UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXII CICLO DEL DOTTORATO DI RICERCA IN

Earth Science and Fluid Mechanics

Temporal and spatial analysis of near fault stations in terms of impulsive behavior

Settore scientifico-disciplinare: GEO/10 Geofisica della Terra Solida

DOTTORANDO / A
DENIZ ERTUNCAY

COORDINATORE
PROF. PIERPAOLO OMARI

SUPERVISORE DI TESI
PROF. GIOVANNI COSTA
DR. ANDREA DE LORENZO

ANNO ACCADEMICO 2018/2019

Temporal and spatial analysis of near fault stations in terms of impulsive behavior

By

Deniz Ertuncay

University of Trieste, 2019

A blank page

ACKNOWLEDGEMENTS

I would first like to thank my supervisor Prof. Giovanni Costa from the Department of Mathematics and Geosciences at University of Trieste. He provided me a free and peaceful environment to work and gave me enough support when needed. I would also like to thank Dr. Andrea De Lorenzo from Machine Learning Laboratory of University of Trieste for his constant support to me, whenever I had a hard time accomplishing tasks related with machine learning algorithms.

I would like to thank to Dr. Antonio Scala from RISSC-Lab at University of Napoli Federico II for his help on the computation of synthetic data that have been used in the thesis.

I would also like to thank to Prof. Nurcan Meral Özel from Comprehensive Nuclear-Test-Ban Treaty Organization for her countless efforts for me throughout my postgraduate career.

This project could not have been accomplished without the of support of the SeisRaM working group members, Dr. Antonella Gallo, Enrico Margin, Laura Cataldi and Piero Falconer. I would also like to thank former members of the group, Dr. Blaž Vičič, Dr. Giuliana Zoppè, Dr. Lara Tiberi and Shaula Martinolli and my friends, Ekrem Bekin, Hafize Başak Bayraktar, Jonathan Ford, Rıdvan Örsvuran, Samantha France and Suat Vardar.

This work could not have been done without the support of my spouse Selen Caner Ertuncay and my dear family. I would like to express my gratitude to my wife for her love and her continuous support. I would also like to thank my grandmother for her boundless support to me in pursuing my education. I am thankful to my uncle Mehmet Yücel Önal, who, although no longer with us, convinced me to build a career on earth science. Thank you.

ABSTRACT

Temporal and spatial analysis of near fault stations in terms of impulsive behavior

Increasing number of seismic stations located in close proximity to active faults allows analysis of seismic signals that are recorded in near fault regions. Unique seismic signals, called "impulsive" or "pulse shaped" signals, are captured in velocity waveforms in numerous large magnitude earthquakes. In such waveforms, the earthquake is recorded as a one or several long period high amplitude signals. Long period signals are important in engineering seismology due to their large loads on structures. Ground motion prediction equations and design codes fail to capture the amplitudes in long periods of the impulsive signals. In this thesis impulsive signals and their spatial distribution in near fault regions are investigated. To do that two different algorithms are developed in order to distinguish impulsive signals from non-impulsive signals. Moreover, the probability of the pulse shaped signal occurrence is estimated.

In order to investigate the impulsive signals, near fault records from major crustal earthquakes are merged into a dataset. It contains waveforms that are coming from well-known seismogenic zones. Waveforms in the dataset are also analyzed by implementing several previous studies to make a comparison.

The first pulse shaped signal classification algorithm is developed using wavelet analysis. Wavelet analysis decomposes the signal into time-frequency domain, which provides the energy variation with time and frequency. The wavelet power spectrum of velocity waveforms are analyzed by using Ricker and Morlet wavelets. A threshold of minimum amplitude is applied. A comparison is made between the total energy of a signal and the energy of the time incidence where peak ground velocity is measured. Furthermore time incidence, where maximum spectral energy is located in time, is also taken into consideration. Energy ratios

are used for determination of impulsive signals. It is found that a Ricker wavelet explains the features of the impulsive part of the velocity waveforms more accurately than the Morlet wavelet. It can measure the period of the pulse and the phase shift of the impulsive parts of the waveform. Spectral features of the impulsive signals are also captured successfully using a Ricker wavelet.

The second classification algorithm uses convolutional neural networks. In order to train the convolutional neural networks, synthetic impulsive signals are created. A model is developed using real non-impulsive velocity waveforms from the dataset and synthetic impulsive waveforms. Impulsive signals are manually labeled as impulsive or non-impulsive. The trained model is run on the real manually-picked impulsive signals of the dataset and the performance of the convolutional neural network, the wavelet method, and various previously published methods are benchmarked. The convolutional neural networks approach correctly identifies almost 97 % of the impulsive signals. Accuracy rate of the model is superior than other models.

In order to understand the probability of the impulsive signals on earthquakes, a multivariate Bayes classifier method is implemented on the dataset. Various information about the fault, earthquake and station are analyzed and 3 parameters that are correlated with the impulsive signals are used for the probability calculations. Probability models are developed for normal, reverse and strike slip faults. The validity of this model is tested on the data set. Developed models can provide pulse probability distributions without requiring earthquake-specific parameters. A relation between the period of the pulses and the moment magnitude is also developed.

A blank page

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

A blank page

1. INTRODUCTION

Earthquakes are one of the most destructive natural hazards in the world. According to National Oceanic and Atmospheric Administration (NOAA) around 2.5 million people have died due to earthquakes or earthquake induced disasters such as tsunami, landslide and liquefaction. In order to prevent such human and economic loss, earthquake risk models are created. Earthquake risk can be defined as:

$$\text{Earthquake Risk} = \text{Earthquake Hazard} \times \text{Exposure} \times \text{Structural Vulnerability}$$

Structural vulnerability, earthquake hazard and exposure have clear differences. Exposure can be defined as the number of humans and quantity and quality of the infrastructures in the site of interest. Structural vulnerability is directly related to the infrastructure and methods used to build it. Earthquake hazard, on the other hand, is the combination of natural phenomena that can produce earthquake risk. There are various hazards, some of which are explained above, that can create extra loads on infrastructure, which increases the earthquake risk. To understand the source of earthquake hazards, various methods have been used. For instance, tsunami hazard maps are created in order to predict the maximum wave height from tsunami waves at coasts and inland. In order to create earthquake related hazard maps, fault line related information must be known. Using this information, the maximum amplitude (eg. PGA, PGV) that an earthquake can produce for a site of interest is calculated. This is one of the most used parameters. Ground motion prediction equations (GMPEs) are used in order to predict the maximum amplitude of acceleration, velocity and spectral periods of a site. Prediction GMPEs may fail due to directivity effect. Directivity effects may be seen in seismic recorders as larger amplitudes in the directions of rupture propagation.

The areas where the distance between the site of interest and the fault line that ruptures during an earthquake is less than the length of the ruptured fault are called near fault areas

(regions). The velocity-time history of the seismic stations in near fault regions may have large, pulse-shaped motions. These velocity waveforms have large, long period signals. A pulse-shaped signal is the representation of the effect of the almost all seismic energy from the fault arriving within a short period of time. In this case GMPEs may fail to predict the amplitudes of these signals. In such cases structures may have to deal with higher loads than were predicted in building codes. Some of the seismic design codes fail on long period large amplitude signals (Bertero et al., 1978; Chai et al., 2000; Loh et al., 2002).

A velocity waveform, which contains the effect of this natural phenomena, is called an impulsive signal, a velocity pulse or a pulse shaped signal. Thanks to the denser installation of seismic stations around near fault regions in recent years, waveforms with impulsive features are now recorded. Data centers around the world are now monitoring these near fault regions and have started to collect broad band and strong motion data. The data providers, earthquakes and the frequency of signals have been explained in Section 2.

Starting from the late 1990s, several different indicators of impulsive signals have been identified. Previous studies have inspected the near fault velocity signals and have identified different features of impulsive signals, such as:

  (i) Signals with long periods and large amplitudes (Somerville et al., 1997),
 (ii) Earthquake energy is concentrated in one (or multiple) pulse(s) (Somerville et al., 1997),
(iii) A high PGV/PGA ratio (Bray and Rodriguez-Marek, 2004),
(iv) Unexpectedly high response values at the pulse period (T_p) on the response spectra (Yang and Wang, 2012).

Several methods have been created to identify signals with the above mentioned features from regular, non-impulsive, ones. These methods are explained in Section 1.2. Furthermore, attempts have been made to calculate the probabilities of the occurrence of the pulse shaped signals. Previous studies of impulsive signal occurrences are explained in Section 1.3. The main purpose of such attempts is to incorporate the effect of impulsive signals into earthquake hazard maps in order reduce the earthquake risk.

The spatial distributions of impulsive signals are generalized by analyzing the physical features of the earthquakes that have created impulsive signals, spatial information and local soil conditions of the stations which have recorded impulsive waveforms. It has been observed that impulsive signals are more likely to occur when any of the following conditions are satisfied:

(i) Forward directivity (Somerville, 2003, 2005; Somerville et al., 1997; Spudich and Chiou, 2008),

(ii) Fling step effect,

(iii) Similar rupture velocity and shear-wave velocity of the bedrock of the site of interest.

Directivity effect can be observed when the site of interest is located in the direction of the fault rupture. The fling step effect occurs when the position of the seismic station is dislocated permanently. This effect can be seen in the displacement records as a permanent static displacement. Fling step and directivity effects are explained in Section 1.1.

As mentioned before, impulsive signals may create bigger demands on infrastructure than anticipated by building codes. Impulsive signals may affect the infrastructures as below:

(i) Spectral ratios can be locally amplified in the region where structural fundamental period is closer to the pulse period (Shahi and Baker, 2011a),

(ii) Structure will be loaded with considerable seismic energy in few pulses in the higher modes (Kalkan and Kunnath, 2006),

(iii) Big inelastic demands on intermediate period, multistory structures can be affected (Alavi and Krawinkler, 2004),

(iv) High-velocity pulses can place huge inelastic demands on multistory infrastructures (Hall et al., 1995).

Impulsive signals can be destructive to various types of infrastructure, such as idealized single and multi degree-of-freedom systems (Guo et al., 2018), seismically isolated structures (Mazza, 2018), and bridges (Antonellis and Panagiotou, 2013).

The above mentioned facts mean that it is important to identify impulsive signals and determine their probability of occurrence. In the identification process there are several important parameters — the period, starting and ending point and phase shift of the impulsive part of the velocity waveform. Impulsive signals can be better understood using these parameters. In probability of occurrence calculations, it is important to locate the regions where impulsive signals may occur with given geological settings.

In order to better understand the impulsive signals, two impulsive signal identification methods are created. Wavelet analysis is the first method that has been used. Several wavelets are used in order to identify the impulsive signals and mimic their features using a simpler waveform. The method and the differences between the previous studies are explained in Section 3.1. Impulsive signals are identified by their start and end points and phase shifts. Results are explained and compared with previous studies in Section 4.1.

The second pulse identification algorithm is created by using artificial neural networks. Machine learning algorithms and artificial neural networks have been used in earth science and seismology since the early 1990s. Several studies of the usage of machine learning algorithms in seismology are presented in Section 1.4. Using a large database of real and synthetic waveforms, a model than can detect impulsive signals is created. The methodology is explained in Section 3.2. The main advantage of the neural networks is the lack of strictly used mathematical formulas. As explained in Section 1.2 and Section 3.1, identification of impulsive signals are done by using data-fitting, inspection of the signal on time and frequency domains. Well known wavelets are used and several mathematical constrains are applied to real waveforms. By using machine learning algorithms it is possible to capture the nature of impulsive signals without using predefined constraints and thresholds. The trained neural network is able to detect impulsive signals with high accuracy. Results can be seen in Section 4.2.

In order to determine the probability of impulsive signals, parameters related to the fault and the signals are analyzed. In Section 3.3, parameters that are strongly correlated on impulsive signals are determined. The parameter space is created using features that can be detected for earthquakes or site of interests using the well-defined and previously used

parameters (Section 1.3). After the parameter selection, a multi-variate Bayesian classifier method is used to create models to calculate the probability of a pulse occurrence for given inputs. Probability distribution of observing impulsive signal maps are created using the model. Results are presented in Section 4.3. The spatial distribution of probability distributions of observing impulsive signals for several fault types and moment magnitudes (M_w) are plotted using artificial points on an imaginary fault. The generated maps are also shown for real earthquakes that created impulsive signals. Furthermore, smoothed probability distribution maps are created in order to have more applicable models.

In the final part, Section 5, outcomes of the thesis are summarised.

1.1. Physical Process of Impulsive Signals

Stations in near fault sites record earthquake signals which have great variations. These variations depend on parameters, such as fault mechanism, radiation pattern, directivity, stress drop, attenuation, scattering, basin effect and local soil conditions. As explained in Section 1, fling step effect and forward directivity are two major source of impulsive signals.

Fling step is a result of the permanent displacement due to the tectonic deformation that is created by a ruptured fault. This effect can be seen in very near fault stations. Mathematical representation of fling step effect can be seen below:

$$a(t) = \frac{2\pi D}{T_p^2} sin \left[\frac{2\pi}{T_p}(T - T_i) \right] \qquad (1.1)$$

In Eq. 1.1, a, D, T_p and T_i are acceleration waveform, maximum amplitude in displacement waveform, period of the sinusoidal pulse and pulse arrival times, respectively. Velocity and displacement waveforms can be obtained with integration of the signal. Graphical representation of these signals can be seen in Fig. 1.1.
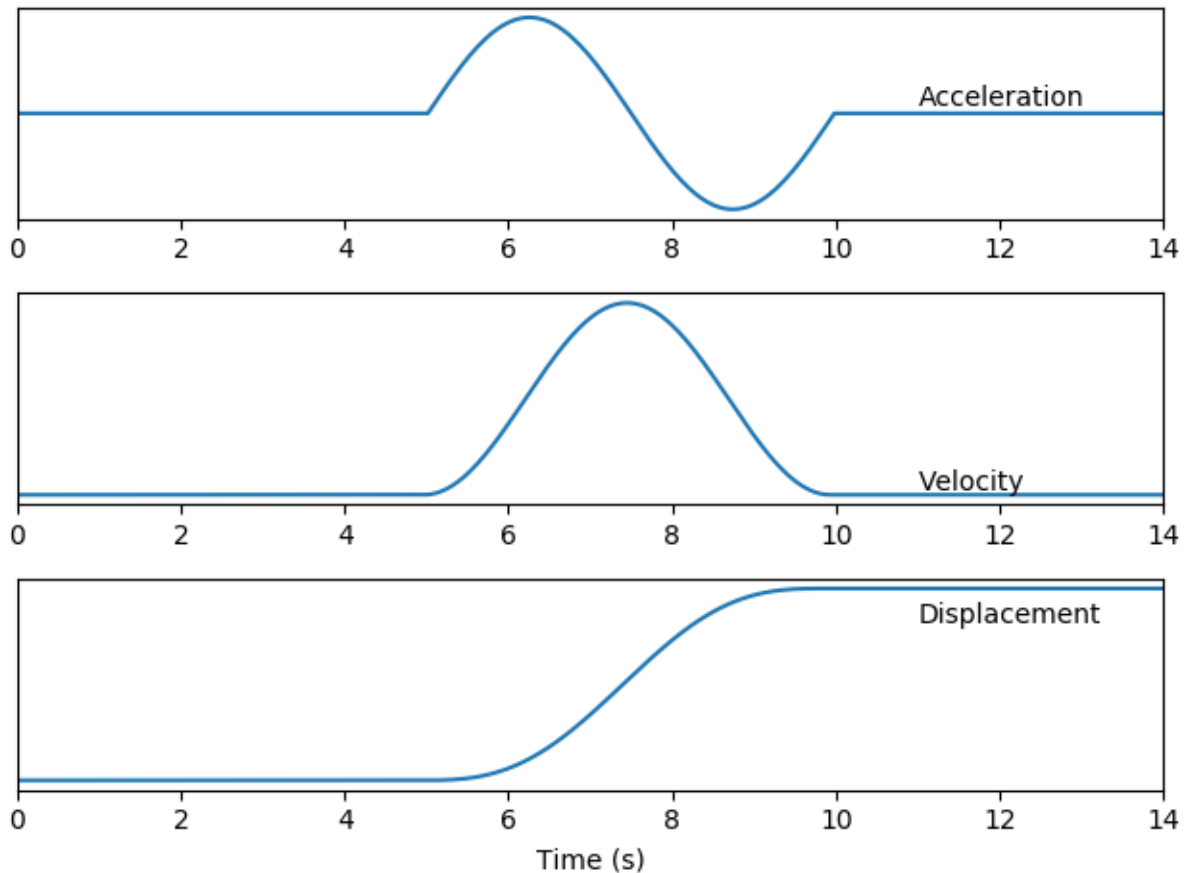
Figure 1.1: Recording of permanent displacement (fling step) in an ideal environment without the interference of other effects in displacement, velocity and acceleration records. Signals are normalized with maximum amplitude of each record.

Permanent displacement can be seen in the third graph in Fig. 1.1. In such a case, velocity waveform is a Gaussian-like signal. In fact Mavroeidis and Papageorgiou (2003) has assumed that velocity pulses can be explained by Gaussian signals with phase shift (Section 1.2.1).

Directivity effect is the second force that can create an impulsive signal. During an earthquake it is assumed that fault will be ruptured, starting from a point on a fault (hypocenter of the earthquake), and will propagate towards a direction over time. When the speed of the propagation is close to the velocity of the shear-wave of the region, source time function of the earthquake on the region where site of interest is located will be the cumulative. Graphical explanation of this phenomena can be seen in Fig. 1.2.
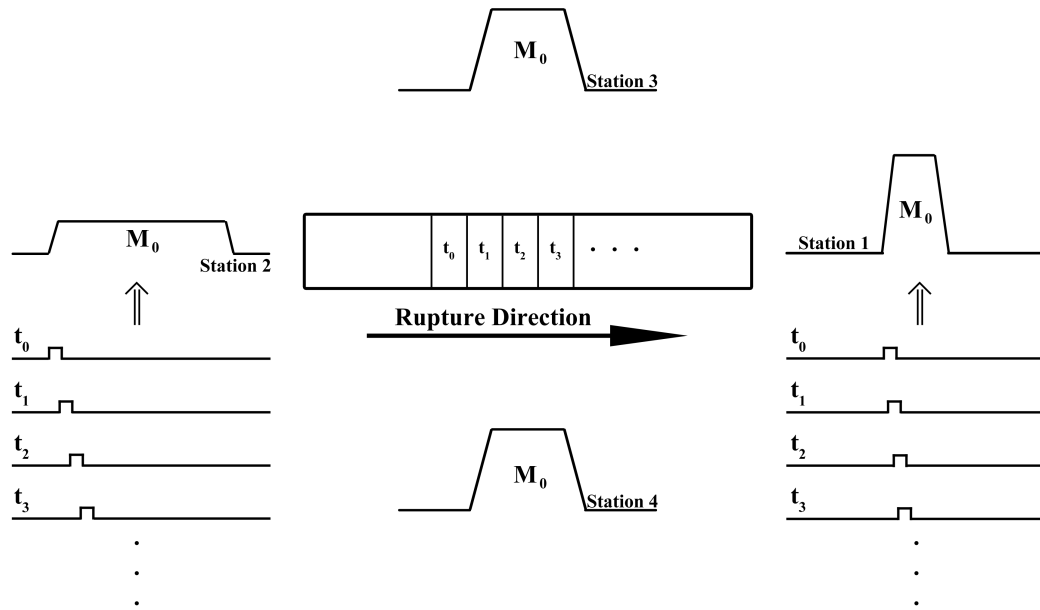
Figure 1.2: Temporal variation of source-time function for stations in different position with respect to the ruptured fault. Source-time function is the the cumulative effect of ruptured sub-faults over time. Figure is the modified versions of Benioff (1955); Lay and Wallace (1995).

The area under each station, which is the seismic moment, M_0 of the earthquake, is the same. Seismic moment is spread over time differently at Station 1 to Station 4. The rupturing process of the fault starts at time t_0 and rupture propagation direction is given with the arrow in Fig. 1.2. Source-time functions at stations are the cumulative effect of each ruptured part of the fault plane. They are given with time stamps, t_n. In Station 1 all seismic energy is concentrated in a short time range. It is due to the cumulative effect of the ruptured fault arriving to the station at the same time. Since rupture speed is close to the shear wave velocity of the region, when the second sub-region (t_1) of the fault is ruptured, the effect (seismic waves) of the first sub-region (t_0) is on the second sub-region. This cumulative effect of the rupture is the reason of unexpectedly high amplitudes for the stations that are located in the rupture propagation direction. In Station 2, effects of sub-regions spread over time. The source-time function will be different. Hence, an impulsive behavior does not occur. In Station 3 and Station 4, distribution of source-time function will be similar. These stations may have impulsive behavior depending on the amplitude of the signal and local soil conditions. Forward directivity effect can be formulated as below:

$$a(t) = \frac{\pi D}{T_p^2} sin\left[\frac{2\pi}{T_p}(T - T_i)\right] for \begin{cases} T_i < T < (T_i + 0.5T_p) \\ \\ (T_i + T_p) < T < (T_i + 1.5T_p) \end{cases} \qquad (1.2a)$$

$$a(t) = \frac{2\pi D}{T_p^2} sin\left[\frac{2\pi}{T_p}(T - T_i)\right] for \begin{cases} (T_i + 0.5T_p) < T < (T_i + T_p) \end{cases} \qquad (1.2b)$$

Graphical explanation of this phenomena can be seen in Fig. 1.3. In forward directivity effect, velocity pulse occurs as two half-cycle signals.
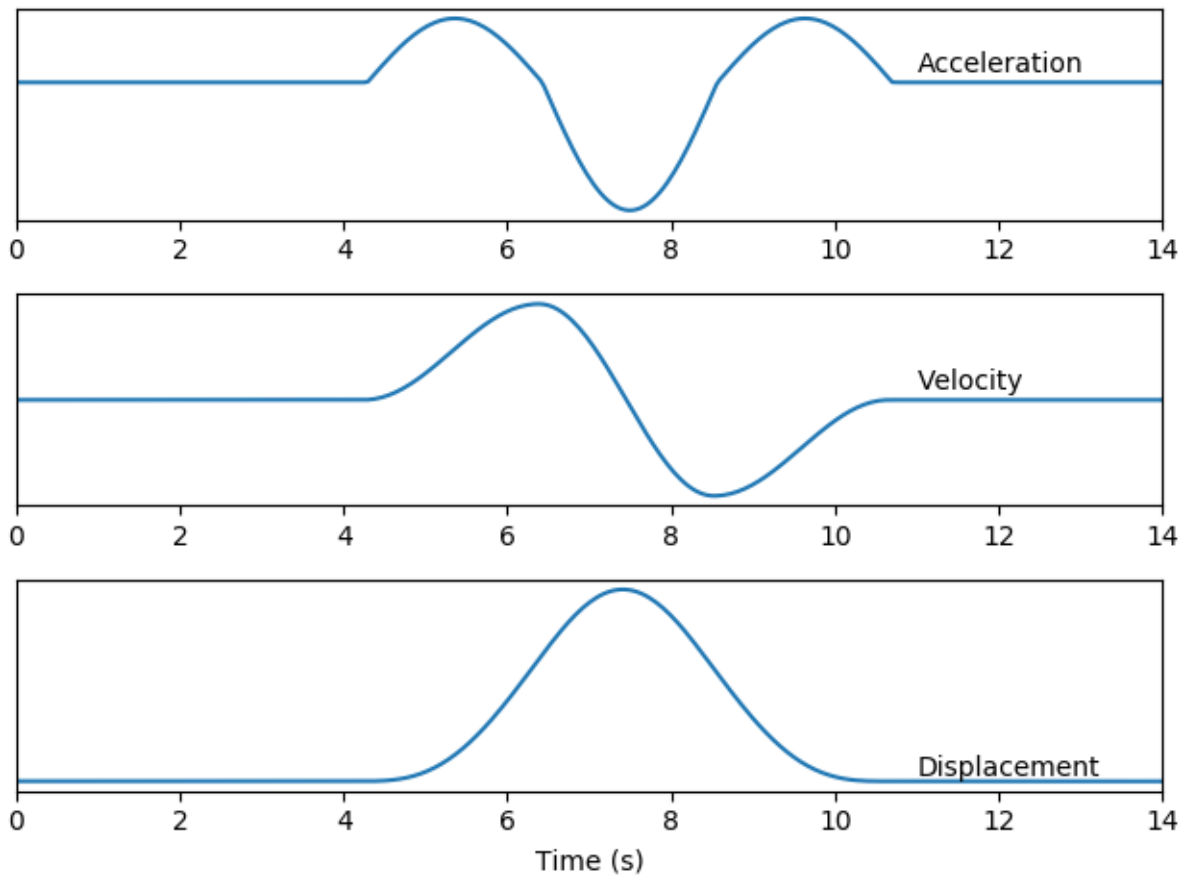
Figure 1.3: Recording of forward directivity in an ideal environment without the interference of other effect in displacement, velocity and acceleration records. Signals are normalized with maximum amplitude of each record.

Forward directivity effect (Station 1 in Fig. 1.2) can be seen in velocity in the latter

graph in Fig. 1.3. In such cases, velocity waveform has two half-cycles. Pulse classification algorithms in Section 1.2 are able to detect directivity related impulsive signals. Algorithm of Zhai et al. (2018) can count the number of half-cycles.

1.2. Previous Studies on Identification of Pulse Shaped Signals

The identification of an impulsive signal is vital since it contains information about the period of the pulse. Pulse shaped signals with long periods can be destructive for man made structures. Therefore various studies have been done in the last 2 decades. The focus of this section is the previous attempts to identify their presence in velocity waveforms. The methodologies of Chang et al. (2019, 2016); Ghaffarzadeh (2016); Kardoutsou et al. (2017); Mavroeidis and Papageorgiou (2003); Mena and Mai (2011); Mukhopadhyay and Gupta (2013); Shahi and Baker (2014); Yaghmaei-Sabegh (2010); Zhai et al. (2018) are explained.

1.2.1. Mavroeidis and Papageorgiou (2003)

Mavroeidis and Papageorgiou (2003) is one of the earliest studies that aimed to identify pulse shaped signals. In this study Gabor wavelets are used to explain impulsive signals. In order to characterize the impulsive signals, 4 key features are determined: amplitude, A, prevailing frequency, f_p, phase angle, v, and oscillatory character of the signal, γ. The proposed near-fault ground velocity pulse is created with the Eq. 1.3

$$v(t) = \begin{cases} A\dfrac{1}{2}\left[1+cos\left(\dfrac{2\pi f_p}{\gamma}(t-t_0)\right)\right]cos[2\pi f_p(t-t_0)+v] & t_0-\dfrac{\gamma}{2f_p} \leq t \leq t_0+\dfrac{\gamma}{2f_p} \text{ with } \gamma \geq 1. \\ 0 & \text{otherwise} \end{cases}$$

$$(1.3)$$

In Eq. 1.3, t_0 specifies the epoch of the envelope's peak. The pulse period, T_p, is calculated as $1/f_p$. The created pulse is manually tuned by checking the pseudo spectral response spectra and changing the parameters in Eq. 1.3.

1.2.2. Baker (2007)

Baker (2007) created the earliest example of automated pulse classification algorithm using wavelet analysis. A 4th order Daubechies wavelet is used in the study. Continuous wavelet analysis is implemented to the velocity waveforms. Wavelets are recursively applied on the velocity waveform. After each wavelet fitting, the fitted wavelet is subtracted from the residual signal. The first continuous wavelet that is applied on the waveform is used in order to calculate the period of the impulse. This method is also able to identify the impulses that has occurred due to the directivity effect. Pulse indicator (PI) is used for the identification of the pulse-shaped signals. PI formulation can be seen in Eq. 1.4

$$\text{PI} = \frac{1}{1 + e^{-23.3 + 14.6(\text{PGV ratio}) + 20.5(\text{energy ratio})}} \tag{1.4}$$

The PGV of the residual divided by the original PGV, and the energy of the residual divided by the energy of the original signal are referred as PGV ratio and energy ratio in Eq. 1.4, respectively. Waveforms with *PI* higher than 0.85 are identified as pulse-shaped signal.

1.2.3. Yaghmaei-Sabegh (2010)

Yaghmaei-Sabegh (2010) used the method of Baker (2007) to identify the pulse shaped signal. The part that differs Yaghmaei-Sabegh (2010) from Baker (2007) is the usage of alternative wavelets which are,

(i) Orthogonal wavelets
    (a) Haar wavelet
    (b) 7th order Daubechies wavelet
    (c) 4th order symlet wavelet
    (d) 2nd order oiflet wavelet

(ii) Biorthogonal wavelet series

    (a) 2nd order reverse biorthogonal wavelet

    (b) 1st order BiorSpline wavelet

In this study, it has been found that Haar wavelet is not appropriate for the identification of impulsive signals. Moreover the study concludes that BiorSpline wavelet is better on identifying pulse shaped signal behaviors than 4th order Daubechies wavelet, which is used in Baker (2007).

1.2.4. Mena and Mai (2011)

Mena and Mai (2011) used time-frequency analysis to describe pulse shaped signals. In this study, window lengths to calculate spectrograms are

$$T_w = -16.1 + 3.08 M_w. \tag{1.5}$$

In the study they conclude that if forward directivity is the driving force, most of the seismic energy is concentrated in a short time arrival, while in backward directivity, seismic energy is distributed all along the waveform. They calculated the region where 15 %, 35 %, 65 % and 85 % of the energy of the spectrogram are accumulated. They selected the 35 % as a threshold for the impulsive signal identifier. Then they fitted sine waves to the location of PGV and found the best fitting sine wave to identify the pulse period, T_p. They also calculated the number of cycles in the impulsive part of the waveform.

1.2.5. Mukhopadhyay and Gupta (2013)

Mukhopadhyay and Gupta (2013) calculated the waveform energy ratio between the entire waveform and the most energetic 20 half cycles in the waveform. Calculation of the energy of the half cycles is shown in Eq. 1.6

$$\text{fracEn}(i) = \frac{\int_a^b v^2(t)dt}{\int_0^T v^2(t)dt} \tag{1.6}$$

In Eq. 1.6, a and b are the starting end ending points of each half cycle, respectively. T is the length of the waveform. They determined a pulse identifier (PI) which can be seen in Eq. 1.7

$$PI = \frac{1}{1 + e^{7.64 - 27\text{fracEn}(1)}} \tag{1.7}$$

In Eq. 1.7, $fracEn(1)$ is the half cycle with the biggest energy. Signals are classified as Eq. 1.8

$$PI = \begin{cases} PI > 0.5 & \text{pulse type} \\ PI = 0.5 & \text{ambiguity} \\ PI < 0.5 & \text{non-pulse type} \end{cases} \tag{1.8}$$

In Eq. 1.8, $fracEn(1)$ values are found as greater than 0.29, less than 0.29 and equal to 0.29, respectively. Relation between half cycles before and after the main half cycle is investigated. It is accomplished by taking the energy ratio between the cycles that are lying on main half cycle's sides. Pulse period T_p is found by taking the Fourier amplitude spectra of the smoothed version of impulsive part of the waveform.

1.2.6. Shahi and Baker (2014)

Shahi and Baker (2014) classification algorithm uses wavelet-based signal processing to detect pulse-shape signals in the area with the largest value of velocity (PGV). This algorithm is the improved version of Baker (2007). The algorithm can differentiate early and late arrival pulses by analyzing the arrival of PGV. Early arrivals of PGV generally indicate directivity effects.

Classification algorithm uses two criteria to determine whether the signal has impulsive or non-impulsive behavior. First criterion is the hazardousness of the signal. If PGV is less than $30\,\mathrm{cm\,s^{-1}}$, it is considered as non-hazardous signal. Second criterion is that the pulse indicator (*PI*) values should be bigger than 0. Calculation of PI has two stages. In the first stage principal component has been found (*PC*) as explained in Eq. 1.9:

$$PC = 0.63 \times (\text{PGV ratio}) + 0.777 \times (\text{energy ratio}) \tag{1.9}$$

PGV ratio indicates the ratio of the PGV value of the residual signal and the PGV value of the original signal, which is calculated by subtracting the original signal from that produced by the 4$^{\text{th}}$ order Daubechies wavelet signal. *energy ratio* is the ratio calculated by dividing the power of the residual signal by the power of the original signal. In the second stage PI is calculated as Eq. 1.10:

$$PI = 9.384(0.76 - PC - 0.00616\text{PGV})(PC + 6.914 \times 10^{-4}\text{PGV} - 1.072) - 6.179 \tag{1.10}$$

If the signal is considered as pulse shaped signal, then 4$^{\text{th}}$ order Daubechies wavelet is fitted to the entire waveform. It is hard to determine where the impulsive part of the signal

starts and ends by using Shahi and Baker (2014) classification algorithm.

1.2.7. Chang et al. (2016)

Chang et al. (2016) uses an energy-based classification algorithm. This algorithm is the improved version of Zhai et al. (2013). The algorithm determines a region around the PGV and the energy ratio between the pulse region and the total energy of the signal by taking the squared values on both signals. The region around PGV is calculated by using a least-square fitting for various pulse periods, then the one with the smallest residual is used for the pulse region. The energy ratio is then calculated as Eq. 1.11:

$$E(t) = \frac{\int_{t_s}^{t_e} v^2(\tau)d\tau}{\int_0^\infty v^2(\tau)d\tau} \tag{1.11}$$

If the ratio between the pulse region energy, the numerator part of the Eq. 1.11, and the total energy, the denominator of Eq. 1.11, exceeds 0.34, the signal is considered as a pulse-shaped signal. t_s and t_e represent the starting and ending points of the impulse part in the time axis and v^2 represents the velocity time history of the signal.

If the signal is considered as pulse shaped, a signal is extracted by using least square fitting. The fitted waveform is created as in Eq. 1.12

$$v_p(t; V_p; T_p; N_c; T_{pk}; \varphi) = V_p exp\left(-\frac{\pi^2}{4}\left(\frac{t-T_{pk}}{N_c T_p}\right)^2\right) \times \cos\left(2\pi\frac{t-T_{pk}}{T_p} - \varphi\right) \tag{1.12}$$

v_p is the extracted velocity pulse. The model has five parameters: peak pulse velocity, V_p; pulse period, T_p; number of cycles in the pulse, N_c; location of the pulse, T_{pk}; and phase of the pulse, φ. In this study, V_p equals PGV and thus T_{pk} is accordingly determined. t_s and

t_e are taken as the time locations of the two local peaks or troughs in the predominant pulse extracted by Eq. 1.12.

1.2.8. Ghaffarzadeh (2016)

Ghaffarzadeh (2016) has implemented S-transform to the recordings of ground motions, and time-frequency plots of the S-transform. S-transform is utilized for selection of the ground motions containing pulses. In the next stage, a method is presented based on the convolution theorem to extract pulse-like motion from the velocity time history of ground motions. This process evokes the main pulses from the ground motion, if they exist.

They introduced an indicator for the initial classification of near-fault earthquakes which is the energy gyration factor. This factor indicates the amount of the concentration of the wave energy in the S-transform. It defines the ratio of the second moment of the S-transform absolute values to the S-transform integral over time and frequency, which can be seen in Eq. 1.13

$$g_E = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d(t,f) \, |S(t,f)| \, dt \, df}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |S(t,f)| \, dt \, df} \tag{1.13}$$

where $|S(t,f)|$ is the absolute value of the S-transform, $d(t,f)$ is the distance between coordinate (t,f), in which the S-transform is calculated at the center of the coordinate system, $(0,0)$, and g_E is the energy gyration factor. A bigger g_E indicates the focusing of waveform energy in a small time and frequency range which is used as an indicator of impulsive signal. Impulsive part of the signal is calculated by applying inverse S-transform on a filtered signal which is filtered by using π-shaped window function. Then pulse period, T_p, is found by using the Fourier amplitude spectra.

1.2.9. Kardoutsou et al. (2017)

Kardoutsou et al. (2017) created a variable which is called the cross-correlation factor of the significant pulse and the original record. The factor can determine whether the signal is pulse-shaped or not. New pulse indicator, *PI*, has three possibilities. If PI is greater than 0.65, the signal is considered as pulse shaped. If it is less than 0.55, it is determined as non-impulsive. When correlation coefficient get a value between 0.65 and 0.55, the signal has an ambiguity.

In this study, the pulse period, T_p, is determined from the peak of the product spectrum $S_d \times S_v$, where S_d is the displacement response spectrum and S_v is the velocity response spectrum for 5% damping. If the signal is pulse shaped, a wavelet which is also used by Mavroeidis and Papageorgiou (2003) is fitted to the velocity waveform. For the determination of the wavelet's amplitude, *A* is found as in Eq. 1.14,

$$A = \frac{4\xi PS_{v,\xi}(T_p)}{(1 - e^{2\pi\gamma\xi}) \times [+1(\gamma-1)\xi]} \tag{1.14}$$

In Eq. 1.14, $PS_{v,\xi}(T_p)$ is the value of the pseudo velocity response spectrum of the ground motion for the period of T_p, where damping (ξ) is 0.05 and the duration (γ) is unknown but varies between 0 up to 5. For a given set of (A, γ) and phase (v) vary between 0° and 360°. Corresponding wavelets are calculated if wavelet amplitude (*A*) is less than PGA, PGV and PGD. Then the wavelet with the largest cross correlation coefficient is selected.

1.2.10. Zhai et al. (2018)

In this study, pulse shaped signals are divided as single, multiple, or irregular pulses. Energy based significant velocity half-cycles are found. Energy formula of this study is the same as Eq. 1.11. If the energy ratio of the half-cycle is larger than 0.1, it is considered as significant half cycle. Ground motions then are categorized with their significant half-cycles

between Type 0 and Type 5. Integers of the categories indicate the number of half-cycles in the waveform. Classifying the pulse-like ground motion based on the energy-threshold levels is explained in Eq. 1.15

$$
\begin{aligned}
\text{Type-1} &= 0.30 \\
\text{Type-2} &= 0.42 \\
\text{Type-3} &= 0.50 \\
\text{Type-4} &= 0.57 \\
\text{Type-5} &= 0.73
\end{aligned}
\tag{1.15}
$$

1.2.11. Chang et al. (2019)

Chang et al. (2019) developed an algorithm that used both velocity and acceleration data to identify the impulsive signals. In order to erase the high frequency content of acceleration data, wavelet packet transform is applied as a filter. In order to detect velocity and acceleration pulses, cumulative squared velocity (CSV) and cumulative squared acceleration (CSA) are calculated. Relative energy of the velocity pulse (EVP) and relative energy of acceleration pulse (EAP) are used in order to make an identification.

$$
\text{CSV}(t) = \frac{\int_0^t v^2(\tau)d\tau}{\int_0^T v^2(\tau)d\tau} = \frac{\sum_{i=1}^n v_i^2}{\sum_{i=1}^N v_i^2}
\tag{1.16}
$$

t_s and t_e are found with the same method that is used by Chang et al. (2016). In Eq. 1.17, T, $v(t)$, N and v_i are duration of the ground motion, velocity waveform, and total number of the velocity time series, respectively. v_i is the amplitude of the velocity waveform in the time location i. Limits of t and n are 0 to T and 0 to N, respectively. EVP is calculated as Eq. 1.17.

$$EVP = CSV(t_e) - CSV(t_s) \tag{1.17}$$

t_s and t_e are calculated with Eq. 1.12. Velocity pulses are found by using Eq. 1.16 and Eq. 1.17.

Acceleration pulses are found in a similar way to velocity pulses. Eq. 1.18 and Eq. 1.19 explain the methodology to determine the acceleration pulses.

$$CSA(t) = \frac{\int_{t_s}^{t} a^2(\tau)d\tau}{\int_{t_s}^{t_e} a^2(\tau)d\tau} = \frac{\sum_{i=1}^{k} a_i^2}{\sum_{i=1}^{K} a_i^2} \tag{1.18}$$

$a(t)$ is the truncated acceleration waveform with $t \in [t_s, t_e]$; $a(i)$ is the amplitude of the acceleration waveform in time i; K represents the number of discretized time series; and k has limits between 0 to K. *EAP* is the indicator for acceleration pulses and is calculated as in Eq. 1.19.

$$EAP = CSS(t_2) - CSA(t_1) \tag{1.19}$$

t_1 and t_2 are zero crossings in time series of the half-cycle pulse in the largest energy area. In general, t_1 and t_2 are different in time from t_s and t_e.

1.3. Previous Studies on Occurrence of Impulsive Signals

Pulse shaped signals have been studied in many aspects such as their probabilities of occurrence with M_w, distance, fault type and so on. In this section, previous attempts to categorize impulsive signals by using different metrics have been explained briefly.

1.3.1. Somerville (2003)

One of the earliest attempts for categorizing impulsive signals has been done by Somerville (2003). Period of the impulsive signal, T_p is explained with the M_w of the earthquake as in Eq. 1.20a and Eq. 1.20b:

$$\log T_p = -3.17 + 0.5 M_w, \sigma_{log(T_{p)}} = 0.80 \qquad (1.20a)$$

$$\log T_p = -2.02 + 0.346 M_w, \sigma_{log(T_{p)}} = 0.62 \qquad (1.20b)$$

Eq. 1.20a and Eq. 1.20b are produced for seismic stations that are located on rock and soil sites, respectively. Furthermore, it is concluded that velocity response spectrum contains peak values on the 0.85 of the T_p value.

1.3.2. Iervolino and Cornell (2008)

Iervolino and Cornell (2008) identifies the total seismic hazard as the linear combination of pulse and no pulse hazard curves. For a single fault line these curves are calculated as in Eq. 1.21.

$$\lambda_{S_a,NS\&\text{pulse}}(x) = \upsilon \int_m \int_r \int_{\underline{Z}} \int_{t_p} P\left[\text{pulse}|m,r,\underline{z}\right] \times G_{S_a|\text{pulse},M,R,\underline{Z},T_p}(x|m,r,\underline{Z},t_p)$$

$$\times f_{T_p|\underline{Z},M,R} f_{\underline{Z}|M,R} f_{M,R} dt_p d\underline{z} dm dr \qquad (1.21\text{a})$$

$$\lambda_{S_a,NS\&\text{no pulse}}(x) = \upsilon \int_m \int_r \int_{\underline{Z}} (1 - P\left[\text{pulse}|m,r,\underline{z}\right]) \times G_{S_a|\text{no pulse},M,R}(x|m,r)$$

$$\times f_{\underline{Z}|M,R} f_{M,R} d\underline{z} dm dr \qquad (1.21\text{b})$$

Pulse occurrence and no pulse occurrence probabilities are calculated as in Eq. 1.21a and Eq. 1.21b, respectively. υ is the mean rate of events, M is the magnitude of an event, R is the source-to-site distance, dt_p and $d_{\underline{Z}}$ are the intervals of the integration of T_p and other parameters that depend on attenuation ($\underline{Z}$), respectively. $G_{S_a|pulse,M,R,\underline{Z},T_p}$ is the cumulative distribution function of spectral acceleration, S_a, for given $M,R,\underline{Z}$ and T_p. $f_{T_p|\underline{Z},M,R}, f_{\underline{Z}|M,R}$ and $f_{M,R}$ are probability distribution functions of T_p, $\underline{Z}$ and M,R for given $M,R,\underline{Z}$ and M,R, respectively.

Simple and multi variable regression are used to determine the occurrence of the impulsive signal on fault-normal direction.

In simple logistic regression, it has been found that R, θ and s can explain the probability of occurrence for strike slip faults. Pulse occurrence is decreasing with increasing R and θ angle, whereas it is increasing with increasing s. Pulse occurrence has the same trends with R, d and ϕ angle as in strike slip faults.

L, s, R and θ are length of the fault, distance from the epicenter and the projection of the site on the fault, distance from the fault line and the station and the angle between the fault line and the station, respectively. For not strike slip faults W, d, R and ϕ are width of the fault, distance from the hypocenter and the projection of the site on the fault in the ground, distance from the fault line in the ground and the station and the angle between the fault line in the ground and the station, respectively. Parameters can be seen in Fig 1.4.
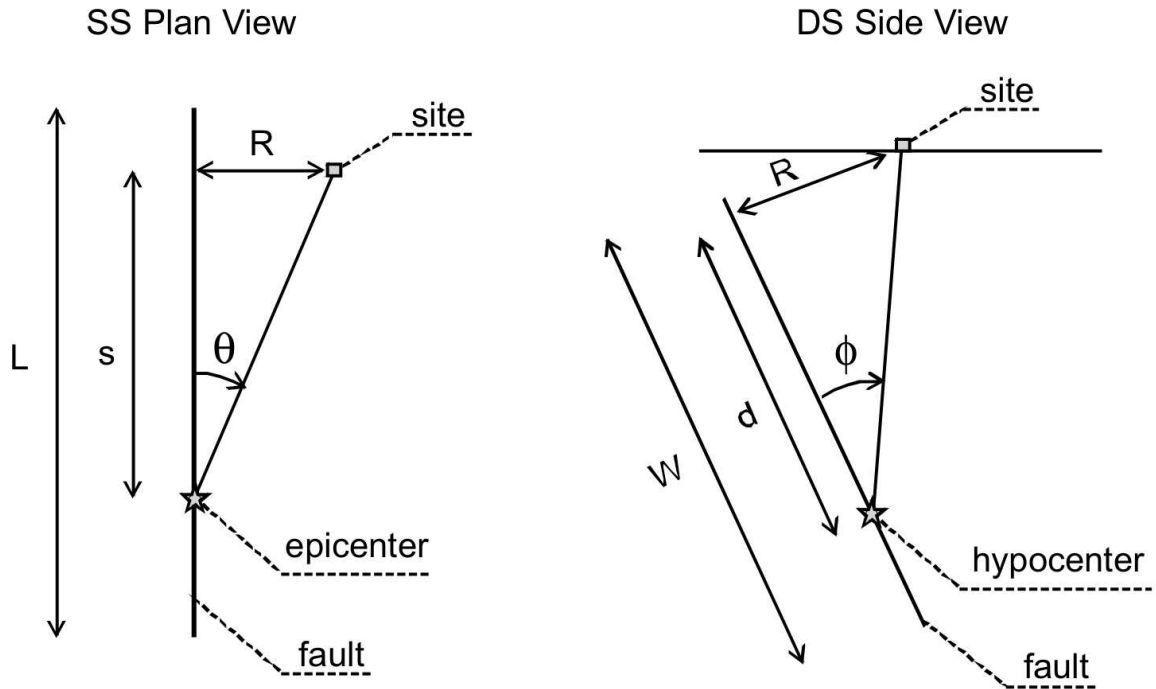
Figure 1.4: Strike slip and not strike slip schematics from Somerville et al. (1997).

In multi variable logistic regression, probability of pulse occurrence can be explained with combination of R, s and θ. Pulse probability formulas are created as in Eq. 1.22.

$$P\big[\text{pulse}|R,s,\theta\big] = \frac{e^{0.859-0.111R+0.019s-0.044\theta}}{1+e^{0.859-0.111R+0.019s-0.044\theta}} \tag{1.22a}$$

$$P\big[\text{pulse}|R,d,\phi\big] = \frac{e^{0.553-0.055R-0.027d-0.027\phi}}{1+e^{0.553-0.055R-0.027d-0.027\phi}} \tag{1.22b}$$

1.3.3. Shahi and Baker (2011a)

Shahi and Baker (2011a) used multi variable regression to determine the probability of pulse occurrence by using both fault-normal and fault parallel data. Method of Baker (2007) is used in order to determine pulse shaped signals. Pulse probabilities of strike slip and not

strike slip cases can be found in Eq. 1.23a and Eq. 1.23b.

$$P\big[\text{pulse}|R,s\big] = \frac{1}{1+e^{0.642+0.167R-0.075s}} \tag{1.23a}$$

$$P\big[\text{pulse}|R,d,\phi\big] = \frac{1}{1+e^{0.128+0.055R-0.061d+0.036\phi}} \tag{1.23b}$$

Furthermore Shahi and Baker (2011a) rotated the signals from 0° to 90° (α_{rot}) and determined if the signal is impulsive or not by using Baker (2007). It is found that probability of impulsive signals are increasing when orientation is normal to the strike ($\alpha = 90^{\text{o}}$) and decreasing when orientation is parallel to strike ($\alpha = 0^{\text{o}}$). Shahi and Baker (2011a) has also updated the relationship between T_p and M_w which is studied by Somerville (2003) by the help of increasing number of near fault recordings and numerical methods to determine impulsive signals. Formulation of this relation is explained in Eq. 1.24.

$$\mu_{\ln T_p} = -5.73 + 0.99 M_w \tag{1.24a}$$

$$\sigma_{\ln T_p} = 0.56 \tag{1.24b}$$

In Eq. 1.24a, lnT_p can be assumed as normally distributed with mean (μ_{lnT_p}) and standard deviation is given in Eq. 1.24b.

1.3.4. Chioccarelli and Iervolino (2013)

In this study, pulse probability equations of Iervolino and Cornell (2008) are found as in Eq. 1.22. Computation of hazard is updated for strike slip and dip slip faults. New methodology is formalized as in Eq. 1.25. Eq. 1.25a and Eq. 1.25c are adapted for pulse occurrence, whereas Eq. 1.25b and Eq. 1.25d are used for no pulse occurrence on strike slip

and dip slip faults, respectively.

$$\lambda_{S_a,ss\_pulse}(s_a^*) = \upsilon \int_{t_p} \int_m \int_l \int_p \int_e P\big[\text{pulse}|l,p,e\big] \times G_{S_a,mod|T_p,M}(s_a^*|t_p,m,l,p,e)$$

$$\times f_{T_p|M}(t_p|m) f_{P,E|L}(p,e|l) f_{L|M}(l|m) f_M(m) dt_p dm dl dp de \qquad (1.25a)$$

$$\lambda_{S_a,ss\_nopulse}(s_a^*) = \upsilon \int_m \int_l \int_p \int_e 1 - P\big[\text{no pulse}|l,p,e\big] \times G_{S_a,mod|M,L,P,E}(s_a^*|m,l,p,e)$$

$$\times f_{P,E|M,L}(p,e|m,l) f_{L|M}(l|m) f_M(m) dm dl dp de \qquad (1.25b)$$

$$\lambda_{S_a,ds\_pulse}(s_a^*) = \upsilon \int_{t_p} \int_m \int_l \int_p \int_h \int_{x_t} P\big[\text{pulse}|l,p,h,x_r\big] \times f_{T_p|M}(t_p|m)$$

$$G_{S_a,mod|T_p,M,X_r,L,P}(s_a^*|m,l,p,x_r)$$

$$\times f_{P,H|L}(p,h|l) f_{X_r|L}(x_r|l) f_{L|M}(l|m) f_M(m) dt_p dm dl dp dh dx_r \qquad (1.25c)$$

$$\lambda_{S_a,ds\_nopulse}(s_a^*) = \upsilon \int_m \int_l \int_p \int_h \int_{x_t} 1 - P\big[\text{no pulse}|l,p,h,x_r\big] \times G_{S_a|M,X_r,L,P}(s_a^*|m,l,p,x_r)$$

$$\times f_{P,H|L}(p,h|l) f_{X_r|L}(x_r|l) f_{L|M}(l|m) f_M(m) dm dl dp dh dx_r \qquad (1.25d)$$

Features of Eq. 1.25 are:

(i) Pulse probability is independent on magnitude (M) and pulse period (T_p);

(ii) Probability density function of pulse period is depending on magnitude;

(iii) Probability density function of rupture (p) and epicenter's position (e) is dependent on the rupture length (l);

(iv) $G_{S_a,mod|T_p,M}$ is the modified version of Eq. 1.21;

(v) Summation of pulse occurrence and not pulse occurrence is 1;

(vi) In Eq. 1.25c and Eq. 1.25d, e, is replaced with hypocenter, h.

(vii) In order to calculate pulse probability, distance of the rupture to fault bounds (X_r) is necessary to solve Eq. 1.25c and Eq. 1.25d.

(viii) Knowledge of probability density function of X_r to l is needed for calculating the Eq. 1.25c and Eq. 1.25d.

(ix) s_a^* is defined as the elastic spectral acceleration at a fixed spectral period exceeding the

chosen intensity value.

(x) Position of the rupture on the fault is given as p.

1.3.5. Shahi and Baker (2014)

Pulse probabilities, pulse periods with M_w are also calculated in the study of Shahi and Baker (2014) along with the pulse identification algorithm. Regression has been done for the parameters of R, s, d, θ and ϕ. Both pulse probabilities and pulse periods with M_w are calculated for the directivity related impulsive signals and for the general case. In the case of the directivity effect, pulse probabilities for strike slip and non strike slip are explained in Eq. 1.26.

$$P(\text{directivity}|R, s, \theta, \text{strike slip}) = \frac{1}{1 + e^{(0.7897 + 0.1378R - 0.3533\sqrt{s} + 0.020\theta)}}$$
(1.26a)

$$P(\text{directivity}|R, d, \phi, \text{non strike slip}) = \frac{1}{1 + e^{(2.118 + 0.129R - 0.895\sqrt{d} + 0.025\phi)}}$$
(1.26b)

Relation between the pulse period and M_w for directivity affected impulses are as in Eq. 1.27.

$$\ln T_p = -6.256 + 1.084M_w$$
(1.27)

Pulse probabilities for the general case are found as in Eq. 1.28.

$$P(\text{directivity}|R, s, \theta, \text{strike slip}) = \frac{1}{1 + e^{(0.457 + 0.126R - 0.244\sqrt{s} + 0.013\theta)}}$$

(1.28a)

$$P(\text{directivity}|R, d, \phi, \text{non strike slip}) = \frac{1}{1 + e^{(0.544 + 0.069 - 0.259\sqrt{d} + 0.019\phi)}}$$ (1.28b)

Relation between the pulse period and M_w for general case is as in Eq. 1.29.

$$\ln T_p = -6.55 + 1.12 M_w$$

(1.29)

Calculation errors in non strike slip probabilities are fixed in Shahi and Baker (2019). Results of this study provide updated coefficients that have been calculated in Shahi and Baker (2011a). Interpretation of the equations are the same with Section 1.3.3.

1.3.6. Fayjaloun et al. (2016)

In this study, spatial dependency of T_p on parameters is used to explain rupture process. In order to calculate the T_p, synthetic signals are created. Pulse period is determined as in Eq. 1.30.

$$T_p = \left(\frac{D}{V_r} + \frac{clsD}{V_s} - \frac{hypD}{Vs}\right) + T_{rise}$$

(1.30)

Variables in Eq. 1.30 are D, the length of the fault that rupture toward the station; *clsD*, the closest distance between station and ruptured fault area; *hypD*, the hypocentral

distance; V_r, the rupture speed; V_s, the shear wave velocity of the bedrock, and T_{rise} the rise time of the earthquake. This formula works only when shear wave velocity is less than rupture speed. This formula works well, if the site of interest is not located away from the termination of rupture and V_r/V_s ratio is between 0.8 to 1. Otherwise found T_p values would be underestimated. In this study, the relation between T_p and M_w is not explicit, but it is related since D is related with M_w.

1.3.7. Scala et al. (2018)

L'Aquila earthquake ($M_w = 6.3$) is used in this study. Both real seismic signals and synthetic seismograms are analyzed. Synthetic seismograms are created with varied hypocenter, V_r, V_s, T_{rise} and slip distributions. In total 1500 different source models are created. It is found that, increasing T_{rise} increases the occurrence of pulse and its periods. Increasing V_r increases the probability of pulse occurrence while decreasing the period of the pulse. Increasing hypocentral depth increases the pulse occurrence which can be related with up-dip directivity. Increasing distance is increasing the pulse occurrence in distances between 0-30 km on both foot wall and hanging wall records, although occurrence of the impulsive signals are more common on the foot wall. Pulse period is explained as in Eq. 1.31. Parameters are graphically presented in Fig. 1.5.

$$T_p \approx T_{rise} + \frac{L_{dir}}{V_r} - \frac{1}{V_s} \int_0^{L_{dir}} \cos\theta(x)dx \approx T_{rise} + \frac{L_{dir}}{V_r} - \frac{L_{dir}}{V_s}\cos(\theta) \qquad (1.31)$$
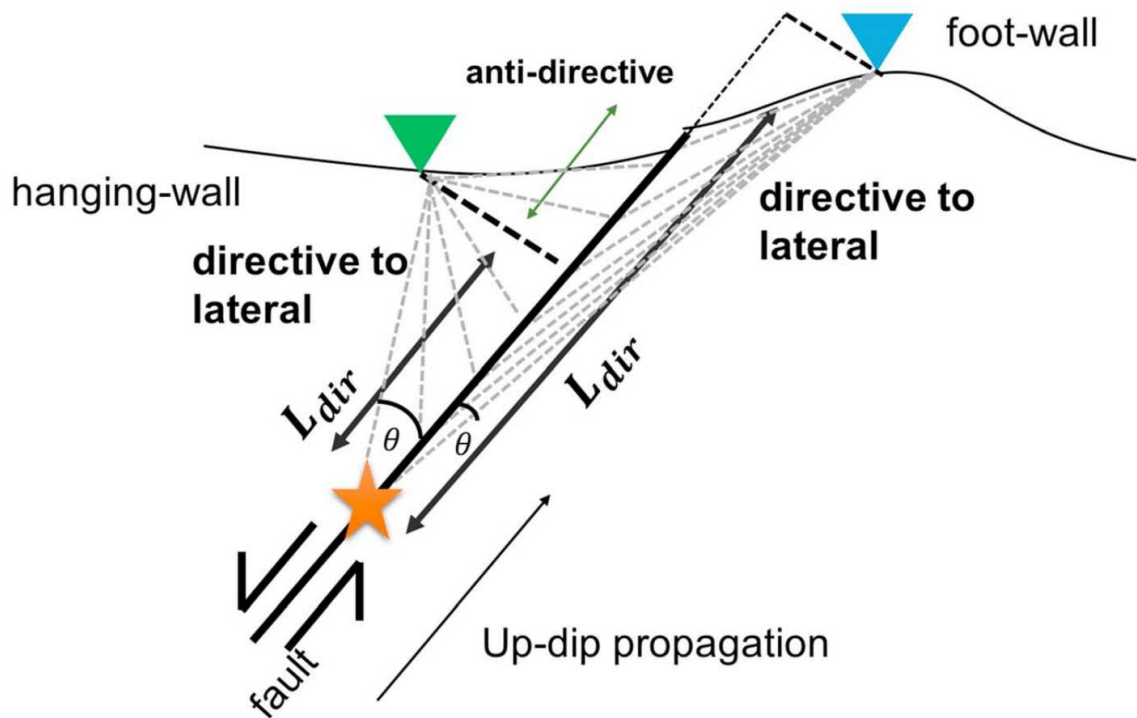
Figure 1.5: Parameters that affect the pulse occurrence that is found by Scala et al. (2018). Orange star, blue reversed triangle and green reversed triangle are representing hypocenter, station placed on foot wall and station placed on hanging wall, respectively. L_{dir} is the closest distance between the hypocenter and the projection of the station on fault line beneath the earth; θ is the angle between the fault line and the line which has the closest distance between the hypocenter and the station.

1.4. Machine Learning in Seismology

Machine learning techniques are implemented in seismological problems in the last 3 decades. Dysart and Pulli (1990) used event classification through the use of trained neural networks. Wang and Teng (1995) used artificial neural detectors recursive STA/LTA time series as input data, and moving window spectrograms as input data to detect earthquake signals. Wang and Teng (1997) used artificial neural networks (ANN) automatic picking of local and regional S phase. Falsaperla et al. (1996) used eight neurons in a single hidden layer to identify the volcanic explosions at the Stromboli volcano, Italy. Shimshoni and Intrator (1998) used ANN to determine seismic signals. Scarpetta et al. (2005) used supervised

neural network to detect local seismic signals and volcano-tectonic earthquakes. Spectral features of the signals and the parameterized attributes of their waveforms have been used as input for this network. Several stations from Mt. Vesuvius monitoring network, for the discrimination of low magnitude volcano-tectonic earthquakes and transient signals caused by either artificial (quarry blasts, underwater explosions) and natural (thunder) sources. A similar study has been done by Del Pezzo et al. (2003) in order to discriminate the volcano-tectonic earthquakes and underwater explosions done by fishermen in Phlegraean Fields on Southern Italy. Ambrožič and Turk (2003) used ANN to classify earthquakes and mining explosions in Velenje Coal Mine, Slovenia.

Machine learning algorithms are used in petroleum geophysics as well. Studies of Al-Kaabi et al. (1993); Ali et al. (1994); McCormack et al. (1993); Mohaghegh et al. (1996, 1995) are some early examples of neural networks and machine learning applications in geophysics.

High amount of data in earth science and improvements on computers allowed scientists to use machine learning algorithms more efficiently. It became a hot topic also in earth science, since there are huge amounts of data in geosciences to be processed. Adeli and Panakkat (2009) used probabilistic neural network (PNN) for predicting the magnitude of the largest earthquake with a given future. Eight parameters are used in this study. The model yields good prediction accuracy for earthquakes of magnitude between 4.5 and 6.0. Asencio-Cortés et al. (2018) used seismicity database of California to forecast earthquakes. Martínez-Álvarez et al. (2011) used b values and made predictions about the major earthquakes. Reyes et al. (2013) used artificial neural networks (ANN) to predict earthquakes in Chile. Florido et al. (2018) implemented several machine learning algorithms to discover seismic patterns. Paolucci et al. (2018) used ANN to predict the response spectral ordinates at short periods. Shang et al. (2017) Principal Component Analysis and ANN to discriminate the microseismic events from quarry blasts. 22 seismic parameters of 1600 events have been used on this study. Titos et al. (2018b) used recurrent neural networks (RNN), long short-term memory (LSTM), and gated recurrent unit (GRU) to detect and classify continuous sequences of volcano-seismic events at the Deception Island Volcano, Antarctica. Titos et al. (2018a) evaluated the classification performance on seven different classes of isolated

seismic events by using two different deep neural networks (DNNs). Dickey et al. (2019) used seismic detection algorithm which is capable of carrying out an array-like performance from a single trace. They achieved this directly, by training their single-trace detector against labeled events from an array catalog, and by utilizing a deep temporal convolutional neural network. Ross et al. (2018a,b) used CNN algorithms to detect P wave arrival and first motion polarity detection.

Machine learning algorithms are also implemented to engineering seismology and civil engineering problems. Wu et al. (1992) used neural networks (NN) in structural damage assessment. Alizadeh et al. (2018); Amirifar and Shafiee (2018); Xie et al. (2017) used machine learning algorithms for loss estimation and vulnerability problems. Rafiei and Adeli (2017) used classification algorithm on earthquake early warning systems. Nejad et al. (2018) used microzonation algorithm that combines NNs and geographic information system (GIS) to investigate shear wave velocity and soil type microzonation. Möller et al. (2009) implemented NN to performance level, for a specific combination of the design parameters. The application is done on a reinforced concrete, multi-story structure with proper seismic demands in Mendoza, Argentina. In Shahi and Baker (2011b), generalized linear models (GLM) were used in combination with a model fitting for predicting the probability of near-fault earthquake ground motion pulses without using the time series data.

2. DATA

In this study, 21 458 waveforms are used. These signals are recorded as a result of 399 earthquakes. Only crustal earthquakes with magnitudes bigger than 5.5 are taken into account. Seismic records from near fault stations (Epicentral distance (R_{ep}) $\leq$ 115 km) are selected for the analysis.

Signals are used in velocity waveform for the pulse identification process. PGA, PGV, PGV/PGA, R_{ep}, Joyner-Boore distance (R_{jb}) and azimuth angle of each station are used with the earthquake information of M_w (Appendix A.1), rake angle (Appendix A.2), focal mechanism (Appendix A.2), slip of the fault (cm), rupture width (Appendix A.3), rupture length (Appendix A.3), rupture area (Appendix A.3), seismic moment (M_0, Appendix A.1) and stress drop (Appendix A.4) for the calculation of spatial distribution of impulsive signal probabilities. Database information is explained in Section 2.1.

Collected waveforms are stored as acceleration waveforms and converted to velocity by using first order integration of data using the trapezoidal rule. Mean of the signal are removed. Sampling rates are not homogenised.

2.1. Spatial Distribution of Database

In this section, distribution of seismic stations are presented. Database that is created for the study is sub-divided into countries (combined or individual), continents and previous databases.

2.1.1. Chile

In Chile, there are 488 earthquake signals, recorded due to 38 earthquakes by 45 stations. All earthquakes have occurred due to the subduction of Nazca plate under the South American plate, which produce reverse focal mechanism (Fig. 2.1).
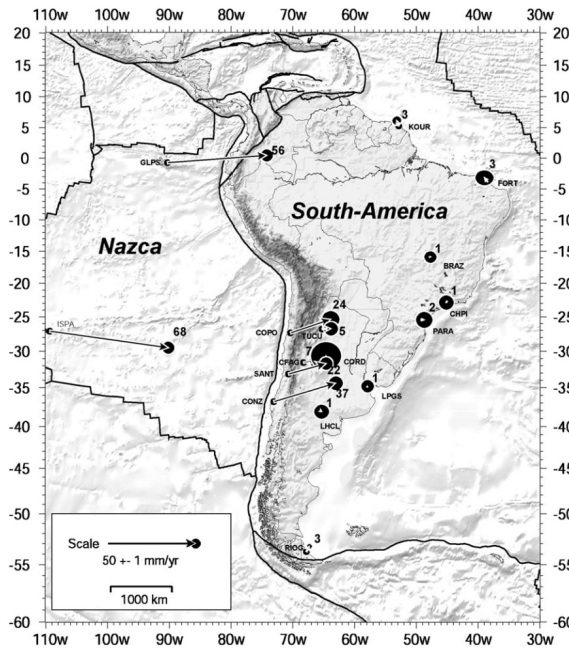
Figure 2.1: Rates of motion between the South American plate and Nazca plate has been covered by Vigny et al. (2009). Dots are the locations of the GPS stations. Arrows indicate the horizontal velocities with respect to a reference frame, which is fixed to South American Plate. Numbers indicate the velocities in mm/year. Ellipses are the 99 % confidences.

Earthquake catalog and the waveform data are collected by National Seismological Center of Chile (CH). Distribution of the stations and earthquakes can be seen in Fig. 2.2.
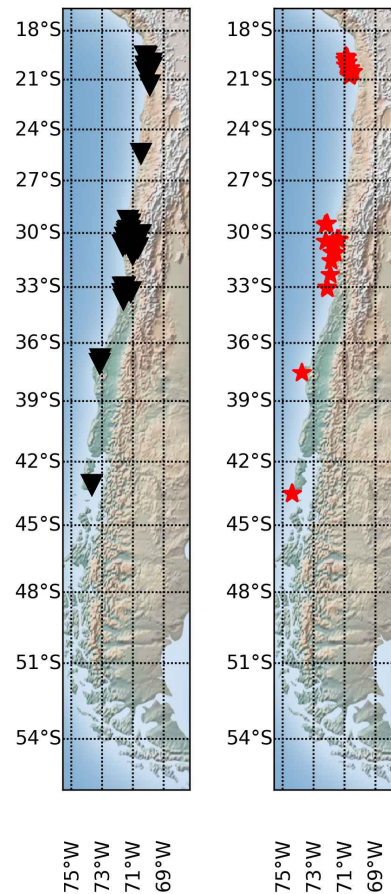
Figure 2.2: Distribution of the seismic stations (left) and epicenters of the earthquakes (right) in Chile. Maps are created with matplotlib package of Python (Hunter, 2007).

2.1.2. Greece and Turkey

In Greece and Turkey, there are 313 earthquake signals recorded due to 15 earthquakes by 30 stations. Major forces that created those are the west-southwest movement of Anatolian microplate, extension of Anatolian microplate on the west of Turkey and Hellenic Arc. There are 15 $M_w \geq 5.5$ earthquakes in which 9 strike slip, 5 normal and 1 reverse focal mechanisms occur. Earthquake catalogs are collected by Disaster and Emergency Management Presidency of Turkey (AFAD) (AFAD, 2019), Hellenic Seismic Network (HL) (National Observatory Of Athens, 1997) and Bogazici University Kandilli Observatory and Earthquake Research Institute Regional Earthquake-Tsunami Monitoring Center (KOERI) (Bogazici University Kandilli Observatory And Earthquake Research Institute, 2001). Wave-

forms are collected by AFAD, HL, Aristotle University of Thessaloniki Seismological Network (HT) (Aristotle University Of Thessaloniki Seismological Network, 1981), Seismological Network of Crete (HC) (Technological Educational Institute Of Crete, 2006), The GE Seismic Network (GE) (GEOFON Data Centre, 1993), ITSAK Strong Motion Network (HI) ((ITSAK) Institute Of Engineering Seimology Earthquake Engineering, 1981) and KOERI. Distribution of the stations and earthquakes can be seen in Fig. 2.3.
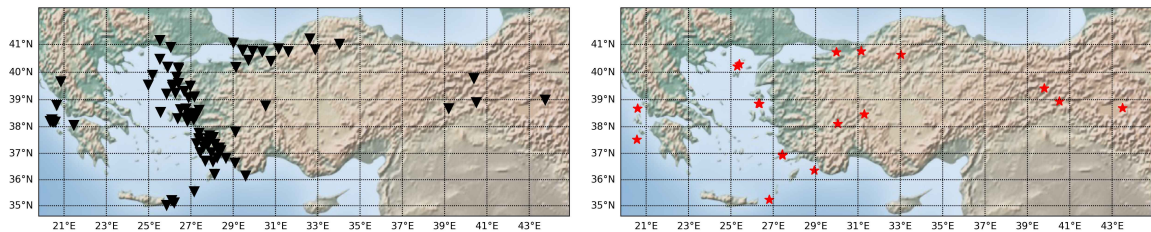
Figure 2.3: Distribution of the seismic stations (left) and epicenters of the earthquakes (right) in Greece & Turkey.

Main motions in Turkey that create earthquakes are the west to southwest movement of Anatolian microplate due to the movement of Arabian plate on the east of Anatolian microplate. Anatolian microplate escape from the stable Eurasian plate and the moving Arabian plate. This movement has created North Anatolian and South Anatolian strike slip faults. Westward motion of Anatolia shifts to the southwest motion on the west of Turkey. It creates normal faults, which are parallel to each other, from north to south of west Turkey. Subduction of African plate underneath the Eurasian plate on the south and southwest of Greece produce reverse faults around Crete Island and reverse and strike slip faults on the West of Peloponnese peninsula. Tectonic settings of Greece and Turkey can be found in Fig 2.4.
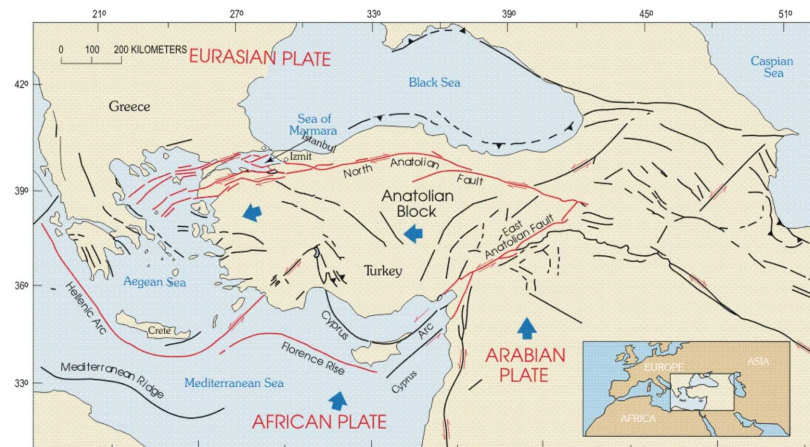
Figure 2.4: Tectonic settings of Greece & Turkey. Map is taken from the modified map of Barka (1992); Rockwell et al. (2001) originally from Holzer (2000).

The westward movement of the Anatolian block results from (1) differences in rates of motion between the Arabian and African plates, (2) different directions of motion between the Anatolian block and Eurasian plate to the north, and (3) subduction of the African plate beneath the Anatolian block at the Hellenic and Florence trenches. The Arabian plate is moving to the north faster than the African plate, both relative to a stable Eurasian plate.

2.1.3. Italy

In Italy, there are 1553 earthquake signals recorded due to 12 earthquakes by 340 stations. There are 1 strike slip, 6 normal and 5 reverse focal mechanism earthquakes in the database. Distribution of the earthquakes can be seen in Fig. 2.5. Waveforms can be found in The Italian Accelerometric Archive (ITACA) (Luzi et al., 2016; Pacor et al., 2011). Distribution of the stations can be seen in Fig. 2.5.
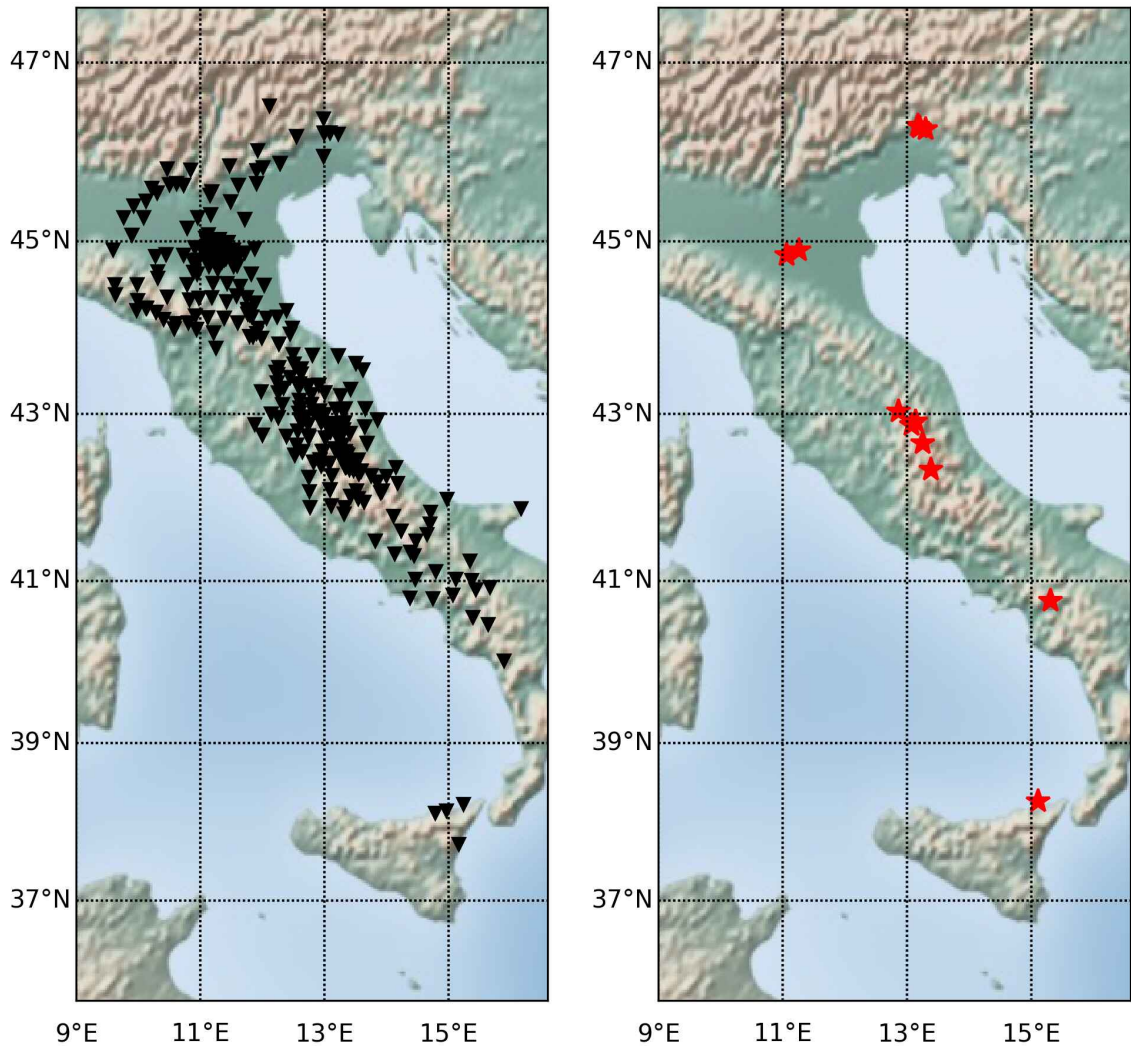
Figure 2.5: Distribution of the seismic stations (left) and epicenters of the earthquakes (right) in Italy.

Majority of the earthquakes that have been used in this study are located in the Apennines. Emilia earthquakes have occurred on North Apennines, Amatrice, L'Aquila and Umbria earthquakes have occurred in Central Apennines and Irpinia earthquake has occurred in South Apennines. 1978 Patti earthquake has occurred in Patti valley in the northeast of Sicily Island. 1976 Friuli earthquakes have occurred where Adriatic microplate, part of African plate, meets Eurasian plate, in the Southern Alps. Earthquake catalog is collected from Italian Strong Motion Network (RAN) (Presidency Of Counsil Of Ministers-Civil Protection Department , 1972). Tectonic settings of Italy can be seen in Fig. 2.6.
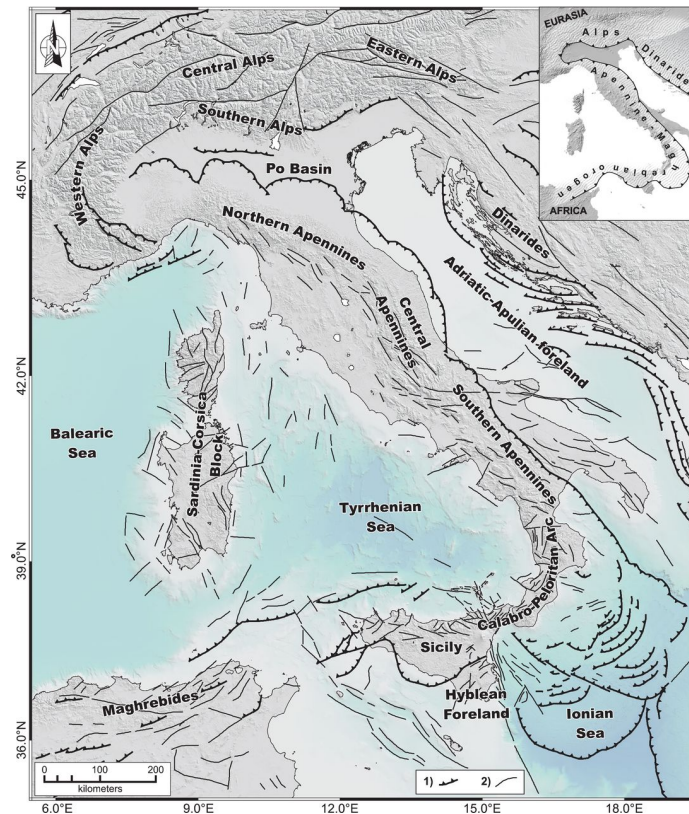
Figure 2.6: Simplified tectonic map of Italy and surrounding areas (main tectonic lineaments are redrawn from Morelli and Barrier (2004); Polonia et al. (2011)) taken from Palano (2014).

2.1.4. Japan

In Japan, there are 10 852 earthquake signals recorded due to 94 earthquakes by 1118 stations. Database contains 19 strike slip, 9 normal, 7 normal-oblique, 46 reverse and 13 reverse-oblique earthquakes from Japan. Earthquake catalog and waveform data are collected by Kyoshin Net (K-Net) and Kiban-Kyoshin Net (KiK-net) (for Earth Science and Resilience, 2019). Distribution of the stations and earthquakes can be seen in Fig. 2.7.
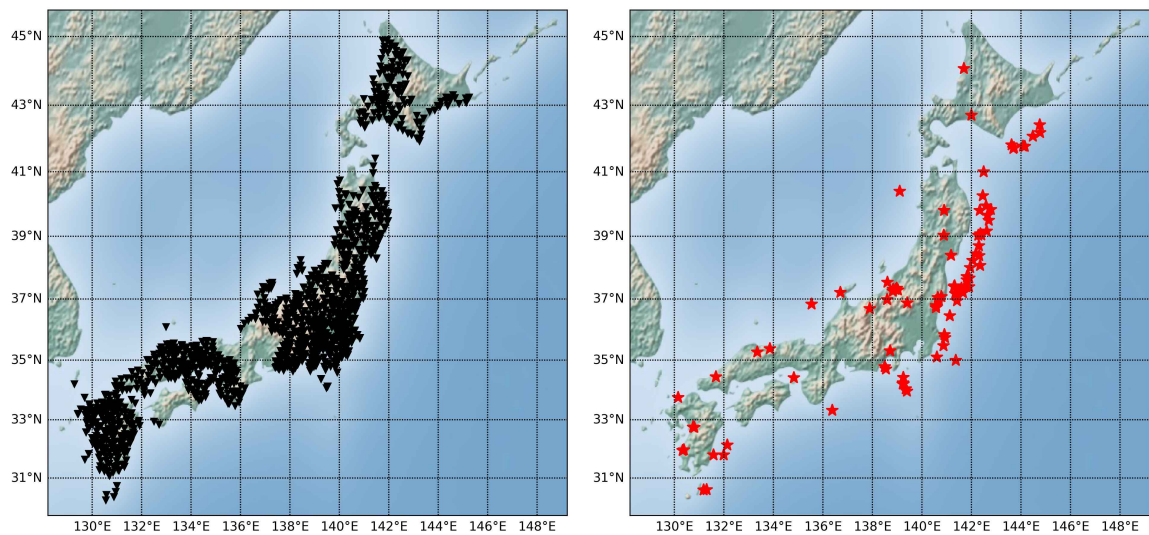
Figure 2.7: Distribution of the seismic stations (left) and epicenters of the earthquakes (right) in Japan.

Japan is one of the most seismically active regions on Earth. Offshore on the southeast of Japan, the subduction of the Philippine Sea plate under the Eurasian plate is observed, while offshore on the east of Japan, Pacific plate subduction under the North American Plate (Okhotsk microplate) in the upper half of the mainland and subduction under the Philippine Sea plate on the far southeast are in constant motion. Philippine Sea plate subduction is also moving towards the American Plate in the south of Tokyo. Even though subduction zones are relatively away from the Japanese mainland, the catalog is dominated by the reverse fault earthquakes that occurred in the shallow depths. Furthermore, back arc tectonics of these subduction zones produce various seismically active fault lines. For instance, Median Tectonic Line that is located in the southern half of the Japanese mainland produces various strike slip earthquakes, such as 1995 Kobe earthquake and 2016 Kumamoto earthquakes. Kumamoto earthquake sequence has been generated by the Futagawa and Hunagu fault zones (Ikeda et al., 2001), which are part of the Median Tectonic line. Tectonic settings of Japan can be seen in Fig. 2.8.
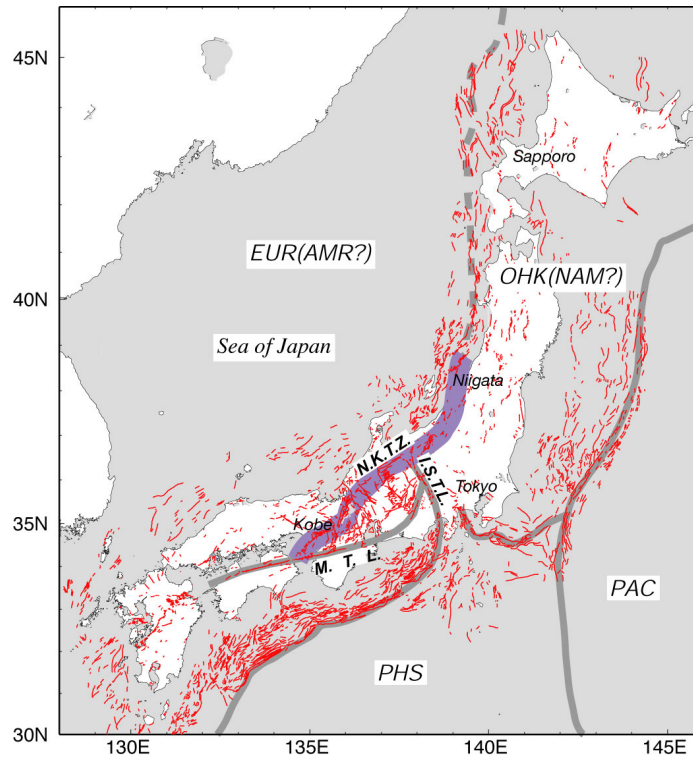
Figure 2.8: Tectonic map of Japan. Thin black lines denote active faults. Gray lines denote plate boundaries and major tectonic lines. Dashed gray line denotes a possible plate boundary between the Eurasian and the Okhotsk (or North American) plates. Thick gray line indicates location of the Niigata-Kobe Tectonic Zone (NKTZ). PAC: Pacific plate, PHS; Philippine Sea plate, EUR: Eurasian plate, AMR; Amurian plate, OHK: Okhotsk plate, NAM: North American plate, ISTL: Itoigawa-Shizuoka Tectonic Line, MTL: Median Tectonic Line. The map is modified from (Sagiya et al., 2000).

2.1.5. New Zealand

In New Zealand, there are 1443 earthquake signals recorded due to 32 earthquakes by 497 stations. These tectonic features created 8 strike slip, 6 normal, 9 reverse and 9 reverse-oblique earthquakes that are used in this study. Earthquake catalog and waveform data are collected by GeoNet project of New Zealand (GeoNet) (Van Houtte et al., 2017). Distribution of the stations and earthquakes can be seen in Fig. 2.9.
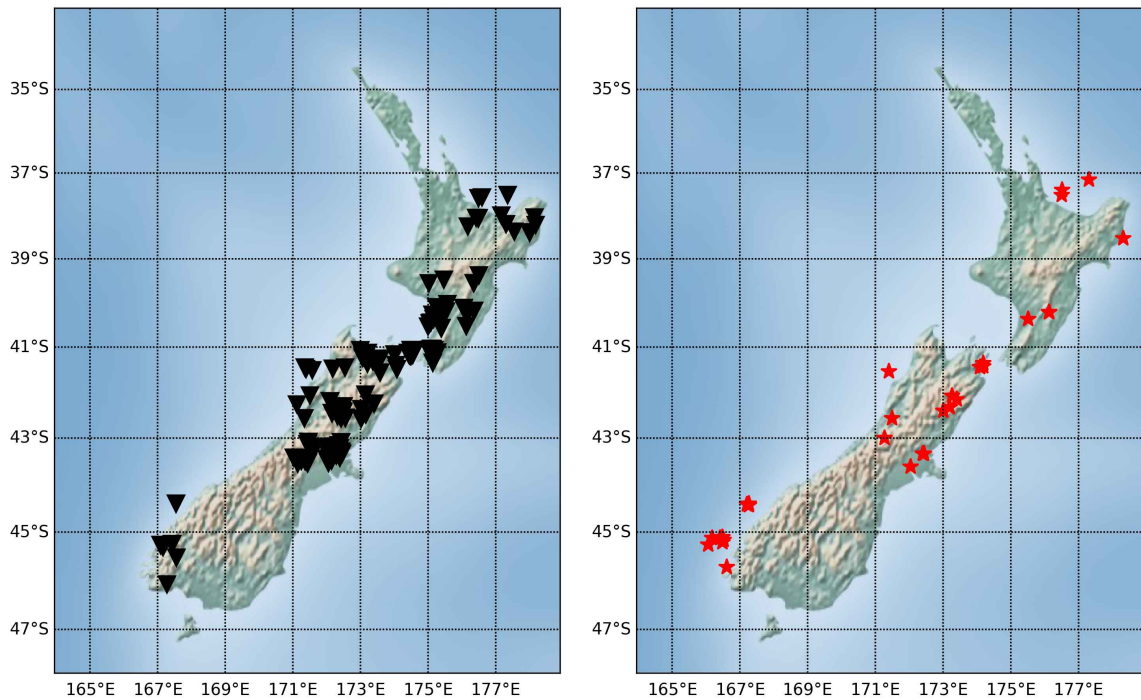
Figure 2.9: Distribution of the seismic stations (left) and epicenters of the earthquakes (right) in New Zealand.

New Zealand is located between the Australian plate and the Pacific plate. Northern part of the North Island has southward motion whereas the south of the island is moving west to southwest direction. South Island is moving towards southwest direction. Strike slip earthquakes that have been used in this study have been generated by the North Island fault system on North Island, Marlborough fault system, and undefined fault system before the Christchurch earthquakes in 2011 in Christchurch, South Island. Earthquakes with normal mechanisms are also located in the North Island fault system and Marlborough fault system. In addition, there are several normal fault earthquakes which are used from the west of Hikurangi subduction zone on the east of New Zealand. Reverse fault earthquakes are concentrated on the southwest of the South Island where Puysegur subduction zone is located. Reverse-oblique faults are in the Marlborough fault system, Alpine Fault, Hope Fault and Christchurch in South Island and North Island fault system in North Island. Tectonic settings of New Zealand can be seen in Fig. 2.10.
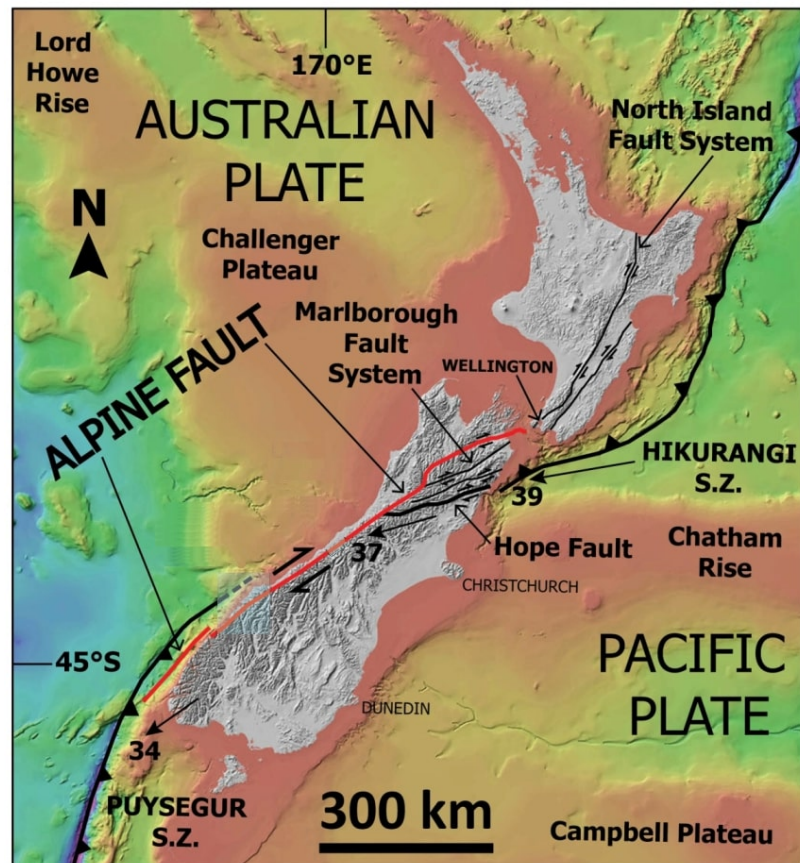
Figure 2.10: Tectonic map of New Zealand modified from Barth (2013). Northwest-illuminated hillshade derived from a digital elevation model (DEM) of Land Information New Zealand (LINZ) topographic data and National Institute of Water and Atmospheric Research (NIWA) 250 m bathymetry data. Australian-Pacific relative plate motion vectors (mm/year) from DeMets et al. (1994). Subduction zone is abbreviated as S.Z.

2.1.6. NGA-West2

In NGA-West2 ground motion database (NGA), there are 5583 earthquake signals recorded due to 100 earthquakes. Data can be found in Pacific Earthquake Engineering Research Center (PEER) ground motion database's NGA (Bozorgnia et al., 2014). Distribution of the stations can be seen in Fig. 2.11.
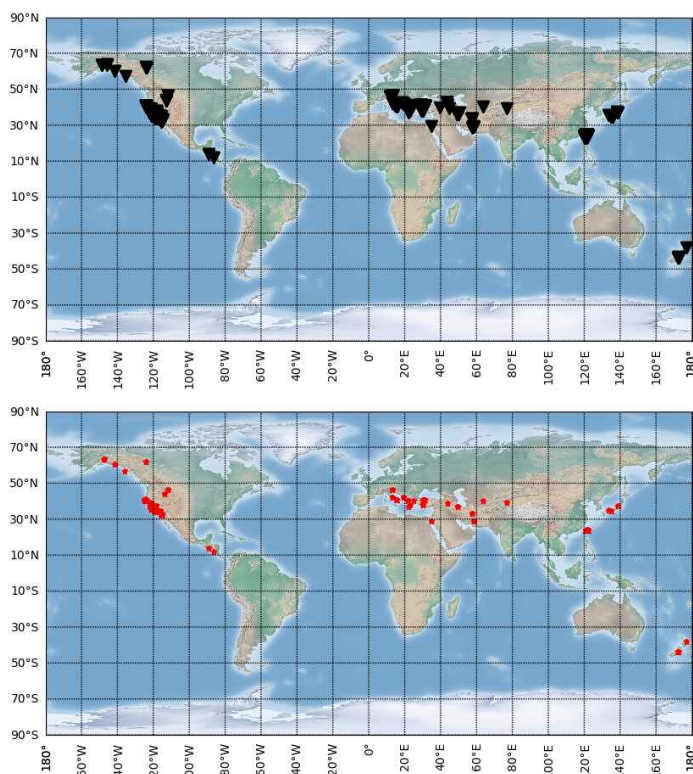
Figure 2.11: Distribution of the seismic stations (up) and epicenters of the earthquakes (down) in NGA database.

2.1.7. North America

In North America, there are 683 earthquake signals recorded due to 64 earthquakes by 196 stations in 4 countries, which are Canada, Costa Rica, Mexico and USA. In total 16 strike slip, 7 normal, 2 normal-oblique, 32 reverse and 5 reverse-oblique earthquakes are included to the database. Catalog is taken from Incorporated Research Institutions for Seismology (IRIS), Canadian National Seismograph Network (CNSN) (Geological Survey Of Canada, 1989) and National Autonomous University of Mexico (UNAM) (SSN, 2017). Data is collected by CNSN, UNAM, Alaska Regional Network (Alaska Earthquake Center, 1987), Alaska Volcano Observatory (Observatory/USGS, 1988), Global Seismograph Network (Albuquerque Seismological Laboratory (ASL)/USGS, 1988), Pacific Northwest Seismic Network (University Of Washington, 1963), Southern California Seismic Network (California Institute Of Technology And United States Geological Survey Pasadena, 1926), Berkeley

Digital Seismic Network (Northern California Earthquake Data Center, 2014), Plate Bound-
ary Observatory Borehole Seismic Network (Bendick, 2012), Red Sismologica Nacional de
Costa Rica (RSN, 2017), Nicoya Seismogenic Zone (Susan Y. Schwartz, 2009), Tomography
Under Costa Rica and Nicaragua (Geoffrey A. Abers, 2003) and IRIS/IDA Seismic Network
(Scripps Institution Of Oceanography, 1986), GEOSCOPE, USArray Transportable Array
and ChiriNet. Distribution of the stations and earthquakes can be seen in Fig. 2.12.
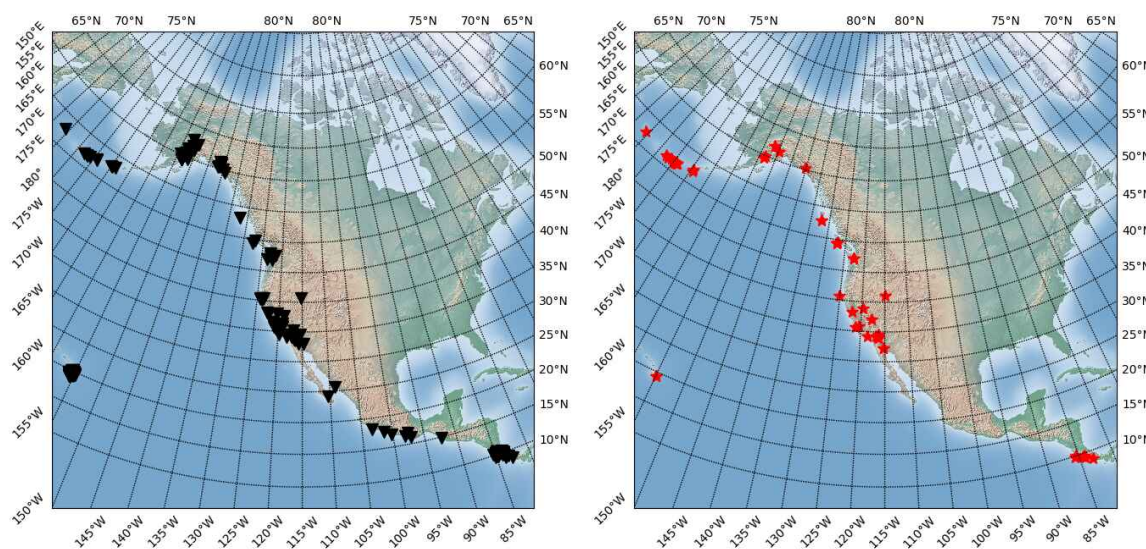
Figure 2.12: Distribution of the seismic stations (left) and epicenters of the earthquakes
(right) in North America.

Both in USA and Canada, earthquakes are located on the west coast of the North
American continent.

Aleutian Islands have a large number of reverse faulting earthquakes. They are the
result of the subduction of the Pacific plate under the North American plate. It is the main
tectonic force in the state of Alaska. Tectonic settings of Alaska can be seen in Fig 2.13.
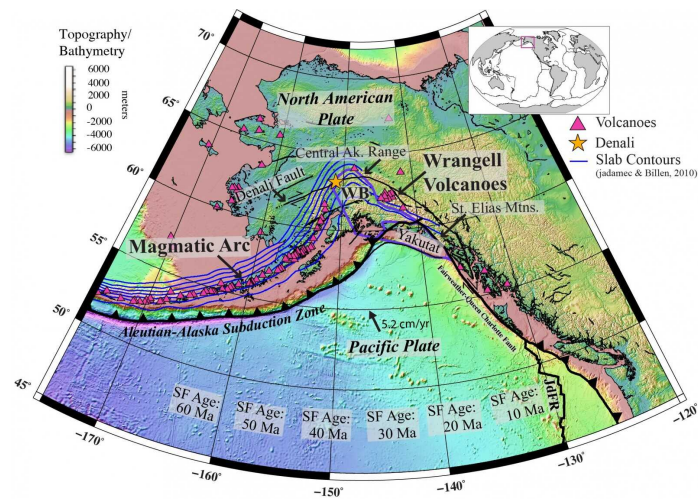
Figure 2.13: Tectonic map of Alaska modified from Haynie and Jadamec (2017). Topography/bathymetry is from Smith and Sandwell (1997) and Seafloor (SF) ages are from Müller et al. (2008). Blue lines are the slab contours of Jadamec and Billen (2010) in 40 km intervals; the thick black line is the plate boundary from Bird (2003); and the thinner black lines are faults from Plafker et al. (1994). The location of Denali is marked by the orange hexagon. Holocene volcanoes are given by the pink triangles (Alaska Volcano Observatory). The purple polygon is the outline of the Yakutat oceanic plateau (Haynie and Jadamec, 2017). WB – Wrangell block fore-arc sliver; JdFR – Juan de Fuca Ridge.

An earthquake in Hawaii has occurred as a result of reverse faulting on the south flank of Kilauea volcano, in the east rift zone. Tectonic settings of Hawaii can be seen in Fig 2.14.
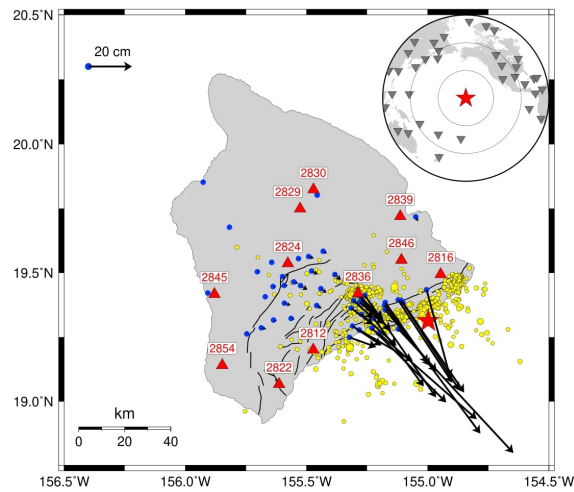
Figure 2.14: Map view of the near-field seismic and geodetic data coverage in Kilauea volcano. The red star denotes the epicenter of the Hawaii earthquake. The map is taken from Liu et al. (2018).

2 earthquakes in Canada are due to the subduction of Juan de Fuca Plate beneath the North American plate. Tectonic settings of Canada can be seen in Fig 2.15.
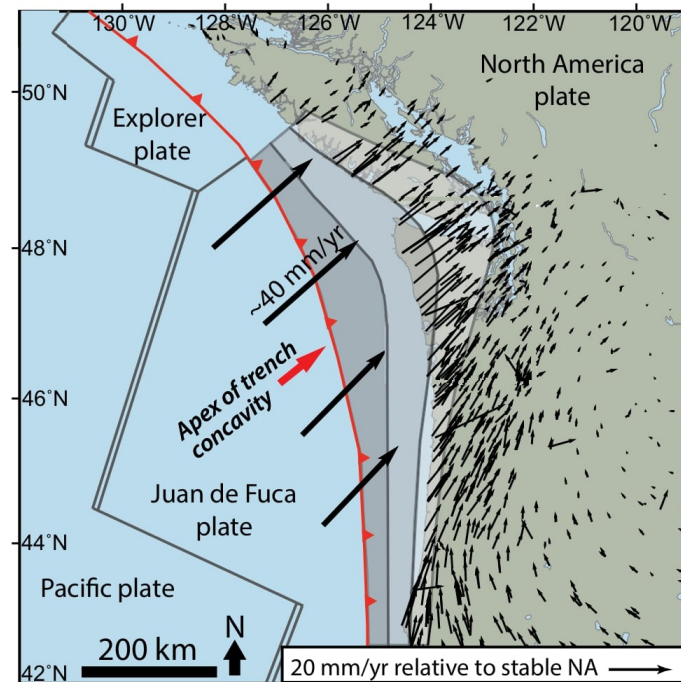
Figure 2.15: Tectonic setting of concave-outboard Cascadia subduction zone, showing Juan de Fuca–North American plate motion (thick black arrows, MORVEL model; DeMets et al. (2010)) and Global Navigation Satellite System velocity vectors (thin arrows; error ellipses (0.43 mm/yr mean standard error) omitted for clarity) relative to stable North America (NA) (UNAVCO (2019); McCaffrey et al. (2013)). Megathrust interseismic locking pattern is from Wang et al. (2003), where locked zone is dark gray and locking decreases downdip through effective transition zone (lighter gray). Maps is modified from Finley et al. (2019).

2008 Nevada earthquake has occurred in Independence Valley fault zone with a mechanism of normal faulting. Most of the strike slip faults in our database for USA is located on the San Andreas fault line. Transverse motion of North American plate and the Pacific plate created the San Andreas fault line. Furthermore, there are several reverse and reverse-oblique earthquakes on the east of San Andreas fault. Tectonic settings of San Andreas Fault and surroundings can be seen in Fig 2.16.
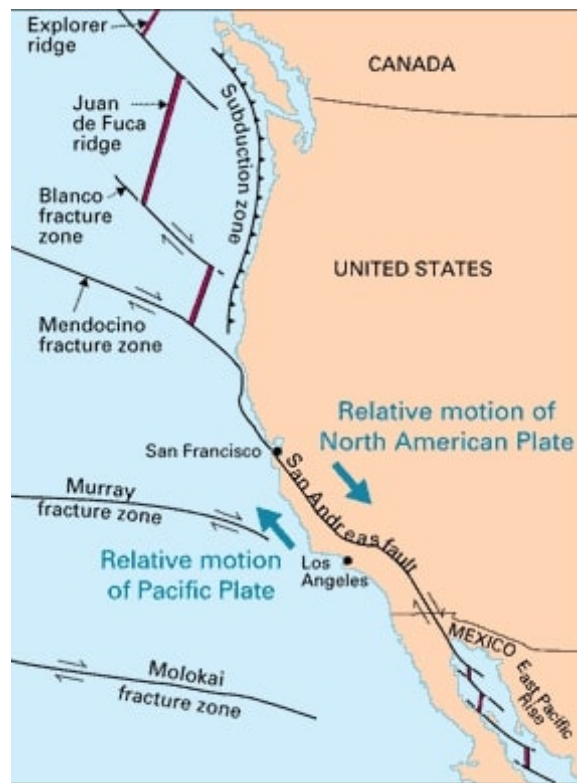
Figure 2.16: Tectonic settings of San Andreas Fault and surroundings. The map is taken from Kious and Tilling (1996).

Subduction of Cocos plate underneath the Caribbean and the Panama microplate is the source of 2 reverse and 1 reverse oblique earthquakes that occurred in Costa Rica. The only strike slip earthquake in Costa Rica is 20th of November 2004 earthquake. It is located on the Central Costa Rica Deformed Belt, which defines the western Panama microplate boundary. Tectonic settings of Costa Rica can be seen in Fig 2.17.
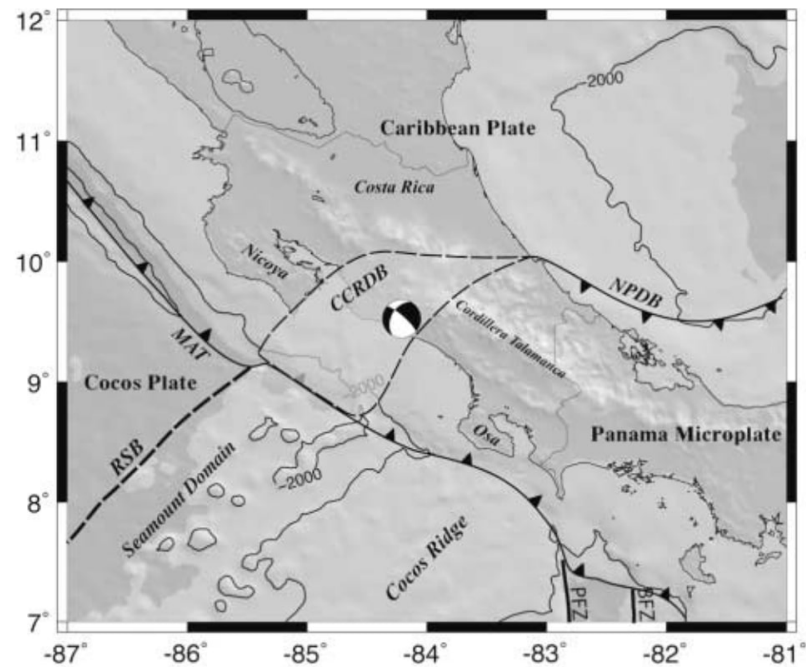
Figure 2.17: Tectonic map of Costa Rica. Middle American Trench (MAT), Central Costa Rica Deformed Belt (CCRDB), North Panama Deformed Belt (NPDB), Rough-smooth boundary (RSB), Panama Fracture Zone (PFZ), Balboa Fracture Zone (BFZ). The map is taken from Pacheco et al. (2006).

All strike slip earthquakes that are coming from UNAM catalog have occurred in Gulf of California. In the Gulf of California, there are strike slip faults with ridges in between. There is also an earthquake with a normal faulting which has occurred on the extension zone. Earthquakes with reverse faulting in Mexico are located in the subduction zone on the southwest of Mexico, where Cocos plate is in subduction under North American plate. Tectonic settings of Mexico can be seen in Fig 2.18.
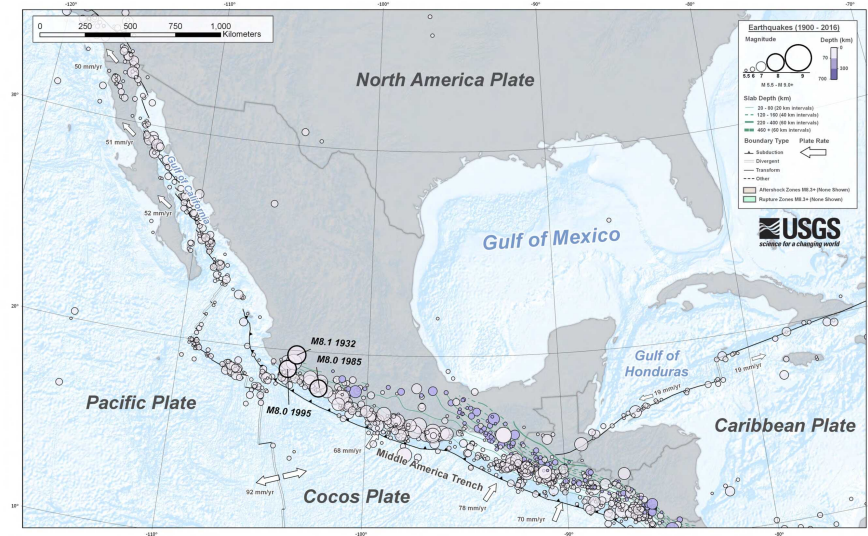
Figure 2.18: Seismicity map of Mexico between 1900 and 2016. Earthquakes with magnitudes bigger than 5.5 are plotted. Tectonic plate names, fault lines, subduction zones and relative plate motions are represented by their names and features, which can be found in the legend. Map is taken from United States Geological Survey (USGS).

2.1.8. Taiwan

In Taiwan, there are 543 earthquake signals recorded due to 46 earthquakes by 21 stations. 13 strike slip, 1 normal, 19 reverse and 13 reverse-oblique earthquakes are used. Earthquake catalog and waveforms are collected by Broadband Array in Taiwan for Seismology (BATS) Institute of Earth Sciences (1996). Distribution of the stations can be seen in Fig. 2.19.
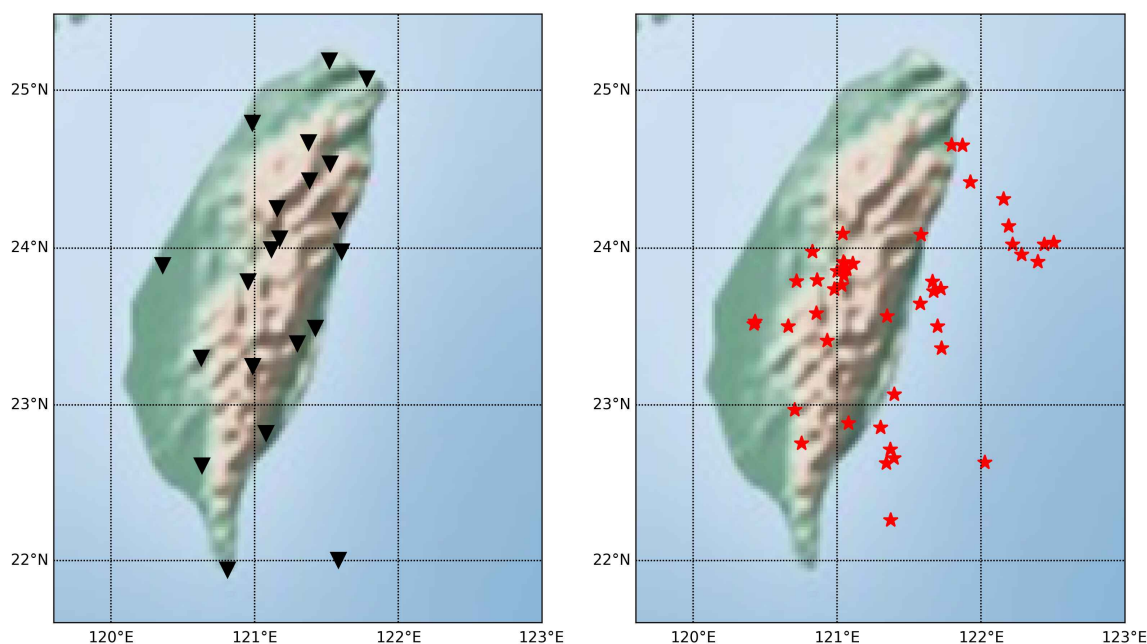
Figure 2.19: Distribution of the seismic stations (left) and epicenters of the earthquakes (right) in Taiwan.

Taiwan has complex tectonic features. The Philippine Sea plate is in subduction beneath the Ryukyu island arc in the east of Taiwan. In the south of Taiwan, the Philippine Sea plate is overriding the Eurasian plate along the Manila trench. Strike slip earthquakes that have been used are mostly located in the Central Mountains and the lateral movement of Ryuku Trench. The single normal fault earthquake from Taiwan is also located in the Central Mountains. Reverse and reverse-oblique fault earthquakes in the database are located in the coastal range and central range. They are located on the east and west of the Central Mountains, respectively. Tectonic settings of Taiwan can be seen in Fig. 2.20.
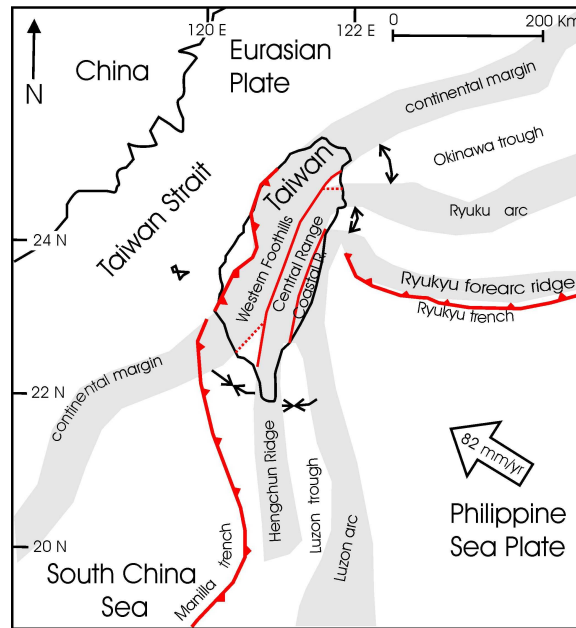
Figure 2.20: Tectonic settings of Taiwan modified from Shyu et al. (2005).

3. METHOD

3.1. Wavelet Method Analysis

Wavelet analysis is a common tool for analyzing the energy of the signal over time (Gurley and Kareem, 1999; Hancock et al., 2006; Shahi and Baker, 2014; Vassiliou and Makris, 2011; Zhou and Adeli, 2003). Time series signals can be decomposed into time-frequency space to identify dominant modes and their changes over time.

Let's assume an earthquake signal (x_n) with time interval δt and $n = 0, \ldots, N-1$. N is the number of samples in the earthquake signal. Define a wavelet function ($\Psi_0(\eta)$) which is a function of non dimensional time parameter, η. The wavelet function should have zero mean. For a Morlet wavelet:

$$\Psi_0(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\eta^2/2}, \tag{3.1}$$

where ω_0 is the non dimensional frequency. The wavelet transform of x_n can be written as the convolution of the time series with the scaled and translated version of the wavelet function.

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \Psi^* \left[\frac{(n'-n)\delta t}{s} \right] \tag{3.2}$$

where $*$ is the complex conjugate. One can vary the wavelet scale, s, and the time index to control the amplitude variation with time. In order to apply the wavelet transform faster, calculations are done in Fourier space. Discrete Fourier transform (DFT) of x_n is,

$$\hat{x}_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-2\pi i k n/N}, \tag{3.3}$$

where k is the frequency index. The Fourier transform of $\Psi(t/s)$ is $\overset{\wedge}{\Psi}(s\omega)$. One can calculate the wavelet transform by getting the inverse Fourier transform, thanks to convolution theorem. It can be seen in Eq. 3.4.

$$W_n(s) = \sum_{k=0}^{N-1} \hat{x}_k \overset{\wedge}{\Psi}{}^{*}(s\omega_k) e^{i\omega_k n \delta t} \tag{3.4}$$

ω_k defines the angular frequency. The wavelet function is normalized with s to be comparable to other earthquake signals. Since the Ψ_η is complex, the wavelet transform, $W_n(s)$, is also complex. Transformation can be divided into real and imaginary or amplitude and phase parts. Since the Ricker wavelet is real-valued, the imaginary part is zero and phase is not defined.

Wavelet analysis package of Torrence and Compo (1998) is used for signal analysis. Two different wavelet types, which are Ricker (Appendix B) and Morlet wavelets (Appendix C), are implemented to the wavelet analysis process. Wavelet power spectra of the signals are calculated by using these wavelets. Morlet wavelet is complex, while Ricker wavelet is real-valued. The complex wavelet function returns both amplitude and phase information, whereas real wavelet function returns only real components. This allows to isolate discontinuities. Since both of the wavelets are giving the same qualitative results on power spectra, both of them can be treated as equal. Ricker wavelet can distinguish the discontinuities (Fig. 3.1) since it is a real-valued function, whereas Morlet wavelet can give more smooth results, which is important when there is high-frequency content in the pulse region (Fig. 3.1).

The resolution of the wavelet function depends on the width of real space and width in Fourier space. A broad function will give a poor time resolution but a good frequency resolution, and vice versa. The width of the wavelet is proportional to the sampling rate of the signal.

As a result of the analysis, one can determine the power spectrum values over time. The maximum power spectrum values at PGV and the biggest power spectrum value of the signal (if it does not occur at PGV) are used in the pulse identification.
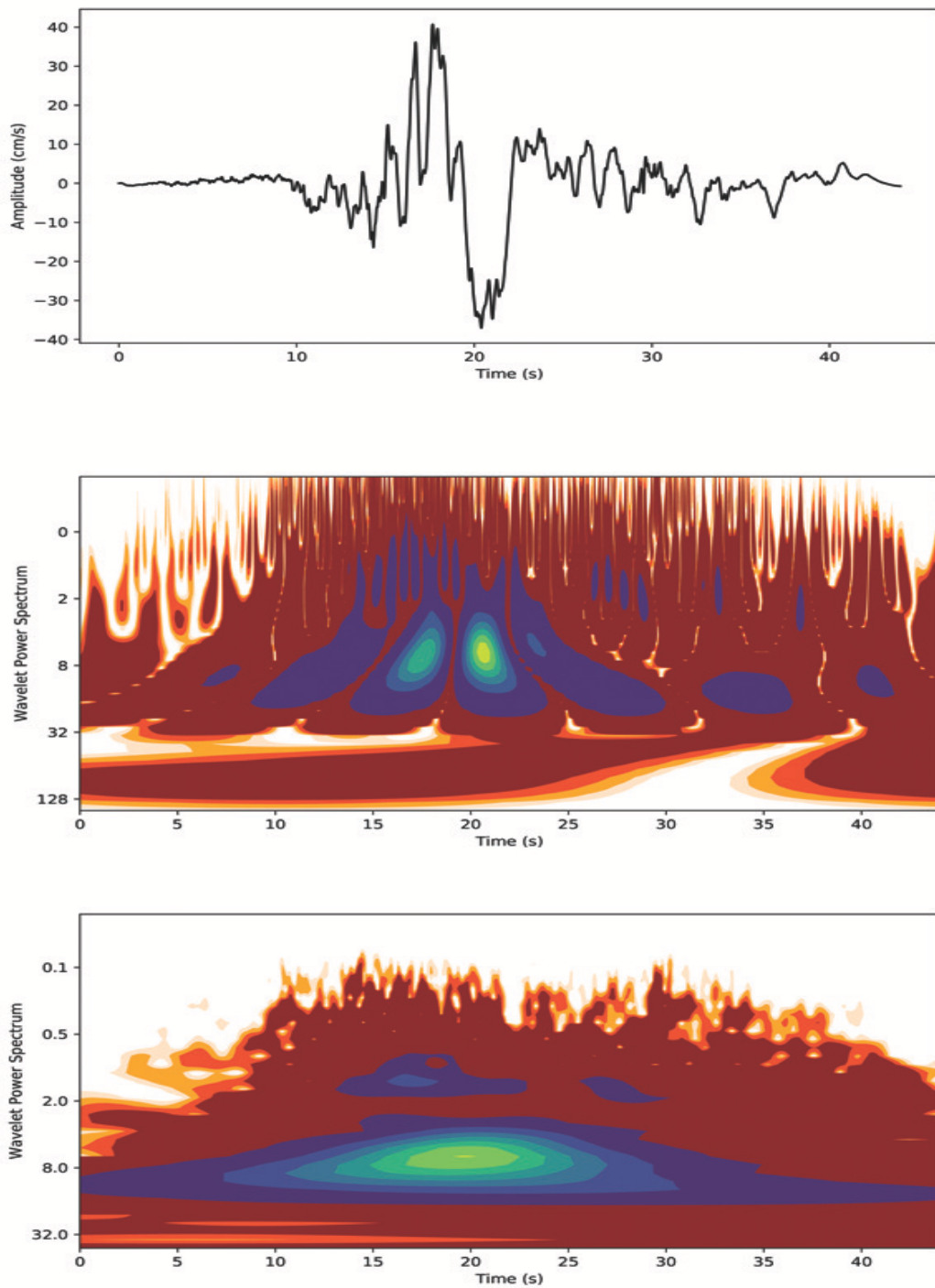
Figure 3.1: Velocity waveform (upper), Ricker wavelet power spectrum (center) and Morlet wavelet spectrum (lower) of 1992 Landers earthquake ($M_w = 7.3$), Yermo Fire Station ($R_{ep} = 85.99$ km). Red and blue colors represent low and high concentration of power, respectively. Waveforms are plotted with the Python package of Obspy (Krischer et al., 2015).

3.1.1. Pulse Identification

Unlike previous studies, wavelet method can identify velocity pulses that occur away from the time interval where the PGV is located. Several decision mechanisms are used to identify pulse shaped signals. The criteria for pulse shaped signals differ with respect to the position of the pulse, as explained in Section 3.1.1.1 and in Section 3.1.1.2.

3.1.1.1. Velocity Pulse at PGV.   Most of the seismic energy is assumed to be concentrated at the position where PGV occurs. One of the logical ways to analyze the signal is to focus around the region of PGV. Wavelet method is similar to Chang et al. (2016) since it analyzes the energy ratio of the waveform. Furthermore, the method also analyzes for spectral energy, which is similar to the method of Mena and Mai (2011). The threshold of the method occurs when the average of these two parameters around the PGV are equal or bigger than 30 % of the whole waveform. The criteria are reported below:

(i) $\text{PGV} \geq 30\,\text{cm}\,\text{s}^{-1}$.

(ii)

$$\frac{\left(\frac{\int_{t_s}^{te} v^2(\tau)d\tau}{\int_0^\infty v^2(\tau)d\tau} + \frac{\int_{t_s}^{te} \text{WPS}(\tau)d\tau}{\int_0^\infty \text{WPS}(\tau)d\tau} \right)}{2} \geq 0.30 \tag{3.5}$$

In Eq. 3.5, t_s and t_e represent the starting and ending points of the pulse, respectively. These points are found by identifying the T_p where the maximum wavelet power spectrum occurs at PGV. The pulse area is then identified as $t_{PGV} \mp T_p/2$ where t_{PGV} represents the time of the PGV. WPS indicates the wavelet power spectrum. The parameters can be seen in Fig. 3.2.

Left side of the numerator in Eq. 3.5 indicates the energy ratio between the impulsive part (velocity time history between t_s and t_e in time axis). Right side of the numerator is the ratio between wavelet spectrum energy of the waveform and impulsive part between the aforementioned area of the signal. Integrals are for summation process and infinity signs indicate the whole waveform.
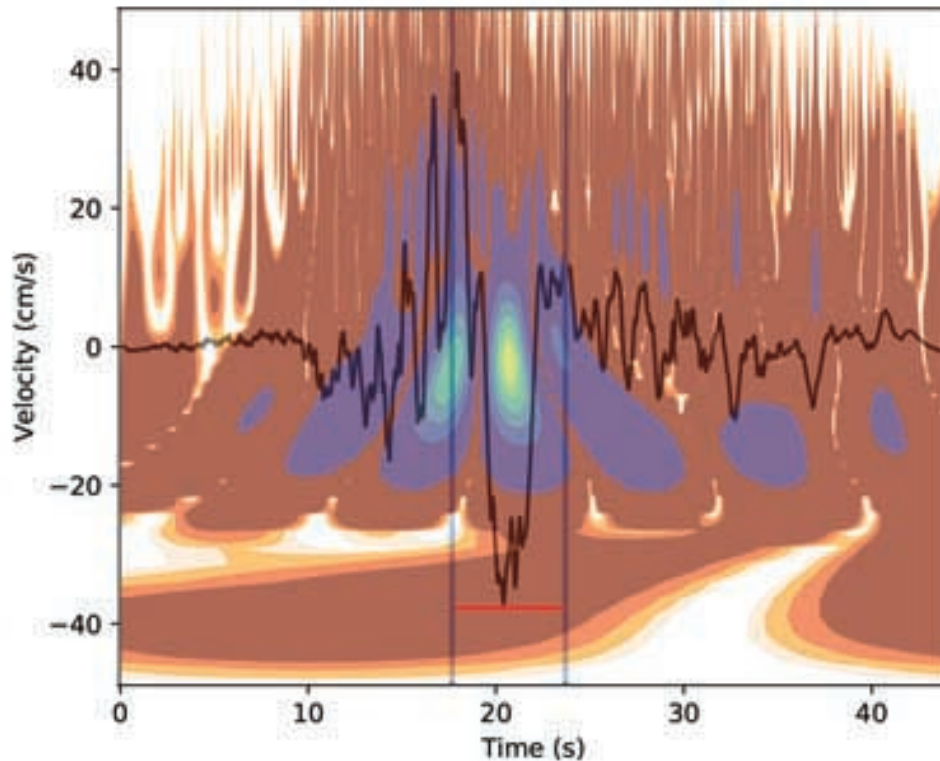
Figure 3.2: 1992 Landers earthquake, Yermo Fire Station velocity waveform. Red line and blue lines represent width (T_p) and borders (t_s and t_e) of the pulse region, respectively. Background image is Ricker wavelet power spectrum of the signal with the same color content of Fig. 3.1.

3.1.1.2. Velocity Pulse Outside the PGV Region.   Unlike previous studies, wavelet method also checks for the biggest energy arrival rather than the position of PGV. The logic behind the energy calculation is the same as in the Section 3.1.1.1. The minimum amplitude is fixed at $25\,\mathrm{cm\,s^{-1}}$. However, this amplitude is not the amplitude of PGV, but the biggest amplitude of the region where the maximum energy is concentrated. Furthermore, the maximum energy of the region should be equal or bigger than 10 % of the energy of the PGV region. The average of the energy of the waveform and wavelet power spectrum of this region should exceed 30 % of the total energy of the signal. The criteria are reported below:

(i) The biggest amplitude, in absolute sense, in the area where the maximum power spec-

trum value occurs, should be equal or bigger than $25\,\mathrm{cm\,s}^{-1}$.

(ii) Difference between PGV and the time where the maximum power spectrum in time axis should be larger than $T_p/4$.

(iii)

$$\frac{\int_{t_{semax}}^{t_{eemax}} v^2(\tau)d\tau}{\int_{t_s}^{t_e} v^2(\tau)d\tau} \geq 1.1 \tag{3.6}$$

(iv)

$$\frac{\int_{t_{semax}}^{t_{eemax}} \mathrm{WPS}(\tau)d\tau}{\int_{t_s}^{t_e} \mathrm{WPS}(\tau)d\tau} \geq 1.1 \tag{3.7}$$

(v)

$$\frac{\left(\dfrac{\int_{t_{semax}}^{t_{eemax}} v^2(\tau)d\tau}{\int_0^\infty v^2(\tau)d\tau} + \dfrac{\int_{t_{semax}}^{t_{eemax}} \mathrm{WPS}(\tau)d\tau}{\int_0^\infty \mathrm{WPS}(\tau)d\tau}\right)}{2} \geq 0.30 \tag{3.8}$$

In Eq. 3.6 and Eq. 3.7 t_{eemax} and t_{semax} represent the starting and ending points of the pulse in the maximum energy area in time axis. These points are found by identifying, in the area where the maximum power spectrum values are located, the maximum pulse period ($T_{p,emax}$) of the signal. The pulse area is then identified as $t_{eemax} \mp T_{p,emax}$ where t_{eemax} represents the time of the biggest value in the $T_{p,emax}$ region. The parameters can be seen in Fig. 3.3.

Eq. 3.6 and Eq. 3.7 describe the threshold for the energy ratios between the area around the PGV and the area around the maximum energy, if exists, for waveform and wavelet power spectrum, respectively. Other parameters have the same meanings that are explained in Section 3.1.1.1. Eq. 3.8 has the same meaning as Eq. 3.5.

The parameters can be seen in Fig. 3.3.
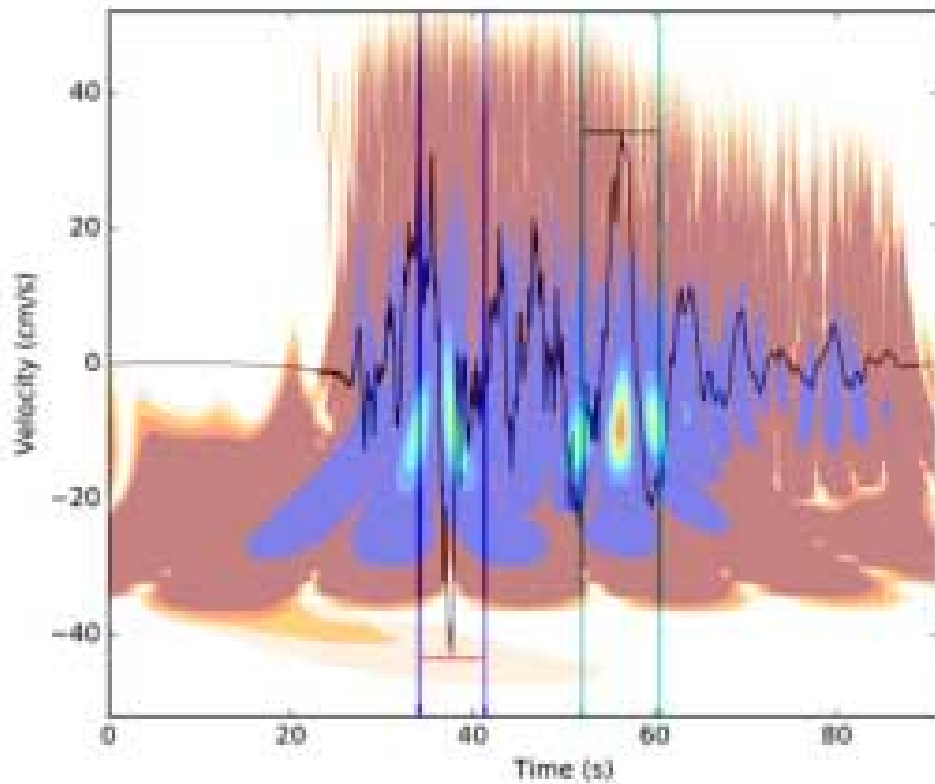
Figure 3.3: 1999 Chi-Chi Taiwan Earthquake ($M_w = 7.6$), TCU051 Station ($R_{ep} = 38.53\,\text{km}$) velocity waveform. Red line and blue lines represent the width (T_p) and borders (t_s and t_e) of the pulse region around PGV, respectively. The green line and cyan lines represent the width ($T_{p,emax}$) and borders ($t_{e\,emax}$ and $t_{s\,emax}$) of the area where the maximum energy is concentrated, respectively. Background image is Ricker wavelet power spectrum of the signal with the same color content of Fig. 3.1.

Both Ricker and Morlet wavelets are fitted to the pulse region when the algorithm detects a pulse shaped signal.

3.2. Convolutional Neural Networks

CNN is a deep learning algorithm which uses the features that it extracts from the given data to differentiate the labels of the inputs. CNN needs a vast amount of examples in order to understand the features of each label. Inputs can be vectors, matrices or tensors.

If images have colors, each element of the matrix contains 3 information about the colors of red, green and blue, which makes this matrix a tensor. If the given image is gray scaled, then the element contains only the gray scale value. If time series is given, then it can be represented as a vector.

In this study, the input is a vector of velocity waveform, for which the the waveforms are picked manually. It is assumed that the human interpretation is the ground truth and the methods that have been developed are not perfect to identify the impulsive signals. In the end of labeling process, 556 of the signals are considered as impulsive and 21 458 of the signals are labeled as non-impulsive. Signals are trimmed from P wave arrival, which is picked manually, to 40 s after the P arrival. P wave arrivals are picked by using the Seismic Analysis Code (SAC) of Goldstein and Snoke (2005). Furthermore, waveforms are down sampled to 20 Hz to reduce the computation time. The data is filtered between 0.05 Hz and 5 Hz in the pre-processing part of the CNN method. Since there is a major difference between the signals with impulsive and non-impulsive labels, synthetic signals are created in order to have 1:1 ratio between labels. The methods of calculation of synthetics are explained in Section 3.2.3.1.

In order to understand the mechanics of the CNN, the terms that are used in CNNs are explained between Section 3.2.1 and Section 3.2.2.6.

3.2.1. What is a Neuron?

In biology, neuron is a nerve cell that is responsible for the transfer of information between brain cells. Neurons let the electric current flow when activated. It allows to let the information pass to the following neuron cell. In a CNN, an artificial set of neurons are linked to each other with weighted connections. Artificial neurons have activation functions, mathematical functions, that lets or prevents the flow of information to the next layer of neurons. In this study, rectified linear unit (ReLU) is used as an activation function (Appendix D). Sigmoid activation function is used in the final layer as an activation function (Appendix E). In the first layer of the CNN, data points of the velocity waveform are the inputs. Neurons give the output depending on the input value(s). After the first layer, input parameters are the

output value(s) of the previous layer(s) inputs. An example neuron can be seen in Fig. 3.4.
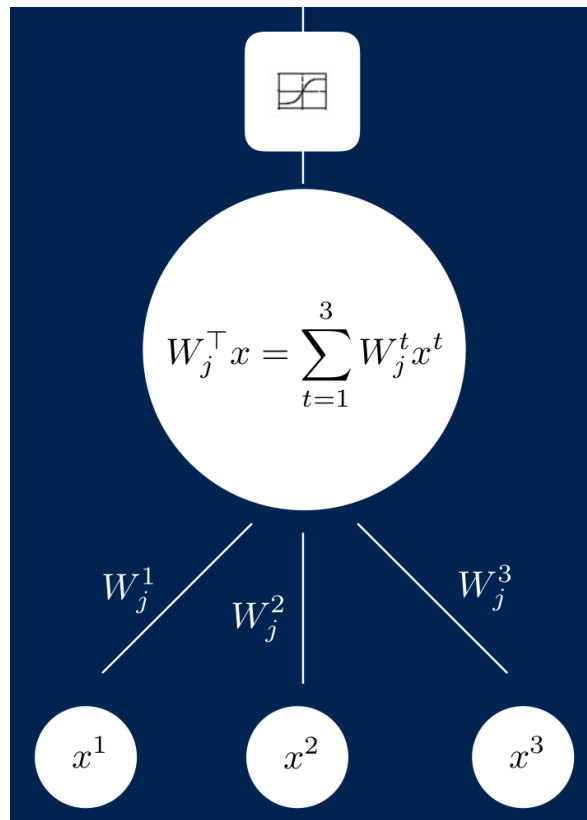
Figure 3.4: Visual representation of an artificial neuron. The result of the inner product of weights with input parameters may or may not be transferred to the following neuron depending on the result of the activation function, which is the Sigmoid in the figure.

In Fig. 3.4, x^1, x^2, x^3 are the input parameters (eg. east, north and vertical components of an earthquake record). A neuron will compute an inner product based on the weight matrix W_j. Weights are vital for the learning process. They are the representation of the strength of the connection between neurons. If the weight between two neurons has bigger magnitude, it means that the neuron that passes the information to the following layer has greater influence on the receiving neuron. If a weight has a value close to 0, it means that the information one neuron carries to the following neuron on the next layer has no effect on the receiving neuron. It can be formulated as in Fig. 3.4. In the end of this linear process, a non-linear operation, activation function, is applied.

3.2.2. Architecture of Convolutional Neural Networks

Layers are the collection of artificial neurons. Each layer give output to the next layer of neurons and take input from the previous ones. An example of a neural network can be seen in Fig. 3.5.
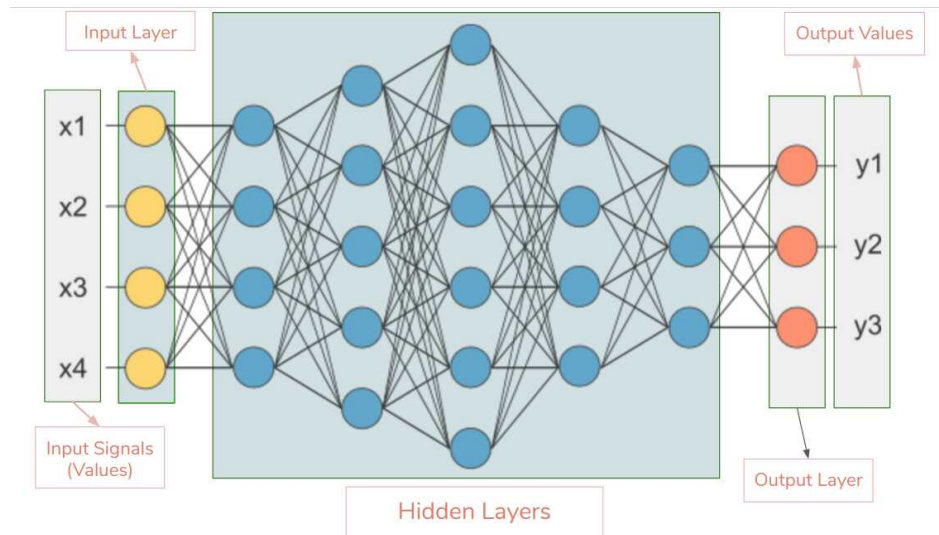
Figure 3.5: Visual representation of a Neural Network which is a basic case of CNN.

In Fig. 3.5, the layer on the far left is the input layer. In that case, this layer has the amplitudes of each data point of each velocity waveform. Layers in the center are called hidden layers. It is named as such since its values are not observable in the training stage. In hidden layers, all neurons of the previous layer are inputs for each neuron of the current layer. This structure is also called dense layers. The layer on the right edge is called output layer. It gives the probability of each label. Since binary classification is used for the pulse identification, the output is just 1 and the label is chosen according to the output value. In Fig. 3.5, arrows represent the mathematical operation for each unit.

In Fig. 3.5, there are 22 neurons, which are located in the hidden layers 1-5 and in the output layer. It makes the model flexible. The number of weights depends on the number of neurons on each layer. In the Fig. 3.5, there are $4 \times 4 + 4 \times 5 + 5 \times 6 + 6 \times 4 + 4 \times 3 + 3 \times 3 = 111$ weights.

Parameters that affect the performance of CNNs are explained between Section 3.2.2.1 and Section 3.2.2.6.

3.2.2.1. Biases.   Each neuron in these layers contains biases in order to delay the triggering of the activation function. In the Fig. 3.5, there are 22 biases. This CNN contains $111 + 22 = 133$ parameters. The weights of the neural network are initialized using the Glorot normal initializer (Glorot and Bengio, 2010). It initializes each weight with a small Gaussian value with a zero mean.

3.2.2.2. Filtering.   Filtering is the convolution between the input signal and the filtering (also known as kernel vector) signal. It can be formulated as in Eq. 3.9,

$$(f * g)(i) = \sum_{j=1}^{m} g(j) \cdot f(i - j + m/2). \tag{3.9}$$

f and g are the original velocity waveform vector and filtering vector.

3.2.2.3. Max pooling.   Max pooling is applied in the end of each hidden layer. Max pooling is a sample-based discretization process. The purpose of this operation is to reduce its dimensions and allow assumptions to be made about the features of the earthquake signal. Maximum value of the portion of the input that is covered by the kernel vector is the return of each neuron in max pooling layer.

3.2.2.4. Loss Function.   Loss function is used to measure the difference between the predicted value and the actual label. In this work, the loss function is defined as the binary cross-entropy. Since the signals are labeled as impulsive and non-impulsive, binary cross-entropy can be used as a loss function. Cross entropy measures the divergence between two probability distributions. Large cross-entropy means that the difference between the distri-

butions are large, whereas small cross-entropy means that two distributions are similar to each other.

CNN tries to produce smaller loss values on each iteration in order to have higher accuracy. Weights are evolved during the learning phase to reduce the loss value. This process is also called "learning".

3.2.2.5. Learning Rate & Optimizer. Choosing a proper learning rate is an important task on machine learning studies. Small learning rate can increase the calculation time, whereas large learning rate can cause the divergence from the minimum loss. For each set of weights, an error space is calculated. Loss function values are used in the creation of these spaces. Basic visualization of this process can be seen in Fig. 3.6. Each set of weights are represented with a contour line. New set of weights are calculated by the help of the gradient that has been calculated in each epoch. By evaluating the gradient at a certain position, one can find the direction of the steepest descent, and move along on that direction to set new weights. In practice, learning rate determines how far we will go in perpendicular direction from the contour that we are sitting. However, using a constant learning rate may create problems, such as diverging away from the minimum.
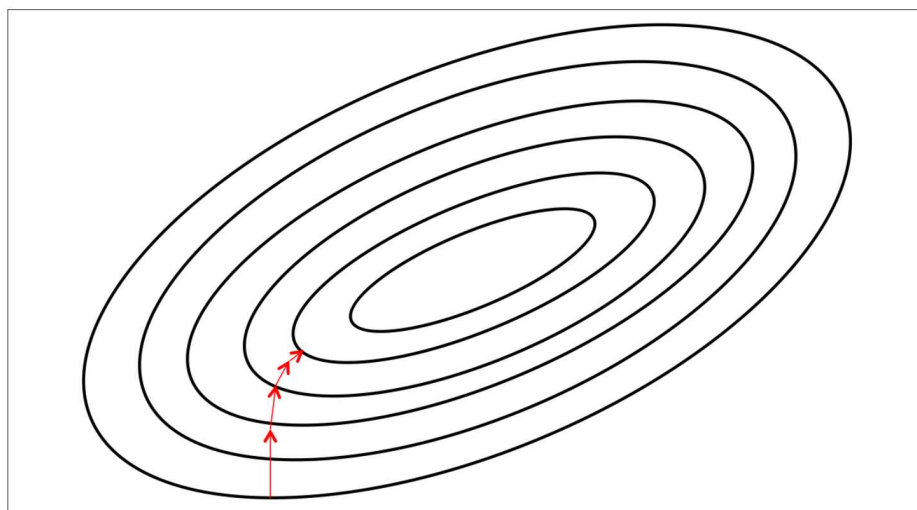
Figure 3.6: Visual demonstration of the error surface with two weights. The mapping of 3D space of error on 2D surface, the results are for a set of two weights. Red arrows indicate the gradient of each calculation point.

Gradient descent can fail when the contours are elliptical, which is the usual case for errors in deep networks. The gradient can fail as 90° away from the true direction. This situation can be seen in Fig. 3.7. Adaptive methods are used in order to overcome this problem. Adam optimization (Kingma and Ba, 2014) is used as a solution.
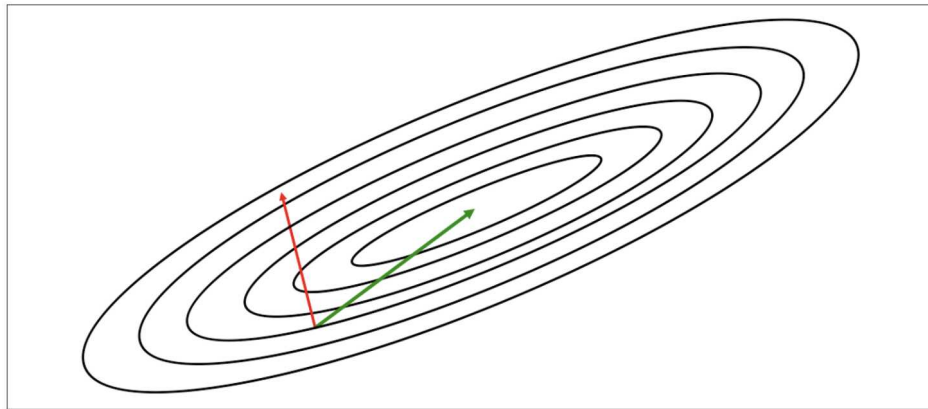
Figure 3.7: Direction with the steepest descent determined by the gradient descent. True direction is represented by the green arrow.

Individual adaptive learning rates are found by using the first and second moments of the gradients (Buduma and Locascio, 2017). Adam optimization algorithms are used for the determination of learning rate. It can roughly be explained as below:

If finding the minimum in the error space is demonstrated as the movements of the ball on a curvy plane, then the acceleration of the ball will be the gradient descent. It may be affected by almost all effects of the surface. Velocity of the ball will be the smoothed time history of the ball, which will the be its momentum. Movement of the ball to the minimum error location can be seen more easily on its velocity time history. Adam optimizer uses first and second order moments to actively change the learning rate over epochs to reach the minimum error location without getting affected.

3.2.2.6. Dropout. Dropout is the stage where ignoring neurons aimlessly during the training process. These neurons are ignored during both forward and backward propagation. This process is used in order to prevent over fitting. Over fitting is a problem in which the model arranges the weights highly relayed to the signals that it has seen during the learning process.

It will create problems when the new, unseen, data is shown to the model. The model tends to have lower prediction rate, if it has over fitting problem. In training phase randomly picked neurons will return zero output.

3.2.3. Data

There are various methods to calculate impulsive signals, some of which are explained in Section 1.2. Wavelet method and other studies use different constraints and mathematical assumptions in the determination process. Because of that, different approaches identify different signals as impulsive. Number of impulsive signals that are determined by previous studies and wavelet method can be seen in Section 4.1. In order to test the CNN method, impulsive signals are identified manually. 556 out of 21 458 earthquake signals are labelled as impulsive. However, impulsive signals are only covering 2.6 % of the dataset. The ratio between impulsive and non-impulsive signals are not good enough for CNN to understand the features of the impulsive signals. In order to overcome this problem synthetic signals are created. If the synthetic signal is labelled as impulsive by Chang et al. (2016); Shahi and Baker (2014) or wavelet method, it is considered as impulsive.

3.2.3.1. Synthetic Creation. Several methods have been developed for the creation of synthetic seismograms. Some of these methods do not take into account the fault plane, but use point sources (Boore, 1983; Wennerberg, 1990), while others use kinematic models and finite faults (Irikura, 1992; Papageorgiou and Aki, 1983). The problem with these models is the assumption of small source with respect to the source to station distance. As a result, they fail to model the near source effects of large earthquakes. In order to get the near source effects on stations, k^{-2} kinematic model is adopted for the seismic source modeling.

Fault planes are divided into numerous sub faults and numerical Green's tractions are calculated for each sub fault (Bianchi et al., 2010; Del Gaudio et al., 2015). Synthetics contain frequency information between 0.05 Hz and 2.2 Hz for 181 different frequency values. AXITRA software is used in order to compute the Green's functions (Coutant, 1989).

Representation Theorem. Displacement measured at the Earth's surface due to the slip is explained by the representation theorem (Aki and Richards, 2002). Let us assume that we want to express the displacement, **u** due to body forces, **f**, on a volume, *V*, with boundary conditions at *S*. During an earthquake (discontinuity), the displacement on the Σ^- side of Σ may move away from Σ^+ side. Displacements across the surface, Σ, are discontinuous. Displacement discontinuity between Σ^+ and Σ^- can be written as $[\mathbf{u}(\xi, \tau)]$. ξ and τ are the applied unit impulse and time, respectively. Graphical explanation of the volume, *V* during a rupture can be seen in Fig 3.8. Displacement on Earth can be computed as the convolution of the slip function with the elastic features of the medium. It can be formulated as in Eq. 3.10.
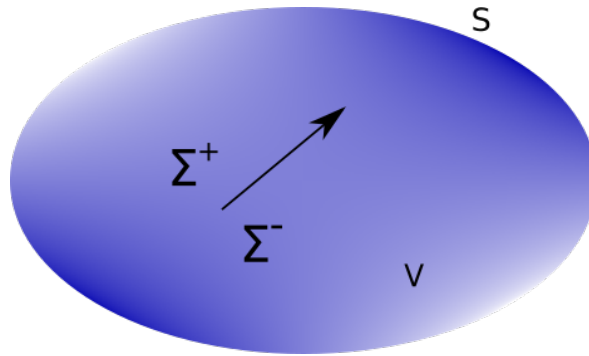
Figure 3.8: A finite body with the surface of S and a body of V. Displacement surface is represented by Σ.

$$u_m(\mathbf{x},t) = \int_{-\infty}^{+\infty} d\tau \int \int_{\Sigma} [u_i(\xi, \tau)] c_{ijpq} n_j \frac{\delta}{\delta \xi_q} G_{np}(\mathbf{x}, t-\tau; \xi, 0) d\Sigma \qquad (3.10)$$

In Eq. 3.10, **c** is the elasticity tensor, **G** is the Green's function. Elasticity tensor can be defined by λ and μ, Lame constants, in isotropic medium. c_{ijpq} can be rewritten as:

$$c_{ijpq} = \lambda \delta_{ij} \delta_{kl} + \mu(\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}). \qquad (3.11)$$

Green's Function. Green's function is a tensor that explains the impulse response of the medium. Green's function in Eq. 3.10 can be explained as the ith component of displacement with the force applied in j^{th} direction at ξ at the position $\mathbf{x}$ and time t - τ. Green's function tensor and elasticity tensor can be combined to calculate the stress on the fault plane. It can be seen in Eq. 3.12.

$$T_{im}^G(\xi, t - \tau; \mathbf{x}) = c_{ijkl}\frac{\delta G_{km}}{\delta \xi_l}(\xi, t - \tau; \mathbf{x}) \tag{3.12}$$

T_{im}^G is the Green's traction on the fault plane generated by the force at $\mathbf{x}$ along m^{th} direction. Eq. 3.10 can be simplified by using the relation in Eq. 3.12.

$$u_m(\mathbf{x}, t) = \int_\Sigma \delta u_i(\xi) * T_{im} * G(\xi; \mathbf{x})d\xi \tag{3.13}$$

Seismic sources are represented with double-couple sources. In that case only the components of $\mathbf{c}$ with $i \neq j$ contribute the Green's tractions. Green's function can also be replaced by the displacement, which is generated by the double-couple forces with an orientation of the slip and the fault normal vector. Eq. 3.13 can be rewritten as,

$$u_m(\mathbf{x}, t) = \int_{-\infty}^{+\infty} d\tau \int_\Sigma \mu(\delta u_i n_j + \delta u_j n_i)\frac{\delta G_{mi}}{\delta \xi_j}d\xi \tag{3.14}$$

$\mu(\delta u_i n_j + \delta u_j n_i)$ is also known as the moment density tensor, m_{ij}. Seismic moment (M_0) can be expressed as the summation of the moment density tensor(s) over the fault line, $\int_\Sigma m_{ij}d\xi = M_0$. In the finite fault approach, this procedure is applied for each sub fault and

the total displacement can be calculated as their combination.

k^{-2} *Kinematic Model.*   Study of Bernard et al. (1996) explains the k^{-2} model. This kinematic model uses the rupture parameters based on several constraints. It describes the f^{-2} high-frequency decay in the far-field displacement spectra (Bernard et al., 1996; Causse et al., 2009). General slip spatial spectrum equation can be seen in Eq. 3.15.

$$\widetilde{\Delta u_L}(k) = C\frac{\Delta\sigma}{\mu}\frac{L}{k^2} \tag{3.15}$$

$\widetilde{u_L}$, C, $\Delta\sigma$, μ, L and k^2 represent mean slip, geometrical factor ($C \simeq 1$), mean stress drop, rigidity of the environment, fault length and wave number, respectively. Slip distributions are represented in the wave number domain, **k**. **k** is decayed by a $|\mathbf{k}|^{-2}$ beyond the corner wave number, k_c.

$$k_c = K/L_c \tag{3.16}$$

K and L_c represent the roughness parameter and smallest fault dimension, respectively. L_c is the fault width W. Rupture evolution is modelled with the slip pulse of width L_o, rise-time, and constant rupture velocity V_r. Rupture started from the hypocenter point.

Velocity Structure.   Synthetic seismograms are created by using the 1-D velocity model for wave propagation (Table 3.1).

Depth of Layer Top (km)	V_p (km s^{-1})	V_s (km s^{-1})	ρ (g/cm^3)	Q_p	Q_s
0.0	3.00	1.70	2.50	200	100
1.0	4.83	2.60	2.84	400	200
2.0	5.76	3.10	2.94	400	200
5.0	6.51	3.50	3.15	400	200
27.0	7.00	3.80	3.26	600	300
42.0	7.80	4.20	3.50	800	400

Table 3.1: 1-D velocity model that has been used for synthetic seismogram creation. V_p and V_s, ρ, Q_p and Q_s represent P and S wave velocities, density, quality factors of P and S wave of the layers, respectively. Velocity structure is taken from Di Stefano et al. (2011).

Results. In order to get the maximum number of impulsive signals, station distribution, fault geometry, and earthquake parameters are exaggerated. Stations are distributed based on both epicentral distance and azimuthal variations. Epicentral distances vary between 5 km to 120 km with 10 km intervals. Stations are placed where they satisfied 360º azimuthal coverage with 30º differences. For each set of fault plane geometry, there are 288 stations (3 component) identified.

Fault planes are defined with various dip and rake angles. 4 different dip angles which are 30º, 45º, 60º and 90º along with 6 different rake angles which are 0º, 90º, 110º, 165º, −80º and −100º are used recursively. In total, 24 different fault lines are created.

10 different slip distributions are calculated by using the aforementioned method. Synthetic seismograms are calculated for each different slip distributions. Earthquake parameters are used as below:

(i) M_w = 7.2

(ii) Mean Slip = 2.0 m

(iii) k = 0.70

(iv) $V_r = 3200 \, \text{m} \, \text{s}^{-1}$

(v) $Z_{tor} = 0.5 \, \text{km}$

(vi) Length & Width of the fault = $60\,000 \, \text{m}$ & $15\,000 \, \text{m}$

(vii) Number of Green's Traction computed along strike and along dip = 20 & 5

Slip distributions on the fault plane can be seen in Fig. 3.9.
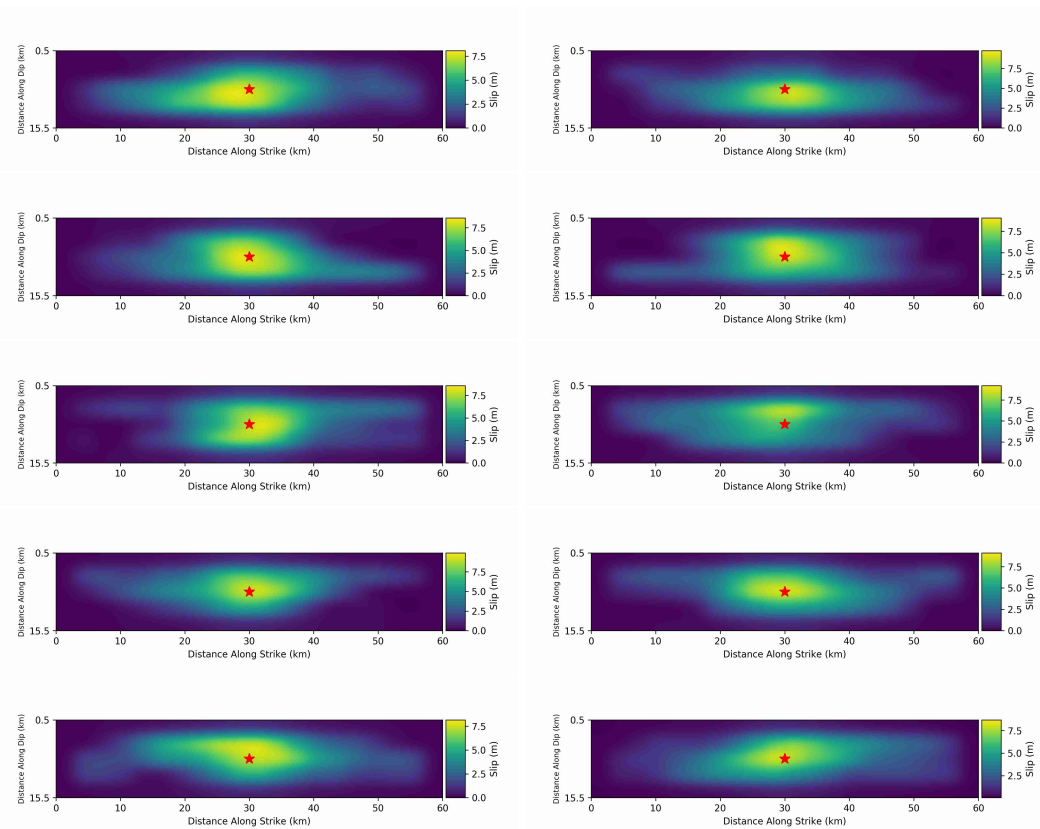
Figure 3.9: Slip models that have been used for the creation of synthetic seismograms. Maximum slip (m) can be seen in the legends of each figure. Red stars indicate the hypocenter of the earthquakes.

In total, 207 360 synthetic waveforms are created. 12 239 of them are labelled as impulsive by Chang et al. (2016); Shahi and Baker (2014) or wavelet method. Some of the impulsive synthetics can be seen in Fig. 3.10.
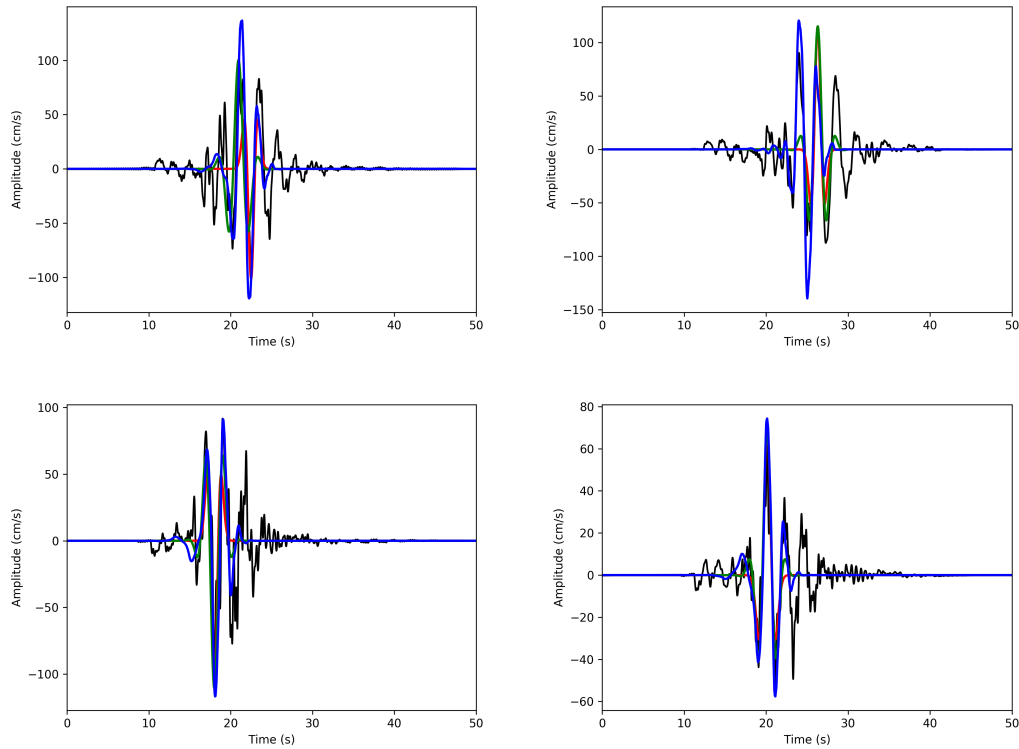
Figure 3.10: Impulsive synthetic velocity waveforms and fitted waveforms on top of them. Epicentral distances, azimuth angles and components of the waveforms are a) 5 km, 210° and East, b) 15 km, 180° and East, c) 20 km, 90° and East d) 30 km, 90° and vertical, respectively. Black, red, blue and green signals represent velocity waveform, Ricker wavelet, 4th order Daubechies wavelet extracted by the algorithm of Shahi and Baker (2014), and extracted waveform by the algorithm of Chang et al. (2016), respectively.

3.3. Pulse Probability

The occurrence of impulsive signal depends on several parameters, described in Section 1. Previous studies on this topic are summarized in Section 1.3. Impulsive signals can cause damage to structures. For instance, the ratio between the structural fundamental period (T_s) and pulse period, T_p, is considered relevant (Iervolino et al., 2012). When the response spectra of pulse-like ground motions are plotted with respect to the period ratio T_s/T_p, the spectral accelerations would be locally amplified around $T_s/T_p = 1$ (Shahi and Baker, 2011a), meaning that larger than expected seismic demands are observed in this particular period region. Seismic hazard studies are important for the characterization of an area or a site of an

interest in terms of the possible effects of earthquakes (Blaikie et al., 2005; Giardini, 1999), but they do not always contain the information about the velocity pulses. By quantifying the probability of pulse occurrence, hazard studies can incorporate this when characterising an area.

In this section, a methodology to create probability maps of pulse occurrence by using selected variables in Section 3.3.1 is explained. To get the relation between the parameters and the pulse probabilities, a naïve Bayesian approach is used.

Preliminary results are presented in Section 3.3.1. These results are used as the basis for upcoming analysis for the probability calculations. Since the main topic of the probability calculation is not the determination of the parameters but the creation of the probability models, they are presented in the method section instead of the results section.

3.3.1. Preliminary Analysis

The preliminary analysis of pulse occurrence uses the following parameters:

(i) Epicentral distance (R_{ep}),

(ii) hypocentral distance (R_{hyp}),

(iii) distance to the surface projection of the rupture (R_{jb}),

(iv) distance from the ruptured fault (R_{rup}),

(v) the horizontal distance to the surface projection of the top edge of the rupture measured perpendicular to the fault strike (R_x),

(vi) depth to top of rupture (Z_{tor}),

(vii) PGA,

(viii) PGV,

(ix) PGV/PGA,

(x) M_w,

(xi) M_0,

(xii) stress drop,

(xiii) rupture area.

The relationship between each variable and the various pulse detection algorithms is shown in Fig. 3.11.
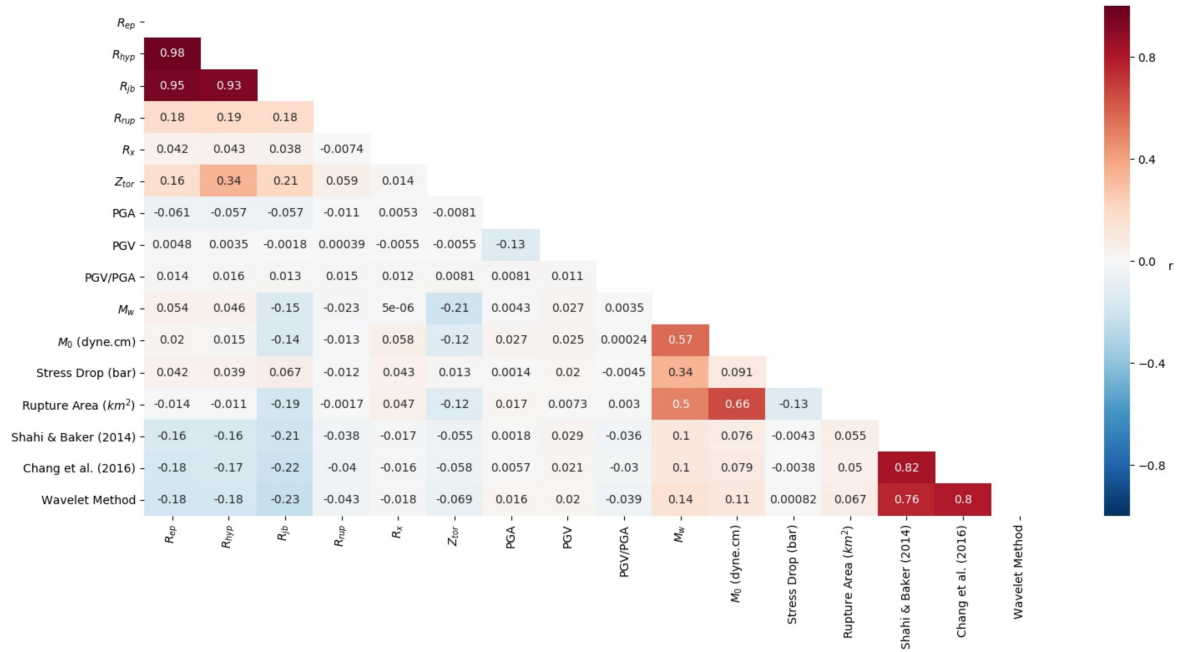
Figure 3.11: Correlation coefficient between the parameters that have been used in the preliminary analysis.

The Pearson correlation coefficient, PCC, (r) measures the linear correlation between two variables (Eq. 3.17), x and y. The correlation coefficient is bounded between -1 and 1. A correlation coefficient of ± 1 implies that x and y have perfect correlation (positive or negative depending on the sign). 0 means no correlation. The PCC has been calculated for each variable. Signals and/or earthquakes without variables are excluded from the analysis.

$$r = \frac{\sum x_i y_i - \dfrac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i^2 - \dfrac{(\sum x_i)^2}{n}}\sqrt{\sum y_i^2 - \dfrac{(\sum y_i)^2}{n}}} \tag{3.17}$$

rs are calculated for all the parameters that can be seen in Fig 3.11, which are calculated as in Eq. 3.17.

In Fig. 3.11, positive values indicate variables which are directly proportional. Conversely, negative values indicate variables which are inversely proportional. For instance, for three different pulse classification algorithms that have been used in the parameter selection, the bigger the PGV, the higher the chance of having impulsive signals (positively correlated); the lower the R_{jb}, the higher the chance of having impulsive signals (negatively correlated). Several parameters are highly correlated with the pulse detection algorithms. R_{jb} has the highest correlation with impulsive signals. Shahi and Baker (2014), Chang et al. (2016) and the wavelet method represent the impulsive signals without the period information in Fig. 3.11. R_{ep} and R_{hyp} have the second and the third highest correlation, but R_{jb} is already providing enough information about position between the fault plane and the station. Thus R_{rup}, R_x and Z_{tor} are excluded from further analysis. The depth of the earthquake has high uncertainty due to it being strongly dependent on the velocity model used in time-travel calculation, the distribution of seismic stations, and picking errors in the waveforms (De Natale et al., 1984). Therefore it is not considered as one of the candidate parameters.

The moment magnitude, M_w, has the 4th highest correlation with impulsive signals (2nd if other distance information is excluded) and it is selected as the 2nd parameter to be used for further analysis. M_0 is strongly correlated with M_w. Hence it is also eliminated. Even though the PGV value is used as a threshold parameter for both Chang et al. (2016); Shahi and Baker (2014) and the wavelet method, it does not show a significant correlation with impulsive signals. Even if it were to show a good correlation with impulsive signals, it would not be used because determining accurate PGV values in near fault areas using GMPEs is not possible. GMPEs predict lower PGV values than observed in the near fault areas (Appendix F). The PGV/PGA ratio is also taken into account in Bray and Rodriguez-Marek (2004)'s study. Since PGA and PGV/PGA are dependent on each other, and PGV is represented by the impulse identification methods, neither of these parameters are used.

Stress drop has insignificant effects on impulsive signals. Rupture area have more significance on impulsive signals, but in most cases rupture area is calculated by using the method of Wells and Coppersmith (1994). Due to that fact, it is not used in order to prevent a false interpretation. None of them are selected for further analysis.

There are other parameters that are briefly explained on Section 1.3. The parameters which are mostly used in previous studies are related with directivity parameters such as R, s, d, θ and ϕ. These parameters are used in Chioccarelli and Iervolino (2013); Iervolino and Cornell (2008); Shahi and Baker (2011b) in order to explain the probability of pulse occurrence on strike slip and non strike slip faults. Those parameters are calculated for several earthquakes and can be found in NGA-West 2 database. Previous studies are used in this database to explain the probability of pulse occurrence. Most of the earthquakes that are part of NGA-West 2 database are well-known and densely studied. However, a bigger database which includes the NGA-West 2 is created in order to explain the impulsive behaviors. In order to calculate above mentioned directivity related parameters (Fig. 1.4), greater knowledge about the earthquakes, rupture behaviors and fault plane information are needed. Parameters such as d and ϕ require precise information about the edges of the ruptured fault plane and velocity structure of the earth. In order to avoid wrong interpretations due to the over simplification and fixation of various parameters, R, s, d, θ and ϕ are not implemented to the pulse probability calculations.

Previous studies of Fayjaloun et al. (2016); Scala et al. (2018) are associated T_p with d, ϕ, R, R_{jb}, R_{hyp}, rise time, rupture velocity and velocity of the bedrock. In order to calculate rise time, kinematic features of each earthquake must be found. Determination of the bedrock velocity of each station requires intensive studies for each station. Over simplification, such as bed rock velocity maps of regions, may create false interpretations. Since all the stations are close to each other, stations that recorded an event would get the same bed rock velocity value due to lack of resolution. Furthermore, T_p correlation with M_w is already done. Results can be seen in Section 4.3.10. Due to the relatively hard estimation of these parameters, none of the above mentioned parameters are used in the probability calculations.

Moreover, the distribution of impulsive signals with EC8 soil class are also investigated. Explanation of soil classes can be seen in Appendix G. Number of impulsive signals that occurred on different soil classes can be seen in Fig. 3.12. Most of the impulsive signals occur in type C and type D soil classes. Type A, type B and type E do not contain sufficient number of impulsive signals. Even though weak soils create majority of the impulses, occurrence of the impulsive signals are not necessarily connected to local soil classes.

Moreover, local soil classes are determined by various methods when they are not available (Appendix G). Heterogeneity of the local soil class determination process makes further interpretation harder. Weak soil classes may increase the amplitudes. Due to the increase, some of the signals may increase the PGV threshold and be labelled as impulsive. However, it is not possible to say that weak soil class changes the waveform and creates impulsive signals. Thus it is not the focus of the thesis. As a result, soil class information is not implemented to the probability models.
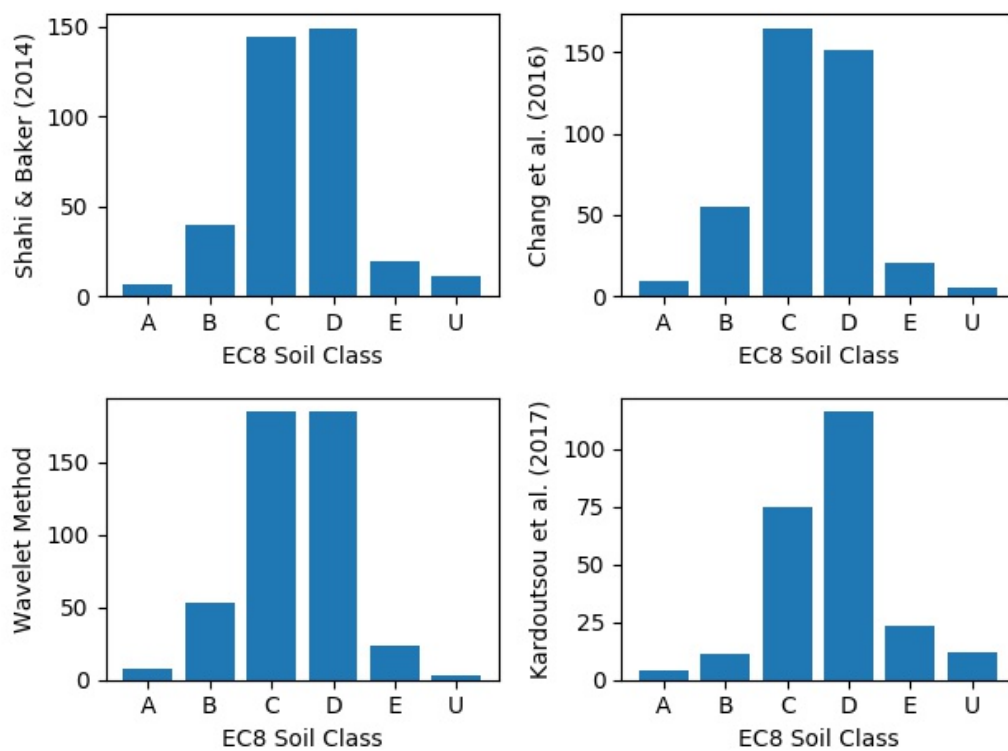
Figure 3.12: Number of impulsive signals determined by the wavelet method, Chang et al. (2016); Kardoutsou et al. (2017); Shahi and Baker (2014) on each EC8 soil class. Unknown indicates (U) the lack of EC8 information from local data providers.

3.3.2. Bayes Theory

Bayes' theorem is widely used in probabilistic theory. It is an advanced application of conditional probability. Conditional probability is the probability of an event happening depending on another event. Bayes' theorem describes the probability of an event with a given prior information, which is the information about the event without having the data of

that event. It can be formulated as in Eq. 3.18.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3.18}$$

A and B are events, and P states the probability. $P(B)$ should be bigger than zero. $P(A|B)$ and $P(B|A)$ are the conditional probabilities, which are the likelihood of events A and B occurring for a given true B and A, respectively. $P(A)$ and $P(B)$ are independent from each other. It means that event A has no effect of event B.

In Eq. 3.18, $P(A)$ is the probability of the event before getting any information about B. $P(A|B)$ is the probability of the event after getting the information about B. Because of that $P(A)$ is called the prior probability and $P(A|B)$ is called posterior probability. The ratio between $P(B|A)/P(B)$ is called the likelihood ratio. $P(B)$ is called the evidence. Eq. 3.18 can be defined as the posterior probability equals the prior probability time, the likelihood ratio.

$P(A)$ is called the prior, that is known through an understanding of the nature of the event A. $P(A|B)$ is the posterior which is the observation. Posterior probability increase, decrease or remain as before after each new introduced data. An example of Bayes' theorem can be seen in Appendix H. In other words Eq. 3.18 can be written verbally as below:

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}} \tag{3.19}$$

3.3.3. Naïve Bayes Classifier

In Bayes classifier A will be the classes, which are impulsive and non-impulsive classes. B is the feature that is going to be used to determine the impulsive and non-impulsive signals. A and B will get indices and become A_j, and B_i. Binary classification will be applied on the data and A_0 and A_1 will be the probabilities of non-impulsive and impulsive signals, respectively. B_i will have the dataset of M_w, source-to-site azimuth and R_{jb}. They are also called predictor features. $P(B|A)$ can be rewritten as:

$$P(B_i|A_j) = P(B_1|A_j) \times P(B_2|A_j) \times \ldots \times P(B_n|A_j) \qquad (3.20)$$

The probability of class A_j with the set of values at B_i, is explained as in Eq. 3.20. Eq. 3.18 can be modified with the information in Eq. 3.20 and rewritten as in Eq. 3.21.

$$P(A_j|B_{1,\ldots,n}) = \frac{P(B_1|A_j)P(B_2|A_j)\ldots P(B_n|A_j)P(A_j)}{P(B_1)P(B_2)P(B_3)\ldots P(B_n)} \qquad (3.21)$$

Eq. 3.21 can be put into words as, the probability of the class (A_j), for a given predictor feature combination $(B_{1,\ldots,n})$, is as a function of the probability of the predictor feature combination $(P(B_{1,\ldots,n}|A_j))$ with the given category (A_j). It is the likelihood term times the probability of the category $P(A_j)$. The prior (numerator part of Eq. 3.21) all divided by the evidence term, which is the combination of the predictor features $(P(B_{1,\ldots,n}))$.

Prior is independent of the predictor features (new data observations). There are impulsive and non-impulsive signals on the classes. Sum of the prior would be 1 which is the summation of the probability of impulsive and non-impulsive probabilities. A naïve prior would use 50% to 50% probability of each class.

Evidence term does not consider the categories. It is the combination of the predictor features. It can be seen as the standardizer of the probabilities to make sure that they sum to 1.

In the likelihood term, the conditional relationship between all of the predictor features are ignored. It is called conditional independence. In other words, it means that $P(B_1|A_j)$ and $P(B_2|A_j)$ must affect the $P(A_j|B_{1,...,n})$ separately. High or low correlation between $P(B_1|A_j)$ and $P(B_2|A_j)$ does not mean that the correlation have a meaningful effect on $P(A_j|B_{1,...,n})$. This assumption of conditional independence is the reason why this method is called Naïve Bayes classifier.

Likelihood terms can be calculated by assuming conditional distributions are parametric distribution. It is called Naïve Gaussian Bayes approach. It is assumed that conditional distributions are Gaussian. In such a case, conditional probability, $P(B_n|A)$ can be written as in Eq. 3.22. μ and σ are the average and standard deviation of each indices of B_n, and exp is the abbreviation of e^x.

In classes, which are impulsive and non-impulsive, μ and σ are calculated for each feature, which are source-to-site azimuth, M_w and R_{jb}.

$$P(B_n|A) = \frac{1}{\sqrt{2\pi\sigma_A^2}}exp\left(-\frac{(B_n-\mu_A)^2}{2\sigma_A^2}\right) \qquad (3.22)$$

Since B has higher dimensions, multivariate Gaussian distribution is applied to it. Multivariate Gaussian distribution is a vector with multiple Gaussian distribution variables. Any linear combination of the variables is also Gaussian distribution. One should modify the Eq. 3.22, since it is actually valid for the normal distribution in 1 variable.

In normal distribution with 1 variable, $-(B_n-\mu_A)^2/2\sigma_A^2$ is a parabola. In the multivari-

ate case, Eq. 3.22 should be modified with a quadratic version of B_n. Modified multivariate Gaussian distribution formula can be seen in Eq. 3.23.

$$P(B_n|A) = \frac{1}{\sqrt{2\pi\sigma_A^2}} exp\left(-\frac{1}{2}(B_n - \mu_A)^T \sigma^{-1}(B_n - \mu_A)\right) \tag{3.23}$$

Denominator of Eq. 3.21 is constant with all the values in the dataset. One can remove the denominator and introduce the proportionality as in Eq. 3.24.

$$P(A|B_1, \ldots, B_n)\alpha P(A_j)\Pi_{i=1}^{n}P(B_n|A) \tag{3.24}$$

Since there are two different classes, A with the maximum probability is the answer that needs to be found. One can obtain the class by using the Eq. 3.25.

$$A = argmax_A P(A)\Pi_{i=1}^{n}P(B_n|A) \tag{3.25}$$

Conditional probabilities and priors are assigned accordingly. After that, the model is used for making predictions for the impulsive and non-impulsive signals by giving input parameters of source-to-site azimuth, M_w and R_{jb}.

3.3.4. Application

Parameters that have been selected in Section 3.3 are used in the naïve Bayes method. Multivariate Gaussian distributions are created for M_w, source-to-site azimuth and R_{jb}. Likelihood function is selected as multivariate Gaussian distribution.

Dataset is subdivided into 3 categories which are the strike slip, normal and reverse faults. Fault types are defined as in Appendix A.2. Normal and normal-oblique and reverse and reverse-oblique fault types are merged to normal and reverse faults, since there is no significant number of data on the fault types with oblique features. Data on each fault line are also sub categorized with their position with respect to ruptured fault. Determination of hanging wall and foot wall has been done as in Appendix A.5. Python package of Pomegranate (Schreiber, 2018) is used for the multivariate naïve Bayesian approach on selected parameters to create a model for each fault type. Results of the pulse occurrence can be found in Section 4.3.

4. RESULTS

4.1. Wavelet Method Results

There are four main features of the pulse-shape signals as explained in Section 1. Position, amplitude, phase and period of the pulse are the focus of the method. One way to determine the validity of the method is to compare spectral response of the original and created signals. Large amplitudes at long periods in velocity response spectrum is also an indicator of the pulse shaped signal. The wavelet signal is expected to imitate the behavior around the pulse period. In order to do that, spectral responses of the original strong motion data and the wavelet that is expected to mimic the pulse are visually compared . Some of the spectral response graphs can be seen in Fig. 4.1. Acceleration waveforms are filtered with band-pass filter between $0.05\,\text{Hz}$ to $10\,\text{Hz}$.
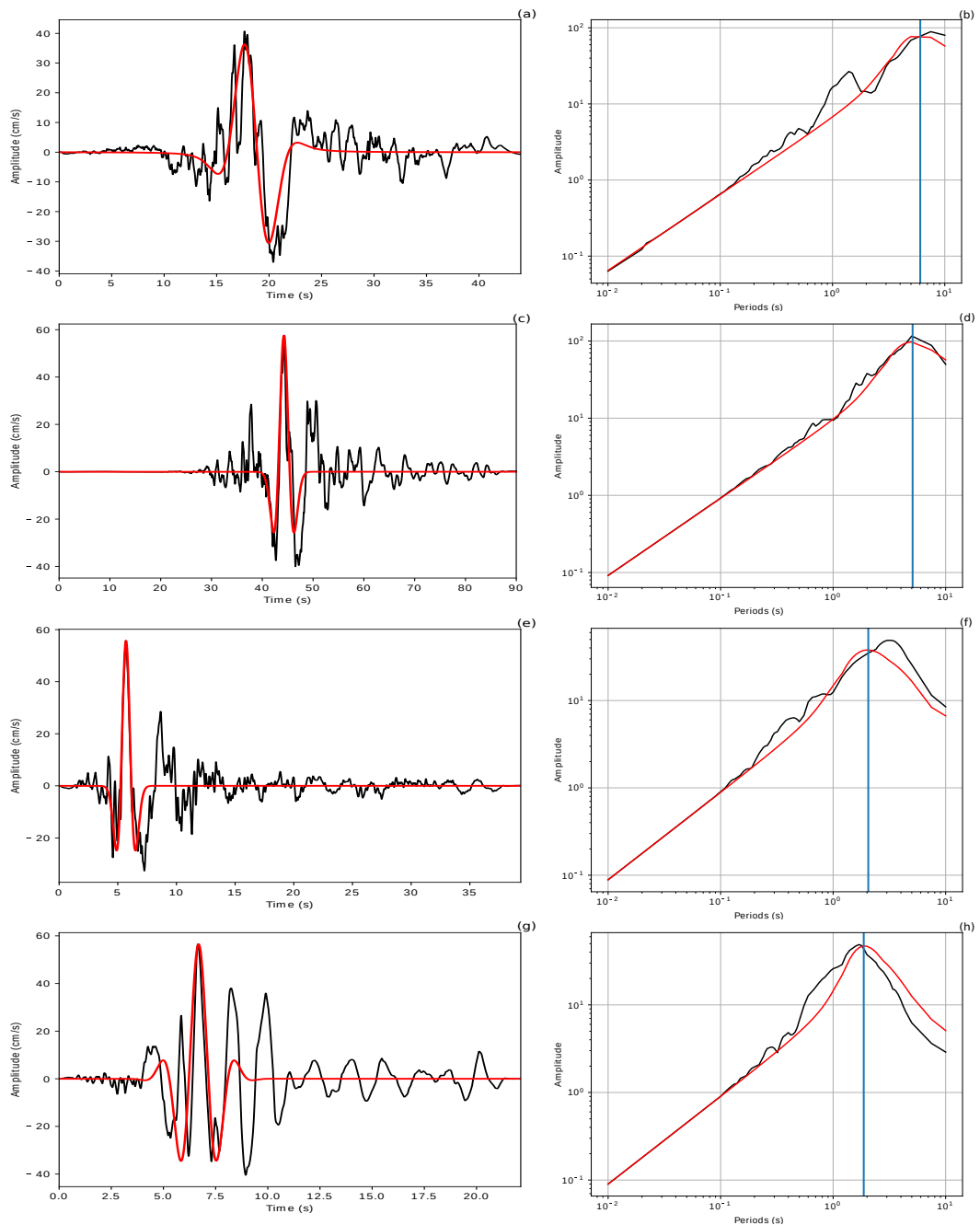
Figure 4.1: Velocity waveform and fitted wavelets (left column) and pseudo spectral velocity graphs (right column) of 1992 Landers Earthquake, Yermo Fire Station signal and obtained Ricker wavelet signal (a,b), 1999 Chi-Chi Taiwan Earthquake, TCU039 Station signal and obtained Ricker wavelet signal (c,d), 1980 Irpinia Earthquake (M_w = 6.9), STN Station (R_{ep} = 30.35 km) signal and obtained Ricker wavelet signal (e,f) and 1994 Northridge Earthquake (M_w = 6.7), SCE Station (R_{ep} = 24.97 km) signal and obtained 3rd order Morlet wavelet signal (g,h). In all figures, the blue line represents the period of the pulse. Red and black colors indicate the velocity waveform and fitted wavelet signal, respectively.

Phases of the impulsive part of the velocity waveform are also analyzed. Impulsive signals that can be identified with Ricker wavelet are analyzed, since Ricker wavelet dominated the representation of the impulsive part of the signals. Ricker wavelet can be fitted to the original waveform to visualize the impulsive part more easily, however it is not providing further information about the impulsive part. Method for phase determination is explained in Appendix I.

Chang et al. (2016); Kardoutsou et al. (2017); Shahi and Baker (2014); Zhai et al. (2018) and wavelet method identified 406, 241, 371, 151 and 458 waveform as pulse-shape signals out of 21458 waveforms, respectively. 22 of the signals are identified as pulse-shape signals by five of these studies. 290 signals are determined as pulse shaped signal by Chang et al. (2016); Shahi and Baker (2014) methods and wavelet method. Location of the impulsive part and the wavelet that mimic the features are also investigated. 57 of the pulses are located outside the region of PGV. 445 of the pulse-shaped signals are mimicked better by using a Ricker wavelet, whereas only 13 of them are mimicked better by using a 3rd order Morlet wavelet. A 4th order Morlet wavelet is not suitable to mimic any of the pulse-shape signals.

It is important to mention that, impulsive signals that have occurred outside of the PGV region are mostly influenced by local soil conditions. Physical conditions that produce impulsive signals are explained in Section 1.1. Impulses that are located other than PGV region are not following the sources of impulsive signals, which are fling step and directivity effects. On the other side, impulsive signals where impulse is located in PGV region are not necessarily fueled by these two effects. PGV threshold of $30\,\mathrm{cm\,s^{-1}}$ is applied in the decision making process. Local soil conditions may amplify the amplitudes of signals. Hence signals may be identified as impulsive due to the amplification. On the contrary, signals with impulsive shapes may be identified as non-impulsive because of low amplitude of PGV. Due to the lack of impulsive signals in the dataset, it is not possible to distinguish the primary (fling step, directivity) and secondary (basin effect, local site condition) parameters on impulsive signals.

4.1.1. Comparison With Previous Studies

Database is analyzed with the algorithms of Chang et al. (2016); Kardoutsou et al. (2017); Shahi and Baker (2014); Zhai et al. (2018) and manual picking done visually. The resutlts of these methods can be found in the Table 4.1 and Table 4.2. In Table 4.1, number of signals with an impulsive part that occurred on PGV region and occurred other than PGV region is given.

	Shahi and Baker (2014)	Chang et al. (2016)	Kardoutsou et al. (2017)	Zhai et al. (2018)	Manual Picking	Wavelet Analysis
Pulse at PGV	277	312	154	68	334	401
Pulse Other than PGV	29	34	12	10	38	57
Total	316	346	166	78	372	458

Table 4.1: Number of impulsive signals that are detected by both wavelet method and previous studies of Chang et al. (2016); Kardoutsou et al. (2017); Shahi and Baker (2014); Zhai et al. (2018). Manual picking done by hand is also given.

In Table 4.2, impulsive parts of the signals that can be mimicked by using Ricker, 3rd and 4th order Morlet wavelets are presented. These results are in the case of the impulsive signal that identified as wavelet method.

	Shahi and Baker (2014)	Chang et al. (2016)	Kardoutsou et al. (2017)	Zhai et al. (2018)	Manual Picking	Wavelet Analysis
Ricker Wavelet	315	343	161	71	366	446
3rd Order Morlet Wavelet	1	3	5	7	6	12
4th Order Morlet Wavelet	0	0	0	0	0	0
Total	316	346	166	78	372	458

Table 4.2: Number of impulsive signals that are mimicked properly by using Ricker, 3rd order Morlet and 4th order Morlet wavelets with their comparisons with respect to Chang et al. (2016); Kardoutsou et al. (2017); Shahi and Baker (2014); Zhai et al. (2018) algorithms. Manual picking done by hand is also given.

Differences between previous methods and wavelet method in terms of numerical results are also analyzed. One can notice that some signals are not identified as impulsive

signal by one study whereas considered as impulsive in another one (Fig. 4.2a, Fig. 4.2b). Numerical results of Shahi and Baker (2014), Chang et al. (2016) and the wavelet method (Eq. 3.5) are explained in Table 4.3.

Waveform Name	T_p Shahi and Baker (2014)	PI	T_p Chang et al. (2016)	E_p	T_p	Waveform Energy	Wavelet Power Spectrum Energy
TCU078	0	-0.71	1	0.50	3.60	74.15	63.79
Vineyard Canyon 1E	0	-1.63	1	0.50	1.27	47.90	37.02
Brawley Airport	0	-2.4	0	0.29	6.05	57.91	47.53
D08C	3.89	1.90	0	0.30	0	29.74	23.41
AQK	2.04	0.69	1.7	0.38	0	34.76	24.33
Pacoima Dam	0.78	7.69	0.7	0.38	0	39.78	19.19
KJMA	1.09	5.82	1	0.35	0	38.34	19.24
Port Island	2.7	5.94	2.1	0.39	0	32.35	18.21

Table 4.3: Numerous signals with dissimilar impulsive results. T_p is pulse period with studies that are used to calculate. E_p is explained in Eq. 1.11. Waveform Energy and Wavelet Power Spectrum Energy is calculated as in Eq. 3.5 and Eq. 3.8 depending on the position of the impulsive part, respectively. T_p gets 0, if the signal is not found impulsive by the method that is used for the calculation. Waveforms are 25th of September 1999 Chi-Chi Taiwan Earthquake (M_w = 6.3) TCU078 Station (R_{ep} = 17.94 km), 28th of September 2004 Parkfield California Earthquake (M_w = 6.0) Vineyard Canyon 1E Station (R_{ep} = 18.76 km), 15th of October 1979 Imperial Valley California Earthquake (M_w = 6.5) Brawley Airport Station (R_{ep} = 43.15 km), 17th of May 1976 Gazli USSR Earthquake (M_w = 6.8) Karakyr Station (R_{ep} = 12.81 km), 21st of February 2011 Christchurch New Zealand Earthquake (M_w = 6.2) D08C Station (R_{ep} = 3.30 km), 6th of April 2009 L'Aquila Italy Earthquake (M_w = 6.1) AQK Station (R_{ep} = 1.76 km), 17th of January 1994 Northridge California Earthquake (M_w = 6.69) Pacoima Dam Station (R_{ep} = 20.36 km), 16th of January 1995 Kobe Japan Earthquake (M_w = 6.9) KJMA (R_{ep} = 18.27 km) and Port Island (R_{ep} = 19.25 km) Stations, respectively.

An example of the regions that are considered by Chang et al. (2016) and wavelet method can be seen in Fig. 4.2. Velocity pulse has occurred outside of the PGV region. One can notice that, Chang et al. (2016) covers large part of the waveform. This feature makes it harder to analyze the impulsive part of the waveform since it is spoiled by the non-impulsive

parts of the waveform.
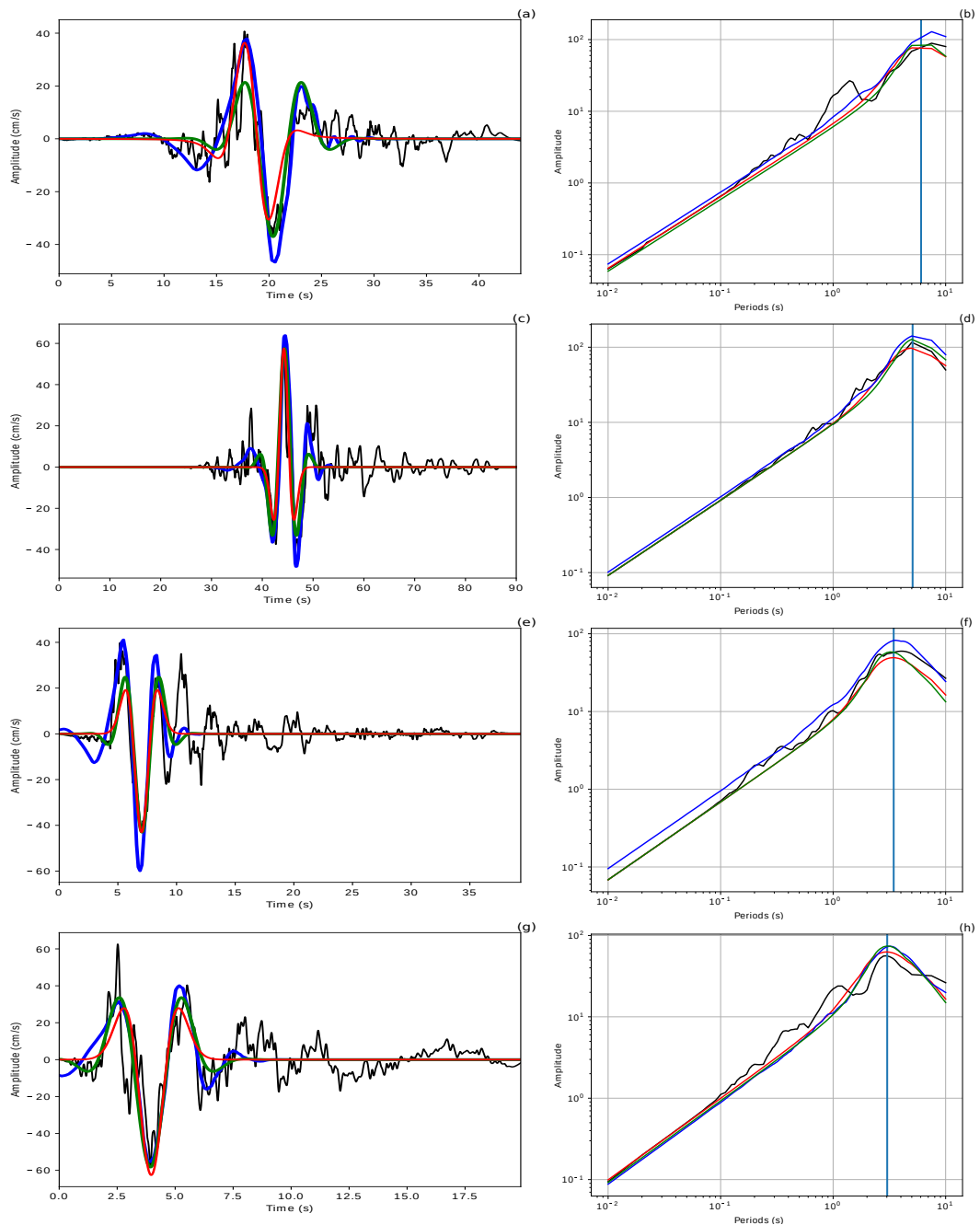
Figure 4.2: Velocity waveform and fitted wavelets (left column) and pseudo spectral velocity graphs (right column) of 1992 Landers Earthquake, Yermo Fire Station signal (a,b), 1999 Chi-Chi Taiwan Earthquake, TCU039 Station signal (c,d), 1980 Irpinia Earthquake (M_w = 6.9), STN Station (R_{ep} = 30.35 km) signal (e,f) and 1994 Northridge Earthquake (M_w = 6.7), Rinaldi Reveiving Station (R_{ep} = 9.30 km) signal (g,h). Black, red, blue and green signals represent velocity waveform, Ricker wavelet, 4th order Daubechies wavelet extracted by the algorithm of Shahi and Baker (2014) and extracted waveform by the algorithm of Chang et al. (2016), respectively. Vertical blue line represents the period of the pulse.

Shahi and Baker (2014) is not able to identify some of the signals that are considered as impulsive by both Chang et al. (2016) and wavelet method. *PI* is very close to the threshold of 0 on these examples. It is also valid for Chang et al. (2016). Threshold of 0.34 for Eq. 1.11 is almost exceeded at D08C Station. Brawley Airport Station is also just below the thresholds of Eq. 1.10 and Eq. 1.11, which gets a long pulse period by wavelet method (Fig. 4.3). On the other hand, the wavelet method fails when the pulse period is short. In waveform energy parameter, which is the left side of the numerator of Eq. 3.5, the threshold is exceeded by almost all non-impulsive signals, which are AQK, Pacoima Dam, KJMA and Port Island Stations (Fig. 4.3). However, wavelet power spectrum energy (right side of the numerator of Eq. 3.5) is so small that it makes the signal not impulsive. Common feature of these signals that are identified as impulsive is the fact that they have very short impulses. One of the features of impulsive signals is their long periods (Section 1). Wavelet method filters short period signals thanks to wavelet power spectrum energy.
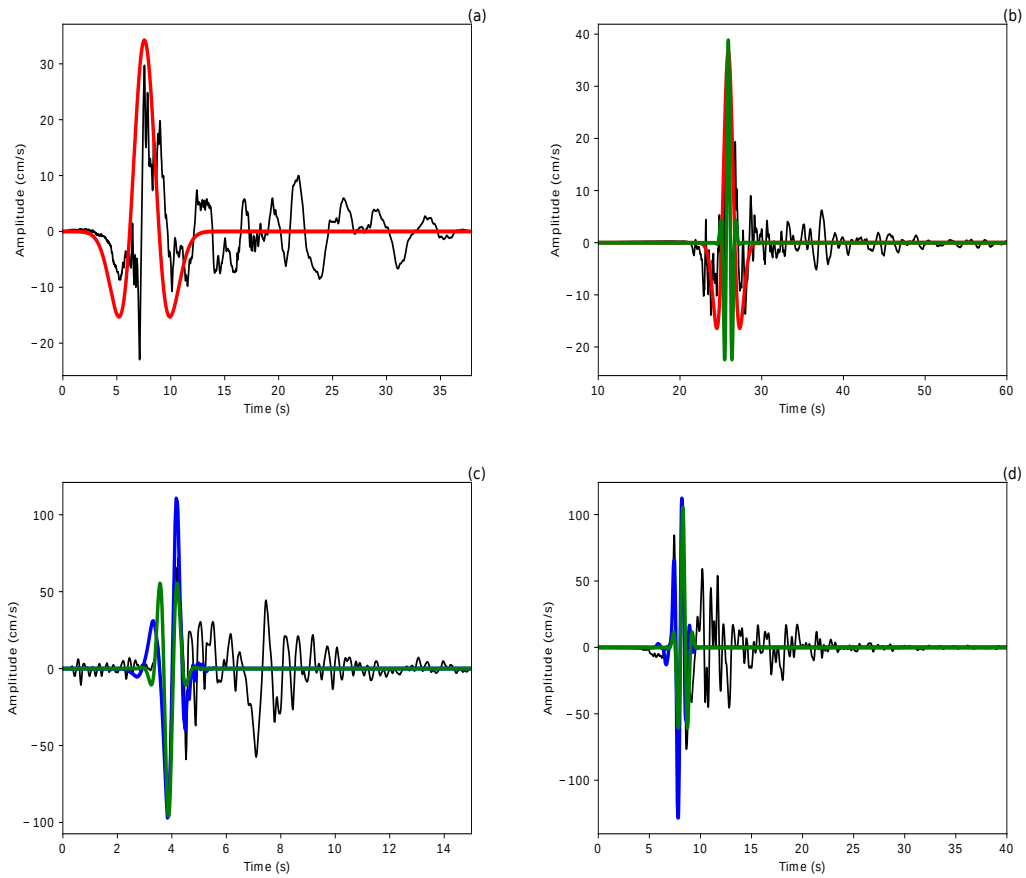
Figure 4.3: Velocity waveforms and fitted wavelets of Brawley Airport (a), TCU078 (b), Pacoima Dam (c) and KJMA (d) stations. Colors represent the studies as in Fig. 4.2. Brawley Airport is labeled as impulsive by only wavelet method (T_p = 6.04), TCU078 has been considered as impulsive by both wavelet method (T_p = 3.60) and Chang et al. (2016) (T_p = 1.00), Pacoima Dam and KJMA are identified as impulsive by both Shahi and Baker (2014) and Chang et al. (2016) with pulse periods of 0.78, 0.70 and 1.09, 1.00, respectively.

Due to the relatively lower correlation between the studies of Kardoutsou et al. (2017); Zhai et al. (2018) and wavelet method, these two methods are excluded in the following analysis.

4.1.2. Pulse Distribution With Fault Type

Distribution of the impulsive and non-impulsive signals on fault types are presented in this section.
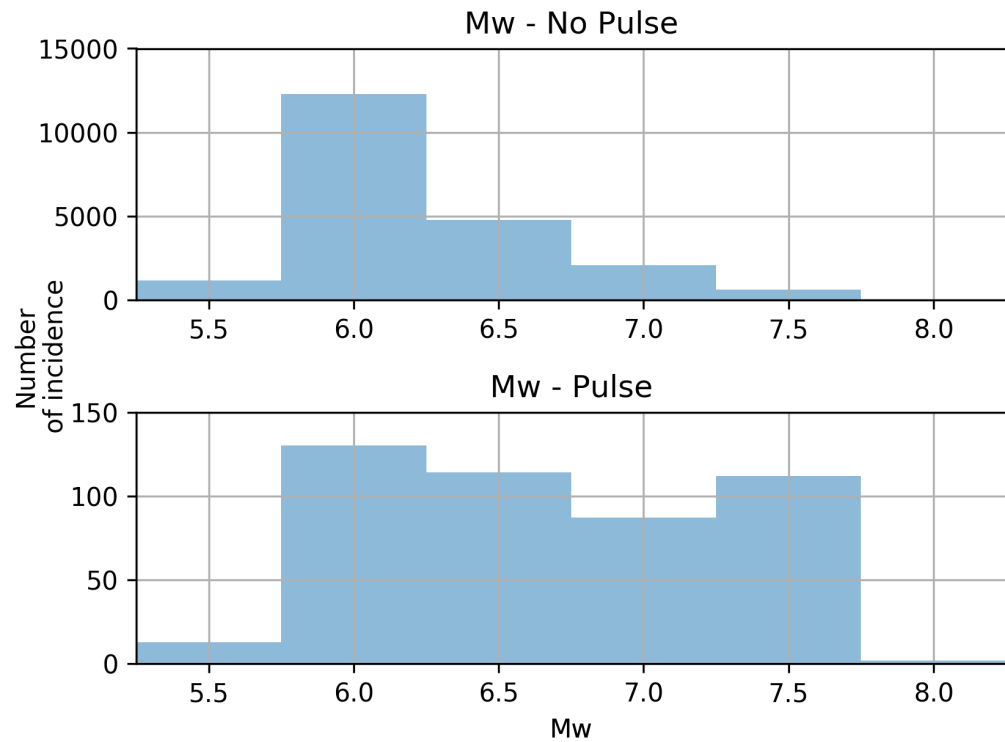
Figure 4.4: Distribution of impulsive and non-impulsive signals with M_w. Magnitudes are binned with 0.5 interval from 5.5 to 8.5 for this section and Section 4.1.3.

Number of signals with impulsive and non-impulsive behavior for different M_w bins can be seen in Fig. 4.4. Even though the highest number of impulsive signals are occurring between M_w of 6.0 and 6.5, ratio of impulsive to non-impulsive signals are higher in M_w and range between 7.5 and 8.0. Ratio between impulsive and non-impulsive signals are increasing with M_w.
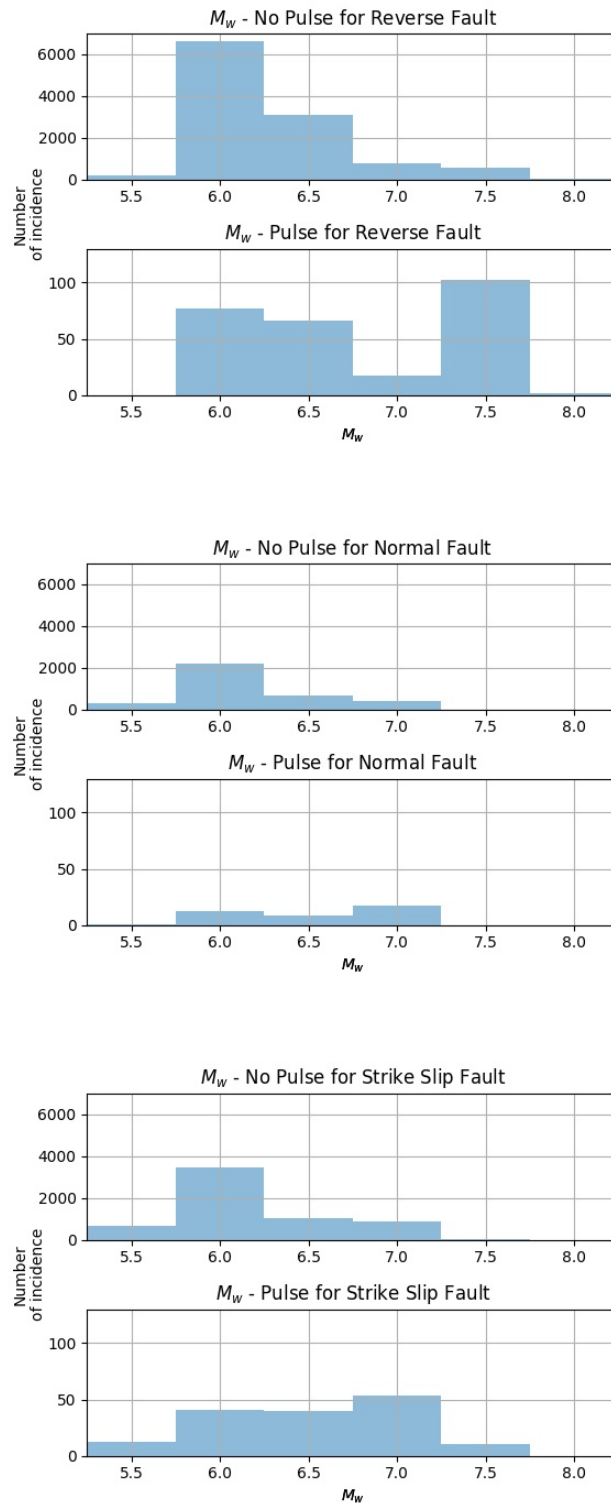
Figure 4.5: Distribution of impulsive and non-impulsive signals with M_w for reverse (top), normal (middle) and strike slip faults (bottom). Reverse-oblique and normal-oblique faults are merged into reverse and normal faults, respectively.

Number of signals with impulsive and non-impulsive behavior for different M_w bins for normal, reverse and strike slip, respectively, which can be seen in Fig. 4.5. Trend of having more impulsive signals with increasing M_w applies for each fault type.

4.1.3. Pulse Distribution With Source-to-site Azimuth

Distribution of the impulsive and non-impulsive signals with source-to-site azimuth are presented in this section. Determination of the azimuth angles with respect to ruptured fault area is done as explained in Appendix A.5. Stations that are located in the site-to-source azimuth less than $0°$ are on the foot wall whereas the ones in site-to-source azimuth greater than $0°$ are on the hanging wall. There is no foot wall, hanging wall separation in strike slip faults due to the absence of vertical displacement between plates.
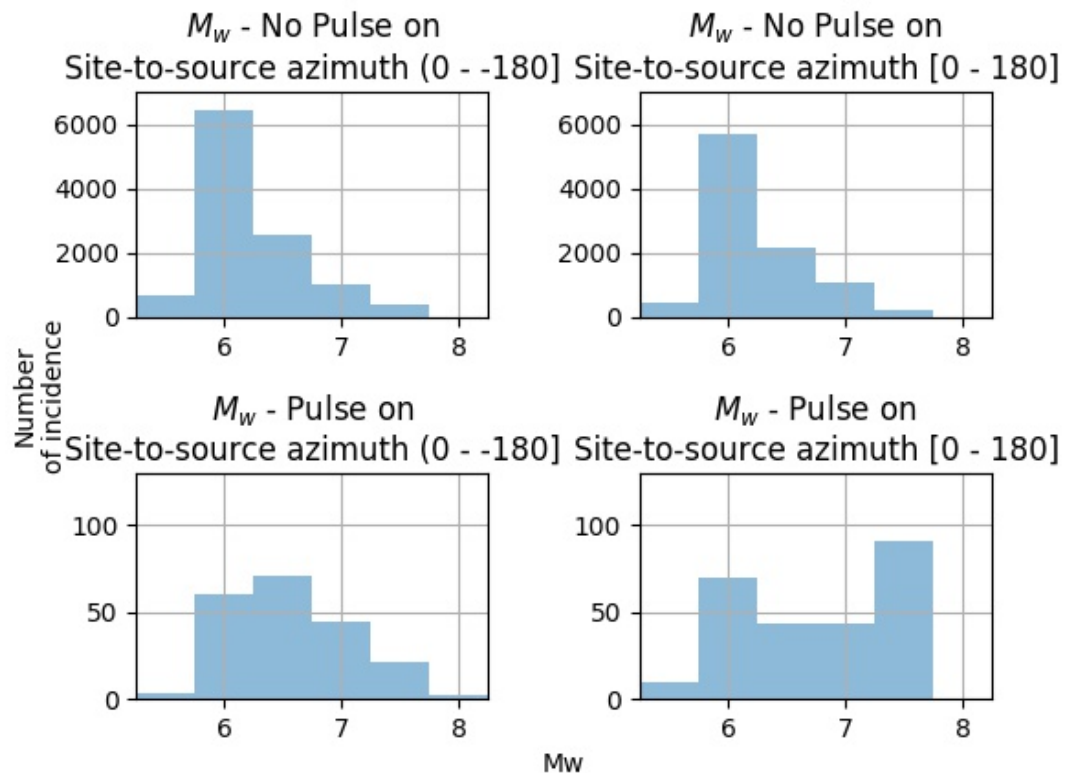
Figure 4.6: Distribution of impulsive and non-impulsive signals with M_w for stations with positive and negative site-to-source azimuths.

Number of signals that are recorded on positive and negative source-to-site azimuths with impulsive and non-impulsive behavior for different M_w bins can be seen in Fig. 4.6. Number of impulsive signals are slightly bigger in positive source-to-site azimuths than negative source-to-site azimuths. Number of signals in foot wall is bigger than the hanging wall.
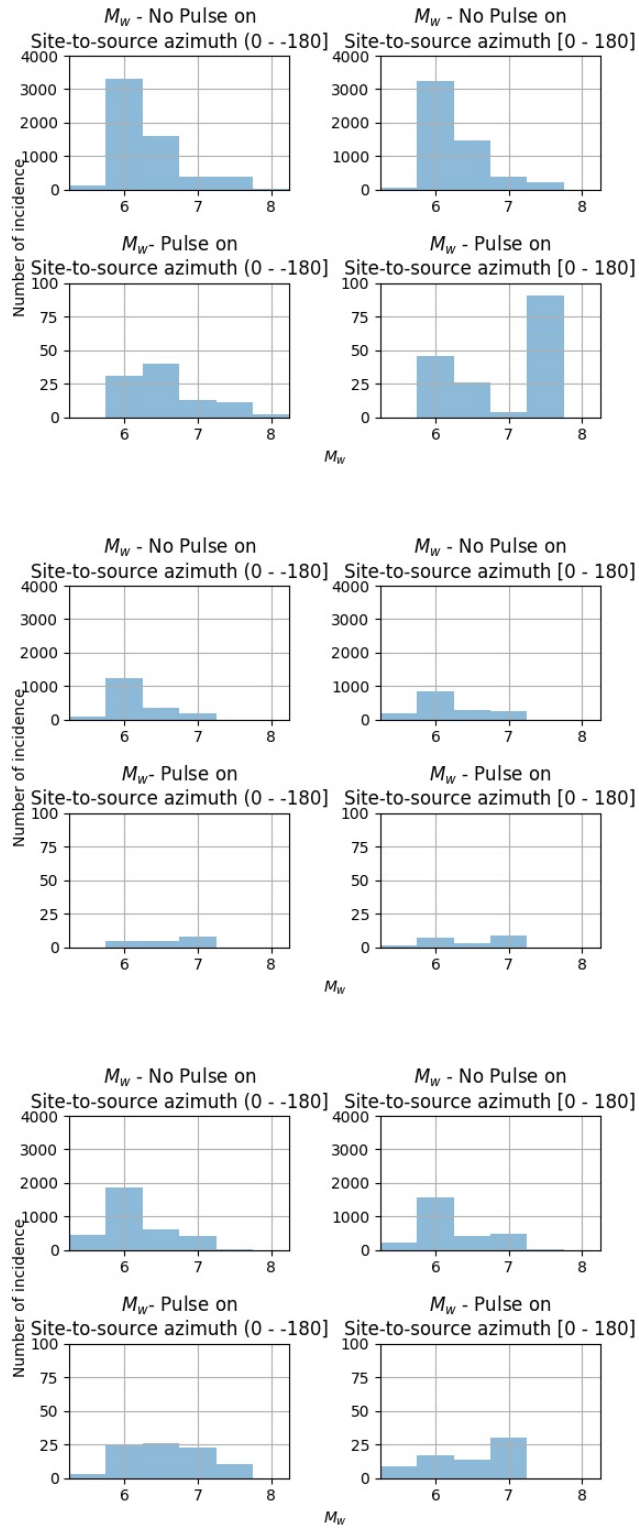
Figure 4.7: Distribution of impulsive and non-impulsive signals with M_w for stations with positive and negative source-to-site azimuths for reverse, normal and strike slip faults, respectively.

Number of signals that are recorded on positive and negative source-to-site azimuths with impulsive and non-impulsive behavior for different M_w bins for normal, reverse and strike slip can be seen in Fig. 4.7. Number of signals that are recorded in reverse earthquakes are bigger than normal faults and strike slip faults. Increasing number of impulsive signals with increasing M_w can be seen in Fig. 4.7 as well.

Developed wavelet mode (Ertuncay and Costa, 2019) can be found at along with the dataset that has been used in the study.

4.2. Convolutional Neural Networks

4.2.1. Experimental procedure

A neural network is created in order to identify impulsive and non-impulsive signals. Features that are explained in Section 3.2 are implemented in the architecture of the neural network. Velocity waveforms are processed by 4 1D convolutional layers that are connected to each other with 1D max pooling layers. ReLU activation function is used between the layers. A dropout layer is added in the end of the 4th 1D max pooling layer. Then, the PGV of the input velocity waveform is merged into the output vector of the waveform as a number. 3 dense layers are applied to the merged vector and a sigmoid activation function finalized the architecture. The created model gives the probability of the impulsiveness of a given signal by using its data points and the highest amplitude.

The architecture of the CNN model can be seen in Fig. 4.8. When a waveform is given to the model, it has 800 data points, which are the points of the velocity signal. 64 neurons are used for the first 1D convolutional layer (the input layer). The output is a 789×64 matrix. Each column of the matrix has the weight of a single filter. Each filter has 789 weights. After that, max pooling layer reduces the size of the matrix to prevent overfitting. Following these, 1D convolutional layers and max pooling layers act as the first 1D convolutional layer and max pooling layer. 32, 16 and 16 neurons are used in the second, third and forth convolutional layers, respectively. Dropout randomly gives 0 value to weights to prevent the problem of overfitting. Dense layers take the vectored high level information about the ve-

locity waveform and reduce the dimension by doing another matrix multiplication and using the sigmoid activation function. At the end of the process, the output value is the probability of impulsiveness of a given signal. A loss function is calculated according to the result and weights are updated in order to reduce the loss value. This procedure carries out iterations for batches of signals.

Waveforms that have been used to train the data are randomly divided into folds. Each fold has been used to train the model and the model with the highest accuracy rate is selected as the final model. Intel® Xeon® Gold 6140 CPU @ 2.30GHz with 34 cores and equipped with 196GB of RAM along with Tesla V100 with 16GB of RAM GPU are used for the process. Duration of the process is in order of dozens of seconds.
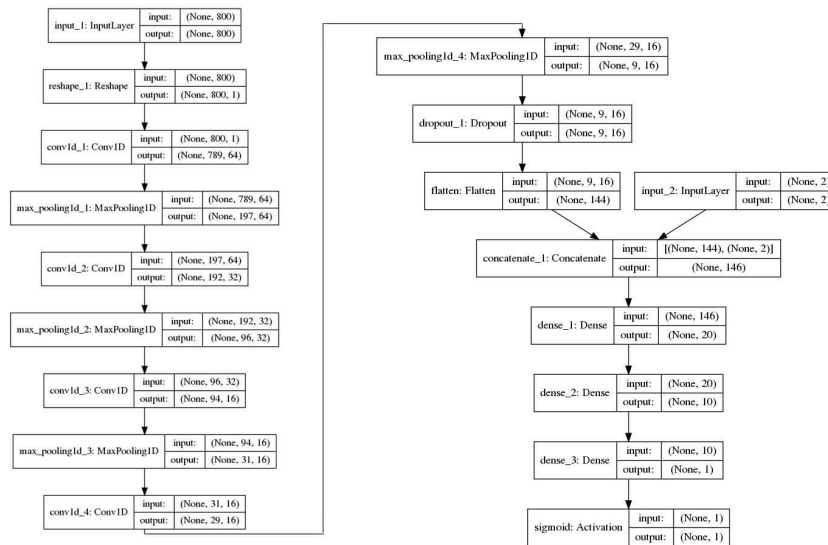
Figure 4.8: Architecture of the CNN. Layers of Reshape and Flatten are for the mathematical manipulations to change data types and their lengths. Hence, they have no effect on the process. None values inside the input and output vectors mean that the batch size for that layer is ignored. Batch sizes are automatically assigned in the fitting process. 2nd and 3rd dimensions are the length of the input that have been processed and the number of filters in 3D input and output vectors, respectively. Number inside the 2D vectors are the length of the input.

4.2.1.1. Cross-Validation.   Cross validation is a way of measuring the performance of the model by using a new, unseen, dataset. It is a way to observe the performance of the trained

model in practice. Cross validation is important to generalize the model without being influenced by the input data.

Cross-validation can be used to measure overfitting and underfitting. Overfitting is the memorization of features of the signals in the training set, which are not present in the unseen data. Underfitting means not capturing enough patterns in the training data.

4 fold cross-validation divides the data to 4 subsets. It is done by using 1 of the subsets as the test set and validation set, while keeping the rest as training set. The error is calculated by averaging the error of 4 different trials. By doing that, one can be sure that each data is used as a validation just once and used as a training data for the rest of the procedure. Biases (Section 3.2.2.1) are reduced by using 4 fold cross-validation method.

Training and testing data is chosen randomly. Ratio between the training and the testing set is determined as 75 % (3 folds) and 25 % (1 fold) of the all signals. Ratio between the labels in each set is the same. At the end of the cross-validation, results are presented as a table of all the outputs, which is called confusion matrix. An example table can be seen in Table 4.4.

		Predicted	
		Negative	**Positive**
Real	**Negative**	True Negative	False Positive
	Positive	False Negative	True Positive

Table 4.4: An example confusion matrix. True Positive (TP): Prediction result is positive and the real result is positive. True Negative (TN): Prediction is negative and the real result is negative. False Positive (FP): Prediction is positive but the real result is negative (also known as a Type 1 error) . False Negative (FN): Prediction is negative but the real result is positive (also known as a Type 2 error).

Performance of the CNN method can be seen in Table 4.5. Randomly chosen single fold is the source of non-impulsive results. Positive examples are the manually picked real signals.

		Predicted	
		Negative	**Positive**
Real	**Negative**	99.91 %	0.09 %
	Positive	3.66 %	96.34 %

Table 4.5: Performance of CNN method in terms of percentages of true and false predictions. Negative examples are taken from one fold.

4.2.2. Predicted Waveforms

Manually picked impulsive signals are investigated. Waveforms that are identified by Chang et al. (2016); Shahi and Baker (2014) and wavelet method are compared with the results of CNN. Output of the CNN is the probability of impulsiveness of the signal. Since the signal can only be impulsive or non-impulsive, results that are larger than 0.5 are considered as impulsive. Various waveforms are shown in Fig. 4.9, Fig. 4.11, Fig. 4.13 and Fig. 4.14. In Fig. 4.9, waveforms are identified as impulsive by Chang et al. (2016); Shahi and Baker (2014), wavelet method, and CNN. In Fig. 4.11, waveforms are labeled as impulsive by at least one of the 3 studies and CNN. Waveforms that are only characterized as impulsive can be seen in Fig. 4.13. Waveforms that are identified as impulsive by at least one of 3 studies but not identified as impulsive by CNN is presented in Fig. 4.14.
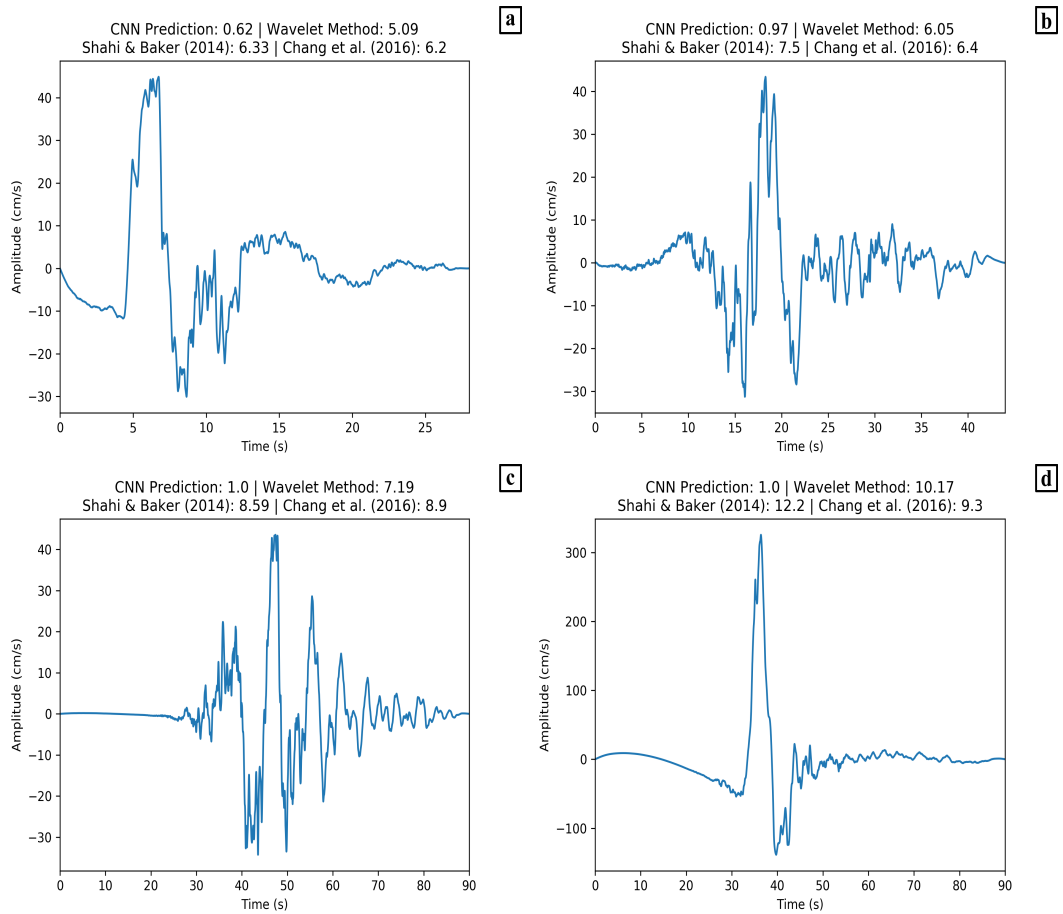
Figure 4.9: Velocity waveforms that are identified as impulsive by Chang et al. (2016); Shahi and Baker (2014), wavelet method and CNN method. Prediction is the normalized probability of the impulsiveness of the signal, determined by the CNN method. Chang et al. (2016); Shahi and Baker (2014) and wavelet method are the T_p of the signal calculated by the methods (Information in Fig 4.11, Fig. 4.13 and Fig. 4.14 are the same). a) Gebze station ($R_{ep} = 47.03\,\text{km}$) in Izmit, Turkey earthquake ($M_w = 7.5$), b) Yermo Fire station ($R_{ep} = 85.99\,\text{km}$) in Landers, USA earthquake ($M_w = 7.3$), c) TCU036 station ($R_{ep} = 67.81\,\text{km}$) in Chi-Chi, Taiwan earthquake ($M_w = 7.6$) and d) TCU068 station ($R_{ep} = 47.86\,\text{km}$) in Chi-Chi, Taiwan earthquake.

Training of the CNN is done by using synthetic seismograms. The synthetics are not able to capture the long period impulses. However, CNN detects long period signals with high accuracy. The parameters are investigated in order to understand the relation between the prediction outputs and the pulse periods. Results can be seen in Fig 4.10.
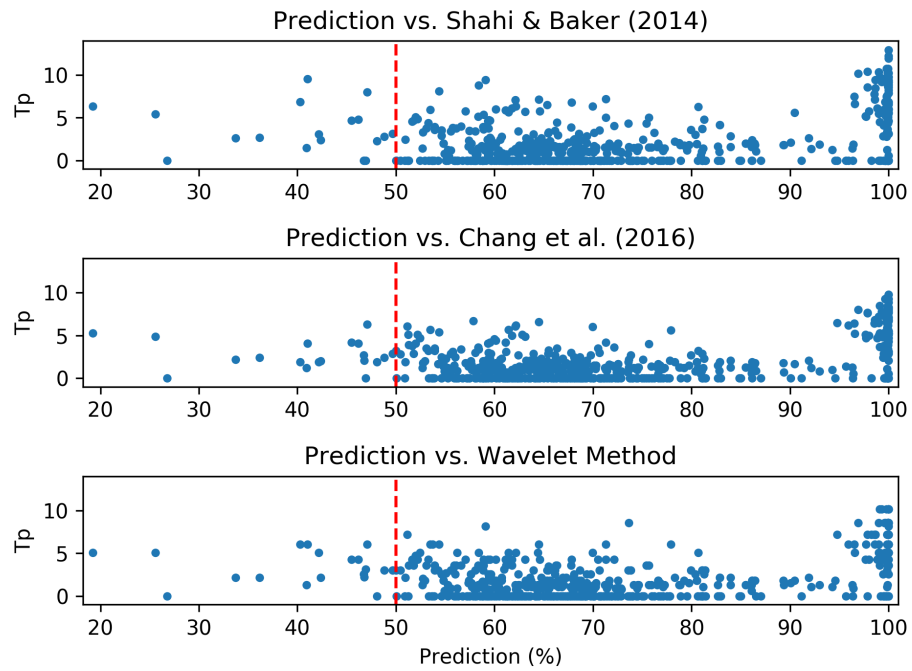
Figure 4.10: Pulse period (T_p) and the probability of impulsiveness of manually picked impulsive signals. non-impulsive signals determined by any of the studies have pulse period of 0 s.

CNN identified the signals that are not identified as impulsive by all 3 methods. Several incidences can be seen in Fig. 4.11. One can see that CNN can detect the impulsive signals with various pulse periods. Spectral response of these stations can be seen in Fig 4.12. It is clear that CNN can detect impulsive signals with very long period to very short period of T_p.
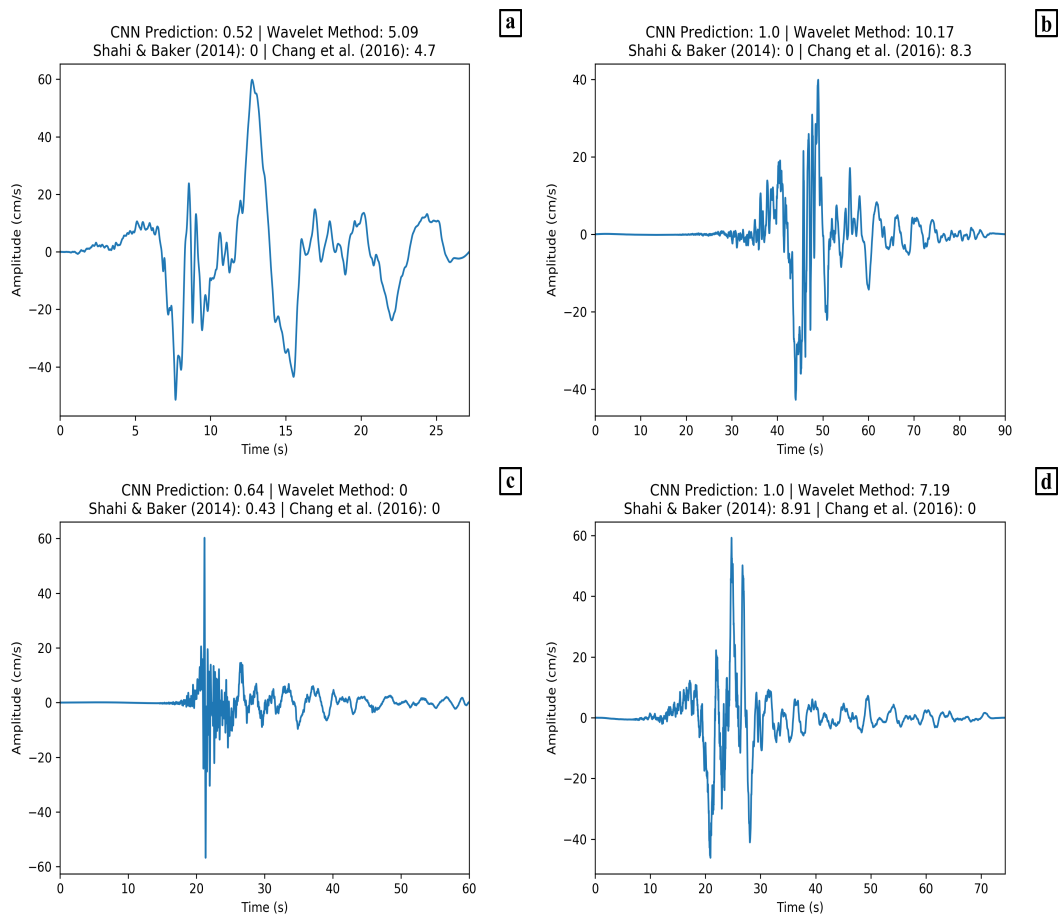
Figure 4.11: Velocity waveforms that are identified as impulsive by CNN method and at least one of the methods of Chang et al. (2016); Shahi and Baker (2014) and wavelet method. Chang et al. (2016); Shahi and Baker (2014) and wavelet method are the T_p of the signal calculated by the methods. a) Duzce station ($R_{ep} = 98.22$ km) in Izmit, Turkey earthquake ($M_w = 7.5$), b) TCU047 station ($R_{ep} = 86.39$ km) in Chi-Chi, Taiwan earthquake ($M_w = 7.6$), c) NIG021 station ($R_{ep} = 21.52$ km) in Nigata, Japan earthquake ($M_w = 6.6$) and d) LINC station ($R_{ep} = 33.76$ km) in Darfield, New Zealand earthquake ($M_w = 7.0$).
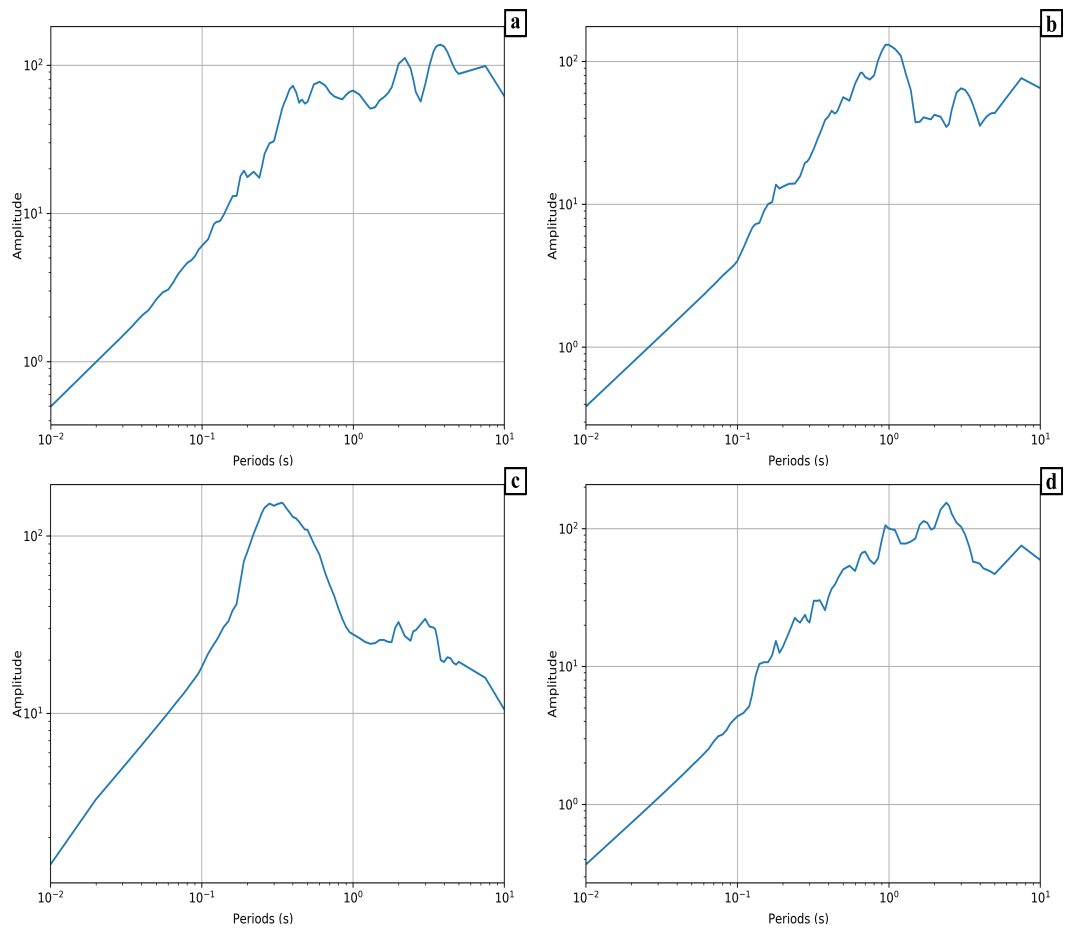
Figure 4.12: Spectral responses that are identified as impulsive by CNN method and at least one of the methods of Chang et al. (2016); Shahi and Baker (2014) and wavelet method. Chang et al. (2016); Shahi and Baker (2014) and wavelet method are the T_p of the signal calculated by the methods. a) Duzce station in Izmit, Turkey earthquake, b) TCU047 station in Chi-Chi, Taiwan earthquake, c) NIG021 station in Nigata, Japan earthquake and d) LINC station in Darfield, New Zealand earthquake.

Various velocity waveforms are not considered as impulsive by none of the previous studies or the wavelet method. These signals are identified as impulsive in manual selection, however most of these signals are considered as non-impulsive since the PGV value is less than the threshold. Shape of the signals are similar with the impulsive signals. CNN method is able to detect the impulsive-like signals with less amplitude. PGV threshold has been implemented to the studies since signals with less amplitude are considered as non destructive. However, they can well be destructive for the building constructed without proper engineering. Several examples can be seen in Fig. 4.13.
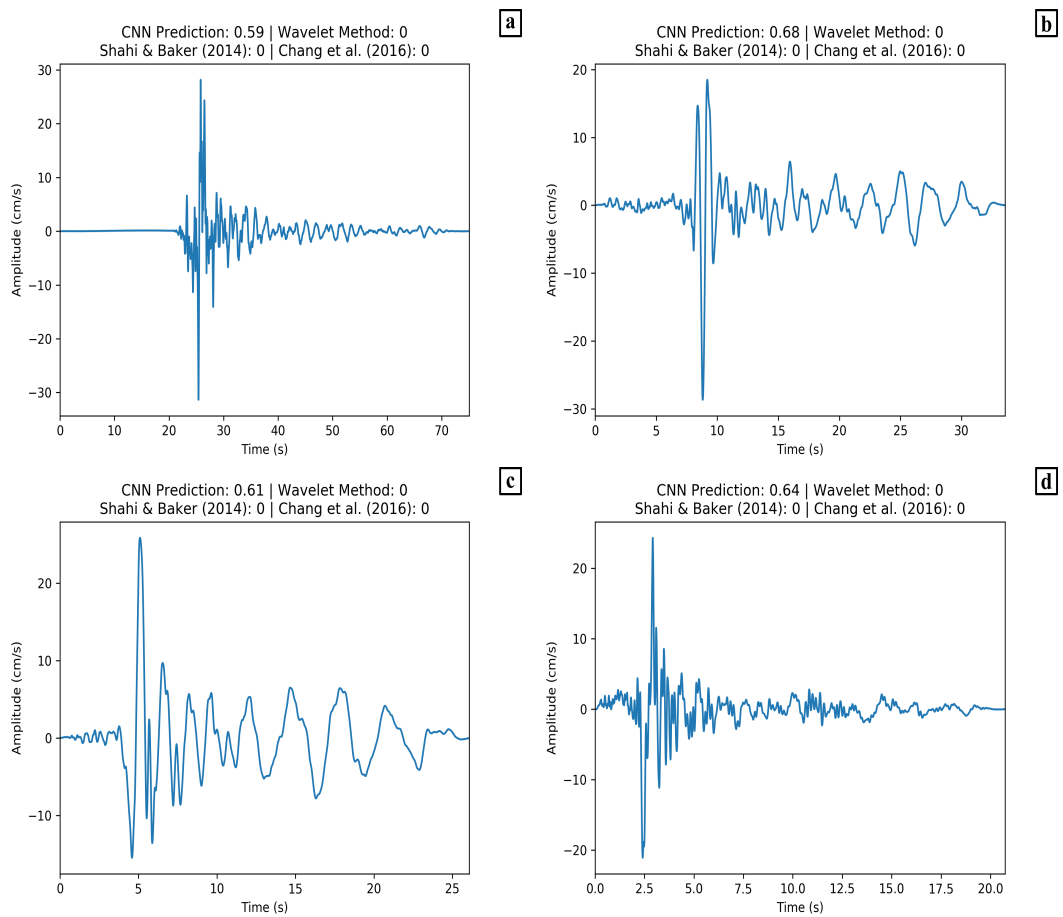
Figure 4.13: Velocity waveforms that are identified as impulsive by CNN method but not by Chang et al. (2016); Shahi and Baker (2014) or wavelet method. Chang et al. (2016); Shahi and Baker (2014) and wavelet method. a) TCU078 station ($R_{ep} = 17.94$ km) in Chi-Chi, Taiwan earthquake ($M_w = 6.3$), b) SMART1 M02 station ($R_{ep} = 69.19$ km) in Chi-Chi, Taiwan earthquake ($M_w = 6.3$), c) SMART1 O03 station ($R_{ep} = 69.30$ km) in Chi-Chi, Taiwan earthquake ($M_w = 6.3$) and d) Parkfield - Cholame 2E station ($R_{ep} = 12.06$ km) in Parkfield, USA earthquake ($M_w = 6.0$).

A part of the impulsive signals is not considered as impulsive, unlike the above mentioned methods (Fig. 4.14). Since the exact logic behind the CNN is hard to interpret by humans, it is hard to tell the reason of the false identification. One possible reason would be the lack of long period incidences on synthetics. Even though it captures several long period signals (Fig 4.9), it sometimes fails to detect one. An obvious method to improve it is to add long period examples. One idea would be to add all long period examples. However it may cause overfitting problem and even though CNN will predict the signals after the updated

model, one cannot know if it is due to the fact that it memorizes the training set (overfitting) or it actually learns the nature of long period signals. Another idea would be to add some of the impulsive signals, but due to lack of long period signals, CNN might not be able to learn by analyzing small amount of long period signals. It is possible to update the model by giving new incidences to update the model. Longer period signals should be given as an input in order to update the model.
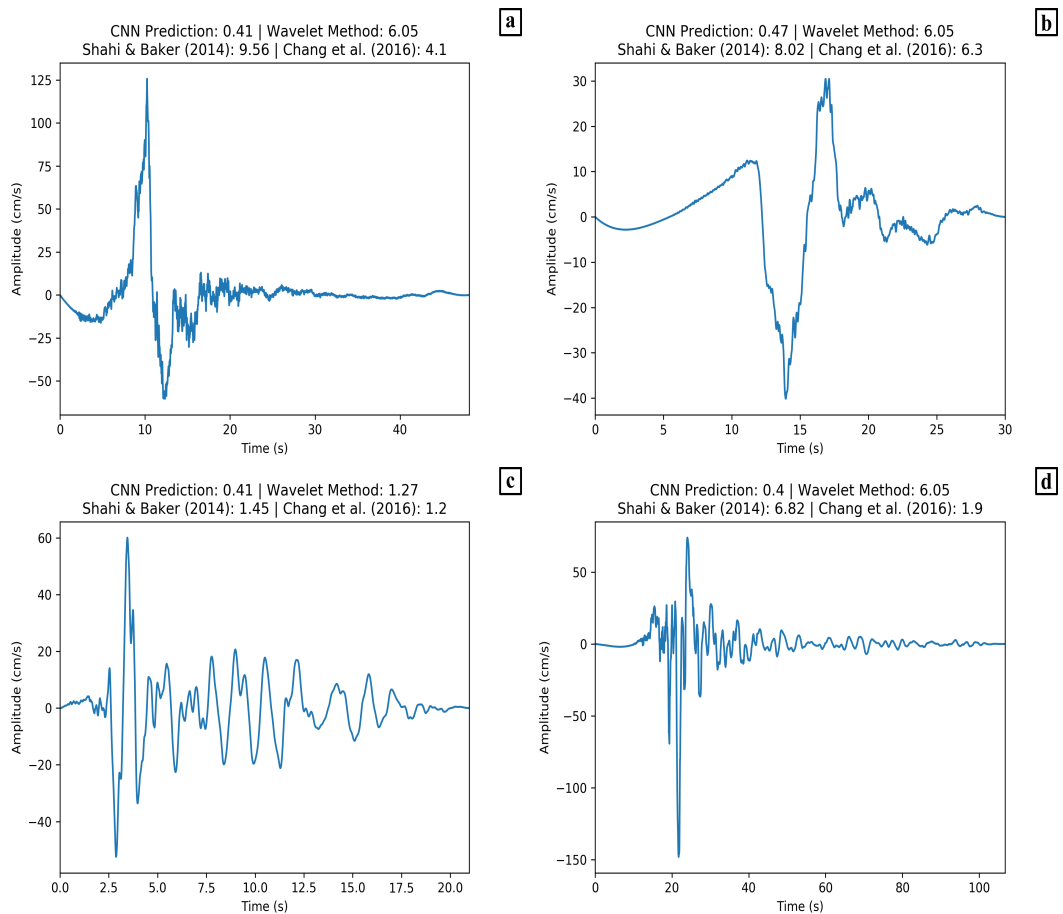
Figure 4.14: Velocity waveforms that are not identified as impulsive by CNN method but by Chang et al. (2016); Shahi and Baker (2014) or wavelet method. Chang et al. (2016); Shahi and Baker (2014) and wavelet method. a) Lucerne station ($R_{ep} = 44.02\,\text{km}$) in Landers, USA earthquake ($M_w = 7.3$), b) Arcelik station ($R_{ep} = 53.68\,\text{km}$) in Izmit, Turkey earthquake ($M_w = 7.5$), c) Parkfield - Fault Zone 1 station ($R_{ep} = 8.40\,\text{km}$) in Parkfield, USA earthquake ($M_w = 6.0$) and d) GDLC station ($R_{ep} = 4.42\,\text{km}$) in Darfield, New Zealand earthquake ($M_w = 7.0$).

CNN method has the highest accuracy rate. Chang et al. (2016); Shahi and Baker

(2014) and wavelet method also have high accuracy rates. Accuracy of these studies can be seen in Table 4.7. 526 out of 546 manually picked impulsive signals are labelled as impulsive by the CNN. Chang et al. (2016); Shahi and Baker (2014) and wavelet methods detect 382, 345 and 372 of the impulsive signals, respectively. However, PGV threshold that has been implemented to those studies is not a part of the manual labelling. In order to investigate the signals with large PGV values, only the signals with PGVs larger than $30\,\mathrm{cm\,s^{-1}}$ are considered. Results can be seen in Table 4.6.

Algorithm	No. of Impulsive Signal
Manual Picking	415
Shahi and Baker (2014)	329
Chang et al. (2016)	352
Wavelet Method	362
CNN Method	395

Table 4.6: Number of impulsive signals with PGV larger than $30\,\mathrm{cm\,s^{-1}}$ for various pulse classification algorithms.

CNN is good for detecting the signals where the amplitudes are less than the thresholds. Moreover, CNN model has learned the nature of the impulsive signals by analyzing vast amount of impulsive and non-impulsive signals. It can be used as a reliable impulsive signal identification tool with a high accuracy rate.

Algorithm	Accuracy Rate
Shahi and Baker (2014)	63.19%
Chang et al. (2016)	70.00%
Wavelet Method	68.13%
CNN Method	96.34%

Table 4.7: Accuracy rates of Chang et al. (2016); Shahi and Baker (2014), wavelet method, and CNN method. Manually picked impulsive signals are used as the ground truth.

14 out of 16 096 manually labelled real non-impulsive velocity waveforms are identified as impulsive by CNN method. Accuracy rates of Chang et al. (2016); Shahi and Baker (2014), wavelet method, and CNN method for non-impulsive signals are given in Table 4.8.

Algorithm	Accuracy Rate
Shahi and Baker (2014)	99.90%
Chang et al. (2016)	99.96%
Wavelet Method	99.80%
CNN Method	99.91%

Table 4.8: Accuracy rates for manually picked non-impulsive signals. Negative examples are taken from a single fold.

4.3. Pulse Probability Results

The focus of this section is calculating the probability of observing an impulsive signal. The probability is dependent on the fault mechanism, moment magnitude, R_{jb} and source-to-site azimuth. To calculate the probability distributions of impulsive signals, the dataset is divided into 3 different fault types: normal, reverse and strike slip faults. A Multivariate Naïve Bayes Classifier method is implemented on the dataset.

The distribution of impulsive and non-impulsive signals over the source-to-site azimuth for Chang et al. (2016); Shahi and Baker (2014) and for the wavelet method is shown in Table 4.9.

Normal & Normal-Oblique Fault						
Source-to-site	**Shahi & Baker (2014)**		**Chang et al. (2016)**		**Wavelet Method**	
Azimuth (^{o})	**No Pulse**	**Pulse**	**No Pulse**	**Pulse**	**No Pulse**	**Pulse**
[0 90)	679	3	678	4	679	3
90	374	11	374	11	372	13
(90 180]	818	2	818	2	818	2
(0 -90)	628	1	628	1	628	1
-90	311	15	308	18	308	18
(-90 -180]	617	2	617	2	618	1
Reverse & Reverse-Oblique Fault						
Source-to-site	**Shahi & Baker (2014)**		**Chang et al. (2016)**		**Wavelet Method**	
Azimuth	**No Pulse**	**Pulse**	**No Pulse**	**Pulse**	**No Pulse**	**Pulse**
[0 90)	2834	22	2825	31	2831	25
90	922	49	915	56	908	63
(90 180]	2104	8	2106	6	2103	9
(0 -90)	2000	68	1999	69	1988	80
-90	954	46	947	53	932	68
(-90 -180]	2420	12	2421	11	2413	19
Strike Slip Fault						
Source-to-site	**Shahi & Baker (2014)**		**Chang et al. (2016)**		**Wavelet Method**	
Azimuth (^{o})	**No Pulse**	**Pulse**	**No Pulse**	**Pulse**	**No Pulse**	**Pulse**
[0 90)	1541	21	1542	20	1538	24
90	628	41	623	46	616	53
(90 180]	1244	6	1242	8	1241	9
(0 -90)	1171	16	1171	16	1169	18
-90	522	43	519	46	517	48
(-90 -180]	1027	5	1026	6	1028	4

Table 4.9: Distribution of impulsive and non-impulsive signals with source-to-site azimuth categorization for normal and normal-oblique faulting, reverse and reverse-oblique faulting and strike slip faulting.

Previously published methods find a similar frequency of impulsive signals between different fault types and source-to-site azimuth angles (Chang et al., 2016; Shahi and Baker, 2014). Here, Shahi and Baker (2014) is chosen for the development of the model since it is well-known and a widely-used method for impulsive signal characterization. To develop the model, spatial variations of the impulsive signals are analyzed. The ratio between the impulsive and non-impulsive signals over the source-to-site azimuth angle and R_{jb} can be seen in Fig. 4.15.
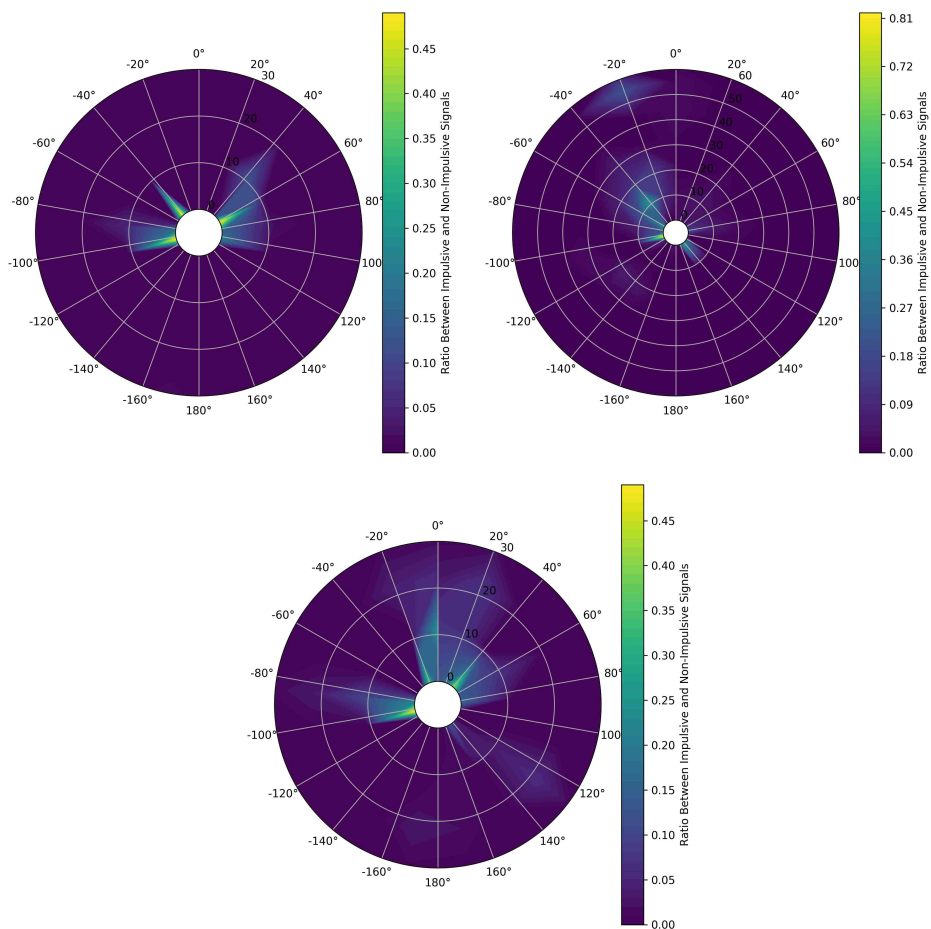
Figure 4.15: Ratio between impulsive and non-impulsive signals over source-to-site azimuth angles and R_{jb} distance for normal (upper left), reverse (upper right) and strike slip (down) faults. Source-to-site azimuth and R_{jb} are binned at 20° and $10\,$km. Impulsive signals are identified using Shahi and Baker (2014).

Fig. 4.15 shows a heterogeneous distribution of impulsive signals with respect to non-impulsive signals both by distance and by source-to-site azimuth angle. One flaw of the

dataset is the low ratio of impulsive to non-impulsive signals for all source-to-site azimuths. Here, the ratio is increased by removing the earthquakes that did not create any impulsive signals.

Normal faults have an equal frequency of impulsive signals on both sides of the fault plane where source-to-site azimuth $\neq \pm 90$. However, the number of impulsive signals on the foot wall and hanging wall can be different as explained by Scala et al. (2018). Because the difference is small for this dataset, however, this effect is not accounted in these probability calculations.

Rupture of strike slip faults can create a directivity effect at the extreme ends of the fault plane (Somerville et al., 1997). However, here the dataset does not contain information about the directivity effect. Therefore we use a conservative approach and assume that directivity effect can occur on both sides of the fault plane. Due to this, rupture direction has no effect on the probability distributions.

Reverse faults show relatively unevenly distributed impulsive signals. However most of the impulsive signals between source-to-site azimuths 0° - 90° are recorded from one event, the Chi-Chi Taiwan earthquake. In order to prevent a single event from dominating the dataset, fault plane information is also removed for normal and strike slip faults. Absolute values of source-to-site azimuth angles are used.

The spatial distribution of observing impulsive signals is investigated using a model with a fixed fault geometry. Contour maps indicate the probability of observing impulsive signals at different locations around the fault plane in this section. The probability distribution maps are plotted centered on the hypothetical fault line. A hypothetical epicenter is fixed at $(0,0)$ position on the maps. Fault plane geometry and station locations relative to the epicenter are given in Table 4.10. For strike slip faults there is no R_{jb} zone since dip angle is assumed to be 90°.

Fault Type	Normal	Reverse	Strike Slip
Length	50 km		
Width	30 km		
Source-to-site Azimuth	1 degree interval		
Epicentral Distance	0-100 km with 2 km interval		

Table 4.10: Fault plane and arbitrary station information.

Maps showing the probability distribution of observing impulsive signals are calculated for 5 different case study earthquakes with various fault types and M_w that have produced impulsive signals. These results are shown in Section 4.3.4.1 to Section 4.3.4.5.

Due to the limitations of the dataset (explained above) the probability distributions are not smooth throughout the source-to-site azimuth. The main reason for this is the heterogeneous distribution of seismic stations. The biggest effect of this is the lack of resolution when directivity effects generate the impulsive signals. Directivity can focus most of the seismic energy in the direction of the rupture. If the coverage of seismic installations is not dense at the edges of the ruptured fault (the case for most of the earthquakes) the spatial variability of the impulsive signals cannot be determined accurately. Furthermore, it is hard to predict the initial point and the direction of the earthquake rupture. This ambiguity can be seen in the probability distributions.

The dataset is affected by the heterogeneous distribution of stations, local soil conditions around the stations, the basin effect, the rupture propagation direction (directivity effect) and the rupture speed. These effects have a major impact on impulsive signals. Effect of the location of the station and the features of the earthquakes must be kept in mind throughout this section.

To overcome this problem, models are generalized by using only the most representative range of source-to-site azimuths (based on the factors discussed above). From here this is referred to as smoothing the probabilities. Smoothed maps of the probability distribution of observing impulsive signals are calculated with the same fault plane and station distribu-

tion. The results are presented in Section 4.3.5, Section 4.3.6 and Section 4.3.7. Results of the case study examples are presented in Section 4.3.8.1 and Section 4.3.8.2. In Section 4.3.9 the results are compared with probability distribution maps from the previous studies.

4.3.1. Normal & Normal-Oblique Faulting

The spatial distribution of impulsive and non-impulsive signals with respect to the source-to-site azimuth is presented in Table 4.11.

Source-to-site	Shahi & Baker (2014)	
Azimuth (^{o})	No Pulse	Pulse
[0 90)	524	4
90	370	26
(90 180]	424	4

Table 4.11: Distribution of impulsive and non-impulsive signals with source-to-site azimuth categorization for normal and normal-oblique faulting.

Probability distribution maps are created for moment magnitudes from $M_w = 6.0$ to $M_w = 7.3$ with increment 0.1. The uppermost interval corresponds to the largest magnitude earthquake in the dataset that has been observed to generate an impulsive signal. Results are shown in Fig. 4.16.
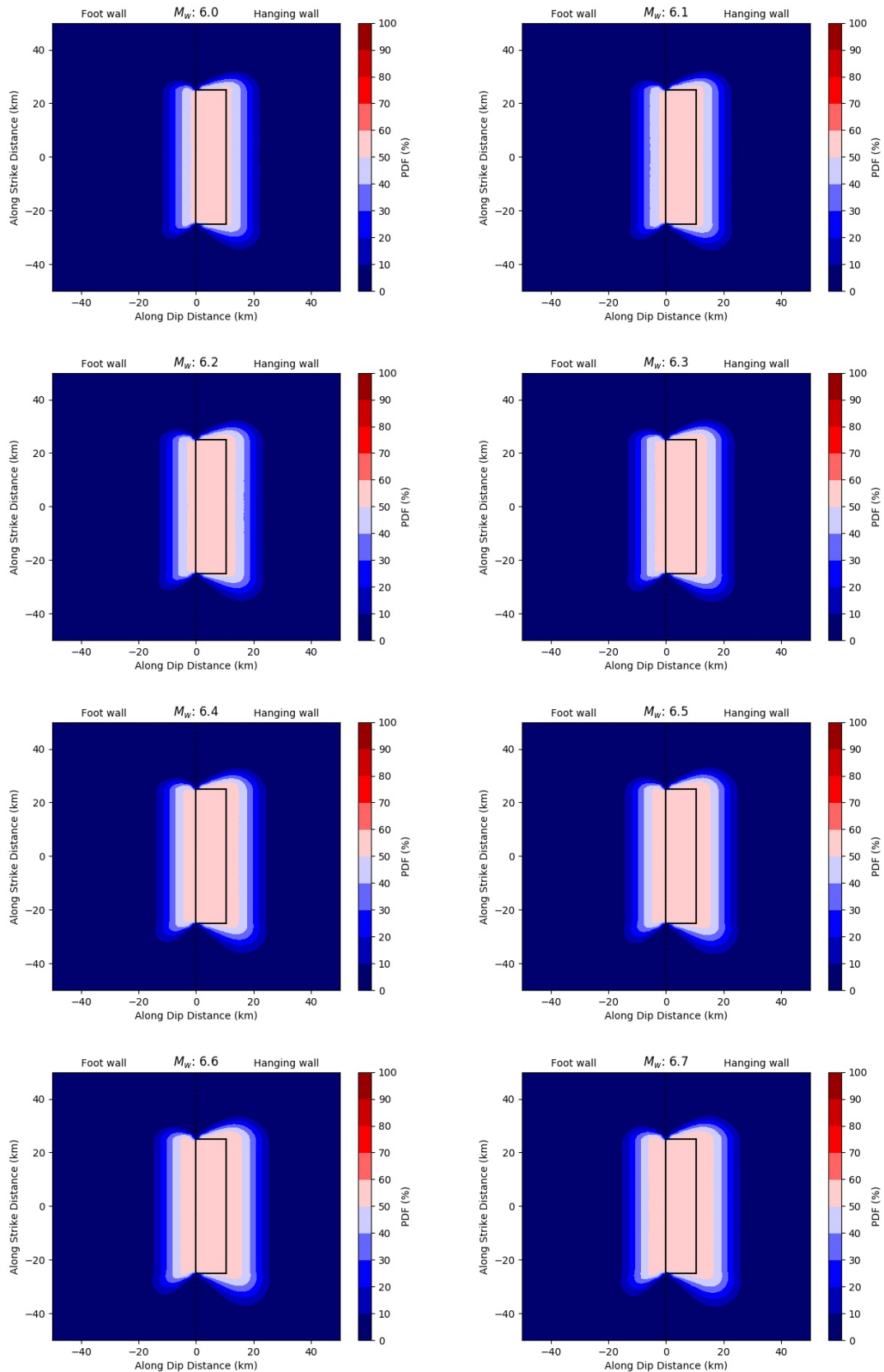
Figure 4.16: Probability distribution of observing impulsive signals with varying M_w for normal and normal oblique faults. Parameters are presented in Table 4.10. Foot wall and hanging wall are divided with vertical dots in this and upcoming probability distribution maps.
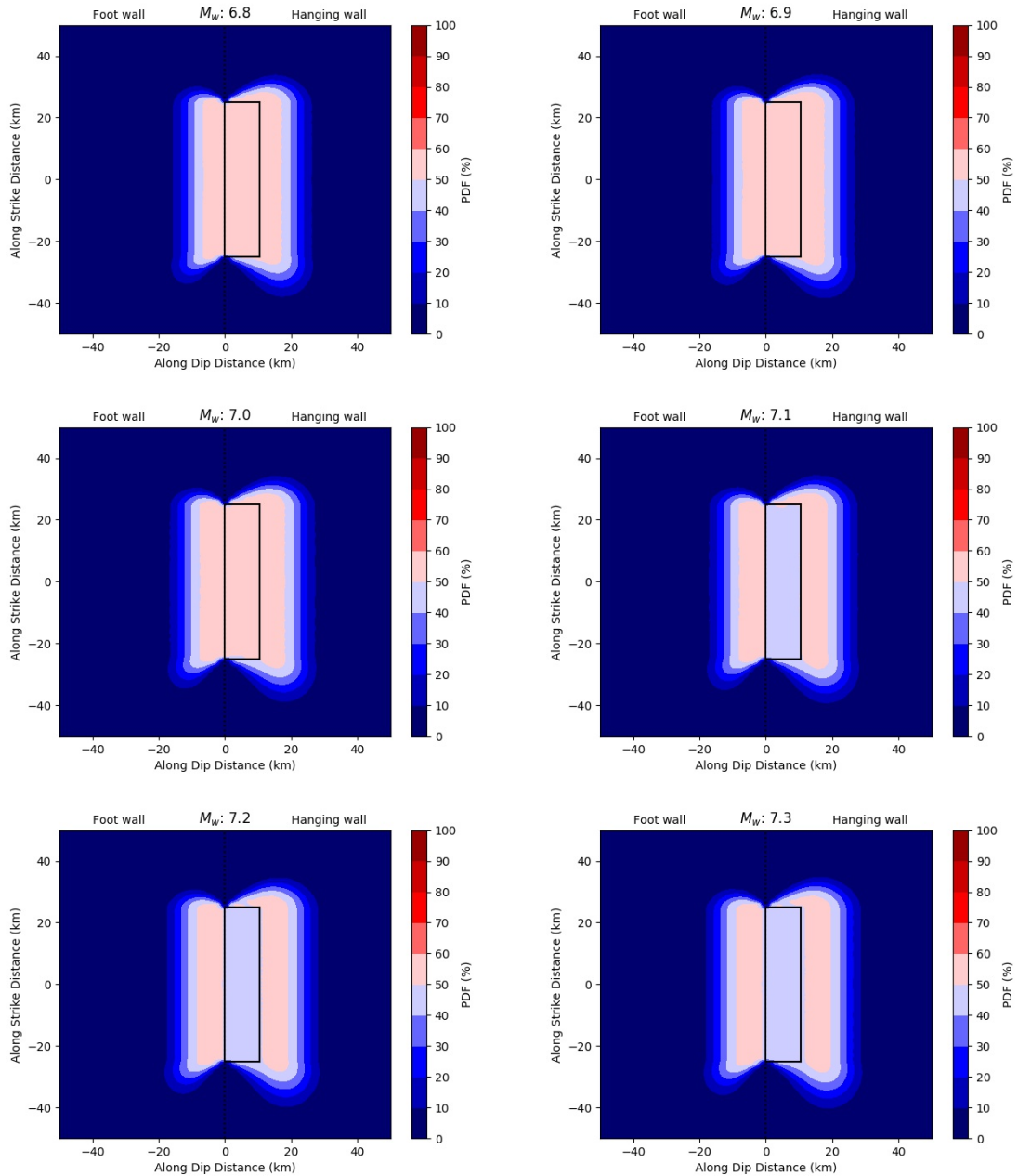
Figure 4.16: Probability distribution of observing impulsive signals with varying M_w for normal and normal oblique faults (cont.)

Most of the impulsive signals are along the source-to-site azimuth $\pm 90^{\circ}$ in the dataset. Because of that, there is not much variation of occurrence probability for source-to-site azimuth $\neq \pm 90^{\circ}$. The maximum probability does not increase with increasing moment magnitude but the area of the high probability zone increases. The high probability zone extends to the greater distances for source-to-site azimuths between 90° an 180°. This is due to a single

impulsive signal recorded at 30 km R_{jb} distance on a source-to-site azimuth of 178°. On the other side, source-to-site azimuth between 0° an 90°, the longest R_{jb} distance is only 13 km. Probabilities inside the R_{jb} zone start to decrease after $M_w \geq 7.0$. This is due to fact that there are no impulsive signals with moment magnitudes larger than 7.0 that occurred inside the R_{jb} area in the normal fault dataset.

4.3.2. Reverse & Reverse Oblique Faulting

The spatial distribution of impulsive and non-impulsive signals with respect to the source-to-site azimuth is shown in Table 4.12.

Source-to-site Azimuth (°)	Shahi & Baker (2014)	
	No Pulse	Pulse
[0 90)	2124	95
90	898	90
(90 180]	2083	20

Table 4.12: Distribution of impulsive and non-impulsive signals with source-to-site azimuth categorization for reverse and reverse-oblique faulting.

As in Section 4.3.1, probability distribution maps are generated for intervals of moment magnitude from $M_w = 6.0$ to $M_w = 7.6$ with increment 0.1. The upper limit on M_w is the reverse earthquake with the biggest magnitude (1999 Chi-Chi, Taiwan earthquake) that have created an impulsive signal. Generated maps can be seen in Fig. 4.17.
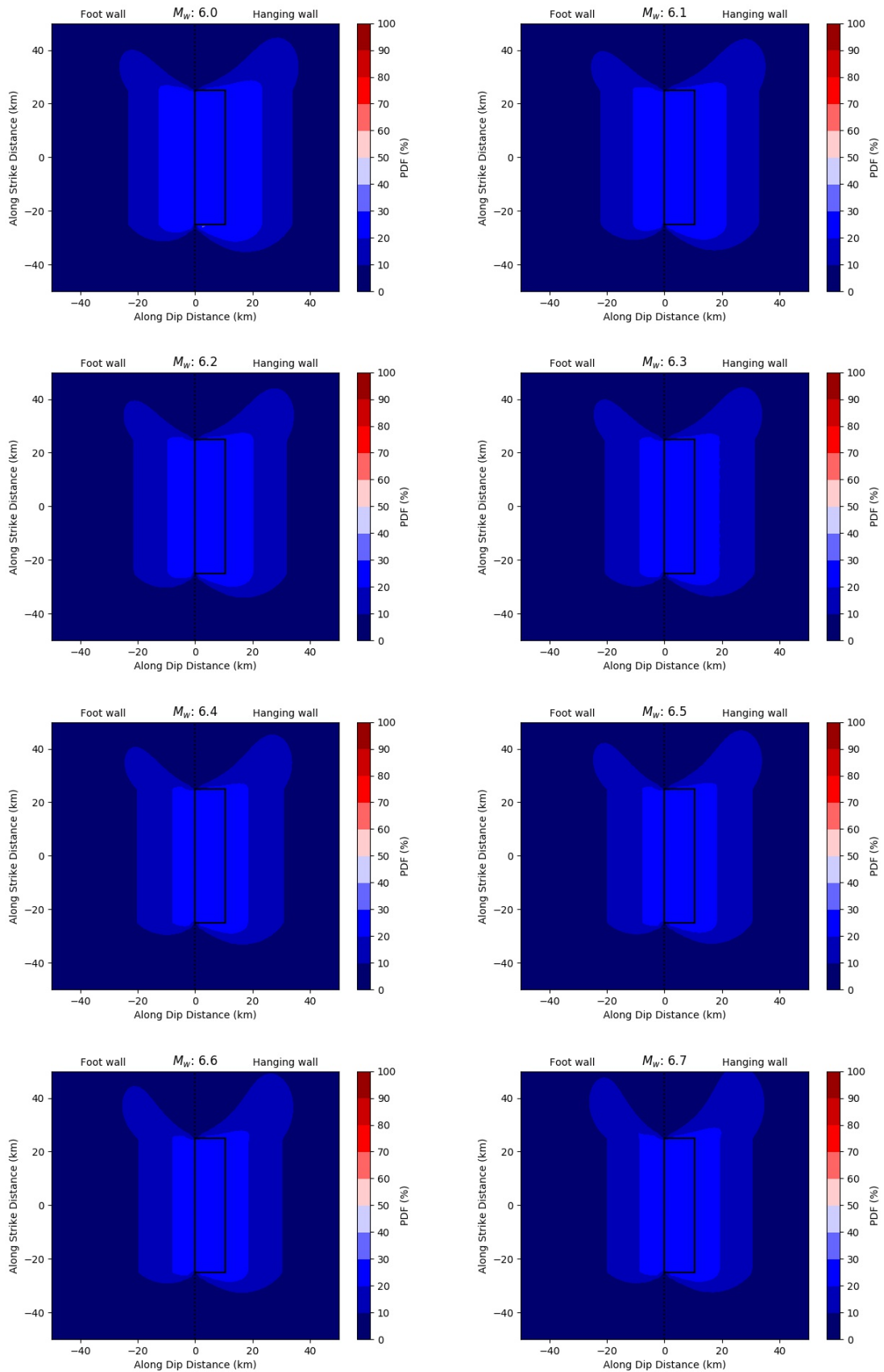
Figure 4.17: Probability distribution of observing impulsive signals with varying M_w for reverse and reverse oblique faults. Parameters are presented in Table 4.10.
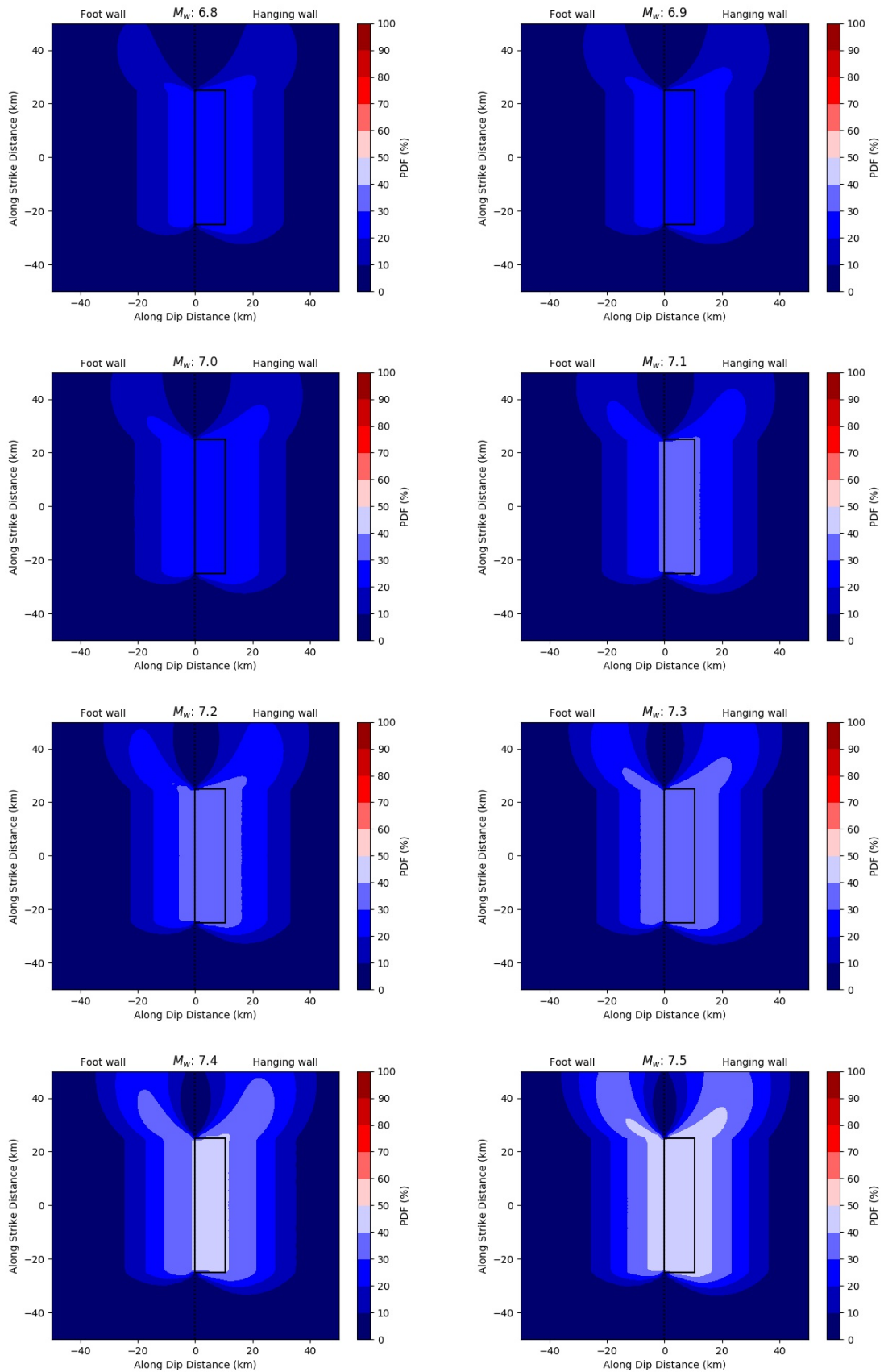
Figure 4.17: Probability distribution of observing impulsive signals with varying M_w for reverse and reverse oblique faults (cont.)
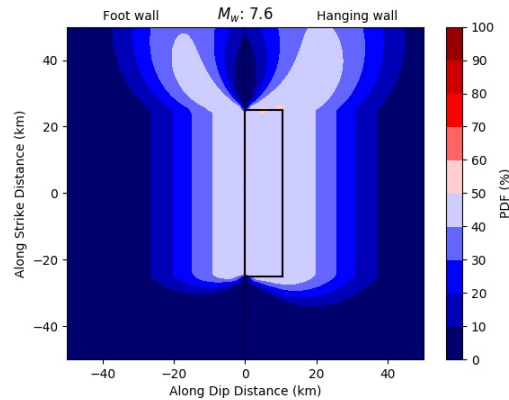
Figure 4.17: Probability distribution of impulsive signals with varying M_w for reverse and reverse oblique faults (cont.)

Probability distribution maps of observing impulsive signals for reverse and reverse oblique faulting have more complex features than normal faulting. An unequal distribution of impulsive signals with respect to M_w can be seen in Fig.4.17. Probabilities are decreasing between moment magnitudes 6.0 and 6.7. Between these magnitudes, the relatively high probability zone is shrinking towards the R_{jb} zone with increased magnitude. Starting from M_w 6.8 high probability zone is spreading away from the R_{jb} zone. There is a significant difference between the source-to-site azimuths 0° to 90° and 90° to 180°. For source-to-site azimuths 0° to 90°, further distances have higher probabilities of observing impulsive signals. This is caused by the large amount of impulsive signals (N =26) recorded by the SMART1 network during the 20th of May 1986 earthquake in Taiwan (M_w of 6.3). The SMART1 network is located on a soft soil site with a Vs_{30} around $270 \, \mathrm{m \, s^{-1}}$. Soil amplification is the main source of the impulsive signal on that location. Impulsive signals recorded by the SMART1 network dominate the source-to-site azimuth between 0° and 90°.

There are various impulsive signals for source-to-site azimuths between 90° and 180° in the dataset. However they are mostly generated by earthquakes with magnitude between 6.0 and 6.5, or the 1999 Chi-Chi Taiwan earthquake ($M_w = 7.6$). Moreover the main source of the higher probabilities of observing impulsive signals in magnitudes between 6.0 and 6.5 is from the SMART1 network.

Two types of directivity effect can affect the waveform: up-dip directivity and along strike directivity. Both of these directivity effects are present in the Chi-Chi Taiwan earthquake (Aagaard et al., 2004). Higher probabilities at source-to-site azimuth between 0° and 90° at large moment magnitudes for reverse and reverse-oblique faults can be linked to the data coming from the Chi-Chi Taiwan earthquake.

4.3.3. Strike Slip Faulting

The spatial distribution of impulsive and non-impulsive signals with respect to the source-to-site azimuth can be seen in Table 4.13.

Source-to-site Azimuth (^{o})	Shahi & Baker (2014)	
	No Pulse	Pulse
[0 90)	1063	39
90	652	82
(90 180]	615	11

Table 4.13: Distribution of impulsive and non-impulsive signals with source-to-site azimuth categorization for strike slip faulting.

Probability distribution maps of observing impulsive signals for moment magnitudes 6.0 to 7.9 (0.1 interval) are presented in Fig. 4.18.

Figure 4.18: Probability distribution of observing impulsive signals with varying M_w for strike slip faults. Parameters are presented in Table 4.10.
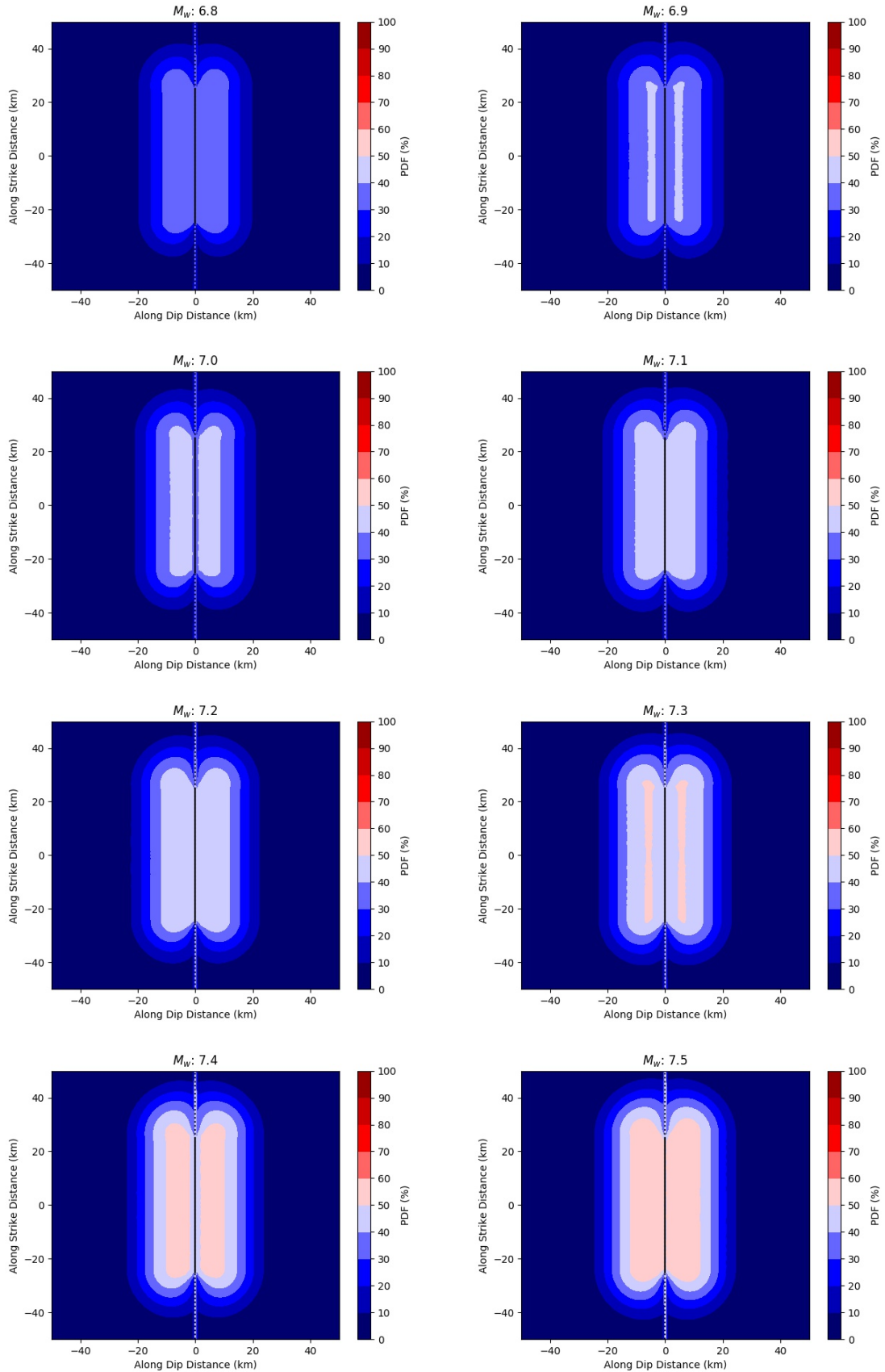
Figure 4.18: Probability distribution of observing impulsive signals with varying M_w for strike slip faults (cont.)
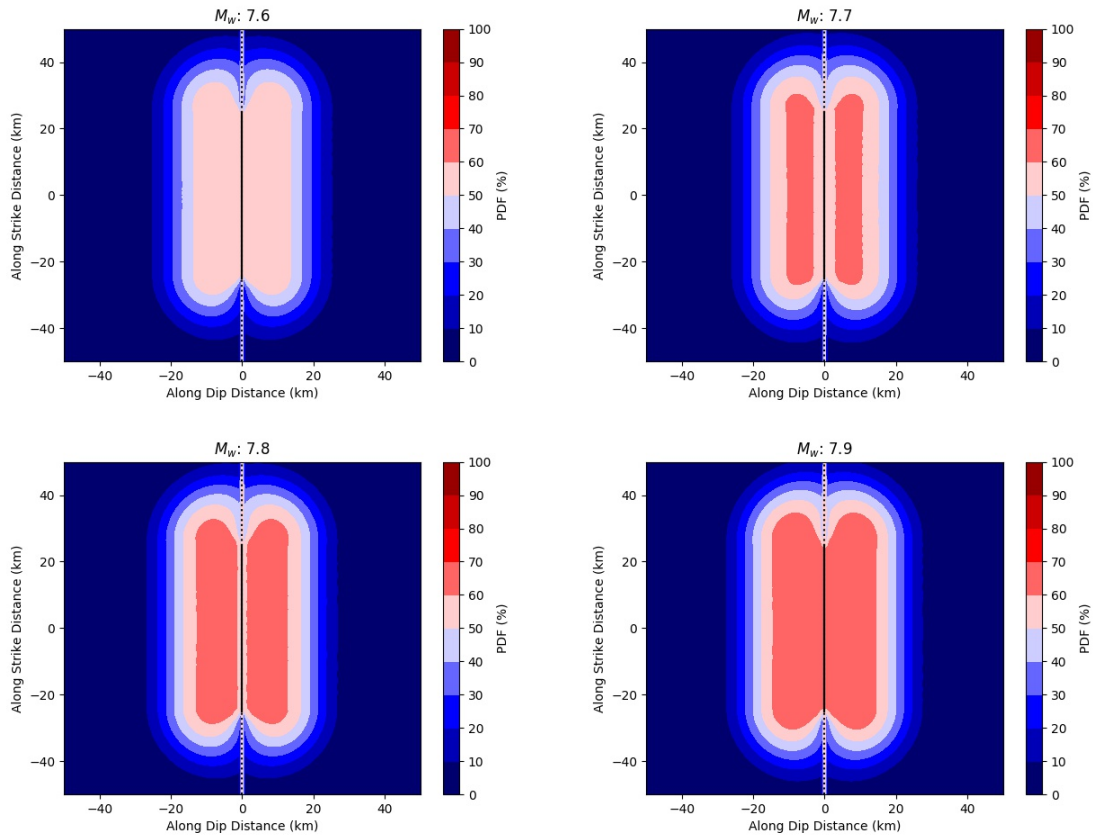
Figure 4.18: Probability distribution of observing impulsive signals with varying M_w for strike slip faults (cont.)

The probability of an impulsive signal being generated by a strike-slip fault increases smoothly with magnitude, distance and source-to-site azimuth angle. Source-to-site azimuths between 0° and 90° have higher probability of observing impulsive signals at slightly further distances with respect to the azimuths greater than 90°. The ratio between the impulsive and non-impulsive signals is also higher in that region. Hence, probability of observing impulsive signals on source-to-site azimuth between 0° and 90° represents the nature of the impulsive signals better. The directivity effect can be seen as a higher probability "butterfly wings" effect at extreme ends of the fault line. The tips of the "butterfly wings" are the directivity effect.

4.3.4. Probability Distributions on Real Earthquakes

The generated models are implemented on case studies which are presented in this section. 5 case study earthquakes are selected. Information about the earthquakes can be seen in Table 4.14. Probability distribution maps created by using the method explained in Section 3.3 can be seen from Section 4.3.4.1 to Section 4.3.4.5.

Earthquake Name	M_w	Fault Type	Source-to-site Azimuth (°)											
			[0 90)		90		(90 180]		[-0 -90)		-90		(-90 -180]	
			Pulse	No Pulse	Pulse	No Pulse	Pulse	No Pulse	Pulse	No Pulse	Pulse	No Pulse	Pulse	No Pulse
2016 Kumamoto, Japan	7.3	Normal - Oblique	0	48	3	99	2	31	0	0	5	94	1	32
1999 Chi-Chi, Taiwan	7.6	Reverse - Oblique	2	126	0	54	0	105	0	0	6	57	1	80
2018 Hokkaido Eastern Iburi, Japan	6.7	Reverse	2	12	4	4	2	71	0	33	0	16	0	54
1999 Izmit, Turkey	7.4	Strike Slip	0	21	1	5	1	5	0	0	0	9	0	0
2010 Canterbury, New Zealand	7.0	Strike Slip	4	2	2	4	0	0	0	0	8	100	0	24

Table 4.14: Earthquake information and impulsive and non-impulsive signals distribution over source-to-site azimuth.

4.3.4.1. 2016 Kumamoto, Japan Earthquake. The probability distribution map of observing impulsive signals for the Kumamoto, Japan earthquake is presented in Fig. 4.19. The focus is the main shock of Kumamoto earthquake sequence (16th of April 2016, 1:25 in Japan standard time). The earthquake has oblique-normal faulting with a high strike slip component. One can see that 2 of the impulsive signals are lying in the regions with 0 % probability. There are also non-impulsive signals in the high probability areas. The probability of observing impulsive signals is between 50 % and 60 % for 15 km distance normal to the ruptured fault. However, there are 4 impulsive and non-impulsive signals in that region. Other signals are located in 0 % probability regions.

In Kumamoto earthquake, there is a strong rupture propagation effect to the northeast of the fault plane (Kobayashi et al., 2017). Impulsive signals that has occurred to the northeast of the fault plane can be explained by the rupture propagation direction.
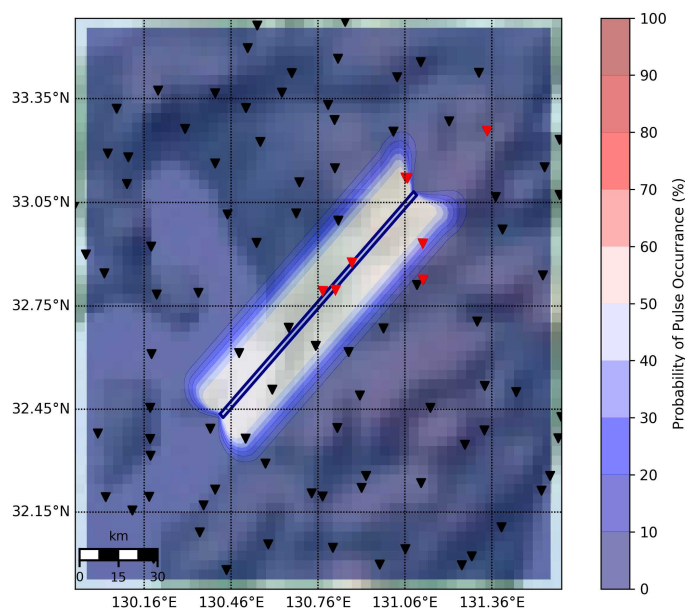
Figure 4.19: Probability distribution of observing impulsive signals on Kumamoto earthquake ($M_w = 7.3$). Legend of the probabilities are on the right side of the figure. Ruptured fault is indicated with blue rectangle. Impulsive and non-impulsive signals are represented by red and black reverse triangles, respectively.

4.3.4.2. 1999 Chi-Chi, Taiwan Earthquake. The probability distribution map of observing impulsive signals for the Chi-Chi, Taiwan earthquake can be seen in Fig. 4.20. Aagaard et al. (2004) observe that there is an up-dip directivity effect that created impulsive signals that are located in the foot wall part of the R_{jb} area. A fling step effect is also recorded (Kalkan and Kunnath, 2006).

Most of the impulsive signals are located in the foot wall part and they are located in the region with high probabilities (50 % - 60 %). 3 impulsive signals lie in the hanging wall part with source-to-site azimuth close to zero. They are also in the region with relatively high probabilities.

Even though there are plenty of impulsive signals observed in the foot wall part, there are also a large number of non-impulsive signals. In fact, the ratio between the impulsive and non-impulsive signals in foot wall part where source-to-site azimuth is -90°, is lower than 50 %.
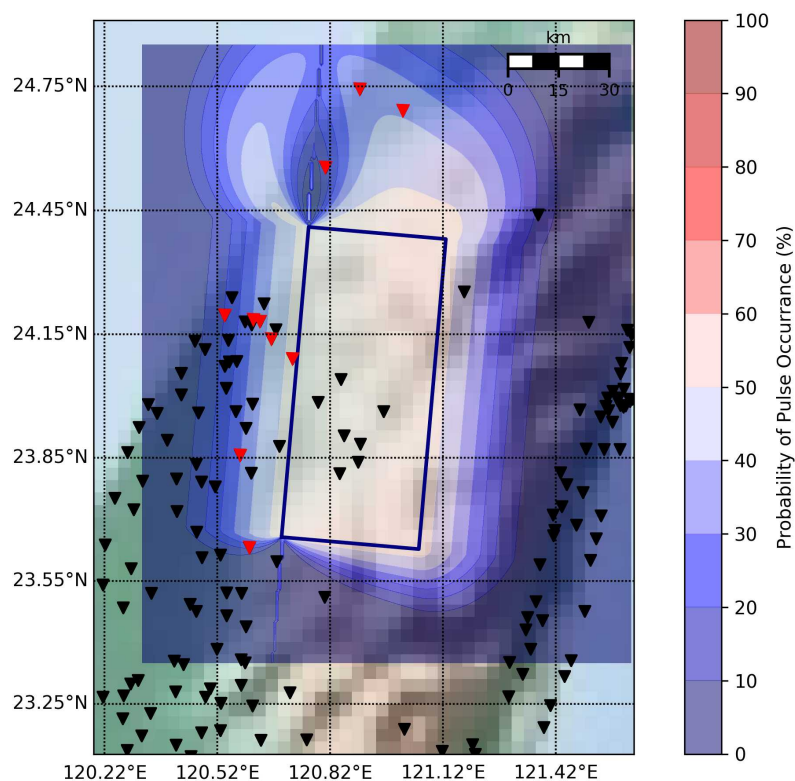
Figure 4.20: Probability distribution of observing impulsive signals on Chi-Chi earthquake. Features on the map are the same as in Fig. 4.19. Fault plane information is taken from Zeng and Chen (2001).

4.3.4.3. 2018 Hokkaido Eastern Iburi, Japan Earthquake. The probability distribution map of observing impulsive signals for the Hokkaido, Japan earthquake can be seen in Fig. 4.21. There are 3 stations that have recorded impulsive signals during the earthquake (approximately 20 km and 10 km away from the R_{jb} area). At this distance, there are an almost equal number of non-impulsive signals. Hence, the probability distribution of impulsive signals is underestimated for this earthquake. An up-dip rupture occurs on the earthquake (Asano and Iwata, 2019; Kobayashi et al., 2019). Both Asano and Iwata (2019) and Kobayashi et al. (2019) found similar rupture kinematics by using different fault plane information. In both studies, the maximum upward slip is in the Southwest direction. The slip propagated towards the 2 impulsive signals that are located to the West of the ruptured fault. Dhakal et al. (2019) connects the impulsive behavior of the HKD126 station to its position, which is in the direction of up-dip propagation. Kobayashi et al. (2019) also states that the pulse occurrence at the HKD126 station is to the Southwest of the ruptured fault. In that study, however, the

occurrence of the pulse for that station is connected with the weak soil condition. EC8 soil class is registered as type D in the dataset.

Figure 4.21: Probability distribution of observing impulsive signals on Hokkaido Eastern Iburi earthquake. Features on the map are the same as in Fig. 4.21.

4.3.4.4. 1999 Izmit, Turkey Earthquake. The probability distribution map of observing impulsive signals for the Izmit, Turkey earthquake is shown in Fig. 4.22. The Izmit earthquake has a moment magnitude of 7.6 with a fast rupture speed of ($3.50 \, \text{km s}^{-1}$) and showed directivity effects in the region (Gomberg et al., 2003). A fling step is also recorded in various displacement waveforms (Bouchon et al., 2002).

There are two stations that recorded impulsive signals for this earthquake. One of these

stations is inside the area with high probabilities whereas the other is not. There is a station located very close to the ruptured fault, yet it does not have impulsive features. Probabilities are estimated correctly normal to the ruptured fault direction. However, an impulsive signal lies outside the high probability area along strike direction to the east of the fault.

Figure 4.22: Probability distribution of observing impulsive signals on Izmit earthquake. Features on the map are the same as in Fig. 4.19. Fault plane information is taken from Bouchon et al. (2002).

4.3.4.5. 2010 Canterbury, New Zealand Earthquake.   The probability distribution map of observing impulsive signals for Canterbury, New Zealand earthquake is shown in Fig. 4.23. Probabilities in the areas very near the fault ($R_{jb} \leq 3 \, \text{km}$) normal to the ruptured fault are underestimated. There are only 2 impulsive signals in the very near fault ($R_{jb} \leq 3 \, \text{km}$). However, the probability of observing impulsive signals normal to the ruptured fault is between 30 % and 40 %. Probabilities, surprisingly, increased between 4 km and 15 km normal to the ruptured fault. Within these distances there are two impulsive and non-impulsive signals. Even in that case, the probabilities are less than 50 %.
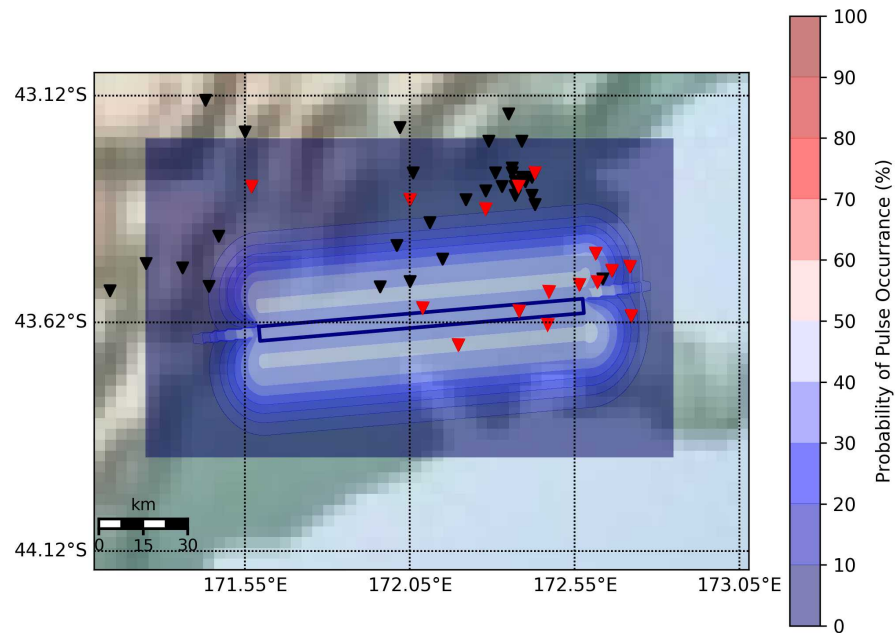
Figure 4.23: Probability distribution of observing impulsive signals for the Canterbury earthquake. Features on the map are the same as in Fig. 4.19. Fault plane information is taken from Mai and Thingbaijam (2014).

Several impulsive signals are located to the east of the fault plane. The main source of these impulses are the eastward rupture propagation and the forward directivity effect (Bradley, 2012). The directivity effect causes a "butterfly wings" effect in the probability distribution map (Section 4.3.3). Bradley (2012) associates impulsive signals on the north east of the fault plane to the basin effect. Impulsive signals can be observed at R_{jb} distance greater than 30 km.

4.3.5. Smoothed Normal & Normal-Oblique Faulting

Smoothed probability distribution maps of observing impulsive signals for moment magnitudes 6.0 to 7.3 (0.1 interval). Results are shown in Fig. 4.24.
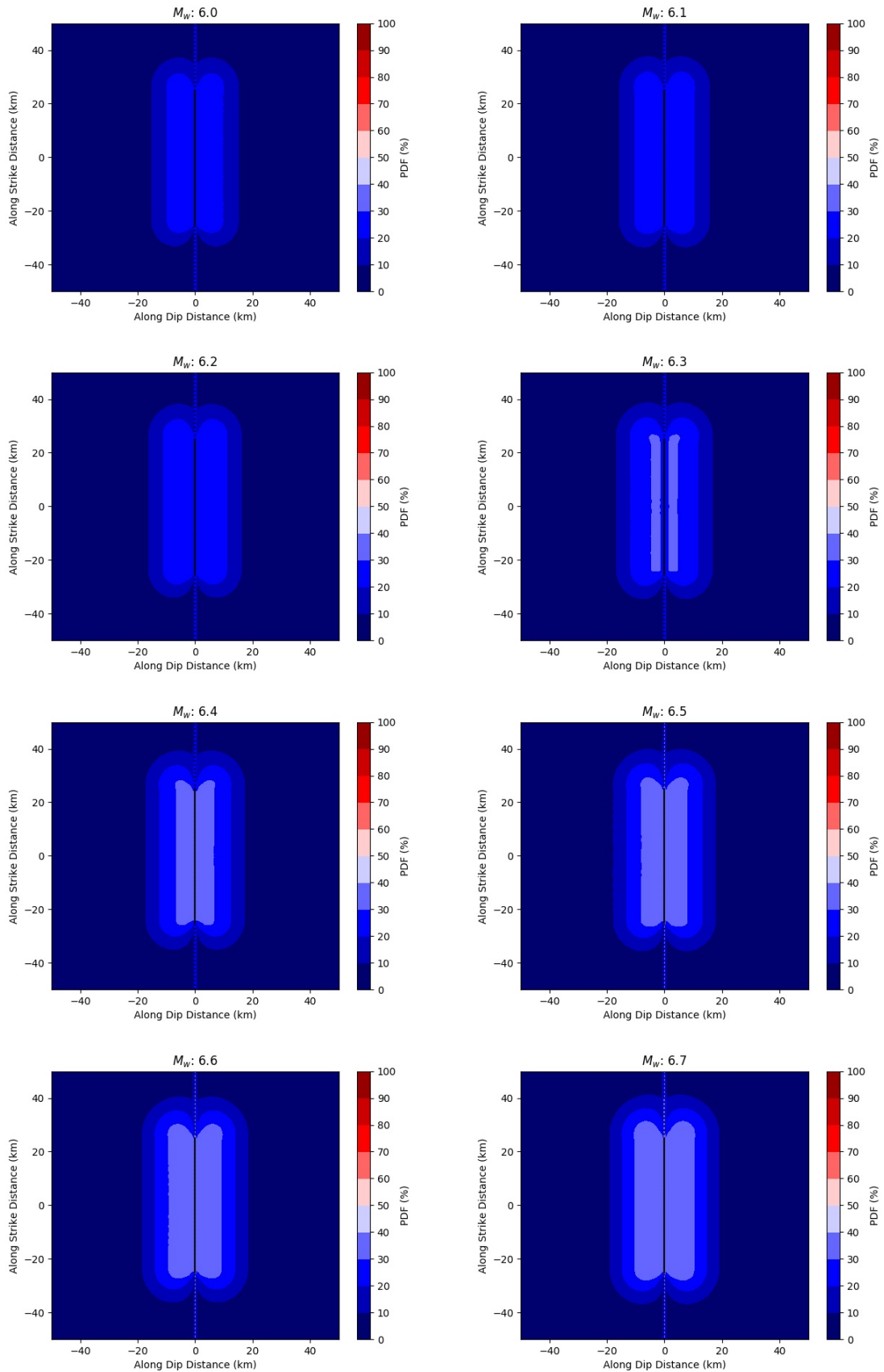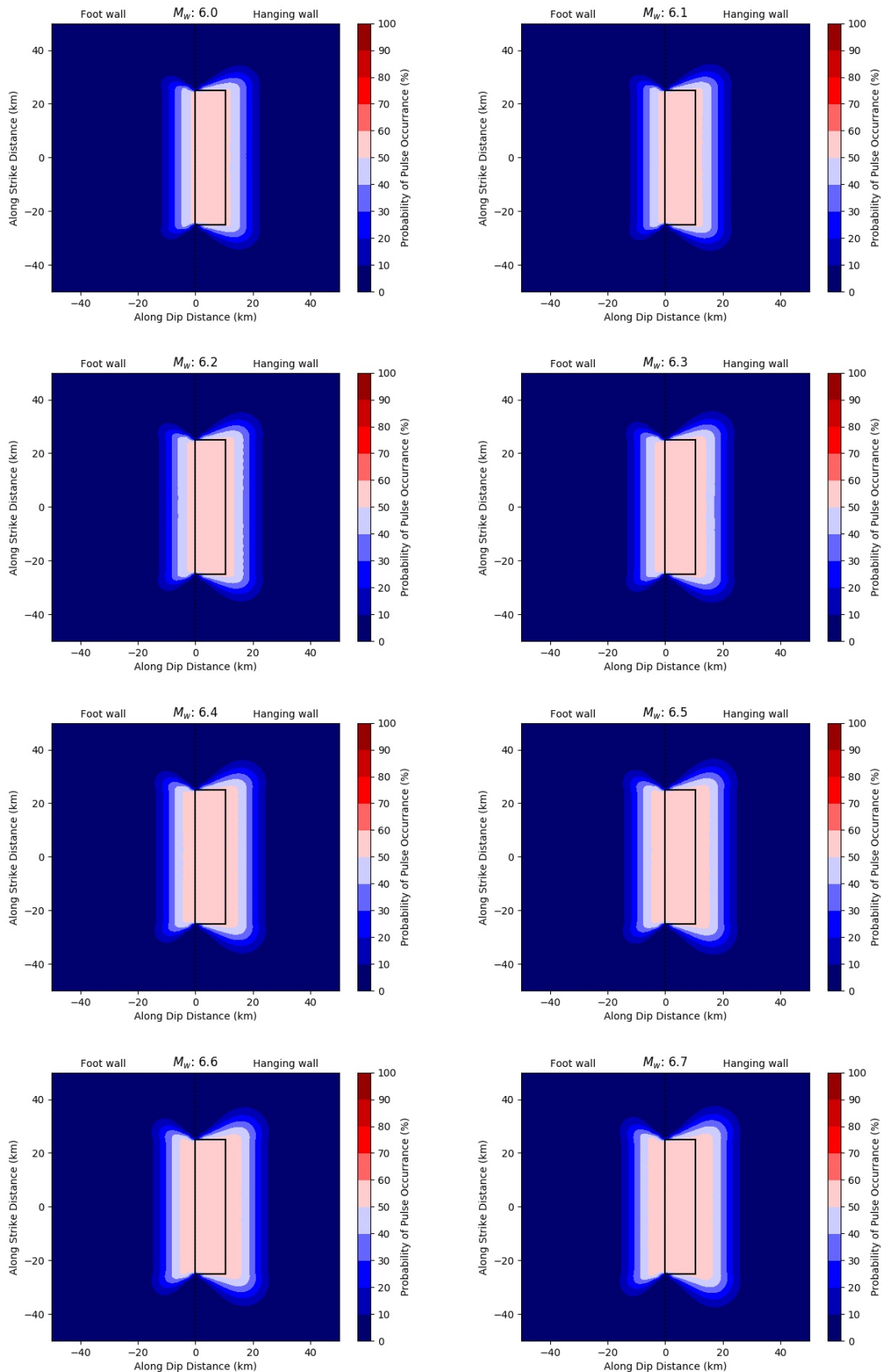
Figure 4.24: Smoothed probability distribution of observing impulsive signals with varying M_w for normal and normal oblique faults. Parameters are same as in Table 4.10.
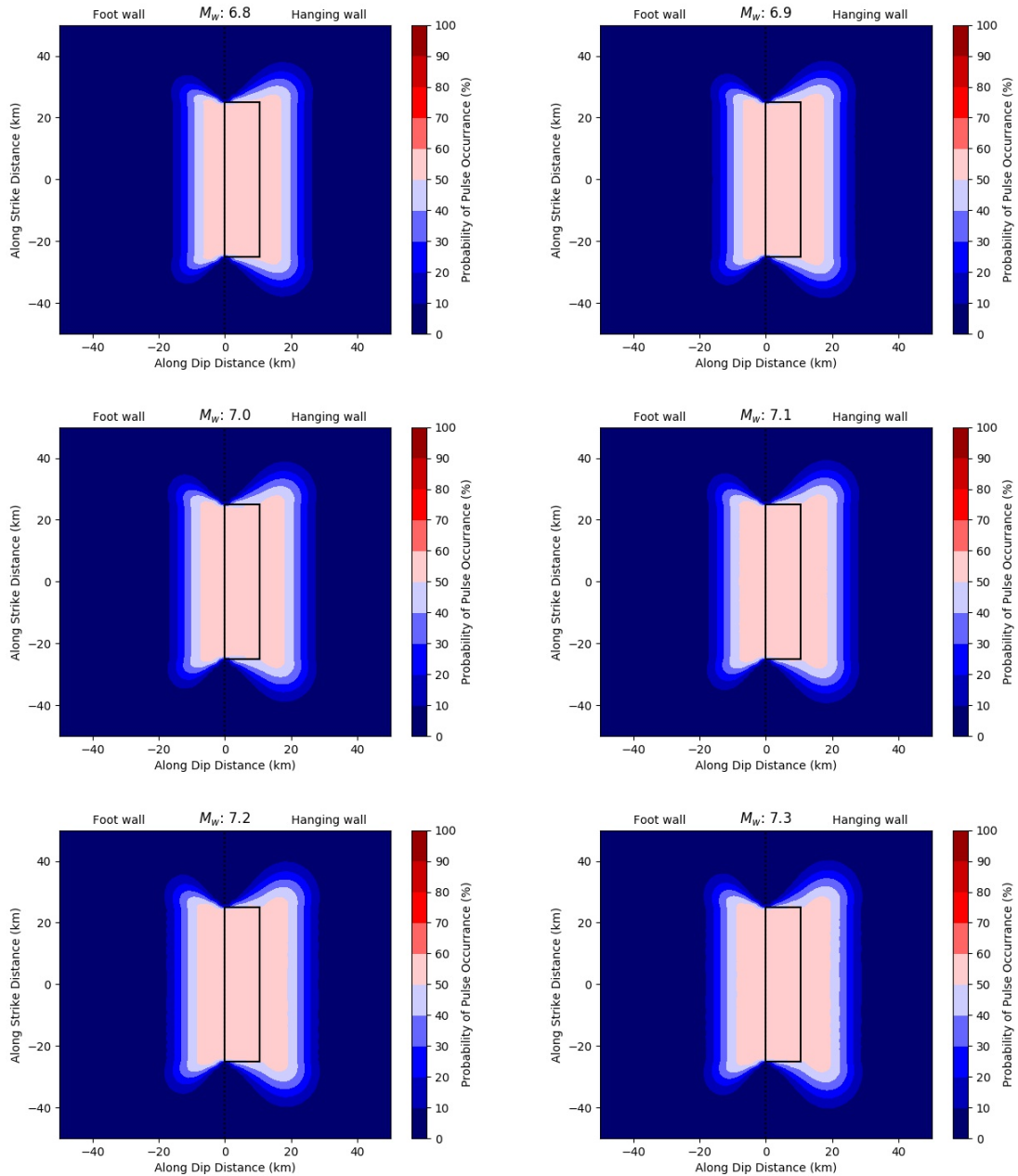
Figure 4.24: Smoothed probability distribution of observing impulsive signals with varying M_w for normal and normal oblique faults (cont.)

A smoothing procedure has been applied in order to create probability maps that are not affected by the unbalanced distribution of impulsive signals recorded in the near-fault regions. For normal faults, source-to-site azimuths from 90º to 180º are chosen for the smoothing. The ratio between impulsive and non-impulsive signals (Table 4.10) is considered during the creation of these maps. Lower probabilities in the R_{jb} area due to the lack of

impulsive signals in that region in bigger magnitude earthquakes ($M_w \geq 7.0$) are corrected by assigning the maximum probability of occurrence value to the R_{jb} zone.

4.3.6. Smoothed Reverse & Reverse-Oblique Faulting

Smoothed probability distribution maps of observing impulsive signals for moment magnitudes 6.0 to 7.6 (0.1 interval). Results are shown in Fig. 4.25.
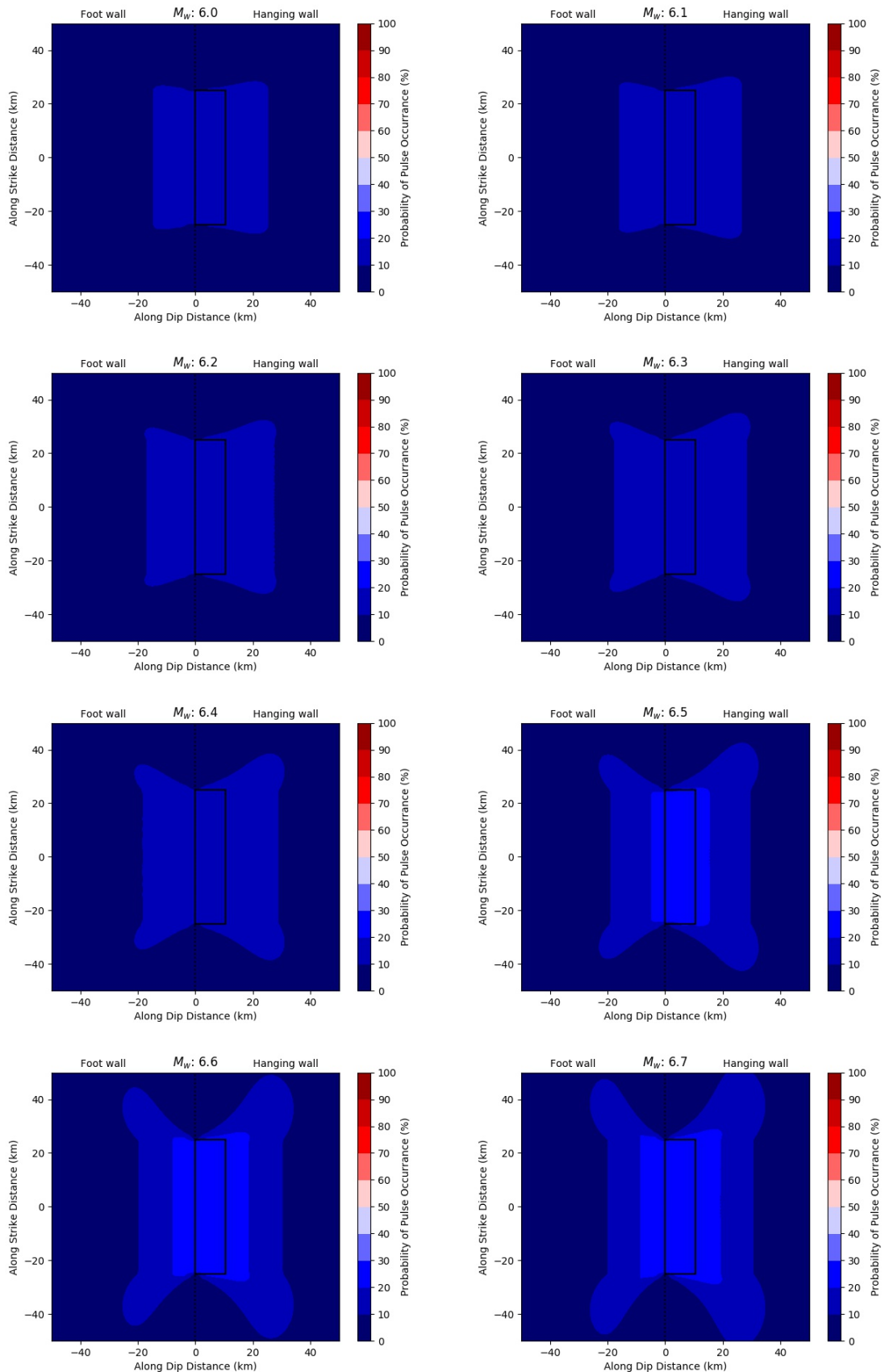
Figure 4.25: Smoothed probability distribution of observing impulsive signals with varying M_w for reverse and reverse oblique faults. Parameters are same as in Table 4.10.
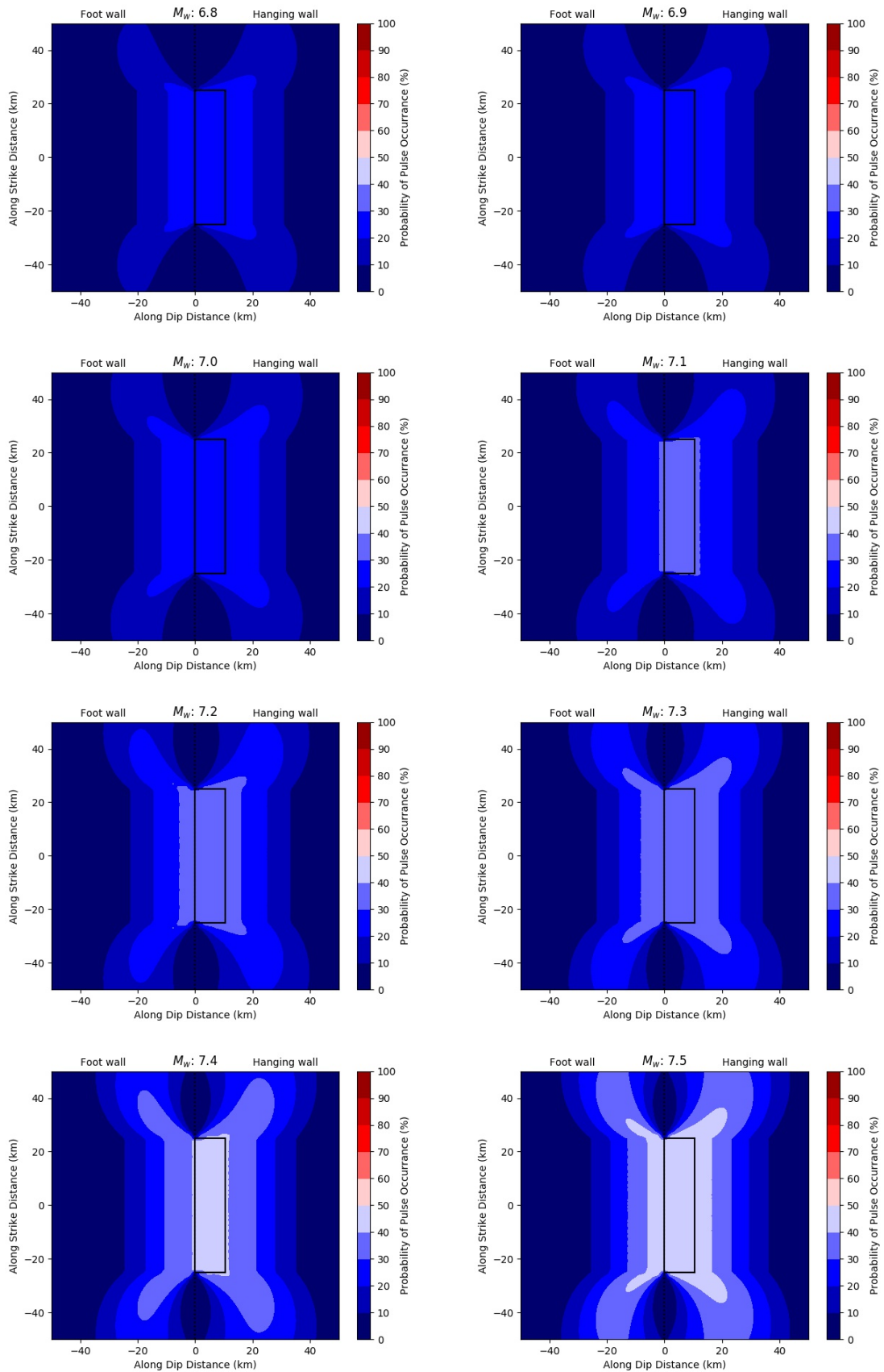
Figure 4.25: Smoothed probability distribution of observing impulsive signals with varying M_w for reverse and reverse oblique faults (cont.)
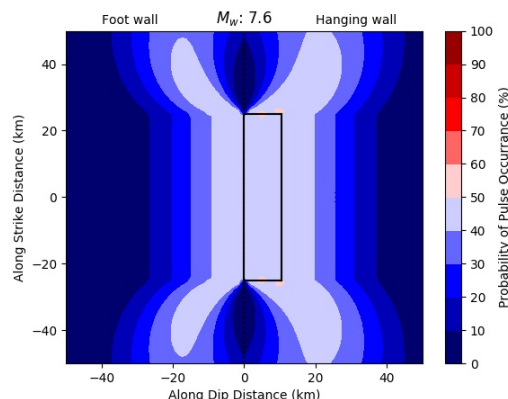
Figure 4.25: Smoothed probability distribution of impulsive signals with varying M_w for reverse and reverse oblique faults (cont.)

A major revision in reverse and reverse-oblique faulting is done for the magnitudes from 6.0 to 6.5. Due to the effects that are explained in Section 4.3.2, probabilities are not increasing with the increasing M_w. It created probability distribution maps of observing impulsive signals, in which the high probability area is shrinking with increasing moment magnitude. In order to overcome this problem, a first order line is fitted to probabilities on 0.1 interval of M_w from 6.6 to 7.2 for 1° of interval of source-to-site azimuth and 1 km interval of R_{jb} distance. Probability distributions for moment magnitudes between 6.0 and 6.5 are calculated using the fitted lines. After applying the smoothing procedure, a smoothly increasing probability of observing pulse occurrence with moment magnitude is guaranteed.

Since the smoothing process does not affect the earthquakes that have been used as examples for reverse and reverse-oblique faults in Section 4.3.4.2 and Section 4.3.4.3, smoothed versions of these earthquakes are not presented.

4.3.7. Smoothed Strike Slip Faulting

Smoothed probability distribution maps of observing impulsive signals for moment magnitudes 6.0 to 7.9 (0.1 interval). Results are shown in Fig. 4.26.
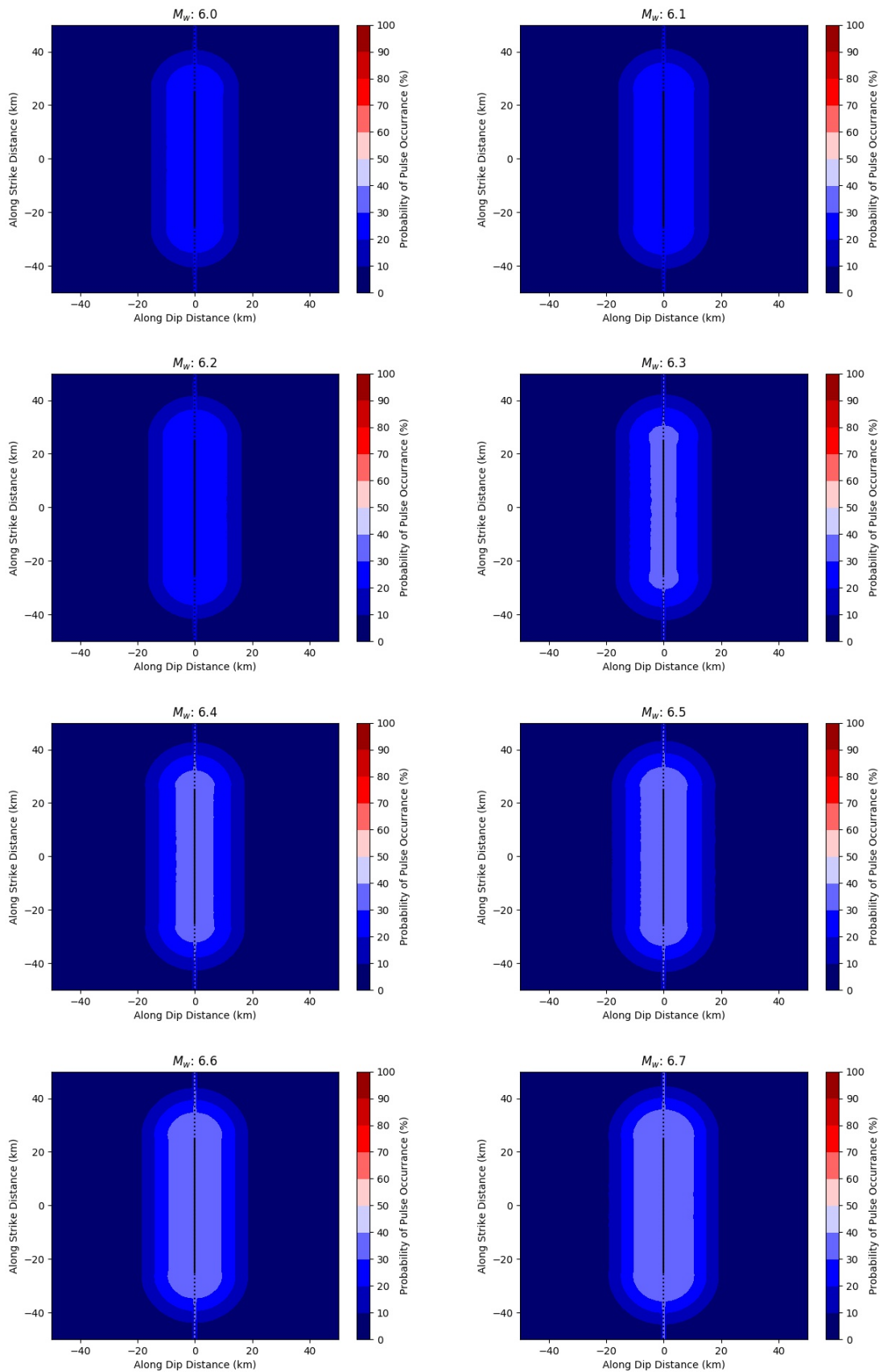
Figure 4.26: Smoothed probability distribution of observing impulsive signals with varying M_w for strike slip fault. Parameters are same as in Table 4.10.
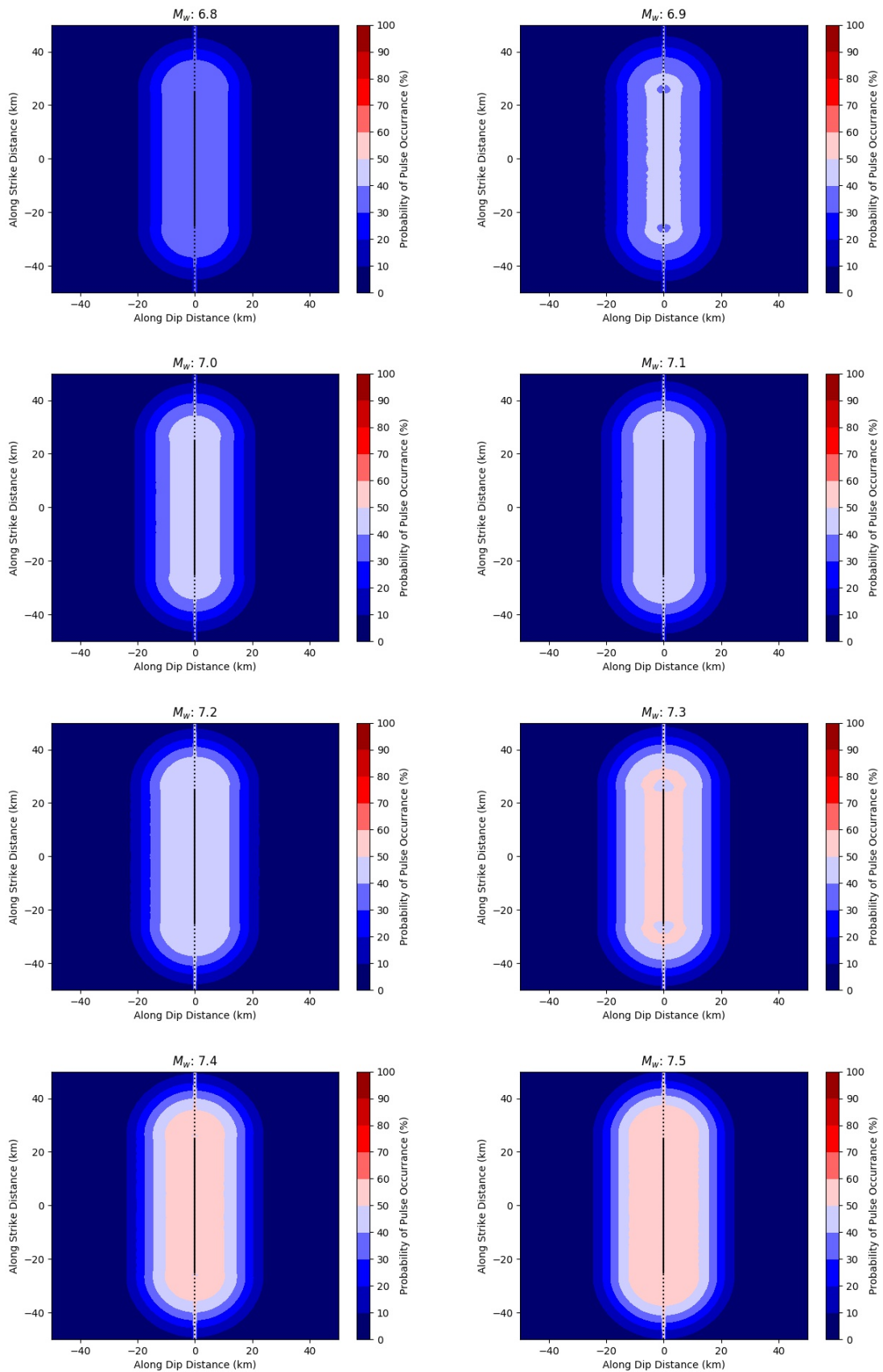
Figure 4.26: Smoothed probability distribution of observing impulsive signals with varying M_w for strike slip fault (cont.)
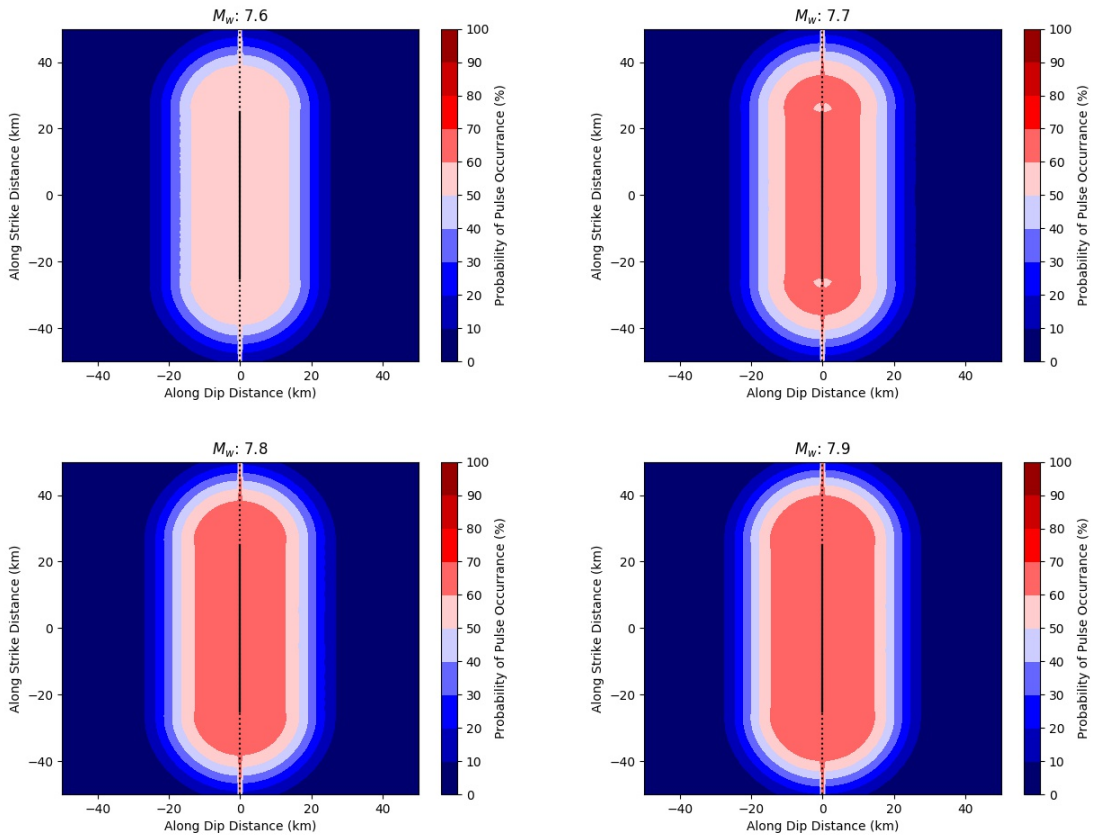
Figure 4.26: Smoothed probability distribution of observing impulsive signals with varying M_w for strike slip fault (cont.)

To create the smoothed probability distributions of observing impulsive signals for strike slip faults, the area with source-to-site azimuth from 0° to 90° is chosen for the locations, where source-to-site azimuths are not 90°. This part has the highest impulsive to non-impulsive ratio in the dataset. The lack of impulsive signals for source-to-site azimuths less than 10° and larger than 170° in the dataset, which results in the probability of observing impulsive signals decreasing rapidly. In order to remove this artefact, the maximum probability along the source-to-site azimuth for the same epicentral distance in the parameterization (Table 4.10) is assigned to very low ($\leq 10^o$) and very large ($\geq 170^o$) source-to-site azimuths. This creates a more realistic geometry for the high probabilities of observing impulsive signals area at the top and bottom edges of the fault. Gaps in the edges of moment magnitudes 6.9 and 7.3 are due to the binning of the probabilities. In these areas, probabilities are extremely close to the upper bin (the one that dominates the surroundings of the gaps). The main purpose of the smoothing in strike slip faults is the proper representation of the higher

probabilities for very close distances normal to the fault rupture. Relatively lower probabilities are present very close to the ruptured faults. The lower probabilities are replaced with the highest probability value for the source-to-site azimuth = 90°. The improved, smoothed probability distributions are shown in Section 4.3.8.2.

4.3.8. Smoothed Probability Distributions on Real Earthquakes
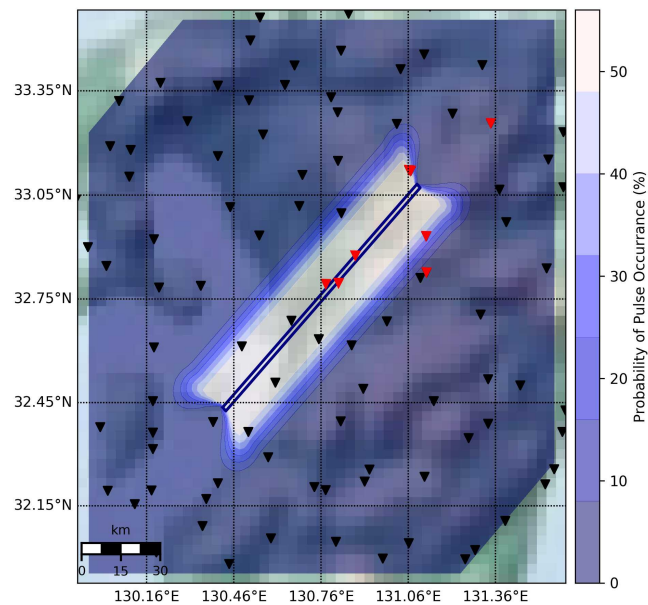
Figure 4.27: Smoothed probability distribution of observing impulsive signals for Kumamoto earthquake. Features on the map are the same as in Fig. 4.19.

4.3.8.1. 2016 Kumamoto, Japan Earthquake.   Unlike the Fig. 4.19, the probability of observing impulsive signals is higher than 10 % until 22.5 km to the Southwest of the ruptured fault. Although there is no improvement in the correctly identified impulsive signals for this earthquake, the probability distribution map in Fig. 4.27 is more realistic with respect to the map in Fig. 4.19.
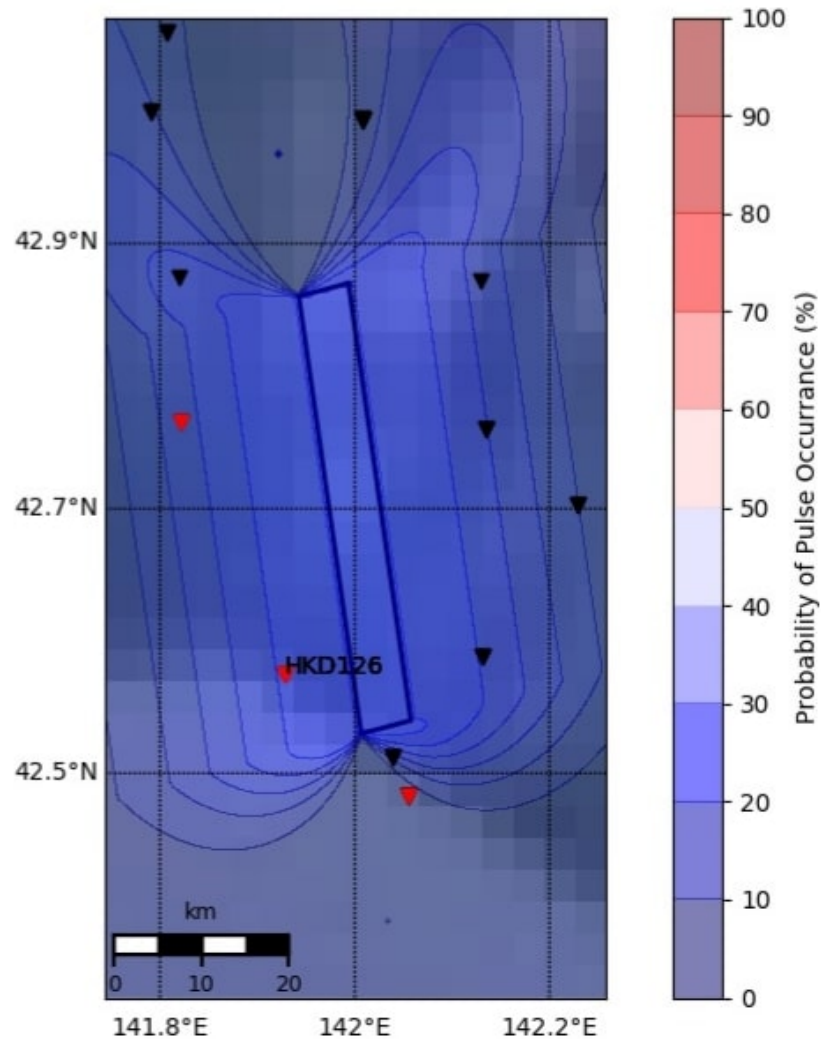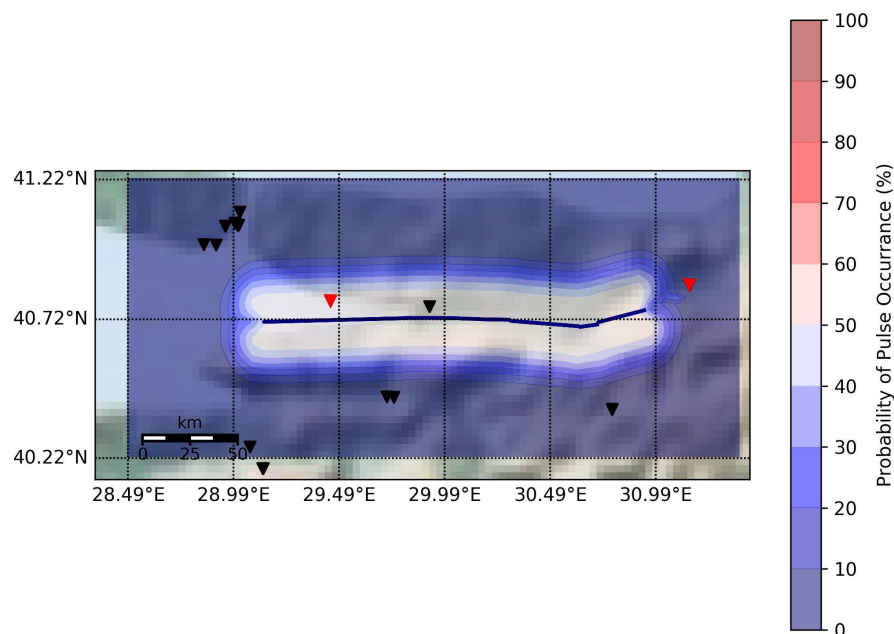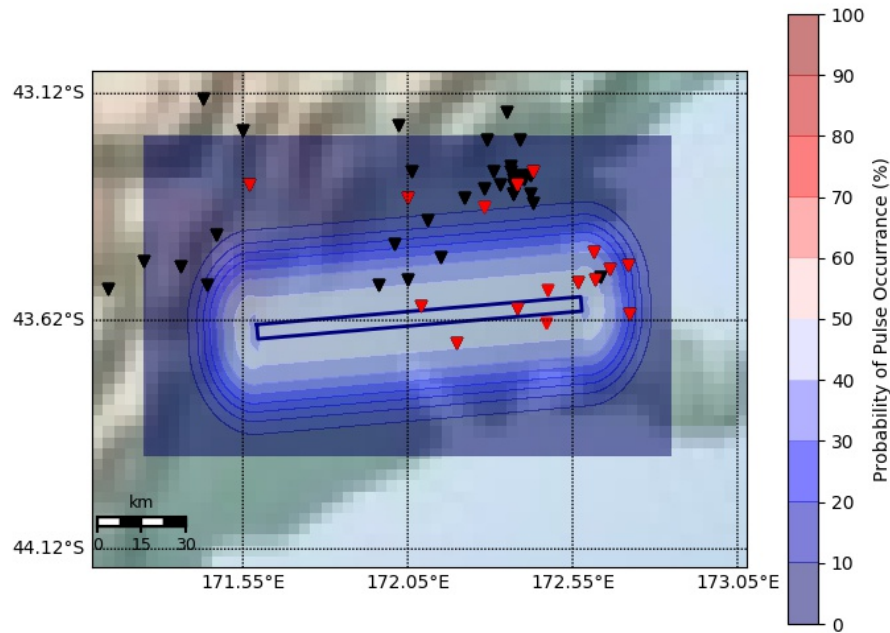
Figure 4.28: Smoothed probability distribution of observing impulsive signals for Canterbury earthquake. Features on the map are the same as in Fig. 4.19

4.3.8.2. 2010 Canterbury, New Zealand Earthquake.   The probability distribution map of observing impulsive signals for the Canterbury earthquake has become smoother. Relatively lower probabilities for the 3 km nearest to the fault are increased by 10 %. The probability of observing an impulsive signal to the Northeast of the rupture area (due to the directivity effect) is still underestimated and only partially captured.

4.3.9. Accuracy of the Model

In order to calculate the accuracy of the models, a simple approach has been taken. Calculation of the accuracy rate can be seen in Eq. 4.1.

$$\tilde{x} = \begin{cases} 1, & \text{if } pred(M_w, R_{jb}, \text{source-to-site azimuth}) \geq 50 \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

Eq. 4.1, $pred(M_w, R_{jb}, \text{source-to-site azimuth})$ identifies the prediction of the model for the given M_w, R_{jb} and source-to-site azimuth. If it is larger than 50%, the outcome is considered as impulsive, otherwise it is considered as non-impulsive. The accuracy is the average for the given dataset. Accuracy rates are calculated for each fault type. For each fault type, general accuracy and accuracy of impulsive and non-impulsive signals are calculated separately. Results can be seen in Table 4.15.

Fault Type		Normal & Normal - Oblique	Reverse & Reverse - Oblique	Strike Slip
Accuracy Rate (%)	Impulsive	76.19	35.71	77.78
	Non-impulsive	95.58	91.16	86.54
	General	93.93	90.66	86.41

Table 4.15: Accuracy rate of normal, reverse and strike slip faults for subgroups of impulsive and non-impulsive signals and general accuracy.

On all fault types the accuracy rate of non-impulsive signal predictions are higher than those of impulsive signals. The main reason for this is that recording impulsive signals at greater than 50 km R_{jb} distance from the rupture is unlikely. The dataset is dominated by records with R_{jb} distance larger than 50 km, almost all of which are non-impulsive signals. Thus the probability of observing impulsive signals in R_{jb} distance larger than 50 km is always 0%. Hence, the accuracy rates of non-impulsive signals will be significantly higher than the accuracy rates of observing impulsive signals because non-impulsive signals are better represented in the dataset. The presence of impulsive signals is quite rare also in the very near-fault regions because of their specific requirements (Section 1). Since not all of the earthquakes and the sites fulfill these requirements, some of the earthquakes do not create any impulsive signals. Due to this fact, non-impulsive signals in very near fault regions are predicted as impulsive and this acts to decrease the accuracy rate of non-impulsive signals.

Accuracy rates are lower for the impulsive signal predictions. Impulsive signals can occur at longer distances, if there is a soft soil effect (eg. SMART1 network in Taiwan) or strong directivity effect (Section 4.3.4.2). Even though these impulsive signals are implemented to the probability calculations, they created only minor increases to the probabilities due to their scarcity. Since the threshold is set at 50 %, they are all wrongly labelled. It also explains the low accuracy rate of impulsive signals for reverse faults.

The probability distribution of observing impulsive signals is given for strike slip and non strike faults in previous studies (Section 1.3). In this study, non-strike slip faults are divided into normal and reverse faults. In normal faults, there are few impulsive incidences. Merging normal and reverse faults into a single category would not change the accuracy rate of the new non-impulsive signals since the dataset would be dominated by the reverse signals. Moreover, impulsive signals of reverse faults are dominated by the high directivity effect of the Chi-Chi Taiwan earthquake and the soft soil conditions of SMART1 network in Taiwan. Even though the number of impulsive signals are lower in normal faults, in terms of the effects that created these signals they are more diverse. On the other hand, impulsive signals on normal faults are mostly concentrated in the source-to-site azimuth ± 90. Both normal and reverse faults have their own upsides and downsides in terms of data points. Hence, they are analyzed separately. A 3D representation of impulsive and non-impulsive signals for strike slip, normal and reverse faults can be seen in Fig. 4.29.
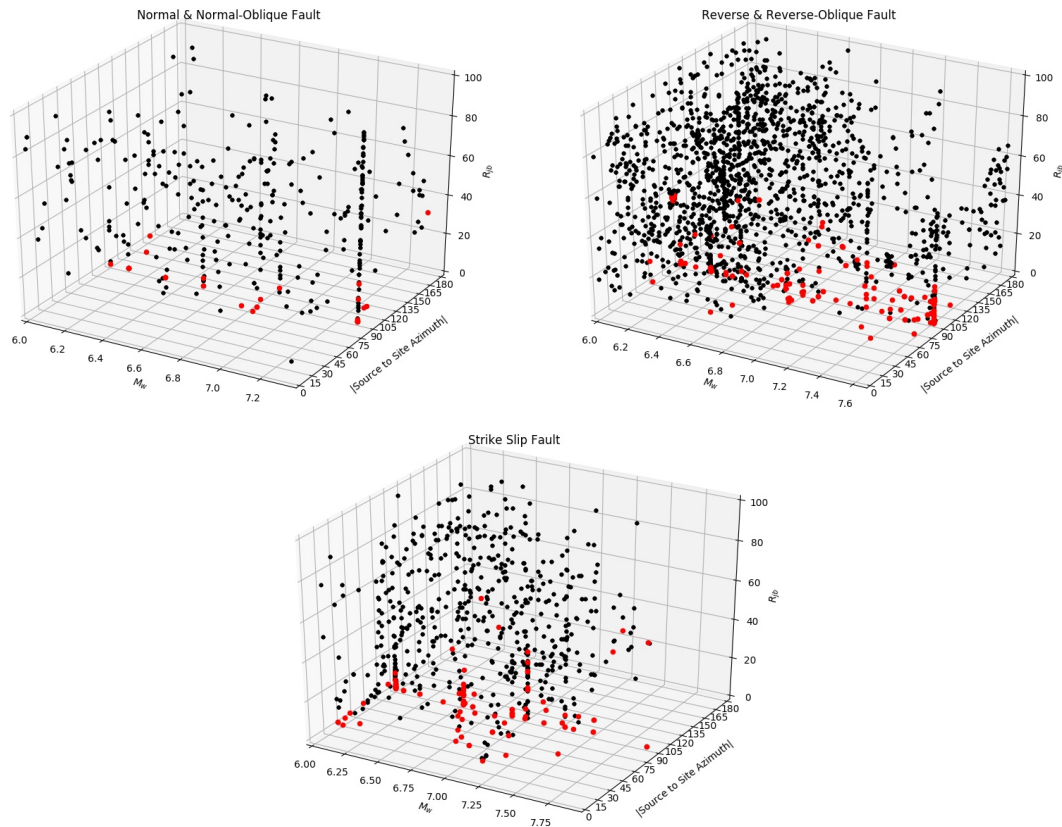
Figure 4.29: 3D representation of impulsive (red dots) and non-impulsive (black dots) signals of strike slip, normal and reverse faults with M_w, R_{jb} and |source-to-site azimuth|.

Impulsive signals for normal faults are increasing with increasing M_w where source-to-site azimuth $\neq \pm 90$. This produces relatively smooth distributions for normal fault earthquakes. There is a complex impulsive - non-impulsive data point distribution for reverse faults (Fig. 4.29). Impulsive signals for moment magnitudes from 6.0 to 6.7 are suppressed by the non-impulsive signals on all moment magnitudes, source-to-site azimuths and R_{jb}. This effects the entire model. If the model is created by using only $M_w \geq 6.8$, probabilities will be higher. The Multivariate Bayes classifier method creates smoother results by analyzing all parameters as a continuous parameter. The downside of this method is the underestimated probability distributions for reverse faults. This explains the lower accuracy rate for reverse faults. Strike slip faults have few data points after $M_w \geq 7.5$. Impulsive and non-impulsive signals are better separated with respect to reverse faults. Impulsive signals on source-to-site azimuth < 90 are more diverse in both M_w and R_{jb}. Moreover the accuracy rate of strike slip faults is higher than reverse faults.

Shahi and Baker (2011a) created a probability distribution of impulsive signals for the Imperial Valley earthquake ($M_w = 6.5$) using a method that is explained in Section 1.3.3. Probability distribution maps for this event can be seen in Fig. 4.30. Shahi and Baker (2011a) shows that the probability of an impulsive signal is increasing with increasing s (Eq. 1.23a). It results in higher probabilities where the directivity effect is focused.
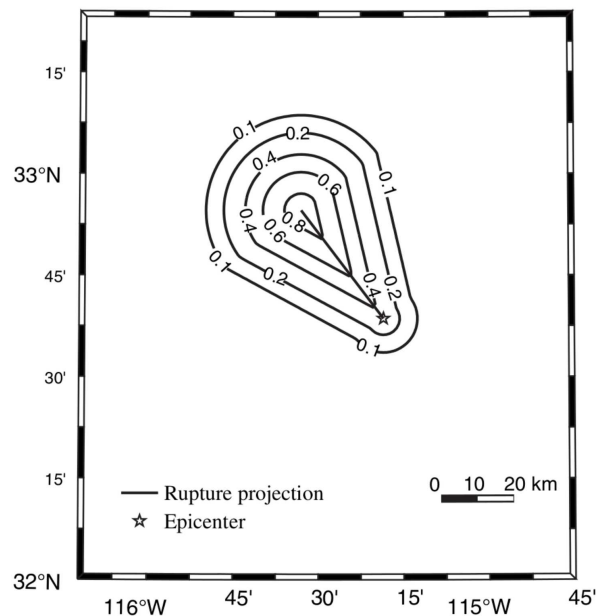
Figure 4.30: Contour map of the Imperial Valley earthquake with strike slip fault mechanism that is created by Shahi and Baker (2011a).

The initial point of an earthquake rupture cannot be predicted. For the general case, the initialization point must be taken in the middle of the fault which reduces the probabilities in one end whilst increasing probabilities at the other end. Models are created in the thesis as generalised solutions for the probability of observing impulsive signals. Even though the probabilities are lower in Fig. 4.26 with respect to Fig. 4.30, impulsive signals that occurred due to the directivity effect create higher probabilities at the edges of the fault in the model. The maximum distance of the lowest contour (10 %) is also similar. Furthermore, models that have been developed for all fault types give realistic results with moderate to high accuracy rates. Site specific and event related parameters are successfully covered by the created model.

Furthermore, probabilities are magnitude independent while Shahi and Baker (2014)

uses a PGV threshold, which is quasi dependent on M_w. Theoretically earthquake ruptures with shorter length and longer width and/or larger displacement can create bigger magnitude earthquakes. In such cases, the probability of observing impulsive signals that are calculated by Multivariate Bayesian approach is higher than Shahi and Baker (2011b, 2014) due their higher dependency on s which is related to the length of the fault. The same effect is valid for non-strike-slip faults.

4.3.10. Pulse Period vs. Moment Magnitude

The relation between M_w and T_p is calculated both for previous methods and for the wavelet method. This gives a fundamental and easy to calculate relation between the M_w and T_p that can be used for design codes. Results are shown in Fig. 4.31.
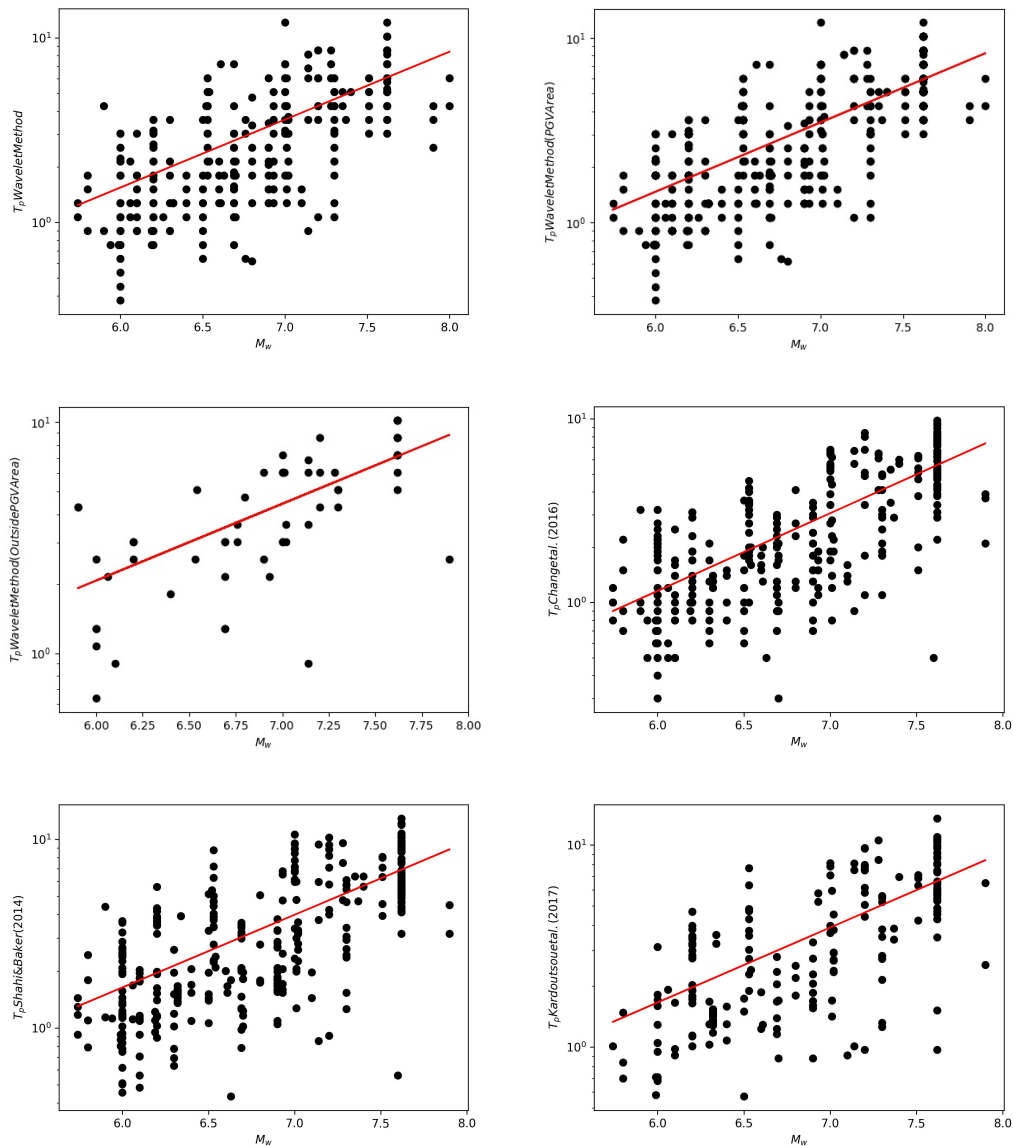
Figure 4.31: Best fitted line to the pulse periods that are determined by the wavelet method with all impulsive signals, wavelet method with impulsive signals in PGV area, wavelet method with impulsive signals outside PGV area, Chang et al. (2016); Kardoutsou et al. (2017); Shahi and Baker (2014), respectively.

Lines are fitted to the data used in this study. Equations that describe the relation between M_w and T_p can be seen in Eq. 4.2.

$$\begin{aligned}
\text{Wavelet Method} \mid \ln(T_p) &= 0.85 M_w - 4.65, \sigma_{ln(T_p)} = 0.51 \\
\text{Wavelet Method PGV Area} \mid \ln(T_p) &= 0.86 M_w - 4.80, \sigma_{ln(T_p)} = 0.52 \\
\text{Wavelet Method Outside PGV Area} \mid \ln(T_p) &= 0.76 M_w - 3.87, \sigma_{ln(T_p)} = 0.50 \\
\text{Chang et al. (2016)} \mid \ln(T_p) &= 0.97 M_w - 5.70, \sigma_{ln(T_p)} = 0.56 \\
\text{Shahi \& Baker (2014)} \mid \ln(T_p) &= 0.89 M_w - 4.83, \sigma_{ln(T_p)} = 0.50 \\
\text{Kardoutsou et al. (2017)} \mid \ln(T_p) &= 0.85 M_w - 4.61, \sigma_{ln(T_p)} = 0.48
\end{aligned} \tag{4.2}$$

A comparison between the previous studies of Baker (2007); Shahi and Baker (2014); Somerville (2003) and results from Eq. 4.2 can be seen in Fig. 4.32.

Equations used by Somerville (2003) can be seen in Eq. 1.20a and Eq. 1.20b. Equations used by Baker (2007); Shahi and Baker (2014) can be seen in Eq. 4.3.

$$\begin{aligned}
\text{Shahi \& Baker (2011)} \mid \ln(T_p) &= 1.02 M_w - 5.78, \sigma_{ln(T_p)} = 0.56 \\
\text{Shahi \& Baker (2014) Directivity Related} \mid \ln(T_p) &= 1.08 M_w - 6.26, \sigma_{ln(T_p)} = 0.61 \\
\text{Shahi \& Baker (2014) General} \mid \ln(T_p) &= 1.12 M_w - 6.55, \sigma_{ln(T_p)} = 0.57
\end{aligned} \tag{4.3}$$
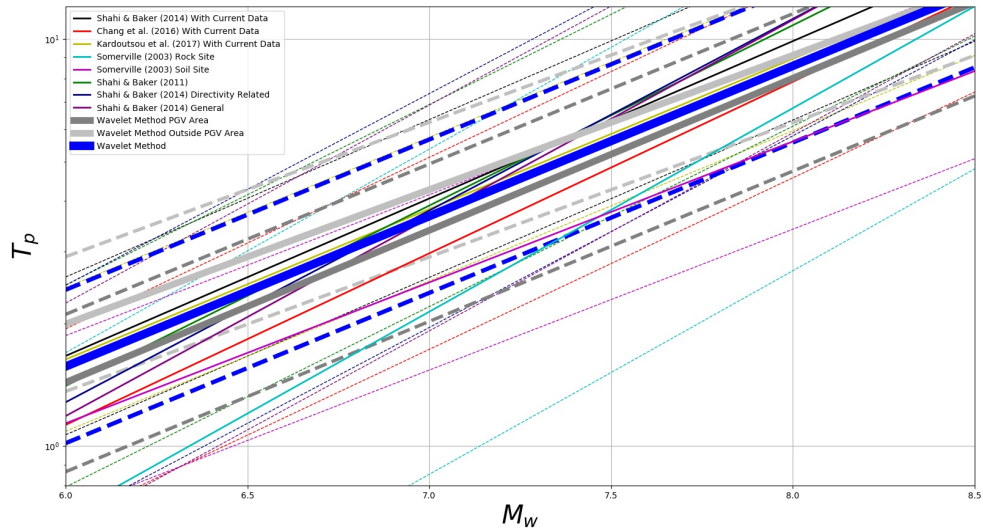
Figure 4.32: Pulse periods (s) for given M_w. Methods are given in the legend.

Periods calculated by Somerville (2003) have shorter periods with respect to the wavelet method. Shahi and Baker (2011b, 2014) calculated almost the same periods but with slightly longer periods for larger magnitudes. Lines fitted using our dataset and the previous methods and the wavelet method are also very close to each other. The wavelet method has, in general, high correlation with previous studies.

Relation between M_w and T_p of wavelet method is divided into two different formulas since the engine that produces these impulsive signals are different (Section 4.1). Pulse periods of impulsive signals outside of PGV area are larger for the same M_w with respect to the ones at PGV area. Local soil conditions are the main sources for impulses outside of PGV area. Similarly, Somerville (2003) found that pulse periods are larger for soil sites with respect to the rock sites (Eq. 1.20a and Eq. 1.20b).

5. CONCLUSION

In order to understand the features of impulsive signals in the time domain, two different algorithms are developed. Wavelet analysis and convolutional neural networks are used to create these algorithms. Spatial (probability distribution of observing impulsive signal) features of the impulsive signals are analyzed using a multivariate Bayesian classifier approach. Key results of the thesis are given below:

Impulsive signals are detected by wavelet method. Detection is done on velocity waveforms by using two different ratios. The first ratio is calculated in frequency domain. Ricker and Morlet wavelets are used for the wavelet transform. It is the ratio of wavelet power spectrum energy of the earthquake and wavelet power spectrum energy of the location where the maximum spectral power is measured. The second ratio is calculated in the time domain. It is the ratio of the squared power of velocity waveform of the earthquake and the area where maximum power spectrum energy is measured. 458 impulsive signals are identified with the wavelet method. A Ricker wavelet represents the features of 446 of impulsive signals better than a Morlet wavelet. It is found that,

(i) A Ricker wavelet can be used to identify the impulsive part of velocity waveforms in both time and frequency domain.

346 and 316 of the impulsive signals that are detected by the wavelet method are also identified as impulsive by the previous studies of Chang et al. (2016); Shahi and Baker (2014).

(ii) Results using a Ricker wavelet show a strong correlation with previous studies of Chang et al. (2016); Shahi and Baker (2014), which used alternative methods to detect impulsive signals.

Unlike the previous methods which consider only the area around the arrival of PGV for the analysis, wavelet method investigates also the arrival of maximum power spectrum energy in frequency domain. It is found that there are 57 signals where the impulsive part of the waveform is not located in the arrival of PGV.

(iii) There are significant cases where impulse-like features are located away from PGV incidence.

A law between the moment magnitude and pulse period is created which explains the relation between these two parameters with an equation of $ln(T_p) = 0.85M_w - 4.65$.

(iv) Wavelet analysis shows that a relation between pulse period, T_p, and moment magnitude, M_w exists.

A CNN method is developed in order to detect impulsive velocity waveforms. Synthetic impulsive velocity signals are created. Recorded non-impulsive signals and synthetic impulsive signals are used for the training of the model. Real impulsive signals are detected manually, and performance of the CNN method, wavelet method and previous methods are measured by using these waveforms as ground truth.

(v) The CNN method can detect impulsive signals. It has the highest accuracy rates compared to previous studies (96.34 %).

(vi) The CNN method is able to determine the weights between neurons in layers to predict the impulsive signals by training only on synthetic velocity waveforms. In other words, it understands the nature of the impulsive signals.

Unlike the wavelet method and the previous studies, CNN method makes predictions by learning the features of the impulsive signals that are given in the training phase. There are no ratios, signal fittings or thresholds in the model. It "learns" the features of the impulsive and non-impulsive signals by using a set of activation functions and convolutions. Moreover,

(vii) CNNs do not need any mathematical formulations or thresholds to make decisions about the impulsiveness of a waveform, but instead depend only on the used training set.

Parameters of R_{jb}, R_{hyp}, R_{ep}, M_w, M_0, Z_{tor}, R_{rup}, PGV/PGA, PGV and R_x have the 10 highest correlation coefficients with impulsive signals. In order to create probability of observing

impulsive signals without using complex or earthquake physics dependent parameters, only several of these parameters are implemented to the created pulse probability calculations.

(viii) The probability of observing impulsive signals can be explained by using only moment magnitude and Joyner-Boore distance, R_{jb}. These parameters are easier to determine than the parameters that have been used in various previous studies.

Models that predict the probability of observing impulsive signals are created for fundamental fault types. Only the earthquakes that created at least 1 impulsive signal are used for the model creation. Models are created by multivariate naive Bayes classifier approach. Probability distribution maps of observing impulsive signals for the fundamental fault types are generated.

(ix) Probability distributions of observing impulsive signals are calculated for strike-slip, normal and reverse faults. High probabilities for observing impulsive signals can reach up to 25 km away from the R_{jb} area.

Validity of the model is measured by using a binary classification method. Real results are compared with the developed models for each fault type. If the probability of observing impulsive signals exceeded 50 %, it would be classified as impulsive, otherwise non-impulsive. Normal and strike slip faults have above 75 % accuracy rates on the identification of impulsive signals, whereas reverse faults has accuracy rate of only 35.71 %. Low accuracy rate of reverse faults is explained by the complex features of the reverse faulting earthquakes and the stations that record them. Due to various earthquakes with high directivity effect and various stations on weak soil conditions, impulsive signals are observed in far away locations from the ruptured fault area. However, they are not common enough to create higher probabilities. As a result, probability of observing impulsive signals on reverse faults are underestimated.

(x) The probability distribution models for normal and strike-slip faults have high accuracy rates (76.19 % and 77.78 %). Probability of observing reverse faults are underestimated with the accuracy rate of 34.71 % due to rare occurrence of directivity effect and weak local soil conditions of various stations.

The results presented here give rise to further topics to be investigated. The CNN method would give more advanced information about the impulsive signal such as its pulse period, T_p, and start and end points of the impulsive part. In order to do that, another machine learning algorithm, that can interpret the information about the anomaly on the data, should be developed. The wavelet method can be used as a way of comparison, since it can identify the edges of the impulsive part of a signal. This will provide a more complete analysis of the signals. Furthermore, previous studies show the possibility of multiple impulses in the signal Zhai et al. (2018). Presence of multiple impulsive signals in a single earthquake signal can be determined when the starting and ending positions of impulses are located.

It is presented that, probabilities can be explained by using moment magnitude, Joyner-Boore distance and source-to-site azimuth. The probability distribution of the impulsive signals can be improved as well. For instance, probability distribution can be analyzed by using synthetic seismograms. There are advanced algorithms to create synthetic ground motion signals by using more complex earth structures and fault mechanics (Komatitsch and Tromp, 2002; Motazedian and Atkinson, 2005). With such powerful tools, effects of the fault orientation, local soil conditions, basin effects and rupture parameters can be observed. Spatial distributions can be investigated more deeply by implementing algorithms that include fault plane parameters such as dip and rake angles,and rupture parameters such as rise time and rupture velocity and local soil conditions. These parameters will expand the knowledge on the pulse occurrence with earthquake physics and local soil conditions, and therefore, will allow the creation of models for more complex cases of fault planes and stations.

APPENDIX A: Physical Information of the Earthquake

A.1. M_0 & M_w

M_w and M_0 are calculated automatically by almost all the seismic networks. M_0 is taken from USGS (Sipkin, 1994) when it is not provided by the local data provider (Section 2). If there is no information for the earthquake from neither the local seismological service nor USGS, parameters are calculated by using the methods that are explained below. Formulas to calculate M_w and M_0 are given in Eq. A.1.

$$M_w = \frac{3}{2} log M_0 - 16.05$$
$$M_0 = \mu L W \bar{D}$$

$$(A.1)$$

where μ indicates the rigidity of the crust and used as a constant number, 3.58^{11}. M_0 is in a unit of *dyne.cm*.

BATS provide local magnitude, M_l, on their database. M_l is converted to M_w according to the study of Chen and Tsai (2008).

A.2. Fault Mechanism

Strike, ϕ, dip, δ, and rake, λ, angles of each earthquake are used in order to calculate various parameters. These parameters are taken from the local data providers. Rake angle is an important parameter to identify the physical features of the earthquake. Rake is the angle measured on the fault plane counterclockwise from the reference strike direction to the average slip direction. In the database, fault mechanisms are divided into 5 different categories based on the rake angle. Division rules are explained in Table A.1.

Fault Type	Identifier	Lower Limit (º)	Fault Plane Parameter	Upper Limit (º)
Strike Slip	0	-180		-150
		-30		30
		150		180
Normal	1	-120		-60
Reverse	2	60	$\leq Rake \leq$	120
Reverse - Oblique	3	30		60
		120		150
Normal - Oblique	4	-150		-120
		-60		-30

Table A.1: Division of fault mechanisms with rake angle. Division is done according to study of Ancheta et al. (2014).

A.3. Fault Geometry

Fault parameters of *L* and *W* are calculated when they are not retrieved from previous studies. These parameters are calculated according to the method developed by Wells and Coppersmith (1994), which is explained in Table A.2.

Equation	Fault Type	Coefficients	
		a	b
$logW = a + b * M_w$	Strike Slip	-0.76	0.27
	Reverse	-1.61	0.41
	Normal	-1.14	0.35
$logRA = a + b * M_w$	Strike Slip	-3.42	0.90
	Reverse	-3.99	0.98
	Normal	-2.87	0.82

Table A.2: Determination of *W* and rupture area (km^2), *RA*, are explained by Wells and Coppersmith (1994). *L* is calculated by dividing *RA* to *W*. Reverse oblique and normal oblique fault mechanisms are treated as reverse and oblique faults.

A.4. Stress Drop

In the context of earthquake rupture, stress-drop is the difference between the stress across a fault before and after an earthquake rupture. Stress drop is related to the energy released as a consequence of an earthquake rupture (expressed by its seismic moment or magnitude) and to the dimensions of the rupture. A constant stress drop with increasing magnitude implies a specific amount of energy released per unit area of fault rupture as rupture size increases. One factor that may cause stress drop to vary from earthquake to earthquake or with the dimensions of an earthquake is the changing physical properties of the Earth, particularly with depth. Stress drop is measured in bars and it can be calculated as Eq. A.2.

$$StressDrop = \frac{7}{16} \frac{M_0}{\frac{(A\frac{10^{10}}{\pi})^{1.5}}{10^6}} \tag{A.2}$$

A.5. Distance Parameters

Several distance parameters are calculated in order to use as an input in multivariate Bayes classification method. These parameters are the horizontal distance to the surface projection of the rupture, R_{jb}, the horizontal distance to the surface projection of the top edge of the rupture measured perpendicular to the fault strike, R_x, the slant distance to the closest point on the rupture plane, R_{rup} and depth to top of rupture, Z_{tor}. Hanging wall, foot wall identification of the station can be done by source-to-site azimuth which can be calculated with known rupture plane geometry. Visual representation of R_{jb}, R_x and α can be seen in Fig. A.1.
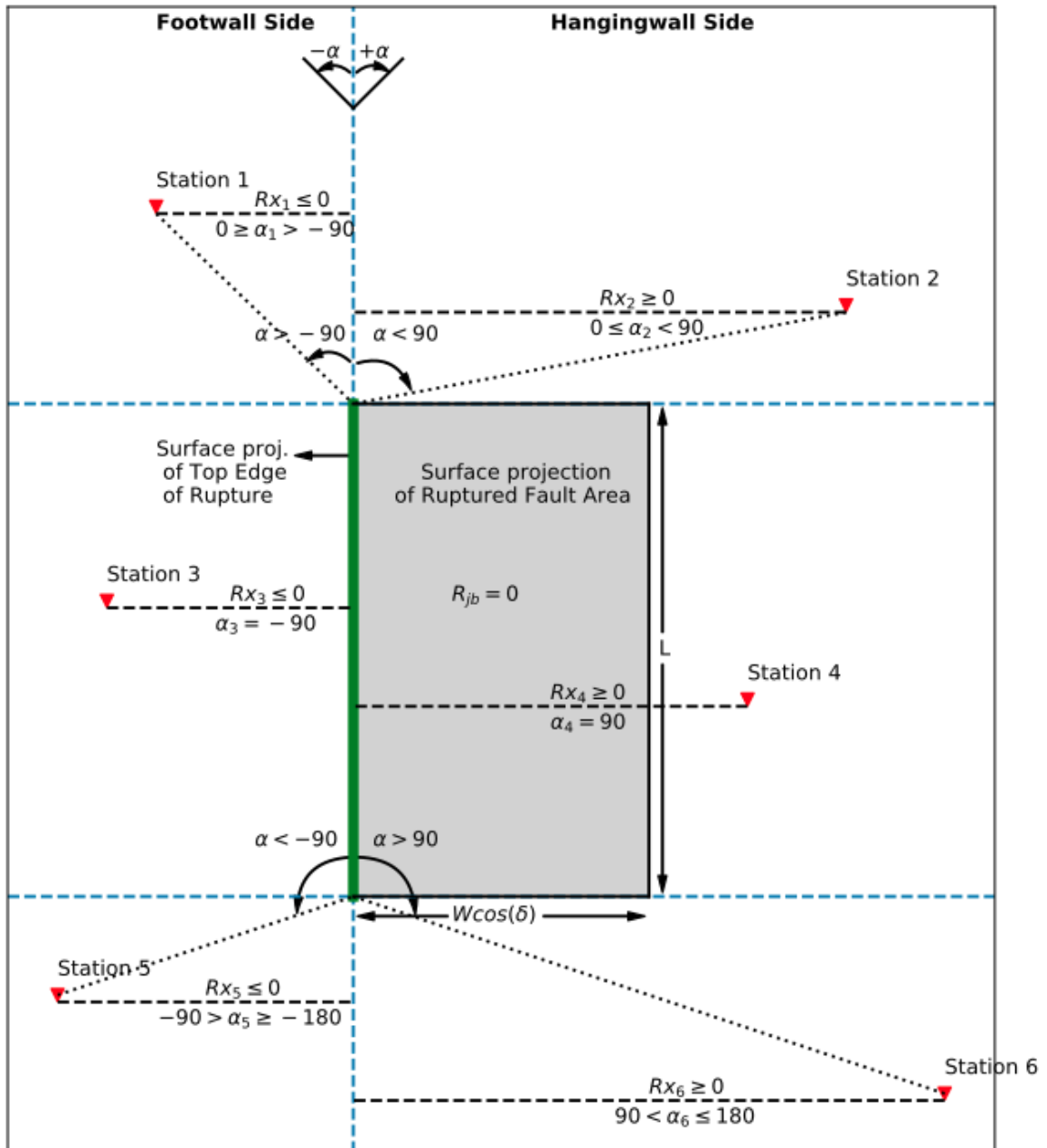
Figure A.1: Distance parameters that are used in pulse probability calculations. Several cases for source-to-site geometries are explained as planar view. α is measured clockwise from the positive source direction. α get 90º on stations that are along the fault rupture and located on hanging wall whereas they get -90º on foot wall. R_x get positive values where α is positive, vice versa. R_{jb} is the closest distance from station to surface projection of the ruptured fault area.

Illustration of R_{rup}, R_{jb}, R_x and Z_{tor} can be seen in Figure A.2. Calculation of R_{rup}, R_x

and Z_{tor} are calculated with the formulas that are explained in Kaklamanos et al. (2011).
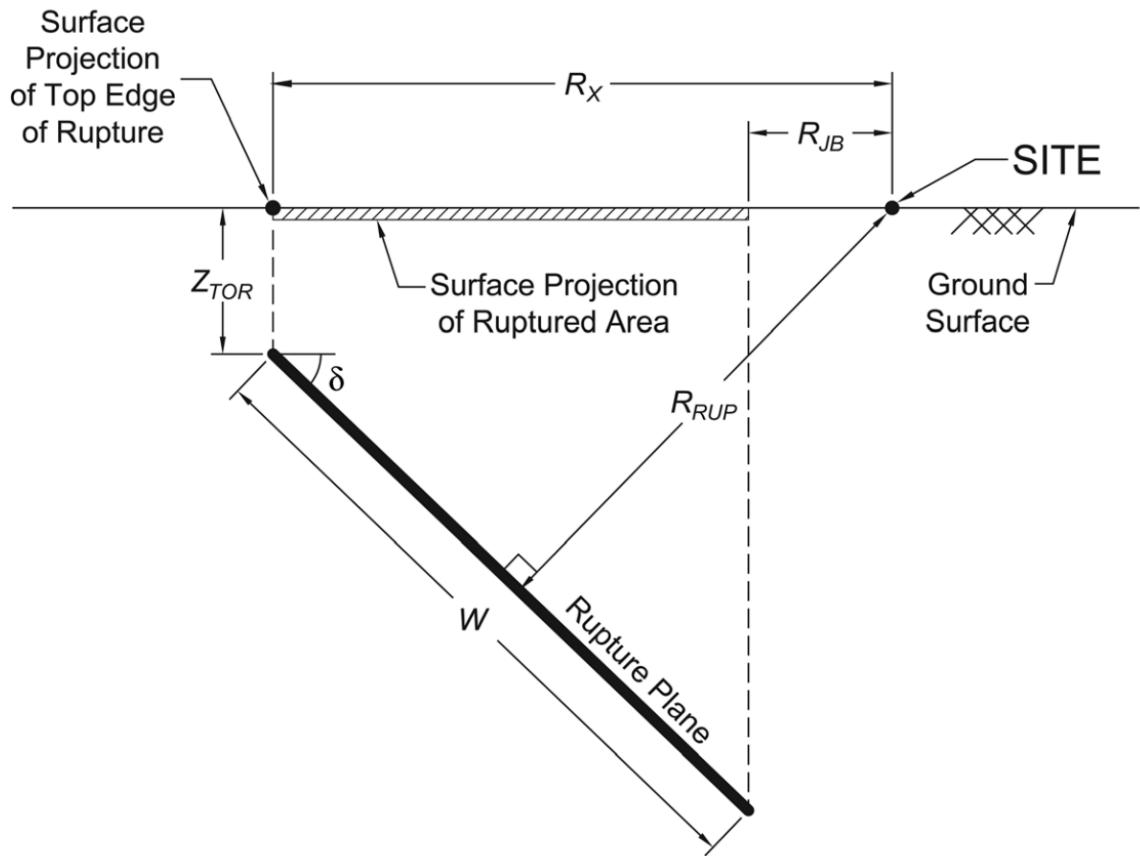
Figure A.2: Visual explanation of parameters of R_x, Z_{tor} and R_{rup}. Figure is taken from Kaklamanos et al. (2011).

APPENDIX B: Ricker Wavelet

A Ricker wavelet is created by using Eq. B.1,

$$A = (1 - 2\pi^2 f^2 t^2)e^{-\pi^2 f^2 t^2} \tag{B.1}$$

A, *f* and *t* represent a Ricker wavelet, the frequency and time stamp of the wavelet, respectively. An example of this wave can be seen in Fig. B.1.
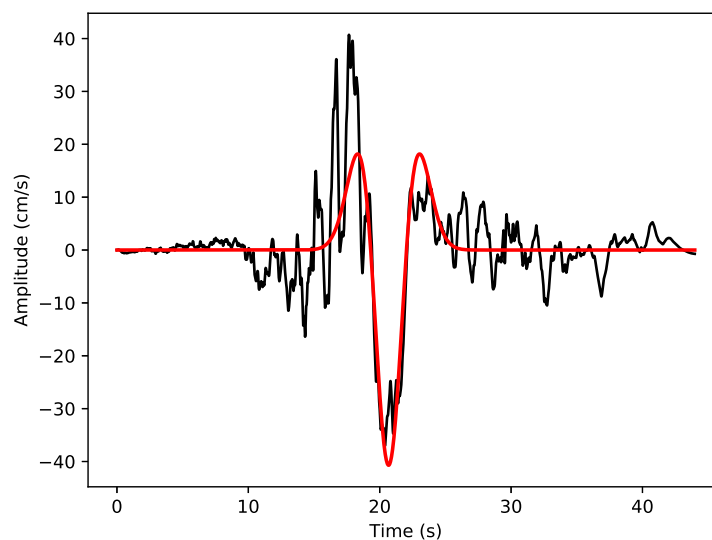
Figure B.1: 1992 Landers earthquake, Yermo Fire Station velocity waveform (black) and Ricker wavelet (red) aligned at the position of PGV.

APPENDIX C: Morlet Wavelet

A Morlet wavelet is constructed by multiplying a Gaussian wave with a sinusoidal wave. The sine wave is created as Eq. C.1,

$$SW = e^{\pi i f t} \tag{C.1}$$

SW, *i*, *f* and *t* represent the sine wave, complex number, frequency and time stamps, respectively.

The Gaussian wave is created as Eq. C.2,

$$\text{Gaussian} = e^{\frac{-t^2}{2s^2}} \tag{C.2}$$

Gaussian, *t* and *s* indicate Gaussian wave, time-stamp and width of the Gaussian, respectively. *S* value sets the order of the wavelet.

$$\text{Morlet} = SW \cdot \text{Gaussian} \tag{C.3}$$

In Eq. C.3, *Morlet* is the Morlet wavelet created by using the multiplication operation with *SW* and *Gaussian*. The Morlet wavelet is used in 3$^{\text{rd}}$ and 4$^{\text{th}}$ order. It can be seen in Fig. C.1.
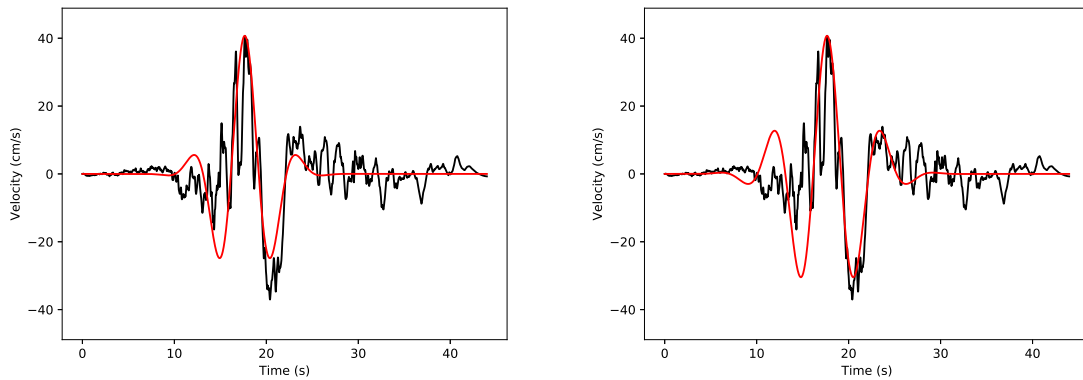
Figure C.1: 3$^{\text{rd}}$ and 4$^{\text{th}}$ order Morlet wavelet (red) on 1992 Landers earthquake, Yermo Fire Station velocity waveform (black) and aligned at the position of PGV, respectively.

Above mentioned wavelets are subtracted from the original signal, separately. The wavelet that fits best, rendering the energy of the residual signal minimum, is recognized as the wavelet that mimicked the pulse most accurately.

APPENDIX D: ReLU Activation Function

ReLU function returns 0, if the input has negative value, otherwise returns the input as it is. ReLU can be written as $f(x) = max(0,x)$, where x is the input. It can be visualized as in Fig. D.1.
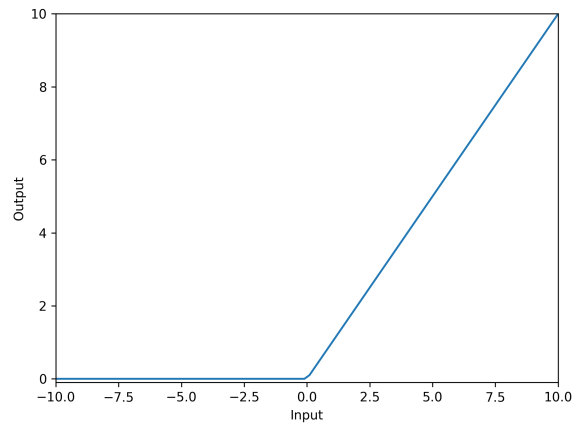
Figure D.1: Visual representation of ReLU activation function.

APPENDIX E: Sigmoid Function

Sigmoid function tends to return the value of 0, if the input is closer to the negative infinity. It tends to return 1, if the input is closer to he positive infinity. Mathematical representation of the sigmoid function is given in Eq. E.1.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{E.1}$$

Output values for sigmoid function can be seen in Fig. E.1.
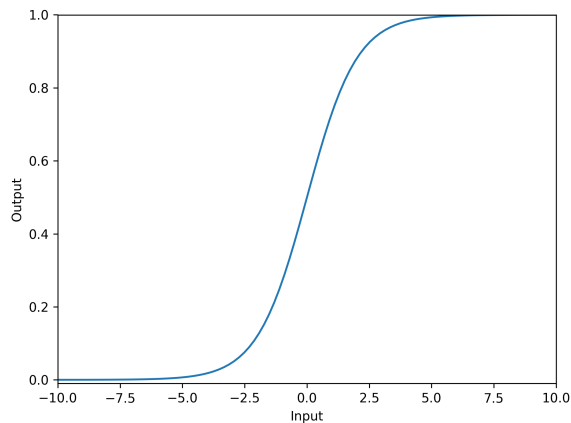
Figure E.1: Visual representation of sigmoid activation function.

The reason for the usage of this function is that it can identify the non-linearity in the data. Moreover it is differentiable, which makes it possible to do backpropagation for the calculation of the weights of different layers.

APPENDIX F: Differences Between Measured and Predicted PGVs

Predicted PGV values from GMPE of Bozorgnia et al. (2014) is compared with the measured values. Bozorgnia et al. (2014) is used NGA-West 2 database to calculate the GPME. Prediction of PGV values are done by using the EC8 classes of each station. Absolute values of measured and predicted PGVs are used. If soil classification information is not determined, type A soil class is assigned to the station. Region specification is not done and global results are used by using the basin effect correction for California (Day et al., 2008). Basin depth is chosen as 0.2 km. Impulsive signals that are identified by Shahi and Baker (2014) are plotted. Differences between measured and predicted PGV vales can be seen in Fig. F.1.
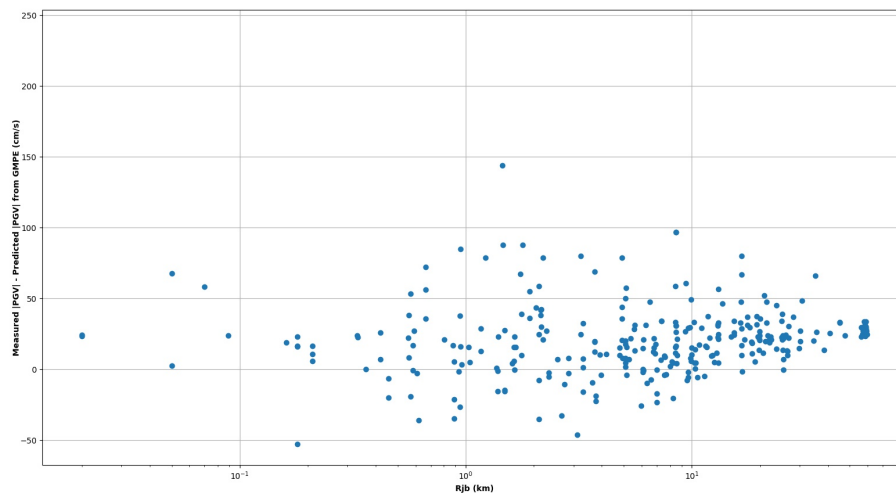
Figure F.1: Amplitude difference between measured PGV and predicted PGV from GMPE of Bozorgnia et al. (2014).

APPENDIX G: Vs_{30}

Soil classes can directly effect soil amplification and frequency content of the seismic signal. Depending on the soil classification, design spectra of the construction that is, hypothetically, going to be built on the site is changed with a constant factor. Soil classification is determined by using Vs_{30} information of the ground. Vs_{30} information is calculated as Eq. G.1

$$Vs_{30} = \frac{30}{\sum \frac{d_i}{Vs_i}} m/s \tag{G.1}$$

EC8 information is calculated by using Vs_{30} value as Table G.1:

EC8 Code	Vs_{30} $(\mathrm{m\,s^{-1}})$
A	>800
B	360-800
C	180-360
D	<180

Table G.1: EC8 Soil Classification

EC8 (Standard, 2005) codes can be explained as below

  (i) Rock or other rock-like geological formation, including at most 5 m of weaker material at the surface.

 (ii) Deposits of very dense sand, gravel, or very stiff clay, at least several tens of meters in thickness, characterized by a gradual increase of mechanical properties with depth.

(iii) Deep deposits of dense or medium dense sand, gravel or stiff clay with thickness from tens to hundreds of meters.

(iv) Deposits of loose-to-medium cohesion-less soil (with or without some soft cohesive layers), or of predominantly soft-to-firm cohesive soil.

Most of the Vs_{30} information of AFAD, CH, GeoNet, ITACA, NGA-West 2, K-NET and KiK-Net are collected from their websites. Several UNAM station information are found from Solano Hernandez (2006). BATS stations soil type information are determined from the study of Kwok et al. (2018). Rest of the soil type information are retrieved from the global slope based Vs_{30} studies of Allen and Wald (2007).

APPENDIX H: Bayes' Theorem Example

An example of Bayes' theorem can be given by using medical data as below:

Let's assume that 5 ($P(A)$) out of 100 people contain a disease. The test for the determination of the disease has an accuracy of 90 % ($P(B|A)$) meaning that the test gives the correct result in 90 % of the cases. If one person in the group tests positive, what is the probability of this person having the disease?

It is not 90 %, which is the accuracy of the test, but it is less likely as shown in Fig. H.1.
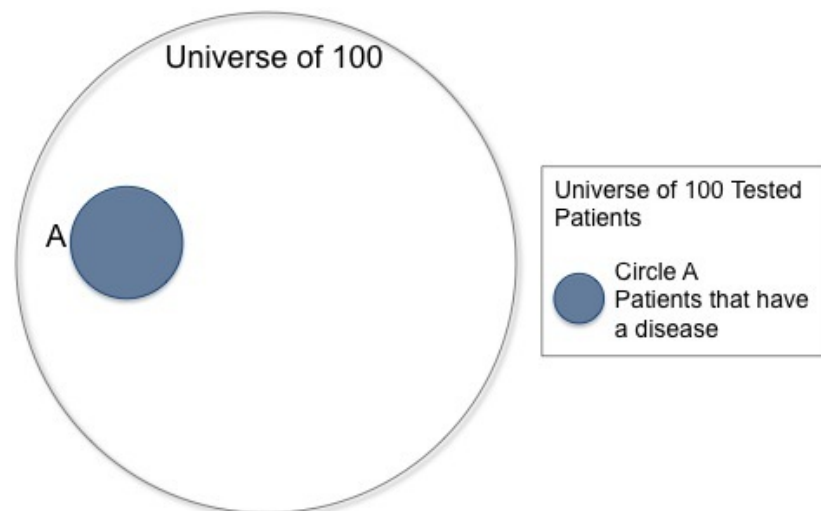
Figure H.1: Total population and the 5 people who have the disease. The circle A represents 5 people with a disease out of 100.

Now, one can overlay a circle to visualize the number of people who get a positive result on the test. Circle B will cover 10 % ($P(B)$) of the population, which is ratio of the population that is identified as having a disease by the test. Circle B will cover 90 % of the circle A since the accuracy of the test is 90 %.
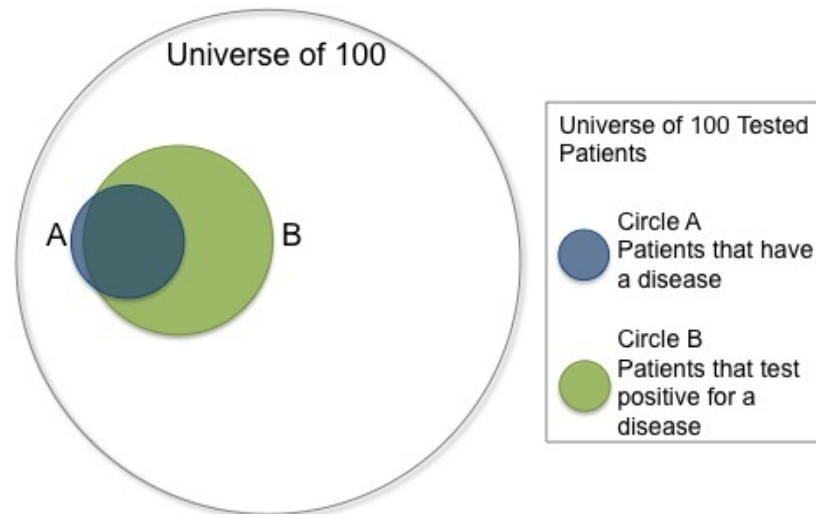
Figure H.2: Total population, 5 people who have the disease and 10 people who are detected as diseased by the test. Circle B is the identifier of the 10 % of the total population.

As one can easily identify, circle B covers more area than the people who are actually having a disease. It is due to the fact that 14 out of 100 (90 % of the 5 people with the disease + 10 % of the 95 people without the disease) will receive a positive result. Even though the accurate rate of the test is 90 %, only the 32.14% ($P(A|B)$) of the population inside the circle B will have the disease.

APPENDIX I: Phase Identification for Ricker Wavelet

Ricker wavelet that can represent the impulsive part of the waveform is rotated between $0°$ and $355°$. In order to rotate the Ricker wavelet, first the Hilbert transformation is applied on Ricker wavelet in Eq. I.1.

$$x + iy = F^{-1}(F(x)2U) \tag{I.1}$$

In Eq. I.1, x and iy, represents the real and imaginary part of the Hilbert transformed signal, respectively. $F(x)$ is the Fourier transform of the velocity waveform and U is the unit step function. Then the signals are rotated by using the real, x, and imaginary part, y, of the transformed signal as in Eq. I.2.

$$R(\theta) = cos(\theta) * x - sin(\theta) * y \tag{I.2}$$

Rotated Ricker wavelet is represented with $R(\theta)$ where θ is the rotation angle. Correlation coefficients are calculated between original waveform and rotated Ricker wavelet. Rotated wavelet with the highest correlation coefficient is considered as the representative of the impulsive part of the velocity waveform with a rotation angle. An example of a rotated Ricker wavelet can be seen in Fig. I.1.
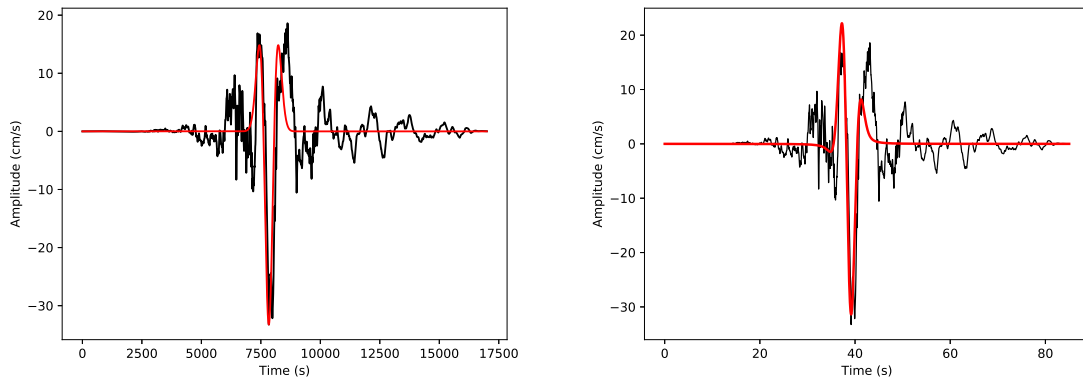
Figure I.1: 1999 Chi-Chi Taiwan Earthquake, TCU046 Station signal (R_{ep} = 78.17 km) and the Ricker wavelet without phase shifting (upper) and with phase shifting of θ =310° (lower). Black and red waveforms are indicating the original velocity waveforms and Ricker wavelet signals, respectively.

REFERENCES

Aagaard, B. T., Hall, J. F., and Heaton, T. H. (2004). Effects of fault dip and slip rake angles on near-source ground motions: Why rupture directivity was minimal in the 1999 Chi-Chi, Taiwan, earthquake. *Bulletin of the Seismological Society of America*, 94(1):155–170.

Adeli, H. and Panakkat, A. (2009). A probabilistic neural network for earthquake magnitude prediction. *Neural networks*, 22(7):1018–1024.

AFAD (2019). AFAD - Turkey Earthquake Data Center System. Accessed: 2019-10-25.

Aki, K. and Richards, P. G. (2002). *Quantitative seismology*. University Science Books.

Al-Kaabi, A., Lee, W. J., et al. (1993). Using Artificial Neural Networks To Identify the Well Test Interpretation Model (includes associated papers 28151 and 28165). *SPE formation evaluation*, 8(03):233–240.

Alaska Earthquake Center, U. O. A. F. (1987). Alaska Regional Network.

Alavi, B. and Krawinkler, H. (2004). Behavior of moment-resisting frame structures subjected to near-fault ground motions. *Earthquake engineering & structural dynamics*, 33(6):687–706.

Albuquerque Seismological Laboratory (ASL)/USGS (1988). Global Seismograph Network (GSN - IRIS/USGS).

Ali, J. et al. (1994). Neural networks: a new tool for the petroleum industry? In *European petroleum computer conference*. Society of Petroleum Engineers.

Alizadeh, M., Ngah, I., Hashim, M., Pradhan, B., and Pour, A. (2018). A Hybrid Analytic Network Process and Artificial Neural Network (ANP-ANN) Model for Urban Earthquake Vulnerability Assessment. *Remote Sensing*, 10(6):975.

Allen, T. I. and Wald, D. J. (2007). Topographic slope as a proxy for seismic site-conditions (VS30) and amplification around the globe. Technical report, Geological Survey (US).

Ambrožič, T. and Turk, G. (2003). Prediction of subsidence due to underground mining by artificial neural networks. *Computers & Geosciences*, 29(5):627–637.

Amirifar, L. and Shafiee, H. (2018). Estimating of Loss Human Life Caused Through Earthquake Employing Neural Network. *Journal of Advances in Computer Research*, 9(2):71–89.

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., and Donahue, J. L. (2014). NGA-west2 database. *Earthquake Spectra*, 30(3):989–1005.

Antonellis, G. and Panagiotou, M. (2013). Seismic response of bridges with rocking foundations compared to fixed-base bridges at a near-fault site. *Journal of Bridge Engineering*, 19(5):04014007.

Aristotle University Of Thessaloniki Seismological Network (1981). Permanent Regional Seismological Network operated by the Aristotle University of Thessaloniki.

Asano, K. and Iwata, T. (2019). Source rupture process of the 2018 Hokkaido Eastern Iburi earthquake deduced from strong-motion data considering seismic wave propagation in three-dimensional velocity structure. *Earth, Planets and Space*, 71(1):101.

Asencio-Cortés, G., Morales-Esteban, A., Shang, X., and Martínez-Álvarez, F. (2018). Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure. *Computers & Geosciences*, 115:198–210.

Baker, J. W. (2007). Quantitative classification of near-fault ground motions using wavelet analysis. *Bulletin of the Seismological Society of America*, 97(5):1486–1501.

Barka, A. (1992). The north anatolian fault zone. In *Annales tectonicae*, volume 6, pages 164–195.

Barth, N. C. (2013). *A tectono-geomorphic study of the Alpine Fault, New Zealand*. PhD thesis, University of Otago.

Bendick, R. (2012). Southwest Montana GPS Network.

Benioff, H. (1955). Mechanism and strain characteristics of the White Wolf fault as indicated by the aftershock sequence. *Bull., Calif. Div. Mines*, 171:199–202.

Bernard, P., Herrero, A., and Berge, C. (1996). Modeling directivity of heterogeneous earthquake ruptures. *Bulletin of the Seismological Society of America*, 86(4):1149–1160.

Bertero, V. V., Mahin, S. A., and Herrera, R. A. (1978). Aseismic design implications of near-fault San Fernando earthquake records. *Earthquake engineering & structural dynamics*, 6(1):31–42.

Bianchi, I., Chiarabba, C., and Piana Agostinetti, N. (2010). Control of the 2009 L'Aquila earthquake, central Italy, by a high-velocity structure: A receiver function study. *Journal of Geophysical Research: Solid Earth*, 115(B12).

Bird, P. (2003). An updated digital model of plate boundaries. *Geochemistry, Geophysics, Geosystems*, 4(3).

Blaikie, P., Cannon, T., Davis, I., and Wisner, B. (2005). *At risk: natural hazards, people's vulnerability and disasters*. Routledge.

Bogazici University Kandilli Observatory And Earthquake Research Institute (2001). Bogazici University Kandilli Observatory And Earthquake Research Institute.

Boore, D. M. (1983). Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. *Bulletin of the Seismological Society of America*, 73(6A):1865–1894.

Bouchon, M., Toksoz, M. N., Karabulut, H., Bouin, M.-P., Dietrich, M., Aktar, M., and Edie, M. (2002). Space and time evolution of rupture and faulting during the 1999 Izmit (Turkey) earthquake. *Bulletin of the Seismological Society of America*, 92(1):256–266.

Bozorgnia, Y., Abrahamson, N. A., Atik, L. A., Ancheta, T. D., Atkinson, G. M., Baker, J. W., Baltay, A., Boore, D. M., Campbell, K. W., Chiou, B. S.-J., et al. (2014). NGA-West2 research project. *Earthquake Spectra*, 30(3):973–987.

Bradley, B. A. (2012). Strong ground motion characteristics observed in the 4 September 2010 Darfield, New Zealand earthquake. *Soil Dynamics and Earthquake Engineering*, 42:32–46.

Bray, J. D. and Rodriguez-Marek, A. (2004). Characterization of forward-directivity ground motions in the near-fault region. *Soil dynamics and earthquake engineering*, 24(11):815–828.

Buduma, N. and Locascio, N. (2017). *Fundamentals of deep learning: Designing next-generation machine intelligence algorithms*. " O'Reilly Media, Inc.".

California Institute Of Technology And United States Geological Survey Pasadena (1926). Southern California Seismic Network.

Causse, M., Chaljub, E., Cotton, F., Cornou, C., and Bard, P.-Y. (2009). New approach for coupling k- 2 and empirical Green's functions: application to the blind prediction of broad-band ground motion in the Grenoble basin. *Geophysical Journal International*, 179(3):1627–1644.

Chai, J.-F., Loh, C.-H., and Chen, C.-Y. (2000). Consideration of the near-fault effect on seismic design code for sites near the Chelungpu fault. *Journal of the Chinese Institute of Engineers*, 23(4):447–454.

Chang, Z., De Luca, F., and Goda, K. (2019). Automated classification of near-fault acceleration pulses using wavelet packets. *Computer-Aided Civil and Infrastructure Engineering*.

Chang, Z., Sun, X., Zhai, C., Zhao, J. X., and Xie, L. (2016). An improved energy-based approach for selecting pulse-like ground motions. *Earthquake Engineering & Structural Dynamics*, 45(14):2405–2411.

Chen, K.-P. and Tsai, Y.-B. (2008). A Catalog of Taiwan Earthquakes (1900-2006) with Homogenized Mw Magnitudes. *Bulletin of the Seismological Society of America*, 98(1):483–489.

Chioccarelli, E. and Iervolino, I. (2013). Near-source seismic hazard and design scenarios. *Earthquake engineering & structural dynamics*, 42(4):603–622.

Coutant, O. (1989). Programme de simulation numerique AXITRA. *Rapport LGIT*.

Day, S. M., Graves, R., Bielak, J., Dreger, D., Larsen, S., Olsen, K. B., Pitarka, A., and Ramirez-Guzman, L. (2008). Model for basin effects on long-period response spectra in southern California. *Earthquake Spectra*, 24(1):257–277.

De Natale, G., Zollo, A., Del Gaudio, C., Ricciardi, G., and Martini, M. (1984). Error analysis in hypocentral locations at Phlegraean Fields. *Bulletin volcanologique*, 47(2):209–218.

Del Gaudio, S., Causse, M., and Festa, G. (2015). Broad-band strong motion simulations coupling k-square kinematic source models with empirical Green's functions: the 2009 L'Aquila earthquake. *Geophysical Journal International*, 203(1):720–736.

Del Pezzo, E., Esposito, A., Giudicepietro, F., Marinaro, M., Martini, M., and Scarpetta, S. (2003). Discrimination of earthquakes and underwater explosions using neural networks. *Bulletin of the Seismological Society of America*, 93(1):215–223.

DeMets, C., Gordon, R. G., and Argus, D. F. (2010). Geologically current plate motions. *Geophysical Journal International*, 181(1):1–80.

DeMets, C., Gordon, R. G., Argus, D. F., and Stein, S. (1994). Effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motions. *Geophysical research letters*, 21(20):2191–2194.

Dhakal, Y. P., Kunugi, T., Kimura, T., Suzuki, W., and Aoi, S. (2019). Peak ground motions and characteristics of nonlinear site response during the 2018 Mw 6.6 Hokkaido eastern Iburi earthquake. *Earth, Planets and Space*, 71(1):56.

Di Stefano, R., Chiarabba, C., Chiaraluce, L., Cocco, M., De Gori, P., Piccinini, D., and Valoroso, L. (2011). Fault zone properties affecting the rupture evolution of the 2009 (Mw 6.1) L'Aquila earthquake (central Italy): Insights from seismic tomography. *Geophysical Research Letters*, 38(10).

Dickey, J., Borghetti, B., and Junek, W. (2019). Improving Regional and Teleseismic Detection for Single-Trace Waveforms Using a Deep Temporal Convolutional Neural Network Trained with an Array-Beam Catalog. *Sensors*, 19(3):597.

Dysart, P. S. and Pulli, J. J. (1990). Regional seismic event classification at the NORESS array: seismological measurements and the use of trained neural networks. *Bulletin of the Seismological Society of America*, 80(6B):1910–1933.

Ertuncay, D. and Costa, G. (2019). An alternative pulse classification algorithm based on multiple wavelet analysis. *Journal of Seismology*, 23(4):929–942.

Falsaperla, S., Graziani, S., Nunnari, G., and Spampinato, S. (1996). Automatic classification of volcanic earthquakes by using multi-layered neural networks. *Natural Hazards*, 13(3):205–228.

Fayjaloun, R., Causse, M., Voisin, C., Cornou, C., and Cotton, F. (2016). Spatial variability of the directivity pulse periods observed during an earthquake. *Bulletin of the Seismological Society of America*, 107(1):308–318.

Finley, T., Morell, K., Leonard, L., Regalla, C., Johnston, S. T., and Zhang, W. (2019). Ongoing oroclinal bending in the Cascadia forearc and its relation to concave-outboard plate margin geometry. *Geology*, 47(2):155–158.

Florido, E., Asencio-Cortés, G., Aznarte, J. L., Rubio-Escudero, C., and Martínez-Álvarez, F. (2018). A novel tree-based algorithm to discover seismic patterns in earthquake catalogs. *Computers & Geosciences*, 115:96–104.

for Earth Science, N. R. I. and Resilience, D. (2019). NIED K-NET, KiK-net. Accessed: 2019-11-18.

Geoffrey A. Abers, K. M. F. (2003). Tomography Under Costa Rica and Nicaragua.

GEOFON Data Centre (1993). GEOFON Seismic Network.

Geological Survey Of Canada (1989). Canadian National Seismograph Network.

Ghaffarzadeh, H. (2016). A classification method for pulse-like ground motions based on S-transform. *Natural Hazards*, 84(1):335–350.

Giardini, D. (1999). The global seismic hazard assessment program (GSHAP)-1992/1999. *Annals of Geophysics*, 42(6).

Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256.

Goldstein, P. and Snoke, A. (2005). SAC availability for the IRIS community. *Incorporated Research Institutions for Seismology Newsletter*, 7(UCRL-JRNL-211140).

Gomberg, J., Bodin, P., and Reasenberg, P. A. (2003). Observing earthquakes triggered in the near field by dynamic deformations. *Bulletin of the Seismological Society of America*, 93(1):118–138.

Guo, G., Yang, D., and Liu, Y. (2018). Duration effect of near-fault pulse-like ground motions and identification of most suitable duration measure. *Bulletin of Earthquake Engineering*, 16(11):5095–5119.

Gurley, K. and Kareem, A. (1999). Applications of wavelet transforms in earthquake, wind and ocean engineering. *Engineering structures*, 21(2):149–167.

Hall, J. F., Heaton, T. H., Halling, M. W., and Wald, D. J. (1995). Near-source ground motion and its effects on flexible buildings. *Earthquake spectra*, 11(4):569–605.

Hancock, J., Watson-Lamprey, J., Abrahamson, N. A., Bommer, J. J., Markatis, A., Mc-COY, E., and Mendis, R. (2006). An improved method of matching response spectra of recorded earthquake ground motion using wavelets. *Journal of earthquake engineering*, 10(spec01):67–89.

Haynie, K. and Jadamec, M. (2017). Tectonic drivers of the Wrangell block: Insights on fore-arc sliver processes from 3-D geodynamic models of Alaska. *Tectonics*, 36(7):1180–1206.

Holzer, T. L. (2000). *Implications for earthquake risk reduction in the United States from the Kocaeli, Turkey, earthquake of August 17, 1999*, volume 1193. US Government Printing Office.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

Iervolino, I., Chioccarelli, E., and Baltzopoulos, G. (2012). Inelastic displacement ratio of near-source pulse-like ground motions. *Earthquake Engineering & Structural Dynamics*, 41(15):2351–2357.

Iervolino, I. and Cornell, C. A. (2008). Probability of occurrence of velocity pulses in near-source ground motions. *Bulletin of the Seismological Society of America*, 98(5):2262–2277.

Ikeda, Y., Chida, N., Nakata, T., Kaneda, H., Tajikara, M., and Takazawa, S. (2001). 1: 25,000-scale active fault map in urban area "Kumamoto". Technical report, Technical report of the Geographical Survey Institute, Ibaraki, D1.

Institute of Earth Sciences, Academia Sinica, T. (1996). Broadband array in Taiwan for seismology.

Irikura, K. (1992). The construction of large earthquake by a superposition of small events. In *Proc. of the 10th World Conf. on Earthquake Engineering*, pages 727–730.

(ITSAK) Institute Of Engineering Seimology Earthquake Engineering (1981). ITSAK Strong Motion Network.

Jadamec, M. A. and Billen, M. I. (2010). Reconciling surface plate motions with rapid three-dimensional mantle flow around a slab edge. *Nature*, 465(7296):338.

Kaklamanos, J., Baise, L. G., and Boore, D. M. (2011). Estimating unknown input parameters when implementing the NGA ground-motion prediction equations in engineering practice. *Earthquake Spectra*, 27(4):1219–1235.

Kalkan, E. and Kunnath, S. K. (2006). Effects of fling step and forward directivity on seismic response of buildings. *Earthquake spectra*, 22(2):367–390.

Kardoutsou, V., Taflampas, I., and Psycharis, I. N. (2017). A new pulse indicator for the classification of ground motions. *Bulletin of the Seismological Society of America*, 107(3):1356–1364.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kious, W. J. and Tilling, R. I. (1996). *This dynamic Earth: the story of plate tectonics*. DIANE Publishing.

Kobayashi, H., Koketsu, K., and Miyake, H. (2017). Rupture processes of the 2016 Kumamoto earthquake sequence: Causes for extreme ground motions. *Geophysical Research Letters*, 44(12):6002–6010.

Kobayashi, H., Koketsu, K., and Miyake, H. (2019). Rupture process of the 2018 Hokkaido Eastern Iburi earthquake derived from strong motion and geodetic data. *Earth, Planets and Space*, 71(1):63.

Komatitsch, D. and Tromp, J. (2002). Spectral-element simulations of global seismic wave propagation—I. Validation. *Geophysical Journal International*, 149(2):390–412.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C., and Wassermann, J. (2015). ObsPy: A bridge for seismology into the scientific Python ecosystem. *Computational Science & Discovery*, 8(1):014003.

Kwok, O. L. A., Stewart, J. P., Kwak, D. Y., and Sun, P.-L. (2018). Taiwan-Specific Model for VS30 Prediction Considering Between-Proxy Correlations. *Earthquake Spectra*.

Lay, T. and Wallace, T. C. (1995). *Modern global seismology*, volume 58. Elsevier.

Liu, C., Lay, T., and Xiong, X. (2018). Rupture in the 4 May 2018 MW 6.9 earthquake seaward of the Kilauea east rift zone fissure eruption in Hawaii. *Geophysical Research Letters*, 45(18):9508–9515.

Loh, C.-H., Wan, S., and Liao, W.-I. (2002). Effects of hysteretic model on seismic demands: consideration of near-fault ground motions. *The Structural Design of Tall Buildings*, 11(3):155–169.

Luzi, L., Puglia, R., Russo, E., D'Amico, M., Felicetta, C., Pacor, F., Lanzano, G., Çeken, U., Clinton, J., Costa, G., et al. (2016). The engineering strong-motion database: A platform to access pan-European accelerometric data. *Seismological Research Letters*, 87(4):987–997.

Mai, P. M. and Thingbaijam, K. (2014). SRCMOD: An online database of finite-fault rupture models. *Seismological Research Letters*, 85(6):1348–1357.

Martínez-Álvarez, F., Troncoso, A., Morales-Esteban, A., and Riquelme, J. C. (2011). Computational intelligence techniques for predicting earthquakes. In *International Conference on Hybrid Artificial Intelligence Systems*, pages 287–294. Springer.

Mavroeidis, G. P. and Papageorgiou, A. S. (2003). A mathematical representation of near-fault ground motions. *Bulletin of the seismological society of America*, 93(3):1099–1131.

Mazza, F. (2018). Seismic demand of base-isolated irregular structures subjected to pulse-type earthquakes. *Soil Dynamics and Earthquake Engineering*, 108:111–129.

McCaffrey, R., King, R. W., Payne, S. J., and Lancaster, M. (2013). Active tectonics of northwestern US inferred from GPS-derived surface velocities. *Journal of Geophysical Research: Solid Earth*, 118(2):709–723.

McCormack, M. D., Zaucha, D. E., and Dushek, D. W. (1993). First-break refraction event picking and seismic data trace editing using neural networks. *Geophysics*, 58(1):67–78.

Mena, B. and Mai, P. M. (2011). Selection and quantification of near-fault velocity pulses owing to source directivity. *Georisk*, 5(1):25–43.

Mohaghegh, S., Arefi, R., Ameri, S., Aminiand, K., and Nutter, R. (1996). Petroleum reservoir characterization with the aid of artificial neural networks. *Journal of Petroleum Science and Engineering*, 16(4):263–274.

Mohaghegh, S. et al. (1995). Neural network: what it can do for petroleum engineers. *Journal of Petroleum Technology*, 47(01):42–42.

Möller, O., Foschi, R. O., Quiroz, L. M., and Rubinstein, M. (2009). Structural optimization for performance-based design in earthquake engineering: Applications of neural networks. *Structural Safety*, 31(6):490–499.

Morelli, A. and Barrier, E. (2004). Geodynamic map of the Mediterranean. Sheet 2: Seismicity and tectonics. *Commission for the geologic map of the world- CGMW/CCGM*.

Motazedian, D. and Atkinson, G. M. (2005). Stochastic finite-fault modeling based on a dynamic corner frequency. *Bulletin of the Seismological Society of America*, 95(3):995–1010.

Mukhopadhyay, S. and Gupta, V. K. (2013). Directivity pulses in near-fault ground motions—I: Identification, extraction and modeling. *Soil Dynamics and Earthquake Engineering*, 50:1–15.

Müller, R. D., Sdrolias, M., Gaina, C., and Roest, W. R. (2008). Age, spreading rates, and spreading asymmetry of the world's ocean crust. *Geochemistry, Geophysics, Geosystems*, 9(4).

National Observatory Of Athens, I. O. G. (1997). National Observatory of Athens Seismic Network.

Nejad, M. M., Momeni, M. S., and Manahiloh, K. N. (2018). Shear wave velocity and soil type microzonation using neural networks and geographic information system. *Soil Dynamics and Earthquake Engineering*, 104:54–63.

Northern California Earthquake Data Center (2014). Berkeley Digital Seismic Network (BDSN).

Observatory/USGS, A. V. (1988). Alaska Volcano Observatory.

Pacheco, J. F., Quintero, R., Vega, F., Segura, J., Jiménez, W., and González, V. (2006). The M w 6.4 Damas, Costa Rica, earthquake of 20 November 2004: Aftershocks and slip distribution. *Bulletin of the Seismological Society of America*, 96(4A):1332–1343.

Pacor, F., Paolucci, R., Luzi, L., Sabetta, F., Spinelli, A., Gorini, A., Nicoletti, M., Marcucci, S., Filippi, L., and Dolce, M. (2011). Overview of the Italian strong motion database ITACA 1.0. *Bulletin of Earthquake Engineering*, 9(6):1723–1739.

Palano, M. (2014). On the present-day crustal stress, strain-rate fields and mantle anisotropy pattern of Italy. *Geophysical Journal International*, 200(2):969–985.

Paolucci, R., Gatti, F., Infantino, M., Smerzini, C., Güney Özcebe, A., and Stupazzini, M. (2018). Broadband Ground Motions from 3D Physics-Based Numerical Simulations Using Artificial Neural Networks. *Bulletin of the Seismological Society of America*, 108(3A):1272–1286.

Papageorgiou, A. S. and Aki, K. (1983). A specific barrier model for the quantitative description of inhomogeneous faulting and the prediction of strong ground motion. Part II. Applications of the model. *Bulletin of the Seismological Society of America*, 73(4):953–978.

Plafker, G., Moore, J., Winkler, G., and Berg, H. (1994). Geology of the southern Alaska margin. *The geology of Alaska*, 1.

Polonia, A., Torelli, L., Mussoni, P., Gasperini, L., Artoni, A., and Klaeschen, D. (2011). The Calabrian Arc subduction complex in the Ionian Sea: Regional architecture, active deformation, and seismic hazard. *Tectonics*, 30(5).

Presidency Of Counsil Of Ministers-Civil Protection Department (1972). Italian Strong Motion Network.

Rafiei, M. H. and Adeli, H. (2017). NEEWS: a novel earthquake early warning model using neural dynamic classification and neural dynamic optimization. *Soil Dynamics and Earthquake Engineering*, 100:417–427.

Reyes, J., Morales-Esteban, A., and Martínez-Álvarez, F. (2013). Neural networks to predict earthquakes in Chile. *Applied Soft Computing*, 13(2):1314–1328.

Rockwell, T., Barka, A., Dawson, T., Akyuz, S., and Thorup, K. (2001). Paleoseismology of the gazikoy-saros segment of the north anatolia fault, northwestern turkey: Comparison of the historical and paleoseismic records, implications of regional seismic hazard, and models of earthquake recurrence. *Journal of Seismology*, 5(3):433–448.

Ross, Z. E., Meier, M.-A., and Hauksson, E. (2018a). P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, 123(6):5120–5129.

Ross, Z. E., Meier, M.-A., Hauksson, E., and Heaton, T. H. (2018b). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, 108(5A):2894–2901.

RSN (2017). Información de la Red Sismológica Nacional de Costa Rica. Technical report, Universidad de Costa Rica.

Sagiya, T., Miyazaki, S., and Tada, T. (2000). Continuous GPS array and present-day crustal deformation of Japan. *Pure and applied Geophysics*, 157(11-12):2303–2322.

Scala, A., Festa, G., and Del Gaudio, S. (2018). Relation Between Near-Fault Ground Motion Impulsive Signals and Source Parameters. *Journal of Geophysical Research: Solid Earth*, 123(9):7707–7721.

Scarpetta, S., Giudicepietro, F., Ezin, E. C., Petrosino, S., Del Pezzo, E., Martini, M., and Marinaro, M. (2005). Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy, using neural networks. *Bulletin of the Seismological Society of America*, 95(1):185–196.

Schreiber, J. (2018). Pomegranate: fast and flexible probabilistic modeling in python. *Journal of Machine Learning Research*, 18(164):1–6.

Scripps Institution Of Oceanography (1986). IRIS/IDA Seismic Network.

Shahi, S. K. and Baker, J. W. (2011a). An empirically calibrated framework for including the effects of near-fault directivity in probabilistic seismic hazard analysis. *Bulletin of the Seismological Society of America*, 101(2):742–755.

Shahi, S. K. and Baker, J. W. (2011b). Regression models for predicting the probability of near-fault earthquake ground motion pulses, and their period. *Applications of Statistics and Probability in Civil Engineering*, 30(4):459.

Shahi, S. K. and Baker, J. W. (2014). An efficient algorithm to identify strong-velocity pulses in multicomponent ground motions. *Bulletin of the Seismological Society of America*, 104(5):2456–2466.

Shahi, S. K. and Baker, J. W. (2019). Erratum to An Efficient Algorithm to Identify Strong-Velocity Pulses in Multi-Component Ground Motions. *Bulletin of the Seismological Society of America*.

Shang, X., Li, X., Morales-Esteban, A., and Chen, G. (2017). Improving microseismic event and quarry blast classification using Artificial Neural Networks based on Principal Component Analysis. *Soil Dynamics and Earthquake Engineering*, 99:142–149.

Shimshoni, Y. and Intrator, N. (1998). Classification of seismic signals by integrating ensembles of neural networks. *IEEE Transactions on Signal Processing*, 46(5):1194–1201.

Shyu, J. B. H., Sieh, K., Chen, Y.-G., and Liu, C.-S. (2005). Neotectonic architecture of Taiwan and its implications for future large earthquakes. *Journal of Geophysical Research: Solid Earth*, 110(B8).

Sipkin, S. A. (1994). Rapid determination of global moment-tensor solutions. *Geophysical Research Letters*, 21(16):1667–1670.

Smith, W. H. and Sandwell, D. T. (1997). Global sea floor topography from satellite altimetry and ship depth soundings. *Science*, 277(5334):1956–1962.

Solano Hernandez, E. A. (2006). Evaluacion del efecto de sitio en estaciones de banda ancha del Servicio Sismologico Nacional.

Somerville, P. G. (2003). Magnitude scaling of the near fault rupture directivity pulse. *Physics of the earth and planetary interiors*, 137(1-4):201–212.

Somerville, P. G. (2005). Engineering characterization of near fault ground motions. In *Proc., NZSEE 2005 Conf.*

Somerville, P. G., Smith, N. F., Graves, R. W., and Abrahamson, N. A. (1997). Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity. *Seismological Research Letters*, 68(1):199–222.

Spudich, P. and Chiou, B. S. (2008). Directivity in NGA earthquake ground motions: Analysis using isochrone theory. *Earthquake Spectra*, 24(1):279–298.

SSN (2017). Servicio Sismologico Nacional de México.

Standard, B. (2005). Eurocode 8: Design of structures for earthquake resistance. *Part*, 1:1998–1.

Susan Y. Schwartz, A. N. (2009). Nicoya Seismogenic Zone.

Technological Educational Institute Of Crete (2006). Seismological Network of Crete.

Titos, M., Bueno, A., García, L., and Benítez, C. (2018a). A Deep Neural Networks Approach to Automatic Recognition Systems for Volcano-Seismic Events. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5):1533–1544.

Titos, M., Bueno, A., García, L., Benítez, M. C., and Ibañez, J. (2018b). Detection and Classification of Continuous Volcano-Seismic Signals With Recurrent Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*.

Torrence, C. and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1):61–78.

UNAVCO (2019). UNAVCO Plate Boundary Observatory database. Accessed: 2019-10-21.

University Of Washington (1963). Pacific Northwest Seismic Network.

Van Houtte, C., Bannister, S., Holden, C., Bourguignon, S., and McVerry, G. (2017). The New Zealand strong motion database. *Bull. New Zeal. Soc. Earthq. Eng.*, 50(1).

Vassiliou, M. F. and Makris, N. (2011). Estimating time scales and length scales in pulse-like earthquake acceleration records with wavelet analysis. *Bulletin of the Seismological Society of America*, 101(2):596–618.

Vigny, C., Rudloff, A., Ruegg, J.-C., Madariaga, R., Campos, J., and Alvarez, M. (2009). Upper plate deformation measured by GPS in the Coquimbo Gap, Chile. *Physics of the Earth and Planetary Interiors*, 175(1-2):86–95.

Wang, J. and Teng, T.-L. (1995). Artificial neural network-based seismic detector. *Bulletin of the Seismological Society of America*, 85(1):308–319.

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

Wang, K., Wells, R., Mazzotti, S., Hyndman, R. D., and Sagiya, T. (2003). A revised dislocation model of interseismic deformation of the Cascadia subduction zone. *Journal of Geophysical Research: Solid Earth*, 108(B1).

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

Wennerberg, L. (1990). Stochastic summation of empirical Green's functions. *Bulletin of the Seismological Society of America*, 80(6A):1418–1432.

Wu, X., Ghaboussi, J., and Garrett Jr, J. (1992). Use of neural networks in detection of structural damage. *Computers & structures*, 42(4):649–659.

Xie, J., Che, S., and Lin, Y. (2017). Earthquake disaster economic loss estimation and prediction based on neural networks with random weights. *Journal of Catastrophology*, 32(1):1–4.

Yaghmaei-Sabegh, S. (2010). Detection of pulse-like ground motions based on continues wavelet transform. *Journal of seismology*, 14(4):715–726.

Yang, D. and Wang, W. (2012). Nonlocal period parameters of frequency content characterization for near-fault ground motions. *Earthquake Engineering & Structural Dynamics*, 41(13):1793–1811.

Zeng, Y. and Chen, C.-H. (2001). Fault rupture process of the 20 September 1999 Chi-Chi, Taiwan, earthquake. *Bulletin of the Seismological Society of America*, 91(5):1088–1098.

Zhai, C., Chang, Z., Li, S., Chen, Z., and Xie, L. (2013). Quantitative identification of near-fault pulse-like ground motions based on energy. *Bulletin of the Seismological Society of America*, 103(5):2591–2603.

Zhai, C., Li, C., Kunnath, S., and Wen, W. (2018). An efficient algorithm for identifying pulse-like ground motions based on significant velocity half-cycles. *Earthquake Engineering & Structural Dynamics*, 47(3):757–771.

Zhou, Z. and Adeli, H. (2003). Wavelet energy spectrum for time-frequency localization of earthquake energy. *International Journal of Imaging Systems and Technology*, 13(2):133–140.

europass Curriculum vitae

PERSONAL INFORMATION	**Deniz Ertuncay**

Via Udine, 32, 34135 Trieste (Italy)

(+39) 339 339 13 78

dertuncay@gmail.com

http://web.units.it/dottorato/esfm/en/content/ertuncay-deniz
https://scholar.google.com/citations?user=OrB7o4IAAAAJ&hl=en   https://github.com/dertuncay
Skype dertuncay

Sex Male | Date of birth 04/05/1990 | Nationality Turkish

EDUCATION AND TRAINING

03/09/2008–21/01/2013	**Bachelor of Engineering** EQF level 6

University of Istanbul, Istanbul (Turkey)

Thesis: Analysis of Earthquake Hazard and Perceptibility Study in Saros Bay, NW Turkey
Supervisor: Prof. Dr. Oguz Ozel
- I was the head of Geophysics Student Club for 1 year.

GPA: 2.86/4.00

01/2014–16/06/2016	**Master of Science** EQF level 7

Bogazici University, Istanbul (Turkey)

Thesis: Variation of Source Parameters in Anisotropic Focal Region
Supervisor: Asst. Prof. Dr. Cagri Diner
Description: Investigation of effect of anisotropic focal regions on moment tensor decomposition.
- I was a part of a scientific project sponsored by Scientific and Technological Research Council of Turkey (TÜBİTAK) with a name 'Determination of Finite Fault Parameters with its Residual Stress Tensors'
- I attended Inverse Theory course from PhD. program.

GPA: 3.28/4.00

01/11/2016–Present	**Philosophy of Doctorate** EQF level 8

University of Trieste, SeisRaM Working Group, Trieste (Italy)

Thesis: Temporal and spatial analysis of near fault stations in terms of impulsive behavior
Supervisor: Prof. Dr. Giovanni Costa

Description: I am focusing on impulsive signals and their spatial and temporal distributions. I am using wavelet analysis and machine learning algorithms to identify. Moreover, I am trying to find the probability of occurrence of impulsive signals by using Multivariate Naive Bayes Method. In order to do that, I have correlated various parameters that are carrying information from the earthquake physics or spatial features of the seismic station.
- I am also part of an European Union project called ARMONIA. In WP4 of ARMONIA project, I am testing different seismic instrumentation installed in Foligno, Italy.

WORK EXPERIENCE

07/2011–07/2011	**Intern**

Bogazici University Kandilli Observatory and Earthquake Research Institute, Istanbul

(Turkey)

I worked on earthquake monitoring service for a month. I located earthquakes on my shifts. Other than that I helped preparing seismic bulletins. I took practical courses on linux systems and GIS.

09/2012–06/2013	**Geophysicist**

Ilkon Engineering, Istanbul (Turkey)

Non destructive tests on buildings.

02/2014–02/2016	**Reserve Board Member**

Chamber of Geophysical Engineers of Turkey, Istanbul (Turkey)

02/2016–02/2018	**Board Member**

Chamber of Geophysical Engineers of Turkey, Istanbul (Turkey)

12/02/2019–Present	**Scientist**

ARMONIA project, Trieste (Italy)

I am part of Work Package 4 of ARMONIA project. The main goal is the improvement and the integration of existing network to enhance the effectiveness in a trasnational prospective, in order to obtain the required quality of the data together with the maximum distribution of the instruments in the territory. The deployment of new instrumentation in near field and in sentinel buildings provides to high level of efficiency adding important information in case of emergency.

PERSONAL SKILLS

Mother tongue(s)	Turkish

Foreign language(s)

UNDERSTANDING		SPEAKING		WRITING
Listening	Reading	Spoken interaction	Spoken production	
English				
C1	B2	B2	B2	B2
TOEFL-IBT 89				
Italian				
A1	A1	A1	A1	A1

Levels: A1 and A2: Basic user - B1 and B2: Independent user - C1 and C2: Proficient user
Common European Framework of Reference for Languages

Organisational / managerial skills	I am a co-supervisor of a Master's student, Antonino Caponnetto, at University of Trieste, Department of Mathematics and Geosciences.
Job-related skills	MathWorks Matlab, Seismic Analysis Code (SAC), Linux, Latex, Python (Numpy, Obspy, Matplotlib, Keras, scikit-learn), Latex
Driving licence	AM, B1, B, D1

ADDITIONAL INFORMATION

Memberships	Chamber of Geophysical Engineers of Turkey (CGET)
	European Geosciences Union (EGU)
	American Geophysical Union (AGU)

Publications

1. Ertuncay, Deniz and Cagri Diner. "Variations of Source Parameters Due to Anisotropic Focal Region" *Bulletin of Geophysics (Jeofizik)* (2016) (Turkish): 18, 36-47. ISSN:0259-1472.

2. Ertuncay, Deniz, and Giovanni Costa. "An alternative pulse classification algorithm based on multiple wavelet analysis." *Journal of Seismology* (2019): 1-14.

Presentations

Poster Presentations:

1. Ertuncay, Deniz, and Çagrı Diner. "Variations in the Fault Plane Solutions and in the Magnitudes of Earthquakes Occurring in the Anisotropic Region." *EGU General Assembly Conference Abstracts*. Vol. 18. 2016.

2. Ertuncay, Deniz, Blaž Vičič, and Giovanni Costa. "Seismicity of Eastern Alps and North western Dinaric Alps." *EGU General Assembly*. EGU General Assembly 2017, 2017.

3. Ertuncay, Deniz, and Giovanni Costa. "Identification of Possible Velocity Pulses in Earthquake Near Fault Regions by Using Machine Learning." *EGU General Assembly Conference Abstracts*. Vol. 20. 2018.

4. Ertuncay, Deniz, and Giovanni Costa. "An Alternative Pulse Classification Algorithm Based on Multiple Wavelet Analysis" *37th GNGTS Meeting.* Bologna, Italy, 2018.

5. Ertuncay, Deniz, and Giovanni Costa. "Identification Of Pulse Shape Signals On Near Fault Stations With Convolutional Neural Network Algorithms: Preliminary Results." *New Trends in Geophysics and Engineering International Symposium*. Istanbul University-Cerrahpasa Engineering Faculty Department of Geophysical Engineering, 2018.

6. Ertuncay, Deniz, and Giovanni Costa. "Identification of Possible Velocity Pulses in Earthquake Near Fault Regions by Using Machine Learning Algorithms" *Machine Learning in Bosporus - Ismail Ari Summer School 2018: Speech, Language Processing and Bioinformatics.* University of Bogazici, 2018.

7. Ertuncay, Deniz, et al. "Identification of Near Fault Pulse Shaped Signals With Machine Learning Algorithms." *EGU General Assembly*. EGU General Assembly 2019, 2019.

8. Ertuncay, Deniz, et al. "ARMONIA Project: Transfrontier Strategy in the Management of Earthquakes." *27th IUGG General Assembly*. IASPEI, 2019.

Oral Presentations:

1. Ertuncay Deniz. "Identification of Near Fault Pulse Shaped Signals with Wavelet Analysis and Machine Learning Algorithms", *Antelope User Group meeting 2019*, Taormina, Italy, 2019.

2. A. Gallo, G. Costa, L. Tiberi, L. Filippi, D. Ertuncay et al. "Improvement for Near Real Time Accelerometric Data Analysis. Case of Study: Italian Strong Motion Network", *27th IUGG General Assembly*. IASPEI, 2019.

Conferences

1. EGU General Assembly 2017 in Vienna, Austria on 23 - 28 April 2017

2. European Antelope User Group Meeting 2017 in Vienna, Austria on 29 - 31 May 2017

3. PSHA Workshop Future directions for probabilistic seismic hazard assessment at a local, national and transnational scale in Lenzburg, Switzerland on 5 - 7 September 2017

4. Social Media and Science Communication in Trieste, Italy on 11th September 2017

5. A road-map for geology based seismic hazard models: Current practices, advantages, limitations, and what still lies ahead in Ljubljana, Slovenia on 21st October 2017

6. Seismology: facts, fantasy and the just plain weird In memory of Marco Mucciarelli in Trieste, Italy on 13th November 2017

7. Technical Workshop on Internet Macro-seismology in Ljubljana, Slovenia on 14 - 15 November 2017

8. EGU General Assembly 2018 in Vienna, Austria on 8 - 13 April 2018

9. Machine Learning in Bosporus - Ismail Ari Summer School 2018: Speech, Language Processing and Bioinformatics in Istanbul, Turkey on 2 - 5 July 2018

10. New Trends in Geophysics and Engineering in Istanbul, Turkey on 7 - 9 November 2018

11. 37th Gruppo Nazionale di Geofisica della Terra Solida in Bologna, Italy on 19 - 21 November 2018

12. ARMONIA Group Meeting in Venice, Italy on 11 - 12 February 2019

13. EGU General Assembly 2019 in Vienna, Austria on 7 - 12 April 2019

14. Seismotectonics of the Eastern and Southern Alps and their transition to the Dinarides in Trieste, Italy on 8 - 9 May 2019

15. Antelope User Group Meeting in Taormina, Italy on 23 - 30 May 2019

16. Machine Learning Crash Course 2019 in Genova, Italy on 17 - 21 June 2019

17. 27th General Assembly of The International Union of Geodesy and Geophysics in Montreal, Canada on 8 - 18 July 2019

18. SERA European Seismic Risk Model Workshop and OpenQuake-engine Training Workshop in Istanbul, Turkey on 12 - 13 September 2019

19. ARMONIA Group Meeting in Udine, Italy on 18 - 19 September 2019