# Localization of subsurface seismic events on the San Jacinto faultusing a dense array of sensors

Chloé Gradon

Communauté
UNIVERSITÉ Grenoble Alpes

# THÈSE

Pour obtenir le grade de

## DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Sciences de la Terre et de l'Univers et de l'Environnement (CESTUE)**

Arrêté ministériel : 25 mai 2016


Présentée par

**Chloé GRADON**

Thèse dirigée par **Philippe ROUX, Chercheur,CNRS**
et codirigée par **Ludovic MOREAU, Enseignant-Chercheur, UGA**

préparée au sein de **Institut des Sciences de la Terre**
dans l'école doctorale **Terre Univers Environnement**


# Localisation d'évènements sismiques en proche surface sur la faille de San Jacinto à l'aide d'un réseau dense de capteurs

# Localization of subsurface seismic events on the San Jacinto Fault using a dense array of sensors

Thèse soutenue publiquement le **15 Janvier 2019**,
devant le jury composé de :

**Monsieur JEROME MARS**
Professeur, Université Grenoble Alpes, Président
**Monsieur NIKOLAI SHAPIRO**
Directeur de Recherches CNRS délégation Paris-Villejuif, Rapporteur
**Madame ALESSIA MAGGI**
Professeure, Université de Strasbourg, Rapporteur
**Monsieur YEHUDA BEN-ZION**
Professeur, University of Southern California, Los Angeles, Examinateur
**Monsieur JEAN VIRIEU**
Professeur émérite,Université Grenoble Alpes, Examinateur
**Madame STEPHANIE DURAND**
Professeure, Université de Lyon, Examinateur
**Monsieur LUDOVIC MOREAU**
Maitre de Conférence, Université Grenoble Alpes, Co-Directeur de thèse
**Monsieur PHILIPPE ROUX**
Directeur de Recherches CNRS délégation Alpes, Directeur de thèse

# Thanks

I would like to thank my directors, Ludovic and Philippe for their guidance and support during these three years. Yehuda, for his help throughout this PhD, as well as his team for welcoming me at USC and sharing their scientific inputs. Many thanks to my friends from the lab, especially Philippe, Capucine, Blandine and Anne who shared my office and all the joys and hardships of this PhD adventure. Thanks to my rugby teammates who provided a much needed outlet outside of the lab. I also thank my family and Ben, of course, for the constant support, especially during the writing months.

# Résumé

Cette thèse traite de la détection et de la localisation de sources autour de la faille de San Jacinto. Son but était de détecter des sources dans la croute superficielle, sur des profondeurs de l'ordre de quelques kilomètres. Ces sources ont une faible énergie et émettent principalement dans les hautes fréquences. Les sources à la surface autour et sur le réseau possèdent les mêmes caractéristiques et sont aussi étudiées afin de pouvoir les séparer des évènements en profondeur.

Une méthode basée sur le traitement d'antenne, le Match Field Processing (MFP), est utilisée pour détecter et localiser de faibles évènements à faible profondeur et à la surface. Le MFP est appliqué a des données mesurées grâce a un réseau dense de capteurs une composante déployés sur une zone de 600mx600m sur la faille de San Jacinto. La méthode a d'abord été testée sur un ensemble d'évènements à la surface et en profondeur. Nous appliquons ensuite la technique sur 26 jours de données, afin de déterminer si des évènements sont présents en proche surface. Pour cela, seules la position en surface de la source et la vitesse apparente des ondes émises sont utilisés comme paramètres d'inversion. L'utilisation de ces trois paramètres permet de réaliser une première étude à moindre coût de calcul. Cependant cette première inversion ne permet pas de conclure sur la présence de sources en proche surface. L'information sur la postions de la source en profondeur est nécessaire. Les résultats de localisations qui incluent la profondeur comme paramètre étant peu concluants lorsque le modèle classique de vitesse homogène est utilisé, nous étudions ensuite différentes stratégies pour améliorer la résolution en profondeur sans augmenter le coût de calcul.

# Abstract

The focus of this thesis is the detection and localization of weak sources on the San Jacinto Fault Zone. The primary targets of interest are sources in the shallow crust, with depth down to a few kilometers. For sources at these depths high frequency content and low energy are expected. Surface sources present on and around the array site are also studied in order to discriminate them from weak seismic sources at depth.

We rely on a methodology based on array processing to detect and localize shallow and weak seismic events in the fault zone complex environment. We use Match field Processing on data recorded from a dense array of 1108 vertical component geophones in a 600m x 600m area on the Clark branch of the San Jacinto Fault. We first test the method on a set of chosen events at depth and at the surface. Source epicentral positions and associated apparent velocities are then inverted for surface and seismic sources for 26 days, with the intention of determining if shallow sources are present. Inverting only for these three parameters is less expensive in terms of computational cost and is suitable for a first approach. However, this first inversion leaves us unable to conclude on the presence of shallow sources. As the resolution at depth is insufficient when all three source coordinates are inverted with a classical homogeneous velocity model, we finally investigate strategies to improve resolution at depth without increasing computational cost.

# Table of content

# List of figures

# List of tables

# General introduction

## Sub-surface study

The goal of seismology is to study the structure and behavior of the Earth interior, using seismic waves propagating in the ground to gain knowledge on an otherwise inaccessible object.

Seismology can be separated into two main subjects. The first is the study of the sources emitting seismic waves. The main sources of interest are of course earthquakes, but other sources at the surface are also studied (Inbal *et al.* (2018), Meng et Ben-Zion (2018a), Riahi et Gerstoft (2016), Retailleau *et al.* (2018)), mainly because they emit signal that can hinder the study of sources at depth. The second subject of interest of seismology is the imaging of the Earth interior. Among the most popular techniques used are tomography and migration.

Those two subjects are deeply linked as the quality of the results in one often depends on the a priori knowledge of the other. As an example, quality of earthquake tomography depends on the good knowledge of the sources emitting the signal used as an input (Zhang et Thurber (2003)). On the other hand, inverting for source position and mechanism often requires a-priory information on the medium between the source and the receivers measuring the signal of interest (Shearer (1997)).

The study of the Earth using seismic wave covers a wide range of scales. On one hand studies in global seismology go down to the core of the planet thousands of kilometers under our feet. On the other hand, exploration geophysics campaigns can focus on the first tens of meters bellow the surface. Paradoxically, the part of the Earth least studied in academic seismology is the closest to us, i.e. the very shallow part of the crust.

Several factors are responsible for this. The first one stems from material limitations. Historically, the sensor technology did not allow for the detection of small earthquakes that would occur in the shallow part of the crust, or measure the easily attenuated higher frequencies necessary to obtain the resolution necessary to it. The development of more sensitive sensors allowed the measurement of increasingly weaker signals leading to studies at shallower depth up to a point were the signal of interest has the same energy and frequency content as the surface noise around the sensors. At this point, the quality of the sensors cannot improve our capacity to separate signal from depth from surface noise.

Seismic arrays are a good solution to address this issue (Rost et Thomas (2002)), but the higher frequency content of the targeted signal requires dense sampling and a high number of sensors. The deployment of temporary dense arrays has only recently become more popular in the context of academic research thanks to the improvement of sensor technology, such as the development of wireless communication that facilitates installation and the decrease in costs. Those arrays give us access to datasets with dense spatial sampling approaching exploration geophysics scale. One can take full advantage of the noise-dominated part of the data, either by directly using pre-processed noise as an input for tomography, or by enhancing coherent signal from sources of interest.

Arrays have already been successfully used in high-resolution imaging and monitoring of faults zone and volcanic structures. The improved spatial coherency at higher frequencies has allowed the use of noise-based tomography to image finer details of the Newport-Inglewood Fault (Lin *et al.* (2013)) and the San Jacinto Fault (Roux *et al.* (2016)), with scales ranging from hundreds to tens of meters. Focal spot imaging (Hillers *et al.* (2016))has also been applied on the San Jacinto Fault. The improved resolution provided by dense arrays can also be used to detect smaller earthquakes and other sources of radiation (e.g. tremors and anthropogenic sources) than is possible with regional networks (e.g., Inbal *et al.* (2016) ; Li *et al.* (2018) ;Meng et Ben-Zion (2018a)).

Although the technical and methodological advancements allows the study of weaker and weaker source, no study has been performed to target earthquakes occurring in the top few kilometers of the crust. Because weaker seismic events occur more often, the study of these sources would allow the monitoring of the fault in a quasi-continuous manner. These very shallow earthquakes also occur in a very complex part of the crust. Improved detection and localization of sub-surface events can provide fundamental information on the dynamics and properties of the systems producing the sources, surrounding media and factors contributing to observed ground motion. We take advantage of a very dense temporary array installed on the San Jacinto Fault Zone to look for evidence of such events.

## San Jacinto Fault

California is often referred to as a seismic laboratory, its numerous active faults located around populated area making it a critical area of study and resulting in extensive instrumentation at various scales. In particular, several very dense arrays have been deployed for short periods of time in addition to the permanent regional networks.

Among the faults present in California is the San Jacinto Fault Zone (SJFZ)(Green box in figure 1 ) system that separate from the San Andreas Fault in Southern California at the level of Los Angeles. This fault is the most active fault of Southern California (Hauksson *et al.* (2012)), accommodating a large portion of the plate motion in the region. In particular, the areas known as the trifurcation area show the most seismicity, accounting for more than 10% of the seismic activity of Southern California in the last two decades(Ross *et al.* (2017)). Another feature of interest is the Anza Gap, where

a significant earthquake is expected. Induced tremors have also been reported. Those characteristics are the reason why in addition to the permanent seismic network, several temporary networks have been deployed over the last years on the SJFZ, offering an excellent coverage of the fault.

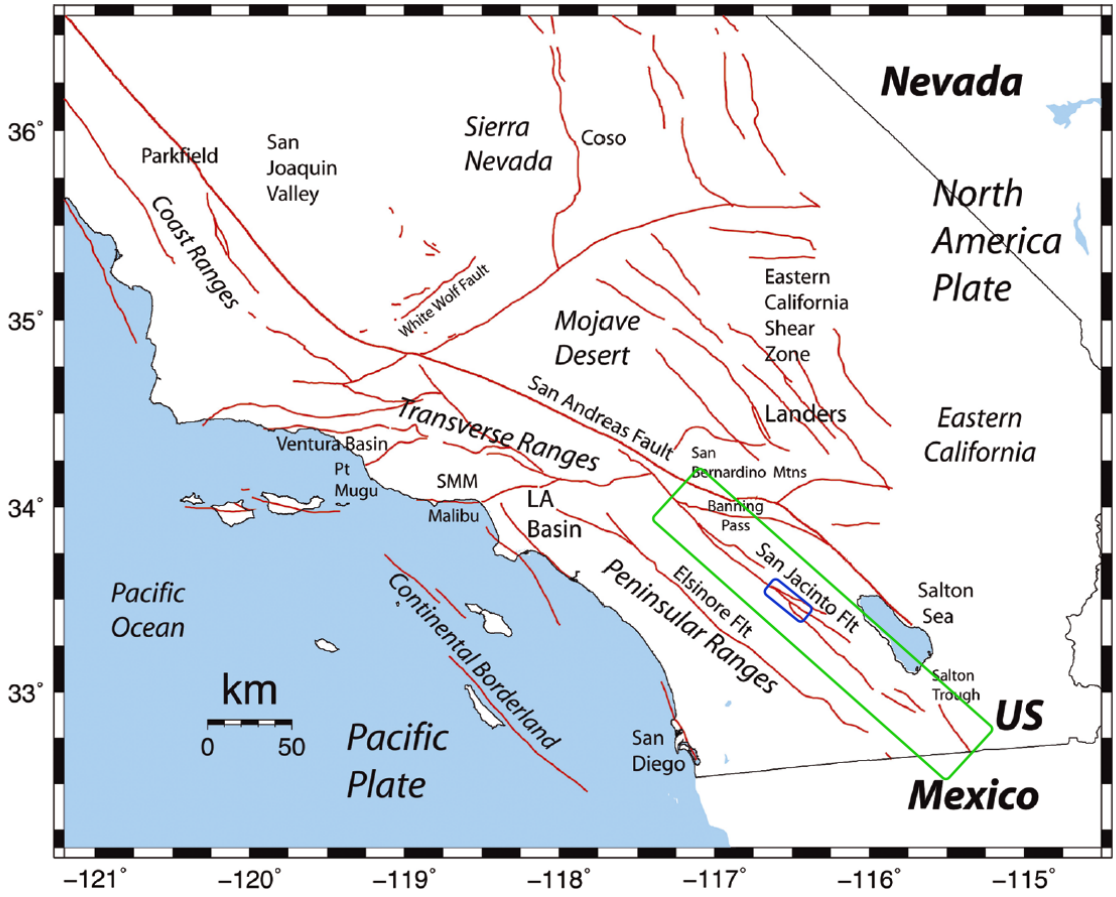


FIGURE 1 – Map showing the locations of major geographical and tectonic features in southern California. The green box surrounds the San Jacinto Fault Zone. The blue box is the Trifurcation area (modified from Hauksson (2011))

Our area of study is the Trifurcation area, a very complex zone where the Clark branch separates into three different offshoots (Blue box, figure 1). This complexity produces a high rate of low magnitude events that increases our chances of finding very shallow events, but also adds to the challenge of locating weak events. The lack of reliable a priori knowledge on the medium will impact our capacity to detect such weak, shallow events.

The goal of this work is to implement a processing method to reliably localize very weak event in a complex medium in order to investigate the presence of sources in the

top few kilometers of the crust.

## Manuscript layout

Chapter 1 presents the different methods developed for the location of weak events in seismology, as well as the optimization methods used to gain the efficiency necessary to process the data from more than a thousand sensors. Chapter 2 introduces the Match Field Processing method and its application to dense array data, with a focus on the efficiency of the data processing. The method is applied to a month of continuous data, and the results are studied chapter 3. Chapter 4 investigates strategies to improve the depth resolution of the proposed Match Field Processing process.

# Chapitre 1

# Chapter 1 : Localization methods

Source localization is a key interest of seismology and seismic exploration. Determining the origin of recorded seismic waves, i.e. determining spatial coordinates and time of origin of a source, is essential to the study of seismic (or volcanic) activity and to the monitoring of underground structures.

The study of micro earthquakes has been a subject of constant interest throughout the last eighty years. Since their detection was made possible by improved sensor technology in the late 50's, they have been studied first because of their greater numbers. Because of the scaling law that says that weaker events occur more often Gutenberg et Richter (1944), microearthquakes happen much more frequently than larger earthquakes and can be used to study the tectonic activity of faults in less time . They also have been studied with the goal of defining precursors in order to predict larger earthquakes. The interest of having many sources of precise known location was also an advantage in term of illumination for seismic imaging. In parallel, in the field of exploration geophysics, microsesismic event have been studied in order to control the integrity of the exploited underground structures such as mines (Young *et al.* (1992), Gharti *et al.* (2010)) and hydrocarbon reservoirs. More recently, the issues revolving around induced seismicity have generated a new interest in microseimic monitoring. The localization of weak events has also been developed for applications on volcanoes (Langet *et al.* (2014), Lomax *et al.* (2001)). Similar techniques have also been applied to glacial earthquakes study (Ekström (2006), Larmat *et al.* (2006)). This chapter will review the most commonly used methods to detect weak events in the fields mentioned above.

Most of the localization methods presented in this chapter rely on a-priori knowledge of the wave propagation in the medium. The first section introduces basic physical properties of waves on which the localization techniques are based. Section 1.3 will mostly focus on methods used for the localization of weak or micro-seismic events, after briefly presenting classical picking techniques. Optimization methods that can be included to improve computational efficiency are presented in section 1.4.

## 1.1 Wave Physics

This section will briefly introduce the key points of waves physics that will be mentioned in this manuscript. More details can be found in Aki et Richards (2002), Bruneau et Potel (2010) and Červený (2001).

### 1.1.1 Elastic wave equation for homogeneous isotropic media

A major part of the earth (i.e. the crust and upper mantle, as well as the core) is constituted of solid materials in which the waves emitted by a source propagate. Seismic waves are called mechanical waves because, contrary to electro-magnetic waves such as light, the propagation of energy is also accompanied by a local displacement of the medium they are passing through. In the case of elastic media, the equation describing this displacement is derived from the fundamental principle of dynamics and the generalized Hooke's law. In homogeneous solids, the displacement of any given particle in the medium is solution to the following wave equation :

$$\frac{\partial^2 \mathbf{u}}{\partial t^2} = \frac{\lambda + 2\mu}{\rho} \nabla.(\nabla.\mathbf{u}) - \frac{\mu}{\rho} \nabla \times \nabla \times \mathbf{u} + s \tag{1.1}$$

with $\mathbf{u}$ the displacement and $s$ the source term. The medium properties are described by the Lamé coefficient $\lambda$ and $\mu$ and the density $\rho$. As for every vector, $\mathbf{u}$ can be separated into scalar and vector potentials $\Phi$ and $\mathbf{\Psi}$ :

$$\mathbf{u} = \nabla.\Phi + \nabla \times \mathbf{\Psi} \tag{1.2}$$

with $\nabla.\mathbf{\Psi} = 0$ and $\nabla \times \Phi = 0$. We will show that those potentials can be associated with two different waves propagating in the medium.

### 1.1.2 Solutions of the elastic problem : Wave groups

Combining equation 1.1 and 1.2 leads to the following simplified forms of the wave equation, also called the d'Alembert equations.

$$\frac{1}{{v_p}^2} \frac{\partial^2 \Phi}{\partial t^2} - \Delta\Phi = 0 \tag{1.3}$$

$$\frac{1}{{v_s}^2} \frac{\partial^2 \mathbf{\Psi}}{\partial t^2} - \Delta\mathbf{\Psi} = 0 \tag{1.4}$$

$v_p = \sqrt{\frac{\lambda+2\mu}{\rho}}$ and $v_s = \sqrt{\frac{\mu}{\rho}}$ are respectively the velocity associated with the scalar potential, and the velocity associated with the vector potential. From these potentials we can derive two displacements $\mathbf{u_p}$ and $\mathbf{u_s}$ :

$$\mathbf{u_p} = \nabla.\Phi \tag{1.5}$$

$$\mathbf{u_s} = \nabla \times \mathbf{\Psi} \tag{1.6}$$

$\mathbf{u_p}$ and $\mathbf{u_s}$ are generated by two different waves constituting the body wave family. As $\nabla \times \Phi = 0$, $\mathbf{u_p}$ is irrotational and correspond to pressure waves (P-waves) traveling at velocity $v_p$. The displacement generated by these waves is parallel to the direction of propagation. Similarly, as $\mathbf{\Psi}$ has a null divergence, we can associate $\mathbf{u_s}$ with shear waves (S-waves) of velocity $v_s$, for which the displacement is normal to the propagation. S-waves can be seen as a combination of vertical (SV) and horizontal (SH) shear waves.

When the propagation medium is a half space, another wave family, the surface wave group, is present. They are called as such because, contrary to body waves, their amplitudes are effectively concentrated close to the surface. They are the results of interaction between P- and S-waves and their reflection at the boundary of the medium. The two different surface waves propagating in a semi-infinite medium are Rayleigh and Love waves. Surface waves usually have higher amplitudes than body waves. They are especially dominant when emitted by surface sources. Their velocity is lower than S-waves by a few percent, and they are dispersive in media with depth dependent properties, i.e. their velocity depends on their frequency.

Remark : Waves of different types (body waves or surface waves) are sometime referred as phases, especially in the context of classical localization techniques. In this manuscript we will call the different arrival on a seismogram "seismic phases" while "phase" will correspond to the phase of the Fourier transform of seismic data.

### 1.1.3 Harmonic waves

A set of solution (a base) to the d'Alembert equation is the harmonic waves group. As the wave equation is linear, it is possible to decompose each of its solutions as a sum of sines and cosines with varying frequencies f. If we consider the complex notation we can write :

$$\mathbf{u_{p,s}} = \sum_{\omega} \tilde{\mathbf{u}}(t, \mathbf{x}, \omega) = \sum_{\omega} \mathbf{A}(\mathbf{x}) e^{i(\omega t - \mathbf{kx})} \tag{1.7}$$

with $\mathbf{k}$ the wave number, $\omega = 2\pi f$ the pulsation and $\mathbf{x}$ the position The expression for the spherical wave, i.e. wave emitted by a point source in a 3D homogeneous isotropic medium is the following :

$$\tilde{\mathbf{u}}(t, r, \omega) = \mathbf{A}(r) e^{i(\omega t - kr)} \tag{1.8}$$

where $k$ is a constant scalar and $r$ is the distance between a point $M(\mathbf{x})$ of the medium and the source at position $\mathbf{x_0}$ :

$$r = |\mathbf{x} - \mathbf{x_0}|. \tag{1.9}$$

as $k = \frac{\omega}{v}$, for any wave velocity $v$,

$$\tilde{\mathbf{u}}(t, r, \omega) = \mathbf{A}(r)e^{i\omega(t-\frac{r}{v})} \tag{1.10}$$

At a fixed time t ($e^{i\omega t} = e^{i\omega\phi} = Cst$), we can consider only the displacement according to space :

$$\tilde{\mathbf{u}}(t, r, \omega) = \mathbf{A}(r)e^{-i\omega(\frac{r}{v}-\phi)}, \tag{1.11}$$

or,

$$\tilde{\mathbf{u}}(t, r, \omega) = \mathbf{A}(r)e^{-i\omega T(r)} \tag{1.12}$$

$T(r)$ is the time that the wave will need to propagate between two given points, called the travel time. Harmonic waves are also solution of the wave equation in inhomogeneous media. In this case, the wave groups presented previously are not easily defined. At high frequency, waves approximating P and S body waves are present. Using the high frequency assumption, travel time can be computed using the ray theory.

## 1.2 Ray theory

Ray theory describes the propagation of a wavefront. In the high frequency and large bandwidth limit, the waves can be represented by rays that follow a particular path in the medium and are described by the Eikonal equation :

$$|\nabla T|^2 = \frac{1}{v^2}, \tag{1.13}$$

where $v$ is either the local P-wave speed or the local S-wave speed and T is the travel time. This equation is derived from equation 1.1 using an infinite frequency approximation. Analytical solutions are available for specific and relatively simple cases. We will present here the case of a constant gradient with depth z such as

$$v(z) = v_0 + G(z - z_0). \tag{1.14}$$

$G$ is the gradient in the z-direction, $z_{src}$ is the depth of the source and $v_{src}$ is the velocity at source depth . In this case, the solution for the travel time is (Fomel *et al.* (2009)) :

$$T(x) = \frac{1}{G}arcosh(1 + \frac{1}{2}S(x)S_0G^2|x - x_{src}|), \tag{1.15}$$

with $S(\mathbf{x}) = \frac{1}{v(\mathbf{x})}$ and $S_s = \frac{1}{v_{src}}$, the slowness at position $\mathbf{x}$ and the slowness at the source position $\mathbf{x_{src}}$ respectively. The ray parameter $p$, which is constant along the ray can be derived from the epicentral distance X between the source and a point at position $\mathbf{x}$ (Meng et Ben-Zion (2018b)) :

$$p(\mathbf{x}) = \frac{2GX}{\sqrt{[(GX)^2 + {v_{src}}^2 + v(\mathbf{x})^2]^2 - (2v_{src}v(\mathbf{x}))^2}} \tag{1.16}$$

This ray parameter can be used to compute the travel time in the case of layered medium. From this equation we can also express the epicentral distance as a function of $p$, $v_s$ and $v(\mathbf{x})$ :

$$X = \frac{1}{Gp}\sqrt{(1 - pv_{src})^2 - (1 - pv(\mathbf{x}))^2}. \tag{1.17}$$

Using these solutions, it is possible to compute analytically, and thus efficiently, the travel time using more realistic medium than the homogeneous case discussed in the previous section. These modeled travel times are essential to most of the localization techniques described in section 1.3.

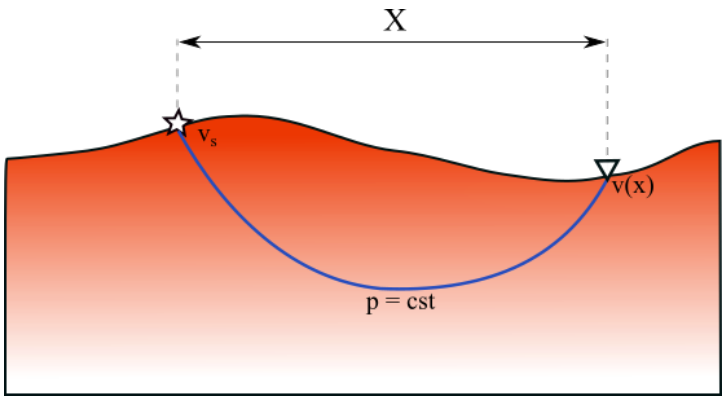


FIGURE 1.1 – Schematics of ray propagation in a 1D vertical gradient medium

## 1.3 Localization techniques

This section will go over a scope of localization techniques separated into phase picking techniques and array techniques. The two families of methods distinguish themselves by their need for clear arrivals on one hand and for dense spatial coverage (with many sensors) on the other. While this manuscript is centered on data from a dense array, skewing our attention towards array techniques, their driving principles often stem from more classical approaches and are often similar for both families of methods.

### 1.3.1 Phase picking techniques

The most common localization methods rely on the measurement of time of arrival of phases on different stations. From these time measurements, called picks, it is possible to determine the hypocenter position and time of origin of a seismic event. One advantage of those techniques compared to array processing methods presented in section 1.3.2 is the fact that only a few stations are necessary. These stations can be far away from each other as long as they all record events of interest. One the other hand, phase picking techniques require the prior detection of the events, which may raise some issues in terms of weak events or poor signal-to-noise ratio recordings.

#### 1.3.1.1 Absolute localization

The first historical localization method, described by Geiger (1912), is the following : for a possible source position, arrival times are estimated for each stations and compared to the measured times from the data. The possible position of the source is corrected iteratively so that the difference between computed and measured arrival times, also called residuals, is minimized. Because the starting point of the localization is supposed to be close to the true position of the source, it is possible to linearize the problem in order to converge quickly to a solution.

This method is based on the computation of synthetic travel times, the accuracy of which depends on the capacity of the velocity model to describe the true medium. If the model is too different from the reality, the final solution will not correspond to the true position of the source. The quality of the localization depends then directly on the knowledge of the studied area. In order to bypass this issue, relative localization methods, such as the popular Double Difference method (Waldhauser et Ellsworth (2000)), Figure 1.2A-b), have been developed.

#### 1.3.1.2 Relative localization

Contrarily to the absolute localization method, for which an absolute source position is determined separately for each event, the events are localized simultaneously in order to obtain their relative positions. If the events are close enough to one another, we can assume that the waves mainly follow the same path when propagating between the source and a given station. Using the difference between the arrival times of two neighboring events allows us to do away with the knowledge of the medium where the waves share a common path. These methods are based on the minimization of the residuals computed from the measured and synthetic difference of arrival time of two neighboring events. Thanks to this, the relative position is greatly improved. Such improvement is displayed figure 1.2A-a and 1.2A-b corresponding to absolute and Double Difference relocation respectively. One step further is to also correct for heterogeneities close to the sensor where rays do not share a common ray path. This can be done in the form of static

station terms. They are correction obtained from the misfits of all the events at one station. The Source Specific Station Term technique (Lin et Shearer (2005) ; Richard-Dinger et Shearer (2000)) computes correction terms associated with neighboring events for each station, obtaining corrections that depend on the relocated source position. An example of such relocation on earthquakes of the Imperial Valley is displayed figure 1.2B-b.

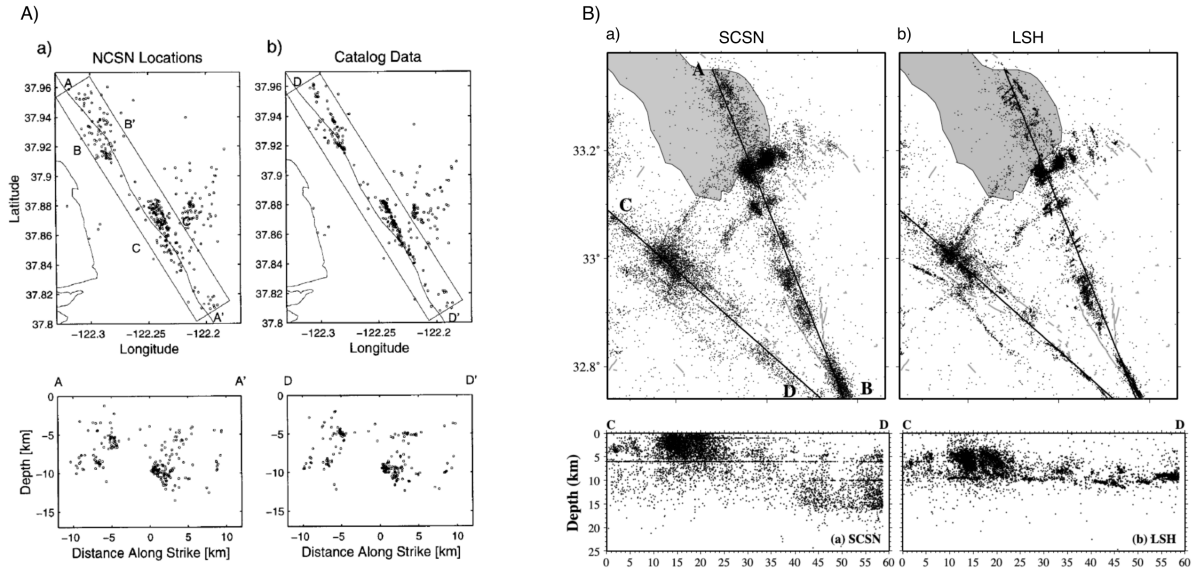


FIGURE 1.2 – Comparison of localizations obtained with absolute picking methods (A-a and B-a) with localizations obtained with the double difference method (A-b) and the Source Specific Station Term method (B-b). The study areas are respectively the Hayward fault and the Imperial Valley (Waldhauser et Ellsworth (2000) and Lin *et al.* (2007))

Despite these improvements, the quality of the localization is still dependent on the quality of the measurement of the time of arrival on the entirety of the stations of the network. The wrong labeling of a phase, or the absence of a visible seismic phase on some of the recordings of the stations lead to localization errors. The first can happen in the case when event are too close in time, in the case of longer duration signals such as tremors or events that occur with high frequency, and the second when seismic sources are too weak to be recorded by station farther away or when there is a strong noise source at the station. The use of cross-correlation can ensure better picking and detect events buried in noise, as long as they are similar to reference events called Master events or templates (Gibbons et Ringdal (2006)).

#### 1.3.1.3 Matched Filter Analysis

The Matched Filter Analysis (Shearer (1994) ; Shelly *et al.* (2007), relies on such principle. In this case, carefully chosen templates are cross-correlated against the studied dataset. The first way to localize events with Match Filter Analysis is to restrict the cross-correlation lags to a set that corresponds to travels distances associated with a spatial grid of possible source position. Grid points that are associated with high correlations are the most probable source localization (Meng *et al.* (2013) ; Shearer (1994)). Figure 1.3, from Shelly *et al.* (2007) shows the results of such technique on tremor in Japan. A simpler approach is to localize the slave event at the position of the master event, as the theory is that waveforms are similar if the sources are close in space.

Match Filter can also be used as a detector by continuously scanning the data (Gibbons et Ringdal (2006)). This process allows the detection of all events that have a waveform similar to the master event. Relative localization can then be performed with the master event as a reference, (Master Event technique (Ito (1985) ; Zollo *et al.* (1995))) or using other relative localization schemes. This method is really efficient in term of detection of low signal to noise ratio events. However, the localizations are limited to the areas where there is a known master event. The waveform of each template must also be sufficiently different so that slave events are not associated to more than one master event. And more importantly, this requires a pre-existing catalog of the targeted events. On the other hand, array techniques do not require a priori knowledge of the seismicity and are not restricted to a specific area.

### 1.3.2 Array processing techniques

The development of array techniques has been driven by the need for a sensitive, yet robust way of localizing seismic events. The amplitude of the small events that are the target of detection of such techniques is very weak, with signal-to-noise ratio (SNR) around and below 1. In addition to SNR issues, sources can be closely grouped in time, or are not necessarily impulsive, rendering traditional localization techniques based on phase picking ineffective. Array-based techniques that scan continuous data and utilize spatial coherency of the wavefield across the array provide extra gain that can allow detection and localization of weak events. A distinction can be made between time domain and frequency domain techniques. In the time domain, broadband signal are typically used, allowing the study of impulsive events with higher resolutions. The frequency domain approach is more adapted to longer duration signals or repetitive overlapping events where longer windows can be used.

Before detailing the array techniques used in geophysics, we will introduce some notions relative to array processing.

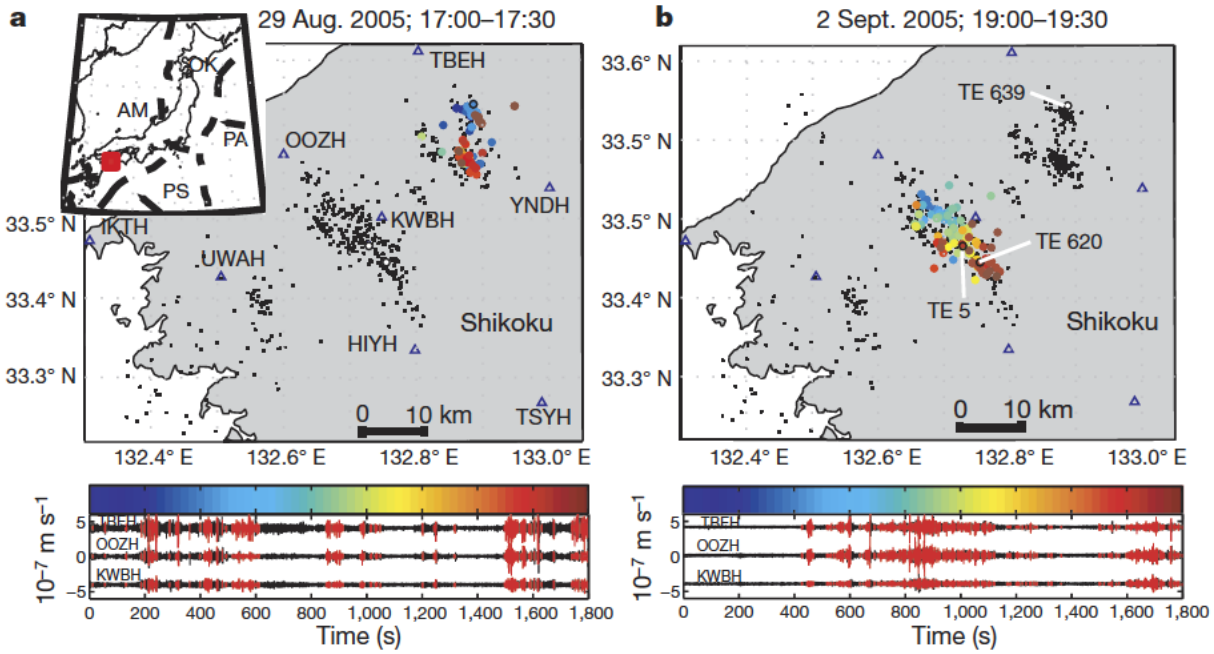


FIGURE 1.3 – Detection of LFE swarms forming tremor. a, Top panel ; map view of westernmost Shikoku, showing areas active (colored circles) during the 30-min period on 29 August 2005 beginning at 17 :00, color coded with time (see color scale below). Only 'strong' detections are shown. Note the clear spatial coherence of detected events with time. The spatial distribution of positive detections is not built into the detection algorithm, but emerges from the data. Black dots show epicentral locations of template LFEs used in this study. The depth of these events corresponds to the plate interface at 30–35km (ref. 6). Blue triangles denote station locations. Bottom panel shows east-component waveforms at three Hi-net stations, bandpass filtered between 1 and 8 Hz. Portions plotted in red indicate times with a detected event ("strong" or 'weak') similar to a template event. Inset, the regional tectonics, with the red box indicating the area shown in the main panel. PA, Pacific plate ; PS, Philippine Sea plate ; AM, Amur plate ; OK, Okhotsk plate. b, Same as a, but beginning 2 September 2005 at 19 :00. In this episode, a clear up-dip migration of the tremor source can be seen. The locations of template events (TE) are also labeled. (from Shelly *et al.* (2007))

#### 1.3.2.1 Array characteristics

The aim of this section is to introduce key points of array processing such as spatial coherency, spatial sampling and apparent velocity that will be mentioned in the rest of this chapter or in the following part of the manuscript.

### Spatial Coherency

The strength of array processing lays in the dense spatial sampling of the wave fields. For a range of appropriate frequencies, neighboring sensors can record similar waveforms if the signal is energetic enough. The measure of this similarity of the waveform on different sensors is the spatial coherency. If a signal is highly coherent on the sensors of the array, the contribution of all the sensors will add constructively. On the other hand, the contribution of incoherent noise will be reduced, improving the signal to noise ratio. The choice of distance between sensors is a trade of between measuring highly coherent signal of interest, while keeping unwanted, higher frequency signal incoherent.

### Spatial sampling

Observing a wavefield with an array corresponds to spatially sampling the ground movement at the array location. Consequently, the geometry of the array determines the range of wavelength that can effectively be recovered by the array. This has an impact on the localization performance of array techniques, particularly on resolution, which depends of the wavelength of the signal. The aperture of the array, i.e. its extent, constrains the longest wavelength that can be recovered. The actual boundary for the longest measurable wavelength depends on the desired resolution. However if the wavelength is much bigger than the aperture then the array behave like a single station.

The inter-sensor spacing constrains the shortest resolvable wavelength. Spatial aliasing will occur if the wavelength is smaller than the intersensor distance. This results in the strengthening of the side lobes in the array response. Those side lobes are local optima that do not correspond to the true position of the source, but are artifacts directly linked to the geometry of the array. As stated in the Niquist theorem, the wavefield should be sampled by more than two point by wavelength to be correctly recovered. In the case of sparser arrays, coherent array processing, where emphasis on the coherent part of the signal recorded by the array is made by processing cross-correlated data, is a way to improve the resolution.

Arrays are mostly composed of surface sensors. Only the ground motion at the surface is sampled. This means that the wavelengths observed at the array are not necessarily the true wavelengths of the wavefield but can be apparent wavelengths, if the incident wavefield arrives with a non-zero tilt (body waves). These apparent wavelengths are the projection of the true wavelengths onto the sensor plane (Figure 1.4).

### Apparent wavelength and apparent velocity

The wavefronts of a field emitted by a source will have a curvature in the near field. If the source is at depth, sensors at the surface will sample the projection of the wavefront

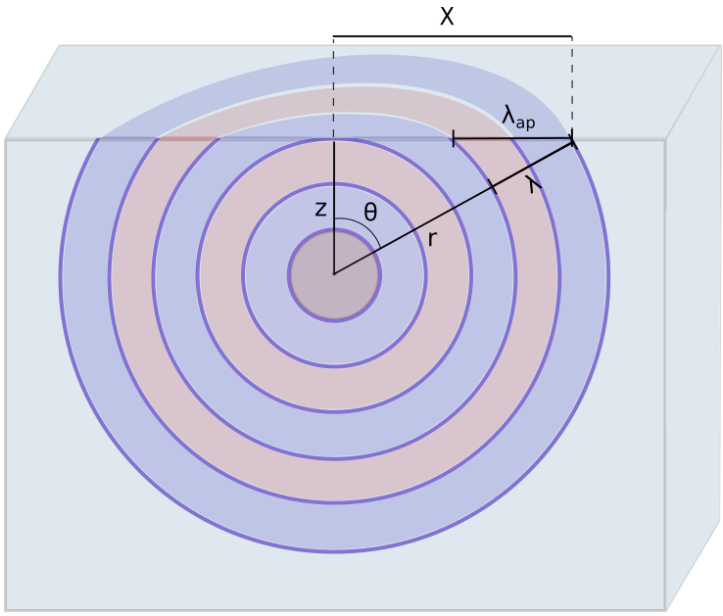


FIGURE 1.4 – 3D view of a wavefield emitted by a source at depth and its projection at the surface

onto the sensor plane. The variation in apparent wavelength $\lambda_{ap}$ depends on the angle $\theta$ between the direction of propagation and the normal to the surface.(Figure 1.4).

$$\lambda_{ap} = \frac{\lambda}{cos(\theta)} \tag{1.18}$$

This result induces increased apparent wavelength near the epicenter of the source when the array at the surface is close to the x-y position of the source epicenter ($cos(\theta) \sim 0$). This effect is reduced once the distance from the epicenter is greater than the depth of the source. As wavelength and velocity are linked through frequency, we can define an apparent velocity dependent on depth $z$, epicentral distance $X$ and the true velocity $V$. Indeed, with $cos(\theta) = \frac{1}{\sqrt{(1+(\frac{z}{X})^2}}$ :

$$v_{ap} = \lambda f \sqrt{1 + (\frac{z}{X})^2} = v\sqrt{1 + (\frac{z}{X})^2} \tag{1.19}$$

In the rest of this manuscript apparent velocity will refer to $v_{ap}$ , which is the velocity in direction of propagation of the wave projected at the surface.

#### 1.3.2.2 Graph clustering

Graph clustering consists in creating clusters according to the coherency of direct neighboring stations (Riahi et Gerstoft (2016)). A cluster contains all the stations that

are coherent with their closest neighbors at a given time. Cluster construction can be performed through network-based techniques, in which a network is defined by its vertices (here the array stations) and by edges that account for the connection between two given vertices. This connection is driven by the value of the coherency between two stations at a given time window. If the coherency between two stations meets the chosen acceptance criteria, they are connected. In short, graphs are all the vertices that are connected, directly and indirectly, i.e., all the neighboring stations that are sufficiently coherent. If we are in presence of weak sources with quickly decaying signal, we can assume that the position of the graph is analog to the source position.

While not commonly used in seismology, graph-clustering methods are a good example of the power of dense arrays. Taking advantage of the dense spatial samples they provide, this method is purely data-driven and does not require extensive a priori knowledge of the medium. As mentioned in the previous section, the errors on the velocity models necessary to most localization techniques can lead to localization errors. Riahi et Gerstoft (2016) used graph clustering to localize weak sources within the Long Beach array (Figure 1.5). While this method shows satisfying results in the case of dense array deployment and unknown or very complex medium, the resolution is much lower than for techniques relying on velocity models.

#### 1.3.2.3 Time reversal

In a completely different approach, time reversal relies heavily on models and the computation of full waveform synthetics. It is a broadband approach that consists of back-propagating time-domain recorded signals in a velocity model after they have been numerically time-reversed (Artman *et al.* (2010) ; Larmat *et al.* (2006), Larmat *et al.* (2008)). Due to reciprocity, the back-propagation focuses at the source position when the model is appropriate, with resolution capabilities limited by the diffraction limit even in complicated media. In theory, the source mechanism at the focusing point should also be retrieved. Such results were obtained with different level of success by Larmat *et al.* (2006) for glacial earthquakes data in Greenland. An example of one of the resulting locations is displayed Figure 1.6-b and 1.6-c with the data traces shown Figure 1.6-a. One of the main advantages of time reversal is that it exploits the entirety of the information contained in the waveform. Issues may arise with time-reversal techniques when several wave types (P-, S-waves and surface waves) all contribute to the recorded waveforms. In that case an elastic (rather than acoustic) model is required. The first issue is that the efficiency of time reversal focusing of seismic waves strongly depends on the 3D model accuracy (Kremers *et al.* (2011)). In the case of small events emitting at higher frequencies, the necessary high-resolution velocity model are not typically available. Another drawback is the fact that the use of 3D elastic models generally requires time-consuming numerical computations.

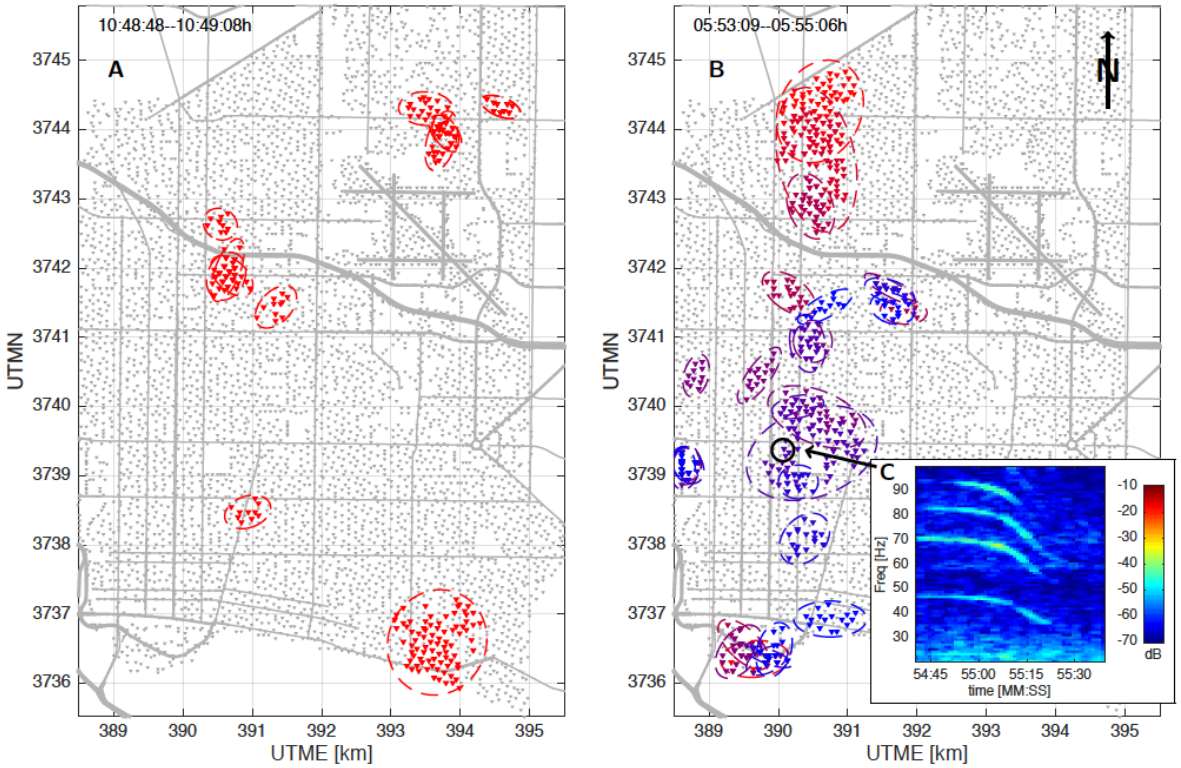


FIGURE 1.5 – Connected components of the array network were used to find coherent sensor clusters in the Long Beach geophone array. (A) The clusters at 20 Hz from four 10.2 s windows after 10.48h on March 11th. The spatial extend of the clusters (0 :5) is indicated by dashed ellipses (10). (B) A North-to-South helicopter transect is captured in a sequence of coherent clusters at 47 Hz over consecutive analysis time periods (starting at 05.53h). The colors change from red to blue as the analysis windows advance in time. The arrow points to the receiver from which the spectrogram (C) was computed around the time of increased coherence. See text for details. Coordinates are given in the Universal Transverse Mercator (UTM) coordinate system. (From Riahi et Gerstoft (2015))

#### 1.3.2.4 Time-migration/Stacking techniques

Stack-based methods can be seen as a simplified back-propagation of the signals, where the propagation is reduced to the time-shifting of an original trace and where the waveform variations are discarded. This simplification decreases considerably the computational costs. "Stack-based" techniques exploit the enhanced spatial coherency of arrays in order to reconstruct signal buried in noise. Two different approach can be used :

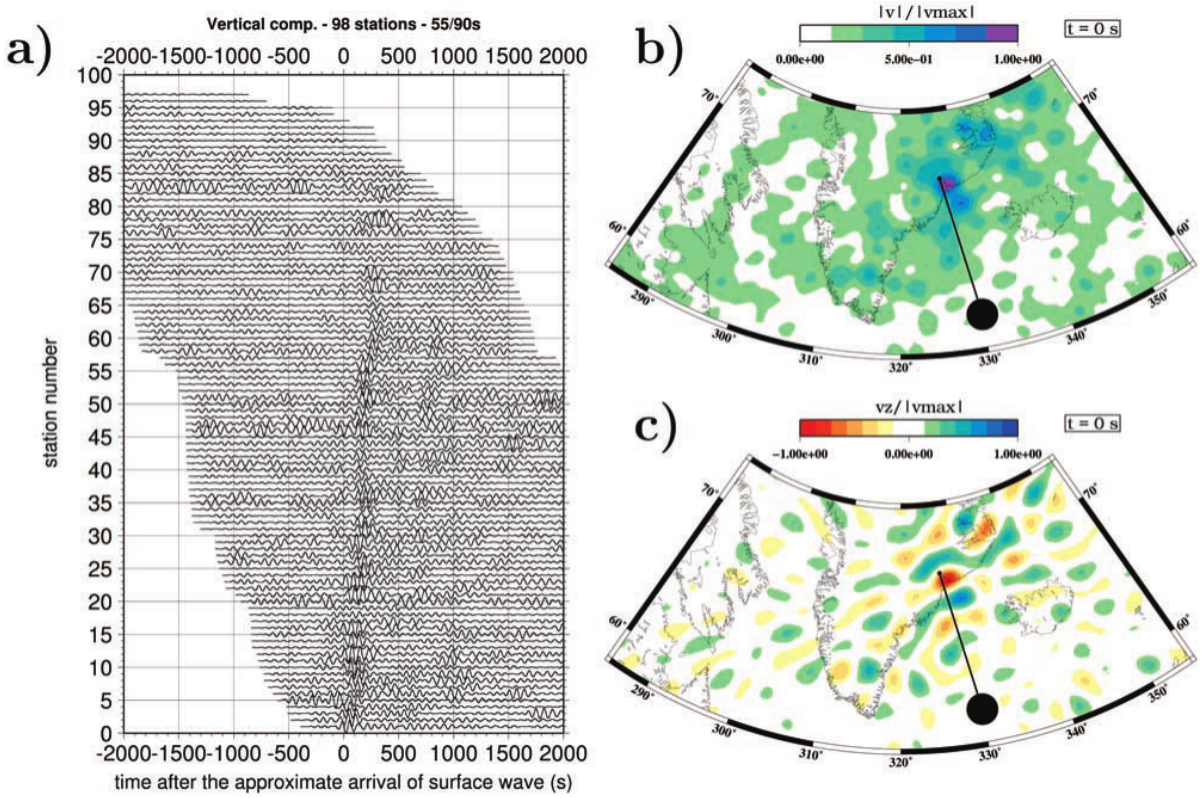


FIGURE 1.6 – The 28 December 2001, M = 5.0 Greenland event. (a) Data. (b) and (c) Adjoint wavefield (norm and vertical component of the velocity) when sending back vertical component data from 98 stations. The black beach ball denotes the Ekström et al. (2003) location. (Larmat *et al.* (2008))

1) Evaluate and sum the absolute amplitude of a single point per trace. 2) Shift and sum entire traces. The choice of trace point or time shifts depends on synthetic travel times associated with possible source localizations. These travel times are computed according to the nature of the problem and the distance from the sources to the array.

One of the main examples of the stack-based techniques is beamforming. Beamforming consists of scanning the angles of arrivals for planar incident waves and applying corresponding phase shifts to signals recorded at the array to identify the arrival angle and wave slowness that maximizes the coherency of the shifted signals (e.g., Landès *et al.* (2010) ; Rost et Thomas (2002)). Other methods such as back-projection (Ishii *et al.* (2005),Satriano *et al.* (2012)), which has been used to image the rupture of large earthquakes, seek to directly determine the source epicentral position and time of origin. In this case, potential longitude and latitude, and time of origin around the known source position are scanned and travel times from the possible source locations are derived through numerical velocity models. When doing so, the plane wave assumption used for beamforming methods can be discarded. This kind of approach is also suitable for

the study of micro-seismicity (Grigoli *et al.* (2016) ; Kao et Shan (2004) ; Langet *et al.* (2014) ; Poiata *et al.* (2016)), where arrays must be much closer to the sources in order to record relevant signals. Figure 1.7 shows one localization from Langet et al, with results of the inversion of tremor data measured on Piton de la Fournaise volcano for both source position and origin time.

Such methods are effective since the array gain classically compensates for low SNR, but they require dense spatial sampling to meet the Nyquist's criterion. Sufficient conservation of the wavefield on the array sensors is also necessary. Traces must be similar enough, i.e. not too distorted, to add constructively. Thus, these methods are less efficient in highly heterogeneous and attenuating media such as fault zones. The use of Characteristic Functions (CF) can counterbalance the effect of waveform distortion. CF have been introduced in order to enhance specific signal features and increase the sensitivity of the method.

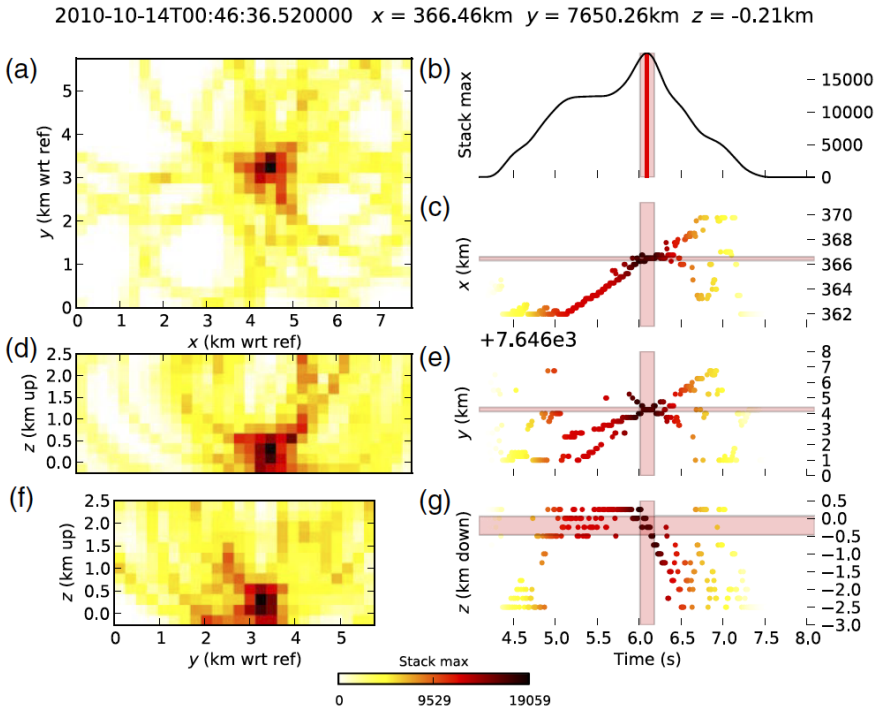


FIGURE 1.7 – Migration and localization using summary information of a volcano-tectonic event on Piton de la Fournaise (14 October 2010, 00 :46 :36 UTC). Layout of figure as in Figure 5. Uncertainty bounds are indicated by shading in (b), (c), (e), and (g). (from Langet *et al.* (2014))

The data is transformed into a characteristic function chosen to put an emphasis of certain features of the signal before being staked. Envelope, short-term average over long-term average (STA/LTA) or higher order statistical moments are typical transforms. Figure 1.8, from Poiata *et al.* (2016)) shows two common CF, kurtosis and envelope,

along with their spectral and multi-band filtered variants that were developed during the authors' study. The choice of the transform depends on the studied type of event. For example, envelopes have been used for tremor localization (Kao et Shan (2004) ; Poiata *et al.* (2016)), while STA/LTA and kurtosis are sensitive to impulsive events (Drew *et al.* (2013) ; Langet *et al.* (2014)). CF are an effective tool to improve the quality and the number of localization. However, they are often limited to targeting specific types of events. Using them for broader targets requires more complex processing (Poiata *et al.* (2016)) .

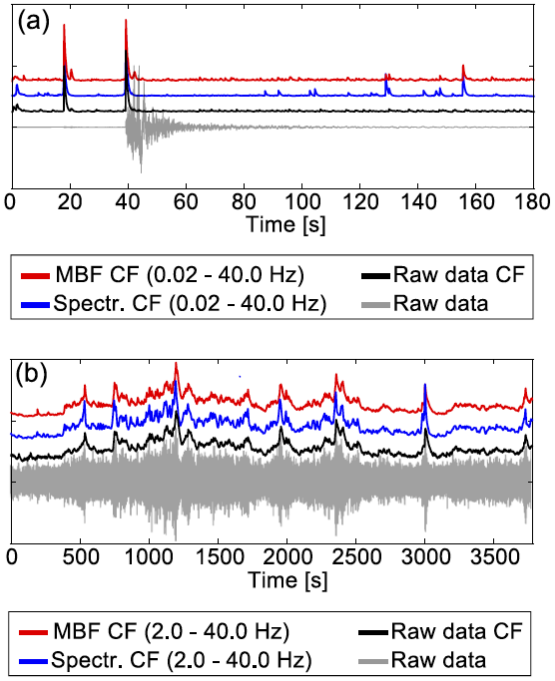


FIGURE 1.8 – (a) Comparison of the kurtosis CF (black line) for a raw recording of an earthquake (grey line), and the broad-band kurtosis CFs composed from the transformed spectrogram (blue line), and the MBF analysis (red line). (b) Comparison of the RMS envelope CF (black line) of a raw recording of a tectonic tremor (grey line), and the broad-band envelope CFs composed from the transformed spectrogram (blue line), and the MBF analysis (red line). (From Poiata *et al.* (2016))

#### 1.3.2.5 Match field Processing

This section briefly introduces the principle of Match Field Processing (MFP). More details on this method can be found in chapter 2.

MFP has originally been developed for monochromatic sources in ocean acoustics in order to take full advantage of the spatial coherency of a wavefield through an array and

include modal velocity models of the sea to improve the localizations. With the recent increase of dense seismic arrays, it has become possible to also apply this approach in the context of exploration (Corciulo *et al.* (2012)), hydrothermal geophysics (Cros (2011) ; Vandemeulebrouck *et al.* (2013)) and global detection (Gerstoft et Tanimoto (2007) ; Neale *et al.* (2017)).

MFP can be seen as an equivalent to the stack-based techniques in the frequency domain. It is a model-based inversion method where the phase and amplitude of a wavefield in array data is compared to the phase of a synthetic wavefield referred to as a replica. The best match between the data and the replica provides the best description of the seismic event given the model. Replica are commonly computed using closed form solution such as plane waves in an homogeneous model or, in the case of near field application, spherical waves in a homogeneous model. MFP can be seen as a relative localization method, as the data input corresponds to the phase difference and shared energy between sensors. Because of this feature, the inversion of the time of origin is not needed. It also diminishes the impact of errors in the velocity model, as waves emitted from a source will have mostly the same path before reaching neighboring sensors.

**Remarks** : All other methods described in the chapter are time-domain techniques. Working in the frequency domain, and especially in narrow frequency bands allows the separation of temporal and spatial structure of the signal. This results in improved localization performance compared to time migration techniques when there is a poor conservation of the waveform across the array, without resorting to using CF that limits the targets of study. The choice of the frequencies of study can also be a tool for source discrimination, as will be explained in the next chapter. While MFP was designed for numerical replica, empirical replica have been used in seismology (Harris et Kvaerna (2010) ; Wang *et al.* (2015)). The use of empirical replica does not require a velocity model to locate new event. In that case, the set of possible localization is limited by the positions of the known events.

As for many of the source localization methods presented in this section, MFP require the scanning of possible sources to determine the position that best matches the data. While each possible source position can successively tested, methods can be implemented to accelerate the localization process.

## 1.4 Optimization methods

Most of the localization methods described above are iterative processes where different possible source positions are evaluated. In order to find the most probable source position, we seek to find the maximum (or minimum) of an objective function. This function contains the source position likelihood for all the parameter space and is evaluated discretely for each candidate source position. The strategy for choosing the candidate sources is a trade of between a complete test of all the inverted parameters and computational efficiency, because of the non-linearity of the seismic source localization problem.

This non-linearity comes mostly from the complexity of the medium and the surface position of the sensors. There are three main approaches for choosing sources candidates : test as much of the parameter space as possible by uniformly sampling the possible source positions (grid search), use local or global optimization methods. The diversity of optimization methods used in geophysics is important. We will focus here on the techniques most commonly used in earthquake localization. More details on the methods presented here and other methods typically used in geophysics are available in Press *et al.* (2007) and Sen et Stoffa (2013).

### 1.4.1 Local optimization methods

Local optimization methods are the fastest way to converge toward a final solution. They are the historical optimization method for source localization because of their computational efficiency. They are used with the assumption that the unknown objective function can be approximated by simpler known functions e.g. linear or quadratic functions. This assumption is correct if the position of the source is close to the starting point of the optimization, so that the only possible solution will be the true position. Local optimization has been used extensively in phase picking methods, which often solve a simplified source localization problem where the travel-time residuals are linearly related to perturbations in the source position and origin time. The linearized equation is then solved using matrix factorization algorithms. Only a few iterations are necessary to reach a stable solution.

The main local optimization methods are gradient-based (Gill *et al.* (1981)). A schematic summing up their principle is displayed Figure 1.9, with several examples of 1D objective functions. In the case of the steepest descent method, the iterative steps toward the minimum are made in the direction in which the objective function changes the most rapidly. This corresponds to the direction of the gradient (or slope) of the objective function. Newton methods minimize a quadratic approximation of the objective function using the slope and the curvature (second derivative) at the current step.

The simplex method (or Nelder-Mead method (Lagarias *et al.* (1998) ; Nelder et Mead (1965)), that do not require the use of derivative, have been used in exploration geophysics. This method relies on the evaluation of the objective function at N+1 points forming a simplex (with N the number of dimensions of the parameter space). The worst point of the simplex, associated with the highest value of the objective function, is replaced by either its reduction or its reflection with respect to the barycenter of the simplex. The alternative with the lowest value is chosen. Figure 1.10 shows the different outcomes of an iteration of the simplex method. While slower than the previous methods, this technique does not make any assumption on the shape of the objective function, and is less expensive in term of memory use. More details on local optimization can be found in dedicated books.

Local optimization has been favored in earthquake localization, in particular with phase picking techniques that are mostly weakly non linear (Buland (1976)). However,

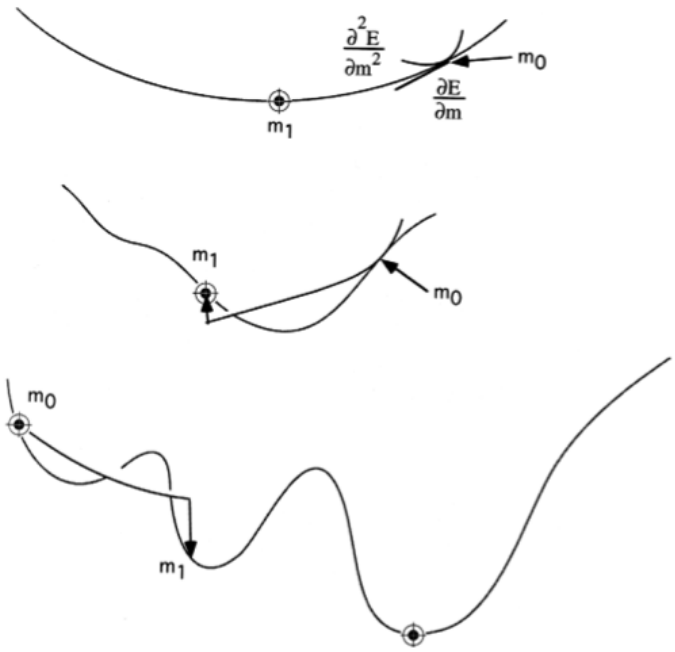


FIGURE 1.9 – Three 1D objective functions of varying complexity that are to be minimized. Gradient methods use the slope to find the nearest minimum, and Newton's method uses the slope (first derivative) and curvature (second derivative) to make a parabolic fit. For increasingly non-parabolic functions (middle), Newton's method may require more than one iteration to converge, and for extremely non-parabolic functions (bottom), it will fail unless the process is repeated from many different starting points. (From Sen et Stoffa (2013))

in the case of array techniques, the objective functions contain several optima, some of them artifacts coming from the so-called array response, which is linked to the geometry of the array. If the starting point is ill chosen, the convergence will occur around a local maxima, that is not necessarily the true source. The initial source candidate must then be chosen carefully, which can be problem in the case of monitoring, where detection and localization are performed together without strong a priori knowledge of the true source position.

### 1.4.2 Grid search

Another way to find the most probable position of the source, which, in practice, corresponds to finding the global optimum of our objective function, is to evaluate it for a set of position that evenly sample the scope of all the candidates. In this case, the chosen candidate sources are a grid of points that constitute an image of the objective function through all the parameter space. The spacing of the points corresponds to the

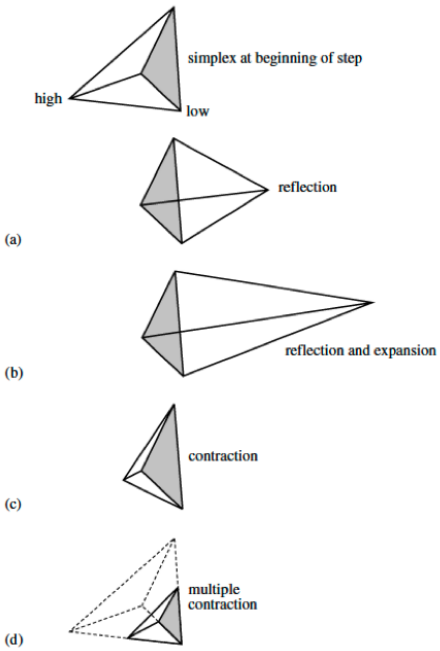


FIGURE 1.10 – Possible outcomes for a step in the downhill simplex method. The simplex at the beginning of the step, here a tetrahedron, is shown, top. The simplex at the end of the step can be any one of (a) a reflection away from the high point, (b) a reflection and expansion away from the high point, (c) a contraction along one dimension from the high point, or (d) a contraction along all dimensions toward the low point. An appropriate sequence of such steps will always converge to a minimum of the function. (Press *et al.* (2007))

resolution of the process. The advantage of this method is that it evaluates the entirety of the parameter space within the limits of the chosen resolution (see figure 1.7 for an example). The global optimum can then be found among all the other optima of the objective function. This method is commonly used in array processing techniques, but has also been applied to classical phase picking techniques (Shearer (1997)). The main drawback of the grid search is that the resolution must be enough to properly sample the maxima. This usually corresponds to an important computational cost, especially when the area of study is large or when the extent of the main maximum is small. To circumvent this problem, Kao et Shan (2004)) performed a preliminary study for each specific event and then used a reduced grid search approach round their first approximation. However this process cannot be employed in localization scheme where events are detected and located simultaneously, as it requires a selection of the events beforehand. A compromise between local optimization and grid searches is the global optimization methods.

### 1.4.3 Global optimization methods

Global optimization techniques are methods that seek to find the global optimum of an objective function without evenly sampling the parameter space. The goal is to reduce the number of iterations needed to find the global optimum by reducing the sampling in area of the parameter space where there is no optimum. Global optimization relies for a great part on stochastic methods, the so-called Monte Carlo methods, and in particular Markov Chain Monte Carlo (MCMC) methods (Sambridge et Mosegaard (2002). The MCMC sampling algorithm is an iterative process that follows a Markov Chain to determine candidate solutions in multi-dimensional parameter space and evaluate their likelihood (PDF of parameters) given the data. Candidates solutions are subsequently accepted or rejected, based on a Bayesian criterion that compares this likelihood to that of the last accepted solution. The most well known MCMC scheme is the Metropolis-Hasting method that comprises 3 steps :

**First step : Candidate draw.** The candidate for the next step is drawn randomly within a chosen step size around the current point.

**Second step : Computation of the likelihood of the candidate.** The likelihood of the candidate is computed using a zero-mean Gaussian likelihood function :

$$L(X^{cand}) = e^{-\frac{F(X^{cand})}{(2\sigma^2)}}$$

where $L(X^{cand})$ is the likelihood of the set of candidate parameter $X^{cand}$, $F(X^{cand})$, is the value of the function to minimize, and $\sigma^2$ is the variance.

**Third step : Assignation of the next point.** The value of the likelihood of the candidate is compared to the likelihood of the current point $X^n$ to determine the value of $X^{n+1}$ :

$$\begin{cases} X^{n+1}=X^{cand} & \text{and} \quad \frac{L(X^{cand}}{L(X^n)} > r \\ X^{n+1}= X^n & \text{Otherwise} \end{cases}$$

Where $r$ is a random number between 0 and 1. The Bayesian criterion does not automatically reject "worse candidates". This means that the optimization process can explore local minima without getting stuck. Note that if $L(X^{cand}) > L(X^n)$, the quotient of the two is superior to one and is then automatically accepted.

The three steps are repeated over a number of chosen iteration is or a convergence level is reached. Metropolis-Hasting sampling has been implemented in the NonLinLoc package and where it shows a better efficiency than grid search (Lomax *et al.* (2001), Lomax *et al.* (2000)). The convergence of this method is still quite slow, and it can

be associated with other techniques to accelerate the process. The main techniques are evolutionary algorithm and simulated annealing.

Evolutionary algorithms are based on the theory of adaptation. The choice of the next samples is derived from candidates of the current population through random mutations and crossovers. Evolutionary algorithms, such as Genetic algorithms (Holland (1975)) and Differential Evolution (DE) (Storn et Price (1997)) have been used in source localization for source localization in seismology and seismic exploration (Gharti *et al.* (2010) ; Ružek et Kvasnička (2001) ; Sambridge et Gallagher (1993)). A special feature of DE comes from its differential mutation scheme, where the mutation of a given specimen of the current population depends on the difference of two other randomly chosen individuals. Ružek et Kvasnička (2001)) used DE to locate events with a phase picking method and Gharti *et al.* (2010) used the same optimization method to localize microseisms in the crust using a stack-based technique. Figure 1.11 shows the results of the localization of an event along an ore chamber performed by the authors. The evolutionary algorithm approach can provide a satisfactory sampling of the objective function without a priori knowledge of the source position, and is faster than the classical genetic algorithm approach. However its computational efficiency is still low compared to other methods such as simulated annealing (Barros *et al.* (2015)).

Simulated annealing (SA), (Kirkpatrick *et al.* (1983) ; Moreau *et al.* (2014)) is a MCMC-like algorithm where the tolerance for unlikely candidates is gradually lowered. The likelihood of the candidate is regulated by a temperature parameter that decreases with the number of iterations according to a cooling schedule. Because of this, the likelihood of worse candidates is diminished, lowering the rate of acceptance of unlikely solution as the number of iteration increases. SA has proven adapted to source localization problem in underwater acoustics MFP studies, where velocity profiles or multiple sources where inverted (Baer et Collins (2006) ; Dosso *et al.* (2015)). It has also been employed in seismic source localization (Billings (1994) ; Sun *et al.* (2016)). Very Fast Simulated Annealing (VFSA, Ingber (1989)) is a variant of the SA where the cooling schedule is different and the choice of the candidates is operated according to a temperature-dependent Cauchy distribution instead of a Gaussian. This allows for a broader range of moves at higher temperature that is then restricted as the cooling unfolds. VFSA shows significantly better results than SA in the case of complex optimization problem (Garabito *et al.* (2012)). However in the case of source localization, where only 3 to 4 parameters need to be inverted, a Simulated Annealing scheme with carefully chosen control parameters (number of iteration, speed of the temperature decrease) shows good results. An example of a localization using SA is displayed figure 1.12, showing the inverted epicentral position and apparent velocity for a Betsy gun shot source.

## 1.5 Conclusion

This chapter introduced some of the methods used for seismic source localization. Most of them use feature of the time domain traces as inputs, and rely on priory know-

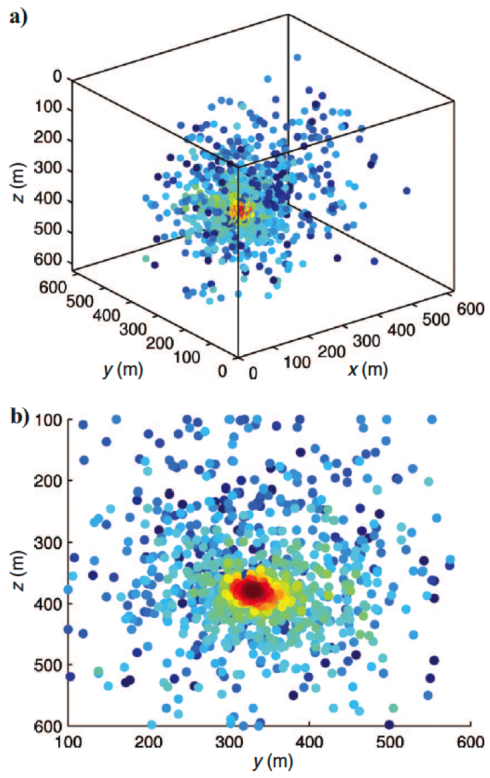


FIGURE 1.11 – All trial locations visualized in (a) 3D space and (b) yz-plane. Similar features are observed in xy- and zx-planes. Colors represent the value of the objective function (high values in red and low values in blue). The spot with the maximum value represents the best source location. (From Gharti *et al.* (2010))

ledge of wave propagation in the studied medium. There are many other methods some of them combining array techniques with most classical methods. Grigoli *et al.* (2016), combined for example stacking techniques with a Master event approach to improve relative localization accuracy. The convergence of the localization methods can be accelerated using global optimization searches. While all optimization methods presented here require tuning of the "control parameters" as a trade off between computational cost and convergence Li *et al.* (2018) ; Sambridge et Mosegaard (2002)), simulated annealing is more computationally efficient and is the preferred method for this thesis.

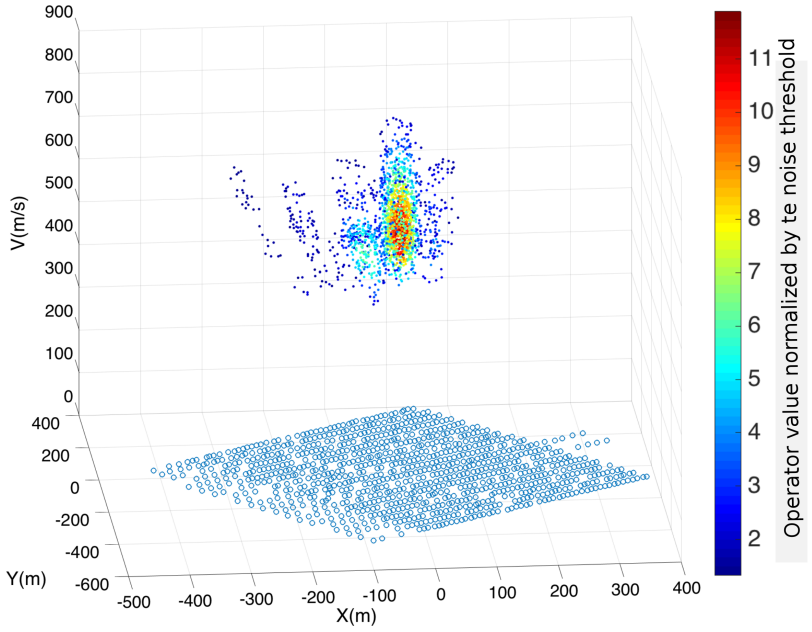


FIGURE 1.12 – 3D representation of the output of a MCMC scheme implemented with simulated annealing for a Betsy shot source. The inverted parameters are epicentral position and apparent velocity

# Chapitre 2

# Match Field Processing : introduction and 2D application

## 2.1 Introduction

Here we introduce a methodology based on array processing of continuous seismic waveforms to detect and locate events producing weak ground motion. The study employs a 30-days-long dataset recorded by a dense array of 1108 vertical component geophones (Figure 2.1) covering an area of about 600 m x 600 m around the Clark branch of the San Jacinto Fault Zone (SJFZ) in the trifurcation area southeast of Anza, CA (Ben-Zion *et al.* (2015)).

## 2.2 Match field Processing

MFP can be considered the frequency domain equivalent of a shift-and-stack technique. It is a model-based inversion method where the phase of a wavefield in array data is compared to the phase of a synthetic wavefield referred to as a replica. The best match between the data and the replica provides the best description of the seismic source given the model.

## 2.3 Replica

The main advantage of using replicas over stacking traces comes from the fact that, for a given frequency, they can be parameterized to describe the physical problem with any suitable degree of complexity, e.g. by including a search of the velocity profile in the medium for improved data fitting. Obviously, the efficiency of this approach is limited by the time required to compute the replicas, but this can be made very fast when a closed-form solution is available. In many geophysical studies, a simple homogeneous

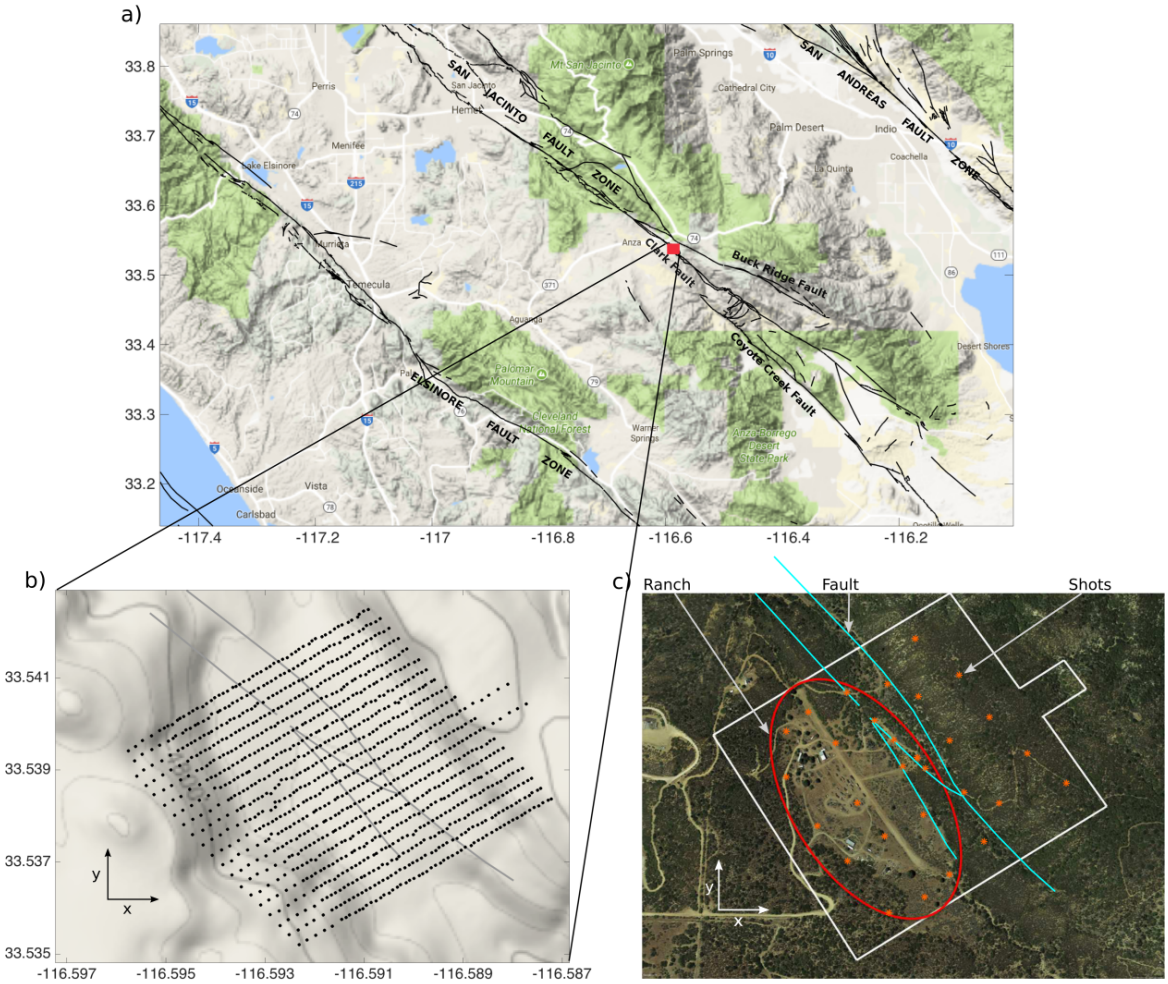


FIGURE 2.1 – a) Geographical situation of the Sage Bush Flat array on the San Jacinto Fault Zone (red square). b) Geometry of the 1108 sensor array and situation with respect to fault traces (black lines) and local topography (Map data, Google 2017). Each black dot corresponds to a station. c) Positions of expected main sources. Seismic source should mainly be located around the geological fault trace (marked with blue lines) and cultural sources from human activity and wind excitation are expected in the ranch area (within the red circle). The orange spots represent shots that were fired within the array.)

model of a spherical wave is considered (Chmiel *et al.* (2016), Corciulo *et al.* (2012), cite Cros (2011), Vandemeulebrouck *et al.* (2013)). For efficient evaluation of the replicas, the same assumption is made here for initial calculations and the propagation between the source and receivers is specified by the source coordinates, $\mathbf{x}$, and medium velocity, $v$. This strong assumption and its limitations are discussed in section 2.8.

In the following, we denote by $\mathbf{b}$ the replica computed from a candidate source. For

a given frequency,**b** is a vector that contains the value of the wavefield at all sensor positions. For example, at sensor $j$ the replica has the following expression :

$$b_j(\omega, \mathbf{a})) = A_j(\omega, \mathbf{a}) e^{i\omega t_j(\mathbf{a})} \tag{2.1}$$

where $A_j$ is an amplitude term, $\omega = 2\pi f$ is the angular frequency and $\mathbf{a} = [\mathbf{x}, v]$ is the set of parameters to be inverted. $t_j(\mathbf{a})$ is the travel time between the source and sensor j,

$$t_j(a) = \frac{r_j(\mathbf{x})}{v} \tag{2.2}$$

with $r_j(\mathbf{x})$ being the source-sensor distance.

In the study region, the SJFZ exhibits strong lateral variations of seismic velocities (Qin *et al.* (2018), Roux *et al.* (2016)) and amplitude variations in the data are mainly a consequence of the structure of the medium. Therefore, using a simple geometrical decay in the replica is not sufficient to properly describe amplitude variations. To avoid ambiguities when comparing the replica with the measurements, $A_j(\omega, \mathbf{a})$ is set to 1 and the data are normalized. The comparison is therefore only based on phase values. As explained in chapter 1, one advantage of MFP is that the origin time of an event does not need to be known, because the source localization only depends on the relative phase differences between sensors.

### 2.3.1 Cross Spectral Density Matrix

These phase differences are generally described in the form of the so-called Cross Spectral Density Matrix (CSDM). The CSDM, denoted by **K**, is the frequency domain equivalent of the time-domain broadband cross-correlations between all sensors. It is estimated from the normalized data as :

$$\hat{\mathbf{K}}(\omega) = E[\mathbf{d}^*(\omega)\mathbf{d}(\omega)] \tag{2.3}$$

where $\mathbf{d}(\omega)$ is a complex vector obtained from the Fourier transform of windowed time series records at each sensor and the star denotes complex conjugation. $E[]$ signify the averaging over several time windows. As we are interested in impulsive signals of short duration, we use only one time window to compute our CSDM. This results in less stability, and prevents us to use non-linear techniques for which the matrix must be inverted, but still shows satisfactory results when used with the Bartlett operator.

Figure 2.2 illustrates values of **d** on the whole array for various sources and two different frequencies. Because we use normalized data, these examples correspond to spatial phase variations for a given windowed time series. The time window is short enough to preserve the contribution from impulsive signals but long enough to ensure the stability of the

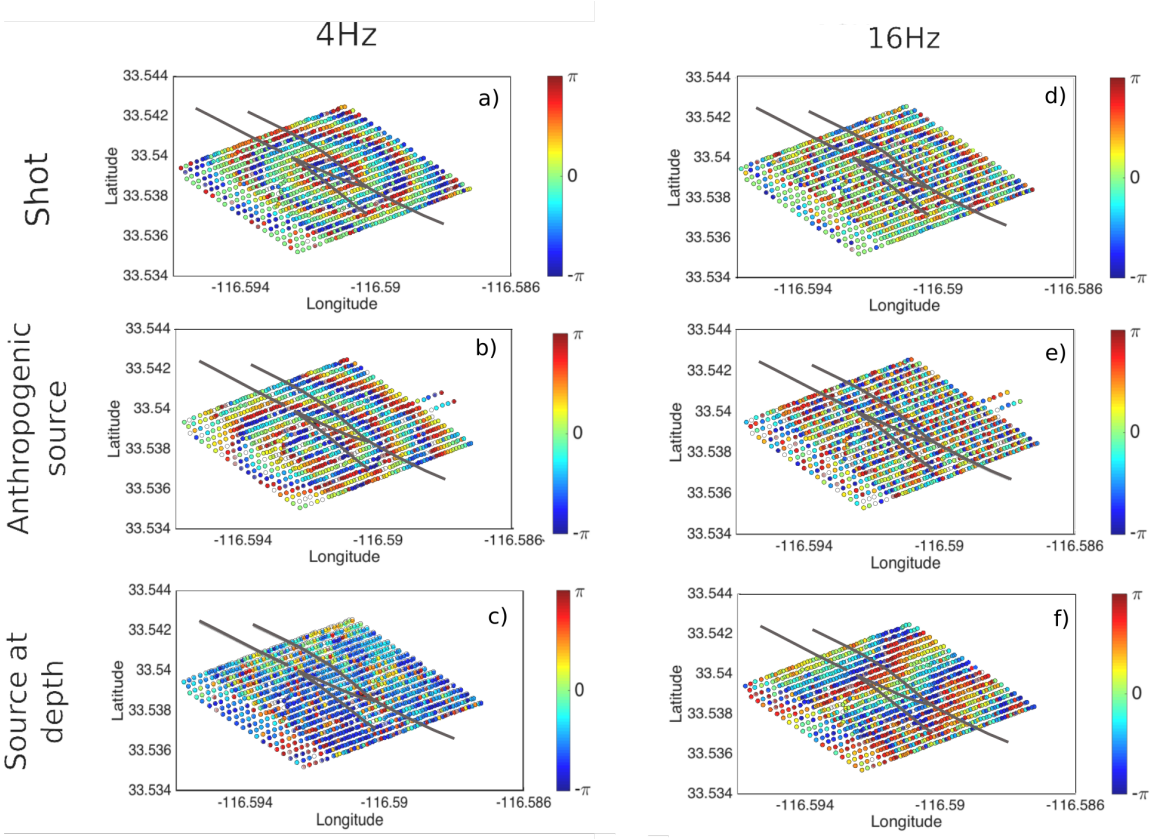


FIGURE 2.2 – a, b, c) Phase patterns recorded at the array for 4Hz. d, e, f) Phase patterns recorded at the array for 16Hz. The patterns are obtained from the Fourier transform of 1-second (4Hz) and 0.25-second (16Hz) windows containing the events. The phases are highly coherent in space at 4Hz for the first two sources (a and b) and at 16Hz for the last one (f) and display clear propagation patterns. The signal from the source at depth shows little spatial variation at 4Hz. The upper right panels show spatial aliasing of the phase, inducing a loss of spatial coherency

Fourier transform. In practice at a given frequency, we use a time window of duration T=4 periods. This means that we obtain one localization for the dominant seismic source for each time interval T.

The choice of the traces that will be used to compute the CSDM is also a key feature in terms of quality of localization. If repetitive and extremely weak surface events are only seen at the sensor they are the closest to, they will degrade the spatial coherency of the array and result in a lower output. In the same way, geophone malfunction can also affect the quality of the final result. In order to avoid this, the median of the energy is computed for all sensors over half hour segments. Sensors that have a median energy 10 times higher or lower than the global median will be discarded for the corresponding half hour. Except for the first and last 3 days, where many sensors were down because

of initialization or loss of battery, around a 100 sensors are discarded on average.

### 2.3.2 Bartlett operator

The comparison between the replica and CSDM can be performed using various operators, the choice of which depends on a trade-off between robustness and resolution. An operator with higher resolution, such as the Capon algorithm (Capon (1969), is more sensitive to the accuracy of the model and requires an inversion of the CSDM. In our case, because of the complexity of the subsurface structure, the transient nature and low SNR of micro-seismic events, we use the more robust Bartlett operator (Bucker (1976),Corciulo *et al.* (2012), Cros (2011)), given by :

$$B(a,\omega) = \sum_{\omega} |\mathbf{b}^*(\mathbf{a},\omega).\mathbf{K}(\omega).\mathbf{b}(\mathbf{a},\omega)| \tag{2.4}$$

We also divide this operator by the number of sensors squared ($N^2$) to scale its output between 0 and 1, where 1 accounts for perfect agreement between the data and the replica and 0 for a complete decorrelation of the two. When the Bartlett operator reaches a maximum, the physical parameters (source location and medium velocity) provide the best description of the dominant source given the model.

As we demonstrated before, MFP allows some flexibility in terms of velocity and source modeling, including numerical models with elastic wave propagation and more complex velocity structures. However, the computational cost must be taken into account, especially when processing large amounts of data. Indeed, in order to work on impulsive events, MFP must be applied to short time windows. This makes the scanning of the complete 30-day dataset computationally expensive, especially when considering classic grid search approaches. The use of optimization techniques is then necessary to our study. The next section will briefly present some results obtained with the different methods considered during my PhD. The results were obtained using the data from shots that were fired during the experiment. As we are using surface sources only epicentral position and velocity at the surface are inverted.

## 2.4 Optimization

### 2.4.1 Choice of the optimization method

Grid searches are classically implemented with array techniques that require the test of several candidate source positions and medium properties. As explained in chapter 1.4, their main disadvantage is the trade off between the resolution of the parameter space sampling and the computational cost.

We compare three optimization methods to the original grid search approach to find a better performing alternative. The first method is a simplex optimization, the principle of

which has been explained in chapter 1.4. The second is a simplex optimization combined with a grid search. We iterate the simplex process starting from different points on a coarse grid (a point every 200m in each direction, the starting velocity stays the same). Depending on the shape of the Bartlett output, each iteration can converge to a different point. The point with the maximum value is chosen as the result of the gridded simplex method. The last method is MCMC sampling combined with a simulated annealing global optimization. As stated in chapter 1.4, the output of such as scheme is a point cloud, the density of which is proportional to posterior probability distribution of the parameter.

Remark : The MCMC algorithm generally requires a burn-in period before reaching a stable state where the PDF is efficiently sampled. As mentioned before, we use a simulated annealing global optimization before the MCMC as suggested in Moreau et al. (2014) to speed up this burn-in process. As a consequence of the forced convergence, the output of the simulated annealing process is not a PDF. Because of this, the results shown in section 2.7 only represent the second part of the process, i.e. the results of the MCMC that come after the simulated annealing run.

Figure 2.3 shows the results of the three optimization methods applied on shots data, for which the positions and occurrence time are known. The inverted parameters are the epicentral position and apparent velocity of the source. Here, for clarity, only the maximum value of the MCMC output is displayed. The results of the three optimizations are represented respectively by a circle, a diamond and a triangle.

Four shots are studied here, their respective positions shown figure 3-a. The backgrounds of Figure 2.3-b,2.3-c and 2.3-d are a slice of the outputs of the grid search, i.e. the image of the Bartlett operator output, for a fixed velocity. This velocity corresponds to the velocity of the maximum of the grid search output. All three examples demonstrate the complexity of the Bartlett output, which comprises several local optima around the main lobe.

Error on epicentral location can occur because optimization methods converge to a local minimum (Figure 2.3-b and 2.3-c and 2.3-d). We see that in these three examples, the simplex method (circle) always fails to converge to the global maximum. The gridded simplex method shows better results, with only one failed convergence (shot 14). It is still less accurate than the SA optimization.

Another cause for location error is the inaccuracy of the Bartlett operator itself, as is shown Figure 2.3-b and 2.3-c. In this case, obtaining a quantitative measure of the confidence in our results is key. The grid search approach gives us access to the shape of the main lobe, which can be used to assess the quality of our inversion. For example, the shape of the output for shot 9 (Figure 2.3-c), with several maxima of similar values, is an indicator of poor performance of the process. In practice we can evaluate uncertainties by measuring the width, or shape, of the main lobe at -3dB. While the output of the simplex method does not allow such trivial estimation of the uncertainties, SA combined with MCMC finely samples the global maxima of the function. Thus, evaluating the uncertainties by measuring the width of the main lobe is possible. In addition we can

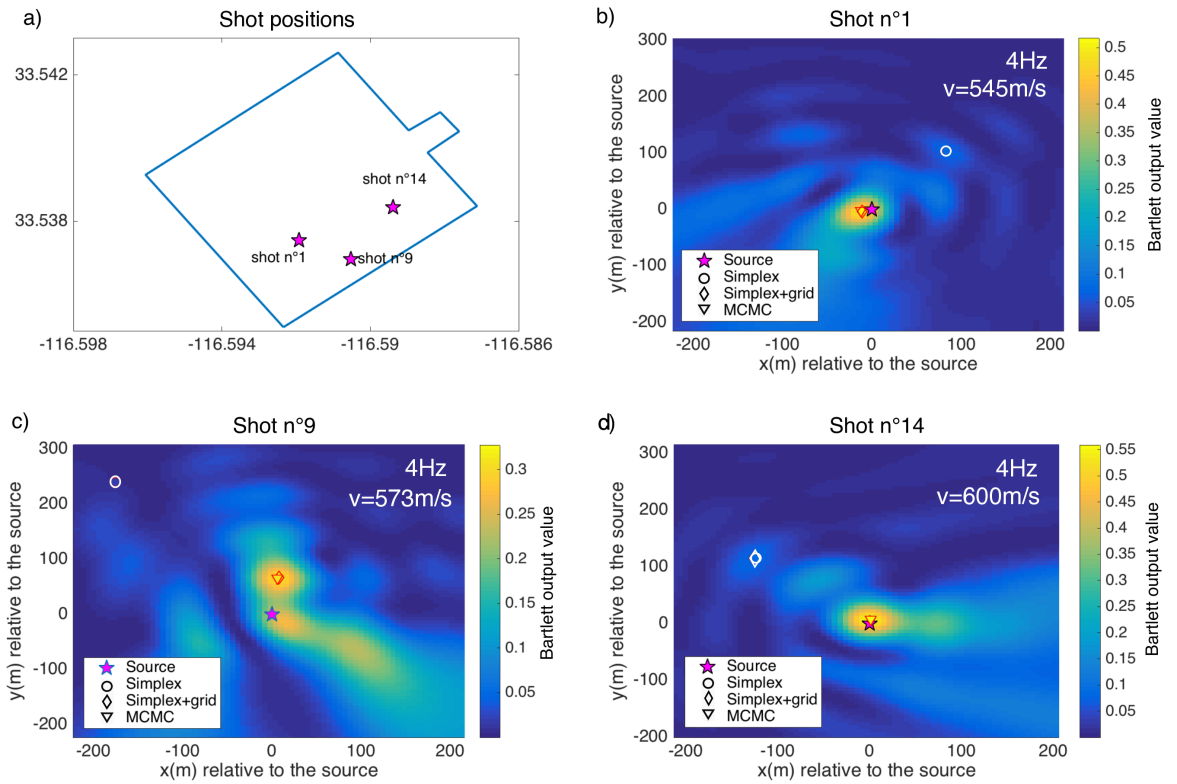


FIGURE 2.3 – a) Position of the shots used to test the optimization method. b,c,d)Location results for the three shots. All results are displayed relatively to the source location (purple star). The circle, the diamond and the triangle represent the results of the simplex method, the gridded simplex method and the MCMC optimization respectively. The background is the Bartlett output (i.e. the objective function) sampled through a very fine grid search for the velocity displayed in the top right corner

measure a confidence interval directly through the dispersion of the posterior distribution, by estimating the variance of the output for each parameter.

Another advantage of MCMC is its increased efficiency compared to grid search processes. The latter requires a number of iterations that grows exponentially with the size of the parameter space of interest. A grid sampling of a 100 points per parameters for N parameters will require $100^N$ iteration. As an example, 400s were necessary to sample the 3-D Bartlett output with the resolution presented figure 2.3, even with a grid restricted to 300m around the known position of the source. This computational cost is too high. Such a fine grid cannot be used while processing the 30-day dataset.

Figure 2.4 compares the resolutions of both the SA optimization and the grid search at equivalent computational cost ($\sim 30s$) for shot number 9. The very complex output corresponding to this shot requires a fine sampling in the area surrounding the main

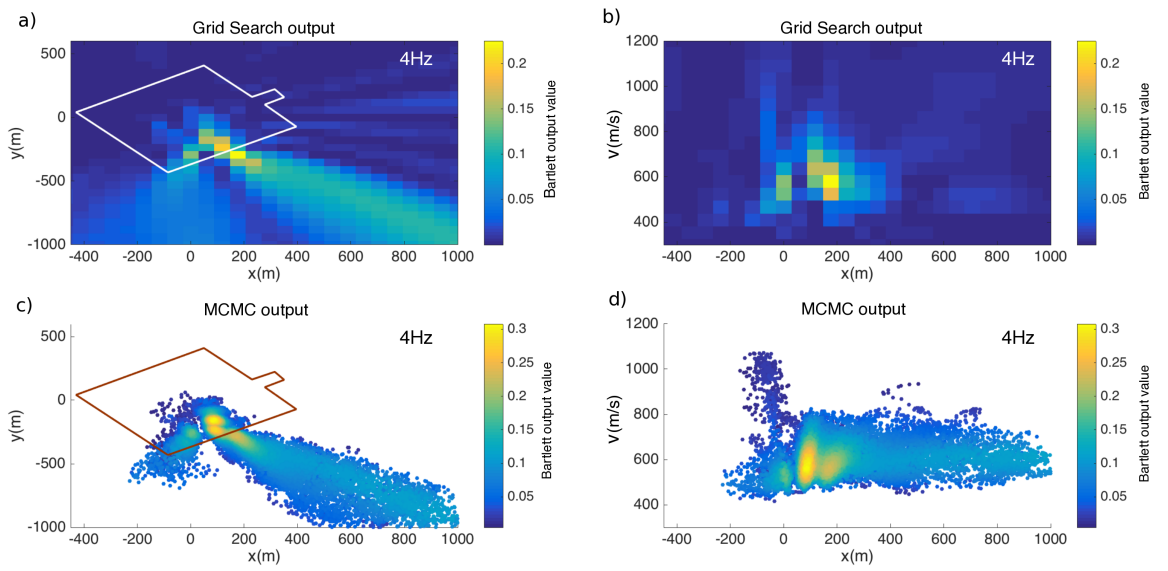


FIGURE 2.4 – Outputs of the grid search and MCMC at equivalent computational time for shot n9. a) x-y view of the output of the grid search,b) x-v view of the output of the grid search, c) x-y view of the output of the MCMC,b) x-v view of the output of the MCMC

lobe. For each method two slice are shown, displaying the output variation according to $x$ and $y$ and $x$ and $v$ respectively. We see that the resolution of the grid search is greatly reduced compared to Figure 2.3. As a result, several lobes are not distinguishable. On the other hand, they are clearly visible in the case of the MCMC output. This is due to the fact that only the area around the global optimum is finely sampled.

| Method | Computational cost (s) for 1 iteration (8 cores) |
|---|---|
| Grid search | 598 |
| Grid search + library | 513 |
| Simplex method | 2 |
| Gridded simplex | 70 |
| MCMC + SA | 29 |

TABLE 2.1 – Computational cost of the different optimization methods

Table 2.1 compares the performances of grid search with the simplex and SA optimization methods. The tests where also performed on shot data. The computational cost is estimated as the mean of all processing times over 15 shots, with Matlab codes running on 8 cores, for a parameter space ranging for -/+1000m around the center of the array for the epicentral position and from 240 to 1600m/s for the velocity. Grid search methods are too time consuming to be used in our study. The use of pre-computed library,

that avoids the computation of replica at each iteration, does not sufficiently improve the computation time. The most efficient method in terms of computational costs is the simplex method, however it is not reliable enough to be used on this application. Overall, the MCMC method proves to be the best compromise between reliability, assessment of quality of the localization and computational costs.

The methods in table 2.1 are listed following the chronological order of the improvements and tests performed during my PhD. The running of the codes was first performed on a personal computer using Matlab before being uploaded to the CIMENT server (the results shown in Table 1 correspond to this step). Working with Albanne Lecointre, simplex and MCMC optimization were added to the existing Fortran grid search code. This code was then altered to allow parallelization on the University grid (CIGRI). This step reduced the computing time to 152 elapsed hours for the processing of 25 days of data, in the case of the MCMC.

### 2.4.2 Implementation of the MCMC

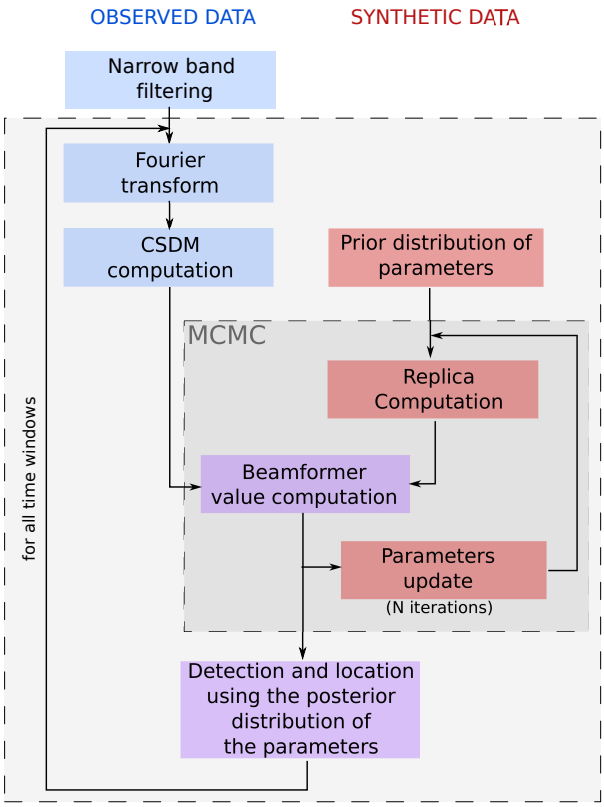


FIGURE 2.5 – General workflow of the method, describing the implementation the MFP method with a Markov Chain Monte Carlo scheme. The key steps of the latter are detailed in the darker gray box labeled MCMC

The inclusion of the MCMC in the classical MFP process is illustrated Figure 2.5. The first step of the method is application of a narrow band filter with a bandwidth of 4Hz centered on a chosen frequency. Since MFP involves a frequency-domain phase-based optimization problem, the choice of frequency is a key step that will be detailed in the next section. Once the frequency is set, the filtered waveforms are segmented into successive time windows of length T. For each window, the CSDM is calculated using equation (3) and output is obtained through the MCMC sampling process (Figure 2.5). This output is an ensemble of sets of physical parameters associated with their corresponding Bartlett operator value. For a candidate detection, the value of the maximum of the output must be higher than a statistical threshold $T_{th}$. This is selected here as the theoretical limit of incoherent noise level for an array of N sensors :

$$T_{th} = \frac{1}{N} \sim 0.001 \tag{2.5}$$

All MCMC output values will be normalized by this theoretical threshold $T_{th}$ for the remainder of this chapter, in order to have a relative measure of the quality of our localizations compared to incoherent noise.

MCMC schemes are usually implemented to estimate the PDF of the parameters, but in our case we mostly use the method capacity to sample more finely the region of parameters around the global optimum and increase the computational efficiency. Another possible improvement is to reduce the number of parameters. When using the spherical wave in a homogeneous medium approximation, we need to invert for the velocity in the medium to have consistent location in spite of heterogeneities. This means that 4 parameters must be inverted to determine the hypocentral position of a source. As a first approach we can invert only for latitude, longitude and velocity. However this simplification also implies that there will be no direct mean to discriminate between surface and deep sources.

This issue needs to be taken into account, because while the dense array was deployed in a relatively remote ranch, recorded waveforms in the area are still affected by air-traffic events, and other non-seismic sources. Among them, the interaction of the wind with trees, man-made structures and machines in the ranch located mainly on the western side of the fault (Figure 2.1-c, (Johnson *et al.* (2018), Meng et Ben-Zion (2018b)). More precisely, the discrimination between surface and deep sources is of importance because of the frequency content of both surface event and deeper event signals. Indeed, the weak events have energy mostly at frequencies above 5Hz ; hence we focus on signals filtered between 0.5 and 20Hz. In this frequency band the noise signal is dominated by surface waves (Roux *et al.* (2016)). This makes the detection of impulsive sources at depth difficult. Also, the existence of additional, non-seismic sources such as the anthropogenic activities or atmospheric sources mentioned above can be an issue, because they can be coherent on part of the array (Riahi et Gerstoft (2015) and complicate signal interpretation. The next section will demonstrate how we can use the choice of frequency of study to limit the detection to one type of event, while keeping only longitude latitude and apparent velocity as inversion parameters.

## 2.5 Choice of the frequency of study as a tool for source discrimination

The geometry of the array limits the range of wavelength that can be sampled. On one hand, a minimum aperture of about twice the wavelength is necessary to allow for a good resolution when locating the source. On the other hand, the Nyquist's criterion requires having at least 2 sensors per wavelength to avoid spatial aliasing. In practice, we require 4 sensors per wavelength to properly reconstruct the wavefield.

This wavelength depends on the frequency and the velocity of the wave, or the apparent velocity when sensors are at the surface. This means that at a given frequency, the range of apparent velocity that can be detected by the array is fixed.

These limitations of the detection capability can be turned into a practical way of separating between sources at depth and the surface, because the apparent velocity measured at the surface is higher for sources at depth. The reasons for this are twofold. First, velocities of P- and S-waves are larger than surface wave velocities. Second, the projection at the surface of a body wave propagating from sub-surface sources results in a higher apparent velocity when measured at the array. Consequently, the choice of frequencies translates directly into a filter for the measureable apparent velocities, and through that for the source depth.

While the depth of the sources of interest drives the choice of specific frequencies, the range of potential frequencies of study is limited by the medium and sensor characteristics. In theory, the highest frequency that can be used is only limited by the spatial sampling of the array, which sets the measurable phase coherency between sensors. However, initial tests on events outside the array show that above 20Hz, the phase measured on the array appears highly distorted due to medium heterogeneities. At the other end of the spectrum, the lower frequency boundary is limited to 1 Hz due to both poor SNR of the recorded signals (related to the 5Hz corner frequency of the geophones) and the limited aperture of the array.

The dense array used for the recording of the data comprised 1108 5-Hz geophones set in 20 rows separated by 30m with an inter-sensor distance of around 10m in each row (Figure 2.1). This translates to a diagonal aperture of 800m and a mean inter-station spacing of 25m. In this configuration, all waves propagating with apparent wavelengths between 75m and 400m can be processed. The choice of central frequencies determines the range of apparent velocities for which detections are possible. In this study we use 4Hz and 16Hz, chosen as they provide a sufficient gap to distinguish between sub-surface and surface sources while staying below the 20Hz phase distortion limit. The apparent velocities that can be processed at 4Hz are in the range 260-1600m/s, while at 16Hz we can seek velocities in the range 920-6400m/s.

Figure 2.2 shows the phase distribution across the array for detected events with waveforms filtered around 4 and 16Hz. We distinguish two surface sources and a third type involving sub-surface sources, associated with wavefields with estimated apparent velocities under 900m/s for the first two and above 4000m/s for the third type. On the one

hand, surface sources at 4Hz (Figure 2.2-a and Figure 2.2-b) exhibit a clear propagation pattern, whereas for a source at depth (Figure 2.2-c) the array is not wide enough to record the phase oscillation patterns. On the other hand, at 16Hz the wavelength of surface sources is too small (Figure 2.2-d and Figure 2.2-e), resulting in aliasing and loss of spatial coherency in the array, whereas the wavelength of deep sources is now reduced and can be seen within the array (Figure 2.2-f). In other words, the choice of the frequency sets the range of apparent velocities that can be measured by the array, which directly translates to discrimination between surface and deep sources.

## 2.6 Application to the dense deployment data

The array described in the previous section recorded continuous waveforms for 30 days at a 500Hz sampling frequency, providing high quality data at frequencies between 1 and 250 Hz. The duration of the recordings, large frequency range as well as the location and density of the array, are expected to provide a wide variety of sources to be detected and located. As mentioned, in addition to the seismic activity from the fault under the array, air-traffic, interaction of wind with obstacles above the surface, along with car-traffic and other activities produce ground motion recorded by the array (Figure 2.1-c). Small Betsy gunshots fired at different locations of the array during the experiment (Ben-Zion *et al.* (2015)) provide good benchmarks on surface source localization, as their position and time of occurrence is known.

Figure 2.6 shows 10s long time series filtered between 0.5 and 49.5Hz that correspond to three different events. For each event, the traces of five sensors are displayed : one for each corner plus one at the center of the array. Some events, such as shots, are easily identifiable with a clear propagation across the array, even on these five traces only. However, times of arrival and wave propagation for events of longer duration such as the anthropogenic sources are more difficult to distinguish in the traces, even with a good SNR. In some cases, the SNR is too low and events cannot be detected on single traces (for sources at depth). Figure 2.6-d, figure 2.6-e and figure 2.6-f show representations of the input to the MFP technique. Each panel is the wrapped phase of the Fourier transform of the time window in the corresponding time series.

In these examples, the epicentral locations of the sources are readily identifiable because of the high array density. The data representation in figure 2.6-d and figure 2.6-f is equivalent to considering the spatial variations of the wavefield, which will then be matched against different replica. The choice of the model used to compute the replica is a crucial step, given the complexity of the medium and its possible impact on the spatial variation of the wavefield.

The primary goal of this study is to localize shallow seismic sources in the top 3 kilometers of the crust. In most such cases, the source-to-array distance is within one wavelength of signal propagation near the sources, which prevents separating P- and S-wave contributions in the recorded wavefield at the surface. Moreover, as mentioned above we approximate the propagation with a Green's function of a scalar wave in a

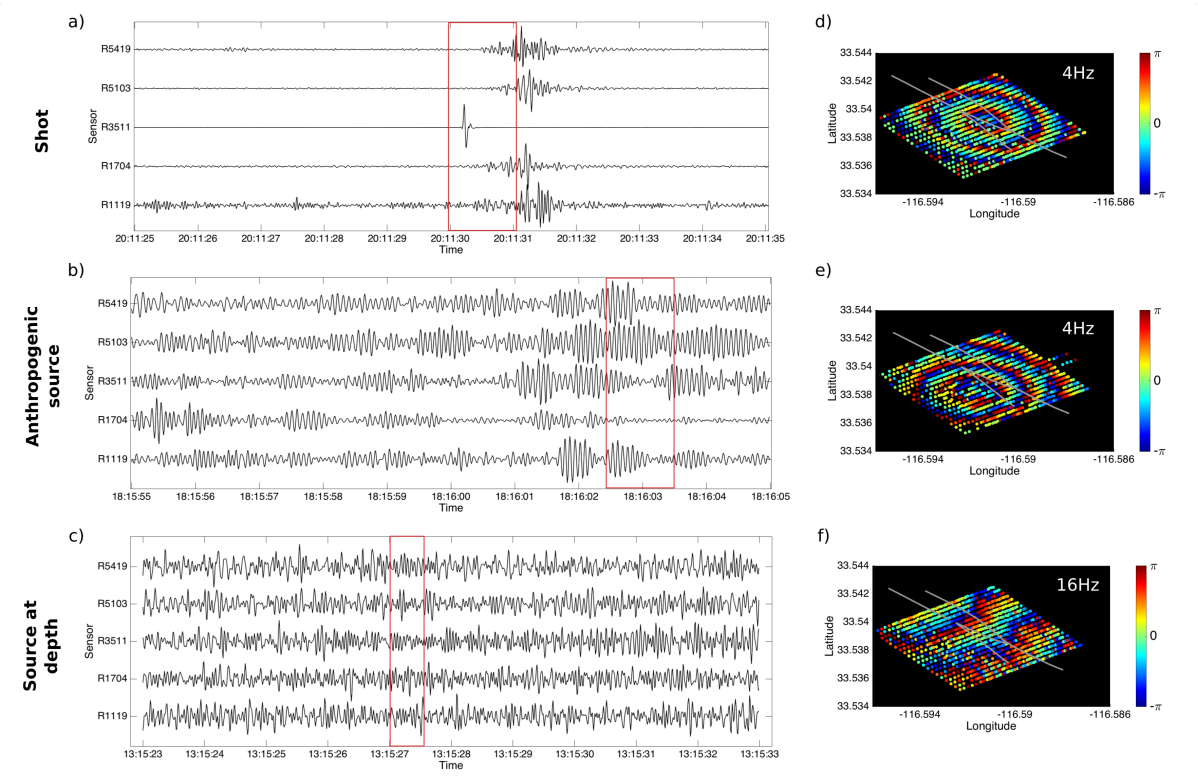


FIGURE 2.6 – a, b, c) show 10 seconds time signals measured by 4 sensors at the corners of the array and one at the center. The traces are filtered between 0.5 and 49.5Hz. Three different types of signal, coming from two surface sources (a and b) and a source at depth (c), are represented. The red boxes represent the 1-second and 0.25-second time windows, used to obtain the phase patterns in the right panels. d, e, f) are the corresponding wrapped phase of the Fourier transform at each sensor

homogeneous medium for fast computation of the replicas. This is far from representing the complexity of the medium under the array, but it is nonetheless an appropriate model for initial testing of the methodology.

Three different sets of inversion parameters are used in the example data analyzed. In the first set, only three parameters are taken into account : the source epicentral position x and y, and the apparent velocity at the array $v_ap$. Since the apparent velocity is linked to the depth of the source, this description of the problem still provides a general information on depth, while being extremely time-efficient. Once the time-windowed signals that correspond to non-surface sources are identified, a follow up study that includes a higher number of parameter or a more elaborate model can be performed to better constrain depth. In the second inversion, the parameters are the x, y and z coordinates of the source, and the velocity of the medium is set. Lastly, we invert for both source position (x,y, and z) and homogeneous velocity. In the next section, the method is tested first on

controlled sources (shots) and then on the several types of events detected in the data.

## 2.7 Results

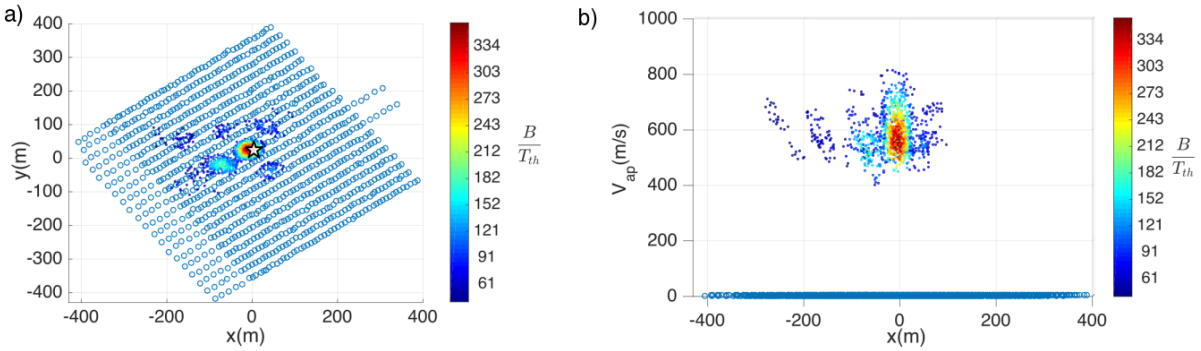


FIGURE 2.7 – MCMC output for a shot according to x and y positions an apparent velocity. The dots correspond to the candidate source position explored by the MCMC scheme after convergence of the simulated annealing. Colors represent the output value associated with each trial source position normalized by our detection threshold. The dots with the maximum values give the most probable source location and apparent velocity. a) 2D view showing the epicentral position of the output. The diameter at -3dB is around 40m. A black star represents the known position of the source. b) 2D view of the output according to apparent velocity and x direction

The first inversion is tested on Betsy shots, for which the positions (figure 2.1) and time of occurrence are known. Figure 2.7 presents results for a shot that was fired at the center of the array, during Julian day 154, (see 1 s time windows in Figure 2.6-a). Because shots are surface sources, the frequency of investigation in this case is 4 Hz as discussed previously.

The maximum output is eleven times higher than the noise threshold. This is a high value considering the simple homogeneous model used in the calculations. The radius of the main spot gives an uncertainty of +/- 20 m on source position. The source localization is accurate within the uncertainty limit, the actual position of the shot being in the spot around the most probable source position. Given the uncertainties, the apparent velocity associated with the position of the maximum output, i.e. 580 m/s, is consistent with the phase velocity expected at the surface of the SGB site at 4Hz (Roux *et al.* (2016)). Side lobes associated with the array response are also present. The method shows encouraging results for controlled sources such as shots.

Figure 2.8 presents a part of a series of events detected in successive 1-second time windows. The specific example showed in figure 2.8 corresponds to the time window represented as a red box in figure 2.6-b. Detection and localization were performed with

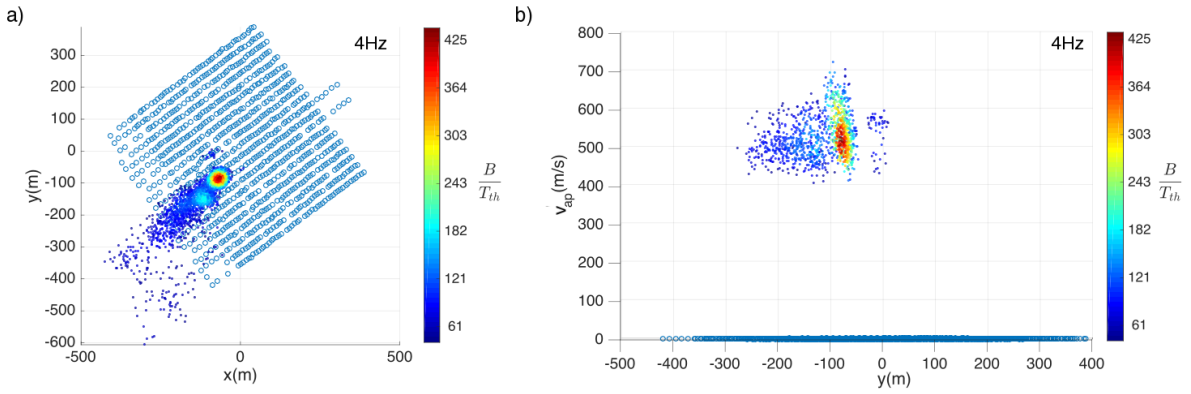


FIGURE 2.8 – MCMC output for an anthropogenic source according to x and y positions and apparent velocity. a) 2D view of the epicentral position of the output. As before the diameter is roughly 40m at -3dB. b) 2D view of the output according apparent velocity and y direction. The velocity range is consistent with a surface source

the first inversion parameters. The localizations indicate a motion of the source position in the consecutive examined time-windows, with a mean velocity of 20km/h. The source, possibly a moving vehicle is thus considered to be anthropogenic. This signal is different than the shot discussed previously in terms of duration and frequency content. However, the uncertainties for the anthropogenic source are similar to those of the shot. This is expected, because we are working with narrow frequency bands and short time windows. The sizes of the spots around the most probable position remain comparable (around 40m in diameter) and the associated apparent velocities are between 500 and 600 m/s.

A 3D localization of this source was performed with the other inversion parameters to check consistency in our results. The homogeneous velocity was fixed at 550 m/s, corresponding to the mean velocity at the surface at 4Hz. The inversion output represented in figure 2.9, only for values higher than 50% of the maximum output value for clarity, confirms that the source is located at the surface. Comparing the spots size for the three dimensions indicates that the uncertainty on depth is three times larger than the uncertainty on the epicentral location.

Figure 2.10 shows results associated with a subsurface source analyzed using a frequency of 16Hz. The MFP output was computed for 0.25s time window in Figure 2.6 (deep source) with the first inversion parameters. We observe a smearing of the PDF for the source position in the radial direction. This phenomenon is classical in array analysis, and is due to the fact that the event is located outside the array. The velocity associated with the maximum output is 4000m/s. This value is far too high to describe wave propagation near the surface, leading to the conclusion that this source is at depth. This illustrates that it is possible to detect and obtain a first approximation of the source position without inverting for depth.

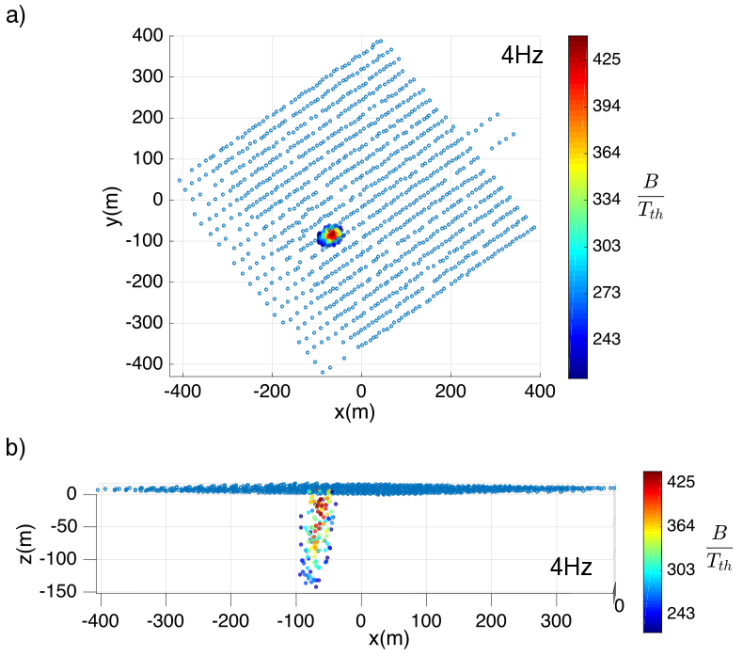


FIGURE 2.9 – MCMC output for an anthropogenic source according to x and y position and depth. The velocity of the model is fixed at 550m/s (mean of the velocity of the x,y and $v_{ap}$ study). The output according to x and y shown in a) is of the same size as the previous study. b) 2D view according to x and z. The uncertainty on the depth of the source displayed in b) is greater than the uncertainties on x and y, due to the position of the sensors at the surface and the simple used homogeneous model.

However, more precise depth estimation is needed to conclude on the presence of seismic sources in the shallow crust. The next logical step is to invert for the three coordinates of the source position as well as the medium velocity.

As a preliminary step before inverting all 4 parameters simultaneously, we computed the MCMC outputs from the second inversion parameters with different medium velocities : 1000m/s 2500m/s and 4000m/s (Figure 2.11). We observe a dipping of the PDF as the velocity gets lower. This can be explained geometrically, because a deeper source produces a planer wavefront at the surface and thus higher apparent velocity. A spherical wave propagating from a source at depth and recorded at the surface will therefore exhibit an apparent velocity that is higher than the medium velocity. Snell's law links the dipping angle of the MFP output, $\alpha$, to the apparent velocity $v_{ap}$, and homogeneous velocity in the medium v, as

$$\alpha = arcos(\frac{v}{v_{ap}}) \tag{2.6}$$

This interdependence will cause some resolution problem when trying to invert for

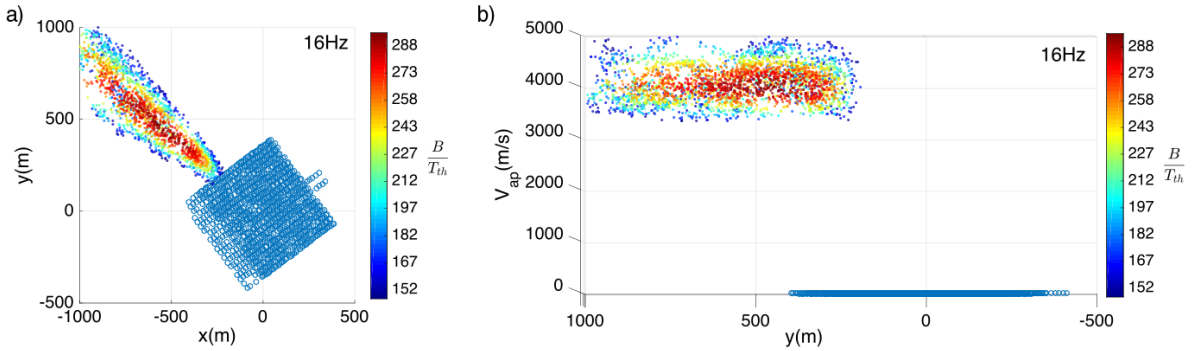


FIGURE 2.10 – MCMC output for a possible deep source according to x and y positions and apparent velocity. The output according to x and y presented in a) is elongated in the radial direction, a consequence of the source being outside the array. The apparent velocity displayed in b) is much higher than for the two previous events, pointing at a source at depth

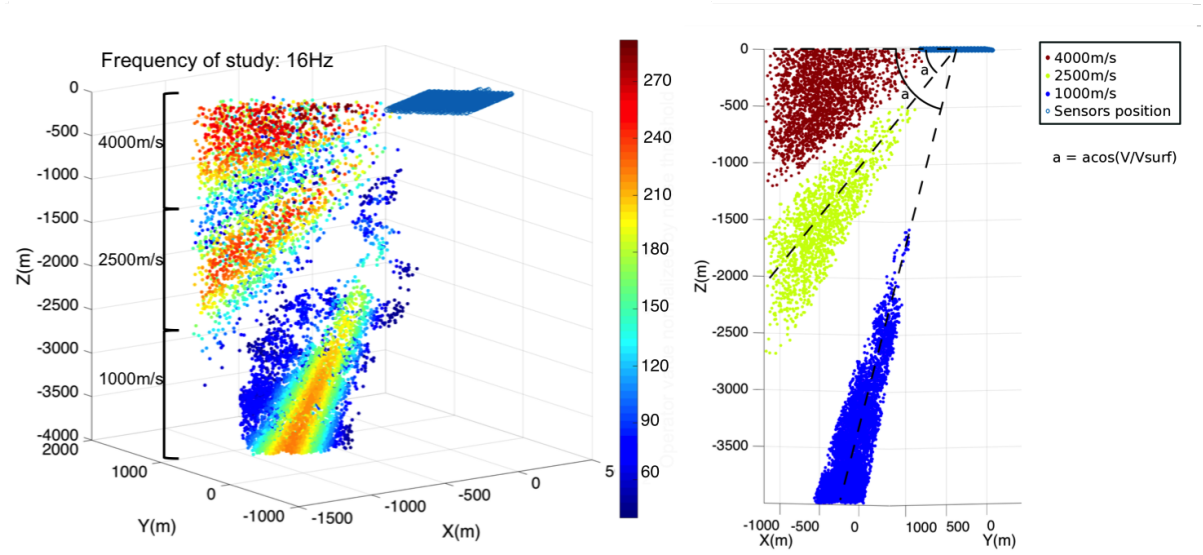


FIGURE 2.11 – Dipping of the output according to the velocity of the medium and associated beamformer value. a) shows three clusters that correspond to three different outputs computed for a deep source. They were obtained by inverting for x, y and z position while successively using a fixed medium velocity of 1000, 2500 and 4000m/s. The dipping angle depends on the ratio between the apparent velocity measured at the surface and the velocity of the medium, as is shown in the side view displayed in b).

depth as well as medium velocity. Figure 2.11 shows only a 25% variation of the MFP

output between 0 and 4000m depth, which is not enough to conclude on the source position.

This example of a source outside the array is not enough to draw conclusions on the limitations of the homogeneous model in the case of a source located under the array. We performed a localization at depth for such a case, using the three coordinates of the source position and the velocity of the medium as inversion parameters. The results are displayed figure 2.12. The MCMC output has an output length of more than 3500m in the z-direction. Because we are interested in the first few kilometers of the crust, uncertainties of this order of magnitude are to important for our study.

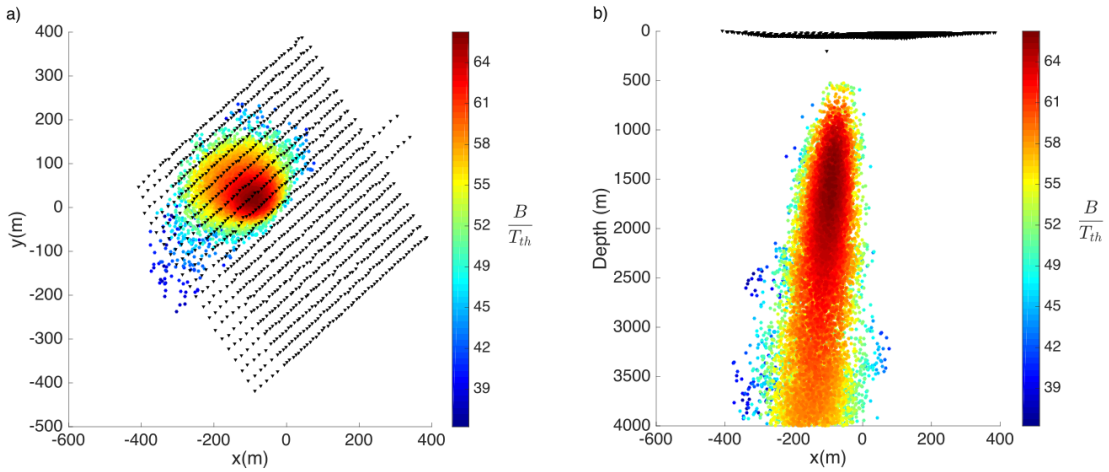


FIGURE 2.12 – MCMC output for a possible event under the array according according to x, y and depth. a) and b) show, respectively, the hypocentral position and a 2D view of the output according x and z.

From the discussed results we can conclude that the resolution obtained when using a replica computed with a homogeneous velocity model is not sufficient for determining the depth of a shallow seismic source. This is essentially due to the trade-off between depth and velocity, but also because the medium under the array is complex. For the 3D inversion, this makes the choice of a single velocity in the model difficult to determine.

## 2.8 Discussion and conclusions

This study on source localization with continuous seismic waveforms recorded by a dense array shows that it is possible to detect and locate both surface and sub-surface events by combining capabilities of the MFP and MCMC sampling. The MFP technique was developed originally for simpler applications on acoustic waves propagating in the ocean (e.g., Kuperman et Turek (1997) ; Baggeroer *et al.* (1993)). Application of the

method for localization of sources recorded by seismic arrays at the Earth surface are considerably more challenging because the velocity structure of the medium is more complex and a wide variety of sources contribute to the recorded ground motion (e.g., Riahi et Gerstoft (2016),Meng et Ben-Zion (2018a),Meng et Ben-Zion (2018b)) The 30-days length of the data at end also require an efficient localization tool. To address these difficulties, we augmented the method for localization of sources at and below the surface of the earth crust by using the frequency of study as a "filter" separating sources at the surface and sources at depth. With this, we can keep an analytic homogeneous model depending on only 3 parameters insuring the efficiency of the algorithm. The use of MCMC further reduces the computational time compared to grid search approaches and provides statistic information that can be useful when interpreting the data.

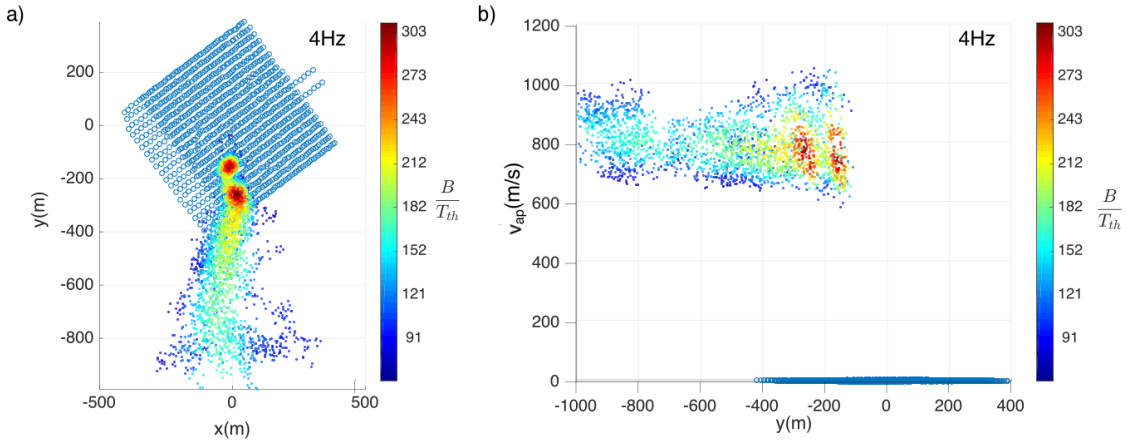


FIGURE 2.13 – MCMC output for another window containing signal from the anthropogenic source according to x, y and apparent velocity. a) and b) show, respectively, the hypocentral position and a 2D view of the output according apparent velocity and y.

This methodology, introduced in section 2.2, gives access to the PDF of parameters that describe the wavefield. It provides an estimate of the source position and medium velocity, as well as a confidence interval for the solution. In addition, our approach also allows the discrimination between surface and deep sources without determining the actual position of the source. Due to spatial sampling requirement, the choice of the frequency has an impact on the depth range of the detections : surface sources are detected at lower frequencies than deep sources, because the apparent velocity measured at the array increases with depth.

Surface sources such as shots and moving vehicles were successfully detected and located when inverting for the epicentral position and the apparent velocity under the array. In those cases we obtain a good resolution for epicentral coordinates when the events are occurring inside the array, and velocities that are consistent with previous studies.

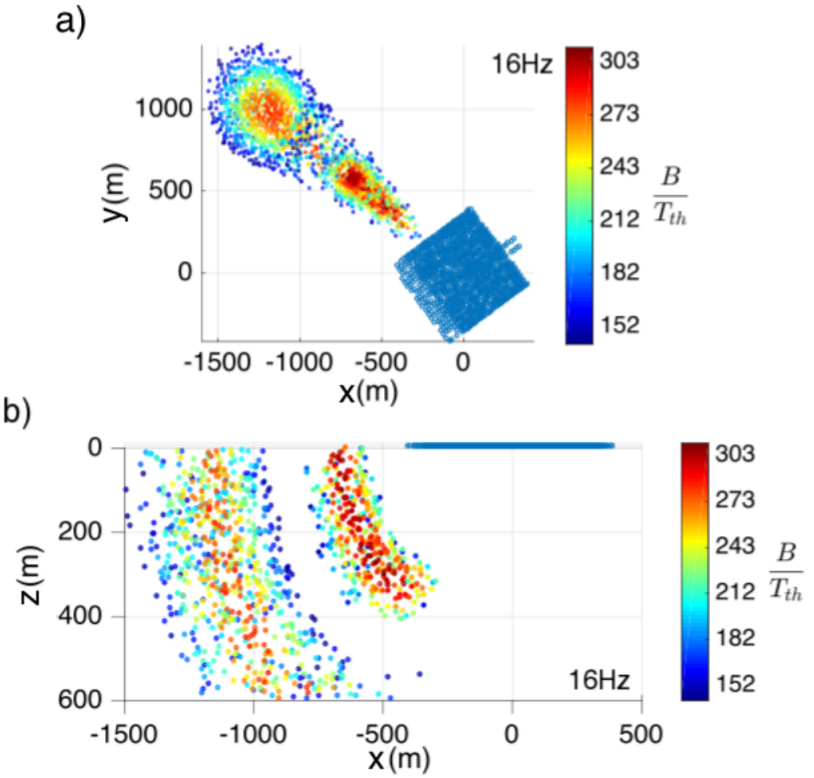


FIGURE 2.14 – MCMC output for another window containing signal from a possible deep source outside the array according to x, y and z, computed with a 1D gradient model.a) and b) show, respectively, the hypocentral position and a 2D view of the output x and z.

A shallow source located outside the array was also detected successfully based on the same model parameters. However, for source localization the use of a homogeneous velocity model to compute the replicas has limitations. First, the resolution along depth is not sufficient to determine the position of sources below the surface. Because of the trade-off between depth and apparent velocity, there is a strong ambiguity on the depth. Secondly, the simplicity of the replica cannot match complex phase patterns that can be produce by strong lateral and horizontal contrasts of seismic properties and wave interferences. In the case of anthropogenic sources, time windows may show outputs with distorted shapes or multiple spots (Figure 2.13). It is not possible to conclude on whether those outputs are due to multiple/moving sources, or the inability of calculations with an overly simple replica to match data associated with one source.

A simple way to address the first issue and improve resolution at depth would be to use a 1D velocity gradient. In such a case, it is still possible to compute the replica analytically and the model is closer to reality. Figure 2.14 shows two outputs computed with velocity gradients of 5s-1 and 9s-1and velocity at the surface of 800 m/s. This value corresponds to the velocity at the elevation of the lowest sensor in the average

1D velocity model below the array derived from data of the Betsy gunshots by Meng et Ben-Zion (2018b). As a simplification, when studying sources at depth, we consider the lowest elevation to be 0-meter depth and do not invert for source in the volume above this boundary. We observe improved resolution and higher value for the maximum of the MFP output when using a model with 9s-1 gradient. This is generally consistent the model from Meng et Ben-Zion (2018b) and imaging results in the area (Hillers *et al.* (2016), Roux *et al.* (2016)). Another way to reduce depth uncertainties would be to use data from the borehole sensor located within the array, to better separate sources at and below the surface and constrain the vertical position of sub-surface sources. This will be investigated in Chapter along with the use of the model of Meng et Ben-Zion (2018b) to compute the replica. Chapter will focus on the localization results for a part of the 30-day dataset.

# Chapitre 3

# Time monitoring of the San Jacinto fault zone

## 3.1 Introduction

The main focus of this chapter is to find evidence of very shallow events under and around the array. In a lesser measure, we are also interested in the position and occurrence time of the surface noise sources present in the area, because they can give us a comparison point to test potential shallow sources.

We applied the method presented chapter 2 to 26 days of the 30-day dataset. The remaining days correspond to the first two and last three days of the experiment were not used because few sensors were functioning (less than 400). Successive time windows of the data were processed at 4Hz and 16Hz. For each scanned window a MCMC-output was computed. Windows with output values above a given threshold are classified. The localizations at 4Hz are then studied to derive the positions of the main surface noise sources inside and outside of the array. Finally, we focus in more detail on localizations at 16Hz in order to determine whether very shallow seismic sources are present.

## 3.2 Choice of the noise threshold

The first step of this study is to determine a localization threshold. Windows with outputs under this threshold are considered to contain only noise. In chapter 2, we used a theoretical statistical threshold, assuming that noise is uncorrelated from station to station. We take advantage of the important number of processed windows to study the distribution of values of the MCMC outputs in order to determine an empirical threshold for detections and verify this assumption.

We use all 26 days of data processed at 4Hz and at 16Hz. This corresponds to 2246400 processed windows of length 1s at 4Hz, and 8985600 processed windows of length 0.25

seconds at 16Hz. We use the distribution of the maximum of the outputs at both frequencies, represented in Figure 3.12, to determine thresholds ($T_{exp4}$ and $T_{exp16}$) above which an event is detected. To reduce false detections, we use the 95th percentile, ensuring that only 5% of the windows are above the thresholds. In this case, $T_{exp4} = 0.1$ and $T_{exp16} = 0.03$. We find that, over the 26 days, 133932 windows have outputs above our threshold at 4Hz, against 334000 at 16Hz (Table 3.1). This constitutes a daily average of 5000 and 12000 detections and localizations for 4Hz and 16Hz respectively.

It is possible to compare the empirical threshold to the theoretical limit of incoherent noise level for an array of N sensors. This corresponds to the power reduction associated with averaging N realizations of incoherent signals and is equal to $T_{th} = 1/N$ (blue line in Figure 3.12). This threshold is one or two order of magnitude below the mean of the 4Hz and 16Hz MFP output distributions. This means that most of the noise recorded by neighbouring sensors is spatially coherent, which can be explained by the density of the array. This also explains the difference between the chosen thresholds at 4Hz and 16Hz. At 4Hz, the wavelength is longer and the noise will be more spatially coherent, resulting in higher MFP values.

| | 4Hz | 16Hz |
|---|---|---|
| Number of windows above threshold | 133932 | 334132 |
| Number of localizations | 64824(48%) | 166888(50%) |
| Number of detections | 69108(51%) | 167244(50%) |

TABLE 3.1 – Distributions of localizations and detections over 26 days at 4Hz and 16Hz

We distinguish detections, which have outputs higher than our chosen threshold but converged at the parameter space boundaries, from localizations. While we cannot conclude on the on relocated source position and apparent velocity in the case of detections, they can have high MCMC output values that suggest a true source outside of the boundaries of our chosen parameter space. Localizations constitute half of the windows with outputs above the chosen threshold. We focus on them for the rest of this chapter.

## 3.3 Localization classification and statistics

We wish to separate localizations outside and inside the array, because the examples provided in chapter 2 demonstrated that while we can locate events outside the array, their epicentral positions are not well constrained, with outputs that are elongated in the radial direction. Moreover, the apparent velocity for events outside the array will greatly depend on the distance between the source and the array, making interpretation difficult. In order to classify the localizations, we use simple criteria based on the envelope of the output.

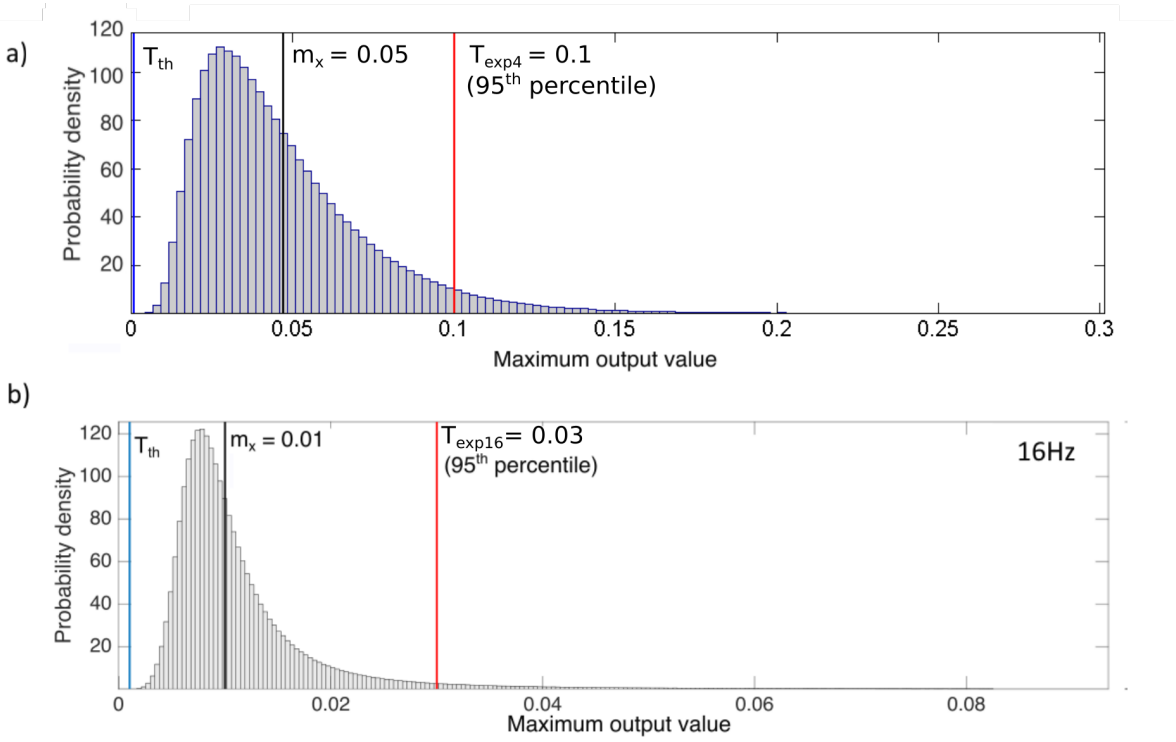


FIGURE 3.1 – Distribution of the maximum values of the MCMC outputs of all windows for 26 days. The theoretical threshold $T_{th}$ and the mean of the distribution $m_x$ are represented by the blue line and the black line, respectively. The chosen threshold $T_{exp}$ ($95^{th}$ percentile) is the red line. a) Distribution of the results at 4Hz. b) Distribution of the results at 16Hz.

### 3.3.1 Classifications of events

We classify our localizations into three different categories. The first two correspond to localizations inside and outside the array, respectively. Localizations in the third category, which are referred to as undetermined, are all the localizations with misshapen or ambiguous outputs that could be due to the sampling of side lobes by the MCMC scheme or aliasing effects and are considered here poor quality localizations.

In order to tag each localization, we rely on the shape of the output as well as on its position. In order to be considered outside the array, a localization must have an elongated MCMC output in the radial direction. To evaluate this automatically, we fit an ellipse to the envelope of the MCMC output at -3dB, i.e. taking only into account the points that have a value greater than half of the maximum. In practice, an event is categorized as outside the array if :

1. $l_{shortaxis} < \frac{1}{2} l_{longaxis}$

2. $l_{longaxis} > 200m = \frac{1}{4}l_{ap}$
3. $l_{shortaxis} < 600m = \frac{3}{4}l_{ap}$
4. Most of the points of the output are located outside the array

With $l_{shortaxis}$ the length of the short axis, $l_{longaxis}$ the length of the long axis and $l_ap$ the maximum aperture of the array.

While condition 4 is technically self sufficient, the other conditions eliminate distorted outputs. Condition 1 and 2 insures that we have an elongated output while condition 3 discards outputs with multiple optima that we cannot interpret. Figure 3.2 illustrates these conditions on 4 examples of outputs, with their envelopes and fitted ellipses. The output in Figure 3.2-a meets the conditions mentioned above and is classified as outside the array. On the other hand, because the output displayed in Figure 3.2-c has two lobes, it was not well fitted. The short axis of the fitted ellipse is greater than 200m. The localization is classified as undetermined.

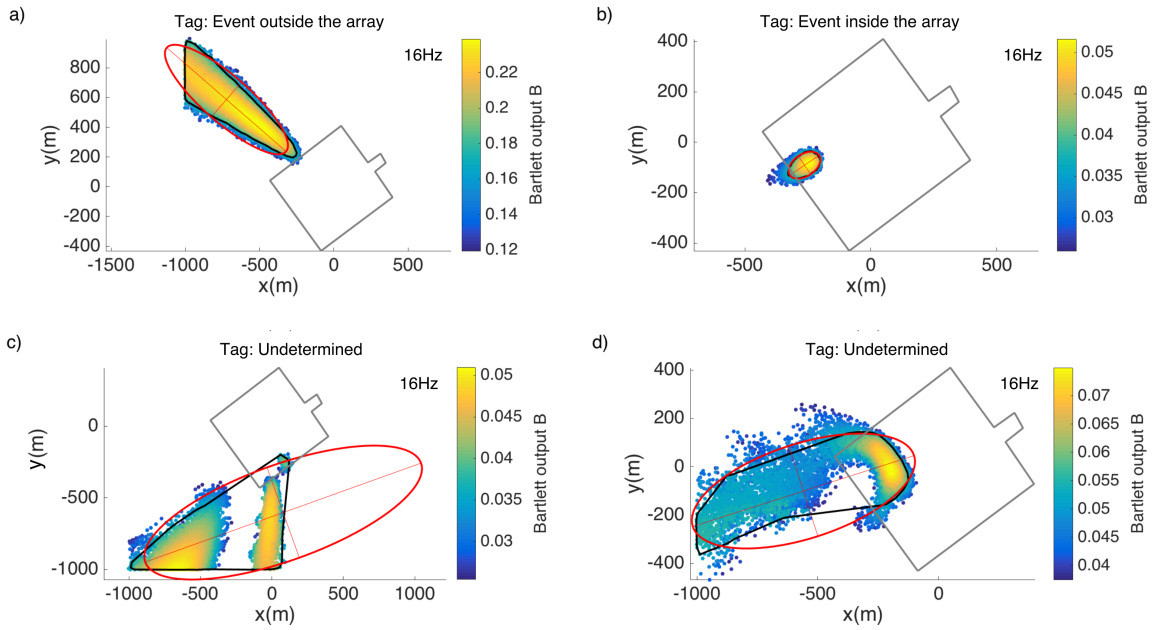


FIGURE 3.2 – Example of different outputs, their envelopes (black) and fitted ellipses (red). a) Output associated with a localization categorized as outside the array. b) Output associated with a localization categorized as inside the array. c) Output associated with a localization categorized as undetermined. d) Other output associated with a localization categorized as undetermined.

Similarly, an event is categorized as inside the array if :

1. $l_{shortaxis} > \frac{1}{2}l_{longaxis}$
2. $l_{shortaxis} < 400m = \frac{1}{2}l_{ap}$

| | 4Hz | 16Hz |
|---|---|---|
| Nb of localizations inside array | 35601 (55%) | 6316 (4%) |
| Nb of localizations outside array | 29159 (45%) | 150555 (90%) |
| Nb of localizations undetermined | 63 (<1%) | 10003 (6%) |

TABLE 3.2 – Distributions of the types of localizations at 4Hz and 16Hz

3. $l_{longaxis} < 400m = \frac{1}{2} l_{ap}$
4. Most of the points of the output are is located inside the array

The output in Figure 3.2-b meets the conditions for a event inside the array and is classified as such. The output displayed in Figure 3.2-d is distorted, resulting in an elongated output that is not expected inside the array. The localization is classified as undetermined.

The conditions of classification are deliberately restrictive. For example, the event displayed in figure 3.2-d, could have been accepted as a event inside the array. Our classification is very simple and we have a great amount of data. Because the main goal is to find strong evidence of events in the shallow crust, we choose to use more restrictive conditions.

An examination of the distribution of the localizations (3.2) reveals that while the localizations outside and inside the array at 4Hz are in equal proportions, the majority of the localizations at 16Hz corresponds to sources outside the array. There are more undetermined localizations at 16Hz, which could be explained by the fact that some of the strong sources at depth are very far from the array, resulting in distorted phases patterns that translate into misshapen outputs.

### 3.3.2 Borehole data

One of the main challenge of the study presented in this chapter is the characterization of the localizations of interest. One of the strategy is to study the time traces of the working sensors present on site at the time of the localizations. Some weak events may not be visible on time traces of individual sensors at the surface. To characterize our localizations, we can also rely on the data of a borehole located in the southwestern quadrant of the array at a depth of 142m (Figure 3.3) that records continuous high quality 3 component data.

### 3.3.3 Localization clustering

Our algorithm allows the localization of the spatial origin of coherent arrivals present in a window. Most of those arrival are part of larger recorded events. We expect events

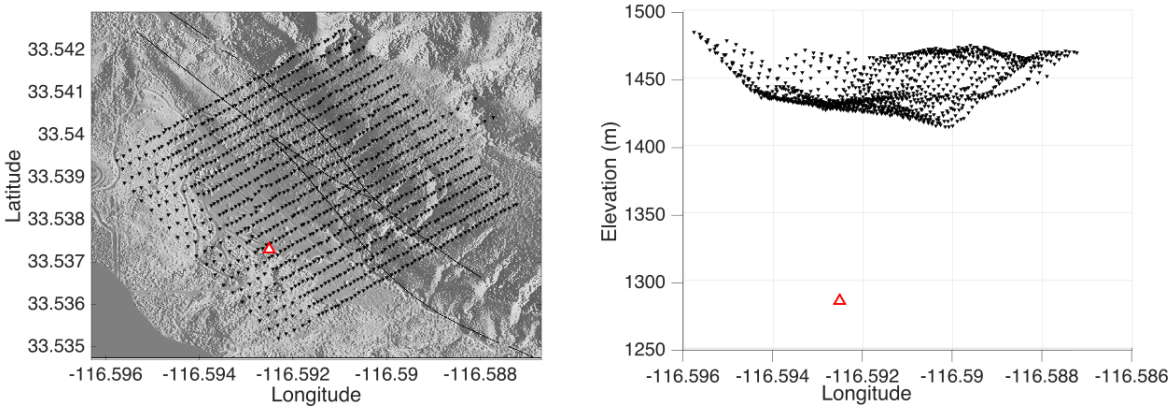


FIGURE 3.3 – Position of the borehole seismometer (red triangle) with respect to the array (black triangle).

from the surface as well as depth, with various duration time depending on their nature and distance from the array.

Figure 3.4 displays two examples of events with their associated localizations. One of them is a regional seismic event and the other is a surface event (Moving vehicle).

In the case of seismic event, we see two arrivals, one with stronger vertical component and another with more energy on the horizontal component of the borehole data. (Figure 3.4-a, top three panels). The signal from the seismic event is strong enough to be recorded above noise level at the surface (Figure 3.4-a bottom panel). The difference between P and S arrival times ($t_p - t_s = 6s$) suggests a regional event. Figure 3.4-b is a 2D view of the envelopes of the outputs corresponding to the event. The set of localizations fits within the duration of the event but the localizations do not correspond to distinctive arrivals on the time series. However we have a detections for both direct P and S-waves (Black blocks in figure 3.4-a). Given that this is a regional event, it is expected that direct waves would have a too high apparent velocity to be localized. This means that what we localize are other coherent seismic phases that we cannot identify in the time series but that are the results of diffraction or reflections on heterogeneities closer to the array.

The signal generated by the moving vehicle (Figure 3.4-c) is very different from the seismic event, but was also successfully localized. In this case, we have successive localizations that follow the landing path besides the ranch (Figure 3.4-d). The signal from the surface is also recorded at the borehole.

While the study of events defined as cluster of localization would make more physical sense, grouping localizations is not trivial, as some events are very weak. The sequence of localizations for a specific event can be disrupted by stronger, more transient signal that is not necessarily coherent at the frequency of interest.

Conditions involving the position and apparent velocity could be devised, because

localizations corresponding to the same event should have similar positions and apparent velocities. However, in the case of regional event localization, the variability of the output position is high (Figure 3.4). This variability is due to the distortion of the phase caused from multiple reflections and the short length of the time windows. In addition to that, moving surface source would also pose a problem. Because we could not find simple criteria to group our localizations, we will focus on studying individual localizations instead of clusters that would represent events.

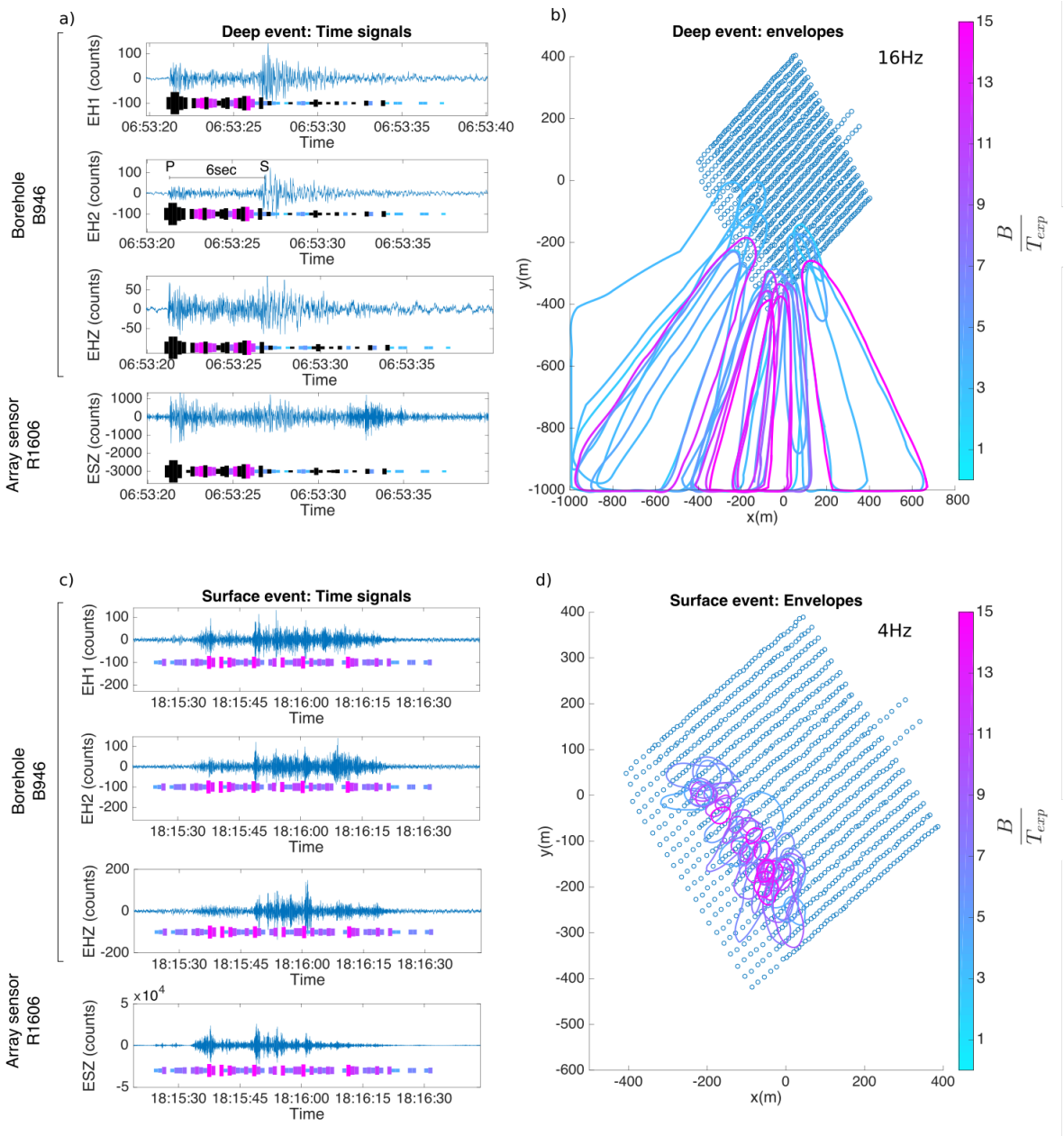


FIGURE 3.4 – Representation of events as clusters of localizations and detections a) Top to Bottom : 1st horizontal component of the borehole, 2nd horizontal component of the borehole, vertical component of the borehole, and vertical component of a sensor of the array. Each trace is filtered between 0.5 and 49.5Hz. The time windows corresponding to the localizations are represented by blocks of varying heights and colors associated with the maximum of the output, black blocks are detections. b) Envelope of the outputs corresponding to each localizations. The color corresponds to the maximum of the output. c) Same as a) for an anthropogenic source. d) Same as b) for an anthropogenic source

We begin our study by observing the overall trends in the 26 days. This will allow the highlighting of the main sources inside and outside the array, for localizations at 4Hz and 16Hz, which are sensitive mostly to surface and deeper sources, respectively.

## 3.4 Localizations at 4Hz

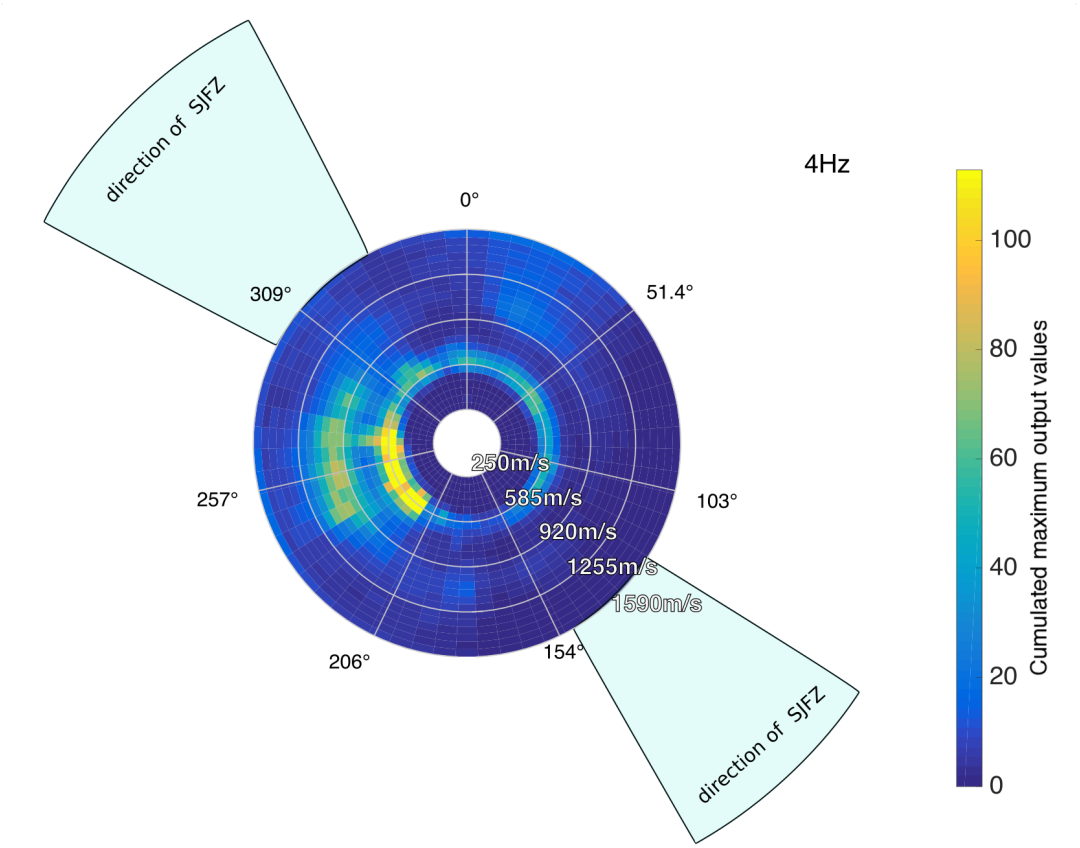


FIGURE 3.5 – Stacked localizations according to the apparent velocity and back azimuth at 4Hz, weighted by the value of the maximum output for each localization. Here we use output values between 0 and 1 instead of the output normalized by the noise threshold defined in section 3.2. The direction of the fault is indicated by the blue cones

### 3.4.1 Events outside the array

It is possible to determine a back azimuth from the MCMC outputs of localizations outside the array. This is achieved by fitting a line to the output and computing the angle between this line and the north. The localizations can then be represented in polar plots with the apparent velocity as the radial coordinate (Figure 3.5).

Figure 3.5 shows the stacks of individual localizations, weighted by the maximum value of their output, according to direction and apparent velocity. The dominant apparent velocities (500m/s to 1000m/s) correspond to surface wave velocity, confirming that the majority of the localizations at 4Hz can be attributed to surface sources. While the dominant localizations are at the west of the array, consistent with the fact that the nearest anthropogenic sources and infrastructures are on this side, we see that a significant amount of localizations at all other azimuths. This shows that in addition to anthropogenic surface sources, natural source make a non-negligible contribution to the recorded signal.

### 3.4.2 Localizations inside the array

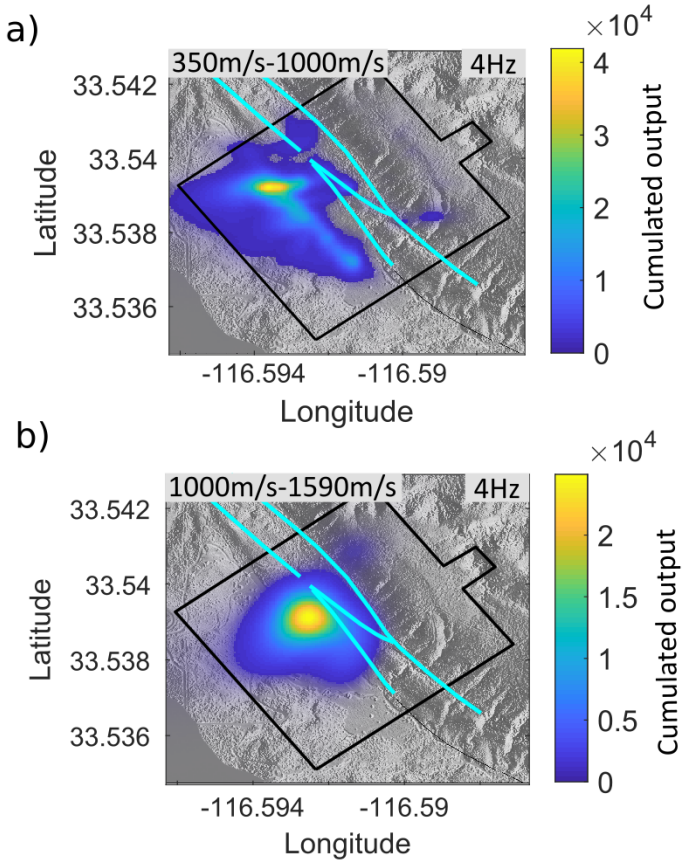


FIGURE 3.6 – Cumulated outputs of localizations under the array at 4Hz stacked into 5mx5m bins and summed over velocities from a) 350m/s to 1000m/s, b) 1000m/s to 1590m/s

A significant number of localization at 4Hz are on the array. In order to find the more dominant sources, we stack all the outputs of individual localizations on the array

into 5mx5mx15m/s bins. The stacks are then summed over different velocity ranges and displayed Figure 3.6. The velocity ranges are [350m/s 1000m/s], which correspond to surface velocities measured at 4Hz (Roux *et al.* (2016)), and [1000m/s 1590m/s] which could be associated with very shallow sources inside the array. The stack of the outputs of all the localizations inside the array at 4Hz exhibits an area with dominant sources that coincides with the main buildings of the ranch and the landing strip for velocities between 350m/s and 1000m/s (Figure 3.6-b)). At higher velocities, we observe a dominant source on the left side of the fault (Figure 3.6-b).

## 3.5 Localizations at 16Hz

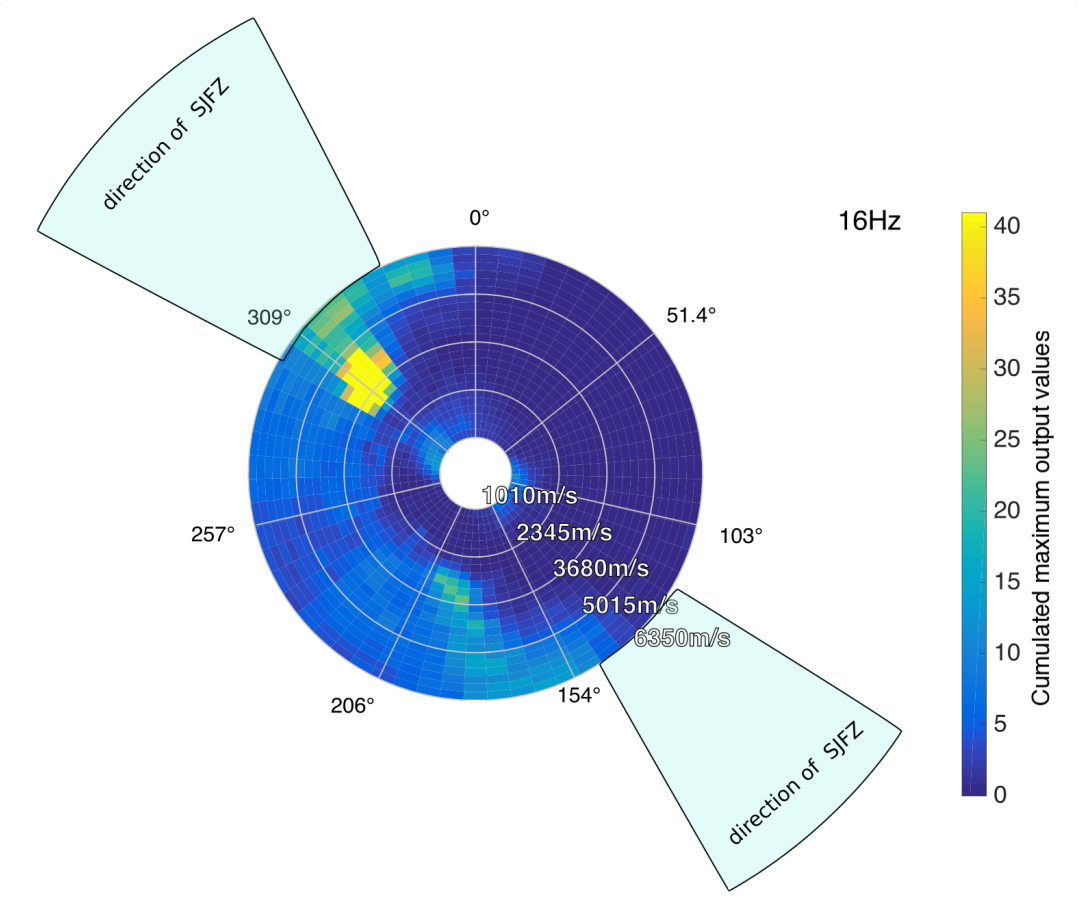


FIGURE 3.7 – Stacked localizations according to the apparent velocity and back azimuth at 16Hz, weighted by the value of the maximum output for each localization. The direction of the fault is indicated by the blue cones.

The study of the localizations at 4Hz confirmed the diversity of the surface sources present on and around the array. A significant number of sources were localized all around

the array. Inside the array, we find sources on both sides of the faults.

### 3.5.1 Events outside the array

#### 3.5.1.1 Dominant source positions

Figure 3.7 is a stack of the all the localizations outside the array from the day 130 to 155. It is possible to see dominant localizations around 309°with apparent velocities ranging from 3500m/s to 4500m/s. This direction coincides with the direction of the SJFZ.

Because apparent velocity can be used as a proxy for depth, we can hypothesize that the dominant sources outside the array, with back azimuths centered on 309°are good candidates for subsurface sources. Their apparent velocities, ranging from 3500m/s to 4500m/s indicate waves coming from depth, but still traveling in the subsurface. However, the position of the sources outside the array could also mean that these waves are diving body waves emitted from a source at the surface. The use of the 16Hz frequency does filter surface sources inside the array, as they emit waves at lower apparent velocities, but diving P-waves emitted by sources far from the array would have higher velocity and could be measured at such frequency. Concluding on these two candidates for subsurface sources require further study and will be investigated in the next section.

There is also a number of localizations with velocities under 1500m/s. These apparent velocities are too low to be associated with diving body waves coming from outside the array, but to high to be associated with surface waves at 16Hz. We suppose they correspond to surface events as they are as they are localized within the lower part of the velocity range of our study, but again, further characterization of such event is necessary in order to draw conclusions.

#### 3.5.1.2 Characterization of the dominant sources outside the array

**Localization with 309°back azimuth**

In order to characterize the localization at 309°, we first studied traces from the array but also from the borehole and the surrounding permanent stations and as well as stacks of the array data, looking specifically for impulsive signal on the borehole or wind induced signal (Johnson *et al.* (2018)) or traffic/machinery on surface sensors. No characteristic time signal was found in the windows corresponding to the localizations. The next step was to examine the time variation of the sources activity by measuring the number of localizations with respect to time.

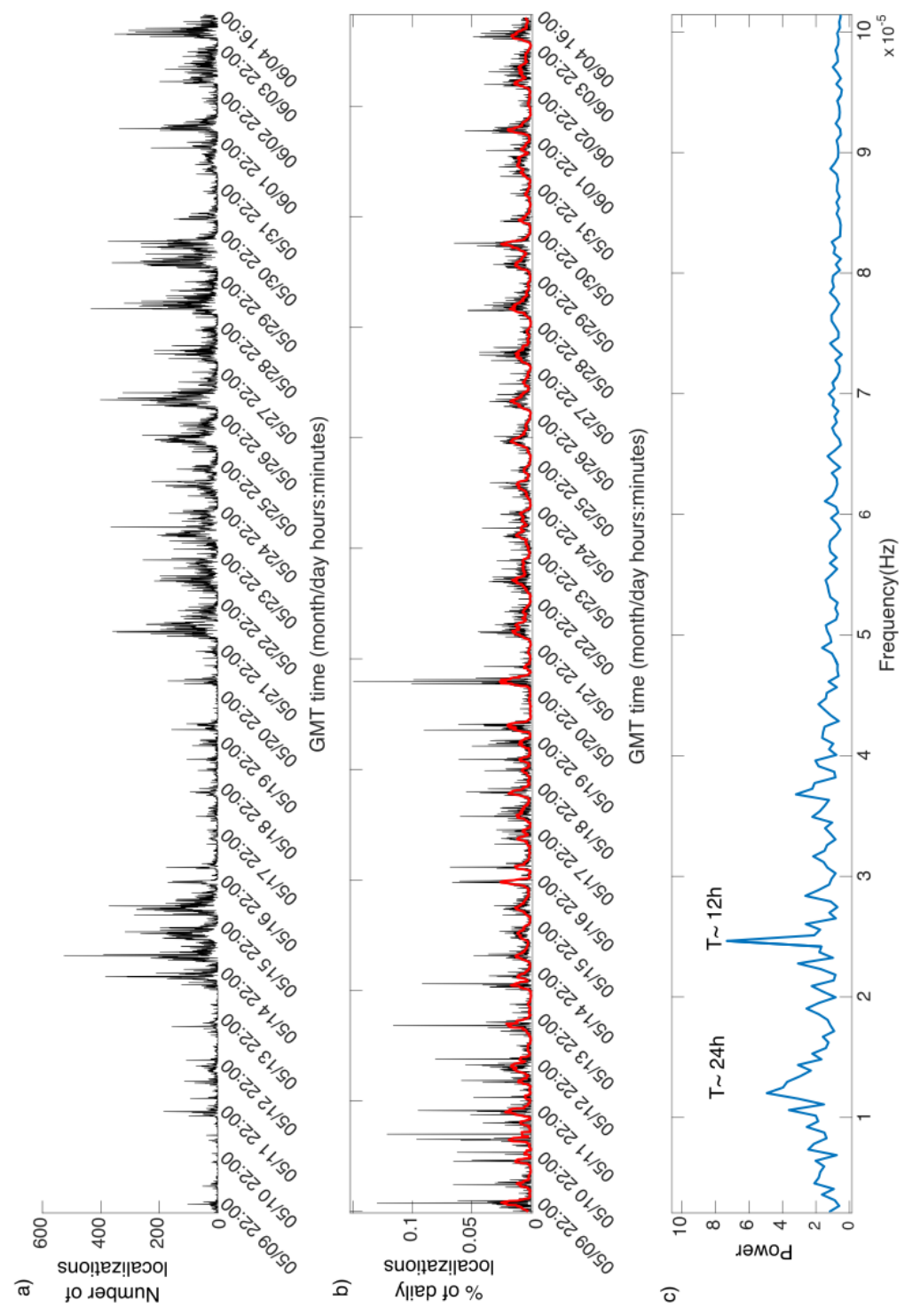


FIGURE 3.8 – a) Time variations of localization counts for back azimuths around 309°and apparent velocities around 3750m/s, according to local time for 26 successive days. b) black line : Time variations of the localization couns normalized by the daily number of localization. The red curve is obtained through a moving mean over 2-hour windows c) Spectrum of the moving mean of normalized localization counts.

Figure 3.8-a shows the time variations of the number of localizations for all 26-day according to the local time. The localizations are stacked in 7.5 minutes bins. These stacks are normalized by the number of events per day in order highlight repeating patterns regardless of the daily activity (Figure 3.8-b).

While the amount of measured activity varies over the days, we see patterns that seem to follow a 24h cycle. This trend is better observed in the normalized localization count Figure 3.8-b and is confirmed by the study of the spectrum, which shows clear peaks at corresponding to periods of 24 hours and 12 hours. The spectrum was computed by performing a FFT on the moving mean signal. Figure 3.9 shows a stack of the localization count over 24 hours for all 26 days, according to local time. The part of the stack from 19h to 24h was duplicated (gray part) to better represent the variations over nights.

This variation in activity, with two peaks at 7am and 7pm and reduced activity during the night could suggest surface events linked to anthropogenic activity. An other argument on the favor of categorizing the studied event as surface events is fact that events at depth should have a better localization rate during the night because there is less surface noise. However, we could counter that the use of higher frequency efficiently filters close surface sources, which are the most problematic, mitigating their effect on the localization rate.

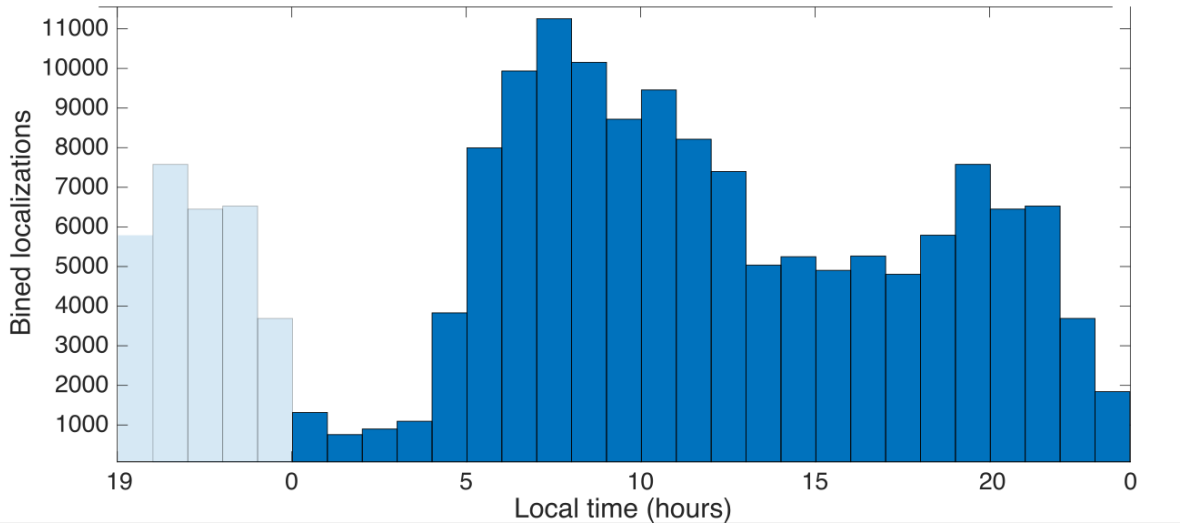


FIGURE 3.9 – Blue : stack of all localizations with back azimuths around 309°and apparent velocities around 3750m/s, over 24 hours according to local time. Grey : duplicate of the stack between 19 :00 and 23 :00, used to highlight the periodicity

Sources at the surface radiating diving body waves would also emit strong surface waves that should be detected with a delay. Finding a consistent delay between localizations at 16Hz (diving waves) and 4Hz (surface waves) could orient us toward surface sources

and give us an idea of the distance of the sources from the array. To test this theory, we compare the 16Hz localization times with the localizations at 4Hz in the same direction. Figure 3.10 represent such delay for all 309°localizations at 16Hz. In practice, we looked for localizations at 4Hz included between 0.5 and 30 seconds after the localizations at 16Hz.This time interval corresponds roughly to distances ranging from 500m to 15km.

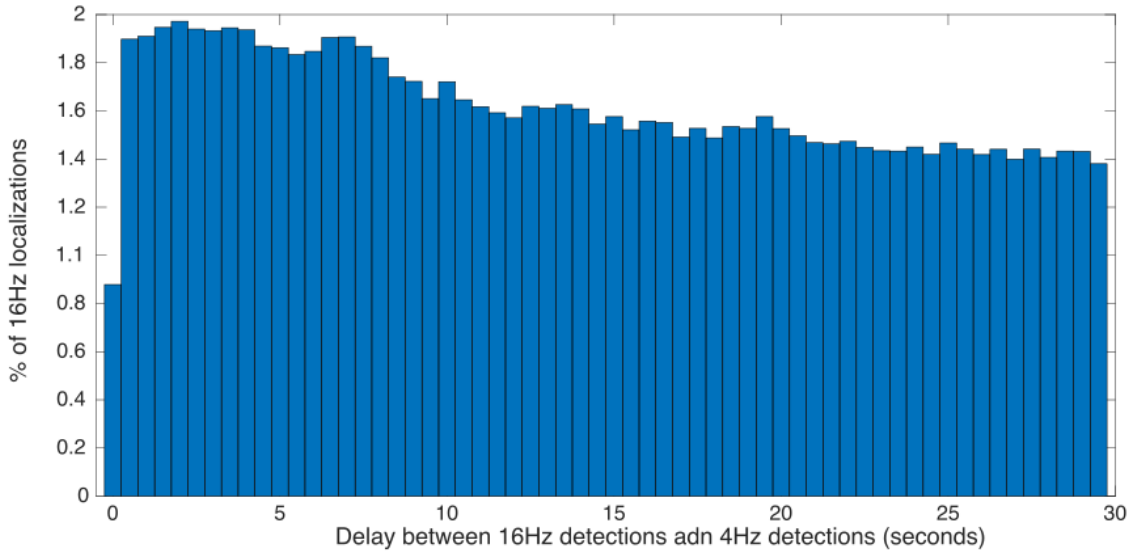


FIGURE 3.10 – Stack of the delay between 16Hz and 4Hz localizations in the 309°direction

We see first that no clear trend is visible in term of delay between 16Hz and 4Hz localizations. Secondly, the accumulation of all the bins adds up 12% of the 309°localizations at 16Hz. This means that at best, only 12% of 16Hz localizations have a corresponding 4Hz localization within the next 30 seconds. Those results are not sufficient to conclude on the nature of the sources. An other possibility is a surface wave to body wave conversion due to the medium structure and topography. In this case, no surface wave would be measured at 4Hz, explaining our results.

We could also be looking at very shallow events induced by temperature changes, as the peaks in activity are consistent with the sunrise and sunset times during this period of the year. We could test that hypothesis by measuring the daily ground level variations through GPS measurement for example.

### Localizations with low velocities at 16Hz

A type of localization that raises questions is the very low apparent velocity localizations located outside the array. Figure 3.7 showed that a non-negligible number of

localization have apparent velocities around 1200m/s. These localizations most likely correspond to surface source outside the array, but their apparent velocity is higher than what is expected of surface waves. We performed a time frequency analysis on a sensor of the array for the times associated to strong localizations with low apparent velocities. Figure 3.11 is a representative example of the 21 time windows studied. We see that the localizations correspond to a very distinctive signal which is generated by a helicopter (Meng et Ben-Zion (2018a)). We suspect we measure an aliased acoustic wavefield, because for each tested signal the localizations only occur after the frequency of the lowest harmonic has lowered around 16Hz. The strong harmonic nature of the signal and its frequency content could result in coherent aliasing that translates into increased apparent velocity.

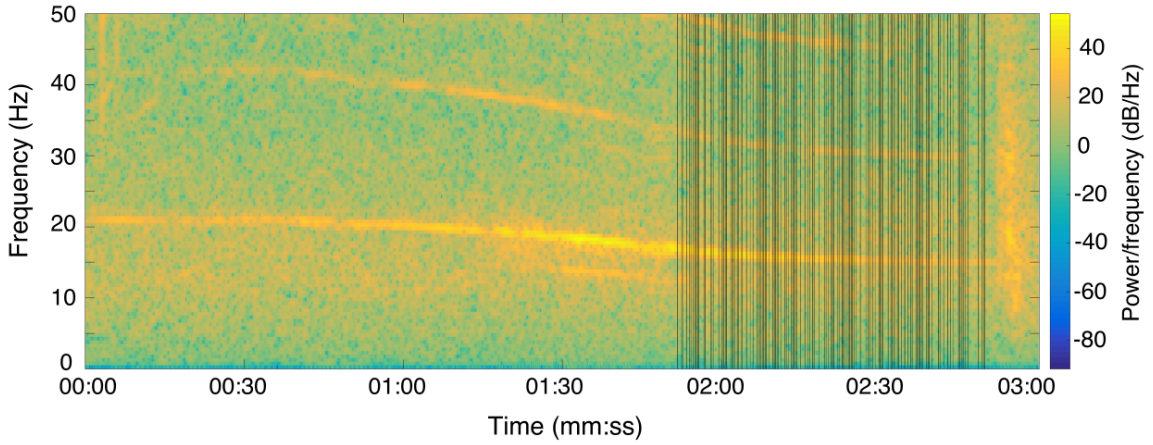


FIGURE 3.11 – Spectrogram of a time window associated with a low velocity localization on Julian day 133. The time of the localizations are represented black lines.

We could not conclude on the presence of sources at very shallow depth from the results of the study of events outside the array. We focus next on sources inside the array.

## 3.5.2 Event under the array

### 3.5.2.1 Dominant source positions

Figure 3.12 shows stacks of the outputs into 5mx5m bins for different velocity ranges. Figure 3.12-a, 3.12-b, 3.12-c correspond respectively to velocities ranging from 970 to 2000m/s, 2000m to 4500m/s, and 4500 to 6350m/s. We see that most of the localizations under 4500m/s are located at the position of the main body of the ranch. This correlation with the position of what is one of the major surface source is surprising and will be

investigated latter on. At higher velocities the dominant localizations are located across the fault.

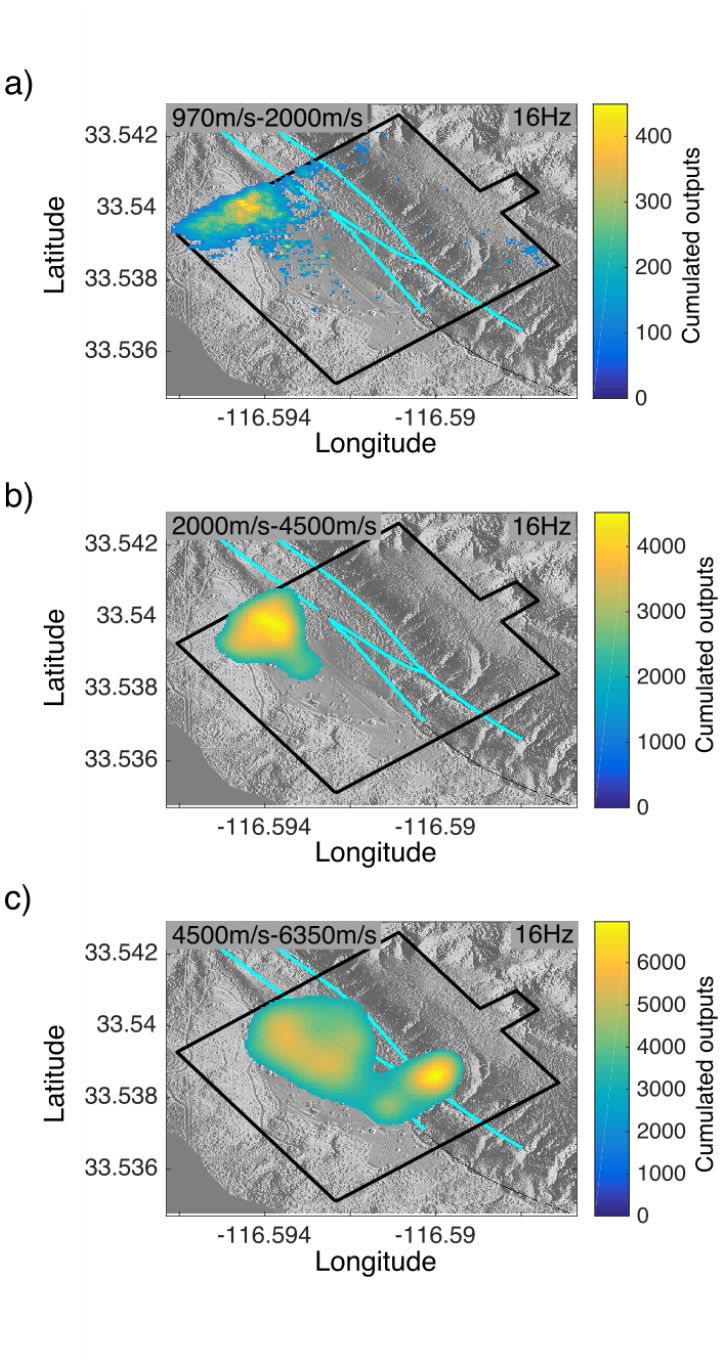


FIGURE 3.12 – Cumulated outputs of localizations under the array at 16Hz stacked into 5mx5m bins and summed over a) [970m/s 2000m/s] b) [2000m/s 4500m/s] c) [4500m/s 6350m/s]

#### 3.5.2.2 Characterization of the dominant sources under the array

The dominant feature for middle range apparent velocities (3000-5000m/s) coincides with the ranch. Such strong correlation with a surface structure is unexpected, as the apparent velocities should be high enough to filter out most surface sources inside the array. However, the position of the ranch also coincides with a low velocity zone present at 200m depth (Mordret *et al.* (2018)). Such heterogeneity could be at the origin of diffraction, resulting in a localization. The study the time series recorded by the borehole shows that localizations under the array often occur during regional events, between P and S arrivals. From this, we suspect that instead of sources, we detected diffracted waves on heterogeneities under the surface.

## 3.6 Conclusion

We obtained a great number of localizations from the scan of 26 days of data at both 4Hz ad 16Hz. Localizations were tagged as in the array, outside the array or undetermined depending on the positions and shape of their associated outputs. We studied the stacks of localizations inside and outside the array separately, for both 4Hz and 16Hz. From this we defined dominant sources of interest that were studied more in details. No major conclusions were made on the nature of the sources outside the array with the results available, arguments can be found for both surface and deeper sources. Events at low apparent velocity were identified as helicopters, and localizations under the array seem to correspond to diffracted waves instead of sources at shallow depths. We demonstrated here that, while the use of apparent velocity is enough as a first approach to highlight sources of interest, it is sometimes not sufficient to draw conclusions on the nature of those sources. The use of 3D localization could help us better characterize unknown sources, in addition to the estimation of their depth. The next chapter will investigate solutions that could allow the invertion for depth with sufficient resolution.

# Chapitre 4

# 3D localization

## 4.1 Introduction

Chapter 2 proved that 3D localization is not successful when using a replica computed with a homogeneous model with the velocity as a parameter. The position at depth is still critical information in terms of insight into near surface seismicity, as was demonstrated by the results of the scan of the data in chapter 3. While several promising dominant sources where identified, in some cases, no conclusion could be made on their nature.

In this part we investigate approaches that improve depth resolution. Two main strategies have been studied. The first one is the use of more realistic models accounting for velocity variation with depth. Two models will be tested against a homogeneous velocity model, a 1D gradient and a 2-layer model that was inverted by Meng et Ben-Zion (2018b). A closed form is directly available for the first case, or can be derived from an approximation in the second, allowing the analytical computation of travel times between the surrogate source at depth and the array at the surface.

The other strategy is to include weighted borehole data to the surface measurement to create a pseudo 3D dataset. The first section presents the two gradient-based models that are investigated in this chapter. The integration of borehole data into the inversion and the choice of weight are then discussed. The different inversions scheme are finally tested on chosen events in order to determine the best strategy to improve the resolution at depth.

## 4.2 Gradient-based models

A preliminary result presented in chapter 2 showed that for an event outside the array, the use of a vertical gradient model reduced significantly the uncertainties at depth. We investigate the use of two different acoustic models to compute our replica. The first one is a 1D vertical gradient model that involves 5 inversion parameters ; the three source coordinates, the surface velocity and the gradient value. The second model

| Name | Model | Inverted parameters [range] |
|---|---|---|
| HOM2DV | Homogeneous, acoustic | x [-1000m 1000m]<br>y [-1000m 1000m]<br>$v_{ap}$(4Hz) [240m/s 1600m/s]<br>$v_{ap}$(16Hz)[960m/s 6400m/s] |
| HOM3DV | Homogeneous, acoustic | x [-1000m 1000m]<br>y [-1000m 1000m]<br>z [0m 4000m]<br>v [350m/s 5000m/s] |
| GDT | Vertical gradient, acoustic | x [-1000m 1000m]<br>y [-1000m 1000m]<br>z [0m 4000m]<br>$v_{surf}$ [350m/s 1200m/s]<br>G $[0.1s^{-1}\ 20s^{-1}]$ |
| 2L | 2-layer model, acoustic, P-wave velocity | x [-1000m 1000m]<br>y [-1000m 1000m]<br>z [0m 4000m] |

TABLE 4.1 – Inversions schemes

is an acoustic model for P-wave velocity comprising two layers, inverted by Meng et Ben-Zion (2018b). In this case only the three source coordinates are necessary for the inversion. For the remainder of this chapter, the inversion schemes relying on the two gradient-based model are called GDT and 2L, respectively. We compare their results to the results of a 4-parameter inversion using an acoustic homogeneous model (HOM3DV). All three inversion schemes are summarized in table , which contains the characteristic of the associated models and the parameters involved.

All the models used for the inversion are acoustic, and an exact or approximated closed form solution can be derived in each case. The use of replica computed numerically from elastic models has also been investigated. However, the computation time required to perform those computations is too important to consider using them as a first approach.

### 4.2.1 Vertical gradient model

In order to compute replica in a vertical gradient model we use the travel time formulation from Fomel *et al.* (2009) expressed in chapter 1, equation 1.2. The velocity of at any given depth $z$ of the medium is derived from the velocity at the surface $v_{surf}$ and the gradient $G$, using :

$$v(z) = v_{surf} + G(z - z_{surf}) \tag{4.1}$$

With $z_{surf}$ and $v_{surf}$ the depth and the velocity at the highest sensor of the array. We use the $G$ and $v_{surf}$ as parameters of our inversion, in addition to the Three source

coordinate, in order to account for 3D heterogeneities and model uncertainties. Once the travel time is computed, we can compute replica using a simple propagator. However, inverting for all 5 parameters will have an impact on the resolution of our localizations. Because of this, we also investigate a model obtained by Meng et Ben-Zion (2018b). As the velocity profile has already been inverted, we only invert for the source position.

### 4.2.2 2-layer 1D model

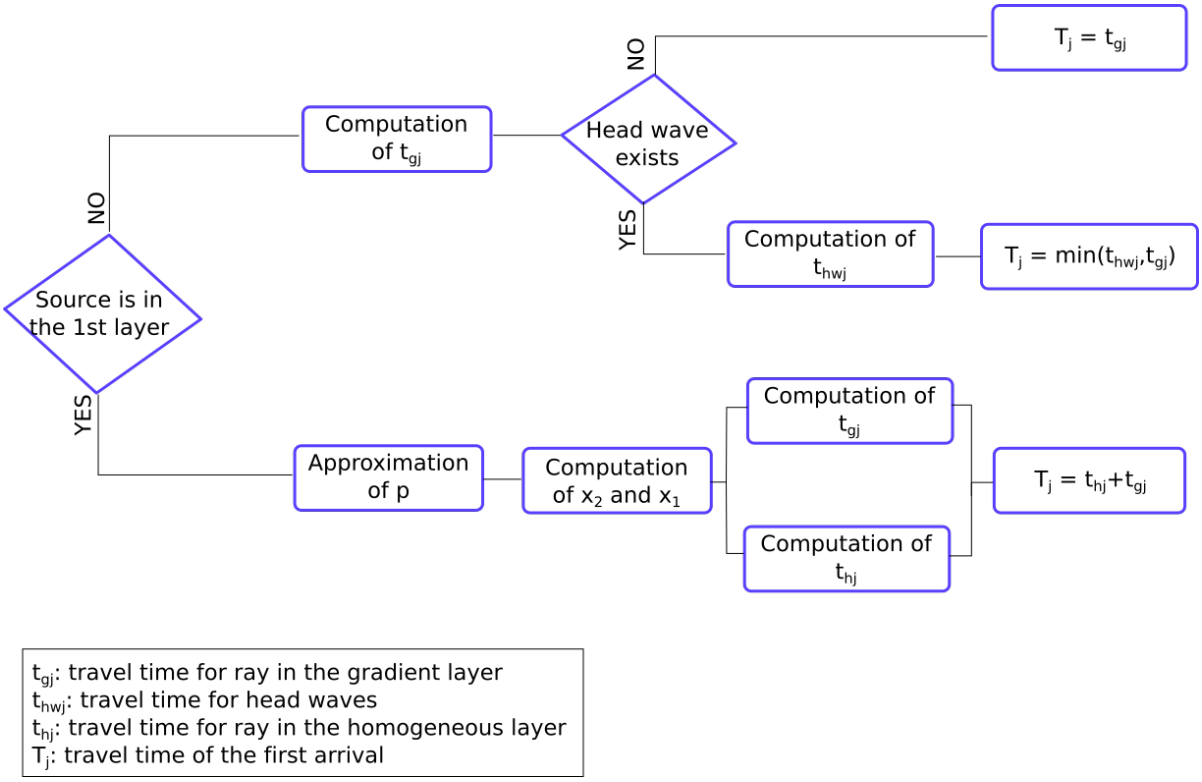


FIGURE 4.1 – Workflow of the computation of the travel time for the 2-layer model

To improve the quality of the model without increasing the computational cost, we investigate the use of a relatively simple model for P-waves from Meng et Ben-Zion (2018b). This model is a two-layer model comprising a vertical velocity gradient and a homogeneous layer with a sharp increase of velocity at their boundary. Travel times for sources in the to different layer are computed differently. Figure 4.1 summarizes the process in both cases and Figure 4.2-a 4.2-b displays the different possible ray paths, figure 4.2-c shows the approximated ray path used to compute travel times for sources in the $2^{nd}$ layer.

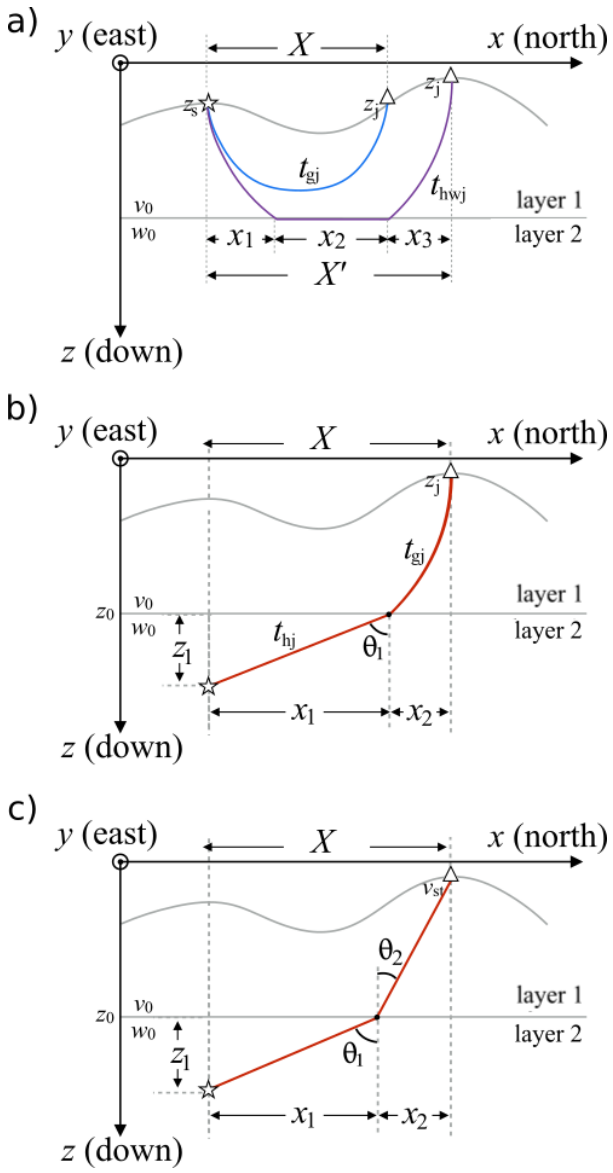


FIGURE 4.2 – a)Possible ray paths for a source in the 2nd layer. The blue ray is a typical bending ray and the purple ray corresponds to head waves b) Ray path for a source in the first layer c) ray path for the approximation $\tilde{p}$.

#### 4.2.2.1 Source in the gradient layer (1st layer)

When the source is in the top layer, two possible ray paths between the source and the sensors are considered.

1. The turning point of the ray is in the gradient layer (Figure 4.2 blue curve). In that case we can compute the travel time using the following expression, derived

from eq. 1.15 :

$$t_{gj} = \frac{1}{G} arcosh\left(1 + \frac{1}{2} S_j S_s G^2 \sqrt{X^2 + (z_j - z_s)^2}\right). \tag{4.2}$$

Here, $G$ is a scalar containing the gradient in the z-direction, $S_j$ is the slowness at sensor $j$ and $z_j$ is its elevation. $Z_s$ is the elevation of the source and $S_s$ is the slowness at the source position. $X$ is the epicentral distance between the source and the sensors.

2. The turning point of the ray is below the gradient layer (Figure 4.2-a purple curve). In that case, for specific ray parameters, head waves travel times can be computed. We follow the methodology proposed by Meng et Ben-Zion (2018b) to compute those travel times. The ray is separated into three parts. Horizontal distances $x_1$ and $x_3$, corresponding to the downgoing and upgoing parts of the ray are computed using eq. 1.16 :

$$x_{1j} = \frac{1}{Gp}\sqrt{(1 - pv_0)^2 - (1 - pv_s)^2}, \tag{4.3}$$

$$x_{2j} = \frac{1}{Gp}\sqrt{(1 - pv_0)^2 - (1 - pv_j)^2}. \tag{4.4}$$

Here $v_0$ is the velocity at the boundary in the gradient layer, $v_s$ is the velocity at the source and $v_j$ is the velocity at the sensor. The ray parameter is $p = \frac{1}{w_0}$, with $w_0$ the velocity in the homogeneous layer. From this we can deduce $x_2$ as :

$$x_2 = X - x_2 - x_3. \tag{4.5}$$

The corresponding travel times for the three part of the ray are :

$$t_{1j} = \frac{1}{G} arcosh\left(1 + \frac{1}{2} S_0 S_s G^2 \sqrt{X^2 + (z_0 - z_s)^2}\right), \tag{4.6}$$

$$t_{3j} = \frac{1}{G} arcosh\left(1 + \frac{1}{2} S_0 S_j G^2 \sqrt{X^2 + (z_0 - z_j)^2}\right) \tag{4.7}$$

and

$$t_{2j} = px_2. \tag{4.8}$$

The total travel time of the head wave between the source and sensor $j$ is

$$t_{hwj} = t_{1j} + t_{2j} + t_{3j}. \tag{4.9}$$

If the ray corresponding to diving P-waves and the ray corresponding to head waves can both be computed, the fist arrival is the shorter of the two.

#### 4.2.2.2 Sources in the homogeneous layer (2nd layer)

There is no analytical solution for travel times of rays traveling through both a homogeneous and a gradient layer. In order to obtain a closed form solution, we have approximated the ray parameter by considering, for this stage, that both layers are homogeneous (see Appendix A and Figure 4.2-c). From this approximated ray parameter $\tilde{p}$, we can derive the horizontal distances traveled by the ray in the homogeneous layer and in the gradient layer, called $x_1$ and $x_2$ respectively (Figure 4.2). $x_2$ is computed using eq. 1.16 :

$$x_2 = \frac{1}{Gp}\sqrt{(1-\tilde{p}v_0)^2 - (1-\tilde{p}v_j)^2}. \tag{4.10}$$

Here $v_0$ is the velocity at the boundary in the gradient layer and $v_j$ is the velocity at the sensor. $x_1$ is the difference between the epicentral distance between the source and the receiver $X$ and $x_2$ (Figure 4.2-c) :

$$x_1 = X - x_2 \tag{4.11}$$

We can then compute $t_{hj}$, the travel time in the homogeneous layer

$$t_{hj} = \frac{x_3}{w_0}, \tag{4.12}$$

and $t_{gi}$ the travel time in the 1-D gradient layer

$$t_{gj} = \frac{1}{G} arcosh\left(1 + \frac{1}{2}S_0 S_j G^2 \sqrt{X^2 + (z_0 - z_j)^2}\right). \tag{4.13}$$

The total travel time corresponding to the first arrival is the sum of the two travel times :

$$T_j = t_{hj} + t_{gj} \tag{4.14}$$

This approach allows the determination of travel time of the first arrival at each sensor for every possible source position in the model. As we cannot use the high frequency approximation in our case, this methodology is only valid for first arrivals. This is a limitation, because we cannot insure that the signal of interest contains clear arrivals. However, we expect impulsive events that could be detected by such process. From the travel times we can compute replica using a simple propagator. Similarly to the replica computed with a homogeneous model, we do not take into account amplitude variations across the array.

Before investigating the performance of the two gradient based models, we discuss the integration of weighted borehole data to the array dataset.

## 4.3 Investigation on the integration of borehole data into the data set

One of the main factors lessening the resolution at depth is the fact that all of our sensors are at the surface. Including the vertical data from the borehole present on site to the dense array data set is a way to improve the coverage and with it the resolution. Because of the greater number of sensors at the surface, the integrated data must be weighted to have an influence on the quality locations. In practice, the weight is applied after the data is transformed into the frequency domain and normalized.

The choice of the weight attributed to the borehole data is a balance between assuring a sufficient contribution of the signal measured at depth and keeping the processing power of our dense array. Putting too much emphasis on the signal recorded at depth would be equivalent to having one sensor. We perform a study based on the probability density function (PDF) of our MCMC output, as it can be a proxy for evaluating the resolution of the location. We estimate the PDF by binning the output according to depth and normalizing it. We use one out of ten points of the output to improve the statistical independence of the samples, which is necessary to estimate the PDF.

### 4.3.1 Preliminary test

In order to test the applicability of such integration, we perform a preliminary numerical test with a monochromatic source in a homogeneous acoustic medium. The source is located at 400m depth and at -50m and -120m from the center of the array in the x and y direction, respectively. The velocity in the medium is 2500m/s. We use the HOM3DV inversion scheme to locate the synthetic source, with varying weights applied to the synthetic borehole data.

The smoothed PDF for a set of chosen weights is displayed Figure 4.3. We see a sharpening of the PDF for weights ranging from 100 to 1000. For greater weights, the PDF collapses, consistent with the fact that the array behaves as a single station located at borehole position after a critical weight. 400 is the optimal weight, i.e. the weight for which the PDF is the most narrow.

We can also study the effect of the weighted borehole data on the relationship between the inverted depth and velocity. Figure 4.4 is a 2D representation of the PDF of the outputs for three different weights according to depth and medium velocity. The relationship between inverted depth and velocity is very clear in the case were no borehole data in integrated (Figure 4.4-a). For the optimum weight, Figure 4.4-b, we see that the maximum of the probability density centers back on the true source depth and velocity of the medium. For weights above 1000, the probability density is more spread out and is not centered on the true source and medium parameter values.

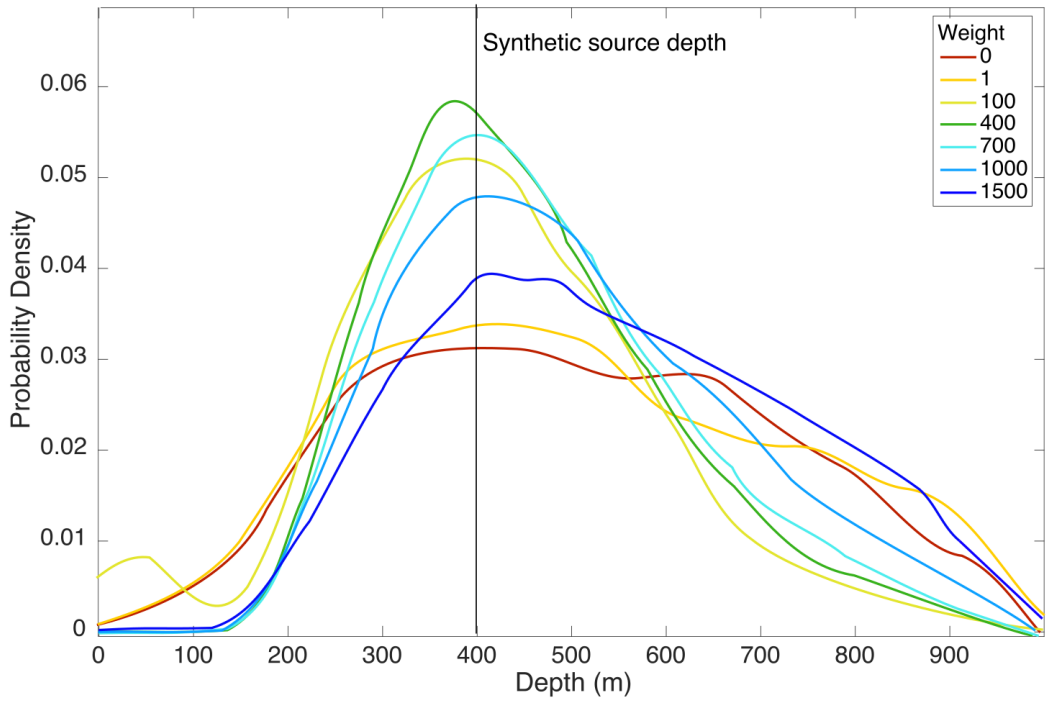


FIGURE 4.3 – PDF of source depth for several weights. The model used for the inversion is the homogeneous model. The synthetic data has been computed through a homogeneous acoustic model. 4 parameters are inverted : x, y, z, and v.

### 4.3.2 Choice of the weight for the HOM3DV, GDT and 2L inversion schemes

This preliminary study showed that for a simple case, the integration of weighted borehole data does improve the depth resolution. We want to combine this integration with all 3 inversion schemes presented in section 4.2. The next step is then to determine the optimal weight for each of the inversion scheme. To do so, and in order to further test the robustness of the process, we use the three elastic models that match each of the acoustic models used in the inversion schemes. Figure 4.5 shows the P-waves velocity variations for the homogeneous elastic model the vertical gradient elastic model and the 2-layer elastic model. The values of the S-waves velocity $v_s$ and the density $\rho$ are deduced from the P-wave velocities $v_p$.

$$v_s = v_p/1.73 \tag{4.15}$$

$$\rho = 310 v_p^{0.25} \tag{4.16}$$

The source position is the same as in section 4.3.1. We compute numerical waveforms with using SEM3D, a finite spectral element code, and derive the CSDM from the Fourier

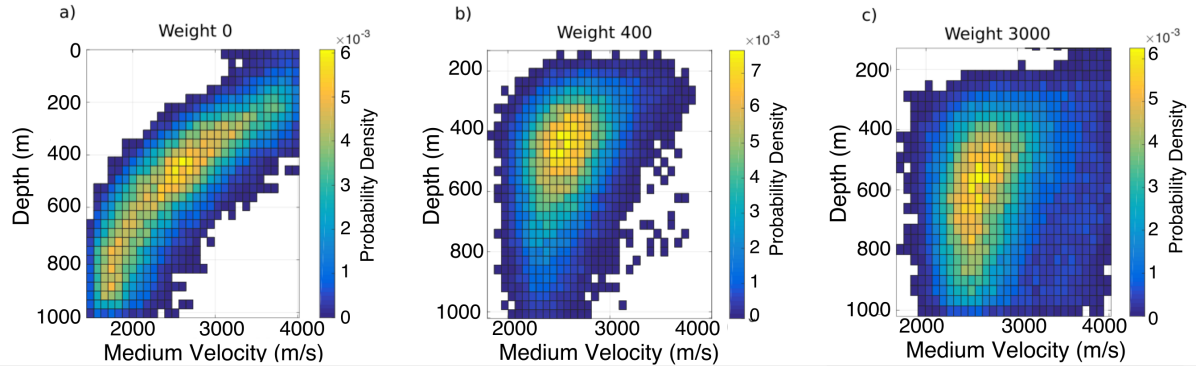


FIGURE 4.4 – a) 2D PDF of source depth and medium velocity without borehole data. b) 2D PDF of depth and medium velocity with borehole data weighted at 400. b) 2D PDF of depth and medium velocity with borehole data weighted at 3000. The model used for the inversion is the homogeneous model. The synthetic data has been computed through a homogeneous acoustic model. 4 parameters are inverted : x, y, z and v.

transform of the synthetic signals.

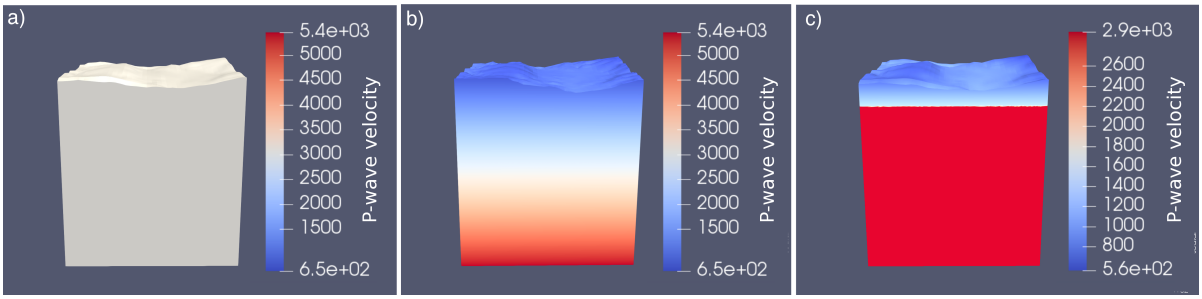


FIGURE 4.5 – Variation of P-wave velocity for the 3 elastic models used for the determination of the optimal weight. a) Homogeneous elastic model, b) 1D vertical gradient elastic model, c) 2-layer elastic model.

The results of the 3 inversions are displayed Figure 4.6. Once again, the smoothed PDF for depth is displayed. The results of the HOM3DV inversion on the synthetic data computed with the homogeneous elastic model show a good performance of the integration of borehole with a important sharpening of the PDF around the depth of the synthetic source for weight greater than 100. We observe an almost flat PDF in the case of no borehole or with a weight equal to 1.

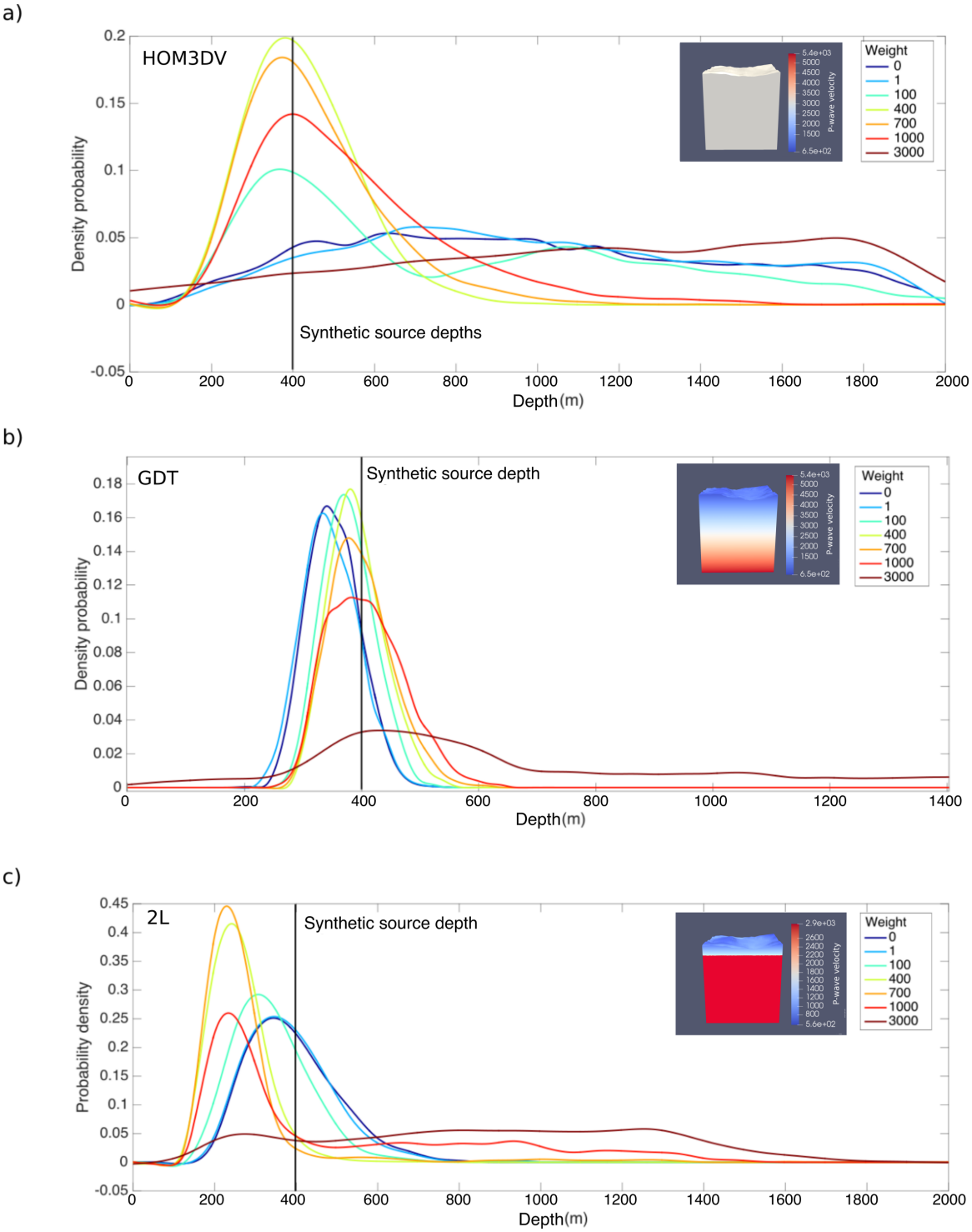


FIGURE 4.6 – a) Smoothed PDF of depth for several weights obtained with the HOM3DV inversion. The synthetic data was computed using the homogeneous elastic model. b) Smoothed PDF of depth for several weights obtained with the GDT inversion. The synthetic data was computed using the vertical gradient elastic model. c) Smoothed PDF of depth for several weights obtained with the 2L inversion. The synthetic data was computed using the 2-layer elastic model.

In the case of the GDT inversion, we see that the integration of weighted borehole data does significantly improve the resolution at depth, but has an effect on the accuracy of the localization. Indeed, the maximum of PDF of the inversion with no borehole data is about 50m above the synthetic source position. For the HOM3DV as well as GDT, the optimal weight is 400.

Figure 4.6 shows that for the 2L inversion, the use of the borehole is detrimental. The increase of the weight induces a shift in the PDF away from the true source position. Given the proximity between the layer boundary and the borehole, we suspect this is due to reflections on the boundary between the two layers that are not taken into account in our acoustic computation. As it is unlikely that such a sharp boundary exists in the true medium, we choose to include weighted borehole data in the 2L inversion during our test on recorded data. We also use a weight of 400.

We have presented two new acoustic gradient-based models that can be used to efficiently compute replica. We also investigated the use of borehole data to improve resolution at depth. The two gradient-based schemes and the integration of borehole data are combined and applied to the events studied in chapter 2 and compared to the HOM3DV inversion.

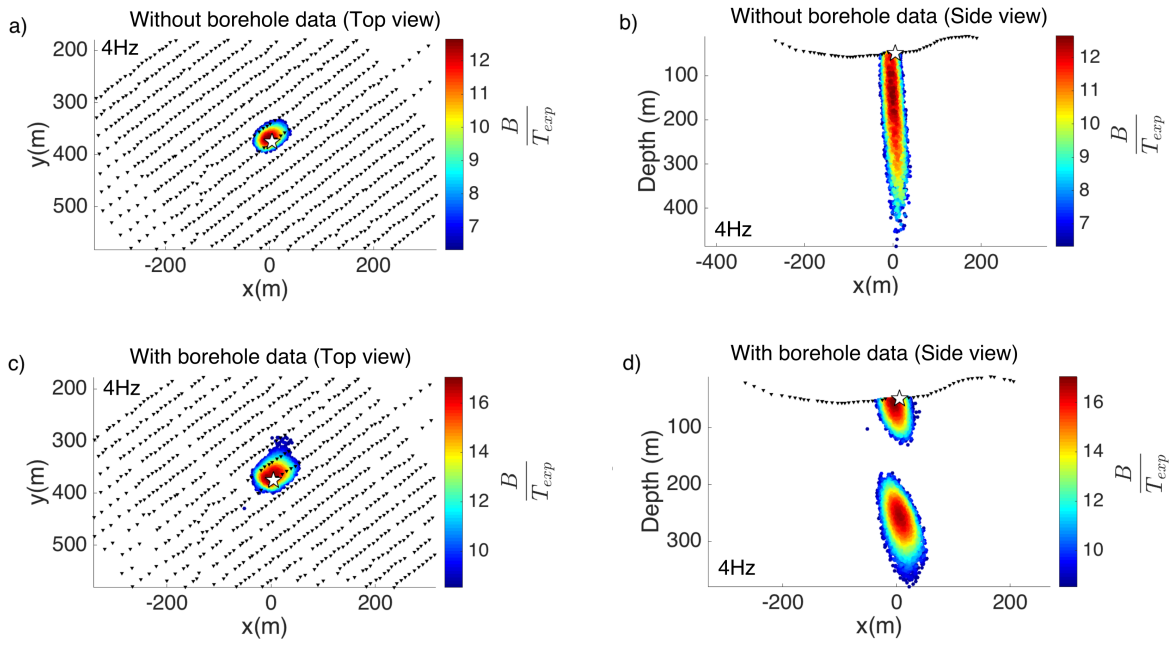


Figure 4.7 – a, b) MCMC output for a shot according to x, y and z for the HOM3DV inversion without weighted borehole data : a) 2D view of the epicentral position of the output. b) 2D view according to x and z. c, d) MCMC output for a shot according to x, y and z for the HOM3DV inversion with weighted borehole data : c) 2D view of the epicentral position of the output. d) 2D view according to x and z.

## 4.4 Investigation on sources of interest

### 4.4.1 Application to Betsy gun shot data

We first apply the three inversion scheme (HOM3DV,GDT and 2L) to shot data at 4Hz. For each of them, we compare outputs computed with and without weighted borehole data. The inversions with borehole data are always performed with a weight of 400.

**HOM3DV :** A top and a side view of the results of the HOM3DV inversion are displayed figure 4.7. Figure 4.7-a and -b and Figure 4.7-c and -d correspond to the inversion without and with borehole data, respectively. We see that for both inversions, the source is correctly localized. The use of weighted borehole data improves the resolution but results in the apparition of a second lobe in the output. Both outputs have a similar radius ($\sim$ 30m), and their size in the z-direction is 450m and 150m, respectively.

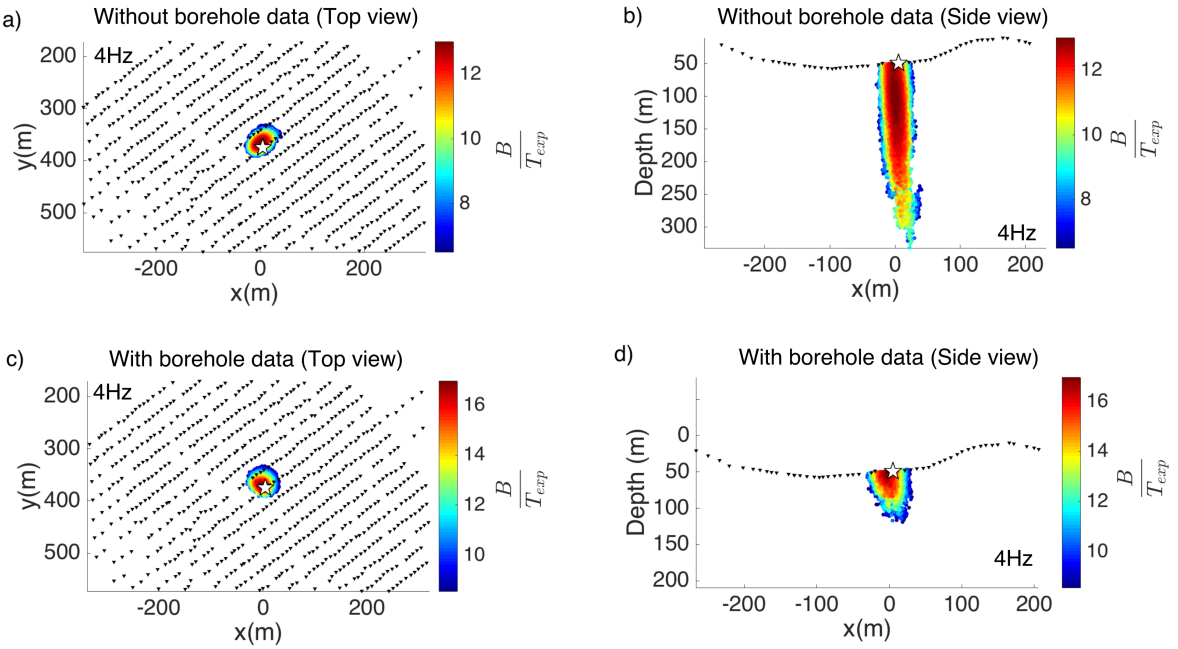


FIGURE 4.8 – a,b) MCMC output for a shot according to x, y and z for the GDT inversion without weighted borehole data : a) 2D view of the epicentral position of the output. b) 2D view according to x and z. c, d) MCMC output for a shot according to x, y and z for the GDT inversion with weighted borehole data : c) 2D view of the epicentral position of the output. d) 2D view according to x and z.

**GDT :** Figure 4.8 shows the 3D outputs corresponding to the GDT inversion scheme for the shot. We see that, even without borehole data, the resolution at depth is improved (Figure 4.8 -b). The length of the output along the z-direction is 270m and 70m for the GDT inversion without and with weighted borehole data respectively. The accuracy and resolution for the epicentral coordinates remain unchanged. The value of the output is greater for the GDT inversions than for the HOM3DV inversions, indicating that the wavefield recorded by the sensors is better matched by the vertical gradient model. The inverted gradient and surface velocity corresponding to the maximum value of the output are 0.571/s and 556m/s. Those values are consistent with surface waves propagating in the subsurface.

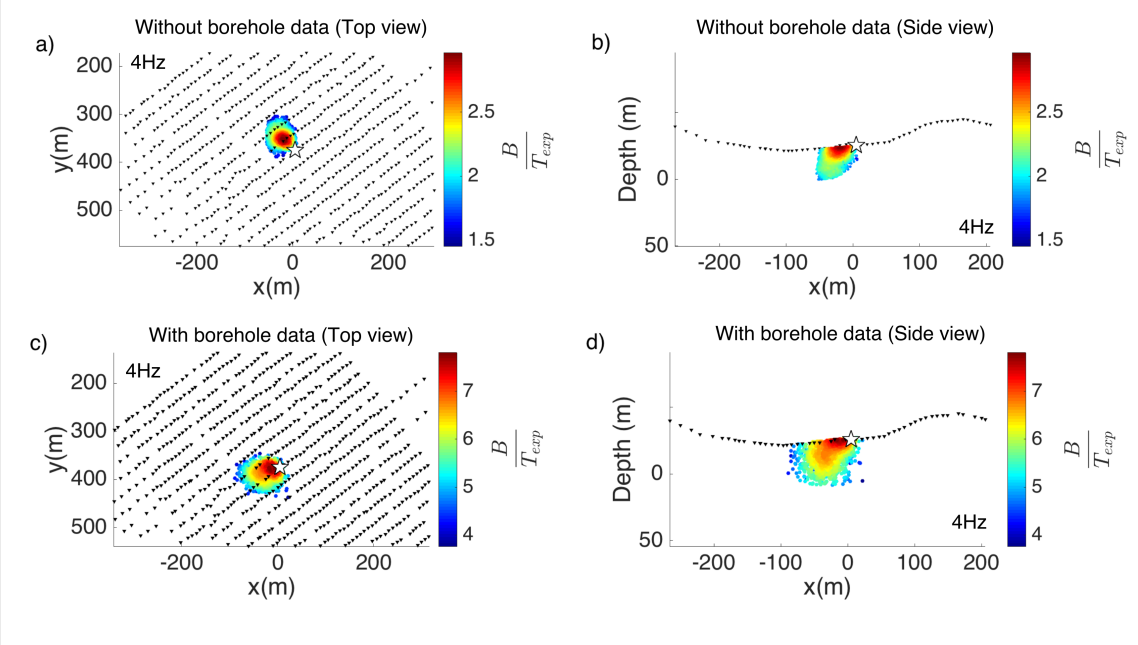


FIGURE 4.9 – a,b) MCMC output for a shot according to x, y and z for the 2L inversion without weighted borehole data : a) 2D view of the epicentral position of the output. b) 2D view according to x and z. c, d) MCMC output for a shot according to x, y and z for the 2L inversion with weighted borehole data : c) 2D view of the epicentral position of the output. d) 2D view according to x and z.

**2L :** No successful localization of the shot was made at 4Hz using the 2L inversion scheme. This is because the model used is an acoustic P-wave velocity model, which cannot match the phase patterns of surface waves that dominate our chosen time window. In order to perform a successful inversion we chose an earlier time-window were P-wave phase patterns are identifiable. In addition, the head waves propagating in the 2-layer model have velocities much higher than the 1200m/s upper velocity boundary for detection at 4Hz. Trying to locate sources at this frequency will yield poor results in

terms of resolution. In the same way, waves that propagate only in the first layer are to slow and will be aliased at 16Hz. We choose 10Hz as a frequency of study in this case. The result of the location of the shot using the time window containing P-waves is the output displayed Figure 4.9-a and b.

The results of the 2L inversion with weighted borehole data are shown figure 4.9-c and d. The resolution is improved compared to the HOM3DV inversion but is similar to the GDT case (70m). The use of weighted borehole data does not impact the resolution at depth, but the resolution along the epicentral coordinates is slightly degraded. However, the accuracy is improved. The white star indicating the shot position is included in the output when borehole data is included.

The inversion of the shot data using the three inversion scheme and the integration of borehole data shows that the use of gradient based model improves the resolution at depth. The effect of the addition of weighted borehole data is only beneficial with the GDT inversion scheme. The use of the 2L inversion for surface source is limited, as only P-waves will be detected. We continue our investigation with the potential deep source outside the array.

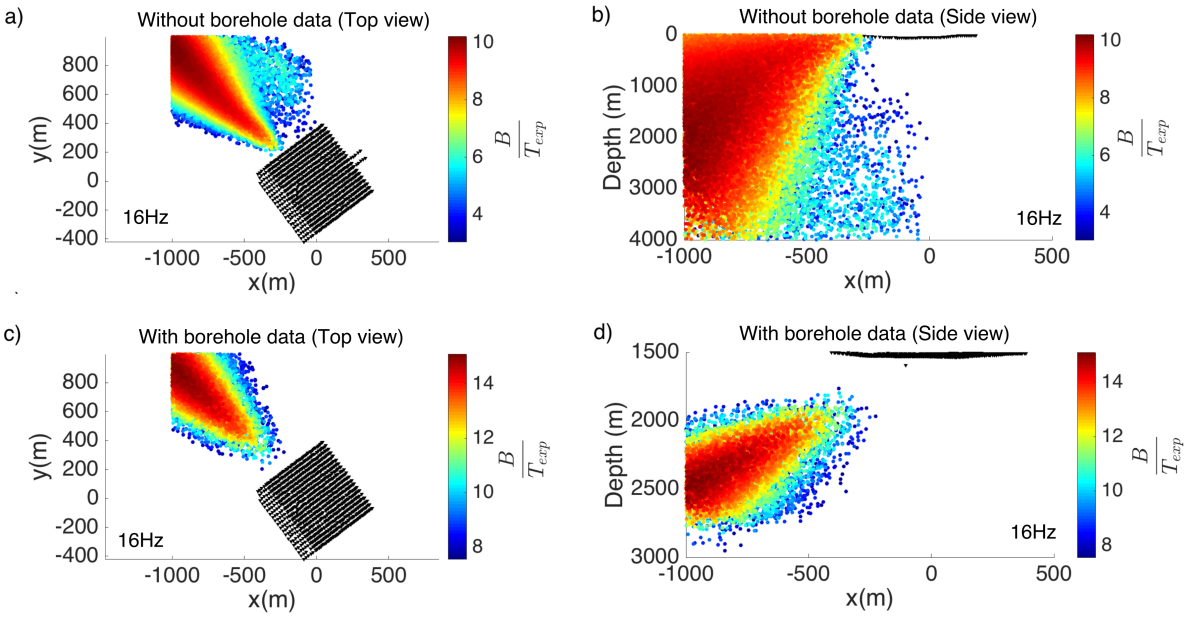


FIGURE 4.10 – a,b) MCMC output for a possible deep source according to x, y and z for the HOM3DV inversion without weighted borehole data : a) 2D view of the epicentral position of the output. b) 2D view according to x and z. c, d) MCMC output for a possible deep source according to x, y and z for the HOM3DV inversion with weighted borehole data : c) 2D view of the epicentral position of the output. d) 2D view according to x and z.

### 4.4.2 Application to a possible deep event outside the array

Events outside the array where successfully detected with a homogeneous 2D replica. The angular position of the event was also retrieved. The inversion of all three position parameters using of a more elaborate model could reduce the source location uncertainties highlighted in chapter 2. We applied the three inversion scheme on the window containing the deep event discussed in section 2.7. Once again all inversions are performed with and without weighted borehole data.

**HOM3DV :** A top and a side view of the results of the HOM3DV inversion are displayed figure 4.10. Figure 4.10-a and -b and Figure 4.10-c and d correspond to the inversion without and with borehole data, respectively. The integration of borehole data improves resolution. In both cases the resolution stays low along the z-direction (>4000m and 1500m, respectively), due to the strong link between inverted depth and velocity.

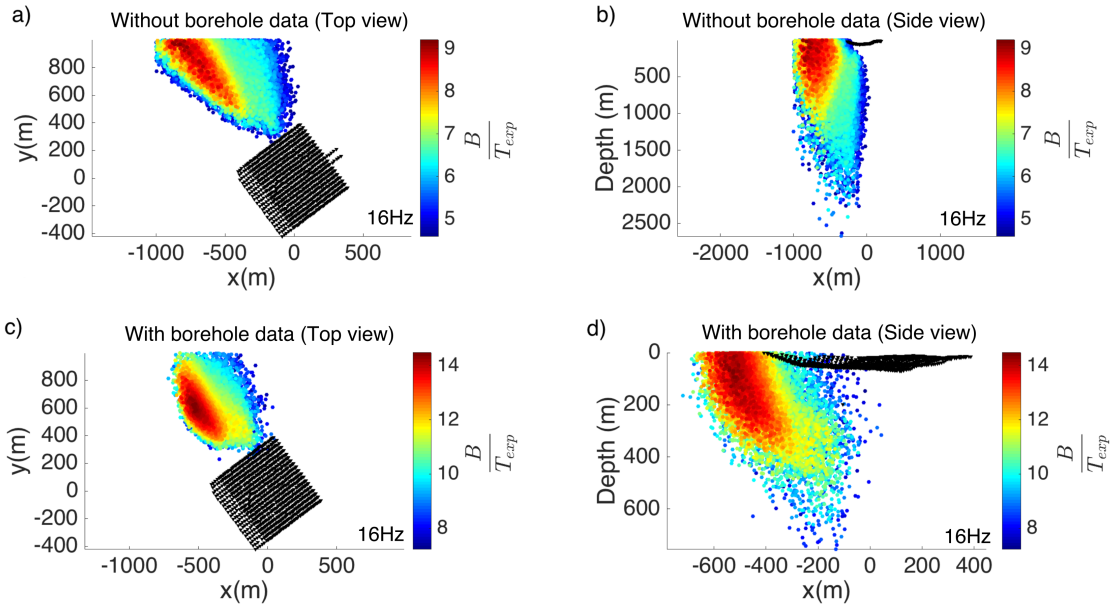


FIGURE 4.11 – a,b) MCMC output for a possible deep source according to x, y and z for the GDT inversion without weighted borehole data : a) 2D view of the epicentral position of the output. b) 2D view according to x and z. c, d) MCMC output for a possible deep source according to x, y and z for the GDT inversion with weighted borehole data : c) 2D view of the epicentral position of the output. d) 2D view according to x and z.

**GDT :** Figure 4.11 shows the 3D outputs corresponding to the GDT inversion scheme for the possible deep event. We see that, even without borehole data, the resolution at depth is improved ( Figure 4.11 -b). The resolution at depth is around 2000m for the

inversion without borehole data and 800m otherwise. The use of borehole data also has a positive impact on the epicentral resolution.

**2L :** The localization of the possible deep event performed using the 2L scheme resulted in the outputs displayed Figure 4.15. The resolution at depth is two times better than the resolution obtained with GDT. This is expected, as the velocity model is fixed, reducing the uncertainties. However, the value of the output from the GDT inversion is higher, indicating a better reconstruction of the measured wavefield. This is because the inversion of the medium properties allows for a better match with the heterogeneities present in the medium. As for the GDT inversion, the use of weighted borehole data improves depth resolution, an output length of 500m in the z-direction (Figure 4.15-c and -d). However the distortion of the shape of the output can does not give us confidence in the inversion results.

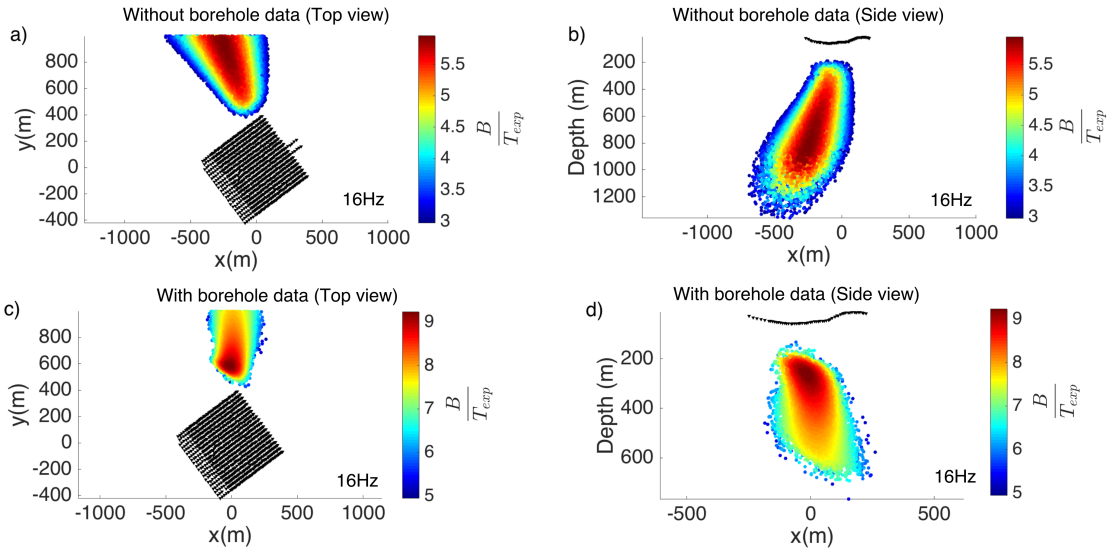


FIGURE 4.12 – a,b) MCMC output for a possible deep source according to x, y and z for the 2L inversion without weighted borehole data : a) 2D view of the epicentral position of the output. b) 2D view according to x and z. c, d) MCMC output for a possible deep source according to x, y and z for the 2L inversion with weighted borehole data : c) 2D view of the epicentral position of the output. d) 2D view according to x and z.

Once again, the use of the gradient based method has a positive effect on the resolution at depth. It should be noted that the output values are grater in the case of the HOM3DV inversion, which means that, in this case, gradient based inversions do not reconstruct as well the wavefield. This especially true for 2L. The use of borehole data further improves the resolution for all three coordinates of the source position. the

synthetic results showed that the resolution at depth should also improve for localization under the array.

### 4.4.3 Application to a localization under the array

We choose a localization from the HOM2DV inversion results presented chapter 3. This localization was chosen because visible patterns are present in the phase of the Fourier transform.

**HOM3DV :** The HOM3DV inversion without borehole data shows a maximum under the array (Figure 4.12-a and -b). The output is also smeared with high values outside the array. The poor resolution can be explained by the position of the maximum, which is very close to the array boundary. The output computed with weighted borehole data has two lobes, one of which is under the array at 300m depth (4.12-c and -d). However, the presence of two lobes hinders interpretation and leaves us unable to conclude on the quality of the localization.

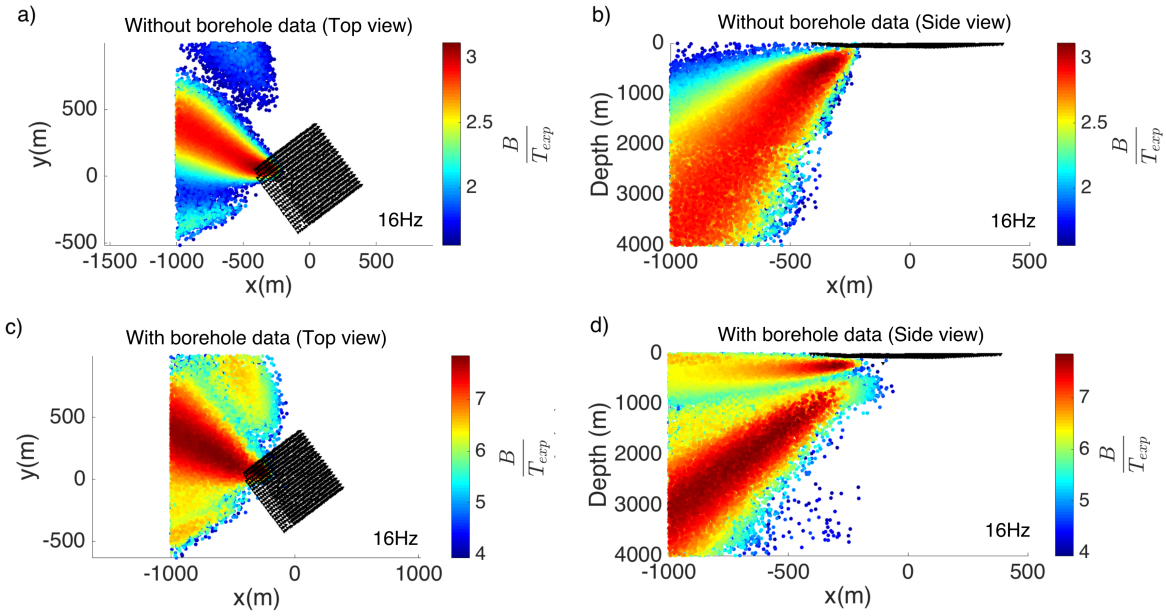


FIGURE 4.13 – a,b) MCMC output for a possible source under the array according to x, y and z for the HOM3DV inversion without weighted borehole data : a) 2D view of the epicentral position of the output. b) 2D view according to x and z. c, d) MCMC output for a possible source under the array according to x, y and z for the HOM3DV inversion with weighted borehole data : c) 2D view of the epicentral position of the output. d) 2D view according to x and z

**GDT :** The inversion performed using the GDT scheme without borehole data shows an improved resolution in all three dimensions of the source position (figure 4.13-a and -b). The inversion converges to a source at the boundary of the array at around 200m depth. The depth resolution is about 700m. The use of borehole data degrades the resolution, with a loss of dynamic in the output (figure 4.13-c and -d).

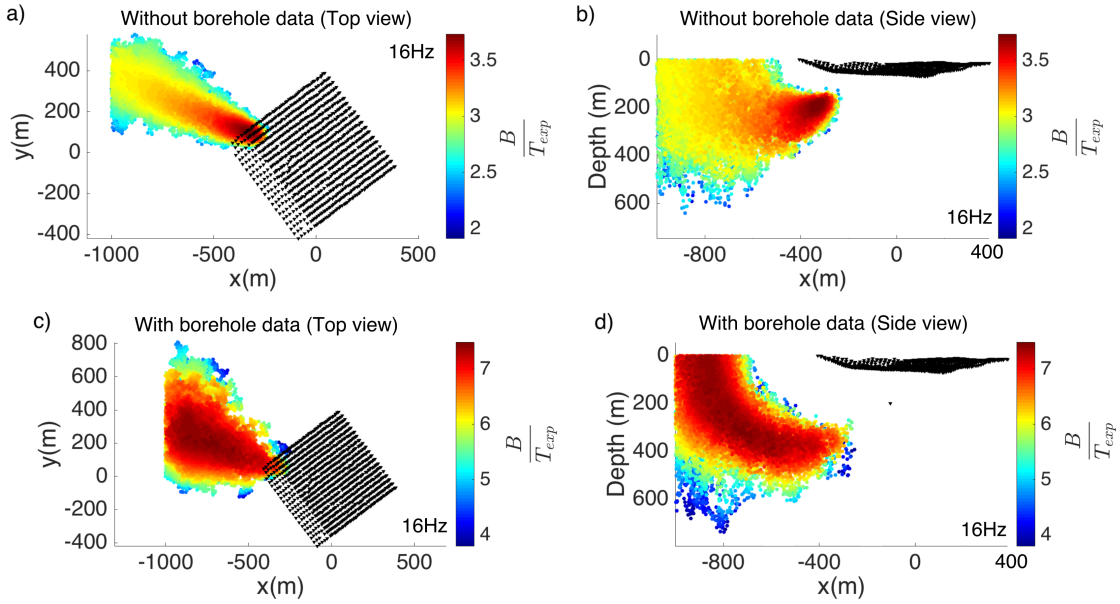


FIGURE 4.14 – a,b) MCMC output for a possible source under the array according to x, y and z for the GDT inversion without weighted borehole data : a) 2D view of the epicentral position of the output. b) 2D view according to x and z. c, d) MCMC output for a possible source under the array according to x, y and z for the GDT inversion with weighted borehole data : c) 2D view of the epicentral position of the output. d) 2D view according to x and z

**2L :** The 2L scheme does not perform well on this example, with output values just above the threshold for the inversion without borehole data. There is no convergence under the array 4.14-a and -b) . This could be the effect of heterogeneities that are not accommodated by the fixed velocity model of the 2L inversion. The distorted shape of the output obtained with the inclusion of borehole data leads us to conclude that the inversion with borehole data was also unsuccessful despite the high values of the output.

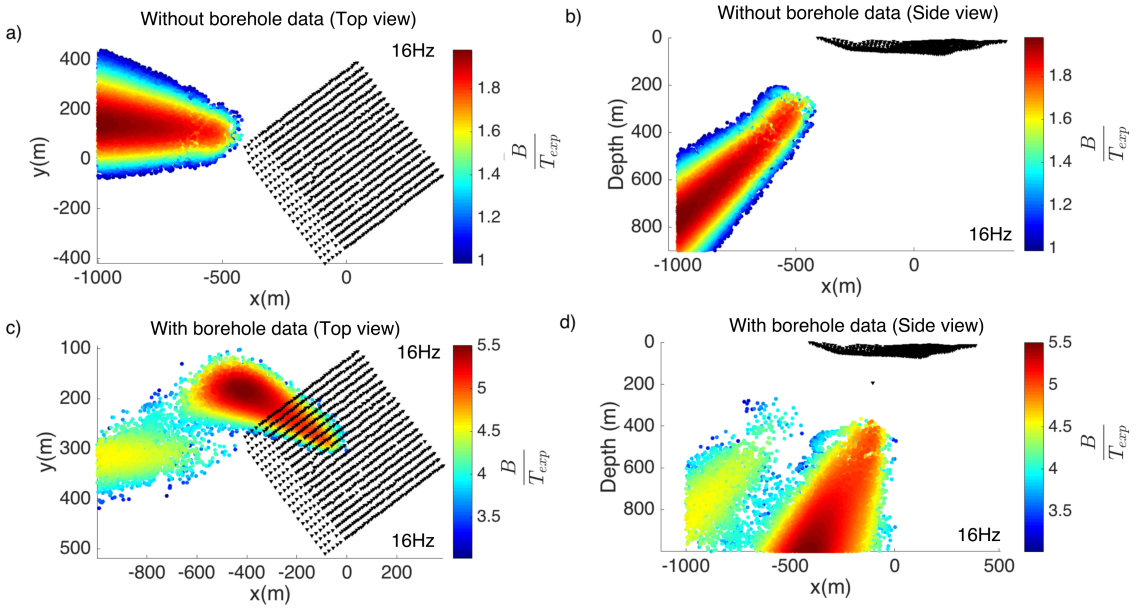


FIGURE 4.15 – a,b) MCMC output for a possible source under the array according to x, y and z for the 2L inversion without weighted borehole data : a) 2D view of the epicentral position of the output. b) 2D view according to x and z. c, d) MCMC output for a possible source under the array according to x, y and z for the 2L inversion with weighted borehole data : c) 2D view of the epicentral position of the output. d) 2D view according to x and z

## 4.5 Conclusion

We devised and tested 3 inversion schemes combined with the integration of weighted borehole data in order to determine the best approach to invert for source position at depth. For the HOM3DV inversion, the use of borehole data significantly improves the resolution, but in two case it created several lobes. The use of the GDT inversion with and without borehole data shows satisfying results, with a global improvement in resolution when weighted borehole data is integrated The shape of the output is also better defined. The 2L inversion is not adapted to surface source localization, and did not yield satisfying results in the case of the possible source under the array. The integration of borehole data to the 2L inversion scheme distorts the its outputs. From this we can conclude that the use of a gradient with borehole data is the strategy to employ in order to improve resolution at depth.

# Conclusions and future works

## Conclusions

The first focus of this work has been to implement a MFP scheme, where synthetics are matched to the data in the frequency domain, to perform source localization subsurface source on continuous dense array data. The end goal was to localize automatically low signal to noise ratio as-of-yet undetected events. As this requires systematic scanning of the 30-days dataset, a great part of this PhD work consists in finding an efficient MFP implementation. In chapter 2, we propose to use a homogenous model with a closed form solution at single frequencies that ensure quick computation of the synthetics. We use apparent velocity as a proxy for depth to invert only 3 parameters and further reduce the computational cost. Several optimization methods where investigated, and simulated annealing was chosen for its efficiency and reliability. We also implemented the process on a grid to perform parallel computation. Even though we do not invert for all 3 coordinates of the source position, the first tests on carefully selected events showed encouraging results, especially when comparing them to 4D (X,Y,Z,V) inversion results that have very low depth resolution.

We then process 26 days of the data set, at 4Hz and 16Hz to localize both surface and deeper sources. A great number of localization is obtained. Localizations are tagged as inside the array, outside the array or undetermined depending on the positions and shape of their associated outputs. We study the stacks of localizations inside and outside the array separately, for both 4Hz and 16Hz. From this we defined dominant sources of interest that are studied more in details. Among them two types of localizations for sources outside the array at 16Hz, with apparent velocities that could correspond to shallow event are investigated. However we were not able to draw decisive conclusions on the position of these sources at depth or at the surface from the available results. Arguments can be found for both surface and deeper sources. Events at low apparent

velocity are identified as helicopters, and localizations under the array seem to correspond to diffracted waves instead of sources at shallow depths. We demonstrate here that, while the use of apparent velocity is enough as a first approach to highlight sources of interest, it is sometimes not sufficient to draw conclusions on the nature of those sources. The use of 3D localization could help us better characterize unknown sources, through the estimation of their depth.

In order to improve resolution at depth, we investigate two strategies. The first strategy is the use of two acoustic gradient-based models to compute replica. We consider first an acoustic 1D vertical gradient for which a closed form is available. The use of this model requires the use of 5 inversion parameters ; the 3 coordinates of the source position, the velocity at the surface and the vertical gradient. We also investigate a simpler inversion where an acoustic 2-layer model for P-waves velocity, inverted by Meng et Ben-Zion (2018b), is implemented. The approximation of the ray parameter allows the derivation of a closed form solution that satisfies our efficiency requirement. In this case only 3 inversion parameters, the source coordinates, are necessary.

The second strategy to improve resolution is to include weighted borehole data in our dataset, to increase the 3D coverage of the array. We studied the impact of including the borehole data using 3 different inversion schemes (HOM3DV, GDT and 2L) that involves a homogeneous model, a 1D vertical gradient model and a 2-layer model, respectively. For each of these inversion scheme an optimal weight of the borehole data was derived from simple synthetic cases computed with elastic models. We found that a weight of 400 should be used in every case.

Inversions with the gradient-based models combined with the integration of borehole data, were performed on shot data and possible sources at depth inside and outside the array. The results showed that the use of the 2-layer gradient model with and without borehole data was not advantageous. Localization with a gradient model has better resolution than with the homogeneous model. In most cases, the addition of borehole data improved the resolution at depth in both HOM3DV and GDT inversion. In the first case, multiple lobes were present in the inverted outputs, complicating the interpretation. A 1D vertical gradient model proves to be the best solution to reduce the uncertainties along depth. The added value of the integration of borehole data is not as clear, as it leads in some cases to ambiguous results.

# Future works

## MFP implementation

**Process of 26 days with gradient model** The next step would be the implementation of a gradient model MFP on the CIGRI grid to process all the dataset and invert for 3D source positions. This could improve our conclusions on the sources of interest presented in chapter 2.

In addition, MFP could be performed at multiple frequencies in a coherent ways. This has not been done because of computational cost, but recent improvement on the code implemented on the CIGRI would allow such computations at reasonable costs.

In the same way, another lead would be the use of an elastic numerical model. While it has been rejected because of its computational cost, the implementation of the MFP process on the CIGRI grid for analytical models has shown a drastic reduction of computational time that could make the use of pre-stored numerically computed replica possible. Such replica could be computed by using waveform modelization codes such as SEM46 (which is an update of SEM3D), allowing us to work in the near-field of the source and to take into account the different source mechanisms. Preliminary study must be performed to assess whether the use of such expensive models could be beneficial.

Very few localizations were found under the array, and none of them could be categorically classified as an event. The values of the outputs for localizations inside the array at 16Hz are also generally lower than localizations outside of it. One of the reasons for this could be the fact that we do not take into account the source mechanism in our inversion. A double couple creates phase opposition in the signal that would degrade the value of the MCMC-output. A solution for this is to invert for the radiation pattern as well as for the source position. This can be done "blindly" by flipping the phase at each sensor in order to maximize the output value. Josefine Umlauft successfully performed this inversion scheme on synthetic data. This method could be applied during a scan of the data.

The results of the 2D source position and apparent velocity inversion showed that the sources localized under the array correspond most likely to heterogeneities diffracting the wavefield. Our MFP approach, and more precisely the computation of our replica with the assumption of a point source, only allows us to take into account diffraction on heterogeneities. The seismic exploration community works on approaches, such as the

Common Reflection Surface, that consider travel times of reflections as well as diffractions. Investigating how to use such approach with our method could be beneficial, as it could yield new insights on the heterogeneities present inside and around the array.

## Results interpretation

The scan of the data using the 2D source position and apparent velocity inversion has resulted in many localizations. In chapter 3, we focused on examples of localization at 16Hz chosen from the dominating localizations highlighted by the overview. A more thorough study could be of interest, with a categorization and characterization of all the localizations at both frequency. This would require the use of automatic and advanced classification tools and the input of other sources of data, such as weather and temperature variations and ground motion measurements.

Few efforts were dedicated to the localized surface sources. Focusing more on them would be beneficial. Firstly, characterizing precisely the different surface sources inside and outside the array could help removing their contribution from the data of interest. Secondly, recent studies on tremors have shown that surface sources such as trains could be erroneously considered as deeper sources (Inbal *et al.* (2018)). Providing a database of known surface sources against which candidate sources of interest could be tested, would be a tool against this issue. Lastly, surface sources cover the majority of the array and could be used for imaging. A simple, direct application of our results could be the study of the variations of the inverted apparent velocity for all the sources in the array, which could allow the imaging of the structure of the medium under the array.

# Bibliographie


Aki, K. et Richards, P. G. (2002). *Quantitative seismology*.

Artman, B., Podladtchikov, I. et Witten, B. (2010). Source location using time-reverse imaging. *Geophysical Prospecting*, 58(5):861–873.

Baer, R. N. et Collins, M. D. (2006). Source localization in the presence of gross sediment uncertainties. *The Journal of the Acoustical Society of America*, 120(2):870–874.

Baggeroer, A., Kuperman, W. et Mikhalevsky, P. (1993). An overview of matched field methods in ocean acoustics. *IEEE Journal of Oceanic Engineering*, 18(4):401–424.

Barros, T., Ferrari, R., Krummenauer, R. et Lopes, R. (2015). Differential evolution-based optimization procedure for automatic estimation of the common-reflection surface traveltime parameters. *Geophysics*, 80(6):WD189–WD200.

Ben-Zion, Y., Vernon, F. L., Ozakin, Y., Zigone, D., Ross, Z. E., Meng, H., White, M., Reyes, J., Hollis, D. et Barklage, M. (2015). Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone. *Geophysical Journal International*, 202(1):370–380.

Billings, S. (1994). Simulated Annealing for Earthquake Location. *Geophysical Journal International*, (September):680–692.

Bruneau, M. et Potel, C. (2010). *Materials and Acoustics Handbook*.

Bucker, H. P. (1976). Use of calculated sound fields and matched-field detection to locate sound sources in shallow water. *The Journal of the Acoustical Society of America*, 59(2):368–373.

Buland, R. (1976). The mechanics of locating earthquakes. *Bulletin of the Seismological Society of America*, 66(1):173–187.

Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418.

Červený, V. (2001). *Seismic Ray Theory*. Cambridge University Press, Cambridge.

Chmiel, M., Roux, P. et Bardainne, T. (2016). Extraction of phase and group velocities from ambient surface noise in a patch-array configuration. *GEOPHYSICS*, 81(6):KS231–KS240.

Corciulo, M., Roux, P., Campillo, M., Dubucq, D. et Kuperman, W. A. (2012). Multiscale matched-field processing for noise-source localization in exploration geophysics. *GEOPHYSICS*, 77(5):KS33–KS41.

Cros, E. (2011). *Etude de la dynamique du Geyser Old Faithful, USA, à partir de méthodes de sismique passive*. Thèse de doctorat.

Dosso, S. E., Dettmer, J. et Wilmut, M. J. (2015). Efficient localization and spectral estimation of an unknown number of ocean acoustic sources using a graphics processing unit. *The Journal of the Acoustical Society of America*, 138(5):2945–2956.

Drew, J., White, R. S., Tilmann, F. et Tarasewicz, J. (2013). Coalescence microseismic mapping. *Geophysical Journal International*, 195(3):1773–1785.

Ekström, G. (2006). Global detection and location of seismic sources by using surface waves. *Bulletin of the Seismological Society of America*, 96(4 A):1201–1212.

Fomel, S., Luo, S. et Zhao, H. (2009). Fast sweeping method for the factored eikonal equation. *Journal of Computational Physics*, 228(17):6440–6455.

Garabito, G., Stoffa, P. L., Lucena, L. S. et Cruz, J. C. (2012). Part I - CRS stack : Global optimization of the 2D CRS-attributes. *Journal of Applied Geophysics*, 85:92–101.

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

Gerstoft, P. et Tanimoto, T. (2007). A year of microseisms in southern California. *Geophysical Research Letters*, 34(20):1–6.

Gharti, H. N., Oye, V., Roth, M. et Kühn, D. (2010). Automated microearthquake location using envelope stacking and robust global optimization. *Geophysics*, 75(4): MA27–MA46.

Gibbons, S. J. et Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, 165(1): 149–166.

Gill, P. E., Murray, W. et Wright, M. H. (1981). *Practical optimization*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London.

Grigoli, F., Cesca, S., Krieger, L., Kriegerowski, M., Gammaldi, S., Horalek, J., Priolo, E. et Dahm, T. (2016). Automated microseismic event location using Master-Event Waveform Stacking. *Scientific Reports*, 6(1):25744.

Gutenberg, B. et Richter, C. F. (1944). Frequency of earthquakes in California*. *Bulletin of the Seismological Society of America*, 34(4):185–188.

Harris, D. B. et Kvaerna, T. (2010). Superresolution with seismic arrays using empirical matched field processing. *Geophysical Journal International*, 182(3):1455–1477.

Hauksson, E. (2011). Crustal geophysics and seismicity in southern California. *Geophysical Journal International*, 186(1):82–98.

Hauksson, E., Yang, W. et Shearer, P. M. (2012). Waveform relocated earthquake catalog for Southern California (1981 to June 2011). *Bulletin of the Seismological Society of America*, 102(5):2239–2244.

Hillers, G., Roux, P., Campillo, M. et Ben-Zion, Y. (2016). Focal spot imaging based on zero lag cross-correlation amplitude fields : Application to dense array data at the San Jacinto fault zone. *Journal of Geophysical Research : Solid Earth*, 121(11): 8048–8067.

Holland, J. H. (1975). Adaptation in Natural and Artificial Systems : An introductory Analysis with Applications to Biology, Control and Artificial Intelligence. *MIT Press*, page 183.

Inbal, A., Ampuero, J. P. et Clayton, R. W. (2016). Localized seismic deformation in the upper mantle revealed by dense seismic arrays. *Science*, 354(6308):88–92.

Inbal, A., Cristea-Platon, T., Ampuero, J., Hillers, G., Agnew, D. et Hough, S. E. (2018). Sources of Long-Range Anthropogenic Noise in Southern California and Implications for Tectonic Tremor Detection. *Bulletin of the Seismological Society of America*, pages 1–20.

Ingber, L. (1989). Very fast simulated re-annealing. *Mathematical and Computer Modelling*, 12(8):967–973.

Ishii, M., Shearer, P. M., Houston, H. et Vidale, J. E. (2005). Extent, duration and speed of the 2004 Sumatra-Andaman earthquake imaged by the Hi-Net array. *Nature*, 435(7044):933–936.

Ito, A. (1985). High resolution relative hypocenters of similar earthquakes by cross-spectral analysis method. *Journal of Physics of the Earth*, 33(4):279–294.

Johnson, C. W., Meng, H., Vernon, F. L., Nakata, N. et Ben-Zion, Y. (2018). Characteristics of ground motion generated by interaction of wind gusts with trees, structures and other obstacles above the surface. *In Abstract to the annual meeting of the Southern California Earthquake Center*.

Kao, H. et Shan, S.-J. (2004). The Source-Scanning Algorithm : mapping the distribution of seismic sources in time and space. *Geophysical Journal International*, 157(2):589–594.

Kirkpatrick, S., Gelatt, C. et Vecchi, M. (1983). Optimization by Simulated Annealing. *Wire*, 220(4598):671–680.

Kremers, S., Fichtner, A., Brietzke, G. B., Igel, H., Larmat, C., Huang, L. et Käser, M. (2011). Exploring the potentials and limitations of the time-reversal imaging of finite seismic sources. *Solid Earth*, 2(1):95–105.

Kuperman, W. A. et Turek, G. (1997). Matched field acoustics. *Journal of Engineering and Applied Science*, 11(1):141–148.

Lagarias, J. C., Reeds, J. A., Wright, M. H. et Wright, P. E. (1998). Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions. *SIAM Journal on Optimization*, 9(1):112–147.

Landès, M., Hubans, F., Shapiro, N. M., Paul, A. et Campillo, M. (2010). Origin of deep ocean microseisms by using teleseismic body waves. *Journal of Geophysical Research*, 115(B5):B05302.

Langet, N., Maggi, A., Michelini, A. et Brenguier, F. (2014). Continuous Kurtosis-Based Migration for Seismic Event Detection and Location, with Application to Piton de la Fournaise Volcano, La Reunion. *Bulletin of the Seismological Society of America*, 104(1):229–246.

Larmat, C., Montagner, J.-P., Fink, M., Capdeville, Y., Tourin, A. et Clévédé, E. (2006). Time-reversal imaging of seismic sources and application to the great Sumatra earthquake. *Geophysical Research Letters*, 33(19):L19312.

Larmat, C., Tromp, J., Liu, Q. et Montagner, J.-P. (2008). Time reversal location of glacial earthquakes. *Journal of Geophysical Research*, 113(B9):B09314.

Li, Z., Peng, Z., Hollis, D., Zhu, L. et McClellan, J. (2018). High-resolution seismic event detection using local similarity for Large-N arrays. *Scientific Reports*, 8(1):1646.

Lin, F.-C., Li, D., Clayton, R. W. et Hollis, D. (2013). High-resolution 3D shallow crustal structure in Long Beach, California : Application of ambient noise tomography on a dense seismic array. *GEOPHYSICS*, 78(4):Q45–Q56.

Lin, G. et Shearer, P. (2005). Tests of relative earthquake location techniques using synthetic data. *Journal of Geophysical Research : Solid Earth*, 110(B4):1–14.

Lin, G., Shearer, P. M. et Hauksson, E. (2007). Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate southern California seismicity from 1981 to 2005. *Journal of Geophysical Research : Solid Earth*, 112(12):1–14.

Lomax, A., Virieux, J., Volant, P. et Berge-Thierry, C. (2000). *Probabilistic Earthquake Location in 3D and Layered Models*, pages 101–134. Springer Netherlands, Dordrecht.

Lomax, A., Zollo, A., Capuano, P. et Virieux, J. (2001). Precise, absolute earthquake location under Somma-Vesuvius volcano using a new three-dimensional velocity model. *Geophysical Journal International*, 146(2):313–331.

Meng, H. et Ben-Zion, Y. (2018a). Characteristics of Airplanes and Helicopters Recorded by a Dense Seismic Array Near Anza California. *Journal of Geophysical Research : Solid Earth*, In Press.

Meng, H. et Ben-Zion, Y. (2018b). Detection of small earthquakes with dense array data : example from the San Jacinto fault zone, southern California. *Geophysical Journal International*, 212(1):442–457.

Meng, X., Peng, Z. et Hardebeck, J. L. (2013). Seismicity around Parkfield correlates with static shear stress changes following the 2003 Mw6.5 San Simeon earthquake. *Journal of Geophysical Research : Solid Earth*, 118(7):3576–3591.

Mordret, A., Roux, P. et Ben-zion, Y. (2018). Shallow 3-D structure of the San Jacinto Fault zone revealed from ambient noise imaging with a dense seismic array. *Geophysical Journal International*, (Submitted).

Moreau, L., Hunter, A., Velichko, A. et Wilcox, P. (2014). 3-D reconstruction of sub-wavelength scatterers from the measurement of scattered fields in elastic waveguides. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 61(11):1864–1879.

Neale, J., Harmon, N. et Srokosz, M. (2017). Monitoring remote ocean waves using P-wave microseisms. *Journal of Geophysical Research : Oceans*, 122(1):470–483.

Nelder, J. A. et Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313.

Poiata, N., Satriano, C., Vilotte, J. P., Bernard, P. et Obara, K. (2016). Multiband array detection and location of seismic sources recorded by dense seismic networks. *Geophysical Journal International*, 205(3):1548–1573.

Press, W. H., Flannery, B. P., Teukolsky, S. A. et Vetterling, W. T. (2007). *Numerical Recipes : The Art of Scientific Computing*. Cambridge University Press.

Qin, L., Ben-Zion, Y., Qiu, H., Share, P.-E., Ross, Z. E. et Vernon, F. L. (2018). Internal structure of the San Jacinto fault zone in the trifurcation area southeast of Anza, California, from data of dense seismic arrays. *Geophysical Journal International*, 213(1):98–114.

Retailleau, L., Landès, M., Gualtieri, L., Shapiro, N. M., Campillo, M., Roux, P. et Guilbert, J. (2018). Detection and analysis of a transient energy burst with beamforming of multiple teleseismic phases. *Geophysical Journal International*, 212(1): 14–24.

Riahi, N. et Gerstoft, P. (2015). The seismic traffic footprint : Tracking trains, aircraft, and cars seismically. *Geophysical Research Letters*, 42(8):2674–2681.

Riahi, N. et Gerstoft, P. (2016). Locating sources in a dense array through network-based clustering. *In 2016 Information Theory and Applications Workshop (ITA)*, pages 1–8. IEEE.

Richard-Dinger, K. M. et Shearer, P. M. (2000). Earthquake locations in southern California using source-specific station terms. *Journal of Geophysical Research*, 105(B5):10939–10960.

Ross, Z. E., Hauksson, E. et Ben-Zion, Y. (2017). Abundant off-fault seismicity and orthogonal structures in the San Jacinto fault zone. *Science Advances*, 3(3):1–8.

Rost, S. et Thomas, C. (2002). Array seismology : Methods and applications. *Reviews of Geophysics*, 40(3):1008.

Roux, P., Moreau, L., Lecointre, A., Hillers, G., Campillo, M., Ben-Zion, Y., Zigone, D. et Vernon, F. (2016). A methodological approach towards high-resolution surface wave imaging of the San Jacinto Fault Zone using ambient-noise recordings at a spatially dense array. *Geophysical Journal International*, 206(2):980–992.

Ružek, B. et Kvasnička, M. (2001). Differential Evolution Algorithm in the Earthquake Hypocenter Location. *Pure and Applied Geophysics*, 158(4):667–693.

Sambridge, B. Y. M. et Gallagher, K. (1993). Earthquake hypocenter location using genetic algorithms. *Bulletin of the Seismological Society of America*, 83(5):1467–1491.

Sambridge, M. et Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, 40(3):1009.

Satriano, C., Kiraly, E., Bernard, P. et Vilotte, J. P. (2012). The 2012 Mw 8.6 Sumatra earthquake : Evidence of westward sequential seismic ruptures associated to the reactivation of a N-S ocean fabric. *Geophysical Research Letters*, 39(15):1–6.

Sen, M. K. et Stoffa, P. L. (2013). *Global Optimization Methods in Geophysical Inversion*. Cambridge University Press, Cambridge.

Shearer, P. M. (1994). Global seismic event detection using a matched filter on long-period seismograms. *Journal of Geophysical Research*, 99(94):13713–13725.

Shearer, P. M. (1997). Improving local earthquake locations using the L1 norm and waveform cross correlation : Application to the Whittier Narrows, California, aftershock sequence. *Journal of Geophysical Research : Solid Earth*, 102(B4):8269–8283.

Shelly, D. R., Beroza, G. C. et Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, 446(7133):305–307.

Storn, R. et Price, K. (1997). Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359.

Sun, L., Zhang, M. et Wen, L. (2016). A new method for high-resolution event relocation and application to the aftershocks of Lushan Earthquake, China. *Journal of Geophysical Research : Solid Earth*, 121(4):2539–2559.

Vandemeulebrouck, J., Roux, P. et Cros, E. (2013). The plumbing of Old Faithful Geyser revealed by hydrothermal tremor. *Geophysical Research Letters*, 40(10):1989–1993.

Waldhauser, F. et Ellsworth, W. L. (2000). A Double-difference Earthquake location algorithm : Method and application to the Northern Hayward Fault, California. *Bulletin of the Seismological Society of America*, 90(6):1353–1368.

Wang, J., Templeton, D. C. et Harris, D. B. (2015). Discovering new events beyond the catalogue-Application of empirical matched field processing to Salton Sea geothermal field seismicity. *Geophysical Journal International*, 203(1):22–32.

Young, R. P., Maxwell, S. C., Urbancic, T. I. et Feignier, B. (1992). Mining-induced microseismicity : Monitoring and applications of imaging and source mechanism techniques. *Pure and Applied Geophysics PAGEOPH*, 139(3-4):697–719.

Zhang, H. et Thurber, C. H. (2003). Double-difference tomography : The method and its application to the Hayward Fault, California. *Bulletin of the Seismological Society of America*, 93(5):1875–1889.

Zollo, A., Matteis, R. D., Capuano, P., Ferulano, F. et Iannaccone, G. (1995). Constraints on the shallow crustal model of the Northern Apennines (Italy) from the analysis of microearthquake seismic records. *Geophysical Journal International*, 120(3):646–662.

# Appendix A

# Appendix : travel time computation for 2 layer medium

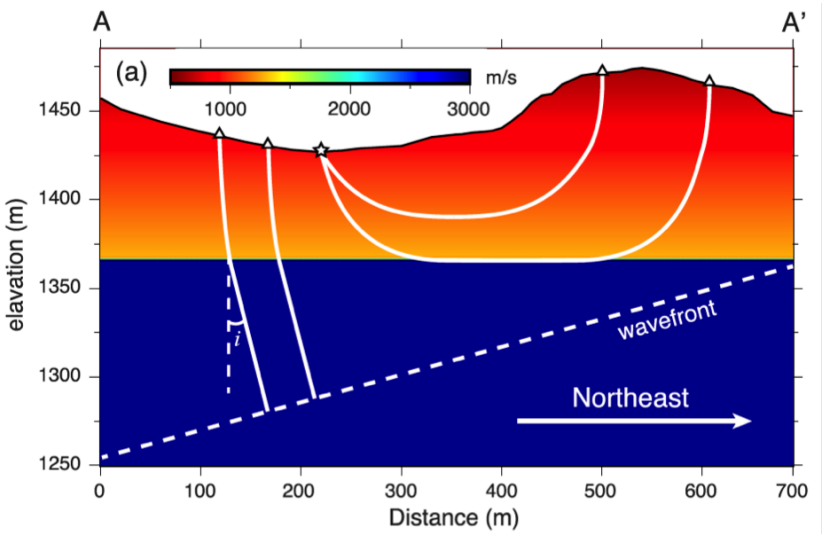


FIGURE A.1 – A derived P 1D velocity model along a cross-section. The star and triangles represent a Betsy gunshot and geophones. From Meng et Ben-Zion (2018b)

We used the ray tracing method and model presented in Meng et Ben-Zion (2018b) (Figure A.1) as a basis to compute our travel time in the entire model. Meng et Ben-Zion (2018b) only computed travel times for sources at the surface, as it was sufficient for their inversion.

However, in the case of MFP, we need to be able to compute travel times for sources in the homogeneous layer. For the sake of computational efficiency, we are looking for an

analytical solution for these travel times.

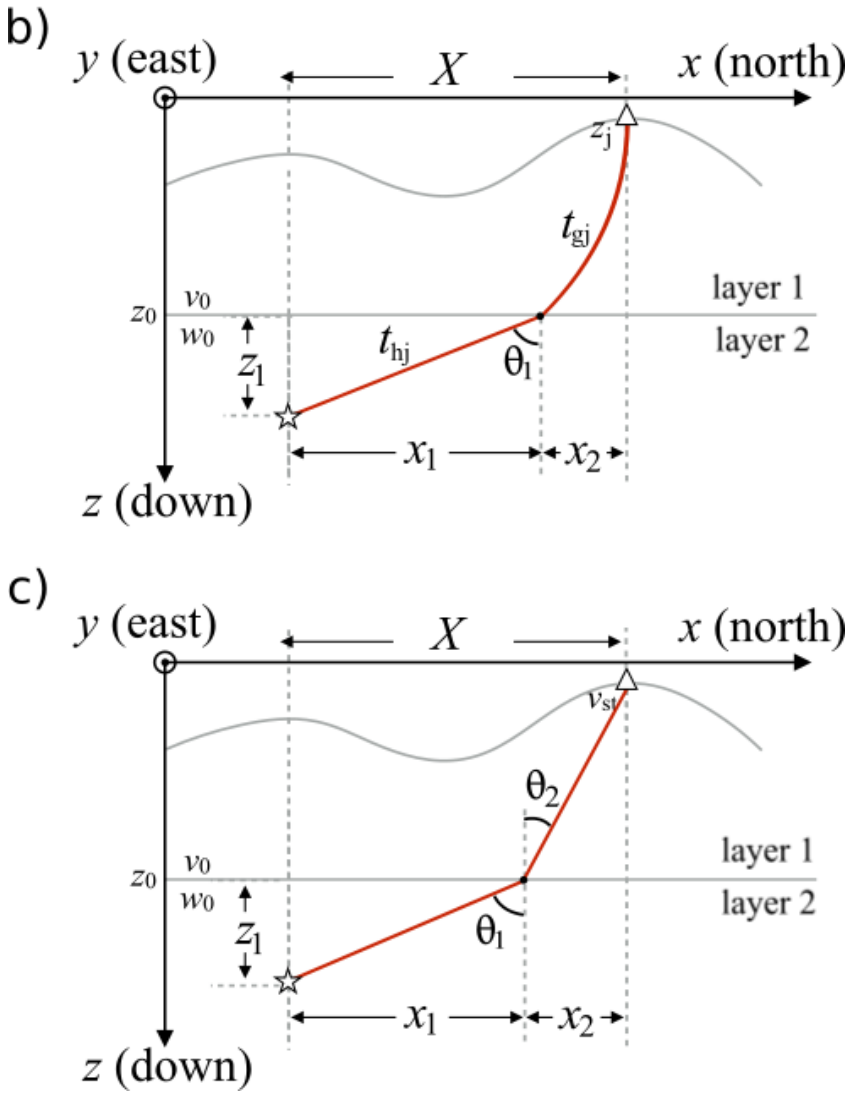


FIGURE A.2 – a) Ray path from a source in the $2^{nd}$ layer (star) to a receiver at the surface (triangle). b) Ray path from a source in the $2^{nd}$ layer (star) to a receiver at the surface (triangle) for the 2 homogeneous layer approximation

In the case of the 1D gradient layer, the ray parameter in the homogeneous 1D gradient media is (equation 1.16) :

$$p_g = \frac{2Gx_2}{\sqrt{[(Gx_2)^2 + {v_2}^2 + v(\mathbf{x})^2]^2 - (2v_2 v(\mathbf{x}))^2}} \tag{A.1}$$

Were G is the gradient, $p_g$ is the ray parameter, $v_2$ is the velocity in the 1D gradient

layer at the interface and $v(\mathbf{x})$ is the velocity at the receiver, as shown figure A.2.

The ray parameter in a homogeneous media is :

$$p_h = \frac{sin(\theta_1)}{v_1} \tag{A.2}$$

where $\theta_1$ is the incidence angle of the ray, $v_1$ is the velocity in the medium. $sin(\theta_1)$ can be expressed as a function of the horizontal distance $x_1$ between the source and the position of the ray at the interface and $z_1$ the vertical distance between the source and the interface :

$$sin(\theta_1) = \frac{x_1}{\sqrt{x_1^2 + z_1^2}}. \tag{A.3}$$

As :

$$p_h = p_g, \tag{A.4}$$

we have :

$$\frac{2Gx_2}{\sqrt{[(Gx_2)^2 + {v_2}^2 + v(\mathbf{x})^2]^2 - (2v_2 v(\mathbf{x}))^2}} = \frac{x_1}{v_1\sqrt{x_1^2 + z_1^2}}. \tag{A.5}$$

If we use $x_2 = X - x_1$, with $X$ the horizontal distance between the source and receiver, this equation can be transformed into a 6th order polynomial, but cannot be solved analytically. In order to simplify the problem, we consider two layer of homogeneous media of velocity $v_1 = w_0$ and $v_2 = v_0$.

$$x_i = tan(\theta_i) z_i \quad i = \{1, 2\} \tag{A.6}$$

if we take into account eq. A.2, $x_1$ and $x_2$ can be expressed as :

$$x_i = \frac{z_i p v_i}{\sqrt{1 - (pv_i)^2}}. \quad i = \{1, 2\} \tag{A.7}$$

we can say $X = x_1 + x_2$, or

$$X = \frac{z_1 p v_1}{\sqrt{1 - (pv_1)^2}} + \frac{z_2 p v_2}{\sqrt{1 - (pv_2)^2}} \tag{A.8}$$

This equation can be rewritten as a 4th order polynomial expression of the form :

$$ap^4 + bp^2 + c = 0 \tag{A.9}$$

with

$$a = (X^2 + z_1^2 + z_2^2)v_0^2 w_0^2 \tag{A.10}$$

$$b = v_0^2(X^2 + z_1^2) - w_0^2(X^2 + z_2^2) \tag{A.11}$$

$$c = X^2 \tag{A.12}$$

for which we can find analytical solutions. Two solutions are physically correct :

$$p_1 = \sqrt{\frac{-b - \sqrt{b^2 - 4ac}}{2a}} \tag{A.13}$$

and

$$p_2 = \sqrt{\frac{-b + \sqrt{b^2 - 4ac}}{2a}} \tag{A.14}$$

Comparing the two solutions to the numerically solved solution of the 6th order polynomial, we find that $p_1$ matches the results. We will call $\tilde{p} = p_1$ the approximation of p. Once the ray parameter is estimated, the values of $x_1$ and $x_2$ can be computed and from that, we can derive the travel times for the two layers and the total travel time.

We compared the travel times $t_{\tilde{p}}$ obtained from $\tilde{p}$ to the true travel times t computed numerically for all station positions and for and for a set of source positions contained in the $\sim$ 1200m x 1200m x 500m area centered on the array. For each source position we have a set of 1108 values corresponding to $t - t_{\tilde{p}}$ at each sensors. For clearer representation, Figure A.3 shows only the maximum of the absolute difference $t - t_{\tilde{p}}$ at each source position.

As the difficulties for computing travel time only arise when the source is in the $2^{nd}$ layer, we only show the maximum difference between $t_{\tilde{p}}$ and t for depth greater than 118m. The maximum difference between the travel times is located close to the boundary, where the assumption of two homogeneous layers is the weakest. This maximum difference is lower that $1.10^{-3}$ which is below the sampling of the data in our study. Figure A.3 represents the phase difference computed from the travel time difference at 16Hz. We see that in this case the phase difference is small enough to be overlooked, even close to the boundary. Using this method for sources in the homogeneous layer and

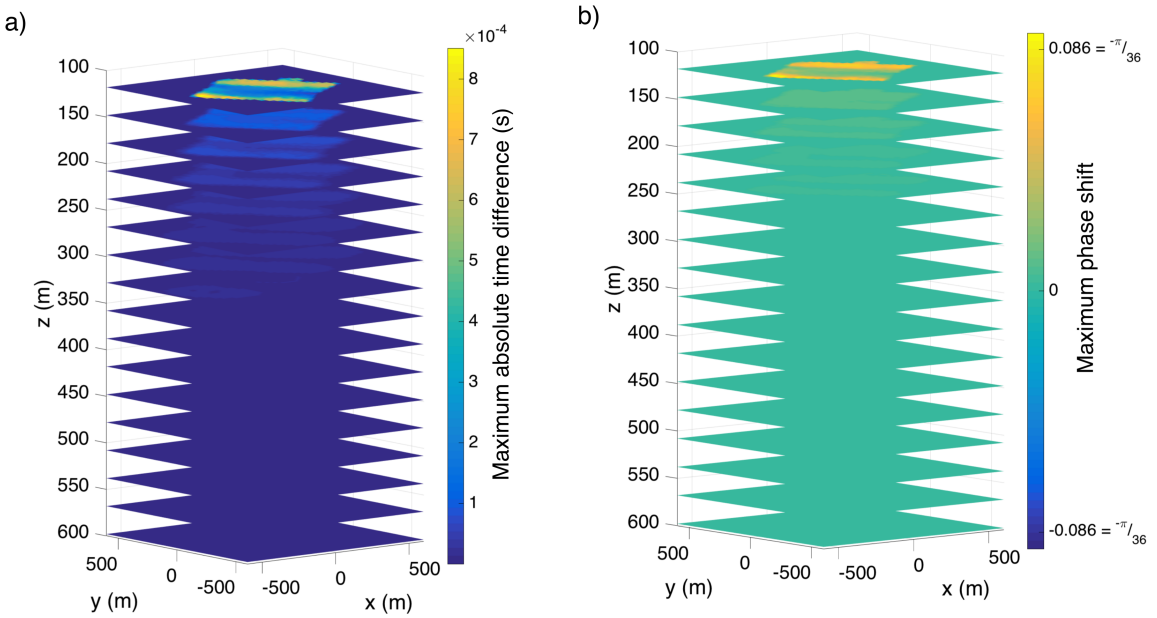


FIGURE A.3 – a) Absolute time difference between true travel times computed numerically and our approximations. b) Phase difference obtained from the time differences in a)

the method described in Meng et Ben-Zion (2018b) and in chapter 4 for the source in the $2^{nd}$ layer, we are able to estimate travel times for the complete range of sources of interest.

We can compare this approximation with a simplified problem where we determine the coordinates of the boundary crossing using a straight ray assumption. Figure A.4 shows the maximum absolute travel time difference when using this simplification.

In this case the travel time difference is more important at depth. Those differences are above 0.05s for most of the volume of interest, resulting in phase differences of the order of $\pi$. We conclude that, while the 2 homogeneous layer approximation is satisfying in terms of travel time computation, the straight path approximation cannot be used as an approximation for the 2-layer model.

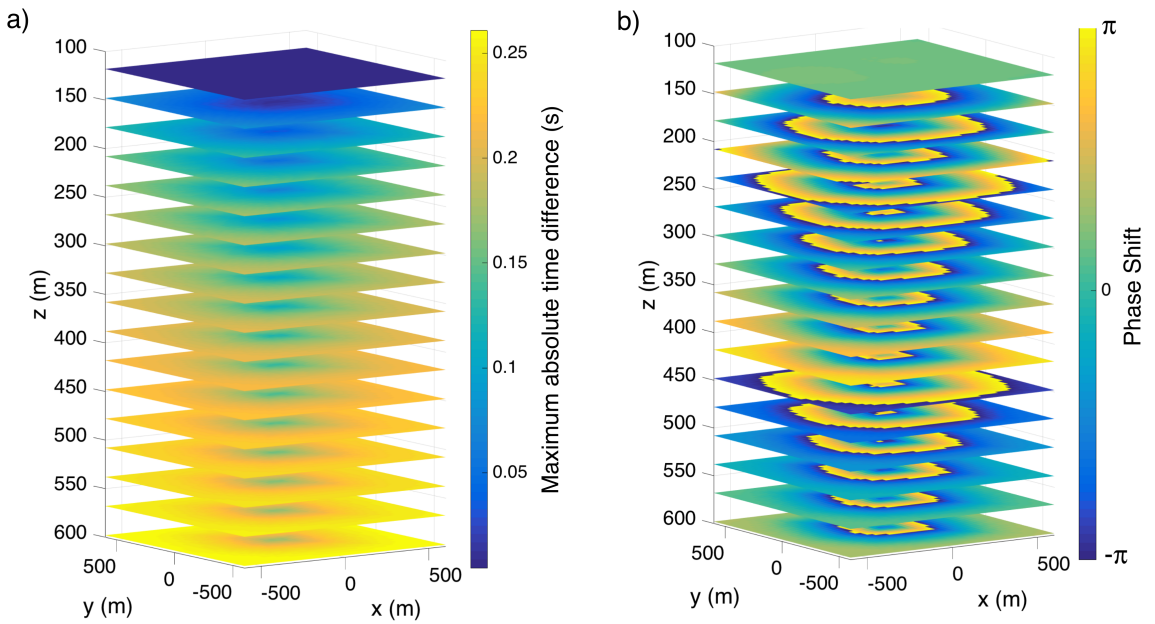


FIGURE A.4 – a)Absolute time differences between true travel time computed numerically and straight ray approximation. b) Phase difference obtain from the time difference in a)