epicenter of the event could be estimated through the computation of apparent velocity and back azimuth (conventional method). Once the epicenter is known the idea about shallow focus depth of event could be obtained from the presence of LR phase on the seismogram recorded by a 3-component broad band station.

these limitations by the proposed method, which depends not only on magnitude but also source mechanism, has also been addressed in this chapter. This chapter ends by drawing an outline of the thesis.

Chapter 2 provides detailed literature review. The cause and physical characteristics of tsunami have been discussed. Past records indicate that 75% of tsunamis are generated by shallow (depth < 70 km) undersea earthquakes. The tsunami generation by these earthquakes and tsunami propagation as well as inundation have been described. The methods involved in tsunami warning such as event detection, location, magnitude estimation, use of tsunami models and sea level gauges have been reviewed along with their limitations. The chapter ends by reviewing spectral analysis of seismic phases relevant to tsunami warning.

Chapter 3 describes the proposed method of identifying tsunamigenic earthquakes using broad band data of a single 3-componenet station located at a regional distance from the source. This method consists of four steps, namely, detection, location, magnitude estimation and identification of tsunamigenic earthquakes. Since detection and location of seismic events using seismic arrays are generally superior to that using single 3-component stations the proposed method introduces seismic array (preferably L shaped and short period) based detection and location of seismic events. The methods of magnitude estimation and tsunamigenic earthquake identification are based on single 3-component broad band records. For detection a circular geographical region with center at seismic array and radius much more than regional distance may be considered. The data recorded on the array could be collected corresponding to a large number of seismic events which occurred in this region with varying magnitudes. Identifying the event onsets in a good channel, the smoothened amplitude spectra at event and noise windows could be at regional distance from the recording station the epicenter, depth and magnitude could be estimated within 8 minutes of occurrence. This much delay is same as that taken by present tsunami warning systems. Confirming $M_w > 6$ and occurrence of a shallow focus (depth < 70 km) event from the presence of LR wave in the seismogram, root mean square (rms) amplitudes of seismic phases P, S and LR may be computed. These amplitudes together with location and magnitude can be used as inputs to train an ANN with earthquake category (i.e., Tsunamigenic (T)/ Non-tsunamigenic (NT)) as desired output. The trained ANN thereafter can be used to predict the tsunami potential of an earthquake [4]. This whole prediction procedure could be completed within 15 minutes of occurrence of an earthquake. This prediction method has been validated via 1) computation of water volume by estimating focal parameters using another ANN and 2) mapping between rms amplitudes of seismic phases (along with location and magnitude) and earthquake category using Support Vector Machine (SVM) [5]. The single station based proposed tsunamigenic earthquake identification technique eliminates the requirement of the multi-station data and complements the pre-computed database approach or tsunami forecast model based approach. The proposition is validated with seismic data recorded at short period array GBA (India) and 3-component broad band station PALK (Sri Lanka) for the earthquakes originated from the tsunami prone region, namely, Sumatra and is expected to be both faster and economic.

4. Organization of the thesis

The thesis comprises five chapters.

Chapter 1 introduces the research work. It states the objective and motivation of the work undertaken under the PhD program. The limitations of existing tsunami warning systems have been discussed. The elimination of

3. Brief description of the research work

To circumvent the above difficulties faced by the existing tsunami warning systems a novel method of tsunami identification has been proposed using seismic data of a single 3-component broad band station. This will involve four tasks, namely, detection, location, magnitude estimation and identification of tsunamigenic earthquakes in near real time. Since the accuracy of detection and location of an earthquake obtained by a single 3-component station is not very good, the first stage i.e., the detection and location of seismic event is carried out using the short period data recorded by a seismic array. The technique involved in detection is that the features extracted in the time-frequency domain using a suitable technique such as Short-Time Fourier Transform (STFT) for both signal and noise windows are compared with the pre-computed features using a popular pattern matching tool, Artificial Neural Network (ANN) to obtain onset of the event at recording station [2]. From the time lags of the seismic event recorded by various sensors of a seismic array, the epicenter of the event could be estimated through the computation of apparent velocity and back azimuth (conventional method). However, estimation of depth and true size (i.e., moment magnitude, M_w) of an earthquake is quite difficult or not possible using short period data. Therefore, the depth information (i.e. shallow focus depth) of an earthquake could be inferred from presence of LR wave (surface wave) in the seismogram recorded by a single 3-component station. The true size could be estimated shortly after the arrival of P wave in the same seismogram by computing the displacement spectral amplitude at P onset in a suitable window. An ANN may then be used to map the computed displacement spectra and moment magnitude [3]. For an earthquake occurring analysis will be initiated. The more accurate source parameters from this analysis will be available about 25 minutes after the origin time. Third step is to drive a run of tsunami forecast model based on these source parameters. The forecast coastal amplitudes become available about 35 minutes after the origin time. Lastly, the predicted costal amplitudes will be cross verified with the actual data available from the sea level gauges and the tsunami warning will be upgraded or downgraded accordingly. The whole procedure requires the involvement of multi-station data which may not be accessible by the poorly instrumented regions of the world. The computation delay of about 35 minutes may be prohibitively long for issuing alert at nearby coastal regions. Moreover, coastal gauges and deep-ocean gauges which are used to monitor sea level data to measure tsunami amplitudes have their own limitations. For example, tide gauges which are generally installed near the coast and are used as confirmation of tsunami could detect the tsunami only after the wave reaches near the coast and no time is left for warning. Alteration of local seafloor bathymetry and harbor shapes could also affect the tide gauge performance. The open sea buoys which are being used to detect the passage of tsunamis through deep seas are too expensive and difficult to maintain. Furthermore, a tsunami wave in deep-ocean has very small wave amplitude and a long wavelength making its detection difficult.

Because of these limitations, most of the initial tsunami warnings tend to generate false alarms that cause panic and questions about warning methods. Hence, alternative means of identifying tsunamigenic earthquake is necessary and desirable.

determination of actual fault rupture characteristics is a time consuming process and often requires data from multiple stations distributed azimuthally around the source. Availability of such data within a short span of time after the earthquake is often difficult. To circumvent this difficulty, the present thesis aims to utilize the seismic data from a single 3-component station and develop a robust method of tsunami warning which is faster, more accurate (free from false alarms) and reliable. The central idea behind this method is that seismic signal is essentially the end result of the earthquake magnitude, fault rupture characteristics at the source, the propagation path effects and instrument response at the recording site. Thus, even if the source-station geometry remains fixed, the seismic phase amplitudes in the seismograms of two nearby earthquakes of comparable magnitude may differ due to the source rupture characteristics. The facts that the seismic wave travels faster than tsunami wave and the tsunamigenic potential of an earthquake could be determined from seismic data with reasonable accuracy motivates the present study to design a robust rapid identifier of tsunamigenic earthquake which will reduce the number of false tsunami alarms so that an effective and accurate early tsunami warning could be issued to reduce the hazards.

2. Identification of the research problem

Currently, there are a few world-wide tsunami warning systems like PTWC and couple of regional warning centers like JMA, IOTWS etc. A typical tsunami warning procedure (as followed by PTWC) is as follows. First is the computation of epicenter, depth and magnitude of the earthquake with the seismic data available within 8 minutes of occurrence of earthquake. Secondly, if these parameters indicate the presence of a shallow focus (depth< 100 km) undersea earthquake with magnitude greater than 6.5, a preliminary threat message will be issued and W-phase Centroid Moment Tensor (WCMT)

SYNOPSIS

1. Introduction and Motivation

Tsunami is one of the most destructive forces in nature and it can cause colossal loss of life and damage to property. Majority of tsunamis are generated by earthquake events which occur under the sea. Prompt identification of the earthquake that may generate tsunami, using a minimum of one 3-component station, is the main focus of this thesis. This is important because only a few earthquakes, located under the sea, produce severe tsunamis in a land mass located at a regional distance (~2000 km) from the source. For example the December 2004 Indian Ocean earthquake caused large scale devastation at Indian coast while the April 2012 Indian Ocean earthquake generated only minor tsunami waves. Thus, the knowledge, whether an earthquake will generate significant tsunami is critical for issuing an early tsunami alert. The mechanism that builds up the tsunami is very complex and most of the current methods based on magnitude and epicenter location alone are prone to large false alarm rate. This is evident from the fact that more than 50% warnings issued by Pacific Tsunami Warning Center (PTWC) are later found to be false [1].

The main reason for large number of false alarms in the existing methods is that they are essentially based on epicentral location and magnitude alone. It has been perceived that the tsunami generation not only depends on magnitude and epicenter location but also affected by other factors such as fault rupture characteristics (i.e. the amount of slip on the fault plane and focal mechanism), focal depth, water depth at the source, etc. However the

List of Tables

List of Figures

Motivation for choice of various techniques and tools, sensitivity of results to the choice of analyzing parameters have also been discussed.

Chapter 4 describes the application of the proposed method to a specific region. The results obtained by various techniques involved in the method are presented considering Sumatra region as tsunami source and GBA and PALK as recording stations. The fair accuracy of the method confirms that a prompt alert system could be developed based on proposed method.

Important conclusions and findings drawn from this research study are summarized in Chapter 5 along with the possible direction for future work.

extremely important for the possible devastation in coastal regions of India. The detection and location using seismic array such as GBA (Gauribidanur Array, India) proved to be very accurate. The use of seismic records of a single 3-component station, namely, PALK (Pallekele, Sri Lanka) makes the proposed method an effective mean of identifying tsunamigenic earthquakes. It may be noted here that the proposed method for identification of tsuanmigenic earthquakes which seems to be very effective for the chosen source-station pair in the present work, namely, GBA/PALK-Sumatra, would be equally applicable for other regional source-station pairs in the world with ANNs trained specifically for those pairs. It may also be mentioned here that in case the training is done with the recordings of more than one station located in close vicinity, the classification results will hold and should further improve robustness of the prediction.

1.5 Outline of the thesis

This thesis is on the identification of tsunamigenic earthquakes from the analysis of the seismic signals of the earthquakes. The thesis comprises five chapters.

Chapter 1 introduces the objective and motivation of thesis work. The existing tsunami warning systems and their limitations have been discussed. The limitations overcome by the proposed method have also been discussed. An outline of the thesis work is given at the end of the chapter.

Chapter 2 provides details on literature review. The cause and physical characteristics of tsunami wave have been discussed. The concept of tsunami generation, propagation and inundation has been stated. The chapter ends with a detailed discussions on existing tsunami warning methods.

Chapter 3 presents the proposed method of tsunami warning in detail.

ANN with earthquake category (i.e., Tsunamigenic (T)/ Non-tsunamigenic (NT)) as desired output. The trained ANN thereafter can be used to predict the tsunami potential of an earthquake [4]. This whole prediction procedure could be completed within 15 minutes of occurrence of an earthquake. This prediction method has been validated via 1) computation of water volume by estimating focal parameters using another ANN and 2) mapping between rms amplitudes of seismic phases (along with location and magnitude) and earthquake category using Support Vector Machine (SVM) [5]. The single station based proposed tsunami identification technique eliminates the requirement of the multi-station data and complements the pre-computed database approach or tsunami forecast model based approach. The proposition is expected to be both faster and economic.

At a station, located at regional distance from the seismic source, a number of seismic phases are recorded. The onsets and amplitudes of these phases will be different. In the chosen problem domain, three phases, namely, P, S and LR have clear onsets. Out of these, P is the fastest phase whereas LR is the slowest one. As far as severity is concerned LR is the most damaging phase because of its largest amplitude. Therefore, LR is the phase which carries more information about tsunami. As far as rapid estimation of tsunamigenic potential is concerned, P phase appears as a first choice due its higher speed of travel through earth and earlier arrival at the detector compared to other phases. However, the present method utilizes all three waves (i.e., P, S and LR) for efficient mapping between seismic phase amplitudes and earthquake category.

The proposed method has been applied to one of the known sources of worst tsunamis, namely, Sumatra region. The tsunamis originating in this region are location of an earthquake obtained by a single 3-component station is not very good, the first stage i.e., the detection and location of seismic event has been carried out using the short period data recorded by a seismic array. The technique involved in detection are the features extracted in the time-frequency domain using a suitable technique such as Short-Time Fourier Transform (STFT) for both signal and noise windows. These are compared with the pre-computed features using a popular pattern matching tool, Artificial Neural Network (ANN), to obtain onset of the event at recording station [2]. From the time lags of the seismic event recorded by various sensors of a seismic array the epicenter of the event could be estimated through computation of apparent velocity and back azimuth (conventional method). However, estimation of depth and true size (i.e., moment magnitude, M_w) of an earthquake is quite difficult or often is not possible using short period data. Therefore, the depth information (i.e. shallow focus depth) of an earthquake could be inferred from presence of LR wave (surface wave) in the seismogram recorded by a single 3-component station. The true size could be estimated shortly after the arrival of P wave in the same seismogram by computing the displacement spectral amplitude at P onset in a suitable window. An ANN may then be used to map the computed displacement spectra and moment magnitude [3]. For an earthquake occurring at regional distance from the recording station, the epicenter, depth and magnitude could be estimated within 8 minutes of occurrence. This delay is nearly the same as that taken by present tsunami warning systems. Confirming $M_w > 6$ and occurrence of a shallow focus (depth < 70 km) event from the presence of LR wave in the seismogram, root mean square (rms) amplitudes of seismic phases P, S and LR may be computed. These amplitudes together with location and magnitude can be used as inputs to train an number of stations are required it inherently involves computational delay. Besides, there is also delay involved in the computation of the forecast of coastal amplitudes by running a tsunami forecast model using the accurate parameters obtained from moment tensor inversion and comparing those amplitudes with the actual sea level data monitored by the sea level gauges.

1.4Proposed method of identification of tsunamigenic earthquakes

To overcome the limitations of the existing tsunami warning methods a novel method of tsunami warning is proposed in this thesis. As far as seismic means of evaluation of tsunami potential of an earthquake is concerned, there are four challenges, viz., reliable detection, accurate location estimation, magnitude estimation and identification, in case the earthquake is tsunamigenic. Proper onset detection is an important task because the accuracy of location estimation heavily depends on accurate estimation of P phase arrival. This estimation of the seismic source location, in turn, indicates the coastal regions to be affected by the impending tsunami. Being a crucial parameter, the moment magnitude also demands accuracy in its estimation for evaluation of tsunami potential of a shallow focus, under-sea earthquake. The real challenge is to categorize an earthquake to be tsunamigenic or not, based on data recorded by a single station or more than one station located in a narrow band of azimuthal angle. It takes some time to lay hands on the data records from all azimuthal angles. Further, in a poorly instrumented region, it is desirable to categorize a regional earthquake promptly, based on a single station data as the arrival time of tsunami may be too short to wait for the full azimuthal data.

In the proposed method tsunamigenic earthquakes are identified using seismic data of a single 3-component broad band station. Since the accuracy of detection and

From Fig. 1.1 it is seen that the preliminary tsunami threat message which is generated typically 8 minutes after the occurrence of earthquake can be verified either by the measurement of sea level data or performing W-phase Centroid Moment Tensor (WCMT) analysis and subsequently by driving a run of tsunami forecast model. However, these verifications take about 35 minutes from the origin of earthquake and may be prohibitively long for issuing alert at nearby coastal regions. Even though the more accurate estimation of earthquake's location, depth, magnitude and focal parameters using WCMT analysis complements the preliminary estimation of tsunami threats, it requires the involvement of multiple stations which may not be accessible by the poorly instrumented regions of the world.

Moreover, coastal gauges and deep-ocean gauges which are used to monitor sea level data to measure tsunami amplitudes have their own limitations. For example, tide gauges which are generally installed near the coast and are used as confirmation of tsunami could detect the tsunami only after the wave reaches near the coast and no time is left for warning. Alteration of local seafloor bathymetry and harbor shapes could also affect the tide gauge performance. The open sea buoys which are being used to detect the passage of tsunamis through deep seas are too expensive and difficult to maintain. Furthermore, a tsunami wave in deep-ocean has very small wave amplitude and a long wavelength making its detection difficult.

The large computation delay involved in the existing tsunami warning method is mainly due to the computation of accurate source parameters (e.g., location, magnitude, focal and parameters) using moment tensor inversion technique. It basically generates synthetic seismogram and compares it with the observed one for a large number of stations using least-squares inversion method. Since data from a large accurate early tsunami warning could be issued to reduce the hazards. The motivation behind using single 3-component seismic data is to develop a fairly accurate tsunami warning method which uses minimum resources. This will assist the poorly instrumented regions of the world to assess tsunami potential.

1.3Limitations of the existing tsunami warning systems

Currently, there are a few world-wide tsunami warning systems like PTWC and couple of regional warning centers like JMA, IOTWS etc. A typical tsunami warning procedure (with timeline) as followed by PTWC is shown in Fig.1.1.

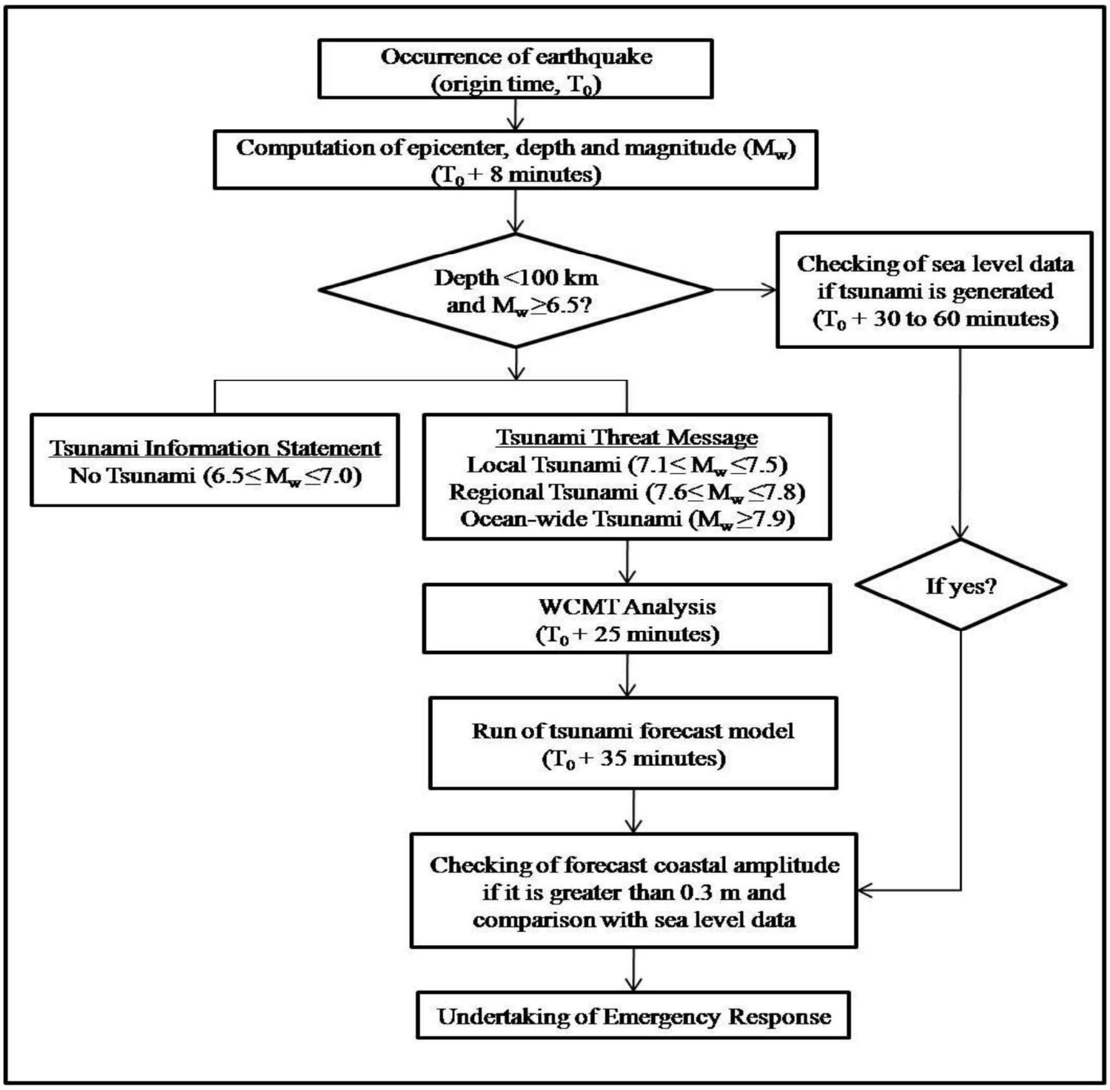


Fig. 1.1: A typical tsunami warning procedure with timeline.

(M_w 9.0~9.3) Indian Ocean earthquake caused one of the deadliest tsunamis in the recorded history. The mechanism that builds up the tsunami is very complex and most of the current methods based on magnitude and epicenter location alone are prone to large false alarm rate. This is evident from the fact that more than 50% warnings issued by Pacific Tsunami Warning Center (PTWC) are later found to be false [1].

It has been perceived that the tsunami generation not only depends on magnitude and epicenter location but also affected by other factors such as fault rupture characteristics (i.e. the amount of slip on the fault plane and focal mechanism), focal depth, water depth at the source, etc. However the determination of actual fault rupture characteristics is a time consuming process and often requires data from multiple stations distributed azimuthally around the source. Availability of such data within a short span of time after the earthquake is often difficult. To circumvent this difficulty, the present thesis aims to utilize the seismic data from a single 3-component station and develop a robust method of tsunami warning which is faster, more accurate (free from false alarms) and reliable. The central idea behind this method is that seismic signal is essentially the end result of the earthquake magnitude, fault rupture characteristics at the source, the propagation path effects and instrument response at the recording site. Thus, even if the source-station geometry remains fixed, the seismic phase amplitudes in the seismograms of two nearby earthquakes of comparable magnitude may differ due to the source rupture characteristics. The facts that the seismic wave travels faster than tsunami wave and the tsunamigenic potential of an earthquake could be determined from seismic data with reasonable accuracy motivates the present study to design a robust rapid identifier of tsunamigenic earthquake which will reduce the number of false tsunami alarms so that an effective,

Chapter-1

Introduction

1.1Objective

Tsunami is one of the most destructive events in nature triggered primarily by undersea earthquakes. It has caused catastrophic devastations all over the globe, several times in the history of mankind. Since tsunami occurs suddenly, without any prior warning, lives of the coastal communities are put to extreme danger. The main objective of this thesis is to present a purely seismological method by which a tsunamigenic earthquake can be identified promptly using seismic data of at least one 3-component station located at a regional distance (~2000 km) as well as an effective and accurate warning could be issued early enough to reduce the hazards. This method would prove to be extremely useful for the regions that are not adequately instrumented for azimuthal coverage.

1.2 Motivation

About 75% of the tsunamis are caused by subduction zone earthquakes in the ocean (United States Geological Survey (USGS)). Not all undersea earthquakes, even with sufficiently large magnitudes cause tsunamis. Located within a close proximity and with comparable magnitudes, some earthquakes produce very severe tsunamis than others in a regional land mass. For example, the April 2012 (M_w 8.6) Indian Ocean earthquake did not generate any significant tsunami, while the December 2004 frequency energy compared to non-tsunamigenic earthquakes using wavelet transform of all phases in the seismogram. Chamoli et al. [61], using the wavelet transform of first few minutes of the seismogram, have established that tsunamigenic earthquakes do not show any significant amplitude for frequencies higher than 0.33 Hz. Ewing et. al. [62] first proposed T waves as an analyzing tool for predicting an earthquake as tsunamigenic or non-tsunamigenic. Hiyoshi et al. [63] showed that tsunamigenic earthquakes have larger T-phase spectral strength and M_0 than non-tsunamigenic earthquake. However, Okal et al. [64] showed that tsunamigenic earthquakes have smaller M_0 and T-phase spectral strength compared to the non-tsunamigenic earthquakes. They attributed this deficiency in T-waves in tsunamigenic earthquakes as a direct result of the slow rupture velocities. Therefore, the research on the T-waves with regard to prediction of an earthquake as tsunamigenic is not conclusive and even contradictory. Furthermore, the data for T-waves is limited by the availability of hydrophones.

To circumvent the difficulties faced by the existing tsunami warning system, the present thesis aims to develop an alternative method to reach a decision in the quickest possible time for a more accurate, fast and reliable tsunami warning that has been described in the next chapter.

2.3.5 Review on sea level gauges

To verify the existence of a tsunami and to calibrate models the continuous sea level data from coastal tide gauges and where available, data from Deep Ocean Assessment and Reporting of Tsunamis (DART™) buoys are monitored. Four types of tide gauges namely, 1) stilling well and float, 2) pressure systems, 3) acoustic systems and 4) radar systems are used to measure sea level variations. Tide gauges that initially detect tsunami waves provide little advance warning at the actual location of the gauge. Open ocean buoy systems equipped with bottom pressure sensors are now a reliable technology that can provide advance warning to coastal areas that will be first impacted by an ocean-wide tsunami, before the waves reach them and nearby tide gauges. Open ocean buoys often provide a better forecast of the tsunami strength than tide gauges at distant locations. However, open sea buoys are too expensive and difficult to maintain.

2.3.6. Review on spectral analysis of seismic phases

Researchers have also tried to estimate the tsunamigenic potential of an earthquake analyzing the spectral contents of either a particular seismic phase or all phases in the seismogram. For example, W-phase, which is interpreted as superposition of overtones of long-period Rayleigh waves or superposition of multiple-reflected phases like PP, PPP, etc. is used as an indicator of tsunami potential. Lockwood and Kanamori [59] reported that tsunamigenic earthquake generates a W-phase of significantly greater amplitude compared to the non-tsunamigenic earthquake. However, the W-phase is registered only at very far stations and thus put a constrain for early warning. Chew and Kuenza [60] have shown that the tsunamigenic earthquakes are depleted in high-frequency energy but rich in low

2.3.4 Review on tsunami models

It is well known that the tsunami generation depends not only on magnitude and location but also on other factors such as the amount of slip on the fault plane and focal mechanism, etc. To complement the tsunami warning method based mainly on location and magnitude, pre-computed database approach has evolved. For example, Indian Tsunami Early Warning System (ITEWS) operated by Indian National Center for Ocean Information Services (INCOIS) uses a numerical modeling technique based on finite difference code of TUNAMI-N2 [19] to determine the potential run-ups and inundation for local or distant tsunami. The TUNAMI-N2 model takes the seismic deformation as input to predict the run-up heights and inundation levels at coastal region for a given tsunamigenic earthquake. At the time of earthquake, only location and magnitude are available immediately. These are subsequently used to search the pre-computed data base developed on the basis of worst case scenario for given regions to issue a first alert of an impending tsunami. However, pre-computed database models are confined to specific ocean basins and cannot be used to forecast ocean-wide tsunamis, such as the 2004 Indian tsunami [58]. Therefore it would not be possible with database approach to forecast tsunamis in real-time for earthquake in any location with any focal mechanism using rapidly derived earthquake parameters. To complement the pre-computed database approach PTWC has developed Real-time Inundation Forecast of Tsunamis (RIFT) model which include 40 pre-defined ocean basins and major marginal seas [58]. This model can account for both local and global domains and can handle multiple events in different or in the same ocean basins. Nevertheless this forecast which is typically available in about 35 minutes after the occurrence of earthquake is sensitive to the model inputs such as earthquake magnitude and focal mechanism.

relationship:

$$\log_{10} M_0 = M_m + 13 \tag{2.7}$$

where M_0 is in Nm. The M_m (a variable period magnitude) is potentially available within about 20-50 min after origin time of the event at 30-90° from the source. M_m can be used only for distant tsunami warning.

P -wave Duration-amplitude magnitude (M_{wpd})

Lomax et al. [52] developed a duration-amplitude procedure to obtain earthquake moment magnitude, M_{wpd} from P wave recordings. This procedure determines apparent source durations (i.e. rupture duration), T_0 from high frequency P wave records and then estimates moments through integration of broad-band displacement waveforms over the interval t_p to $t_p + T_0$, where t_p is the P-arrival time. M_{wpd} which is obtained from moment through scaling relation can be estimated within 20 min or less after the earthquake origin time. This magnitude does not face the problem of saturation. Nevertheless the tsunamigenic potential obtained from M_{wpd} is not confirmatory.

Presently the accurate estimation of location and magnitude are carried out through moment tensor analysis using multi-station data. However, the computational delay which is of the order of 25 minutes from the occurrence of earthquake may put constrain for early warning.

$$M_o = C \times Max(|p_1|, |p_1 - p_2|) \qquad (2.4)$$

where p_1 and p_2 are the first peak and the second peak values in the integrated displacement seismogram. Then P-wave moment magnitude, M_{wp}, is computed at each station with no correction for the radiation pattern using the standard moment formula:

$$M_{wp} = \frac{2}{3}\log M_0 - 6.07 \qquad (2.5)$$

Where M_0 is in Nm. M_{wp} is calculated at three or more stations and an average value is obtained. Finally corrected M_{wp} is obtained by adding 0.2 to the averaged M_{wp} where 0.2 accounts for the double couple radiation pattern, $\gamma^{p}_{\phi\phi_s}$. However, M_{wp} underestimates M_w for very large earthquakes and saturates at about M_w 8.3. This indicates that estimation of tsunamigenic potential using M_{wp} is not satisfactory for large earthquakes.

Mantle wave magnitude (M_m)

The mantle magnitude (M_m) is computed from the mantle wave, which is a very long-period surface wave (Rayleigh) with corresponding wavelength of several hundred to about 1000 kilometers. Mantle waves are generated by the large earthquakes. The M_m [55-57] is computed from Fourier spectral amplitude $X(\omega)$ of mantle Rayleigh waves at variable long periods (>50 s) from teleseismic stations.

$$M_m = \log_{10}(X(\omega)) + C_d + C_s - 3.9 \qquad (2.6)$$

Where C_d is distance correction and C_s is source correction (both these constants are period dependent). Then the seismic moment (M_0) is calculated using a very simple from a network. Arrays provide estimates of the station-to-event azimuth and the apparent velocity of seismic signals (generally using frequency-wavenumber (f-k) analysis). These estimates are used to locate event. However, array processing techniques require high signal coherency across the array.

2.3.3.Review on magnitude estimation

To improve the accuracy of tsunami forecasts, the magnitude estimation methods based on seismic information such as the P-wave moment magnitude, M_{wp} [48-49], mantle magnitude, M_m [50-51], P-wave duration-amplitude magnitude M_{wpd} [52] are used. These are reviewed here.

Moment magnitude (M_{wp})

Tsuboi et al. [48, 53] developed M_{wp} method to determine magnitude rapidly from teleseismic P-waves (in the period range 10 to 60 s). This method was developed for shallow earthquakes (<70 km) for which the seismic waves experience less attenuation. It estimates P-wave seismic moment, M_o using a broadband vertical displacement (BHZ channel) seismogram, as given below:

$$M_o = C\left|Max(\int_{t_0}^{t_0+\tau_r} u_z(x,t)dt)\right|; C = \frac{4\pi\rho\alpha^3 r}{2\gamma_{\phi\phi_s}^p} \tag{2.3}$$

where t_0 is the P wave arrival time, τ_r is fault rupture duration and $u_z(x,t)$ is the attenuation-corrected vertical displacement. C is a constant that depends on density (ρ) of rocks at the source, P-wave velocity (α), double-couple radiation pattern for P-wave ($\gamma_{\phi\phi_s}^p$) and epicentral distance (r). The radiation pattern depends on the take-off angle (ϕ) and fault strike angle (ϕ_s). $\gamma_{\phi\phi_s}^p$ is 0.63 [54]. By integrating, the above equation reduces to:

2.3.2. Review on event location

It is possible to get an approximate location of an earthquake manually. If the seismic network is dense and the earthquake is originated inside the network, the station with the earliest arrival can be approximated as the epicenter. This is called earliest station method, and it gives an estimate with a probable error of the same order as the station spacing. A pretty good approximation of the epicenter can be made by arc method or circle method with three or more S-P times.

It is, however, possible to use a single seismic station to obtain a crude estimate of earthquake location [45]. Single-station method requires three-component recordings of ground motion. Since P waves are vertically and radially polarized, the vector P-wave motion and the amplitude ratio on two horizontal components (AE/AN) can be used to infer the back azimuth (against North) from station to the source. If the vertical motion of the P-wave is upward, its radial component is directed away from the epicenter. If the vertical component of the P-wave is downward, the radial component is directed toward the epicenter.

When several stations are available, the earthquake location problem is resolved mathematically by least squares method which gives a precise hypocenter location of the earthquake. Most of the location algorithms in common use are based on Geiger's method [46]. Genetic algorithms which work on minimizing some misfit criteria of the data also find application in estimating hypocenter of earthquake [47].

Event location using an array is superior to those using single 3-component stations. A seismic array differs from a local network of seismic stations mainly by the techniques used for data analysis. In principle, a network of seismic stations can be used as an array and on the other hand data from an array can be analyzed as data irrelevant ones (noise) decay quickly at larger scale indices. As a further development many other methods such as hybrid techniques [38-39], the cross correlation [39] and the higher-order statistics [40-43]) have also been emerged as onset pickers. The hybrid techniques combine the results of three methods namely 1) energy analysis, 2) instantaneous frequency, and 3) autoregressive representation of the trace to pick onset time of arrivals that have different dependences on noise and signal content. The cross correlation works on the quantification of the existing similarity between a reference event signals, considered as a template, and the successive incoming data in a time window. The algorithm based on higher-order statistics, involves computation of the characteristic functions (CF) which are derived by kurtosis over several frequency bandwidths, window sizes, and smoothing parameters. The algorithm determines the onset type (P or S) using polarization parameters, removes bad picks using a clustering procedure and assigns a pick quality index based on the signal to noise ratio (SNR). However all these methods share a common disadvantage that they require one or more of the following information (a) detection interval, (b) threshold settings, (c) an approximate knowledge of the move out character for an event, and (d) tuning of picking parameters specific to intensity and/or frequency content of the signal a priori [44]. Despite the vast amount of research in this field, the signal processing algorithms for the detection of seismic events have not yet fully come of age.

[20]. Typically these warning systems follow a common procedure - detection, location, magnitude estimation, computation of tsunami amplitudes and travel times using either pre-computed database approach or real-time tsunami forecast model and finally comparison with the actual data of sea level gauges. A brief review on these stages is stated below.

2.3.1. Review on event detection

Accurate identification of onset of an event in seismic data is extremely important so far as source parameter estimation is concerned. An onset picking task has two steps: detecting a phase and reading its arrival time. A number of detecting methods using either the recordings of single 3-component stations or multi-station or seismic array have been proposed in the past. These methods are broadly classified into four categories: time domain, frequency domain, particle motion processing, and pattern matching [21]. The most widely used method of onset picking is the time-domain short-term average to long-term average (STA/LTA) algorithm and its variants [22-28]. The method uses the ratio STA/LTA to detect phases. This technique is well suited for detecting amplitude changes. However, its accuracy heavily depends on the detection interval and threshold settings. It is prone to pick up pulse-like noises, especially those contained in micro tremors, unrelated to the earthquake signal [29]. In recent years, Artificial Neural Network (ANN) [30-34] has found potential application in the design of onset picking tools. ANN works on positive correlation with patterns of some known earthquakes and noise signals. Furthermore Wavelet Transform (WT) [35-37] has also been used for picking onset. WT decomposes the signal at different scales and adaptively characterizes its components at different resolutions. The primary features in the signal (phase arrivals) are retained over several scales and viz. seismic moment, M_0 and tsunami amplitude decay, Q which accounts for geometrical spreading and frequency dispersion at coastline. Actually M_0 determines the initial tsunami wave height profile at the earthquake source and Q determines the tsunami strength at the distant coastal areas. The Q factor depends on seafloor bathymetry and presence or absence of barriers (like islands) that could reflect and refract the water waves. Due to geometrical spreading tsunami wave amplitude drop with distance (r) approximately as $r^{-0.5}$ whereas due to frequency dispersion near the shoreline tsunami wave amplitude drop as $r^{-\chi}$, where the decay factor χ is in the range 1/8 to 1/2 depending on the frequency content of the tsunami wave [7]. Larger magnitude earthquakes produce greater tsunami at the same epicenter distance and generate longer period waves that are less affected by dispersion, so waves from them decay more slowly with distance.

In the context of tsunami early warning systems, tsunami simulations can be used to provide information about expected arrival time, maximum wave height and inundation at the risked coastlines. In the next section, tsunami warning systems are reviewed.

2.3 Review on existing tsunami warning systems

A tsunami warning system has existed in the Pacific Ocean since the late 1940s. Current tsunami warning systems can be grouped into a Pacific-wide system and regional (or local) systems. The Pacific Tsunami Warning Center (PTWC), located in Hawaii, monitors seismic and tsunami waves and issues tsunami warnings. The French Polynesia Tsunami Warning Center developed and adopted TREMORS (Tsunami Risk Evaluation through seismic Moment in a Real-time System) in 1987

But the most popular model for tsunami generation modeling is the Okada [14] model for finite rectangular source.

However all these models are based on the assumption that seafloor displacement is smooth and the rupture is instantaneous. But in reality the actual seafloor displacement is complicated and for earthquake of large magnitude instantaneous rupture assumption may fail. Although several complex numerical models have evolved but they are not yet disseminated in tsunami research circle [15].

2.2.2 Tsunami propagation

Tsunami propagation is the second phase of tsunami evolution. From the generation point the tsunami wave propagates in both the directions. As a result, one side of the fault experiences receding while other side observe sudden rise in water. There are 3 prominent computational tsunami propagation models:

(1) The Method Of Splitting Tsunami (MOST): It was developed originally by researchers at the University of Southern California [16], (2) The Cornell Multi-grid Coupled Tsunami Model (COMCOT), which was developed by researchers at Cornell University [17], and (3) The TUNAMI which was developed at Tohoko University in Japan [18-19]. All these models solve the same Nonlinear Shallow Water Equations (NSWE) with different finite-difference algorithms. They can simulate tsunami propagation over a long distance with a fairly high accuracy, provided that the initial wave profile and the sea bottom bathymetry data are accurate.

2.2.3 Tsunami inundation

The inundation of land is the third phase of tsunami evolution. It is difficult to predict the largest tsunami wave height expected at a distance r from an earthquake of certain magnitude M_w. It can be addressed with the help of two important factors area, respectively. It is known that larger the earthquake's seismic moment, the larger is the tsunami, if all other conditions remain same.

Focal mechanism specifies the orientation of the earthquake fault and the direction of slip on the fault plane with faults idealized as rectangular planes. Three angles the strike (ϕ_s), the dip (δ) and slip (λ) determine the type of faulting and the direction of tsunami propagation. Tsunami occurs mainly due to the thrust (dip-slip) faults that cause vertical uplifting of the water column above the plate while the strike-slip faults, which involve no vertical displacement, are less likely to produce tsunamis. However, Tanioka et al. [9] showed that it is also possible for a strike-slip fault to generate tsunamis, where horizontal displacement of a steep slope failure leads to a significant displacement of the water column.

The far-field tsunami depends mainly on the seismic moment [10]. But for local tsunamis (for which tsunami can strike the coastline within few to 10 minutes), it is difficult to obtain accurate, near real-time M_0 information. Usually, the local (or near-field) tsunami is too rapid for any meaningful attempt to prediction.

The tsunami generation process can be modeled numerically. One such model is the point source model formulated by Steketee [11]. He explicitly derived the expressions of displacement fields at an observation point due to a slip across a fault surface by a point source in an isotropic semi-infinite solid medium for 3 different fault types (i.e. Strike-slip, Dip-slip and Tensile faults).Based on the point source model several theoretical formulations have been developed which describe the deformation of an isotropic homogeneous semi-infinite medium. For example, Maruyama [12] gave the expressions of surface displacements due to vertical and horizontal tensile faults in a semi-infinite Poisson solid and Davis [13] derived an expression of the vertical displacement due to an inclined tensile fault in a half space.

potential of accumulating stress resulting in most of the tsunamis generation. The basic mechanism of tsunami generation by earthquake is depicted in Fig. 2.6 [8].

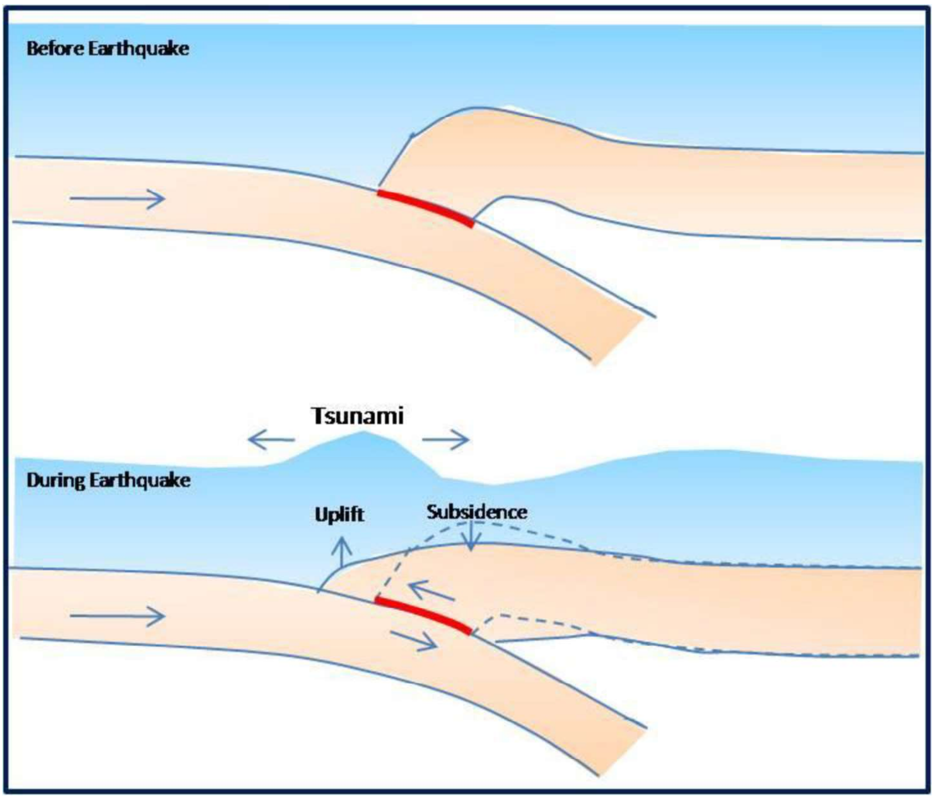


Fig. 2.6: Idealized cross-section through subduction zone showing tsunami generation from earthquake. Relationship of subducting plate (left) to overriding plate (right). Sudden release of strains accumulated over centuries result in a large seafloor uplift causing tsunami.

The Fig.2.6 shows that the interplate contact exhibiting stick-slip friction that drags the overriding plate down with it and at the same time deforming it. Eventually the stress building up on this interplate contact exceeds the strength of the rocks, and the plate simply pops back up into position. When it does that, a mass of water is pushed vertically upward, and that is what causes a tsunami.

Tsunami generation is affected by the seismic moment, focal mechanism, focal depth and other factors. The earthquake size is expressed in terms of seismic moment, $M_0 = \mu DA$, where μ, D and A are rigidity of rocks, average slip and fault surface

2.4). Depending on the type of boundary, two lithospheric plates can separate from each other, slip along each other or override each other (Fig. 2.5)

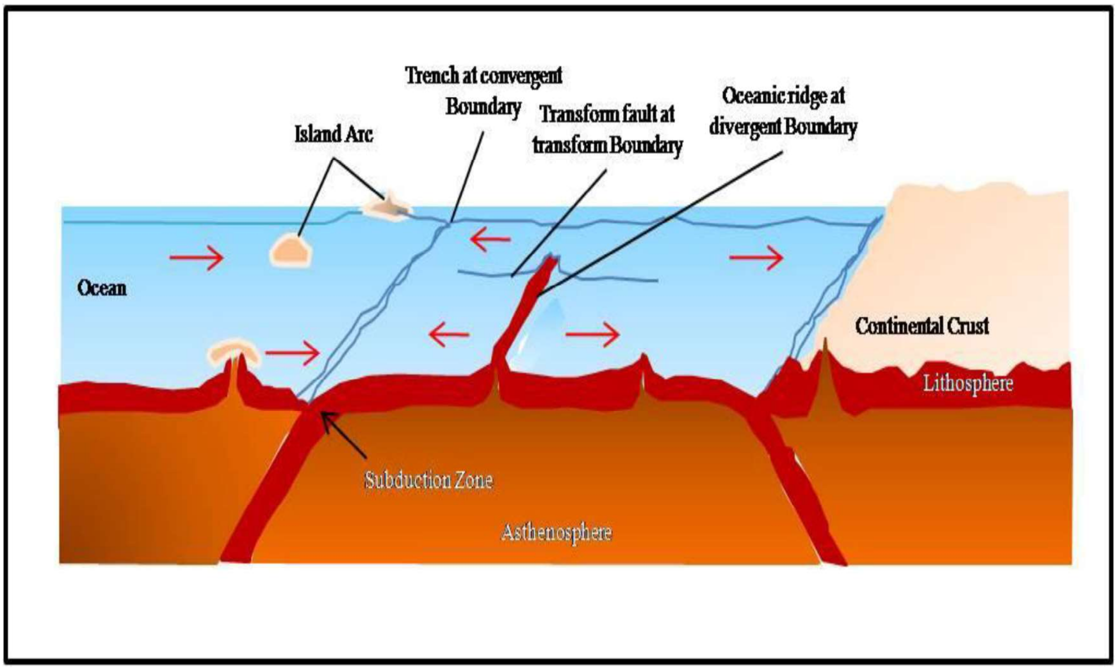


Fig. 2.4: The three principal types of tectonic plate margins and various associated features.

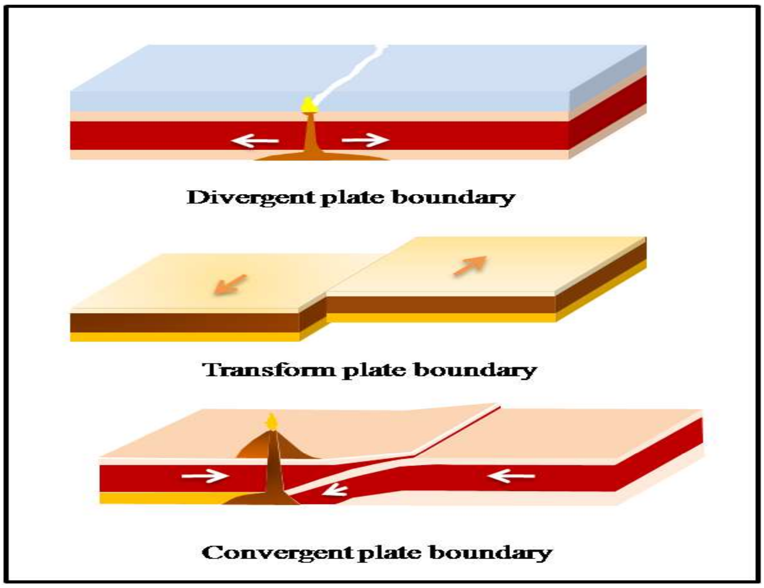


Fig.2.5: Activity at lithospheric plate boundary.

The subduction type of convergent plate boundary is known to have the largest onshore above sea level, called a run-up height, of about 30 meters. A notable exception is the landslide generated by tsunami in Lituya Bay, Alaska in 1958 which produced a 60 meter high wave.

2.2 Review on tsunami generation and propagation and inundation

Tsunamis evolve through three quite distinct physical processes: (1) generation by a force (earthquake, volcano, submarine landslide etc.) that disturbs the water column, (2) propagation from deeper water near source to shallower coastal areas, and (3) inundation of dry land. Of these three phases, propagation is best understood, whereas generation and inundation are more complex and difficult to model with computer simulations. This thesis focuses mainly on the first process that is the initial tsunami generation by earthquake and its identification. However, the later processes are also briefly discussed here for complete understanding of the tsunami problem.

2.2.1 Tsunami generation by earthquake

Past records indicate that 75% of tsunamis are generated by shallow undersea earthquakes as shown in Fig.2.3.

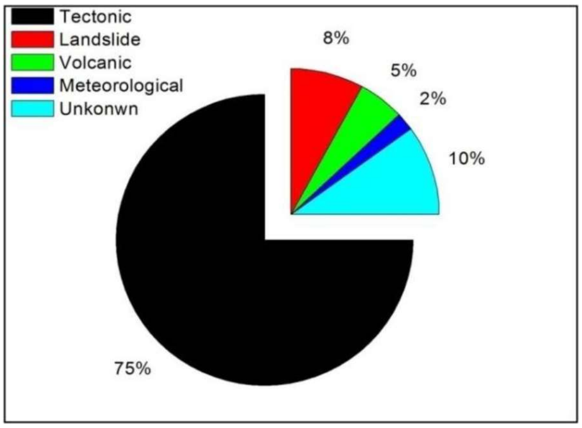


Fig. 2.3: Tsunami sources highlighting earthquakes as the dominating source.

Tsunamis are generally generated at the lithospheric plate boundaries. These plate boundaries are of three types: divergent, strike-slip or transform and convergent (Fig.

reduction in kinetic energy from the retarding propagation of the wave is transferred to potential energy which causes an increase in the amplitude η of the wave leading to significant inundation of coastlines (Fig. 2.2).This phenomenon, called shoaling effect, causes a tsunami, imperceptible at sea, to grow several meters or more in height near the coast. If the trough of the tsunami wave reaches the coast first, this causes a phenomenon called drawdown, where it appears that the sea level has dropped

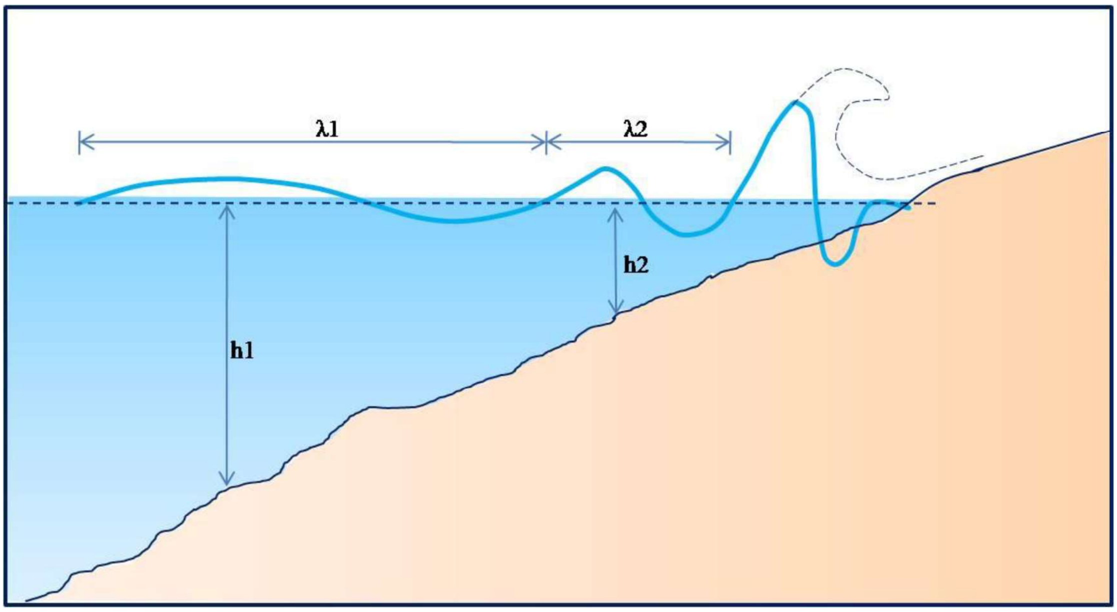


Fig. 2.2: Tsunami waves drag on sea bottom near coastline, becoming shorter in wavelength (λ) but higher in wave amplitude (η) before breaking at the shore. Here $h1>h2$ and $\lambda1>\lambda2$.

considerably. When the crest of the wave hits, sea level rises (called run-up). Run-up is usually expressed in meters above normal high tide. Run-ups from the same tsunami can be variable because of the influence of the shapes of coastlines. The flooding of an area can extend inland up to 300 m or more, covering large areas of land with water and debris. Flooding tsunami waves tend to carry loose objects and people out to sea when they retreat. Tsunami may reach a maximum vertical height quantities vary both as a function of water depth and wave period. However, at T>100 sec both phase and group wave velocities approach $\sqrt{gh}$. This happens in the open ocean, where the wavelength of a tsunami is usually of the order of 100 km, wave amplitude (η) is less than 1 m, and the average water depth (*h)* is few kilometers (e.g. 4 km for Indian Ocean). Thus, the relative depth ($\frac{h}{\lambda}$) and wave steepness ($\frac{\eta}{\lambda}$) are much smaller than unity. The ratio $\frac{h}{\lambda} << 1$ indicates non-dispersive nature of wave while $\frac{\eta}{\lambda} << 1$ gives linearity of the tsunami wave. Therefore, under these two conditions (which is satisfied in open ocean), linear non-dispersive wave theory is valid and can be used to model tsunami wave propagation. Non-dispersive wave motion refers to the motion of wave in which the wave disturbance does not change shape as it propagates. For this to occur, all components of the wave must travel with the same speed. The rate at which a wave loses its energy is inversely related to its wavelength. Since a tsunami has a very large wavelength, it loses little energy as it propagates. Hence, in very deep water, a tsunami will travel at high speeds across great transoceanic distances with limited energy loss.

However, as the tsunami waves move into the shallow waters at the coastlines, dispersion and non-linearity set in. In the shallower water at the coastlines, the wave speed ($c_p = \sqrt{gh}$) decreases and its period ($T = \frac{\lambda}{c_p}$) remains more or less the same as that in deep waters. Consequently, its wavelength decreases. As wavelength decreases and becomes comparable to water depth (i.e. $\frac{h}{\lambda} \to 1$), dispersion sets in and non-linearity also needs to be considered as the waves move into the shallow water. The uniform depth (h) are given by [7]:

$$c_p(\omega) = \sqrt{\frac{gh \ \tanh[k(\omega)h]}{k(\omega)h}} \approx \sqrt{gh} \tag{2.1}$$

$$c_g(\omega) = c_p(\omega)\left[\frac{1}{2} + \frac{k(\omega)h}{\sinh[2k(\omega)h]}\right] \tag{2.2}$$

where g is acceleration due to gravity and $k(\omega)$ is the wavenumber associated with an ocean wave of frequency ω. Wavenumber connects to wavelength $\lambda(\omega)$ as $k(\omega) = \dfrac{2\pi}{\lambda(\omega)}$. Fig. 2.1 shows tsunami wave phase velocity (c_p), wave group velocity

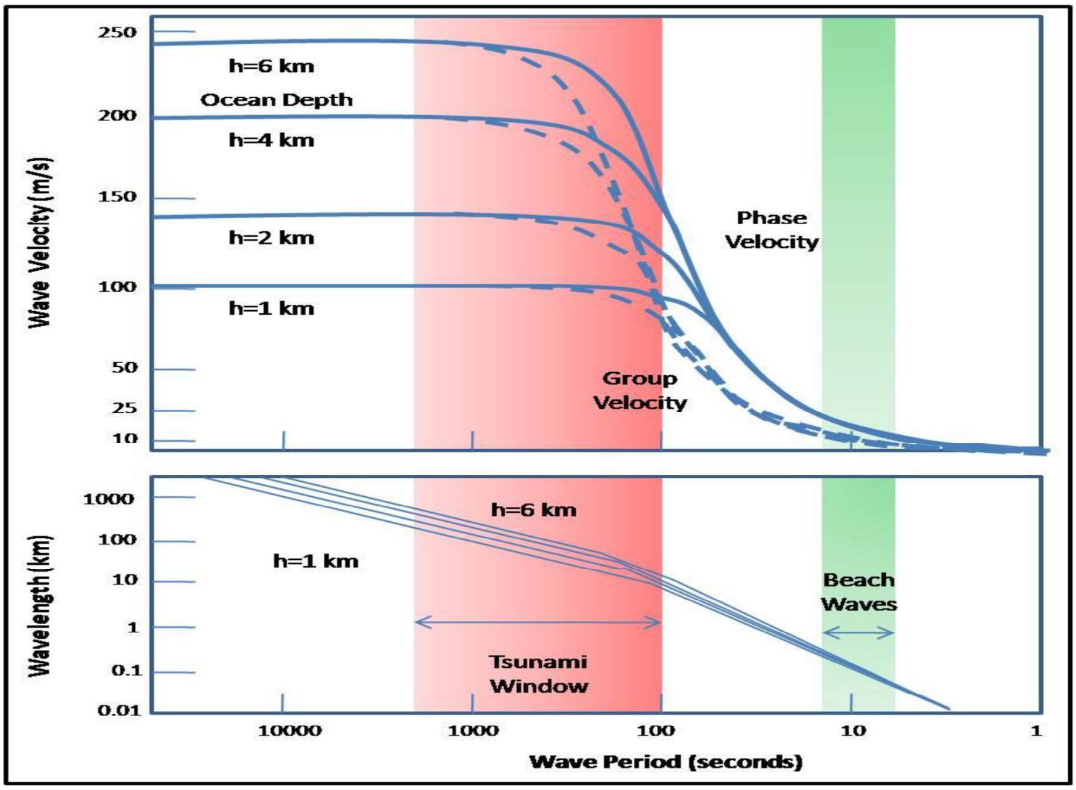


Fig.2.1: Characteristics of tsunami wave: (a) Phase velocity, c_p (solid lines) and group velocity, c_g (dashed lines) of tsunami wave on a flat earth with ocean depths of 1, 2, 4 and 6 km and (b) Wavelength decreases with wave period.

(c_g) and wavelength (λ) against wave period (T) for different ocean depths (h).These

Chapter-2

Literature Review

This chapter reviews the literature and research work related to identification of tsunamigenic earthquakes. Section 2.1 reviews the physical characteristics of a tsunami wave. Section 2.2 reviews the tsunami generation, propagation and inundation. Reviews on existing tsunami warning systems are presented in section 2.3.

2.1Physical characteristics of tsunami wave

A tsunami is a wave train, or series of waves, generated in a body of water by an impulsive disturbance that vertically displaces the water column. Normal ocean waves are caused by the wind, weather, tides, and currents, whereas tsunamis are powered by a geological force. Tsunami waves are surface gravity waves that are formed as the displaced water mass moves under the influence of gravity and radiate across the ocean. Regular wind-driven beach waves involve motion of the uppermost layer of the water only, but tsunami waves move the entire water column from surface to seafloor. Tsunami waves travel with wave speeds of 0.1 to 0.24 km/s with wave periods of 100 to 2000 s and wavelengths of 10 to 500 km in water depths of1 to 6 km [6] whereas the wind driven beach waves travel at speed of about 10 m/s and have wavelength around 100 m and period near 10 s. Tsunami wave height may not be great in the open sea (e.g. < 1 m), but it turns into a giant, rising as high as 30 m at the shore. The regular wind generated beach waves have height of about 3m.

The tsunami phase velocity $c_p(\omega)$ (i.e. the velocity of a monochromatic wave) and group velocity $c_g(\omega)$ (i.e. the velocity of entire group of wave) on a flat ocean of and satisfies Mercer's condition which requires that

$$\int K(x,z)\, g(x) g(z) dx dz \geq 0 \tag{3.39}$$

for any square integrable function $g(x)$.

The same inputs (as mentioned in section 3.5.2 and 3.5.3) could be used to train SVM with target as binary decision (i.e., T/NT). The trained SVM would then be used to get the category of new earthquake from the study region and results may be compared with the previous methods to check the consistency of the results obtain by direct mapping using ANN-I.

3.4.3 Identification of tsunamigenic earthquake using Support Vector Method (SVM)

This section briefly introduces functioning of SVM in the context of non-linear function mapping. The basic SVM deals with two-class problems in which the data are separated by a hyperplane defined by a number of support vectors [91]. Support vectors are a subset of training data used to define the boundary between the two classes. In situations where SVM cannot separate two classes, it solves the problem by mapping input data into high-dimensional feature spaces using a *'kernel'* function. In high-dimensional space it is possible to create a hyperplane that allows linear separation which corresponds to a curved surface in lower-dimensional input space. Accordingly, the *'kernel'* function plays an important role in SVM. In practice, various kernel functions can be used, such as linear, polynomial or Gaussian.

In SVM, classification of a new data sample *'x'* is performed according to the sign of discriminant function or separating function

$$y(x) = \sum_{k \in SV} \alpha_k y_k K\left(x, x_k\right) + b \qquad (3.37)$$

where SV denotes set of support vectors, α_k are the Lagrange multipliers, b is the bias and K is kernel function. The x_k denotes support vectors (training data points with $\alpha_k > 0$) and $y_k \in \{+1, -1\}$ denotes class labels [92]. If $y(x) \geq 0$, then x is classified as a member of the first class, otherwise it is classified as a member of the second class. The kernel function is defined by

$$K\left(x, x_k\right) = \varphi(x)^T \varphi(x_k) \qquad (3.38)$$

Assuming $\{f_{nl}^{i}(\phi_s,\delta,\lambda),\ i=1,2,...n\}$ in Eq.(3.34) are unique invertible functions, it should be possible to estimate fault parameter (ϕ_s,δ,λ) from the phase amplitudes recorded by a single three component station. To obtain the inverse map, an ANN, namely ANN-II (Fig. 4.1) could be trained with seismic phase amplitudes along with other parameters such as location and moment magnitude as inputs and the focal parameters as desired targets. The trained ANN-II could then be used to estimate the focal parameters of a new shallow focus earthquake.

Mapping of absolute water volume and earthquake category

Given the parameters $(M_w,\phi_s,\delta,\lambda)$ of shallow focus earthquakes, the absolute volume (V_T) of displaced water at the tsunami source can be approximately computed by using the relation given by [89]

$$V_T = \frac{M_0 \sin(\delta)\sin(\lambda)}{\mu} = \frac{10^{(M_w + 10.7)/0.67} P}{10^7 \mu} \tag{3.36}$$

Where V_T is in m^3 and the value of μ, the rigidity of the earth at the source, is about $7\times10^{10}\,Nm^{-2}$ for interplate earthquakes [90]and P is the product of $\sin(\delta)$ and $\sin(\lambda)$. Computing and comparing the water volumes with the corresponding categories of the past recorded earthquakes from the study region a tsunami-volume scale may be established.

When new earthquake originates from the same region, first its magnitude and focal parameters would be estimated and then the volume of water displaced at the source could be computed using Eq. (3.36). Finally the earthquake could be categorized by comparing this water volume with tsunami-volume scale.

classification algorithm such as ANN-I may be used. ANN-I could be trained with rms phase amplitudes along with location and magnitude as inputs and the category (i.e., 0/1 corresponding to tsunamgenic/non-tsunamigenic (T or NT)) as desired output for a large number of both tsunamigenic and non-tsunmaigenic earthquakes. Subsequently the trained ANN-I could be used to identify new tsunamigenic earthquake originating from the same geographical region based on the recordings from the given station.

3.4.2 Identification of tsunamigenic earthquake using ANN-II

This method accounts for the theoretical basis behind the direct mapping between seismic phase and earthquake category.

The force system causing an earthquake is modeled as a double couple point source which gives the radiation pattern of P waves as a compressional and a dilatational lobe in alternate quadrants [82]. The polarity of S waves alternates in a similar way. However, S waves have highest amplitude along the nodal planes orthogonal to those of P waves. For a given fault geometry, apart from P and S waves, there could be other phases present in the recorded signal. Each of these phases is generated by interaction of P and S waves taking off from the source at unique angles. For a given *'source-station path'*, as the fault geometry changes, take-off angles of P and S waves change, thereby generating a seismogram at the receiver with a distinctly different set of phase amplitudes. Conversely a given set of phase amplitudes at the receiver at a fixed azimuthal location with respect to a fault will correspond to unique fault geometry and it would be possible to estimate source parameters from them, provided a large number of phases are included in the set. An ANN could be trained using seismic phase amplitudes corresponding to a large number of undersea earthquakes from a given region, to estimate focal parameters.

displaced at the source and eventually deciding category of the earthquake. In addition, a Support Vector Machine (SVM) has also been trained to estimate the earthquake category. Idea behind this was to check the performance of a competitive mapping method vis-à-vis that of an ANN. The details about these techniques are discussed below.

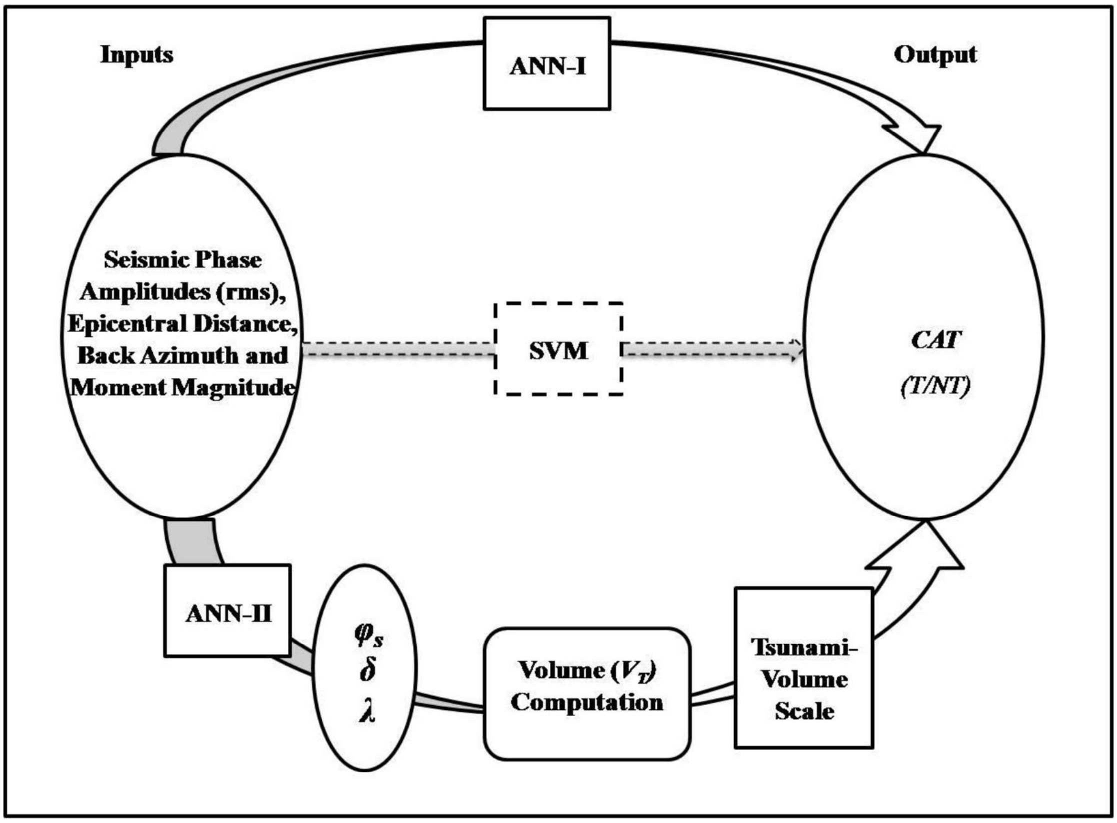


Fig. 3.9: Method of estimating earthquake category (CAT).

3.4.1 Identification of tsunamigenic earthquake using ANN-I

The amplitudes of the seismic waves recorded at a 3-component station at regional distance from a given geographical area are different for a tsunamigenic (slow rupture earthquake) and a non-tsunamigenic earthquake originating from that area, even if other conditions remain same. To classify the earthquakes (i.e.,tsunamigenic/non-tsunamigenic) based on this change in amplitudes, a supervised

$$\left|\mathbf{u}^i(\mathbf{x},t)\right| = K_3 M_w \gamma_{nl}^i(\phi_s,\delta,\lambda),\ i=1,2,...n,\ K_3 \neq K_1 \qquad (3.33)$$

The rms velocity amplitude can then be expressed as

$$\left|\mathbf{v}^i(\mathbf{x},t)\right| = K M_w f_{nl}^i(\phi_s,\delta,\lambda),\ i=1,2,...n,\ K \neq K_3 \qquad (3.34)$$

Eq.(3.34) states that there exists a set of unique non-linear maps between four parameters strike, dip, slip, magnitude and seismic phase amplitudes. In the expanded form Eq.(3.34) can be written as

$$\begin{aligned}
\left|\mathbf{v}^1(\mathbf{x},t)\right| &= K M_w f_{nl}^1(\phi_s,\delta,\lambda) \\
\left|\mathbf{v}^2(\mathbf{x},t)\right| &= K M_w f_{nl}^2(\phi_s,\delta,\lambda) \\
\left|\mathbf{v}^3(\mathbf{x},t)\right| &= K M_w f_{nl}^3(\phi_s,\delta,\lambda) \\
&\vdots \\
\left|\mathbf{v}^n(\mathbf{x},t)\right| &= K M_w f_{nl}^n(\phi_s,\delta,\lambda)
\end{aligned} \qquad (3.35)$$

Eq. (3.35) essentially indicates that for a given *'source-station path'*, the seismic phase amplitudes are function of source strength (magnitude) and focal parameters. Hence a hypothesis may be formally stated as below.

Hypothesis–I: Seismic record of a single 3-component broad-band station contains sufficient information required for categorizing a potentially tsunamigenic earthquake and hence azimuthal coverage of stations may not be necessary.

The proposed hypothesis could be verified through formalization as depicted in Fig. 3.9. The direct mapping (the main method) of root mean square (rms) amplitudes (along with the moment magnitude (M_w), location parameters) with earthquake category has been performed using an ANN (say ANN-I). The direct mapping has been corroborated by an alternating mapping technique. This technique involves computation of focal parameters (using ANN-II), estimation of water volume

Where bw_i stands for body waves i.e., P, SV and SH waves (for i =1, 2, and 3 respectively).Incase the seismic waves are recorded by long period sensor, one can write wave amplitudes as

$$\left|\mathbf{u}^{bw_i}(\mathbf{x},t)\right| = K_1 M_0 f^{bw_i}(\phi_s,\delta,\lambda) \tag{3.29}$$

Where K_1 is a constant for fixed source to sensor geometry. Similarly amplitude of surface wave displacement is given by [86]

$$\left|\mathbf{u}^{sw_i}(\mathbf{x},t)\right| = \frac{1}{2\pi}\int_{-\infty}^{\infty} \chi^{sw_i}(\omega)e^{i\omega t}d\omega \tag{3.30}$$

Where $\chi^{sw_i}(\omega)$ is the Fourier spectrum of $\mathbf{u}^{sw_i}(\mathbf{x},t)$ and sw_i stands for Love and Rayleigh waves (for i= 1 and 2 respectively). Now $\chi^{sw_i}(\omega)$ can be expressed as [88]

$$\chi^{sw_i}(\omega) = M_0 f^{sw_i}(\phi_s,\delta,\lambda) \tag{3.31}$$

From Eqs. (3.29, 3.30and 3.31)it is seen that seismic far field displacement amplitude for body/surface wave may in general takes the form

$$\left|\mathbf{u}^{i}(\mathbf{x},t)\right| = K_2 M_0 \gamma_{nl}^{i}(\phi_s,\delta,\lambda),\ i = 1,2,...n \tag{3.32}$$

where K_2 is a constant for a sensing station monitoring a specific region, n is the number of seismic phases (body or surface waves) and $\gamma_{nl}^{i}(.)$ is the radiation pattern function of i^{th} seismic phase. Subscript nl indicates that the function $\gamma_{nl}^{i}(.)$ is a non-linear function of ϕ_s, δ and λ. Using the relation between M_0 and M_w, Eq. (3.32) can be rewritten as amplitude of the source signal is half of that of the plateau. In a double logarithmic plot, the corner frequency appears roughly at the intersection of two asymptotes: one for the plateau and the other for the high-frequency region of the spectrum. For an earthquake of large magnitude, the corner frequency shifts toward lower frequency in the displacement spectrum. Traditionally magnitude spectrum is used to compute moment magnitude. The Fourier transform is defined over $L^1(\mathbb{R})$ space for integrable and continuous functions [87] which is the case here. Extension of Fourier transform to the $L^2(\mathbb{R})$ space is adopted for addressing the functions which are continuous but not square integrable.

Estimating the M_w by the proposed technique and confirming that it is greater than or equal to 6.0, the tsunamigenic capability of the earthquake could be predicted using broad band records of a single 3-component station as described in the next section.

3.4 Method of identification of tsunamigenic earthquake

For a double-coupled point source in an infinite homogeneous elastic medium, the far-field P wave displacement at receiver position $\mathbf{x}$ and time t can be written (using Eq. 3.22) as

$$\mathbf{u}^{\mathbf{P}}(\mathbf{x},t) = \frac{M_0 R^P}{4\pi\rho\alpha^3 r} \tag{3.27}$$

Although exact relation for radiation patterns (R^{SV} and R^{SH}) of vertical and horizontal components of shear wave are different, the general nature of the relationship for seismic body waves coming from a known region (with identical i_ξ and ϕ) may be written as

$$R^{bw_i} = f^{bw_i}(\phi_s, \delta, \lambda) \tag{3.28}$$

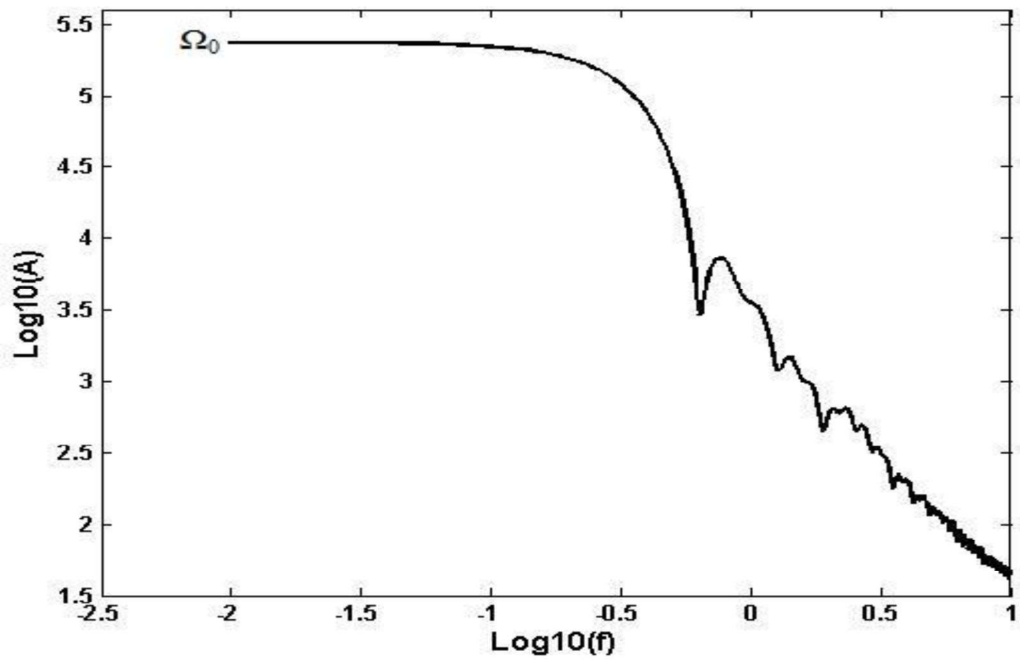


Fig. 3.8: Plot of logarithm of displacement amplitude spectrum (A) vs. logarithm of frequency (f) for a typical earthquake signal at P onset. Ω_0 is the low frequency saturation value of the amplitude spectrum.

The above constraints may be eliminated, if the complete P wave amplitude spectrum which is essentially a combined effect of all the above mentioned factors, be mapped directly to M_w using an ANN. The direct mapping between amplitude spectrum and M_w by ANN may be achieved by training it using seismograms recorded in a station at regional distance for large number of seismic events that occurred in the past. This technique is likely to give a better estimate of M_w because magnitude of an earthquake not only affects amplitude but also the frequency content of the recorded signal.

The displacement spectrum (i.e. magnitude/amplitude spectrum), obtained by computing the Fourier transform of P wave displacement waveform [84-86]. It essentially contains a flat low frequency level and a high frequency region in which the spectral amplitude decays rapidly with increasing frequency. The transition region is commonly known as the corner frequency (f_c) at which the spectral

Since the Fourier transform of a boxcar function is a sinc function, the far-field P wave displacement source spectrum can be modeled as

$$A(f) = F(\mathbf{u}^{\mathbf{P}}(\mathbf{x},t)) = \frac{M_0 R^P}{4\pi\rho\alpha^3 r}\frac{\sin(\pi f T_D)}{\pi f T_D}\frac{\sin(\pi f T_R)}{\pi f T_R} \tag{3.23}$$

Where the symbol F represents Fourier transform. The moment magnitude M_w is related to M_0 as [83]

$$M_w = \frac{2}{3}\log M_0 - 10.7 \tag{3.24}$$

Where M_0 is in dyne-cm. At very low frequency Eq. 3.22 can be approximated as

$$A(f \approx 0) = \Omega_0 = \frac{M_0 R^P}{4\pi\rho\alpha^3 r} \tag{3.25}$$

from which we can write

$$M_0 = \frac{4\pi\rho\alpha^3 r\Omega_0}{R^P} \tag{3.26}$$

Where Ω_0 is the value of P wave amplitude at very low frequency (Fig. 3.8). Ω_0 can be measured directly from the amplitude (displacement) spectrum or can be estimated by fitting a source model spectrum to amplitude spectrum. Eqs. (3.24 and 3.26) show that M_w can be estimated from the knowledge of P wave velocity, source-receiver distance, P wave radiation pattern, and amplitude spectrum. However, the above methods may not provide good estimation of M_w due to the uncertainties involved in approximating radiation pattern, velocity and rock density. Further, M_w estimation by above mentioned methods needs to have a prior knowledge of location of the event as well as recordings of multi-station data.

wave displacement (Eq. 3.17) due to finite source of length x and width w can be written using Eq.(3.19) as

$$\mathbf{u}^{\mathbf{P}}(\mathbf{x},t) = \frac{1}{4\pi\rho\alpha^3}\frac{R^P}{r}\int_0^x \mu\dot{\bar{u}}(t-\frac{x}{v_R})wdx \tag{3.20}$$

$$= \frac{1}{4\pi\rho\alpha^3}\frac{R^P}{r}\mu w\dot{\bar{u}}(t)\int_0^x \delta(t-\frac{x}{v_R})dx;\quad \dot{\bar{u}}(t) \text{ is independent of } x$$

$$= \frac{1}{4\pi\rho\alpha^3}\frac{R^P}{r}\mu w\dot{\bar{u}}(t)\int_t^{t-\frac{x}{v_R}} \delta(z)(-v_R dz);\quad \text{Let } t-\frac{x}{v_R}=z$$

$$= \frac{1}{4\pi\rho\alpha^3}\frac{R^P}{r}\mu w v_R\dot{\bar{u}}(t)\int_{t-\frac{x}{v_R}}^{t} \delta(z)dz$$

$$= \frac{1}{4\pi\rho\alpha^3}\frac{R^P}{r}\mu w v_R\dot{\bar{u}}(t)H(z)\Big|_{t-\frac{x}{v_R}}^{t};\quad H(z) \text{ is a Heaviside step function}$$

$$= \frac{1}{4\pi\rho\alpha^3}\frac{R^P}{r}\mu w\frac{L}{T_R}\dot{\bar{u}}(t)B(t,T_R);\quad At\ x=L,\ v_R=\frac{L}{T_R} \text{ and}$$

$$B(t,T_R) = H(t)-H(t-T_R)$$

=Boxcar function

Let us consider the displacement history of a particle as a ramp function, given below

$$\bar{u}(t) = \begin{cases} 0; & t \le 0 \\ \dfrac{\bar{u}}{T_D}t; & 0 < \mathrm{t} < T_D \\ \bar{u}; & t \ge T_D \end{cases} \tag{3.21}$$

Where T_D is rise time and $\bar{u}$ is the final displacement. Using Eq. (3.21), Eq. (3.20) can be rewritten as

$$\mathbf{u}^{\mathbf{P}}(\mathbf{x},t) = \frac{1}{4\pi\rho\alpha^3}\frac{R^P}{r}\mu wL\bar{u}\frac{B(t,T_D)}{T_D}\frac{B(t,T_R)}{T_R};\quad B(t,T_D) \text{ is a} \tag{3.22}$$

Boxcar of duration T_D

$$= \frac{M_0 R^P}{4\pi\rho\alpha^3 r}\frac{B(t,T_D)}{T_D}\frac{B(t,T_R)}{T_R};\quad M_0 = \mu wL\bar{u}$$

= seismic moment

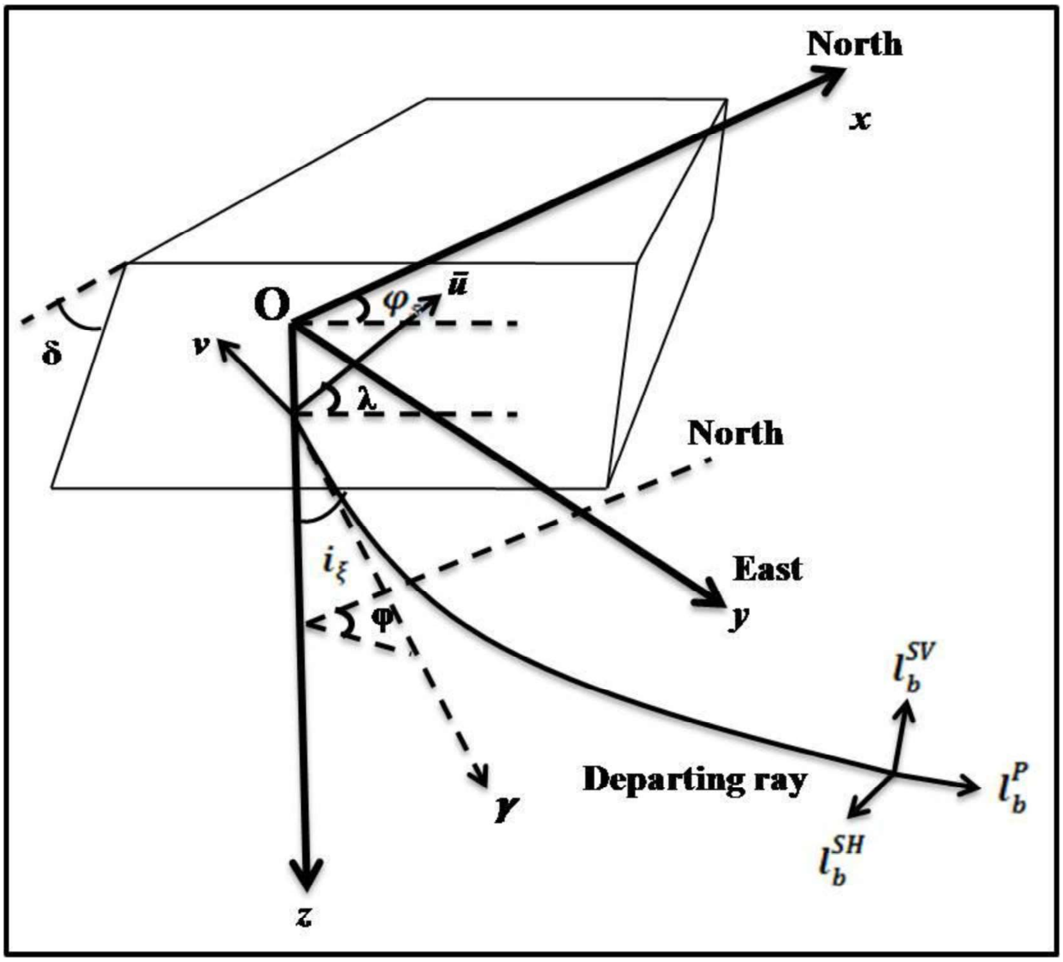


Fig. 3.6: Rupture geometry showing relative orientation of the fault plane parameters.

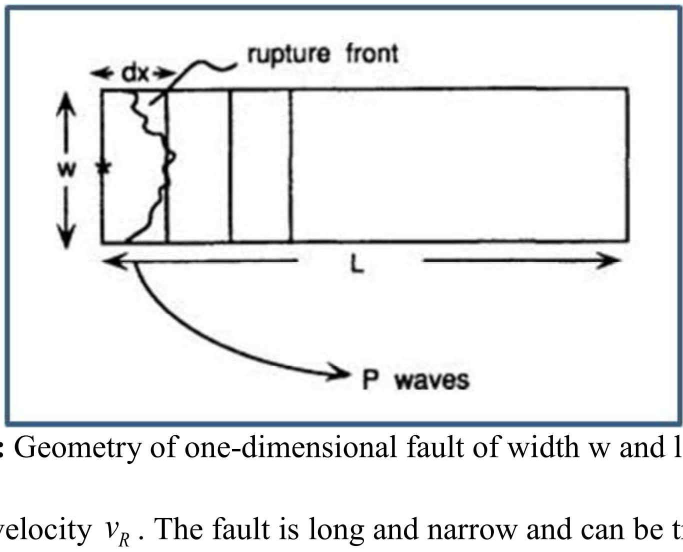


Fig. 3.7: Geometry of one-dimensional fault of width w and length L.

with a constant velocity v_R. The fault is long and narrow and can be treated as a series of small segments that individually approximate point sources. Therefore far-field P

$$R^P = \cos\lambda \sin\delta \sin^2 i_\xi \sin 2(\phi - \phi_s) - \cos\lambda \cos\delta \sin 2i_\xi \cos(\phi - \phi_s) \\ + \sin\lambda \sin 2\delta (\cos^2 i_\xi - \sin^2 i_\xi \sin^2(\phi - \phi_s)) \\ + \sin\lambda \cos 2\delta \sin 2i_\xi \sin(\phi - \phi_s) \tag{3.18}$$

The strike (ϕ_s), dip (δ), slip (λ), take-off angle (i_ξ) and azimuth (ϕ) are shown in Fig. 3.6. The time-dependent seismic moment $M_0(t)$ can be expressed in terms of slip vector $\overline{u}$ (i.e. particle displacement at the source, averaged over fault area A') as

$$M_0(t) = \mu \overline{u}(t) A'(t) \tag{3.19}$$

Where μ is rigidity of the Earth. In order to take into account the finiteness of the fault, let us consider the geometry (Fig. 3.7) of a one-dimensional fault of width w and length L. The individual segments of the fault are of length dx. The fault ruptures formation using the sign of τ_r and τ_b. All the angles are in degrees. The value of ϕ is 90^0.

Knowing Δ and Z, the source location ($slat, slon$) can be computed with respect to the station coordinates ($stnlat, stnlon$) using the relations as shown in Eq.(3.16).

$$\begin{cases} slat = \sin^{-1}[\sin(stnlat)\cos(\Delta / R) + \cos(stnlat)\sin(\Delta / R)\cos(Z)] \\ slon = stnlon + \tan^{-1}[a / b] \\ a = \sin(Z)\sin(\Delta / R)\cos(stnlat) \\ b = \cos(\Delta / R) - \sin(stnlat)\sin(slat) \end{cases} \tag{3.16}$$

where $slat, slon, stnlat, stnlon, Z$ are in radian, Δ is in km. R is radius of the earth (6378.1 km).

Estimating the location of an event its magnitude could be computed using broad band data recorded by a single 3-component station as described in the next section.

3.3Method of estimation of moment magnitude

The far-field P wave displacement from a double-coupled point source with a moment $M_0(t)$ in an infinite homogeneous elastic medium at receiver position $\mathbf{x}$ and time t is given by [82]

$$\mathbf{u}^P(\mathbf{x},t) = \frac{1}{4\pi\rho\alpha^3}\frac{R^P}{r}\dot{M}_0(t - \frac{r}{\alpha}) \tag{3.17}$$

Where ρ is the density of the earth at the source, α is the P wave velocity, r is the hypocentral distance and R^P is the P wave radiation pattern which is expressed as

In actual practice, lag along an arm can be computed by observing the onsets (T) at i^{th} and $(i+1)^{th}$ detector traces and calculating the time differences $\tau\,(=T_i - T_{i+1})$. A lag is taken to positive if τ is positive. From the knowledge of τ_r and τ_b the apparent velocity (V) could be obtained using Eq.(3.14). Thus velocity being known, the epicentral distance (Δ) can be calculated using the polynomial fit of distance–velocity relations. With the knowledge of lags τ_r and τ_b back-azimuth (Z) could be estimated using Eq.(3.15) as shown in Fig.3.5.

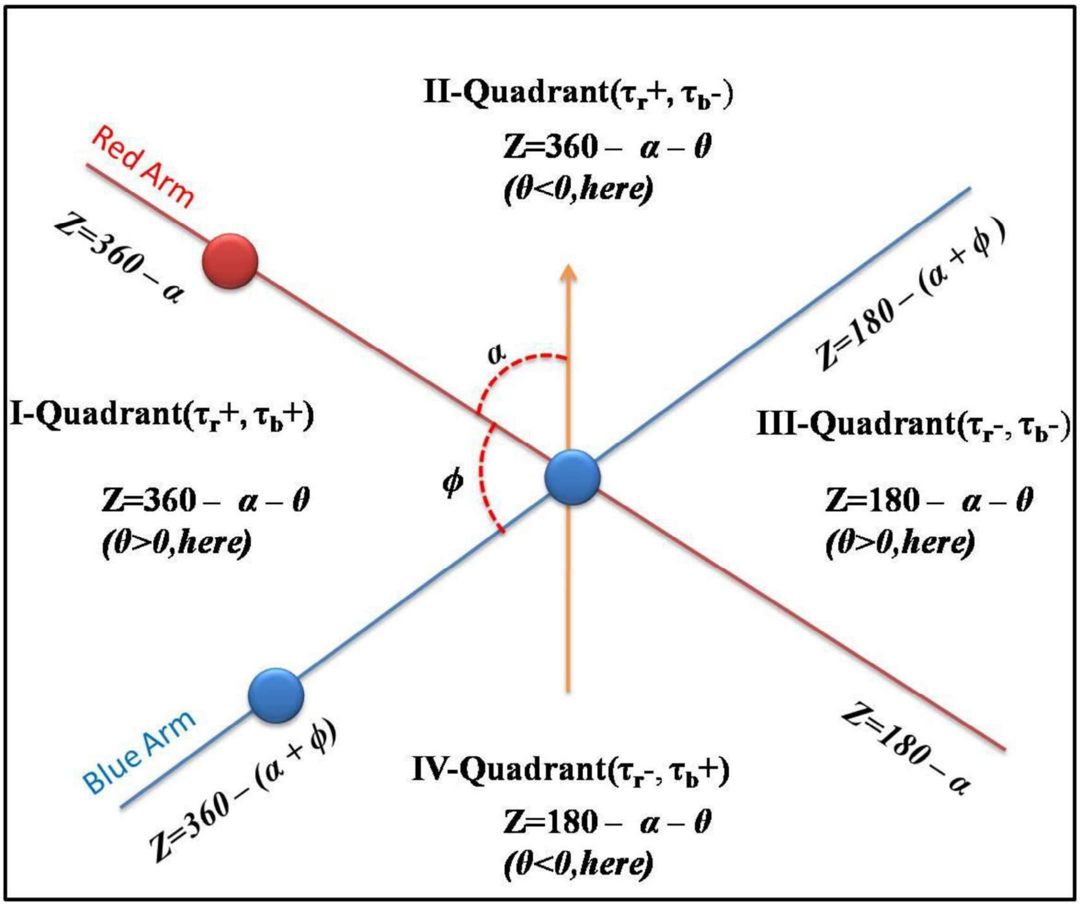


Fig. 3.5: Estimation of Back-azimuth (Z) from ray inclination angle (θ) by quadrant

arm are $d\cos\theta$ and $d\sin\theta$ respectively, with d as spacing between two consecutive sensor elements. Now τ_r and τ_b can be computed as follows:

$$\begin{cases} \tau_r = \dfrac{d\cos\theta}{V} \\ \tau_b = \dfrac{d\sin\theta}{V} \end{cases} \tag{3.13}$$

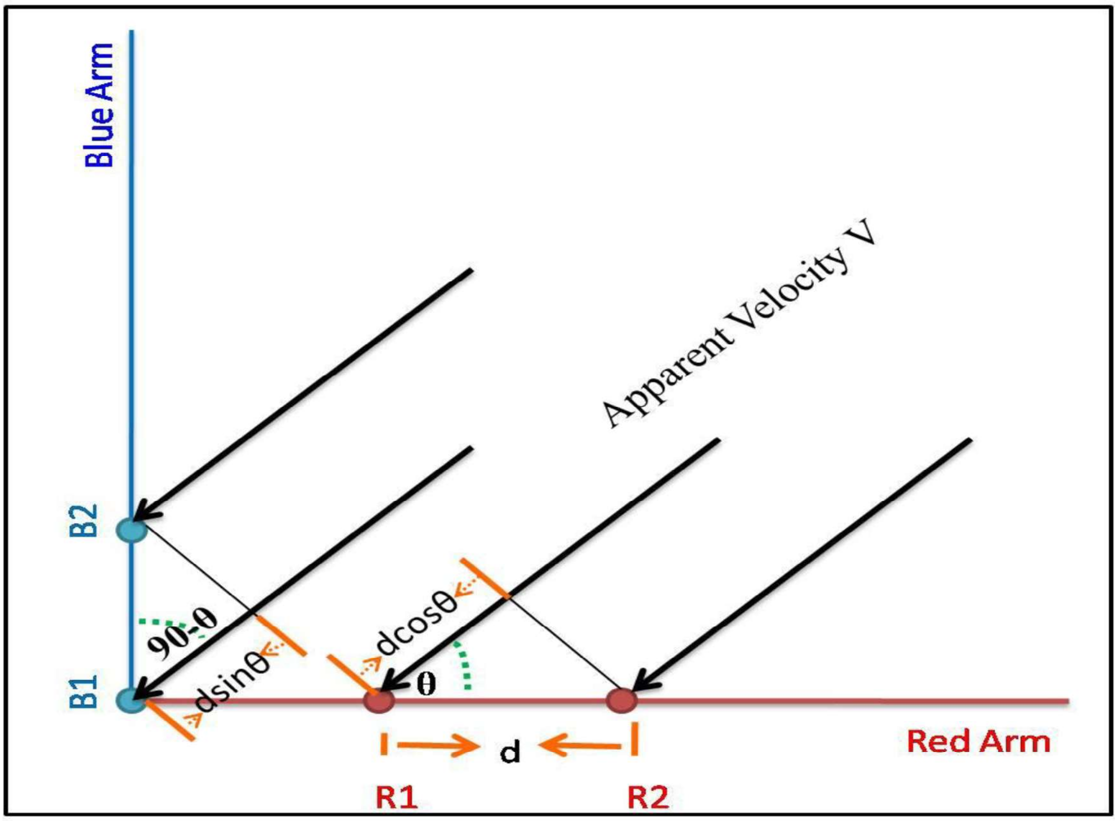


Fig. 3.4: Ray diagram to compute time delays along red and blue arms.

Squaring and adding (3.13) we get

$$V = \frac{d}{\sqrt{\tau_r^{\,2} + \tau_b^{\,2}}} \tag{3.14}$$

Dividing one by the other in (3.13) we get

$$\theta = \tan^{-1}\left(\frac{\tau_b}{\tau_r}\right) \tag{3.15}$$

stationary seismic noise (with less irregular fluctuations) allows lower STA/LTA trigger threshold level, whereas completely irregular behavior of seismic noise demands higher values.

The STA/LTA de-trigger threshold level determines the termination of a recording. It determines how well the coda waves of recorded earthquakes will be captured in data records. To include as much coda waves as possible, low value of threshold is required. On the other hand, a high value of STA/LTA de-trigger is preferred if coda waves are not wanted.

Once an event is detected by the proposed technique it could be located using the time lags between sensors in the seismic array as described in the section.

3.2 Method of location of seismic event

When a seismic ray traverses an array of sensors it reaches different sensors at different instants, the velocity and direction of arrival being fixed by the source with respect to the array. Consequently a phase difference in signal waveform is introduced from sensor to sensor. On the actual record a relative shift in P arrivals is noticed in time domain all along the array channels. These significant time delays along two arms of L shaped array, called the blue and red arms, may be conveniently called as the blue lag or τ_b and red lag or τ_r respectively. Consider a seismic wave traversing the array with an apparent velocity V (Fig. 3.4).

Suppose the direction of velocity makes an angle θ to the red arm. As seen in the diagram the signal approaches R2 first and R1 later. Similarly, it reaches B2 first and B1 later. The path differences between two consecutive elements on red and blue

$$\mathrm{R}(i) = \frac{\mu_s(i)}{\mu_l(i)} \tag{3.12}$$

When the ratio exceeds a predetermined threshold, detection is declared. The trigger is active until the ratio falls below trigger-off threshold. The most important STA/LTA trigger algorithm parameters are thus the STA and LTA window lengths and the detection threshold [72].

STA window measures average amplitude of a seismic signal. Generally, STA duration must be longer than a few periods of typically expected seismic signal. If the STA window is too short, averaging of the seismic signal will not function properly. The shorter the duration selected, the higher the trigger's sensitivity to short lasting local earthquakes compared to long lasting distant earthquakes giving rise to lower frequency. The longer the STA duration selected, the less sensitive it is for local earthquakes. The STA duration is also important from the aspect of false triggers.

The LTA window measures average amplitude seismic noise. It should last longer than a few periods of typically irregular seismic noise fluctuations. A short LTA duration allows that the LTA value to more or less adjust to the slowly increasing amplitude of emergent seismic waves whereas a long LTA window duration increases trigger sensitivity to emergent earthquakes because the LTA value is not so rapidly influenced by the emergent seismic signal.

The STA/LTA trigger threshold level has the largest contribution that determines which events will be recorded. The higher the value is set, the more number of earthquakes will be missed, but the fewer false-triggers will result. The lower the STA/LTA trigger threshold level is selected, more events will be recorded with a penalty of more frequent false triggers. Optimal STA/LTA trigger threshold level depends mostly on amplitude and type of seismic noise at the site. Statistically computational difficulty becomes prominent if T is less than W i.e., for a problem dealing with limited number of training examples as in the case of present study.

To embed learning from past history in the network, it is important to choose an appropriate training algorithm. Resilient propagation [81] algorithm may be chosen for training the ANN. The weights could be adapted based on local gradient information. The trained ANN could be used to detect new seismic events. The detection accuracy of STFT and ANN method could be compared by using the popular detection algorithm STA/LTA (Short Term Average/Long Term Average) which has been described briefly in the following section.

3.1.3 Detection using STA/LTA method

The definition of short term average (μ_s) of a seismic data channel V is

$$\mu_s(i) = \frac{1}{N_S}\sum_{j=1}^{N_S} V(i-j) \tag{3.10}$$

where i is sample number and N_s is the product of sampling rate(samples/sec) and STA window length (sec). The definition of long term average (μ_l) of the seismic data channel V is

$$\mu_l(i) = \frac{1}{N_L}\sum_{j=1}^{N_L} V(i-j) \tag{3.11}$$

where N_L is the product of sampling rate and LTA window length. Here overlapping STA and LTA windows have been considered. The short term average represents the average of the shortest period over which an event of interest could occur and the long term average represents the average of the longest period to assess the background noise. The ratio between them is defined as layers, results in under-fitting, indicating inadequacy of the network to detect the pattern in a complicated dataset. On the other hand, using too many neurons in the hidden layers may cause over-fitting indicating that the neural network has so much information processing capacity that the limited amount of information contained in the training set is not enough to train all of the neurons in the hidden layers. Further, even when the training data is sufficient, an inordinately large number of neurons in the hidden layers can increase the training time to the point that it becomes impossible to adequately train the neural network. Obviously, some trade-off must be reached between too many and too few neurons in the hidden layers. Although there are many rule-of-thumb methods for determining the correct number of neurons to use in the hidden layers, ultimately the selection of an architecture of the neural network for a particular problem comes down to trial and error.

The computational difficulty may arise either from sample complexity or training time. Sample complexity is the number of training examples (T) required to learn the class and the training time is the computation time required to learn the class. For a three layer neural network, the number of interconnecting weights (W) is given by

$$W = (N + Q)H \tag{3.9}$$

For a good generalization of performance the number of training examples should grow at least linearly with the number of adjustable parameters (i.e., W) in the network. The time complexity can be expressed by the term nTW where n is the number of epochs. More specifically, the computational complexity for training can be expressed as N^2C, where C is the number of classifiers. This suggests that the computational complexity is highly sensitive to the number of input features. Thus the

Thus for an input vector X ($X^0 = X$), the network estimated output vector $\hat{Y}_m$ can be determined as

$$\hat{Y}_m = \hat{f}(X;W) = \left[X^l\right]_{l=L-1} = \left[G_l(A^l)\right]_{l=L-1} = \left[G_l(W^{l-1}X^{l-1})\right]_{l=L-1} \qquad (3.7)$$

Where G_l is a vector function which applies a non-linear activation function $g_l(.)$ to each component of its vector argument A^l.Eq. (3.7) in the parametric model, takes the form

$$\hat{f}_k(x_1, x_2, ..., x_{N;}W) = \sum_{j=1}^{H} w_{jk} g_j \left(\sum_{i=1}^{N} w_{ij} x_i \right), k = 1, 2, ..., Q \qquad (3.8)$$

Where N is the number of nodes in input layer, H in the hidden layer and Q in the output layer of a three layer network. The weights w_{ij} connect i^{th} neuron in the input layer to j^{th} neuron in the hidden layer, the weights w_{jk} connect j^{th} neuron in the hidden layer to the k^{th} neuron in the output layer. $g_j(.)^{'}s$ are the activation functions in the hidden layer. The activation functions for the output nodes are considered as linear. Essentially (3.8)states that arbitrary non-linear functions could be approximated by a linear combination non-linear activation functions.

The choice of number of hidden layers and the number of neurons in these layers depends on the nature of the problem. It is well known that an ANN with one or more hidden layers can approximate an arbitrary function that contains a dense mapping from one finite space to another. Deciding the number of neurons in the hidden layers is an important part of designing the overall neural network architecture. Though these layers do not directly interact with the external environment, they have tremendous influence on the final outcome. Using too few neurons in the hidden

The smoothened spectra extracted using STFT need to be mapped with event or noise category for detection of an event as mentioned earlier. However this mapping function is not linear in nature. This non-linear mapping could be achieved using an ANN as stated below.

3.1.2 Non-linear function mapping using ANN

Multilayer feedforward neural networks have been proposed as a tool for non-linear function approximation from a given set of input-output pairs [77-79].It has also been proven that a three layer network can approximate any Borel measurable function provided sufficiently many hidden layer units are used [77]. In fact, multilayer networks could be referred to as universal approximators.

Let us assume that an unknown function f has to be approximated from a training set of input-output pairs, say, $S = \{(X_m, Y_m); m = 0,1,...M\}$ with input vector $X_m = (x_{m1},...,x_{mN})^T$ and output vector $Y_m = (y_{m1},...,y_{mQ})^T$ such that $Y_m = f(X_m) + e_m$. Here M is the total number of input-output pairs in the training set and e_m is random error. The job of ANN is to estimate $\hat{f}(.)$ of $f(.)$ obtained by minimization of approximation error.

Multilayer networks are highly non-linear models for function approximations [80]. In general a L -layer feed forward network with fixed non-linear activation functions can be parameterized by a set of $L-1$ weight matrices, $W = \{W^l, l = 0,1,...,L-2\}$. The weight matrix W^l relates the l^{th} layer output vector X^l to $(l+1)^{th}$ layer activation vector $A^{(l+1)}$ by

$$A^{(l+1)} = W^l X^l, 0 \le l \le L-1 \tag{3.6}$$

$$H(f) = \frac{\sin(\pi f M)}{M \sin(\pi f)} \tag{3.4}$$

Filtering with M point moving average filter in frequency domain leads to pre-multiplying the signal with a modified Tukey window $w'(t)$ which is defined by

$$\begin{aligned} & w'(t) = w(t)H(t); \\ & \text{where } H(t) = F^{-1}[H(f)] = \int H(f) e^{j2\pi f t} df \end{aligned} \tag{3.5}$$

Thus it can be stated that the smoothening of STFT based amplitude spectra is equivalent to computing amplitude spectra using STFT with a modified Tukey window $w'(t)$ (Fig.3.3).

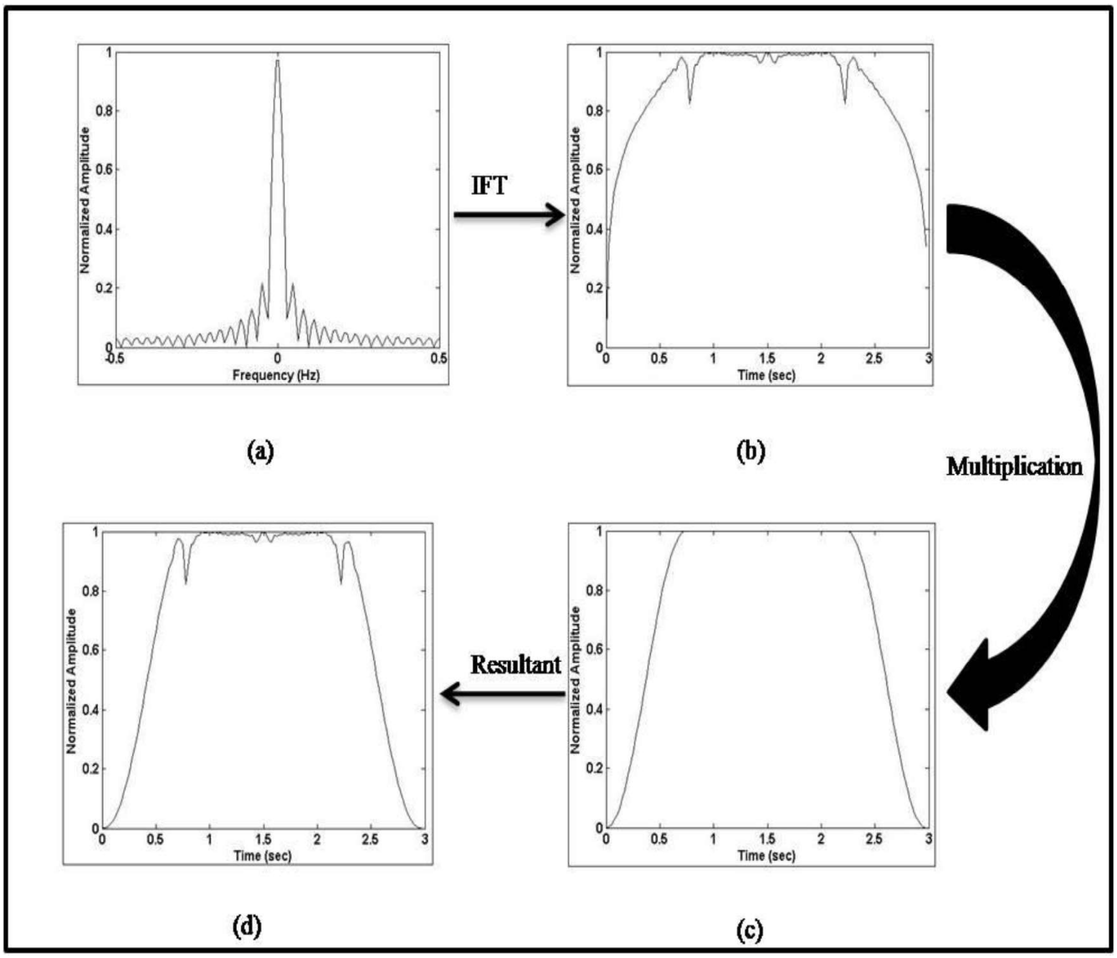


Fig.3.3: (a) Frequency responses of a 31 point moving average filter. (b) Inverse Fourier Transform (IFT) of the response of moving average filter. (c) Time domain plot of Tukey window defined by Eq. (3.2). (d) Time domain product of Fig. (b) and (c), called as modified Tukey window $w'(t)$.

sliding window of 3 sec centered at times τ (i.e., 2 sec,3 sec, etc.). Through consecutive movement of window and performance of Fourier transform, the Fourier transform of entire signal can be performed. The signal segment within the window function is assumed to be approximately stationary. This way, decomposing the time domain signal into time-frequency representation, STFT reveals the variations of the frequency content of the signal within the window function. The window $w(t)$ may be considered as cosine-tapered (Tukey) window which is defined by [71]

$$w(t)=\begin{cases}1.0, 0\le|t|\le\alpha\frac{S}{2}\\ 0.5\left[1.0+\cos\left[\pi\frac{t-\alpha\frac{S}{2}}{2(1-\alpha)\frac{S}{2}}\right]\right], \alpha\frac{S}{2}\le|t|\le\frac{S}{2}\end{cases} \tag{3.2}$$

where α is percentage of cosine taper and S is window length. This taper basically suppresses the side-lobe spectral leakage and reduces the bias of spectral estimates [72-73]. To remove the short term variations to reveal the important underlying features of the signal [74] (as shown in Fig 3.2) the STFT based spectra of the signal need smoothening which could be performed using a moving average filter. The M point moving average filter output (say, $\bar{Y}$) for equally spaced spectral data (Y) is defined as [75]

$$\bar{Y}(f)=\frac{1}{2M+1}\sum_{f=-M}^{M}Y(f) \tag{3.3}$$

This smoothening is equivalent to low pass filtering whose frequency response can be expressed as [76] view of these characteristics, STFT has been adopted to characterize and distinguish seismic event and noise. Mathematically STFT can be expressed by [65]

$$Y(\tau, f) = \int y(t)w(t-\tau)e^{-j2\pi ft}dt \tag{3.1}$$

which can be viewed as measure of similarity between the signal $y(t)$ and the time-shifted and frequency-modulated window function $w(t)$.The feature extraction using STFT has been shown in Fig. 3.2. As shown in the figure, the STFT employs a

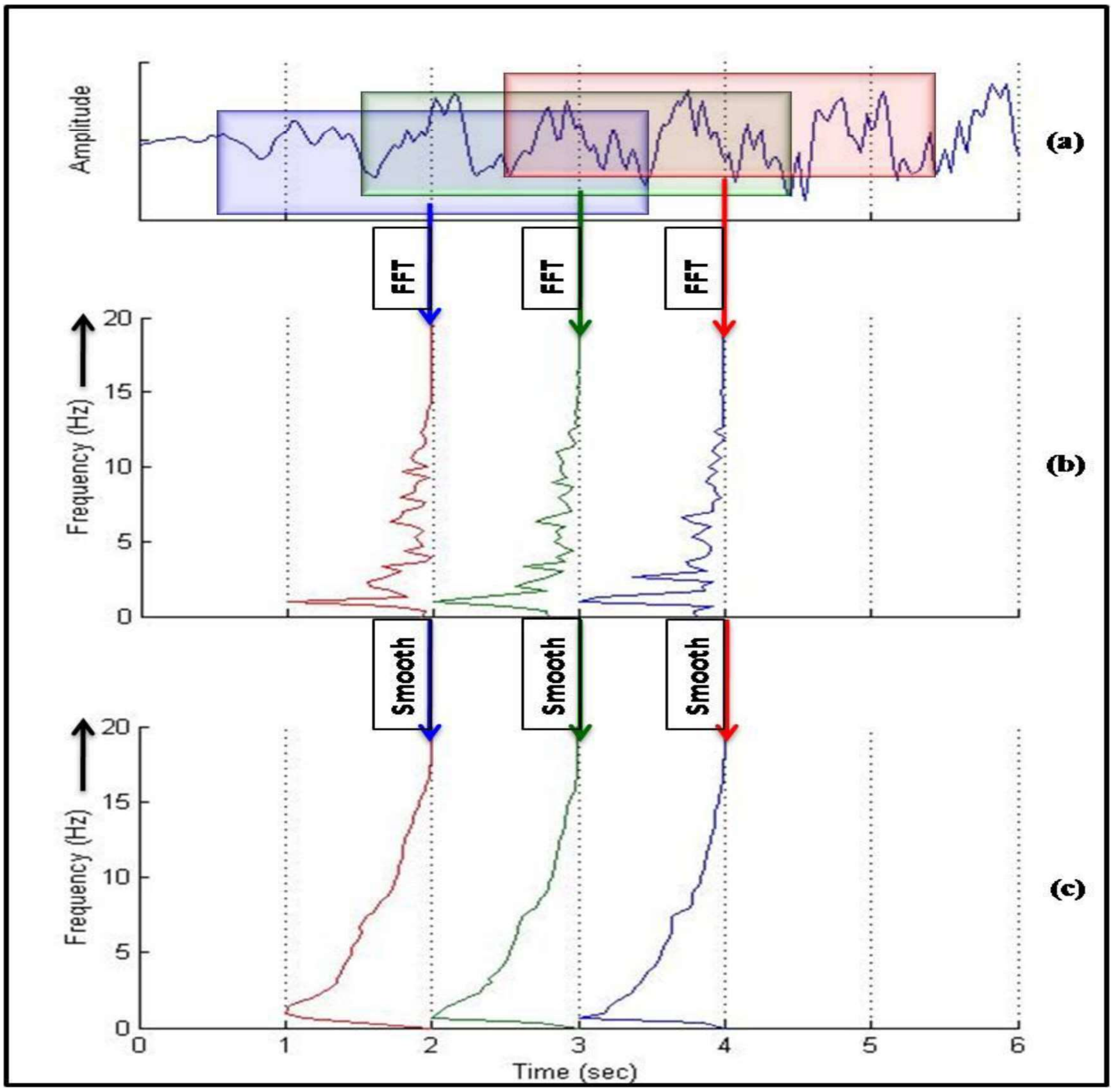


Fig. 3.2: Extraction of spectral features using STFT: (a) Plot of a typical signal. (b) FFT of the signal for window length of 3 sec at three different instants with an overlap of 2 sec. (c) Corresponding smoothened amplitude spectra.

where n is number of event onsets, μ and σ are mean and standard deviation of X and p is a constant to be chosen judiciously. The actual onset time is then found by removing the outliers and averaging the remaining onsets. For online detection of the event the above procedure could be repeated by moving the window in steps of 1 sec. Motivations for the choice of STFT to extract feature (i.e., amplitude spectra) and ANN to map these features with category (i.e., event/ noise) are explained in the following sections.

3.1.1 Feature extraction using STFT

Seismic signal is non-stationary in nature. Analysis of seismic signal only in time or in frequency domain is not an appropriate approach. Analysis in both domains simultaneously provides deeper insight to a non-stationary signal. Some typical examples of this kind of analysis are the Gabor transform or STFT [65-66], the S-Transform [67], bilinear transforms like the Wigner-Ville [68] or wavelet transforms [69-70]. The main difference among these methods concerns their dual time–frequency resolution capability. Having a fixed window, STFT provides uniform time and frequency resolutions in the entire time-frequency domain whereas in wavelet analysis by shifting and reshaping (dilating or contracting) wavelets at different scales multi-resolution analysis of the signal is achieved. This has inspired the wide use of the wavelet transforms recently. However, wavelet transform is computationally intensive and it also needs proper choice of wavelets for a specific application. In contrast, fixed resolution STFT provides the information about the spectral content of the signal over smaller time windows, capturing local variations. It is actually a compromise between exposition of information and computational complexity. In the ANN shows decision as event. If the decisions are found to match K number of channels, the occurrence of an event will be declared. In order to obtain the onset time, the event onsets (i.e., sample numbers/sampling rate) will be stored in an array. From the stored array (say, X) the outliers will be found by satisfying the inequality

$$X(n) - \mu(X) \geq p\sigma(X), 1 \leq n \leq K \quad (3.1)$$

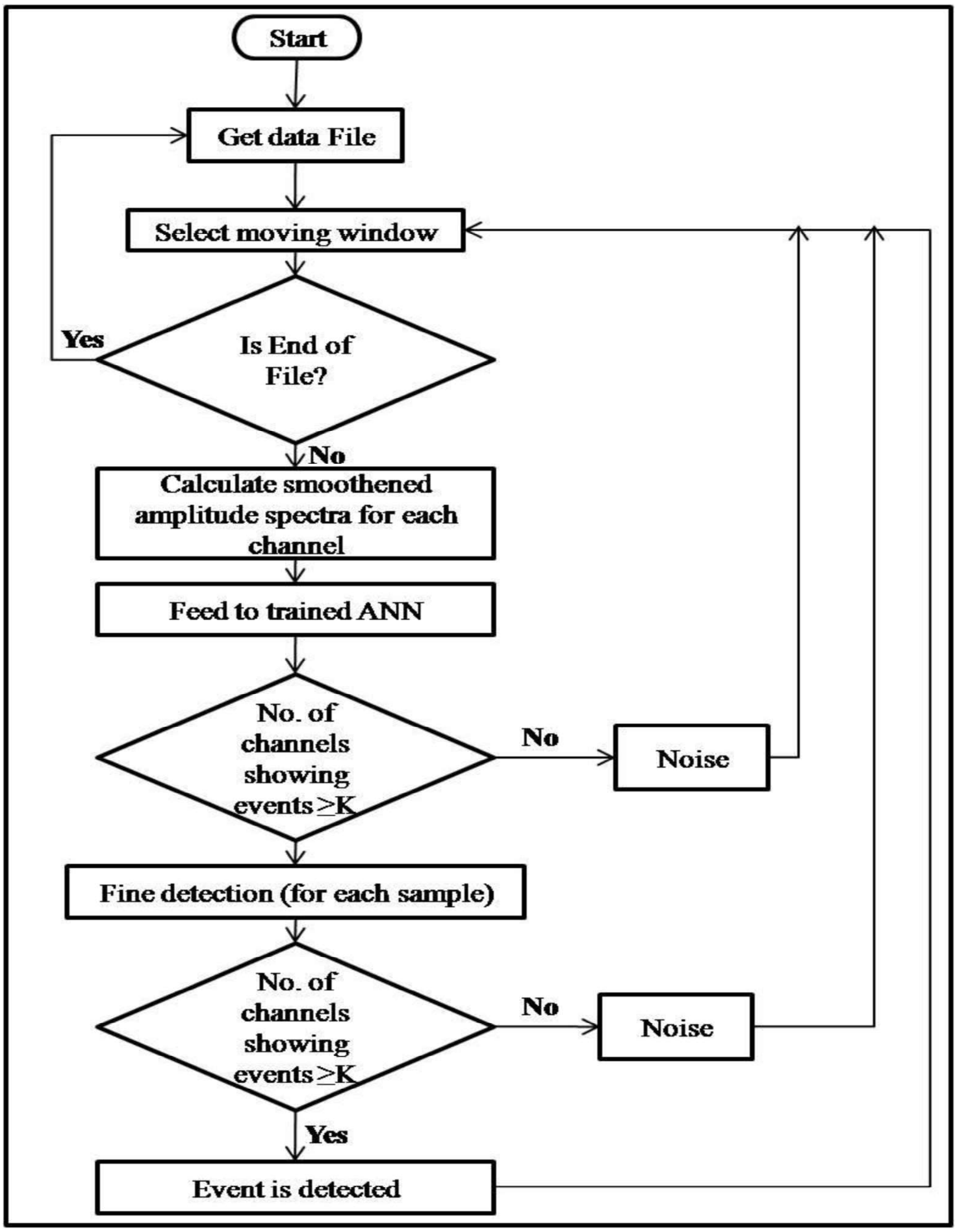


Fig. 3.1: A flow chart for online detection of seismic event using ANN.

collected corresponding to a large number of seismic events which occurred in this region with varying magnitudes. Identifying the event onsets (i.e, onset of P phase) in a good channel, the smoothened amplitude spectra at event and noise windows could be computed using spectral technique such as STFT (Short-Time Fourier Transform). The width of the window for spectral analysis plays an important role. It actually reveals the spectral content of the analysed window. It is determined by the dominant period of the phases being analysed. Here, the window length needs to be selected, such that it accommodates the onset times at two farthest channels along each of the arms of L shaped array corresponding to maximum lag. To map the spectral amplitudes with event or noise category, a machine learning algorithm such as ANN (Artificial Neural Network) may be used. For this, an ANN has to be trained with spectral amplitudes as inputs and event/noise (i.e. binary decision) as desired output for large number of events. After training, the trained ANN could be used to find the onset of new event online in a manner as shown as shown in Fig. 3.1. Getting the data file, smoothened spectral amplitudes will be computed in a predefined time window for all the channels of the seismic array. These spectral amplitudes would then be fed to the trained ANN to obtain decisions corresponding to all the channels. If certain number say, K out of all the decisions show event for the given window it would be considered as a seismic event, otherwise it would be classified as noise. Once an event is declared, the starting time of the window, called the preliminary onset is noted and the procedure for fine detection is started over a data block starting from preliminary onset minus one second to preliminary onset plus 3 sec. Within this data block corresponding to each channel, amplitude spectra will be computed for the same time window which will slide in steps of one sample to find the sample number at which

Chapter-3

Proposed Method of Identification of Tsunamigenic Earthquakes

It is proposed that the tsunamigenic earthquakes could be identified reliably using broad band data of a single 3-component station located at a regional distance from the source. This process involves four steps, namely, detection, location, magnitude estimation and identification of tsunamigenic earthquakes. The proposed method has been developed for poorly instrumented region of the world where multi-station data may not be available. Since detection and location of seismic events using seismic arrays are generally superior to that using single 3-component stations the proposed method introduces seismic array (preferably L shaped and short period) based detection and location methods in sections 3.1 and 3.2 respectively. The methods of magnitude estimation and tsunamigenic earthquake identification which are based on single 3-component broad band records are presented in sections 3.3 and 3.4 respectively.

3.1Method of detection of seismic event

Seismic signal has a specific signature on the seismogram recorded at a given station depending upon its geographical origin and wave propagation path. This signature is different from noise to event. By analyzing the change in signal characteristics of noise and event, the occurrence of seismic event may be detected.

Let us consider a circular geographical region with center at seismic array and radius much more than regional distance [2]. The data recorded on the array could be station at a regional distance from the earthquake source. Comparison of the above results shows that the results obtained by ANN-I are marginally better than those obtained using ANN-II and SVM. The close agreement among the three mapping methods indicates that mapping between rms seismic phase amplitudes (along with other parameters such as location and magnitude) and binary decision by ANN-I is reliable based on which a single station based tsunami alert system could be achieved reasonably. It may be noted that input vector is available for use immediately after the arrival of LR phase. Nominal computational load of ANN enables quick analysis of the event. Promptness (within 5 min after the arrival of LR phase) of the proposed technique makes it a very useful choice as an initial tsunami estimator for full-fledged tsunami alert systems. The methodology has been tested with the seismograms recorded at IRIS station, PALK, for categorizing earthquakes originating from Sumatra region with epicentral distance from 10.62^0 to 37.63^0. This method would prove to be extremely useful for the regions that are not adequately instrumented for azimuthal coverage.

linearly separable using kernel function. However, ANN employs multi-layer connection and various activation functions to deal with non-linear problems. In fact, single layer ANN can only generate linear boundaries, and the 2nd layer can combine the linear boundaries together; while at least three layers are required to produce boundary of arbitrary shapes. Using the gradient descent learning algorithm, ANN tends to converge to local minima. As a result, it suffers from the over-fitting problem. On the other hand, SVM tends to find a global solution during the training as the model complexity has been taken into consideration as a structural risk in SVM training. In other words, ANN minimizes only the empirical risk learnt from the training samples, but SVM considers both empirical risk and the structural risk. Consequently, the training results from SVM have better generalization capability than those from ANN. However ANN outperforms SVM when there is learning from imbalanced datasets with more data samples from one class than other. The present study comes under that learning where the ratio of data samples corresponding to non-tsunamigenic and tsunamigenic categories is of about 8:1.

4.4Summary

Two ANN based direct mapping algorithms have been studied for detection of seismic event using short period data and rapid estimation of M_w of earthquakes using broad band data. The study establishes that onset detection and M_w estimation are fairly accurate. Knowing the epicentral location (estimated by conventional method using short period data) and M_w, the tsunamigenic potential of the earthquake has been studied using another ANN (i.e. ANN-I) along with two alternative mapping techniques (ANN-II and SVM) based on the recordings of a single 3-component empirical Tsunami-Volume scale as stated below may be formulated:

$$\begin{cases} V_T < 1.10\times10^8 : NT \\ 1.10\times10^8 \le V_T \le 1.10\times10^9 : NT/T \\ V_T > 1.10\times10^9 : T \end{cases}$$

Using the estimated P and M_w, V_T and CAT have been computed using Eq.(3.36) and Tsunami-Volume scale . These are listed in last two columns of Table 4.5. A comparison of estimated CAT (column 10 of Table 4.5) with those reported by USGS (column 4 of Table 4.5) shows 95% (19/20) agreement between the estimated and the actual CAT which is noteworthy.

4.3.3 Estimation of earthquake category using SVM

SVM has also been trained with the same 46 earthquakes that were used to train ANN-I and ANN-II. The same input parameters viz. rms phase amplitudes, epicentral distance, back azimuth and moment magnitude were used to train SVM. The desired outputs in this case also were NT or T. The trained SVM has been tested with 22 events (marked * and ** in Table A.2). The categories obtained by SVM for the test dataset are compared with those reported by USGS in Table 4.5. The comparison shows that SVM has successfully identified about 91% (20/22) of the new events. Though SVMs are often considered superior to ANNs in the context of pattern classification over a small data set, the present study shows a marginal deviation from the general trend. It may be possible that ANN outperforms SVM particularly for the present set of data under consideration.

The two different algorithms, SVM and ANN share the same concept of linear learning model for pattern recognition. The difference is mainly on how non-linear data is classified. Basically, SVM utilizes non-linear mapping to make the data

Water volumes of 66 earthquakes (both tsunamigenic as well as non-tsunamigenic) of magnitude 6 and above that occurred in Sumatra region during the period 2001– 2016 (Table A.2) have been estimated using Eq. (3.36) based on the parameters reported by USGS. These volumes were plotted against magnitudes as shown in Fig. 4.23.

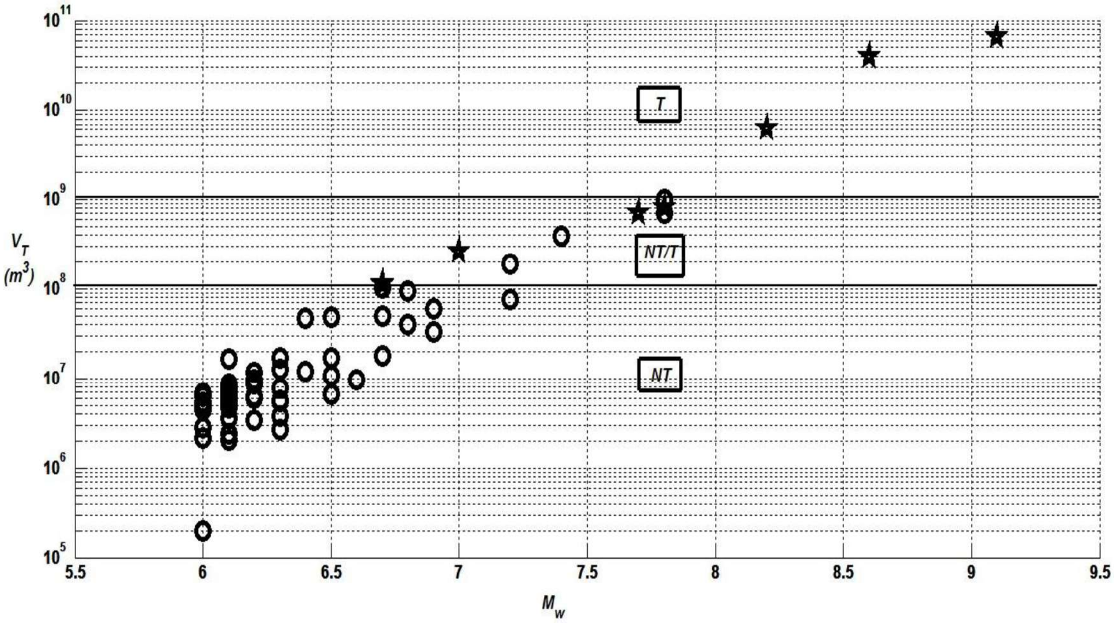


Fig. 4.23: Plot of water volumes vs. moment magnitude. T corresponds to Tsunamigenic (pentagram) and NT to Non-Tsuanmigenic (circle). Two horizontal dark lines indicate region of uncertainty.

It may be noted that in Fig. 4.23 an earthquake which is reported to have generated a tsunami (T) has been marked as pentagram while those which have not generated tsunami (NT) are marked as circles. The area between the two horizontal bars in the middle of Fig. 4.23 is the area of uncertainty limited by minimum and maximum volumes corresponding to which a tsunami is reported and not reported, respectively. Thus the values of water volumes lying between these two extremes correspond to uncertain category of the earthquake. Fig. 4.23 clearly indicates that there exists a relation between the earthquake category (CAT) and (V_T). Hence an back azimuth and moment magnitude were used to train ANN-II. The desired outputs in this case were ϕ_s, δ and λ which are also listed in Table A.1. ANN-II has been tested with the test set containing 20 earthquakes (marked * in Table A.2). The focal parameters obtained using ANN are used to compute the product P (i.e., $\sin(\delta)\sin(\lambda)$) which are listed in Table 4.5 along with those obtained using focal parameters reported by USGS. A histogram of number of events against absolute error between USGS reported P and ANN-II estimated P is plotted in Fig. 4.22.

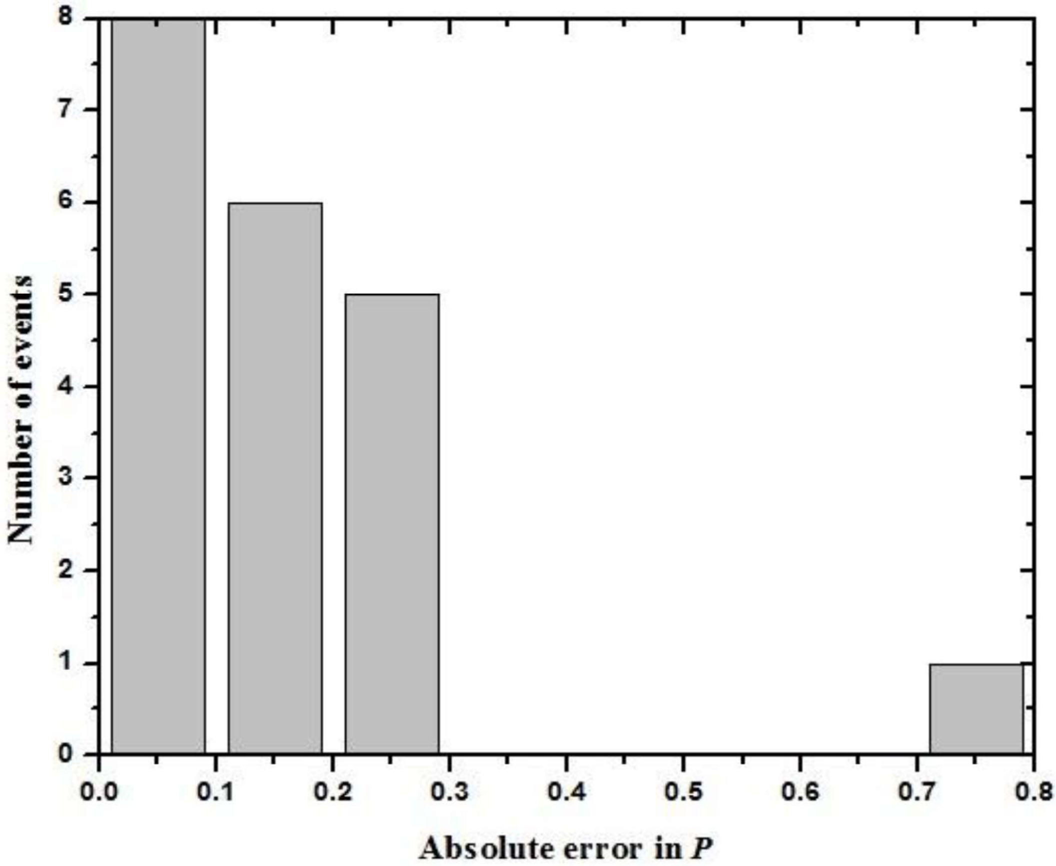


Fig. 4.22: Plot of number of events vs. absolute error in product term.

The figure shows that the number of events rapidly reduces as the absolute error in P increases. This clearly indicates that the estimation of P which in turn is used to estimate water volume is unbiased and reliable.

rate is 100% and false negative rate is 0% for both the classes. Overall, 100% of the predictions are correct and 0% is wrong.

The confusion matrix for the test dataset (NT=19 and T=3) is shown in Fig. 4.21. The confusion matrix shows that overall 100% of the predictions are correct. This result is in fair agreement with that obtained by training ANN with resilient propagation algorithm. It indicates that the ANN-I based method of identifying tsunamigenic earthquake is highly accurate enough.

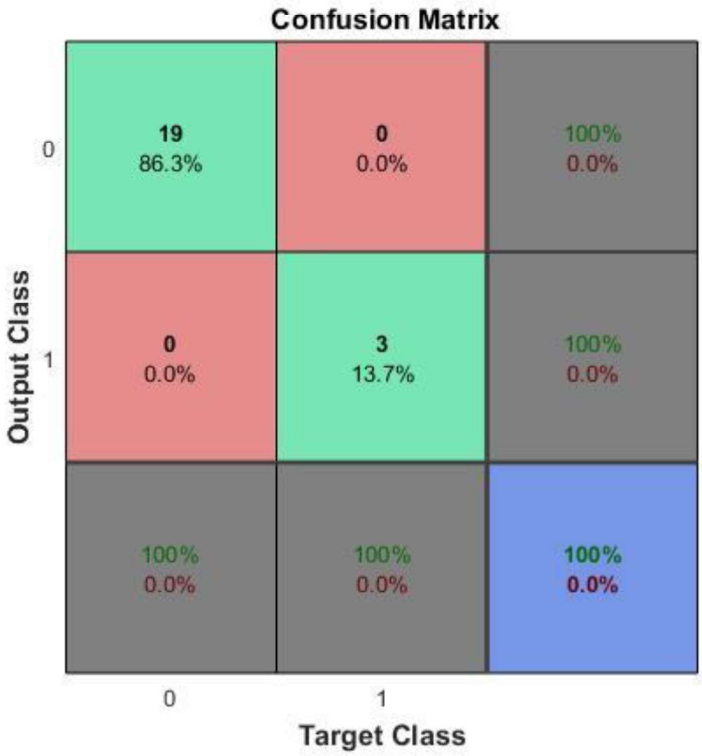


Fig. 4.21: Confusion ` matrix for the test dataset in identifying tsunamigenic earthquake**.**

4.3.2 Estimation of earthquake category from water volumes derived using focal parameters estimated by ANN-II

ANN-II has also been trained with the same 46 earthquakes that were used to train ANN-I. The input parameters viz. rms phase amplitudes, epicentral distance,

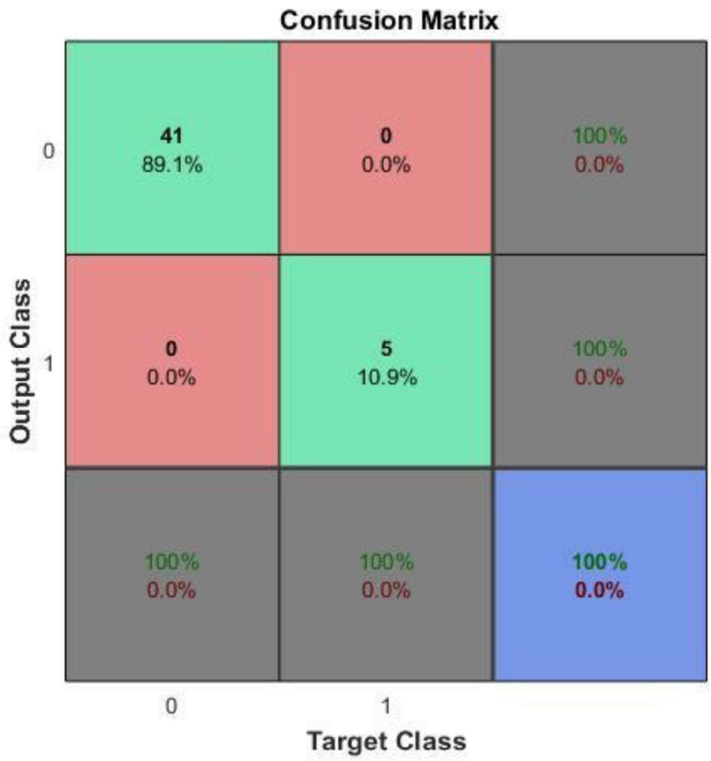


Fig. 4.20: Confusion matrix for the training data set in identifying tsunamigenic earthquake**.**

In Fig.4.20, the first two diagonal cells (green) show the number and percentage of correct classifications by the trained network. In this case, 41earthquakes are correctly classified as non-tsunamigenic. This corresponds to 89.1% of all 46 earthquakes. Similarly, 5 cases are correctly classified as tsunamigenic. This corresponds to 10.9% of all earthquakes. The off-diagonal cells (pink) correspond to incorrectly classified observations. Both the number of observations and the percentage of the total number of observations are shown in each cell. In this case, 0 of the tsunamigenic earthquakes are incorrectly classified as non-tsunamigenic and this corresponds to no error for all 46 earthquakes in the data. Similarly, 0 of the non-tsunamigenic earthquakes are incorrectly classified as tsunamigenic and this corresponds to 0% of all data. The precision rate is 100% and false discovery rate is 0% for both the classes. The recall

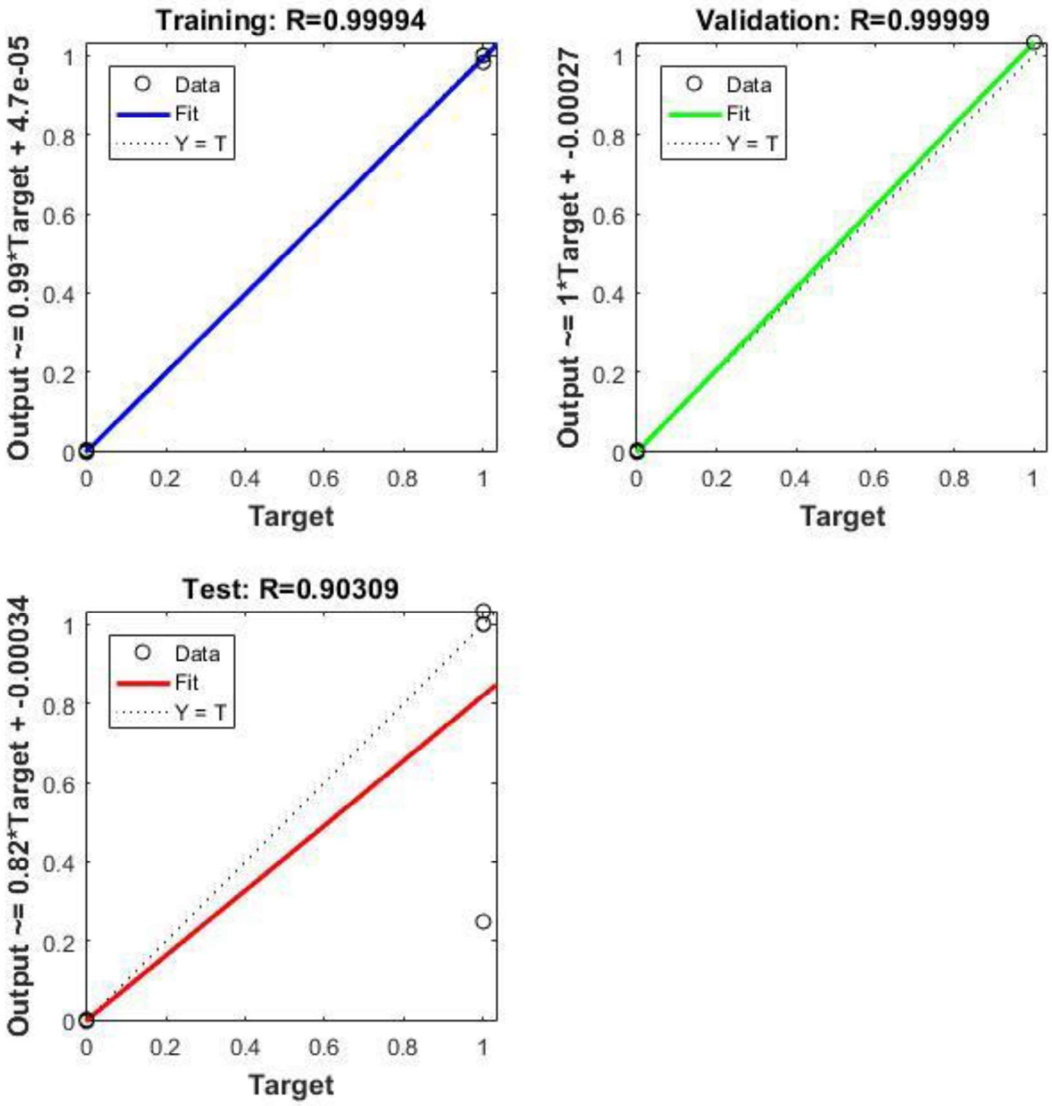


Fig. 4.19: Linear regression plots during training of ANN in tsunamigenic earthquake identification**..**

The confusion matrix for the training set (Non-tsunamigenic events (NT) = 41 and Tsunamigenic events (T) = 5) is shown in Fig. 4.20. Here 0 represents NT and 1 reprsents T.

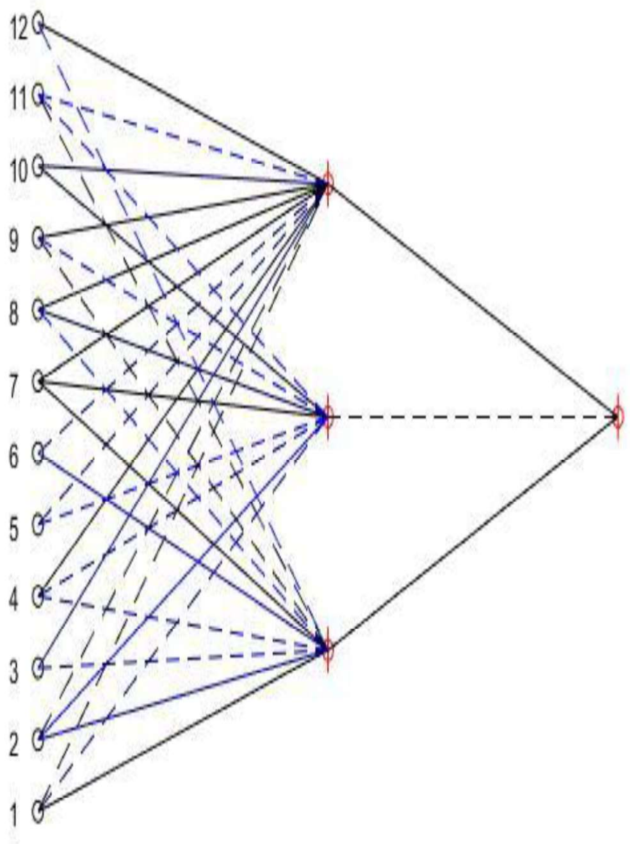


Fig. 4.18: Configuration of the pruned network for tsunamigenic earthquake identification**.**

The linear regression plots of output vs target for training data (46 in no.) which is divided into three sets, namely, Training (70%), Validation (15%) and Test (15%), are shown in Fig. 4.19.

Table 4.5: Comparison among USGS reported *CAT* and that estimated using ANN-I, ANN-II and SVM

S.No.	Evt. No.	M_w (USGS)	*CAT* (USGS)	*CAT* (ANN-I)	*CAT* (SVM)	P (USGS)	P (ANN-II)	V_T (ANN-II) m^3	*CAT* (ANN-II)
1	1	6.5	*NT*	*NT*	*NT*	--	--	--	--
2	2	6.6	*NT*	*NT*	*T*	--	--	--	--
3	3	6.1	*NT*	*NT*	*NT*	0.453	0.441	7.49×10^6	*NT*
4	5	6.1	*NT*	*NT*	*NT*	0.28	0.205	3.47×10^6	*NT*
5	7	6.2	*NT*	*NT*	*NT*	0.142	0.128	3.05×10^6	*NT*
6	10	8.6	*T*	*T*	*T*	0.438	0.606	5.54×10^{10}	*T*
7	13	6.1	*NT*	*NT*	*NT*	0.407	0.51	8.65×10^6	*NT*
8	14	6.7	*NT*	*NT*	*NT*	0.769	0,682	9.10×10^7	*NT*
9	32	6	*NT*	*NT*	*NT*	0.444	0.171	2.06×10^6	*NT*
10	38	7.7	*T*	*T*	*NT*	0.17	0.234	9.69×10^8	*NT/T*
11	39	6.8	*NT*	*NT*	*NT*	0.51	0.23	4.32×10^7	*NT*
12	42	6.5	*NT*	*NT*	*NT*	0.253	0.24	1.61×10^7	*NT*
13	43	6.7	*NT*	*NT*	*NT*	0.133	0.24	3.19×10^7	*NT*
14	44	6.1	*NT*	*NT*	*NT*	0.324	0.333	5.64×10^6	*NT*
15	45	6.9	*NT*	*NT*	*NT*	0.126	0.24	6.35×10^7	*NT*
16	50	6.4	*NT*	*NT*	*NT*	0.253	0.24	1.14×10^7	*NT*
17	52	6.4	*NT*	*NT*	*NT*	0.983	0.244	1.16×10^7	*NT*
18	53	6.5	*NT*	*NT*	*NT*	0.729	0.45	3.02×10^7	*NT*
19	55	6.1	*NT*	*NT*	*NT*	0.215	0.361	6.12×10^6	*NT*
20	57	6	*NT*	*NT*	*NT*	0.234	0.442	5.32×10^6	*NT*
21	60	6.2	*NT*	*NT*	*NT*	0.49	0.25	5.99×10^6	*NT*
22	63	9.1	*T*	*T*	*T*	0.131	0.234	1.19×10^{11}	*T*

Additionally, ANN-I has also been trained with the same inputs using Levenberg-Marquardt algorithm. In this case the tangent hyperbolic activation function is used in the hidden layer and linear one is used in the output layer. For this task the network has been pruned for 46 training data with 12 input neurons, 3 hidden neurons and one output neuron. The configuration of the pruned network has been shown in Fig. 4.18. Here, insignificant inputs are partially connected with hidden neurons, whereas significant inputs are fully connected.

regularization. On the other hand SVM for estimating earthquake category has been designed to have 12 input data points with polynomial kernel of degree 6. The training of SVM involves solving a quadratic optimization problem. The training and testing of ANNs and SVM and their outputs are discussed below.

4.3.1 Estimation of earthquake category using ANN-I

ANN-I has been trained with 46 earthquakes. The rms amplitudes of the clearly identifiable seismic phases P, S and LR along with epicentral distance, back azimuth and moment magnitude were fed as input to ANN-I for training. The desired output is binary decision i.e. either 0 (for NT) or 1 (for T). After training, the trained ANN-I has been tested with test dataset of 22 events (marked * and ** in TableA.2). The *CAT* obtained by ANN-I for the test data-set are compared with those reported by USGS in Table 4.5. The comparison shows that ANN-I has successfully identified 100% (22/22) of the new events.

against only those earthquakes which have generated noticeable tsunami at a land mass. Remaining events in Table A.2 have been marked by us as NT . Randomly chosen 22 events (marked * and **) have been reserved for testing and remaining 46 events have been used for training both ANN-I and SVM for estimating the potentially tsunamigenic earthquake category. It may be noted here that only 20 events (marked * only) has been used for testing of ANN-II for focal parameter estimation because 2 events (marked **) have been included later. Both ANNs and SVM have been trained with rms amplitudes of seismic phases P, S and LR along with the magnitude and location parameters (represented by distance and back azimuth) of events listed in Table A.2.For the source-station pair under consideration, these three phases are unambiguously identifiable on the seismograms recorded at PALK (see Fig. 4.13). The moment magnitude and location parameters have been considered as inputs to take into account the effect of path on the phase amplitudes. The window lengths from onsets for the computation of rms amplitudes of P, S and LR phases have been taken as 5, 10 and 120 s respectively. These window lengths have been chosen to invariably include the maximum peak amplitudes of these phases. ANN-I for estimating earthquake category has been designed to have input layer with 12 neurons, one hidden layer with 20 neurons and one output layer with single neuron while ANN-II for estimating focal parameters and in turn water volumes has input layer with 12 neurons, one hidden layer with 30 neurons and output layer with three neurons. The transfer function used in the hidden layer is sigmoid and that used in the output layer is linear. As mentioned earlier, the training algorithm is resilient back-propagation [81]. The learning function is gradient descent with momentum weight and bias. The performance function is the mean squared error with

Fig. 4.17 indicates that the error in M_w estimated by ANN trained with Levenberg-Marquardt algorithm is in good agreement with that obtained by ANN trained with resilient propagation algorithm. This proves that the ANN based mapping between spectral amplitude and moment magnitude is highly reliable.

4.3 Identification of tsunamigenic earthquake using PALK data

Within the same geographical area (as mentioned in section 4.2) there were a total of 119 shallow focus earthquakes as reported by USGS during 16 Jan 2001 to 6 Dec 2016 with magnitudes 6 and above. Out of these, epicenters of 10 earthquakes were found to be on land far from the sea, 4 were with no moment tensor parameters mentioned in the database and 37 were not recorded properly at PALK station. That leaves us with 68 shallow focus (depth < 70 km) earthquakes that have been considered to demonstrate the efficacy of the methodology proposed in this work. Out of 68, only 8 events were reported to have generated tsunami. Broad-band seismograms recorded at PALK station (7.273^0 N, 80.702^0 E), corresponding to these events have been downloaded from IRIS website (www.iris.edu) and are listed in Table A.2. The location, magnitude and focal parameters of the events, as reported by USGS are also listed in Table A.2. A plot of epicenters of these events on geographical map of Sumatra region is shown in Fig. 4.12. It is seen from Table A.2 that epicentral distances of these events range from 10.62^0 (~1180 km) to 37.63^0 (~4184 km) while the station-source azimuths (i.e., back azimuth) range from 97.1^0 to 131.57^0 with respect to PALK station. The entries in CAT column in Table A.2 have been marked as T (for Tsunamigenic) and NT (for Non-Tsunamigenic) based on the reports published by USGS. It may be noted that USGS generally writes Tsunami

Table 4.4: Comparison between USGS reported M_w and that estimated by ANN trained with Levenberg-Marquardt algorithm.

S.No.	Evt.No.	Mw(USGS)	M_w (ANN)	Absolute error in M_w
1	2	6.6	6.5	0.1
2	10	6.2	6.3	0.1
3	15	6.7	6.9	0.2
4	16	6.1	6.3	0.2
5	19	6	5.8	0.2
6	21	7.8	7.5	0.3
7	22	7.2	7.3	0.1
8	29	7	7.1	0.1
9	32	6	6.1	0.1
10	42	6	6.1	0.1
11	43	6.1	6.1	0.0
12	46	6.2	6.1	0.1
13	49	6.1	6.2	0.1
14	51	6.3	6.1	0.2
15	54	6.2	6.1	0.1
16	56	6.7	6.3	0.4
17	57	6.1	6.1	0.0
18	64	6.1	6.2	0.1
19	66	6.5	6.2	0.3
20	71	6.3	6.1	0.2
21	82	6.3	6.1	0.2
22	84	6.9	6.7	0.2

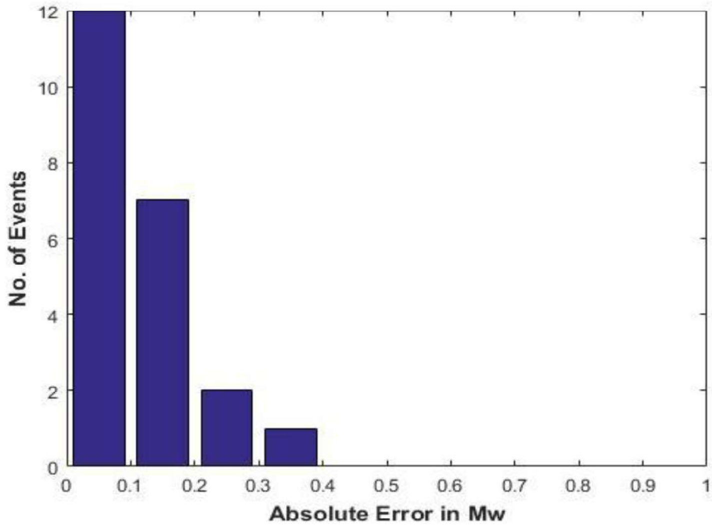


Fig. 4.17: Plot of number of events vs. absolute error in M_w estimated by ANN trained with Levenberg-Marquardt algorithm.

The linear regression plots of output vs target for training data (62 in no.) which is divided into three sets, namely, Training (70%), Validation (15%) and Test (15%), are shown in Fig. 4.16.

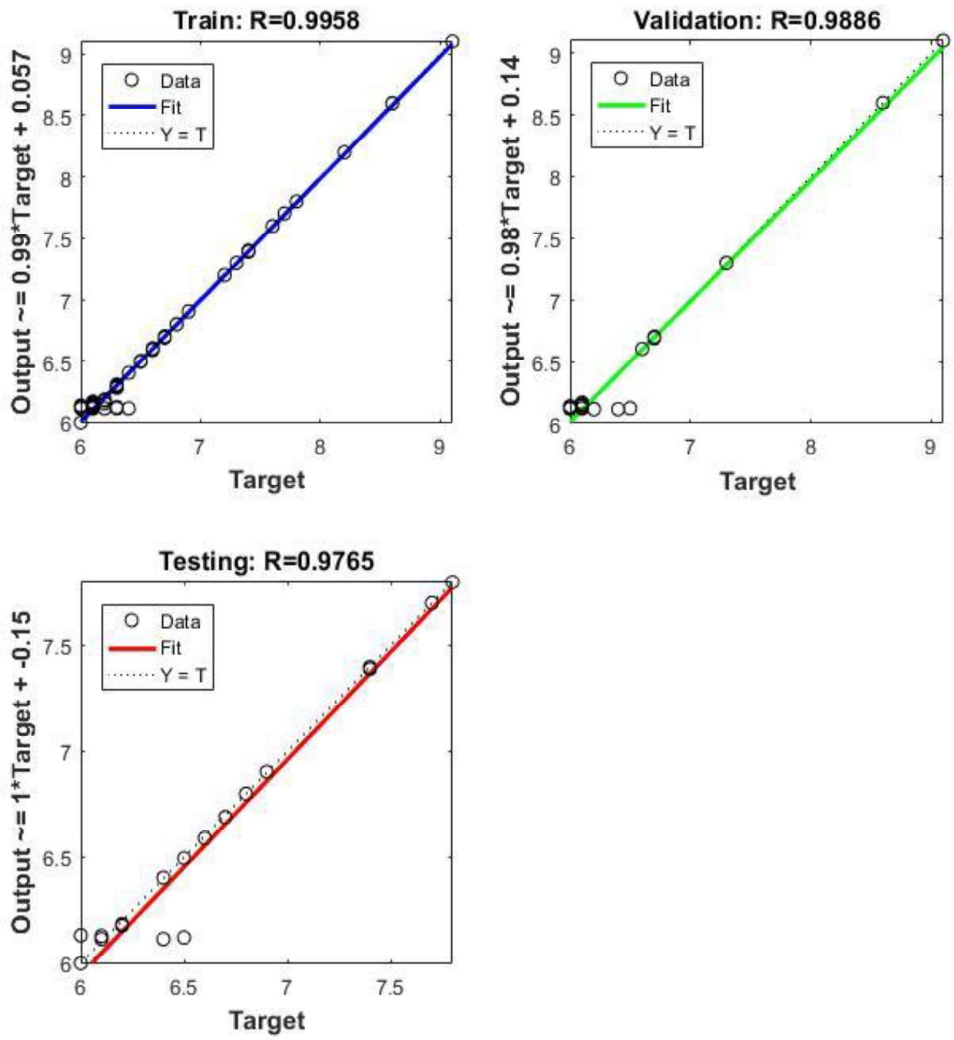


Fig. 4.16: Linear regression plots during training of ANN in magnitude estimation.

The M_w obtained by the ANN for the test dataset are compared with those reported by USGS in Table 4.4. A histogram of number of events against absolute error between USGS reported M_w and that estimated using the ANN is plotted in Fig. 4.17.

To check the reliability of the proposed method of magnitude estimation, the ANN has also been trained the Levenberg-Marquardt algorithm. In this case the tangent hyperbolic activation function is used in the hidden layer and linear one is used in the output layer. Here, instead of 147 values, only 19 values of amplitude spectrum in the frequency range from 0.0098 to 10 Hz have been used as inputs to ANN. For this task the network has been pruned for 62 training data with 19 input neurons, 2 hidden neurons and one output neuron. The configuration of the pruned network has been shown in Fig. 4.8. Here, insignificant inputs are partially connected with hidden neurons, whereas significant inputs are fully connected.

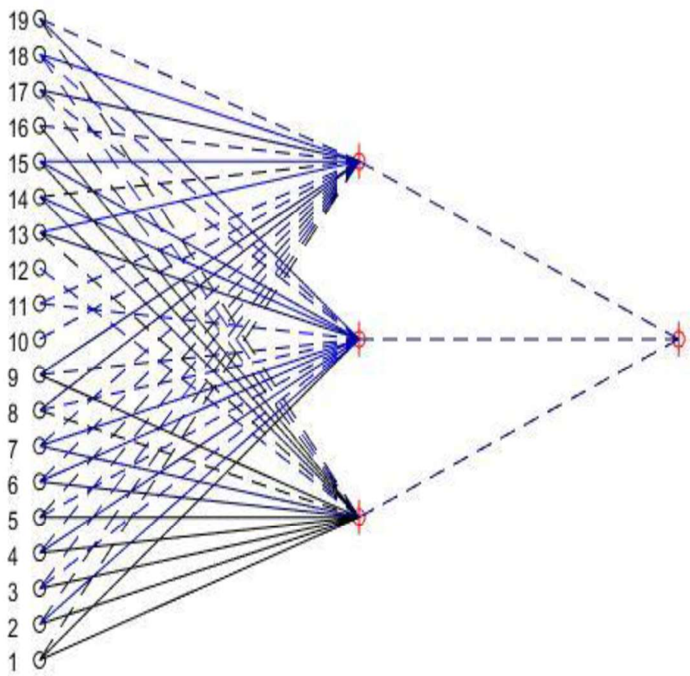


Fig. 4.15: Configuration of the pruned network for magnitude estimation.

Table 4.3: Comparison between USGS reported M_w and that estimated by ANN

S.No.	Evt.No.	M_w (USGS)	M_w (ANN)	Absolute error in M_W
1	2	6.6	6.7484	0.1484
2	10	6.2	6.2839	0.0839
3	15	6.7	6.7345	0.0345
4	16	6.1	6.3339	0.2339
5	19	6	6.3275	0.3275
6	21	7.8	7.7070	0.0930
7	22	7.2	7.0430	0.1570
8	29	7	7.0972	0.0972
9	32	6	6.0992	0.0992
10	42	6	6.2832	0.2832
11	43	6.1	6.0074	0.0926
12	46	6.2	6.1392	0.0608
13	49	6.1	6.2768	0.1768
14	51	6.3	6.2474	0.0526
15	54	6.2	6.1901	0.0099
16	56	6.7	6.7621	0.0621
17	57	6.1	6.0178	0.0822
18	64	6.1	6.0994	0.0006
19	66	6.5	6.5403	0.0403
20	71	6.3	6.2269	0.0731
21	82	6.3	6.3418	0.0418
22	84	6.9	6.828	0.0720

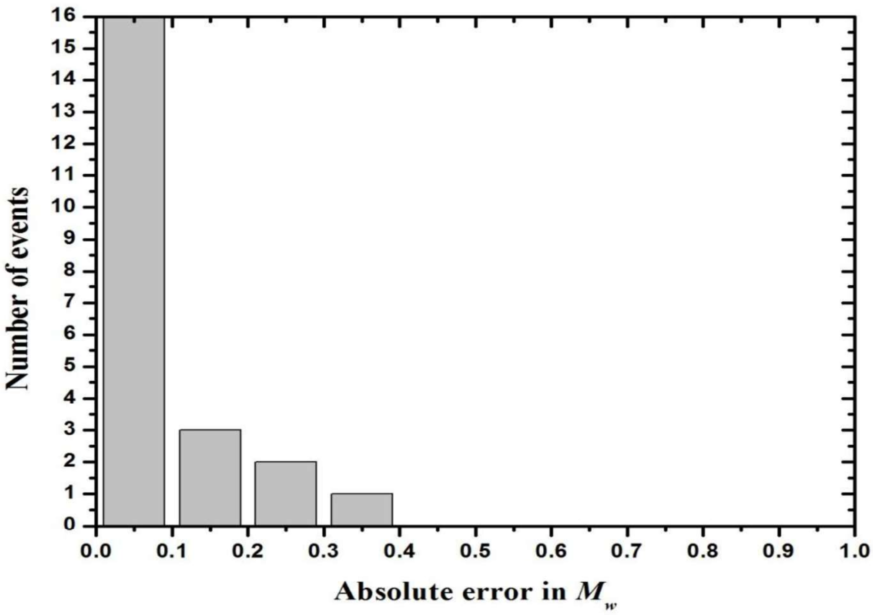


Fig. 4.14: Plot of number of events vs. absolute error in estimated M_w.

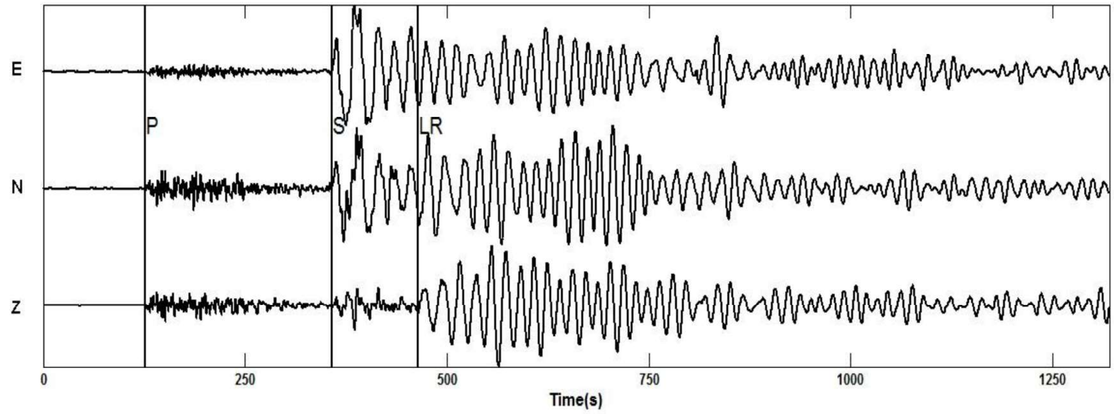


Fig. 4.13: Plot of typical 3-component (E, N and Z) seismogram with the onset markings of P, S and LR (surface wave) phases.

The training algorithm is resilient back-propagation [81]. The learning function is gradient descent with momentum weight and bias. The performance function is the mean squared error with regularization. The training, testing and validation results of the ANN and its outputs are discussed below.

The spectral amplitudes computed at P onset of the events for a window length of 3 s have been fed as inputs to ANN for training. The desired outputs are corresponding M_w. After training, the trained ANN has been tested with test dataset of 22 events (marked * in Table A.1). The M_w obtained by the ANN for the test dataset are compared with those reported by USGS in Table 4.3. A histogram of number of events against absolute error between USGS reported M_w and that estimated using the ANN is plotted in Fig. 4.14. The figure shows that the number of events rapidly decreases with increase of absolute error in M_w. The figure also shows that M_w estimated by the ANN for about 73% of test events deviates from the reference by less than 0.1 and that for 100% of the events by less than 0.35. This clearly indicates that the estimation of M_w is unambiguous and reliable. These results are very significant in the context of tsunami warning.

as reported by USGS are also listed in Table A.1. A plot of epicenters (shown in pentagram) of these events on geographical map of Sumatra region is shown in Fig. 4.12. In Table A.1 randomly chosen 22 events (marked *) have been reserved for testing while remaining 62 events have been used for training the ANN. A typical 3-component seismogram as recorded at PALK station is shown in Fig. 4.13. Vertical component (Z) has been used for computing the amplitude spectrum of P wave in this work.

As discussed in section 3.4, amplitude (displacement) spectra are mapped to M_w using ANN. The input layer of the ANN has 147 neurons corresponding to 147 values of amplitude spectrum in the frequency range from 0.0098 to 10 Hz. The choice of number of hidden layers and the number of neurons in hidden layers depend on the nature of the problem [95]. For the present study the ANN has been designed to have one hidden layer with 20 neurons. The output layer has one neuron. The transfer function used in the hidden layer is sigmoid and that used in the output layer is linear.

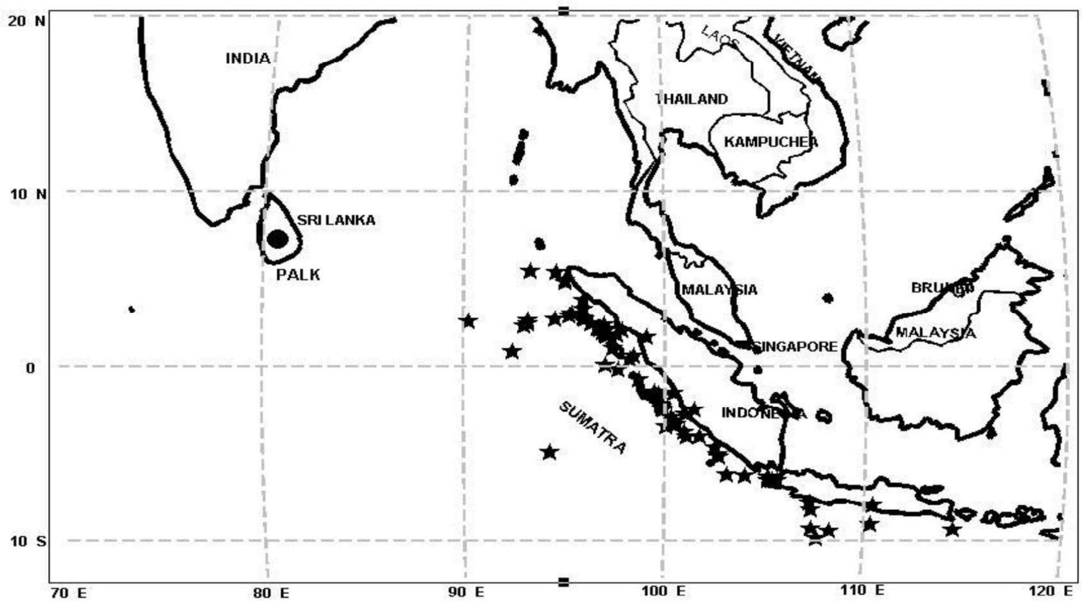


Fig. 4.12: USGS reported epicenters (shown in pentagram) of earthquakes from the study area (Sumatra region) and the location of recording station PALK (black circle).

Fig. 4.11: Confusion matrix of the test data set for seismic event detection.

4.2Moment magnitude estimation using PALK data

For estimation of moment magnitude, the region that has been considered is from 6^0 N to 10^0 S, and 90^0 E to 115^0 E. This area is seismically very active and within this area there had been 129 earthquakes as reported by USGS during 16 Jan 2001 to 19 Oct 2016 with magnitudes 6 and above. Out of these, 45 earthquakes were not recorded properly at PALK station. That leaves us with 84 earthquakes that have been considered to demonstrate the efficiency of the methodology proposed in this work. Broad-band seismograms recorded at PALK station (7.273^0 N, 80.702^0 E), corresponding to these events have been downloaded from IRIS website (www.iris.edu) and are listed in Table A.1. The location and magnitude of the events, shown in each cell. In this case, 5 of the events are incorrectly classified as noises and this corresponds to 1.0% of all 496 signals in the data. Similarly, 4 of the noises are incorrectly classified as events and this corresponds to 0.8% of all data.

The column metrics on the far right (grey) of the plot are often called the precision (or positive predictive value) and false discovery rate, respectively. In this case, out of 248 noise predictions, 98.0% are correct and 2.0% are wrong. Out of 246 event predictions, 98.4% are correct and 1.6% is wrong. Therefore, precisions are 98.0% and 98.4% and false discovery rate are 2.0% and 1.6% for both the classes respectively. The row metrics at the bottom (grey) of the plot are often called the recall (or true positive rate) and false negative rate, respectively. In this case, out of 247 noises, 98.4% are correctly predicted as noises and 1.6% is predicted as event. Out of 247 events, 98.0% are correctly classified as events and 2.0% are classified as noises. Therefore the recall rates are 98.4% and 98.0% and false negative rates are 1.6% and 2.0% for both the classes respectively. The cell in the bottom right (blue) of the plot shows the overall accuracy. In this case, overall, 98.2% of the predictions are correct and 1.8% is wrong.

The confusion matrix for the test set (events=162 and noise=162) is shown in Fig. 4.11. Here 0 represents noise and 1 reprsents events.The confusion matrix shows that overall 97.2% of the predictions are correct and 2.8% of the predictions are wrong. This result is in close agreement with that obtained by training ANN with resilient propagation algorithm. It indicates that the ANN based seismic event detection is accurate enough for prediction.

The confusion matrix for the training set (events=247 and noise=247) is shown in Fig. 4.10. Here 0 represents noise and 1reprsents events.

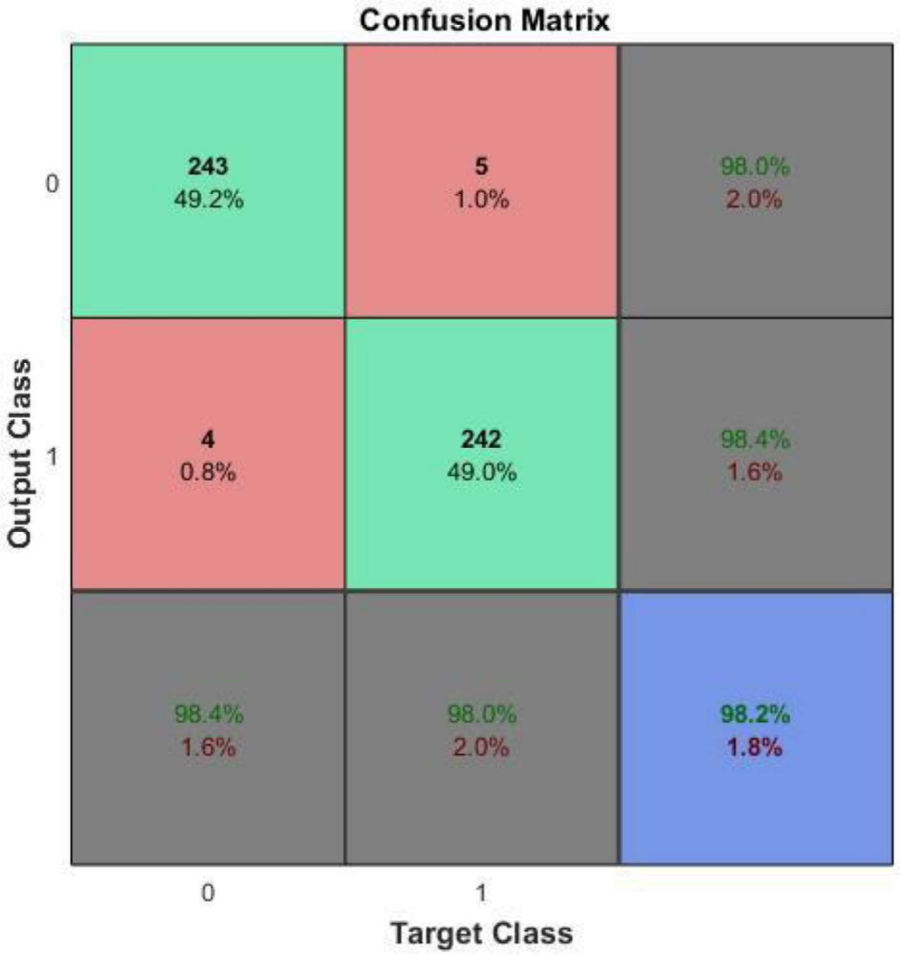


Fig. 4.10: Confusion matrix of the training data set for seismic event detection.

In Fig. 4.10, the first two diagonal cells (green) show the number and percentage of correct classifications by the trained network. In this case, 243 signals are correctly classified as noises. This corresponds to 49.2% of all 494 signals. Similarly, 242 signals are correctly classified as events. This corresponds to 49.0% of all signals. The off-diagonal cells (pink) correspond to incorrectly classified observations. Both the number of observations and the percentage of the total number of observations are

The linear regression plots of output vs target for training data (494 in no.) which is divided into three sets, namely, Training (70%), Validation (15%) and Test (15%), are shown in Fig. 4.9.

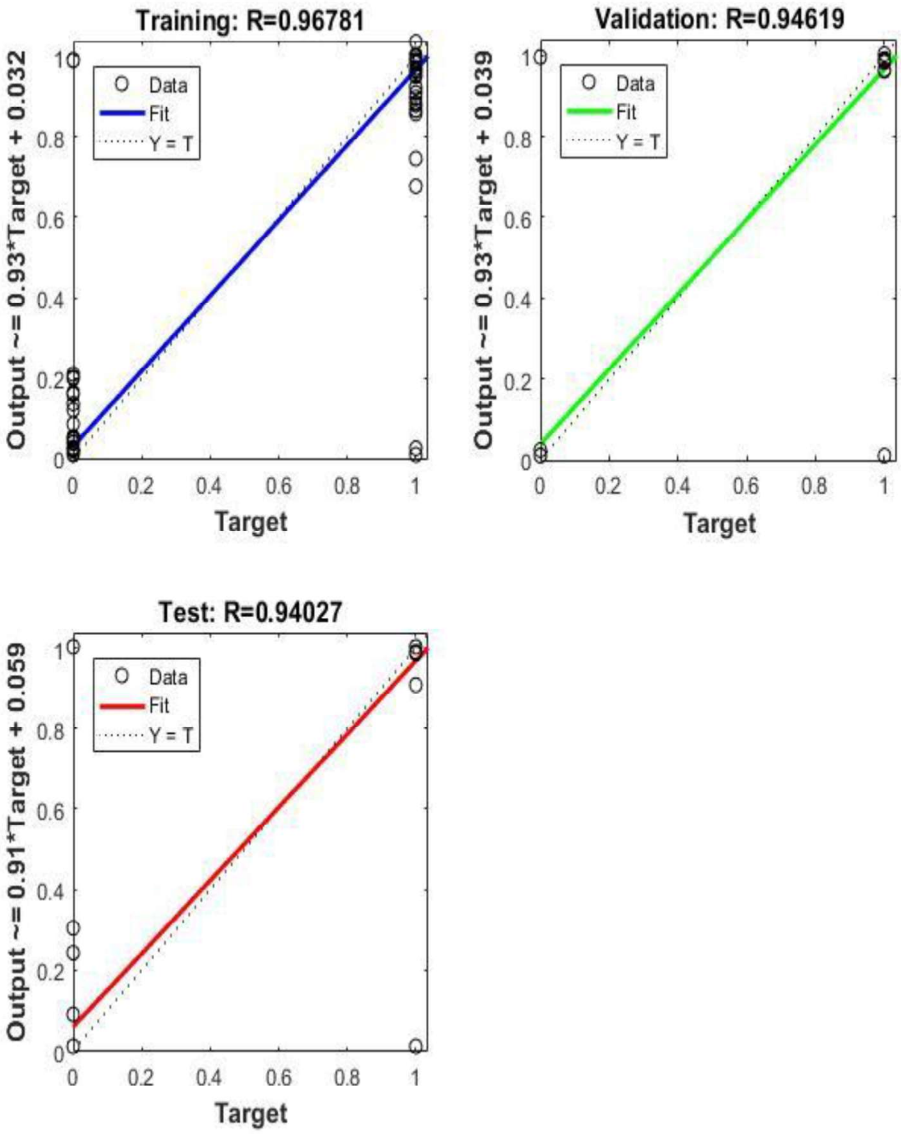


Fig. 4.9: Linear regression plots during training of ANN in seismic event detection.

activation function is used in the hidden layer and linear one is used in the output layer. For this task the network has been pruned for 494 training data (events=247 and noise=247) with 50 input neurons, 2 hidden neurons and one output neuron. The configuration of the pruned network has been shown in Fig. 4.8. Here, insignificant inputs are either completely disconnected or partially connected with hidden neurons, whereas significant inputs are fully connected.

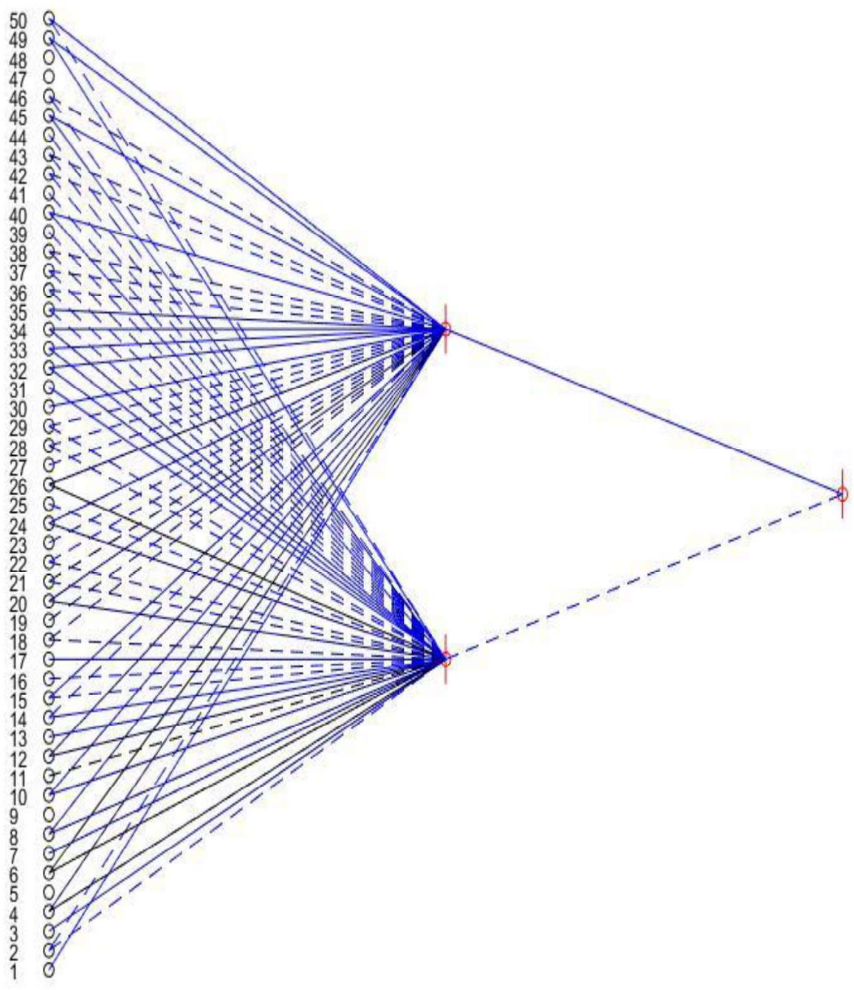
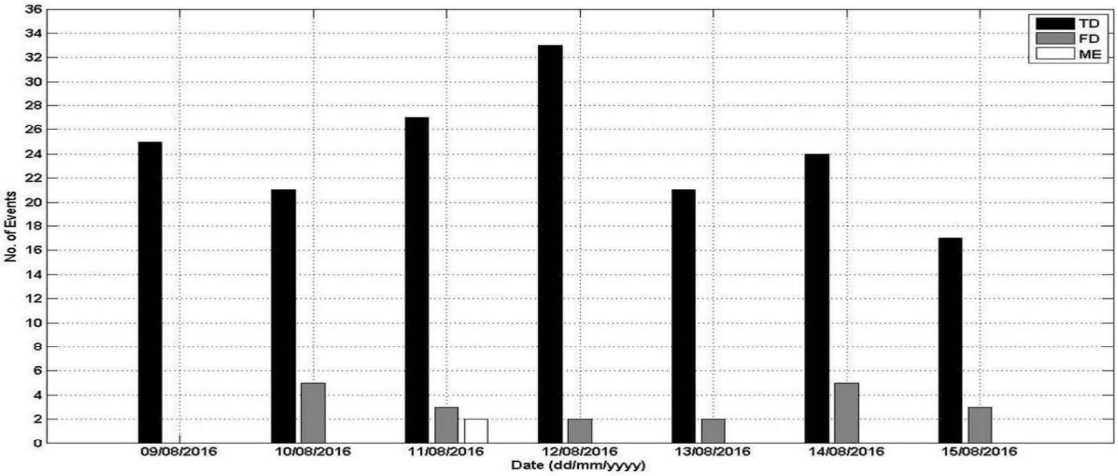
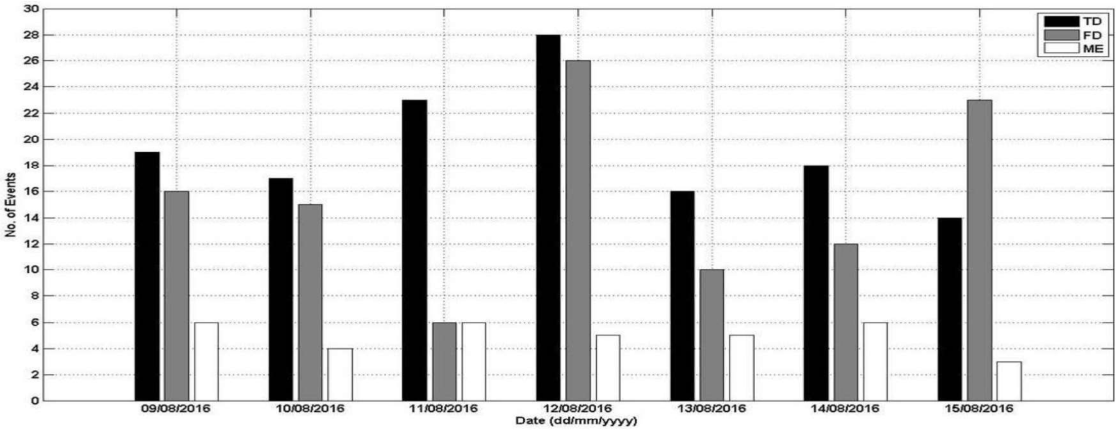


Fig.4.8: Configuration of the pruned network for seismic event detection.



(a)



(b)

Fig. 4.7: Histogram plot of number of events truly detected (TD), missed (ME) and falsely detected (FD) by (a) ANN and (b) STA/LTA through 7 days continuous detection.

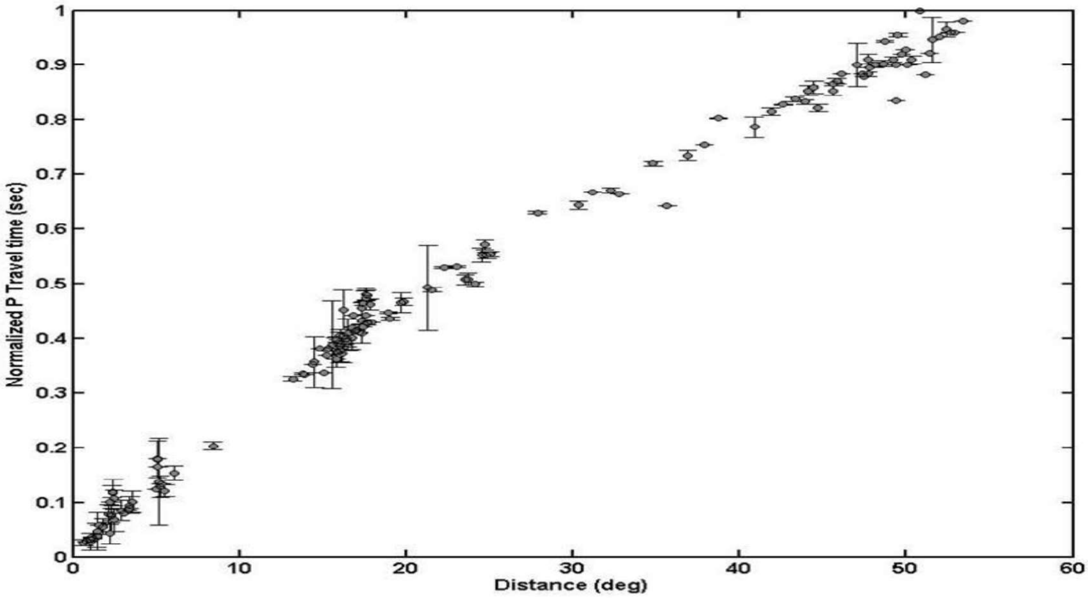
Table4.2.Results obtained by ANN and STA/LTA through 7 days continuous detection.

Methods	TD (%)	ME (%)	FD (per day)
ANN	99	1	3
STA/LTA	79	21	15

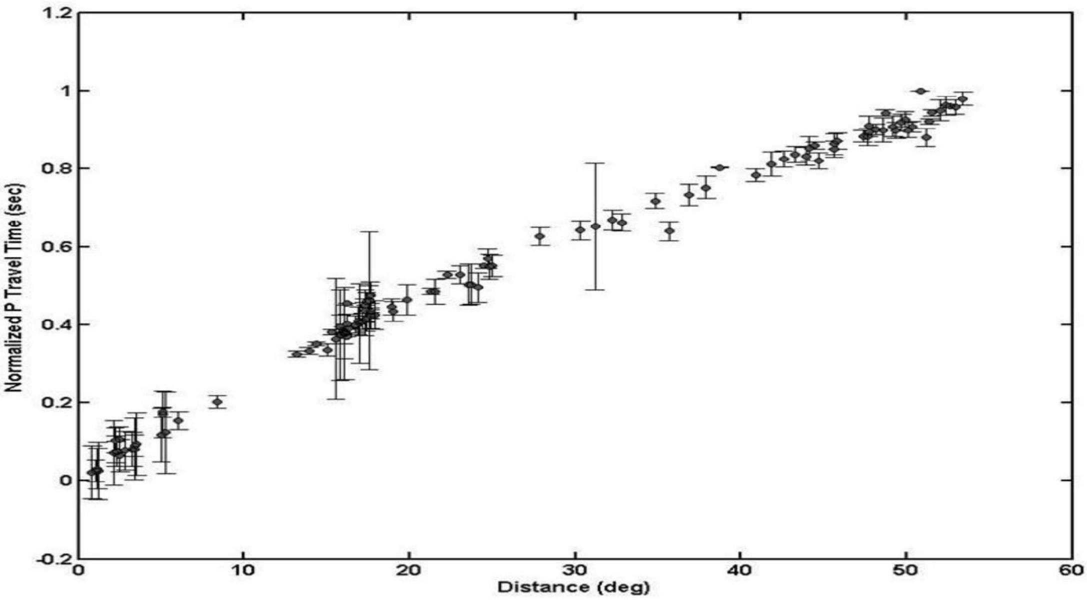
Note: TD refers Truly Detected events, ME represents Missed Events and FD indicates False Detection

Additionally, to check the potential of the proposed method, the seismic event detection, using ANN has also been performed choosing another training algorithm, namely, Levenberg-Marquardt algorithm. In this case the tangent hyperbolic that picked up by analyst. The error bars have further been scaled by a factor of 10 for clear visibility. The average errors in P travel times (and hence the P onsets) using ANN is ± 0.036 sec and whereas that using STA/LTA is ± 1.73 sec. This indicates the superior detection capability as well as detection accuracy of ANN demanding that the proposed technique could be exploited as an online detector.

To ensure the feasibility of the proposed method to exploit as a potential online detector both ANN and STA/LTA algorithm have been tested online with 7 days 09/08/2016 to15/08/2016) continuous data. The results obtained by this continuous detection are shown in Fig. 4.6. The figure shows the histogram plots of number of Truly Detected (TD) events, Missed Events (ME) and Falsely Detected (FD) events against the number of days of detection for detection by ANN and that by STA/LTA respectively. From Fig. 4.7 it is seen that the detection capability of ANN ($\sim 99\%$) is superior to that ($\sim 79\%$) of STA/LTA as listed in Table 4.2. Table 4.2 clearly indicates that the proposed STFT and ANN based method could be used as an efficient online detector. The detection capability (i.e. 84%) of STA/LTA for database of selected events drops to 79% for online test data because P-onsets of large number of events were not very prominent.



(a)



(b)

Fig. 4.6: (a) Plot of Normalized P travel times of 147 events detected by ANN with error bar indicating the deviation with respect to analytical time. The error bars have been computed by taking the difference between normalized travel times obtained by ANN and the normalized travel times obtained by analyst. (b) Plot of Normalized P travel times of 125 events detected by STA/LTA algorithm with error bar indicating the deviation with respect to analytical time. The error bars have been computed by taking the difference between normalized travel times obtained by STA/LTA and the normalized travel times obtained by analyst. The error bars in both cases have been scaled by a factor of 10 for clear visibility.

between the time from the model and that picked by the analyst may be attributed to the difference in geographical structure adopted in the velocity model. Considering 148 events picked up by the analyst as true events, ANN based method has detected 147 events. To detect the same 148 events by STA/LTA method, the setting parameter that has been chosen are listed in Table 4.1.

Table 4.1: Parameters used for the STA/LTA method.

STA/LTA Parameters	Local Events	Regional Events	Teleseismic Events
STA window length (sec)	**3**	**2**	**2**
LTA window length (sec)	**60**	**60**	**60**
STA/LTA trigger on threshold	**1.8**	**2.0**	**3.5**
STA/LTA trigger off threshold	**1.0**	**1.0**	**1.0**
Minimum event duration (sec)	**2**	**2**	**2**

These parameters have been selected in such a way that they give the best balance between the false alarms and missed events. With the best parameter setting (obtained by trial and error) STA/LTA has detected only 125 events. Thus the detection capability of ANN is $\sim 99\%$ (147/148) and that of STA/LTA is $\sim 84\%$ (125/148). To ascertain the accuracy in onset detection, the normalized P travel times computed using both ANN and STA/LTA are plotted against epicentral distance in Fig. 4.5. Here actual travel times are normalized with respect to maximum for better visibility of the error. The error bar in Fig. 4.6 (a) indicates the difference between normalized P travel times as obtained by ANN and that picked up by analyst. The error bar in Fig. 4.6 (b) represents the difference between P travel times obtained by STA/LTA and trained ANN has been used to detect events with value of K (minimum number of channels required for showing decision as event) equal to 13. This value has been chosen by trial and error to achieve the best balance between the false alarms and missed events. The trained ANN has been used to detect seismic events as given below.

P onsets of 162 test events have been computed theoretically using AK135 velocity model [94]. Out of 162 test events, analyst (duty scientist) has picked up 148 events correctly. To know the accuracy in picking up of P onsets by an analyst, the analyzed P wave travel times are plotted with those obtained from model against epicentral distance for these 148 events as shown in Fig. 4.5.

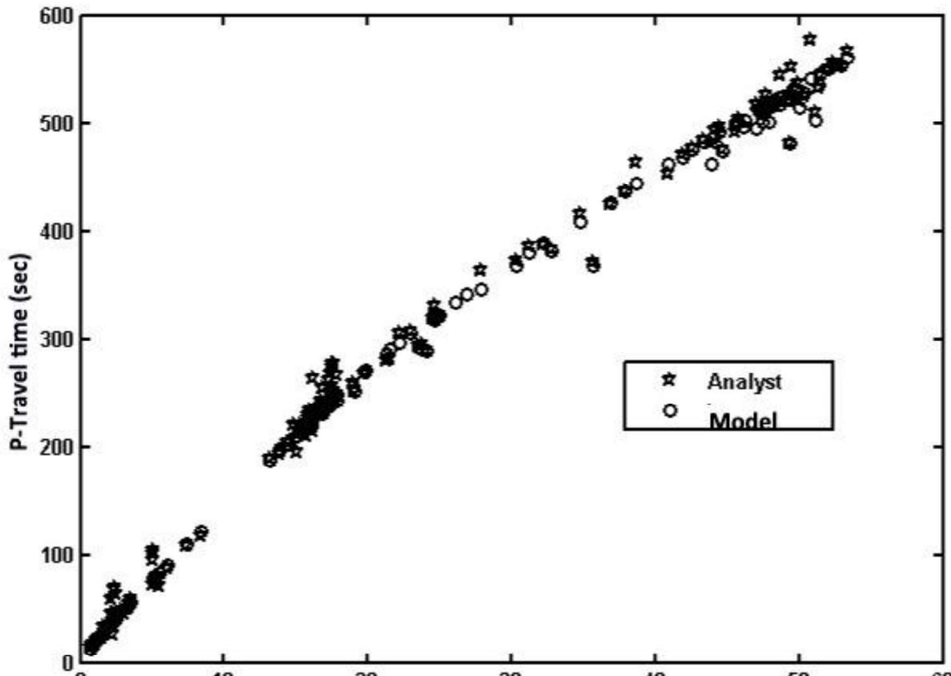


Fig.4.5: Comparison of P travel times computed using AK135 velocity model and that picked up by analyst for 148 events.

The average of deviations between P travel times obtained from the model and that by analyst is ± 4.26 sec. Assuming, the analyzed times as standard, this difference

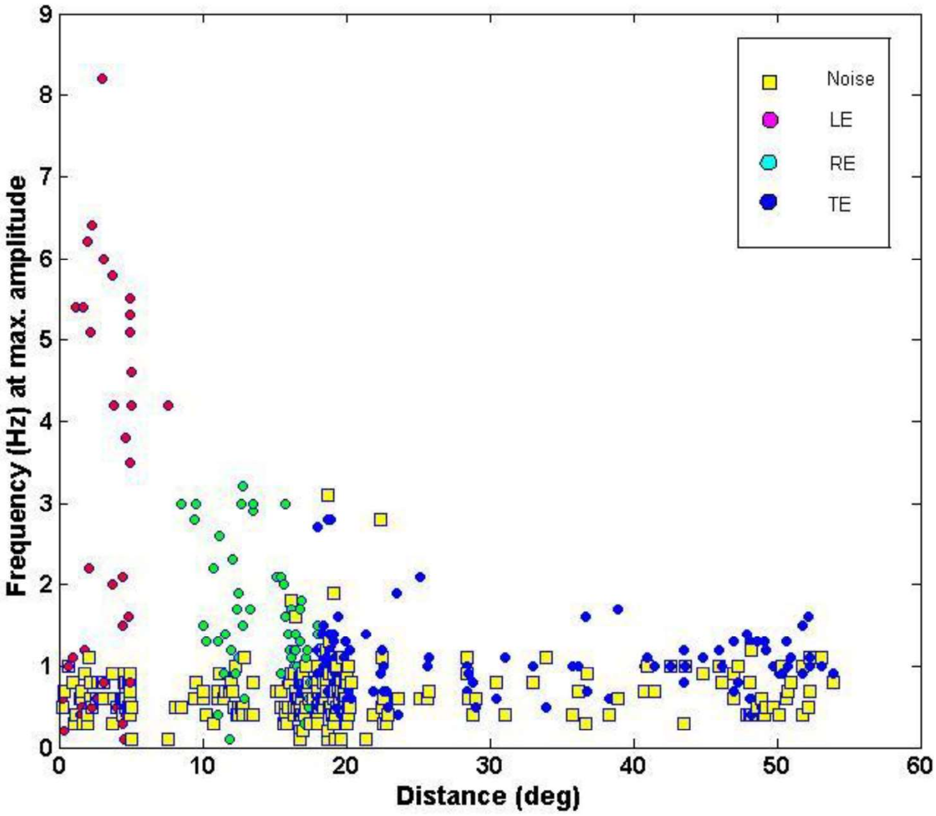


Fig. 4.4: Plot of frequency corresponding to maximum spectral amplitude vs epicentral distance for Noise and Local Event (LE), Regional Event (RE) and Teleseismic Event (TE).

To classify the event (i.e., LE or RE or TE) from noise based on spectral amplitude a mapping between spectral amplitudes and binary decision (event/noise) is required. However this mapping is not linear in nature. Therefore non-linear mapping capability of ANN has been used. In order to achieve this, a set of spectral amplitudes (typically 50) corresponding to the frequency range 0.2 to 20 Hz at an interval of 0.4 Hz have been computed using STFT at noise and event windows for all the 247 seismic events and fed to ANN as inputs. The desired output for the ANN is either 0 or 1 corresponding to noise or event. For training the ANN resilient propagation [81] algorithm has been used. Using trial and error basis the number of neurons in the hidden layer is set to 25 which correspond to minimum approximation error. The

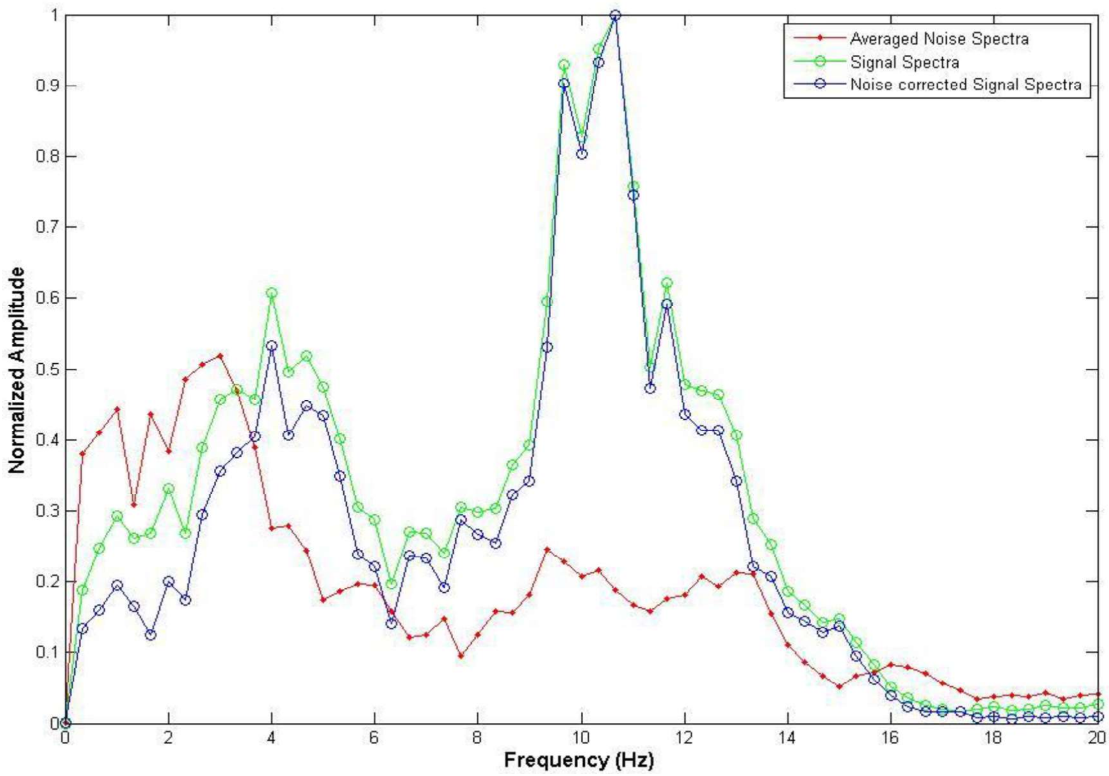


Fig. 4.3: The plot of normalized amplitudes vs. frequency.The green curve represents spectra of a typical earthquake signal for a time window of 3 sec. The red one represents averaged spectra of multiple noise samples and the blue curve is the noise corrected signal spectra.

To demonstrate further, in addition, amplitude spectra of 247 events (shown as pentagrams in Fig.4.1) and that of noise just preceding the events are computed. From these spectra the frequencies at which the spectral amplitude becomes maximum are noted and plotted against epicentral distances as shown in Fig. 4.4. It is seen from Fig. 4.4 that events are not distinguished unambiguously from noise considering the frequencies corresponding to maximum amplitude as the only separating parameter particularly for regional and teleseismic events (distance $\geq 5^0$).It is therefore perceived that to distinguish an event from noise it is essential to increase dimensionality. For example, event classification may consider the complete spectrum instead of a single frequency corresponding to maximum amplitude.

Corresponding smoothened (by 31 point moving average filter) amplitude spectra (sampling rate is 40 samples/sec) which are computed at the onsets of event and noise in a time window of 3 sec using STFT (with 50 percent cosine-tapered Tukey window) are shown in Fig.4.2(d), (e) and (f). Window width has been selected to be 3 sec to accommodate the onset times at the farthest channels along each of the arms (CH1 and CH10 or CH11 and CH20) of events corresponding to maximum lag of 3 sec. From Fig. 4.2 it is seen that noise spectra peak at higher amplitude than events. This makes detection based on threshold on amplitude unsuccessful.

For achieving a good estimation of signal spectrum (velocity), the effect of noise spectrum (average) is removed from signal spectrum as given by the following relation:

$$S(f)= \sqrt{(|X(f)|^2 - |N(f)|^2)}\ e^{i\Phi(f)} \tag{4.1}$$

where $S(f)$ is the noise corrected signal spectrum, $|X(f)|^2$ represents the squared magnitude of the signal spectrum, $|N(f)|^2$ represents the squared magnitudes of averaged noise spectra, and $\Phi(f)$ indicates the phase value of the signal spectrum. The phase value of the noise spectra has not been taken into consideration due to its randomness. The effect of this operation is shown in the Fig. 4.3.

It has been observed that for events with high SNR, smoothened amplitude spectra are clearly distinguishable from the noise spectra. However, detection based on amplitude thresholding in frequency domain, is quite difficult in case of events with low SNR. To demonstrate this, three typical low SNR events of local, regional and teleseismic origin have been considered. Signals of these events are shown in Fig. 4.2(a), (b) and (c).

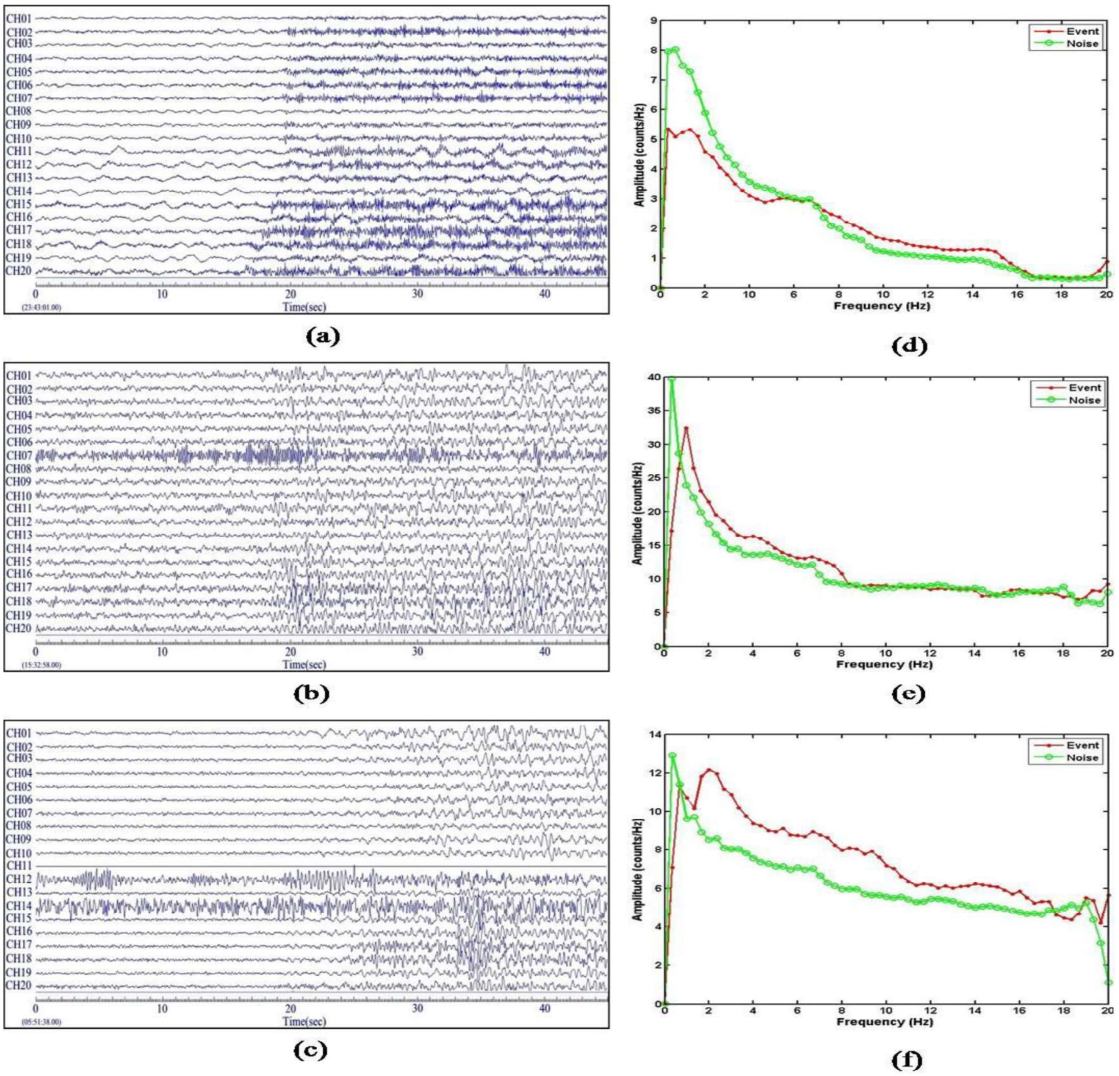


Fig. 4.2: The plot of seismic signals recorded by 20 channels of GBA for (a) local , (b) regional and (c) teleseismic events with low SNR. The corresponding plots of spectral amplitudes computed at noise (shown in hollow circle) and at P-onset of events (shown in dot) for time window of 3 sec as a function of frequency for a particular channel are shown in (d), (e) and (f). Here the spectral amplitudes are scaled by a factor of (1/S) where S is window length.

during the period Jan 2016 to Nov 2016 have been kept for testing.

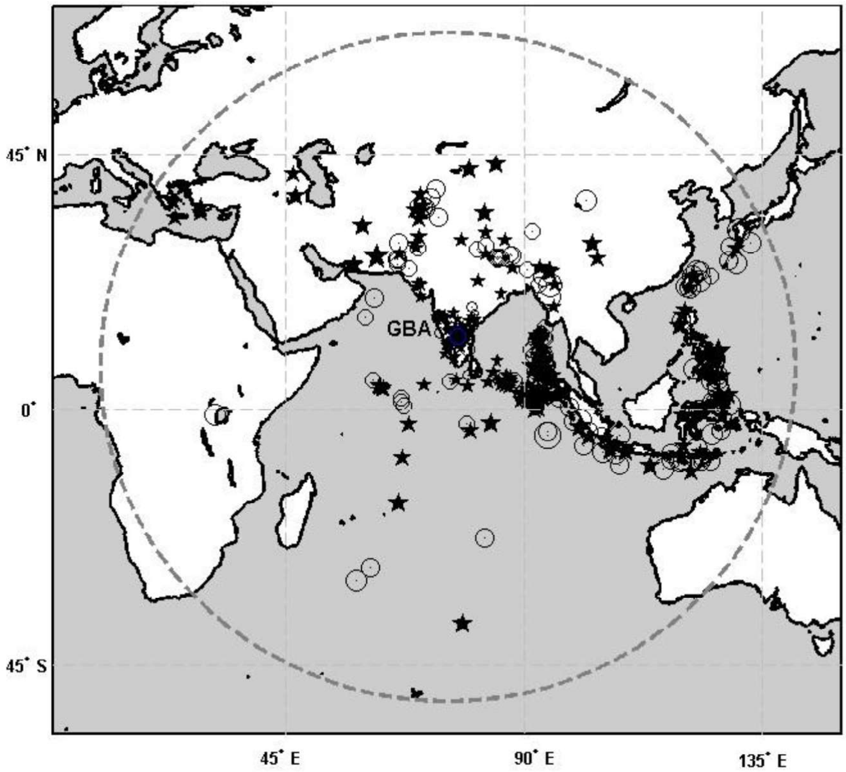


Fig. 4.1: Epicenters (shown in pentagram) of 247 events used for training ANN from the study area (area inside the dashed circle of radius ~54^0) and the location of recording station GBA. The tested events from the same area are 162 (shown by hollow circle).

It may be noted that parameters for events with magnitude greater than 4.5 have been taken from USGS reported event lists and those corresponding to lower magnitude (local events) have been picked up from GBA event list. This data set of 409 events which includes parameters from both USGS and GBA catalogues is considered as Truly Detected events (TD) during the above period. The training dataset contains events with magnitude 2.8 and above whereas the test dataset contains events with magnitude 0.8 and above. The various types of magnitudes have been considered as defined by USGS *(https://earthquake.usgs.gov/earthquakes/eventpage/terms.php)*.

Chapter-4

Application to Sumatra Region

In the Indian Ocean, Sumatra region is one of the prominent sources of tsunamis and it is located at a regional distance from eastern Indian coast, Sri Lanka and many other habitable islands. It has been considered as a study area for application of the proposed method of identification of tsunamigenic earthquakes. For detection of seismic event, short period array (GBA) of India has been considered and the results obtained are discussed in section 4.1. Location estimation results which have been computed using conventional method (using GBA data) have not been included here. For estimating moment magnitude and for identifying the earthquake category, a single 3-component broad band station (PALK) of Sri Lanka has been chosen. The results obtained are quoted in sections 4.2 and 4.3 respectively. It may be noted here that the seismic waves share nearly the same path from Sumatra to PALK as that from Sumatra to GBA. The chapter is ended by summarizing in section 4.4.

4.1Detection using GBA data

Considering GBA [93] as centre (Fig. 4.1), a total of 409 seismic events within a radial distance of $\sim 54^0$ (shown by dashed circle) and azimuthal coverage of 360^0 have been collected from the GBA observatory. Out of 409 events, 247 events (shown in Pentagram) which occurred during the period 2010 to 2013 has been used for training the ANN and 162 events (shown by hollow circle in Fig. 4.1) which occurred the development of Indian tsunami warning system. In this study, the near real-time data have been used to demonstrate the identification of tsunamigenic earthquakes originating from Sumatra. Despite being proven successful in the research, further work towards implementation of the proposed methods for a real tsunami warning system is necessary and desirable.

methods. For identifying the event as tsunamigenic, seismic data of a single 3-component broad band station, namely, PALK located at regional distance has been used. From the vertical component record, the moment magnitude (an important parameter for tsunami warning) is estimated using ANN, as fast as 3 s after the P onset. The ANN has been found to estimate M_w of about 73% of the new events with absolute error less than 0.1 and of 100% of the events with absolute error less than 0.35. Using 3- component records rms amplitudes of seismic phases namely P, S and LR have been computed. Subsequently an ANN has been used to identify the event as tusnamigenic or non-tusnamigenic by mapping these amplitudes along with location and magnitude parameters to earthquake category. This identification could be as fast as 5 min after the arrival of LR phase. The ANN based mapping method has been found to categorize 100% of the new earthquakes successfully as tsunamigenic or non-tsunamigenic. This mapping technique has been corroborated by two alternative mapping techniques namely 1)computation of water volume by estimating focal parameters using ANN and 2) mapping between rms amplitudes of seismic phases (along with location and magnitude) and earthquake category using SVM. These corroborative techniques have been found to identify 95% and 91% respectively of the new earthquakes successfully as tsunamigenic and non-tsunamigenic. The close agreement among the mapping techniques ensure the strength and reliability of the proposed algorithm in this thesis work which could be implemented as an additional tool to conventional tsunami warning system, used to alert Indian coast against tsunamis originating in Sumatra.

Scope of future study

All the major research findings in the present work can be incorporated into

Chapter-5

Conclusion and Future Scope

This research work has been carried out with the objective that the tsunamigenic earthquakes could be identified promptly using a minimum of one 3-component seismic station located at a regional distance from the tsunami source. The motivation behind this work was that the tsunamigenic potential could be determined from the 3-components records of seismic wave which travels faster than tsunami wave keeping in mind that seismic signal is the end result of earthquake magnitude, fault rupture characteristics at the source, the propagation path effects and the instrumentation effects at the recording site. The method proposed in this work has been validated with seismic data recorded at PALK station, Sri lanka for earthquakes originating from tsunami source region, namely, Sumatra. All the important findings from this research are summarized in the following sections.

Major achievement of this work has been in establishing that rms amplitudes of seismic phases like P, S and LR along with location and moment magnitude computed from seismograms recorded at a single 3-component station can be mapped to earthquake category. This has been achieved by the following process. An ANN has been used to detect seismic events by extracting features from short period recordings of Gauribidanur Array via STFT. The STFT and ANN based method has been found to detect 99% of the new events with an average error of ±0.036sec in onset pick up. After detection the events have been located using the conventional

28	25/2/2008	21:02:18	-2.245	99.808	25	21.31	115.84	6.7	320	8	106	*NT*
29	25/2/2008	18:06:04	-2.332	99.891	25	21.42	115.95	6.6	318	6	103	*NT*
30	25/2/2008	8:36:33	-2.486	99.972	25	21.56	116.23	7.2	317	6	102	*NT*
31	24/2/2008	14:46:21	-2.405	99.931	22	21.49	116.08	6.5	322	6	107	*NT*
32*	4/1/2008	7:29:18	-2.782	101.032	35	22.64	115.66	6	323	27	102	*NT*
33	22/12/2007	12:26:17	2.087	96.806	23	16.86	107.13	6.1	295	9	65	*NT*
34	21/9/2006	18:54:50	-9.05	110.365	25	33.77	118.44	6	276	24	85	*NT*
35	27/7/2006	11:16:40	1.707	97.146	20	17.31	107.98	6.3	336	5	115	*NT*
36	19/7/2006	10:57:37	-6.535	105.389	45	28.23	118.7	6.1	311	35	116	*NT*
37	17/7/2006	15:46:00	-9.42	108.319	21	32.18	120.84	6.1	283	75	-89	*NT*
38*	17/7/2006	8:19:27	-9.284	107.419	20	31.35	121.48	7.7	290	10	102	*T*
39*	16/5/2006	15:28:26	0.093	97.05	12	17.81	113.07	6.8	358	82	-31	*NT*
40	25/4/2006	18:26:17	1.994	96.995	21	17.07	107.22	6.3	293	7	66	*NT*
41	19/4/2006	20:36:46	2.643	93.226	17	13.31	109.74	6.2	19	86	-24	*NT*
42*	19/11/2005	14:10:13	2.164	96.786	21	16.82	106.9	6.5	306	15	78	*NT*
43*	5/7/2005	1:52:03	1.819	97.082	21	17.21	107.69	6.7	329	8	107	*NT*
44*	8/6/2005	6:28:11	2.17	96.724	23.5	16.76	106.94	6.1	308	19	85	*NT*
45*	19/5/2005	1:54:53	1.989	97.041	30	17.12	107.19	6.9	290	8	65	*NT*
46	18/5/2005	11:37:42	5.439	93.357	2.5	12.71	97.56	6.1	283	85	157	*NT*
47	14/5/2005	5:05:18	0.587	98.459	34	18.93	109.9	6.7	326	22	88	*NT*
48	10/5/2005	1:09:05	-6.226	103.139	17	26.14	120.55	6.3	315	33	111	*NT*
49	28/4/2005	14:07:34	2.132	96.799	22	16.84	106.99	6.2	301	15	75	*NT*
50*	16/4/2005	16:38:04	1.812	97.662	31	17.76	107.09	6.4	344	19	129	*NT*
51	11/4/2005	6:11:12	2.169	96.759	24	16.79	106.91	6.1	308	18	81	*NT*
52*	10/4/2005	17:24:39	-1.591	99.717	30	20.94	114.32	6.4	297	80	87	*NT*
53*	10/4/2005	11:14:20	-1.714	99.779	30	21.05	114.55	6.5	293	49	105	*NT*
54	10/4/2005	10:29:11	-1.644	99.607	19	20.86	114.59	6.7	323	56	91	*T*
55*	8/4/2005	5:48:38	-0.215	97.731	20.9	18.56	113.08	6.1	336	56	165	*NT*
56	3/4/2005	3:10:56	2.022	97.942	36	17.96	106.16	6.3	329	24	111	*NT*
57*	3/4/2005	0:59:21	0.368	98.319	30	18.88	110.69	6	325	14	105	*NT*
58	30/3/2005	16:19:41	2.993	95.414	22	15.26	105.53	6.3	298	14	72	*NT*
59	9/2/2005	13:27:25	4.797	95.117	44.5	14.55	98.99	6	310	25	89	*NT*
60*	26/1/2005	22:00:43	2.699	94.602	22.2	14.58	107.58	6.2	300	31	72	*NT*
61	9/1/2005	22:12:57	4.926	95.108	40	14.51	98.49	6.1	311	22	88	*NT*
62	27/12/2004	9:39:07	5.348	94.65	35	14	97.1	6.1	320	21	97	*NT*
63*	26/12/2004	0:58:53	3.295	95.982	30	15.72	103.86	9.1	329	8	110	*T*
64	16/4/2004	21:57:05	-5.214	102.718	44.5	25.27	119.01	6	223	11	5	*NT*
65	22/2/2004	6:46:27	-1.559	100.488	42	21.63	113.35	6	224	54	-45	*NT*
66	15/1/2002	7:12:58	-6.314	105.205	10	27.97	118.48	6.1	74	84	-7	*NT*
67	13/2/2001	19:28:30	-4.68	102.562	36	24.87	118.09	7.4	315	16	103	*NT*
68	16/1/2001	13:25:10	-4.022	101.776	28	23.87	117.59	6.9	321	14	111	*NT*

Note: 22 events (marked * and **) kept for testing ANN-I and SVM while 20 events (marked *) for testing ANN-II.

68	08/04/2005	05:48:37.88	-0.215	97.731	20.9	6.1
69	03/04/2005	03:10:56.47	2.022	97.942	36	6.3
70	03/04/2005	00:59:21.42	0.368	98.319	30	6
71*	30/03/2005	16:19:41.10	2.993	95.414	22	6.3
72	28/03/2005	16:09:36.53	2.085	97.108	30	8.6
73	09/02/2005	13:27:25.34	4.797	95.117	44.5	6
74	26/01/2005	22:00:42.57	2.699	94.602	22.2	6.2
75	09/01/2005	22:12:56.51	4.926	95.108	40	6.1
76	27/12/2004	09:39:06.80	5.348	94.65	35	6.1
77	26/12/2004	00:58:53.45	3.295	95.982	30	9.1
78	25/07/2004	14:35:19.06	-2.427	103.981	582.1	7.3
79	16/04/2004	21:57:05.41	-5.214	102.718	44.5	6
80	22/02/2004	06:46:27.04	-1.559	100.488	42	6
81	15/01/2002	07:12:58.03	-6.314	105.205	10	6.1
82*	25/05/2001	05:06:10.68	-7.869	110.179	143.1	6.3
83	13/02/2001	19:28:30.26	-4.68	102.562	36	7.4
84*	16/01/2001	13:25:09.83	-4.022	101.776	28	6.9

Note: * indicates events (22 in no.) kept for estimating M_w by ANN

A.2: List of events from Sumatra which occurred between 16 Jan 2001 and 6 Dec 2016 with magnitude 6 and above.

Evt. No.	Date	Or. Time hh: mm :ss	Lat deg	Lon deg	Depth km	Distance deg	Azimuth deg	M_w	ϕ_s deg	δ deg	λ deg	*CAT* USGS
1**	6/12/2016	22:03:33	5.283	96.168	13	15.5	96.48	6.5	243	81	33	*NT*
2**	1/6/2016	22:56:00	-2.097	100.665	50	21.41	115.24	6.6	150	90	86	*NT*
3*	6/4/2016	14:45:30	-8.204	107.386	29	30.78	119.7	6.1	301	29	111	*NT*
4	2/3/2016	12:49:48	-4.952	94.33	24	18.29	131.57	7.8	96	84	-170	*NT*
5*	3/3/2015	10:37:30	-0.779	98.716	28	19.69	113.41	6.1	323	17	107	*NT*
6	5/7/2014	9:39:28	1.934	96.939	20	17.04	107.48	6	296	11	69	*NT*
7*	14/9/2012	4:51:47	-3.319	100.594	19	22.5	117.43	6.2	330	9	115	*NT*
8	15/4/2012	5:57:40	2.581	90.269	25	10.62	115.74	6.2	109	78	-158	*NT*
9	11/4/2012	10:43:11	0.802	92.463	25.1	13.39	118.37	8.2	109	80	-164	*T*
10*	11/4/2012	8:38:37	2.327	93.063	20	13.27	111.27	8.6	289	89	154	*T*
11	10/1/2012	18:36:59	2.433	93.21	19	13.37	110.61	7.2	12	78	-15	*NT*
12	13/10/2011	3:16:30	-9.35	114.587	39	37.63	115.71	6.1	291	83	16	*NT*
13*	22/08/2011	20:12:21	-6.282	104.054	29	26.95	119.63	6.1	301	24	89	*NT*
14*	3/4/2011	20:06:40	-9.848	107.693	14	31.87	122.13	6.7	307	54	-72	*NT*
15	26/1/2011	15:42:30	2.205	96.829	23	16.85	106.72	6.1	304	19	76	*NT*
16	17/1/2011	19:20:57	-5.03	102.647	36	25.12	118.72	6	310	22	95	*NT*
17	25/10/2010	14:42:22	-3.487	100.082	20.1	22.13	118.46	7.8	316	8	96	*T*
18	9/5/2010	5:59:42	3.748	96.018	38	15.64	102.2	7.2	308	15	88	*NT*
19	5/5/2010	16:29:03	-4.054	101.096	27	23.29	118.47	6.5	326	10	111	*NT*
20	6/4/2010	22:15:02	2.383	97.048	31	17	105.91	7.8	307	7	88	*NT*
21	5/3/2010	16:07:01	-3.762	100.991	26	23.06	117.95	6.8	324	13	109	*NT*
22	2/9/2009	7:55:01	-7.782	107.297	46	30.49	119.06	7	198	50	65	*T*
23	16/8/2009	7:38:22	-1.479	99.49	20	20.69	114.31	6.7	338	59	89	*T*
24	15/4/2009	20:01:35	-3.115	100.471	22	22.3	117.11	6.3	324	10	109	*NT*
25	29/3/2008	17:30:50	2.855	95.296	20	15.19	106.17	6.3	300	10	75	*NT*
26	15/3/2008	14:43:26	2.708	94.596	25	14.57	107.56	6	305	33	77	*NT*
27	3/3/2008	2:37:27	-2.18	99.823	25	21.29	115.66	6.2	323	17	114	*NT*

Appendix A

Data Tables

A.1: USGS reported parameters of events from Sumatra region which occurred between 16 Jan 2001 and 19 Oct 2016 with magnitude 6 and above.

Evt. No.	Date	Or. Time (hh:mm:ss)	Lat (deg)	Lon (deg)	Depth (km)	Mw
1	19/10/2016	00:26:01.09	-4.8626	108.1627	614	6.6
2*	01/06/2016	22:56:00.80	-2.0967	100.6654	50	6.6
3	06/04/2016	14:45:29.62	-8.2036	107.3857	29	6.1
4	02/03/2016	12:49:48.11	-4.9521	94.3299	24	7.8
5	03/03/2015	10:37:30.05	-0.7789	98.7161	28	6.1
6	05/07/2014	09:39:27.79	1.9335	96.9388	20	6
7	14/09/2012	04:51:47.07	-3.319	100.594	19	6.2
8	25/07/2012	00:27:45.26	2.707	96.045	22	6.4
9	23/06/2012	04:34:53.18	3.009	97.896	95	6.1
10*	15/04/2012	05:57:40.06	2.581	90.269	25	6.2
11	11/04/2012	10:43:10.85	0.802	92.463	25.1	8.2
12	11/04/2012	08:38:36.72	2.327	93.063	20	8.6
13	10/01/2012	18:36:59.08	2.433	93.21	19	7.2
14	13/10/2011	03:16:30.16	-9.35	114.587	39	6.1
15*	05/09/2011	17:55:11.22	2.965	97.893	91	6.7
16*	22/08/2011	20:12:20.95	-6.282	104.054	29	6.1
17	03/04/2011	20:06:40.39	-9.848	107.693	14	6.7
18	26/01/2011	15:42:29.59	2.205	96.829	23	6.1
19*	17/01/2011	19:20:57.21	-5.03	102.647	36	6
20	25/10/2010	19:37:31.15	-2.958	100.372	26	6.3
21*	25/10/2010	14:42:22.46	-3.487	100.082	20.1	7.8
22*	09/05/2010	05:59:41.62	3.748	96.018	38	7.2
23	05/05/2010	16:29:03.21	-4.054	101.096	27	6.5
24	06/04/2010	22:15:01.58	2.383	97.048	31	7.8
25	05/03/2010	16:07:00.68	-3.762	100.991	26	6.8
26	16/10/2009	09:52:50.83	-6.534	105.223	38	6.1
27	01/10/2009	01:52:27.32	-2.482	101.524	9	6.6
28	30/09/2009	10:16:09.25	-0.72	99.867	81	7.6
29*	02/09/2009	07:55:01.05	-7.782	107.297	46	7
30	16/08/2009	07:38:21.70	-1.479	99.49	20	6.7
31	15/04/2009	20:01:34.68	-3.115	100.471	22	6.3
32*	19/05/2008	14:26:45.02	1.64	99.147	10	6
33	29/03/2008	17:30:50.15	2.855	95.296	20	6.3
34	15/03/2008	14:43:26.50	2.708	94.596	25	6
35	03/03/2008	02:37:27.12	-2.18	99.823	25	6.2
36	25/02/2008	21:02:18.42	-2.245	99.808	25	6.7
37	25/02/2008	18:06:03.90	-2.332	99.891	25	6.6
38	25/02/2008	08:36:33.03	-2.486	99.972	25	7.2
39	24/02/2008	14:46:21.47	-2.405	99.931	22	6.5
40	20/02/2008	08:08:30.52	2.768	95.964	26	7.4
41	22/01/2008	17:14:57.95	1.011	97.442	20	6.2
42*	04/01/2008	07:29:18.30	-2.782	101.032	35	6
43*	22/12/2007	12:26:17.47	2.087	96.806	23	6.1
44	01/12/2006	03:58:21.65	3.39	99.079	204	6.3
45	21/09/2006	18:54:50.05	-9.05	110.365	25	6
46*	11/08/2006	20:54:14.37	2.403	96.348	22	6.2
47	27/07/2006	11:16:40.37	1.707	97.146	20	6.3
48	19/07/2006	10:57:36.88	-6.535	105.389	45	6.1
49*	17/07/2006	15:45:59.82	-9.42	108.319	21	6.1
50	17/07/2006	08:19:26.68	-9.284	107.419	20	7.7
51*	26/05/2006	22:53:58.92	-7.961	110.446	12.5	6.3
52	16/05/2006	15:28:25.92	0.093	97.05	12	6.8
53	25/04/2006	18:26:17.15	1.994	96.995	21	6.3
54*	19/04/2006	20:36:46.40	2.643	93.226	17	6.2
55	19/11/2005	14:10:13.03	2.164	96.786	21	6.5
56*	05/07/2005	01:52:02.95	1.819	97.082	21	6.7
57*	08/06/2005	06:28:10.92	2.17	96.724	23.5	6.1
58	19/05/2005	01:54:52.85	1.989	97.041	30	6.9
59	18/05/2005	11:37:41.74	5.439	93.357	2.5	6.1
60	14/05/2005	05:05:18.48	0.587	98.459	34	6.7
61	10/05/2005	01:09:05.10	-6.226	103.139	17	6.3
62	28/04/2005	14:07:33.70	2.132	96.799	22	6.2
63	16/04/2005	16:38:03.90	1.812	97.662	31	6.4
64*	11/04/2005	06:11:11.82	2.169	96.759	24	6.1
65	10/04/2005	17:24:39.40	-1.591	99.717	30	6.4
66*	10/04/2005	11:14:19.62	-1.714	99.779	30	6.5
67	10/04/2005	10:29:11.28	-1.644	99.607	19	6.7

[89] H. Kanamori. Mechanism of tsunami earthquakes.*Phys.Earth Planet. Inter*. 6: 346–359, 1972.

[90] J. Steele, S. A. Thorpe and K. K. Turekian.Elements of Physical Oceanography A derivative of the Encyclopedia of Ocean Science.*Academic Press*, 2010.

[91] W. H. Chen, S. H. Hsu and H. P. Shen.Application of SVM and ANN for intrusion detection. Computers and Operations Research, 32: 2617-2634, 2005.

[92] V. N. Vapnik. The nature of statistical learning theory. New York: *Springer*, 1995.

[93] F. Roy, S. K. Arora andT.K.Basu.Detection of weak seismic signals using stationarity of noise and coherency of signals at a seismic array.*Phys. Earth Planet. Inter.,* 73(3-4): 264-273, 1992.

[94] B. Kennett, E. Engdahl and R. Buland.Constraints on seismic velocities in the earth from travel times. *Geophysical Journal International* 122(1):108-124, 1995.

[95] D. K. Chaturvedi, Soft computing: Techniques and its Applications in Electrical Engineering. Springer, 2008.

[74] K.-Y. Chun, Y. Wu and G. A. Henderson.Lg attenuation near the north korean border with china, part i:Model development from regional earthquake sources. *Bulletin of the Seismological Society of America*, 99 (5) : 3021-3029, 2009.

[75] S. Shionoya, W. M. Yen and H. Yamamoto. Phosphor handbook: Second Edition. CRC press, 2006.

[76] S. W. Smith, et al. The scientist and engineer's guide to digital signal processing: First Edition.California, California Technical Publishing, 1997.

[77] K. Hornik, M. Stinchcombeand H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*. 2: 359-366, 1989.

[78] C. Ji, R. R. Snapp and D. Psaltis.Generalizing smoothness constraints from discrete samples.*Neural Computation*, 2: 188-197, 1990.

[79] T. Poggio and F. Girosi.Networks for approximation and learning.*Porc. IEEE*, 78(9): 1481-1497, 1990.

[80] D. S.Chen and R. C. Jain.A robust backpropagation learning algorithm for function approximation.*IEEE Trans. Neural Netw*. 5 (3): 467–479, 1994.

[81] M. Riedmiller and H. Braun. A direct adaptive method for faster backpropagation learning: The rprop algorithm. *Proceedings of the IEEE International Conference on Neural Networks*,586–591, 1993.

[82] K. Aki and P. G. Richards.Quantitative Seismology: Theory and Methods. W.H. Freeman and Company, San Francisco, 1980.

[83] T. C. Hanks and H. Kanamori.A moment magnitude scale.*Journal Geophysical Research*, 84: 2348–2350, 1979.

[84] F. Scherbaum. Of Poles and Zeros, Fundamental of Digital Seismology. *Kluwer Academic Publishers. 1996.*

[85] S. Stein and M. Wysession.An introduction to seismology, earthquakes, and earth structure. *Blackwell Publishing Ltd., 2003*

[86] T. Lay and T. C. Wallace.Modern Global Seismology.*Academic Press*, California, 1995.

[87] S.Mallat. A Wavelet Tour of Signal Processing.*October 9, 2008.*

[88] H.Kanamori and J. W. Given.Use of long-period surface waves for rapid determination of earthquake-source parameters.*Phys. Earth Planet. Inter*. 27: 8–31, 1981.

[60] S. H. Chew and K. Kuenza.Dtectingtsunamigenesis from undersea earthquake signals. *Journal of Asian Earth Sciences*, 36: 84-92, 2009.

[61] A. Chamoli, V. S. Rani, K. Srivastava, D. Srinagesh and V. P. Dimri.Wavelet analysis of the seismograms for tsunami warning.*Nonlin. Processes Geophysics*, 17: 569-574, 2010.

[62] M. Ewing, I. Tolstoy, F. Press. "Proposed use of the T-phase in tsunami warning systems.*Bulletin of the Seismological Society of America,* 40: 53-58, 1950.

[63] Y. Hiyoshi, D. A. Walker and C. S. McCreery.T-phase data and regional tsunamigenesis in Japan, *Bulletin of the Seismological Society of America*, 82:2213-2223, 1992.

[64] E. A. Okal, P. J. Alasset, O. Hyvernaud and F. Schindele.Thedeficient T-waves of tsunami earthquakes, *Geophysical Journal International*, 152: 416-432, 2003.

[65] D. Gabor. Theory of communication.J.*Inst. Electr. Eng*. 93: 429–441, 1946.

[66] J. Allen, L. Rabiner.A unified approach to short-time Fourier analysis and synthesis, *Proc. IEEE* 65: 1558–1564, 1977.

[67] M. Schimmel and J. Gallart.The inverse S-transform in filters with time–frequencylocalization.*IEEE Trans. Signal Process*. 53: 4417–4422, 2005.

[68] H.A. Pedersen, J.I. Mars and P.O. Amblard.Improving surface-wave group velocitymeasurements by energy reassignment, *Geophysics* 68: 677–684, 2005.

[69] I. Daubechies. Ten Lectures on Wavelets, in: *CBMS-NSF Reg. Conf. Series Appl.Math*.,. 61:SIAM, Philadelphia, PA, 1992.

[70] M.V. Wickerhauser.Adapted Wavelet Analysis from Theory to Software.*A.K. Peters, Ltd., Wellesley*, 1994.

[71] F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE* , 66 (1): 51-83, 1978.

[72] Y. Vaezi and M. V. Baan.Comparison of the sta/lta and power spectral density methods for microseismic event detection, *Geophys. J. Int.*, 203: 1896-1908, 2015.

[73] J. Park, C. R. Lindberg and F. L. Vernon. Multitaper spectral analysis of high-frequency seismograms.*Journal of Geophysical Research: Solid Earth*, 92 (B12):12675-12684, 1987.

[49] S. Tsuboi. Application of M_{wp} to Tsunami Earthquake, *Geophysical Research Letters*, 27(19): 3105-3108, 2000.

[50] J. Talandier and E. A. Okal.An algorithm for automated tsunami warning in French Polinesia based on mantle magnitudes, *Bulletin of Seismological Society of America*, 79: 1177-1193, 1989.

[51] F. Schindelé, D. Reymond, E. Gaucher and E. A. Okal. Analysis and Automatic Processing in Near-Field of Eight 1992-1994 Tsunamigenic Earthquakes: Improvements towards Real-Time Tsunami Warning, *Pure and Applied Geophysics*, 144(3-4): 381-408, 1995.

[52] A. Lomax and A. Michelini. M_{wpd} : a duration-amplitude procedure for rapid determination of earthquake magnitude and tsunamigenic potential from P waveforms*, Geophys. J. Int*., 176: 200-214, 2009.

[53] S. Tsuboi, P. M. Whitmore and T. J.Sokolowski.Application of M_{wp} to deep and teleseismic earthquakes.*Bull. Seismol. Soc. Am*, 89: 1345-1351, 1999.

[54] D. M. Boore and W. B.Joyner.W.B. Ground motions and response spectra at soil sites from seismological models of radiated spectra.*Proceedings, Eighth World Conference on Earthquake Engineering II*: 457- 464, 1984.

[55] E. A. Okal and J. Talandier. M_m : A variable-period mantle magnitude, *Journal of Geophysical Research*, 94(B4): 4169-4193, 1989.

[56] A. V. Newman and E. A. Okal. Tele-seismic estimates of radiated seismic energy: The E/Mo discriminent for tsunami earthquakes.*Journal of Geophysical Research*, 103(26): 885- 898, 1998.

[57] S. Weinstein, C. McCreery, B. Hirshornand P. Whitmore. Comment on a Strategy to Rapidly Determine the Magnitude of Great Earthquakes by W. Menke and V. Levin, Eos Trans. AGU 86, 263, DOI: 10.1029/2005EO280005, 2005.

[58] D. Wang, N.C. Becker, D. Walsh, G.J. Fryer, S. A. Weinstein, C. S. McCreery, V. Sardiña, V. Hsu, B. F. Hirshorn, G. P. Hayes, Z. Duputel, H. Kanamori, K. K. Koyanagi and B. Shiro. Real-time forecasting of April 11, 2012 Sumatra tsunami.*Geophysical Research Letters*, 39: L19601, 2012.

[59] O. G. Lockwood and H. Kanamori.Wavelet analysis of the seismograms of 2004 Sumatra-Andaman earthquake and its application to tsunami early warning, Geochemistry Geophysics Geo-systems, 7: 2006.

[37] A. G. Hafez, M. T. A. Khan and T. Kohda. Clear p-wave arrival of weak events and automatic onset determination using wavelet filter banks.*Digital Signal Processing*, 20 (3): 715-723, 2010.

[38] C. Bai and B. Kennett, Automatic phase-detection and identification by full use of a single three-component broadband seismogram.*Bulletin of the Seismological Society of America*, 90 (1): 187-198, 2000.

[39] E.-S. Akhouayri, D. Agliz, A. Atmani, et al. Automatic detection and picking of p-wave arrival in locally stationary noise using cross-correlation.*Digital Signal Processing*,26 :87-100, 2014.

[40] C. D. Saragiotis, L. J. Hadjileontiadis, S. M. Panas, Pai-s/k.A robust automatic seismic p phase arrival identification scheme.*IEEE Transactions on Geoscience and Remote Sensing*,40 (6): 1395-1404, 2002.

[41] L. Küperkoch, T. Meier, J. Lee, W. Friederich, E. W. Group, et al. Automated determination of p phase arrival times at regional and local distances using higher order statistics.*Geophysical Journal International.* 181(2): 1159-1170, 2010.

[42] C. Baillard, W. C. Crawford, V. Ballu, C. Hibert and A. Mangeney. An automatic kurtosis-based p-and s-phase picker designed for local seismic networks. *Bulletin of the Seismological Society of America*, 104(1): 394-409, 2014.

[43] Z. E. Ross and Y. Ben-Zion.Automatic picking of direct p, s seismic phases and fault zone head waves. Geophysical Journal International, 199(1): 368-381, 2014.

[44] E. Kalkan.An automatic p-phase arrival-time picker.*Bulletin of the Seismological Society of America*, 106(3): 971-986, 2016.

[45] R.G. Roberts, A.Christoffersson and F. Cassidy.Real time event detection, phase identification and source location estimation using simple station three component seismic data,*Geophys. J. Roy. Astro. Soc.,*97 : 471-480,1989.

[46] L. Geiger.Probability method for the determination of Earthquake epicenters from the arrival time only.*Bull.St.Louis Univ.,* 8: 60-71, 1912.

[47] M. Sambridge and K. Gallagher.Earthquake hypocenter location using genetic algorithms.*Bull.seismol. Soc. Am.*,83(5): 1467–1491, 1993.

[48] S. Tsuboi, K. Abe, K. Takano and Y. Yamanaka.Rapid determination of M_w from broadband P waveforms.*Bulletin of the Seismological Society of America*, 85: 606-613, 1995.

[25] M. Baer and U. Kradolfer.An automatic phase picker for local and teleseismic events.*Bulletin of the Seismological Society of America*,77 (4): 1437-1445, 1987.

[26] B. O. Ruud and E. S. Husebye. A new three-component detector and automatic single-station bulletin production.*Bulletin of the seismological society of America*, 82 (1) : 221-237, 1992.

[27] M. Tarvainen. Automatic seismogram analysis: statistical phase picking and locating methods using one-station three-component data. *Bulletin of the Seismological Society of America*, 82 (2): 860-869, 1992.

[28] P. S. Earle and P. M. Shearer.Characterization of global seismograms using an automatic-picking algorithm.*Bulletin of the Seismological Society of America*, 84 (2) : 366-376,1994.

[29] T. Akazawa. A technique for automatic detection of onset time of p-and s-phases in strong motion records. *Proc. of the 13th World Conf. on Earthquake Engineering*, 2004.

[30] H. Dai and C. MacBeth.Automatic picking of seismic arrivals in local earthquake data using an artificial neural network.*Geophysical journal international*, 120 (3): 758-774, 1995.

[31] E. Mousset, Y. Cansi, R. Crusem and Y. Souchet.A connectionist approach for automatic labeling of regional seismic phases using a single vertical component seismogram.*Geophysical Research Letters*, 23 (6): 681-684, 1996.

[32] J. Wang and T. l. Teng. Identification and picking of s phase using an artificial neural network.*Bulletin of the Seismological Society of America*, 87(5): 1140-1149,1997.

[33] Y. Zhao and K. Takano.An artificial neural network approach for broadband seismic phase picking.*Bulletin of the Seismological Society of America*, 89 (3): 670-680, 1999.

[34] S. Gentili and A. Michelini. Automatic picking of p and s phases using a neural tree.*Journal of Seismology*, 10 (1): 39-63,2006.

[35] K. S. Anant, F. U. Dowla, Wavelet transform methods for phase identification in three-component seismograms. Bulletin of the Seismological Society of America, 87 (6): 1598-1612, 1997.

[36] H. Zhang, C. Thurber and C. Rowe. Automatic p-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings. *Bulletin of the Seismological Society of America*, 93 (5): 1904-1912, 2003.

[12] T. Maruyama. Statistical elastic dislocations in an infinite and semi-infinite medium.*Bulletin of Earthquake Research Institute, Tokyo Univ.* 42: 289–368,1964.

[13] P. M. Davis. Surface deformation associated with a dipping hydro-fracture.*Journal of Geophysical Research*, B88(7): 5826-5834, 1983.

[14] Y. Okada. Surface deformation due to shear and tensile faults in a half-space.*Bulletin of the Seismological Society of America*, 75(4): 1135-1154, 1985.

[15] E. L. Geist. Source characteristics of the July 17, 1998 Papua New Guinea Tsunami. EOS, Transactions of the American Geophysical Union, 79 (supplement): 571, 1998.

[16] V.V. Titov and F.I. Gonzalez. Implementationand testing of the Method Of Splitting Tsunami (MOST) Model. NOAA Technical Memorandum ERL PMEL-112 (PB98-122773), PMEL, Seattle, WA, 11, 1997.

[17] P. L. F. Liu, C. Yong-Sik ,M. J. Briggs, U. Kanoglu , C. E. Synolakis. Runup of solitary waves on a circular island. JFM 302:259 – 285, 1995.

[18] F. Imamura and H.J.Lee. State-of-arts of numerical simulation for tsunamis generation and propagation, *Bull. on Coastal Oceanography*, 36: 91-101,1988.

[19] F. Imamura. Review of tsunami simulation with a finite difference method, Long-wave Runup Models edited by H.Yeh, P.Liu and C.Synolakis, *World Scientific (*ISBN981-02-2909-7): 25-42, 1996.

[20] D. Reymond, O.Hyvernaud and J.Talandier. Automatic detection, location and quantification of earthquake: application to tsunami warning, *Pure Applied geophysics*, 135: 361-382, 1991.

[21] M. Withers, R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, J. Trujillo. A comparison of select trigger algorithms for automated global seismic phase and event detection.*Bulletin of the Seismological Society of America, 88 (1): 95-106, 1998.*

[22] R. V. Allen. Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America,* 68 (5): 1521-1532,1978.

[23] R. V. Allen. Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, 72 (6B) : S225-S242,1982.

[24] R. R. Blandford.Seismic event discrimination.*Bulletin of the Seismological Society of America,*72: 69-87,1982.

Bibliography

[1] E. N. Bernard.Program Aims to Reduce Impact of Tsunamis on Pacific states. *Earth in Space*, 11(2): 6–10, 1998.

[2] A. kundu, Pratap Mane, Y. S.Bhadauria, S. Basu, and S. Mukhopadhyay. Detecting seismic events using short-time Fourier transform and artificial neural network.*Report No. BARC/2017/I/019*.

[3] A. kundu, Y. S.Bhadauria, S. Basu, and S. Mukhopadhyay. Artificial neural network basedestimation of moment magnitude with relevance to Earthquake Early Warning.*International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*,2017.

[4] A. kundu, Y. S. Bhadauria, S. Basu, and S. Mukhopadhyay. Prompt identification of tsunamigenic earthquakes from 3-component seismic data. *Physics of the Earth and Planetary Interiors*, 259: 10– 17, 2016.

[5] A. kundu, Y. S.Bhadauria, S. Basu, and S. Mukhopadhyay. Application of ANN and SVM foridentification of tsunamigenic earthquakes from 3-component seismic data.*2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2017.*

[6] L. L. Noson, A..Qamar and G. W.Thorsen.Washington state earthquake hazards.*Washington Division of Geology and Earth Resources Information Circular*,85: 77, 1988.

[7] S. N. Ward. Tsunamis.*Encyclopedia of Physical Science and Technology-Academic Press*, 2005.

[8] S. Stein and E. A. Okal. The 2004 Sumatra earthquake and Indian Ocean tsunami:What happened and why? *Geoscience*, 1-6, 2005.

[9] Y. Tanioka and K. Satake.Tsunami Generation by Horizontal Displacement ofOcean Bottom.*Geophysical Research Letters*, 23(8): 861-864, 1996.

[10] K. Abe. Estimate of tsunami run-up heights from earthquake magnitudes, in Tsunami: Progress in Prediction, Disaster Prevention and Warning.Edited by Y. Tsuchiya and N. Shuto, Kluwer.*Academic Publishers, Dordrecht*, 21–35, 1995.

[11] J. A. Steketee. On Volterra's dislocation in semi-infinite elastic medium. *Canadian Journal of Physics*, 36: 192–205, 1958.