Long Period – Long Duration Seismic Event Location Using a 3D Fast Marching – BackProjection Imaging Method

Author:
CRISTIAN ALEXIS MURILLO MARTÍNEZ

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Geophysics

Advisor:
William Mauricio Agudelo Zambrano
PhD. In Geophysics

Co-Advisors:
Carlos Andrés Niño Niño
Master in Electronics Engineering
Daniel Alfonso Sierra Bueno
PhD. In Biomedical Engineering

Universidad Industrial de Santander
Faculty of Science
School of physics
Master in Geophysics
Bucaramanga
2019

Acknowledgements

First and foremost, God thank you for giving me the necessary strength, wisdom and discipline to overcome all the challenging moments experienced during this thesis. “I will thank God for the day and the moment I have”.

I express my sincere gratitude to my thesis advisors PhD. William Agudelo Zambrano and MIE. Carlos Niño Niño for their guidance, continuous supervision and support in completing this thesis. Thanks for the valuable experience and knowledge I have gained beside them.

I must express my profound gratitude to my parents (John and Hilda), my sister and my nephew, who have provided me through emotional support in my life. Thanks for your encouragement and motivation to never give up despite the hard times. Because of them, I see the opportunities in every difficulty.

I would like to thank professor José David Sanabria, my friends Luis Carlos Prada and Rubén Mauricio Vargas for their help during my master studies.

To all members and staffs from GIEP, CEMOS, Faculty of Science and ICP-Ecopetrol, for providing facilities and resources to carry out the experiments for this study.

A huge “thank you” to everyone who has been a part of my life and has helped me to growth personally and professionally

Table of Contents

List of Tables

List of Figures

List of Appendices

Abbreviation	Explanation
AmI	**Arithmetic-mean Imaging**
BPI	**Back-Projection Imaging**
CIP	**Common Image Point Gathers**
ETS	**Episodic Tremor and Slip**
FD	**Finite Differences**
FMM	**Fast Marching Method**
FWI	**Full Waveform Inversion**
GmI	**Geometric-mean Imaging**
ISSA	**Improved Source Scanning Algorithm**
KRL	**Kirchhoff Reconstruction Location**
LFE	**Low Frequency Events**
LPLD	**Long Period – Long Duration**
MATF	**Maximum Amplitude Time Function**
NAV	**Normalized Absolute Value**
PbAS	**Brightness PDF-based Amplitude Stacking**
PDF	**Probability Density Function**
PrAS	**Spatial PDF-based Amplitude Stacking**
P-wave	**Pressure/Acoustic Wave**
RT	**Ray Tracer**
RTM	**Reverse Time Migration**
SNR	**Signal to Noise Ratio**
SRV	**Stimulated Reservoir Volume**
SSA	**Source Scanning Algorithm**
S-wave	**Shear/Elastic Wave**
t_0	**Origin Time**
t_p & t_s	**P and S -wave Arrival Time**
TRI	**Time-Reversal Imaging**
TTT	**Travel-Time Table**
V_p & V_s	**P and S -wave Velocity**
VSP	**Vertical Seismic Profile**

Resumen

TITULO: LOCALIZACIÓN DE EVENTOS SÍSMICOS DE LARGO PERIODO – LARGA DURACIÓN (LPLD) EMPLEANDO EL MÉTODO MIGRACIÓN – FAST MARCHING 3D[*]

Autor: Cristian Alexis Murillo Martínez[**]

Palabras Clave: Back-Projection Imaging (BPI), localización de sismos, microsismos, tremores, análisis de sensibilidad, incertidumbre.

Descripción:

Generalmente se emplean técnicas cinemáticas para localizar sismos mediante un proceso iterativo que minimiza residuales de tiempos; no obstante, baja relación Señal-Ruido (SNR) genera incertidumbre en el picado que degrada la solución. Asimismo, existen fenómenos sísmicos complejos sin fases distinguibles que dificultan el uso de técnicas puramente cinemáticas. Este trabajo implementa la técnica BPI que utiliza información cinemática y dinámica de la onda a fin de localizar la fuente del evento al apilar las señales registradas en las estaciones, sin resolver la ecuación de onda completa, haciéndola adecuada para análisis de sensibilidad e incertidumbre.

Aunque los métodos BPI convencionales son apropiados para la localización espaciotemporal, estos determinan el tiempo de origen a partir del pico de la Curva de Máximo Brillo (CMB) lo cual puede ser incorrecto, especialmente para eventos de larga duración o con altos niveles de ruido. Consecuentemente, dos nuevos métodos basados en centroide se proponen y comparan con los BPI convencionales para demostrar su desempeño.

Este trabajo está constituido por: capítulo 1: revisión de métodos relevantes, destacando características principales, ventajas y desventajas. Capítulo 2: metodología empleada en la generación y retropropagación de sintéticos. Capítulo 3: perturbaciones del nivel de ruido, frecuencia pico, velocidades, y geometrías de sensores para evaluar su impacto sobre la localización. Capítulo 4: retropropagación de sintéticos realistas y datos reales de 2 eventos para determinar su fuente dentro del volumen de estudio y la probabilidad de inducción por actividad humana (fracturamiento hidráulico). Basados en la comparación con métodos BPI convencionales, los métodos propuestos presentan mayor tolerancia a incertidumbre en la velocidad, además, no requieren una discretización fina para generar resultados confiables. En conclusión, para eventos de larga duración o baja SNR, se podrían desarrollar CMBs sin picos claros, haciendo imprácticos los BPI clásicos, mientras que los métodos basados en centroide podrían dar mejores resultados.

[*] Trabajo de grado – Tesis de investigación

[**] Facultad de Ciencias. Maestría en Geofísica. Director: PhD William M. Agudelo Z., ICP – Ecopetrol. Codirectores: MIE. Carlos A. Niño N. & PhD Daniel A. Sierra B., Escuela de ingeniería Eléctrica – UIS.

Abstract

TITLE: LONG PERIOD – LONG DURATION SEISMIC EVENT LOCATION USING A 3D FAST MARCHING – BACKPROJECTION IMAGING METHOD.*

Author: Cristian Alexis Murillo Martínez**

Keywords: Back-Projection Imaging (BPI), Location of earthquakes, microearthquakes, tremors, sensitivity analysis, uncertainty.

Description:

Seismology typically requires picking of arrivals and phase association to locate earthquakes by minimization of time residuals; however, some seismic phenomena generate complex seismograms, with no distinguishable phases, which does not allow accurate picking and make location difficult to perform. Back-Projection Imaging (BPI) methods, similar to migration in active seismic, can successfully locate sources with enough resolution, and they do not require to solve the full wave equation, but the Eikonal equation in this work. Therefore, BPI can be used to conduct sensitivity analyses and location uncertainty assessment efficiently.

Although conventional BPI methods are suitable for spatiotemporal location, they determine the origin time of an event from the peak of the Maximum Brightness Curve (MBC) which may be an erroneous assumption, especially for long duration or highly noisy events. Subsequently, the 2 proposed centroid-based methods are compared to conventional BPI methods to demonstrate their accuracy.

This thesis is structured as follows: Chapter 1 presents an overview of some relevant methods including their main features, advantages and disadvantages. Chapter 2 describes the methodology followed for the design of velocity models, synthetics and algorithms. In chapter 3, parameters such as background noise level, frequency of the signal, acoustic velocities and recording geometries are changed to evaluate their effects on the spatiotemporal locations. In chapter 4, realistic synthetic seismograms and field data of 2 candidate events were back-projected to determine their location and the probability for these 2 events to have been triggered inside the volume of study by a man-made activity (hydraulic fracturing). In conclusion, conventional BPI methods are stable when there exists a clear peak on the MBC, however, a poor discretization alongside complex seismograms, low SNR or unusual geometries may generate MBC without clear peaks, making conventional BPI impractical for use. In those cases, centroid-based methods may provide better results.

---

* Thesis

** Faculty of science. Master in Geophysics. Advisor: PhD William M. Agudelo Z., ICP – Ecopetrol. Co-Advisors: MIE. Carlos A. Niño N. & PhD Daniel A. Sierra B., School of electrical engineering – UIS.

Introduction

In September of 2015, unusual seismic waveforms were registered on multicomponent 3C recording sensors during a mini-hydraulic fracturing monitoring carried out in the San Francisco field – Upper Magdalena Valley Basin (UMVB). Mini-fracturing stimulation of sand reservoirs has been a common practice for several years worldwide, therefore a multidisciplinary group of geoscientists conducted microseismic monitoring operations during a routinary mini-fracturing to get a better understanding of the deformation mechanisms going on in the subsurface. During the process, the geoscientists identified anomalous events whose spectral content, shape and timing resembled those events reported by (Das & Zoback, 2011) as Long Period- Long Duration (LPLD) possibly triggered by the injection of high pressure fluids (fracturing) in unconventional (shale) reservoirs. Consequently, the team endeavored to provide a valid explanation to those events, however, due to the lack of an adequate location method and other difficulties, this phenomenon could not be fully addressed, and the question was left open.

In this thesis, we propose a location algorithm based on Back-Projection Imaging, willing to solve some of the early questions. For this reason, it is necessary to study the nature of induced seismic events and their spatiotemporal distribution to shed some light on the previous concerns, however, there is a large room for study; especially in Colombia, where microseismic monitoring is a new field, since little or no studies have been carried out on these topics, therefore, we expect this thesis to be an interesting contribution to our Colombian geophysical community because it provides a location method for seismic signals in general, including LPLD, microseisms and tremors, therefore, it will help to gain some insight into the LPLD and tremor-like phenomena.

1. A Review of Seismic and Microseismic Event Location Techniques

1.1. Overview

According to Reid's *elastic rebound theory (1906),* an earthquake is commonly understood as a sudden release of energy, accumulated in form of stress, which builds up until the maximum shear strength is reached; beyond that point, movement along a fault plane occurs and energy is released until a new equilibrium is reached (Shearer, 2009, pp. 301-302), as shown in **figure 1**.

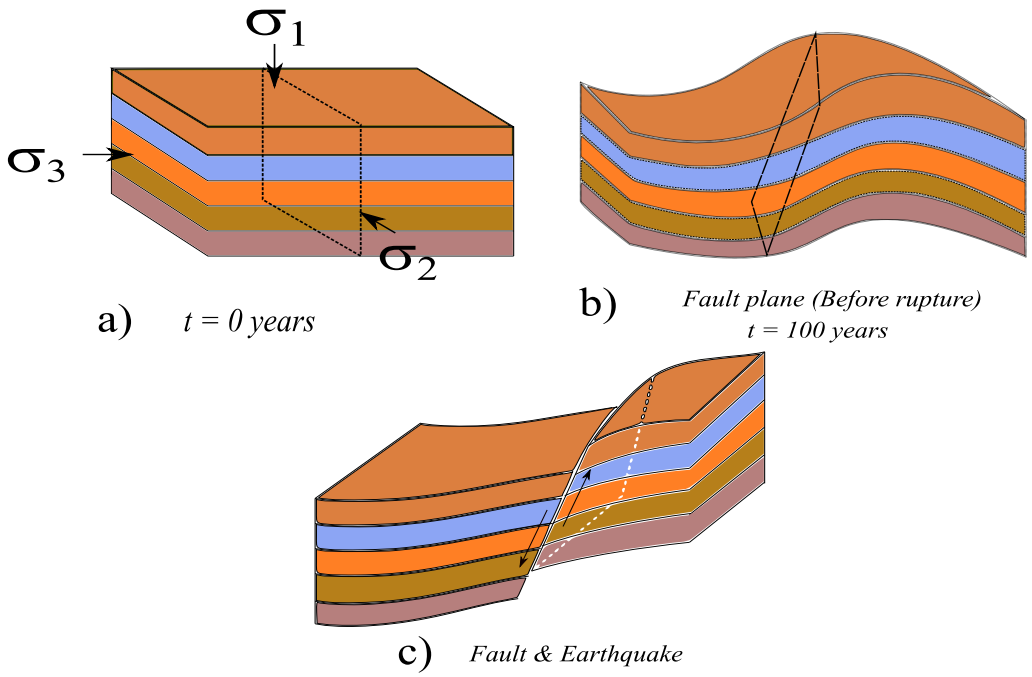


Figure 1. Origin of Earthquakes. (a) Initial state (Undeformed), (b) Compressive stress acting over time, (c) Displacement of faults (pre-existing or new ones).
Adapted from (Shearer, 2009, p. 302)

In 2001, an intriguing type of long period-long duration non-volcanic earthquakes that cannot be completely explained by the elastic rebound theory was identified along the strike of the subducting Philippine Sea plate in Japan (Obara, 2002); such events, now known as ***tectonic tremors***, were found to last from several days to weeks, to occur at an average focal depth of 30km, to have a low amplitude signal with a frequency content of 1-10Hz and no clear P and S wave onsets. Additional research on their nature indicates that tectonic tremors may be related to the ***slow slip*** phenomena in a complex system referred to as ***ETS (Episodic Tremor and Slip)*** (Obara, Hirose, Yamamizu, & Kasahara, 2004). (Brudzinski, 2011) suggests a model of a subduction zone to explain the relation of tremor activity and slow slip; **figure 2** illustrates the occurrence of ETS just below the locked zone between the subducting and overriding plates. Even though this zone can release massive amounts of elastic strain (equivalent to a magnitude 7 earthquake), this process is extremely slow and generates low amplitude – long duration signals. Moreover, ETS might be a trigger for great earthquakes to occur at shallower depths, thus, understanding ETS is a powerful alternative to gain insight into the regional stress accumulation process, boundary plate conditions and crustal deformation.

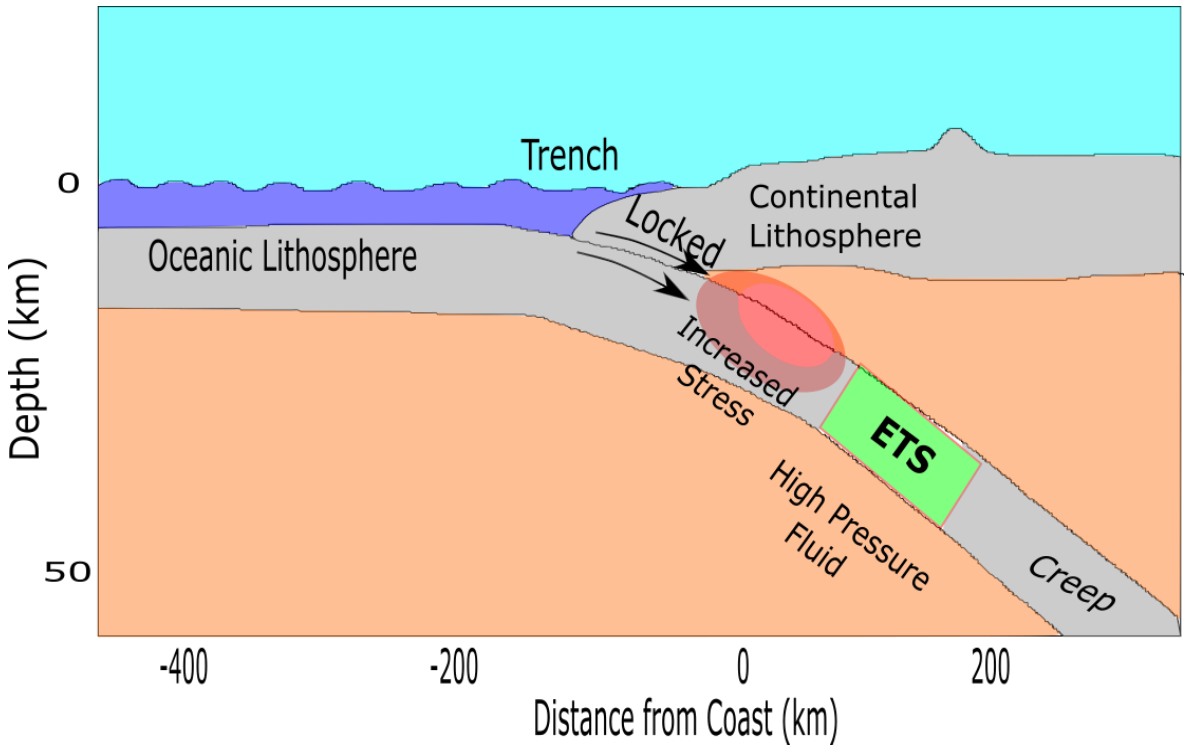


Figure 2. Occurrence of tectonic tremor and slip (ETS) phenomena on a subduction zone. Image adapted from (Brudzinski, 2011, p. 8)

A few years later, (Das & Zoback, 2011) analyzed seismic datasets from the Barnett shale formation recorded during hydraulic fracturing operations (see **Appendix A** for more information regarding this stimulation process) and identified a series of seismic events with similar features to those of tectonic tremors. Afterwards, they localized the events and suggested that the seismicity was associated to the reservoir stimulation process; additionally, they stated that understanding the nature of these phenomena and their spatial distribution, would greatly help to comprehend how reservoir deformation is associated with fault reactivation and permeability enhancement; hence, this thesis aims to make a small contribution to understand the ***Long Period – Long Duration*** (***LPLD***) phenomena by implementing a source location method suitable for this type of seismic signals.

1.2. LPLD and microseismic events

As (Das & Zoback, 2013b, p. KS107) stated: "Although hydraulic fracturing is successfully used to stimulate production from hydrocarbon reservoirs, the principal deformation mechanisms responsible for the stimulation are still poorly understood". Among them, the most widely studied deformation mechanism are ***microseisms*** – i.e. high frequency and low moment magnitude events

Mw [-3 to 3] (Wang, Li, & Shang, 2016, p. 4) - around the stimulated well, produced by opening of new and pre-existing fractures. Nowadays, the spatial distribution of the induced microseisms is commonly used to estimate fracture geometries and Stimulated Reservoir Volume (SRV) even though this correlation is still unsettled (Kumar, Zorn, Hammack, & Harbert, 2017).

(Maxwell & Urbancic, 2001) Outlined some of the applications of passive microseismic monitoring in the oil & gas industry including well stimulation, well/casing failure, fault mapping, fluid movements, compaction strains, and others. (Perez Altamar & Marfurt, 2015) Presented a study where microseismic data in conjunction with well logs, production data and mineralogy were used to generate brittleness maps to improve well completion and enhance gas production. Several other applications are encountered in rock mechanics, geothermal and geotechnical engineering studies.

In addition to the occurrence of microseismicity (microearthquakes) during hydraulic fracturing operations, higher magnitude events with similar characteristics to tectonic tremors have been identified as ***Long Period – Long Duration (LPLD)*** (Das & Zoback, 2011). These events comprise coherent wave trains, predominantly composed of low amplitude S-waves, which makes it arduous to distinguish impulsive P- and S- wave arrivals from background noise (**figure 3**).

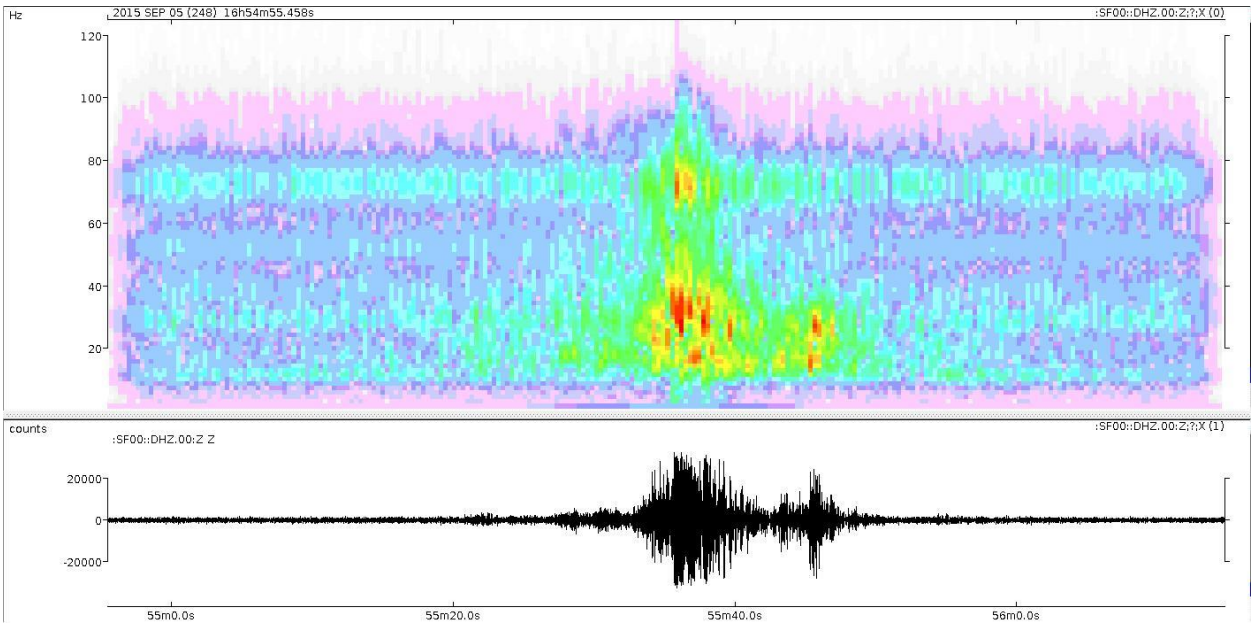


Figure 3. Example of LPLD seismic event. A) Spectrogram, b) Seismogram.

1.2.1. An insight into the nature of LPLD events. Although the nature and mechanisms of LPLD events are still unclear, they are commonly interpreted as the superposition of Low Frequency Events (LFEs) occurring in a cascade effect because of slow slip along large faults, resulting in an overall duration of tens or hundreds of seconds and a predominant frequency band of 10 - 80Hz (Das & Zoback, 2011, p. 1473). In an attempt to explain how these events happen, some complementary information about their nature has been published in the literature; for instance, (Royle, 2006) where tremor sequences identified in the Northern Cascadia subduction zone were studied and located, showing different characteristics for 2 sets of tremors registered in different periods (February and September - 2002); the first type, classified as *ETS*, correlated both in space and time with recursive *slow slip episodes*, in which stress accumulates and releases as a result of the interaction between the overriding and subducting slabs; in this case, slip appears to migrate along the strike of the subduction zone and occurs mainly at a depth range of 10 - 40 km . The second type is *non-episodic tremors* which are even more mysterious, a first approximation suggests they can be understood as product of fluid migration inducing oscillations in conduit walls, perhaps, altering the stress state of the media. Even though it is uncertain, we may think of hydraulic fracturing as an operation which involves both deformation mechanisms, supported by the presence of events of different characteristics (e.g. microseisms and LPLD), contributing together to maximize the overall Stimulated Reservoir Volume (SRV). Now, a few questions arise:

How different is an LPLD from a microearthquake? In short, microearthquakes are sudden releases of small amounts of energy due to unstable rupture of the rock, hence, their frequency content is usually high, very low magnitude, and rupture propagates at the same velocity as typical earthquakes do (0.8 – 0.9 of the shear wave speed *Vs*). On the other hand, LPLD are composed of Low frequency earthquakes that occur in conjunction with slow slip, therefore, rupture propagates at a more stable and slower rate (0.2 *Vs*) which translates into longer duration and lower stress drops (Thomas, Beroza, & Shelly, 2016).

What determines the occurrence of one or the other? Besides the aforementioned reasons, there are other important factors to be evaluated (for instance, clay content and fault orientation) as presented in (Zoback, Kohli, Das, & McClure, 2012). They suggested that shales consisting of less than about 30% clay content and critically stressed faults are expected to generate fractures

whose rupture propagates unstably thus resulting in microseismicity. In contrast, LPLDs seem to be triggered by the injection of high-pressure fluids into misoriented faults with respect to the current stress field, which in conjunction with a clay content of approximately 30% originates a stable and slow slip of the fault, producing low frequency and long duration seismic and aseismic signals that may sometimes not be recorded in the frequency band used in microseismic monitoring. A more detailed comparison of both phenomena is presented in (Das & Zoback, 2012).

1.2.2. Cases of LPLDs in shale plays. Posterior to the identification of LPLDs in 2011, a series of studies in hydraulically fractured shale plays have been carried out seeking the presence of these types of events; some of them are briefly reviewed below:

(Das, 2013a) Analyzed seismic datasets of the Barnett shale plays during and after hydraulic fracturing stimulation, identifying unusual waveforms which seem to have the typical characteristics of LPLD events as described in (Das & Zoback, 2011). Then, (Das & Zoback, 2013b) compared the main features of these recently discovered events to tectonic tremors heavily distributed along subducting plates. Among their findings, they indicated that LPLDs might play a crucial role as a reservoir deformation mechanism, increasing formation permeability due to a larger stimulated volume and hydraulic conductivity enhancement by reactivation of pre-existing faults and fractures. Finally, (Das & Zoback, 2013c) tested the performance of two source location techniques on real data from Barnett. First, the *seismic array method* locates the source by triangulating the back-azimuths obtained from the apparent velocity, for a given signal and seismic array, thus allowing to estimate a region of probability of occurrence, nonetheless, this technique provides no extra constraints regarding the source depth when a single sensor array is used, as it was the case there. On the other hand, the *Source Scanning Algorithm (SSA)* (Kao & Shan, 2004) was implemented using 2 sets of vertical receivers, also resulting in high uncertainty about the depth of the source due to inconsistency between both set of sensors. However, borrowing extra information from well logging, geology and drilling engineering, it was possible to estimate possible fault planes, indicating correlation between the hydraulic fracturing stimulation, reactivation of pre-existing fractures and enhanced hydraulic conductivity.

Later on, a study of 117 impulsive – high amplitude events and 473 LPLDs identified in 3-component surface seismic datasets of Marcellus shale plays was conducted by (Kumar, Zorn, Hammack, & Harbert, 2016) and (Kumar, Zorn, Hammack, & Harbert, 2017) to determine their relation to the hydraulic fracturing stimulation. Whilst the occurrence of the impulsive events

showed no spatiotemporal correlation, the number of LPLD events increased as the pressure and rate of injection reached its maximum during the formation stimulation. Furthermore, waveform signals typical of these events were found in high agreement with those reported by (Das & Zoback, 2011); therefore, possibly indicating the reactivation of pre-existing fractures, creation of new ones and productivity enhancement triggered by the hydraulic fracturing.

1.3. Introduction to seismic event location

Several seismic source location methods have been developed over the years, seeking to improve the accuracy with which hypocenters can be calculated. According to (Wang, Li, & Shang, 2016, p. 19) most microseismic event location methods can be classified into one of the following two main categories: kinematic, that uses arrival times only, and dynamic, which is based on full waveform information.

1.3.1. Travel time-based methods. These methods employ picking of phase arrivals (observed data) and the computation of theoretical arrivals (calculated data) to minimize travel time residuals at each recording station in a given velocity model. Most of these methods belong in one of the following 3 sub-categories: ray tracing algorithms, grid-based schemes, and the derivative approach. According to (Wang, Li, & Shang, 2016, p. 19), most of the techniques currently used in hydraulic fracturing monitoring have been introduced from seismology as modifications of the traditional approaches:

Method of circles (for homogeneous medium). This graphic method determines the origin time of an earthquake as the x-intercept on a graph of arrival time differences between the S and P – waves ($\boldsymbol{t_s - t_p}$) against $\boldsymbol{t_p}$ at different seismograph stations (this plot is called a Wadati diagram **figure 4a**). Once computed its origin time ($\boldsymbol{t_0}$), and for a given velocity model, it is straightforward to calculate hypocentral distances and use them to draw spheres around each station (**figure 4b**); the estimated location of the source lies somewhere within the intersection volume (Lowrie, 2007, pp. 149-150).

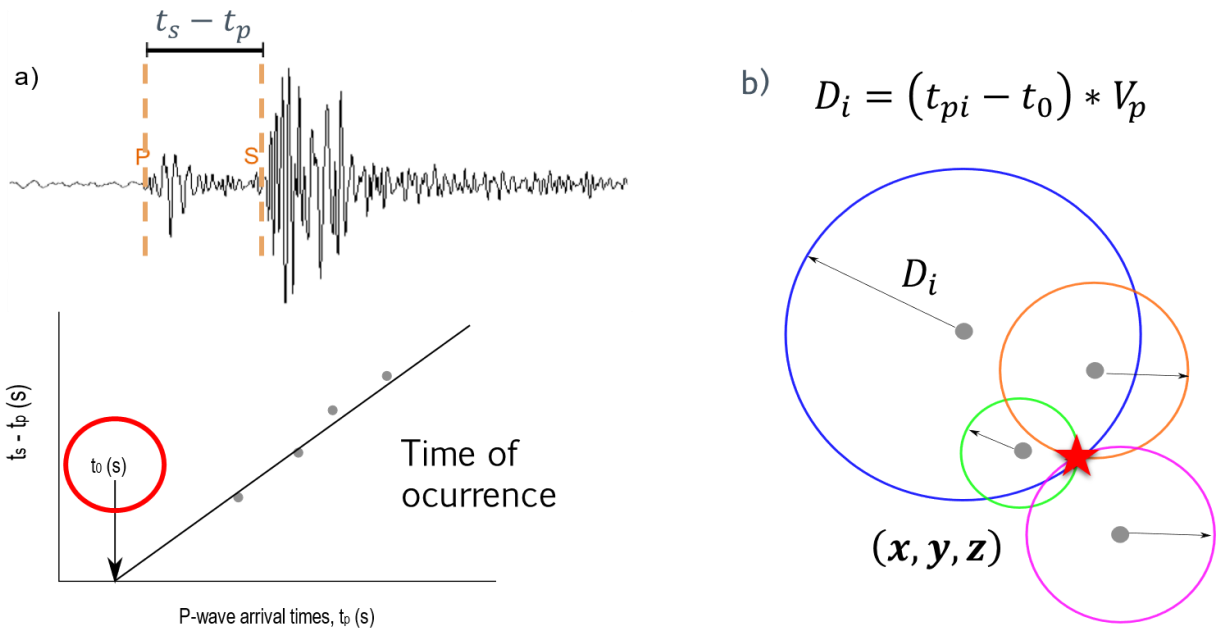


Figure 4. Method of circles. A) Wadati diagram for t_0 determination, b) hypocenter location.

The main disadvantage of these time-based methods is, indeed, inaccurate travel time picking under low SNR conditions, where uncertainty in phase arrivals may degrade the reliability of the location (Wang, Li, & Shang, 2016). Since technology evolves at a fast pace, many techniques and procedures for locating seismic sources have emerged. Sometimes, vast arrays of data requires to be processed in a short time, making manual picking impractical thus automatic picking algorithms are implemented to overcome this challenge; however, manual picking is still used as a means to verify and improve the accuracy of automatic pickers (Rentsch, Buske, Lüth, & Shapiro, 2007, p. S33).

Derivative approach. The generally preferred way to locate earthquakes consists in an iterative least-squares technique attributed to (Geiger, 1912). Those methods use information of the derivatives to approximate the nonlinear behavior of the problem and update a trial solution iteratively. The Geiger's method (1910) is a local inversion algorithm which uses the first 2 terms of Taylor series to approximate the non-linear problem of hypocenter location (**Eq.1 & 2**).

$$\boldsymbol{d} = \boldsymbol{G}(\boldsymbol{m}),$$

Where: $\boldsymbol{m} = (x, y, z, t)$ $(Eq.\,1)$

$\boldsymbol{d} = Measured\ arrival\ times$

$$\Delta \boldsymbol{m} = (\boldsymbol{G^T G})^{-1} \boldsymbol{G^T} \Delta \boldsymbol{d},$$

(*Eq.* 2)

$(\boldsymbol{G^T G})^{-1} \boldsymbol{G^T}$: *Generalized inverse*.

The algorithm begins assuming an initial model (**figure 5**) comprising 4 unknown parameters: spatial coordinates of the hypocenter (x, y, z) and origin time (t_0). Then, theoretical travel times are computed using this starting point; finally, an iterative process is carried out to minimize the residuals between the observed and predicted data at each recording station until a preset tolerance is met (Lay & Wallace, 1995, pp. 221-223) (Stein & Wysession, 2003, pp. 416-424).

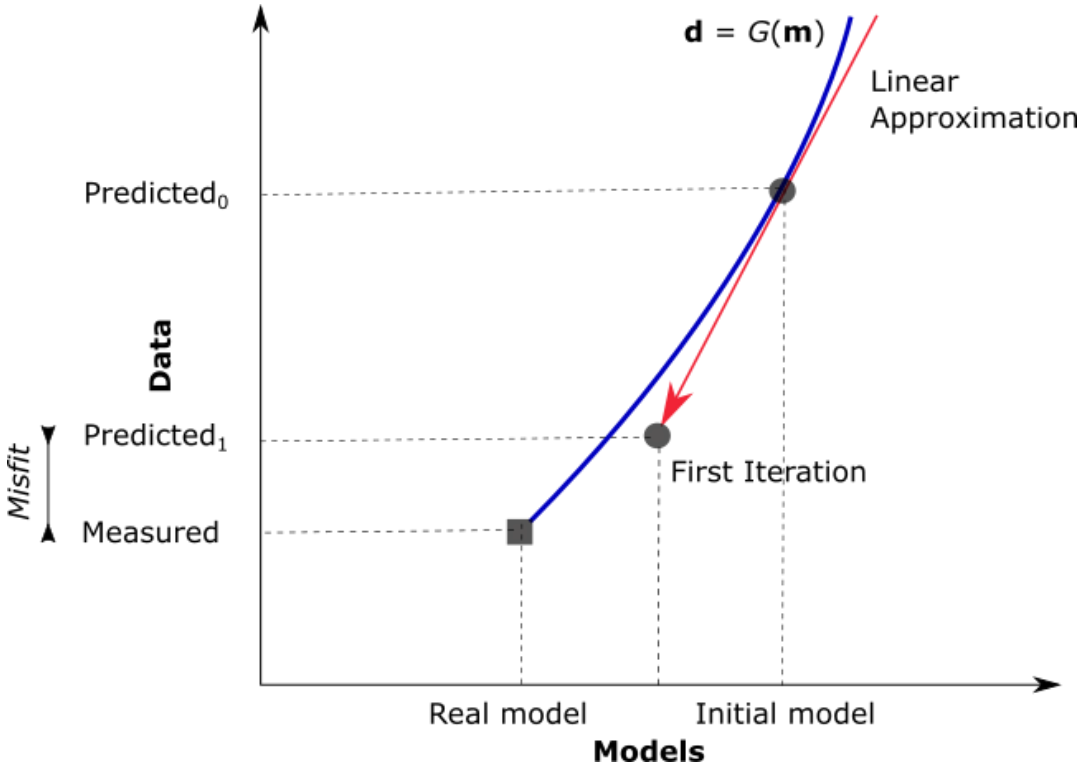


Figure 5. Linearization of a complex problem and its subsequent inherent error (misfit).

Unfortunately, the least-squares solution proposed by Geiger is ill-conditioned and may suffer divergence (Ge, 2003). Consequently, numbers of researchers have implemented modifications to the method to overcome such difficulties. Recent proposals can solve for hypocenter location, magnitude, first motion, source parameters and others. Some of the most popular methods derived from Geiger are ***Hypo71*** (Lee & Lahr, 1972), originally designed for large numbers of local earthquakes, this algorithm uses a step-wise multiple regression which converges more rapidly to a solution and eliminates the ill-conditioned matrix; ***Hypoellipse*** (Lahr J. , 1979) (Lahr J. C., 1999) adapted for local and regional earthquakes; ***Hypoinverse-2000*** (Klein F. W., 1978) (Klein F. W.,

2002); ***Hyposat*** (Schweitzer, 2001), and a more recent software ***Seisan*** (Ottemoller, Voss, & Havskov, 2017) includes the main features of the previous programs, besides, location of teleseismic events, addition of more seismic phases, and other improvements.

"Linearized solutions can be a poor representation of the complete solution, and it may be unstable when the complete solution is irregular or has multiple solutions" (Husen & Hardebeck, 2011). In other word, linearized methods (i.e. derived from Geiger) depend on the choice of a starting point and, therefore, can be stuck in a local minimum or be unstable for poorly constrained earthquake locations. Additionally, there are 3 main sources of uncertainty: first, *measurement errors of seismic arrival times*, typically related to SNR and dominant frequency of the signal. These are formal errors and can be represented by error ellipsoids. Second, modelling error of calculated travel times, determined by the quality of the seismic velocity model and method used to calculate the travel times. Proper handling of these errors is difficult. Third, Nonlinearity of the earthquake problem, this one is only relevant for linearized, least-square solutions. In conclusion, different linearized location algorithms use different statistics to compute location uncertainties and some of them may yield misleading results. Therefore, care should be taken when choosing earthquake location methods, as they do not represent the exact nature of the phenomena and each of them is affected by random and systematic errors. **Note:** Since picking and phase identification are extremely difficult or impossible in LPLD events and tremor-like signals, these methods are not considered in this thesis.

1.3.2. Full Waveform-based methods. A series of methods that utilize the waveform information have been developed to solve for earthquake locations in presence of noisy data. What makes these methods attractive is the fact that they follow the kinematic and dynamic characteristics of wave propagation, as opposed to time-based methods, that only use kinematic information. According to (Kao & Shan, 2004, p. 589): "the complete source configuration can be recovered from the constructive and/or destructive interference observed at stations along different azimuths/distances".

Although these methods produce high resolution results, their efficiency can be compromised and therefore, they may require the use of auxiliary algorithms to compute travel time-tables for each station, eliminating the repetitive calculations. Such time-tables can be determined using different techniques as suggested by (Baker, Granat, & Clayton, 2005, p. 700); in this book we use

the Fast-Marching Method (FMM) to obtain such travel time tables because this technique has been proved to yield accurate estimates (see **Appendix B)**.

1.4. Advances in seismic and microseismic event location algorithms.

Nowadays, spatiotemporal seismic and microseismic location methods have undergone several modifications to estimate event sources for real time monitoring applications with lower uncertainty; moreover, automatic and semiautomatic picking algorithms are being utilized to reduce the time it takes to process the increasing amounts of continuous seismic data. All these technological advances have improved our understanding of the seismic phenomena and deformation mechanisms present in the earth. In this book, however, no automatic picker is used since LPLD events imply very low SNR conditions, yet a brief description about automatic picking algorithms can be found in (Wang, Li, & Shang, 2016).

Techniques like kinematic ray tracing, aiming to diminish the residuals between observed and theoretical arrival times for different seismic phases (mainly P and S), used to be the preferred way to locate earthquakes, nonetheless, since technology and computing power have evolved, other possibilities have risen including: more robust ray tracers, finite difference, finite element and many other grid-based schemes. Hence, forming a vast variety of tools and approaches to accomplish the task.

The diagram in **figure 6** schematically summarizes the mainly used methods for seismic event location. Two big families can be seen: the first one (left branch) is dependent on arrival time picking, and the second one (right branch) uses full waveform information and is based on the principle of backpropagation.

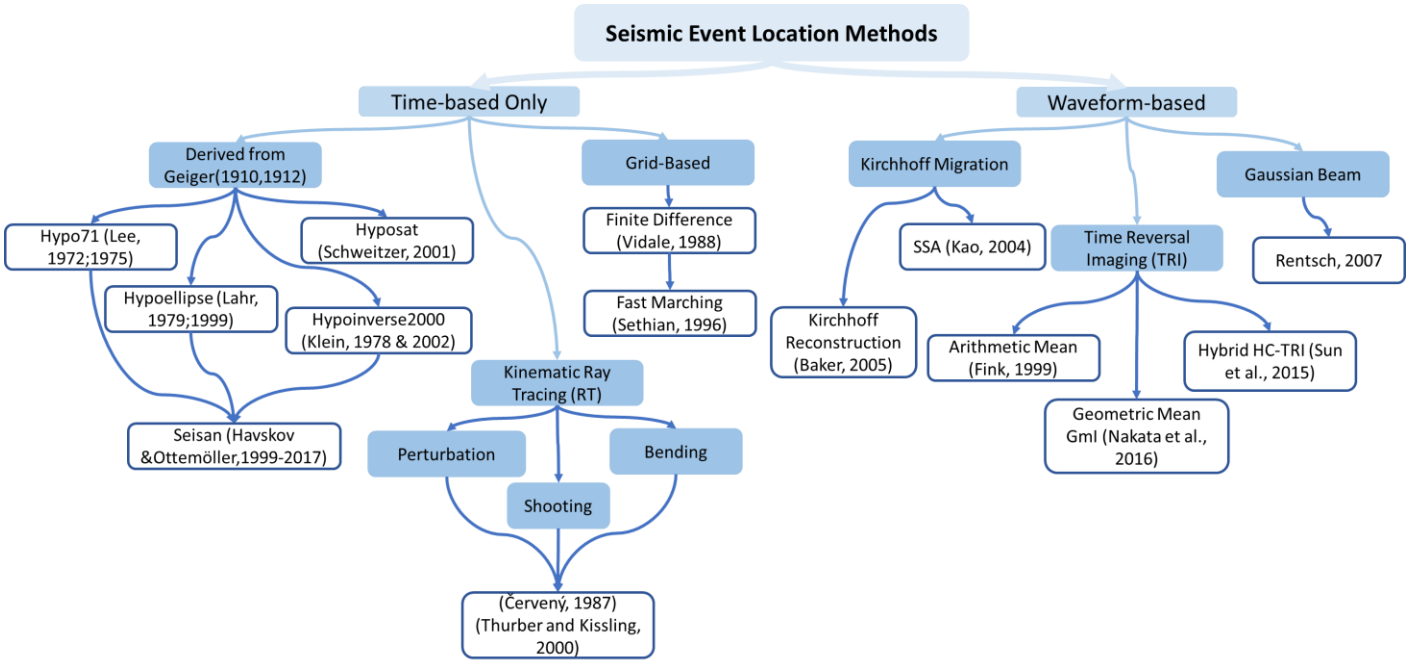


Figure 6. Common seismic hypocenter methods. (Left branch) Only kinematic information. (Right branch) Both kinematic and dynamic information.

A detailed review of some of the most relevant methods is presented to analyze their strengths and weaknesses for earthquake hypocenter location under very low Signal to Noise Ratio as found in LPLD events. Additionally, their main ideas are summarized in **table 1.**

1.4.1. Travel time-based methods. After the introduction of Finite Difference (FD) schemes to approximate the 2D and 3D Eikonal equation for travel time computation (Vidale J. , 1988) (Vidale J. E., 1990), various schemes were developed seeking to decrease the computational cost of FD; one that stood out was the Fast-Marching Method (FMM) introduced by (Sethian J. A., 1996) which solves the Eikonal equation accurately and faster than most others FD schemes. Afterwards, (Sethian & Popovici, 1999) used the FMM for the first time on a geophysical application when computing first arrival travel times on a 3D orthogonal coordinate system for a model with sharp changes in the velocity gradient. Some remarkable and worth mentioning characteristics of the FMM are its efficiency, precision and robustness which, altogether, make it a powerful tool to improve more complex methods, as seen in posterior chapters.

Generally, two different approaches can be taken to solve for the travel time field: one using differential equations to solve the ray tracing problem in 2D and 3D producing “exact” solutions, but unfortunately, experiencing difficulties when dealing with complex velocity models (e.g.

shadow zones). On the other hand, grid-based methods approximating the solution and being able to handle high velocity gradients; nonetheless, acceptable accuracy may require a dense grid, which in turn increases computation time exponentially (Thurber & Kissling, 2000, p. 85).

Note: due to the high sensitivity ray tracers (RT) have to erroneous phase picking on seismograms and the special characteristics of LPLD events, they are not considered in this book as an appropriate option for precise location; however, a FD method, more specifically FMM, is used to compute travel time tables and improve the overall efficiency of the main backpropagation algorithm.

1.4.2. Full waveform-based methods. Besides RT and Grid-based algorithms, a more robust family of techniques that utilizes full waveform information for hypocenter location has emerged and gained supporters in the last decades. This family is based on the concept of "*Time-Reversal Imaging (TRI)*" proposed in (Fink, 1999) where an acoustic wave field can be reversed in time propagating through a given model and thus, allowing to determine its initial conditions or source properties. Although this approach is accurate, and commonly used in active seismic applications nowadays, it is computationally expensive because it implies solving the full wave equation, scalar or vector depending on the circumstances, as well as defining imaging conditions for the forward and back propagated wave fields, moreover, one of the main goals of this book is to provide a perhaps less accurate but faster and handful algorithm; therefore, the TRI technique is not selected in this book.

(Kao & Shan, 2004) proposed the ***Source Scanning Algorithm (SSA***) (see **Appendix C.2**) for determining spatiotemporal seismic source distributions based on a brightness function built by stacking the normalized amplitudes from each station according to predicted travel times for trial solutions (similar in concept to Kirchhoff migration used in active seismic for imaging purposes). The region where the brightness function reaches its maximum value, or perhaps a few peaks, is considered as the location(s) of one or more candidates, depending on the number of events occurring simultaneously. Some remarkable features of this algorithm are: 1. There is neither need for phase picking nor assuming source geometries, 2. The brightness function is not computed for the whole temporal-spatial domain but for target sub-volumes and time windows where a particular event is identified, and 3. It can solve sources using information either of P or S waves, however, S-waveforms are typically preferred due to their higher amplitudes which makes them easier to recognize over the background noise.

After (Kao & Shan, 2004), some research has been done following the same general idea of amplitude stacking but with some substantial differences; among these studies, it is worth mentioning:

(Baker, Granat, & Clayton, 2005) Developed a method for real time location of earthquakes based on a Kirchhoff reconstruction derived from the one used in migration of seismic data. Along the lines of SSA and other similar techniques, Kirchhoff Reconstruction Location (KRL) does not require first arrival picking, instead, it backpropagates seismogram amplitudes from each station towards the volume of study where the source is identified through global searching. One noticeable difference with respect to SSA is that KRL is designed for real time applications using P phase information only, since S and converted waves travel at a lower velocity.

(Rentsch, Buske, Lüth, & Shapiro, 2007) Proposed a location method for multicomponent seismic (3C) data that combines initial-value ray tracing with Gaussian-beam migration. This method consists in propagating the energy recorded on each sensor along beams, then, the candidate source is determined by the region of maximum focused energy. Since this technique requires P-wave identification, there is an implicit sensitivity to incorrect picking; however, such an uncertainty is taken into consideration by adding Gaussian distributions to the beam widths.

In (Kao & Shan, 2007) the SSA method is again presented with a few modifications to image the rupture plane created during the stress release of a typical earthquake. For this purpose, the brightness function is constructed from the normalized envelops of the P- phase only, from stations distributed at different epicentral distances. This application of the SSA is an alternative to conventional nodal plane analysis which often leads to ambiguity and requires extra information that may not always be available. Using SSA seems advantageous because it does not need any information of the source configuration to produce an image of the rupture process; unfortunately, it is strongly affected by the geometry and coverage of sensors, besides its poor resolution because it utilizes only P-phase information.

(Liao, Kao, Rosenberger, Hsu, & Huang, 2012) Introduced an Improved Source-Scanning Algorithm (ISSA) derived from the SSA to describe the spatiotemporal distribution of aftershocks. Some of their improvements are: the total absence of phase picking, use of both P and S-wave information to discriminate true events from artefacts and automatic adjustment of time windows depending upon the scanning resolution. As drawbacks and matter of future work are: ISSA cannot determine the magnitude of individual events due to the assumptions made to construct the

brightness function, and there is a lack of uncertainty analysis accounting for the effects of an incorrect velocity model, array of receivers and other factors.

(Anikiev, Valenta, Staněk, & Eisner, 2014) utilized a backpropagation imaging method, where the *image function* (their energy function) is converted into a Gaussian probability density function centered around the point of maximum energy. In addition, the introduction of a polarity correction factor permits the constructive or destructive interference of signals. This new perspective allowed to calculate the spatial coordinates on a centroid-based approach rather than a grid search one.

As summarized in **figure 6**, there is a branch of methods that employs full waveform information and relies on the concept of Reverse Time Migration or Imaging (RTM) attaining the highest degree of accuracy among the above-mentioned methods, but extremely computationally expensive. A remarkable contribution in this area was presented in (Gajewski, Anikiev, Kashtan, Tessmer, & Vanelle, 2007) where a 2D seismic event location method based on diffraction stacking of the squared amplitudes at each receiver was presented. Some of the assumptions made in this study are: 1) Source point approximation, 2) acoustic wave only and, 3) designed for low magnitude events. One troublesome aspect is the use of a homogeneous velocity model, which makes it partially unrealistic. However, it is still a good starting point to gain insight into the phenomena. A subsequent modification was presented in (Zhebel, Gajewski, & Vanelle, 2011) where the method was extended to 3D in both homogeneous and heterogeneous media and then, tested on field data proving the potential of the method.

In the past few years, most effort has been focused on RTM based techniques due to their accuracy; some relevant studies are: extended Time-Reversal Imaging (TRI) to source location and characterization using both P and S waves (Artman, Podladtchikov, & Witten, 2010); use of the extended Common Image Point Gathers (CIP's) as an efficient imaging condition for Reverse Time Migration (RTM) (Sava & Vasconcelos, 2011); a new imaging condition called Geometric-mean Imaging (GmI) to increase spatial resolution, unfortunately, an increase in computational cost as well (Nakata, Beroza, Sun, & Fomel, 2016); and reformulating passive source imaging as a Full Waveform Inversion (FWI) problem using *Least-Squares Time-Reversal Imaging (LSTRI)* and allowing to jointly invert seismic source information and velocity model (Sun, Xue, Fomel, Zhu, & Nakata , 2016). As mentioned earlier however, TRI (derived from RTM) is more computationally expensive than BPI (similar to Kirchhoff migration). A ratio of 15:1 in running time was obtained when migrating a salt model by RTM and Kirchhoff (Yoon, Shin, Suh, Lines,

& Hong, 2003); the difference in time is mainly because the former computes numerical solutions to the complete wave equation while the latter uses a high-frequency ray approximation. Other studies reported a ratio of 8:1 (RTM – Kirchhoff) (Zhu & Lines, 1998). Two more papers showing the higher computational cost of RTM are (Sihna, Vishnoi, Basu, & Singh, 2009) and (Jiang, Bonham, Bancroft, & Lines, 2010). Consequently, RTM (TRI) methods are impractical for LPLD location applications where seismograms of several hundreds of seconds need to be analyzed quickly.

Table 1. Summarizes fundamental characteristics of some of the most relevant methods described earlier with the intention of serving as a guide and motivation for the reader, perhaps new in this field of seismology, who looks forward to beginning his research career.

Table 1. Miscellaneous of seismic source location methods.

Branch	Name	Approach	Advantages	Disadvantages	Class	Reference
Travel Time Based[1]	*Geiger's Method*	Local inversion algorithm which uses truncated Taylor Series to solve a non-linear problem $F(x, y, z, t)$.	Fast and straightforward.	It requires an initial model close to the real one.	Theoretical / Analytical	(Lay & Wallace, 1995, pp. 221-223)
	Method of circles (2D) or Spheres (3D)	Plotting $(t_s - t_p)$ vs t_p at different stations, and drawing circumferences centered on each station, thus, localizing the hypocenter inside an intersection region.	Fast and straightforward. Very intuitive.	It is designed for homogeneous models.	Theoretical / Analytical	(Lowrie, 2007, pp. 149-150)

[1] These methods have problems with erroneous picking and phase association at different recording stations.

Branch	Name	Approach	Advantages	Disadvantages	Class	Reference
	Ray Tracers	Solving ray equations to find the ray path between two points, with the minimum time residual.	Useful and accurate for non-complex velocity models. Computationally Affordable.	It experiences difficulties when dealing with complex velocity models (Shadow zones, incorrect paths). Rays do not fully reproduce the wave propagation phenomena.	Ray tracing algorithms: - Bending - Shooting - Perturbation	(Rawlinson, Hauser, & Sambridge, 2008) (Červený, 1987)
	Vidale, FMM (Fast Marching Method)	Use of finite difference or others grid-based schemes for travel time computation.	Determination of the global minimum. Efficient and acceptable accuracy.	Significant trade-off between accuracy and computation speed (High resolution scenarios).	Grid-Based	(Vidale J. , 1988) (Sethian J. A., 1996)

Branch	Name		Approach	Advantages	Disadvantages	Class	Reference
Waveform Based[2]	Backprojection Imaging	*SSA (Source Scanning Algorithm)*	Construction of a brightness function by stacking normalized amplitudes from each station.	There is no need for phase picking nor associations at different stations.	Uses information of only a single phase (Typically S-phase).	Kirchhoff Migration	(Kao & Shan, 2004)
		ISSA (Improved Source Scanning Algorithm)		Incorporates both P and S- phases information into the brightness function. Reduces the number of artefacts.	Lack of uncertainty analysis accounting for: - Incorrect velocity model. - Acquisition geometry. - Variable SNR		(Liao, Kao, Rosenberger, Hsu, & Huang, 2012)
		Location of seismic events by diffraction stacking	Consists on stacking the squared amplitudes from each receiver to	Straightforward to implement.	Developed for 2D cases. It uses homogeneous velocity models.		(Gajewski, Anikiev, Kashtan, Tessmer, & Vanelle, 2007)

[2] These methods require few to none previous picking of arrivals.

Branch	Name		Approach	Advantages	Disadvantages	Class	Reference
		Location of seismic events in 3D media by diffraction stacking	determine the source location.	Application for 3D cases in presence of heterogeneous media and tested on field data.	It is still affected by complexly heterogeneous media.		(Zhebel, Gajewski, & Vanelle, 2011)
		KRL (Kirchhoff Reconstruction Location)	Similar to Kirchhoff migration used in active seismic data, but with some modifications for passive source studies.	KRL utilizes the P-wave phase since it travels faster than others do. Hence, suitable for real time applications.	Uses only P-phase information, thus, producing several artefacts and ambiguity in the images.		(Baker, Granat, & Clayton, 2005)
		Joint location and source mechanism inversion of microseismic events.	Converts the image function into a PDF to compute the spatial centroid.	Source location is not restricted to grid-nodes.	Restricts the PDF to a Gaussian distribution.		(Anikiev, Valenta, Staněk, & Eisner, 2014)

Branch	Name		Approach	Advantages	Disadvantages	Class	Reference
		Fast location of seismicity: A migration-type approach	Stacking beams from each sensor and identifying candidates where peaks of energy are spotted.	Multicomponent (3C) seismograms. Beams use a Gaussian distribution to accounts for some uncertainties in picking.	It is susceptible to incorrect P-wave identification during the picking procedure (picking dependent).	Gaussian Beam Migration	(Rentsch, Buske, Lüth, & Shapiro, 2007)
	Full Wave Equation[3]	*Time-Reversal Imaging (TRI)*	A source can be recovered by reversing a wavefield in time, making use of the superposition principle.	Accuracy and novelty. It handles complex waveforms and low SNR.	Acoustic wave only. Computational expense.	Time Reversal Imaging	(Fink, 1999)
		Extended TRI	Extension of TRI to account for other phases.	- Solution to the Elastic Wave equation. -It handles complex waveforms and low SNR.	Increased computational load as more phases are involved.	Time Reversal Imaging and Imaging Condition	(Artman, Podladtchikov, & Witten, 2010)

[3] Solving the full wave equation can be highly computationally expensive

Branch	Name		Approach	Advantages	Disadvantages	Class	Reference
		Extended CIP's	Imaging involves two steps: wave-field reconstruction and imaging condition. Proposed the Extended Common Image Point Gathers (CIP's)	-Solves complex waveforms and low SNR. -Improvement to the RTM. -Enables accurate velocity analysis.	Not suitable for real time applications.		(Sava & Vasconcelos, 2011)
		HC-TCI	A series of improvements to the imaging conditions used in RTM	-It solves complex waveforms and low SNR. -High spatial resolution.	Not suitable for LPLD event location in real time applications.		(Nakata & Beroza, 2016a) (Sun, Zhu, Fomel, & Song, 2015)
		Full Waveform Inversion		-High resolution.	- Highly computationally expensive.		

1.5. **Main Ideas**

Numerous algorithms have been developed in seismology to accurately locate earthquakes in space and time, most of which have been adapted to active and hybrid seismic sources, as well as microseismic monitoring.

The most commonly used location methods require picking and association of the seismic phases at different recording stations to localize the event source. These kinds of methods aim to minimize the residuals between predicted and measured times until a certain tolerance is met. However, there are some phenomena which generate complex seismograms making picking impractical; in those cases, it is preferred to use methods that employ both the kinematic (time and depth) and dynamic (amplitude and phase) wave information to gain more insight into the subsurface structures and consequently, reducing the uncertainty of the outcome.

BackProjection Imaging (BPI) is a simplified version of Time-Reversal Imaging (TRI) in which the full wave equation is not computed but less computationally demanding equations like Eikonal. BPI makes use of those auxiliary equations to calculate travel-time tables (isochrons) along which amplitude stacking from the n-number of stations takes place; for this reason, BPI arises as a more computationally efficient though less accurate alternative to TRI. Despite the resolution loss BPI may experience, it is less susceptible to velocity and other sources of uncertainty; feature that makes it a qualified method for location of seismic events. Nonetheless, each one is suitable under different circumstances thus the choice must be made based on the type of seismic signal under study, the computing power available, reliability of the input data, required resolution, personnel's expertise and many other factors.

2. Methodology

2.1. **Overview.**

The design of a seismic event location algorithm that is suitable for applications to conventional earthquakes, microseismicity and tremor-like signals is explained step by step, describing the resources (software) used during the process. Additionally, we propose a simple methodology for the quick generation of synthetic seismograms that consists of the following steps: first, construction of the 3D structural model, surface sensors acquisition geometry and model properties

(P and/or S velocities) derived from well logs and other geophysical information available for the selected oil field. Then, computation of the response of the media from an impulsive source, deep in the reservoir, to a set of sensors placed on surface; finally, convolution is applied between a representative tremor-like signal and the previously computed impulse responses to generate synthetic seismograms to be backpropagated using the selected algorithm.

2.2. **Introduction**

Location of seismic events has always been a major goal in seismology, for this reason, several techniques have been developed to produce results with the highest possible accuracy. Earthquake and other types of seismic events have been conventionally located using techniques that require picking of arrivals and association of phases; however, there are many other seismic events in which energy behaves in extremely complicated ways that no clear phases can be distinguished thus revealing the need for different approaches to handle a wider variety of events.

Back-Projection Imaging (BPI) has recently become a powerful tool for location and detection of seismic signals due to its ability to use the dynamic and kinematic information of the waveform to recover the seismic source without solving the full wave equation as Reverse-Time Migration (RTM) does. In other words, RTM is highly computationally expensive and requires both accurate velocity models and densely recorded data. Fortunately, "RTM can be simplified in the form of BPI by ignoring amplitude changes due to wave propagation effects, as well as polarity changes due to earthquake radiation pattern" (Beskardes, et al., 2018, p. 1987) thus becoming a robust technique which requires neither phase picking nor extensive wavefield computations but rather accurate travel-time computations to localize the seismic source. Therefore, the here-proposed location algorithm is based on the concept of Backpropagation/Backprojection which explains how a set of receivers can be treated as point sources, to re-emit a previously recorded signal, in order to reproduce the behavior of a particular wavefield and determine the actual source from which the wave had formerly originated (Fink, 1999).

Although the foundations of the algorithm have been borrowed from previous studies (Kao & Shan, 2004) (Gajewski, Anikiev, Kashtan, Tessmer, & Vanelle, 2007), significant changes have been introduced by incorporating techniques and observations other authors have made thus producing a somewhat different approach. **Figure 7** summarizes the methodology followed in this thesis to create a reliable location and perhaps detection algorithm, from the construction of the geophysical model to the determination of candidate sources.

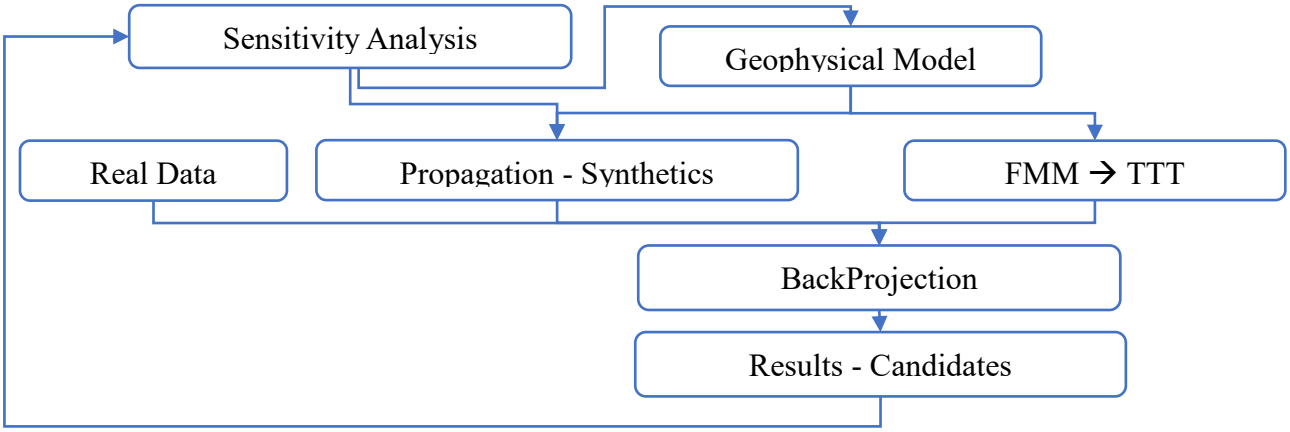


Figure 7. General workflow for the design and assessment of the location method.

Note: All the algorithms used in this thesis are written in Python, MATLAB, SCONSTRUCT structure and the Madagascar library which comprises numbers of geophysical scripts also written in different languages including c, c++, fortran, python, etc., becoming a powerful tool for academic and research purposes.

2.3. **Earth model construction (Geophysical model)**

A realistic earth model is essential in any geophysical study, because it reproduces the structures, heterogeneities and anomalies of the subsurface. Thus, the closer this replica resembles reality, the more accurate results can be obtained.

Here, a program called ***sfunif3*** from the *MADAGASCAR* package[4] was used to build the 3D model. *Sfunif3* enables to create layered velocity models with interfaces defined by algebraic equations such as horizontal and dipping layers (**figure 8**); besides, more complex structures can also be reconstructed by splines interpolation. Note: other parameters including the x, y and z dimensions, spatial discretization, and velocity values are also set at this point.

4 Madagascar available at http://www.ahay.org/wiki/Main_Page

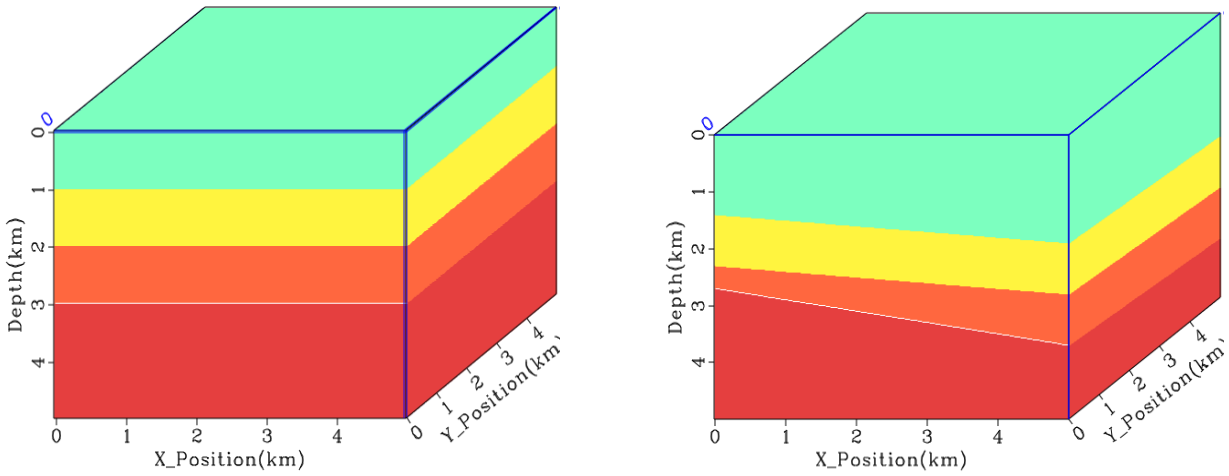


Figure 8. Earth model built by sfunif3. Left) Horizontal layers. Right) Dipping layers.

The outcome of this step is an rsf-file with information of velocities and interfaces to be fed to the back-propagation algorithm later on. In case a different program is used, the generated sgy-files can be converted into rsf-files through the command ***sfsegyread***.

2.4. **Fast Marching Method (FMM)**

Computing travel time tables (*TTT*) before backpropagating the seismograms, is an efficient way to reduce computational cost and time during the process. Several finite difference schemes can be used to compute travel time fields from an initial velocity model. In this work, the fast-marching method is chosen due to its ability to calculate fast and accurate results by solving the Eikonal equation.

Once the velocity model is available, it is fed into a Madagascar program called ***sfeikonal*** in which travel time tables are calculated for each receiver location (**figure 9**) specified in a file called "*shots.rsf"*. Note: Another program called ***sfeikonalivt*** can be used in case VTI anisotropy is desired.

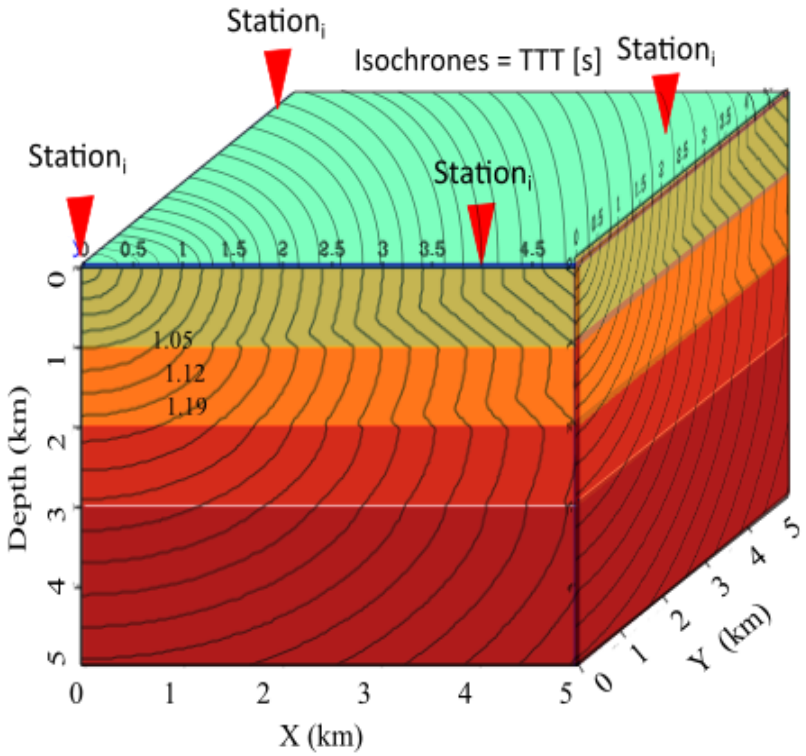


Figure 9. Travel time table computed by fast-marching method for station at (0,0,0). Each station must have its own TTT prior to backpropagation.

Readers are encouraged to have a look at **appendix B** for a better explanation of the fast-marching method.

2.5. **Generation of Synthetic Data**

Synthetic seismograms are powerful tools to study the wave propagation phenomena under controlled conditions, enabling one to make sensitivity analysis on a set of independent variables and their impact on dependent variables. In this study, two types of synthetic traces are generated to study the capability of the algorithm to handle both impulsive and emergent signals:

2.5.1. **Gaussian envelopes which resemble impulsive signals.** These kinds of signals are the products of bandpass filtering and calculating absolute values of unity spikes placed at times equal to the travel times between the source and their corresponding receiver (**figure 10**).

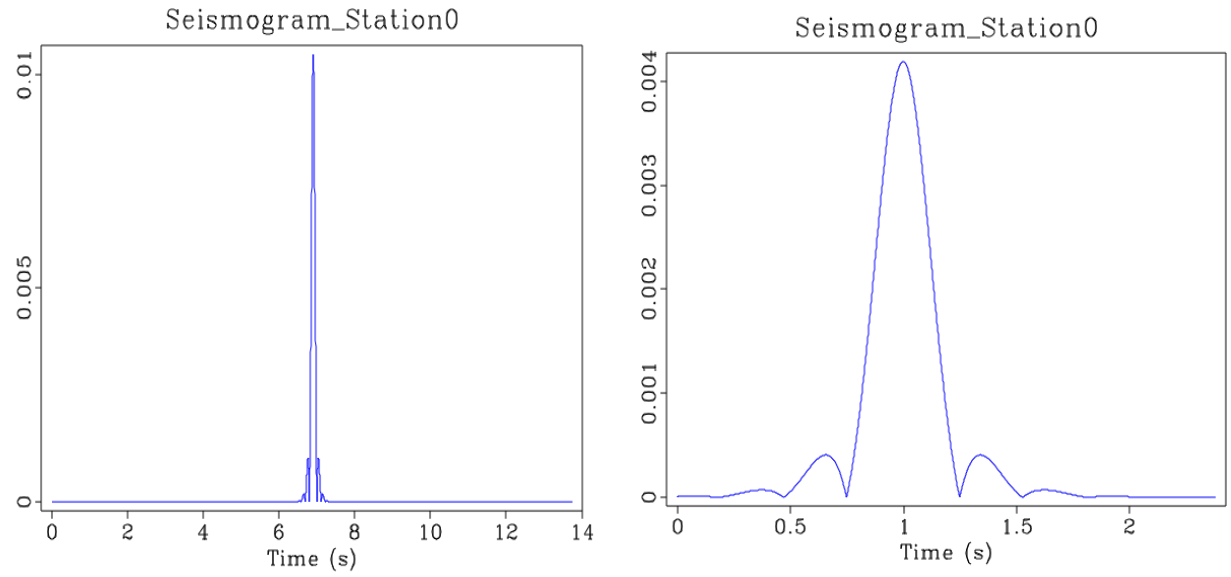


Figure 10. Noise-free Impulsive signals after using a low-pass filter and calculating the absolute values of the traces. Left: 5Hz. Right: 2Hz.

Among the advantages of using such simple signals is that they are handful for sensitivity analysis, because they allow to individually change certain parameters and evaluate their direct impact on the solution. Besides, more complicated signals can be interpreted as composed of several impulsive signals, therefore, understanding simple cases help to infer information in complex circumstances.

2.5.2. Realistic tremor-like signals. First, it is important to mention that the generation of synthetic tremor-signals itself is not the main purpose of the thesis but using them to evaluate the location algorithms. Therefore, we present a quick methodology for generation of synthetics that avoids detailing the nature and rupture process of *LPLD* seismic events (i.e. a type of tremor-like signal), because covering this open research thoroughly might cover the time of a PhD thesis on its own.

Consequently, a simple methodology (see **figure 11**) similar to using empirical Green's functions is implemented to produce synthetic seismograms to be backpropagated using the chosen location algorithm.

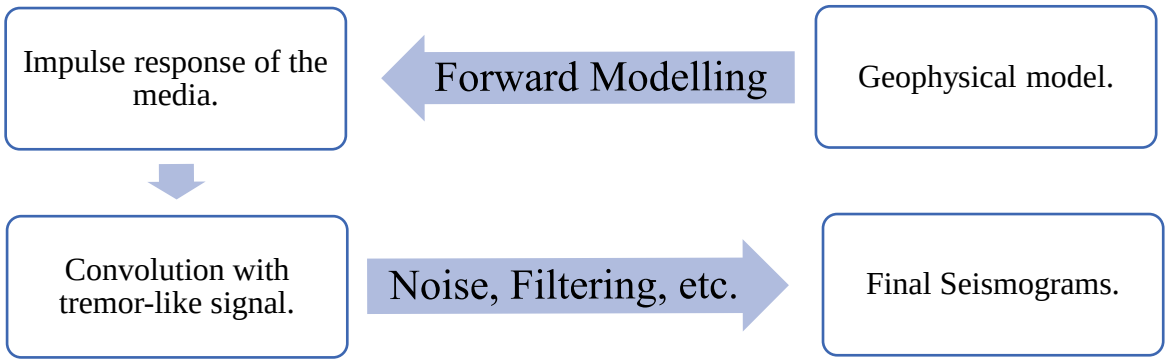


Figure 11. Proposed workflow for synthetic seismogram generation.

Impulse response of the media. Once the geophysical model is created (see section 2.3), it is possible to compute the response of the media employing different methods including ray tracers and finite differences. In this case, unity spikes were generated using the program sfspike according to the previously computed first arrival travel times from source(s) to the n receivers. Figure 12 is an example of such type of signals.

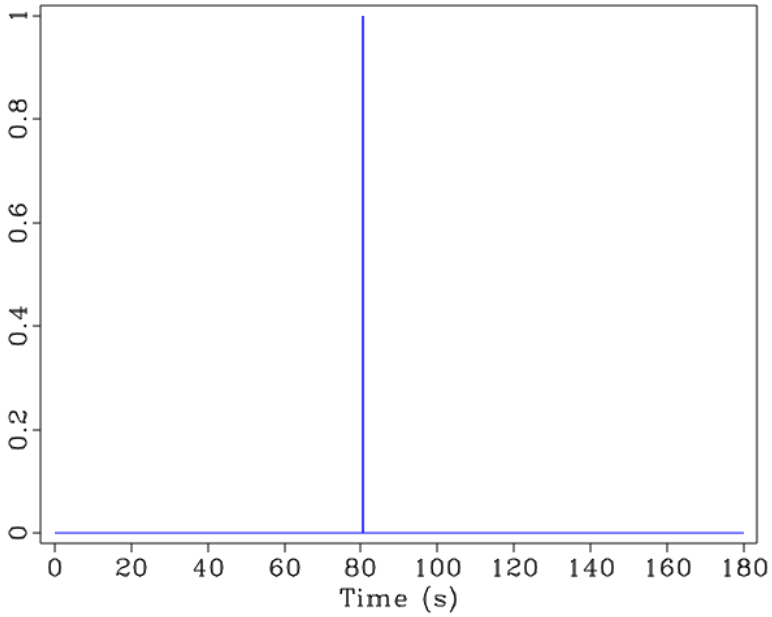


Figure 12. Spike at arrival time from source to station.

Convolution with tremor-like waveform. Convolution is applied between the previously calculated impulse responses and LPLD signals extracted from real data. After filtering, noise addition and other techniques, the seismograms are ready to be backprojected.

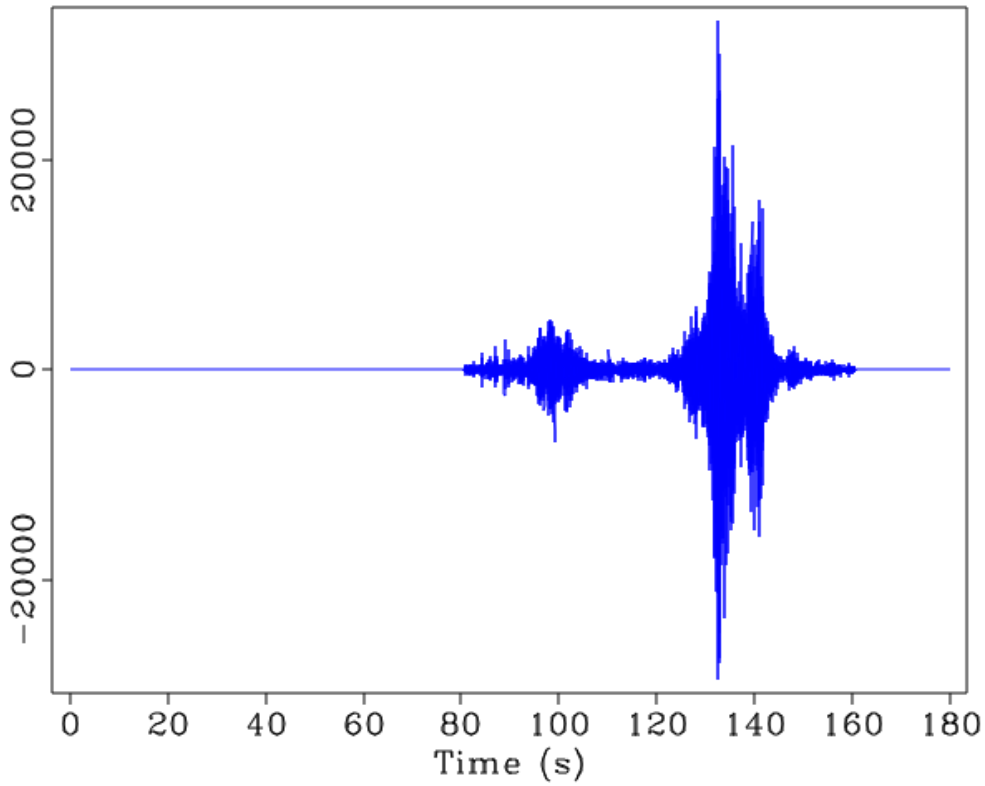


Figure 13. Synthetic LPLD seismic event obtained through convolution.

2.6. **Designing an LPLD Seismic Event Location Algorithm Using the BPI method.**

At this point, all the inputs inside the red box in **figure 14** are ready to be fed to the Back-Projection algorithm. Now it is time to proceed with **Part B – figure 14**. **Back-Propagation** is commonly used in techniques like RTM in which the full wave equation is solved by finite difference schemes to account for several seismic phases and wave propagation phenomena; nonetheless, these kinds of techniques are computationally expensive. A simplified approach called BPI allows to reconstruct the source configuration without the need to compute wavefields but travel times instead, then, these travel-time tables are used together with seismograms to locate the source in a less accurate although faster way. The main steps followed in the location algorithm are presented in **table 2**.

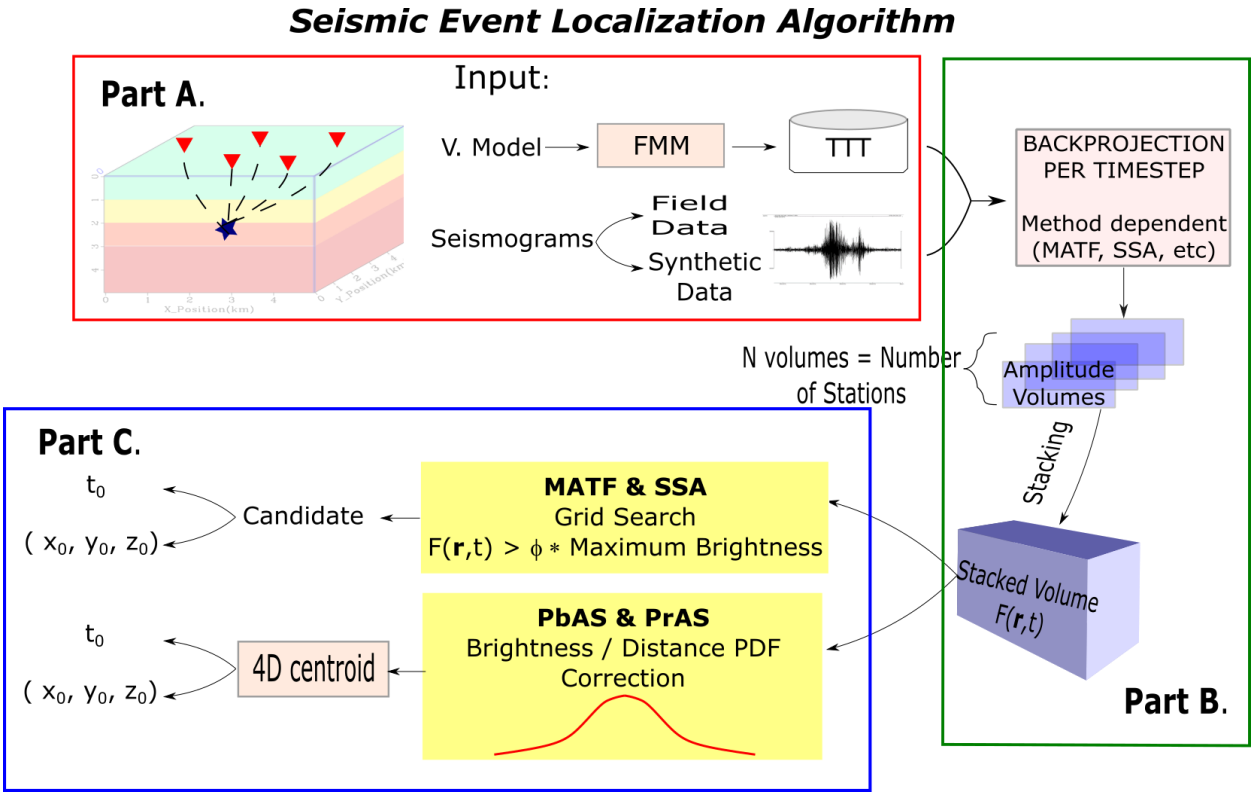


Figure 14. General workflow of the location method. Part A: Input data. Part B: BackProjection approach. Part C: Location Method (Results).

Table 2. Location algorithms: Pseudocode.

- Input (**Part A**):
  - Pre-Processing (frequency analysis, filtering, deconvolution, stacking, etc.)
  - Travel-time tables.
- **Backpropagation (Part B)**
  - Seismogram normalization & Absolute values.
  - $F(\boldsymbol{r}, t)$ → *Method-Dependent* → *(MATF, SSA, etc.).*

    For each time step t

    For each station

    Interpolation and stacking of amplitudes

    Pre-candidates for each time step (Compared to a given threshold)
- Final candidates Results (**Part C**) → Method dependent → (Maximum, Centroid, etc.).

This process starts by feeding a velocity model to a fast-marching program from the Madagascar library, which solves the Eikonal equation to calculate travel-time tables (***TTT***) from each receiver (acting as new sources) to every grid-point. Since these tables are computed only once for a given velocity model, they can be stored on disk to eliminate repetitive computations and thus, increasing the algorithm efficiency. Later, a spline interpolation is used to assign amplitudes from each seismogram to each grid point according to their travel-times (**figure 15**).

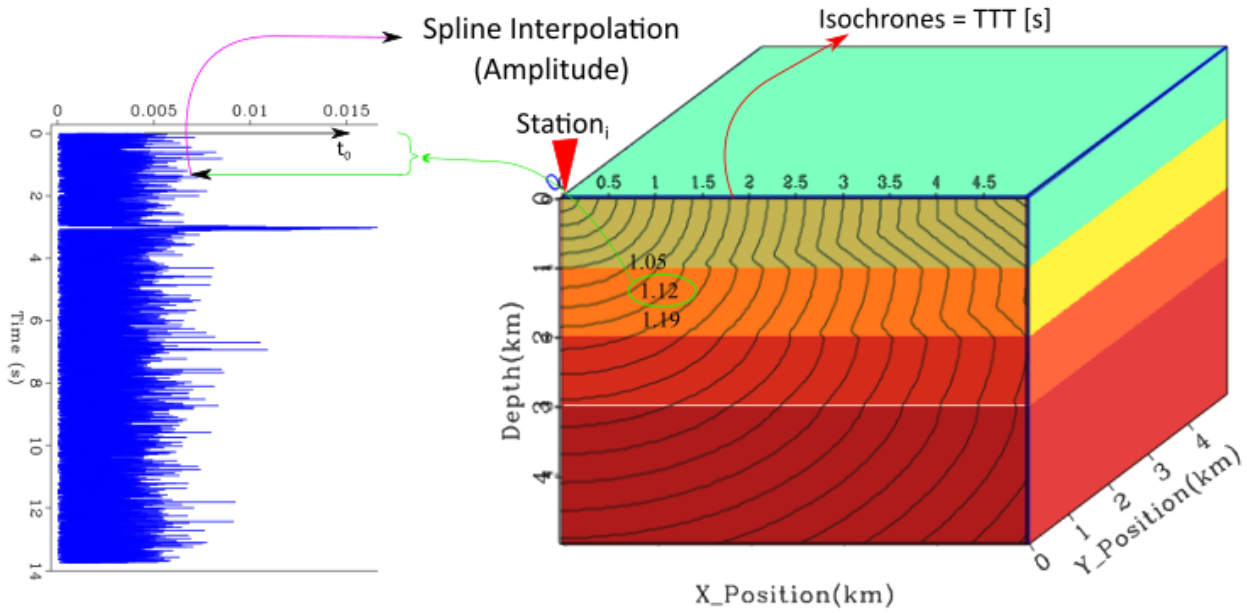


Figure 15. Spline interpolation (left) and travel time-table computations (right).

As mentioned in **table 2** the Back-Projection operation is method-dependent because there are several ways to define the brightness function $F(\boldsymbol{r}, t)$. In this case, the MATF[5] (**Eq. 3**) and the SSA[6] (**Eq. 4**) methods are used.

When backpropagating, the energy at each station is sent back to subsurface using **Eq. 3** which relates amplitudes from normalized seismograms u_i recorded at station $i = 1, \dots, N$ to grid-points according to their travel-times $TTT_i(\boldsymbol{r})$; therefore, converting the travel-time volumes into amplitude volumes for every station.

$$F(\boldsymbol{r}, t) = \frac{1}{N}\sum_{i=1}^{N}\left|u_i\left(t + TTT_i(\boldsymbol{r})\right)\right|, \quad (Eq.\,3)$$

[5] See **Appendix C.1** for information about the MATF method.
[6] **Appendix C.2**

Then, amplitude volumes from the N stations are stacked together along the time axis (t), resulting in a stacked energy volume per timestep function $F(\boldsymbol{r}, t)$, commonly known as *brightness/energy/image function*. Since the true occurrence time is generally unknown, this procedure must be repeated for consecutive time windows of the same length (**figure 16**).

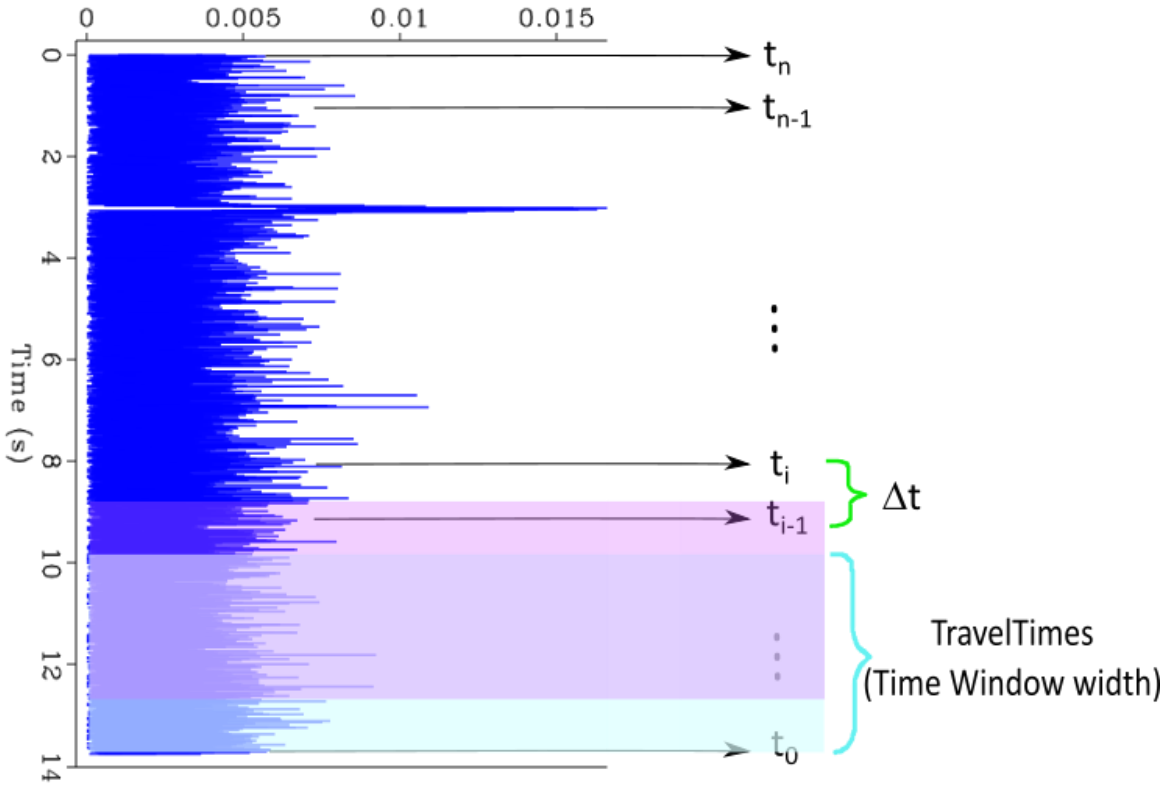


Figure 16. Seismograms reversed in time. The time window width is dependent on the maximum value of the TTT. Once a window is back propagated, traces are moved backwards a given time-step Δt, and the process is repeated for the next time window.

Although equation 2.1 is theoretically correct and used in some research works including (Baker, Granat, & Clayton, 2005), it has been modified to consider the contribution from multiple phases (Liao, Kao, Rosenberger, Hsu, & Huang, 2012) and several other improvements (Pesicek, Child, Artman, & Cieślik, 2014) (Anikiev, Valenta, Staněk, & Eisner, 2014). Probably one of the most important BPI methods is the *Source Scanning Algorithm (**SSA**)* (Kao & Shan, 2004) in which a brightness function $br(\boldsymbol{r}, t)$ is built similarly to equation 2.1, however, a time window $(m\delta t)$ is included to account for velocity uncertainty (**Eq. 4**)[7].

[7] See **Appendix C.2** for more information regarding the SSA method.

$$br(\boldsymbol{r},t)=\frac{1}{N}\sum_{n=1}^{N}\left\{\frac{\sum_{m=-M}^{M}W_m|u_n(t+ttt_i(\boldsymbol{r})+m\delta t)|}{\sum_{m=-M}^{M}W_m}\right\},\quad (Eq.\,2.2)$$

Part B is complete once $F(\boldsymbol{r},t)$ or $br(\boldsymbol{r},t)$ is computed. Now, it is time to advance to **PART C** where the actual location algorithm takes place.

2.6.1. **Location algorithm (Part C).** As can be seen in the blue rectangle in figure 14, two ways to localize event sources are tested in this thesis. *First*, a *Grid Search-Based* approach – to which MATF and SSA belong – where grid-points with energy higher than a given threshold are considered candidates, furthermore, the grid-point with the highest amount of focused energy indicates the most likely hypocenter and origin time. However, one of the main disadvantages of methods like MATF and SSA is that their results are completely dependent upon the discretization of the grid, therefore, they can become computationally demanding because of the inherent tradeoff between resolution and computing time. Note: Brief descriptions of these methods are presented in appendix C, however, the reader is encouraged to look at the original papers for more details.

Second approach. One way to overcome the difficulties mentioned above is by implementing a ***Centroid-Based*** approach in which the image function $F(\boldsymbol{r},t)$ is transformed into a probability density distribution (*PDF*) and later used to calculate weighted averages of the spatial coordinates (x_0,y_0,z_0) as done by (Anikiev, Valenta, Staněk, & Eisner, 2014). The most remarkable benefit of this approach is that solutions are no longer restricted to grid-nodes. Hence, such a centroid-based method is adapted in this thesis to a 4-dimensional version from which two slightly different subkinds are introduced (***PbAS*** & ***PrAS***).

Brightness PDF-based Amplitude Stacking (PbAS). The most noticeable difference between PbAS and the 3D centroid presented in (Anikiev, Valenta, Staněk, & Eisner, 2014) is that its applicability has been extended to 4-Dimensions in which the origin time $\boldsymbol{t_0}$ is included.

Origin time. The normalized Maximum Brightness Curve (i.e. maximum stacked energy value in volume per timestep) is used to calculate a weighted average $\bar{t}$ which represents a first approximation to the solution (**Eq. 5**).

This average can be interpreted as the time around which the energy is uniformly distributed thus this average is valid for cases where there are no clear arrivals but bursts of energy without identifiable phases, as for tremors.

$$NmaxF(t_i) = \frac{maxF(t_i)}{\sum_{i=1}^{nt} maxF(t_i)}, \quad with: \ maxF(t) = \max_{\mathbf{r}}(F(\boldsymbol{r}, t)),$$

$NmaxF(t_i)$: *Normalized maximum brightness curve.*

$$\bar{t} = \sum_{i=1}^{nt} \left(t_i * NmaxF(t_i)\right), \quad (Eq. 5)$$

$\bar{t}$: *Weighted mean time;* nt: *Number of timesteps.*

This approach to calculate the time of occurrence will be referred to as ***T.Centroid*** for clarity from here on.

Later, the standard deviation of the normalized maximum brightness curve σ_t can be used as an indicator of coherent summation of energy (focusing), which also reflects the type of signal being backprojected; either impulsive arrivals (low σ_t) or rather emergent/weak signals (high σ_t).

$$\sigma_t = \sqrt{\sum_{i=1}^{nt} (NmaxF(t_i) * (t_i - \bar{t})^2)},$$

Afterwards, the normalized maximum brightness curve is powered by an exponent n_{exp} (**Eq. 6**), which is related to σ_t, in order to reduce or enhance specific characteristics of the signal as follows:

$$n_{exp} = \begin{Bmatrix} \propto C \ \sigma_t \ \rightarrow Maximum \ (impulsive \ signals - earthquakes) \\ \propto C \ \frac{1}{\sigma_t} \ \rightarrow \ Centroid \ (emergent \ signals - \ tremors) \end{Bmatrix}$$

C: *Proportionality constant.*

For instance, if only the locations with highest stacked energy are desired, then n_{exp} should be define in such a way that it highlights strongly impulsive events while suppressing noise (**figure 17**), as in the case of earthquakes where it is common to determine the origin time of the arrival of highest amplitude. On the other hand, if the seismograms show no clear arrivals, one would prefer

a more centroid-like approximation, as for tremors and LPLD events. Finally, a different mathematical expression may be used to corelate n_{exp} to σ_t, however this decision must be made based on personal expertise and provide the best fit for your data.

$$t_0 = \sum_{i=1}^{nt} (t_i * NmaxF(t_i)^{n_{exp}}), \quad (Eq. 6)$$

t_0: *Origin time corrected for the exponent* n_{exp}.

This way to compute the origin time is referred to as the **T.Peak** method in subsequent sections, regardless of how n_{exp} is defined.

Hypocenter. The spatial location of the event r_0 is found by computing the centroid of the energy in space and time simultaneously (see **Eq. 7**):

$$r_0 = \sum_{j=1}^{nr} \left(r_j * \sum_{i=1}^{nt} \left(NmaxF(t_i)^{n_{exp}} * P(\boldsymbol{r}, t_i) \right) \right), \quad (Eq. 7)$$

Where r_j is the position vector of every grid in the domain, $NmaxF(t_i)^{n_{exp}}$ is a weighting factor which varies according to the maximum value of energy per timestep, and $P(\boldsymbol{r}, t)$ represents *the transformation of the image function* $F(\boldsymbol{r}, t)$ *into PDFs for each timestep.* The main objective of computing $P(\boldsymbol{r}, t)$ is noise suppression thus highlighting the most likely candidates. $P(\boldsymbol{r}, t)$ is calculated as follows:

$$P(\boldsymbol{r}, t) = \frac{e^{\left(-\frac{(F(\boldsymbol{r},t) - maxF(t))^2}{2\sigma_{F(r,t)}^{m_{exp}}} \right)}}{\sum_{i=1}^{nt} e^{\left(-\frac{(F(\boldsymbol{r},t_i) - maxF(t_i))^2}{2\sigma_{F(r,t_i)}^{m_{exp}}} \right)}},$$

Where: $\sigma_{F(r,t)}$ is the standard deviation of the image function, and $\sigma_{F(r,t)}^{m_{exp}}$ is a factor which controls the spatial resolution. The exponent m_{exp} is equal to 2 for a Gaussian distribution, however, this value can be modified to directly affect the width of the distribution according to the kinds of signals being studied; therefore, low values of m_{exp} are better for events of long periods, while high values are preferred for events of short periods and durations.

$$m_{exp} = \begin{cases} < 2 \rightarrow & Centroid(Emergent\ or\ long\ period\ and\ duration\ signals - tremors) \\ 2 \rightarrow & Gaussian\ Distribution \\ > 2 \rightarrow & Maximum\ Amplitude\ Arrival(Impulsive\ signals - earthquakes) \end{cases}$$

***Spatial PDF-based Amplitude Stacking (PrAS)*[8].** Similarly, **PrAS** solves for the 4 variables $(\boldsymbol{x_0}, \boldsymbol{y_0}, \boldsymbol{z_0}, \boldsymbol{t_0})$ at once. The origin time is determined following the exact same procedure as in PbAS (see Eq. 5 & 6) and thus, its explanation is skipped.

Hypocenter. Unlike PbAS where the image function $F(\boldsymbol{r}, t_i)$ is transformed into a series of PDFs, PrAS uses $F(\boldsymbol{r}, t_i)$ directly in the computation of the spatial centroid, and introduces a correction factor $w_r(\boldsymbol{r}, t_i)$ to suppress noise and false candidates as their distance from the point of maximum energy increases (**Eq. 8**). In other words, $w_r(\boldsymbol{r}, t_i)$ removes moderate energy values when they occur away from the most likely location. **Note:** the mean value must be removed from the image function $F(r, t_i)$ in order to prevent incoherent noise from contributing to the centroid computation.

$$r_0 = \sum_{j=1}^{nr} \left(r_j * \sum_{i=1}^{nt} \left(NmaxF(t_i)^{n_{exp}} * F(r, t_i) * w_r(r, t_i) \right) \right), \quad (Eq. 8)$$

Where:

$w_r(r, t)$: *Amplitude correction factor based on distance from the maximum brightness source.*

$\sigma_{(r*f,t)}$ is the standard deviation of the image function times the distances from the grid-point of maximum energy and all the rest $\|\mathrm{r_{max}} - \mathrm{r_i}\|$. Notice $\sigma_{(r*f,\ t)}^{m_{exp}}$ controls the spatial resolution of the method thus it can be adjusted based on given requirements (e.g. a fixed-constant value for desired resolution).

$$w_r(r, t) = \frac{e^{\left(\frac{-\left(||r - r_{max}(t)|| \right)^2}{2\sigma_{(r*f,\ t)}^{m_{exp}}} \right)}}{\sum_{i=1}^{nt} e^{\left(\frac{-\left(||r - r_{max}(t_i)|| \right)^2}{2\sigma_{(r*f,\ t_i)}^{m_{exp}}} \right)}},$$

[8] Where **r** stands for "radial".

Notice that PrAS *highlights only the most energetic event per timestep.*

The above-mentioned methods are summarized in **figure 17** for purposes of clarity.

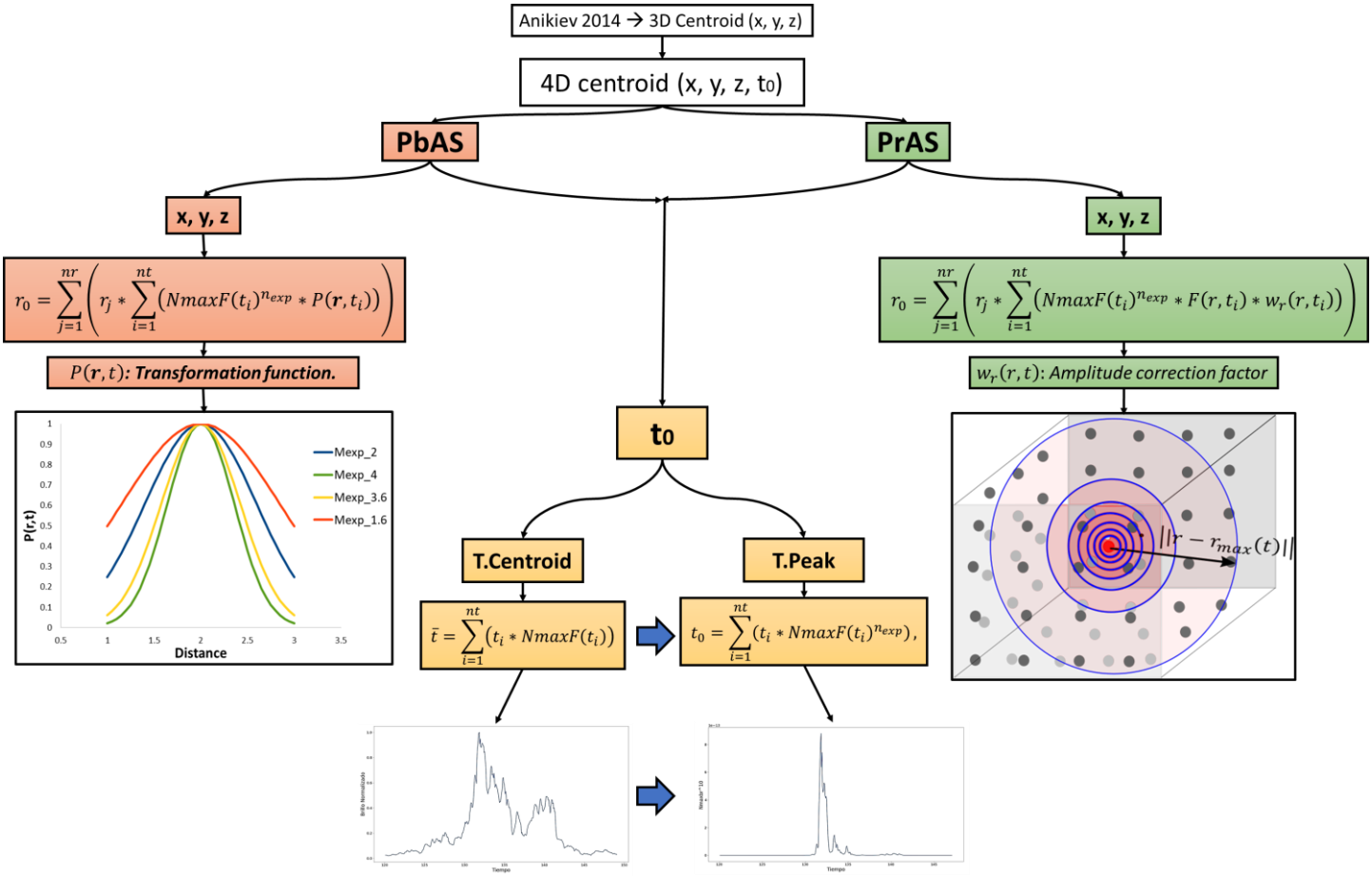


Figure 17. Schematic representation of the proposed methods.

2.7. Sensitivity Analysis

The effect of changes in the velocity model, SNR, frequency content, and surface sensor layouts on the spatiotemporal location of the event is analyzed, concentrating especially on the impact of velocity uncertainty on the solution. For this purpose, a Monte Carlo algorithm coupled with the Fast-Marching technique is implemented to evaluate hundreds of velocity models and produce results in a reasonable time (see **Chapter 3**).

2.8. Main Contributions

We propose a seismic event location algorithm based on Back-Projection Imaging, able to compute simultaneously the spatiotemporal location of diverse seismic signals like conventional

earthquakes, microearthquakes, LPLD, tremors and others. Besides, this method allows to conduct sensitivity analysis thanks to its computational efficiency.

Most BPI methods assume the origin time t_0 as the time when the Maximum Brightness Function (MBF) reaches its peak, but that might not be a proper assumption for more complex events characterized by long duration bursts of energy with no defined amplitude maxima. Subsequently, here we introduce a 4D adaptive centroid-based location method which could sometimes be considered a more reasonable location coordinates and origin time estimation.

This chapter also reminds geoscientists of the existence of powerful free sources, like Madagascar, which are available for the scientific community in general to use in their projects.

3. Assessment of Seismic Event Location in Realistically Simulated Conditions

3.1. Overview

Location of seismic signals implies high levels of uncertainty that may dramatically affect the accuracy of the solution, in other words, the effectiveness of the algorithm is highly dependent upon the quality and reliability of the input data. Several sources of uncertainty are present when locating seismic event hypocenters including but not limited to incorrect velocity models, low Signal to Noise Ratio (SNR), poor coverage and others. Here, a sensitivity analysis is performed on the above-mentioned parameters to assess the inherent uncertainty of an algorithm based on *BackProjection Imaging (BPI).*

In this work, each synthetic seismogram consists of a single impulsive signal filtered at a given frequency. The purpose of using such simplified signals is to completely understand the mechanics of the location method in idealized and controlled scenarios, on which each parameter can be freely perturbed (while the others remain unaffected) to ensure that their individual effects are easily manifested on the outcome.

Results show the need for a time-frequency analysis in order to discard as much noise as possible and prevent random noise from adding up, which could lead to an increasing number of artefacts. It was also observed that a good angular coverage reduces the uncertainty with which

the location of a candidate is determined; additionally, depth is the most sensitive spatial coordinate to velocity and geometry changes, according to our finding.

Since a totally accurate velocity model is never available, there is an implicit velocity uncertainty that needs to be assessed for every location method. Such estimation is carried out using a Monte Carlo Algorithm which randomly samples velocity models according to *Probability Density Functions* – PDFs constructed from well sonic logs and Vertical Seismic Profiles (VSPs); then, each realization is used to backpropagate the synthetic seismograms thus obtaining distributions of the spatiotemporal location (x, y, z, t) for each method. Results indicate a correlation between large velocity uncertainty and location errors; however, some methods are more susceptible to perturbations than others. For instance, an incorrect velocity model may keep the stacked energy from focusing and reaching a reasonable maximum value, perhaps, resulting in missed events when using methods that search for the exact grid-point with the highest stacked energy, whereas the centroid-based methods may experience negligible effects.

3.2. **Introduction**

Location of seismic signals has been commonly performed by methods which require picking of arrivals and correct association of the seismic phases among stations, nonetheless, there are several other phenomena which generate complex waveforms with no distinguishable phases thus promoting the use of other methods able to handle both the dynamic and kinematic wave properties. Among the various proposals, BPI has risen as a relatively simple but efficient way to image and characterize active and passive sources without the demanding computation of the full wave equation (e.g. Time-Reversal Imaging - TRI) which requires both accurate velocity models and densely recorded data (Beskardes, et al., 2018).

Wave propagation in real life is complex, and the assumptions commonly made in seismology such as isotropic layered media may become only one in several sources of uncertainty. For instance, location of seismic and microseismic events is typically performed by ray tracing methods which require accurate picking of P- and S- wave arrival times, additionally, since raypaths are bent by velocity heterogeneities, too complex models may result in shadow zones. Furthermore, combinations of high levels of noise, bandwidth limitations, poor receiver coverage (large azimuthal gaps) and complex source mechanisms, usually encountered during field

acquisitions, lead to large errors from which it is difficult to separate the contribution of each factor (Eisner, Duncan, Heigl, & Keller, 2009).

Among the advantages BPI have over others approaches, we can mention ability to produce accurate results while being simple to use and not so computationally expensive, which makes it suitable for uncertainty analysis. These characteristics are so useful since the main purpose of this chapter is to evaluate the behavior of the algorithm and its results when the values of some parameters are changed.

3.3. **Model Configuration**

A correct analysis of the independent and joint effects of the parameters requires the evaluation of idealized scenarios where each variable can be properly controlled. Thus, the most relevant assumptions made in this initial analysis are: i) a horizontally layered earth model, ii) constant velocity in each layer, iii) velocity increases with depth, iv) an impulsive signal of a single event with no random noise, v) explosive point source and vi) only one event occurs at a time. These assumptions are valid along this study unless otherwise stated.

The volume of study is a cube of 5kms. The chosen geometry consists of 4 recording stations deployed on a square surface array at equal distance in the (x, y, z) coordinate system as shown in **figure 18**. Position of receivers, velocity values and other information about the base model is presented in **table 3.**

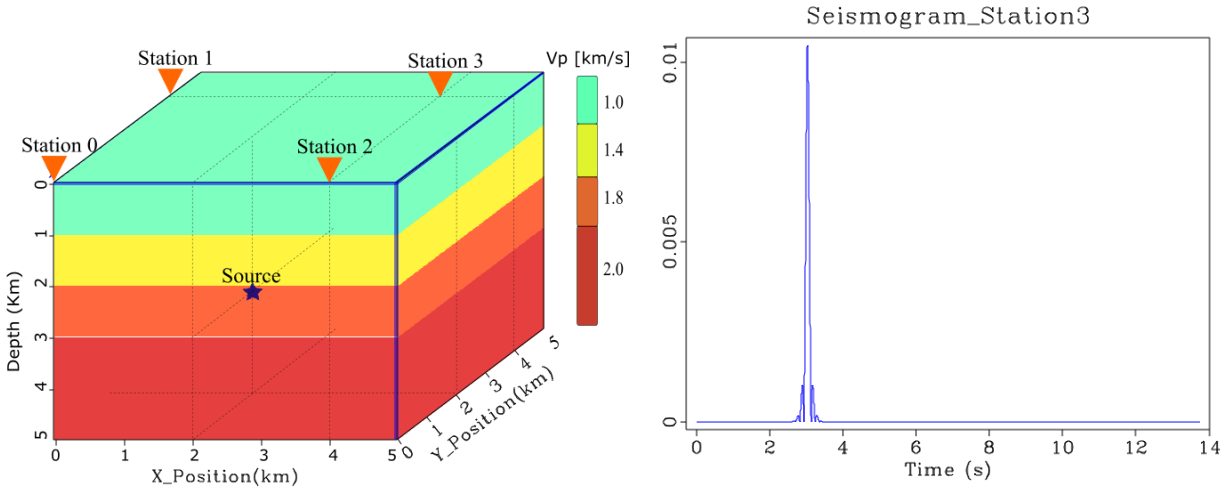


Figure 18. Layer cake velocity model and recording geometry (left), impulsive signal for station (right).

Table 3. Information of base case scenario for sensitivity analysis.

Field	Element	Value
Location (x, y, z) [km]	Station 0	(0, 0, 0)
	Station 1	(0, 4, 0)
	Station 2	(4, 0, 0)
	Station 3	(4, 4, 0)
	Source	(2, 2, 3)
Depth range [top-base] [km]	Layer 1	0 - 1
	Layer 2	1 - 2
	Layer 3	2 - 3
	Layer 4	3 - 5
P Velocity [km/s]	Layer 1	1
	Layer 2	1.4
	Layer 3	1.8
	Layer 4	2
Volume dimensions [km]	X, Y, Z	5, 5, 5
Low pass filter [Hz]	All seismograms	5
Noise level %	All seismograms	0
1 single event.		

Furthermore, to make a clear analysis for each abovementioned parameter, wave propagation phenomena including scattering, dispersion, attenuation and anisotropic media are not considered in this study.

3.4. **Signal to Noise Ratio (SNR)**

Methods using both the dynamic and kinematic wave information are quite stable in presence of high level of noise, however, decreasing the SNR may increase the number of artefacts and affect its resolution in the image domain. Hence, it becomes useful to choose a high threshold value (Φ around 85%) in order to discard highly energetic artefacts but without missing true sources with low energy (**figure 19**).

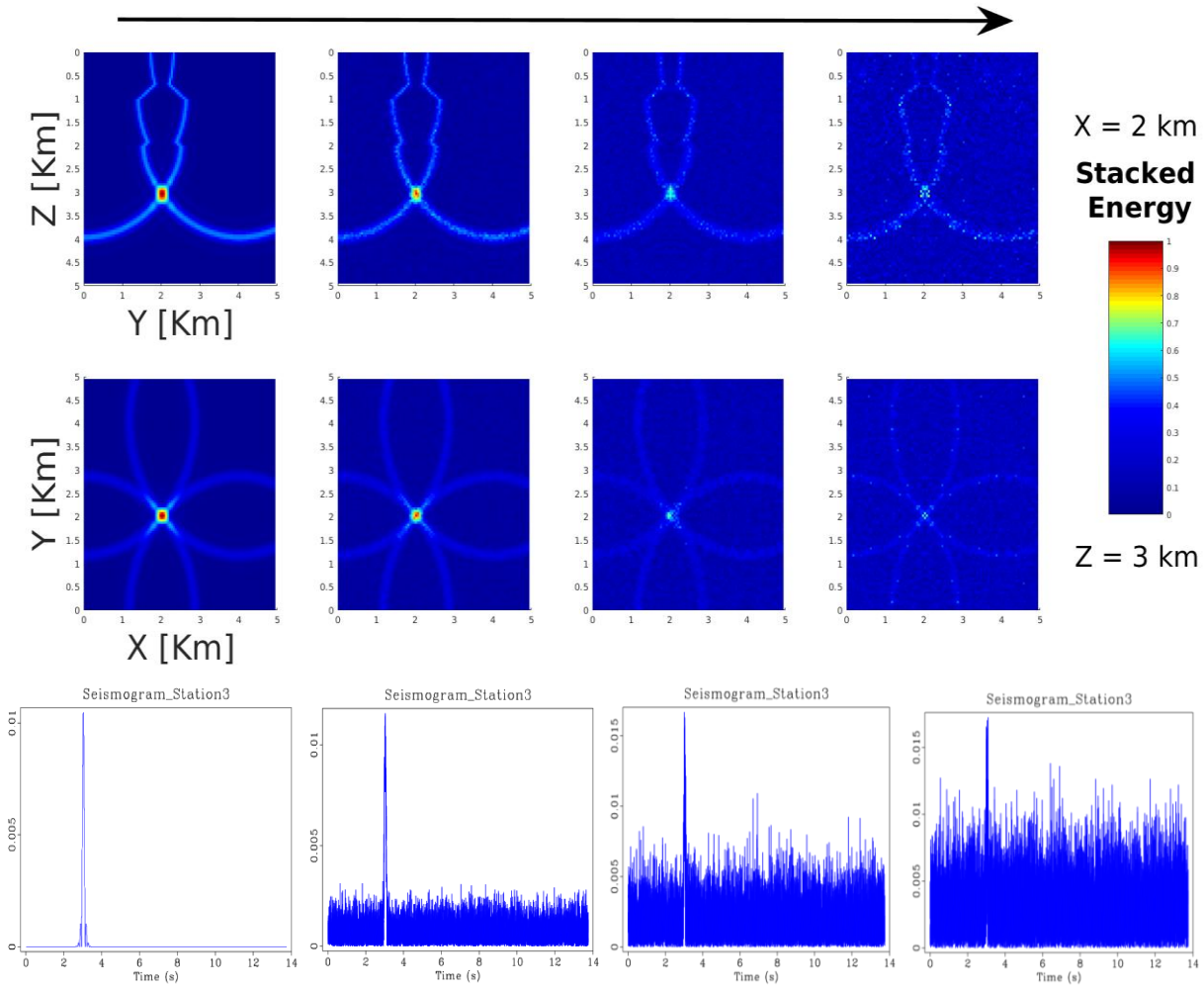


Figure 19. Influence of noise on spatial location of the event. Upper panel shows a horizontal x-plane, a view of a vertical slice in middle panel and their respective seismograms in lower panel. Backpropagating noisy seismograms may lead to increasing numbers of local maxima in the 3D space (x, y, z) as the level of noise rises. Nevertheless, the hypocenter is correctly located at (2, 2, 3) km when an appropriate threshold is used.

3.5. **Frequency content**

As in the case of noise, frequency content plays a key role on the resolution with which the source can be resolved. According to **figure 20**, the lower the frequency of the signal, the larger the zone within which the source can be located. This happens because lower frequencies have longer wavelengths resulting in gradually less capability to detect small underground structures.

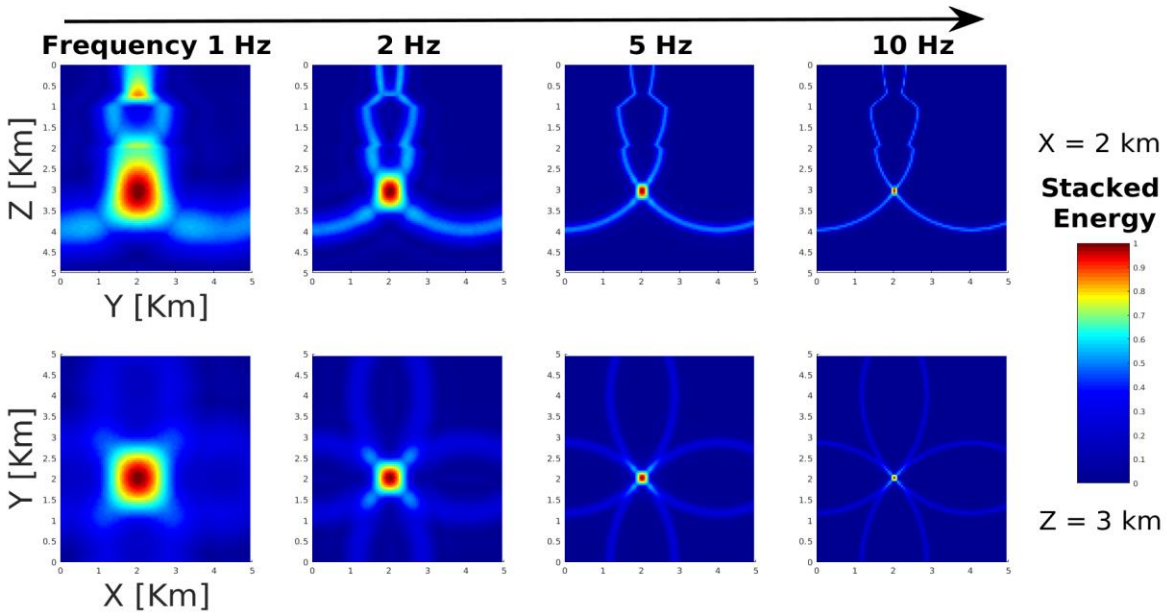


Figure 20. Influence of frequency on spatial location of the event.

Another important observation made after various tests is that switching between different frequency ranges of the signal did not affect the location of the most energetic grid-point (2, 2, 3 km), except for a special case when the event originates close to the boundaries of the model, as it is explained later.

3.6. **Velocity variations**

The knowledge of an accurate velocity model is an essential in seismic and seismology since it determines if a structure or an event can be correctly located in depth. Unfortunately, there is never a complete understanding of the geology and physics of the subsurface, leading to large uncertainty in the distribution of properties such as rock velocity, density, resistivity, permeability, and many other. This issue is partly overcome by the inversion of measure data and prior information, thus obtaining reasonable property values, nonetheless, results are never perfect. **Note:** Velocities in this section are expressed as percentages of the true value (100%).

3.6.1. **Effects of velocity on spatial location.** For a known origin time $\boldsymbol{t_0}$ $[\boldsymbol{s}]$, reducing the values of the velocity model by a given percentage will focus energy at a shallower position than the true hypocenter, whereas overestimating the velocity model will deepen the event (**figure 21**). Besides, in a plane at $\boldsymbol{z = 3km}$ (true depth) the wavefronts travelling at lower velocities have not yet reached the hypocenter, while at higher velocities they have already passed through it.

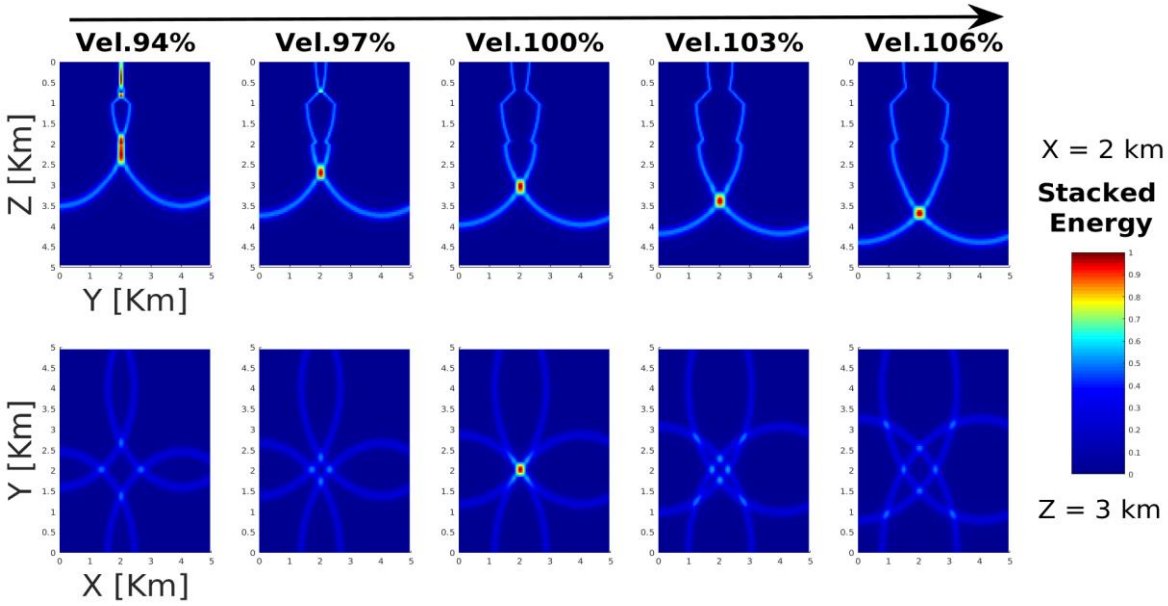


Figure 21. Velocity influence on spatial location of an event at the center of a square array.

3.6.2. Effects of velocity on origin time. For an unknown origin time $\boldsymbol{t_0}$ $[\boldsymbol{s}]$, the search of the global maximum must be carried out not only in a 3D volumen, but also in different timesteps. Now, imagine the spatial location of the source was known, an interesting way to find the origin time would be to plot the energy at such location as a function of time; then, one would expect the energy to increase steeply as the time window approaches the origin time **(figure 22).**

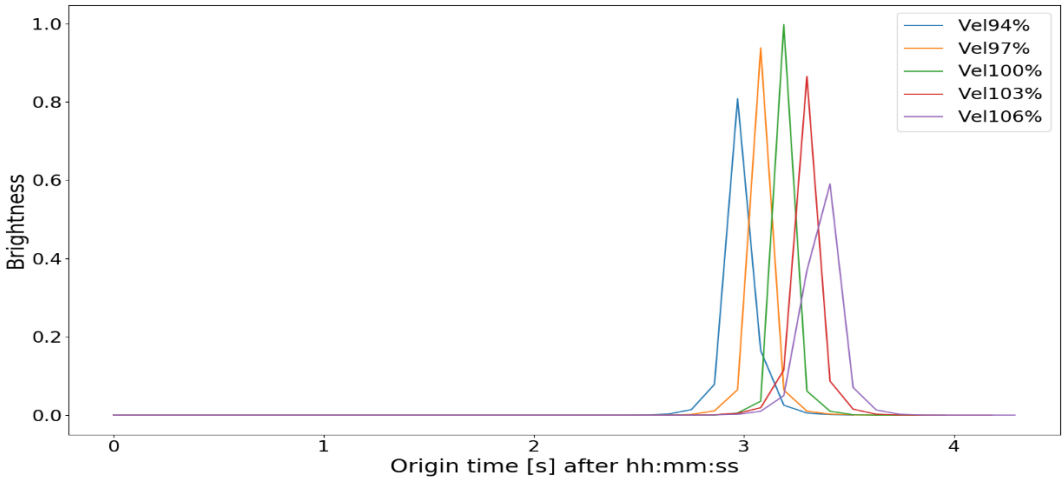


Figure 22. Energy (brightness function) at source vs origin time.

It is worth noting that lower velocities result in earlier origin times than higher velocities, because they travel slower and need more time to reach a place at a specific time (i.e. think of a bus that must reach its destination at 12:00, its departure time t_0 would have to be earlier if traffic were

heavy, resulting in lower mean velocity). Another important feature of these curves is that using incorrect velocity models produce a mismatch between signals such that the highest stacked energy can only be achieved with the true model.

A similar curve can be drawn from the value of stacked energy at the most energetic gridpoint of the volume versus time; this graph is sometimes called *maximum brightness/amplitude time function* (**figure 23**). The interesting part of this plot is that, it provides a range for the origin time. As in figure 3.5, when approching the true ocurrence time of the event the energy function maximizes.

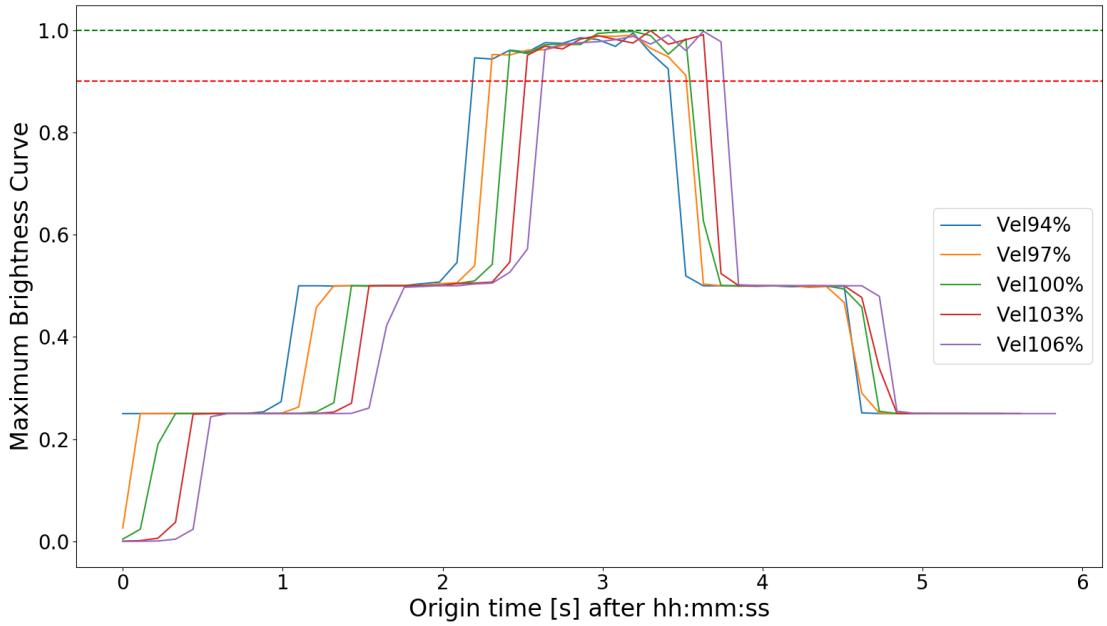


Figure 23. Maximum brightness in volume vs time for 5 different velocity models. Red dashed line represents an arbitrary threshold to distinguish between candidates and false locations.

It is again evident that lower velocities imply earlier origin times. Another interesting feature is the presence of a plateau that raises some uncertainty about the origin time; the length of such a plateau and shape of the *Maximum Brightness Curve (MBC)* in general depend on numerous factors including: i) wavelength of the peak frequency, ii) frequency content, iii) spatial discretization of the grid, iv) geometry of receivers, v) noise level, among others.

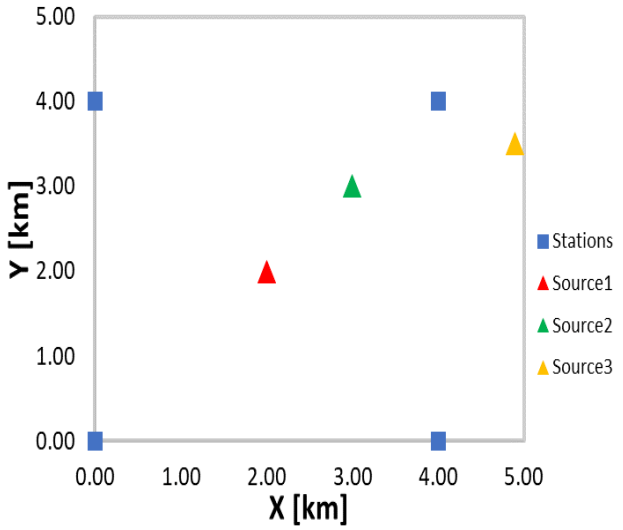
3.7. **Geometry of receivers**

Accuracy on each of the spatial coordinates (x, y, z) is dependent upon the geometry of the array and relative position of the event source. Thus, a good surface coverage is mandatory for correct

identification, location and characterization of events originated from both active and passive sources.

A huge number of configurations can be deployed to register seismic waves on field, and each one will provide better constraints in some directions than others. In this study, relative position of the event, number of receivers and array shape will be varied to analyze their impact on the outcome.

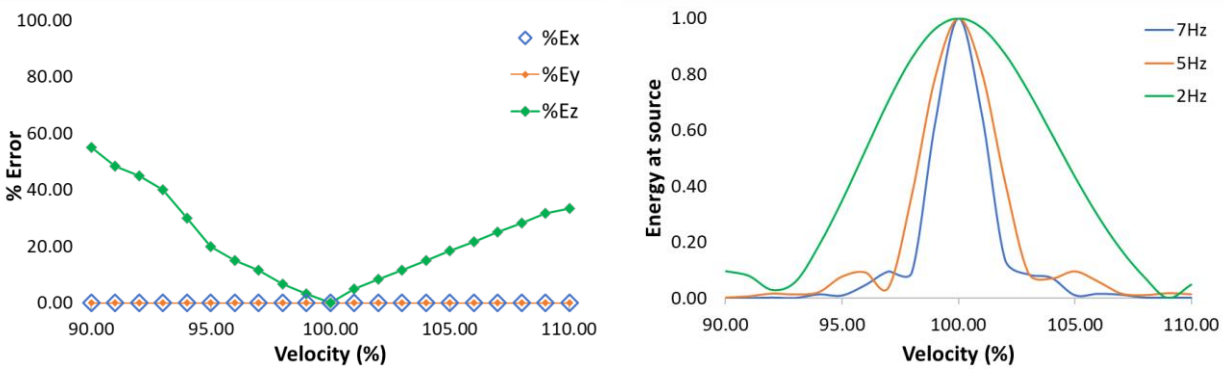
3.7.1. Square geometry (Relative position of event). Three experiments are carried out as shown in **figure 24**, to evaluate the influence of the relative position of the event with respect to the recording array.



Item	X	Y	Z
Station0	0	0	0
Station1	4	0	0
Station2	0	4	0
Station3	4	4	0
Source1	2	2	3
Source2	3	3	3
Source3	4.9	3.5	3

Figure 24. Geometry of receivers and source position for 3 experiments.

Centered source. In this case the event is placed at the center (red triangle) and, due to symmetry, the X and Y coordinates are perfectly retrieved (0% error) regardless of the velocity and frequency used for the backpropagation. On the other hand, error in the Z-component varies significantly with velocity changes as seen in **figure 25a**, another way to study the role of velocity on the correct location is presented in **figure 25b** where brightness at source vs velocity curves are plotted for 3 frequencies. Interestingly, energy at source reaches it maximum value as a more precise velocity model is used.



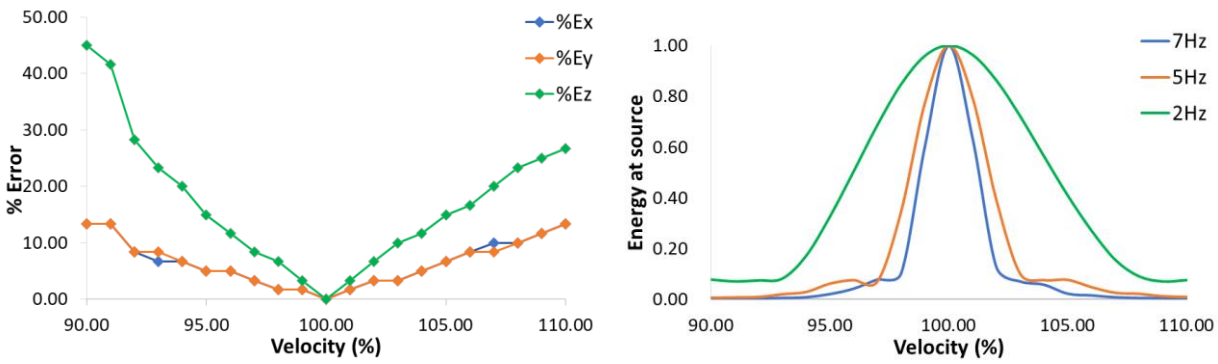
Frequency	7 Hz	5 Hz	2 Hz
S. St. Dev.	0.272	0.305	0.362
P. St. Dev	0.265	0.298	0.353

Figure 25. Centered case: accuracy of the solution vs Velocity variations; a) Percentage error of the spatial coordinates, b) Brightness value at source vs velocity for 3 different frequencies and their respective standard deviations.

Besides, this last set of curves reinforces a previous observation that higher frequencies provide better resolution of the source and require more accurate velocity information, whereas lower frequencies can focus more energy in wider range of velocities; consequently, yielding a larger uncertainty in location associated to velocity variations, as expressed by higher standard deviations. Note: readers should look back at **figure 21** to see the combined effects of this recording geometry and velocity variations on the spatial location.

Non-centered source. As seen in the previous case, a symmetric array with respect to the hypocenter of the event does not enable to analyze the robustness of the algorithm to recover the X and Y coordinates; therefore, moving the source to the position represented by the green triangle in figure 3.7 is useful for this purpose.

Figure 26a shows symmetric X and Y errors increasing linearly with velocity changes, for example, variations of $\pm 10\%, \pm 5\%$ ΔV produce percentage errors of approximately $14\%, 6\%$ respectively. On the other hand, depth (in green dots) is severely affected by small velocity variations; a $\pm 10\%$ ΔV results in an error as large as 45%. **Figure 26b**, explains the influence of velocity on energy at source for 3 frequencies. Higher frequencies are characterized by a concentration of energy in a narrower range of velocities, whereas low frequencies may sum energy even with less accurate models.



Frequency	7 Hz	5 Hz	2 Hz
S. St. Dev.	0.268	0.300	0.352
P. St. Dev	0.262	0.293	0.343

Figure 26. Non-centered case: accuracy of the solution vs Velocity variations; a) Percentage error of the spatial coordinates, b) Brightness value at source vs velocity for 3 different frequencies and their respective standard deviations.

X and Z slices for different velocities are plotted in **figure 27**. Unlike the centered case, lateral event location does change, and energy may not focus with incorrect velocity models. Additionally, the event deepens diagonally instead of vertically because the source is closer to one of the corners of the square.

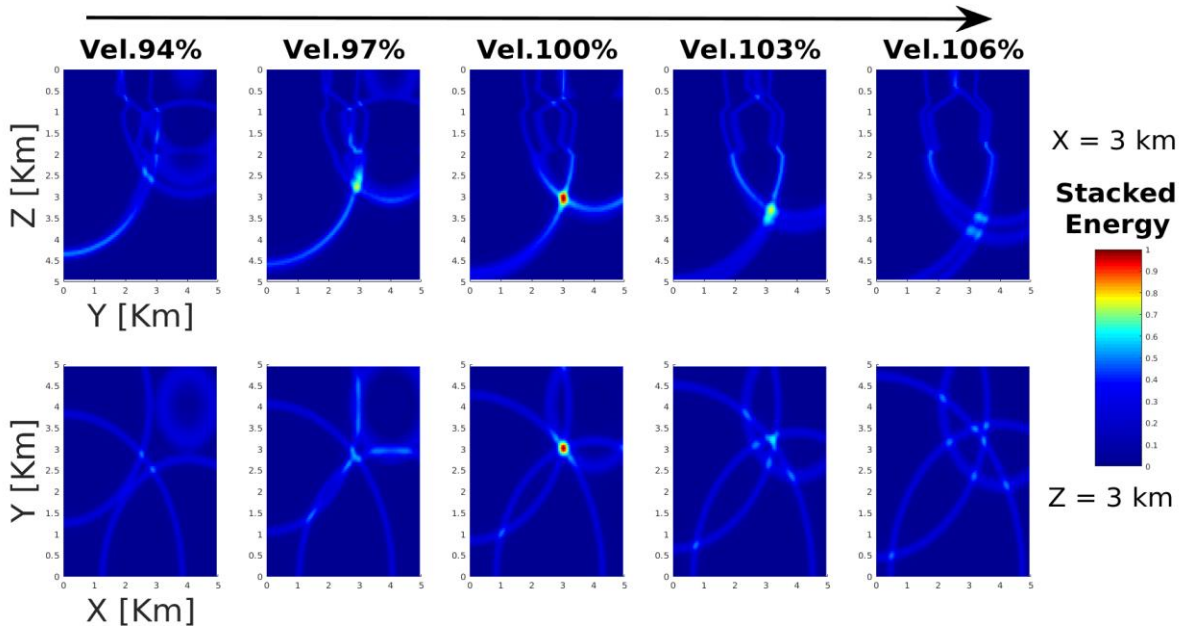
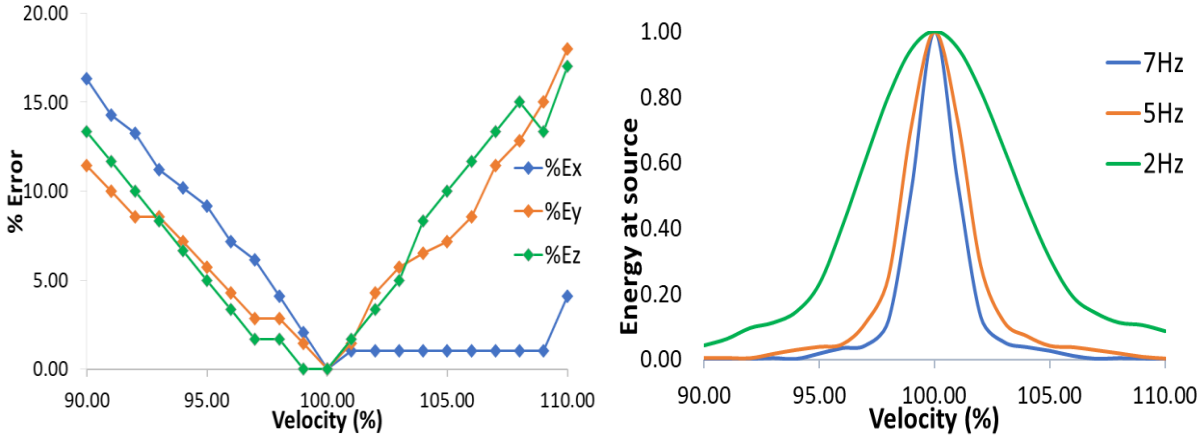


Figure 27. Square Geometry: Velocity influence on spatial location for a non-centered event

Source outside of array. This case is particularly interesting because in seismology, regional events may sometimes be misclassified as local ones due to an incorrect velocity model. Although

this is very rare for conventional earthquakes, it has been reported by (Zecevic, Daniel, & Jurick, 2016) that, under certain conditions, low magnitude regional earthquakes may exhibit waveforms and other features similar to local LPLD events (Our subject of study in **Chapter 4**).

Since velocity plays a more critical role at higher frequencies, as explained in **section 3.5**, an anomalously over estimated velocity model prevents energy from focusing inside the volume of study, producing high errors even when the event originated inside it (**figures 28a)**. Stacked energy at source (**figure 28b)** shows that erroneous velocities do not focus energy, however, energy summation may happen at low frequencies.



Frequency	7 Hz	5 Hz	2 Hz
S. St. Dev.	0.254	0.285	0.339
P. St. Dev	0.248	0.278	0.331

Figure 28. Accuracy of the solution vs Velocity variations for a source outside of the array; a) Percentage error of the spatial coordinates, b) Brightness value at source vs velocity for 3 different frequencies and their respective standard deviations.

An interesting observation can be made from the *maximum brightness/amplitude time function*, which is a plot of maximum brightness value in volume vs time (**figure 29**). The staircase shape reflects when one or more stations start or stop contributing to the energy summation. Additionally, a notorious drop of energy between 3.5 and 4 [s] can be observed on every curve, consequence of the mismatch induced by high velocities.

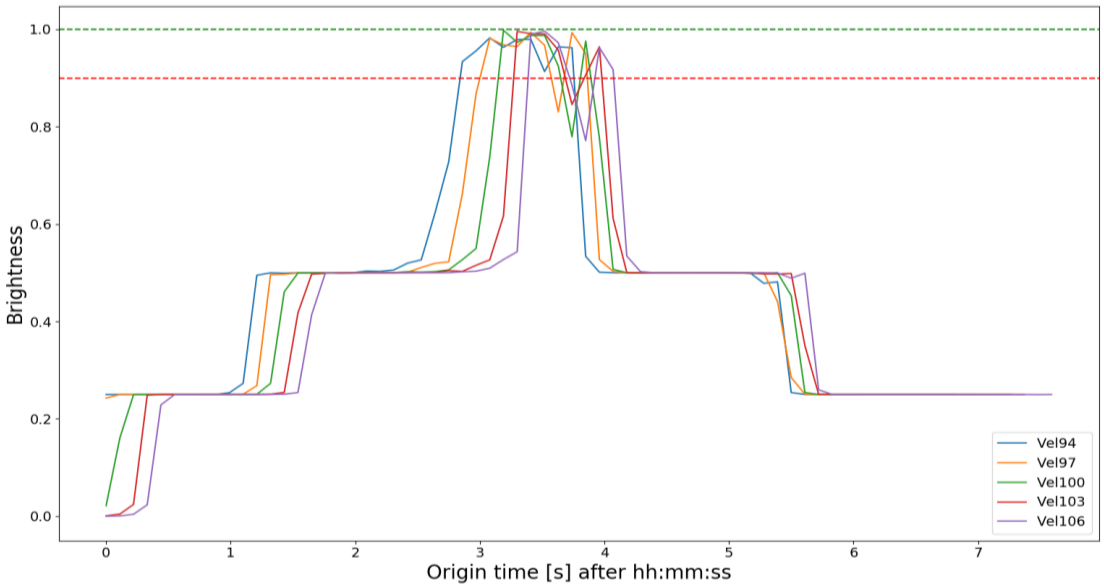


Figure 29. MBCs for an event originated outside the recording array.

Vertical (Y = 3,5 km) and horizontal (Z = 3 km) slices of the volume have been displayed in **figure 30** to see what happens when backpropagating with different velocities. The upper panel shows how velocity affects the event location at depth; lower velocity estimates generate either shallower sources or false candidates, while upper estimates keep energy from adding up. On the lateral location, high velocity errors could incorrectly classify a local event as regional or conversely.

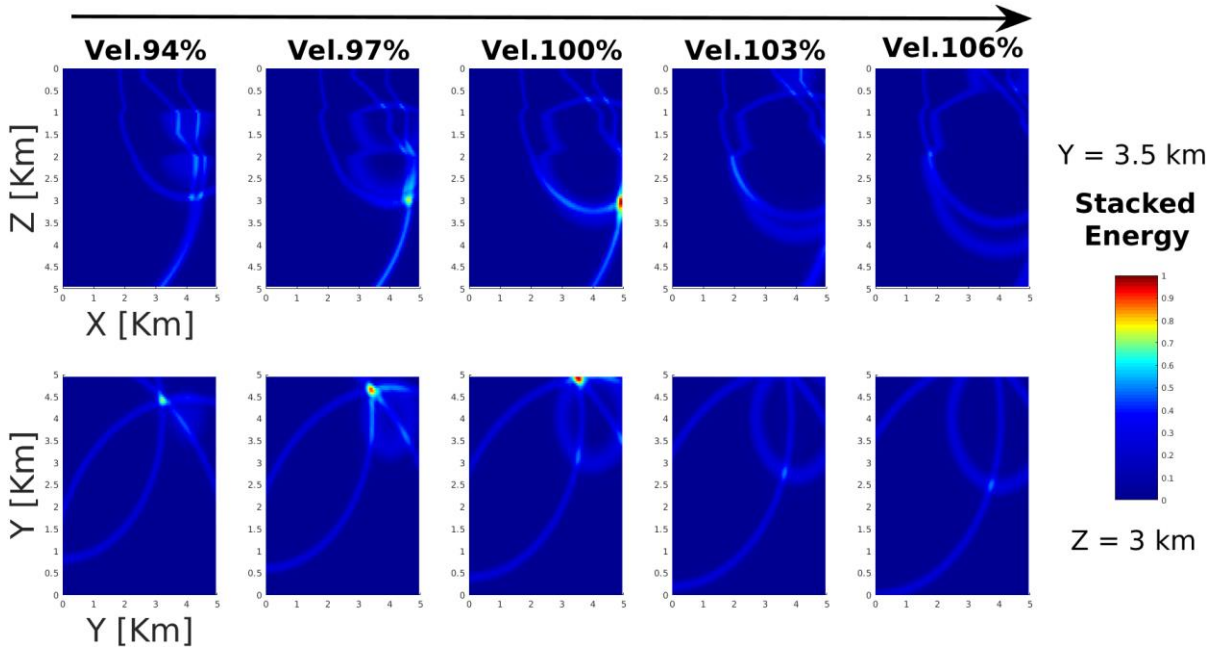
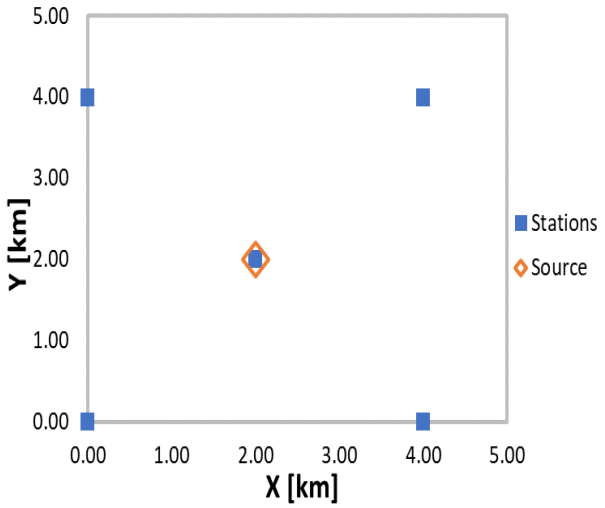


Figure 30. Square Geometry: Velocity influence on spatial location for a nonlocal seismic event.

3.7.2. 5 stations (square + central station). In practice, some seismologists believe that placing a recording sensor right above the source (if it is known) helps constrain the Z coordinate. For this reason, the configuration in **figure 31** attempts to evaluate this idea.



Item	X	Y	Z
Station0	0	0	0
Station1	4	0	0
Station2	0	4	0
Station3	4	4	0
Station4	2	2	0
Source	2	2	3

Figure 31. Base case (Square) + extra–station above source.

Percentage errors for spatial location and curves of energy at source (**figure 32**) do not vary compared to the case in **figure 25**, however, this behavior is induced by the symmetry of the array. In presence of an irregular array, this extra-station could have a positive impact on the z-coordinate and origin time.

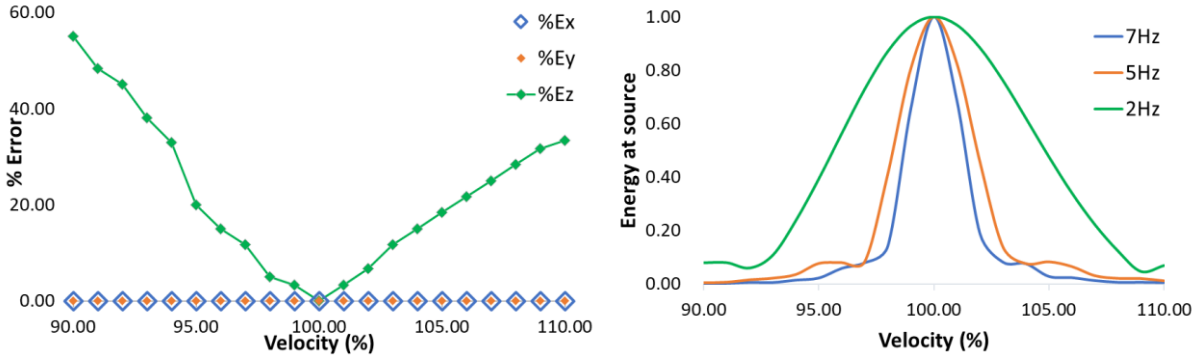


Figure 32. Accuracy of the solution vs Velocity variations; a) Percentage error of the spatial coordinates, b) Brightness at source vs velocity for 3 different frequencies.

Despite the similarity of this case and the base case, there is a clear difference generated by the central station. As it can be seen in **figure 33,** all maximum brightness curves exhibit peaks, making it easier to visually and computationally identify the origin time.

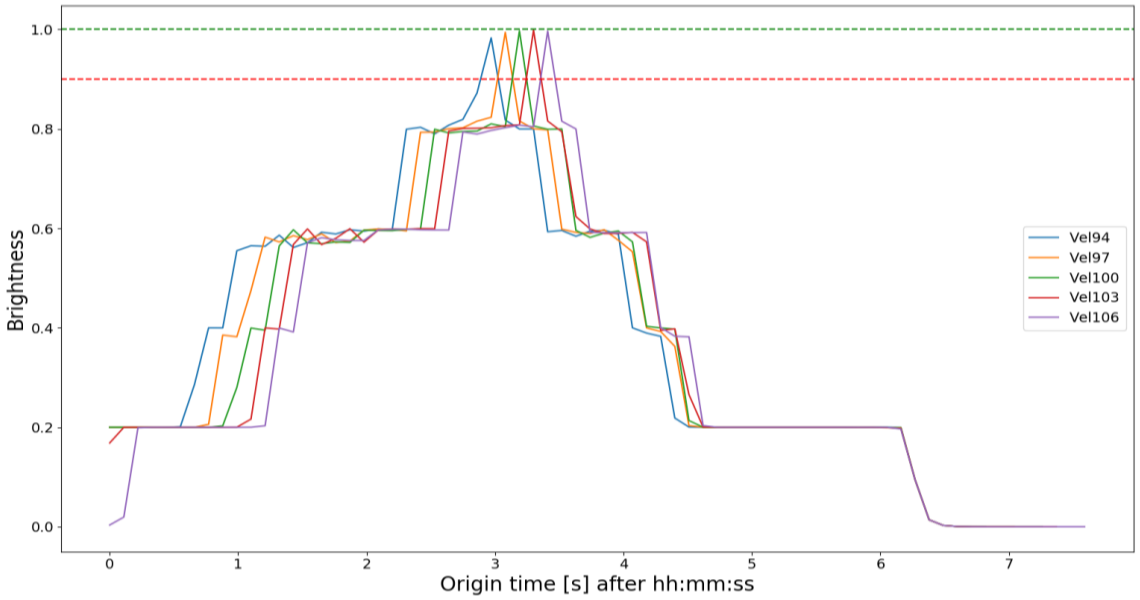


Figure 33. MBCs for a square recording array + one station right above source.

To sum up, if the velocity model is accurate, the spatial location could take advantage of the extra station for cases with presence of high noise level. Nonetheless, if the SNR is high enough, as in **figure 34**, there are no appreciable benefits with respect to the case base.

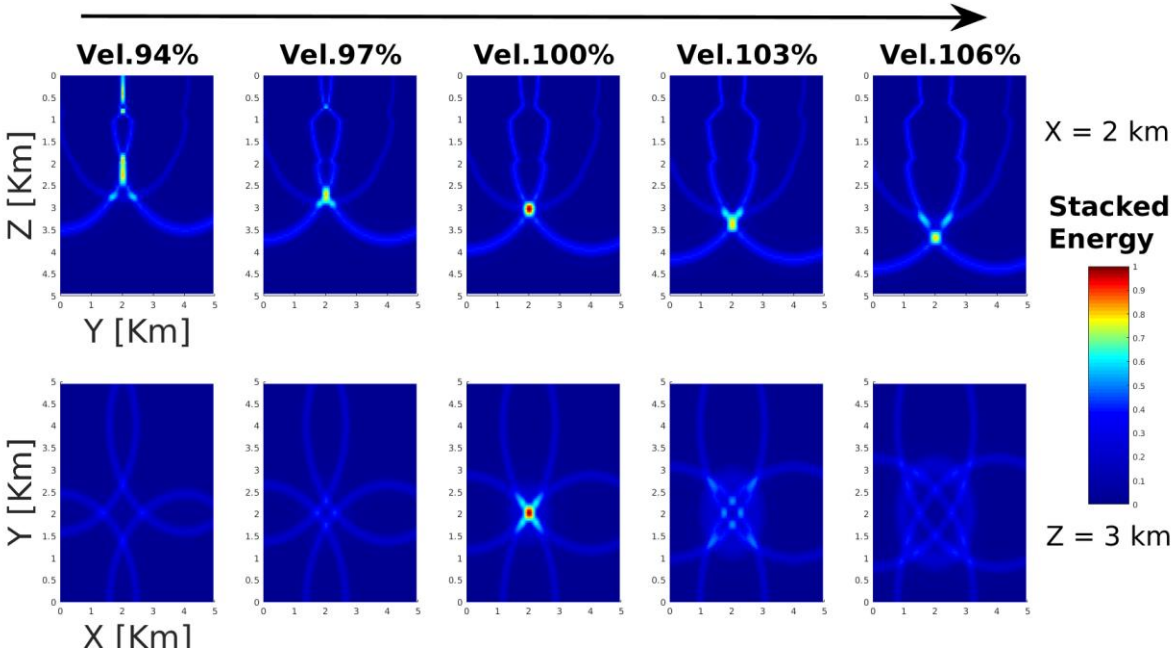
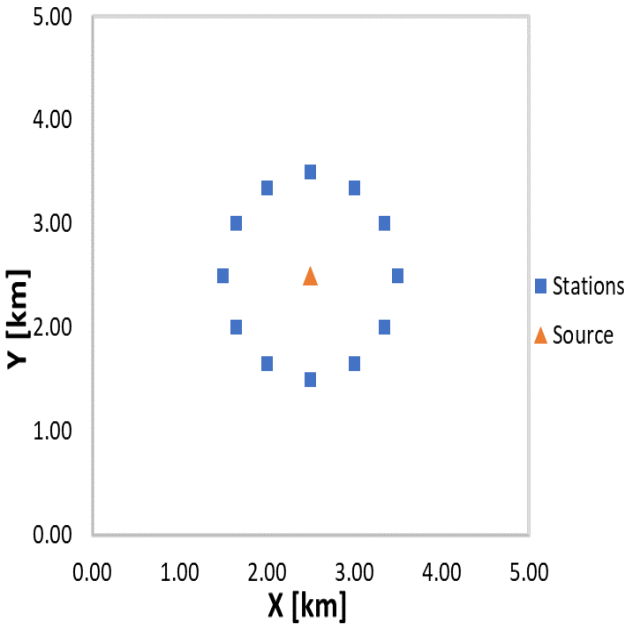


Figure 34. Square + Central St: Effects of velocity on spatial location for a local event.

3.7.3. Circular geometry (12 stations). A geometry commonly deployed on field consists of a circular array around the source (**figure 35**) which in theory facilitates its location. In seismology, a smaller azimuthal gap implies a lower uncertainty in such directions; for this experiment, a 12-station array gives an azimuthal gap of 30°.



Item	X	Y	Z
Station0	3.50	2.50	0
Station1	3.35	3.00	0
Station2	3.00	3.35	0
Station3	2.50	3.50	0
Station4	2.00	3.35	0
Station5	1.65	3.00	0
Station6	1.50	2.50	0
Station7	1.65	2.00	0
Station8	2.00	1.65	0
Station9	2.50	1.50	0
Station10	3.00	1.65	0
Station11	3.35	2.00	0
Source	2.50	2.50	3.00

Figure 35. Circular array configuration.

The above-configuration allows one to effectively recover the X and Y coordinates with total accuracy (it is also a consequence of the symmetry of the array); on the other hand, error in depth (Z-coordinate) behaves linearly with velocity, showing the lowest error (17%) for a $\pm 10\%\ \Delta V$ among all tested scenarios (**figure 36a**). In this case, the MBCs in (**figure 36b**) are not useful for origin time estimation since these curves have large plateaus due to the multiple local maxima.

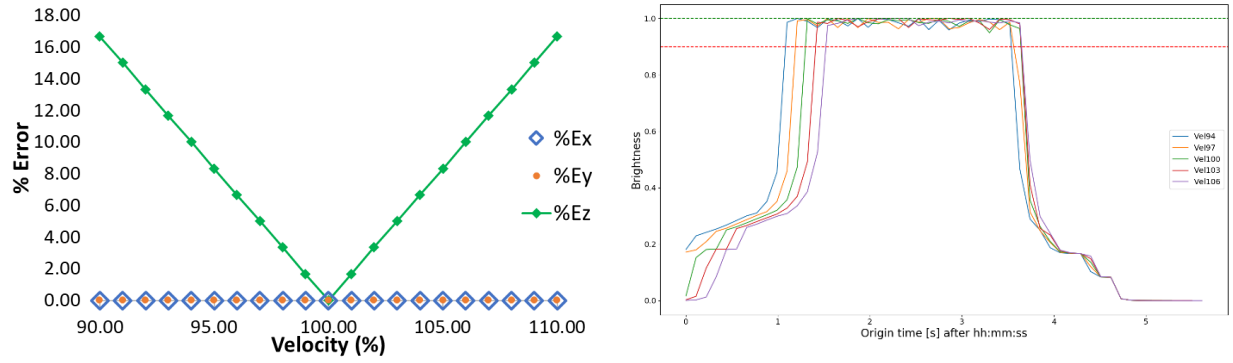


Figure 36. a) Percentage error of spatial coordinates vs Velocity, b) MBCs vs Velocity.

For a centered event in a circular array (**figure 37**), the location is perfectly constrained laterally, and the source deepens with velocity increments. Like the square array, this configuration is recommended for spatial location, although the origin time may require some extra analysis for more precise calculations.

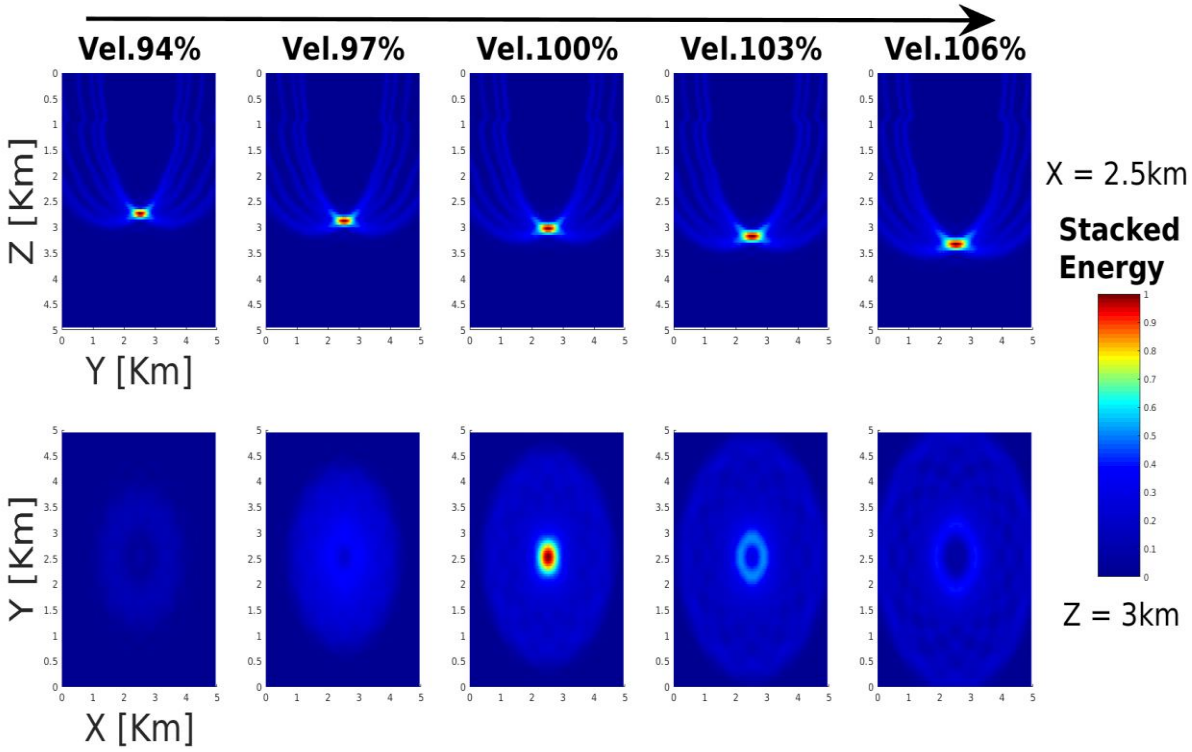
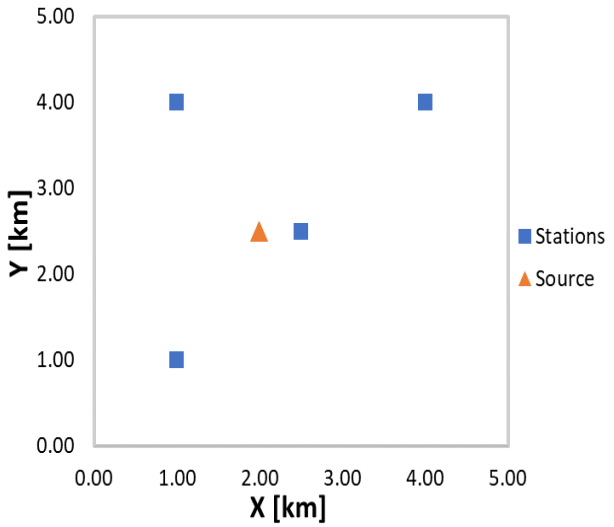


Figure 37. Stacked energy and event location for a 12-receiver-circular array.

3.7.4. Irregular case. The irregular array in **figure 38** resembles one used during a microseismicity monitoring operation in Colombia, accordingly, it is essential to understand the effects on location induced by such recording geometry.



Item	X	Y	Z
Station0	1.0	1.0	0
Station1	1.0	4.0	0
Station2	2.5	2.5	0
Station3	4.0	4.0	0
Source	2.0	2.5	3.0

Figure 38. Irregular sensor configuration.

This geometry produced relatively low errors in X and Y (<2%) and a 10% error in Z, for velocities within an ±5% interval around the true value, beyond this point, results are more complicated to interpret, however, Z is the coordinate with the highest error of the 3 (**figure 39a)**. As in the case in **subsection 3.7.2**, a clear maximum brightness value is reached, and an origin time estimate can be easily made (**figure 39b**).

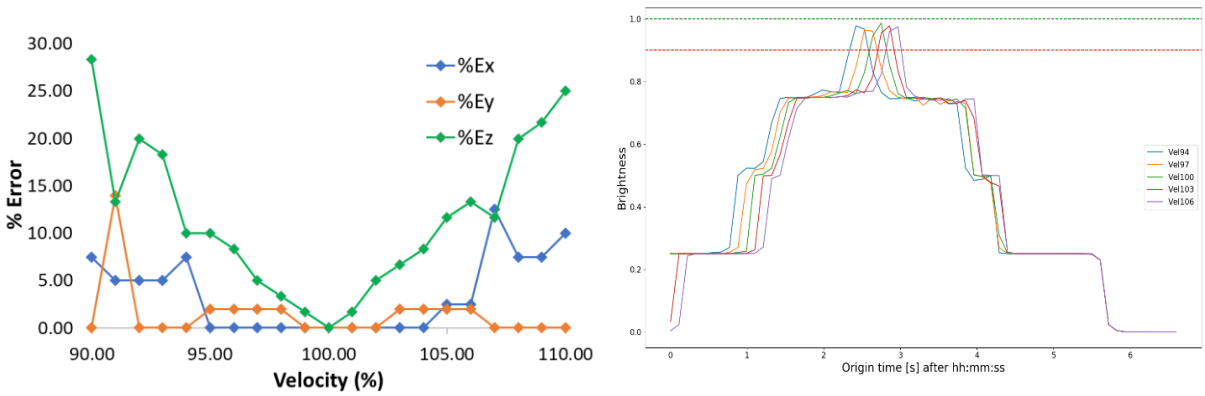


Figure 39. Accuracy of results vs velocity variations. A) Error in spatial coordinates, b) maximum brightness curves exhibiting clear zeniths.

Plotting X and Z slices of the volume (**figure 40**), illustrates how well the event is recovered in depth (upper panel), and poorly constrained laterally under high velocity errors (lower panel).

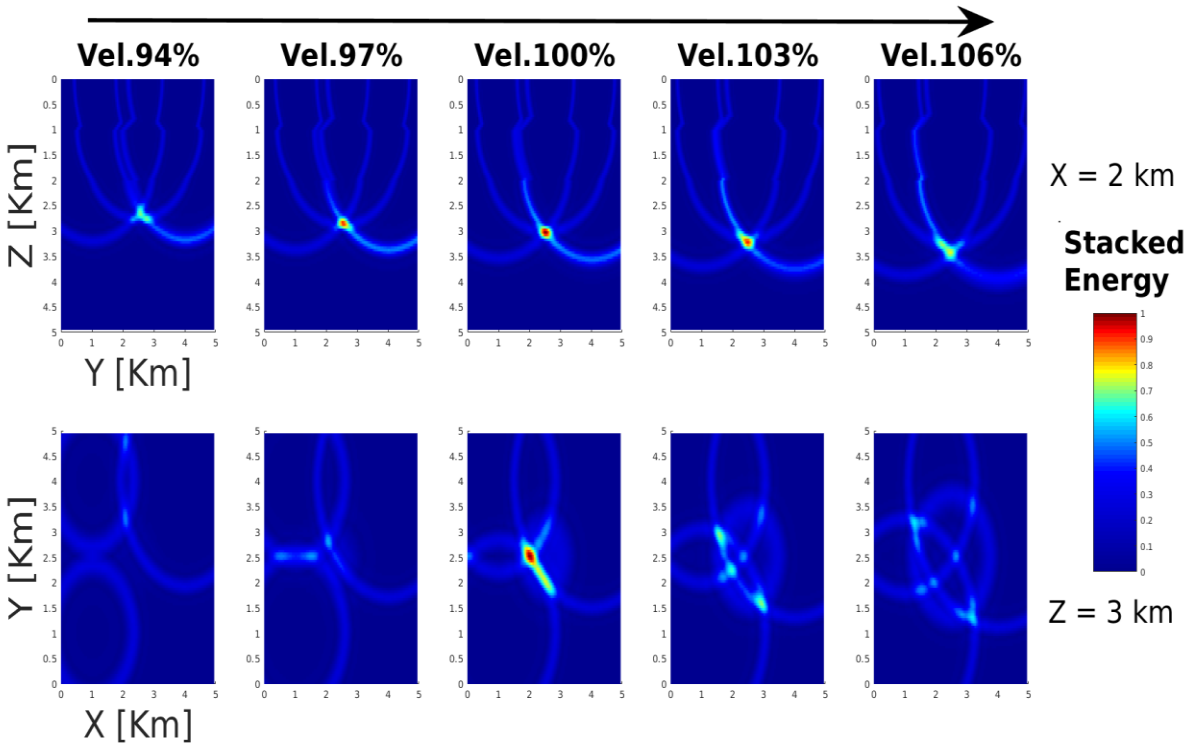


Figure 40. Stacked energy and event location for an irregular geometry.

3.8. **Others**

Combinations of the above factors and other scenarios are presented in this section.

3.8.1. Non-centered square array + Variations in Velocity and Frequency content. This example summarizes the combined effects of frequency and velocity on stacked energy for a non-centered event on a square geometry. Keep in mind that trial velocities are presented as percentages of the correct velocity model (***vel* 100%**).

For a given frequency, lower velocities produce a shallower source while higher velocities yield deeper events. In addition, due to the asymmetry of the array-source configuration (see source 2 in figure 24), it is noticeable that the hypocenter does not deepens vertically but diagonally toward the closest station.

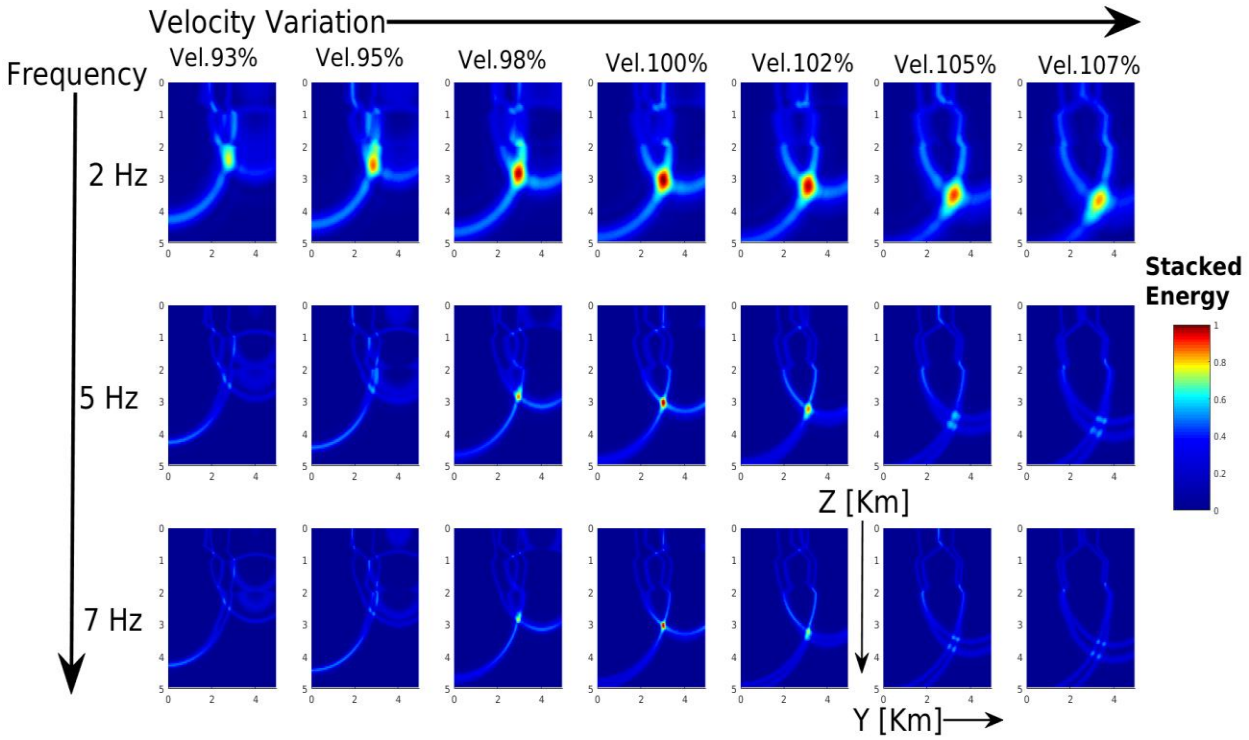


Figure 41. Non-centered event: Effects of frequency content and velocities on stacked energy. Vertical slices for 2Hz, 5Hz and 7Hz in upper, middle and lower panels, respectively.

It is evident from **figure 41** that low frequencies may produce false candidates; however, they will disappear as frequency content is higher, thus gaining in resolution. Nonetheless, a more accurate velocity model is needed for energy to focus at higher frequencies.

3.9. Chapter 3 – Part 2: Velocity Uncertainty Estimation – Statistical Analysis

One last evaluation is made to examine more deeply the importance of an accurate velocity model in passive and active seismic source location. In this experiment, velocity uncertainty of the 4 methods is assessed and compared by a Monte Carlo algorithm which helps to randomly select 100 velocity models according to *Probability Density Functions* derived from sonic logs recorded in 3 wells (WL1, WL2, WL3), VSPs of 4 wells (WL1, WL4, WL5, WL6) and other geological information such as formation tops (**figure 42**). The procedure to generate such PDFs starts by collecting the required velocity data, then using a statistical tool/software to find the best fit to the data from several distributions. Afterwards, backpropagation is performed using each of the previous velocity models (realizations) and their results are presented in histograms with their respective best fits to observe the direct effects of velocity uncertainty on location.

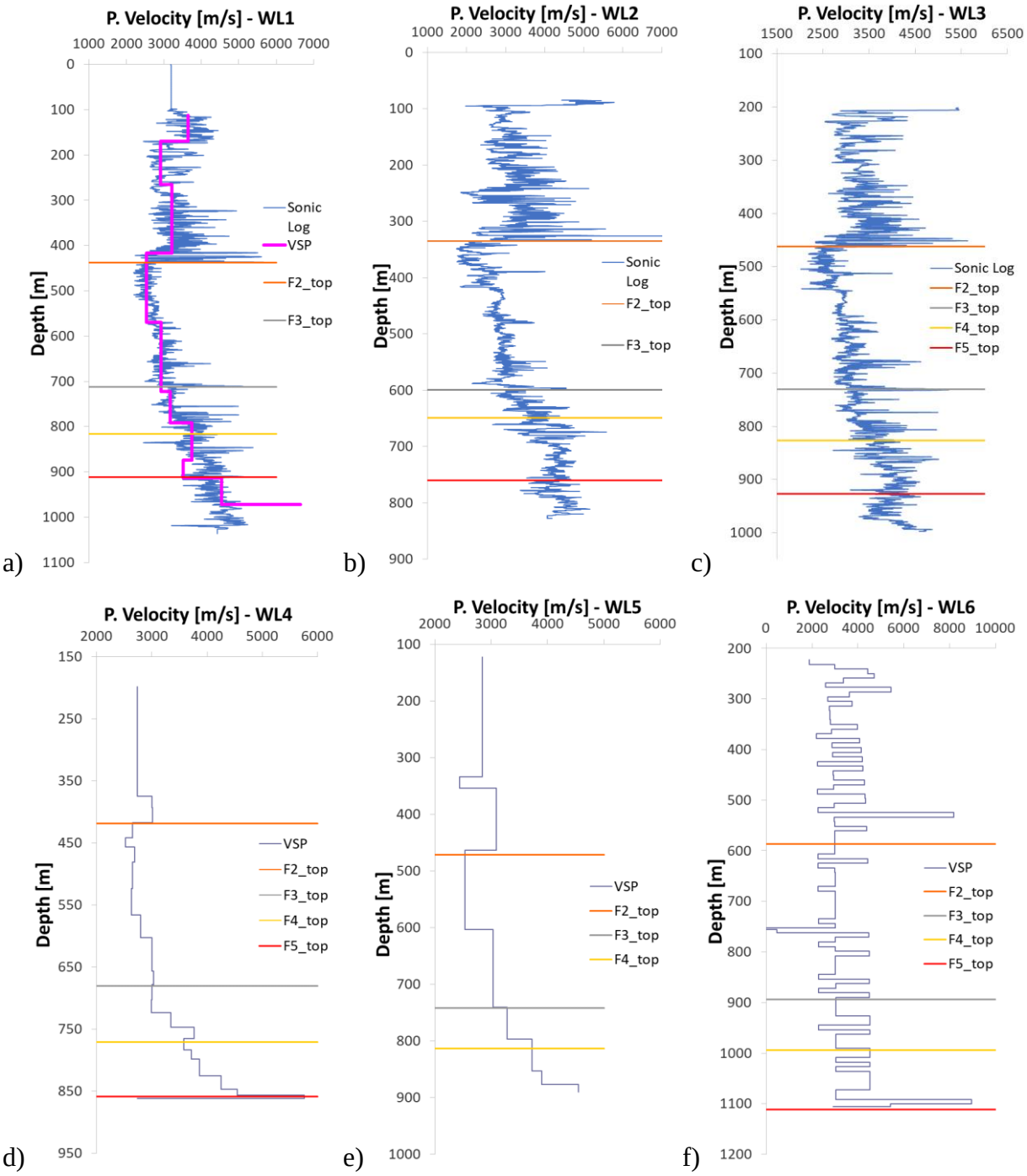


Figure 42. P-Velocity information recorded in different wells. A) WL1 – Sonic log and VSP, b) WL2 – Sonic log, c) WL3 – Sonic log, d) WL4 – VSP, e) WL5 – VSP, f) WL6 – VSP.

Note: The name of formations and wells in this section have been intentionally modified for confidentiality reasons.

3.9.1. **Analysis by formations.** Probability Density Functions (PDF), which best fit the data, are generated for each formation (layer) using a statistical function on MATLAB (called dfittool)[9], so that every random velocity model that may be selected must follow such distributions. **Figures 43-47** show a histogram for each formation and some fitting curves among more than 20 different possible distributions including: Normal, t-Location-Scale, Beta, Gamma, Rayleigh, Exponential, Logarithmic and others. In each case, a Non-parametric PDF (table 4) was chosen as the one which provided the best fit to data and subsequently, such a PDF was preferred for sampling purposes in the Monte Carlo algorithm explained later.

Formation1: The uppermost layer consists of a few thin formations of similar properties thus they are treated as a single formation for convenience. **Figure 43** is a histogram of P-velocities which exhibits a denser concentration of data toward the range of (3150-3200 m/s). A further statistical analysis shows that the mode (3195m/s) is the most adequate value to be used in the forward modelling.

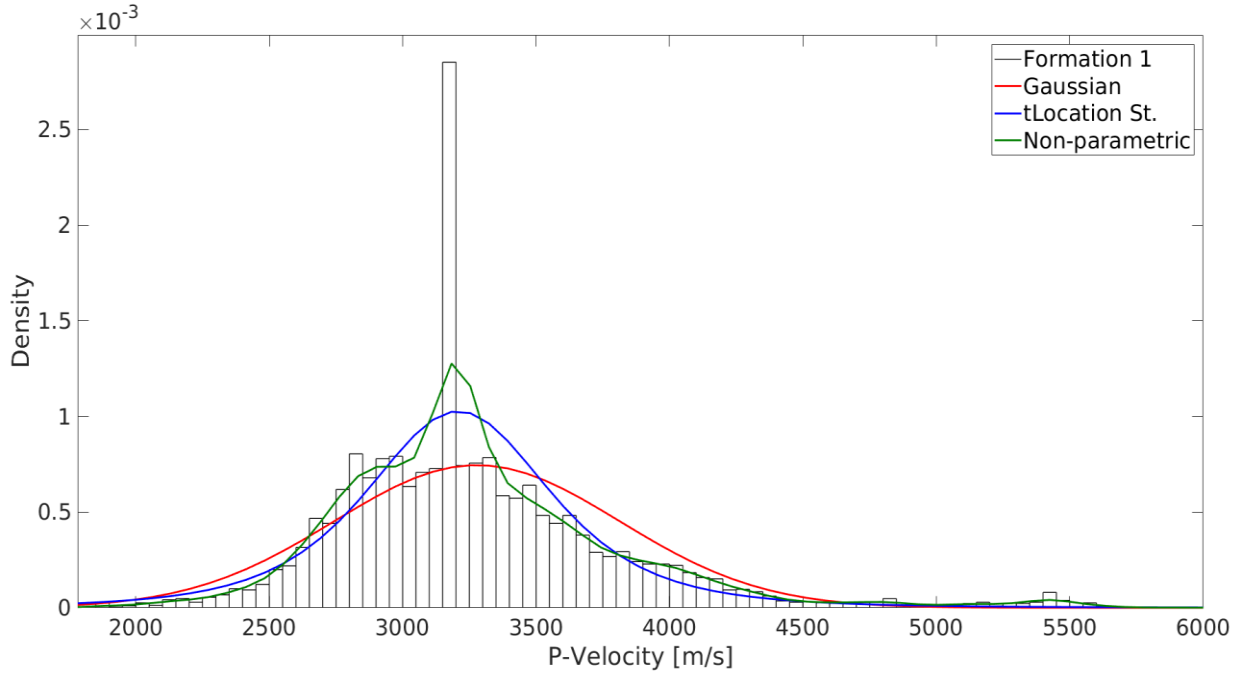


Figure 43. P-velocities of Formation1. Three distributions are displayed: Gaussian (red line), t-location scale (blue line) and non-parametric (green line).

The figure above shows a better fit with the green curve, because the non-parametric PDF has less restrictions to fit the data, whereas the normal and t-location scale distributions are more limited

[9] MATLAB and Statistics Toolbox Release 2017b, The MathWorks, Inc., Natick, Massachusetts, United States.

and do not appropriately describe the histogram. For this reason, the Non-Parametric Distribution is used for sampling purposes during the Monte Carlo analysis.

Formation2 (the second formation downward): the histogram in **figure 44** reveals that the data is not correctly represented by a normal distribution, on the other hand, a t-Location-Scale or a non-parametric distribution can provide a good fit. However, the Non-parametric distribution (green) is a somewhat better option.

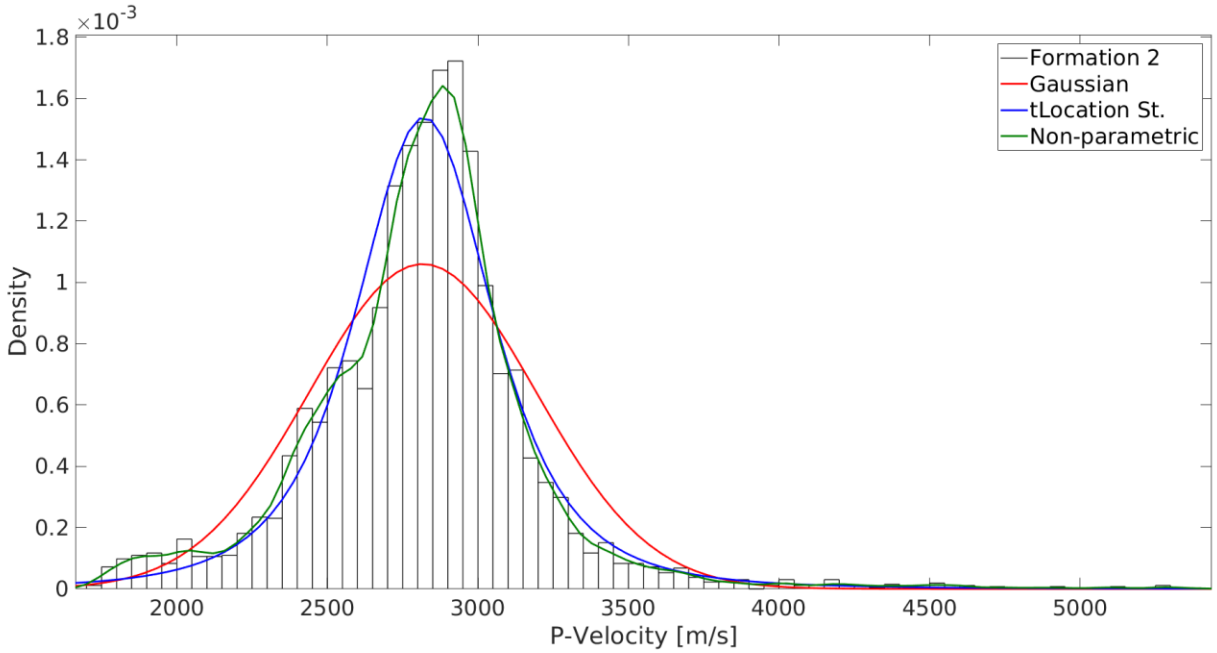


Figure 44. P-velocities of Formation2. Three distributions are presented: Gaussian (red line), t-location scale (blue line) and non-parametric (green line).

Although the mean value of such a histogram is 2818 m/s (**table 4**), the mode is preferred for the forward modelling because it is the most common value, hence, a P-velocity of 2900 m/s is chosen for this formation. **Note**: All of the 5 histograms in figures 1-5 are acceptably well fitted by a Non-parametric PDF and therefore, such distribution is used in the velocity uncertainty analysis.

Formation3: The statistical analysis of P-wave velocities for this layer shows a distribution that cannot be fitted by regular PDFs including Gaussian (red curve), t-Location-Scale (blue), Burr (fuchsia) and others not shown in **figure 45**. Nonetheless, a non-parametric fit (green) shows good agreement with the data. Then, there is difficulty in choosing an appropriate P-value between its mean (3420m/s) and its mode (3250m/s), nonetheless, a quick glance at the histogram is enough to notice that the latter is a wiser option because this value is the most frequently encountered in the sonic logs.

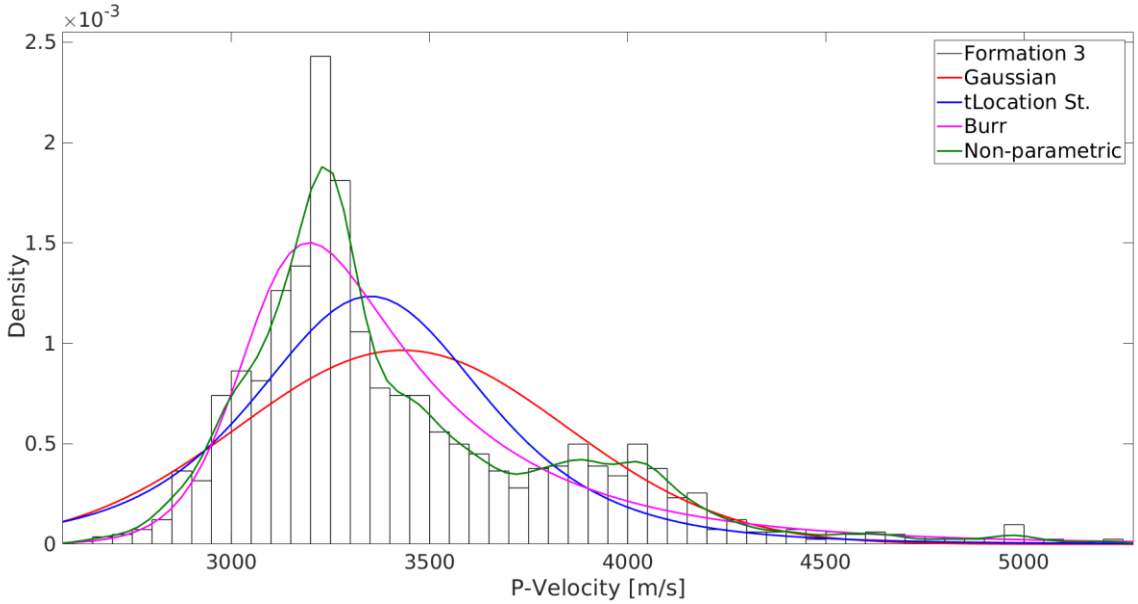


Figure 45. P-velocities of Formation3. Four distributions are presented: Gaussian (red line), t-Location-Scale (blue line), Burr (fuchsia line) and non-parametric (green line).

Formation4: The process to obtain the velocity value for the forward modelling is the same as in the previous formations (**figure 46**). Interestingly, similar curves were obtained with the normal, t-location scale and kernel (non-parametric) distributions, therefore, to be consistent with the previous selection criterion, the mode of the data (4275 m/s) is used as the acoustic wave velocity of the layer.

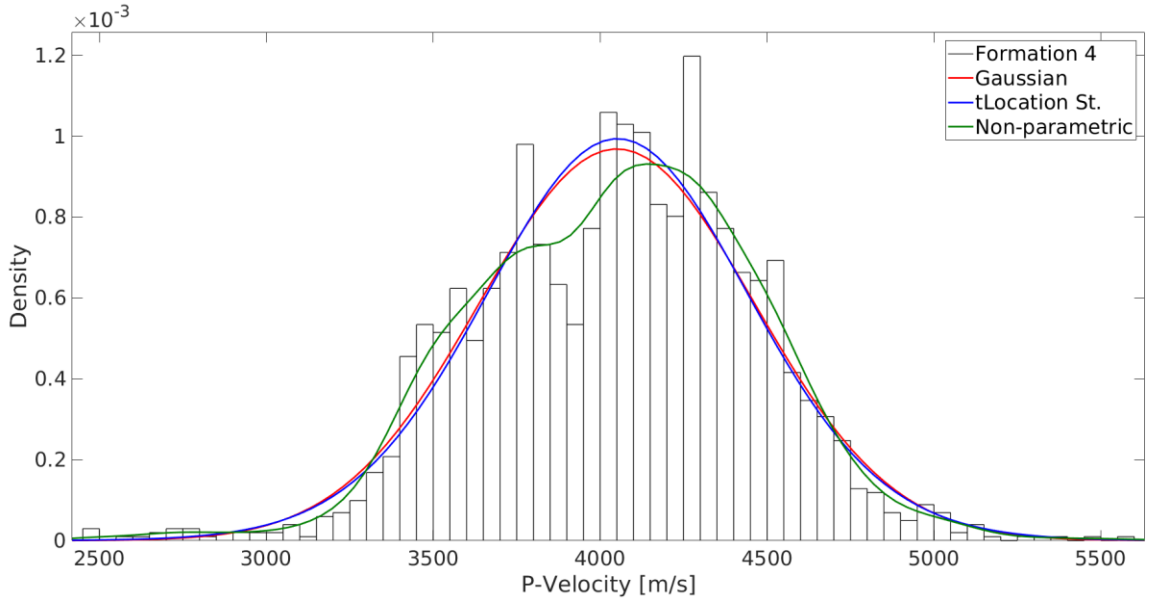


Figure 46. P-velocities of Formation4. Three distributions are presented: Gaussian (red line), t-location scale (blue line) and non-parametric (green line).

Formation5: The lowermost layer of the list is the basement in which acoustic velocities are higher than in the overlying layers, as shown in **figure 47**. At first glance, the 3 distributions fit the data fairly well and the mean value could be used as the p-velocity **(**table 4), however, further inspection helps determine that a velocity of 4430 m/s (its mode) is more suitable.

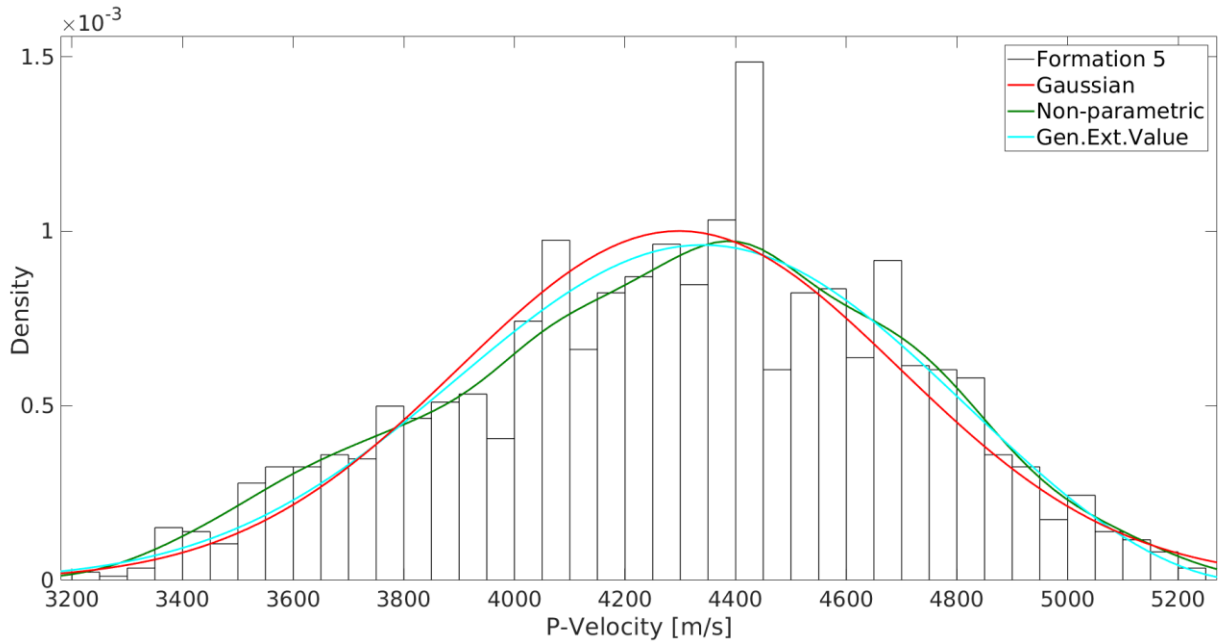


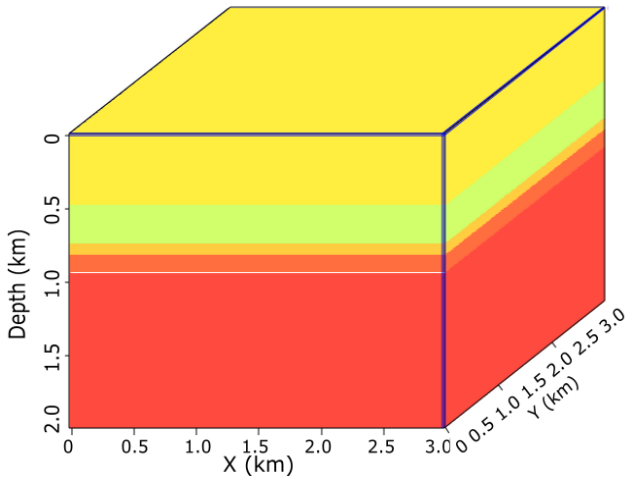
Figure 47. P-velocities of Formation5. Three distributions are presented: Gaussian (red line), t-location scale (blue line) and non-parametric (green line).

Table 4. Non-Parametric Distributions used in the Monte Carlo Analysis.

Formation1	**Formation2**	**Formation3**	**Formation4**	**Formation5**
Bandwidth = 76.9 Estimated Mean=3270m/s	Bandwidth = 49.1 Estimated Mean=2818m/s	Bandwidth = 50.0 Estimated Mean=3420m/s	Bandwidth = 99.2 Estimated Mean=4050m/s	Bandwidth = 98.5 Estimated Mean=4298m/s
PDF[10] = *Non-parametric – Gaussian Kernel*		$f_h(x) = \frac{1}{n\,h}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$		

[10] Any question regarding the dfittool function, PDF or their formulas can be found on Mathworks official website: https://la.mathworks.com/help/stats/

3.9.2. **Earth model.** The volume of study is a rectangular prism whose dimensions are X = 3km, Y= 3km, Z= 2km and a total of 100 gridpoints along each axis with a resolution of 30m for X, Y and 20m for Z. Forward modelling and backpropagation were carried out on a horizontally layered velocity model with representative velocity values and formation tops (**figure 48**). Assumptions for the propagation and backpropagation are the same as in section 3.3.



	Formation Top [m]	**Average P-Velocity [m/s]**
Form.1	0	3195
Form.2	460	2900
Form.3	720	3250
Form.4	820	4275
Form.5	920	4430
Dimensions: X= 3km, Y= 3km, Z = 2km		

Figure 48. Layer cake model derived from sonic log and VSP data.

Then, performance of 4 location methods[11] are tested on two arrays of stations: a square geometry with a non-centered event source and an irregular geometry comprised of 7 sensors with a non-centered event. These two geometries have been intentionally selected since each of them favors different kinds of methods.

3.9.3. **Comparison of BPI methods.** The model space, which is composed of 5 parameters (V_1, … , V_5), is randomly sampled 100 times according to the *PDF* for each layer (**table 4**), and each realization (velocity model) is used in backpropagation to locate the event and study the effect of such distributions on the outcome. Results of the spatiotemporal locations obtained by the 4 methods (MATF, SSA, PbAS, PrAS) are plotted in histograms to facilitate their analysis.

[11] These methods are briefly explained in Chapter 2 and Appendix (C).

Square geometry. Square geometries are commonly in seismology used for recording arrays, therefore, it is interesting to evaluate the location methods with this 4-station configuration (**figure 49a**); besides, this geometry generates a maximum brightness curve with a large plateau (**figure 49b**) that may cause trouble to those algorithms that need a clear maximum peak of energy to localize the source while the centroid-based methods should perform without inconveniences.

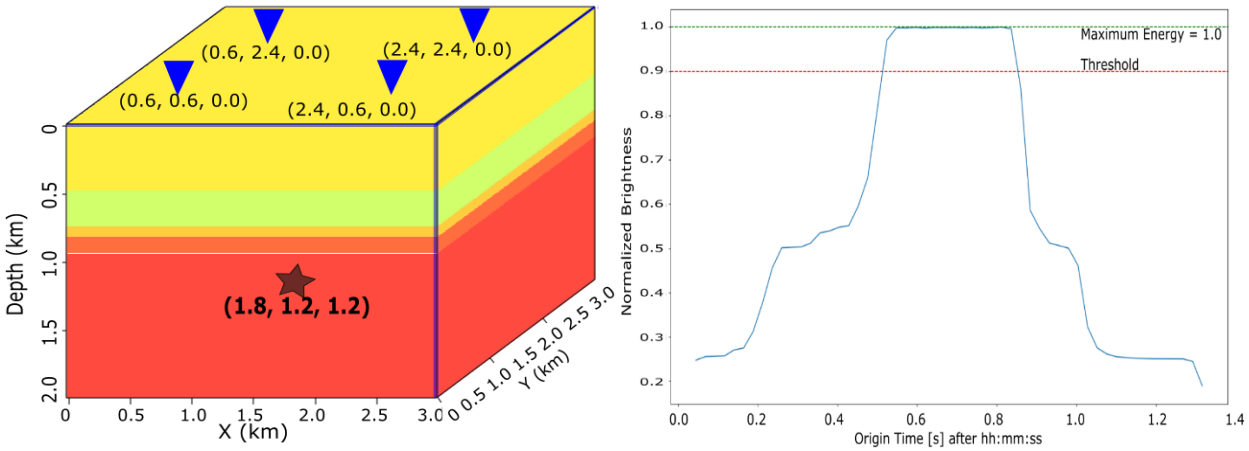


Figure 49. A) Square Geometry + true event location. B) Normalized MBC. This curve plots the maximum energy value per timestep and allows to visually identify events as coherent summation of energy (signals).

Results from this study are presented in **figures 50-58.** A color code is used to facilitate comparison of the methods: red is used when referring to the Maximum Amplitude Time Function Method (MATF), black represents the Source Scanning Algorithm (SSA), blue for the Brightness PDF-based Amplitude Stacking (PbAS) method and light green for the Spatial PDF-based Amplitude Stacking (PrAS) method. Finally, both the T.Centroid and T.Peak methods are used for origin time determination based on the calculation of the centroid: the former (in fuchsia) uses the Maximum Brightness Curve (MBC) for this purpose while the latter (in yellow) uses a different version of such a curve, powered by its standard deviation, in order to enhance the largest arrivals as well as reduce noise (T.Peak).

Time determination.

The 4 location methods can be divided into 2 categories depending upon the approach employed to compute the time of occurrence t_0. In the first category, MATF and SSA, the time of

occurrence is selected from the MBC as the time when the maximum peak of energy is reached (assuming a single event happens); in other words, t_0 is found by grid-searching for the time with the maximum focused energy. The second category, T.Centroid and T.Peak, makes use of the MBC "$\max_{\mathbf{r}}\left(F(\boldsymbol{r},t)\right)$" to calculate the temporal centroid of the energy. One of the advantages of computing centroids is that t_0 is not restricted by the discretization, however, poor acquisition conditions, low signal to noise ratio and other factors may degrade the solution and give erroneous time values.

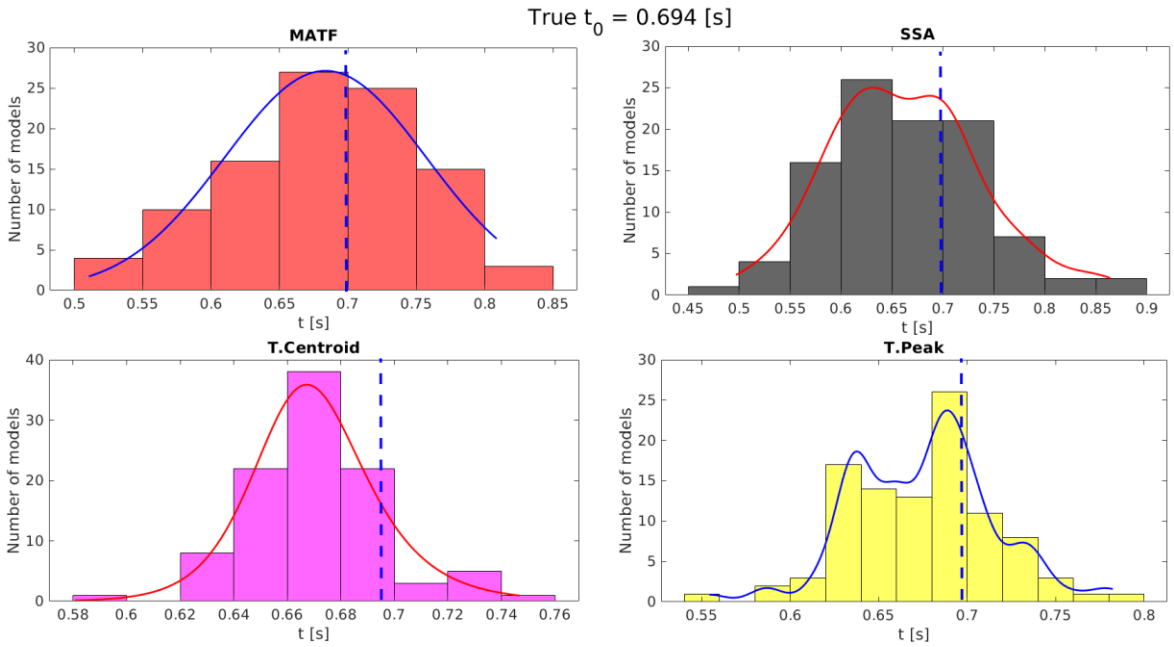


Figure 50. Origin time determination by 4 methods. MATF (red bins), SSA (black bins), T.Centroid (fuchsia bins) and T.Peak (yellow bins).

As previously mentioned, the MATF and SSA methods are restricted to the temporal discretization (timestep = 0.08 [s] in this case). The upper left plot in **figure 50** shows a histogram of the results obtained by MATF; such data is fitted with a Gaussian function obtaining a mean value of 0.684[s] and *Root Mean Square Error (RMSE)* of 0.0739[s] (see **table 5**). The upper right plot shows the results of backpropagating with SSA, after curve fitting, a Kernel regression provided the best fit to the data with an estimated mean value and RMSE of 0.663[s] and 0.0803[s], respectively. In the second category, the temporal-centroid (lower left corner) was fitted with a Burr distribution with a mean of 0.671[s] and 0.035[s] RMSE. Finally, a Kernel distribution is used to fit data of the T.Peak approach, giving a mean of 0.675[s] and 0.0440[s] RMSE.

Table 5. Origin time statistical parameters for square geometry.

True Origin Time (t_0 = 0.694 [s])		
Method – PDF	**Parameter**	**RMSE with respect to (t_0)**
MATF – Normal Distribution.	Mean (μ) = 0.684 Variance = 0.005 σ = 0.074	0.0739
SSA – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.025 Estimated mean (μ) = 0.663	0.0803
T.Centroid – Burr Distribution.	Mean = 0.671 Variance = 0.0007 α = 0.665 c = 51.683 k = 0.794	0.0350
T.Peak – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.009 Estimated mean (μ) = 0.675	0.0440

Notice that bin widths in histograms are not the same which may be misleading. For instance, MATF seems to have achieved the lowest error in t_0 based on the figure only, but that is not true because different bin widths were used for each distribution. In fact, T.Peak and T.Centroid generated better results, in a narrower range and with a RMSE about half that of the MATF and SSA. In conclusion, all methods produced a good estimate of the time of occurrence, however MATF and SSA might result unreliable due to discretization dependency and complex brightness curves.

Space.

Results of the spatial location are displayed in **figures 51-53**. The histogram in the upper left corner in **figure 51** shows the outcome of the MATF method for the X coordinate and the best fit that could be obtained among the available options. A kernel distribution gives a mean value of 1.793 [km] with a 0.0794[km] RMSE, nonetheless, it is important to point out that the data are not well centered around the mean. A similar situation is observed with SSA (upper right corner) for which the data is distributed in the range of 1.65-1.95 [km]; moreover, the mean and its error are

estimated from a Kernel distribution and presented in **table 6**. Besides MATF and SSA, two spatial-centroid methods are introduced and tested for a robustness analysis. The lower panel shows that both methods (PbAS and PrAS) produce remarkably similar results with RSME of 0.0269[km] and 0.0274[km], respectively, and their datasets are well fitted with Generalized Extreme Value functions as witnessed in the plots below.

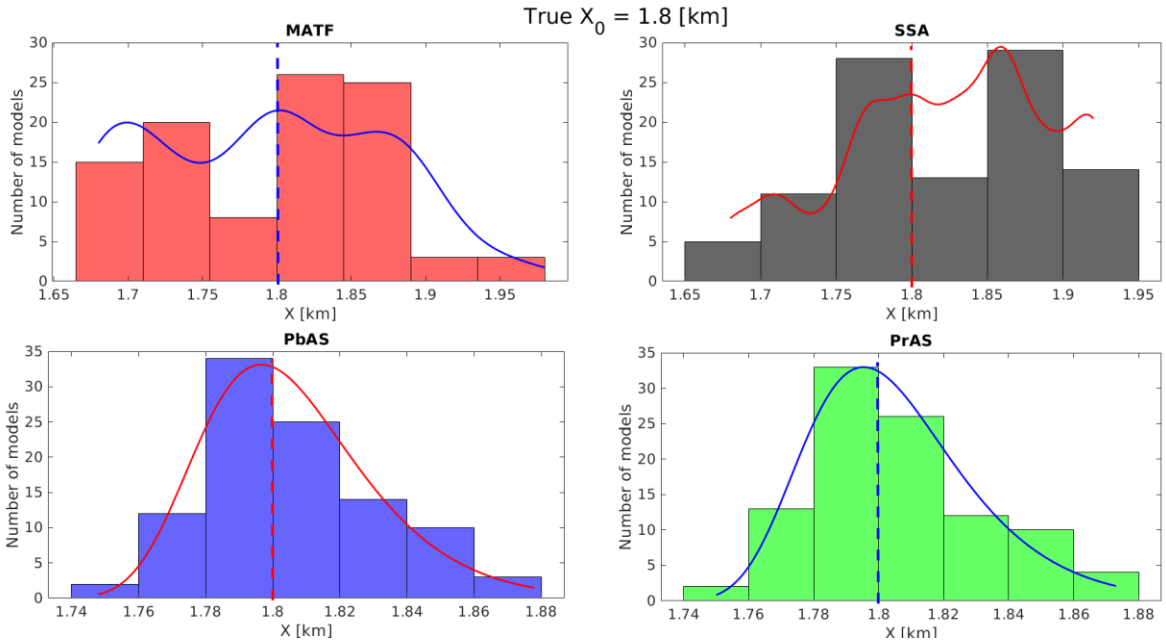


Figure 51. X coordinates using 4 location methods. MATF (red bins), SSA (black bins), PbAS (blue bins) and PrAS (light green bins).

Table 6. Statistical parameters for X-coordinate of square geometry.

Method – PDF	**Parameter**	**RMS Error with respect to X_0=1.8 [km]**
MATF – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.015 Estimated mean (μ) = 1.793	0.0794
SSA – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.015 Estimated mean (μ) = 1.822	0.0720
PbAS – Generalized	Mean = 1.806 Variance = 0.0007	0.0269

Extreme Value Distribution.	σ = 0.022 k = - 0.083 μ = 1.795	
PrAS – Generalized Extreme Value Distribution.	Mean = 1.806 Variance = 0.0007 σ = 0.0224 k = - 0.060 μ = 1.794	0.0274

Now, **figure 52** shows the outcome of the same 4 location methods for the Y-coordinate and their respective fitting curves. In the upper left corner, a Kernel distribution is used to fit the MATF data producing a mean value of 1.207 [km] which represents a 0.0794[km] RMSE in relation to the true value, nevertheless, it can be noticed that the curve poorly fits the data which is dispersed in a large range [1 -1.4] [km], thus, MATF does not produce a clear tendency (**table 7**). Second, SSA generates results with a mean of 1.177 [km] and 0.0720[km] RMSE. Notice that its distribution exhibits a clearer tendency. Finally, PbAS and PrAS are comparable with all data in the range of [1.12 – 1.26 km] and RMSE of 0.0269[km] and 0.0273[km]. Again, the centroid-based methods produce more consistent and reliable outcomes.

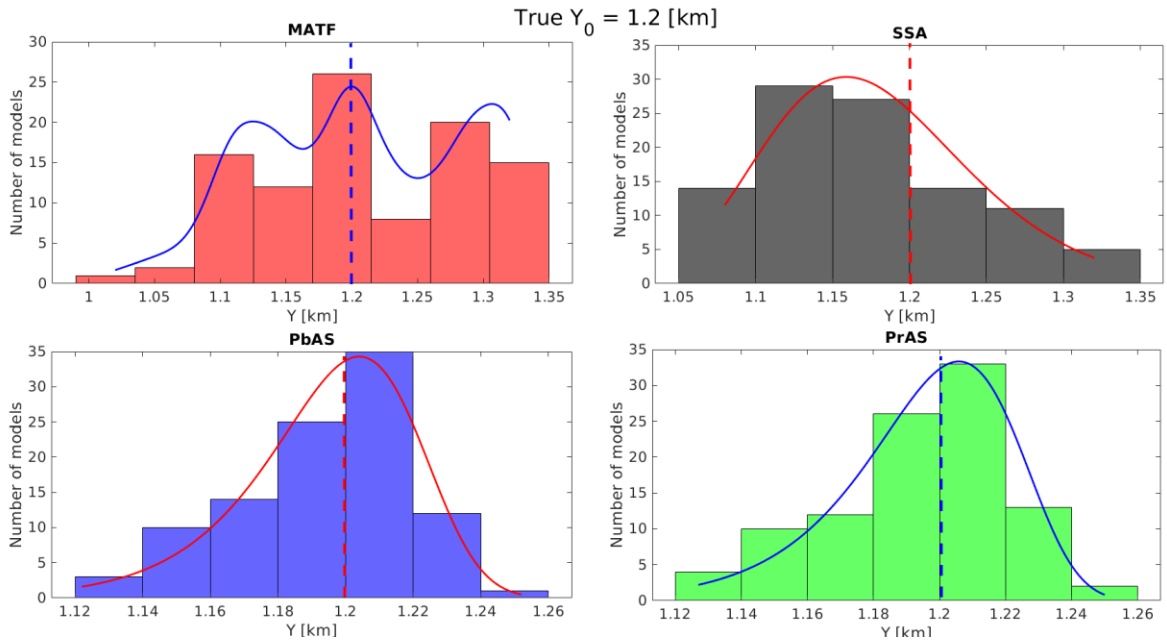


Figure 52. Y coordinate using 4 location methods. MATF (r), SSA (bk), PbAS (bl) & PrAS (g).

Table 7. Statistical parameters for Y-coordinate of square geometry.

Method – PDF	**Parameter**	**RMS Error Y_0=1.2[km]**
MATF – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.018 Estimated mean (μ) = 1.207	0.0794
SSA – Generalized Extreme Value Distribution.	Mean = 1.177 Variance = 0.0045 σ = 0.061 k = - 0.144 μ = 1.150	0.0720
PbAS – Burr Distribution.	Mean = 1.194 Variance = 0.0007 α = 1.253 c = 58.792 k = 10.390	0.0269
PrAS – Burr Distribution.	Mean = 1.195 Variance = 0.0008 α = 1.281 c = 55.619 k = 27.949	0.0273

Lastly, **figure 53** shows the behavior of depth (Z) for the 4 methods. MATF y SSA (upper panel) have RMSE of 0.5227[km] and 0.4568[km], respectively, it is appreciable from the plots that they do not provide appropriate constraints in the vertical direction because some of the events were located at surface or at the bottom of the model. In contrast, PbAS and PrAS compute depths closer to the true value (1.2 km) with lower errors (0.1069[km] and 0.1111[km]) (see **table 8**).

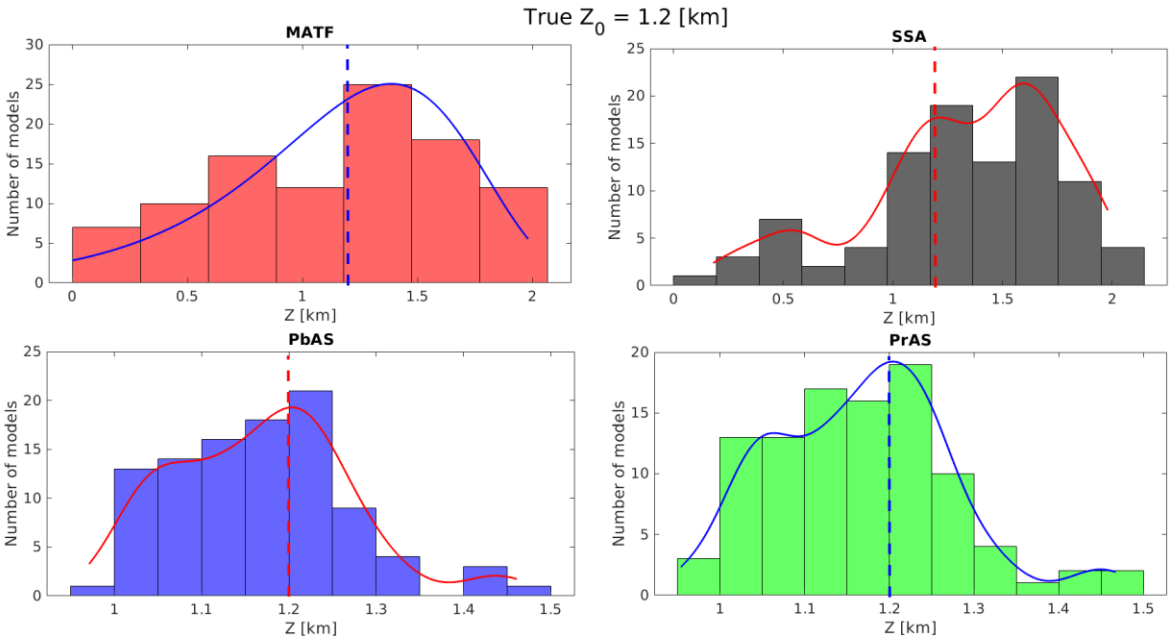


Figure 53. Depth (Z) using 4 location methods. MATF (r), SSA (bk), P_bAS (bl) & P_rAS (g).

Table 8. Statistical parameters for the Z-coordinate of square geometry.

Method – PDF	**Parameter**	**RMS Error (Z_0 = 1.2[km])**
MATF – Extreme Value Distribution.	Mean = 1.130 Variance = 0.319 μ = 1.384 σ = 0.441	0.5227
SSA – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.130 Estimated mean (μ) = 1.320	0.4568
PbAS – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.035 Estimated mean (μ) = 1.168	0.1069
PrAS – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.030 Estimated mean (μ) = 1.168	0.1111

Some of the disadvantages of MATF and SSA is that both the spatial and temporal location are restricted to the grid nodes (discretization); additionally, they need a peak of maximum energy in the maximum brightness curve to work at their best. It is important to keep in mind that this first geometry was intentionally tested because it generates a brightness curve with a large plateau (**figure 49b**), which may make MATF and SSA experience difficulties whereas PbAS and PrAS are able to handle it relatively easy.

Irregular geometry. Now, the experiment is carried out on an irregular geometry according to **figure 54a**. This second case aims to evaluate the performance of the above-mentioned methods under a different geometry of receivers which generates a minor maximum in the brightness curve (**figure 54b**) enough, in principle, to make MATF and SSA more suitable in principle than PbAS and PrAS.

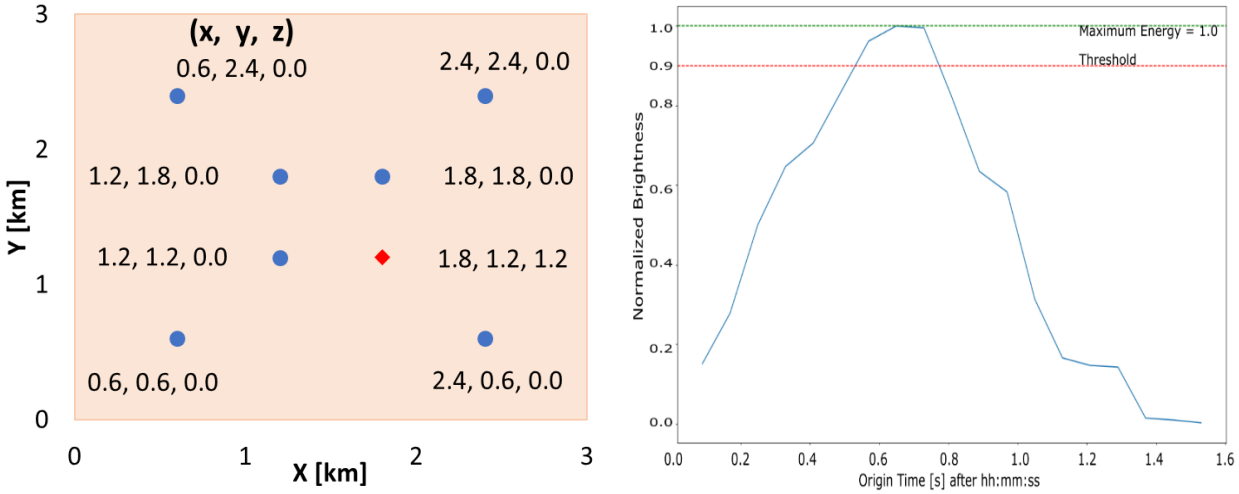


Figure 54. Irregular Geometry: a) View map and coordinates, b) MBC.

Time.

In this opportunity, the temporal distributions for MATF and SSA (**figure 55**) seem to be more densely concentrated around the true value (0.694 [s]) than in the previous case (**figure 50**) and their statistical parameters can be seen in **table 9**. On the other hand, T.Centroid and T.Peak experienced an increase of approximately 0.02[s] in the error with respect to the other geometry. This indicates that when it comes to origin time determination, MATF and SSA produce similar results to T.Centroid and T.Peak provided that a peak of energy exists on the maximum brightness curve.

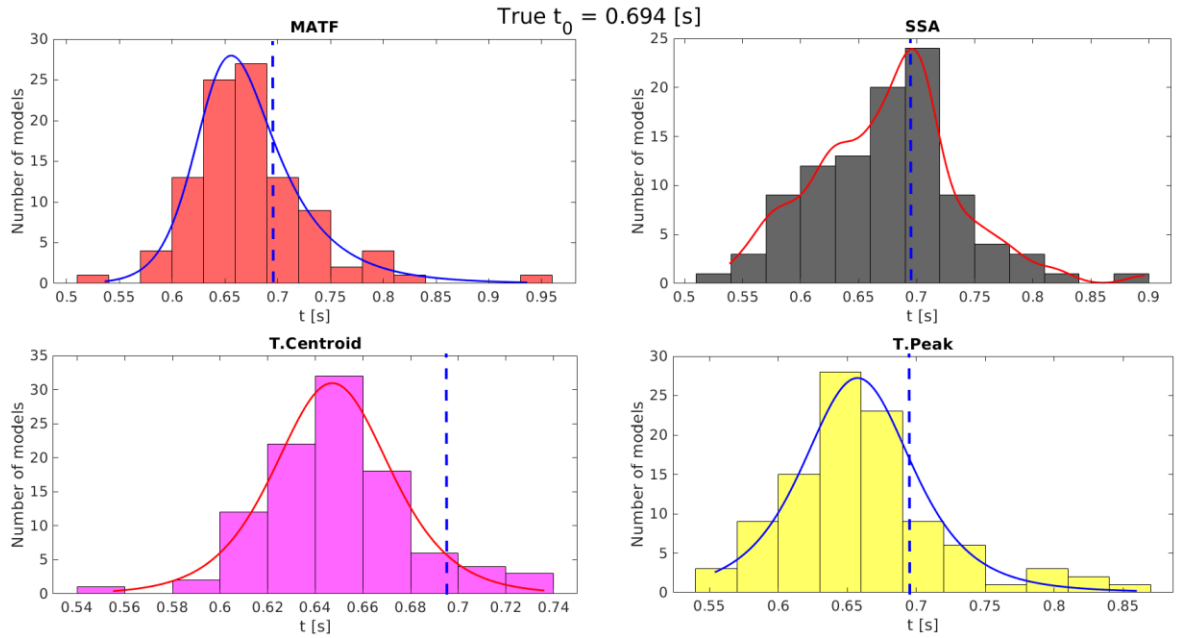


Figure 55. Origin time determination by 4 methods. MATF (red bins), SSA (black bins), T.Centroid (fuchsia bins) and T.Peak (yellow bins).

Table 9. Origin time statistical parameters for irregular geometry.

Method – PDF	**Parameter**	**RMS Error with respect to (t_0=0.694[s])**
MATF – Burr Distribution.	Mean = 0.675 Variance = 0.003 α = 0.644 c = 31.970 k = 0.495	0.0607
SSA – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.015 Estimated mean (μ) = 0.674	0.0654
T.Centroid – Logistic Distribution.	Mean (μ) = 0.647 Variance = 0.009 σ = 0.162	0.0542

T.Peak – tLocation Student Distribution.	Mean (μ) = 0.657 Variance = 0.0035 σ = 0.041 ν = 3.948	0.0637

Space.

Placing extra recording stations transform the formerly large plateau (**figure 49b**) into a slight peak on the MBC (**figure 54b**) that, despite its moderate shape, it makes a huge difference in the outcome of the location methods. For instance, the upper panel of **figure 56** shows the results of MATF and SSA for the X coordinate; it can be seen that most models (60 out of 100) give a value of X= 1.8km which agrees with the true one, besides, the other 40 events are located in the grid nodes right next to the solution (ΔX = 0.03 km). In contrast, PbAS and PrAS (lower panel) experienced a minor decrease in accuracy as their results were overestimated (**table 10**).

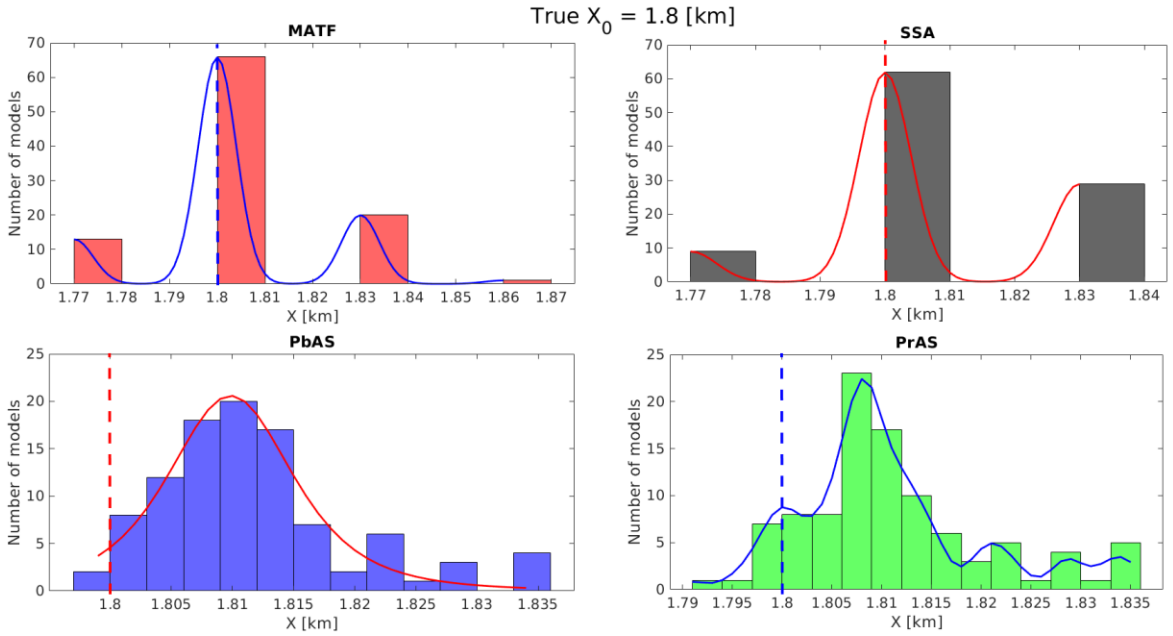


Figure 56. X coordinates using 4 location methods. MATF (red bins), SSA (black bins), PbAS (blue bins) and PrAS (light green bins).

Table 10. Statistical parameters for the X coordinate of the irregular geometry.

Method – PDF	**Parameter**	**RMS Error with respect to X_0=1.8 [km]**
MATF – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.004 Estimated mean (μ) = 1.803	0.0182
SSA – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.004 Estimated mean (μ) = 1.806	0.0185
PbAS – tLocation Student Distribution.	Mean (μ) = 1.801 Variance = 0.00008 σ = 0.005385 ν = 3.22556	0.0136
PrAS – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.0015 Estimated mean (μ) = 1.811	0.0142

Similar distributions were obtained for the Y coordinate (**figure 57**). MATF and SSA (upper panel) exhibit exactly the same behaviour as for the X Coordinate, distributions comprised of only 3 bins and most data concentrated on the central bin right at the solution (Y = 1.2 km). On the other hand, PbAS and PrAS (lower panel) underestimate the Y coordinate in most cases, although their RMSE of less 0.0136 indicate high accuracy (**table 11**).

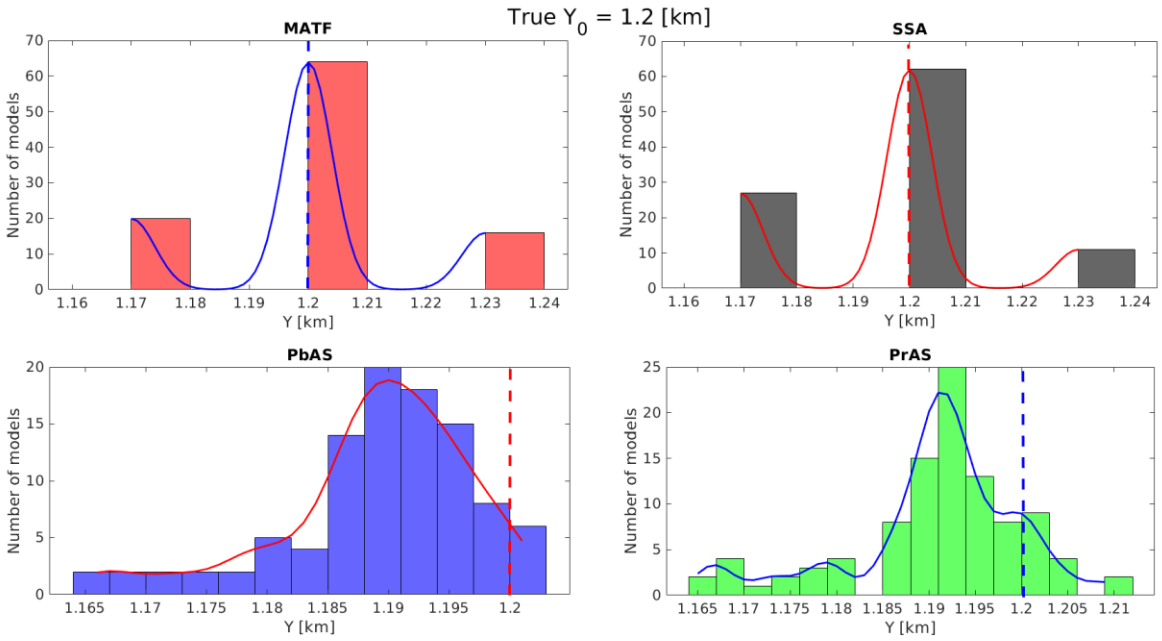


Figure 57. Y coordinate using 4 location methods. MATF (r), SSA (bk), PbAS (bl) & PrAS (g).

Table 11. Statistical parameters for the Y coordinate of the irregular geometry.

Method – PDF	**Parameter**	**RMS Error with respect to Y_0=1.2 [km]**
MATF – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.004 Estimated mean (μ) = 1.199	0.0180
SSA – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.004 Estimated mean (μ) = 1.195	0.0185
PbAS – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.0025 Estimated mean (μ) = 1.189	0.0136
PrAS – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.0018 Estimated mean (μ) = 1.191	0.0133

Lastly, the outcomes for the Z coordinate (depth) are way more intriguing compared to the ones obtained with the square array (**figure 53**) in which MATF and SSA completely failed to estimate

the depth of the event. Alternatively, this second geometry helps MATF and SSA to overcome some of those issue and to produce better results (0.1868[km] and 0.2103[km] RMSE) with a marked tendency toward the true value (**figure 58**). On the other hand, notice that neither PbAS nor PrAS give good estimates, in fact, their RMSE have increased up to 0.1839[km] and 0.1840[km] (**table 12**). A more thorough observation suggests that the overestimation in depth most methods make is consistent with an earlier origin time estimated by the same methods, which might imply a tradeoff between depth and time.

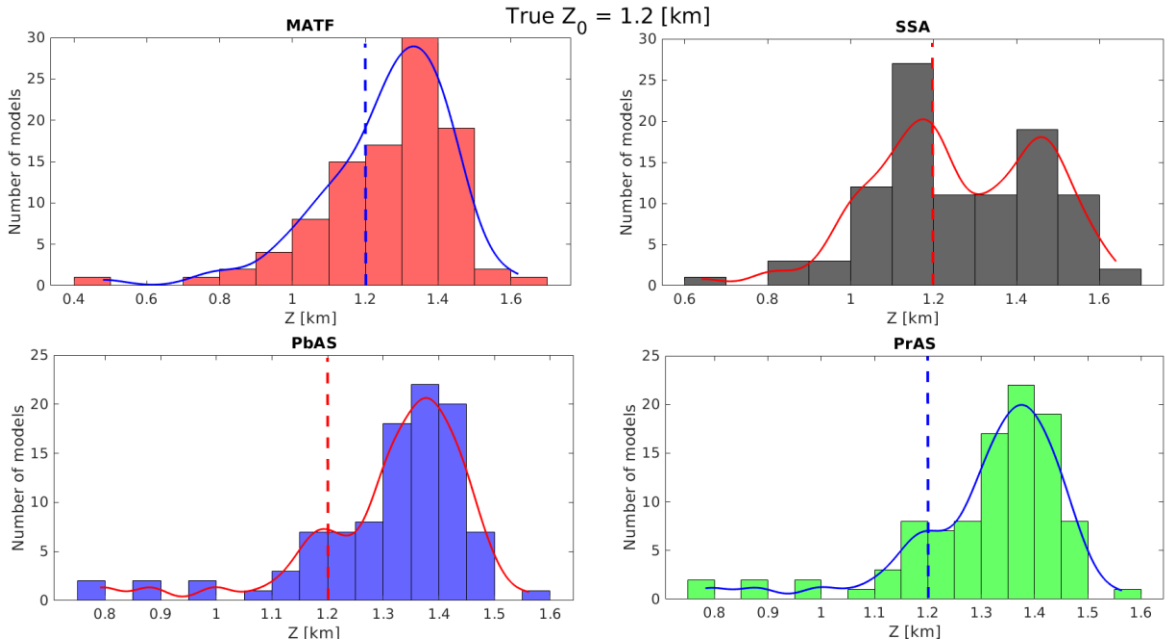


Figure 58. Depth (Z) using 4 location methods. MATF (r), SSA (bk), PbAS (bl) & PrAS (g).

Table 12. Statistical parameters for the Z coordinate of the irregular geometry.

Method – PDF	**Parameter**	**RMS Error Z_0=1.2 [km]**
MATF – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.063 Estimated mean (μ) = 1.257	0.1868
SSA – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.05 Estimated mean (μ) = 1.258	0.2103

PbAS – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.030 Estimated mean (μ) = 1.313	0.1839
PrAS – Kernel Normal Distribution.	Non-parametric Bandwidth = 0.035 Estimated mean (μ) = 1.312	0.1840

3.10. **Discussion &Conclusions**

T.Centroid and T.Peak are the most accurate for t_0 computation; MATF and SSA, on the other hand, are less stable and limited by discretization which implies a tradeoff between accuracy and computational load as can be seen in **figure 59**. The computational domain comprises the number of points in the 3D spatial domain multiplied by the number of timesteps (3) and receivers (4). Therefore, it can sometimes become extremely large.

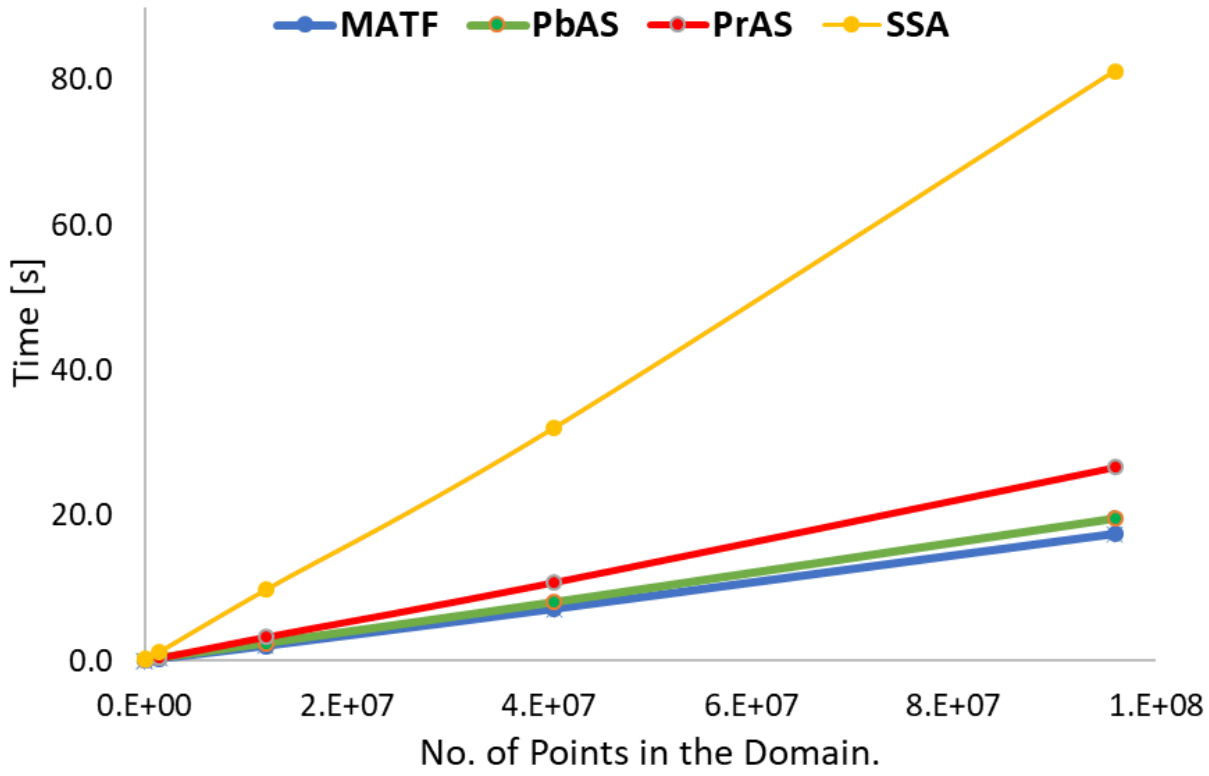


Figure 59. Comparison of the computational cost of the 4 location methods.

This figure shows that SSA takes 4 – 5 times[12] as long as the others to locate a given seismic source; the reason being that SSA incorporates a time window with 2M number of points (Equation 2.2) to account for velocity uncertainty, which translates into elevated computation time. On the other hand, it may seem contradictory that MATF exhibits the highest efficiency of the 4 methods, despite our previous claim. However, remember that MATF requires a large domain to produce acceptable results, while PbAS and PrAS can estimate similar solutions with considerably fewer points (coarse grid-spacing).

The existence of a clear peak on the maximum brightness curve will enhance the performance of location algorithms that are based on grid-searching for the maximum of energy, including MATF and SSA; nevertheless, inadequate timestep and spatial discretization might produce erroneous locations.

Sometimes, the different methods may not produce clear results or give several candidates for the location, therefore, *a-priori information* would be useful to choose the most likely source and discard all other candidates.

Estimating t_0 with the maximum brightness curve is generally a good approximation to the solution, however, some conditions may give rise to complex brightness curves making them hard to use or producing unreliable results.

Switching between different frequency ranges of the signal did not affect the location of the most energetic grid-point. However, higher frequencies require more reliable velocity models, whereas lower frequencies can focus more energy in wider range of velocities; consequently, yielding a larger uncertainty in location associated to velocity variations.

Keep in mind that all 100 scenarios where idealized to evaluate the isolated effects of velocity models; nonetheless, if more complex earth models, lower levels of SNR or unusual geometries are simultaneously studied, a centroid-based method (PbAS & PrAS) may provide better results.

Results show that large velocity uncertainty correlates to location uncertainty, however, some methods are more susceptible to velocity changes than others. For instance, an incorrect velocity value displaces the point of maximum energy from the true position, nonetheless, the centroid of the energy may remain unchanged or slightly affected.

[12] The experiments were performed on an Intel Core i7-6500U CPU @ 2.50GHz x 4 and 12 Gb RAM running Ubuntu Linux 16.04.

3.11. **Main Contributions**

- We have studied the effects of parameters such as SNR, frequency, geometry and velocity model on the outcome of Back-Projection Imaging based location methods.
- We have implemented a Monte Carlo algorithm to generate random realizations according to previously computed PDFs; later, 4 location methods are compared using such realizations.
- We have provided a detailed workflow for the deep analysis of the impact of velocity uncertainty on seismic event location algorithms.

4. Location of tremor-like signals – Real Case

4.1. **Abstract**

Representative synthetic seismograms generated by convolution are backpropagated in a realistic 3D velocity model built from sonic logs and Vertical Seismic Profiles gathered in a Colombian oil field. Employing realistic synthetics enable to determine the capability of the algorithm to handle more complex scenarios that are closer to the conditions encountered in real life. Results support that PrAS (and consequently PbAS) computes source locations with an acceptable margin of error even when considerably incorrect velocity models are used in the backpropagation.

A Monte Carlo algorithm is implemented to evaluate the uncertainty of two location methods when using more complex synthetic data, in addition to that, two scenarios are studied, in the first case, the event was placed inside the area of stimulation during the forward modelling, in the second one, the event originates away from the volume of interest. These two case scenarios help to quantify the certainty with which an event originated in the stimulated interval is correctly located and what percentage of external events are misclassified due to incorrect velocity models, therefore, allowing to construct confidence intervals out of these tests.

Finally, 2 field datasets are backpropagated and compared to synthetic experiments to discard possibilities of such events being distant earthquakes or local noise instead of actual events induced by the stimulation. These synthetic scenarios alongside some criteria indicate a probability of approximately 60% for these 2 events to be generated by the injection of high-pressure fluids; however, solving their exact spatiotemporal locations by BPI methods is still a complicated task since each event was apparently registered only on one station.

4.2. Introduction

Microseismic monitoring has been a common practice in recent years because it helps determine the zone of the underground being stimulated. Typically, this technique locates the sources of induced seismic signals to build a map of the underground and its evolution over time. Although localizing microseismic sources may be an arduous task due to the quantity of events occurring simultaneously, the location of a single microseism is relatively simple (once identified) since they share most of the characteristics of conventional earthquakes and thus traditional location methods (i.e. methods which require picking of first arrivals and minimization of the residual between the computed and measured times) can be employed (Warpinski, 2009).

Microseismicity during hydraulic fracturing is understood as a consequence of the reservoir deformation that breaks the rocks thus releasing seismic energy in the form of large numbers of high frequency - low moment magnitude events (Mw [-3 to 3]) (Wang, Li, & Shang, 2016, p. 4) around the treatment well. Nowadays, the spatial distribution of such induced microseisms is commonly used to estimate fracture geometries and stimulated reservoir volume (SRV) even though this correlation is still under study (Kumar, Zorn, Hammack, & Harbert, 2017). In the last decade, a different type of seismic event occurring concurrently with microseismicity was identified by (Das & Zoback, 2011) after analyzing seismic datasets from the Barnett shale recorded during hydraulic fracturing operations. Such events were found to comprise coherent wave trains, typically interpreted as the superposition of Low Frequency Events (LFEs) occurring in a cascade effect because of the slow slippage along large faults, resulting in an overall duration of tens or hundreds of seconds and a predominant frequency band of 10 - 80Hz (Das & Zoback, 2013b). Last, but not least, the inherently low Signal-to-noise-Ratio (SNR) of these so-called ***Long Period, Long - Duration (LPLD)*** seismic events alongside the difficulty to distinguish between the various seismic phases, make source location a highly challenging task. As suggested by (Das & Zoback, 2013c) such events may be associated to the reservoir stimulation process; hence, understanding the nature of these phenomena and their spatial distribution, would greatly help to comprehend the relation between reactivation of fault and pre-existing fractures and permeability enhancement.

This chapter comprises several stages: first, a geological context of the zone and a brief description of the field data are presented; second, a quick methodology for synthetic data generation is explained; third, our proposed centroid-based location method (PrAS) is compared

to a classical grid-search based method (MATF) using a Monte Carlo algorithm that allows to evaluate numerous velocity models within ±25% errors, from which confident intervals for each method are derived. Finally, the field data is backpropagated and compared to some synthetic scenarios to determine if the two candidates may have been caused by the hydraulic fracturing stimulation.

4.3. Field Information

The present section encompasses information related to the oil field under study including its geographic location, stratigraphic column, and a concise characterization of the reservoir from which the hydrocarbons are extracted. Afterwards, a brief description of the hydraulic fracturing and microseismicity monitoring operations carried out on field. Finally, recorded seismograms which resemble those identified in Barnett shale as LPLD events (Das & Zoback, 2011) are analyzed, focusing mainly on two seismic events reported by the geoscientists in charge during the treatment.

4.3.1. Generalities of the area. The San Francisco field was discovered in 1985. It is located 24 km NW of the city of Neiva (Colombia) in the Upper Magdalena Valley Basin (UMVB). This field produces from a highly heterogeneous reservoir in a fractured anticlinal at a mean depth of 3000 ft. The initial pressure and bubble point pressure are 1100 and 950 psia, respectively. Due to depletion, the field has undergone waterflooding which has also raised the water cut over 90% (Suárez, Gaviria, Pavas, & Frorup, 2005).

This field is currently producing from the Caballos formation which is divided into the Upper (KCUD) and the Lower (KCLD) sands, whose petrophysical properties are summarized in **table 13.** This table shows that KCUD has better properties than KCLD, not only for storage capacity but also for hydraulic conductivity. The generalized stratigraphic column including the reservoir (***R***), source (***F***) and seal (***S***) formations is presented in **figure 60**.

Table 13. Petrophysical properties of the Caballos formation.

Formation	Porosity [%]	Permeability [mD]	Net Pay [ft]	Reservoir Depth [ft]
KCUD	14 – 18	200 – 2000	65	2700 - 3800
KCLD	9 – 13	50 – 200	46	

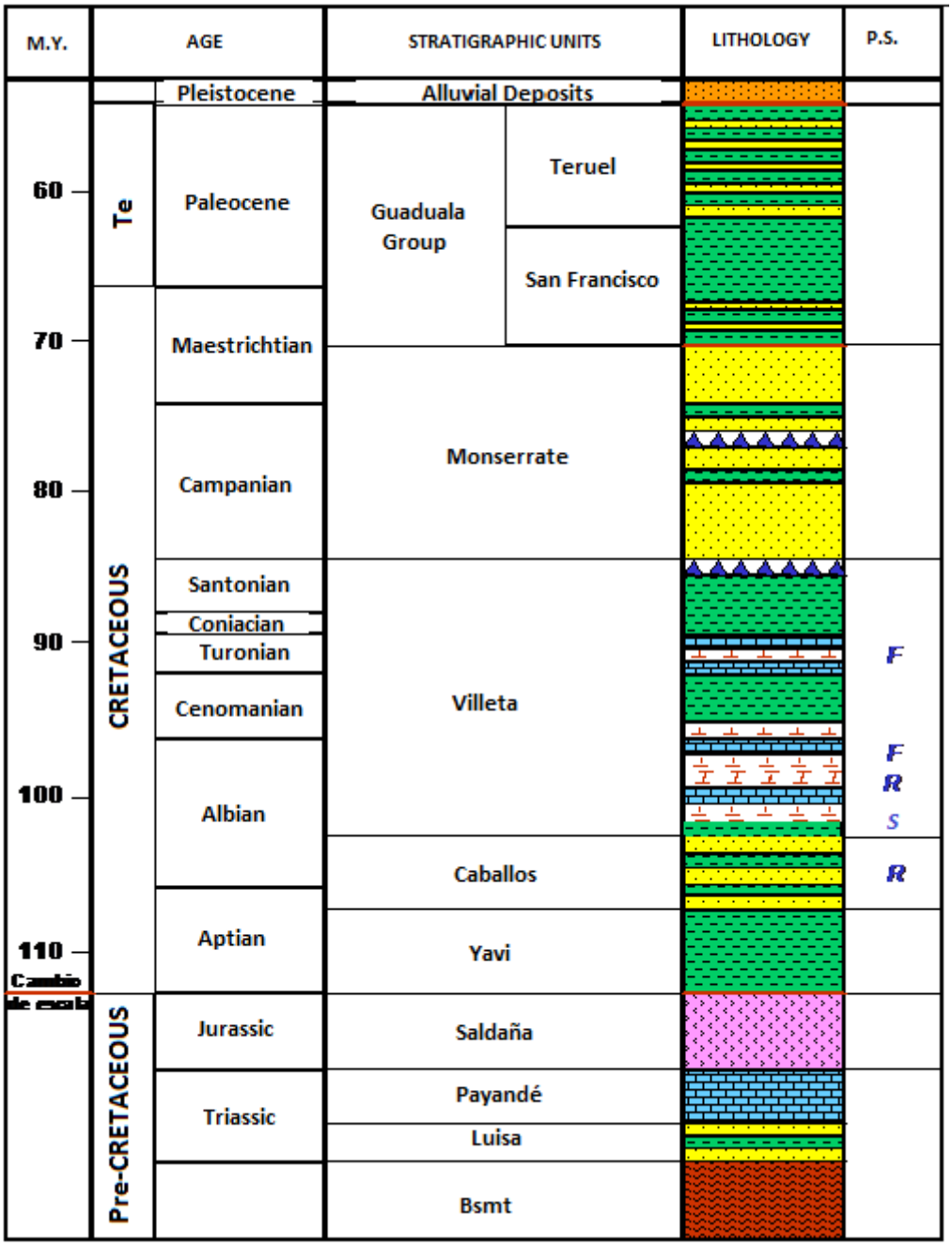


Figure 60. Stratigraphic column of San Francisco Field.

Taken from (Valencia, Pereira, Medina, & Herrera, 2011)

Recall **section 3.9** when a realistic velocity model was constructed from sonic logs and VSPs information. By that time, 5 layers were labeled anonymously as Formation 1, 2, 3, 4 and 5. From here forward, they will be referred to as KM (Monserrate Formation), KVB (Villeta-Bambuca Formation), KVT7 (Villeta-Tetuan Formation), KCUD (Upper Caballos) and BSMT (Basement),

respectively. Let's proceed to analyze our layer cake model (**figure 65**): first, all those thin formations from Monserrate to surface have been replaced by a single thick formation of homogeneous properties named "KM" since they are of no direct interest to this study. Below KM, the source rock (Villeta Group) is represented by 2 layers (KVB and KVT7) of similar composition consisting of marine organic-rich shales mainly, however slightly different acoustic velocities probably related to compaction processes. Further down, the reservoir rock is composed of quartz sandstones intercalated with dark gray shales; although the reservoir encompasses two sections (KCUD and KCLD), there was no appreciable difference in acoustic velocities between them and therefore, only KCUD is drawn in the model. Lastly, the basement is the thickest layer with the highest velocity values at the bottom of the model.

4.3.2. **Generalities: Fracturing Information – San Francisco Field.** In September 2015, hydraulic fracturing was performed in the San Francisco field targeting the sandstones of the Caballos formation (stimulated interval 3318-3328 ft equivalent to 1011.3-1014.4 m) not only to increase the hydrocarbon production but also to decrease the effective permeability to water. For scaling purposes, the area of interest has been enclosed by a 3x3 km green square in **figure 61**. The 4 yellow circles represent the recording stations (SF00, SF18, SF21, SF66) and the red circle is the treatment well (SF83).

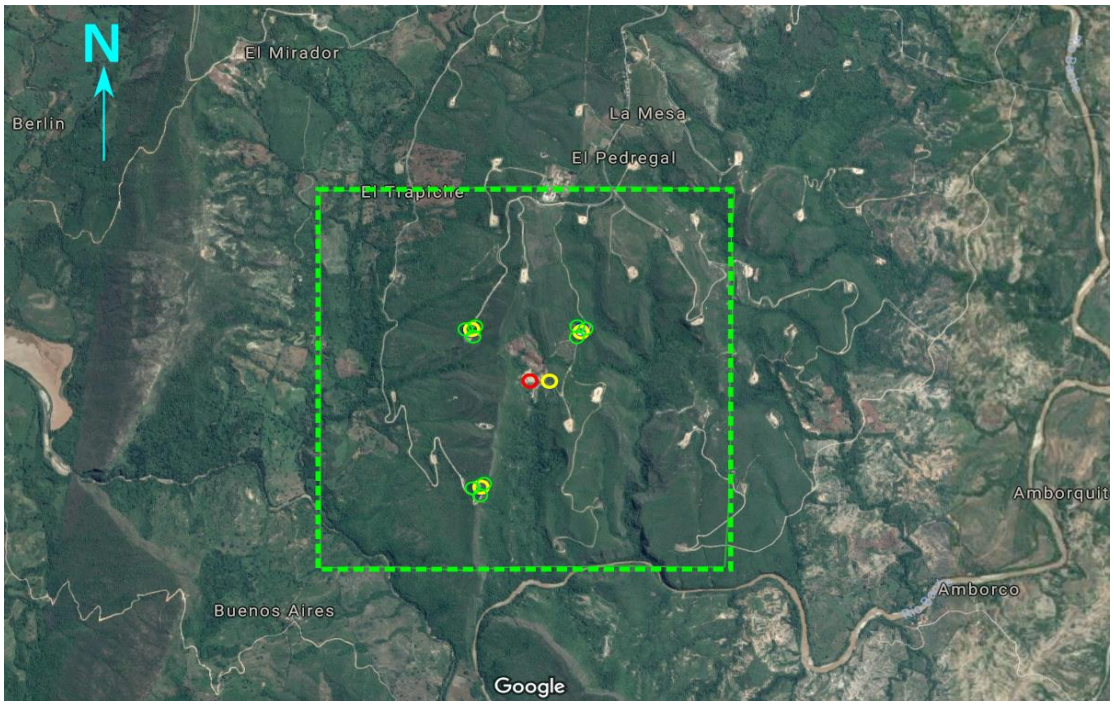


Figure 61. Microseismic monitoring network deployed for hydraulic fracturing in SF83.

A clearer image of the monitoring network is shown in **figure 62**, 3 microarrays (SF18, SF21, SF66) and one extra multicomponent sensor (SF00) were deployed. Each microarray comprises 4 geophones (3 vertical sensors on the sides and 1 (3C) in the middle), for a total of 13 sensors (squares – 4 multicomponent and 9-vertical). Additionally, the red triangle represents the treatment well (SF83), purple circles represent wells where Vertical Seismic Profiles (VSPs) were gathered, and green circles symbolize wells where sonic logs were recorded.

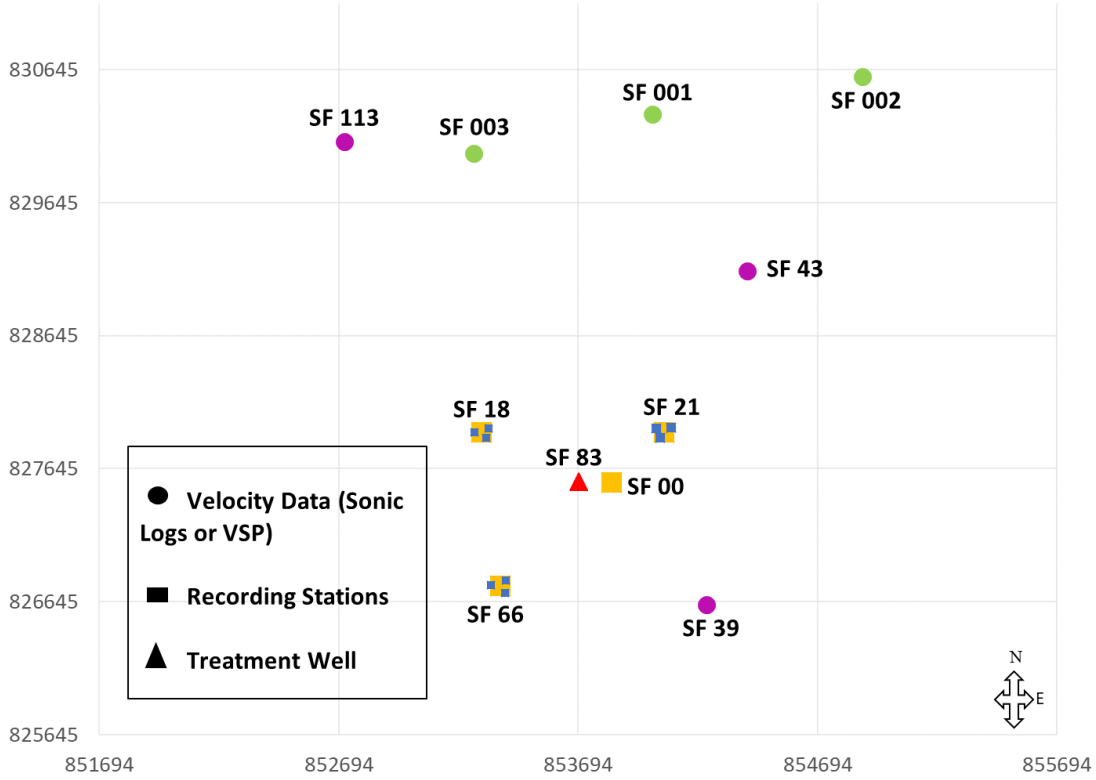


Figure 62. Location of wells where Sonic logs and VSP data were acquired.

4.3.3. **Field data analysis.** Microseismic monitoring was conducted during a CW-Frac (Control Water incursion when Fracturing) service and other minor operations on the Caballos Formation. In this process, multicomponent (3C) seismic data were continuously recorded at 4 stations (SF00, SF18, SF21 an SF66) spanning 2 weeks from September 2nd (18:15) to September 16th, 2015 (08:01).

Microseismic monitoring operations are typically used to assess the impact of the industrial activity on the underground, by locating low magnitude earthquakes (microearthquakes Mw<0) and building real-time models that help to describe the stimulated zone including features like geometry, density and length of fractures, direction of the current stress field, fault reactivation

and other applications (Maxwell, 2011). However, apart from those common practices, engineers and geophysicists also identified seismic events with waveforms similar to those previously reported in a few shale plays that had undergone fracturing operations; unfortunately, there was neither enough evidence to determine if such waveforms belonged to real events nor a method to locate their origin.

The reported seismic events were recorded by the station SF00 on September 5th at 16:55:28 and 22:49:36 UTC exactly during and after the treatment, which implies that they might have been triggered by the stimulation. Consequently, this current work is intended to determine whether such events can be located inside the volume of study using the proposed *Back-Projection Imaging* (BPI) technique.

September 5th at 16:55:28 UTC.

The time at which this event occurs coincides with a series of pumping stages of gradually denser fluid. **Figure 63** is an extracted segment of the seismogram recorded on the vertical component of station SF00. It can be noticed in the spectrogram that most of the energy is concentrated on what would be called a burst of energy with a frequency range of 15Hz and 80Hz, as well as the absence of recognizable seismic phases in the seismogram which agrees with some characteristics reported in the literature for LPLD events.

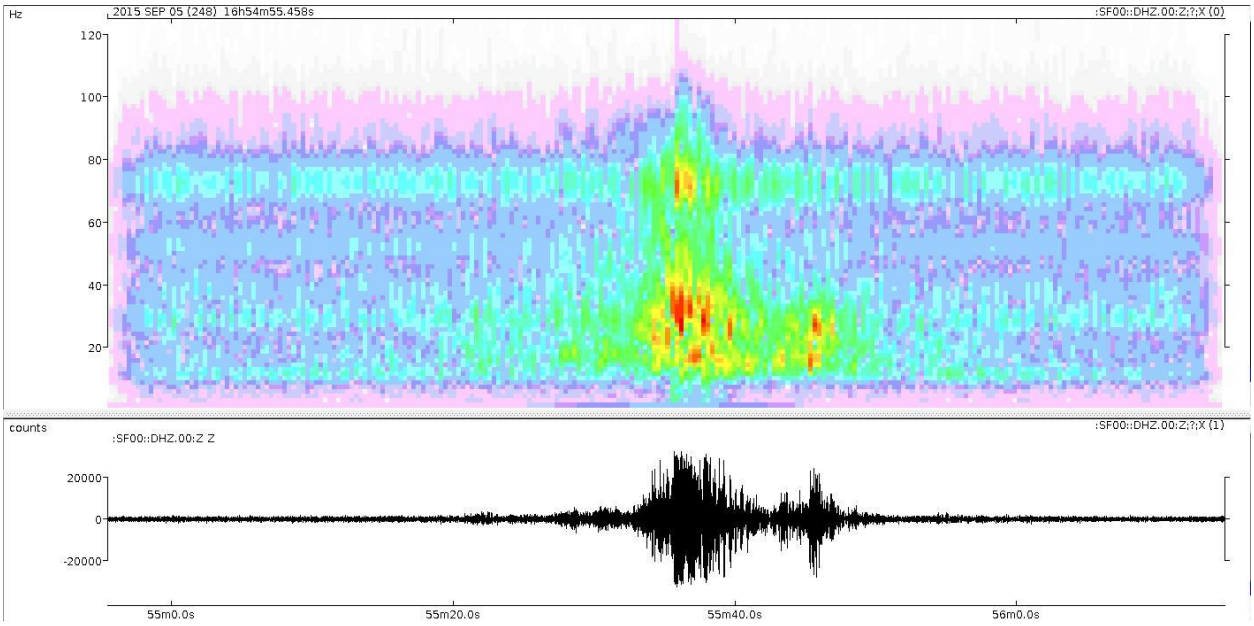


Figure 63. Spectrogram (Above) and Seismogram (Below).

Three more reason to believe this event was triggered by the stimulation process are: first, the time at which it was recorded (16:55:28 UTC), inside the treatment window [16:42 – 17:37 UTC]. Second, no regional earthquake or teleseism was reported on the Colombian earthquake catalogue, and third, not all the stations recorded the event, as it would be the case for regional earthquakes due to the proximity of the stations.

September 5th at 22:49:36 UTC.

This event occurs seemingly post-operation because no treatment procedures were reported at that time. According to **figure 64**, one could arguably recognize the arrival of the P and S phases, nonetheless, the delay between phases is approximately 40 seconds which, roughly speaking, would imply an earthquake 270 kilometers away (assuming Vp=5km/s), and therefore, such event should have been recorded by the other stations, but surprisingly it was not!

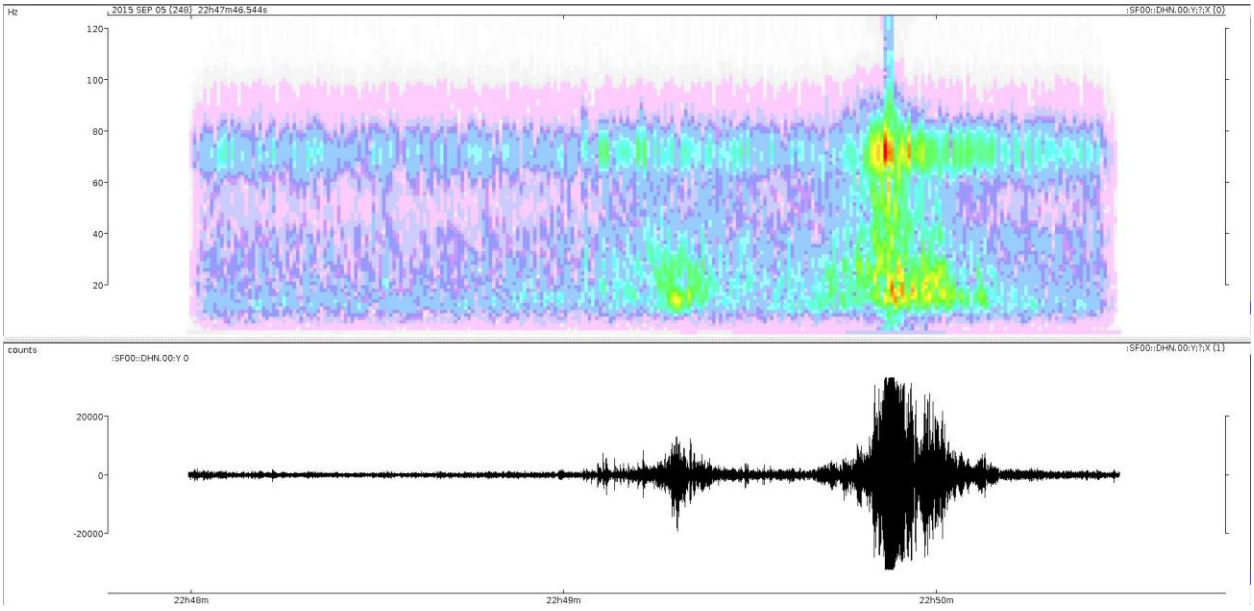


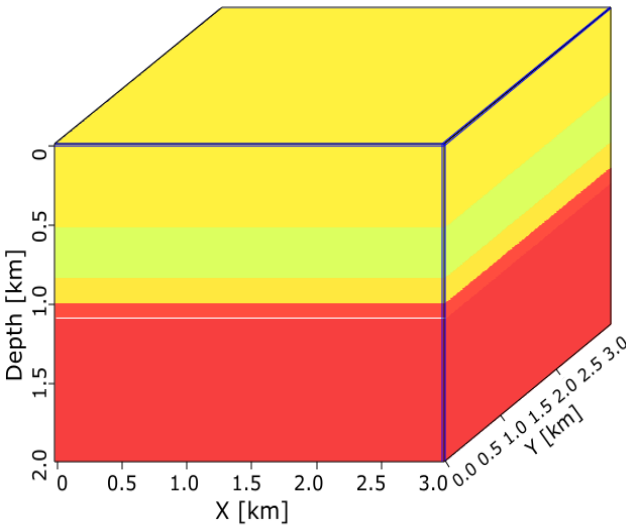
Figure 64. Seismogram and Spectrogram.

The spectrogram also exhibits energy concentrated on two zones around 70Hz and 20Hz. Although this fact does not ensure the event is a true LPLD, it indicates that it should not be immediately discarded.

4.4. **Intermediate Step: Realistic Synthetics**

This synthetic experiment aims to be an intermediate step between the idealized cases studied in **chapter 3** and the more complex application to field data in **section 4.5**. Moreover, it will help to determine the confidence interval of the solution for further analysis of the field data.

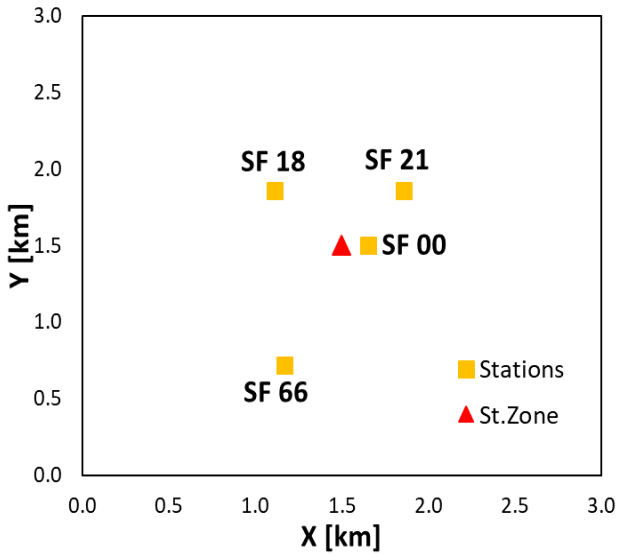
The first step for computing synthetics is to construct a geophysical model in which the forward modelling can be performed. Consequently, the velocity model derived from sonic logs in **section 3.9** is used here with some modifications on the depth of each layer to make them match the formation tops registered in well SF83 (treatment well).



Formation	Top [m]	P-Velocity [m/s]
Monserrate	0	3195
KVB	520	2900
KVT7	840	3250
KCUD	980	4275
BSMT	1080	4430
Dimensions: X = 3[km], Y = 3[km], Z = 2[km]		

Figure 65. Modified 3D Layer cake model.

The second step is defining a recording geometry. For our purpose, a simplified version of the one deployed on field (**figure 62**) suffices. In such a simplified version (**figure 66**), microarrays are replaced by single vertical sensors, however, multicomponent sensors could be used instead.



Item	X	Y	Z
SF00	1.65	1.50	0.00
SF18	1.11	1.86	0.00
SF21	1.86	1.86	0.00
SF66	1.17	0.72	0.00
Stimulated Zone	***1.50***	***1.50***	***1.02***

Figure 66. Recording geometry used for backpropagating the synthetic and field data.

Then, this velocity model is used for synthetic generation under controlled conditions through the convolution between a representative tremor-like signal and previously computed response of the media following the methodology explained in **section 2.4**. Finally, proceed to the backpropagation process.

4.4.1. Backpropagation – Monte Carlo Analysis. The purpose of implementing a Monte Carlo algorithm is not to directly measure the effects of velocity uncertainty on the event location, which was address in **section 3.9**, but to estimate the confidence interval of the methods when incorrect velocity models within a ± 25% range are used.

Besides that, applying the methodology explained in **section 2.4** allows to quickly generate synthetics that are representative of the phenomenon and suitable for our academic purposes. **Figure 67** shows an example of such seismograms after convolution, filtering, random noise addition and other techniques are applied.

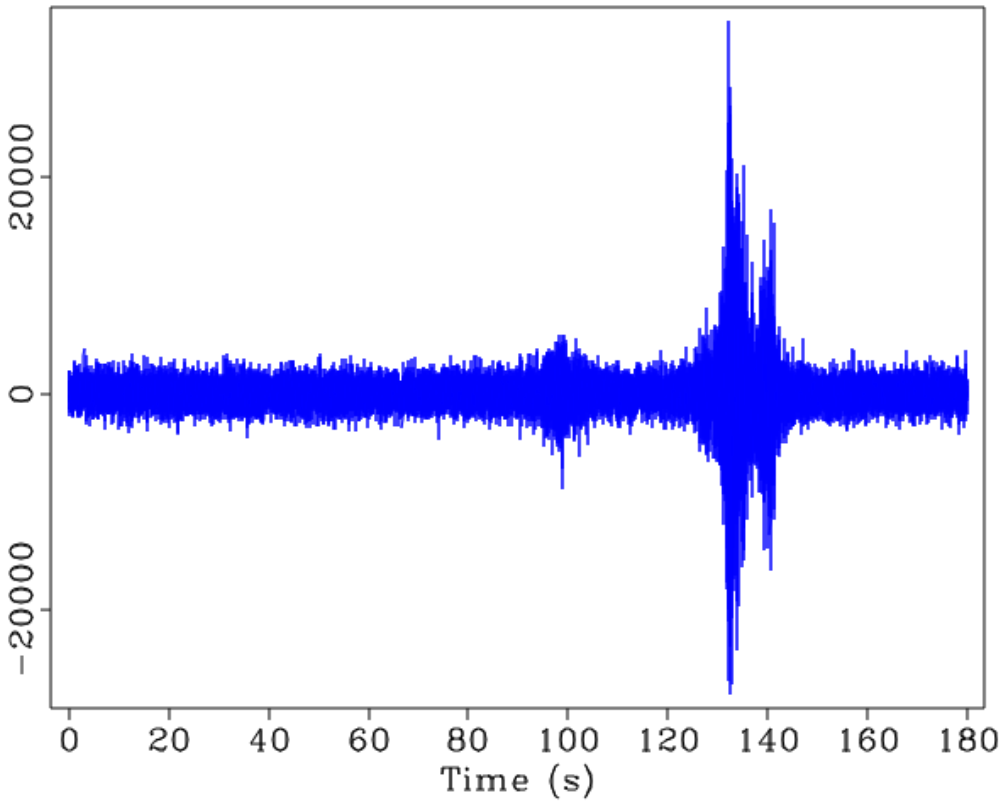


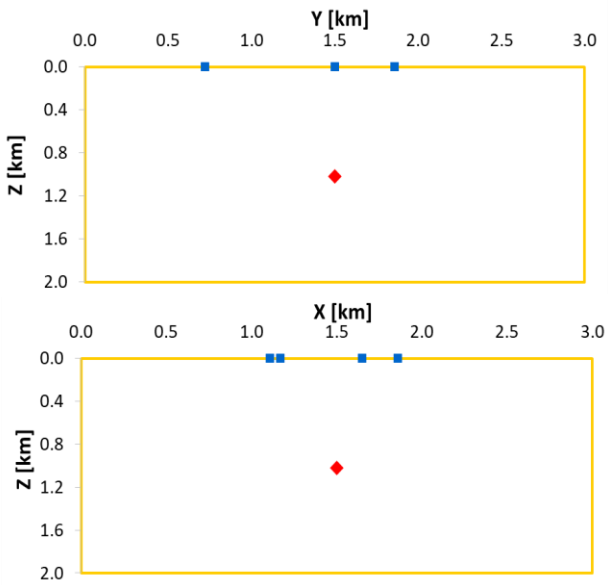
Figure 67. Synthetic trace obtained for station SF00.

This time only two methods are tested and compared: first, *Maximum Amplitude Time Function (**MATF**)* on behalf of classical backpropagation methods, and the proposed *Spatial PDF-based Amplitude Stacking (**PrAS**)* representing the centroid-based methods[13]. Besides, the Monte Carlo analysis is performed twice: in the first case, the event originates in the exact zone where the hydraulic fracturing took place (**figure 68**), and second one, the event originates away from the recording array (**figure 69**) – this case might be extrapolated to non-local earthquakes.

Seismic event source inside treatment interval.

This first case scenario aims to measure the efficacy of the 2 methods to correctly locate the event hypocenters in the region under stimulation. In other words, it determines the probability of a local event to be correctly classified as such.

[13] It was determined from numbers of experiments, not reported here, that PbAS and PrAS produce extremely similar results thus only one of them is used in this chapter.



Item	X	Y	Z
St0	1.86	1.86	0.00
St1	1.17	0.72	0.00
St2	1.11	1.86	0.00
St3	1.65	1.50	0.00
Source	1.50	1.50	1.02

Figure 68. Local event: Experimental set-up.

Table 14 summarizes the results obtained in this experiment. The first column indicates the radial distance from the predicted location of the event to its true hypocenter. The second and third columns show the number of times the event source is located within that tolerance.

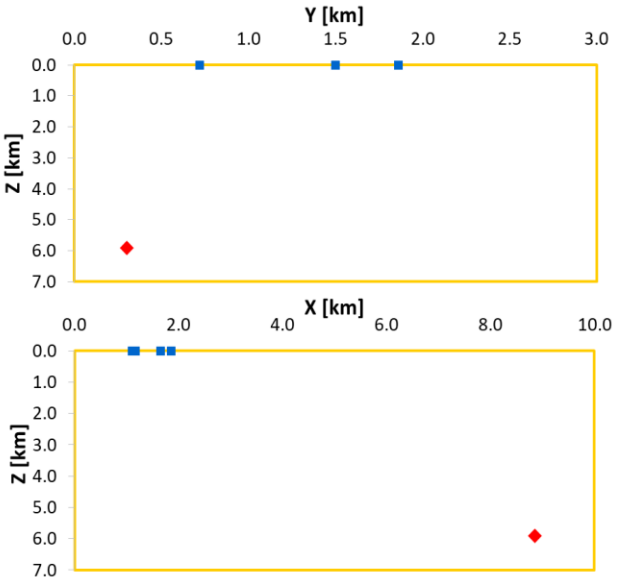
Table 14. Number of local event sources correctly localized.

Dist. [km]	MATF	PrAS
0.10	28	29
0.15	41	47
0.20	52	64
0.25	63	81
0.30	**72**	**91**
0.35	81	97
0.40	88	100
0.45	94	100
0.50	97	100
0.55	99	100
0.60	99	100
0.65	100	100
0.70	100	100

Results show that 95% confidence intervals are achieved within 0.45km and 0.35km from the stimulation zone **(x= 1.5km, y=1.5km, z=1.02km)** for MATF and PrAS, respectively. Alternatively, if a specific distance is required for design purposes, for example 0.2 km, MATF correctly locates 52 events whereas PrAS 64. This Monte Carlo analysis serves not only to validate the effectiveness of the proposed algorithm but also to design acquisition geometries (i.e. The most appropriate geometry for a microseismic monitoring could be derived from the expected length and width of the induced fractures, as well as the Stimulated Reservoir Volume SRV).

Distant Event.

This second scenario studies the possibility of an external event being misclassified as a local one. To put it another way, it is important to determine what percentage of external events will be correctly identified as such and how many might be confused with events triggered by the stimulation. Notice that the model in **figure 69** has been extended to simulate a source away from the recording array.



Item	X	Y	Z
Station0	1.86	1.86	0.00
Station1	1.17	0.72	0.00
Station2	1.11	1.86	0.00
Station3	1.65	1.50	0.00
Source	8.85	0.30	5.90

Figure 69. Distant event: Experimental set-up.

Table 15 shows that no event is located inside a radius of 0.35km from the stimulation zone **(x= 1.5km, y=1.5km, z=1.02km)**, however, MATF begins to encounter false positives beyond that point. Furthermore, the longer the distance, the more prone both methods are to misclassify event sources due to erroneous velocity models.

Table 15. Number of non-local event sources incorrectly localized.

Dist [km]	MATF	PrAS
0.10	0	0
0.15	0	0
0.20	0	0
0.25	0	0
0.30	**0**	**0**
0.35	0	0
0.40	1	0
0.45	3	0
0.50	7	0
0.55	7	0
0.60	7	0
0.65	7	0
0.70	7	0
0.75	7	1
0.80	7	3
0.85	7	8

***Note*:** Keep in mind that these results were obtained after backpropagating through velocity models with a ± 25% error, however, such interval may vary depending on the model reliability.

Binary Classification Test.

This kind of test helps measure the effectiveness of a method, that is to say, to determine whether or not a given method or technique does what it is supposed to do (see **Appendix D**). A more detailed statistical analysis was conducted on the above scenarios to evidence how both MATF and PrAS behave under real conditions.

Example.

Let's proceed to construct and interpret the *confusion matrix* of MATF and PrAS for radial distance of 0.30km.

- Number of events occurring in the stimulation zone → Positives [P] = 100

- Number of non-local events → Negatives [N] = 100
- Local Events correctly classified → True Positives [TP]
- Local Events incorrect classified → False Negatives [FN] = [P] – [TP]
- Non-local events classified as such → True Negatives [TN]
- Non-local events incorrectly classified → False Positives [FP] = [N] – [TN]

Classifier	**MATF**	**PrAS**
TP	72	91
FN	28	9
TN	100	100
FP	0	0

Now, we can continue to calculate parameters such as Sensitivity [TPR], Specificity [TNR], Precision [PPV] and Accuracy [ACC]:

Parameter	**MATF**	**PrAS**
$\boldsymbol{TPR = \frac{TP}{TP + FN}}$	$\frac{72}{72 + 28} = 0.72\ or\ 72\%$	$\frac{91}{91 + 9} = 0.91\ or\ 91\%$
$\boldsymbol{TNR = \frac{TN}{TN + FP}}$	$\frac{100}{100 + 0} = 1\ or\ 100\%$	$\frac{100}{100 + 0} = 1\ or\ 100\%$
$\boldsymbol{PPV = \frac{TP}{TP + FP}}$	$\frac{72}{72 + 0} = 1\ or\ 100\%$	$\frac{91}{91 + 0} = 1\ or\ 100\%$
$\boldsymbol{ACC = \frac{TP + TN}{TP + TN + FP + FN}}$	$\frac{72 + 100}{72 + 100 + 0 + 28} = 0.86\ or\ 86\%$	$\frac{91 + 100}{91 + 100 + 0 + 9} = 0.96\ or\ 96\%$

Finally, the confusion matrices are as follows:

MATF

P	N	
100	100	
TP	FP	PPV
72	0	100%
FN	TN	ACC
28	100	86%
TPR	TNR	
72%	100%	

PrAS

P	N	
100	100	
TP	FP	PPV
91	0	100%
FN	TN	ACC
9	100	96%
TPR	TNR	
91%	100%	

Now, from the confusion matrix of MATF it can be inferred that:

- A *Sensitivity* of 0.72 means that 72% of the local events were correctly identified as such.
- A *Specificity* of 1.00 means that 100% of the non-local events were correctly identified as such.
- A *Precision* of 1 means that 100% of the events classified as local events were actually local.
- An *Accuracy* of 0.86 means that 86% of the total events were properly classified (either local or non-local).

As the reader can see, a lot of valueable information can be extracted from confusion matrices and we could continue generating the matrices for the rest of the distances, however, this process is time consuming. Therefore, an easier way to present the results is by plotting them as in **figure 70** in which both methods can be directly compared.

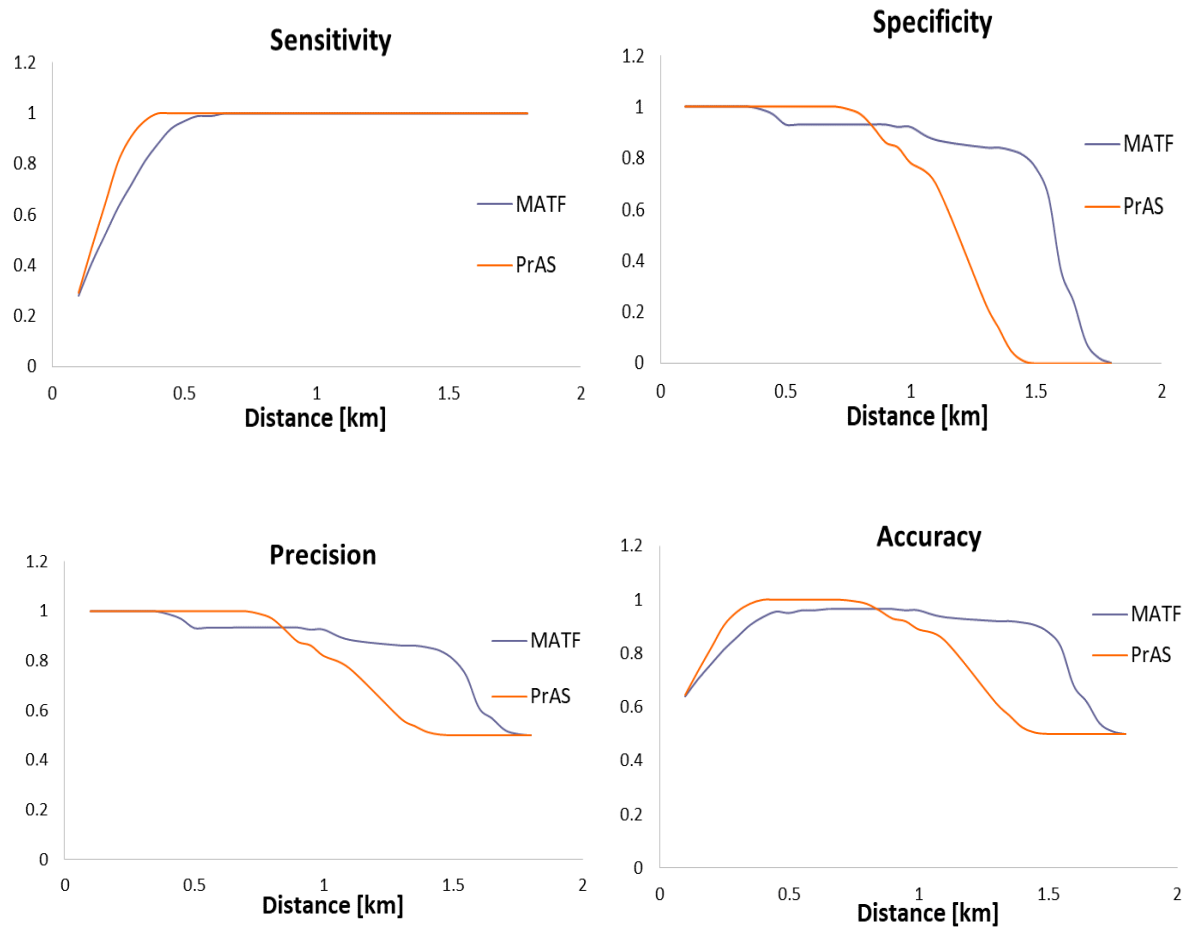


Figure 70. Receiver Operating Characteristic (ROC) Curves.

The *sensitivity* curve shows that PrAS quickly identifies local events while MATF needs a longer radial distance to localize them all what makes PrAS more suitable for local events. However, the *specificity* curve of PrAS exhibits an earlier descent which implies that as the radial distance from the stimulated zone increases, this method begins to misclassify non-local events earlier than MATF. Besides, the *Precision* and *Accuracy* of both methods indicate that although PrAS is more adequate for local events, it loses consistency for distant events.

Another useful way to display the results consists of matrices (**figure 71**) in which the color intensity [0 – 1] indicates the value of the parameter in question and the number inside the box indicates the radial distance in kilometers from the stimulated zone.

For example, at 1.60km from the stimulated zone, MATF has Sensitivity 1.0, Specificity 0.36, Precision 0.61 and Accuracy 0.68, whereas PrAS has 1.0, 0.0, 0.5 and 0.5 for the same indicators respectively.

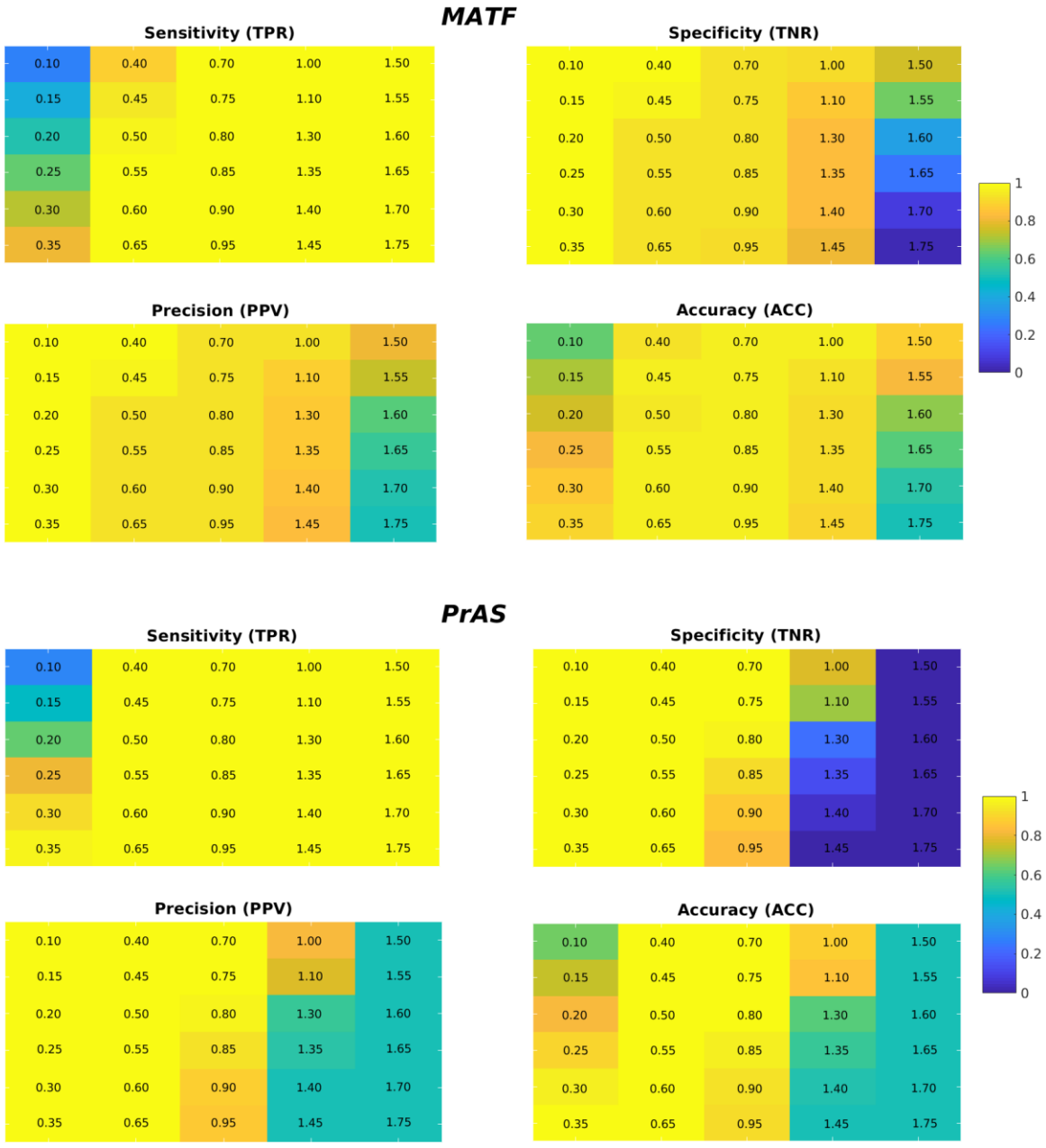


Figure 71. Binary classification matrices of MATF and PrAS.

4.5. **Real Data**

Thanks to all those synthetic tests performed in both chapters 3 and 4, we are now aware of the functionality of the location algorithms and the way they are affected by incorrect velocity models. At this point, the very field data (seismograms) recorded when hydraulic fracturing the KCUD formation can be backpropagated and analyzed to determine whether the events were triggered by the stimulation.

4.5.1. **Velocity model.** Even though there is a more realistic model available, we have experienced difficulties during the well tie process after applying a time-to-depth conversion for the RMS velocity model, hence, making it impractical for use since we do not intend to add more sources of uncertainty, therefore, the layer cake model in **figure 65** is considered a more adequate option.

4.5.2. **Backpropagation.** Once the velocity model is constructed, it is time to backpropagate the field data using both the Raw and the Normalized Absolute Value (NAV) versions of the seismograms for comparison purposes.

September 5th at 16:55:28 UTC.

As previously introduced, this event was recorded with multicomponent sensors, however, only their vertical components (**figure 72**) are extracted because S-waves are not covered in this study.

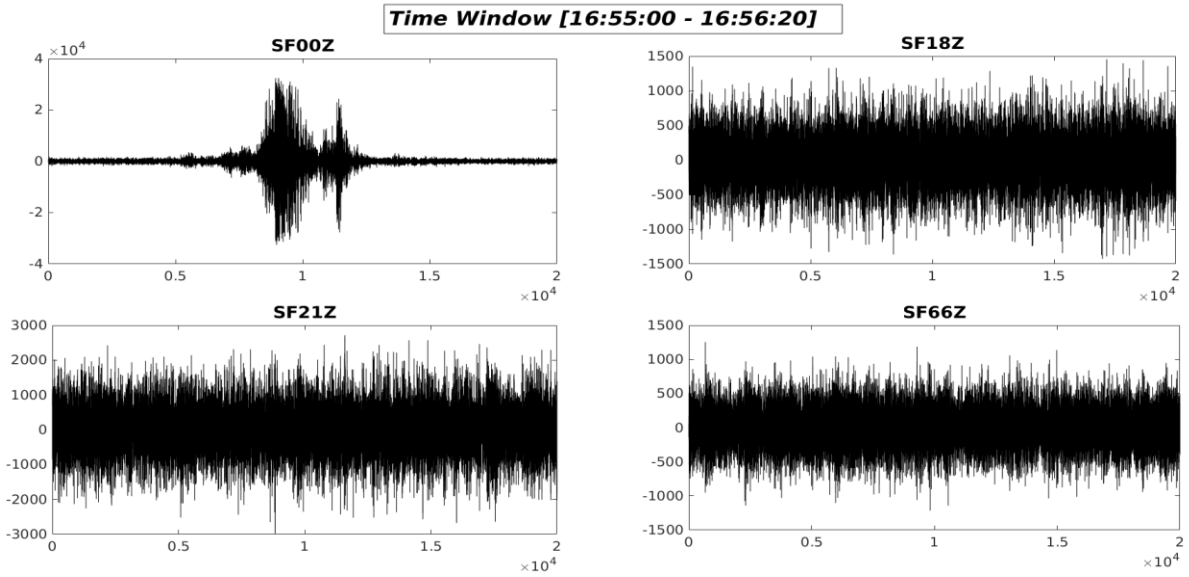


Figure 72. Vertical components of the event recorded at 16:55:28 in stations SF00, SF18, SF21 and SF66.

Notice that the event is apparently recorded only by SF00, and there was no evidence of it on the other stations even after rotating and filtering the traces. This fact makes conventional location a complicated to impossible task without employing polarization analysis and similar techniques.

As explained in **chapter 2**, backpropagation begins by constructing a brightness or image function - $F(\boldsymbol{r}, t)$ - and its corresponding curve of maximum stacked energy - $\max_{\mathbf{r}}(F(\boldsymbol{r}, t))$ – which is a crucial step for BPI methods. **Figure 73** shows the effects of using pre-processed and raw data on such a curve; the former generates varying levels of energy related to noise and signal summation indistinctly, on the other hand, raw seismograms together with a precise velocity model provide a remarkable noise suppression effect.

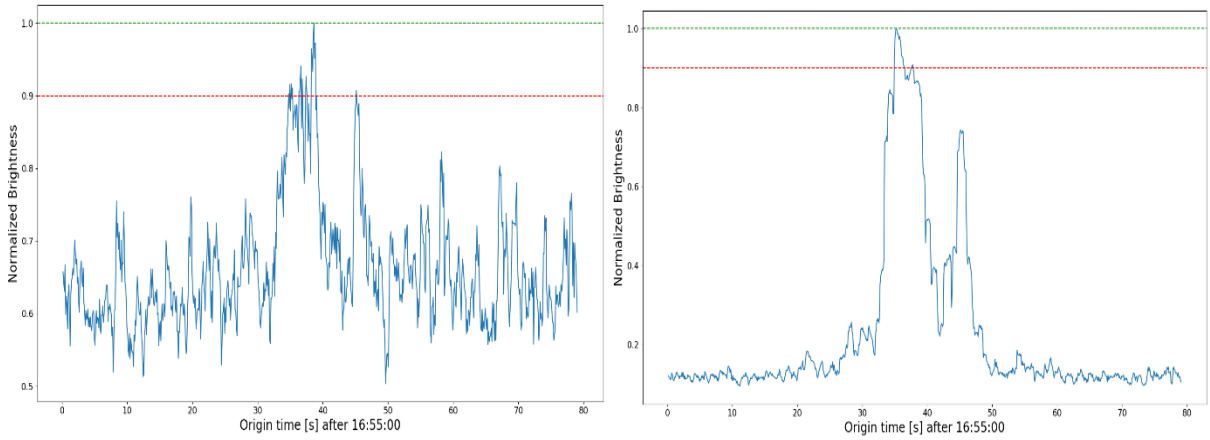


Figure 73. Normalized maximum brightness function for NAV (left) and Raw (right) seismograms.

An appropriate colormap is used in **figure 74** to highlight an apparent concentration of energy at a depth of 0.84km, which is relatively close to the stimulated zone, considering that the stimulation process creates a network of fractures along the well that can extend for dozens or hundreds of meters depending of the well geometry and fracturing design.

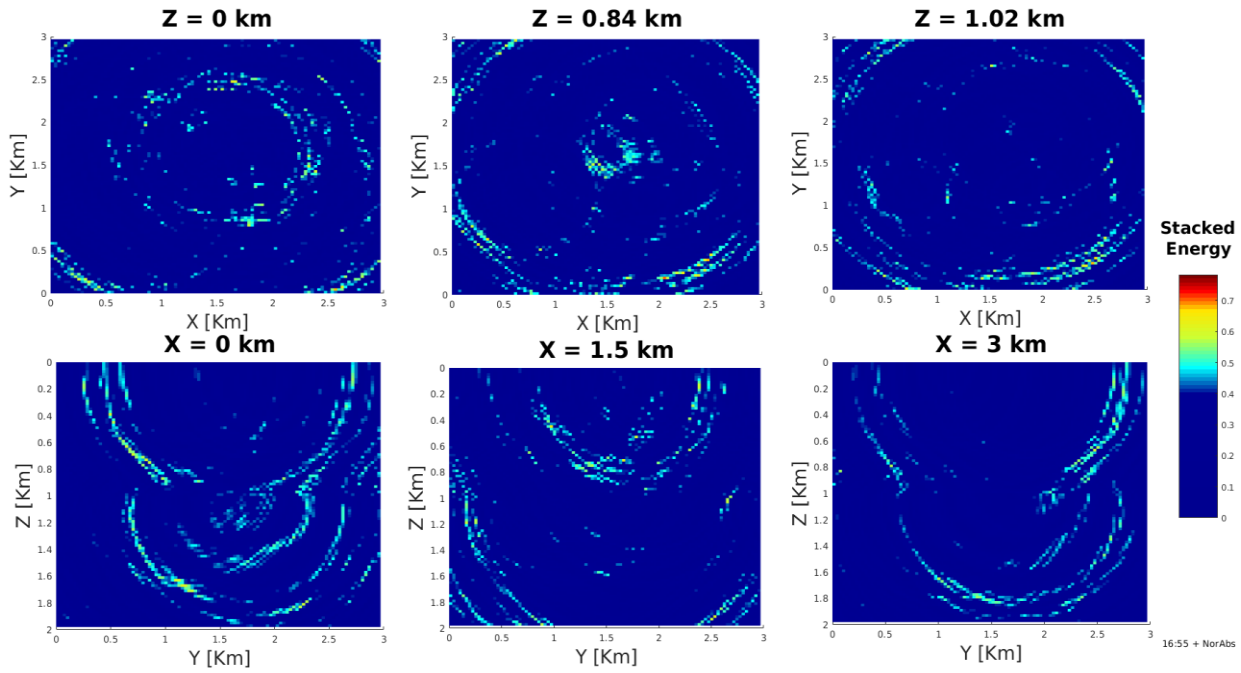


Figure 74. Energy volume obtained with NAV seismograms.

Although the figure above indicates the possible existence of either an event or an artefact in that location, there is a need for a more precise velocity model which would increase the coherence among signals, thus either highlighting this event or completely removing the artefact. This is a clear example of the main disadvantage grid-search methods (such as MATF) have when unsuitable models are used, even with minor errors, they fail to identify the event. **Table 16** summarizes the results obtained by MATF and PrAS. The former finds no substantial stacked energy, at least no close to the stimulated zone. On the other hand, the latter computes the centroid within the zone of interest when the raw seismogram is used.

Table 16. Spatiotemporal location: Event of September 5th at 16:55:28 UTC. A) Results of MATF, B) Results of PrAS.

MATF							
	T_0	**X_0[km]**	**Y_0[km]**	**Z_0[km]**	**% Error**		
					X	**Y**	**Z**
Raw	16:55:35.25	2.28	2.52	0.98	52	68	3.92
NAV	16:55:38.65	0.09	2.52	0.48	94	68	52.94

PrAS										
	T_{cent}	T_{peak}[s]	X_0[km]	Y_0[km]	Z_0[km]	% Error			M_{exp}	N_{exp}
						X	Y	Z		
Raw	16:55:39.10	16:55:35.50	**1.46**	**1.49**	**1.07**	2.80	0.73	5.20	8	40
NAV	16:55:39.92	16:55:38.39	0.62	2.26	0.72	58.87	50.73	29.51	8	40

Although the results are not clear enough, there is a possibility for this event to be associated to the stimulation process, however, it would require more advanced techniques and a better velocity model to be precisely located.

September 5th at 22:49:36 UTC.

Even though this second event was also recorded by a single station (**figure 75**), it was backpropagated for academic purposes. We aim not only to understand how the location methods behave under adverse conditions, but also to find evidence for the existence or not of such events.

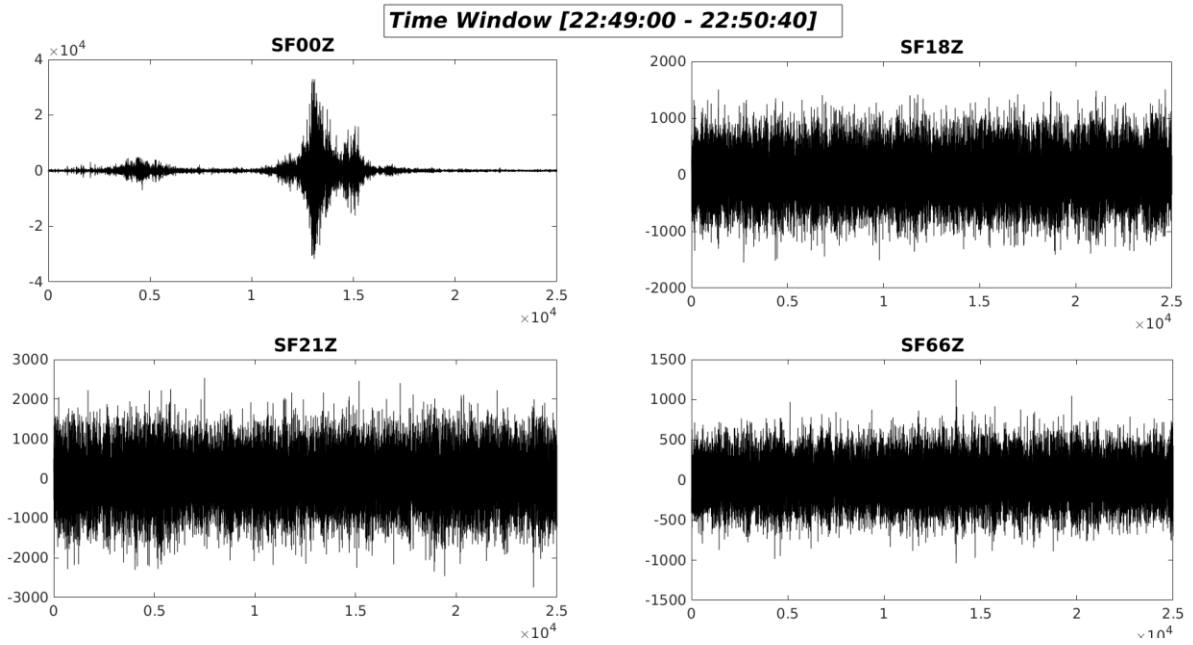


Figure 75. Vertical components of the event recorded at 22:49:36 in stations SF00, SF18, SF21 and SF66.

Despite the clear peak on the maximum brightness curve (**figure 76**), results are also diffuse for this second event. One may spot a decent concentration of energy at a depth of 0.86km (**figure**

77), or even a region of high energy at 1.14km, however, nothing can be concluded due to the non-negligible uncertainty in the velocity model.

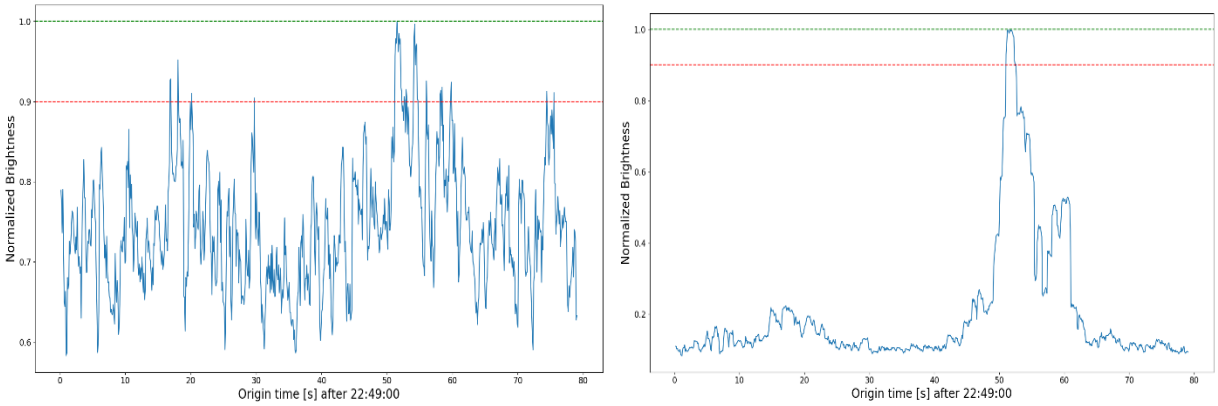


Figure 76. Normalized maximum brightness function for NAV (left) and Raw (right) seismograms.

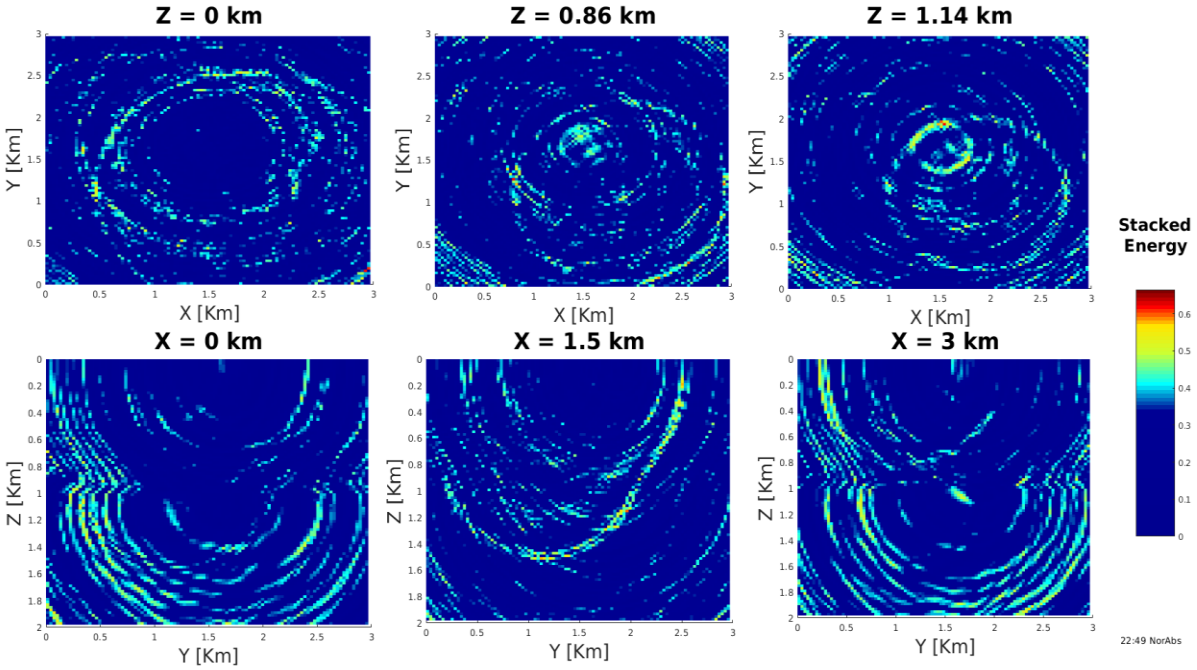


Figure 77. Energy volume obtained with NAV seismograms.

Once more, MATF fails to locate a maximum of energy in the stimulated region (**table 17**), while PrAS is consistent with its earlier results.

Table 17. Spatiotemporal location: Event of September 5th at 22:49:36 UTC.

MATF							
	T_0	**X_0[km]**	**Y_0[km]**	**Z_0[km]**	**% Error**		
					X	**Y**	**Z**
Raw	22:49:51.65	1.02	0.99	0.26	32	34	74.51
NAV	22:49:51.55	0.21	2.91	0.38	86	94	62.75

PrAS										
	T_{cent}	**T_{peak}[s]**	**X_0[km]**	**Y_0[km]**	**Z_0[km]**	**% Error**			**M_{exp}**	**N_{exp}**
						X	**Y**	**Z**		
Raw	22:49:44.28	22:49:51.69	**1.53**	**1.51**	**0.90**	1.93	0.47	11.37	8	40

After obtaining similar results for both events using seismic traces that we consider unsuitable for location (apparently only one station recorded the event), we need to discard the possibility of artefacts being generated by the method due to the absence of coherent summation among traces. Consequently, two extra synthetic tests are performed:

Unidentified surface event registered only on station SF00 + noise on the other stations.

A Gaussian-envelope-like impulsive event is simulated at (1.5, 1.5, 0) km, so close to station SF00 to be the only one to record it, while the other sensors registered only white noise.

Backprojection of NAV seismograms exhibits dispersed energy at shallow depths (**figure 78**), however, there is no focused energy as expected. On the other hand, using raw seismograms causes destructive interference thus producing a volume with no relevant energy concentrations. Yet, MATF and PrAS estimate focal depths between 0.2km and 0.56km (**Table 18)** as a consequence of random noise addition.

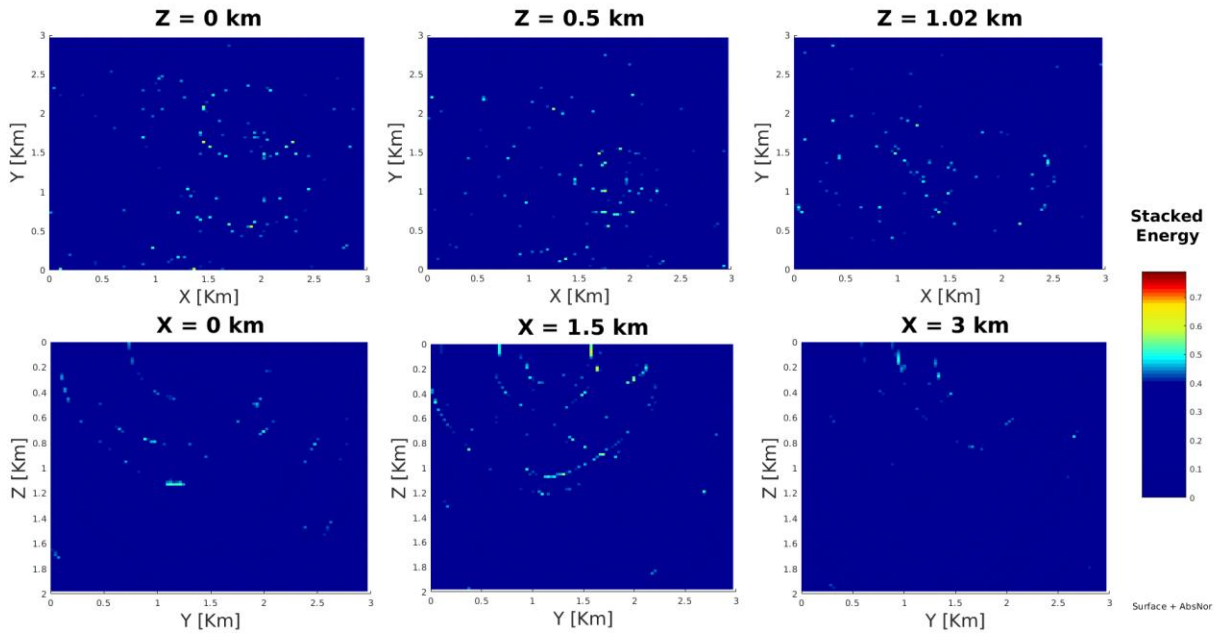


Figure 78. Energy volume obtained for surface event using NAV seismograms.

Table 18. Spatiotemporal location: Noise on surface.

MATF							
	T_0[s]	X_0[km]	Y_0[km]	Z_0[km]	% Error		
					X	Y	Z
Raw	3.85	1.71	2.1	0.22	14	40	Div/0
NAV	8.15	1.62	1.68	0.36	8	12	Div/0

PrAS										
	T_{cent}	T_{peak}[s]	X_0[km]	Y_0[km]	Z_0[km]	% Error			M_{exp}	N_{exp}
						X	Y	Z		
Raw	4.54	3.76	**1.93**	**1.98**	**0.40**	28.67	32.00	Div/0	8	40
NAV	4.62	6.41	**1.76**	**1.91**	**0.56**	17.00	27.40	Div/0	8	40

Then, this experiment helps to reject the idea of noise generated on surface, thus there is a need for a last test.

Injection induced event registered only on station SF00 + noise on the other stations.

This experiment is similar to the preceding one, however, the simulated event occurs at stimulation depth instead (1.5, 1.5, 1.02) km**.** Again, backprojection showed no coherent summation of energy (**figure 79**), which indicates that even when an event does originate at a certain place, its location cannot be recovered by BPI if only one station records it, unless the event is actually recorded by the other stations but concealed in high noise levels.

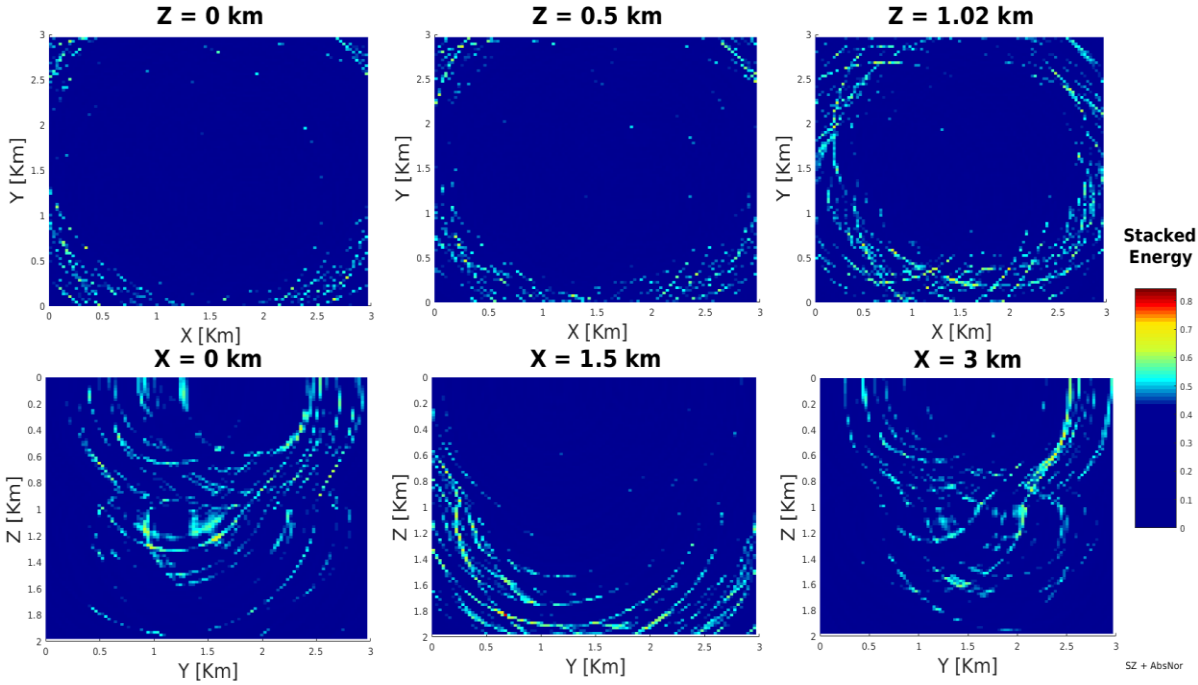


Figure 79. Energy volume obtained for event in stimulated zone using NAV seismograms.

It is worth mentioning that although only one station registered the event, the algorithm gives a rough idea of its location according to results in **table 19**, more specifically for raw traces.

Table 19. Spatiotemporal location: Event inside stimulated zone.

MATF							
	T_0[s]	**X_0[km]**	**Y_0[km]**	**Z_0[km]**	**% Error**		
					X	**Y**	**Z**
Raw	3.85	1.5	1.35	1.44	0	10	41.18
NAV	3.75	0.93	1.32	1.96	38	12	92.16

PrAS										
	T_{cent}	T_{peak}[s]	X_0[km]	Y_0[km]	Z_0[km]	% Error			M_{exp}	N_{exp}
						X	Y	Z		
Raw	4.58	3.85	**1.41**	**1.38**	**1.43**	5.27	7.73	40.20	8	40
NAV	4.63	3.90	**0.99**	**1.26**	**1.62**	34.27	15.80	58.33	8	40

4.5.3. **Results.** After studying the 2 synthetic scenarios, the probability of the hypocenters being actual sources and not artefacts increased. However, there was still not enough evidence to guarantee that with total certainty. Therefore, we decided to follow the approach suggested by (Davis, Nyffenegger, & Frohlich, 1995) in which a list of criteria (**table 20**) is introduced to decide if an event has been induced by fluid withdrawal. Subsequently, here we adapted such questions to analyze the possibility of our candidate events to have been induced by fluid injection:

Table 20. Criteria for induced seismicity determination.

#	Question	Event
1	Timing	
1a	Are these events the first known earthquakes of this character in the region?	Yes/No
1b	Did the events only begin after the hydraulic fracturing had commenced?	Yes/No
1c	Is there a clear correlation between fluid injection and seismicity?	Yes/No
2	Location	
2a	Are epicenters within 5km of wells?	Yes/No
2b	Do some earthquakes occur at or near treatment depths?	Yes/No
2c	Do epicenters appear spatially related to the stimulated region?	Yes/No
3	Fluid Pressure, Etc	
3a	Did fluid injection cause a significant change in fluid pressures?	Yes/No
3b	Did seismicity begin only after the fluid pressures had increased significantly?	Yes/No
3c	Is the observed seismicity explainable in terms of current models relating fluid injection to fault activity?	Yes/No
Total		

Now, let us proceed to answer each question for both reported events.

September 5th at 16:55:28 UTC.

Question 1a.

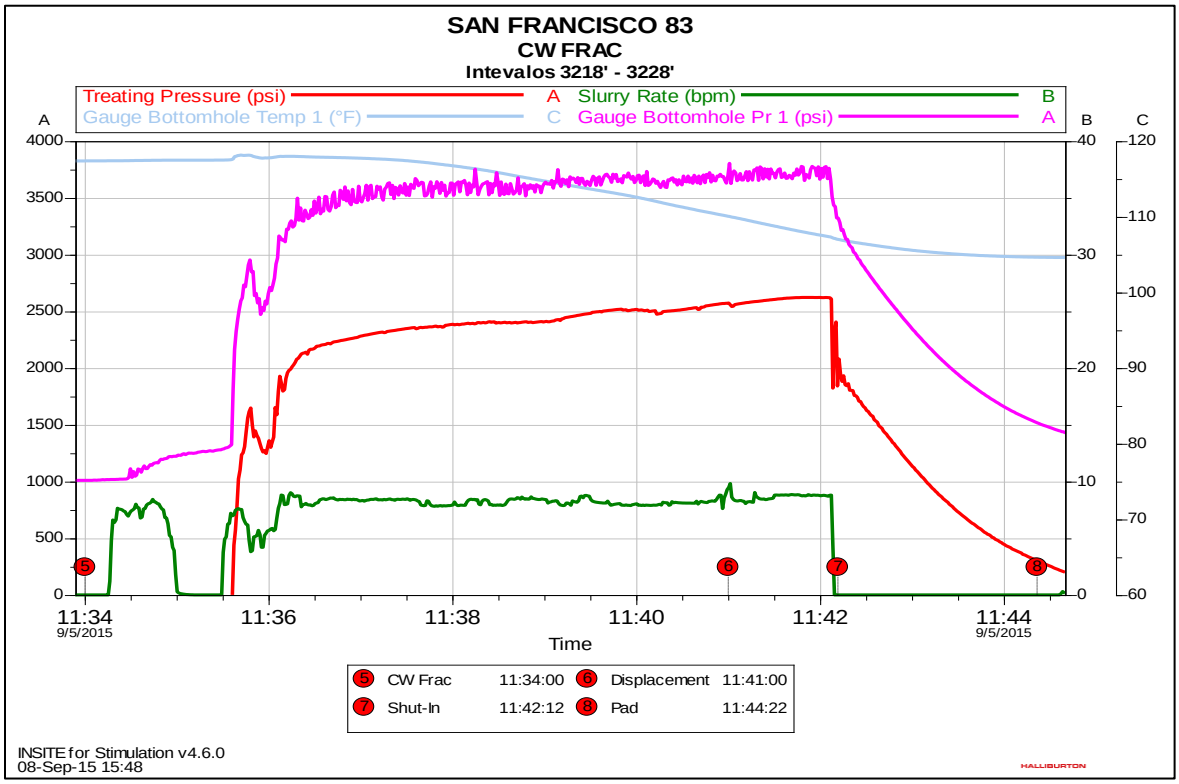
(?). It is impossible to give a proper answer to this question because the national seismological network of Colombia does not have the required instruments to record events of such a low magnitude. On the other hand, the microseismic monitoring operation lasted only 2 weeks, which is not long enough to gather evidence for seismicity in the zone.

Question 1b.

Yes (?). It is important to mention that there are no permanent monitoring stations on field, however, some of them were deployed a few days before the stimulation took place, without success in recording seismicity until the fracturing began.

Question 1c.

Yes. Notice in **figure 72** that the event is recorded on station SF00 at 16:55:28 UTC (11:55:28 in Bogota – Col). This coincides with the injection of high-pressure fluids as it can be seen in **figure 80a&b**.



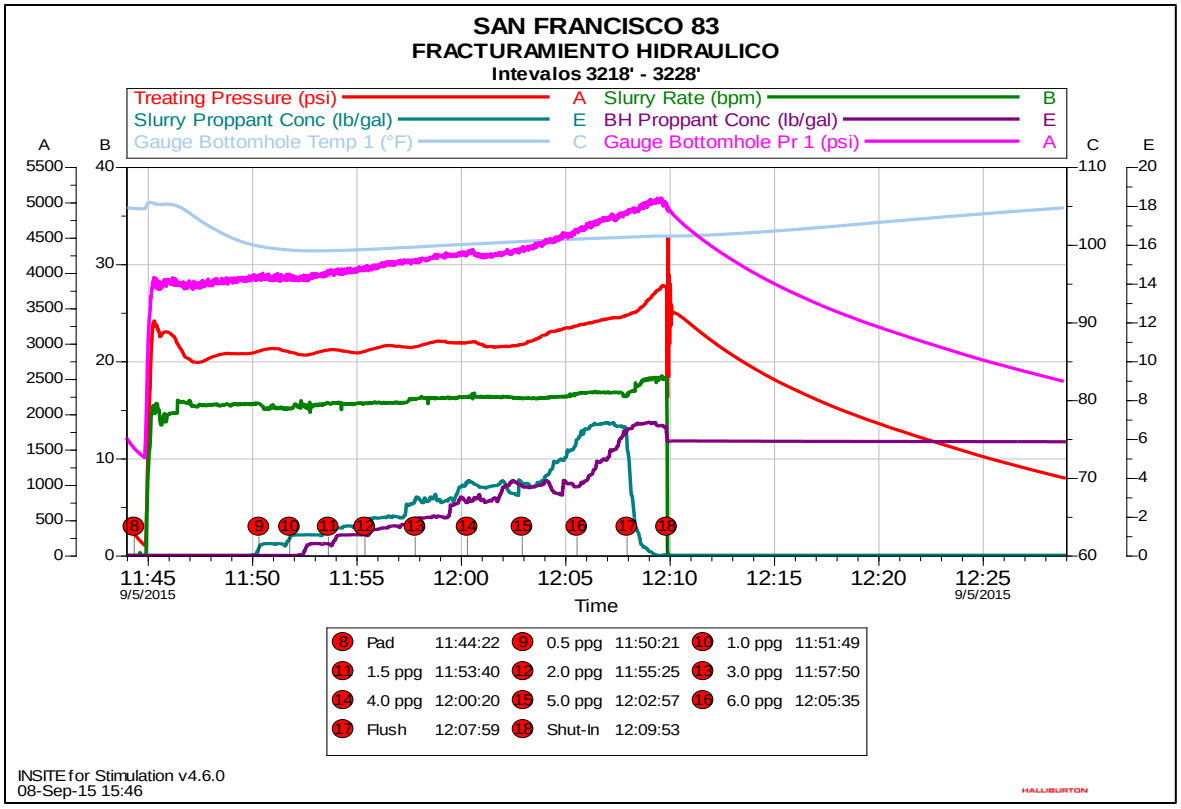


Figure 80. Pressure curves vs time during the hydraulic fracturing operation. Images provided by Ecopetrol S.A.

Question 2a.

Yes (?). Although there is still uncertainty about the results obtained in **subsection 4.5.2**, they suggest the possibility of an event originated at a depth of approximately 1km, which may be related to the stimulation process.

Question 2b.

Yes (?). At least the 2 events presented here, occur near the stimulated zone.

Question 2c.

Yes (?). Our results compute the hypocenter within a 0.1km radius from the treatment zone. However, one should notice that geometries and distribution of fractures (**figure 81**) dictate the extent of Stimulated Reservoir Volume (SRV) and consequently, the region in which microseismicity and LPLD events can occur.

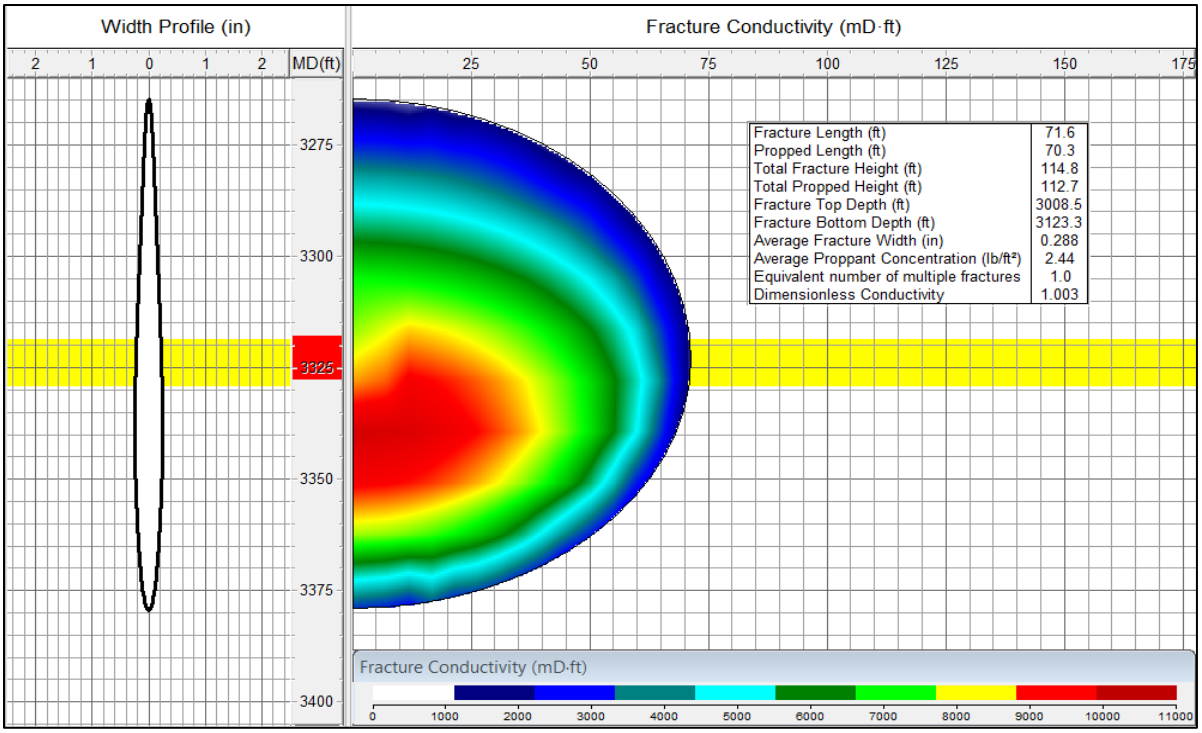


Figure 81. Characteristics of hydraulically induced fractures.

Question 3a.

Yes. Whereas the initial formation pressure does not exceed 1100 psi (Suárez, Gaviria, Pavas, & Frorup, 2005), Curves in **figure 80b** indicate treating and bottomhole pressures over 3000 psi during hydraulic fracturing.

Question 3b.

Yes. The event was recorded during stimulation at 11:55:28 (local time) when the treating and bottomhole pressures had reached 2800 and 4000 psi respectively.

Question 3c.

Yes (?). When fracturing shales, the injection of high-pressure fluids may generate seismicity not only by creating new fractures but also by reactivating pre-existing fractures and faults (Zoback, Kohli, Das, & McClure, 2012). However, the mineralogy of Barnett shales and Caballos formation are notoriously different since the latter is composed of quartz sandstones intercalated with dark gray shales; despite this, we believe such a model would still be applicable to our case.

September 5th at 22:49:36 UTC.

Question 1a.

(?). There is not enough evidence as for the other event.

Question 1b.

Yes (?). This second event is recorded 6 hours after the hydraulic fracturing.

Question 1c.

Yes (?). Although the treatment had finished about 6 hours before the event happened, the injected fluid certainly affected the stress state of the zone, thus this event may have been a consequence of the local stress field reaching equilibrium conditions.

Question 2a.

Yes (?). Both events are thought to have originated within the stimulation region.

Question 2b.

Yes(?). For the same reasons expressed for event 16:55:28.

Question 2c.

Yes (?). Please read the reasons given for event 16:55:28.

Question 3a.

Yes (?). For the same reasons expressed for event 16:55:28.

Question 3b.

Yes (?). Since this event occurred a few hours after treatment, it may have been related to drastic pressure changes undergone during fracturing stimulation.

Question 3c.

Yes (?). Please read the reasons given for event 16:55:28.

Finally, **table 21** summarizes the answers given for the 2 candidate events.

Table 21. Results of the events recorded at 16:55:28 and 22:49:36.

#	Question	16:55:28	22:49:36
1	Timing		
1a	Are these events the first known earthquakes of this character in the region?	(?)	(?)
1b	Did the events only begin after the hydraulic fracturing had commenced?	Yes (?)	Yes (?)
1c	Is there a clear correlation between fluid injection and seismicity?	Yes	Yes (?)
2	Location		
2a	Are epicenters within 5km of wells?	Yes (?)	Yes (?)
2b	Do some earthquakes occur at or near treatment depths?	Yes (?)	Yes (?)
2c	Do epicenters appear spatially related to the stimulated region?	Yes (?)	Yes (?)
3	Fluid Pressure, Etc		
3a	Did fluid injection cause a significant change in fluid pressures?	Yes	Yes
3b	Did seismicity begin only after the fluid pressures had increased significantly?	Yes	Yes
3c	Is the observed seismicity explainable in terms of current models relating fluid injection to fault activity?	Yes (?)	Yes (?)
Score (“Yes” = 1.0, “Yes ?” = 0.5, “No or ?” = 0)		5.5	5.0
% (score/ 9.0)		61	55

The number of positive answers together with the results of the location method indicate that the event 1 (16:55:28 UTC) has a probability of approximately 61% to have been induced by the hydraulic fracturing operation, whereas the second event (22:49:36 UTC) has a lower probability (55%).

4.6. **Discussion**

One of the advantages of using tremor-like synthetics is that it enables to evaluate the behavior and performance of the algorithm in a more realistic environment where traces are no longer clean impulsive signals but noisy train waves in which the different phases are not easily distinguishable.

BPI methods work at their best when events are properly recorded at several stations, or at least two in the worst-case scenario. In our case study, apparently only one station registered each event, consequently, we applied some techniques such as rotation and filtering to seismograms to expose coherent signal masked by noise, unfortunately without success.

Using the absolute value of the seismograms is a common practice in seismology to avoid polarity artefacts, because all energy sums regardless of their polarity, nonetheless, noise also adds up. Besides, using RAW seismograms may help in noise suppression due to destructive superposition, however, incorrect velocity models may prevent true events from adding up.

If seismograms have not been previously corrected for geometrical spreading, normalization of the traces may be a quick solution for such a purpose, however, this process will also raise noise levels coming from stations where the event was not detected.

Confusion matrixes are powerful methods not only to summarize information effectively but also to classify results in such a smart way that induces a perspective shift of the problem, therefore, facilitating a much better interpretation.

4.7. **Conclusions**

PrAS showed good results in the synthetic scenarios, outperforming a classical method like MATF for location of local events. PrAS was able to localize 50% of the events within a radial distance of 0.15km and 100% within 0.40km when backpropagating through velocity models with a ± 25% error. Whereas MATF had radial distances of 0.20km and 0.65km, respectively.

Centroid-based methods are more robust to incorrect velocity models and discretization dependency than grid-search-based methods, however, they should be supervised to avoid artefact generation (although the rate is low).

Confusion matrixes are powerful methods not only to summarize information effectively but also to classify results in such a smart way that induces a perspective shift of the problem, therefore, facilitating a much better interpretation.

The 2 recorded events have around 60% chance of having been induced by the fracturing stimulation, however, the question cannot be completely answered using only the methodology proposed here. Therefore, we suggest the use of techniques such as polarization analysis, reverse

time migration or full waveform inversion, which account for all seismic phases, to either increase or reduce the uncertainty in these results.

4.8. Main contributions

- Construction of a valid layer cake velocity model for the San Francisco Field.
- Application of the proposed Back-Projection Imaging method to Real data.
- Adaptation of Binary Classification Tests (Confusion matrix) to quantify velocity uncertainty and determine velocity confidence intervals for location of seismic events.

5. General Conclusions

An extensive overview of the techniques (see **figure 82**) for seismic and microseismic event location available in literature indicates that, although Back-Projection Imaging is a less accurate version of Time Reversal Imaging, it is more suitable for sensitivity analysis and real time applications since it does not require solving the full wave equation, therefore, being less computationally expensive. On the other hand, the absence of distinguishable phases in LPLD and tremor-like waveforms presents a strong limitation for pure kinematic techniques and makes them impractical.

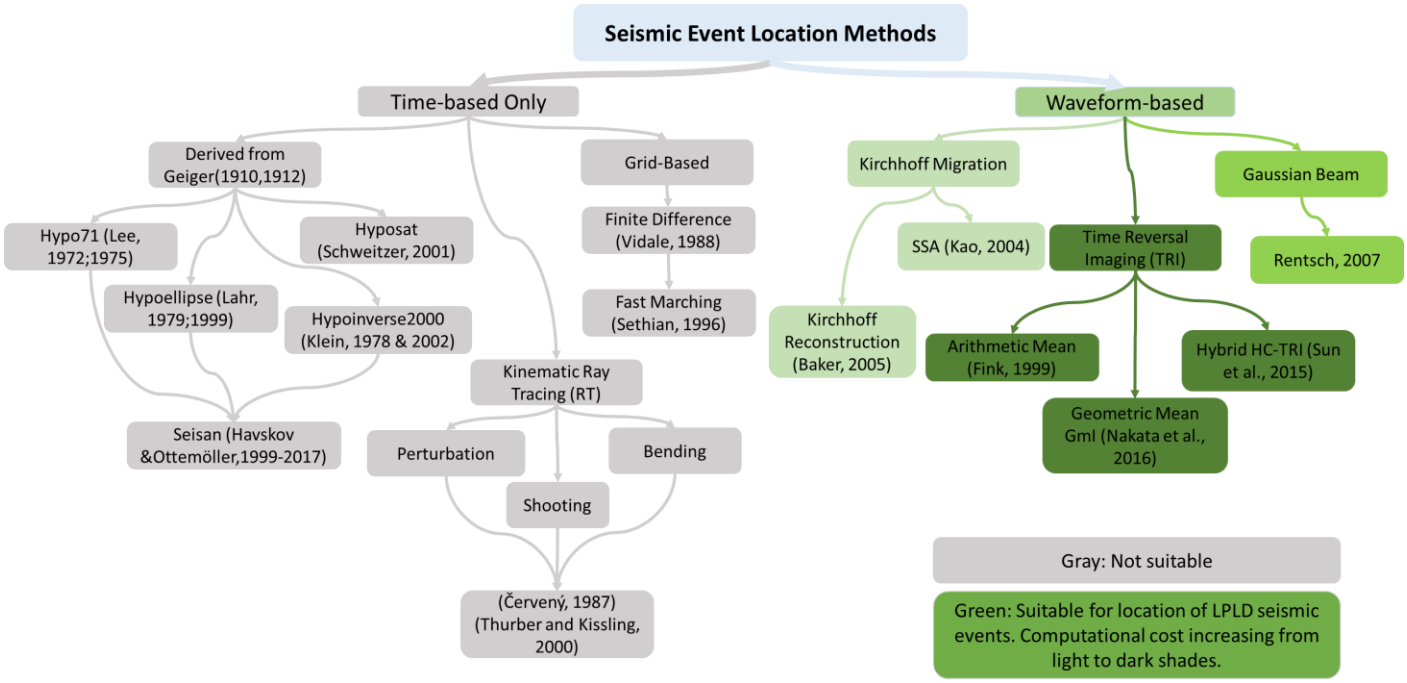


Figure 82. A review of suitable and not suitable methods for LPLD event location.

BPI was chosen as the basis for the design of the location algorithm. Such an algorithm allows to successfully locate the origin time and hypocenter of a wide variety of seismic sources including noisy and noise-free explosive sources as well as more complex waveforms. The algorithm can also determine the spatiotemporal location of LPLD events in which the absence of identifiable phases is a huge challenge. Furthermore, this method can produce a decent estimate of the solution even with a coarse discretization.

BPI have proven to be highly flexible for sensitivity analyses. First, results indicate that this method can successfully process noisy data, in addition, seismic signals with SNR <1.0 can be

located if a proper term to account for radiation pattern is included. Second, the peak frequency controls the attainable resolution (e.g. the higher the frequency, the more resolution; however, a more precise velocity model is required). Third, the sensor layout plays a crucial role in location; in general, the z-coordinate showed the highest error in location when the event originated inside the recording array; nonetheless, the error in the horizontal directions (x ,y) became more relevant at long distances away from the array. Last, incorrect velocities may prevent energy from focusing thus resulting in missed events. Additionally, an origin time – depth tradeoff was often observed in the results.

Two BPI methods (MATF and PrAS) were used to backpropagate synthetic LPLD events in order to evaluate realistic conditions in terms of high noise, absence of clear onsets and velocity uncertainty (using a Monte Carlo algorithm to sample random velocity models within a ± 20% range with a uniform distribution). Despite these unfavorable conditions, MATF and PrAS located 95% of the events within a radius of 0.46km and 0.32km from the true source, respectively. PrAS, the proposed method, outperformed MATF when locating local events, however, it was less effective as the source-sensors distance increased. Finally, it was noticed that the use of raw seismograms or their absolute value does affect the solution. The former causes noise suppression due to destructive superposition, provided that a correct velocity model is available, while the latter ensures that true events coherently add up even with opposite polarity (due to radiation pattern), nonetheless, noise also adds to the total.

Beyond the scope of this thesis, real field data containing 2 candidate LPLD events were backpropagated using the location methods. Although there is high uncertainty in the velocity model and the event was not recorded by enough stations to allow proper location, some interesting observations were made: first, the events represent neither distant events nor local noise, second, there is some energy focused in the volume of study which suggests that these events may have originated inside it, and third, there is a temporal correlation between the occurrence of the events and the injection of high pressure fluids. After analyzing these 3 criteria and others explained in the text, we conclude that there is around a 60% probability for the 2 events to have been triggered during the fracturing stimulation operation.

Grid searching based BPI methods such as MATF and SSA are stable and accurate when there exists a clear peak on the maximum brightness curve, however, a poor discretization alongside

complex brightness curves may make them impractical for use. Consequently, two new centroid based location methods are proposed in this thesis to deal with situations where more complex earth models, low SNR or unusual geometries are simultaneously studied. Furthermore, they introduce an unusual way to define the origin time of seismic phenomena composed of bursts of energy without clear phase arrival, in which a centroid-based approach may be a good estimate of the solution. The two methods (PbAS & PrAS) were proved to successfully compute the spatiotemporal location of single explosive sources, complex synthetics, and real data, while exhibiting a higher tolerance to velocity errors than SSA and MATF.

Please refer to each chapter for more detailed conclusions.

6. Recommendations for Future Work

Concerning the location algorithm:

- Extending the algorithms proposed here to detection and location in real time.
- Applications of the methods for multicomponent (3C) data to allow for the contribution of different seismic phases.
- A wider statistical uncertainty analysis especially related to recording geometries and SNR is suggested to better constrain the results and reliability of this method for further applications on field data.
- A similar workflow is encouraged to be done using the full wave equation to better understand its pros and cons compared to BPI methods.

Concerning LPLD seismic events:

- An exhaustive study on the source mechanism of LPLD events needs to be integrated in order to explain precisely the real phenomena occurring in the underground, which in turn will enable to generate more representative synthetic data.
- An integrated study between geophysical data and engineering data such as well logs, production curves, injection rates & stages, wellbore integrity, geomechanics and completion, would improve the results of the location algorithm.

References

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

Anikiev, D., Valenta, J., Staněk, F., & Eisner, L. (2014). Joint location and source mechanism inversion of microseismic events: benchmarking on seismicity induced by hydraulic fractuing. *Geophysical Journal International*, 249-258.

Artman, B., Podladtchikov, I., & Witten, B. (2010). Source location using time-reverse imaging. *Geophysical Prospecting*, 861-873.

Baker, T., Granat, R., & Clayton, R. W. (2005). Real-time Earthquake Location Using Kirchhoff Reconstruction. *Bulletin of the Seismological Society of America, 95*(2), 699-707.

Beskardes, G. D., Hole, J. A., Wang, K., Michaelides, M., Wu, Q., Chapman, M. C., . . . Quiros, D. A. (2018). A comparison of earthquake backprojection imaging methods for dense local arrays. *Geophysical Journal International*, 1986-2002.

Bland, H. C., & Hogan, C. (2005). A hypocentre location method for microseismicity in complex regions. *CREWES Research Report, 17*, 1-6.

Brudzinski, M. R. (2011). Episodic Tremor and Slip Potential Clues to the Earthquake Process and How Faults Slip. *The Earth Scientist, 27*, 7-12.

Červený, V. (1987). Ray tracing algorithms in three-dimensional laterally varying layered structures. In G. N. (Ed.), *Seismic Tomography* (pp. 99-133). Dordrecht: D. Reidel Publishing Company.

Coulter, G. R. (1976). Hydraulic Fractuing - New Developments. *The journal of Canadian petroleum*, 36-40.

Das, I. (2013a). *Long period long duration seismic events during hydraulic stimulation of gas reservoirs.* Stanford: Stanford University.

Das, I., & Zoback, M. D. (2011). Long Period, Long Duration seismic events during hydraulic fracture stimulation of a shale gas reservoir. *SEG*, 1473-1477.

Das, I., & Zoback, M. D. (2012). Microearthquakes Associated With Long Period, Long Duration Seismic Events During Stimulation of a Shale Gas Reservoir. *In SEG Technical Program Expanded Abstracts 2012*, 1-5.

Das, I., & Zoback, M. D. (2013b). Long-period, Long-duration seismic events during hydraulic stimulation of shale and tight-gas reservoirs _ Part 1: Waveform characteristics. *Geophysics, 78*(6), KS107-KS118.

Das, I., & Zoback, M. D. (2013c). Long-Period long-duration seismic events during hydraulic stimulation of shale and tight-gas reservoirs - Part 2: Location and mechanisms. *Geophysics*, KS97-KS105.

Davis, S. D., Nyffenegger, P. A., & Frohlich, C. (1995). The 9 April 1993 Earthquake in South-Central Texas: Was it Induced by Fluid Withdrawal? *Bulletin of the Seismological Society of America*, 1888-1895.

Eisner, L., Duncan, P. M., Heigl, W. M., & Keller, W. R. (2009). Uncertainties in passive seismic monitoring. *The Leading Edge, 28*(6), 648-655.

Fawcett, T. (2006). An Introduction to ROC Analysis. *Pattern Recognition Letters,* 861-874.

Fink, M. (1999). Time-Reversed Acoustics. *Scientific American*, 91-97.

Gajewski, D., Anikiev, D., Kashtan, B., Tessmer, E., & Vanelle, C. (2007). Localization of Seismic Events by Diffraction Stacking. *SEG Technical Program Expanded Abstracts*, 1287-1291.

Ge, M. (2003). Analysis of source location algorithms: Part II. Iterative methods. *Journal of Acoustic Emission, 21*(1), 29-51.

Geiger, L. (1912). Probability method for the determination of earthquake epicentres from the arrival time only. *Bull. St. Louis Univ., 8*, 60-71.

Hobro, J., Williams, M., & Calvez, J. L. (2016). The finite-difference method in microseismic modeling: Fundamentals, implementation and applications. *The Leading Edge, 35*(4), 362-366.

Hubbert, K., & Willis, D. G. (1957). Mechanics of Hydraulic Fracturing. *AIME Petroleum Transactions*, 153-168.

Husen, S., & Hardebeck, J. L. (2011). Understanding seismicity catalogs and their problems. *Community Online Resource for Statistical Seismicity Analysis.*, 1-35.

Jiang, Z., Bonham, K., Bancroft, J. C., & Lines, L. R. (2010). Overcoming computational cost problems of reverse-time migration. *GeoCanada2010 - Working with the Earth*.

Kao, H., & Shan, S. (2007). Rapid identification of earthquake rupture plane using Source-Scanning Algorithm. *Geophysical Journal International*, 1011-1020.

Kao, H., & Shan, S. J. (2004). The Source-Scanning Algorithm: mapping the distribution of seismic sources in time and space. *Geophysical Journal International,* 589-594.

Klein, F. W. (1978). Hypocenter location program HYPOINVERSE: Part I. Users guide to versions 1, 2, 3, and 4. Part II. Source listings and notes (No. 78-694). *U.S. Geological Survey*.

Klein, F. W. (2002). User's Guide to Hypoinverse-2000, a Fortran program to solve to earthquake locations and magnitudes (No. 2002-171). *US Geological Survey*, 1-123.

Kool, d. M., Rawlinson, N., & Sambridge, M. (2006). A practical grid-based method for tracking multiple refraction and reflection phases in three dimensional heterogeneous media. *Geophysical Journal International,* 253-270.

Kumar, A., Zorn, E., Hammack, R., & Harbert, W. (2016). Surface seismic monitoring of hydraulic fracturing activity in Pennsylvania and West Virginia. *Unconventional Resources Technology Conference (URTEC)*.

Kumar, A., Zorn, E., Hammack, R., & Harbert, W. (2017). Long-period, long duration seismicity observed during hydraulic fracturing of the Marcellus shale in Greene County, Pennsylvania. *The Leading Edge*, 580-587.

Lahr, J. (1979). Hypoellipse: A computer program for determining local earthquake hypocentral parameters, magnitude, and first motion pattern. *U.S. Geological Survey Open-File Report*, 79-431.

Lahr, J. C. (1999). Hypoellipse: A computer program for determining local earthquake hypocentral parameters, magnitude, and first-motion pattern. *US Geological Survey*, 1-119.

Lay, T., & Wallace, T. C. (1995). *Modern Global Seismology* (Vol. 58). San Diego: Academic Press.

Lee, W. H., & Lahr, J. C. (1972). Hypo71: A computer program for determining hypocenter, magnitude, and first motion pattern of local earthquakes (No. 72-224). *US Geological Survey*, 1-100.

Lelièvre, P. G., Farquharson, C. G., & Hurich, C. A. (2011). Inversion of first-arrival seismic traveltimes without rays, implemented on unstructured grids. *Geophysical Journal International*, 749-763.

Liao, Y.-C., Kao, H., Rosenberger, A., Hsu, S.-K., & Huang, B.-S. (2012). Delineating complex spatiotemporal distribution of earthquake aftershocks: an improved Source-Scanning Algorithm. *Geophysical Journal International*, 1753-1770.

Lowrie, W. (2007). *Fundamentals of Geophysics* (2 ed.). New York: Cambridge University Press.

Maxwell, S. C. (2011). What does microseismic tell us about hydraulic fracture deformation. *Annual Technical Conference and Exhibition* (pp. 30-45). Calgary: Society of Petroleum Engineers.

Maxwell, S. C., & Urbancic, T. I. (2001). The role of passive microseismic monitoring in the instrumented oil field. *The Leading Edge*, 636-639.

Nakata, N., & Beroza, G. C. (2016a). Reverse time migration for microseismic sources using the geometric mean as an imaging condition. *Geophysics*, KS51-KS60.

Nakata, N., Beroza, G. C., Sun, J., & Fomel, S. (2016). Migration-based passive source imaging for continuous data. *SEG International Exposition and 86th Annual Meeting*, 2607-2611.

Noble, M., Gesret, A., & Belayouni, N. (2014). Accurate 3-D finite difference computation of traveltimes in strongly heterogeneous media. *Geophysical Journal International*, 1572-1585.

Nolen-Hoeksema, R. (2013). Elements of Hydraulic Fracturing. *Oilfield Review Schlumberger*, 51-52.

Obara, K. (2002). Nonvolcanic Deep Tremor Associated with Subduction in Southwest Japan. *Science, 296*, 1679-1681.

Obara, K., Hirose, H., Yamamizu, F., & Kasahara, K. (2004). Episodic slow slip events accompanied by non-volcanic tremors in southwest Japan subduction zone. *Geophysical Research Letters, 31*(23).

Ottemoller, L., Voss, P., & Havskov, J. (2017). *SEISAN. Earthquake Analysis Software for Windows, Solaris, Linux and MacosX*.

Parikh, R., Mathai, A., Parikh, S., Sekhar, G., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophtahlmology*, 45-50.

Perez Altamar, R., & Marfurt, K. J. (2015). Identification of brittle/ductile areas in unconventional reservoirs using seismic and microseismic data: Application to the Barnett Shale. *Interpretation*, T233-T243.

Pesicek, J. D., Child, D., Artman, B., & Cieślik, K. (2014). Picking versus stacking in a modern microearthquake location: Comparison of results from a surface passive seismic monitoring array in Oklahoma. *Geophysics*, KS61-KS68.

Rawlinson, N., Hauser, J., & Sambridge, M. (2008). Seismic ray tracing and wavefront tracking in laterally heterogeneous media. *Advances in Geophysics, 49*, 203-273.

Rentsch, S., Buske, S., Lüth, S., & Shapiro, S. A. (2007). Fast location of seismicity: A migration-type approach with application to hydraulic-fracturing data. *Geophysics, 72*(1), S33-S40.

Royle, G. T. (2006). *Comparison of episodic and non-episodic non-volcanic tremors in the northern Cascadia subduction zone*. Department of Earth Sciences-Simon Fraser University.

Sava, P., & Vasconcelos, I. (2011). Extended imaging conditions for wave-equation migration. *Geophysical Prospecting*, 35-55.

Schweitzer, J. (2001). HYPOSAT - An enhanced routine to locate seismic events. *Pure and Applied Geophysics, 158*(1-2), 277-289.

Scotchman, I. (2016). Shale gas and fracking: exploration for unconventional hydrocarbons. *Proceedings of the Geologists' Association*, 535-551.

Sethian, J. A. (1996). A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences, 93*(4), 1591-1595.

Sethian, J. A. (1998). Fast Marching Methods and Level Set Methods for Propagating Interfaces. *Computational Fluid Dynamics, Annual Lecture Series, 29th*.

Sethian, J. A. (1999). Fast marching methods. *Society for Industral and Applied Mathematics, 41*(2), 199-235.

Sethian, J. A., & Popovici, A. M. (1999). 3-D traveltime computation using the fast marching method. *Geophysics*, 516-523.

Shearer, P. M. (2009). *Introduction to Seismology.* Cambridge: Cambridge University Press.

Sihna, D. P., Vishnoi, D. K., Basu, S., & Singh, V. P. (2009). A brief comparison of the efficacy of four migration algorithms - a sub-basalt example. *Geohorizons*, 24-27.

Sleeman, R., & Eck, T. V. (1999). Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Physics of the Earth and Planetary Interiors*, 265-275.

Stein, S., & Wysession, M. (2003). *An Introduction to Seismology, earthquakes and Earth Structure.* Malden MA: Blackwell Publishing.

Suárez, A. F., Gaviria, W., Pavas, J., & Frorup, M. (2005). Beating the Marginal Well Performance in a Mature Field: San Francisco Field in Colombia. *Society of Petroleum Engineers*.

Sun, J., Xue, Z., Fomel, S., Zhu, T., & Nakata , N. (2016). Full waveform inversion of passive seismic data for sources and velocities. *SEG Technical Program Expanded Abstracts*, 1405-1410.

Sun, J., Zhu, T., Fomel, S., & Song, W. Z. (2015). Investigating the possibility of locating microseismic sources using distributed sensor networks. *SEG Technical Program Expanded Abstracts*, 2485-2490.

Thomas, A. M., Beroza, G. C., & Shelly, D. R. (2016). Constraints on the source parameters of low-frequency earthquakes on the San Andreas Fault. *Geophysical Research Letters*.

Thurber, C., & Kissling, E. (2000). Advances in travel-Time calculations for three-dimensional structures. *Research Gate*, 71-99. doi:10.1007/978-94-015-9536-0_4

Valencia, E. E., Pereira, J. L., Medina, C. A., & Herrera, M. C. (2011). High Resolution Stratigraphic Controls on Rock Properties Distribution and Fluid Flow Pathways in Reservoir Rocks of The Upper Caballos Formation, San Francisco Field, Upper Magdalena Valley, Colombia. *Society of Petroleum Engineers*.

Vidale, J. (1988). Finite-Difference calculation of travel times. *Bulletin of the Seismological Society of America*, 2062-2076.

Vidale, J. E. (1990). Finite-difference calculation of traveltimes in three dimensions. *Geophysics*, 521-526.

Wang, H., Li, M., & Shang, X. (2016). Current developments on micro-seismic data processing. *Journal of Natural Gas Science & Engineering*. doi:10.1016/j.jngse.2016.02.058

Warpinski, N. (2009). Microseismic monitoring: Inside and out. *Journal of Petroleum Technology*, 80-85.

Yoon, K., Shin, C., Suh, S., Lines, L. R., & Hong, S. (2003). 3D reverse-time migration using the acoustic wave equation: An experience with the SEG/EAGE data set. *The Leading Edge*, 38-41.

Zecevic, M., Daniel, G., & Jurick, D. (2016). On the nature of long-period long-duration seismic events detected during hydraulic fracturing. *Geophysics*, KS113-KS121.

Zhebel, O., Gajewski, D., & Vanelle, C. (2011). Localization of seismic events in 3D media by diffraction stacking. *EAGE Conference and Exhibition incorporation SPE EUROPEC 2011*.

Zhu, J., & Lines, L. R. (1998). Comparison of Kirchhoff and reverse-time migration methods with applications to prestack depth imaging of complex structures. *Geophysics*, 0066-1176.

Zoback, M. D., Kohli, A., Das, I., & McClure, M. (2012). The importance of slow slip on faults during hydraulic fracturing stimulation of shale gas reservoirs. *SPE Americas Unconventional Resources Conference*.

Appendices

Appendix A. **Hydraulic Fracturing**

This technique was introduced to the petroleum industry 1948; however, it has been developed on a massive scale since the discovery of unconventional reservoirs such as shale gas & oil. Hydraulic fracturing or "fracking" is a stimulation technique used to increase hydrocarbon production from low permeability reservoirs which could not be produced otherwise. Two interesting papers about this topic are: (Hubbert & Willis, 1957) which explains the mechanics of hydraulic fracturing: how and why a rock breaks, orientation of the induced fractures, effects of injected fluids on the rock and more. (Coulter, 1976) discusses the theory and objectives of hydraulic fracturing, as well as considerations in designing this kind of treatment.

"A hydraulic fracture is a pressure-induced fracture caused by injecting fluid into a target rock formation" (Nolen-Hoeksema, 2013). The injection fluid or slurry is generally composed of 99.95% water and *proppant* (i.e. solids such as bauxite, resin-coated silica or ceramic that maintain fractures open once fluid injection is halted) and other chemicals (0.05%) including surfactants, corrosion inhibitors and biocides (Scotchman, 2016); this slurry is pumped into the formation at pressures exceeding the fracture pressure, then the rock breaks and the new fracture propagates parallel to the maximum principal compressive stress (i.e. either vertical or maximum horizontal stress) (**figure 83**). The continuous injection of high-pressure fluid drives fracture growth, allowing more fluid to penetrate the rock, consequently, creating an artificial reservoir with enhanced permeability.

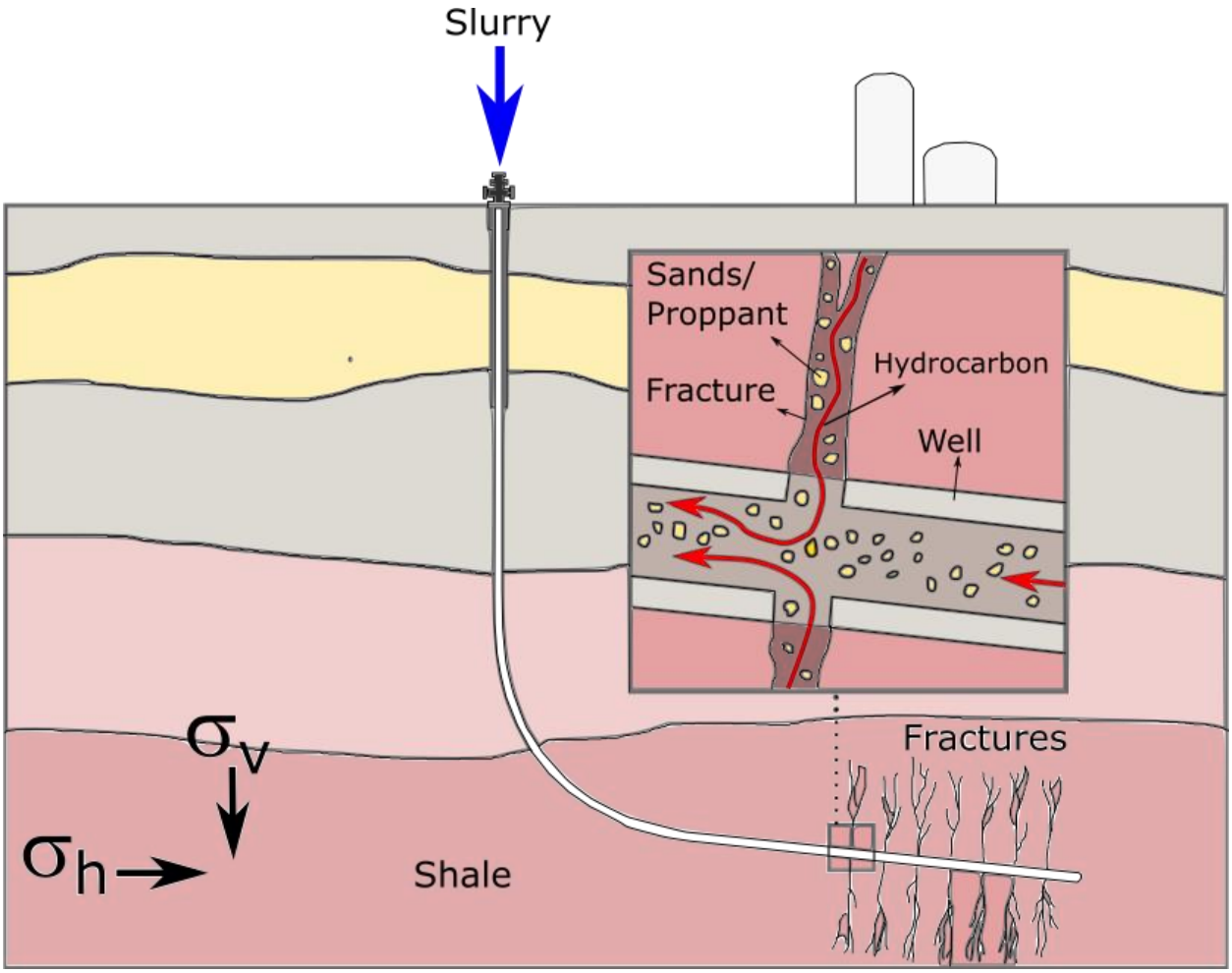


Figure 83. Schematic of hydraulic fracturing operation in a horizontal well.
Adapted from (Scotchman, 2016, p. 538)

Appendix B. **The Fast Marching Method.**

FMM are numerical schemes created by (Sethian J. A., 1996) for solving the nonlinear static Hamilton-Jacobi equations, among them, the Eikonal equation of the form:

$$|\nabla T(x,y,z)| = s(x,y,z), \qquad \text{Eq. 9}$$

Where: the slowness function, $s(x,y,z) > 0$ is provided as input to the equation and dictates how much the wavefront advances each time step. This method is unconditionally stable and produces consistent solutions for arbitrarily large gradient jumps in velocity.

It is possible to compute the travel time field $T(x, y, z)$ in a multi-dimensional grid using an upwind approximation to the gradient (Sethian J. A., 1999), thus obtaining the following scheme for the 3D case:

$$[max\left(D_{ijk}^{-x}T, -D_{ijk}^{+x}T, 0\right)^2 + max\left(D_{ijk}^{-y}T, -D_{ijk}^{+y}T, 0\right)^2 +$$

$$max\left(D_{ijk}^{-z}T, -D_{ijk}^{+z}T, 0\right)^2]^{\frac{1}{2}} = S_{ijk},$$

Eq. 10

Where, D_i^{-x}, D_i^{+x} the forward and backward operators, for a first order approximation, are given by:

$$D_i^{-x}\varphi = \frac{\varphi_i - \varphi_{i-1}}{\Delta x}, D_i^{+x}\varphi = \frac{\varphi_{i+1} - \varphi_i}{\Delta x},$$ Eq. 11

With φ_i being the value of the property under consideration (time in this case) in the i^{th} cell of the grid and Δx the grid spacing between two consecutive cells. Note: Similar nomenclature applies to the rest of variables in a multidimensional grid.

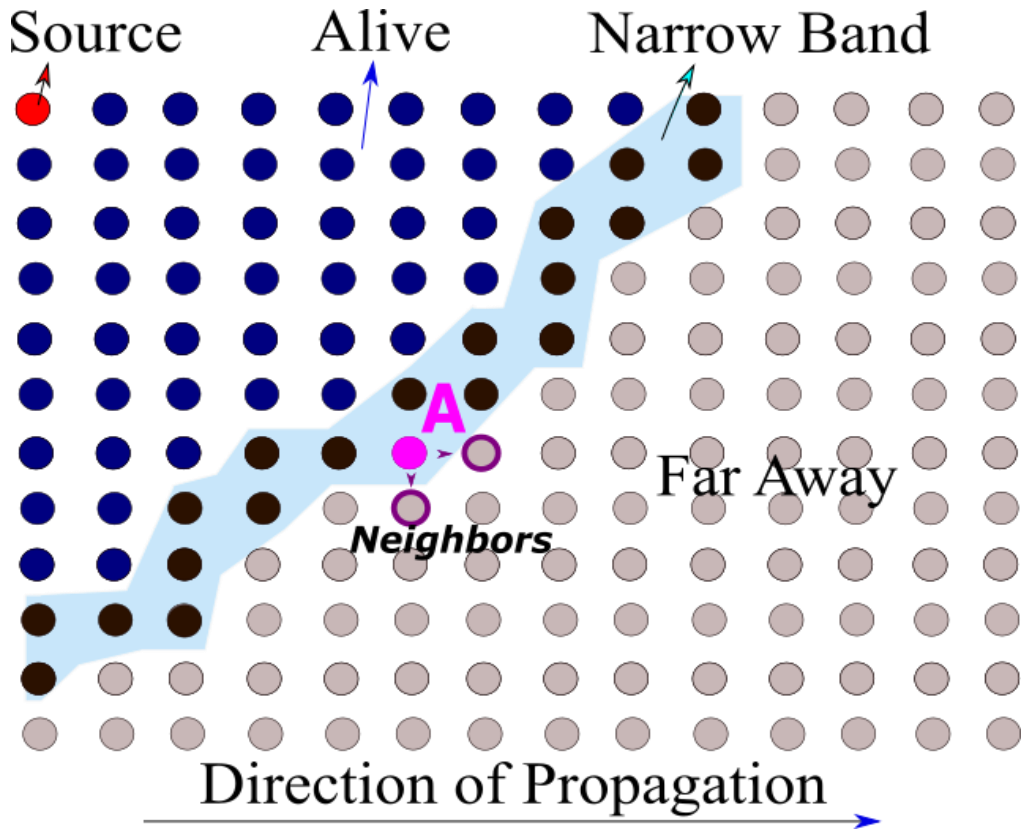


Figure 84. Wave-front propagation using upwind schemes.

A brief explanation of the process followed to compute the travel times in the whole domain is shown in **figure 84.** As the perturbation travels outward from the source, three regions can be easily identified: the first one (called ***alive***) comprises all the grid points whose values have been computed, accepted and will not change during the rest of the simulation; a second region, ***Narrow band***, mimics the propagating wave-front in which a minimum heap tree (see **figure 85**). acts to efficiently select the node with the minimum travel time value (for instance, ***A***), remove it from the list and add it to *alive.* Finally, the direct neighbors of A are removed from the third region ***Far away*** (or unknown) and added to *narrow band* where the travel-times are recomputed, and the loop is repeated until all points belong to *alive*.

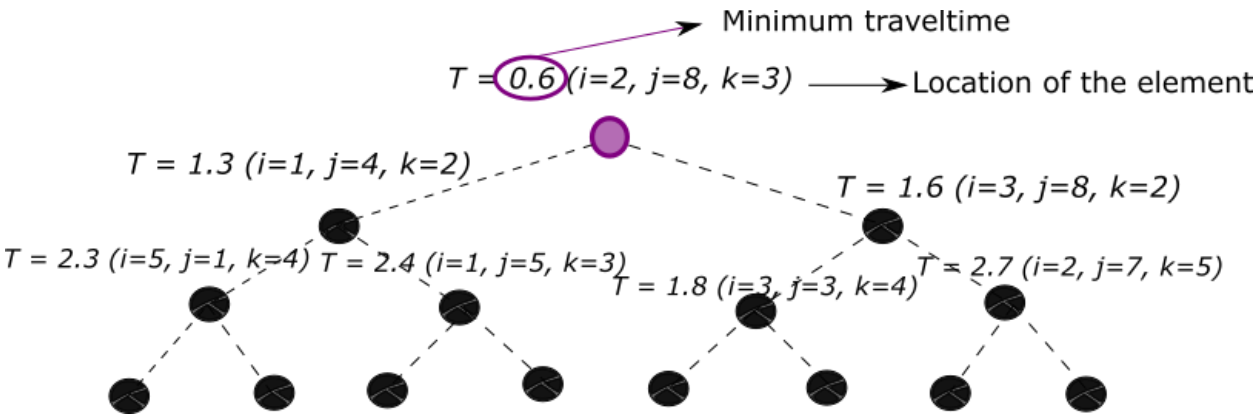


Figure 85. Min-heap tree for minimum traveltime determination.

According to (Sethian J. A., 1998): "The key to an efficient version of the technique lies in a fast way of locating the grid point in the narrow band with the smallest value for *u*". In that study, a minimum heap tree is used; such binary tree has the property that the value at any given node (travel times in our case) is less than or equal to the values at its children; therefore, the minimum time of the wavefront will always be on top of the tree and its removal is straightforward

A remarkable benefit of using this algorithm is that it allows for any shape of the initial wavefront, which can be useful for cases where the source is not a point source. Readers are encouraged to look at Sethian's work to gain a better understanding about the derivation of the fast-marching method, its physical meaning and how to use it in geophysical applications (Sethian & Popovici, 1999).

Appendix C. **Common Back-Projection Imaging (BPI) Methods.**

In general terms, BPI is based on time-shifting and stacking observed seismograms according to travel-time computation on a given velocity model. This back-projection step is usually carried out according to **Eq. 12** although there are several others including **Eq. 13**.

C.1. Maximum Amplitude Time Function (MATF)

$$F(\boldsymbol{r},t)=\frac{1}{N}\sum_{n=1}^{N}|u_n(t+ttt(\boldsymbol{r})_n)|, \quad (Eq.12)$$

Where $F(\boldsymbol{r},t)$: *Brightness/Image or Stacked Amplitude Function.*

u_n: *Normalized seismogram at station n=1, …, N.*

$ttt(\boldsymbol{r})_n$: *Travel times.*

t: *Start time of the time-window under evaluation.*

Finally, the origin time and location of the event are defined as the t and grid-point at which the "brightness function" reaches its maximum value.

$$maxF(t)=\max_{\mathrm{r}}\big(F(\boldsymbol{r},t)\big)$$

$\max_{\mathrm{r}}\big(F(\boldsymbol{r},t)\big)$: *Maximum stacked value per timestep.*

$maxF(t)$: *Maximum stacked amplitude curve.*

Origin time.

Using a grid-search technique to look for the maximum stacked energy.

$$B_{t0}=max\big(maxF(t)\big),$$

$$t_0=t[B_{t0}]$$

B: *Brightness or energy.* B_{t0}: *Highest energy value at origin time* (t_0).

Hypocenter.

Once t_0 *is known,*

$$B_{(r_0,t_0)} = max\big(F(\boldsymbol{r},t_0)\big) = B_{t0},$$

$$r_0 = r\big[B_{(r_0,t_0)}\big]$$

Other candidates

$$if\ F(\boldsymbol{r},t) \geq (B_{t0} * \Phi) \rightarrow (\boldsymbol{r},t), \qquad \Phi\text{: } is\ a\ given\ threshold$$

C.2. Source Scanning Algorithm (SSA) (Kao & Shan, 2004)

$F(r,t)$: Employs a time windows to account for erroneous time shifts due to uncertainty in the velocity model.

$$F(r,t) = \sum_{n=1}^{N} \left\{ \frac{\sum_{m=-M}^{M} W_m |u_n(t + ttt(r)_n + m\delta t)|}{\sum_{m=-M}^{M} W_m} \right\}, \qquad (Eq.13)$$

δt: Sampling Interval.

$2M$: Number of points within time window centered around the predicted arrival time.

W: Weighting factors which varies according to how far the arrival time of the energy departs from the predicted arrival time (PDF).

Afterwards, SSA follows the exact same procedure as MATF to search for the grid with the highest energy.

Origin time.

$$B_{t0} = max\big(maxF(t)\big); \qquad t_0 = t[B_{t0}],$$

Hypocenter.

$$B_{(r_0,t_0)} = max\big(F(\boldsymbol{r},t_0)\big) = B_{t0}; \qquad r_0 = r\big[B_{(r_0,t_0)}\big],$$

Other candidates

$$if\ F(\boldsymbol{r},t) \geq (B_{t0} * \Phi) \rightarrow (\boldsymbol{r},t), \qquad \Phi\text{: } is\ a\ given\ threshold$$

Appendix D. **Binary Classification Test or Binary Function.**

A common practice in the medical decision-making community, and recently extending to machine learning and data mining research is the use of statistical measures that allow not only to present the data but also to classify the results according to certain parameters such as success and failure rates. In other words, this kind of statistical tool helps to validate new procedures or techniques and to determine whether or not they measure what they are supposed to measure.

But, how does it work?

Well, consider a group of 200 people, a half of them have a particular disease (let's say cancer) while the others are disease-free, thus the first group are labeled as *Positives [P]* and the other *Negatives [N]*. When a new diagnostic test is applied to the individuals, there are four possible outcomes. If the subject **has** the disease and it is correctly identified, it is counted as a *true positive [TP]*, if not, it is counted as *false negative [FN]*. On the other hand, if the subject **does not have** the disease and it is correctly identified, it is counted as *true negative [TN]*, otherwise, *False Positive [FP]*. After the test is performed, results can be summarized in a two-by-two *confusion matrix* (**figure 86**) which helps to evaluate the usefulness of the diagnostic test (Fawcett, 2006).

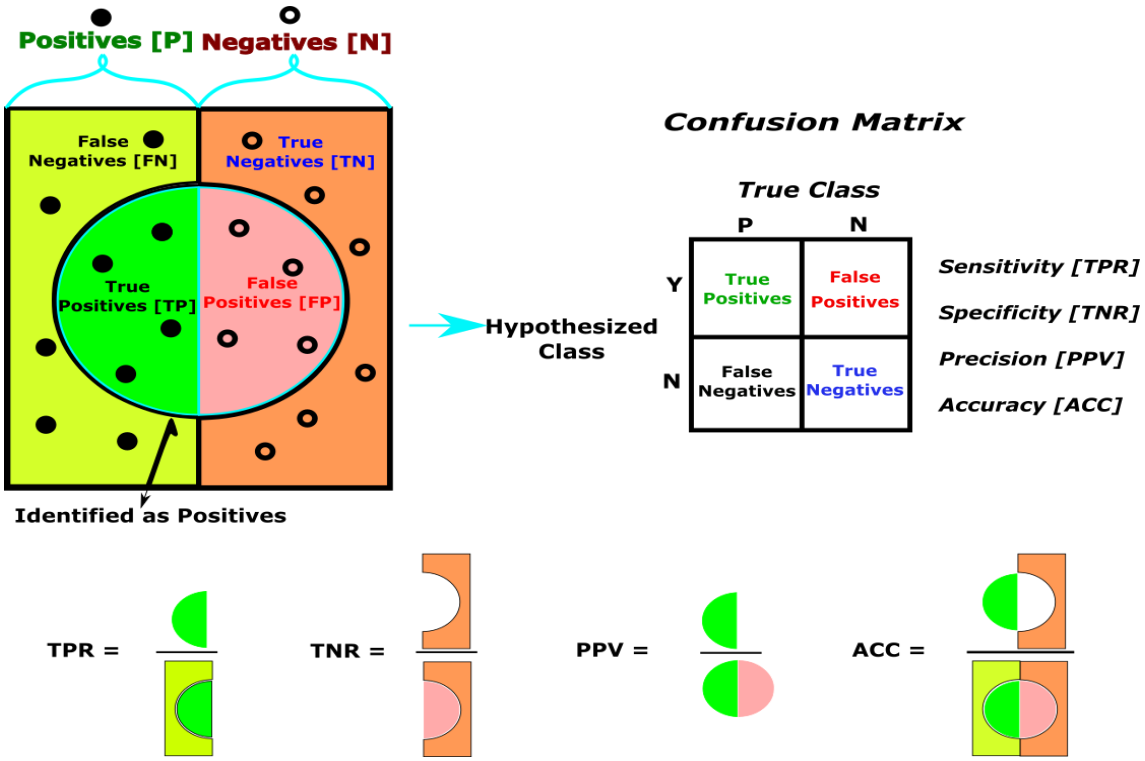


Figure 86. Confusion Matrix.

According to (Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008) the advantage of constructing such a matrix is that it allows to calculate several indicators to measure the validity of the diagnosis including:

Sensitivity (hit rate or True Positive Rate – TPR)

It is the ability of a test to correctly classify a Positive as such. Resuming the previous example, it measures the number of detected ill patients who do have the conditions.

$$TPR = \frac{TP}{TP + FN}$$

Specificity (or True Negative Rate – TNR)

The ability of a test to correctly classify individuals a disease-free (e.g. the percentage of healthy people identified as such).

$$TNR = \frac{TN}{TN + FP}$$

Precision (or Positive Predictive Value – PPV)

It measures how many of the subjects identified as ill actually are.

$$PPV = \frac{TP}{TP + FP}$$

Accuracy (ACC)

It measures the percentage of individuals correctly classified (either ill or not).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Finally, these parameters can be plotted in what are known as Receiver Operating Characteristics (ROC) graphs for visualization, design and/or decision-making purposes.