UNDERWATER EARTHQUAKE CHARACTERIZATION BY ACOUSTIC RADIATION ANALYSIS

Bernabe Gomez Perez

A thesis submitted for the degree of

Doctor of Philosophy

School of Mathematics

Cardiff University

December 2021

Dedicated to my parents

Bernabe and Maria Teresa.

Abstract

A common cause of tsunami waves in the ocean is underwater seismic events, which also generate acoustic radiation. Three fundamental aspects of the introduced physical process are that the sound-induced by underwater tectonic events carries information about its source, it can be recorded by distant hydrophones and it travels much faster than tsunamis in the ocean. To predict tsunami propagation, it is first necessary to have reliable data relating to the rupture characteristics such as uplift speed, duration, fault geometry and epicentre location. In this thesis, particular attention is given to the development of a semi-analytical inverse approach that can be employed in near real-time on acoustic signals, assuming the fault is single, slender and uniform and the seabed is flat. Moreover, a methodology to apply the presented model to real hydrophone recordings was developed. To infer the remaining characteristics (slip type and magnitude of the event) of the studied recorded events, an algorithm that combines digital signal processing techniques with machine learning techniques has been produced and tested by analysing hydrophone recordings of 201 tectonic events, located in the Pacific and the Indian Ocean. Finally, tectonic event scenarios reported by NOAA to have triggered tsunamis were chosen to be replicated using the inverse problem model solutions. The earthquake parameters retrieved by the inverse problem model were used as input to the COMCOT numerical model which, in turn, output surface wave elevations (tsunami) to be validated against DART buoy data and tide gauges.

Acknowledgements

I would like to thank my supervisor, Dr. Usama Kadri, for this opportunity and for his time and help during my PhD stage. I am grateful to all the PhD colleagues, academics and staff members in the School of Mathematics at Cardiff University for making me feel accepted, especially to Hassan, Bertrand, Byron and Mohammed who are not only great professionals but also good friends.

My biggest gratitude goes to all my family, without my parents Maria Teresa Perez and Bernabe Gomez, my brother Miguel Gomez and my grandparents Faustino Perez (Tino) and Fermina Gutierrez (Chiqui), this project would have not been completed, their infinite support has motivated me even on the hardest moments.

I am grateful to CTBTO and Canada Ocean Network for providing the analysed hydrophone data. In addition, I would like to thank Dr. Usama Kadri for providing MATLAB functions that made possible the development of the inverse problem model for acoustic waves. This work has been supported by the Engineering and Physical Sciences Research Council EP/N509449/1.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

A considerable fraction of the human population inhabits the coasts of the World [114] and a big part of these areas are at risk of tsunamis, which can be catastrophic. It has been estimated by the United Nations that about 60000 people and 4 billion USD in assets are exposed every year to tsunami hazards [25]. Tsunamis are large and destructive water waves that can be triggered by earthquakes, landslides or other abrupt disturbances as a response to a sudden vertical displacement of a large volume of water. Tsunami waves have very long wavelengths, which range from tens to hundreds of kilometres and their period can vary from minutes to hours. In addition, tsunamis are hard to detect in deep waters, where they travel very fast (they can reach speeds up to 200 ms^{-1}) and their size is only a few centimetres high. However, they increase their wave height as they reach shallower waters, this process is called shoaling. Fault movements responsible for submarine earthquakes causing a vertical movement of the seafloor over a large area are one of the most common causes of tsunami waves.

We can find clear examples of the destructive nature of tsunami waves triggered by distant underwater earthquakes in recent history, e.g. Chile 22 May 1960, Sumatra 26 December 2004 and Tohoku 13 March 2011. On the 22 May of 1960, an earthquake with an estimated magnitude of 9.5 M_w was detected, located off the coast of Chile, which

triggered a tsunami that radiated out across the Pacific ocean [65]. The tsunami led to approximately 800 fatalities in Chile, 60 in Hawaii and 200 in Japan and had a maximum recorded coastal run-up in Chile of 25 m [55]. The damages in Chile were estimated at 550 million USD, in Hawaii at 24 million USD, and at approximately 1 million USD on the United States west coast. Moreover, the waves were reported to be more than 6 m high in Japan, causing damages estimated at 50 million USD [55].

A more recent case of a distant triggered tsunami is the 26 December of 2004 great Sumatra earthquake. In less than three hours from the earthquake eruption time, a tsunami devastated the coasts of Thailand, Sri Lanka, and India. The death toll reached approximately 250000 people and around 1.7 million people became homeless [12]. Moreover, the tsunami killed a few hundred people in Somalia, which is situated on the other side of the Indian ocean and where the waves arrived several hours after the earthquake happened [4]. Within five hours of the tectonic event, the Harvard Centroid Moment Tensor (CMT) project inferred a magnitude of 9.0 M_w, resulting in a high risk of tsunami [12]. M_w is considered the most appropriate estimate to assess tsunami potential, nevertheless, it is the most difficult to determine rapidly, because most seismological techniques are sensitive to the shorter period components [12]. In the case of the tsunami associated with the 2004 Sumatra tectonic event, the lack of warning was attributed to the absence of a reliable tsunami warning system for the Indian Ocean. Moreover, inadequate remote sensing technology and analysis systems for large earthquakes and the resulting tsunamis was a contributing factor [74].

Another remarkable case study of a devastating tsunami in recent history is associated with the tectonic event that occurred on the 11 March of 2011 off the Pacific coast of Japan. The reported moment of magnitude for this event was 9.1 M_w. The first waves of the so-called Tohoku earthquake-induced tsunami hit nearby shorelines within 30 minutes of the beginning of the tectonic event [74]. The tsunami and earthquake caused about 20000 fatalities or missing people and the total caused damages were estimated at 200 billion USD [74, 139, 102]. Furthermore, the tsunami induced the failure of the Fukushima nuclear reactor, which continues to be an environmental disaster [74].

It is emphasised by the scientific community the need for a more rapid and accurate assessment of earthquake source characteristics [4] since many fatalities and a big part of the damage associated with the introduced tsunami scenarios (and other tsunami cases) could have been mitigated with reliable tsunami early warning systems. However, developing reliable and efficient tsunami early warning systems offers serious challenges, tsunamis can be triggered by various types of sources such as underwater earthquakes, impacting objects on the surface of the sea, landslides or volcanoes; the occurrence of earthquakes is unpredictable; not all earthquakes produce tsunamis (tsunamigenic potential); Tsunamis travel at extremely high speeds in open waters, where they are hard to detect exhibiting small amplitudes; for tsunamis, the rapidity needed to assess the risk implies that the real-time estimates of the source are often affected by large uncertainties [4]; and the capability of the warning systems to reach people at risk is limited. Tsunami warning systems have to be capable to form a collaborative communication infrastructure of distributed tsunami warning systems to implement monitoring and warning strategies [128].

Tsunami warning systems should focus on seismically generated tsunamis since they are the most common source of tsunamis in the oceans [4], once an earthquake happens the efficient assessment of the source characteristics is crucial to evaluate the risk of tsunami and its potential size. Current tsunami detection systems include various techniques such as (1) DART buoys, based on sea-level changes, where the tsunami has to physically reach the location of the measuring devices [10] (National Oceanic and Atmospheric Administration, NOAA); (2) satellite-based sea surface roughness measurements, that require extensive data processing [47]; (3) seismic recordings, that can shed light on the epicentre and the strength of the earthquake [67]; (4) cabled observatories, cable-based systems that use the same sensor as DART buoys, thus measurements are compatible [9]; and (5) Differential GPS buoys, buoy-based technology with non-standard sensor, measurements are not compatible with DART and cabled observatories sensors [9]. Some countries like Russia use earthquake moment magnitude criterion for tsunami hazard estimation, resulting in a high rate of false alarms [104].

It has been established that it is possible to detect tsunami waves in near real-time through analysis of hydrophone recordings [82]. This creates the opportunity for detecting tsunamis using hydrophone networks, which can be located below the sea surface reducing the risk of damage or loss of the sensor compared with those located at the surface [38]. The sudden vertical motion that has the potential to trigger tsunamis is also associated with a slight compression of the water layer that generates compression-type waves, known as acoustic-gravity waves (AGWs) [135]. AGWs travel at the speed of sound in water, $c \simeq 1500$ ms^{-1}, which far exceeds the maximum tsunami phase speed. For example, in a water layer of depth $h = 2000$ m, a tsunami would propagate at a speed that is more than 10 times slower than the phase speed of acoustic-gravity waves (at a speed approximately of $\sqrt{gh}$, where g is the gravitational acceleration). Thus, acoustic-gravity waves have been proposed for early warning systems, as originally noted by Yamamoto (1982) [135].

In this thesis, we propose a set of techniques, conceived as complementary to already existing early tsunami warning technology, capable to analyse hydrophone recordings and evaluate the effective eruption dynamics and fault properties of tectonic events in almost real-time. The theory behind some of the techniques here developed is based on the fact that tsunamis in the ocean are often generated by submarine earthquakes that can be approximated as slender rectangular shapes [82] [63].

1.2 Background

At this point, it is clear that the generation of tsunamis by earthquakes in the ocean is intrinsically linked to the generation of sound. Acoustic waves have been studied for many decades, but the field of acoustics was not widely accepted by the scientific community until the 19th century when Rayleigh and Reynolds used ray concepts to explain acoustic phenomena. They studied the sound generated by vibrating strings and organ pipes.

Sound is a mechanical compression wave that travels through a medium (solid, liquid,

or gas). Human hearing can detect a limited range of frequencies. Thus, sounds are commonly divided into three groups based on the frequency range: infra-sound (0.001-16 Hz), audible sound (16 Hz-20 kHz) and ultra-sound (above 20 kHz). A key property of sound is that it moves at a faster speed in water than in air because of differences between the mechanical properties of both media. The sound speed in the ocean is an increasing function of temperature, salinity, and pressure, but even so, the sound is often approximated to a constant, $c = 1500ms^{-1}$ which correction is usually less than 2 %. Sound waves behave like longitudinal waves in liquids or gases, travelling with periodic compressions and rarefactions that are in the same direction as the propagation. In acoustics, the amplitude is the maximum pressure of the wave, measured in Pascals [Pa].

A key tool for the study of acoustic waves is the Fourier series. They are based on the idea that it is possible to construct any waveform by superimposing a sufficient number of simple harmonic waves with chosen frequencies and amplitudes. Nevertheless, some conditions have to be met before the Fourier series representation of an actual wave can exist: (1) the wave has a finite number of discontinuities, maxima, and minima in any period; (2) the wave has to be integrable in any period [17].

Acoustic Gravity waves

The importance of the compressibility of the water in tsunami generation models has been studied in the literature, revealing the existence of AGWs [93, 94]. As mentioned before, these wave modes have a great potential for serving as tsunami precursors due to their high propagation speeds.

Back in 1950, Ref. [78] described the non-linear interaction of two opposite ocean waves in a compressible fluid, which was observed to result in a compression wave. A few years later in 1954, Ref. [86] described the generation of tsunamis by a disturbance, which may occur at the surface of the sea and oscillates in a compressible water layer under the assumptions of the rigid sea bottom and uniform water depth and density (constant speed of sound in all domain). With this research, the foundation for the importance of the AGWs was laid [86], providing a more accurate description of gravity waves and allowing

the propagating modes that emerge from the compression waves. Ref. [111] developed a model in 1965, where an infinite length strip, situated on the bottom of a slightly compressible ocean is uplifted generating both gravity and acoustic waves, once again the need for the inclusion of the compressibility of the water in the tsunami generation models was made patent. In a similar fashion, in 1982 the two-dimensional problem for the generation of gravity and acoustic waves by the motion of a submarine block situated on a rigid sea-bed under a compressible water layer was solved by Ref. [135]. In this study, the block oscillates vertically inducing both gravity and acoustic waves. It is remarkable that the assumption of incompressible fluid results in longer wavelengths for the gravity wave and permits no propagating acoustic waves. Furthermore, neglecting the gravitational effects can lead to lower cut-off frequencies for the acoustic modes. Additionally, for a specific frequency, some of the acoustic modes decay very fast (named evanescent modes), thus, they can be neglected in the far-field [135].

Another relevant piece of work was produced in 1999 by Ref. [93], where the generation of tsunamis in a compressible ocean was described and compared against results without the consideration of the water compressibility. It was emphasized in the same study the relevancy of water compressibility in the tsunami generation and propagation stages to achieve an accurate description of the process. The concept of gravito-acoustic waves was then introduced, accepting that gravity can play an important role in the description of acoustic waves. In Ref. [94] near-bottom pressure records collected by the Japanese Agency for Marine–Earth Science and Technology (JAMSTEC) gages during the Tokachi-Oki seismic event in 2003 were studied. These records revealed very low frequency elastic vibrations of the water column. Moreover, the low-end of the frequency spectrum associated with the recorded signals was observed to be correlated with the maximum ocean depth in the region of the tsunami source. On the other hand, the authors noted that the high-end of the frequency spectrum was defined by the maximum duration of the bottom deformation. It was not until 2009 that Ref. [22] studied the propagation of acoustic waves, generated by the vertical motion of the sea-floor, over a porous sedimentary layer, revealing that the porous sediment layer can attenuate both

the tsunami and acoustic waves, acting as a low-pass filter for the acoustic waves.

In 2010 Ref. [115] derived an analytical solution for the two-dimensional problem of an infinite length strip that uplifts with constant speed and generates compression waves. The first few generated AGWs (these modes have typical values of 10 s of period and 40 km of wavelength) were observed to have wave-lengths that exceeded the width of the SOFAR channel, hence they cannot be trapped in it. For AGWs, the whole ocean depth serves as a waveguide. Additionally, it was shown that, since tsunamis propagate much slower than AGWs, for a great distance between the source and the observation point, the AGWs are decoupled from the gravity waves and they can be studied separately [115]. AGWs were then reported to leave measurable bottom pressure signals in the far-field that can be recorded by hydrophones [115].

Ref. [61] in 2012 studied the behaviour of propagating AGWs over a sea bottom with a step, which simulates a shelf break, in a two-dimensional domain, observing that at the point of the step, part of the energy carried by the AGWs is transmitted and part of it reflected due to the variation in depth. This verifies that the existence of AGW modes depend on the depth of the ocean and the propagating frequency. In 2013, Ref. [48] extended the description of the generation of AGWs to a three-dimensional domain, where the motion of a cylindrical shaped disturbance in a compressible water layer was analysed and an inverse theory capable to retrieve the earthquake characteristics from the calculated acoustic signals in the far-field was proposed but, at a high computational cost [48]. Further studies about the behaviour of AGWs were published in the same year, which emphasizes the growing attention from the scientific community to this field [30, 8, 7, 62].

AGWs generation and propagation analysis have additionally been approached by numerical solutions in the literature, such as in the works of Refs. [20, 1] where the reproduction of historical catastrophic tsunamigenic earthquake scenarios was performed, which allowed the consideration of real bathymetry. Ref. [109] in 2014, studied pressure changes in the ocean surface, which were reported to produce compression type waves that carry information on the source to distant locations [2], thus, can act as early

precursors of surface gravity waves, being proposed for coastal flooding warning systems [109]. Later works further studied the behaviour of AGWs, such as Ref. [56] in 2015, who extended the solution for the non-linear interaction of two opposite waves in a compressible fluid, to include the general interaction of waves of different wavelengths and Ref. [60] in 2016, where an asymptotic theory for resonant triad interactions between acoustic-gravity waves was developed. Later works would further tackle the study of resonant conditions for AGWs [136, 64]. It was also shown in 2016 that an AGW interacting with a surface-gravity wave can generate a second AGW [57]. In 2017, an inverse problem model to retrieve the location of ocean disturbances was presented [63], in this study several underwater acoustic signals recorded by hydrophones were analysed and compared against acoustic signals obtained from experiments of spheres impacting on the surface of a water tank, finding strong similarities. Additionally, Ref. [58] proposed in 2017 the possibility of gravity waves mitigation and energy redistribution by their interaction with AGWs.

Finally, in Ref. [82] (2018), a new approach to the generation of tsunamis in a compressible ocean was developed by using multiple scales theory, which allows the description of the generation of acoustic waves more efficiently. This approach is based on the assumption that earthquakes in nature often have slender geometry. It is remarkable that in 2019, it was reported that AGWs can penetrate through the sea-bottom, leading to a modification of the propagation speed and arrival times, especially in the far-field [59]. In the same year Ref. [3] utilised a semi-analytical model to study the effects of water compressibility, elasticity and gravity effects on tsunami phase speed, finding speed propagation differences against the models that neglect the mentioned effects and concluding that for a more realistic description of the process elasticity has to be included in the description of AGWs propagation.

More recent studies tackle other simplifications taken in the past, such as homogeneous speed of the sound in the water layer [83], the consideration of varying sound speed vertical profiles leads to discrepancies in the dynamic pressure distributions of the acoustic waves. Moreover, Ref. [132] studied multi-fault scenarios accounting compressibility and

gravity effects [132], which allows the analysis of more complex scenarios.

Instrumentation and data

To analyse sound waves from underwater tectonic events, first, there is the need to record and digitalize them. Hydro-acoustic monitoring involves recording signals that show changes in water pressure generated by sound waves. In this thesis, data collected by The Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) hydrophones have been studied. CTBTO (established in 1996) is an international organization that monitors nuclear weapons testing and encourages the disarmament thereof. The International Monitoring System (IMS), which belongs to the CTBTO, is a worldwide network comprised of monitoring facilities, which consist of 50 primary seismic stations, 120 auxiliary seismic stations, 11 hydroacoustic stations (6 hydrophone stations and 5 T-phase stations), 60 infrasound stations and 80 radionuclide stations.

To detect acoustic energy in the oceans, the CTBTO network has 11 IMS hydroacoustic stations, from which only six stations are equipped with hydrophones. An IMS hydroacoustic station can consist of either ocean deployed hydrophone sensors and data acquisition systems (hydrophone station) or one or more island-deployed seismometer sensors and data acquisition systems (T-phase station).

Hydrophone

To understand the functioning of a hydrophone it is first necessary to introduce what a microphone is. A microphone is a device that converts acoustical energy into electrical energy, hydrophones are microphones designed to work underwater. When the hydrophone station is located on an island, two distinct cables and hydrophone sensors are deployed off opposite shores. The hydrophone sensors are placed at the axis of the Sound Fixing and Ranging (SOFAR) channel, minimum sound speed channel on the ocean, using a subsurface float and an ocean-bottom anchor. Each cable has three sensors with wet-end digitizers. The digital signals are transmitted to the shore facility via fibre

optic cables for processing and transmission to the International Data Centre (IDC), Vienna. To provide the station with some directional capabilities, the three hydrophones are placed in a triangular configuration and each sensor is separated horizontally by a distance of approximately two kilometres. A single cable is used to bring the signals from the hydrophone sensors to shore.

1.3 Outline of this thesis

This thesis has been structured in five chapters plus their corresponding supplementary materials presented in Appendix A. The present chapter, Chapter 1, introduces the studied problem and offers the background related to the research carried out.

Chapter 2, details the development of an inverse problem model for acoustic waves and its further application on synthetic and real acoustic signals. Moreover, a sensitivity analysis was carried out for the model, which is detailed in Appendix A. In Chapter 3, a dataset built with acoustic signals extracted from the CTBTO database is analysed with the purpose of training and validating machine learning algorithms to infer the type of slip and magnitude of the associated tectonic events, these parameters can be used as input for the inverse problem model developed in Chapter 2. Further analysis on the machine learning algorithms and feature extraction techniques is provided in Appendix A.

Chapter 4 shows the validations carried out with a numerical tsunami propagation model (COMCOT) for the solutions provided by the inverse problem model, where various past earthquake scenarios that triggered tsunamis were replicated and the solutions obtained were compared against real measurements in deep and shallow waters. The conclusions extracted from this thesis, the related discussion and the potential future steps that can be taken regarding the approached lines of research are given in Chapter 5. Additionally, in Appendix A the solutions for the application of the inverse problem model to 112 underwater vertical earthquakes are provided.

1.4 Scientific contributions

This thesis has made novel contributions across various fields. A list of the works already published that sprang from this thesis is provided below:

- Bernabe Gomez and Usama Kadri. "Near real-time calculation of submarine fault properties using an inverse model of acoustic signals." Applied Ocean Research 109 (2021): 102557.
- Bernabe Gomez and Usama Kadri. (2021). "Earthquake source characterization by machine learning algorithms applied to acoustic signals." Scientific Reports, 11(1), 1-15.

Manuscripts in preparation:

- Bernabe Gomez and Usama Kadri. "Numerical validation of effective slender fault source solution for past tsunami scenarios." Based on Chapter 4.

Conference posters:

- Bernabe Gomez and Usama Kadri. "Acoustic Waves of Tsunami: Inverse Approach." EGU General Assembly Conference Abstracts. 2018.
- Bernabe Gomez and Usama Kadri. "Inverse problem model for acoustic waves". 10th Oxford SIAM-IMA Student chapter conference poster presentation, 1st price.
- Bernabe Gomez and Usama Kadri. "Acoustic Waves of Tsunami: Inverse Approach". SIAM UKIE Annual Meeting, 2018.
- Bernabe Gomez and Usama Kadri. "Towards an early tsunami warning system: Enhanced inverse solution of acoustic radiation from a slender fault". South China Sea Tsunami Workshop SCSTW-10, Singapore, 2018.
- Bernabe Gomez and Usama Kadri. "Near Real-time Calculation of Submarine Fault Properties". South China Sea Tsunami Workshop SCSTW-11, Hangzhou, 2019.

Conference talks:

- "Inverse theory for tsunami acoustic waves". SIAM-IMA Student Chapter three minute thesis, 2018.
- "Near real-time calculation of submarine fault properties using an inverse model of acoustic waves". Welsh Mathematics Colloquium, 2019.
- "Slender fault inverse problem model of acoustic waves". Underwater Acoustics International Conference and Exhibition (UACE2021).

Chapter 2

Calculation of underwater fault properties from acoustic signals

2.1 Introduction

[1]As mentioned in Chapter 1, it is believed that AGWs can leave measurable bottom pressure signals far from the fault [115, 61], which allows them to be recorded by distant hydrophones [99] providing insight into frequency components of both the tsunami and the AGWs even at the low end of the spectrum [101].

Ref. [48] developed a standard inverse approach capable of retrieving some of the source characteristics from pressure recordings by employing a piston model with cylindrical geometry and a three-dimensional integral description of the pressure field. However, such a model requires extensive computations and applying it in early warning systems is rather challenging. To overcome this difficulty, slender body theory can be employed to derive a closed-form analytical solution of the three-dimensional pressure field [82], which allows near real-time inverse calculations. The model implies several simplifications, (1) gravity effects were neglected since gravity and acoustic modes are virtually decoupled after a long propagation distance (see Ref. [82] and references within);

[1]This chapter is based on the work that was published as an article in Applied Ocean Research journal (ElSevier) [43].

(2) the fault is assumed to be slender with width $2b$ and length $2L$, see Fig. 2.1, which is a reasonable assumption as illustrated in Table 1 of Ref. [82]; (3) the fault moves vertically upwards with a constant speed (W_0) for a time duration ($2T$); (4) the water depth h is considered constant, which is acceptable as long as the characteristic length-scales of the sea bed topography are several orders of magnitude smaller than the water depth (i.e. in deep ocean far from deep tranches or shelf breaks); (5) the sea floor being rigid is a fine assumption considering that the water layer is deep enough for the analysed acoustic mode, otherwise the phase speed of the mode would change dramatically and the elasticity of the sea floor needs to be considered as illustrated in Fig. 2a of Ref. [30].

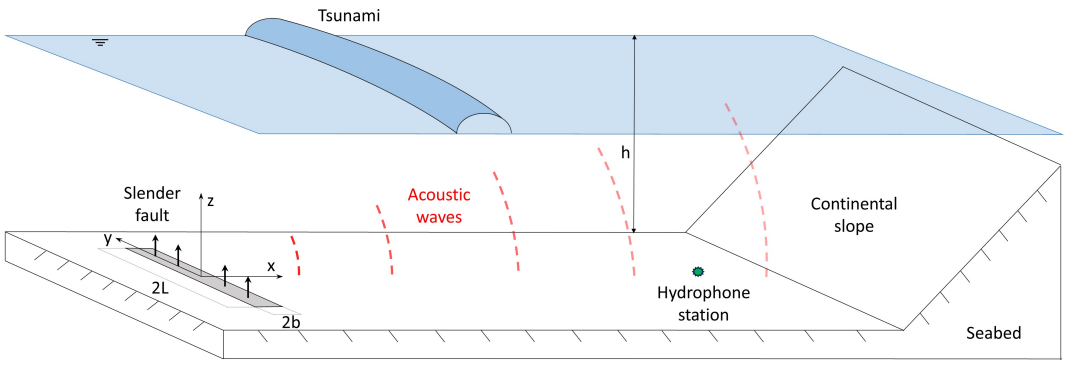

Figure 2.1: Generation of tsunami and acoustic waves by an uplifting slender fault, leaving measurable pressure signals at distant locations.

In this chapter, we developed a probabilistic inverse problem model based on the slender fault solution [82] and the model developed by Ref. [63]. Since the model is based on an analytical solution comprising trigonometric functions, some properties may have multiple solutions that are out of the usual range. To avoid such solutions, we confine the range of the fault properties following the literature. As a range guideline, we consider

fault approximations found in the literature [71, 11], which can be correlated with the pressure signal [21], and the moment magnitude of the event [131]. It is worth noting that, the slender fault geometry and the dynamics considered here are not a simplification of the earthquake dynamics itself, but rather represent an effective vertical motion caused by the much more complicated rupture dynamics. Thus, the horizontal component of the rupture is ignored, e.g. while a rupture can last, say, for tens of minutes, the effective uplift we consider here may only have a duration of a few seconds. Note that effective rupture properties for tsunami generation have been studied extensively, in particular in major submarine earthquakes. For example, the rupture length in the case of the 26 December 2004 Sumatra earthquake was approximately 900 km long [35]; the effective surface vertical uplift of the 22 May 1960 Chile earthquake was about 2 m [88]; and the effective vertical uplift of the Maule 27 February 2010 Chile earthquake was 3-5 m and its width 20-50 km [79].

2.2 Background

Acoustic radiation from slender fault, analytical solution

The propagation of acoustic-gravity waves in a slightly compressible fluid is governed by the three-dimensional wave equation [115]

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}, \tag{2.1}$$

where ϕ is the velocity potential, the velocity field is defined by $\mathbf{u} = \nabla\phi$ and c is the speed of sound in water. x, y and z define the three-dimensional cartesian coordinate system, as shown in Fig. 2.1 and t stands for the time dimension.

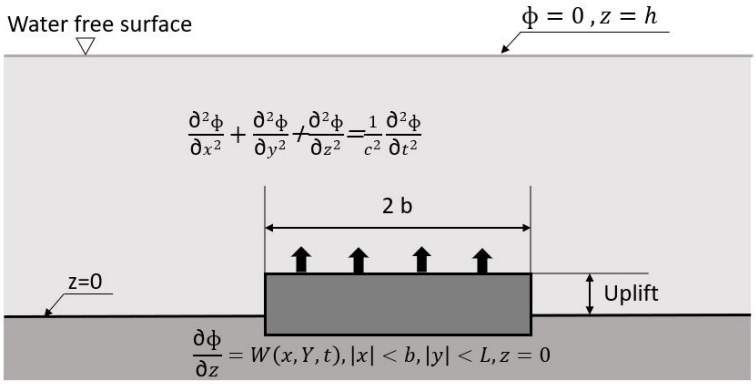

Figure 2.2: Profile sketch of the uplifting slender fault.

The partial differential equation (PDE), Eq. (2.1), is built from the equations of conservation of mass and momentum under the assumptions of irrotational flow, which makes the vorticity 0. Another assumption taken is that water is considered here an inviscid flow.

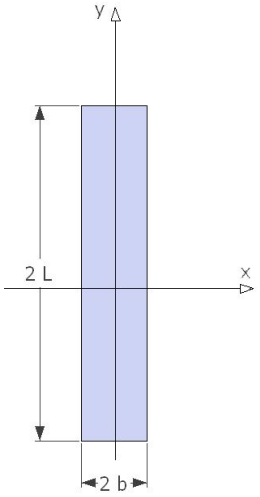

Figure 2.3: Top view of the slender fault.

Since gravity effects are neglected, at the free surface ($z = h$), see Fig. 2.2, the pressure is assumed to be uniform and zero

$$\phi = 0 \quad \text{at} \quad z = h, \tag{2.2}$$

and thus we can confine our analysis to pure acoustic signals, neglecting effects of gravity. On the seabed, located at $z = 0$, a piston model simulates the vertical displacement of the fault by

$$\frac{\partial \phi}{\partial z} = \begin{cases} W_0 \tau(t) & |x| < b,\ |y| < L \\ 0 & \text{elsewhere} \end{cases}, \qquad \tau(t) = \begin{cases} 1 & -T < t < T \\ 0 & |t| > T \end{cases} \tag{2.3}$$

Multiple scale coordinates are introduced, x, z, $X = \epsilon^2 x$, $Y = \epsilon y$, where $\epsilon = b/L \ll 1$ is the slenderness parameters. T is the duration of the disturbance and W_0 is the constant uplift speed. The earthquake is assumed to have a rectangular slender shape, see Fig. 2.3, with a total length of $2L$ and width $2b$ and b is supposed to be much smaller than L

Multiple scales theory is then introduced to the coordinates:

$$x, \quad z, \quad X = \epsilon^2 x, \quad Y = \epsilon y. \tag{2.4}$$

Once the multiple scales coordinates are applied to the wave equation, Eq. (2.1), the following expression is reached:

$$\left(\frac{\partial^2}{\partial x^2} + 2\epsilon^2 \frac{\partial^2}{\partial x \partial X} + \epsilon^2 \frac{\partial^2}{\partial Y^2} + \frac{\partial^2}{\partial z^2} \right) \phi = \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2}, \quad 0 < z < h. \tag{2.5}$$

From Eq. (2.5) with the leading order terms, we have the description of the phenomenon in two dimensions,

$$\frac{\partial^2 \phi_0}{\partial x^2} + \frac{\partial^2 \phi_0}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 \phi_0}{\partial t^2} = 0, \quad 0 < z < h. \tag{2.6}$$

And at $\mathrm{O}(\epsilon^2)$ the slowly varying envelope

$$\frac{\partial^2 \phi_2}{\partial x^2} + \frac{\partial^2 \phi_2}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 \phi_2}{\partial t^2} = -\left\{ \frac{\partial^2 \phi_2}{\partial Y^2} + 2 \frac{\partial^2 \phi_2}{\partial x \partial X} \right\}, \quad 0 < z < h. \tag{2.7}$$

In order to obtain an analytical solution for the pressure using the above PDEs, double Fourier transform is applied, in the domain of the time to frequency and in the domain of space (x) to the wavenumber domain, Φ is the potential transformed twice.

$$\Phi = \int_{-\infty}^{\infty} e^{i\omega t}\bar{\phi}dt, \quad \bar{\phi} = \int_{-\infty}^{\infty} e^{-ikx}\phi dx. \tag{2.8}$$

After transforming the PDE, the eigenvalue problem is solved for λ with the boundary conditions Eq. (2.2) and (2.3) and the corresponding double inverse Fourier transform is applied, which results in a countable infinity of acoustic modes.

The first acoustic modes are progressive, followed by evanescent modes that decay exponentially with distance. For the sake of brevity, we consider the leading (first) progressive mode only, which contains most of the energy [82]. Moreover, we are interested in bottom pressure signatures that are recorded sufficiently far from the fault, at distances in the range of hundreds to a few thousands of kilometres, to avoid the contribution of both evanescent and surge modes (see Eq. (38) of[135]). Note that the water density, ρ, is assumed to be constant so that the SOFAR channel is not considered here. Nevertheless, it is important to note that the leading mode frequency is extremely low, thus, the whole water depth serves as a wave-guide [115], such that it cannot be trapped in the SOFAR channel even if water density variations were taken into account. Finally, applying the stationary phase approximation the bottom pressure is derived (see [82]),

$$\begin{aligned} p = \rho W_0 |A| \frac{2^{5/2} c^3 t^{1/2}}{h\pi^{1/2}\omega^{3/2} x_0} [1-(x_0/ct)^2]^{1/4} \sin\left(\frac{\omega}{c}\frac{x_0/ct}{\sqrt{1-(x_0/ct)^2}} b\right) \\ \times \sin\left(\frac{\omega}{\sqrt{1-(x_0/ct)^2}} T\right) \cos\left(kx_0 - \Omega t - \frac{\pi}{4}\right), \end{aligned} \tag{2.9}$$

where (x_0, y_0) is the epicentre coordinates of the fault relative to the point where the pressure is computed (or measured), k is the wavenumber, Ω is the frequency of the pressure signal and ω is the frequency associated with the first mode,

$$\omega = \frac{\pi c}{2h}; \tag{2.10}$$

and $A(K, X, Y)$ is the two dimensional envelope given by

$$\begin{aligned} A = \frac{1-\mathrm{i}}{2}\left(C\left(\sqrt{\frac{K}{\pi X}}(\epsilon L + Y)\right) + C\left(\sqrt{\frac{K}{\pi X}}(\epsilon L - Y)\right)\right) \\ + \frac{1+\mathrm{i}}{2}\left(S\left(\sqrt{\frac{K}{\pi X}}(\epsilon L + Y)\right) + S\left(\sqrt{\frac{K}{\pi X}}(\epsilon L - Y)\right)\right), \end{aligned} \tag{2.11}$$

with Fresnel cosine and sine integrals denoted as C and S, with the frequency and wavenumber [82],

$$k(\Omega) \equiv K = \frac{x}{c^2 t}\Omega, \quad \Omega = \frac{\omega}{\sqrt{1-(x/ct)^2}}. \tag{2.12}$$

Note that we are interested in signals measured briefly after the critical arrival time ($t_{cr} = x/c$). This helps to avoid reverberations (that arrive at a later time) when analysing real signals.

$\nu = X_0/k$, $\chi = \nu/2$, $l = \epsilon L$, $2Y_+ = l + Y$ and $2Y_- = l - Y$, and $\Omega = \omega$ is defined at the point of stationary phase, where

$$\partial g/\partial\omega = 0, \quad g(\omega) = k\frac{x}{t} - \omega \quad \text{and} \quad k = \frac{\sqrt{\omega^2 - \omega^2}}{c}, \tag{2.13}$$

and,

$$\Omega = \frac{\omega}{\sqrt{1-(x_0/ct)^2}} \quad \text{and} \quad k_j(\Omega) = \frac{1}{c}\sqrt{\Omega^2 - \omega^2} = \frac{\hat{\Omega}_j x_0}{c^2 \hat{t}_j}, \tag{2.14}$$

where $\hat{\Omega}_j$ is the frequency measured from the given pressure signal at a given point j, and $\hat{t}_j$ is the corresponding time at the point. k_j is the wavenumber associated with $\hat{\Omega}_j$.

Inverse solution for acoustic radiation from a slender fault

We considered an arbitrary slender fault eruption to create a numerical synthetic pressure signal of the leading acoustic mode using Eq. (2.9). The signal can be considered as a recorded pressure signal. The inverse problem model is then applied to retrieve the properties of the eruption, hence the slender fault geometry and dynamics. Initially, the fault location (x_0, y_0) and eruption time (t_0) can be calculated using Eqs. (8.2) and (8.3)

of Ref. [82], as summarized below:

$$x_0 = \frac{(\hat{t}_2 - \hat{t}_1)c}{\left\{1 - \left[\frac{\pi c}{2h\hat{\Omega}_{\hat{t}_2}}\right]^2\right\}^{-1/2} - \left\{1 - \left[\frac{\pi c}{2h\hat{\Omega}_{\hat{t}_1}}\right]^2\right\}^{-1/2}};$$
$$t_0 = \hat{t}_j - \frac{x_0}{c}\left\{1 - \left[\frac{\pi c}{2h\hat{\Omega}_{\hat{t}_j}}\right]^2\right\}^{-1/2}; \quad y_0 = \sqrt{(t_0 c)^2 - x_0^2}. \tag{2.15}$$

Similarly, the orientation of the fault can be calculated by Eq. (8.4) of Ref. [82], i.e. $\theta = 90^\circ - \tan^{-1}(x_0/y_0)$, which can be compared against tectonic plate boundaries and estimations made by seismic recordings analysis. Next, the dimensions of the fault and the vertical uplift speed and duration are calculated as detailed below.

Width

The fault width ($2b$) modulates the pressure signal through $\sin(kb)$, as shown in Eq. (2.9). Hence, by selecting pressure points, that are in close proximity to the envelope, e.g. near pressure amplitude extrema from the signal wave packet, the width can be calculated from $\sin(kb) = 1$, see Fig. 2.4.

The wavenumber k is calculated using relation (2.12), which by substitution in $\sin(kb) = 1$, yields

$$b_m = \frac{\pi(m - 1/2)c^2\hat{t}_j}{\hat{\Omega}_j x_0}, \quad m = 1, 2, 3..., \tag{2.16}$$

that corresponds to a countable infinity of possible solutions for b. The process described in this subsection is repeated, using different points j from the signal, resulting in a probability density function composed of potential solutions that correspond to the different possible modes, where $m = [1 - 20]$ are considered. The solutions are analysed and averaged, leading to the most probable value $\bar{b}$ as illustrated in Fig. 2.5.

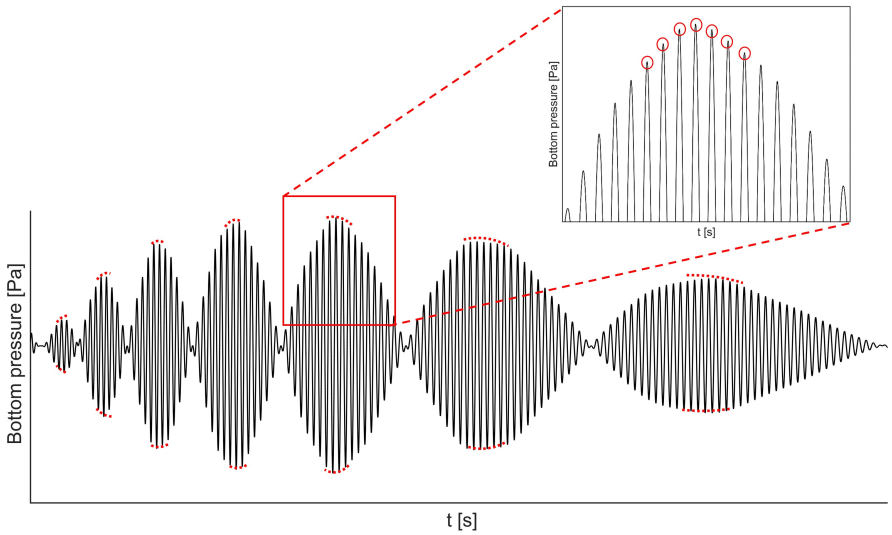

Figure 2.4: Synthetic sound signal from an slender fault. Points that meet the condition $\sin(kb) = 1$ are highlighted in red.

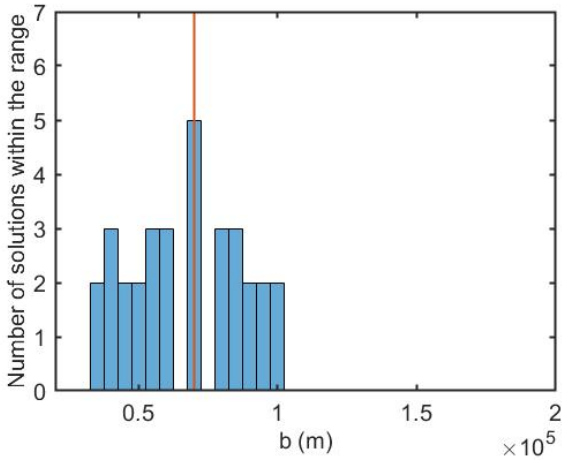

Figure 2.5: Convergence diagram of solutions for the calculation of b. The wavenumbers k_j are associated with 5 different points in time along the signal. The average value of the half width $\bar{b}$ is 70 km; the red vertical line represents the actual solution.

Duration

To find the duration of the event ($2T$), the ratio between two different pressure points i and j from the signal described by Eq. (2.9) is taken to reduce the number of unknowns

$$\frac{\hat{p}_i}{\hat{p}_j} = \frac{|A_i|/\sqrt{k_i}}{|A_j|/\sqrt{k_j}} \frac{\sin(k_i\bar{b})}{\sin(k_j\bar{b})} \frac{\sin(\hat{\Omega}_i T)}{\sin(\hat{\Omega}_j T)}. \tag{2.17}$$

Eq. (2.17) is independent of W_0 and comprises two unknowns, T and L. The latter dictates the behaviour of the envelope A, which is non-linear, as defined in Eq. (2.11).

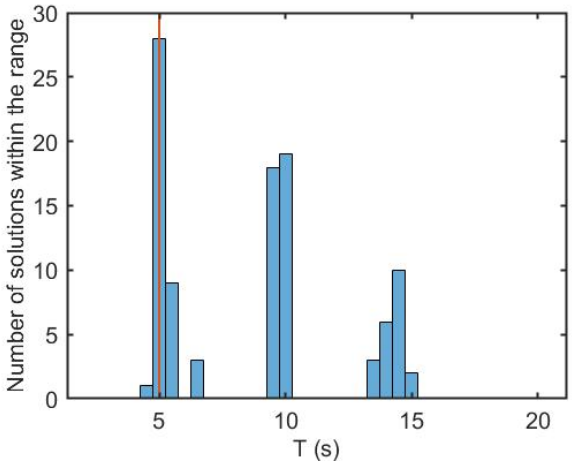

Figure 2.6: Convergence diagram for the potential combinations of L and T that satisfy Eq. (2.17). The highest density of solutions is located around lines with the same value for T, therefore, the number of possible solutions for T is reduced to three in this case, where the final solution for half the duration is identified at $\bar{T} = 5$ s, i.e. full duration $2\bar{T} = 10$ s. The red vertical line represents the actual solution.

There are infinitely many possible combinations of L and T that satisfy Eq. (2.17). We choose the most probable value ($\bar{T}$) that has the highest number of solutions within a given range, as illustrated in Fig. 2.6.

Uplift speed and length

Next, attention is focused on W_0 and L. Rewriting Eq. (2.9), W_0 can be expressed explicitly as

$$W_0 = \frac{\hat{P}_j}{\rho|A_j|\frac{2^{7/2}c}{\sqrt{\pi^3 x_0 k_j}}\sin(k_j\bar{b})\sin(\hat{\Omega}_j\bar{T})}, \tag{2.18}$$

where $|A_j|$ is the value of the envelope factor associated with the time $\hat{t}_j$. Eq. (2.18) has two degrees of freedom, W_0 and L. The envelope factor $|A_j|$ contains the parameter L, see Eq. 2.11. Although, there are several potential W_0 solutions for each possible L that satisfy Eq. (2.18), the solutions can be constrained within ranges defined by previous documented effective earthquake dimensions, e.g. see Ref. [131]. For each combination of L and W_0, within the established ranges, a synthetic pressure envelope is generated by Eq. (2.9) and compared with the actual (recorded) pressure envelope of the signal. This procedure is done for two different inputs of k_j and $\hat{\Omega}_j$ $(j = 1, 2)$, leading to two three-dimensional surfaces, $S_1(L, W_0, err_1)$ and $S_2(L, W_0, err_2)$, where err_1 and err_2 are the absolute errors between the pressure generated by the potential combinations of W_0 and L and the actual pressure in the studied signal. Both surfaces reside in a three-dimensional space so that they intersect on a curve, which we average in both W_0 and L directions, producing a unique final solution for W_0 and L.

A note on the far-field solution

The stationary phase approximation was applied by Ref. [82] to reach an analytical solution for the acoustic-gravity bottom pressure, enabling a near real-time analysis of the signal. Note that the far-field is at $x/h \gg 1$. However, due to the fact that b is typically much larger than h, a few tens of kilometres away would already fall in the far-field of the mathematical solution, yet they are relatively in the near field of the fault. Table 1 of Ref. [63] shows an error of 10% at a distance of 50 km, which drops to 0.3% at a distance of 500 km.

2.3 Inverse problem model

We propose an inverse problem model that processes acoustic pressure signals induced by tectonic events and provides a probabilistic estimation of the effective fault geometry and dynamics in near real-time. The steps required to apply the model are described in this section in detail for both synthetic and real hydrophone recordings.

Inverse problem model application: Synthetic signals

A general flow chart of the inverse problem model process is presented in Fig. 2.7.

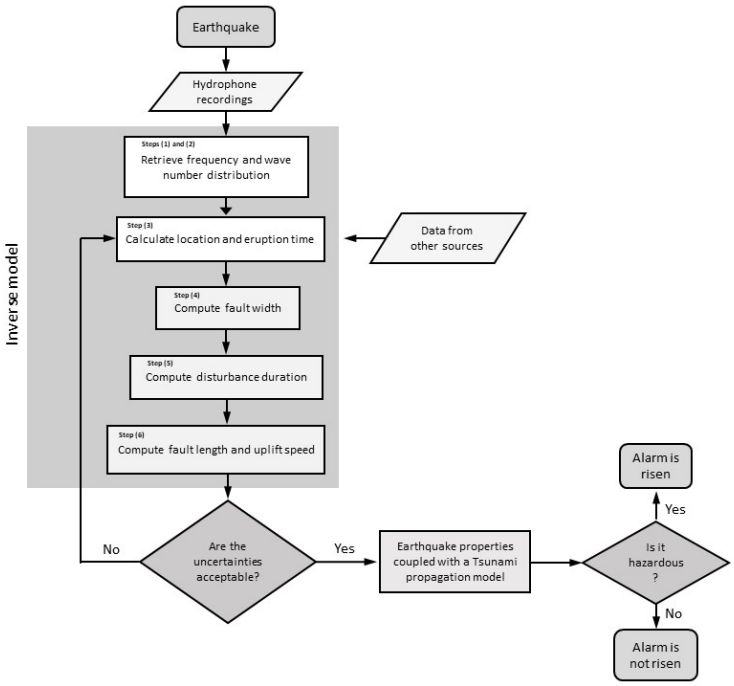

Figure 2.7: Inverse problem model application flowchart - from acoustic pressure signal arrival to probabilistic calculation of source properties.

Step 1: Frequency distribution

The first step, prior to analysing the pressure signal, is to identify regions composed of points that are likely to lead to minimum errors in the solution, see Fig. 2.8. Subsequently,

we calculate the frequency distribution; time differences between five consecutive peaks from the signal, $j-2$, $j-1$, j, $j+1$ and $j+2$, are obtained and the weighted average is computed, leading to the period associated with j, see Fig. 2.9. The weighted average is defined by

$$\hat{T}_{\hat{t}_j} = \frac{c_1\hat{T}_1 + c_2\hat{T}_2 + c_3\hat{T}_3 + c_4\hat{T}_4}{c_1 + c_2 + c_3 + c_4}, \tag{2.19}$$

where $\hat{T}_{\hat{t}_j}$ is the period associated with a time point $\hat{t}_j$, the remaining parameters in Eq. (2.19) are defined by:

$$\hat{T}_m = \hat{t}_{j+m-3} - \hat{t}_{j+m-2}, \quad c_m = \frac{\hat{T}_m}{\hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \hat{T}_4}, \quad m = 1, 2, 3, 4 \tag{2.20}$$

where $\hat{T}_1 < \hat{T}_2 < \hat{T}_3 < \hat{T}_4$ and c_1, c_2, c_3, c_4 are the weighting coefficients.

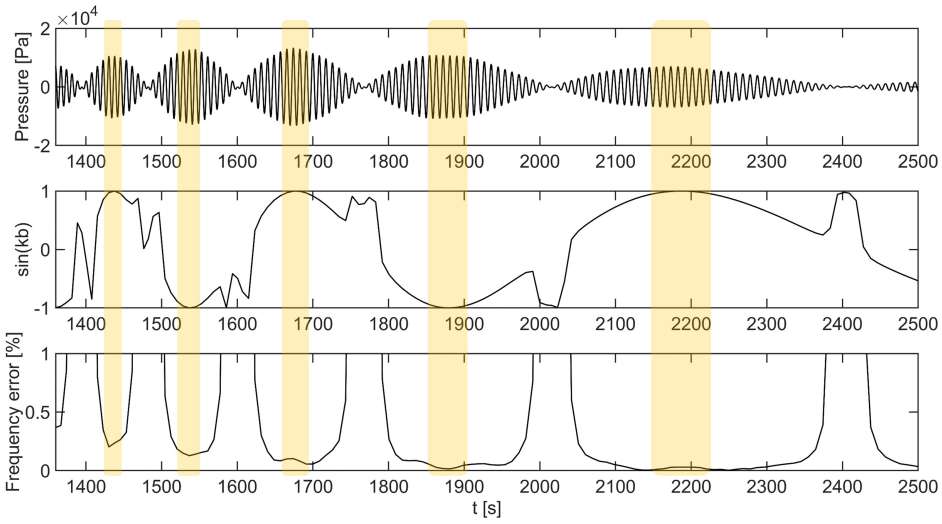

Figure 2.8: Top: Synthetic pressure signal at the relative location $x_0 = 1500$ km, $y_0 = 500$ km induced by a slender fault, where $L = 700$ km, $b = 140$ km, $W_0 = 0.1\ ms^{-1}$ and $T = 10$ s, the average sea depth is $h = 4$ km. Middle: Distribution of $sin(k_j b)$ along the signal. Bottom: Errors in frequency calculation by Eq. (2.21). The highlighted regions correspond to areas related to low errors due to the pressure points being close to the envelope and meeting the condition $sin(kb) = 1$.

In Fig. 2.8, Eqs. (2.19) and (2.20) are computed for every identified pressure peak in the signal and the associated frequency distribution is calculated by applying $\hat{\Omega}_{\hat{t}_j} = 2\pi/\hat{T}_{\hat{t}_j}$. Errors retrieving the frequency distribution are minimised if points that fall near pressure amplitude extrema are selected, see Fig. 2.9. The highlighted regions in Fig. 2.8 comprise points with associated small errors,

$$err_{\hat{\Omega}_{\hat{t}_j}} = \frac{|\hat{\Omega}_{\hat{t}_j} - \Omega_{\hat{t}_j}|}{\Omega_{\hat{t}_j}} 100[\%]. \tag{2.21}$$

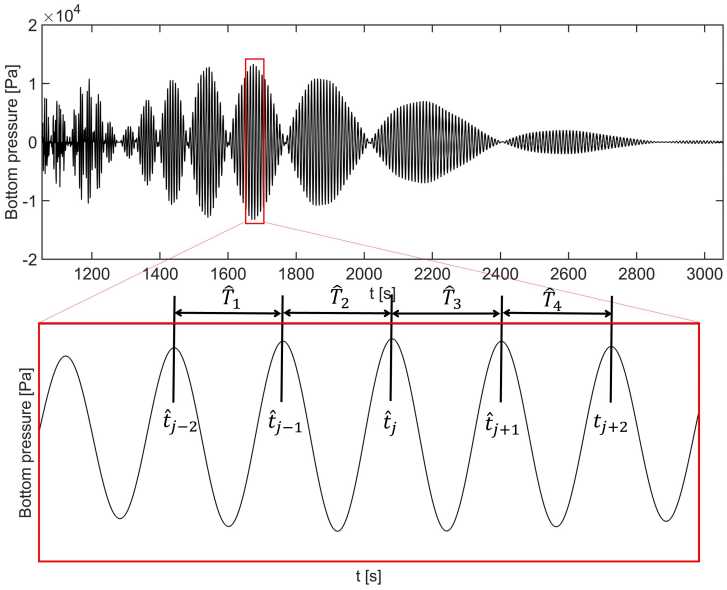

Figure 2.9: Above: Pressure time series induced by vertical slender fault motion. Below: Closer look to the region where consecutive peaks are used to calculate the period related to the central peak, $\hat{T}_{\hat{t}_j}$.

Step 2: Wavenumber distribution

The wavenumber distribution (k_j) associated with the analysed signal can be now calculated by the application of Eq. (2.14) to the previously calculated frequencies ($\hat{\Omega}_j$) in Step 1.

Step 3: Location and eruption time

Eq. (3.8) is sensitive to errors in frequency measurement, in particular, when selecting two pressure points that lie in close proximity to each other. Consequently, the denominator of Eq. (3.8) tends to zero, which increases the instability of the solution and induces larger errors - see highlighted areas on Fig. 2.10. Therefore, these scenarios are avoided by the model.

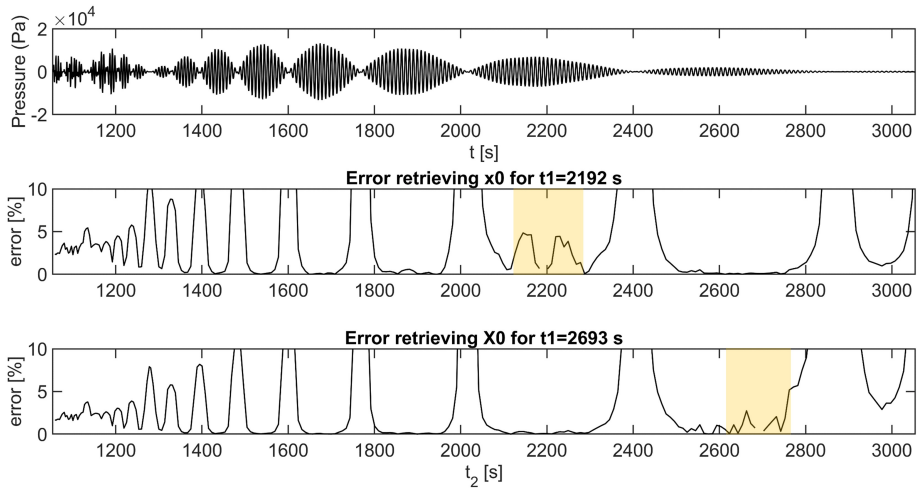

Figure 2.10: Top: Synthetic pressure signal. Middle and Bottom: Eq. (3.8) is solved for a fixed t_1 and every possible t_2 along the signal. The error is calculated by comparing the calculated x_0 with the actual solution. Two different scenarios for t_1 from the same signal are shown. t_1 lies in the coloured areas, demonstrating that points in close proximity induce unnecessary errors in the solution.

Although, differences in pressure between the peaks in the signal and the corresponding points that lie on the actual envelope can induce uncertainties in the results, errors in the frequency are found to be more significant. With these observations in hand, the frequency distribution, fault location, and eruption time are calculated and used to compute the wavenumber distribution by Eq. (A.7).

Step 4: Fault width

The model selects regions in the signal associated with minimum uncertainties, as previously described. The closer these regions to the envelope, the more accurate the approximation $\sin(kb) = 1$ becomes, see Fig. 2.4. The wavenumber (k) is obtained by using the retrieved frequencies in Eq. (2.12). Then, Eq. (A.7) is used for each selected pressure point and its associated wavenumber (k) with a predefined number (n) of solutions. This leads to several sets of n solutions that are constrained in ranges based on estimations of effective earthquake surface fault widths by Ref. [131]. The model analyses the local density of solutions by generating a grid with a fixed step, identifying the step containing the highest number of solutions and selecting it as the weighted average solution for the half width (b), see Fig. 2.5.

Step 5: Disturbance duration

Eq. (2.17) has two unknowns, T and L. Thus, these can be solved numerically by selecting two sets of points, each comprising two pressure points, $\hat{p}_i$ and $\hat{p}_j$, two frequencies $\hat{\Omega}_i$ and $\hat{\Omega}_j$, and two wavenumbers k_i and k_j. Each solution of, say L, leads to an infinite number of solutions for T. However, we limit the solution to realistic ranges associated with the specific region, e.g. $L = [10, 900]$ km [131], and $T = [3, 20]$ s [29, 24]. This leads to a two-dimensional matrix containing possible combinations of L and T, that satisfy Eq. (2.17).

The periodic nature of Eq. (2.17) leads to a periodic distribution of solutions for T. A probability density function for the number of solutions within a given range is established. Pairs of points i, j are considered as needed until convergence of the possible solution is reached, see Fig. 2.6. We consider all possible duration T_c that are associated with high convergence of solutions around them, such that $T_c \geq \alpha T_{max}$, where T_{max} is duration with the highest number of solutions, and α is an arbitrary fraction, e.g. $\alpha = 0.6$ is found efficient in terms of computation time and the number of solutions. Eq. (2.17) is computed for all the considered solutions of T and all possible solutions for L. For each chosen T, the number of combinations with L that satisfy Eq. (2.17) are analysed

and the duration with the highest number of solutions is selected as the converged value for half the uplift duration.

Step 6: Length and uplift speed

The last two unknown parameters to be retrieved are L and W_0. To address this challenge, several combinations of W_0 and L are chosen in regular intervals within realistic ranges. Pressure envelopes, with low sampling frequency (to reduce CPU time), are generated for each combination of L and W_0 using Eq. (2.17). The generated pressure envelopes and the corresponding pressure points from the signal are compared against each other, so the average difference in pressure is given by

$$E_a = \frac{1}{N_{peak}} \left\{ \sum_{n=1}^{N_{peak}} \frac{|p_{env} - \hat{p}_{peak_n}|}{\hat{p}_{peak_n}} \right\}, \tag{2.22}$$

where N_{peak} is the number of chosen peaks used, p_{env} is the pressure point on the envelope, $\hat{p}_{peak_n}$ is the measured pressure of each of the chosen peaks. By minimising E_a, two combinations of W_0 and L, that correspond to the minimum average pressure differences obtained using Eq. (2.22), are chosen and their arithmetic mean is calculated, and thus L_c is obtained.

Considering that the solution for L has some degree of uncertainty, in order to calculate the last parameter W_0 five different values for the length are considered and Eq. (2.18) is calculated for each one of them: $L_1 = L_c - (0.1L_c)$; $L_2 = L_c - (0.05L_c)$; $L_3 = L_c$; $L_4 = L_c + (0.1L_c)$; $L_5 = L_c + (0.05L_c)$. This procedure leads to five solutions for W_0, that can be arithmetically averaged, in order to achieve mean solution for the uplift speed. The choice of five L values about L_c is arbitrary, and more values can be considered, though the CPU time can increase substantially.

Inverse problem model application: real hydrophone recordings

The waves induced by the tectonic event are firstly detected by seismometers (P-waves), which can shed light on the epicentre location and magnitude of the event, thus,

the mentioned data can be used to reduce the uncertainties associated with the inverse problem model results.

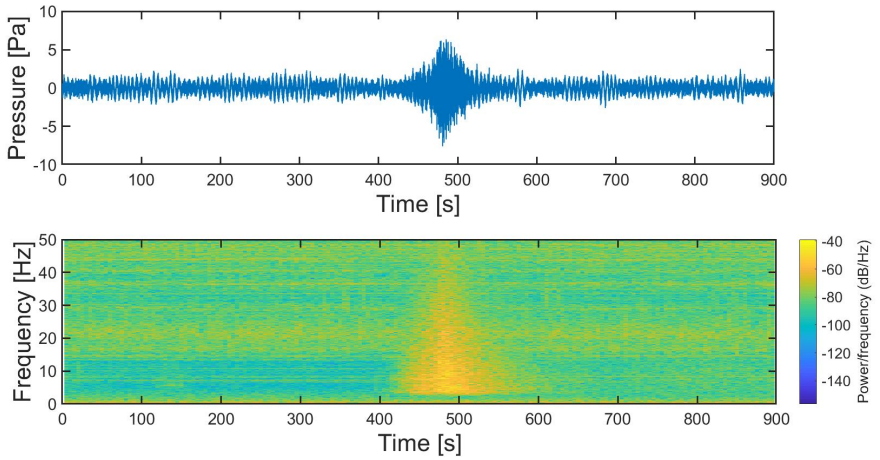

Figure 2.11: Top: Calibrated pressure signal, induced by a 6.9 M_w earthquake recorded by a icListen LF 224. Bottom: Corresponding spectrogram.

As demonstrated in detail in the Results section, the inverse problem model can provide an estimation of the source characteristics when applied to hydrophone recordings. Here, we focus on the application to the real scenarios side of the model, which requires additional steps due to the presence of noise and signal distortion of the real data. In particular, attention is focused on the time-pressure series of acoustic signals recorded in the Cascadia Basin by a hydrophone of Ocean Networks Canada [96], after a 6.9 M_w earthquake in the southern East Pacific ridge on October 9th 2014, see Fig. 2.11.

Step 1: Potential location of the source

Initially, the frequency distribution of the leading mode is calculated, which is a function of time (t, relative to the eruption time t_0), depth (h), and orientation of fault (x_0 and y_0), see Eq. (2.12). The total distance between the hydrophone and the earthquake is approximately 9000 km, which can be given by various combinations of x_0 and y_0, i.e. different possible orientations. The potential coordinates are constructed as two vectors: the first comprises the possible solutions for x_0, ranging from 0 to 9000 km on regular

intervals of 50 km; and the second has the solutions y_0 corresponding to the possible x_0 in the first vector. Eq. (2.12) is computed for each possible coordinate (x_0, y_0) that lead to frequency distributions with minimum value higher than the minimum frequency threshold. The lower limit of the signal frequencies in the power spectrum is identified, in this case at 3 Hz. This procedure leads to only five potential solutions for the location of the earthquake in the presented case, $x_{0,1}$=8750 km and $y_{0,1}$=2100 km, $x_{0,2}$=8800 km and $y_{0,2}$=1887 km, $x_{0,3}$=8850 km and $y_{0,3}$=1636 km, $x_{0,4}$=8900 km and $y_{0,4}$=1337 km, $x_{0,5}$=8950 km and $y_{0,5}$=947 km. A comparison between the spectrogram and the potential frequency distributions is made in Fig. 2.12.

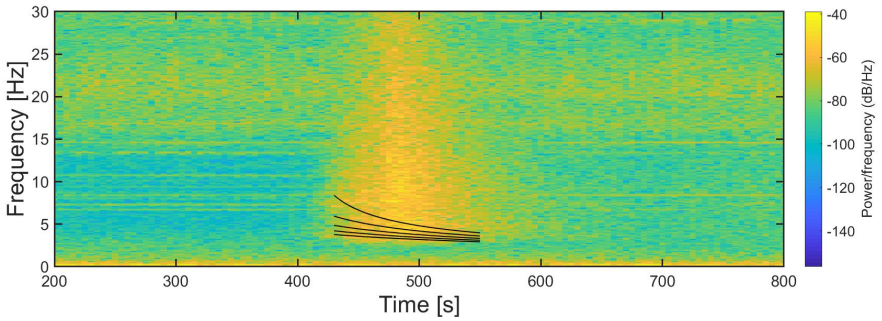

Figure 2.12: Spectrogram associated with the studied pressure signal. The black lines represent the five possible synthetic ideal distributions of the first acoustic mode.

Step 2: Signal detection

To identify the beginning of the relevant signal content, the short-time energy [52] distribution is calculated using

$$E_f = \sum_{n=0}^{N_e-1} x^2(n), \tag{2.23}$$

where $n = 0$ and $(N_e - 1)$ represent the limits of each frame, $N_e - 1$ is the length of the frames in samples. The time-frames were chosen to be five seconds long ($N_e = 4000$ samples). The threshold for the identification of the signal disturbance is set at 1.2 times the average value of the short-time energy distribution for the first ten seconds in the recorded time series, see in Fig. 2.13.

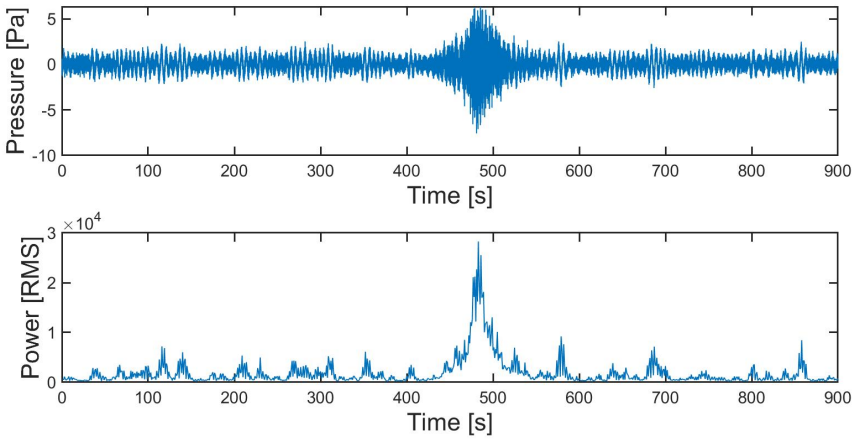

Figure 2.13: Above: Acoustic disturbance originated by the studied tectonic event. Below: Short-time energy distribution.

Step 3: Envelope tracking

In order to minimise the deviation between selected pressure points at the actual signal and the corresponding approximated points at the envelope, we seek points at the vicinity of the extrema.

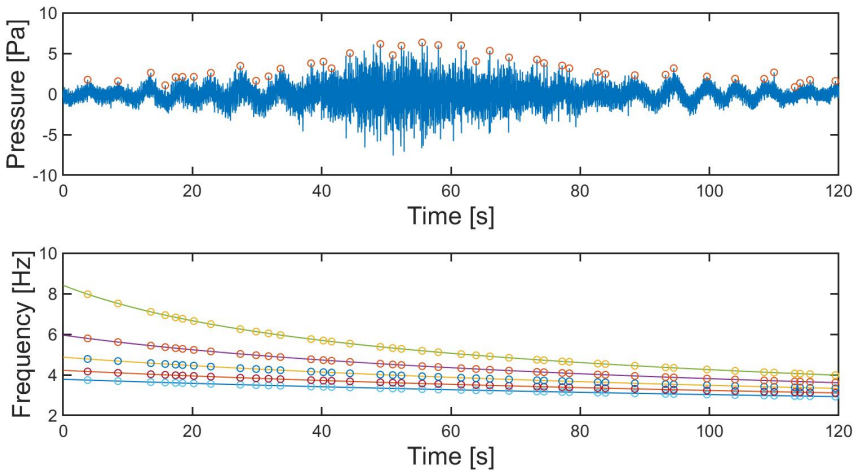

Figure 2.14: Above: Acoustic signal. The identified pressure maxima are highlighted. Below: Five potential first mode frequency distributions for the signal. The points related to the highlighted maxima are circled.

Applying an iterative identification method of local maxima (or minima), the total difference between selected pressure points and the theoretical envelope is reduced. The method is repeatedly applied until the relation $2\pi/T_{peaks}$ is verified to lie below the previously obtained theoretical frequency distributions for the first mode for all the identified peaks, where T_{peaks} are the averaged times between consecutive peaks. This is a necessary condition in order to associate the computed ideal frequency distributions to the identified local maxima in the pressure signal. Each identified pressure point is associated with a frequency point of each of the potential first mode frequency distributions, see Fig. 2.14.

Step 4: Estimation of characteristics range

Ref. [131] established correlations between magnitude, surface rupture length, width and ground displacement. Using these correlations an earthquake, say of magnitude M_w=6.9, is associated with length $2L \approx 35$ km (figure 9 in Ref. [131]) and area $A \approx 620$ km^2 (figure 16 in Ref. [131]), thus the width (in case of a rectangle) is $2b \approx 17.71$ km. An estimation of the average vertical displacement is $d \approx 0.75m$ (figure 11 in Ref. [131]). It is worth mentioning that the used correlations are based on empirical values which include significant scatter [131]. Using these correlations (and scattering) we establish reasonable ranges for the earthquake characteristics for the analysed earthquake, e.g. see ranges listed in Table 2.1.

Property	**Range**
Length [km]	1-100
Width [km]	1-50
Duration [s]	0-15
Uplift speed $[ms^{-1}]$	0.005-0.500

Table 2.1: Input ranges of source characteristics for the inverse problem model, based on Ref. [131].

After the calculation of the solutions the model verifies that the effective uplift of the fault (d) lies in the same order of magnitude as the estimated by the parameterizations,

in this case, $d \approx 0.75m$ [131].

Step 5: Calculation of the source characteristics

Steps 4, 5 and 6 of the inverse problem model applied to the theoretical scenarios (Inverse problem model application: synthetic signals subsection in the Inverse problem model section) are applied to each potential frequency distribution of the first acoustic mode and the associated pressure points from the signal. As a result, we obtain 30 sets of solutions for every frequency distribution, which leads to a total of 150 solutions for each fault characteristic. The distributions of the fault characteristics are shown in the histograms in Fig. 2.15.

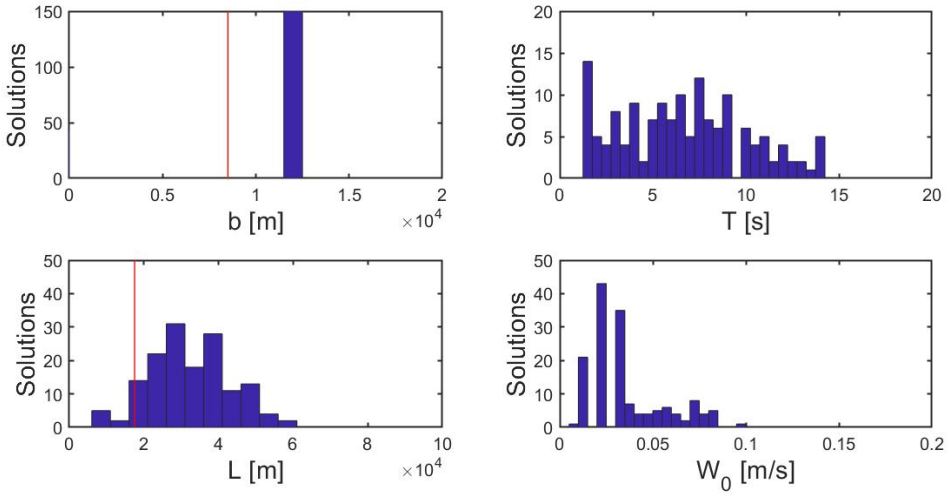

Figure 2.15: Histograms of calculated solutions for the half width, half duration, half length and uplift speed of the source. The average values are: $\bar{b}$= 12.000 km, $\bar{T}$= 6.396 s, $\bar{L}$= 24.873 km and $\bar{W}_0$=0.049 ms^{-1}. The vertical red lines indicate the estimated characteristics by the relations found in Ref. [131].

2.4 Sensitivity Analysis

Since the proposed model requires approximating the selected pressure amplitudes and the associated frequencies, it is important to evaluate the deviation that small errors

in the approximations may lead to. In particular, the model was tested using synthetic data with different levels of noise and errors. For the sake of brevity. Two main sources of error are identified in the results: (1) frequency calculation; and (2) envelope tracking (assumption of $\sin(kb) = 1$). Additionally, errors can be related to the difference between the points in the pressure signal used as input to the model and the corresponding points on the theoretical envelope. Nevertheless, this type of uncertainty can be minimised by selecting pressure points associated with local maxima as mentioned above.

$x_0[km]$	$y_0[km]$	$2b[km]$	$2T[s]$	$2L[km]$	$W_0[ms^{-1}$
1000	500	160	10	1600	0.100

Table 2.2: Theoretical slender fault properties used for the sensitivity analysis.

To proceed with the analysis, a theoretical case scenario was generated, see Table 2.2. Initially, sensitivity to errors in frequency calculation of Eq. (3.8) was tested by introducing random errors up to a maximum level of 100% of the calculated frequency distribution, see Figs. 2.16 and 2.17. In this case, the solution for the relative location of the slender fault showed high sensitivity to errors, especially, the solution for y_0, see Figs. 2.16 and 2.17. Practically, the location calculation is not an issue as it is normally obtained with high accuracy using seismometers. Thus, in practice, input data from seismometers can be used to tune the parameters leading to y_0.

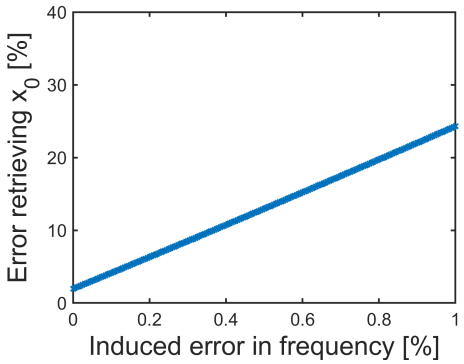

Figure 2.16: Error in x_0 calculation for different levels of induced errors in frequency measurements

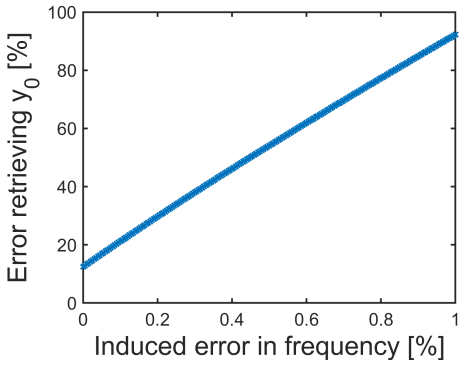

Figure 2.17: Error in y_0 calculation for different levels of induced errors in frequency measurements

As stated before, the model utilises several pressure points to calculate the disturbance properties. However, additional points around them can be chosen in order to increase the stability of the model. Three configurations for the number of points used to compute each set of solutions were tested. The first tested configuration consisted of the initially selected points (one set) only; the second configuration considered two additional sets of points (three sets of points in total), where an extra point can be located at each side of the point initially selected in the first configuration; the third configuration included four additional sets of points (5 sets in total), considering two extra points at each side of the points selected in the first configuration.

The application of the model to each set of points results in a set of solutions to give a probability density of the property and can be averaged to deliver a final unique solution. The number of chosen points used to compute each set of solutions was analysed. A range from two points, which is the minimum, up to eight points was considered. Note that, more points are not considered in the analysis as the computations become extensive. Errors in the location calculations (x_0, y_0) are assumed to be under 5%, which is a reasonable assumption considering that data from seismometers are at hand. Random errors in frequency and pressure amplitude were introduced independently to evaluate their influence on the solutions. It is remarkable that the choice of five points to calculate each set of characteristics, and using three sets of points, provided a good balance between computational efficiency and accuracy, see Figs. 2.18, 2.19, 2.20, 2.21 and 2.22.

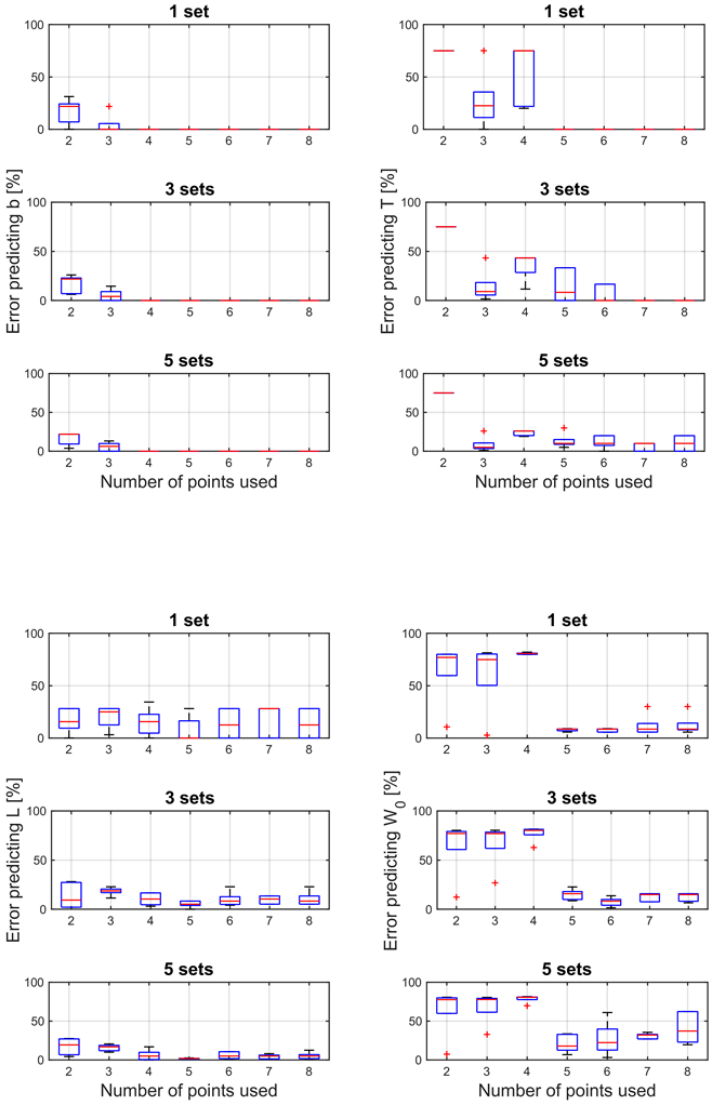

Figure 2.18: Error retrieving different fault properties by considering points that meet $\sin(kb) > 0.99$ and errors in frequency lower than 1%. From up to down the number of sets of solutions increase.

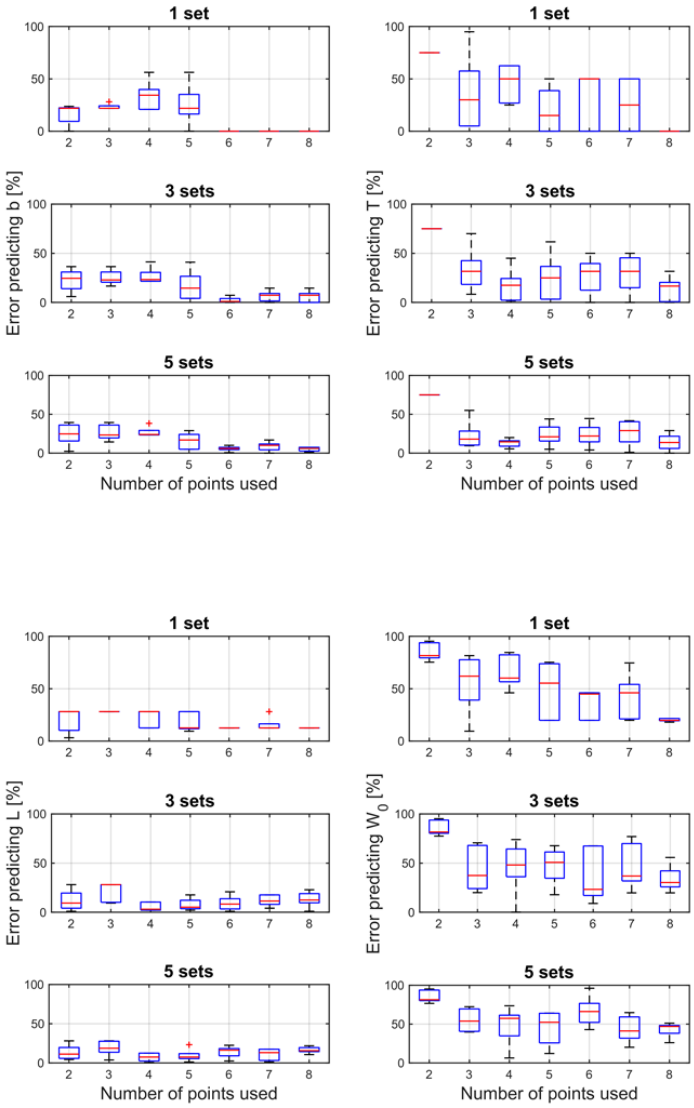

Figure 2.19: Error retrieving different fault properties by considering points that meet $\sin(kb) > 0.95$ and errors in frequency lower than 1%. From up to down the number of sets of solutions increase.

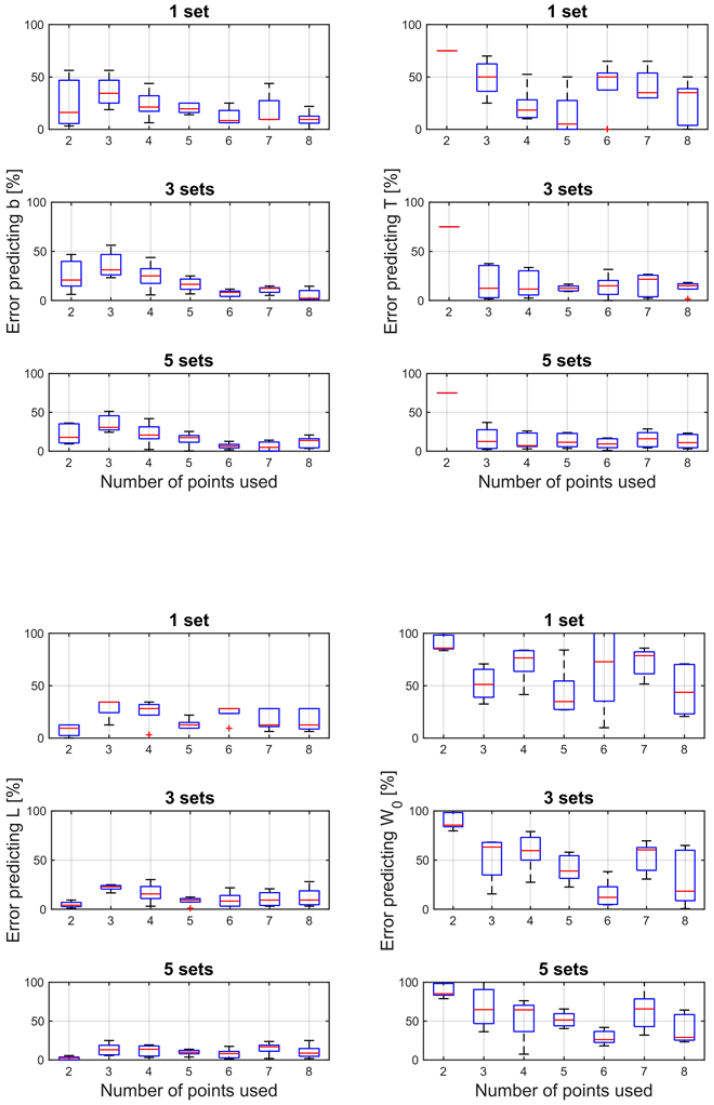

Figure 2.20: Error retrieving different fault properties by considering points that meet $\sin(kb) > 0.90$ and errors in frequency lower than 1%. From up to down the number of sets of solutions increase.

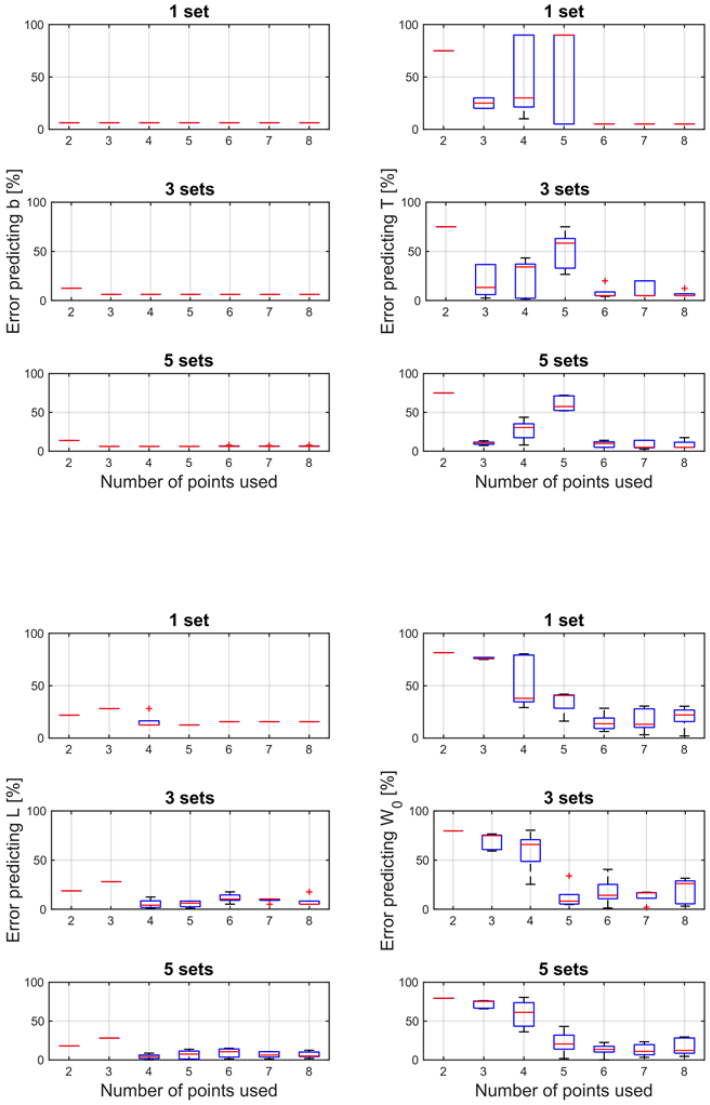

Figure 2.21: Error retrieving different fault properties by considering points that meet $\sin(kb) > 0.99$ and errors in frequency lower than 5%. From up to down the number of sets of solutions increase.

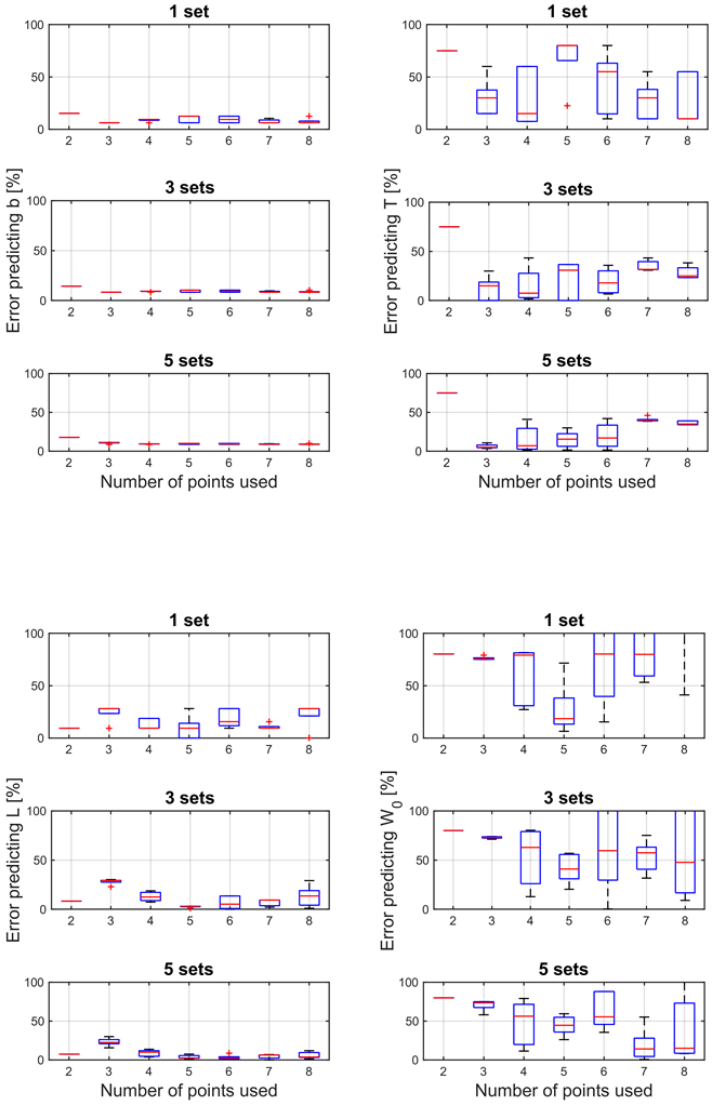

Figure 2.22: Error retrieving different fault properties by considering points that meet $\sin(kb) > 0.99$ and errors in frequency lower than 10%. From up to down the number of sets of solutions increase.

Next, the number of points chosen to compute each set of solutions was fixed to five, and the number of calculated sets of solutions was fixed to ten, which were averaged leading to a final solution for each characteristic. Then, different levels of random errors in pressure and frequency measurements were introduced and the model was applied 30 times for each synthetic noise scenario. We observed that errors in frequency induced a higher variance in the solutions obtained by the model, which resulted in higher uncertainties, see Figs. 2.23 and 2.24.

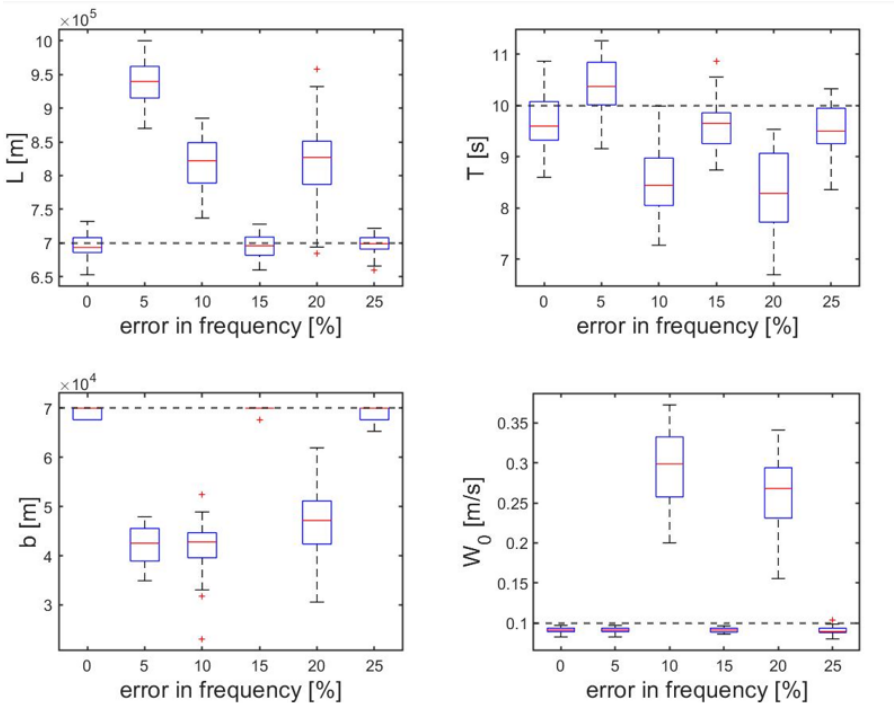

Figure 2.23: Error retrieving the fault properties with errors introduced in the frequency distribution.

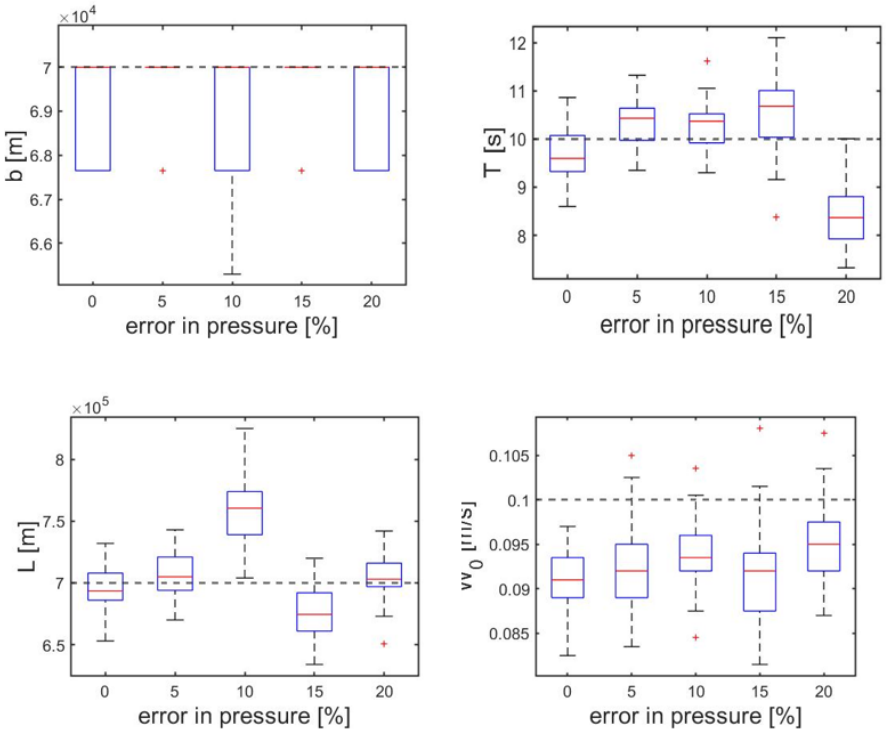

Figure 2.24: Error retrieving the fault properties with errors introduced in the pressure measurement.

The CPU times required by the model increase (almost) linearly with the number of solutions demanded by the user (Fig. 2.25). Moreover, the computational effort is related to the size of the ranges chosen to constrain the potential solutions used as input to the model.

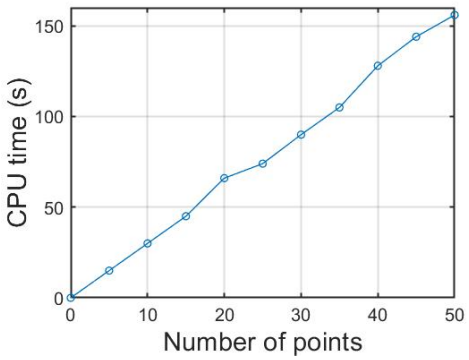

Figure 2.25: CPU times for different amounts of points used to compute each set of solutions requesting five sets of solutions. The processor used is Intel(R) Core(TM) i5-4690 CPU, speed [3.5 GHz-3.5 GHz] and 8 GB RAM.

2.5 Results

Analysis of synthetic data

In theory, to estimate the slender fault geometry and dynamics, only a few points are required from the signal. Nevertheless, the large amount of points (and combinations of points) that can be applied, leads to a probability density of solutions. This allows applying the inverse problem model in scenarios with noise and distorted pressure signals.

The model has been tested using synthetic signals generated by Eq. (2.17) under different input configurations to quantify its accuracy and computational efficiency along with its response to induced Gaussian noise. Different model parameters have been considered, such as the number of sets of points utilised and the number of points used to compute each set of characteristics. The choice of five points with two additional sets of points to compute each set of solutions, as shown in Sensitivity Analysis section, leads to an acceptable relation between model accuracy and computational effort. Thus, the set-up parameters can be configured to minimise the uncertainties associated with the solutions.

Case		x_0 **[km]**	y_0 **[km]**	b **[km]**	T **[s]**	L **[km]**	W_0 $[ms^{-1}]$	**CPU [s]**
A	Input	1000	300	80.0	35.0	250	0.100	183
	Retrieved	998	283	76.5	28.7	316	0.069	
B	Input	800	100	70.0	45.0	750	0.100	126
	Retrieved	796	100	70.1	49.5	681	0.138	
C	Input	1500	500	100	30.0	400	0.100	171
	Retrieved	1464	464	96.5	34.3	471	0.072	

Table 2.3: Comparison between input parameters used to generate synthetic pressure signals and retrieved inverse process parameters. The computational effort needed by the model to calculate 30 sets of solutions using 5 points to compute each set with a PC of characteristics, processor Intel(R) Core(TM) i5-4690 CPU, speed [3.5 GHz-3.5 GHz] and 8 GB RAM is reported.

Three synthetic cases, based on real earthquakes, were generated. Case A: 11 March 2011, 9.1 M_w Tohoku earthquake. Case B: 26 December 2004, 9.1 M_w Sumatra earthquake. Case C: 22 May 1960, 9.5 M_w Chile earthquake. Note that even though the synthetic cases presented here are based on real earthquakes, the main target was to generate numerical scenarios to study the behaviour of the inverse problem model, hence some properties may deviate from the real ones. The eruption duration in cases A and B was obtained from the Global CMT catalog [29, 24], the duration in case C was chosen to be in the same order of magnitude as cases A and B. The simplified geometric properties (length and width) of the effective slender faults were based on previous research related to the mentioned tectonic events. In case A, rupture length and width of 440 km and 180 km were reported [138], respectively. The rupture length in case B was estimated to be over 1000 km and the width between 120 and 140 km [37]. Case C has a 800 km rupture length and 200 km width [66]. The rest of the properties were estimated using scaling relations found in Ref. [131], see Table 2.3.

Analysis of real data

We considered an earthquake that was triggered on 9th October 2014, $M_w = 6.9$ and was detected on the southern East Pacific ridge. The magnitude, centroid time, depth of

the hypocentre, half duration and location of the event are reported in the Global CMT catalog [29, 24]. The tectonic event occurred at 2:14:42.4 UTC at the coordinates 32.34 S and 110.81 W, the depth of the hypocentre was 12 km and the half duration was 7.3 seconds. The angles of the fault planes are 90° and 296° Strike; 39° and 54° Dip; and 69° and 106° Slip, which suggests that the earthquake had an effective vertical motion component. The earthquake triggered a tsunami that was recorded in the Easter Island region with a maximum amplitude of 47 cm [91].

The hydrophone data was provided by Canada Ocean Network [96]; the recording hydrophone is the model icListen LF 224 (sampling frequency 4 kHz), deployed in the Cascadia Basin (47.76 N and 127.75 W) at a depth of 2662 m. The distance between the epicentre of the earthquake and the hydrophone was approximately 9000 km, resulting in 100 minutes of travel time for the acoustic radiation, at a constant speed of $c = 1500$ ms^{-1}. The hydrophone recordings, shown in Fig. 2.11, start at 03:48:2.3 UTC and the pressure disturbance, induced by the above described earthquake, was identified approximately at 03:55:0 UTC, matching the expected travel time (100 min). The pressure signal, displayed in Figs. 2.11 and 2.13, was calibrated using the sensitivity file provided by Canada Ocean Network for the corresponding hydrophone (the provided signal was initially calibrated between 1 and 1600 Hz).

The model delivered a unique solution for b, due to the fact that ideal frequency distributions without noise were assumed, see Inverse problem model section. On the other hand, the solutions for T and L are distributed symmetrically around the mean, see in Fig. 2.15, whereas the distribution of solutions for W_0 present positive skewness. The calculated values of b, L, and vertical displacement are in the same order of magnitude as the values based on Ref. [131], see Fig. 2.15. The calculated duration of the vertical uplift ($2\bar{T}$=13.9 s) is in agreement with the duration provided by the Global CMT (14.6 seconds, see Table 2.4).

Property	**Inverse problem model**	**Ref. [131]**
Length ($2L$) [km]	49.7	35.0
Area [km^2]	1193	620
Width ($2b$) [km]	24.0	17.7
Vertical displacement [m]	0.626	0.750

Table 2.4: Comparison between the calculated effective earthquake properties by the inverse problem model and the estimated properties by relations found in Ref. [131].

2.6 Concluding remarks

This chapter aimed to investigate the possibility of relating the acoustic radiation signature induced by underwater earthquakes with the geometric and dynamic properties of the source. To achieve this purpose a methodology is presented, where a system of equations is built by selecting different pressure points from recorded acoustic signals and the analytical solution for the pressure signature in the far-field produced by an uplifting slender fault [82, 63]. The unknowns to be calculated are the dynamic and geometric effective characteristics of the slender fault. Then, a set of numerical and analytical techniques are introduced to efficiently solve the presented system of equations and provide a description of effective tectonic events in almost real-time. Different configurations of the developed inverse problem model can be utilised, resulting in variations for the uncertainties and computational times, which are analysed and reported in Sensitivity Analysis section. Sensitivity Analysis section can be utilised as a reference guide for the efficient use of the introduced inverse problem model.

Note that, some simplifications were assumed by Ref. [82] to derive an analytical solution for the generation and propagation of acoustic radiation by an uplifting slender fault. The assumption of a rectangular slender fault may correspond to many earthquakes, though we employ that as a mechanism for the generation of an *effective* water uplift rather than an actual description of the full complex dynamics of the rupture. Hence, one could argue that any earthquake that generates a tsunami can be associated with an effective vertical uplift of the water volume. Noting that the leading acoustic mode

length scale is much longer than the water depth, the assumption of a flat rigid seabed is justified as long as the water depth is above some critical value, below it the interaction with the elastic bottom becomes relevant. In addition, the effect of sound speed variations is minor, as the relevant leading acoustic mode has an extremely low-frequency [54] and the whole ocean column acts as a guideline. However, the more realistic uplift with acceleration may modify the absolute amplitudes in the signal (acceleration will increase the amplitude, whereas deceleration will decrease it) but not the ratios which is what the inverse problem model applies.

Attention was first focused on theoretical scenarios based on real earthquakes, and the sensitivity of the calculated parameters to uncertainties in the signal, say due to ambient noise. Many synthetic signals have been produced and the inverse problem model was tested on them as under different uncertainty conditions as shown in Sensitivity Analysis section, where a sensitivity analysis is detailed. The required computational times to assess the cases increases as more solutions are asked from the model to approach more accurately statistical distributions that will define the uncertainties associated with each final solution. Different uncertainty source parameters have been tested under controlled situations such as errors in frequency and pressure. Moreover, a further methodology to apply the developed inverse problem model to real acoustic signals recorded in the ocean was developed, where assumptions on the frequency distributions of the first mode were taken to minimise the uncertainties and increase the stability of the inverse problem model. A probability density function for the effective fault characteristics is delivered by the inverse problem model.

Chapter 3

Characterization of underwater earthquakes by machine learning algorithms

3.1 Introduction

[1]As stated in previous chapters, underwater seismic events can produce very long compression-type waves (AGWs), that propagate in the water layer travelling long distances with almost no attenuation and can be recorded by distant hydrophones [82]. This property of AGWs allows them to carry information on the sound source [82, 63]. The classification and characterisation of such information are important for the assessment of potential Tsunamis. In order to characterise tectonic events, the source dimensions, dynamics and moment magnitude need to be estimated, which can be approached by automated underwater acoustic signal processing methods.

In the previous chapter as in Ref. [43], we proposed an inverse problem model which calculates the effective fault dimensions and vertical uplift speed and duration induced by underwater earthquakes, using slender fault theory [82]. However, this model can

[1]This chapter is based on work that was published as an article in Scientific Reports journal (Nature) [42]

be applied only when the slip direction is vertical. Thus, there is a need to identify the slip direction prior to applying the model in real-time. To this end, we consider Anderson's faulting theory [6] where the equations of stress produced over the fault planes in an earthquake are analysed and earthquakes are divided into three classes depending on the faulting type: wrench (when the greatest pressure is in the horizontal plane), normal, and reversed. This 3-type classification has been widely accepted and used in the literature [34, 14]. Nevertheless, previous studies indicate that tectonic events can be further grouped in only two types, dip-slip and strike-slip, depending on the direction of the dominant motion component [68]. The discrimination of events with significant vertical slip can be performed by machine learning (ML) techniques. The application of ML algorithms to acoustic signals in the ocean has had increasing notability in recent years, such as the classification of vessels [31], earthquakes [125, 124], tsunamigenic events [73], underwater explosions [125, 124] and marine life [87], to name a few. An important input parameter for the inverse problem model is the earthquake moment magnitude, in this chapter we aim to infer the tectonic event moment magnitude from the acoustic recording analysis. Relations between the maximum amplitude of T-phase waves, the earthquake's energy that propagates through the SOFAR channel [121, 100], and earthquake size [137], or correlations between T-phase power level and seismic moment have been developed in previous studies [26, 137]. Other approaches are based on estimating tectonic event magnitudes employing ML techniques to seismic recordings [89, 108, 97]. In this chapter, relationships between acoustic waves characteristics and underwater seismic event magnitudes are approached using ML algorithms. Once the earthquake moment magnitude and slip type are obtained, relations between fault rupture dimensions and earthquake moment magnitude can be used to estimate seismic hazards [131], where earthquakes are mapped into effective slender geometries [134], which can be verified from aftershock distributions [51].

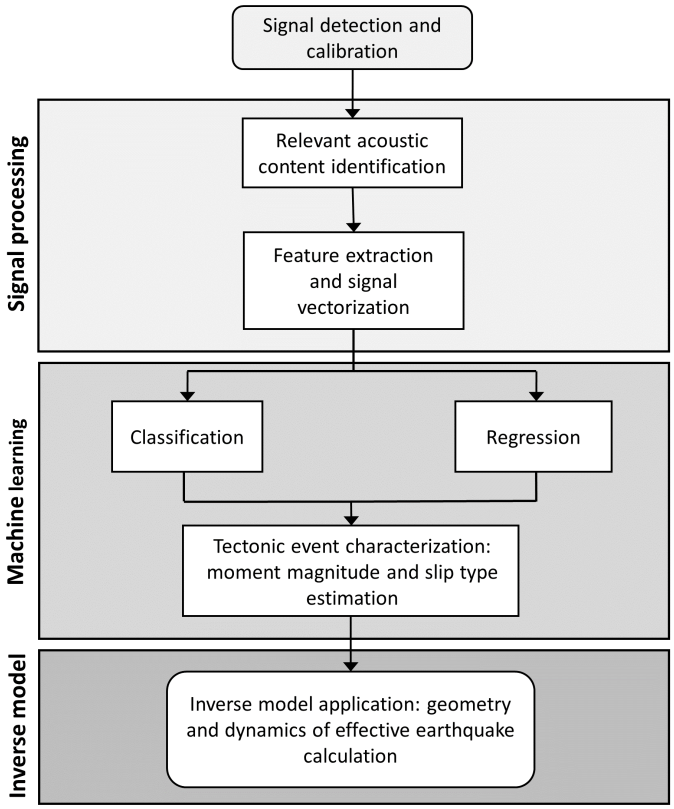

Figure 3.1: Flowchart for the methodology of tectonic event characterization from acoustic recordings analysis.

CTBTO acoustic recordings have been previously studied for identification and detection of T-waves [27, 117] or classification and regression of sound signals related to tectonic events [124, 122, 125]. In this chapter, we analyse 201 acoustic signals related to submarine tectonic events with magnitudes ranging from 5 to 9.1 M_w and different

associated slip types. The signals were recorded by three different CTBTO [33, 45] hydro-acoustic stations located in the Indian and the Pacific Oceans. To characterise tectonic events, first, the associated pressure disturbance is identified in the hydrophone recordings and feature extraction is performed (Fig. 3.1). Feature vectors serve as input to ML algorithms that perform classification of the slip type (existence of significant vertical motion component) and assessment of the magnitude of the event. This information is then used to feed an inverse problem model for acoustic waves that calculates the effective geometry and dynamics of the fault [43].

3.2 Methods

Data and instrumentation

Instrumentation

We used data from three hydrophone stations deployed by CTBTO: *HA01* (Cape Leeuwin), *HA08* (Diego Garcia) and *HA11* (Wake Island). Each station consists of two triplets except *HA01* which has a single triplet [45].

The hydrophones are suspended at a depth corresponding to the SOFAR channel axis and anchored to the seabed via a riser cable, which is kept under tension by a sub-surface buoy [90]. The recordings used in this chapter are extracted from the International Monitoring System (IMS) database, which are originally recorded by the IMS *'H11N1'*, *'H11S1'*, *'H01W1'*, *'H08S1'* and *'H08N1'* hydrophones, see Fig. 3.2. The calibration files for the mentioned instruments, provided by CTBTO, show a steep roll-off below 0.1 Hz in the instrument response curve, and consequently, in this chapter, the analysis of frequency bands below 0.1 Hz is not considered. The data has a sampling frequency of 250 Hz.

Dataset

Long distances from hypocentre to epicentre can induce distortion and attenuation on seismic waves leading to higher uncertainties in measurements. In order to minimise

these effects, we considered only acoustic signals associated with shallow earthquakes. Different values for the maximum hypocentre depth below the seabed are used to define shallow earthquakes in the literature, such as 60 km [105, 34] or 100 km [68]. We only included earthquakes with hypocentre located at less than 60 km deep under the sea bed in the dataset. The dataset was built with 201 acoustic recordings induced by 'shallow' tectonic events, listed in Appendix A. The data was provided by CTBTO.

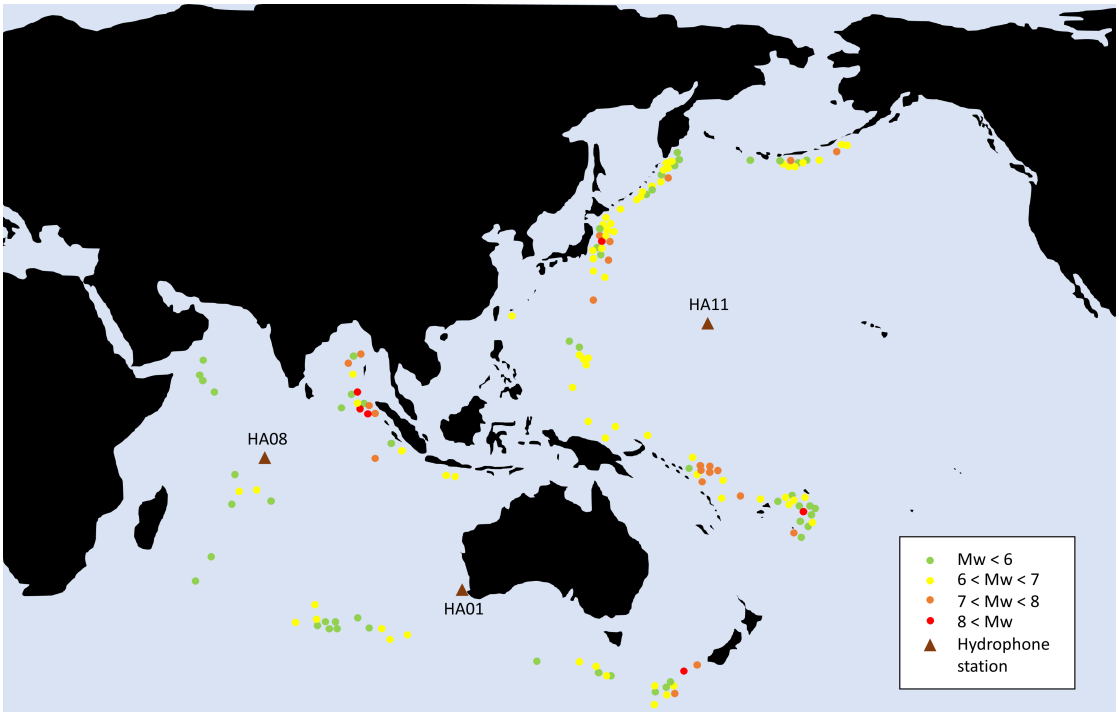

Figure 3.2: Geographic distribution of the earthquake epicentres associated with the signals in the studied dataset. Image generated in Microsoft PowerPoint 2016.

To minimise diffraction effects only earthquakes with shortest transects, that do not cross through lands, were analysed, see Fig. 3.2. In addition to the previously mentioned constraints, only underwater earthquakes are studied, which significantly narrows down the search.

The earthquake's source types of slip and magnitudes were labelled based on data reported by gCMT [28, 29]. The slip type labelling of the dataset depends on the slip angles of the fault planes. For slip angles 0° and 180° the faults were classified as pure

strike-slip, whereas for 90° and -90°, faults were classified as pure dip-reverse and pure dip-slip normal, respectively (±20°).

Digital signal processing methods

Signal vectorization (feature extraction)

Signal vectorization is the process of reducing the size of the acoustic data set by determining category variables that define the sound type or identity of the sound source. It is important to decrease the signal dimensionality, as the training dataset grows exponentially as a function of the number of variables in the feature vector [31].

Previous studies on ML approached the classification of CTBTO acoustic signals by operating on features automatically extracted by the organization [124, 122]. However, the mentioned studies had to handle missing values for some features [125]. To overcome this difficulty, raw acoustic data were analysed and a signal processing algorithm, for signal vectorization was developed, ensuring that there were no missing values in the feature vectors. It has been reported that a single feature cannot train the classifiers efficiently [31]. Thus, we considered and compared five types of features: temporal (obtained directly from the time series) [125, 124, 31], spectral (obtained from the power spectrum) [125, 124, 31], cepstral [125, 23, 87, 124, 31], statistical (statistical moments applied to the times series) [125, 124] and wavelet transform type [118].

Inspired by previous studies [125, 124, 31, 118], four different sets of features were built and tested along with the considered ML algorithms. The first and second sets consist of a combination of four types of features (temporal, spectral, statistical and cepstral); the third set is composed only by wavelet transform extracted features; and the fourth set consists of cepstral features only, as shown in Table 3.1.

	Feature type				
Feature set	**Temporal**	**Statistic**	**Spectral**	**Cepstral**	**Wavelet**
1	Max. Amplitude, total short-time energy, zero-crossing rate.	Kurtosis, skewness.		Maximum, variance.	
2	Max. Amplitude, total energy, zero-crossing rate.	Kurtosis, skewness, std. deviation.	Power spectral density (PSD) mean, PSD coefficient, skewness, roll-off, PSD std. deviation, PSD skewness.	Variance, min kurtosis, max. kurtosis, spectral coefficient, median.	
3					Std. deviation, average short-time energy, power ratio, zero-crossing rate.
4				Variance, min. coefficient, skewness, kurtosis, max. coefficient, median.	

Table 3.1: Studied feature sets, tested on the classification and regression algorithms.

Signal identification

We estimated the signal travel times from source to receiver by dividing the distance between the earthquake epicentre and the recording instrument by the speed of sound, considered constant (c). The centroid times and location coordinates were obtained from the Harvard global CMT database.

For each recording ($x(t)$), the part of the signal carrying most of the information was identified by finding the point with maximum absolute amplitude ($|x(t)|$) and extracting a window composed of N samples at each side of the selected point. Short-time energy [52] analysis was performed to define the potential extracted signal lengths ($2N$), for more details see Appendix A. The expression used to compute the short-time energy (E_f) [52] is 2.23. The time-frames were chosen to be five seconds long ($N_e = 1250$ samples), which was chosen to be shorter than the shortest earthquake duration in the studied dataset.

We set $N = 20000$ samples (160 seconds) as the maximum considered half window length for feature extraction. This choice was made after observing that it exceeds the duration of most signals in the dataset. Several values of N were tested, ranging

from 2500 to 20000 samples, for more details see Appendix A. The minimum tested half window length was 2500 samples, since it was intended to only analyse the frequency behaviour of the signals down to 0.1 Hz.

Frequency behaviour analysis and time series features

To decompose the signals in different frequency bands and study their behaviour at different points of the frequency spectrum, we applied Butterworth band-pass filters, which have been reported to have good performance for classification purposes [50]. Then, the extracted features are statistical moments applied to the filtered time series such as standard deviation, kurtosis, skewness, maximum amplitude and zero-crossing rate. Additionally, the short-time energy distribution was computed for each band, with a five seconds window width, and the maximum and total short-time energy were calculated. Several potential divisions of the frequency spectrum (see Fig. 3.3) were analysed, tested and compared, for more details see Appendix A.

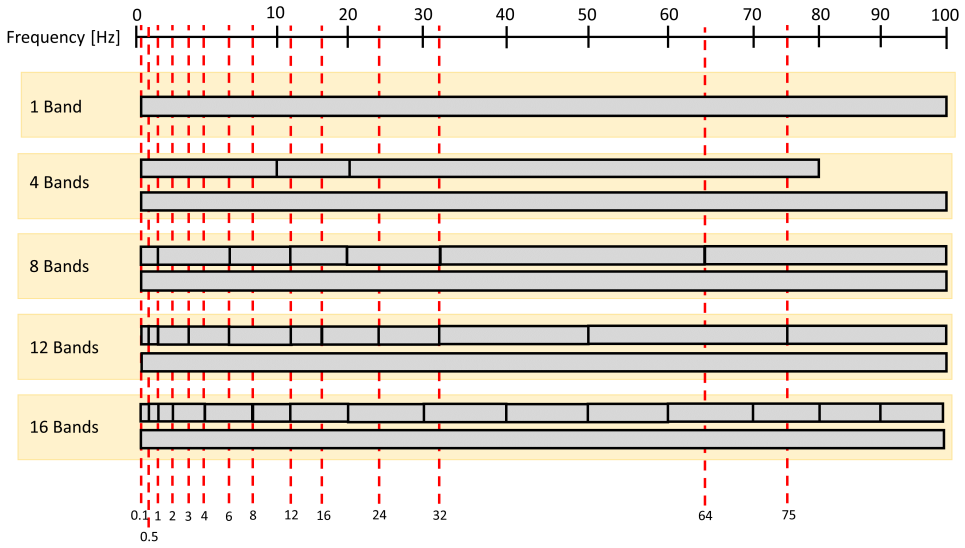

Figure 3.3: Considered frequency spectrum subdivisions.

Based on the fact that earthquakes with vertical motion components excite lower frequencies due to the compression of the water [82], we performed more subdivisions at

lower frequencies than at higher frequencies. Moreover, slow events show lower frequency excitation, in the range of 0.1 to 10 Hz [68]. Some of the tested frequency subdivision approaches taken were inspired by previously published studies [100, 124], see Fig. 3.3.

We analysed the introduced subdivisions of the relevant part of the frequency spectrum and the optimal window size ($2N$) by running the classification models SVM and RFC on the 2-type class labelled dataset. In this analysis, only the set of features '1' was used, along with every combination between the considered window sizes ($2N$) and spectrum division approaches. It can be seen that the standard deviations lie below 15% for SVM and 12% for RFC on the tested cases. In addition, the variations on accuracy along with the different considered scenarios are lower than 3%. Finally, we selected 15000 samples as the half window size (N) and the eight frequency band spectrum division approach for the final set-up. This choice was made based on the results of the carried sensitivity analysis, for more details see Appendix A.

Spectral and Cepstral features extraction

The signals were further processed to extract spectral and cepstral features. The cepstrum can be calculated by taking the inverse Fourier transform after applying a natural logarithm to the Fourier transform of a signal. The logarithm maps convolution in the time domain to addition in the frequency domain.

While for speech the optimum window length for framing is [16-32] ms, for acoustic sounds in the ocean, the window length may be different. Thus, the signal was sliced into frames of ten seconds duration (in order to be able to capture frequencies down to 0.1 Hz). We set 50% overlap [87, 23] amongst the frames and Hamming window was applied [19]. Then, Fourier transform was computed and the power spectrum for each frame was calculated by

$$P = \frac{conj(X(s))X(s)}{NFFT}, \tag{3.1}$$

where $NFFT$ is the transform length in samples of the signal used by the fast Fourier transform, which in this case is 4096 in order to provide enough frequency resolution.

Because of the nature of the FFT algorithm, $NFFT$ was chosen to be multiple of two to make the algorithm more efficient.

The next step was to compute filter banks using triangular filters, which were applied on a scale inspired by the Mel-scale to the power spectrum in order to extract frequency bands. The Mel-scale aims at mimicking the non-linear human ear perception of sound. Cepstral coefficients were used in the literature for the classification of sound sources in underwater environments due to their robustness to noise [125, 124, 31, 23]. Note that the Mel scale was designed for higher frequencies than the bands we are interested in, i.e. 0.1-10 Hz. Therefore, we modified the constants in the conversion formula to have an almost linear mapping from Hertz (f) to Mel frequency (m) in the range of 0.1 to 10 Hz and logarithmic mapping over 10 Hz, see Eq. (A.5). The conversion between Hertz and Mel frequency is done by

$$m = 25.95 \log\left(1 + \frac{f}{7}\right). \tag{3.2}$$

The new constants are obtained from

$$C = \frac{f}{\log\left(1 + \frac{f}{f_0}\right)}, \tag{3.3}$$

where f_0 was chosen to be 7 Hz and $f = 10$ Hz, thus, below 10 Hz the relation between Mel frequency and frequency in Hz is almost linear.

The applied triangular filters are highly correlated due to the overlap, and thus, we used the discrete cosine transform (DCT) to decorrelate the filter coefficients [87]. It is shown in Appendix A that 12 cepstral coefficients capture most of the information carried by the signal. Finally, statistical moments (mean, maximum, kurtosis, skewness and variance) were applied to each coefficient band to reduce the number of features. The spectral features, listed in Fig. 3.3, were extracted from the one-sided power spectrum computed on the calibrated original signals.

Wavelet transform parameters analysis

The power distribution of the signals along with different frequency bands can be analysed by applying the discrete wavelet transform (DWT) [39], of the form

$$W(a,b) = \frac{1}{\sqrt{|a|}} \int_{\infty}^{-\infty} x(t)\psi\left(\frac{t-b}{a}\right) dt, \tag{3.4}$$

where a is the scaling parameter, b is the translational parameter and ψ is the mother wavelet. Here, a given signal is projected onto a space defined by a set wavelets, which are function of frequency and time [53],

Several studies approached the extraction of features by wavelet transform algorithms for different purposes, such as earthquake magnitude prediction by seismic waves analysis [108, 113] or source type classification of acoustic signals [117, 39, 126, 118]. Even though numerous wavelet bases exist, we tested and compared only two discrete wavelets, Daubechies and Symlet, being two of the most popular wavelets in signal processing. Furthermore, the order of the wavelets was also analysed, which indicates the number of vanishing moments and is related to the approximation order and smoothness of the wavelet. After applying the DWT, n sets of detail coefficients and one set of approximation coefficients ($n+1$ times each feature) were produced. Note that $n = 6$ levels, has been used in previous studies [118]. In order to analyse the sensitivity of the ML algorithms to variations in DWT parameters, we tested a different number of levels [4-8] and wavelet orders [2-8], see Appendix A. We found that the accuracy amongst the considered scenarios has a deviation of less than 5%. Nevertheless, we decided to use Symlet wavelet of order eight with six levels of coefficients as the final setup, since it provides a good balance between computational efficiency and accuracy.

Finally, after applying DWT, statistical moments were calculated for each extracted coefficient band or level ($n+1$): standard deviation, average short-time energy, power ratio between the first coefficient band and every other band and zero-crossing rate [126], see Table. 3.1.

Machine learning (ML) algorithms

In essence, ML algorithms learn from data using probability theory and can be grouped into two main categories: supervised learning (labelled dataset) and unsupervised learning (unlabelled dataset). In this study, we applied supervised learning, which can

be further subdivided into classification and regression algorithms based on whether the target outputs are categorical (classification for slip type) or quantitative (regression for magnitude) [70]. In particular, two ML algorithms were explored, SVM and RFC [87, 31, 125, 124].

The first algorithm, SVM, is a classification technique that can additionally be used for regression purposes [127]. It uses a convex cost function, always reaching the global minimum of the cost function [87]. It is able to perform reliable classifications even with small datasets [72]. SVMs differentiate classes by finding the hyper-plane that splits them and maximize the margin between the closest point of each class and the hyper-planes. Each signal is vectorized and described by a point in a n dimensional space and a cost function is fully specified by the subset of training examples called support vectors. The output of the SVM is a set of weights, which in combination will predict the value of the outcome. Because of the non-linear nature of the studied process, a non-linear kernel has been selected, Radial Basis Function (RBF), being this a type of Gaussian kernel, that has been proven to provide good accuracy and efficiency [125]. SVM are commonly found in the literature to classify acoustic signals in the ocean [16, 73, 87, 31, 125, 124, 73, 127].

The second studied algorithm, RFC (introduced in 1995) [49], is a technique based on decision trees that operate as an ensemble. It can perform classification and regression [18] by splitting a dataset into smaller data subsets, while an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. Decision nodes have two or more branches and leaf nodes represent a classification or decision. Each individual tree in the random forest delivers a class prediction and the class with the most votes becomes the model's prediction.

It is important to remark that, regularization (normalization) of the features was carried before the application of SVM, due to significant differences in the order of magnitude between the feature values. However, it is not necessary for the application of RFC. In addition, we used the validation k-fold technique [87] (10-fold) in combination with grid search (5-fold) to identify the best model hyper-parameters and test the accuracy of each algorithm, for more details see Appendix A. Finally, the accuracy of all fold tests

was averaged and precision and standard deviation results were calculated.

Inverse problem model

An array with the potential combinations for x_0 and y_0, which describe the orientation of the fault, can be defined considering that, the detected acoustic radiation induced by the earthquake can be triangulated by the hydrophone triplet to infer the distance to the source. The frequency distribution of the first acoustic mode is then calculated, which is a function of time t (relative to the eruption time t_0), depth h, and every potential orientation of fault (x_0 and y_0). Only the obtained frequency distributions that lie within the range defined by visual inspection of the spectrogram are considered and determine the number of sets of solutions provided by the model.

To identify the beginning and end of the disturbance in the recordings, the short-time energy distribution is analysed. Then, pressure points from the acoustic signal need to be chosen to retrieve the source properties [43]. Envelope tracking algorithms are applied since working with points close to the envelope reduces the associated uncertainties and the obtained pressure points are associated with the calculated potential first mode frequency distributions.

The magnitude estimated by the ML algorithms is used to generate the ranges that feed the inverse problem model and confine the potential solutions for the effective geometry and dynamics of the tectonic event. Finally, the steps described in the Inverse problem model section in Chapter 2, are taken for each of the selected frequency distributions to retrieve b, L, T and W_0. Note that the calculation of ten solutions for each earthquake property and potential frequency distribution is a choice made to balance the computational effort and accuracy and can be modified, see Appendix A. Finally, the solutions are plotted in four probability density functions along with the mean values, see Fig. 3.9. The fault geometry and dynamics are estimated within a few seconds on a standard desktop machine.

3.3 Results

ML algorithms were applied to the acoustic signal properties to estimate two main characteristics of the studied tectonic events: slip type (qualitative) and moment magnitude (quantitative, M_w). Thus, two types of ML algorithms were considered: classification (slip type) and regression (magnitude).

Slip type classification

The primary objective of the classification is to identify the existence of significant vertical motion components in the studied tectonic events. In addition, as a secondary objective, we study the characterization of the type of vertical motion related to the studied earthquakes. Two classification approaches were taken to identify the slip type associated with the tectonic events that generated the acoustic signals composing the dataset: binary and multi-class.

Binary classification

The dataset is divided into two classes. The first class is composed of signals related to tectonic events with vertical motion components (mostly dip-slip), whereas the second is composed of events with relatively small or no vertical motion components (mostly strike-slip). The differentiation between vertical and horizontal events is essential when applying the inverse problem model developed by Ref. [43], which is designed to work for vertical fault displacements. The division of the dataset was made based on source faulting solutions provided by the global CMT catalog [28, 29], resulting in a set distribution of 86 strike-slip earthquakes, which are considered to have mainly horizontal motion components (42.79%) and 115 earthquakes with relevant vertical motion component (57.21%). In that sense, we ensure that the dataset is balanced between events with significant vertical motion components and mainly horizontal slip events.

Feature set	SVM accuracy [%]	RFC accuracy [%]
1	71.62 ± 11.67	71.12 ± 9.75
2	74.10 ± 10.64	75.60 ± 10.64
3	76.64 ± 8.86	78.10 ± 8.74
4	76.12 ± 10.76	73.62 ± 10.76

Table 3.2: Accuracy and standard deviation [%] for binary classification using SVM and RFC on four different sets of features.

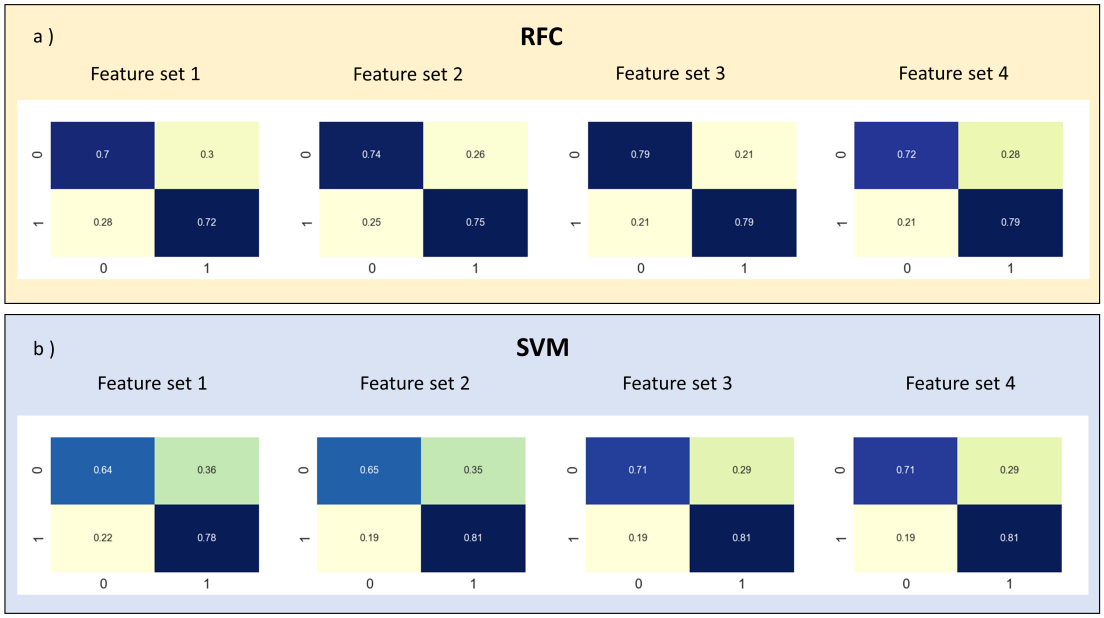

Figure 3.4: Binary classification confusion matrices for the considered feature sets and classification algorithms. '0' stands for events classified as mainly horizontal slip motion and '1' for events with relevant vertical motion component. a) Normalised absolute errors for the RFC application along with 10-fold validation scheme. b) Normalised absolute errors for the SVM application along with 10-fold validation scheme.

In order to characterize the acoustic signals, we applied four different methodologies for feature extraction that are described in detail in the Digital signal processing methods subsection in the Methodology section. In addition, two classification algorithms were applied along with each feature extraction methodology used on the dataset: Random Forest Classifier (RFC) and Support Vector Machines (SVM). In order to test the ML

algorithms, we used a 10-fold validation technique and a 5-fold hyper-parameter grid search, for more details see Appendix A. Results obtained for every fold were averaged to provide final accuracy and standard deviation estimates, see Table 3.2. Moreover, confusion matrices were computed in order to provide classification accuracy for each considered slip type, see Fig. 3.4.

In all cases, there is a higher performance of the ML algorithms for the identification of 'vertical earthquakes' ('1') compared to the identification of 'horizontal earthquakes' ('0'), see Fig. 3.4. This behaviour could be associated with the higher availability of 'vertical earthquakes' in the studied dataset. The overall accuracy of both ML algorithms applied to each feature set is over 70% for every considered scenario. Both ML algorithms show similar accuracy for binary classification of the dataset.

Multi-class classification

The relations between earthquake slip types and their associated acoustic signals are further studied. Here, the dataset was split into three classes [6] based on the source faulting solutions provided by the global CMT catalog [28, 29]. The classes are strike-slip ('0'), thrust or reverse ('1') and normal ('2'). Thus, the 'vertical' class previously defined, was further subdivided into two classes: thrust and normal. This classification led to a dataset distribution composed of 86 strike-slip (42.8%), 62 reverse (30.8%) and 53 normal earthquakes (26.4%). The signal vectorization techniques used in binary classification have been also utilised in this section. RFC and SVM were applied, where a One-versus-all [112] technique was used, generating as many binary classifiers as label types and testing every class against the rest. The resulting classification accuracy and standard deviations are reported in Table 3.3.

The application of RFC to the feature set '3', where only features obtained by applying the DWT are considered, led to the highest observed accuracy results, with an average accuracy for 10-fold validation technique of 64.14% and 8.99%. The normalized confusion matrices suggest that the algorithms classified the majority of the strike-slip and thrust earthquakes accurately ($> 70\%$), though failed to identify most of the normal

earthquakes ($< 30\%$), see Fig. 3.5. The low classification accuracy of normal slip type events is potentially influenced by a low presence of acoustic signals related to this type of earthquake in the dataset, leading to an imbalanced dataset [103]. Nevertheless, the overall classification accuracy ($> 60\%$) for most of the cases reveals statistical significance between the features and the slip types, see Table 3.3.

Features	SVM accuracy [%]	RFC accuracy [%]
1	58.19 ± 7.56	58.69 ± 7.82
2	59.67 ± 6.66	59.14 ± 7.89
3	63.67 ± 11.00	64.14 ± 8.99
4	59.69 ± 10.14	59.62 ± 9.54

Table 3.3: Accuracy and standard deviation [%] for multi-class classification using SVM and RFC.

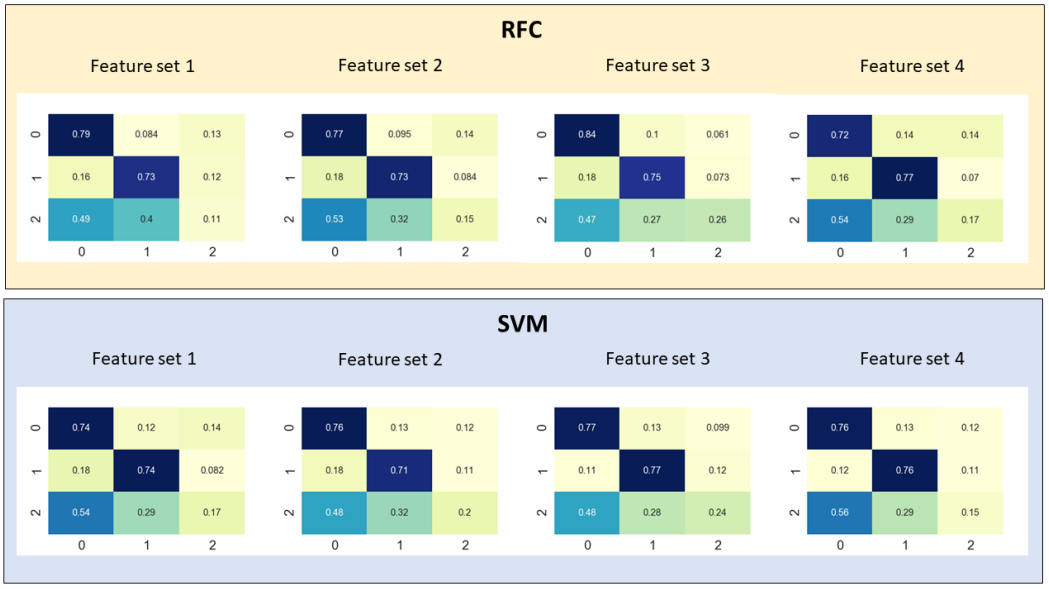

Figure 3.5: Multi-class classification confusion matrices for the tested feature sets and classification algorithms. '0' stands for strike-slip events, '1' for thrust events and '2' for normal events.

In addition, the frequency of each moment magnitude per type of studied earthquake is plotted in Fig. 3.6.

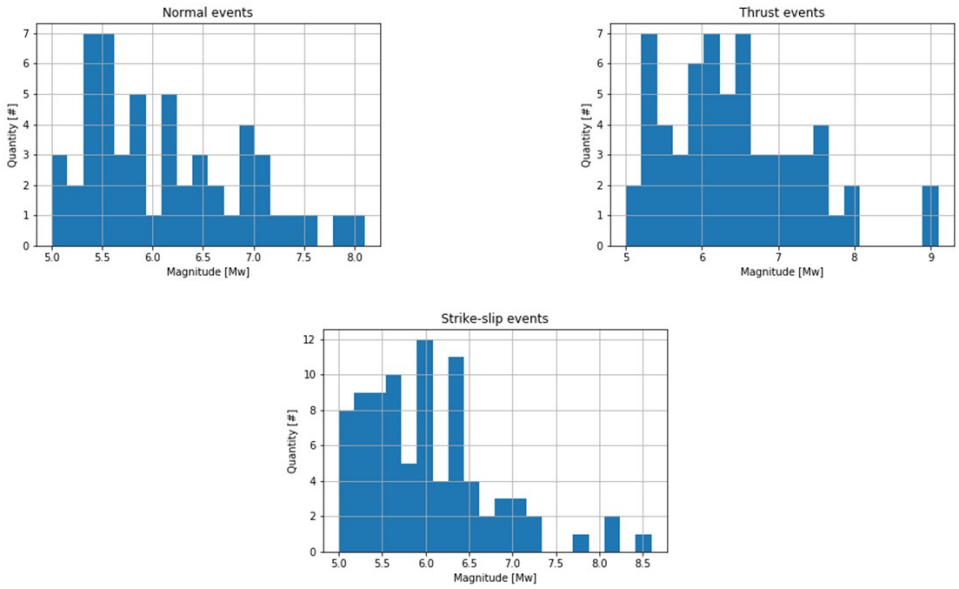

Figure 3.6: Dataset absolute moment magnitude frequency for each studied type of earthquake.

Magnitude regression

In this subsection, we estimate the magnitude of tectonic events by analysing acoustic signals and applying ML algorithms. Support Vector Regressor (SVR) and Random Forest Regressor (RFR) were applied to each of the four different feature sets extracted from the recorded signals that compose the dataset. Additionally, we used 10-fold validation technique and 5-fold hyper-parameter grid search. The magnitudes (M_w) of the events were obtained from the global CMT catalog [28, 29], see Fig. 3.7. Note that since the frequency of occurrence of an Earthquake drops logarithmically with the magnitude, having a perfectly balanced dataset is rather challenging.

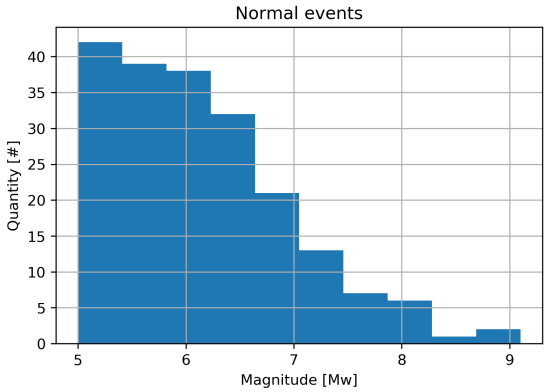

Figure 3.7: Moment magnitude distribution for the studied dataset.

The sum of squared errors (SSE) associated with the results delivered by the ML regression algorithms is calculated by

$$SSE = \sum_{i=1}^{n} (y_i - f(x_i))^2, \tag{3.5}$$

where n is the size of the test set, y_i are the actual values in the test set and $f(x_i)$ are the predictions made by the ML algorithms.

The calculated SSE values for each fold were averaged to provide final estimates for the performance of the regression algorithms, see Table A.7. The SSE, in Eq. (3.5), was applied using the average of the training set as predictor $f(x_i)$, resulting in 13.61, which is about twice the SSE values obtained using the ML models predictions presented in Table A.7. This indicates that the model delivers more accurate predictions than the mean value of the training set. Additionally, the R-squared (R^2) estimator, which represents the part of the variance for a dependent variable that is explained by the independent variables in a regression algorithm, is computed by

$$R^2 = 1 - S_{res}/S_{tot}, \tag{3.6}$$

where S_{res} is the sum of squares of the residual errors and S_{tot} is the total sum of the errors.

	Set 1		Set 2		Set 3		Set 4	
Model	**SSE model**	**R^2 score**	**SSE model**	**R^2 score**	**SSE model**	**R^2 score**	**SSE model**	**R^2 score**
SVR	5.94	0.51	6.07	0.49	7.19	0.42	7.72	0.40
RFR	6.30	0.48	6.54	0.45	7.18	0.42	7.19	0.41

Table 3.4: Calculated SSE for the algorithm estimations against the actual values and R^2 score for each considered ML algorithm and feature set.

The lowest errors were obtained by applying SVR to the set of features '2', which is composed of a combination of temporal, statistic and cepstral features. The observed R^2 scores lie around 0.5, suggesting that the regression algorithms in combination with the extracted features were able to explain some relations between the extracted signal features and the corresponding tectonic event magnitudes, see Table A.7. The calculated and actual values for the magnitudes of the tectonic events are graphically compared in Appendix A. It was observed that the estimations related to magnitude extremes of the dataset ($M_w < 6$ and $M_w > 8$) have higher associated errors.

Earthquake case studies

To demonstrate the applicability of the developed methodology, we have chosen five real case scenarios:

- 16^{th} February 2015 (39.78N 143.22E).
- 14^{th} March 2012 (40.88N, 144.93E).
- 25^{th} October 2013 (27.17N, 144.66E).
- 21^{st} December 2010 (27.10N, 143.76E).
- 29^{th} September 2009 (-15.13N, -171.97E).

The selected earthquakes have a wide range of magnitudes (6.7 to 8.1 M_w) and they are considered independent from the dataset since the ML algorithms applied to them were trained with the other 200 acoustic samples, leaving out one of the studied cases for each

test. Stronger earthquakes were not considered due to the small number of available recordings associated with them. The studied earthquakes in this section are located in the Pacific ocean and their estimated properties are compared against data extracted from the gCMT [28, 29]. Additionally, reports released by the NOAA show that the described earthquakes triggered tsunamis.

The acoustic signals emitted by the earthquakes were recorded by the IMS hydrophone station *'HA11'* situated in the middle of the Pacific ocean and deployed by CTBTO. These earthquake scenarios were chosen to not have abrupt bathymetry changes in the paths between their epicentres and the *'HA11'* hydrophone station, reducing the inverse problem uncertainties. The process taken for retrieving the type of slip, magnitude and effective properties of a tectonic event is detailed in this section for the case of the introduced event 16/02/2015. Then, the results obtained for the remaining four earthquakes are provided in Tables 3.7 and 3.8.

The gCMT [28, 29] reported an earthquake of magnitude 6.7 M_w and coordinates Lat=39.78, Lon=143.22 (Offshore, near the east coast of Honshu), that occurred on 16/02/2015 at 23:06 UTC. The epicentre location is about 2 km deep underwater, with a hypocentre situated 22.2 km under the ground-water interface. The half duration of the event is 5.7 seconds that was catalogued as a thrust earthquake with slip angles: strike=182° and 26°; dip=18° and 73°; slip=68° and 97°. Moreover, the studied event triggered a small tsunami with a wave height of 20 cm, recorded along the coast of Iwate (https://www.ngdc.noaa.gov/hazel/view/hazards/tsunami/event-search) [91].

The distance between the analysed recording hydrophone and the epicentre of the tectonic event is approximately 3200 km, resulting in a calculated travel time for the acoustic waves of 35 minutes before reaching *'H11N1'*, see Fig. 3.8. The extracted hydrophone recordings, displayed in Fig. 3.8, begin at 23:35 UTC and end at 23:50 UTC, the arrival of the disturbance is identified around 23:42 UTC, see Fig. 3.8.

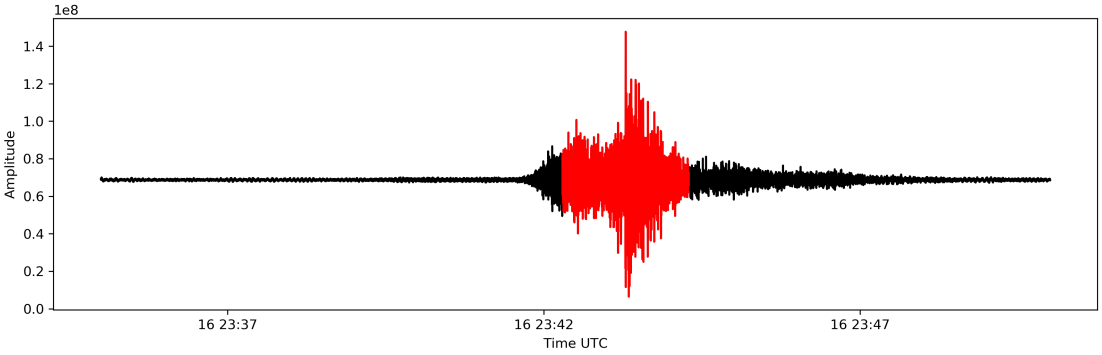

Figure 3.8: Signal related to the studied tectonic event recorded by the *'H11N1'* hydrophone. In red highlighted relevant acoustic disturbance.

The acoustic disturbance was isolated and the four proposed sets of features extracted. The dataset composed of 201 acoustic signal recordings associated with tectonic events was used to train the classification and regression algorithms, since the case study 16/02/2015 is completely external to the dataset, which estimated the slip type of the studied event with 100% accuracy and the magnitude with an associated error lower than 5%, see Tables 3.5 and 3.6.

	Binary				**Multi-class**			
	1st	**2nd**	**3rd**	**4th**	**1st**	**2nd**	**3rd**	**4th**
SVM	1	1	1	1	1	1	1	1
RFC	1	1	1	1	1	1	1	1

Table 3.5: Classification results for the study case using four different feature sets. For the 3-type classification '0' stands for horizontal, '1' for thrust and '2' for normal. For the binary classification '0' represents horizontal, '1' for vertical.

	1st	**2nd**	**3rd**	**4th**
SVR	6.72	6.71	6.79	6.49
RFR	6.79	6.69	6.78	6.40

Table 3.6: Regression results [M_w] for both considered algorithms applied to the four feature sets extracted from the study case signal.

The earthquake's type of motion and magnitude estimations delivered by the ML algorithms, reported in Tables 3.5 and 3.6, are important input parameters for the inverse problem model, see Inverse problem model section in Chapter 2. When the earthquake is classified vertical (binary classification) the inverse problem model can be applied. The potential ranges for the effective fault size and dynamics are calculated by using empirical relations that relate the magnitude of the event with the effective rupture length, width, and ground surface displacement [131]. These ranges are finally fed into the inverse problem model.

Two approaches were taken to generate an input for the inverse problem model and the results were compared, see Fig. 3.9. In the first approach, the magnitude of the event was not taken into account and the inverse problem model was fed with the total observed range of values for each effective earthquake property by Ref. [131]. Thus, for each property the range spans from the minimum to the maximum observed values, half effective length of the fault $L = [1 - 400]$ km; half width $b = [1 - 90]$ km; maximum surface displacement $d = [0.05 - 10]$ m; half duration $T = [0.1 - 25]$ s [28, 29].

The length of the effective displacement $2L$ was estimated from figure 9 in Ref. [131], the effective area from figure 16 in Ref. [131], allowing the further calculation of the width $2b$. The maximum vertical displacement was estimated from figure 11 in Ref. [131]. Note that the predictions for the characteristics of the earthquake are based on empirical values, which may include significant scatter [131]. Ref. [131] regressions were made based on a dataset composed of more than 400 earthquakes with moment magnitude ranging from 5.2 to 8.1 M_w.

The second approach is based on the regressions developed by Ref. [131], that relate moment magnitude (M_w) with effective properties of the earthquakes by:

$$\begin{aligned} \log SRL &= c_1 + c_2 M_w, \qquad c_1 = -3.22, \quad c_2 = 0.69, \\ \log RA &= c_3 + c_4 M_w, \qquad c_3 = -3.49, \quad c_4 = 0.91, \\ \log MD &= c_5 + c_6 M_w, \qquad c_5 = -5.46, \quad c_6 = 0.82, \end{aligned} \tag{3.7}$$

where SRL stands for surface rupture length, RA for rupture area (Table 2A of Ref. [131]), and MD for maximum displacement (Table 2B of Ref. [131]).

The applied parameterizations provided a single value for each effective fault property, which was divided by 2 and multiplied by 2.5 to define the limits for the input ranges of the inverse problem model. These coefficient values are set after visual inspection of the information published by Ref. [131] to simulate the scatter found in the data [43].

In the spectrogram associated with the pressure signal, we identified maximum and minimum limits for the potential values of the frequency distribution of the first acoustic mode at 0.8 and 10 Hz respectively. With this information, the algorithm identified four potential first mode frequency distributions that lie within the established frequency range. Then, the relevant part of the acoustic disturbance was selected by inspection of the short-time energy distribution. Finally, the model was run producing a set of 40 solutions (a set of 10 solutions was calculated for each potential orientation of the fault) for each property of the effective ground uplift associated with the tectonic event: L, b, T and uplift speed W_0, see Fig. 3.9.

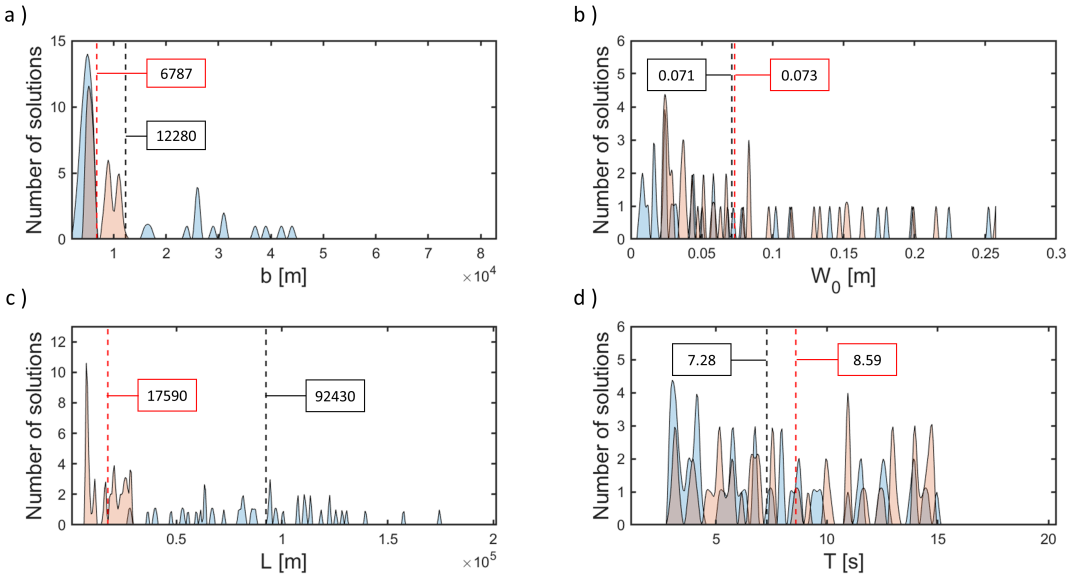

Figure 3.9: Probability density functions for the effective slender fault characteristics calculated by the inverse problem model. In blue the results were calculated without taking into account the moment magnitude of the studied tectonic event and in orange, the inverse problem model was fed with the regressions developed by Ref. [131]. Dashed vertical lines indicate the averaged values for each approach. a) Results for the half width, b, b) Results for the uplift speed, W_0, c) results for the half length, L and d) results for the half duration, T.

The regressions found in Ref. [131] were applied to the estimated magnitude of the event (M_w) leading to: SRL = 25.29 km, RA=404.6 km^2, width b = 15.99 km and MD = 1.08 m. The retrieved properties by the inverse problem model fed with the total observed range of values in Ref. [131] for the effective properties (where the magnitude of the event was not considered as input) are L = 92.43 km, b = 12.28 km, T = 7.28 s and d = 1.03 m ($2TW_0$). For the case where the model is fed with property ranges based on the regressions found in Ref. [131], the retrieved values are: L = 17.59 km, b = 6.78 km, T = 8.59 s and d = 1.25 m, see Fig. 3.9. Both approaches for the input of the model led to sets of results that lie in the same order of magnitude. The calculated effective half length of the fault, in the case of the magnitude of the event not taken into account

as input, is significantly larger than for the case where the magnitude is used as input. For the remaining four introduced earthquakes, only the '1st' and '2nd' sets of features are used along with RFR and SVR, since they led to the best regression accuracy results, and the obtained magnitude results averaged to provide a single final value. The '2nd' and '3rd' sets of features led to the highest classification accuracy. Thus, they are applied and the resulting classifications compared, see Tables 3.7 and 3.8. Note that only the binary classification result is used as input for the inverse problem model. The average absolute binary classification error and the absolute mean regression error are 0% and 2.382%, respectively. In each of the presented case studies, the utilised ML algorithms were trained with the remaining 200 earthquakes from the dataset, to ensure that the studied cases are independent of the ML training set.

Earthquake	M_w	**Retrieved** M_w	**gCMT**	**Multi-class**	**Binary**	**h [m]**	**Distance [km]**
16/02/2015	6.700	6.727	Thrust	(T) 100%	(V) 100%	4000	3200
21/12/2010	7.400	7.072	Normal	(N) 100%	(V) 100%	4500	2500
25/10/2013	7.100	7.017	Normal	(N) 100%	(V) 100%	5000	2900
14/03/2012	6.900	6.740	Normal	(T) 50% - (N) 50%	(V) 100%	5000	3150
29/09/2009	8.100	7.810	Normal	(N) 75%	(V) 100%	4000	4450

Table 3.7: Results delivered by the ML algorithms. Distance refers to the hydrophone-epicentre separation, (V) stands for vertical earthquake, (H) for horizontal, (N) for normal and (T) for thrust. gCMT are the reported slip types. h is the average ocean depth between the epicentre and hydrophone.

Earthquake	**L [m]**	**b [m]**	**T [s]**	W_0 $[ms^{-1}]$	**Time/run [s]**	**Runs [#]**
21/12/2010	26500	9600	17.71	0.035	4.85	4
25/10/2013	26400	9400	27.7	0.02	2.3	3
14/03/2012	14800	7200	23.5	0.013	2.6	3
29/09/2009	93100	11300	12.73	0.135	2.92	5

Table 3.8: Properties retrieved by the inverse problem model. It is reported the number of runs (sets of solutions, 10 solutions each set) carried by the model and the average computational time taken by each of the runs.

Synthetic signal analysis

In this section, we report the performance for the application of classification algorithms on synthetic pressure signals generated by Eq. (3.8) found in Ref. [82]. The applied ML algorithms were trained with real earthquake acoustic signals. The bottom pressure signals induced by the acoustic waves in the far-field are described by

$$p = \rho W_0 |A| \frac{2^{7/2} c}{\sqrt{\pi^3 x_0 k}} \sin(kb) \sin\left(\hat{\Omega} T\right), \tag{3.8}$$

where ρ is the water density, A is the two dimensional envelope[82], k is the wave number and $\hat{\Omega}$ is the frequency. Note that only the pressure induced by the first acoustic mode is taken into account, as it carries most of the energy and information about the source [82]. The analytical solution used to generate the synthetic signals takes simplifications such as constant speed of sound or rigid and flat seabed which are discussed in other studies [43, 82].

The slender fault properties and relative coordinates used to calculate the pressure signals by Eq. (3.8) were randomly generated within the ranges observed by Ref. [131]: water column depth $h = [3000 - 4000]$ m; Total distance from the centre of the fault to the virtual hydrophone $r = [1000 - 4000]$ km; $L = [10 - 200]$ km; $b = [2 - (L/8)]$ km; $d = [0.03 - 10]$ m; $T = [2 - 20]$ s. 20 synthetic signals were generated and windows of 300 seconds extracted after the arrival of each pressure disturbance to the virtual hydrophone. Then, the four types of signal vectorization considered in this chapter were performed and binary classification was applied with a 5-fold hyper-parameter grid search. As a result, the RFC and SVM algorithms identified the synthetic signals as incoming from vertical motion earthquakes ('1') with an accuracy of 100%, as expected.

We studied the effects of adding and shuffling different numbers of synthetic signals into the dataset (composed only by real earthquake recordings) used to train the classification algorithms, for more details see Appendix A. The overall accuracy of the classification algorithms increased due to the addition of synthetic signals to the training dataset. However, when the amount of added synthetic signals became $\approx 10\%$ (20 synthetic signals

in the studied case) of the dataset size, a trend of increased bias was noticed, leading to a higher incorrect classification of the horizontal slip events ('0'). Thus, it is recommended to include a number of synthetic signals smaller than 10% of the total dataset size in order to optimise the classification accuracy for both tectonic event slip types.

3.4 Concluding remarks

In this chapter, we applied a set of techniques capable of analysing acoustic pressure signals induced by underwater earthquakes and calculating the effective fault size and dynamics in almost real-time. To fulfil this goal, we studied a dataset composed of 201 earthquake signals recorded by the IMS hydro-acoustic network. Furthermore, we compared four different methodologies to extract relevant features from acoustic signals incoming from submarine earthquakes, based on statistical moments, time series analysis, power spectrum analysis, wavelet transform coefficients analysis and cepstral coefficients were considered and compared. Along with the vectorization methodologies, we applied two classification ML algorithms (RFC and SVM), which were able to discriminate vertical motion events with over 70% classification accuracy. Amongst the tested methodologies, the wavelet transform feature extraction technique in combination with SVM led to the highest classification results accuracy for both binary and multi-class scenarios.

Regarding the three-type classification, included as a secondary result, there is a low classification of 'normal' events, which can be caused by an unbalanced dataset [103], where the balance between 'horizontal' and 'vertical' events was prioritised, leaving around half of the set to be split into the two new classes (thrust and normal). Additionally, regression ML algorithms were applied to the vectorized signals dataset to estimate the magnitudes of the associated tectonic events. The ML algorithms delivered better predictions than the mean value of the dataset, which was confirmed by the SSE values. It is remarkable that the precomputed vectorized dataset along with the ML algorithms takes less than one second on a standard desktop machine to deliver the source magnitude and slip type estimations. Finally, the magnitude and slip type retrieved by the ML

algorithms can be used to feed an inverse problem model to perform real-time calculations of the fault effective size and dynamics.

Chapter 4

Numerical validation of slender fault solution

4.1 Introduction

Early estimations for earthquake source characteristics coupled with modelling tools, that can accurately predict tsunami wave heights before their arrival to coastal locations, have become important in assisting civil defence agencies in issuing proper warnings during tsunami events. However, there are various challenges that early tsunami warning systems need to overcome, such as the classification of the earthquake type and magnitude and the estimation of the fault dynamics and geometry which require real-time analysis of the source.

Several studies have previously approached the calculation of effective source properties from the generated acoustic signature [48, 63, 82, 43]. In this chapter, we made use of the solutions delivered by the inverse problem model for acoustic waves developed in Chapter 2 [43] coupled with machine learning algorithms capable to estimate the magnitude and the slip type of tectonic events from their acoustic signature [42], see Chapter 3.

The epicentre location and magnitude of tectonic events can be determined relatively accurately and quickly after an earthquake is triggered by seismometer records analysis.

However, other fault plane parameters can be difficult to determine and may remain unknown such as the slip distribution [119]. Since the fault plane parameters are required inputs for the tsunami prediction models, it is important to evaluate the sensitivity of the tsunami prediction to the uncertainty in these parameters. The fault dimensions, the slip displacement, and the strike angle can affect the wave heights significantly, even thousands of miles away from the epicentre [40]. For the evaluation of the initially displaced water volume the dislocation theories developed by Ref. [80] and Ref. [98] are commonly used. These theories are based on linear elasticity theory and predict the seafloor displacement, assuming rectangular rupture, given the fault plane characteristics described by several parameters [40].

On the other hand, improvements in the availability and quality of oceanographic observations and techniques increase the potential for tsunami warnings based on numerical forecasts and statistical algorithms such as machine learning. Tsunami observations are often used in the inversion procedure for constructing finite fault slip distribution models [69, 85] or tsunami source models of initial water elevation [110, 123]. Accurate estimates of fault parameters obtained immediately after a tectonic event occurs are an essential prerequisite for many tsunami early warning systems [76] and inundation models. Once a tsunami is generated by an earthquake, its propagation and inundation can be simulated and studied by hydrodynamic tsunami models [106], such as the COrnell Multi-grid COupled Tsunami model (COMCOT) [77, 129, 107, 5, 76, 133, 40]. Furthermore, it can be found in the literature the application of numerical models to simulate the generation of tsunamis based upon a moving slender-fault geometry [107, 40].

Deep Ocean Assessment and Reporting of Tsunamis (DART) buoys deployed by the National Oceanic and Atmospheric Administration (NOAA) are an array of sensors that measure bottom pressure changes in deep waters. They can be used to track sea-level changes in the open sea during a tsunami event [41]. Furthermore, these recordings can be analysed and used to validate Tsunami propagation numerical models [44, 107, 5, 76].

In this chapter, we considered four past tsunami scenarios to validate the solutions calculated by the introduced inverse problem model for acoustic waves based on a slender

fault approximation to represent real earthquakes [43, 42]. The introduced model provided an almost real-time calculation of the source dynamics. Then, the retrieved properties of the ground uplift were used to calculate the initial water displacement and applied as the initial condition for tsunami propagation simulations. Finally, finer bathymetry and wave non-linearity were included to provide a better prediction of tsunami waveforms in coastal areas.

4.2 Methodology

The inverse problem model developed in Chapter 2 [42] and the ML algorithms studied in Chapter 3 [43] were applied to acoustic signals recorded by distant hydrophones and the solutions were analysed. Note that the inverse problem model considers only the vertical element of the fault dynamics, which is required to elevate the water volume, hence generating a tsunami. Therefore, the fault is referred to as an *effective* rectangular slender fault of a constant vertical speed (see justification in Ref. [82]). To estimate the potential tsunami wave heights at distant locations induced by tectonic events, the solutions of the inverse problem model for acoustic waves are used as input for tsunami propagation models, see Fig. 4.1

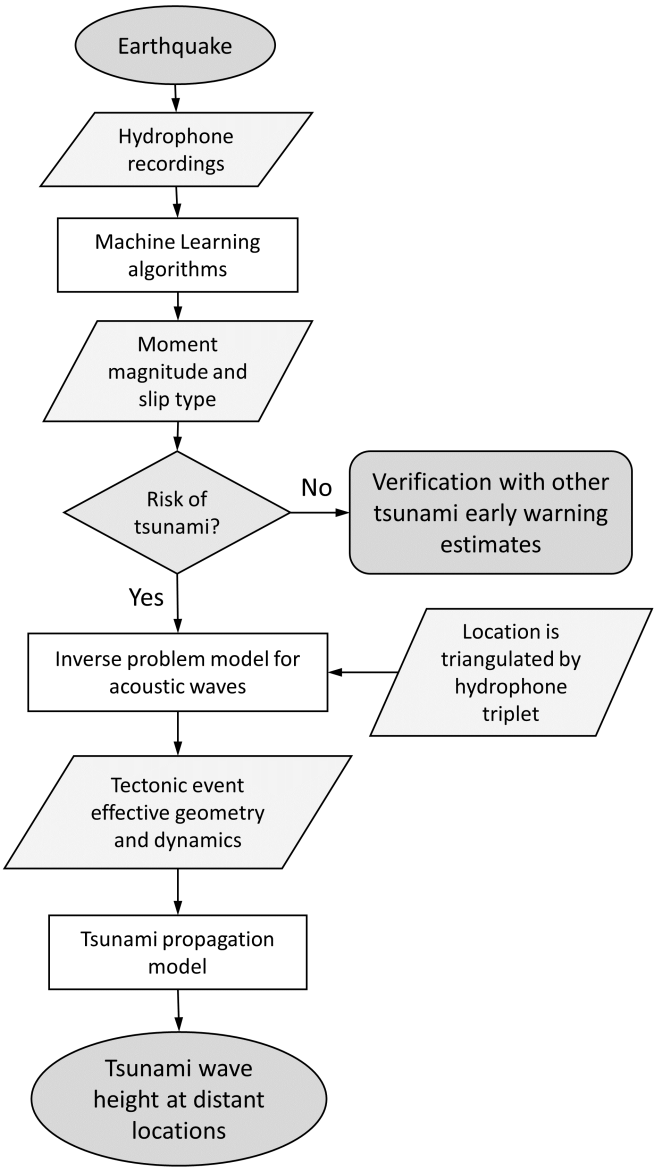

Figure 4.1: Flowchart for the methodology utilised from the recording of the signal to the estimation of the potential wave heights at chosen locations.

Regarding the orientation of the effective slender fault, two potential strike angles are retrieved by the inverse problem model for each study case due to the symmetric nature of the slender fault problem [81], see Fig. 4.2. The strike angle (orientation of the

fault) is an important variable for the numerical modelling of tsunami waves propagation that affects significantly the wave height in specific locations [40]. Thus, both retrieved strike angles must be taken into account and analysed. It is recommended to give more weight to the highest calculated wave height amongst both strike angle scenarios since over-warning is preferable to under-warning for tsunami scenarios [44].

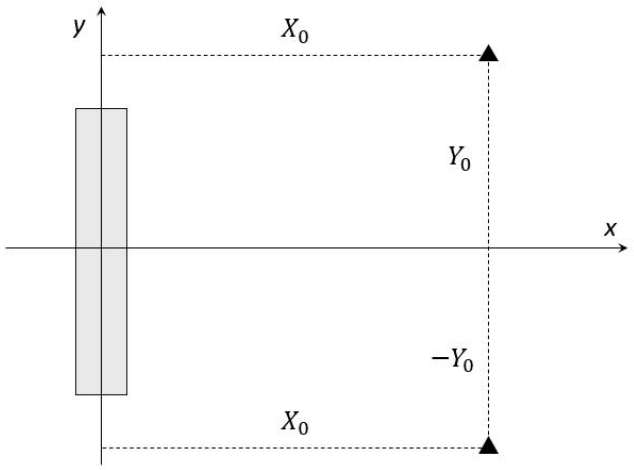

Figure 4.2: Top view sketch of a slender fault. Two black triangles represent the two hydrophones that record the same pressure signal.

Numerical model: COMCOT

To validate the calculated slender fault properties provided by the inverse problem model, we used the COMCOT numerical model. COMCOT is based on the depth-integrated Navier-Stokes equations [77]. It uses a leap-frog finite difference scheme on two variations of the Boussinesq shallow water equation (SWE), linear SWE and non-linear SWE, which are given in Appendix A.

To mimic seafloor deformation, an effective uplift of the ground with constant vertical motion was simulated by using the transient seafloor deformation model in COMCOT [82, 132]. It assumes that the volume of water displaced by the earthquake is equal to the volume of the deformed seabed. Moreover, open boundary conditions were applied

and the setup parameters were chosen following past studies, the Manning coefficient was fixed at 0.03 [5]. The bathymetries were obtained from the Generic Mapping Tools [130]. To compare with DART buoy data (in open ocean locations), non-linear terms were neglected [5, 81], thus, the linear version of the SWE was applied. Spherical coordinates were used with a grid resolution of three arc minutes in both x and y directions and the time step was set at three seconds [44, 120].

When the tsunami waves propagate from offshore to nearshore areas, the non-linear effects increase significantly. Thus, to compare with tide gauge data, where the water depth is shallow, the non-linear version of the SWE was chosen with a finer grid resolution of 15 arc seconds in x and y directions [81]. For this purpose, COMCOT is able to simulate multi-size grids (nested grid method).

4.3 Earthquake case studies

To validate the inverse problem model we considered four past tsunami scenarios from the dataset reported in Appendix A:

- 14^{th} March 2012 (40.88N, 144.93E),
- 25^{th} October 2013 (37.17N, 144.66E),
- 21^{st} December 2010 (27.10N, 143.76E),
- 29^{th} September 2009 (-15.13N, -171.97E).

The selected earthquakes have a wide range of magnitudes amongst the earthquakes that can trigger tsunamis, from 6.9 M_w to 8.1 M_w. Bigger earthquakes were not considered due to the small amount of available data that can be used to train the ML algorithms. The studied earthquakes are located in the Pacific ocean and their estimated properties were compared against data extracted from the gCMT [28, 29]. Moreover, all four selected test cases triggered tsunamis[91]. We only considered shallow earthquakes, which means that the hypocentre is situated less than 60 km below the ground surface to minimise the uncertainties associated with the inverse problem model and ML algorithms [42].

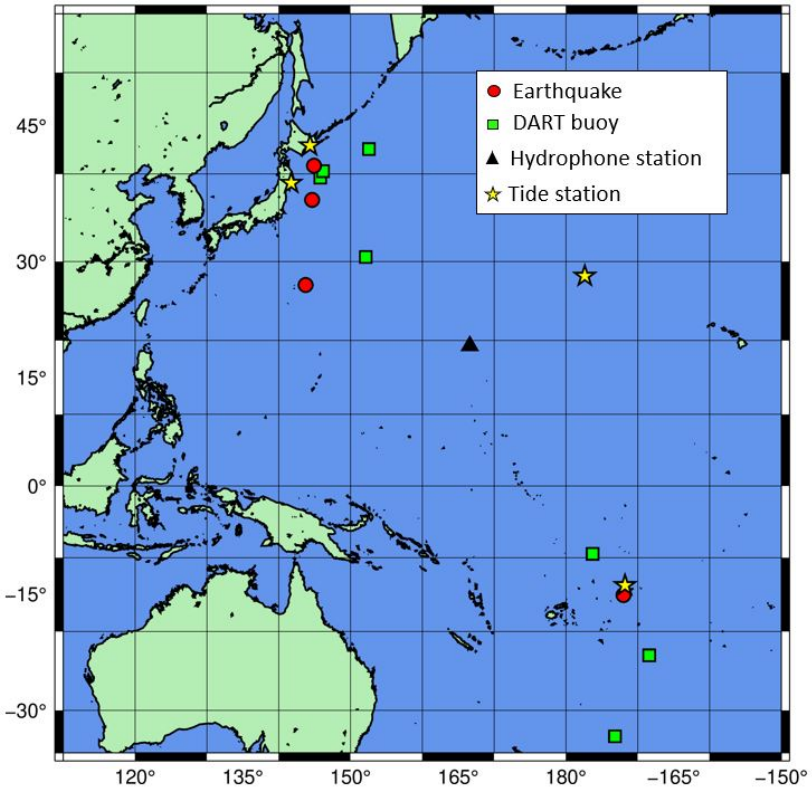

Figure 4.3: Map with the locations of the studied earthquakes, hydrophone stations, tide stations and DART buoys. The map was generated using GMT [130]

The acoustic signals emitted by the earthquakes were recorded by the IMS hydrophone station *'HA11'*, located in the middle of the Pacific ocean and deployed by the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO). Also DART buoy data provided by NOAA [92] was used for the validation, measured by the stations *'21413'*, *'51425'*, *'51426'*, *'21413'*, *'54401'*, *'21346'*, *'21347'* and *'21401'*, located in the Pacific ocean. In addition, four tide stations, which are situated in Kushiro, Ofunato, Midway island and Apia Upolu [32], have been analysed and the extracted water surface time series were compared against the carried multi-grid non-linear tsunami propagation simulations.

The case studies were chosen specifically since there are no abrupt bathymetry changes in the transects connecting the earthquake epicentres and the *'HA11'* hydrophone stations.

Thus, the signal to noise ratio can be maximised and the inverse problem uncertainties are reduced, see Fig. 4.3.

4.4 Results

inverse problem model

The classification results obtained from the application of the ML algorithms and the inverse problem model to the four considered tectonic events in this chapter are utilised to define the initial conditions for a tsunami propagation model, see Tables 3.7 and 3.8. In addition, the inverse problem model retrieved the coordinates of the hydrophone relative to the effective fault centre (X_0 and Y_0), which were averaged to provide two final strike angles, see Appendix A.

Earthquake	X_0 **[km]**	Y_0 **[km]**	**Strike angle** [°]
21/12/2010	1500-2400	2000-700	43 and 175
25/10/2013	2600-2800	1284-754	70 and 200
14/03/2012	2900-3100	1229-559	70 and 208
29/09/2009	4300	1150	254 and 41

Table 4.1: Earthquake retrieved properties by the inverse problem model, X_0 and Y_0 are the ranges for the coordinates of the hydrophone relative to the epicentre that suit the inverse model conditions for the first mode frequency distribution, then two strike angles are provided.

As mentioned before, two strike angles were provided by the model, found in Table 4.1, the first was named 'strike angle 1' and the second, 'strike angle 2'.

The hydrophone coordinates relative to the epicentre of the earthquake were averaged and plotted over a map to calculate the final strike angles, that were used as input in COMCOT, see Appendix A. However, two potential orientations of the fault were obtained for each case study. Thus, both potential orientations had to be considered and used by COMCOT as initial conditions as mentioned above. Finally, the results of the water waves propagation were compared against DART buoy data. Under normal

conditions, DART buoys record the sea level every 15 min. However, during major earthquakes, the buoy stations measure sea-level changes every 15 s.

To produce the initial condition for the model, the points in the bathymetry corresponding to the area of the calculated slender fault were averaged. Then, the vertical constant uplift velocity retrieved by the inverse problem model (W_0) for the calculated uplift duration ($2T$) was applied to the averaged area and snapshots every 0.2 seconds of the deformation of the ground were taken and used as input to the COMCOT numerical model, see Fig. 4.4.

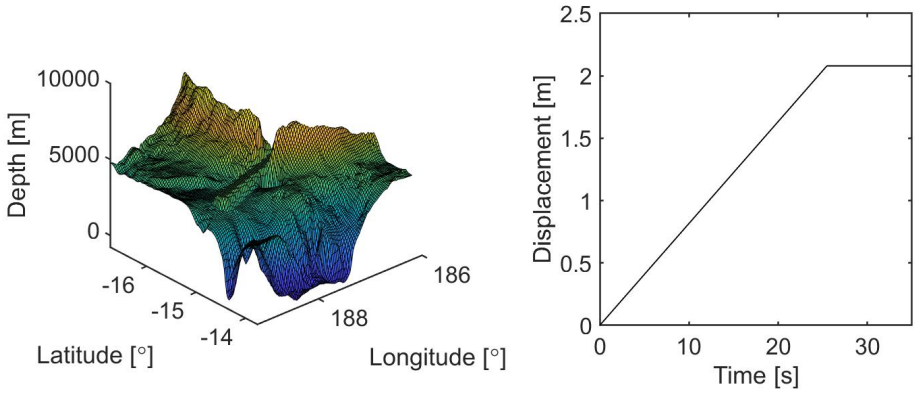

Figure 4.4: Left: The averaged fault area. Right: The averaged area uplift along time.

COMCOT linear simulations

The bathymetries used in the simulations were extracted from the GMT (Generic Mapping Tools) database [130]. The calculated potential fault orientations and dimensions were compared against the actual bathymetry, see Figs. A.26, A.27, 4.7 and 4.8, where both potential fault orientations for each case are displayed in red and yellow color to be differentiated. It has been observed that in the majority of the studied cases the direction of the slender fault is aligned with the tectonic borders.

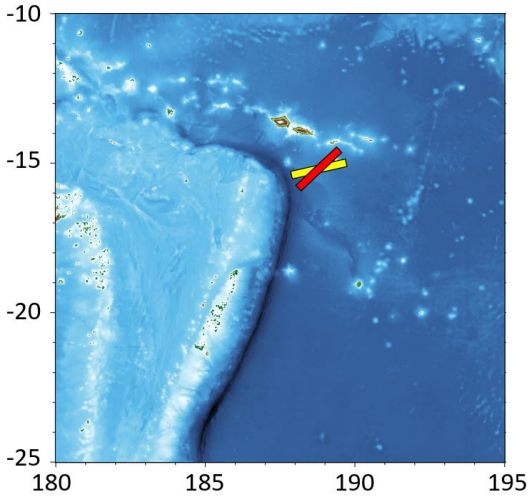

Figure 4.5: Potential orientations of the fault 2009.

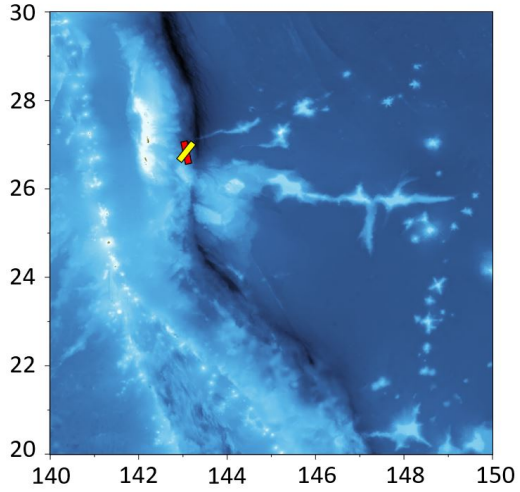

Figure 4.6: Potential orientations of the fault 2010.

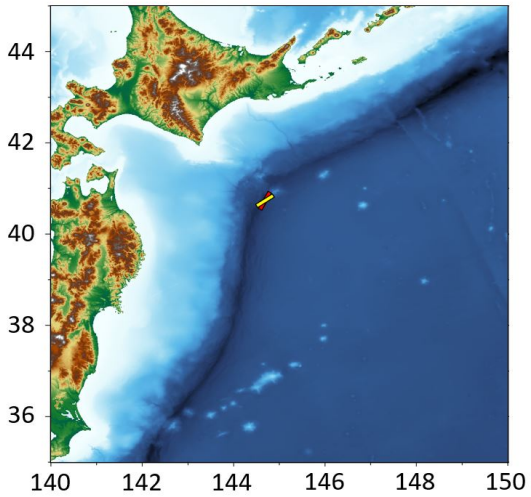

Figure 4.7: Potential orientations of the fault 2012.

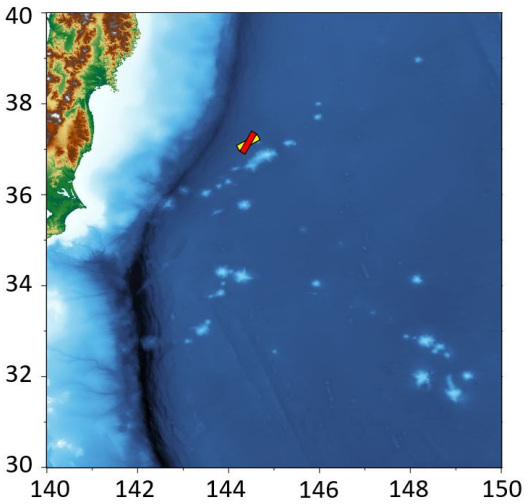

Figure 4.8: Potential orientations of the fault 2013.

The water waves generation associated with the presented tectonic events was simulated using the initial conditions calculated in the previous section and then propagated to the considered DART station locations. The resulting free surface elevation time series were compared against the DART buoy recordings, see Figs 4.9, 4.10, 4.11 and 4.12.

Because tsunami wave amplitudes are much smaller than the water depth in the open sea, the wave non-linearity is negligible and can be ignored (see Ref. [81] pages 13 and 14).

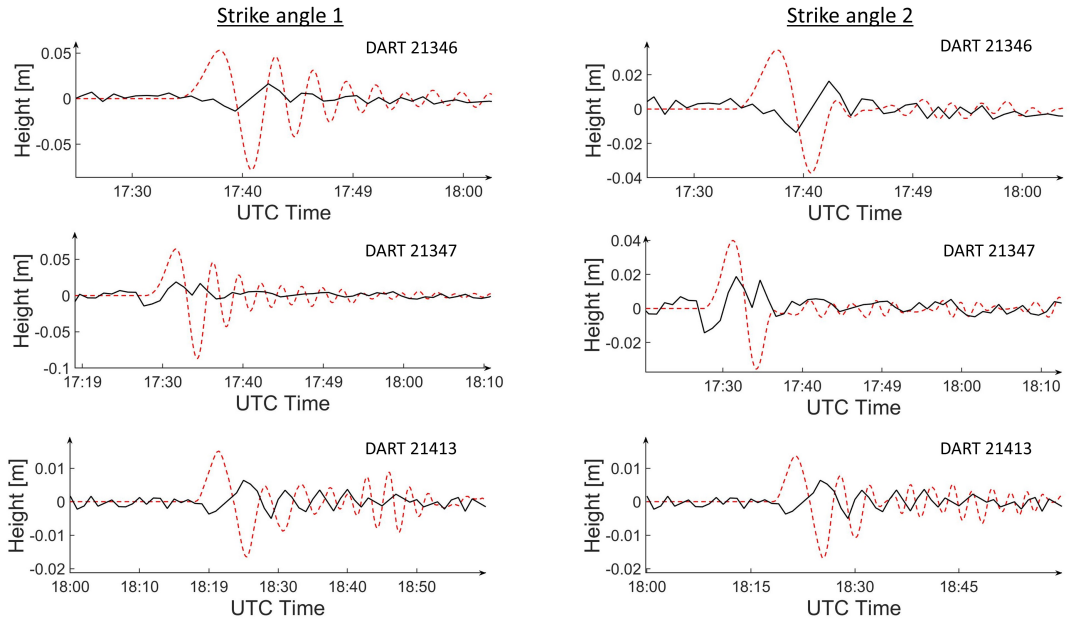

Figure 4.9: Surface water elevation comparison between the DART observations (black) and the numerical (red dashed) simulations for the 29-09-2009 earthquake.

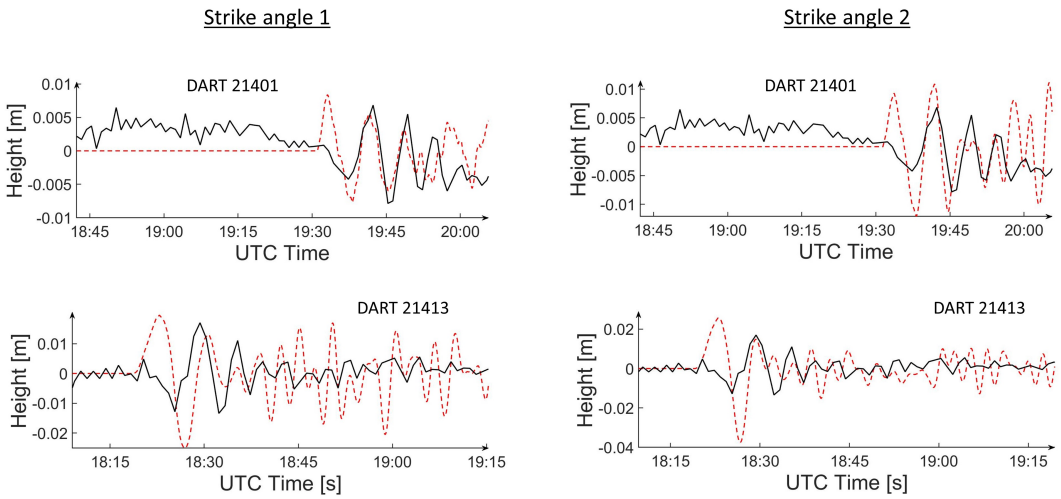

Figure 4.10: Surface water elevation comparison between the DART observations (black) and the numerical (red dashed) simulations for the 21-12-2010 earthquake.

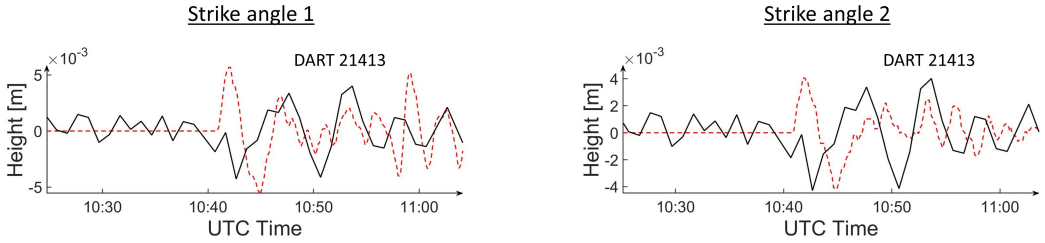

Figure 4.11: Surface water elevation comparison between the DART observations (black) and the numerical (red dashed) simulations for the 14-03-2012 earthquake.

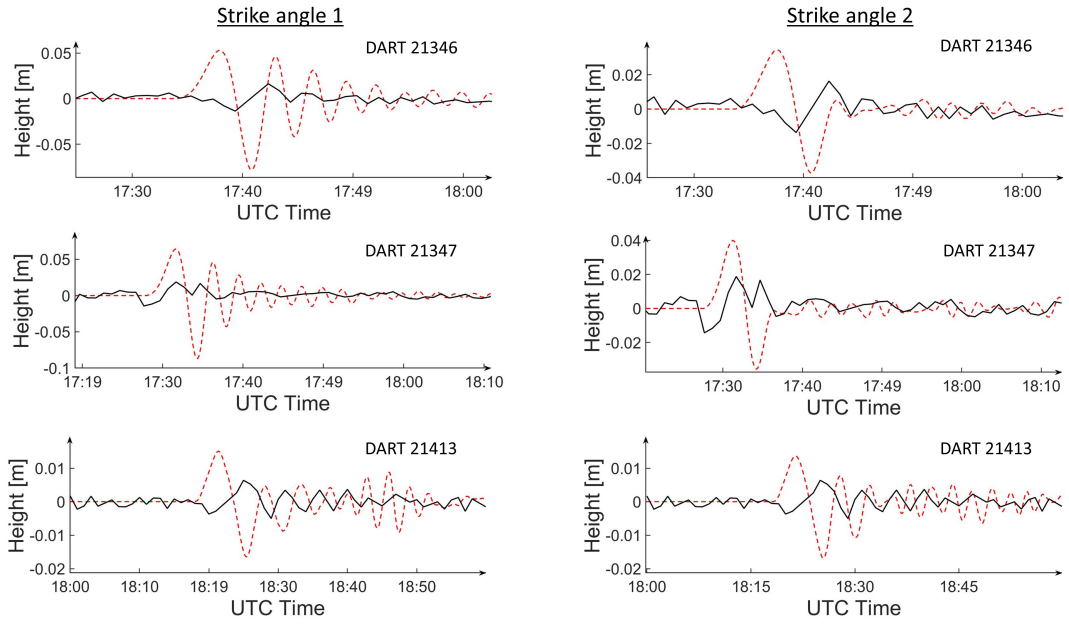

Figure 4.12: Surface water elevation comparison between the DART observations (black) and the numerical (red dashed) simulations for the 25-10-2013 earthquake.

The maximum wave amplitudes recorded by the DART buoys are compared against the calculated maximum waves by the numerical model in Table 4.2.

Earthquake	**DART station**	**DART observation [m]**	**COMCOT (Strike angle 1) [m]**	**COMCOT (Strike angle 2) [m]**
29/09/2009	51425	0.0341	0.053	0.052
	51426	0.058	0.070	0.120
	54401	0.017	0.015	0.021
25/10/2013	21346	0.0163	0.052	0.034
	21347	0.0187	0.064	0.0394
	21413	0.0063	0.0148	0.0135
21/12/2010	21413	0.0195	0.017	0.025
	21401	0.0068	0.0083	0.011
14/03/2012	21413	0.0044	0.019	0.0056

Table 4.2: Maximum wave height [m] comparison between observed DART recordings and numerical simulations.

The strike angle is observed to be a defining factor in the far-field maximum wave height, inducing variations up to 30%. For some of the studied cases, the maximum

wave amplitude was replicated with an uncertainty smaller than 15%, e.g. 20/09/2009, DART 54401, 'Strike angle 2'. However, in the worst-case amongst the tested scenarios, the uncertainty can extend to almost twice the recorded value, e.g. 25/10/2013, DART 21347, 'strike angle 2'.

COMCOT non-linear simulations

To further validate the obtained slender fault solutions, the generated tsunami waves were propagated into shallow water locations. The specific locations were chosen to be near-shore tide stations, which have a sampling period of 60 seconds. For this purpose, nested grids were used in the simulations, where finer grids of 15 arc seconds resolution, that cover the areas around the studied tide stations, were coupled to the general and wider 3 arc minutes resolution grids used for the linear propagation. In this subsection non-linear terms were taken into account. Finally, the simulated water surface time series were compared against tide gauge recordings, see Fig. 4.13.

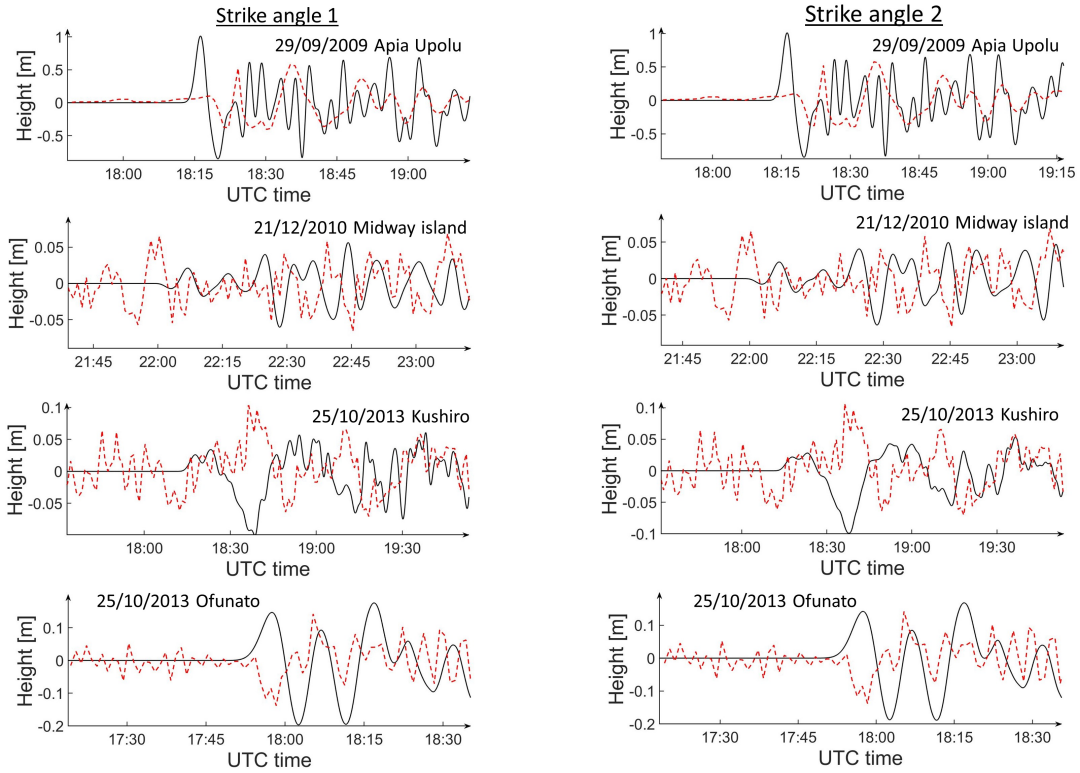

Figure 4.13: Non-linear simulations using 'strike angle 1' and 'strike angle 2' for the 25-10-2013 tsunami recorded in Kushiro, 25-10-2013 tsunami in Ofunato, 21-12-2010 tsunami in Midway island and 29-09-2009 in Apia Upolu. COMCOT results are presented in black lines and tide gauge recordings in dashed red lines.

The simulated maximum wave amplitudes in shallow water locations by the non-linear version of the SWE are observed to lie in the same order of magnitude as the recorded waves at the tide gauge locations, see Fig. 4.13.

4.5 Discussion

For the studied tectonic events, the calculated slender fault effective geometries and dynamics were used as input in a tsunami propagation numerical model (COMCOT). Due to the symmetry in the slender fault solution, two potential orientations were retrieved for each case scenario. Then, both orientations have been considered along with the

rest of the effective characteristics to perform the tsunami simulations in COMCOT. The results show that for each case, the propagation using one of the fault orientations outperforms the other, yielding a better match to the DART data, which highlights the importance of the strike angle as a descriptor of the tsunami source.

In some of the simulated cases, it can be observed that the inverse problem model overestimated the size of the fault, resulting in larger tsunamis. To achieve a more accurate description of the tsunami sources by the analysis of the sound that they produce, it is first necessary to achieve a more realistic description of the process in the direct approach. To this end, the inclusion of elasticity in the seabed [3], gravity effects and the possibility of linear superposition of multiple faults might become necessary in the study of AGWs propagation [132].

Despite the discrepancies in the maximum wave amplitudes, they are observed to lie in the same order of magnitude, which indicates that there is a partial agreement between the calculated effective tectonic event properties and the actual dynamics. Additionally, there is an observed mismatch of the simulated arrival times for the tsunami waves by the numerical model and the actual tsunami arrivals at the DART stations up to 300 seconds (< 5 minutes) in the worst-case scenario. We speculate that the reason behind some of these timing disagreements might be neglecting the elasticity of the ground, which has been reported in the literature to affect the AGWs phase speed [3]. Moreover, the inclusion of the effects of gravity can play a significant role in tsunami propagation [3].

Chapter 5

Conclusions

5.1 Summary

This thesis studies the relation between acoustic radiation induced by an underwater earthquake with the main characteristics of the fault. The real-time calculation is rather challenging due to the highly non-linear nature of the analytical solution for the bottom pressure that acoustic waves generated by an uplifting slender fault leave in the far-field [82]. To overcome this difficulty, in Chapter 2, a set of semi-analytical techniques are presented to retrieve the effective properties of tectonic sources from the long-range induced pressure-time series. The proposed methodology application is extended to real hydroacoustic signals and tested on a real hydrophone recording, under the presence of noise. The further evaluation of data with noise by the inverse problem model may require implementing various techniques (such as cross-correlation between several hydrophones, if applicable) to identify the relevant signal more accurately, considering that oceans are noisy environments. Thus, we anticipate an improvement in the performance of the inverse problem model as more advanced filtering techniques are employed. The inverse problem model has proved to be a computationally efficient tool that can be run on a standard desktop machine. We posit that the availability of more advanced computational resources to the application of the model in real-time to hydrophone recordings could result in almost instantaneous estimations of tectonic source effective

characteristics. Furthermore, it could be coupled with signal detection algorithms and implemented in current hydroacoustic monitoring networks.

The developed inverse problem model for acoustic waves requires external input parameters such as the type of slip, the moment magnitude of the event, and the distance between the hydrophone and the epicentre of the earthquake. In Chapter 3, this problem is approached, where various digital signal processing techniques are tested to extract information from the acoustic waves incoming from tectonic events. Then, artificial intelligence models (regression and classification) are trained with the extracted characteristics from a dataset of 201 earthquake acoustic signals and validated with robust algorithms, which provided reliable estimates on the identification of the characteristics that have a stronger impact when relating acoustic signatures with their originating source. We show that artificial intelligence algorithms can be a powerful tool to infer characteristics of underwater tectonic events from their sound. Furthermore, we provide sensitivity analysis and guidelines about diverse types of acoustic signal vectorization techniques that can be applied to classify earthquake sound signals and how to do it.

The size of the dataset is observed to be a limitation to achieving higher accuracy. Only earthquakes that meet specific conditions were considered in our analysis such as (1) shallow earthquakes, to maximise the signal-to-noise ratio; (2) earthquakes that have a direct path of water to the hydrophone and (3) are located in deep waters, due to the inverse problem model assumption of constant water depth. Thus, this study is presented as a proof of concept to show the applicability of a combination between ML algorithms and semi-analytical solutions to infer the properties of submarine tectonic events from acoustic radiation. The expansion of the dataset by including more signals induced by 'normal' earthquakes or the use of penalised ML algorithms could yield higher accuracy results in the classification stages. Furthermore, the application of feature selection algorithms instead of using feature sets based on previous studies has the potential of further improving the ML algorithms accuracy.

The validity of the inverse problem model solutions is tested and discussed in Chapter 4, where water surface height time series recorded from deep ocean DART buoys and

shallow water tide gauges are compared against numerical linear and non-linear tsunami wave propagation simulations, using the inverse problem model solutions as the initial condition for the generation of the tsunami waves. It is suggested in the literature that the efficiency of the traditional method of specification of the initial condition for tsunami propagation numerical models [95, 15], in which the initial elevation is assumed to be equal to the vertical seabed deformation, can be improved by techniques such as the Laplace smoothing method, leading to more accurate results. This chapter presents a methodology capable of coupling the inverse problem model for acoustic waves with machine learning algorithms and a tsunami propagation model to predict potential tsunami wave heights induced by underwater earthquakes at distant locations by the analysis of the associated recorded acoustic radiation.

Finally, the inverse problem model was applied to 112 vertical motion earthquakes, which results are reported in Appendix A. These solutions are conceived to form a reference dataset for future works that aim to perform tsunami simulations or make comparisons against models that require effective rupture characteristics.

5.2 Future outlook

Acoustic-gravity waves (AGWs) is a topic that has experienced exponentially growing attention from the scientific community, especially in recent history due to their high potential for high impact applications such as early precursors of tsunami waves. In this thesis, we neglected the effects of gravity when analysing acoustic signals. Even though the addition of the gravity effects may only slightly modify the dispersion properties of the leading (low frequency) acoustic modes [132], once gravity is included, retrieving the fault properties will directly result in the solution for the tsunami, hence minimising tsunami model calculation time.

The seabed was assumed to be flat. However, a variable sea depth would allow a more realistic description of the bathymetry of the ocean and shed light on the effect of some processes on AGWs propagation such as shoaling [61]. In addition, we neglected

the elasticity of the sea-floor, which is a potentially relevant factor [3], in particular close to critical depths, where leading acoustic modes couple with the elastic floor boosting its propagation speed to near the speed of sound in the elastic layer, typically over 3000 m/s [94, 30]. This becomes more significant when analysing transects with large variations in water depths, or during shoaling and upon interaction with shelf-breaks.

Although the uplift speed of the fault has been accounted as constant, a variable vertical uplift speed of the effective fault could lead to pressure envelope differences, hence improving the accuracy. Furthermore, the speed of the sound and water density have been considered constant in all the domain, nevertheless, the inclusion of a varying vertical profile of the speed of sound has been reported in the literature to affect the pressure signature in the far-field [83], thus, its might have impact on the direct approach and the inverse problem model solution accuracy.

Finally, the possibility of simulating complex earthquake geometries is a future step that should be considered, which can be achieved by multi-fault scenarios since the proposed model is linear, consisting of several single faults that can be triggered independently [46, 132], and in turn, each fault can be divided into segments following the rupture propagation.

Machine learning algorithms were trained and validated with a dataset composed of 201 samples that are reported to be limited by their size. Increasing the number of training samples is a step that has the potential to lead to higher accuracy and better generalization results in the classification and regression techniques, and allow the application of deep learning models, which are proven to work well on big datasets. However, it is not trivial to collect high-quality data, thus, exploration of data augmentation techniques, which consist of the generation of additional training samples from the already existing samples might be beneficial for the presented methodology. In addition, the deployment of more hydrophone stations in a wider set of locations in the oceans could help to mitigate the problem of data scarcity. The addition of acoustic samples associated with earthquakes with a wider range of moment magnitude and the further balance amongst the moment magnitude frequencies present in the dataset is estimated to potentially increase the

accuracy of the machine learning results.

On the other hand, the depth dependence of the classification accuracy has not been analysed in this thesis (only shallow earthquakes have been selected to reduce uncertainties), which is intended to be done in the future to expand the studied dataset and potentially improve the accuracy of the model. Another matter that we would like to study in the future is the relation between the magnitude prediction results and the type of studied earthquakes, where possible patterns might shed light on the process understanding.

Finally, we speculate that the validation of inverse problem model solutions for a wider range of moment magnitude and geographical location scenarios could reveal overlooked uncertainty sources. Furthermore, coupling the inverse problem model with a high-efficiency tsunami mathematical model capable to produce propagation results in the far-field in almost real-time could offer an advance in current tsunami early warning systems [132].

5.3 Accomplishments of this thesis

The aim of this thesis is the development of a set of tools capable to analyse underwater acoustic waves incoming from tectonic events and provide an estimation of the source characteristics in almost real-time. Below, the achieved targets of this thesis are listed:

- An inverse problem model for acoustic waves was developed and applied on synthetic and real signals [43], based on the model developed by [63] and the analytical solution provided by [82].

- Sensitivity analysis for the inverse problem model was performed [43].

- Digital signal processing techniques were applied and tested to extract features that characterize acoustic signals [42].

- Machine Learning algorithms capable to retrieve the magnitude and type of slip of underwater tectonic events were trained and validated with features extracted from a dataset composed of 201 recorded tectonic events [42].

- Past tsunami scenarios were replicated with a tsunami generation and propagation numerical model (COMCOT) to validate the source solutions provided by the analysis of acoustic waves.

- A dataset with effective source properties for 112 earthquakes with reported vertical motion components has been generated.

Appendix A

Supplementary materials

A.1 Machine learning sensitivity analysis

To understand the behaviour of the introduced classification algorithms under different input and setup conditions, we carried a sensitivity analysis.

K-fold and grid search

10-fold technique was applied for validation purposes on the classification and regression algorithms, see Fig. A.1.

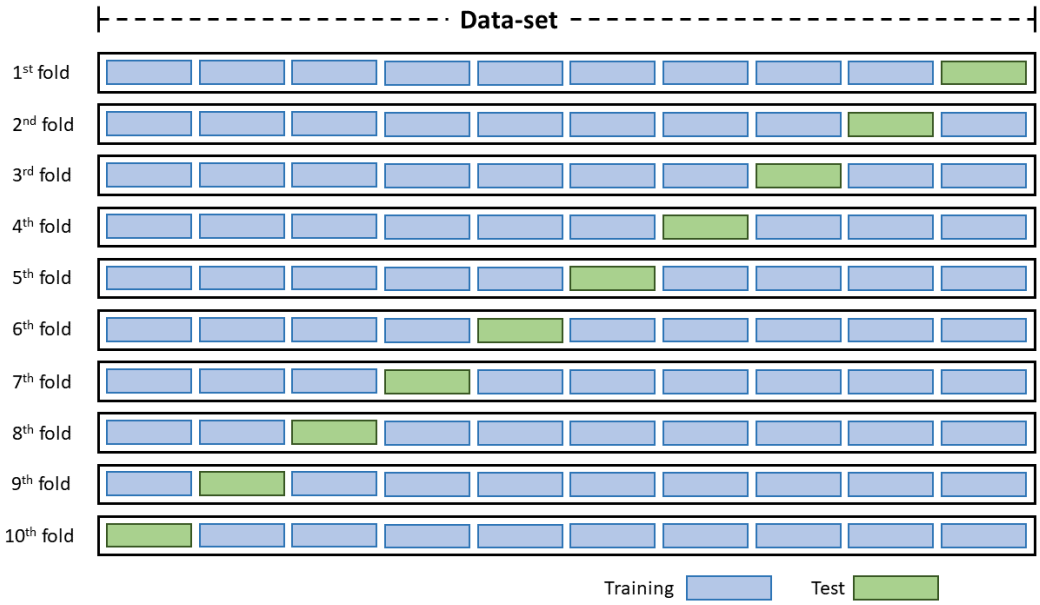

Figure A.1: 10-fold scheme used for validation of the ML algorithms.

Additionally, for each validation fold, we applied 5-fold grid search on the training sets, see Fig. A.2. This technique was used to identify the optimal hyper-parameter setup for the ML algorithms.

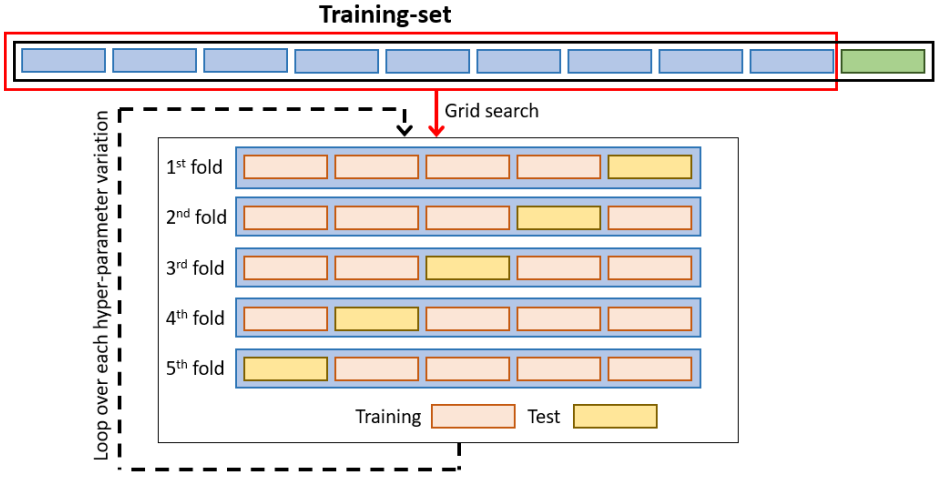

Figure A.2: 5-fold grid search scheme.

For SVM and SVR, the iterated hyper-parameters in the 5-fold grid search are:

- **Kernel**: Radial basis function.

- **C (regularization parameter)**: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 10, 20.

- **Gamma**: 0.01, 0.015, 0.02, 0.025, 0.03, 0.04, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9.

The regularization parameter represents the level of importance that is given to misclassification in the algorithm training stage. The chosen kernel is 'RBF', which is the most generalized form of kernel due to its similarity to the Gaussian distribution. The gamma parameter defines the reach of the influence of a single training example.

List of iterated hyper-parameters in grid search for RFC and RFR:

- **Number of estimators**: 10, 25, 50, 75, 100, 250, 500, 750, 1000.

- **Maximum depth**: 3, 4, 5, 6, 7.

- **Criterion**: 'gini' and 'entropy'.

The number of estimators refers to the number of trees in the forest. The maximum depth represents each tree depth, which is related to the number of splits done for the classification. The function to measure the quality of a split is the criterion, the considered criteria are 'gini' (Gini impurity) and 'entropy' (information gain).

Spectrum bands division analysis

The applied frequency limits for each considered decomposition of the spectrum are reported in Table A.1.

Band	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Lower [Hz]	0.1	0.5	1	2	4	8	12	20	30	40	50	60	70	80	90	0.1
Upper [Hz]	0.5	1	2	4	8	12	20	30	40	50	60	70	80	90	100	100
Lower [Hz]	0.1	0.5	1	3	6	12	16	24	32	50	75	0.1				
Upper [Hz]	0.5	1	3	6	12	16	24	32	50	75	100	100				
Lower [Hz]	0.1	1	6	12	20	32	64	0.1								
Upper [Hz]	1	6	12	20	32	64	100	100								
Lower [Hz]	0.1	10	20	0.1												
Upper [Hz]	10	20	80	100												
Lower [Hz]	0.1															
Upper [Hz]	100															

Table A.1: Tested frequency spectrum divisions.

For the sensitivity analysis, all considered divisions of the frequency spectrum were run along with the potential extraction window sizes for the case of binary classification and the accuracy results analysed, see Figs. A.3, A.4, A.5, A.6, A.7, A.8, A.9, A.10, A.11, A.12, A.13 and A.14.

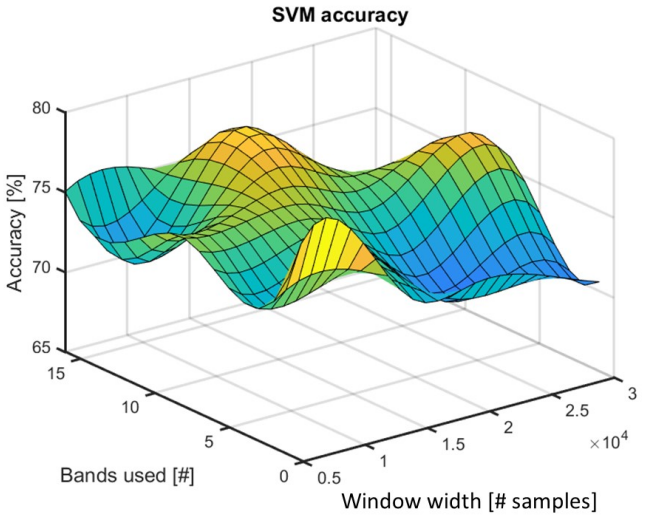

Figure A.3: SVM accuracy [%] results for the potential combinations between different extracted window sizes and considered sets of frequency bands.

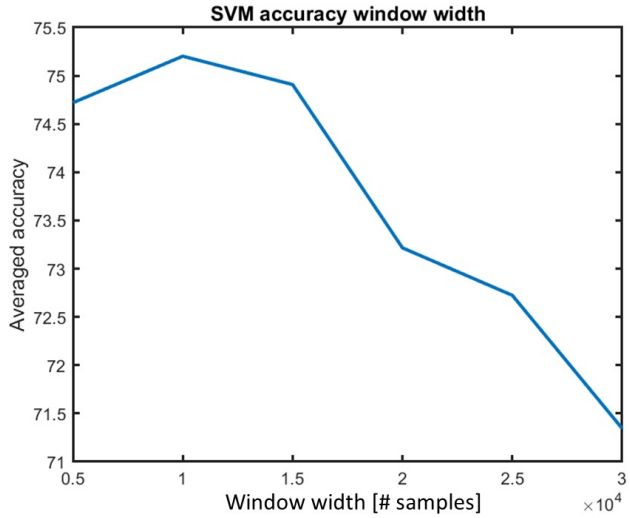

Figure A.4: SVM accuracy [%] results averaged along with the considered spectrum divisions.

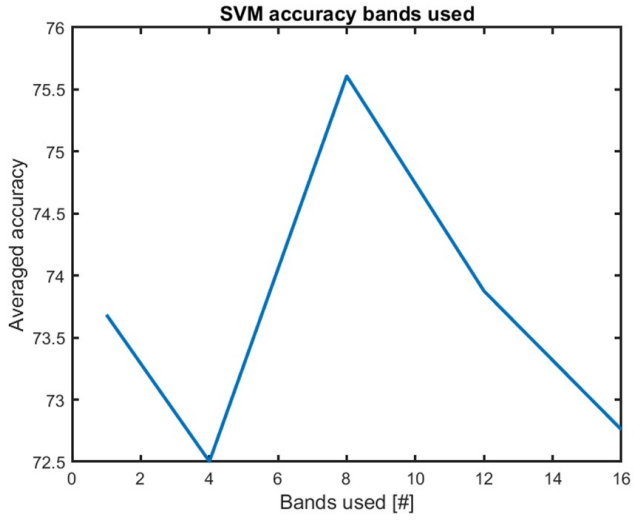

Figure A.5: SVM accuracy [%] averaged results along with the considered extraction window sizes.

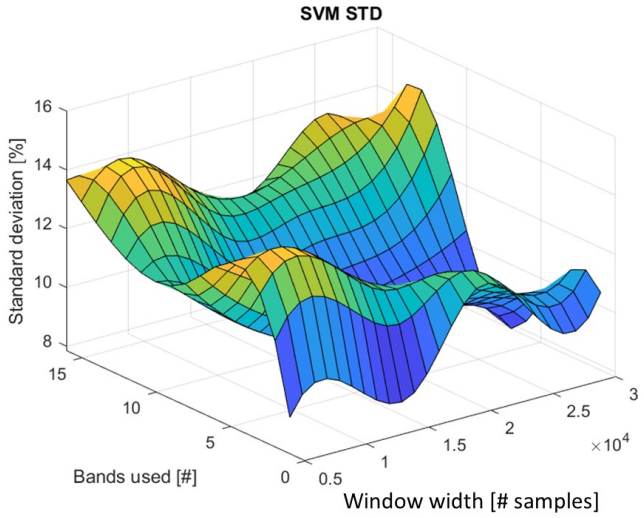

Figure A.6: SVM standard deviation [%] results for the potential combinations between different extracted window sizes and considered sets of frequency bands.

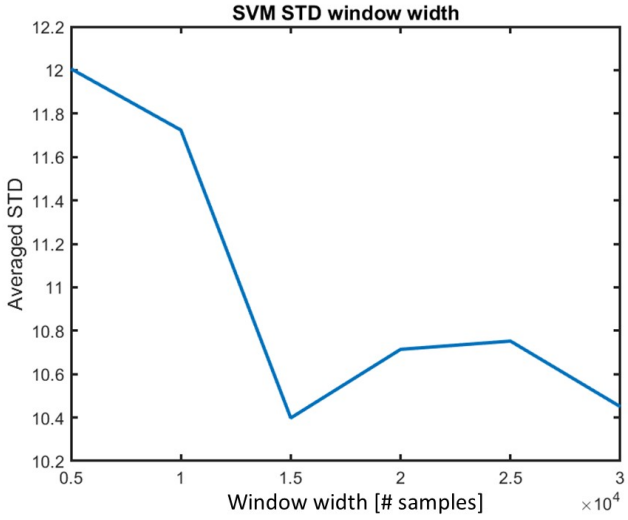

Figure A.7: SVM standard deviation [%] results averaged along with the considered spectrum divisions.

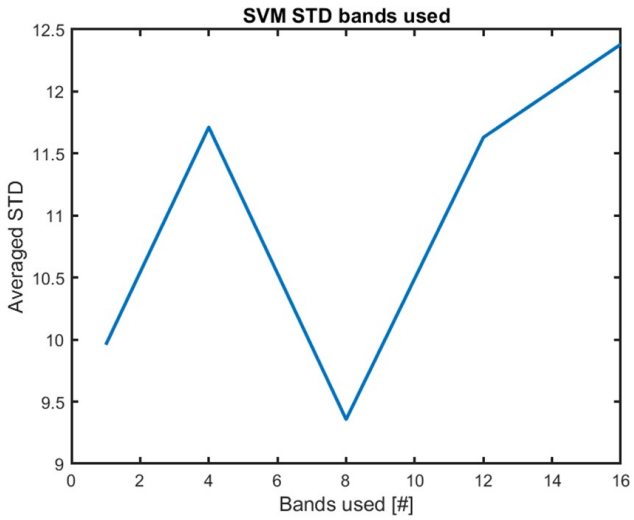

Figure A.8: SVM standard deviation [%] averaged results along with the considered extraction window sizes.

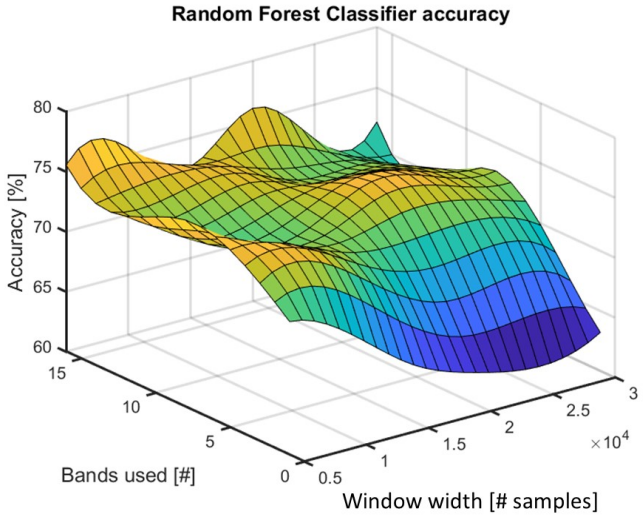

Figure A.9: RFC accuracy [%] results for the potential combinations between different extracted window sizes and considered sets of frequency bands.

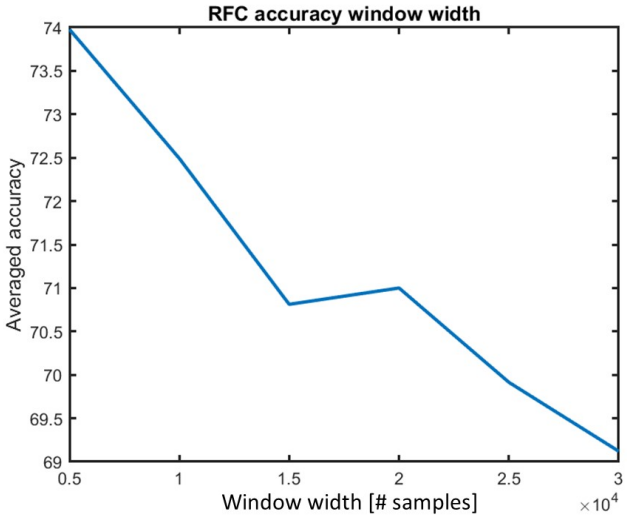

Figure A.10: RFC accuracy [%] results averaged along with the considered spectrum divisions.

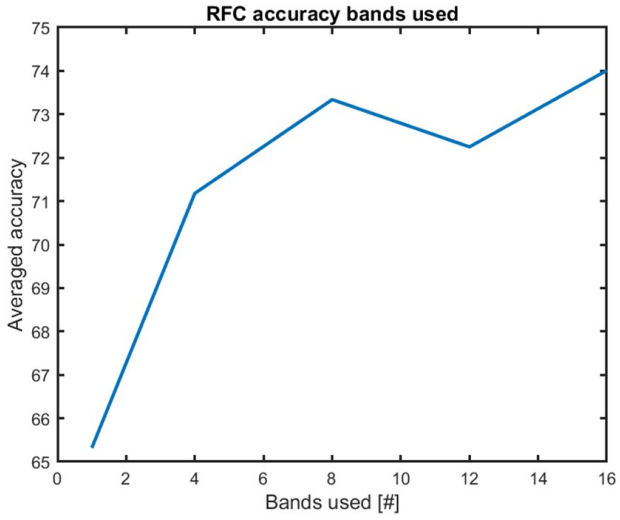

Figure A.11: RFC accuracy [%] averaged results along with the considered extraction window sizes.

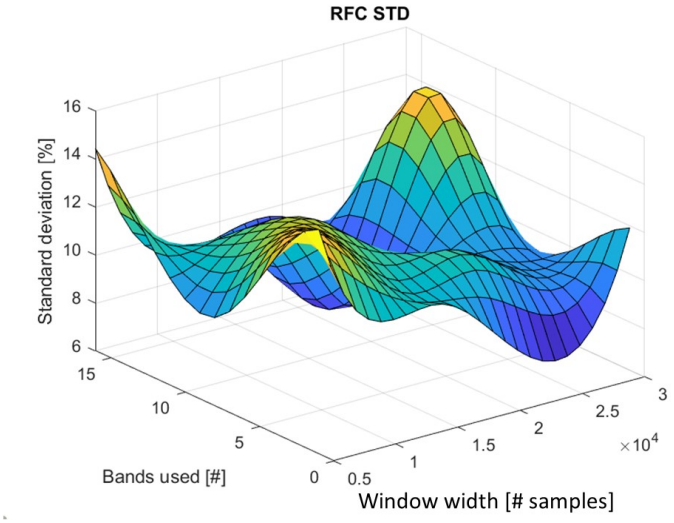

Figure A.12: RFC standard deviation [%] results for the potential combinations between different extracted window sizes and considered sets of frequency bands.

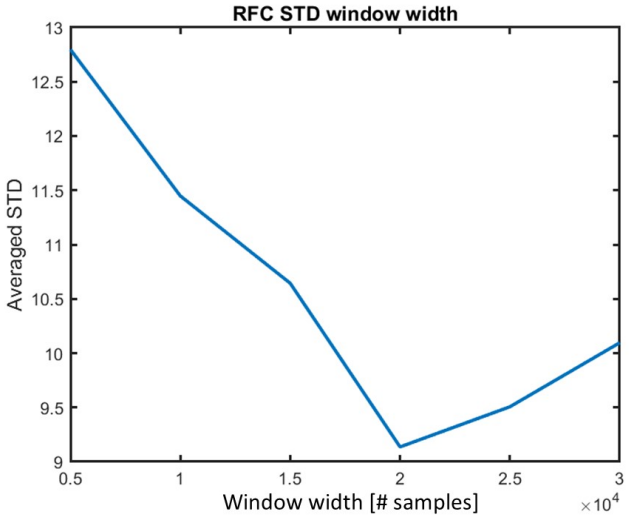

Figure A.13: RFC standard deviation [%] results averaged along with the considered spectrum divisions.

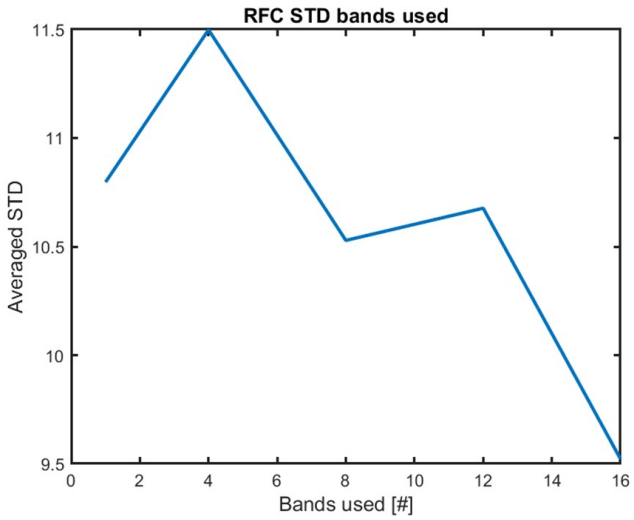

Figure A.14: RFC standard deviation [%] averaged results along with the considered extraction window sizes.

There are variations in the resulted classification accuracy amongst the different applied extraction spectrum divisions and window width, see Figs. A.3, A.4, A.5, A.6, A.7, A.8, A.9, A.10, A.11, A.12, A.13 and A.14.

Cepstral coefficients analysis

We extracted cepstral coefficients from three acoustic signals associated with tectonic events with different moment magnitude (selected from the studied dataset), see Fig. A.15. We consider that, for the studied range of moment magnitude, 12 coefficients can capture most of the information carried by the acoustic signals.

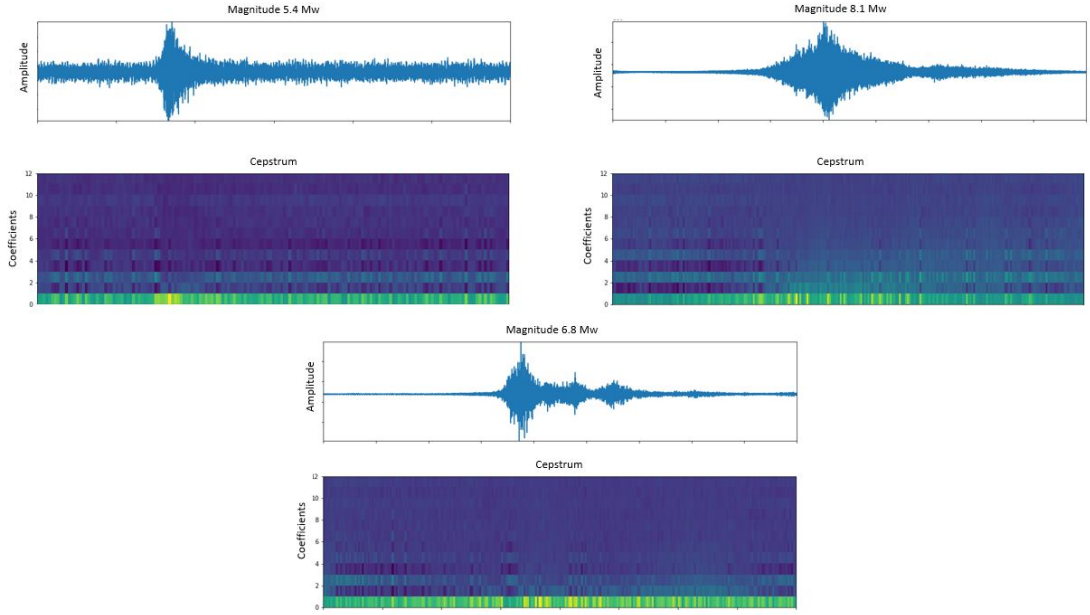

Figure A.15: Cepstrum related to 5.4 M_w earthquake (12/15/2017); 6.8 M_w earthquake (28/02/2013); and 8.1 M_w earthquake (23/12/2004).

For the three studied earthquakes in this subsection, see Fig. A.15, the highest coefficients lie on the first eight cepstral coefficient bands, carrying most of the information.

Wavelet transform parameters analysis

For binary classification and along with the introduced classification algorithms, we tested different types of discrete wavelets, wavelet orders and number of coefficient levels, see Tables A.2, A.3, A.4 and A.5.

Levels	Order 2	Order 5	Order 8
4	76.14 ± 7.58	72.64 ± 9.78	74.62 ± 8.80
6	79.12 ± 6.94	76.64 ± 8.86	78.14 ± 7.99
8	77.07 ± 6.87	75.64 ± 8.42	74.69 ± 10.10

Table A.2: SVM accuracy [%] / ± Standard deviation [%] for wavelet transform features with Symlet wavelet, different wavelet orders and levels.

Levels	Order 2	Order 5	Order 8
4	74.14 ± 11.32	71.67 ± 8.56	75.62 ± 9.78
6	73.64 ± 10.93	78.10 ± 8.74	78.64 ±9.11
8	72.19 ± 7.53	76.12 ± 9.94	74.62 ±11.29

Table A.3: RFC accuracy [%] / ± Standard deviation [%] for wavelet transform features with Symlet wavelet, different wavelet orders and levels.

Levels	Order 2	Order 5	Order 8
4	76.14 ± 7.58	72.64 ± 9.78	75.12 ± 9.49
6	79.12 ± 6.94	76.64 ± 8.58	75.67 ± 8.00
8	77.07 ± 6.87	74.62 ± 7.58	74.62 ± 7.91

Table A.4: SVM accuracy [%] / ± Standard deviation [%] for wavelet transform features with Daubechies wavelet, different wavelet orders and levels.

Levels	Order 2	Order 5	Order 8
4	74.14 ± 11.32	70.14 ±8.95	76.12 ± 9.94
6	73.64 ± 10.93	76.12 ± 9.94	77.12 ± 8.10
8	72.19 ± 7.53	74.62 ± 9.87	76.62 ± 10.25

Table A.5: RFC accuracy [%] / ± Standard deviation [%] for wavelet transform features with Daubechies wavelet, different wavelet orders and levels.

Regression accuracy distribution

The predicted values resulting from the application of the ML regression algorithms on the tested feature sets are plotted against the actual values to identify possible bias, see Figs. A.16 and A.17.

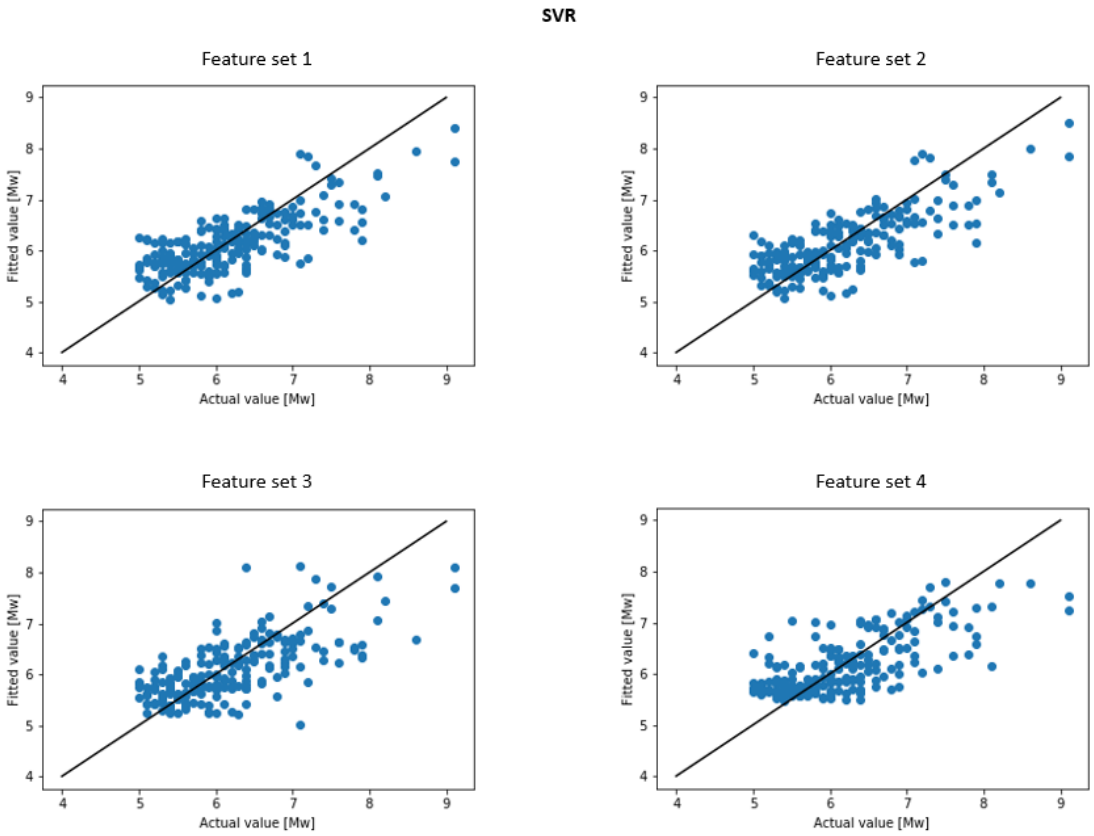

Figure A.16: M_w values comparison between the actual values of the test set and the estimated values by SVR for the different feature sets.

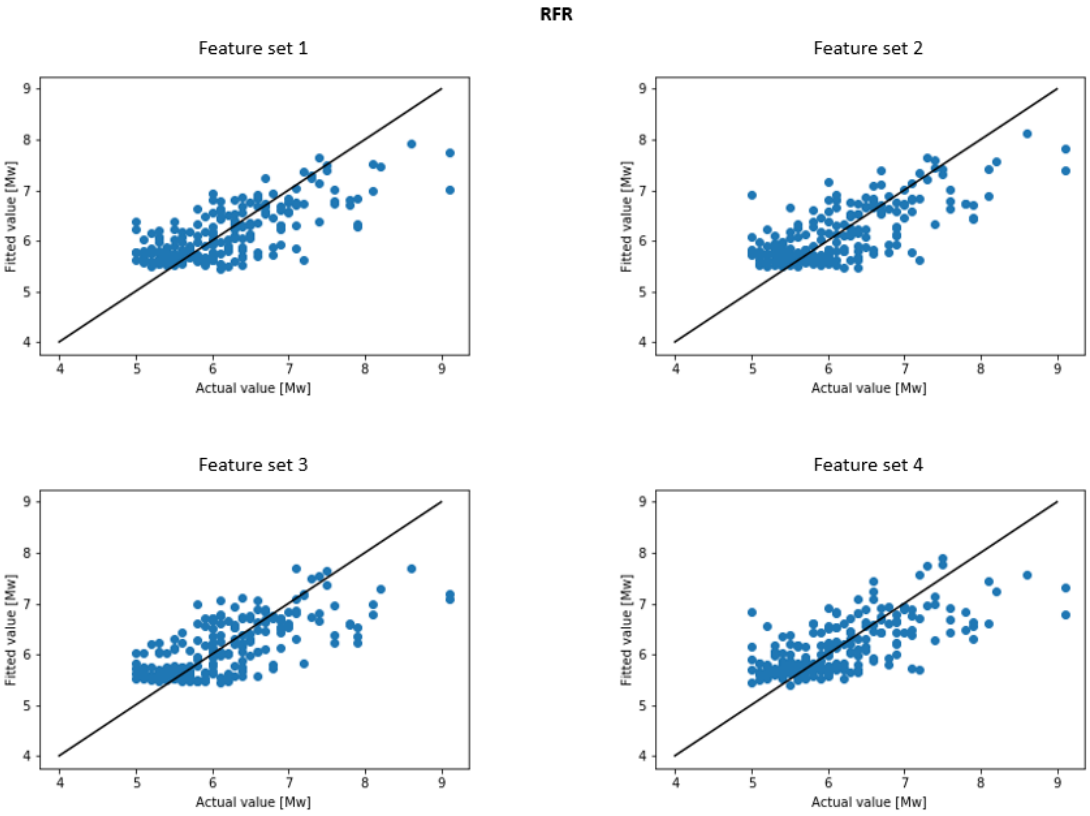

Figure A.17: M_w values comparison between the actual values of the test set and the estimated values by RFR for the different feature sets.

In addition, the mean squared error for 10-fold is plotted for every moment magnitude bin.

	Magnitude [M_w]				
	[5-6]	**[6-7]**	**[7-8]**	**[8-9]**	**[9-10]**
1st	0.474	0.484	0.777	0.831	1.692
2nd	0.495	0.503	0.793	0.833	1.525
3rd	0.523	0.463	0.747	1.317	1.703
4th	0.514	0.456	0.869	1.063	2.058

Table A.6: SVR mean squared errors for each considered feature set.

	Magnitude [M_w]				
	[5-6]	**[6-7]**	**[7-8]**	**[8-9]**	**[9-10]**
1st	0.487	0.434	0.831	0.78	1.073
2nd	0.492	0.433	0.88	0.858	0.48
3rd	0.501	0.434	0.793	1.12	1.852
4th	0.505	0.463	0.873	1.123	2.075

Table A.7: RFR mean squared errors for each considered feature set.

Machine learning application on synthetic signals analysis

Synthetic signals were introduced and randomly shuffled into the dataset. Then, ML classification algorithms were applied with 10-fold validation technique and 5-fold grid search on the first three feature sets, see Figs. A.18 and A.19. Note that the synthetic signals added to the dataset are considered incoming from vertical motion slender faults.

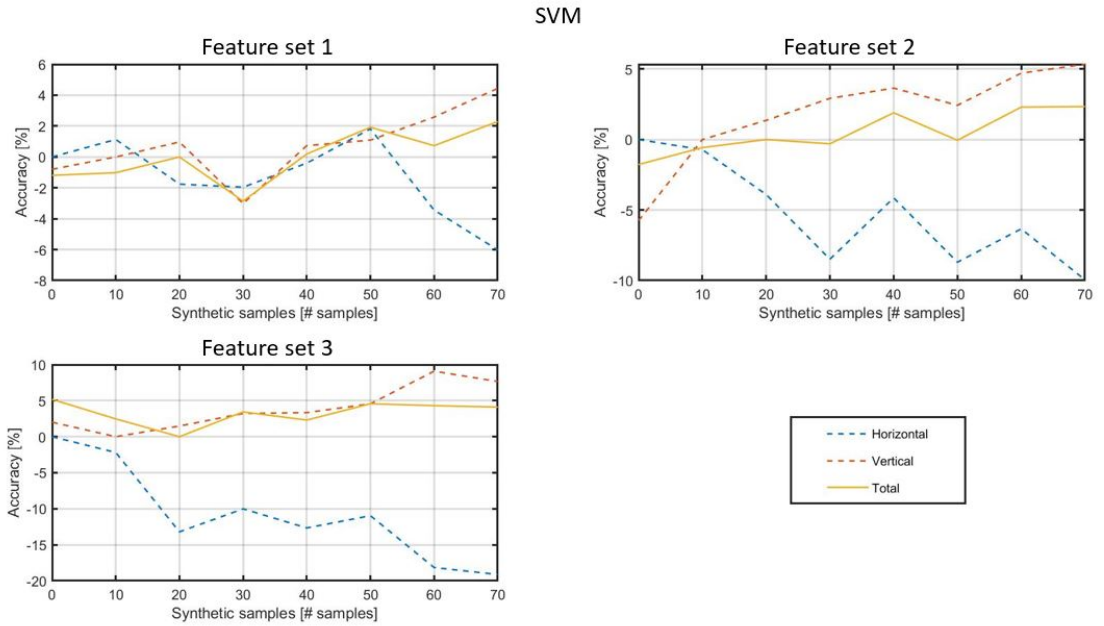

Figure A.18: SVM accuracy [%], for three different feature sets and different amounts of added synthetic signals.

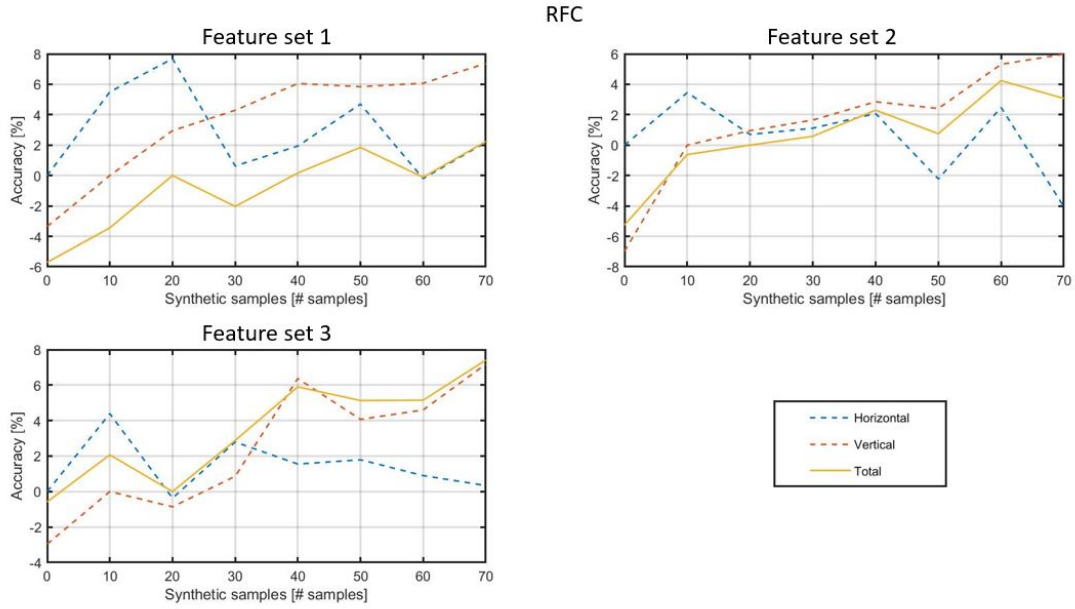

Figure A.19: RFC accuracy [%], for three different feature approaches and different amounts of added synthetic signals.

A.2 Dataset list of earthquakes

List of earthquakes associated with the acoustic signals used in this project. The times are expressed as UTC (Coordinated Universal Time).

Table A.8: In the column 'Type', '0' indicates strike-slip, '1' thrust and '2' normal event. In the column 'Tsunami', '0' indicates that no tsunami was recorded associated with the tectonic event and '1' indicates that there was a reported tsunami.

M_w	**Date**	**Time**	**Location**	**Type**	**Tsunami**
5	20/10/2008	15:51	Southeast of Honshu	0	0
5	02/01/2010	00:21	Mid-Indian ridge	0	0
5	25/01/2011	07:46	Andreanof islands, Aleutian islands	1	0
5	09/01/2013	05:27	Carlsberg ridge	2	0
5	08/04/2017	01:19	Macquarie island region	0	0

5.1	24/08/2008	01:00	Fiji islands region	0	0
5.1	22/02/2009	10:33	East of Kuril islands	2	0
5.1	16/01/2013	18:49	Southeast Indian ridge	0	0
5.1	07/02/2013	02:46	Sta. Cruz islands	2	0
5.1	09/01/2016	20:11	Western Indian-Antarctic ridge	0	0
5.1	17/05/2017	17:34	Macquarie island region	0	0
5.1	25/07/2017	21:12	West of Macquarie island	0	0
5.2	11/11/2008	12:18	Carlsberg ridge	2	0
5.2	16/12/2008	10:19	Mid-Indian ridge	0	0
5.2	17/12/2008	16:07	Southwest of Sumatra	1	0
5.2	01/01/2010	14:31	Mid-Indian ridge	0	0
5.2	31/08/2011	14:26	Mid-Indian ridge	0	0
5.2	15/01/2014	13:02	Mid-Indian ridge	0	0
5.2	26/09/2017	21:16	Southeast Indian ridge	0	0
5.3	05/01/2008	20:01	Near west coast of Sumatra	1	0
5.3	27/11/2008	09:17	Southeast Indian ridge	0	0
5.3	28/05/2010	18:32	Southern Sumatra	1	0
5.3	02/03/2012	10:11	Off coast of Honshu	1	0
5.3	29/07/2015	16:05	Fox Islands	1	0
5.3	04/08/2015	23:21	Kermadec islands region	1	0
5.3	03/09/2015	16:51	Off east coast of Honshu	2	0
5.3	17/01/2016	19:25	Mid-Indian ridge	0	0
5.3	08/10/2016	22:25	Tonga islands	0	0
5.3	03/04/2017	07:19	Mid-Indian ridge	0	0
5.4	14/01/2008	01:20	Indian ocean	2	0
5.4	01/02/2008	10:26	Tonga islands region	0	0
5.4	24/05/2009	07:03	Indian ocean	0	0
5.4	02/12/2010	23:39	Western Indian-Antarctic ridge	2	0
5.4	12/01/2011	13:05	West of Macquarie island	0	0

5.4	14/07/2012	14:22	Western Indian-Antarctic ridge	2	0
5.4	20/07/2015	05:28	Southwest Indian ridge	2	0
5.4	28/07/2015	21:21	Fox islands, Aleutian islands	1	0
5.4	27/10/2015	12:15	Mid-Indian ridge	0	0
5.4	29/01/2017	16:42	Southeast Indian ridge	2	0
5.4	09/05/2017	09:02	Andreanof islands, Aleutian islands	1	0
5.4	15/11/2017	00:23	Indian ocean triple junction	2	0
5.4	25/12/2017	12:02	Mid-Indian ridge	2	0
5.5	26/03/2008	18:33	Kuril islands	1	0
5.5	08/04/2008	02:09	South of Mariana islands	0	0
5.5	12/04/2008	08:50	Mid-Indian ridge	2	0
5.5	19/09/2008	22:49	Sta. Cruz islands region	2	0
5.5	25/02/2009	16:08	Southeast Indian ridge	0	0
5.5	13/01/2010	16:21	Tonga islands	0	0
5.5	07/01/2011	03:09	Off west coast of Sumatra	2	0
5.5	20/07/2012	03:40	Kuril islands	1	0
5.5	12/10/2013	20:01	Mid-Indian ridge	0	0
5.5	05/06/2015	14:54	Mid-Indian ridge	0	0
5.5	24/12/2017	17:33	Samoa islands region	2	0
5.6	07/01/2008	08:14	Aleutian islands region	1	0
5.6	11/03/2008	14:37	Fox islands, Aleutian islands	1	0
5.6	24/05/2009	06:49	Indian ocean	0	0
5.6	07/01/2010	08:29	Macquarie islands region	0	0
5.6	26/03/2012	16:58	Southwest Indian ridge	0	0
5.6	24/04/2012	09:50	Carlsberg ridge	2	0
5.6	06/11/2012	06:17	Carlsberg ridge	2	0
5.6	15/06/2014	18:19	Near east coast of Honshu	2	0
5.6	01/07/2015	14:30	Macquarie islands region	0	0
5.6	25/06/2017	03:01	Mid-Indian ridge	0	0

5.7	09/01/2009	03:44	Carlsberg ridge	0	0
5.7	30/01/2009	03:47	Tonga islands	0	0
5.7	26/03/2009	06:14	Mid-Indian ridge	2	0
5.7	15/04/2009	10:20	Owen fracture	0	0
5.7	05/07/2011	19:02	Fiji islands region	0	0
5.7	27/01/2013	09:59	Tonga islands region	2	0
5.7	04/04/2017	22:08	Andreanof islands, Aleutian islands	1	0
5.7	31/10/2017	04:58	Tonga islands region	2	0
5.7	22/04/2018	16:59	Tonga islands region	0	0
5.8	18/09/2008	11:58	near east coast of kamchatka	1	0
5.8	01/10/2009	06:13	Tonga islands region	1	0
5.8	29/04/2010	23:48	Western Indian-Antarctic ridge	0	0
5.8	14/12/2011	00:48	Tonga islands region	0	0
5.8	02/05/2012	12:18	West of Macquarie island	0	0
5.8	05/02/2014	20:52	Balleny islands region	0	0
5.8	01/06/2014	10:07	North Indian ocean	0	0
5.8	24/05/2017	16:36	Fox islands	2	0
5.8	21/04/2018	19:44	Western Indian-Antarctic ridge	2	0
5.9	27/06/2008	13:07	Andaman	2	0
5.9	21/12/2008	09:16	East coast of Honshu	2	0
5.9	19/03/2011	01:22	Near east coast of Honshu	1	0
5.9	11/05/2014	12:35	Southeast Indian ridge	0	0
5.9	29/06/2014	18:24	Tonga islands region	0	0
5.9	12/01/2016	09:45	Southeast Indian ridge	2	0
5.9	31/05/2016	10:04	Kuril islands region	1	0
5.9	14/10/2018	12:41	Southeast Indian ridge	0	0
6	22/01/2008	07:55	Tonga islands	0	0
6	13/01/2009	01:04	Mid-Indian ridge	0	0
6	23/09/2009	02:59	Macquarie island region	0	0

6	10/11/2009	02:48	Off southeast coast of India	0	0
6	09/06/2010	01:03	Tonga islands region	0	0
6	04/08/2010	23:48	East of Kuril islands	2	0
6	23/04/2011	10:12	Near east coast of Sonshu	1	0
6	23/05/2012	22:59	Western Indian-Antarctic ridge	0	0
6	25/11/2013	05:56	Kuril islands	1	0
6	28/10/2014	03:15	Tonga islands region	0	0
6	13/08/2015	10:39	Mid-Indian ridge	0	0
6	14/10/2015	05:43	East of Kuril islands	1	0
6	19/12/2015	02:10	Vanuatu islands	0	0
6	19/03/2017	15:43	Solomon islands	1	0
6.1	22/01/2008	10:49	Tonga islands region	0	0
6.1	02/11/2008	13:48	Aleutian islands region	1	0
6.1	21/04/2010	17:20	Tonga islands region	2	0
6.1	04/08/2011	13:51	Kuril islands	1	0
6.1	02/08/2012	09:56	New ireland	1	0
6.1	31/01/2013	03:33	Sta. Cruz islands	2	0
6.1	20/04/2013	13:12	Kuril islands region	1	0
6.1	08/12/2013	17:24	Kuril islands region	1	0
6.1	18/02/2015	09:32	Sta. Cruz islands region	0	0
6.1	24/03/2018	19:58	South east Indian ridge	0	0
6.2	07/09/2009	16:12	Java	2	0
6.2	01/02/2010	22:28	Solomon islands	1	0
6.2	05/02/2010	06:59	South east Indian ocean	0	0
6.2	17/08/2011	11:44	Off east coast of Honshu	2	0
6.2	23/09/2016	00:14	Off east coast of Honshu	1	0
6.2	20/09/2017	16:37	Off east coast of Honshu	2	0
6.3	09/08/2008	16:36	West of Macquarie island	0	0
6.3	10/07/2010	11:43	Mariana islands	0	0

6.3	14/08/2010	23:01	South of Mariana islands	2	0
6.3	03/09/2012	18:23	South of Java, Indonesia	2	0
6.3	20/07/2014	18:32	Kuril islands	1	0
6.3	03/11/2014	08:48	Mid-Indian ridge	0	0
6.3	27/08/2017	04:17	Admirality islands region	0	0
6.3	18/04/2019	14:46	Western Indian-Antarctic ridge	0	0
6.4	09/06/2004	22:52	Western Indian-Antarctic ridge	0	0
6.4	01/04/2009	03:55	New Guinea	0	0
6.4	13/10/2009	05:37	Fox islands, Aleutian islands	1	0
6.4	10/11/2010	04:05	Indian Ocean	0	0
6.4	22/03/2011	07:18	Off east coast of Honshu	2	0
6.4	31/03/2011	00:12	Fiji islands region	0	0
6.4	22/09/2011	23:07	Tonga islands	0	0
6.4	04/10/2013	17:26	Mid-indian ridge	0	0
6.4	09/11/2015	16:03	South Alaska	1	0
6.4	08/02/2016	16:19	Solomon islands	1	0
6.4	13/08/2017	03:08	Indonesia	1	0
6.5	03/03/2008	09:31	Kuril islands region	1	0
6.5	03/09/2010	11:16	Andreanof islands, Aleutian islands	1	0
6.5	12/03/2011	01:47	Off east coast of Honshu	2	0
6.5	16/11/2012	18:12	Kuril islands region	1	0
6.5	04/09/2013	02:32	Andreanof islands, Aleutian islands	1	0
6.5	11/07/2014	19:22	Off east coast of Honshu	2	1
6.6	18/04/2009	19:18	Kuril islands region	1	0
6.6	12/08/2009	22:48	Near east coast of Honshu	1	0
6.6	31/07/2011	23:39	Near north coast of New Guinea	0	0
6.6	26/07/2012	05:33	Mauritius	0	0
6.6	09/10/2012	12:32	West of Macquarie island	0	0
6.6	29/06/2014	17:15	Samoa islands region	0	0

6.6	15/08/2018	21:56	Andreanof islands, Aleutian islands	1	0
6.7	01/01/2005	06:25	Off west coast of northern Sumatra	0	0
6.7	09/05/2008	21:51	South of Mariana islands	2	0
6.7	17/08/2009	00:05	Southwestern Ryukyu islands	0	0
6.7	22/06/2011	21:50	Off east coast of Honshu	1	0
6.7	16/09/2011	19:26	Near east coast of Honshu	1	0
6.7	28/04/2012	10:08	Tonga islands region	2	0
6.8	28/02/2013	14:05	Kuril islands	1	0
6.8	03/08/2014	00:22	Caroline islands	0	0
6.8	17/09/2014	06:14	Mariana islands	2	0
6.8	22/05/2015	23:59	Solomon islands	0	0
6.9	30/01/2007	04:54	West of Macquarie island	0	0
6.9	07/04/2009	04:23	Kuril islands region	1	0
6.9	13/08/2010	21:19	South Mariana islands	2	1
6.9	14/03/2012	09:08	Off east coast of Honshu	2	1
6.9	18/07/2015	02:27	Solomon islands	2	1
6.9	27/07/2015	04:49	Fox islands, Aleutian islands	1	1
6.9	03/01/2017	21:52	South Fiji islands	2	1
7	10/07/2011	00:57	Off east coast of Honshu	0	1
7	06/02/2013	01:54	Sta. Cruz islands	0	0
7	08/02/2013	15:26	Sta. Cruz islands	0	1
7	30/08/2013	16:25	Andreanof islands, Aleutian islands	1	0
7.1	22/11/2004	20:26	Off west coast of Stewart island	1	1
7.1	12/04/2008	00:30	Macquarie island region	1	0
7.1	06/02/2013	01:23	Sta. Cruz Islands	2	0
7.1	25/10/2013	17:10	Off east coast of Honshu	2	1
7.1	04/12/2015	22:25	Southeast Indian Ridge	2	0
7.2	26/12/2004	04:21	Nicobar islands, India region	1	0
7.2	10/01/2012	18:37	Off west coast of northern Sumatra	0	0

7.2	12/08/2016	01:26	Loyality islands	0	1
7.3	09/03/2011	02:45	Off east coast of Honshu	1	0
7.3	19/04/2013	03:06	Kuril islands region	2	1
7.4	15/01/2009	17:49	Kuril islands region	1	1
7.4	07/10/2009	23:14	Vanuatu islands	1	0
7.4	21/12/2010	17:19	Bonin islands, Japan region	2	1
7.5	10/08/2009	19:56	Andaman	2	1
7.5	12/06/2010	19:27	Indian Ocean	1	1
7.6	19/03/2009	18:17	Tonga islands region	1	1
7.6	07/10/2009	22:03	Vanuatu islands	1	1
7.6	31/08/2012	12:47	Philippine islands region	1	1
7.8	07/10/2009	22:19	Sta. Cruz islands	1	0
7.8	02/03/2016	12:49	Sumatra	0	0
7.9	06/02/2013	01:12	Sta. Cruz islands	1	1
7.9	30/05/2015	11:23	Bonin islands, Japan Region	2	0
7.9	17/12/2016	10:51	New Ireland	1	1
8.1	23/12/2004	14:59	North of Macquarie Island	0	1
8.1	29/09/2009	17:48	Sumatra	2	1
8.2	11/04/2012	10:43	Sumatra	0	1
8.6	11/04/2012	08:39	Sumatra	0	1
9.1	26/12/2004	01:01	Sumatra	1	1
9.1	11/03/2011	06:15	Tohoku	1	1

A.3 Supplementary materials: Numerical validation of slender fault solution

COMCOT equations

The linear SWE used by COMCOT are:

$$\frac{\partial \eta}{\partial t} + \frac{1}{R\cos\Phi}\left(\frac{\partial P}{\partial \psi} + \frac{\partial (Q\cos\Phi)}{\partial \Phi}\right) = -\frac{\partial h}{\partial t} \tag{A.1}$$

$$\frac{\partial P}{\partial t} + \frac{gh}{R\cos\phi}\frac{\partial \eta}{\partial \psi} - fQ = 0 \tag{A.2}$$

$$\frac{\partial Q}{\partial t} + \frac{gH}{R}\frac{\partial \eta}{\partial \psi} + fp = 0 \tag{A.3}$$

$$f = 2\Omega\sin\phi \tag{A.4}$$

where η is the surface water elevation, P and Q denote the volume fluxes in x (West-East) and y (South-North) directions respectively. ψ and ϕ are the longitude and latitude of the earth, R is the radius of the Earth, g is the gravitational acceleration and h is the water depth. $\dfrac{\partial h}{\partial t}$ refers to the effect of the transient seafloor motion, f represents the Coriolis force coefficient due to the rotation of the Earth and Ω is the rotation rate of the earth.

The non-linear version of the SWE are:

$$\frac{\partial P}{\partial t} + \frac{gh}{R\cos\phi}\frac{\partial}{\partial \psi}\left(\frac{P^2}{H}\right)\frac{1}{R}\frac{\partial}{\partial \Phi}\left(\frac{PQ}{H}\right) + \frac{gH}{R\cos\Phi}\frac{\partial \eta}{\partial \psi} - fQ + F_x = 0 \tag{A.5}$$

$$\frac{\partial Q}{\partial t} + \frac{1}{R\cos\phi}\frac{\partial}{\partial \psi}\left(\frac{Q^2}{H}\right)\frac{1}{R}\frac{\partial}{\partial \Phi}\left(\frac{PQ}{H}\right) + \frac{gH}{R}\frac{\partial \eta}{\partial \phi} - fP + F_y = 0 \tag{A.6}$$

$$F_x = \frac{gn^2}{H^{7/3}}P(P^2+Q^2)^{1/2} \tag{A.7}$$

$$F_y = \frac{gn^2}{H^{7/3}}Q(P^2+Q^2)^{1/2} \tag{A.8}$$

$$H = \eta + h \tag{A.9}$$

where H is the total water depth, F_x and F_y represent the bottom friction in the ψ and ϕ directions respectively, and n is the Manning's roughness coefficient.

Paths and transects

The paths between the earthquake epicentres and the corresponding recording hydrophones were obtained and averaged, see Figs. A.20, A.21, A.22 and A.23.

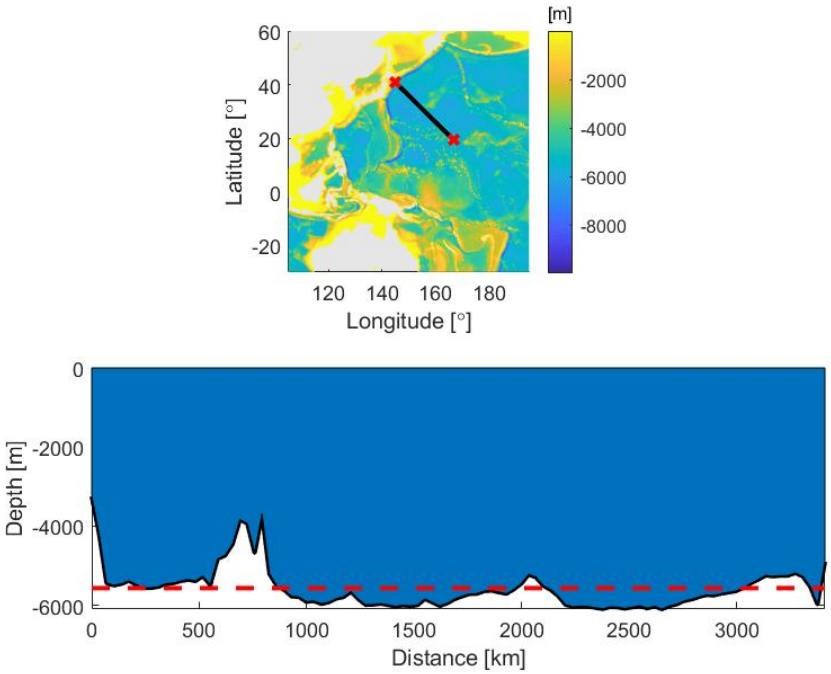

Figure A.20: Transect between hydrophone 'HA11' and 14-03-2012 studied earthquake epicentre.

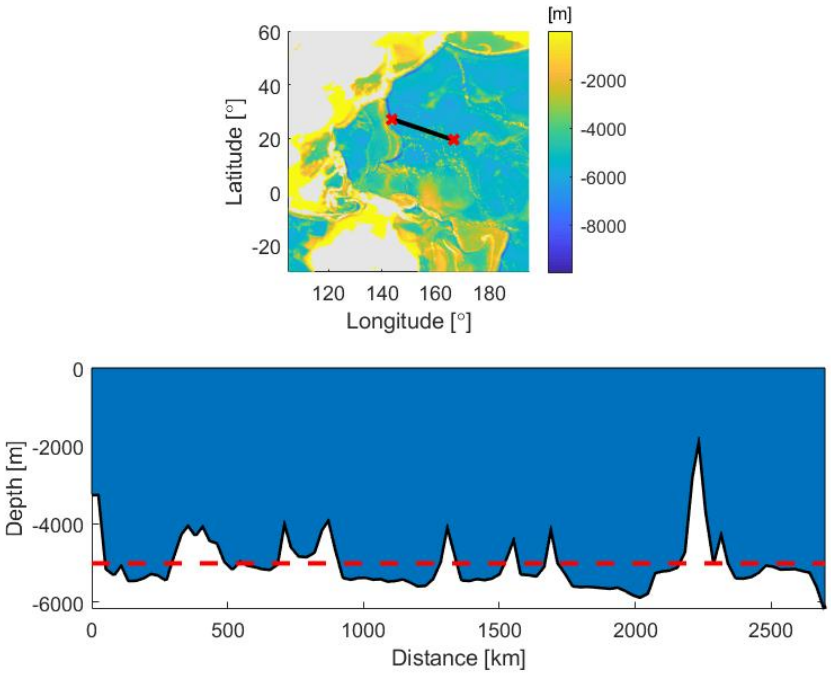

Figure A.21: Transect between hydrophone 'HA11' and 21-12-2010 studied earthquake epicentre.

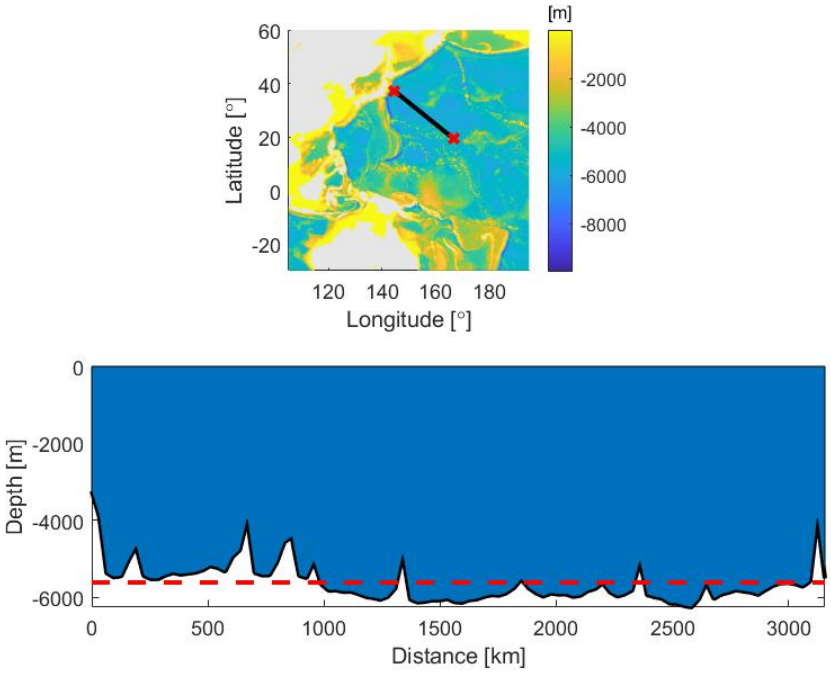

Figure A.22: Transect path between hydrophone and epicentre for 25-10-2013 studied earthquake epicentre.

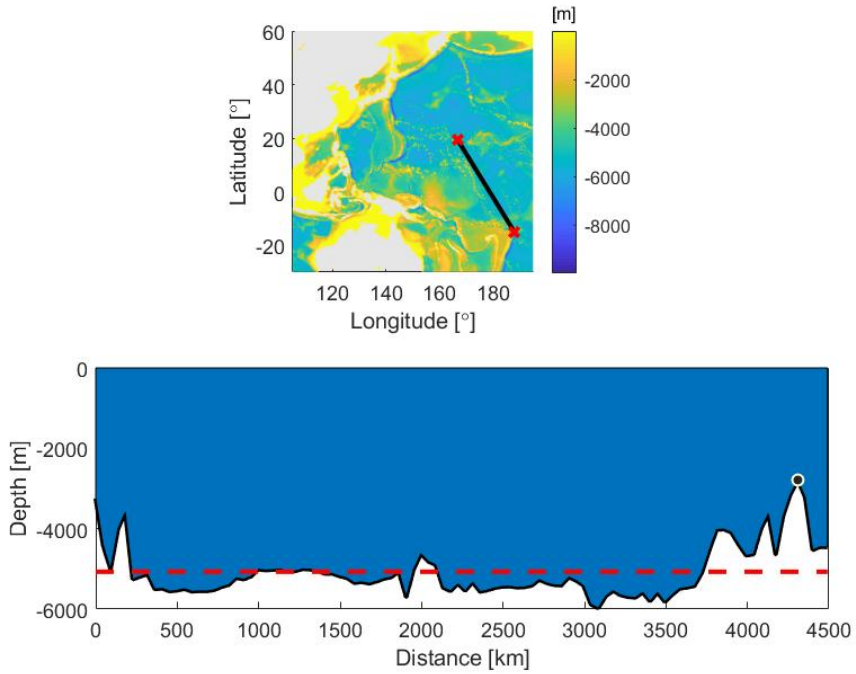

Figure A.23: Transect path between hydrophone and epicentre for 29-09-2009 studied earthquake epicentre.

Strike angles

The potential effective orientations of each studied earthquake were calculated from the relative coordinates of the hydrophone with respect to the fault centre as shown in Figs. A.24, A.25, A.26 and A.27.

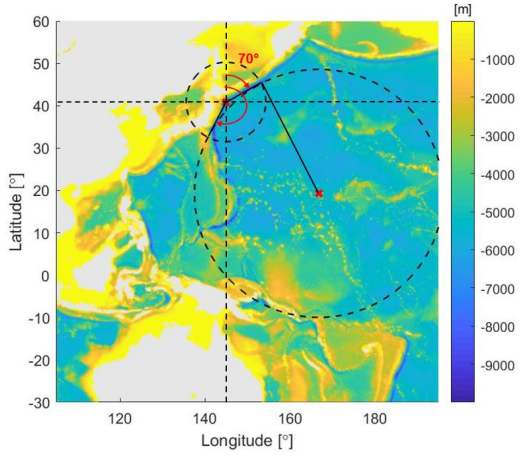

Figure A.24: Location, bathymetry and strike angle calculation.

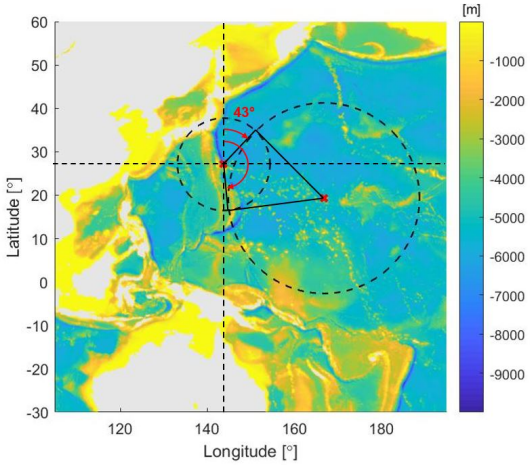

Figure A.25: Location, bathymetry and strike angle calculation.

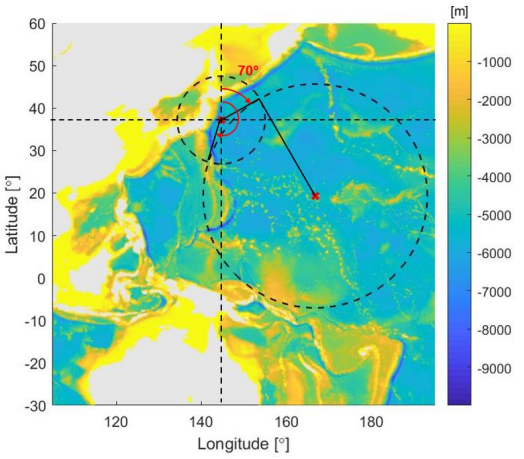

Figure A.26: Location, bathymetry and strike angle calculation.

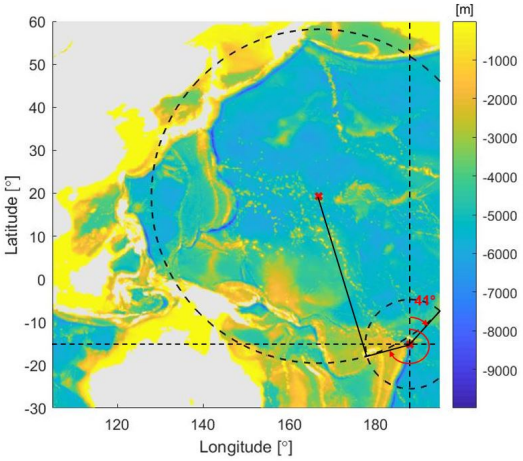

Figure A.27: Location, bathymetry and strike angle calculation.

A.4 Inverse problem model application to the dataset

A set of 112 underwater tectonic events with identified vertical motion components by gCMT catalog has been analysed and the inverse model for acoustic waves developed in Chapter 2 has been applied to it, with the intention of retrieving the effective characteristics associated with each of the earthquakes. This information can be useful to other experts in the field to have reference values of effective slender fault properties associated with a wide dataset of events (For example for tsunami mathematical simulations).

The approximated distance and average depth from the epicentre of every studied tectonic event to the corresponding recording hydrophones are reported, which were utilised in the application of the inverse problem model, see Table A.9. In addition, for the application of the inverse model, 10 sets of solutions were produced for each identified orientation of each slender fault and five pressure points were used in each iteration of the model. The reported solutions consist of the average calculated location (X and Y), dynamics (T and W_0) and geometry (b and L) of the slender faults associated with the considered tectonic events, see Table A.9.

To expand the application of the inverse problem model to earthquakes with associated magnitudes higher than $M_w = 8.1$, a new parameterization has been considered to produce the potential ranges of properties used as input for the model [116], see Chapter 2. Ref. [131] parameterizations are reported to be valid for the magnitude range $M_w = [4.8 - 8.1]$. However, the dataset studied in this thesis includes events with associated moment magnitude up to M_w=9. Thus, the parameterizations found in Ref. [75] have been included for the events with a magnitude higher than $M_w > 8.1$. The mentioned scaling relations are valid for the range $M_w = [5.3 - 9.5]$ [75].

$$\begin{aligned} \log SRL &= c_{L1} + c_{L2} M_w, \qquad c_{L1} = -2.07, \quad c_{L2} = 0.54 \\ \log RW &= c_{w1} + c_{w2} M_w, \qquad c_{w1} = -1.76, \quad c_{w2} = 0.44 \end{aligned} \tag{A.10}$$

where SRL stands for surface rupture length, RW for rupture width.

The upper limit for the effective duration is set to 95 seconds, which coincides with the

reported duration by gCMT of one of the largest recorded tectonic events in recent history, the Sumatra earthquake (24/12/2004). A seafloor vertical displacement of $> 10m$ was reported for the so-called Tohoku earthquake (11/03/2011) [36], which also is considered one of the largest tectonic events ever recorded. thus the potential vertical uplift for earthquakes with magnitudes $> 8.1 M_w$ is set in the range [0.5-15 m].

M_w	Date	Coordinates	Tp	Hd	D [km]	h [m]	h_{eq} [m]	X [km]	Y [km]	B [m]	L [km]	T [s]	W [m/s]
5.0	25/01/2011	51.45 ,-176.07	1	11	3714.65	5279.77	4334.07	3450.00	1377.00	1500.00	4080.00	9.82	0.0026
5.0	09/01/2013	-2.02 , 67.95	2	8	583.37	3375.19	2847.08	500.00	300.53	1500.00	3800.00	14.25	0.0011
5.1	22/02/2009	48.87 , 158.22	2	11	3417.03	5493.34	5905.15	3300.00	886.62	1500.00	3770.00	10.88	0.0023
5.1	07/02/2013	-10.73 , 165.13	2	11	3338.36	4163.47	3437.68	2750.00	1892.65	1550.00	3687.00	12.27	0.0016
5.2	11/11/2008	1.51 , 66.78	2	8	985.37	3345.16	2928.72	900.00	401.19	1750.00	4670.00	12.91	0.0017
5.2	17/12/2008	-6.49 , 103.23	1	8	3577.59	4612.59	1410.08	3450.00	946.92	2050.00	4465.00	11.05	0.0023
5.3	05/01/2008	5.21 , 94.62	1	8	2916.54	3890.12	2489.74	2700.00	1102.83	1991.66	6253.33	12.94	0.0030
5.3	28/05/2010	-4.24 , 100.53	1	8	3285.80	4569.95	3014.60	3150.00	934.86	2125.00	5730.00	12.42	0.0034
5.3	02/03/2012	35.20 , 141.34	1	11	3330.01	5544.35	2155.11	3200.00	921.38	2125.00	5380.00	15.49	0.0020
5.3	29/07/2015	52.00 ,-169.32	1	11	3640.63	5369.07	4703.08	3500.00	1002.10	2500.00	3650.00	13.36	0.0025
5.3	04/08/2015	-27.16 ,-175.99	1	11	5257.41	4056.58	4483.21	4550.00	2633.99	1864.58	3874.17	17.29	0.0018
5.3	03/09/2015	37.15 , 143.55	2	11	3251.86	5614.64	7306.59	3000.00	1254.82	1800.00	3386.66	15.46	0.0017
5.4	14/01/2008	-35.31 , 53.85	2	8	3740.72	3865.92	3474.20	3500.00	1320.22	2641.66	4050.00	17.01	0.0022
5.4	02/12/2010	-49.32 , 120.87	2	1	2085.41	4408.92	3672.56	1850.00	962.51	2625.00	5040.00	13.94	0.0025
5.4	14/07/2012	-49.08 , 124.88	2	1	2296.82	4310.55	4114.48	2200.00	659.83	3000.00	4920.00	16.43	0.0030
5.4	20/07/2015	-34.69 , 54.54	2	8	3642.63	3835.34	3541.92	3300.00	1542.33	2530.00	3806.00	16.69	0.0020
5.4	28/07/2015	52.03 ,-169.31	1	11	3643.87	5370.36	4742.71	3450.00	1172.72	2575.00	3765.00	13.57	0.0022
5.4	29/01/2017	-29.13 , 61.03	2	8	2765.24	3542.63	4338.25	1700.00	2180.95	2300.00	4192.63	16.59	0.0022
5.4	09/05/2017	51.00 ,-176.79	1	11	3689.96	5262.88	5336.41	3550.00	1006.64	2612.50	4390.00	16.26	0.0023
5.4	15/11/2017	-26.11 , 70.89	2	8	2198.94	3391.52	3110.14	1800.00	1263.07	2417.85	3644.28	16.93	0.0016
5.4	25/12/2017	-27.82 , 74.02	2	8	2412.13	3718.56	3299.79	2150.00	1093.55	2506.25	4485.00	15.83	0.0026
5.5	26/03/2008	-56.85 ,-141.75	1	11	8897.55	2285.62	3643.75	7650.00	4543.56	2370.50	6182.62	20.74	0.0022
5.5	12/04/2008	-15.42 , 67.19	2	8	1097.11	3260.28	3149.04	1000.00	451.27	2925.00	6790.00	21.43	0.0014
5.5	19/09/2008	-11.23 , 164.48	2	11	3398.43	4215.42	4067.44	3250.00	993.40	2912.50	3905.00	21.82	0.0015
5.5	07/01/2011	4.20 , 90.37	2	8	2445.65	4073.86	2809.27	2350.00	677.27	2612.50	7350.00	21.23	0.0014
5.5	20/07/2012	49.22 , 156.50	1	11	3511.68	5454.08	3144.68	3350.00	1053.28	2550.00	5375.00	29.20	0.0013
5.5	24/12/2017	-15.33 ,-172.60	2	11	3897.84	4181.56	3432.31	3650.00	1367.71	2525.00	5802.00	26.67	0.0015
5.6	07/01/2008	51.03 ,-179.34	1	11	3787.93	5389.41	2914.56	3700.00	811.40	3275.00	4400.00	24.70	0.0022
5.6	11/03/2008	51.96 ,-169.37	1	11	3636.64	5355.58	4674.33	3350.00	1415.15	2850.00	5320.00	22.10	0.0019
5.6	24/04/2012	5.67 , 61.39	2	8	1703.89	3287.09	3023.87	1400.00	971.20	2785.00	5734.00	26.14	0.0014
5.6	06/11/2012	9.92 , 57.04	2	8	2375.43	3332.64	3655.11	1700.00	1659.11	2615.38	5274.61	26.16	0.0017
5.6	15/06/2014	36.64 , 141.71	2	11	3383.55	5495.85	1721.48	3200.00	1099.29	2841.66	8290.00	22.05	0.0023
5.7	26/03/2009	-27.40 , 73.27	2	8	2355.61	3669.86	3324.98	2200.00	841.97	3025.00	6190.00	23.01	0.0030
5.7	27/01/2013	-16.08 ,-172.89	2	11	3985.31	4149.83	2944.62	3800.00	1201.13	3191.66	6453.00	26.31	0.0017
5.7	04/04/2017	51.23 ,-176.24	1	11	3696.51	5286.50	5148.68	3550.00	1030.38	3525.00	7330.00	25.20	0.0017
5.7	31/10/2017	-15.46 ,-173.01	2	11	3919.84	4203.12	3028.57	3700.00	1294.28	3141.66	6253.33	21.76	0.0024
5.8	18/09/2008	51.78 , 158.86	1	11	3710.92	5411.31	1676.03	3550.00	1080.92	4012.50	10255.0	0 27.14	0.0026
5.8	01/10/2009	-15.11 ,-172.89	1	11	3879.26	4203.24	3336.30	3200.00	2192.86	3300.00	5746.00	21.24	0.0026
5.8	24/05/2017	52.26 ,-166.65	2	11	3658.59	5536.08	3880.28	3500.00	1065.49	4075.00	4900.00	24.37	0.0022
5.8	21/04/2018	-49.07 , 123.76	2	1	2223.75	4320.37	4014.00	2050.00	861.72	4025.00	7580.00	23.55	0.0028
5.9	27/06/2008	11.09 , 91.95	2	8	3022.21	3787.33	1101.16	2750.00	1253.49	3518.75	7942.50	23.42	0.0051
5.9	21/12/2008	36.63 , 142.41	2	11	3319.30	5623.09	3710.49	3050.00	1309.68	4317.00	5955.00	25.72	0.0024
5.9	19/03/2011	39.67 , 143.29	1	11	3451.22	5551.84	2122.66	3150.00	1410.11	4237.50	5842.50	24.02	0.0029
5.9	12/01/2016	-31.39 , 58.15	2	8	3129.00	3701.31	2981.43	2750.00	1492.70	3787.50	7830.00	24.89	0.0039
5.9	31/05/2016	47.58 , 154.49	1	11	3422.69	5359.30	2162.20	3300.00	908.21	3800.00	10430.00	27.31	0.0031
6.0	04/08/2010	46.01 , 153.43	2	11	3314.51	5342.96	7101.91	3100.00	1173.02	4758.33	6560.00	24.36	0.0033
6.0	23/04/2011	39.21 , 142.91	1	11	3450.27	5528.79	1726.58	3200.00	1290.10	4095.00	9598.00	22.02	0.0050
6.0	25/11/2013	45.57 , 151.20	1	11	3390.23	5282.90	1362.88	3200.00	1119.66	4275.00	10793.33	28.36	0.0039

6.0	14/10/2015	48.83 , 156.54	1	11	3469.31	5467.96	5422.49	3400.00	690.01	4000.00	5240.00	27.38	0.0027
6.0	19/03/2017	-8.02 , 161.07	1	11	3098.05	4177.54	3239.67	2950.00	946.25	3300.00	7530.00	25.02	0.0041
6.1	02/11/2008	51.38 ,-174.49	1	11	3662.28	5425.47	5152.81	3500.00	1078.09	3350.00	7080.00	31.17	0.0310
6.1	21/04/2010	-15.21 ,-172.83	2	11	3889.03	4193.29	3209.84	3800.00	827.38	2750.00	9600.00	26.83	0.0047
6.1	04/08/2011	48.70 , 155.28	1	11	3504.17	5348.77	819.04	3350.00	1027.97	3500.00	9925.00	18.17	0.0113
6.1	02/08/2012	-4.93 , 153.14	1	11	3086.40	3624.80	2604.56	2800.00	1298.42	3425.00	10613.33	21.16	0.0044
6.1	31/01/2013	-10.64 , 166.48	2	11	3323.51	4197.54	2473.14	2900.00	1623.48	4182.00	10838.57	24.71	0.0047
6.1	20/04/2013	49.97 , 157.58	1	11	3554.23	5469.10	2657.43	3400.00	1035.66	3525.00	9820.00	22.24	0.0045
6.1	08/12/2013	44.53 ,149.24	1	11	3410.80	5350.55	3012.95	3200.00	1180.48	4383.33	9170.00	21.07	0.0077
6.2	07/09/2009	-10.36 , 110.72	2	8	4431.89	5018.49	5768.97	4250.00	1256.65	3637.50	8135.00	26.67	0.0040
6.2	01/02/2010	-6.39 ,154.35	1	11	3167.84	3602.68	5528.87	2850.00	1383.01	4020.83	14418.33	22.92	0.0066
6.2	17/08/2011	36.80 , 143.84	2	11	3202.68	5596.45	6348.50	2900.00	1359.10	4820.00	8254.00	25.64	0.0043
6.2	23/09/2016	34.47 , 141.74	1	11	3249.72	5574.23	6679.05	2900.00	1466.53	4690.00	8884.00	22.99	0.0045
6.2	20/09/2017	38.02 ,144.69	2	11	3214.03	5629.59	5650.60	3000.00	1153.26	4933.33	5886.67	25.10	0.0041
6.3	14/08/2010	12.24 , 141.37	2	11	2927.35	5132.97	4336.60	2700.00	1131.11	4733.33	5976.67	24.24	0.0056
6.3	03/09/2012	-10.96 , 113.91	2	8	4791.01	5018.44	6086.84	4450.00	1775.20	4487.50	8863.33	22.15	0.0067
6.3	20/07/2014	44.63 , 148.94	1	11	3438.98	5333.40	1068.08	3200.00	1259.61	4025.00	10333.33	21.45	0.0126
6.4	13/10/2009	52.54 ,-166.99	1	11	3689.71	5469.02	4427.63	3550.00	1005.73	4025.00	7775.00	24.98	0.0070
6.4	22/03/2011	37.11 , 144.00	2	11	3209.66	5590.31	6318.72	2900.00	1375.46	5705.00	7726.00	22.29	0.0074
6.4	09/11/2015	51.55 ,-173.02	1	11	3646.14	5484.17	6727.98	3500.00	1021.94	3000.00	7330.00	25.06	0.0080
6.4	08/02/2016	-6.83 , 154.57	1	11	3201.46	3541.24	6145.76	2950.00	1243.73	4272.00	12432.50	23.60	0.0111
6.4	13/08/2017	-3.81 , 101.43	1	8	3389.02	4409.23	877.76	3150.00	1250.18	4075.00	11660.00	23.34	0.0153
6.5	03/03/2008	46.26 , 153.38	1	11	3341.82	5332.94	5237.23	3050.00	1365.75	4881,25	14420.00	28.24	0.0073
6.5	03/09/2010	51.41 ,-175.97	1	11	3707.31	5295.08	4265.53	3550.00	1068.49	4312.50	12670.00	21.66	0.0154
6.5	12/03/2011	37.57 ,142.84	2	11	3342.77	5577.63	2454.68	3150.00	1118.76	5616.67	15213.00	26.40	0.0103
6.5	16/11/2012	49.22 , 155.87	1	11	3535.02	5396.08	1339.67	3350.00	1128.66	4012.50	12275.00	21.68	0.0161
6.5	04/09/2013	51.37 ,-174.82	1	11	3669.93	5396.50	4717.95	3600.00	713.03	4025.00	8780.00	19.88	0.0101
6.5	11/07/2014	36.97 , 142.39	2	11	3343.11	5572.54	2699.21	3100.00	1251.56	5050.00	10543.33	22.71	0.0088
6.6	18/04/2009	45.96 , 151.74	1	11	3397.68	5271.95	1202.86	3200.00	1142.04	4722.22	14566.67	26.44	0.0181
6.6	12/08/2009	32.74 , 140.68	1	11	3258.65	5453.40	2030.40	3050.00	1147.29	4162.50	16585.00	20.15	0.0215
6.6	15/08/2018	51.25 ,-177.81	1	11	3751.84	5290.28	3518.87	3500.00	1351.40	4350.00	16376.67	23.66	0.0194
6.7	09/05/2008	12.36 , 143.28	2	11	2719.87	5391.21	3650.33	2450.00	1181.18	6641.67	14247.50	21.53	0.0117
6.7	22/06/2011	39.99 , 142.51	1	11	3540.03	5439.83	777.79	3300.00	1281.33	6805.56	17533.33	20.72	0.0163
6.7	16/09/2011	40.20 , 143.22	1	11	3495.85	5485.32	1339.07	3300.00	1153.69	7041.67	16490.00	21.10	0.0126
6.7	28/04/2012	-18.79 ,-174.26	2	11	4308.74	4108.66	2891.51	4200.00	961.91	10662.50	13700.00	28.02	0.0107
6.8	28/02/2013	50.83 , 157.93	1	11	3635.08	5417.23	651.25	3400.00	1286.01	6613.89	16416.67	21.51	0.0276
6.8	17/09/2014	13.54 , 144.51	2	11	2553.31	5227.92	2528.31	2300.00	1108.79	5983.33	21616.66	18.79	0.0266
6.9	07/04/2009	46.00 , 151.99	1	11	3388.02	5295.97	2010.34	3250.00	957.16	11066.66	19450.00	22.58	0.0345
6.9	13/08/2010	12.46 , 141.52	2	11	2904.84	5093.50	4003.64	2750.00	935.72	8904.17	25175.00	23.63	0.0218
6.9	14/03/2012	40.88 , 144.93	2	11	3408.84	5578.77	5539.96	3200.00	1174.81	8410.55	22053.33	26.33	0.0267
6.9	18/07/2015	-10.35 , 165.10	2	11	3296.43	4184.96	2131.99	3100.00	1120.91	10413.88	20183.33	22.60	0.0320
6.9	27/07/2015	52.21 ,-169.42	1	11	3664.74	5340.28	4883.93	3350.00	1485.89	7809.38	22782.50	22.32	0.0326
6.9	03/01/2017	-19.28 , 176.00	2	11	4402.97	4172.65	3122.00	4250.00	1150.51	7575.00	20010.00	21.02	0.0286
7.0	30/08/2013	51.44 ,-175.12	1	11	3685.69	5341.88	3612.91	3600.00	790.15	11410.00	29370.00	25.59	0.0316
7.1	22/11/2004	-46.36 , 164.91	1	1	6017.50	3658.46	3966.19	5750.00	1774.20	9428.33	29545.00	24.20	0.0552
7.1	12/04/2008	-55.56 , 158.49	1	1	5710.67	4100.43	2218.03	5500.00	1536.80	8959.17	29216.00	20.07	0.0796
7.1	06/02/2013	-11.32 , 164.86	2	11	3405.50	4218.00	4729.69	3300.00	841.07	11743.33	26240.00	20.50	0.0466
7.1	04/12/2015	-47.74 , 85.23	2	1	3467.59	3992.73	3777.69	3400.00	681.32	11979.16	28480.00	23.69	0.0355
7.2	26/12/2004	6.61 , 92.79	1	8	2811.33	4031.09	2739.99	2500.00	1285.92	10987.50	35724.00	22.42	0.0604
7.3	09/03/2011	38.56 , 142.78	1	11	3415.75	5506.10	1453.67	3250.00	1051.11	12943.33	38610.00	18.00	0.0702
7.3	19/04/2013	46.00 ,150.92	2	11	3447.13	5228.02	1918.04	3300.00	996.36	13018.84	43150.00	20.06	0.0918
7.4	15/01/2009	46.97 , 155.39	1	11	3321.41	5410.34	6486.86	3150.00	1053.21	13741.67	47430.00	25.23	0.0637
7.4	07/10/2009	-13.12 , 166.37	1	11	3598.91	4031.75	3354.09	3350.00	1315.16	13047.39	38645.00	23.45	0.0446
7.5	10/08/2009	14.16 , 92.94	2	8	3329.95	3597.62	981.49	2900.00	1636.64	13940.00	56550.00	24.77	0.0840
7.5	12/06/2010	7.85 , 91.65	1	8	2778.65	3928.11	3740.27	2600.00	980.25	12230.15	62270.00	23.49	0.1125
7.6	19/03/2009	-23.08 ,-174.23	1	11	4776.31	3961.91	3986.94	4550.00	1452.81	11676.74	53202.00	19.73	0.0837
7.6	07/10/2009	-12.59 , 166.27	1	11	3540.24	4053.37	4179.87	3350.00	1144.91	13761.94	50277.50	20.36	0.0688
7.6	31/08/2012	10.85 , 126.97	1	11	4517.36	5011.38	6022.65	4350.00	1218.20	13202.08	59070.00	26.35	0.0730
7.8	07/10/2009	-11.86 ,166.01	1	11	3459.82	4105.31	5184.57	3250.00	1186.52	14347.09	63327.50	19.16	0.1006
7.9	06/02/2013	-11.18 , 165.21	1	11	3387.77	4168.51	5072.32	3100.00	1366.37	15013.06	79832.00	19.87	0.1208
7.9	30/05/2015	27.94 , 140.56	2	11	3064.36	4791.16	3166.43	2650.00	1538.77	14959.61	77203.75	19.93	0.1152
7.9	17/12/2016	-5.55 , 153.76	1	11	3114.03	3653.90	2401.88	2450.00	1922.15	14681.20	73410.00	22.11	0.0939
9.1	26/12/2004	3.09 , 94.26	1	8	2784.31	4104.85	2069.62	2600.00	996.18	91823.82	408193.33	18.55	0.3944
9.1	11/03/2011	37.52 , 143.05	1	11	3320.89	5599.73	3445.06	3200.00	887.86	87383.58	305700.00	23.50	0.2935

Table A.9: Identified vertical motion tectonic events epicentre water depth (D). H_r is the recording hydrophone station and h is the average water depth for the transect between the epicentre and the recording hydrophone. X and Y are the relative coordinates of the hydrophone from the earthquake epicentre and b, L, T and W_0 the geometric and dynamic effective slender fault retrieves properties by the inverse model. Tp stands for the slip type. The coordinates report (Longitude (°), Latitude (°))

Inverse problem model for acoustic waves

[1]An example for the implementation of the inverse model problem for acoustic waves in MATLAB programming language, based on the model developed by Ref. [63], with an user friendly interface is introduced in this subsection. First, the model reads the signal and calculates the associated spectrogram in order to analyse the frequency spectral content in time, see Fig A.28.

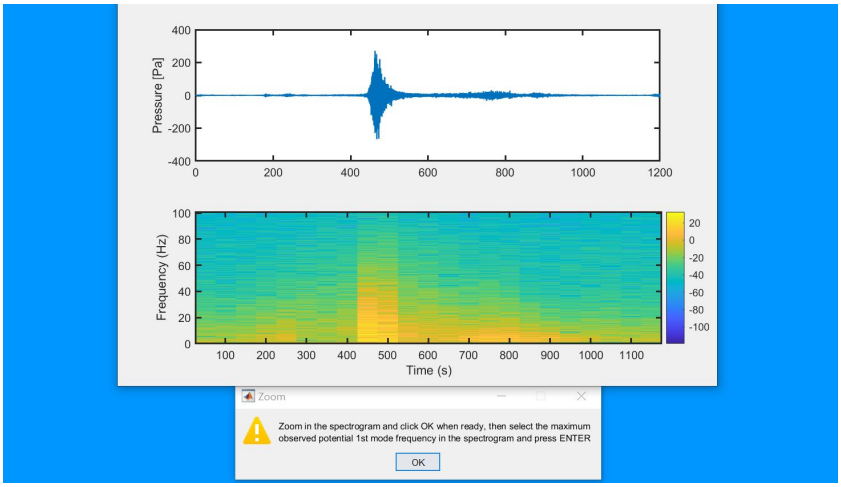

Figure A.28: Inverse model problem user interface, pressure signal and spectrogram.

At this stage, the model requests the minimum and maximum potential frequencies for the first acoustic mode, which have to be selected visually and with the cursor. These points will define the potential orientations of the slender fault that are considered by the model.

[1]The functions utilised for computing the efficient computation of the Fesnel integrals were obtained from the model developed in Ref. [63].

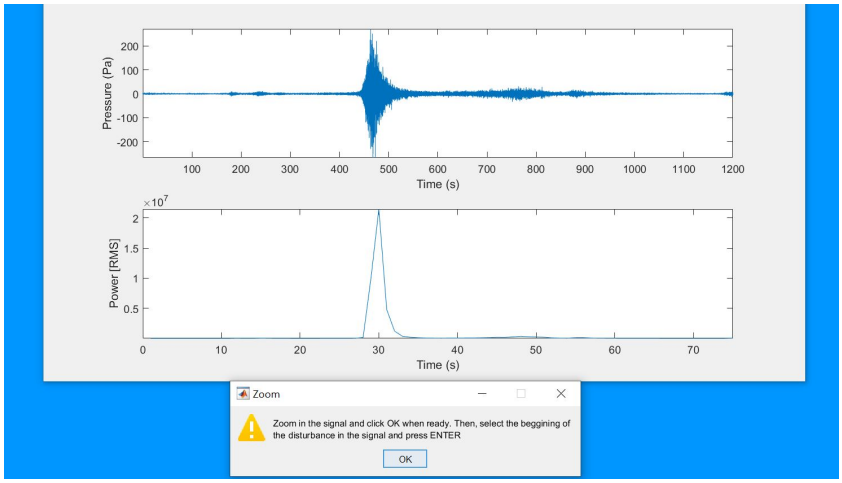

Figure A.29: Inverse model problem user interface, pressure signal and short-time energy.

In the second step, it is necessary to revise the short-time energy distribution that will help to identify the beginning and end of the acoustic disturbance associated with the studied tectonic event, see Fig A.29.

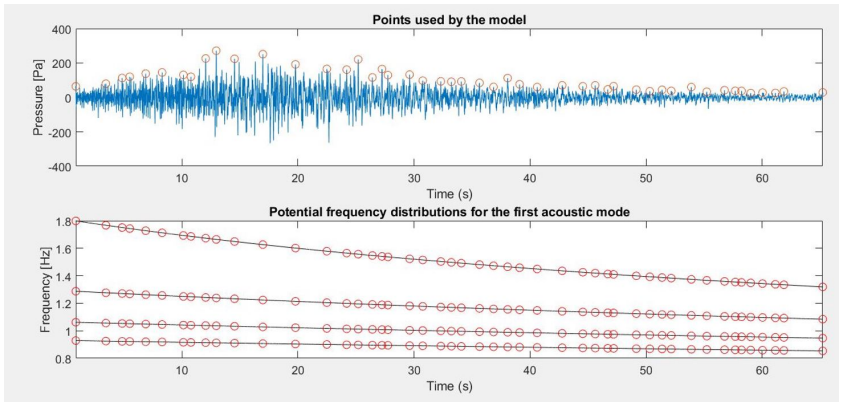

Figure A.30: Inverse model problem user interface, envelope tracking and first mode potential frequency distributions.

The model will track the points close to the envelope and associate them to each of the potential first mode frequency distributions that lie inside the range that we defined previously in the spectrogram, see Fig A.30. Then, we are requested to enter the number of iterations and combinations of points that we want the model to produce, each of them will lead to a set of slender fault characteristics, note that to minimise uncertainties

it is necessary to produce as many solutions as possible. However, it might lead to high computational times, thus, we recommend 10 sets of solutions per potential orientation. We are also asked about the number of points that we require to use per iteration, in the carried sensitivity analysis it was identified that a good balance between CPU time and accuracy is found between four and six points.

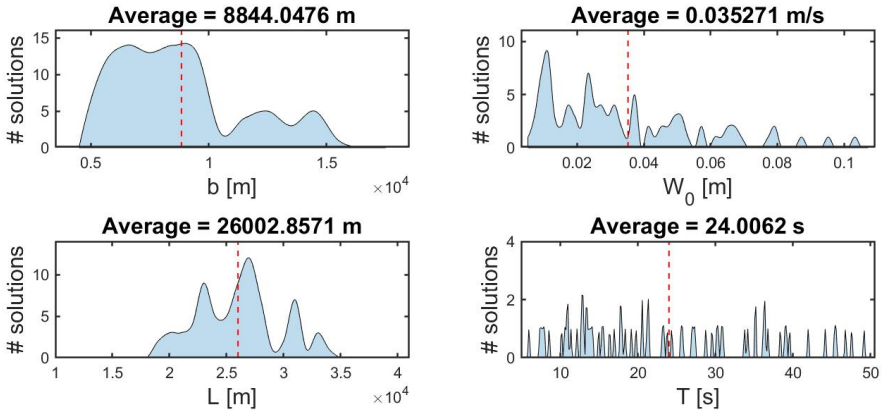

Figure A.31: Inverse model problem output solution distributions.

Finally, the solutions are merged in probability density functions and the mean value is taken as the final result for each characteristic, see Fig A.31.

Bibliography

[1] Abdolali, A, Cecioni, C, Bellotti, G, and Kirby, J. 2015. Hydro-acoustic and tsunami waves generated by the 2012 Haida Gwaii earthquake: Modeling and in situ measurements. Journal of Geophysical Research: Oceans, 120(2), 958-971.

[2] Abdolali A, Cecioni C, Bellotti G, and Sammarco P. 2014. A depth-integrated equation for large scale modeling of tsunami in weakly compressible fluid. Coastal Engineering Proceedings, (34), 9-9.

[3] Abdolali A, Kadri U, and Kirby JT. 2019. Effect of water compressibility, sea-floor elasticity, and field gravitational potential on tsunami phase speed. Sci. Rep. 9 (1), 1–8.

[4] Amato A. 2020. Some reflections on tsunami early warning systems and their impact, with a look at the NEAMTWS. Bollettino di Geofisica Teorica ed Applicata.

[5] An C, Sepúlveda I, and Liu PLF. 2014. Tsunami source and its validation of the 2014 Iquique, Chile, earthquake. Geophysical Research Letters, 41(11), 3988-3994.

[6] Anderson EM. 1905. The dynamics of faulting. Transactions Edinb. Geol. Soc. 8, 387–402.

[7] Ardhuin F, and Herbers THC. 2013. Double-frequency noise generation by surface gravity waves in finite depth: Gravity, acoustic and seismic modes. J. Fluid Mech, 716, 316-348.

[8] Ardhuin F, Lavanant T, Obrebski M, Marié L, Royer JY, d'Eu JF, Howe BM, Lukas R, and Aucan J. 2013. A numerical model for ocean ultra-low frequency noise: Wave-generated acoustic-gravity and Rayleigh modes. The Journal of the Acoustical Society of America, 134(4), 3242-3259.

[9] Bernard E, and Titov V. 2015. Evolution of tsunami warning systems and products. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373(2053), 20140371.

[10] Bernard E, González FI, Meinig C, and Milburn HB. 2001. Early detection and real-time reporting of deep-ocean tsunamis. In International Tsunami Symposium (pp. 7-10).

[11] Bilek SL, and Lay T. 1999. Rigidity variations with depth along interplate megathrust faults in subduction zones. Nature 400 (6743), 443.

[12] Blewitt G, Kreemer C, Hammond WC, Plag HP, Stein S, and Okal E. 2006. Rapid determination of earthquake magnitude using GPS for tsunami warning systems. Geophysical Research Letters, 33(11).

[13] Boersma J. 1960. Computation of Fresnel integrals. Math. Comput., 14(69-72).

[14] Bommer JJ, Douglas J. and Strasser FO. 2003. Style-of-faulting in ground-motion prediction equations. Bull. Earthq. Eng. 1, 171–203.

[15] Borrero, J, Kalligeris, N, Lynett, P, Fritz, H, Newman, A, and Convers, J. 2014. Observations and modeling of the August 27, 2012 earthquake and tsunami affecting El Salvador and Nicaragua. Pure and Applied Geophysics, 171(12), 3421-3435.

[16] Boser BE, Guyon IM, and Vapnik VN. 1992. A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory, 144–152.

[17] Bracewell RN, and Bracewell RN. 1986. The Fourier transform and its applications (Vol. 31999, pp. 267-272). New York: McGraw-Hill.

[18] Breiman L, Friedman JH, Olshen RA and Stone CJ. 2017. Classification and regression trees. Routledge.

[19] Buchan SJ, Mahú R, Wuth J, Balcazar-Cabrera N, Gutierrez L, Neira S, and Yoma NB. 2020. An unsupervised Hidden Markov Model-based system for the detection and classification of blue whale vocalizations off Chile. Bioacoustics, 29(2), 140-167.

[20] Cecioni, C, Abdolali, A, Bellotti, G, and Sammarco, P. 2015. Large-scale numerical modeling of hydro-acoustic waves generated by tsunamigenic earthquakes. Natural Hazards and Earth System Sciences, 15(3), 627-636.

[21] Cecioni C, and Bellotti G. 2018. On the resonant behavior of a weakly compressible water layer during tsunamigenic earthquakes. Pure Appl. Geophys. 175 (4), 1355–1361.

[22] Chierici F, Pignagnoli L, and Embriaco D. 2009. Modelling of the hydro-acoustic signal generated by a tsunami source with account of a porous seabed. In EGU General Assembly Conference Abstracts (p. 8853).

[23] Cipli G, Sattar F. and Driessen PF. 2015. Multi-class acoustic event classification of hydrophone data. In 2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), 473–478.

[24] Decker K, Hirata K and Groudev P. 1992. Guidance document on practices to model and implement earthquake. Bull. Seism. Soc. Am 96, 2181–2205.

[25] Desai B, Maskrey A, Peduzzi P, De Bono A, and Herold C. 2015. Making development sustainable: the future of disaster risk management, global assessment report on disaster risk reduction.

[26] Dziak RP. 2001. Empirical relationship of t-wave energy and fault parameters of northeast pacific ocean earthquakes. Geophys. Res. Lett. 28, 2537–2540.

[27] Dziak RP, Bohnenstiehl DR, Matsumoto H, Fox CG, Smith DK, Tolstoy M, Lau TK, Haxel JH, and Fowler MJ. 2004. P-and t-wave detection thresholds, pn velocity

estimate, and detection of lower mantle and core p-waves on ocean sound-channel hydrophones at the mid-atlantic ridge. Bull. Seismol. Soc. Am. 94, 665–677.

[28] Dziewonski AM, Chou TA, and Woodhouse JH. 1981. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. Journal of Geophysical Research: Solid Earth, 86(B4), 2825-2852.

[29] Ekstrom G, Nettles M, Dziewonski A. 2012. The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes. Phys. Earth Planet. Inter. 200, 1–9.

[30] Eyov E, Klar A, Kadri U, and Stiassnie M. 2013. Progressive waves in a compressible-ocean with an elastic bottom. Wave Motion 50 (5), 929–939.

[31] Feroze K, Sultan S, Shahid S, and Mahmood F. 2018. Classification of underwater acoustic signals using multi-classifiers. In 2018 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST), 723–728.

[32] Flanders Marine Institute (VLIZ); Intergovernmental Oceanographic Commission (IOC), 2021: Sea level station monitoring facility. Accessed at http://www.ioc-sealevelmonitoring.org on 2021-12-07 at VLIZ. DOI: 10.14284/482.

[33] Fox CG, Matsumoto H, and Lau TKA. 2001. Monitoring pacific ocean seismicity from an autonomous hydrophone array. J. Geophys. Res. Solid Earth 106, 4183–4206.

[34] Frohlich C. 2001. Display and quantitative assessment of distributions of earthquake focal mechanisms. Geophys. J. Int. 144, 300–308.

[35] Fujii Y, and Satake K. 2007. Tsunami source of the 2004 Sumatra–Andaman earthquake inferred from tide gauge and satellite data. Bull. Seismol. Soc. Am. 97 (1A), S192–S207.

[36] Fujiwara T, dos Santos Ferreira C, Bachmann AK, Strasser M, Wefer G, Sun T, Kanamatsu T, and Kodaira S. 2017. Seafloor displacement after the 2011 Tohoku-Oki earthquake in the northern Japan trench examined by repeated bathymetric surveys. Geophysical Research Letters 44, no. 23: 11-833.

[37] Gahalaut V, Nagarajan B, Catherine J, and Kumar S. 2006. Constraints on 2004 Sumatra–Andaman earthquake rupture from GPS measurements in Andaman–Nicobar islands. Earth Planet. Sci. Lett. 242 (3-4), 365–374.

[38] Gardner-Stephen P, Wallace A, Hawtin K, Al-Nuaimi G, Tran A, Le Mozo, T, and Lloyd, M. 2019. Reducing cost while increasing the resilience & effectiveness of tsunami early warning systems. In 2019 IEEE Global Humanitarian Technology Conference (GHTC) (pp. 1-8). IEEE.

[39] Ghosh J, Deuser L, and Beck SD. 1992. A neural network based hybrid system for detection, characterization, and classification of short-duration oceanic signals. IEEE J. Ocean. Eng. 17, 351–363.

[40] Gica E, Teng MH, Liu PLF, Titov V, and Zhou H. 2007. Sensitivity analysis of source parameters for earthquake-generated distant tsunamis. Journal of waterway, port, coastal, and ocean engineering, 133(6), 429-441.

[41] Gonzalez FI, Milburn HM, Bernard EN, and Newman JC. 1998, January. Deep-ocean assessment and reporting of tsunamis (DART): Brief overview and status report. In Proceedings of the International Workshop on Tsunami Disaster Mitigation (Vol. 19, p. 2). Tokyo, Japan: NOAA.

[42] Gomez B, and Kadri U. 2021. Earthquake source characterization by machine learning algorithms applied to acoustic signals. Scientific Reports 11.1: 1-15.

[43] Gomez B, and Kadri U. 2021. Near real-time calculation of submarine fault properties using an inverse model of acoustic signals. Appl. Ocean. Res. 109, 102557.

[44] Greenslade DJ, and Titov VV. 2008. A comparison study of two numerical tsunami forecasting systems. Pure and applied geophysics, 165(11), 1991-2001.

[45] Hanson J, Le Bras R, Dysart P, Brumbaugh D, Gault A, and Guern J. 2001. Operational processing of hydroacoustics at the Prototype International Data Center. Pure and Applied Geophysics, 158(3), 425-456.

[46] Hamling IJ, Hreinsdóttir S, Clark K, Elliott J, Liang C, Fielding E, Litchfield N, Villamor P, Wallace L, Wright TJ, and D'Anastasio E. 2017. Complex multifault rupture during the 2016 mw 7.8 Kaikoura earthquake, New Zealand. Science 356 (6334).

[47] Hamlington BD, Leben RR, Godin OA, and Irisov VG. 2010. On the feasibility of tsunami detection using satellite-based sea surface roughness measurements. In 2010 IEEE International Geoscience and Remote Sensing Symposium (pp. 3035-3038). IEEE.

[48] Hendin G, and Stiassnie M. 2013. Tsunami and acoustic-gravity waves in water of constant depth. Phys. Fluids 25 (8), 086103.

[49] Ho TK. 1995. Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition, vol. 1, 278–282.

[50] Ikhsan I, Novamizanti L, and Ramatryana INA. 2014. Automatic musical genre classification of audio using hidden markov model. In 2014 2nd International Conference on Information and Communication Technology (ICoICT), 397–402.

[51] Irikura K, Iwata T, Sekiguchi H, Pitarka A, and Kamae K. 1996. Lesson from the 1995 Hyogo-ken Nanbu earthquake: Why were such destructive motions generated to buildings. J. Nat. Disaster Sci, 17(2), 99-127.

[52] Jalil M, Butt FA, and Malik A. 2013. Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals. In 2013 The international conference on technological advances in electrical, electronics and computer engineering (TAEECE), 208–212.

[53] Jensen A. and la Cour-Harbo A. 2001. Ripples in mathematics: the discrete wavelet transform. Springer Science & Business Media.

[54] Jensen FB, Kuperman WA, Porter MB, and Schmidt H. 2011. Computational Ocean Acoustics. Springer Science & Business Media.

[55] Johnston D, Pettersson R, Downes G, Paton D, Leonard G, Pishief K, and Bell R. 2008. Developing an effective tsunami warning system: Lessons from the 1960 Chile earthquake tsunami for New Zealand coastal communities. Kotuitui: New Zealand Journal of Social Sciences Online, 3(2), 105-120.

[56] Kadri U. 2015. Wave motion in a heavy compressible fluid: revisited. European Journal of Mechanics-B/Fluids, 49, 50-57.

[57] Kadri U. 2016. Triad resonance between a surface-gravity wave and two high frequency hydro-acoustic waves. European Journal of Mechanics-B/Fluids, 55, 157-161.

[58] Kadri U. 2017. Tsunami mitigation by resonant triad interaction with acoustic–gravity waves. Heliyon, 3(1), e00234.

[59] Kadri U. 2019. Effect of sea-bottom elasticity on the propagation of acoustic–gravity waves from impacting objects. Scientific reports, 9(1), 1-7.

[60] Kadri U, and Akylas TR. 2016. On resonant triad interactions of acoustic–gravity waves. Journal of Fluid Mechanics, 788.

[61] Kadri U, and Stiassnie M. 2012. Acoustic-gravity waves interacting with the shelf break. Journal of Geophysical Research: Oceans, 117(C3).

[62] Kadri U, and Stiassnie M. 2013. Generation of an acoustic-gravity wave by two gravity waves, and their subsequent mutual interaction. Journal of Fluid Mechanics, 735.

[63] Kadri U, Crivelli D, Parsons W, Colbourne B, Ryan A. 2017. Rewinding the waves: tracking underwater signals to their source. Sci. Rep. 7 (1), 13949.

[64] Kadri U, and Wang Z. 2021. Approximate solution of nonlinear triad interactions of acoustic–gravity waves in cylindrical coordinates. Communications in Nonlinear Science and Numerical Simulation, 93, 105514.

[65] Kanamori H. 1977. The energy release in great earthquakes. Journal of geophysical research 82.20: 2981-2987.

[66] Kanamori H. and Cipar JJ. 1974. Focal process of the great chilean earthquake may 22, 1960. Phys. Earth Planet. Interiors 9 (2), 128–136.

[67] Kanamori H, and Rivera L. 2008. Source inversion ofWphase: speeding up seismic tsunami warning. Geophysical Journal International, 175(1), 222-238.

[68] Kaverina A, Lander A. and Prozorov A. 1996. Global creepex distribution and its relation to earthquake-source geometry and tectonic origin. Geophys. J. Int. 125, 249–265.

[69] Koketsu K, Yokota Y, Nishimura N, Yagi Y, Miyazaki SI, Satake K, Fujii Y, Miyake H, Sakai SI, Yamanaka Y, and Okada T. 2011. A unified source model for the 2011 Tohoku earthquake. Earth and Planetary Science Letters 310, no. 3-4: 480-487.

[70] Kong Q, Trugman DT, Ross ZE, Bianco MJ, Meade BJ, and Gerstoft P. 2019. Machine learning in seismology: Turning data into insights. Seismological Research Letters, 90(1), 3-14.

[71] Krüger F, and Ohrnberger M. 2005. Tracking the rupture of the m w =9.3 Sumatra earthquake over 1,150 km at teleseismic distance. Nature 435 (7044), 937.

[72] Küçükbayrak M, Güne¸s Ö, and Arica N. 2009. Underwater acoustic signal recognition methods. J. Nav. Sci. Eng. 5, 64–78.

[73] Kundu A, Bhadauria Y, Basu S, and Mukhopadhyay S. 2017. Application of ann and svm for identification of tsunamigenic earthquakes from 3-component seismic data. In 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 10–13.

[74] LaBrecque J, Rundle JB, and Bawden GW. 2019. Global navigation satellite system enhancement for tsunami early warning systems. Global Assessment Report on Disaster Risk Reduction.

[75] Li Z, An C, and Liu H. 2020. Evaluation of different earthquake scaling relations on the generation of tsunamis and hazard assessment. Ocean Engineering, 195, 106716.

[76] Lin SC, Wu TR, Yen E, Chen HY, Hsu J, Tsai YL, Lee CJ, and Philip LFL. 2015. Development of a tsunami early warning system for the South China Sea. Ocean Engineering 100: 1-18.

[77] Liu PLF, Cho YS, Yoon SB, and Seo SN. 1995. Numerical simulations of the 1960 Chilean tsunami propagation and inundation at Hilo, Hawaii. In Tsunami: Progress in prediction, disaster prevention and warning (pp. 99-115). Springer, Dordrecht.

[78] Longuet-Higgins MS. 1950. A theory of the origin of microseisms. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 243(857), 1-35.

[79] Maksymowicz A, Chadwell C, Ruiz J, Trehu A, Contreras-Reyes E, Weinrebe W, Díaz-Naveas J, Gibson J, Lonsdale P, and Tryon M. 2017. Coseismic seafloor deformation in the trench region during the mw 8.8 maule megathrust earthquake. Sci. Rep. 7, 45918.

[80] Mansinha LA, and Smylie DE. 1971. The displacement fields of inclined faults. Bulletin of the Seismological Society of America, 61(5), 1433-1440.

[81] Mei CC. 1989. The applied dynamics of ocean surface waves (Vol. 1). World scientific.

[82] Mei CC, and Kadri U. 2018. Sound signals of tsunamis from a slender fault. J. Fluid Mech. 836, 352–373.

[83] Michele S, and Renzi E. 2020. Effects of the sound speed vertical profile on the evolution of hydroacoustic waves. Journal of Fluid Mechanics, 883.

[84] Mielenz KD. 2000. Computation of Fresnel integrals. II. Journal of research of the National Institute of Standards and Technology, 105(4), 589.

[85] Minson SE, Simons M, Beck JL, Ortega F, Jiang J, Owen SE, Moore AW, Inbal A, and Sladen A. 2014. Bayesian inversion for finite fault earthquake source models–II: the 2011 great Tohoku-oki, Japan earthquake. Geophysical Journal International, 198(2), 922-940.

[86] Miyoshi H. 1954. Generation of the tsunami in compressible water (Part I). Journal of the Oceanographical Society of Japan, 10(1), 1-9.

[87] Mohebbi-Kalkhoran H, Zhu C, Schinault M, and Ratilal P. 2019. Classifying humpback whale calls to song and non-song vocalizations using bag of words descriptor on acoustic data. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 865–870.

[88] Moreno MS, Bolte J, Klotz J, and Melnick D. 2009. Impact of megathrust geometry on inversion of coseismic slip from geodetic data: application to the 1960 chile earthquake. Geophys. Res. Lett. 36 (L16310) https://doi.org/10.1029/ 2009GL039276

[89] Mousavi SM. and Beroza GC. 2020. A machine-learning approach for earthquake magnitude estimation. Geophys. Res. Lett. 47, e2019GL085976.

[90] Munk WH. 1974. Sound channel in an exponentially stratified ocean, with application to sofar. The J. Acoust. Soc. Am. 55, 220–226.

[91] National Geophysical Data Center, NOAA Boulder, CO, USA, NOAA. 2015. National Geophysical Data Center/World Data Service (NGDC/WDS): Global Historical Tsunami Database.

[92] NOAA, 2005. National oceanic and atmospheric administration (NOAA): Deep-ocean assessment and reporting of tsunamis (dart(r)). doi:10.7289/V5F18WNS. Accessed: 2021-06-10.

[93] Nosov M. 1999. Tsunami generation in compressible ocean. Phys. Chem. Earth Part B 24 (5), 437–441.

[94] Nosov M, and Kolesov S. 2007. Elastic oscillations of water column in the 2003 Tokachi- Oki tsunami source: in-situ measurements and 3-D numerical modelling. Nat. Hazards Earth Syst. Sci. 7 (2), 243–249.

[95] Nosov, M, and Kolesov, S. 2011. Optimal initial conditions for simulation of seismotectonic tsunamis. Pure and Applied Geophysics, 168(6), 1223-1237.

[96] Ocean Networks Canada Data Archive. Hydrophone data from 03:48 UTC 2014/09/10 to 04:03 UTC 2014/09/10, Oceans Networks Canada, University of Victoria, Canada. Downloaded on September 2019. Accessed: 2020. http://www.oceannetworks.ca.

[97] Ochoa LH, Niño LF, and Vargas CA. 2018. Fast magnitude determination using a single seismological station record implementing machine learning techniques. Geod. Geodyn. 9, 34–41.

[98] Okada Y. 1985. Surface deformation due to shear and tensile faults in a half-space. Bulletin of the seismological society of America, 75(4), 1135-1154.

[99] Okal EA. 2001. T-phase stations for the international monitoring system of the comprehensive nuclear-test ban treaty: a global perspective. Seismol. Res. Lett. 72 (2), 186–196.

[100] Okal EA. 2008. The generation of T waves by earthquakes. Advances in Geophysics, 49, 1-65.

[101] Okal EA, Talandier J, and Reymond D. 2007. Quantification of hydrophone records of the 2004 Sumatra tsunami. Pure Appl. Geophys. 164 (2-3), 309–323.

[102] Pararas-Carayannis G. 2014. The Great Tohoku-Oki earthquake and tsunami of March 11, 2011 in Japan: A critical review and evaluation of the tsunami source mechanism. Pure and applied geophysics, 171(12), 3257-3278.

[103] Provost F. 2000. Machine learning from imbalanced data sets 101. In Proceedings of the AAAI'2000 workshop on imbalanced data sets (Vol. 68, No. 2000, pp. 1-3). AAAI Press.

[104] Pupatenko VV, and Shestakov NV. 2021. Using Global Navigation Systems for Early Tsunami Warning about Tsunamis Excited by Magnitude 7–8 Earthquakes. Journal of Volcanology and Seismology, 15(1), 26-34.

[105] Purcaru G, and Berckhemer H. 1982. Quantitative relations of seismic source parameters and a classification of earthquakes. Tectonophysics 84, 57–128.

[106] Rakowsky N, Androsov A, Fuchs A, Harig S, Immerz A, Danilov S, Hiller W, and Schröter J. 2013. Operational tsunami modelling with TsunAWI–recent developments and applications. Natural Hazards and Earth System Sciences, 13(6), 1629-1642.

[107] Rasyif TM, and Kato S. 2015. Development of accurate tsunami estimated times of arrival for tsunami-prone cities in Aceh, Indonesia. International Journal of Disaster Risk Reduction, 14, 403-410.

[108] Reddy R, and Nair RR. 2013. The efficacy of support vector machines (svm) in robust determination of earthquake early warning magnitudes in central japan. J. Earth Syst. Sci. 122, 1423–1434.

[109] Renzi E, and Dias F. 2014. Hydro-acoustic precursors of gravity waves generated by surface pressure disturbances localised in space and time. Journal of fluid mechanics, 754, 250-262.

[110] Saito T, Ito Y, Inazu D, and Hino R. 2011. Tsunami source of the 2011 Tohoku-Oki earthquake, Japan: Inversion analysis based on dispersive tsunami simulations. Geophysical Research Letters, 38(7).

[111] Sells CL. 1965. The effect of a sudden change of shape of the bottom of a slightly compressible ocean. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 258(1092), 495-528.

[112] Shawe-Taylor J, and Cristianini N. 2000. An introduction to support vector machines and other kernel-based learning methods (Vol. 204).

[113] Simons FJ, Dando BD, and Allen RM. 2006. Automatic detection and rapid determination of earthquake magnitude by wavelet multiscale analysis of the primary arrival. Earth Planet. Sci. Lett. 250, 214–223.

[114] Small C, and Nicholls RJ. 2003. A global analysis of human settlement in coastal zones. Journal of coastal research: 584-599.

[115] Stiassnie M. 2010. Tsunamis and acoustic-gravity waves from underwater earthquakes. J. Eng. Math. 67 (1-2), 23–32.

[116] Stirling M, Goded T, Berryman K, and Litchfield, N. 2013. Selection of earthquake scaling relationships for seismic-hazard analysis. Bulletin of the Seismological Society of America, 103(6), 2993-3011.

[117] Sukhovich A, Irisson JO, Perrot J, and Nolet G. 2014. Automatic recognition of t and teleseismic p waves by statistical analysis of their spectra: An application to continuous records of moored hydrophones. J. Geophys. Res. Solid Earth 119, 6469–6485.

[118] Sukhovich A, Irisson JO, Simons FJ, Ogé A, Hello Y, Deschamps A, and Nolet G. 2011. Automatic discrimination of underwater acoustic signals generated by teleseismic p-waves: A probabilistic approach. Geophys. research letters 38.

[119] Synolakis C, Liu P, Philip HA, Carrier G, and Yeh H. 1997. Tsunamigenic sea-floor deformations. Science, 278(5338), 598-600.

[120] Titov VV, and Gonzalez FI. 1997. Implementation and testing of the method of splitting tsunami (MOST) model.

[121] Tolstoy I, and Ewing M. 1950. The t phase of shallow-focus earthquakes. Bull. Seismol. Soc. Am. 40, 25–51.

[122] Tschiatschek S, Mutsam N. and Pernkopf F. 2012. Handling missing features in maximum margin bayesian network classifiers. In 2012 IEEE International Workshop on Machine Learning for Signal Processing, 1–6.

[123] Tsushima H, Hirata K, Hayashi Y, Tanioka Y, Kimura K, Sakai SI, Shinohara M, Kanazawa T, Hino R, and Maeda K. 2011. Near-field tsunami forecasting using offshore tsunami data from the 2011 off the Pacific coast of Tohoku Earthquake. Earth, planets and space, 63(7), 821-826.

[124] Tuma M, Igel C, and Prior M. 2012. Hydroacoustic signal classification using support vector machines. In Signal and Image Processing for Remote Sensing, 37–56.

[125] Tuma M, Rørbech V, Prior MK, and Igel C. 2016. Integrated optimization of long-range underwater signal detection, feature extraction, and classification for nuclear treaty monitoring. IEEE Transactions on Geosci. Remote. Sens. 54, 3649–3659.

[126] Tzanetakis G, Essl G, and Cook P. 2001. Audio analysis using the discrete wavelet transform. In Proc. Conf. in Acoustics and Music Theory Applications, vol. 66.

[127] Vapnik V, Golowich SE, and Smola AJ. 1997. Support vector method for function approximation, regression estimation and signal processing. In Advances in neural information processing systems, 281–287.

[128] Wächter J, Babeyko A, Fleischer J, Häner R, Hammitzsch M, Kloth A, and Lendholt M. 2012. Development of tsunami early warning systems and future challenges. Natural Hazards and Earth System Sciences, 12(6), 1923-1935.

[129] Wang X, and Liu PLF. 2006. An analysis of 2004 Sumatra earthquake fault plane mechanisms and Indian Ocean tsunami. Journal of Hydraulic Research, 44(2), 147-154.

[130] Wessel P, Luis JF, Uieda L, Scharroo R, Wobbe F, Smith WHF, and Tian D. 2019. The generic mapping tools version 6. Geochemistry, Geophysics, Geosystems, 20(11), 5556-5564.

[131] Wells DL, Coppersmith KJ. 1994. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. Bull. Seismol. Soc. Am. 84 (4), 974–1002.

[132] Williams B, Kadri U, and Abdolali A. 2021. Acoustic–gravity waves from multi-fault rupture. Journal of Fluid Mechanics, 915.

[133] Wu TR, and Huang HC. 2009. Modeling tsunami hazards from Manila trench to Taiwan. Journal of Asian Earth Sciences, 36(1), 21-28.

[134] Yamada M, Heaton T, and Beck J. 2007. Real-time estimation of fault rupture extent using near-source versus far-source classification. Bull. Seismol. Soc. Am. 97, 1890–1910.

[135] Yamamoto T. 1982. Gravity waves and acoustic waves generated by submarine earthquakes. Int. J. Soil Dyn. Earthq. Eng. 1 (2), 75–82.

[136] Yang X, Dias F, and Liao S. 2018. On the steady-state resonant acoustic–gravity waves. Journal of Fluid Mechanics, 849, 111-135.

[137] Yang Y, and Forsyth DW. 2003. Improving epicentral and magnitude estimation of earthquakes from t phases by considering the excitation function. Bull. Seismol. Soc. Am. 93, 2106–2122.

[138] Yuji Y. and Fukahata Y. 2011. Rupture process of the 2011 Tohoku-Oki earthquake and absolute elastic strain release. Geophys. Res. Lett. 38 (L19307). https://doi.org/10.1029/2011GL048701.

[139] Yun NY, and Hamada M. 2015. Evacuation behavior and fatality rate during the 2011 Tohoku-Oki earthquake and tsunami. Earthquake Spectra, 31(3), 1237-1265.