Passive Deep Ocean Thermometry using Hydroacoustic Signals from Submarine Earthquakes

by

Ludovic Morandini

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended on Friday August 21, 2020 at 3 PM.

Student number:	4912306	
Profect duration:	March 2, 2020 – August 7, 2020	
Thesis committee:	Dr. S. Shani-Kadmiel,	TU Delft, Supervisor
	Prof. Dr. L.G. Evers,	TU Delft, Senior Supervisor
	Prof. Dr. Ir. R. Ghose,	TU Delft, Examiner
	Prof. Dr. F. Wagner,	RWTH Aachen, Examiner

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Abstract

Measuring temperature variations of the deep ocean is necessary to evaluate the heat flux between the atmosphere and the hydrosphere and thus calibrate climate change models. In the last two decades a passive alternative to profiling oceanographic floats has emerged: hydroacoustic thermometry. This method consists of using the oceanic ambient noise field as a source of acoustic waves and hydrophone arrays as receivers. These sensors are part of the International Monitoring System which is in place for the verification of the Comprehensive Nuclear-Test-Ban Treaty. They are positioned at a water depth of approximately 1 km, in the Sound Fixing and Ranging channel. This channel is a low-velocity layer that functions as an acoustic wave guide, thus facilitating very long range propagation with little attenuation. This study analyses transient signals between 2005 and 2018 triggered by submarine earthquakes and detected at station H10, situated near Ascension Island in the South Atlantic Ocean. This station consists of two three-element (triplet) arrays with an aperture of approximately 2 km each. The triplets are 126 km apart. Array processing techniques applied to individual triplets are prone to be biased by local conditions of the array surroundings. We demonstrate that this bias is largely suppressed when jointly processing both triplets as one six-elements array. Due to the malfunction of element S1, data quality decreased after October 2013 and our results are less robust. For the 2005–2013 period, we retrieve a temperature of 4.3 to 4.8 °C.

Contents

List of Figures

List of Tables

1

Introduction

Oceans play a major role in the regulation of global climate. The hydrosphere covers a vast area of the Earth surface and exchanges large heat flux with the atmosphere. Anthropogenic activity raises the level of greenhouse gases in the atmosphere, which warms it up. During the second half of the twentieth century, sea-surface temperature rapidly increased. However, the increase of sea-surface temperature is currently fairly slow as heat is being transported to deeper layers of the ocean, and such deeper heat-sequestration period historically lasted 20 to 35 years (Chen & Tung, 2014). The change in temperature in the oceans affects the water density and specific volume, which in turn affect oceanic phenomena such as the global thermohaline circulation and the sea level. It is therefore important to monitor the temperature in the oceans. Furthermore, it is a crucial step towards quantifying air-sea heat exchanges in order to calibrate climate change models.

The surface temperature of oceans is monitored using satellites. To monitor temperature at depth, free-drifting profiling oceanographic floats are used. However, their efficiency is very limited because the data acquisition process is both very localised and sparse along space and time; it is technologically and economically unfeasible to survey deep ocean temperature over large scales and extended periods of time using this technique. It has been shown that the speed of sound c in the ocean is related to the temperature T, salinity S, and depth z. Medwin (1975) puts in place equation 1.1, also retaken by Medwin and Clay (1997):

$$\begin{aligned} c = 1449.2 + 4.6T - 0.055T^2 + 0.00029T^3 \\ +(1.34 - 0.01T)(S - 35) + 0.016z. \end{aligned} \tag{1.1}$$

By measuring the travel time of an acoustic wave over a known propagation distance in the ocean, either between a source and a receiver or between

two receivers, one can calculate the propagation velocity[1] and thus infer the temperature. This method is known as "hydroacoustic thermometry". There are numerous natural sources emitting sound waves in the oceans. For instance submarine earthquakes, underwater volcanoes, iceberg breaking, whales and others large mammals are some of them. On top of these, there are sources due to human activities: ships, seismic surveys, man-made explosions, etc. All these have an acoustic signature in the ocean.

1.1 The International Monitoring System

In this study, we use data from the hydroacoustic component of the International Monitoring System (IMS). The IMS is a global network of sensors comprising four technologies: seismology, hydroacoustic, infrasound, and radionuclide. It can be seen in figure 1.1. It is in-place for the detection of nuclear explosions for the verification of the Comprehensive Nuclear-Test-Ban Treaty (CTBT). The hydroacoustic component of the IMS is made up of eleven stations submerged in the oceans. These hydroacoustic stations are in fact positioned at a depth of 800-1200 m, in the Sound Fixing and Ranging (SOFAR) channel (see section 2.1). They record the continuous hydroacoustic pressure field at a sampling rate of 250 samples per second.

This study uses signals recorded at H10, a hydroacoustic array located near the British Overseas Territory of Ascension Island. This small volcanic island is situated along the Mid Atlantic Ridge (MAR), in the South Atlantic Ocean (see figure 1.2).

H10 consists of two three-sensor arrays (triplets), situated to the north (H10N) and to the south (H10S) of Ascension Island. The instruments are anchored to the seafloor by a mooring system and suspended in the SOFAR channel at a depth of approximately 850 m by floats. Each triplet has an array aperture of roughly 2 km. In this study, we process each triplet individually as well as all six elements together as one joint-array (see table 1.1).

This array has been in place since the early 2000s and is still active in 2020. This study uses signals from the 24^{th} of March 2005 to the 10^{th} of March 2018. As of October 2013, array element H10S1 (element 1 of the South triplet) has stopped recording data due to a malfunction. This issue seems to have an effect on the other two elements H10S2 and H10S3, and remains unresolved to this day.

1.2 Previous works

Numerous studies have used various processing techniques to approximate the propagation velocity of sound in the ocean and approximate the temperature

[1] *"The sound speed in water does not depend on the direction of the ray. Therefore, the names "sound speed" and "sound velocity" can be used interchangeably"* (Medwin & Clay, 1997).

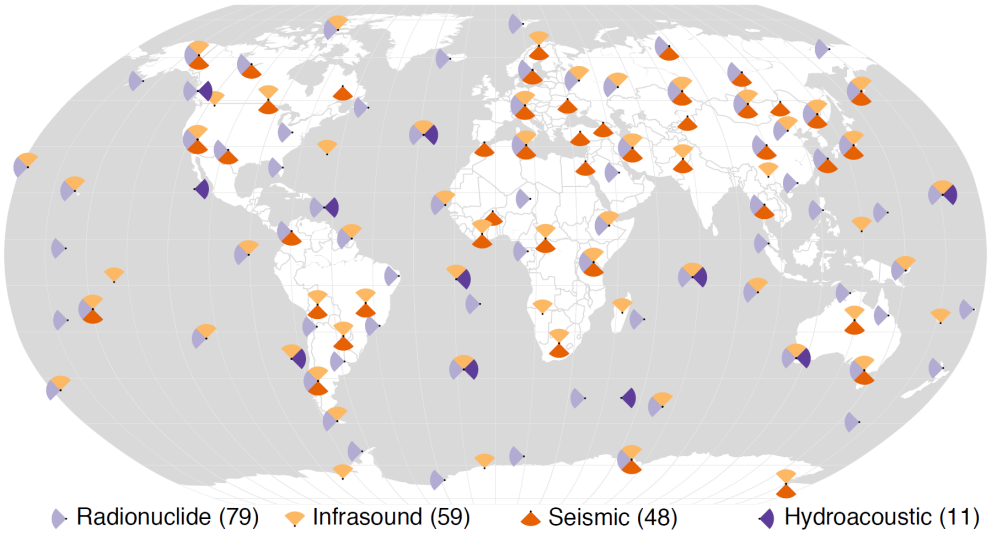

Figure 1.1: *The sensors of the International Monitoring System (IMS). From Shani-Kadmiel (2019).*

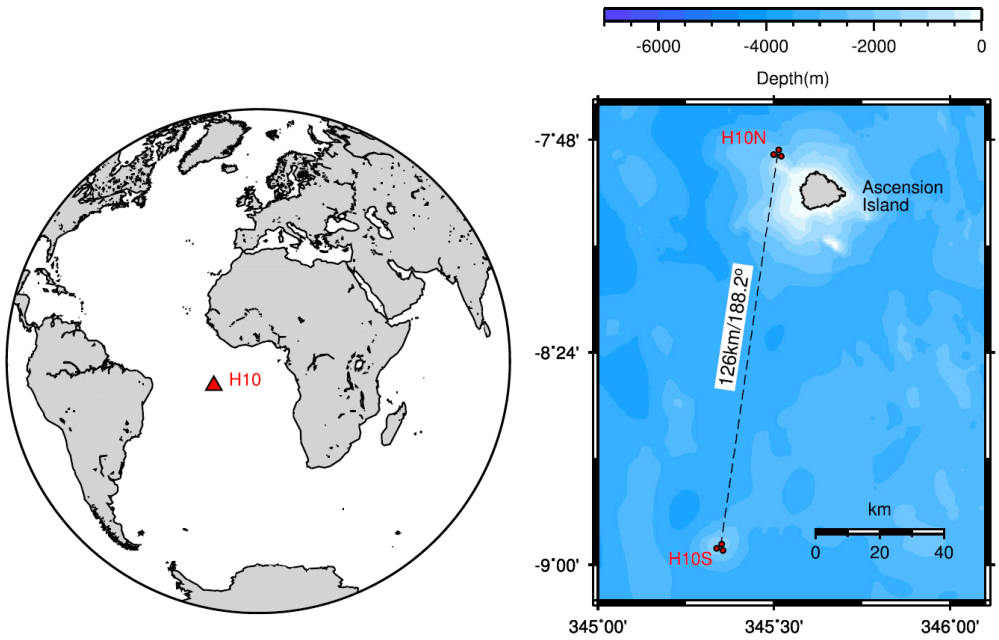

Figure 1.2: *Location and configuration of H10. (left) Location overview. (right) Configuration of H10 around Ascension Island. H10N and H10S are the North and South triplets of H10, separated by a horizontal distance of 126 km. The great circle path connecting H10N and H10S has a bearing of 188°. From Evers et al. (2017).*

Table 1.1: *Geographical and geometrical configuration of H10. Including the North array, the South array and the joint array.*

Array element	Longitude (°)	Latitude (°)	Depth (m)	Aperture (km)
N1	-14.48023	-7.84567	847	2.423
N2	-14.48748	-7.82779	845	
N3	-14.50168	-7.84093	850	
North centre	-14.48985	-7.83813	847	
S1	-14.64843	-8.94118	865	2.058
S2	-14.64531	-8.95915	852	
S3	-14.66295	-8.95274	863	
South centre	-14.65221	-8.95101	860	
Joint centre	-14.57101	-8.39457	854	126.3

based on hydroacoustics. Woolfe et al. (2015) monitored temperature variations using hydroacoustic thermometry over a period of several years. Their method consisted of applying cross-correlation to interarray receiver pairs, between array elements of the North and the South triplets, at Ascension Island and Wake Island stations. They studied the ambient noise field, with polar ice being a main contributor. This resulted in nine cross-correlation functions that were then combined (i.e. beamformed) into a single waveform to continuously track fine variations. The authors found that over the years, the variations in positive lag times (during northwards propagation) were similar to the variations in negative lag times (during southwards propagation). They interpret this as being due to *"reciprocal changes in the environment, such as ocean sound speed fluctuations induced by temperature changes, rather than nonreciprocal changes, such as currents, clock drift, or other signal-processing artifacts"* (Woolfe et al., 2015). Additionally, they derived temperature variations from the lag time fluctuations and found results to be in good agreement with Argo oceanographic floats measurements.

Evers and Snellen (2015) used hydroacoustic detections from earthquakes, listed in the Reviewed Event Bulletin (REB), which is provided by the CTBTO. They evaluated the celerity[2] to infer the temperature. They observed latitudinal celerity variations of the range 1 469 to 1 487 m/s, that the authors attributed to lower temperature in the South Atlantic Ocean than to the North of Ascension Island. It was also noted that any uncertainty in origin time, location, or picking of the phase arrivals, will result in faulty celerity values.

Ball et al. (2016) looked at the possibility of investigating crustal structure using hydrophones instead of ocean bottom seismometers. They cross-correlated the ambient noise field for elements N1 and S1 (at H10) and found two dominant modes: one close to the sound velocity of the SOFAR channel and one around 3 000 m/s. By modelling, the authors show that *"the fundamental mode exhibits strong dependence on the p-wave velocity and thickness of the water layer [i.e.*

[2]Defined as the epicentral distance between the source and receiver divided by the travel time.

the bathymetry], while the higher mode is sensitive to velocity structure at greater relative depths in the crust and upper mantle" (Ball et al., 2016).

Evers et al. (2017) studied the ambient noise field at Ascension Island. The authors used cross-correlation between interarray receiver pairs and showed that the average cross-correlation has a peak at a lag time of 85.5 s which corresponds to a horizontal velocity of 1 477 m/s. Moreover, *"in an attempt to increase the Signal-to-Noise Ratio (SNR) of the image by illuminating it with more acoustic daylight"* (Evers et al., 2017), they beamformed each triplet, and cross-correlated the best beams (more details on beamforming in section 2.2). Firstly, they noticed a difference between the apparent velocity[3] of each triplet retrieved in the beamforming approach and the one retrieved by standard cross-correlation. They interpreted this difference as being due to a certain spatial variability of the sound speed since the velocity retrieved in the beamforming approach is an estimate of the local sound velocity (triplet scale) whereas cross-correlation averages the velocity along the distance separating the two triplets. Secondly, results showed a higher SNR when cross-correlating array beams instead of sensor pairs. This is because the beamforming process suppresses much of the uncorrelated noise. The SNR can be redefined as the ratio between coherent arrivals (signals) and incoherent arrivals (noise).

In a most recent study, Sambell et al. (2019) cross-correlated element pairs of the H03S triplet, the southern part of H03, situated around Robinson Crusoe Island (off the coast of Chile). They studied the ambient noise field and removed signals linked with transient events with one-bit normalisation. In their cross-correlations, they found a signal related to noise from the Antarctic ice field incoming from South-Southwest (Antarctica). The temperature was then estimated from the lag time of the cross-correlations, assuming a homogeneous medium and a maximal lag time propagation (along the receiver pair axis). This analysis resulted in a speed of sound of 1 510 m/s which the authors concluded to be too high to be physically reasonable. Additionally, they applied an array-processing technique to beamform the signals and found two dominant back-azimuths[4]: Antarctica as well and an underwater volcano. The apparent velocity retrieved in the beamforming approach corresponded with the one previously retrieved with cross-correlation (in the back-azimuth range coming from Antarctica). Moreover, the authors found that the apparent velocity changes as a function of back-azimuth. This dependency should not be as the apparent velocity is averaged along the wavefront path, and thus the resulting apparent velocity should remain equivalent regardless on the direction the wavefront crosses the array. The authors interpreted this dependency as a consequence of the surroundings of H03S, most likely related to steep slopes in the local bathymetry. They hypothesised that using an array with a larger aperture would reduce the detrimental effect of the surroundings. Comparing the two processing techniques, it is concluded that beamforming is superior to cross-correlations for the following reasons: firstly it could detect signal coming from

[3]The apparent velocity corresponds to the horizontal component of the real velocity.

[4]The back-azimuth is the bearing of the great circle path from the receiver to the source, measured in degrees clockwise from North.

a narrow back-azimuth range (the volcano) and thus appeared more sensitive to transient events; secondly, wavefront parameters such as back-azimuth and apparent velocity can be extracted, which provide more information than just the lag time.

1.3 Research questions

In this study, we apply an array-processing technique to transient signals generated by submarine earthquakes. Sambell et al. (2019) strongly suggested that beamforming can resolve transient events more effectively than cross-correlation, therefore the use of beamforming to analyse earthquake signals seem more adapted. Moreover, our approach focuses on array-processing to suppress the uncertainties with regard to the earthquake's origin time and location as pointed out by Evers and Snellen (2015). Additionally, in order to minimise local effects, for example caused by the bathymetry, we jointly process all six elements. Hence the choice of H10, which is the station with the largest aperture. Therefore, our goals are:

- To examine if beamforming both array triplets together as one joint array can suppress the bias in apparent velocity due to local effects.
- Can temperature variations be detected from the apparent velocity extracted in the beamforming process?

2

Theoretical background

2.1 The SOFAR channel

In the air, sound waves are much more attenuated than electromagnetic waves (such as light). *"In the sea the reverse is true [...] [and] sound waves can be detected over vast distances and are a much better vehicle for undersea information than light"* (Talley et al., 2011). This is the reason why the IMS can monitor all oceans using only eleven hydroacoustic stations (see for example Heaney et al. (2017) for some basin-scale signals). Sound is a mechanical wave as it propagates via the succession of compression and extension of the elements of matter it goes through. The speed of a mechanical wave depends on the compressibility of the medium; a less compressible water means a higher speed of propagation of a wave. This is due to a quicker transfer of energy as less material deformation and molecule displacement occur. In sea water, two parameters strongly affect the compressibility: the pressure and the temperature (the impact of the salinity is often negligible as we will see in section 2.3). The behaviour of these three parameters is shown in figure 2.1, for a water column in temperate latitudes. There is a trade-off between the effects pressure and temperature have on the speed of sound as a function of depth. In contrast to the pressure, which linearly increases with water depth, the temperature rapidly decreases over approximately the first 500 m and then gradually decreases further. The pressure is strongly correlated with the depth. A high pressure yields a less compressible water due to the fluid being more rigid because its molecules are pushed together. Therefore, a water under high pressure intuitively means a high sound speed. Regarding the temperature, a warm water is less compressible due to the molecules having more kinetic energy and being less willing to be pushed together than cold water. Therefore, a warm water means a high sound speed. The trade-off between the effects of pressure and temperature leads to the formation of a low velocity layer between the surface and the seafloor. This layer

is situated at the depth where *"the decrease in sound speed due to cooler water is overcome by [the] increase in sound speed due to higher pressure"* (Talley et al., 2011). The Sound Fixing and Ranging (SOFAR) channel is the name given to this zone, and the channel axis is centred on the minimum.

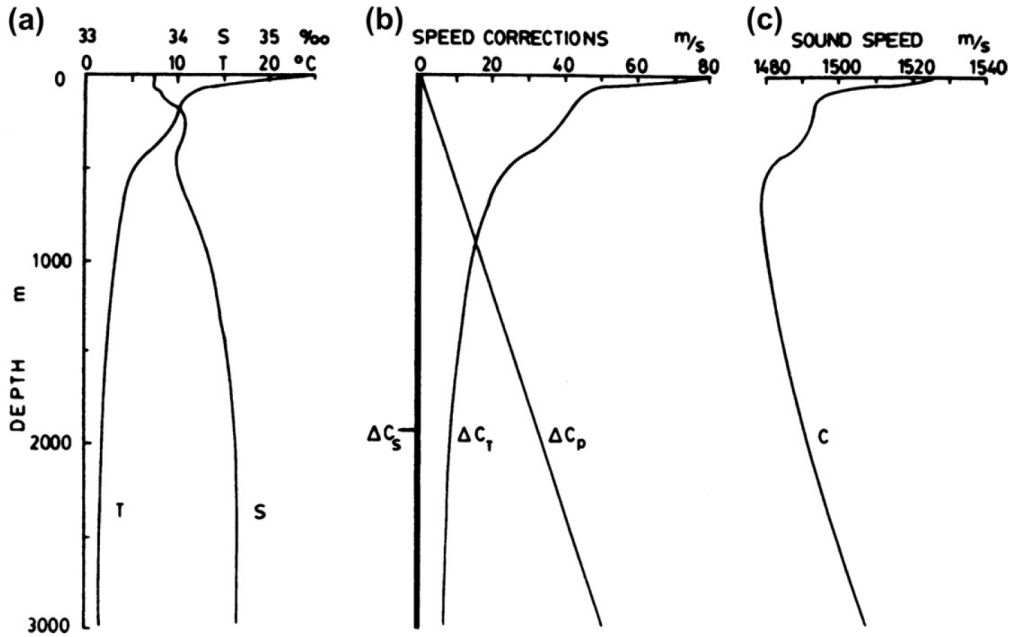

Figure 2.1: *A water profile from the Pacific Ocean at latitude 39°N. (a) temperature T and salinity S profiles, (b) sound velocity correction due to salinity Δc_S, temperature Δc_T, and pressure Δc_P, (c) resultant sound velocity c profile showing sound velocity minimum (SOFAR channel) around 600 m depth. From Talley et al. (2011).*

The SOFAR channel acts as a wave guide that can trap acoustic waves and facilitate their propagation over long distances. Figure 2.2 shows various ray paths propagating from a source in the SOFAR channel. Note how some of the rays are bound to a thin layer around the channel axis. Similarly, sound waves from sources outside the channel can be refracted and become trapped in this low-velocity wave guide. Whatever is the source depth, sound rays propagating upwards (at moderate angle) are refracted downwards, cross the SOFAR channel axis prior to be refracted upwards until crossing the SOFAR channel axis again. Thus, sound waves oscillate by refraction around the channel axis, hence the word "fixing" in the "SOFAR" acronym. It becomes clear that the SOFAR channel has a large acoustic exposure as signals always cross it, come back to it and even stay within it, hence the placement of hydroacoustic instruments in this layer. Moreover, this behaviour causes the sound to *"diverge cylindrically rather than spherically"* (Medwin & Clay, 1997). This focuses the wave energy and diminishes the impact of attenuation, enabling acoustic waves to travel large distances, hence the word "ranging" in the "SOFAR" acronym. Similarly, this horizontal stretch of the ray diminishes the length of their propagation path which also leads to less attenuation via energy absorption and wavefront

expansion. However, despite the shorter distance to cover, these rays are not detected earlier than rays propagating with a larger initial angle of inclination because the lower propagation speed in the SOFAR channel largely compensates for the small difference in distance to travel. Rays with the lowest angle of incidence and which travel the tightest to the channel axis arrive last. Finally, rays with steep initial angles of inclination are reflected on the seafloor and on the ocean surface because of the too large amount of steering required to change their propagation from upwards to downwards and vice versa.

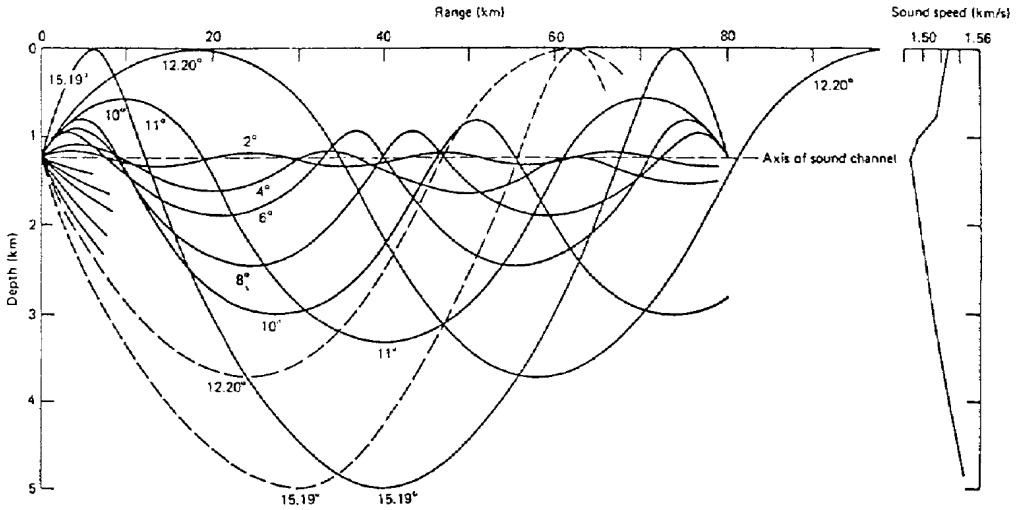

Figure 2.2: *Ray diagram for an Atlantic profile situated in temperate latitudes. The source is on the channel axis, at 1.3 km depth. On the right is the sound velocity profile. The initial inclination angles are given with respect to the channel axis. Steeper rays (omitted here, for simplicity) reflect on the seafloor. From Medwin and Clay (1997).*

Every structure has a natural frequency. Assuming the maximal thickness of the SOFAR channel is 400 m, the natural frequency is 3.75 Hz for the lowest frequency. In other words, frequencies lower than 3.75 Hz have wavelengths longer than the thickness of the channel and thus are not sensitive to the lower velocity. Therefore, they cannot get trapped in it. This value is calculated from the wave equation which links the wavelength λ, the frequency f and the velocity c as such: $c = f\lambda$, and from Talley et al. (2011) we know that *"the speed does not depend on the frequency"*. Moreover, it assumes a velocity close to 1 500 m/s. Nonetheless, 3.75 Hz is for signals propagating vertically, which is not of much interest for us as they do not leave the epicentral area. This low frequency limit of capture is in reality sensibly lower for rays travelling with some initial inclination angle. It is commonly assumed that the SOFAR channel can facilitate frequencies of 3 Hz and above due to its limited thickness (e.g. Evers et al. (2017) and Sambell et al. (2019)).

The SOFAR channel is generally located around a depth of 800 to 1200 m, in tropical and low temperate latitudes. It is at this depth range that the combination of pressure and temperature yields the slowest medium to propagate

acoustic waves. Towards the poles, the initial water temperature at the surface decreases, shifting upwards the sound velocity profile, and causing the SOFAR channel to be shallower. Until, at latitudes beyond 60°N and 60°S, the SOFAR channel reaches the surface because the velocity profile solely depends on the pressure and thus follows its linearly increasing pattern. Therefore, in this polar environment, rays are systematically refracted all the way up towards the ocean surface, where they are reflected (see figure A.1 in appendix).

2.2 Plane wave beamforming

Consider a three-dimensional (3D) planar wave field sampled by an arbitrary array of N receivers on a horizontal xy-plane. The local orientation of this 3D wave is defined by its inclination angle θ relative to the horizontal, azimuth angle ϕ clockwise from North and its horizontal slowness vector $\vec{p}$ defined as:

$$\vec{p} = -c_{app}^{-1} \begin{pmatrix} sin\phi \\ cos\phi \end{pmatrix}. \tag{2.1}$$

Beamforming yields an average array signal, also called "beam". In other words, it drags each array element to the same reference location. To form a beam, recordings are time shifted accordingly to a specific horizontal slowness vector $\vec{p}$. The time shift t_i is defined as follows:

$$t_i = -(p_x \cdot r_x + p_y \cdot r_y), \tag{2.2}$$

where p_x and p_y are the slowness components of the slowness vector $\vec{p}(p_x, p_y)$ and r_x and r_y the offset components of the position of instrument i with regard to the reference point. The coherency of the time shifted waveform recordings is evaluated for a range of slowness vectors of interest. The slowness vector $\vec{p}$ that yields the highest coherency measure is used to retrieve the wavefront parameters θ and ϕ. In this study, we use Fisher statistics to evaluate the coherency of a signal in a sliding processing window in time of size T. The summation process is defined by Melton and Bailey (1957) in equation 2.3. For a matrix of N traces and T samples, we have:

$$F = \frac{T(N-1)}{N(T-1)} \frac{\Sigma_{t=1}^{T}(\Sigma_{n=1}^{N} x_{nt})^2 - \frac{1}{T}(\Sigma_{t=1}^{T}\Sigma_{n=1}^{N} x_{nt})^2}{\Sigma_{t=1}^{T}\Sigma_{n=1}^{N} x_{nt}^2 - \frac{1}{N}\Sigma_{t=1}^{T}(\Sigma_{n=1}^{N} x_{nt})^2}, \tag{2.3}$$

where x_{nt} are the time-shifted samples and F is the Fisher ratio (abbreviated F-ratio). Higher F-ratio indicates higher coherency. The derivation of the F-ratio is an analysis of variance of both noise and signal and is in substance similar to testing a statistical hypothesis. Hypothesis H_0 is that recordings contain only random noise and hypothesis H_1 is that they also contain a common pattern (i.e. a signal perturbed the ambient field and was recorded by more than one instrument). If H_1 is true, a variance appears in the data which cannot be

attributed to the noise as it is common to several instruments. Therefore, data deviates from the mean and have a non-central distribution.

Finally, the apparent velocity and back-azimuth of the most coherent average array signal in each time window are stored for further processing.

The algorithm used in this study is an in-house developed beamforming implementation[1] of Fisher statistics developed in Melton and Bailey (1957). This algorithm was already used in previous studies, e.g. Shani-Kadmiel et al. (2018), and is called via a Python interface: the different codes used as part of our study are in appendix F.

2.3 Temperature–Sound Speed relation

Equation 1.1 in introduction is a simplified version of the empirical formulation of Del Grosso (1974). When compared with his formulation, the velocity calculated has a precision of less than 0.2 m/s, as long as the temperature ranges from 0 to 32 °C, the salinity ranges from 22 to 45 g/l and the depth is not larger than 1000 m. For a given depth, Dushaw et al. (2009) show that change in temperature is the main cause for sound velocity variation. Others before them had remarked the same thing: Munk et al. (1995) and Worcester (2001) show that at mid-latitudes acoustic velocity variations due to change in salinity are about only a few percent of the extend of the ones caused by change in temperature; Lukas (2001) shows that in the subtropical gyre near Hawaii, in the upper thermocline, sound velocity variations due to change in salinity are about 10 times smaller than the ones caused by changes in temperature. With this in mind, Sambell et al. (2019) approaches equation 1.1 with a first order approximation, ignoring higher order terms of the temperature and assuming a constant salinity of 35 g/l. The temperature becomes:

$$T = \frac{c - 1449.2 - 0.016z}{4.6}. \tag{2.4}$$

Equation 2.4 is the one used in this study to convert sound velocities c to temperature T values, with z being the depth at which acoustic velocities are retrieved.

[1] More information on the beamforming algorithm authors in bibliography: Shani-Kadmiel et al. (2020).

3

Methodology

3.1 Analysis of hydroacoustic signals

In this section, we present the main characteristics of an earthquake signal recorded by a hydroacoustic array. This short analysis helps to understand the selection of some of the processing parameters, such as the ones related to the picking of the waveform interval and the bandpass filter. Figure 3.1a shows the raw waveforms recorded by the six hydrophones of H10 following a magnitude 6.2 earthquake 5 422 km away. In addition, frames (b) and (c) show, from top to bottom, the spectogram of the instrument response removed and unfiltered waveform, its corresponding waveform and the filtered waveform. We focus on arrivals with a seismic celerity (b) and on arrivals with a hydroacoustic celerity (c). Seismic waves from a magnitude 6 event and situated more than 5 000 km away will not have any energy at frequencies higher than 1 or 2 Hz. Consequently, the presence of energy above 10 Hz in frames (b) invalid the possibility that they are related to seismic arrivals which shacked the anchorage system. Similarly, the filtered (0.01–2 Hz) waveform reveals that no coherent arrivals has energy below 2 Hz. These signals were certainly generated by interaction of the seismic waves with other local features of H10S, e.g. the bathymetry, topography, sediments. These are secondary sources whose signals can also couple to the water column and get trapped in the SOFAR wave guide.

The two frequency contents show that noise as well as signals from continuous sources mostly have energy in the frequency band below 3 Hz because there is always energy in that band. Regarding acoustic waves (figure 3.1c), it is quite clear that their signal is broadband. It is however more dominant in the 3–12 Hz band. This highlights the lower limit of frequencies that can be facilitated by the SOFAR channel. Moreover, the very small difference between the unfiltered and filtered (between 3 and 20 Hz) waveforms shows that the signal does not have much energy in the frequency band above 20 Hz.

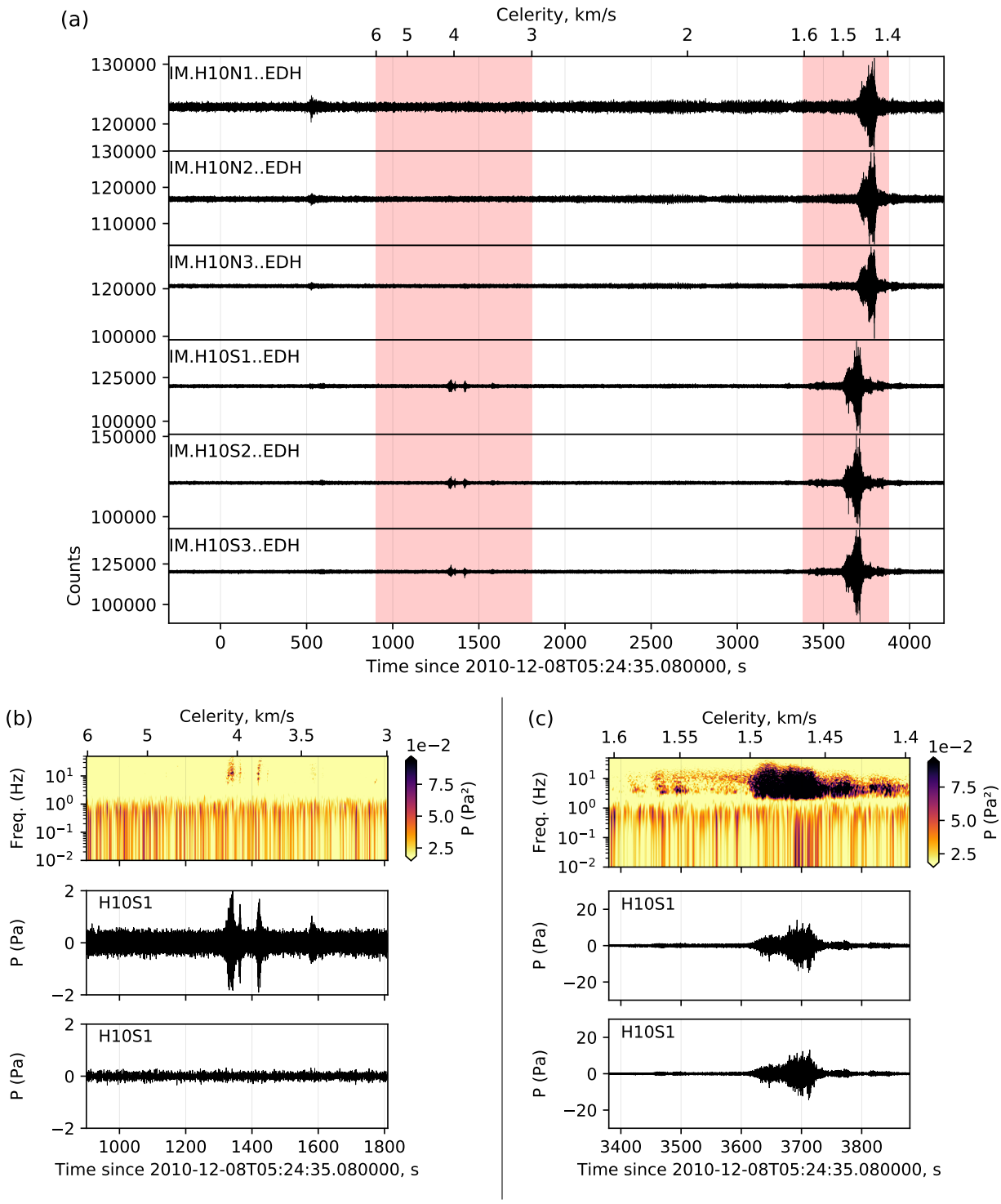

Figure 3.1: *An earthquake signal recorded by a hydroacoustic array. (a) six raw waveforms at H10. Highlighted in red are the seismic (left) and acoustic (right) arrivals. (b) zooms on arrivals with a seismic celerity of element H10S1. From top to bottom, we see the spectogram of the instrument response removed and unfiltered waveform, its corresponding waveform and the filtered (0.01–2 Hz) waveform. (c) is like (b) but for the arrivals with an acoustic celerity. In (c), the bandpass filter used to obtain the lowest frame is between 3 and 20 Hz.*

The shape of the acoustic signal is quite elongated on figure 3.1c. This is mostly due to two reasons. Firstly, the patch of seafloor above the source that vibrates is not punctual and has some size. Some of the patch is closer to the array, and its emissions thus have a shorter travel time, part of the patch is further away and its emissions thus have a longer travel time. The symmetry of the patch also plays a role. Secondly, parts of the wave field also propagated outside the SOFAR channel and thus with a slightly higher celerity. These waves thus have a shorter travel time.

3.2 Data processing

In the first place, a catalogue of earthquakes which happened in the Atlantic ocean basin between 2005 and 2018 was compiled. Figure 3.2 shows the polygon that was constructed to select the events (the polygon coordinates are in the appendix in table B.1). For the purpose of this study, we selected earthquakes with a magnitude larger than 4 to ensure waveforms have a sufficiently large SNR. Afterwards, events whose distance to the array centre is smaller than eight times the array aperture of the joint array are disregarded. From figure 3.5, we see that events closer than this distance lead to a difference in back-azimuth larger than 5° at each triplet. As this difference gets larger, the precision of the retrieved apparent velocity by beamforming diminishes (see section 3.2.1). Moreover, this minimum distance ensures that the plane wave approximation holds.

For each event in the catalogue, a time interval was selected around the acoustic arrivals following equations 3.1 and 3.2. This ensured we are beamforming hydroacoustic signals and that they were generated by one of the events in our catalogue. Moreover, limiting the signal duration leads to a computationally less expensive beamforming. Afterwards, waveforms in this interval were demeaned, tapered (with a max percentage of 5% and a taper length of 20 samples) and had the instruments sensitivity removed. Moreover, they were bandpass filtered between 3 and 20 Hz using a fourth-order Butterworth filter, accordingly to the explanation of the anatomy of the hydroacoustic signal made in section 3.1.

Beamforming was applied to a sliding time window of 2.75 s (687 samples) with an overlap of 50%. The discrete slowness grid was specifically designed for each event to include a back-azimuth range of +/–15° around the theoretical back-azimuth between the array geometrical centre and the epicentre, with an increment of 2°. Sampling only the slowness vectors around the theoretical back-azimuth is computationally less expensive than if the algorithm had to sample the entire grid. An example of slowness grid is visible in figure 3.3. Regarding the apparent velocity, the grid was defined from 1 450 to 1 540 m/s with an increment of 1 m/s. This velocity range is more centred on the low velocities of the SOFAR channel and is also the same as the one chosen by Sambell et al. (2019). Using equation 2.4 we see that every 1 m/s in apparent velocity corresponds to 0.22 °C in temperature. All six elements of H10 were used for beamforming. However, after October 2013, array element S1 started malfunctioning and was discarded.

Figure 3.2: *Geographic zone used for earthquake selection on the interactive bulletin search on the International Seismological Centre (ISC) website.*

3.2.1 A SOFAR channel signal

We have taken several steps to extract the part of the signal that is propagating in the SOFAR channel. Failure to do so would result into beamforming retrieving imprecise apparent velocities, because the velocities sampled by beamforming would correspond to none of the true velocities which yielded the coherent arrivals. Moreover, as we are only interested in the acoustic part of the signal that is propagating in the SOFAR channel, much computational cost can be saved by only beamforming this part of the signal. We extract a window around the acoustic arrival by approximating the start and end time, t_1 and t_2 respectively, as a function of the origin time t_0, the epicentral distance r, and two bounding celerity values chosen to be $c_{max} = 1600$ m/s and $c_{min} = 1400$ m/s in this study. Equations 3.1 and 3.2 show the obtaining of the start time and end time, respectively.

$$t_1 = t_0 + \frac{r}{c_{max}} \tag{3.1}$$

$$t_2 = t_0 + \frac{r}{c_{min}} \tag{3.2}$$

These celerity range bounds were chosen to allow ample time before and after the actual range of interest, which is 1540 m/s and 1450 m/s. This time interval

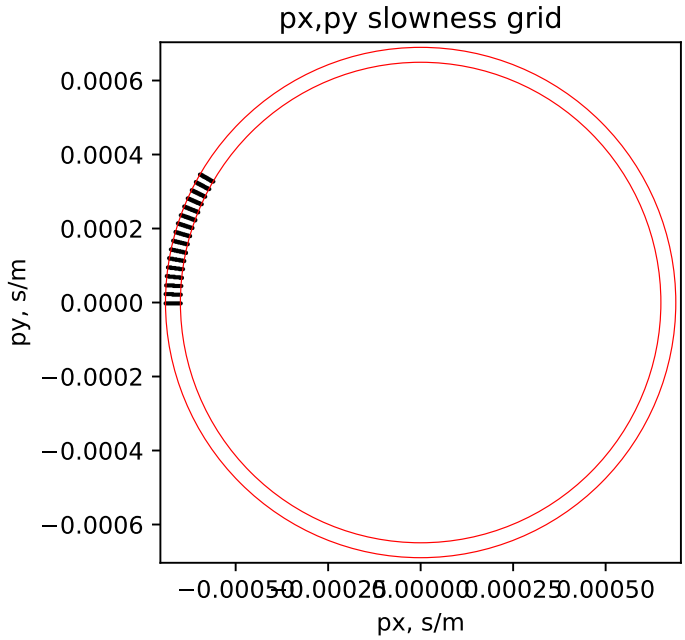

Figure 3.3: *Example of slowness grid for an event whose theoretical back-azimuth is 285°. On the x and y axis are the slowness vectors along the x and y dimension respectively, from a bird's eye view. This grid is built with an increment of 2° in the back-azimuth range 270–300° (i.e. 285°+/–15°) and with an increment of 1 m/s in the velocity range 1450–1540 m/s.*

is broader also to ensure that a maximum of signals travelling across the array in the SOFAR channel are captured. For instance, it can take some time for an acoustic wave to propagate from the seafloor until it gets fixed in the SOFAR channel. During this time span the wave does not propagate at the velocity of the SOFAR channel yet.

Figure 3.4 shows the stacked waveforms of the North and South triplets, for five random earthquakes. In black is the part of the signals that is meant to later be beamformed. Firstly, a back-azimuth difference is visible when the epicentre is situated relatively close to the array. Secondly, the further the epicentre is situated from H10, the broader becomes the time interval and the smaller is the signal amplitude. Thirdly, the signature of the stacked waveforms generated by the passage of the same plan wave over the arrays look similar.

The geometry of the problem we are solving implicates that the back-azimuth of the planar wave field remains constant at each array element so that the apparent velocity can be calculated precisely in the beamforming process. To conform to this, we ensure that the direction of the propagating signal across both triplets is similar (see figure C.1 in appendix). This is achieved by selecting events that are detected at both triplets with similar back-azimuths. Figure 3.5 presents the absolute difference in the theoretical back-azimuth between the North and South triplets as a function of epicentral distance. In this study,

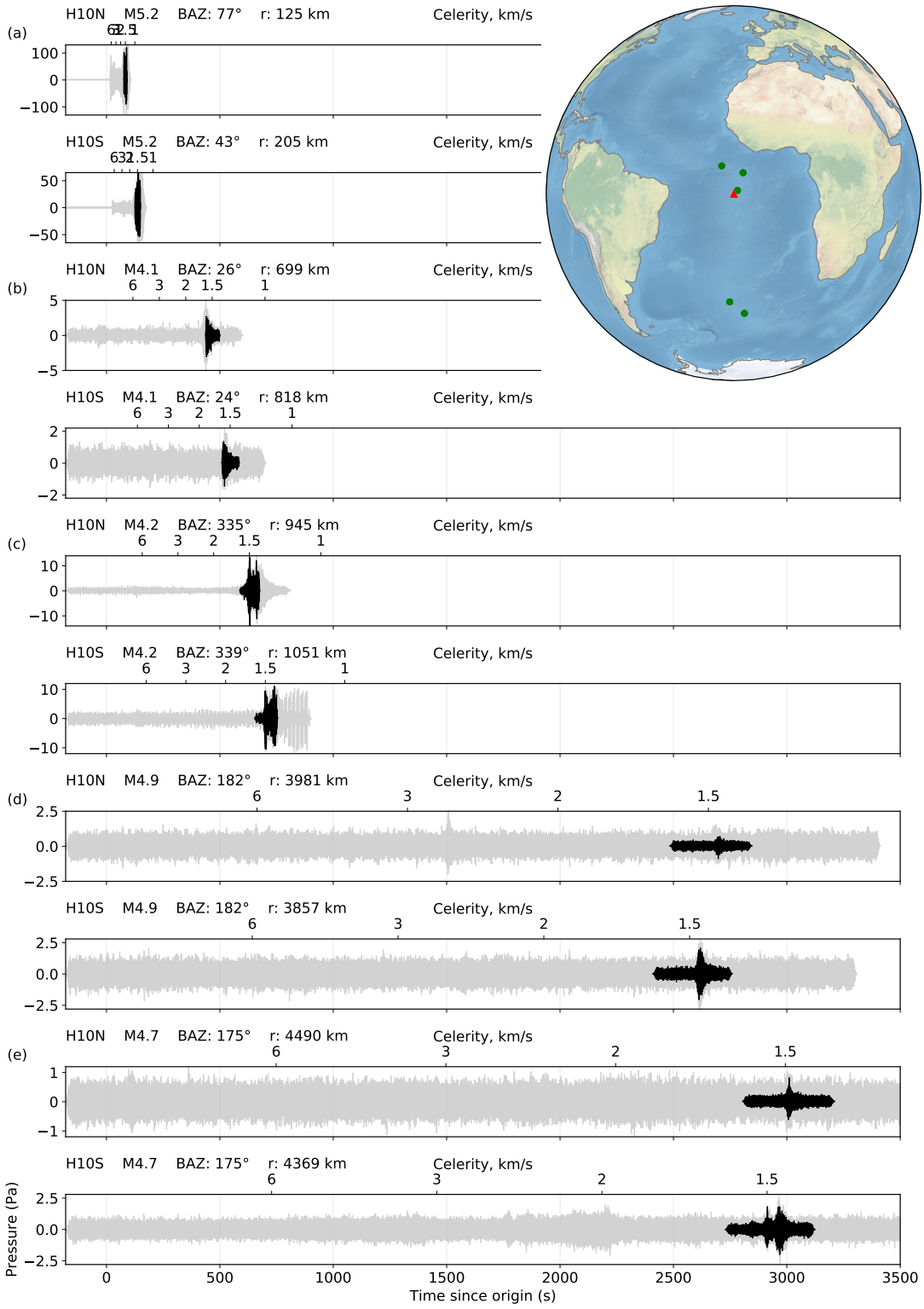

Figure 3.4: *Zero-phase stacked signals of five random events (a to e) from the year 2010, sorted by epicentral distance. For each of them, upper and lower frames correspond to the North and South triplet, respectively. In black is the part of the signal that travelled with a celerity between 1 400 and 1 600 m/s. All waveforms are demeaned, tapered and have the instruments sensitivity removed. Moreover, the black waveforms are tapered a second time and bandpass filtered between 3 and 20 Hz. On the map, green dots show the earthquake epicentres and the red triangle shows H10 location.*

we set a minimum distance criterion of height times the joint array aperture (1 010 km), which ensures the back-azimuth difference is not larger than 5°. As a result of this criterion, we can also assume that the curvature of the wavefront is negligible and that the plane-wave approximation holds.

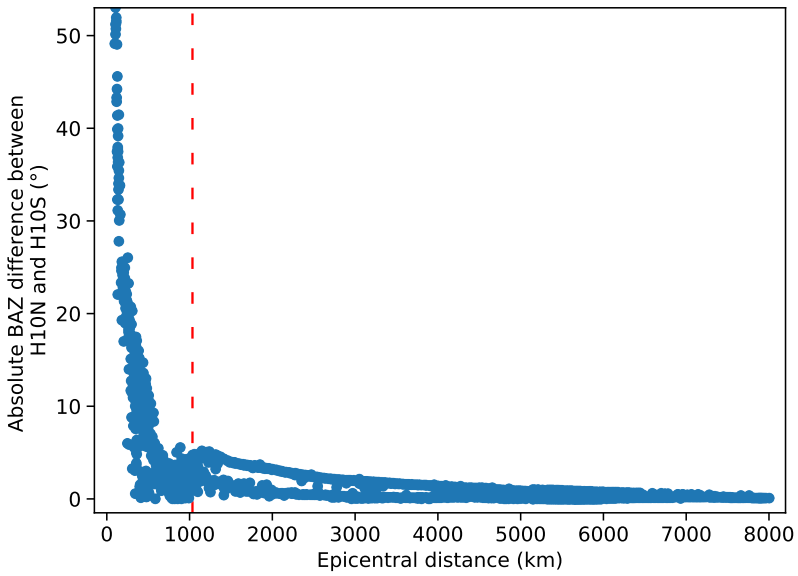

Figure 3.5: *Absolute difference in the theoretical back-azimuth between the North and South triplets in function of epicentral distance. The red dashed line symbolises the minimum epicentral distance (1 010 km) criterion used in this study.*

3.2.2 Beamforming

Beamforming is used in the frame of an inverse problem approach. Indeed, it enables us to retrieve the *"distinctive details of the propagation at a time and space to deduce the parameters and processes of the particular ocean through which the sound has travelled"* (Medwin & Clay, 1997). Medwin and Clay (1997) also call this inverse view "acoustical oceanography". More concretely, beamforming is a method which cancels out incoherent noise and enhances signal by combining traces. The best beam is the name given to the average array signal obtained from the slowness vector yielding the maximum F-ratio.

One of the main advantage of beamforming is that it can evaluate the time shift of a certain range of both apparent velocities and back-azimuths chosen beforehand. This makes beamforming very practical from a computational cost point of view and also to verify prior knowledge. The beamforming algorithm used in this study is called a "time domain Fisher detector". "Time domain" because samples are shifted in time, and "Fisher detector" because the summation process calculating the degree of coherency of each set of delayed traces uses Fisher statistics. For long time periods, traces are not

compared and analysed along the entire signal duration at once, but rather one window of samples at a time. Henceforth, the word "detection" will refer to the beamforming results of the best beam of one window. The three figures in appendix D show some beamforming results for earthquakes with different magnitudes at arrays with different number of elements. The window duration is 2.75 s, and there is a 50% overlap between each window. It is visible that detections with the highest F-ratio are the ones linked with the passage of the signal over the array. More importantly, they also show more constancy in both apparent velocity and back-azimuth and are associated with a higher SNR.

3.3 Individual triplets vs. joint array

From the equation of the F-ratio (equation 2.3), Melton and Bailey (1957) also derived a relation between the SNR on the number of traces N and the F-ratio F:

$$F \;=\; N \cdot SNR^{\,2} + 1. \tag{3.3}$$

It results that a beam has an SNR increase of $\sqrt{N}$ under the assumption of uncorrelated noise. Therefore, the SNR increases if more sensors sample the wave field, as more elements confirm the presence of coherent arrivals. This was similarly discussed by Woolfe et al. (2015), who suggested that both triplets can be jointly process in a beamforming procedure with an associated error e proportional to:

$$e \;\; \alpha \;\; \frac{1}{\sqrt{T/L} \cdot \sqrt{MN}}. \tag{3.4}$$

N and M are the numbers of elements in each array, T is the window duration and L is the distance between the two array centres. Equation 3.4 indicates that the error is inversely proportional to the product of the number of elements in each array. A smaller error obviously means a larger SNR. Regarding the first term, it acts as a normalising factor.

To compare beamforming results using each triplet individually versus processing all six elements as one joint array, we examine one year of earthquakes from the compiled catalogue. Figures 3.6 and 3.7 show detections that were detected at the North and South triplets, respectively, with an SNR larger than 0.6. The top frames (a) show a histogram of the detections in 5° back-azimuth bins. The middle frames (b) show detections as a function of time and back-azimuth, colour-coded by apparent velocity. The bottom frames (c) show apparent velocity and 95% confidence interval in each back-azimuth bin.

Not all events that are detected at H10N (North triplet) are detected at H10S (South triplet) and vice versa. Beamforming at H10N picks up more earthquakes from the North with a back-azimuth range of 300°–5° (figure 3.6a), whereas beamforming at H10S picks up more earthquakes from the South with

a back-azimuth range of 170°–210° (figure 3.7a). Moreover, a close inspection of the average apparent velocity per back-azimuth bin shows that the apparent velocity changes as a function of back-azimuth. Sambell et al. (2019) observed a similar dependency of apparent velocity as a function of back-azimuth at H03S (near Robinson Crusoe Island, off Chile) and attributed it to local effects, most likely related to steep slopes in the local bathymetry.

Figure 3.8 shows all detections that were detected at the joint array with an SNR larger than 0.6. We note that although less events are detected in total, the apparent velocity variation as a function of back-azimuth is almost completely suppressed and that confidence intervals are generally smaller — with the exception of very poorly sampled back-azimuth bin. We therefore conclude that joint-array beamforming is superior to individual triplet beamforming to retrieve an accurate apparent velocity. This method is thus more appropriate for assessing parameters which necessitate high accuracy and precision, such as the water temperature.

3.4 Post-processing

For each event, detections with an SNR higher than 0.6 are selected and the associated wavefront parameters — namely back-azimuth and apparent velocity — are aggregated per back-azimuth bins of 5°. Using an SNR threshold of 0.6 further constrains the detections to the main part of the SOFAR channel and ensures that the F-ratio remains above a value of 2. For instance, by looking at the three figures in appendix D, it is visible that detections with an SNR of 0.6 and above have an increased F-ratio and thus are linked with the passage of a coherent wave field over the array and correspond to a more consistent apparent velocity and back-azimuth. Furthermore, even for a magnitude 6 event (figure D.3), for which a particularly high SNR is expected, 0.6 is a good threshold to conserve only the main part of the signal without the less coherent tails (at least for the joint array, which is the array this study focuses on).

Afterwards, the apparent velocity is averaged per month over each back-azimuth bin and the confidence interval of 95% is calculated. Monthly back-azimuth bins containing less than six detections are disregarded. For the analysis of the temperature over time, only the four back-azimuth bins with the most monthly bins between 2005 and 2018 are retained. In descending order, these are the 5° bins centred on 187.5°, 192.5°, 307.5° and 302.5°. For each of them, monthly average velocities and temperatures are fitted in order to highlight eventual trends over time. We choose a polynomial regression because cyclic temperature variations (e.g. due to seasonal variations) are negligible at the latitude and depth of H10 (Locarnini et al., 2010). Furthermore, we use a polynomial of the second degree. This choice is made upon a trial-and-error process: a second degree polynomial advantageously offers a bit more flexibility in the fitting than a first degree one, without the complexity of higher degree polynomials; our goal is to detect general trends. Moreover, each of the four back-azimuth sets of monthly results is fitted for the 2005–2018 period and the

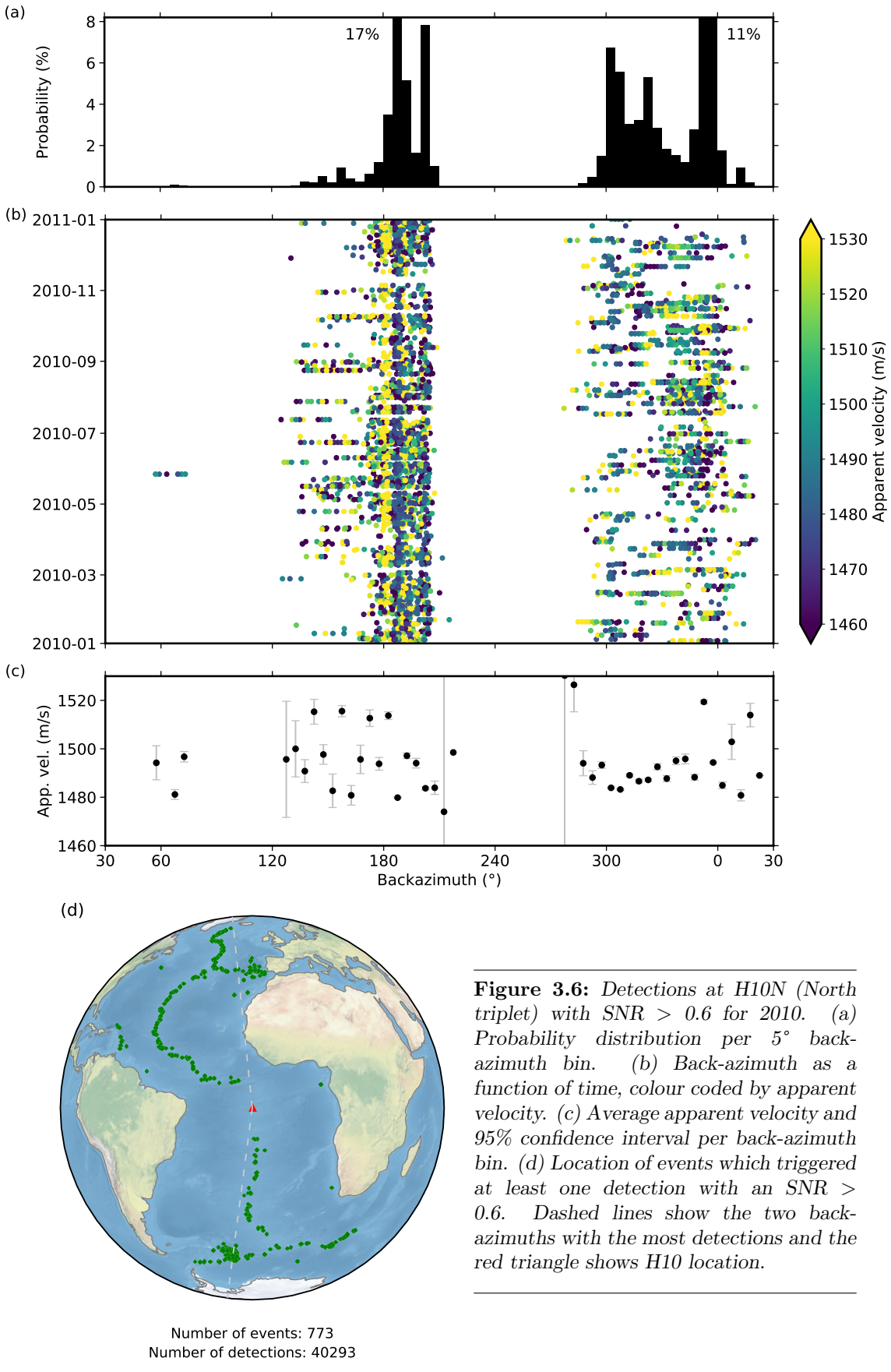

Figure 3.6: *Detections at H10N (North triplet) with SNR > 0.6 for 2010. (a) Probability distribution per 5° back-azimuth bin. (b) Back-azimuth as a function of time, colour coded by apparent velocity. (c) Average apparent velocity and 95% confidence interval per back-azimuth bin. (d) Location of events which triggered at least one detection with an SNR > 0.6. Dashed lines show the two back-azimuths with the most detections and the red triangle shows H10 location.*

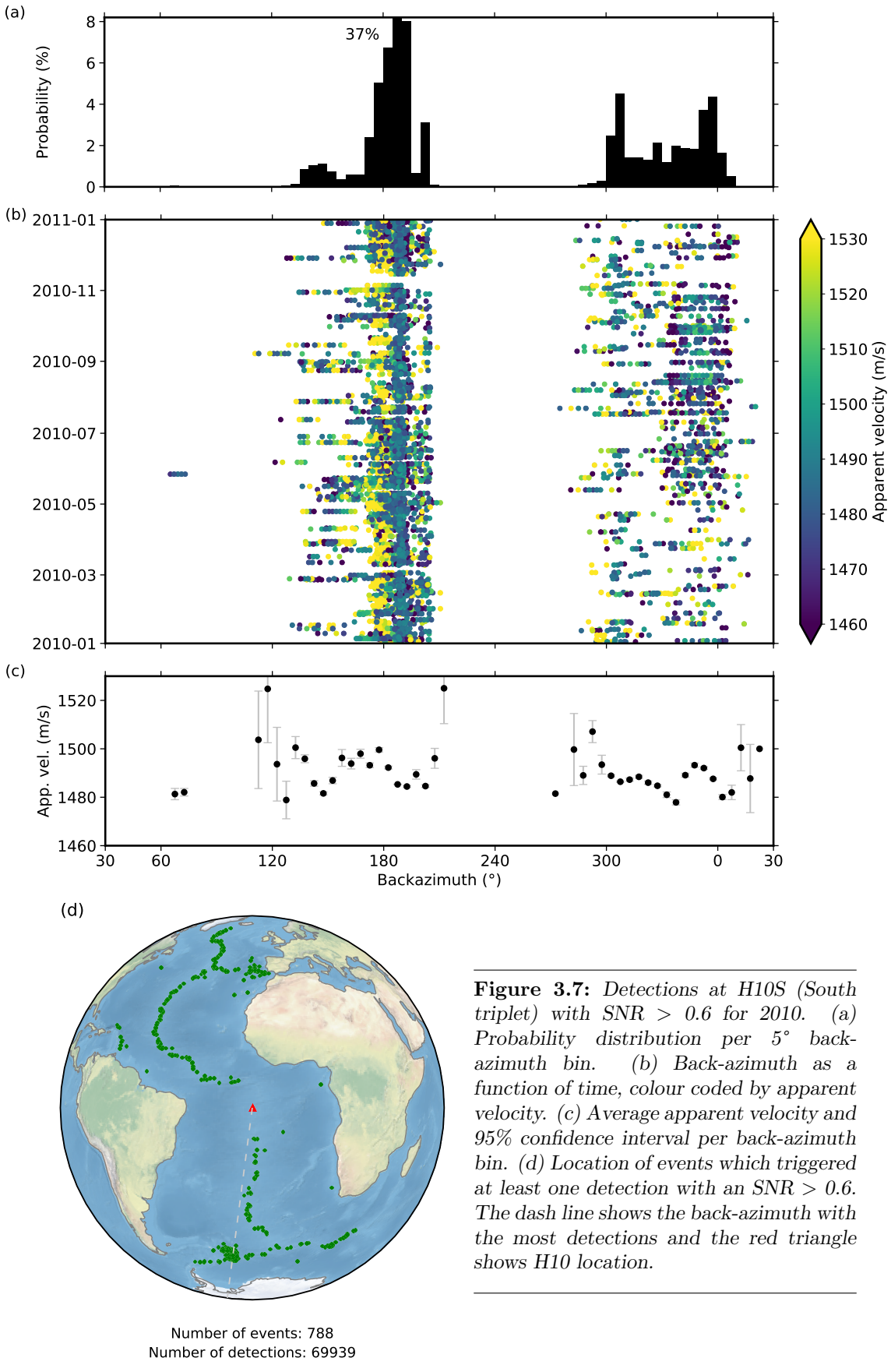

Figure 3.7: *Detections at H10S (South triplet) with SNR > 0.6 for 2010. (a) Probability distribution per 5° back-azimuth bin. (b) Back-azimuth as a function of time, colour coded by apparent velocity. (c) Average apparent velocity and 95% confidence interval per back-azimuth bin. (d) Location of events which triggered at least one detection with an SNR > 0.6. The dash line shows the back-azimuth with the most detections and the red triangle shows H10 location.*

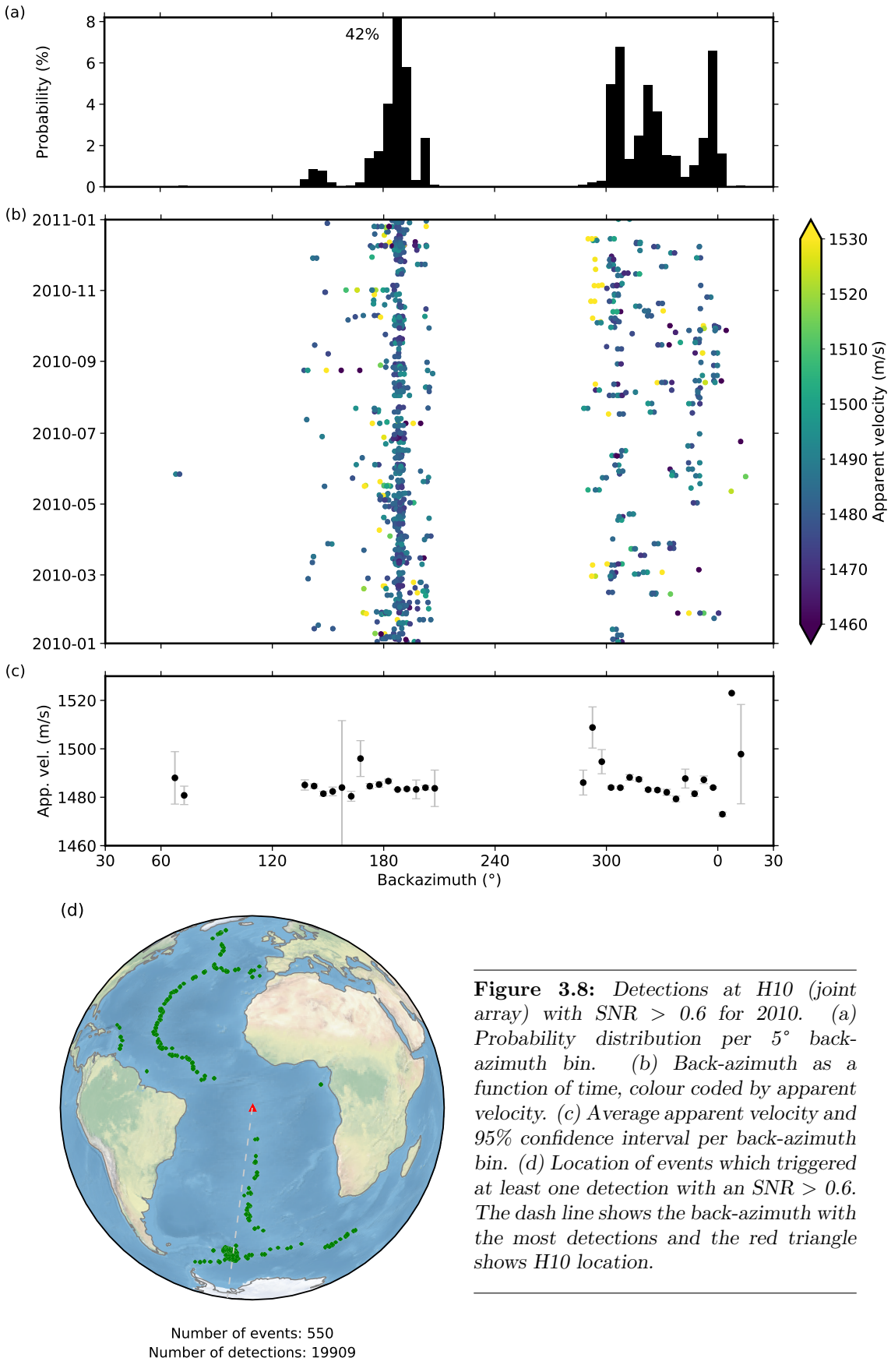

Figure 3.8: *Detections at H10 (joint array) with SNR > 0.6 for 2010. (a) Probability distribution per 5° back-azimuth bin. (b) Back-azimuth as a function of time, colour coded by apparent velocity. (c) Average apparent velocity and 95% confidence interval per back-azimuth bin. (d) Location of events which triggered at least one detection with an SNR > 0.6. The dash line shows the back-azimuth with the most detections and the red triangle shows H10 location.*

2005–2013 one in order to also make an evaluation that is not impacted by the malfunction of element S1. Therewith, the regression is conducted once with no weighting applied on the monthly results, and once with a weighting inversely proportional to the 95% confidence interval of the monthly averages. This means that the larger the confidence interval of a certain month, the less the average apparent velocity or temperature associated to this month contributes to the polynomial regression. In conclusion, each of the four back-azimuth bin is fitted four times.

4

Results

4.1 Description

Figure 4.1 summarises all the earthquakes in our catalogue that were detected at the joint array with an SNR larger than 0.6. As before, the top frame (a) shows a histogram of the detections in 5° back-azimuth bins. The middle frame (b) shows detections as a function of time and back-azimuth, colour-coded by apparent velocity. The bottom frame (c) shows apparent velocity and 95% confidence interval in each back-azimuth bin.

A large proportion of detections (41%) comes from the South Sandwich Islands and ends up in the two bins centred on back-azimuths 187.5° and 192.5°. These bins have apparent velocities considerably lower than their less sampled neighbours: approximately 1 482 m/s compare to 1 493 m/s. Despite being less sampled, these neighbouring bins still have very small confidence intervals. It is also visible that the apparent velocity is higher in the back-azimuth range 280° to 300°, with values above 1 500 m/s (light green and yellow dots). Moreover, more detections occur after October 2013 in this back-azimuth range. On top of this the overall number of detections increases after 2016 and they are spread more continuously across many back-azimuths.

Figure 4.2 shows the monthly average velocities and temperatures of the four back-azimuth bins with the most monthly bins. They almost correspond to the four bins with the largest amount of detections visible in frame 4.1a. A second degree polynomial regression is used to fit the results, for both the 2005–2013 (green) and 2005–2018 (blue) periods, as well as with (solid) and without (dashed) weighting applied on the results. The weighting on a data point is inversely proportional to its confidence interval. These back-azimuth bins conveniently avoid arrivals detected in the back-azimuth range 280–300° which contain rather high apparent velocities.

Firstly, we can see that the monthly averages, before 2013, are mostly

between 4.0 and 5.0 °C. Secondly, they are often within a 0.5 °C interval from the fitting polynomials, and the monthly 95% confidence intervals are mostly 1.0 to 2.0 °C wide. With the exception of back-azimuth bin 187.5° which shows an extraordinarily small variability in both monthly averages and confidence intervals, especially for the 2008–2010 period. All four bins show globally an increase in confidence interval after October 2013, except eventually for back-azimuth bin 302.5° where the confidence intervals are already rather large before 2013. The same goes for the variability of the averages, on top of which an upwards trend emerges. This leads to the 2005–2018 curves (blue) to increase over time, with the exception of back-azimuth bin 192.5° weighted. Regarding the 2005–2013 period, all bins show a slight upwards trend in temperature, and the weighted and unweighted regressions are similar.

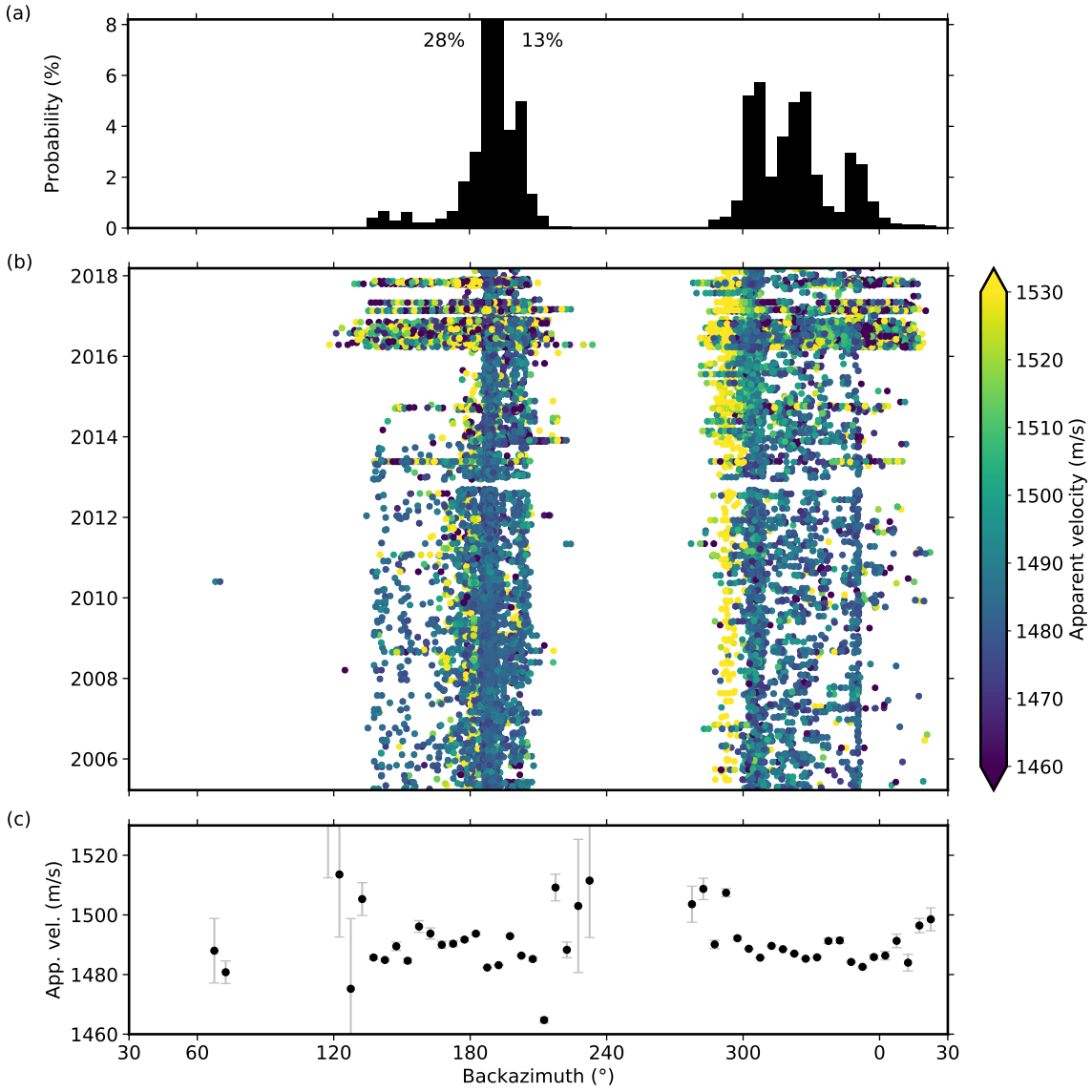

Figure 4.1: *Detections at H10 (joint array) with SNR > 0.6 between 2005 and 2018. (a) Probability distribution per 5° back-azimuth bin. (b) Back-azimuth as a function of time, colour coded by apparent velocity. (c) Average apparent velocity and 95% confidence interval per back-azimuth bin.*

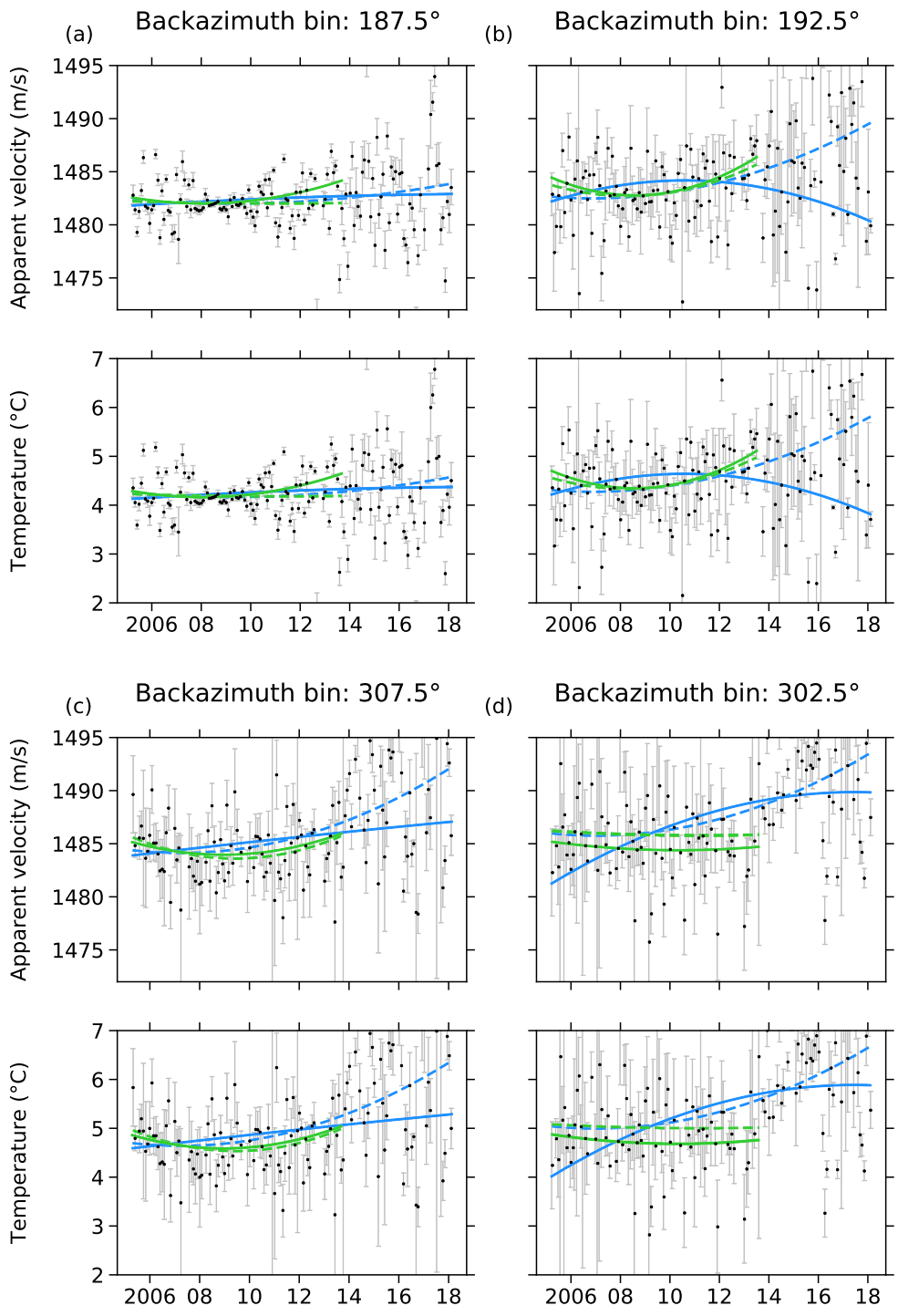

Figure 4.2: *The four back-azimuth bins with the largest coverage through time (from (a) to (d)). The back-azimuth indicates the centre of the bin. For each bin are shown the monthly average apparent velocities and temperatures as well as their 95% confidence intervals. A polynomial regression of the second degree fits the averages for the 2005–2018 and the 2005–2013 periods, in blue and green respectively. In addition, the fitting is applied once with weighting (solid lines) and once without any (dashed lines): the weighting of one data point is inversely proportional to its confidence interval.*

4.2 Interpretation

The relatively high apparent velocities in the back-azimuth range 150° to 180° in figure 4.1c resembles the bias discussed with regard to figures 3.6c and 3.7c. However, we do not think that this bias reappears so strongly in the joint array beamforming results when averaging detections over a longer period of time. Indeed, figure E.1 in appendix, shows that this velocity–back-azimuth dependency is reduced when disregarding detections which occurred after the malfunction of hydrophone S1 (i.e. after October 2013). The same goes for back-azimuth bins situated between 310° and 350°. Therefore, we interpret that when the joint array is reduced to five elements, the retrieved apparent velocity could become more prone to change as a function of back-azimuth — like it is the case for the triplets individually.

To investigate the more numerous detections with a high apparent velocity recorded after 2016, we re-beamformed earthquakes that happened in 2010 but excluding array element S1. Results are in appendix in figure E.2. The apparent velocity is much less constant than in figure 3.8c and confidence intervals are also larger. Moreover, a pattern similar to the one in figure 4.1b after 2016 is visible. Consequently, we interpret that the absence from the beamforming processing of data from element S1 led to retrieving more detections after 2013 — but more especially after 2016 for an unknown reason. We also saw in section 3.3 that the beamforming process for the joint array retrieved less detections than for the individual triplets. We can think of three potential reasons for this: (1) the presence of Ascension Island that could block some signals propagating from one triplet to the other; (2) the F-ratio is divided by the number of array elements which leads to a smaller SNR for detections at the joint array (see also the bottom frames of the three figures in appendix D); (3) the back-azimuth of the planar wave field can vary slightly between each element location, variation that would be larger for the joint array and thus leads to a larger decrease in SNR.

Figure E.3 shows the monthly average temperature across all back-azimuth bins as a function of time. Although this method cannot be applied to retrieve a precise value for the temperature, it has at least the advantage of also putting in evidence that, after the malfunction of element S1, the retrieved apparent velocity and temperature both suddenly increase and also have a larger variability.

Regarding the high apparent velocities in the back-azimuth range 280° to 300°, a velocity of 1 500 m/s corresponds to a temperature of 8 °C. This is too large to be physically reasonable. Indeed, in the World Ocean Atlas (2009), Locarnini et al. give a temperature of 5.0 °C, for a depth of 850 m; as for Evers et al. (2017), cross-correlating the array beams resulted in a velocity of approximately 1 481 m/s (4 °C). Consequently, it seems that joint-array beamforming could not avoid the apparent velocity to be biased for this back-azimuth range. As a result, detections in this back-azimuth range should not be used for retrieving accurate temperature variations over time (they are conveniently not part of the four bins with the largest coverage through time).

We have seen that the malfunction of element S1 leads to considerable changes in the beamforming results. Figure 4.2 shows monthly averages and therefore allows us to study separately the 2005–2013 period, from the 2005–2018 period. Regarding potential dependency between velocity and back-azimuth, they are certainly minor and would be systematic errors as they depend on the back-azimuth. As a result, they would probably shift up or down equivalently all data points but not modify our understanding of the temperature evolution over time. Firstly, before 2013, the monthly average temperatures are mostly between 4.0 and 5.0 °C which is in the range of values given in the literature. For back-azimuth bins 187.5°, 192.5° and 307.5° the blue curves matches the green ones relatively well. It is only after 2013, that the blue curves show a more rapid increase. This increase can well be seen on figure 4.3 which shows the variation in monthly temperature over time and over each back-azimuth bin. For back-azimuth bin 302.5°, the increase is even too rapid for the blue and green curves to match before 2013. This is due to too numerous high velocity detections after 2013 which could be caused by the inclusion in the bin of some of the high velocity detections of the back-azimuth range 280–300°. This could explain why this bin's polynomial fit is the highest and why this bin's results have the largest confidence intervals.

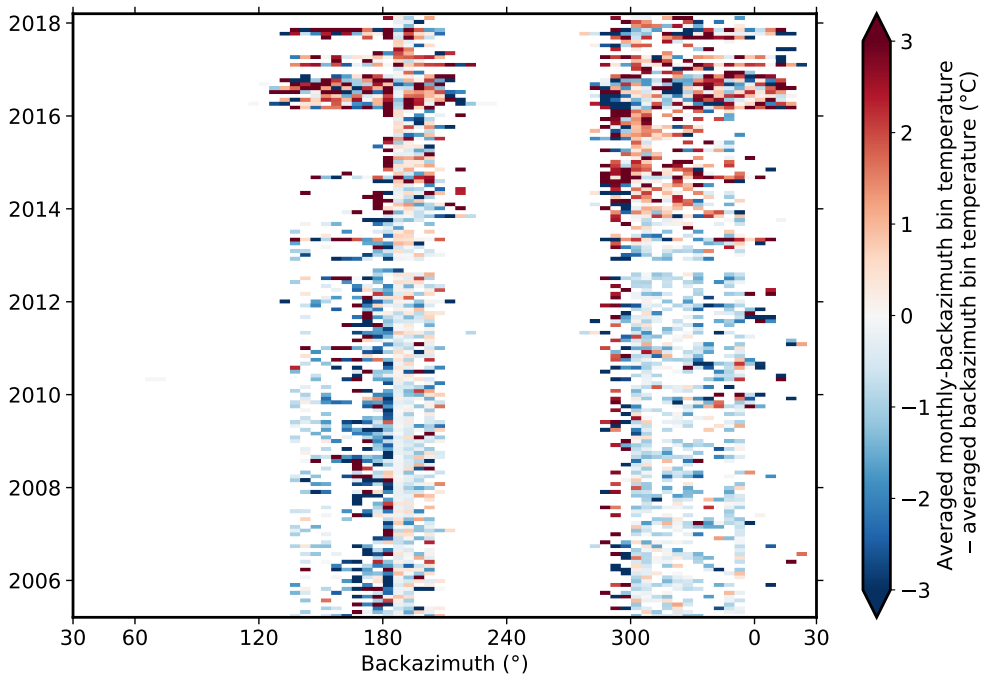

Figure 4.3: *Back-azimuth as a function of time, colour coded by variation of monthly temperature. The colour shows the difference between the bin temperature and the average temperature of its entire back-azimuth bin.*

After 2013, the increase is more rapid for the unweighted fits in all bins: it shows that monthly averages with high values also have large confidence

intervals. The weighted fit of back-azimuth bin 192.5° is the only one to show a decrease in temperature after 2013. Overall, the 2013–2018 period appears not reliable enough to highlight temperature trends, the fault of the introduction of many detections with high apparent velocity and velocity averages with large confidence interval. We think this behaviour is linked with the malfunction of element S1.

Regarding the 2005–2013 periods, the weighted and unweighted fits are similar, except for the back-azimuth bin 302.5° where the weighted regression gives lower values. We interpret this as a possible indication that the weighted fit is less sensitive to the eventual inclusion of high velocities in the bin and thus weighting the results could be more advantageous than not weighting them.

We think that the very small confidence intervals in bin 187.5° could be linked with the fact that this bin has the most detections. However, this would not explain why the temperature varies very little between 2008 and 2010. Neither events nor detections are more numerous during this period. Consequently, it might also be that some particularly similar earthquakes happened along these couple of years in the South Sandwich archipelago.

To summarise, we think that detections after 2013 are not satisfying to retrieve the temperature accurately and that back-azimuth bin 302.5° is the least accurate of the four analysed due to its large confidence intervals. The average temperature of the three others is between 4.3 and 4.8 °C, for the 2005–2013 period. Moreover, there might also be a slight temperature increase of 0.2 °C. However, this is statistically not robust since we estimate having a precision of 0.5 °C (4.8-4.3 °C) and since our beamforming processing has a minimal resolution of 0.22 °C.

5

Conclusion

In conclusion, this study shows that it is possible to retrieve accurately the deep ocean temperature using passively recorded hydroacoustic signals from submarine earthquakes. These signals were recorded at H10, a hydroacoustic array near Ascension Island, which consists of two triplet of hydrophones — whose separation is much larger than their respective aperture. The method used in this study consists of beamforming the six hydrophones together, as part of one large array. As beamforming algorithm, we use a time domain Fisher detector.

After beamforming one year of data, it results that less detections with an SNR larger than 0.6 are retrieved for the joint array than for each individual triplet. We attribute this difference to three reasons: (1) Ascension Island could block some signals propagating from one triplet to another; (2) the Fisher ratio is divided by the number of array elements in the beamforming algorithm; (3) the planar wave field back-azimuth can vary slightly between each element location, variation that would be larger for the joint array and thus leads to a larger decrease in SNR. In any case, for the joint array results, the quality of the retrieved apparent velocity is greatly increased as it appears much more constant across all back-azimuths; i.e. the velocity is much less dependent on the direction the plan wave propagated through the array, compare to the beamforming results of the individual triplets. Therefore, this method is less sensitive to local effects, such as the nearby bathymetry. Moreover, we see that well covered back-azimuths have very small 95% confidence intervals.

Afterwards, detections are divided into 5° back-azimuth bins and monthly average temperatures are derived from the apparent velocity for each back-azimuth bin. When analysing the four back-azimuth bins with the most temporal coverage, we retrieve an average temperature of 4.3 to 4.8 °C, between March 2005 and October 2013. Unfortunately, detections obtained from signals recorded after October 2013 could not lead to well-actionable results, because

of a malfunction in array element S1. A temperature of 4.3–4.8 °C is in good agreement with the literature, where values between 4.0 °C and 5.0 °C are given for a depth of 850 m and a latitude of -8°. Between 2005 and 2013, we may also see a temperature increase of 0.2 °C, however it is not statistically robust. The robustness could perhaps be enhance by using a smaller velocity increment (<1 m/s) in the construction of the slowness grid.

Finally, the research could be taken further by applying the method used in this study on other hydroacoustic stations of the IMS and retrieve the temperature. However, one limitation of this method is the fact that no major spatial obstacle (e.g. an island) must be situated between the two triplets of hydrophones. For instance, H03, near Robinson Crusoe Island is not appropriate whereas H11, near Wake Island in the North West Pacific Ocean, does offer the possibility of jointly processing the two triplets as the island does not create any significant shadow zone. A next step would also be to operate over larger distances in order to determine an average temperature at very large scale, and track its variation over time. Perhaps, cross-beam correlation techniques (cross-correlating array beams retrieved by array processing of the six hydrophones) could be conducted between different IMS stations.

Acknowledgements

The CTBTO station operators are thanked for the high-quality of the IMS data and products. The CTBTO data used for this study are available to member states but can be requested for academic purposes. Hydroacoustic data can be requested at the CTBTO International Data Center (IDC) in Vienna, via the virtual Data Exploration Center.

Moreover, I thank the managers of the International Seismological Centre (ISC) website.[1] for providing a platform from which catalogues of earthquakes can conveniently be extracted.

Finally, I thank Shahar Shani-Kadmiel and his colleagues for building and providing me with the beamforming algorithm used in this study.

[1] Url in bibliography.

Bibliography

Ball, J. S., Godin, O. A., Evers, L. G. & Lv, C. (2016). Long-range correlations of microseism-band pressure fluctuations in the ocean. *Geophys. J. Int.*, *206*, 825–834. https://doi.org/10.1093/gji/ggw110

Chen, X. & Tung, K. (2014). Varying planetary heat sink led to global-warming slowdown and acceleration. *Science*, *345*, 897–903. https://doi.org/10.1126/science.1254937

Del Grosso, V. A. (1974). New equation for the speed of sound in natural waters (with comparisons to other equations). *J. Acoust. Soc. Am.*, *56*, 1084–1091.

Dushaw, B. D., Worcester, P. F., Munk, W. H., Spindel, R. C., Mercer, J. A., Howe, M., Metzger Jr, K., Birdsall, T. G., Andrew, R. K., Dzieciuch, M. A., Cornuelle, B. D. & Menemenlis, D. (2009). A decade of acoustic thermometry in the North Pacific Ocean. *J. Geophys. Res.*, *114*, 2878–2884. https://doi.org/10.1029/2008JC005124

Evers, L. G. & Snellen, M. (2015). Passive probing of the sound fixing and ranging channel with hydro-acoustic observations from ridge earthquakes. *J. acoust. Soc. Am.*, *137*(4), 2124–2136. https://doi.org/10.1121/1.4916267

Evers, L. G., Wapenaar, K., Heaney, K. D. & Snellen, M. (2017). Deep ocean sound speed characteristics passively derived from the ambient acoustic noise field. *Geophys. J. Int.*, *210*(1), 27–33. https://doi.org/10.1093/gji/ggx061

Heaney, K. D., Prior, M. & L., C. R. (2017). Bathymetric diffraction of basin-scale hydroacoustic signals. *The Journal of the Acoustical Society of America*, *141*, 878–885. https://doi.org/10.1121/1.4976052

ISC. (n.d.). *Interactive bulletin search* [On International Seismological Centre website]. Retrieved June 17, 2020, from http://www.isc.ac.uk/iscbulletin/search/bulletin/interactive/

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., Zweng, M. M. & Johnson, D. R. (2010). *World Ocean Atlas 2009* (S. Levitus, Ed.; Vol. 1: Temperature). NOAA Atlas NESDIS 68, Washington, D.C., U.S. Government Printing Office.

Lukas, R. (2001). Freshening of the upper thermocline in the North Pacific subtropical gyre associated with decadal changes of rainfall. *Geophys. Res. Lett.*, *28*, 3485–3488. https://doi.org/10.1029/2001GL013116

Medwin, H. (1975). Speed of sound in water: A simple equation for realistic parameters. *J. Acoust. Soc. Am.*, *58*, 1318–1319.

Medwin, H. & Clay, C. S. (1997). *Fundamentals of acoustical oceanography.* San Diego, USA, Academic Press.

Melton, B. S. & Bailey, L. F. (1957). Multiple signal correlators. *Geophysics*, *22*(3), 565–588.

Munk, W. H., Worcester, P. F. & Wunsch, C. (1995). *Ocean Acoustic Tomography.* Cambridge, UK, Cambridge University Press.

Sambell, K. A. M., Smets, P. S. M., Simons, D. G., Snellen, M. & Evers, L. G. (2019). A study on the ambient noise field at a hydroacoustic array near Robinson Crusoe Island. *Geophys. J. Int.*, *218*, 88–99. https://doi.org/10.1093/gji/ggz124

Shani-Kadmiel, S. (2019). IDEA League proposal. TU Delft, Faculty of Civil Engineering and Geosciences, Delft, Netherlands.

Shani-Kadmiel, S., Assink, J. D., Smets, P. S. M. & Evers, L. G. (2018). Seismoacoustic Coupled Signals From Earthquakes in Central Italy: Epicentral and Secondary Sources of Infrasound. *Geophysical Research Letters*, *45*(1), 427–435. https://doi.org/10.1002/2017GL076125

Shani-Kadmiel, S., Smets, P. S. M., Averbuch, G., Den Ouden, O., Evers, L. G. & Assink, J. (2020). *pysabeam - a python toolkit for Seismo-Acoustic BEAMforming.* Retrieved April 3, 2020, from https://gitlab.com/shaharkadmiel (a closed repository at the moment).

Talley, L. D., Pickard, G. L., Emery, W. J. & Swift, J. H. (2011). *Descriptive physical oceanography: An introduction* (6th ed.). London, UK, Academic Press. [Chap. 3].

Woolfe, K. F., Lani, S., Sabra, K. G. & Kuperman, W. A. (2015). Monitoring deep-ocean temperatures using acoustic ambient noise. *Geophys. Res. Lett.*, *42*(8), 2878–2884. https://doi.org/10.1002/2015GL063438

Worcester, P. F. (2001). Tomography. In J. Steele & al. (Eds.), *Encyclopedia of Ocean Sciences* (pp. 2969–2986). London, UK, Academic.

Appendices

A

SOFAR channel in Arctic environment

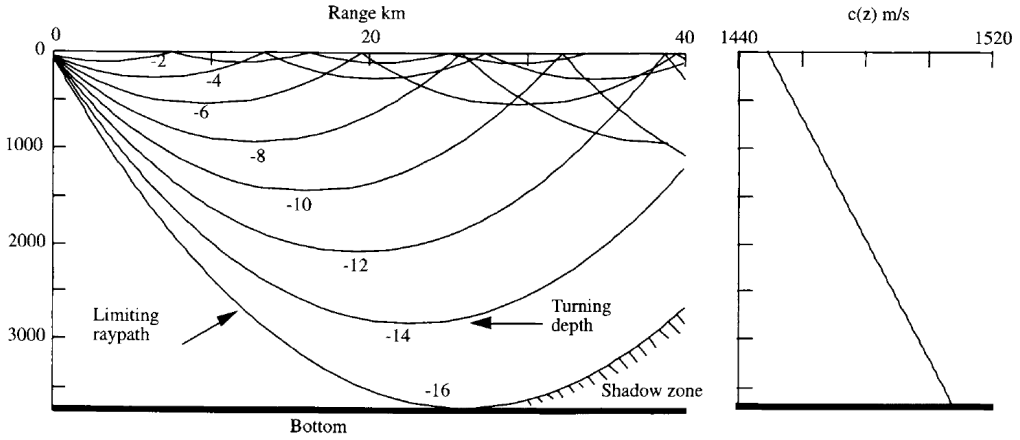

Figure A.1: *Ray diagram for a profile situated in an Arctic environment. The source is on the channel axis, at the surface. On the right is the sound velocity profile. The initial inclination angles are given with respect to the channel axis. Steeper rays (omitted here, for simplicity) reflect on the seafloor. From Medwin and Clay (1997).*

B

Coordinates of the geographic zone used for earthquake selection

Table B.1: *Geographic coordinates of the points used to build the polygon of figure 3.2, used for earthquake selection.*

Point	Latitude (°)	Longitude (°)
1	-68.339	-53.315
2	-67.748	39.674
3	-36.731	35.543
4	-36.327	18.755
5	-18.499	10.767
6	-10.954	11.822
7	2.704	7.340
8	4.546	2.682
9	3.406	-9.008
10	12.278	-18.676
11	25.216	-17.042
12	30.068	-10.934
13	33.100	-10.088
14	35.494	-6.992
15	36.648	-9.831
16	43.900	-9.945
17	44.745	-3.386
18	49.997	-6.699
19	51.762	-12.033
20	63.004	-3.946
21	63.004	-36.993
22	52.708	-54.396
23	46.873	-50.792
24	31.219	-79.884
25	23.253	-72.765
26	18.497	-61.076
27	10.592	-59.054
28	-5.772	-32.862
29	-21.844	-39.103
30	-45.938	-64.415
31	-60.094	-62.218
32	-62.763	-53.341
33	-68.339	-53.315

C

The geometry of the problem

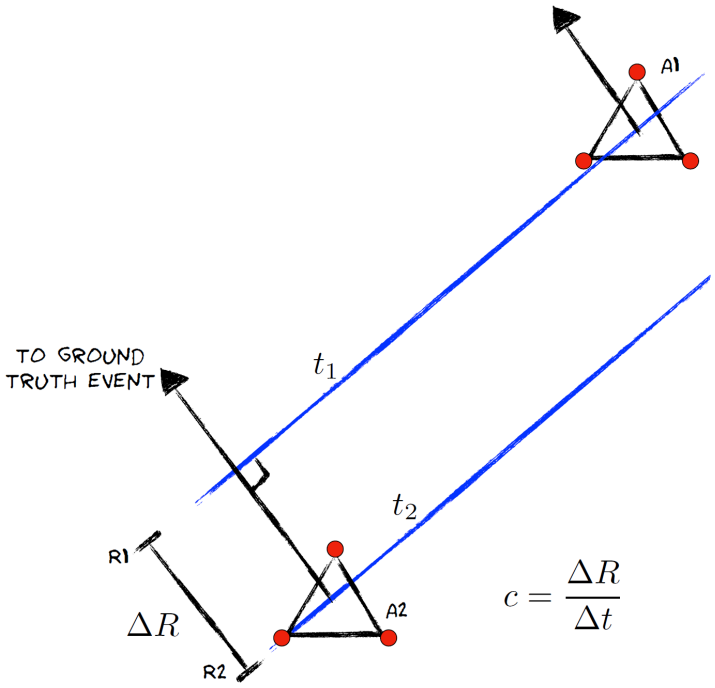

Figure C.1: *A1 and A2 are two arrays and the distance between their centres is measured along the propagation path of the planar wavefront. From Shani-Kadmiel (2019).*

D

Beamforming results for an earthquake per magnitude and per array

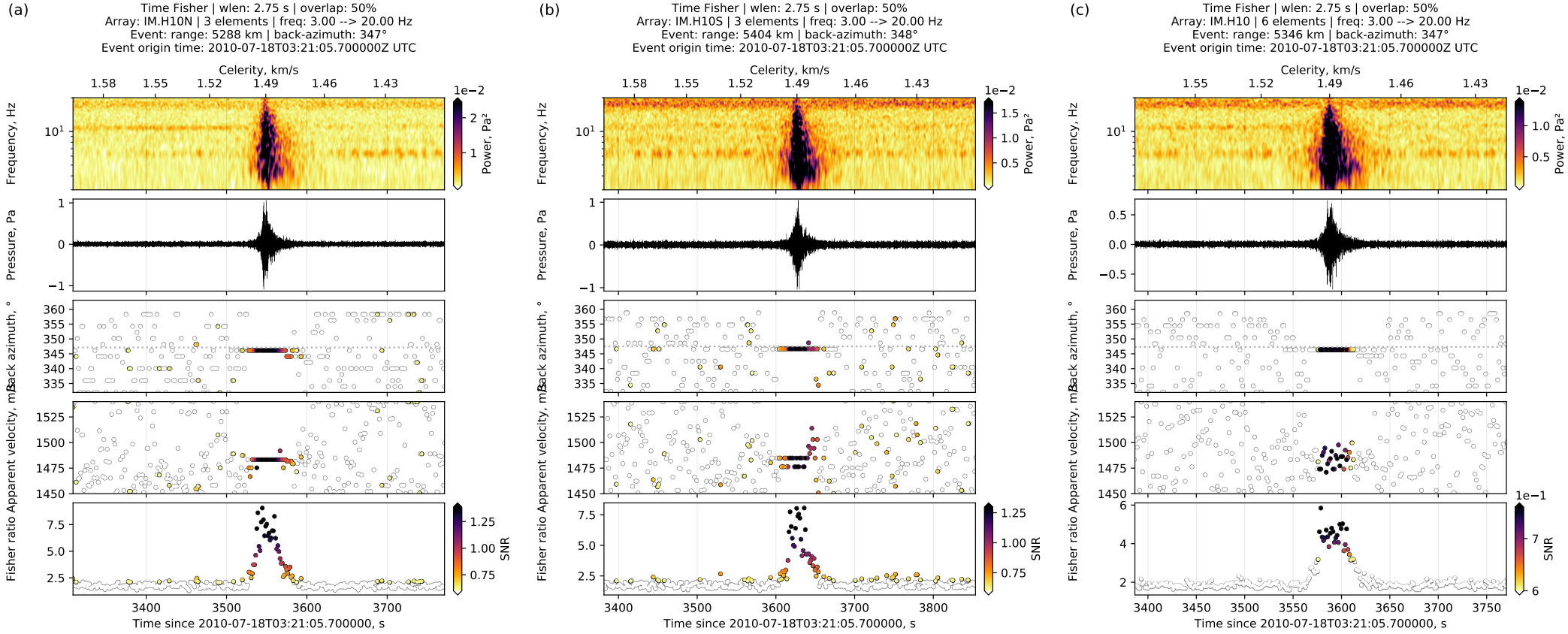

Figure D.1: *Beamforming results of a randomly selected $4 < M_W < 5$ earthquake at (a) the North triplet, (b) the South triplet and (c) the joint array. From top to bottom, frames show: the frequency spectrum, the best beam, the back-azimuth, the apparent velocity and the F-ratio. For the last three, one data point represents one detection. Figures are displayed with a 0.6 SNR threshold (white data point when below).*

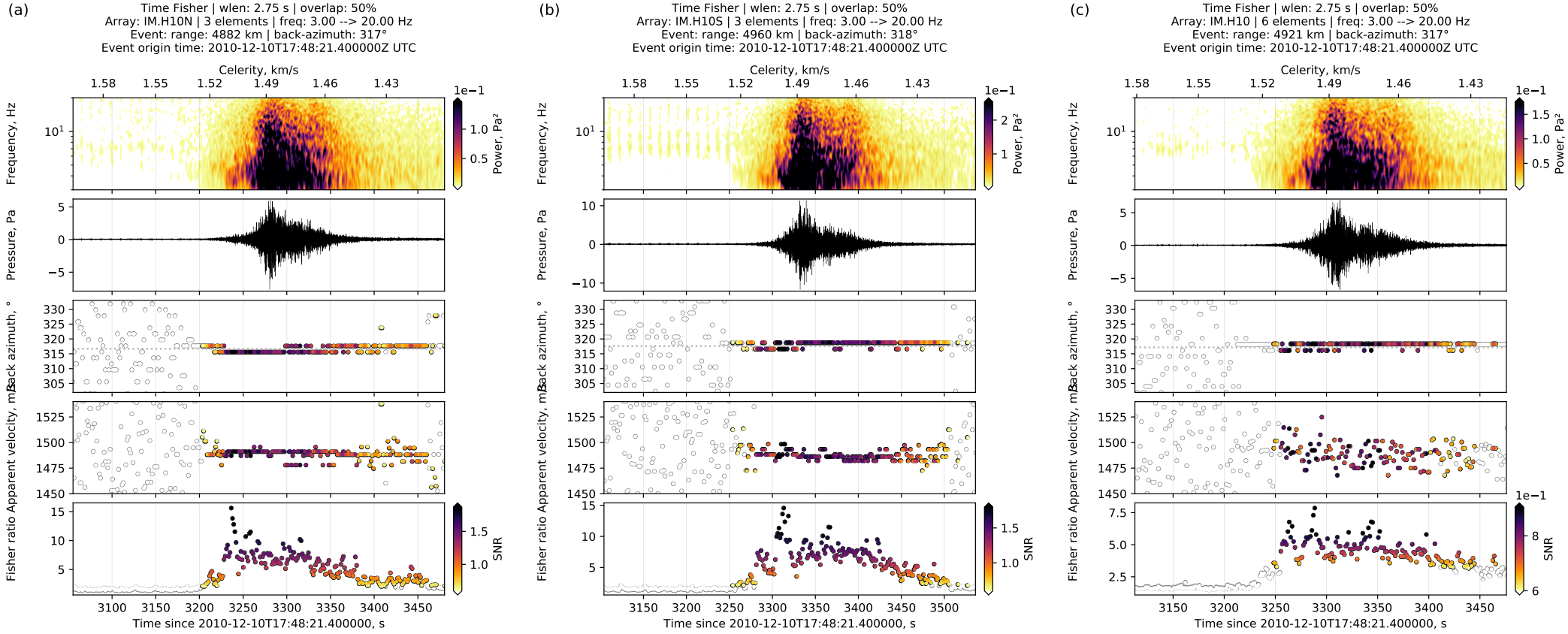

Figure D.2: *Beamforming results of a randomly selected $5 < M_W < 6$ earthquake at (a) the North triplet, (b) the South triplet and (c) the joint array. From top to bottom, frames show: the frequency spectrum, the best beam, the back-azimuth, the apparent velocity and the F-ratio. For the last three, one data point represents one detection. Figures are displayed with a 0.6 SNR threshold (white data point when below).*

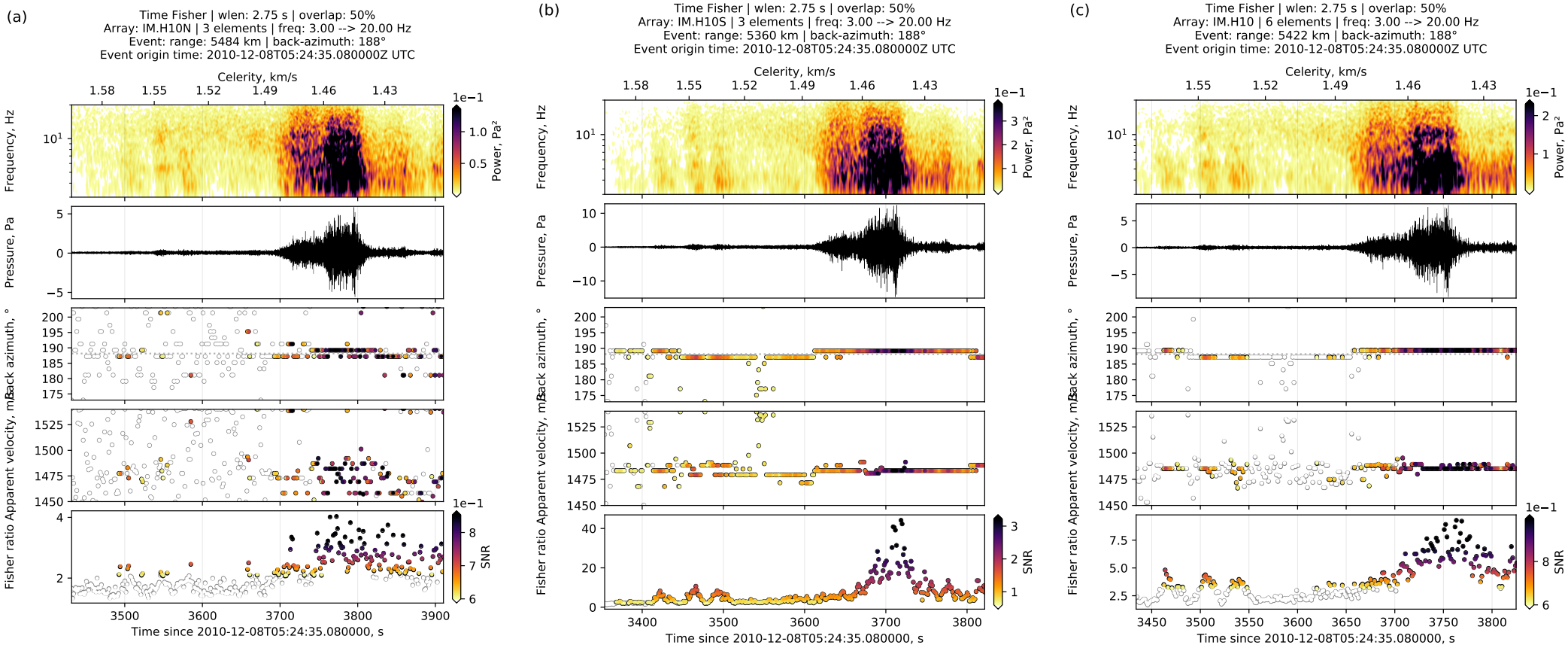

Figure D.3: *Beamforming results of a randomly selected $6 < M_W < 7$ earthquake at (a) the North triplet, (b) the South triplet and (c) the joint array. From top to bottom, frames show: the frequency spectrum, the best beam, the back-azimuth, the apparent velocity and the F-ratio. For the last three, one data point represents one detection. Figures are displayed with a 0.6 SNR threshold (white data point when below).*

E

Change in the results due to the malfunction of array element S1

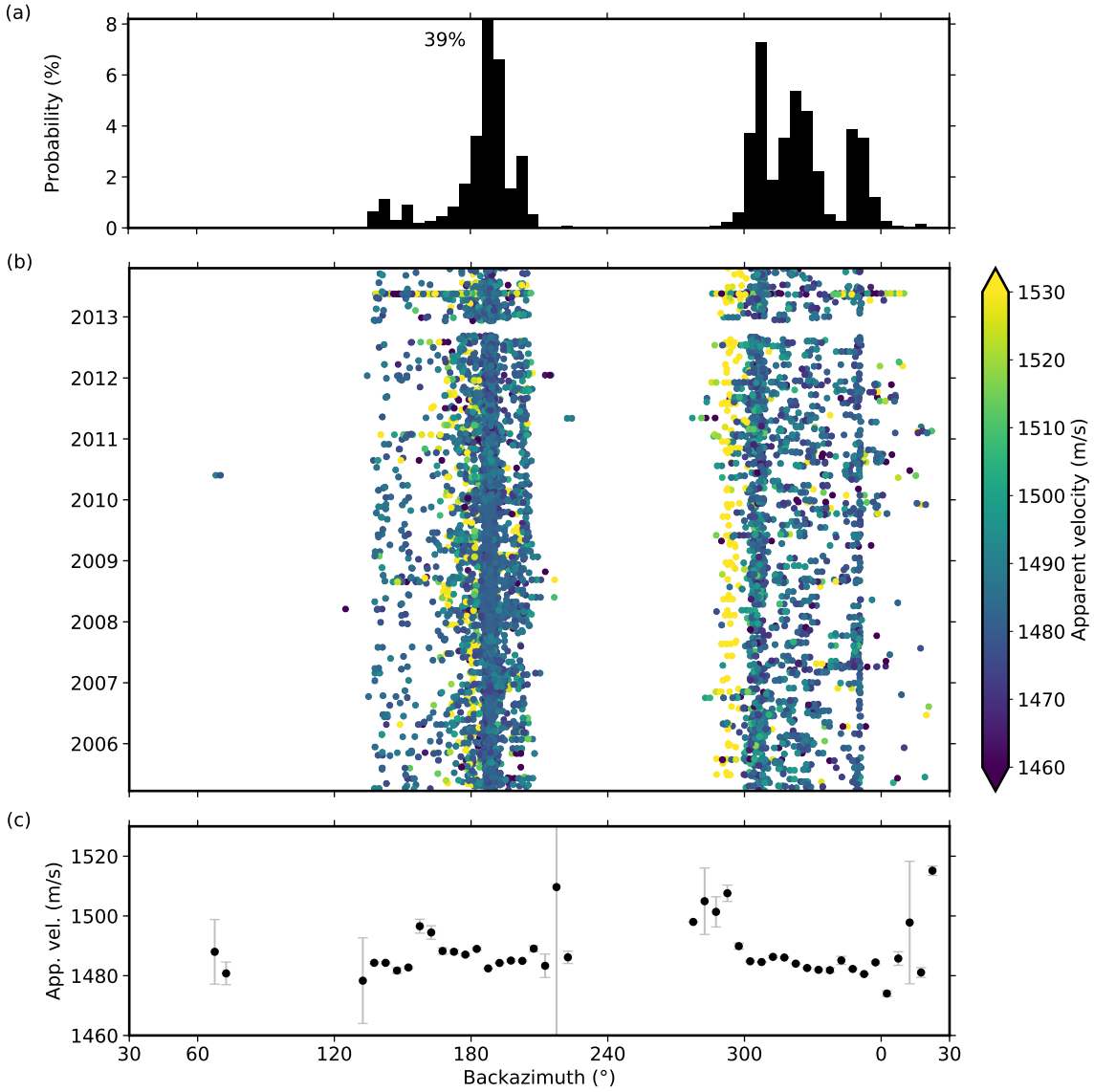

Figure E.1: *Detections at H10 (joint array) with SNR > 0.6 between 2005 and October 2013, i.e. until the malfunction of array element S1. (a) Probability distribution per 5° back-azimuth bin. (b) Back-azimuth as a function of time, colour coded by apparent velocity. (c) Average apparent velocity and 95% confidence interval per back-azimuth bin.*

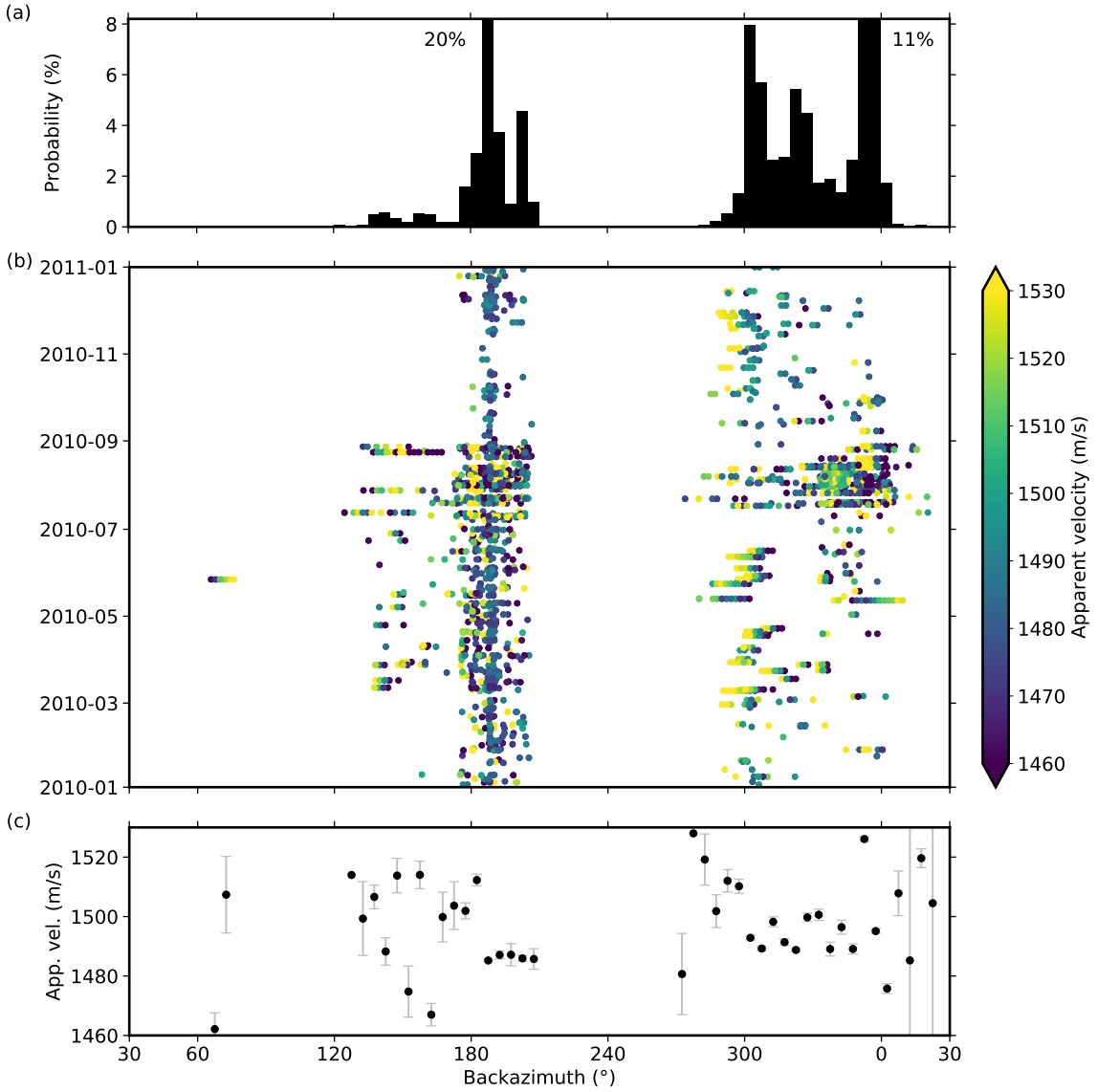

Figure E.2: *Detections at H10 (joint array) with SNR > 0.6 for 2010, and when excluding array element S1 from the beamforming process. (a) Probability distribution per 5° back-azimuth bin. (b) Back-azimuth as a function of time, colour coded by apparent velocity. (c) Average apparent velocity and 95% confidence interval per back-azimuth bin.*

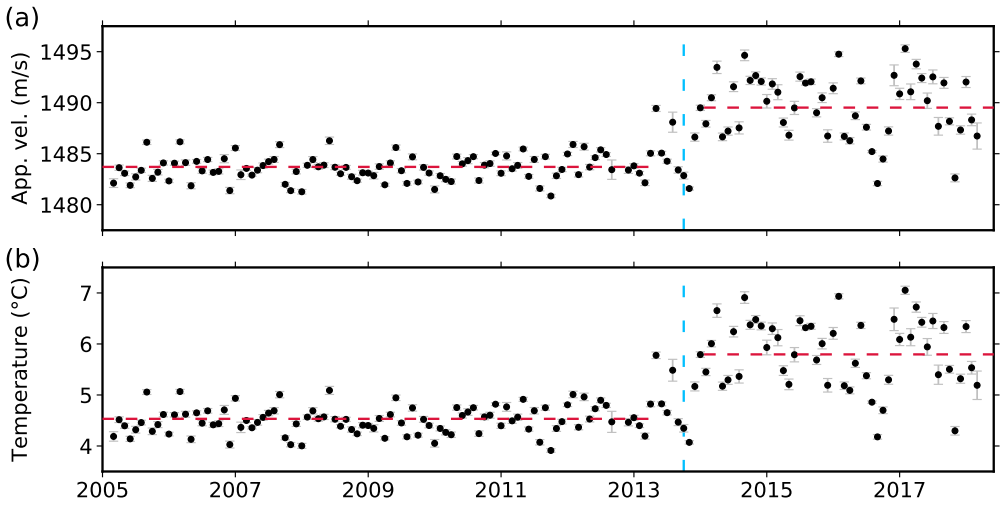

Figure E.3: *Monthly (a) apparent velocity and (b) temperature averaged across all back-azimuth bins as a function of time, at H10. Confidence intervals of 95% are also shown. When averaging, no weighting is applied regarding the number of detections per back-azimuth bin. The dash blue line shows the malfunction of array element S1, and the two dashed red lines show the average before and after.*

F

Python codes

F.1 Data processing/Beamforming

```
## GET PACKAGES AND FUNCTIONS ##
from obspy import read_events
from obspy import UTCDateTime
from obspy.geodetics.base import gps2dist_azimuth
from obspy.core.event import Catalog
from pysabeam import Array, Grid, array_process, read

## CALL EVENTS AND ARRAYS ##
events = read_events('Results/2010/Atlantic_2010_new.xml')
path = '/vardim/home/kadmiel/Data/Hydro'

array_N = Array('H10N.xml')
array_S = Array('H10S.xml')
array_NS = Array('H10.xml')

# Uncomment to exclude element S1 from array geometry
#array_S.exclude(elements='S1', remove_completely=True)
#array_NS.exclude(elements='S1', remove_completely=True)

## DEFINE PARAMETERS ##
# For waveform extraction
sofar_vmin = 1400
sofar_vmax = 1600
freqmin = 3
freqmax = 20

# For slowness grid construction
app_vel_min = 1450
app_vel_max = 1540
app_vel_increment = 1
theta_padding = 15
theta_increment = 2
```

```
33
34 # For beamforming
35 window_length = 2.75
36 overlap = 0.5
37 threads = 4
38
39 ## BEAMFORMING ##
40 bad_events = []
41 bad_events_2 = []
42
43 for event in events:
44
45     # Look for events info
46     lon = event.origins[0].longitude
47     lat = event.origins[0].latitude
48     origintime = event.origins[0].time
49     mag = event.magnitudes[0].mag
50
51     for array in [array_NS, array_N, array_S]:
52
53         try:
54             distance, az, baz = gps2dist_azimuth(lat, lon, array.
                   center[1], array.center[0])
55             if distance < 8 * array.aperture:
56                 break
57
58         except Exception as e_2:
59             print(e_2)
60             bad_events_2.append(event)
61             array.reset_exclude()
62             continue
63
64         # Get waveforms
65         starttime = origintime + distance / sofar_vmax
66         endtime = origintime + distance / sofar_vmin
67
68         # Construct slowness grid
69         grid = Grid(app_vel_params=(app_vel_min, app_vel_max,
                   app_vel_increment),
70                     theta_params=(baz-theta_padding, baz+theta_padding
                           , theta_increment))
71
72         try:
73             # Beamforming
74             result = array_process(
75                 array, path, grid=grid,
76                 starttime=starttime, endtime=endtime,
77                 event=event, wlen=window_length, overlap=overlap,
78                 filter_params=dict(type='bandpass', freqmin=freqmin,
                       freqmax=freqmax, corners=4, zerophase=True),
79                 method='timefisher',
80                 threads=threads
81                 #exclude='S1' # Uncomment to exclude element S1 from
                       beamforming
82             )
83
84         except Exception as e:
```

```
            print(e)
            bad_events.append(event)
            array.reset_exclude()
            continue

        # Write to disk
        result.force_to_netcdf(
            f'Results/{origintime.year}/Beamformed_events/{origintime.
                strftime("%Y%m%d%H%M%S")}_M{mag}_{array.name}.nc'
        )

# Store unbeamformed events which rose an error
catalog = Catalog(bad_events)
catalog.write('Results/2010/Bad_events_atlantic_2010.xml', format='
    QUAKEML')
catalog = Catalog(bad_events_2)
catalog.write('Results/2010/Bad_events_2_atlantic_2010.xml', format='
    QUAKEML')
```

F.2 Post-processing

```
## GET PACKAGES AND FUNCTIONS ##
import numpy as np
import math
import matplotlib.pyplot as plt
import datetime
import xarray as xr
import pandas as pd
from pysabeam import Array, Grid, array_process, read
from scipy.stats import sem, t
from glob import glob
import warnings
warnings.filterwarnings("ignore", category=RuntimeWarning)

## DEFINE TWO FUNCTIONS ##
def confidence_interval(data, confidence=0.95):
    """
    Return the mean and the one sided confidance interval of the ``
        data``.
    """
    h = sem(data) * t.ppf((1 + confidence) / 2, data.size - 1)
    return data.mean(), h

def c2T(c, z):
    """
    Calculate temperature in  C  from sound speed ``c``.
    """
    return (c - 1449.2 - 0.016 * z) / 4.6

## GET INITIAL DATASET ##
results_NS = 'Results/Common_years_results/Datasets/Dataset_H10.nc'  #
     Here, for joint array
array_NS = Array('H10.xml')
ds = read(results_NS)
```

```
ds = ds.where(ds.snr > 0.6, drop=True).sortby('time_detection')

# wrap baz by 30 degrees
wrap_degrees = 30
ds['baz'] = xr.where(ds.baz < wrap_degrees, ds.baz + 360, ds.baz)
# Adjust the time extent of the dataset with line below or else
#ds = ds.where(ds.time_detection < np.datetime64('2013-10-20T12
    :00:00.000000000'), drop=True)

## CALCULATE MEAN APP VEL AND CONF INT PER BAZ BIN ##
baz_step = 5  # 5 degrees bins
baz_bins = np.arange(0, 360 + baz_step, baz_step) + wrap_degrees
baz = baz_bins[:-1] + 0.5 * baz_step  # Label at center of bin

c_app_baz_groups = ds.app_vel.groupby_bins(ds.baz, baz_bins, right=
    False, labels=baz)

c_app_mean = xr.DataArray(np.nan * np.ones_like(baz), coords=[baz],
    dims=['baz'])
h = c_app_mean.copy()

for label, group in c_app_baz_groups:
    app_vel_, h_ = confidence_interval(group)
    c_app_mean.loc[dict(baz=label)] = app_vel_  # Loc is to populate
        an array
    h.loc[dict(baz=label)] = h_
c_app_baz_groups


## TO OVERVIEW THE DATA ##
from mpl_toolkits.axes_grid1 import make_axes_locatable
plt.rc('font', size=15)
plt.rcParams['axes.linewidth'] = 2
fig, ax = plt.subplots(3, sharex=True, figsize=(12, 12),
                       gridspec_kw=dict(height_ratios=[0.4, 1, 0.4]),
                       constrained_layout=True)
ax_histogram = ax[0]
ax_detections = ax[1]
ax_capp = ax[2]

# Histogram
# Use the same baz bins as before
values, bins, patches = ds.baz.plot.hist(
    bins=baz_bins, weights=(100 * np.ones_like(ds.baz) / ds.baz.size),
    color='k', align='mid', ax=ax_histogram
)
ax_histogram.set_title('')
ax_histogram.set_ylim(0,8.2)
ax_histogram.text(  # To annotate out of scope bar
    bins[values.argmax()]-18, 0.19 * values.max(), f'{values.max():.0f
        }%',
    ha='center',
)
#ax_histogram.annotate('13%', (0.48,0.885), xycoords='axes fraction')
ax_histogram.set_xlabel('')
ax_histogram.set_ylabel('Probability (%)')

```

```
84  # Detections
85  vmin, vmax = 1460, 1530
86  sorted_ds = ds.sortby('snr')
87  pc = sorted_ds.plot.scatter(
88      x='baz', y='time_detection', hue='app_vel', s=20, rasterized=True,
89      vmin=vmin, vmax=vmax, ax=ax_detections
90  )
91  cb = fig.axes[-1]
92  cb.set_ylabel('Apparent velocity (m/s)')
93  ax_detections.set_xlabel('')
94  ax_detections.set_ylabel('')
95  ax_detections.set_ylim(datetime.datetime(2005,3,24,0,0,0,0), datetime.
        datetime(2018,3,10,0,0,0,0))
96
97  # Apparent velocity mean and confidance interval
98  ax_capp.errorbar(
99      baz, c_app_mean, yerr=h, fmt='none', ecolor='silver', capsize=4,
            capthick=1,
100     zorder=0
101 )
102 xr.plot.scatter(
103     c_app_mean.to_dataset(name='c_app_mean'), x='baz', y='c_app_mean',
104     s=30, c='k', zorder=1, ax=ax_capp
105 )
106 ax_capp.set_ylim(vmin, vmax)
107
108 ax_capp.set_xlabel('Backazimuth (  )')
109 ax_capp.set_ylabel('App. vel. (m/s)')
110 subscripts = ['(a)', '(b)', '(c)']
111 for axi, subscript in zip(ax, subscripts):
112     axi.yaxis.set_label_coords(-0.08, 0.5)
113     axi.tick_params(top=True, right=True, length=5, width=1)
114     axi.set_xticks(ticks=[30, 60, 120, 180, 240, 300, 360, 390])
115     axi.set_xticklabels(labels=['30', '60', '120', '180', '240', '300'
            , '0', '30'])
116     axi.set_xlim([30, 390])
117     axi.annotate(subscript, (-0.15,1), xycoords='axes fraction')
118
119 #fig.savefig(f'Results/Common_years_results/Figures/H10_2005_to_2018.
        pdf', bbox_inches='tight', dpi=300)
120
121
122 ## GENERATE A MONTHLY TIME VECTOR ##
123 # Dataset ranges from 2005 to 2018
124 time = pd.date_range(
125     pd.to_datetime(ds.time_detection[0].data).date(),
126     pd.to_datetime(ds.time_detection[-1].data).date(),
127     freq='1MS',
128     name='time'
129 )
130 # Initialize a dataset to hold the mean monthly c_app per baz bin
131 monthly_means = xr.Dataset()
132 monthly_means['app_vel'] = xr.DataArray(
133     np.nan * np.ones((time.size, baz.size)), coords=[time, baz], dims
            =['time', 'baz']
134 )
135 monthly_means['h'] = monthly_means.app_vel.copy()
```

```
136 monthly_means['T'] = monthly_means.app_vel.copy()
137 monthly_means['h_T'] = monthly_means.app_vel.copy()
138
139 for i in range(time.size - 1):
140     month = ds.sel(time_detection=slice(time[i], time[i + 1]), drop=
        True)
141     if month.time_detection.size == 0: # In case there is no detection
         in month bin
142         continue
143     for label, group in month.app_vel.groupby_bins(
144         month.baz, baz_bins, right=False, labels=baz
145     ):
146         if group.size < 2: # Because if only 1 or 0 values,
            calculating the mean/conf int don't make sense
147             continue
148         app_vel_, h_ = confidence_interval(group)
149         monthly_means['app_vel'].loc[dict(time=time[i], baz=label)] =
            app_vel_ # Populate the 2 sheets
150         monthly_means['h'].loc[dict(time=time[i], baz=label)] = h_
151
152         # Add temperatures
153         T_, h_T_ = confidence_interval(c2T(group, 850))
154         monthly_means['T'].loc[dict(time=time[i], baz=label)] = T_
155         monthly_means['h_T'].loc[dict(time=time[i], baz=label)] = h_T_
156
157 # Same but dataset ranges from 2005 to 2013 #
158 time = pd.date_range(
159     pd.to_datetime(ds.time_detection[0].data).date(),
160     '2013-10-20',
161     freq='1MS',
162     name='time'
163 )
164
165 # Initialize a dataset to hold the mean monthly c_app per baz bin
166 monthly_means2 = xr.Dataset()
167 monthly_means2['app_vel'] = xr.DataArray(
168     np.nan * np.ones((time.size, baz.size)), coords=[time, baz], dims
        =['time', 'baz']
169 )
170 monthly_means2['h'] = monthly_means2.app_vel.copy()
171 monthly_means2['T'] = monthly_means2.app_vel.copy()
172 monthly_means2['h_T'] = monthly_means2.app_vel.copy()
173
174 for i in range(time.size - 1):
175     month = ds.sel(time_detection=slice(time[i], time[i + 1]), drop=
        True)
176     if month.time_detection.size == 0: # In case there is no detection
         in month bin
177         continue
178     for label, group in month.app_vel.groupby_bins(
179         month.baz, baz_bins, right=False, labels=baz
180     ):
181         if group.size < 6: # Because if only 1 or 0 values,
            calculating the mean/conf int don't make sense
182             continue
183         app_vel_, h_ = confidence_interval(group)
184         monthly_means2['app_vel'].loc[dict(time=time[i], baz=label)] =
```

```
                app_vel_ # Populate the 2 sheets
185         monthly_means2['h'].loc[dict(time=time[i], baz=label)] = h_
186
187         # Add temperatures
188         T_, h_T_ = confidence_interval(c2T(group, 850))
189         monthly_means2['T'].loc[dict(time=time[i], baz=label)] = T_
190         monthly_means2['h_T'].loc[dict(time=time[i], baz=label)] =
                h_T_
191
192
193 ## GENERATE A VELOCITY GRAPH PER MONTHLY-BACKAZIMUTH BIN ##
194 # 3D histogram
195 fig = plt.figure(figsize=(12,8))
196 monthly_means.app_vel.plot(vmin=vmin, vmax=vmax, rasterized=True)
197 ax = plt.gca()
198 plt.rc('font', size=15)
199 plt.rcParams['axes.linewidth'] = 2
200
201 cb = fig.axes[-1]
202 cb.set_ylabel('Apparent velocity (m/s)')
203 ax.set_xlabel('Backazimuth (  )')
204 ax.set_ylabel('')
205 ax.tick_params(top=True, right=True, length=5, width=1)
206 ax.set_xticks(ticks=[30, 60, 120, 180, 240, 300, 360, 390])
207 ax.set_xticklabels(labels=['30', '60', '120', '180', '240', '300', '0'
        , '30'])
208 ax.set_xlim([30, 390])
209
210
211 ## GENERATE A TEMPERATURE GRAPH PER MONTHLY-BACKAZIMUTH BIN ##
212 # 3D histogram
213 fig = plt.figure(figsize=(12,8))
214 T_mean = (monthly_means.T - monthly_means.T.mean(axis=0))
215
216 T_mean.plot(
217     cmap='RdBu_r', center=0,
218     vmin=-3, vmax=+3,
219     rasterized=True
220 )
221 ax = plt.gca()
222 plt.rc('font', size=15)
223 plt.rcParams['axes.linewidth'] = 2
224
225 cb = fig.axes[-1]
226 cb.set_ylabel('Averaged monthly-backazimuth bin temperature\n \u2012
        averaged backazimuth bin temperature (  C )')
227 ax.set_xlabel('Backazimuth (  )')
228 ax.set_ylabel('')
229 ax.tick_params(top=True, right=True, length=5, width=1)
230 ax.set_xticks(ticks=[30, 60, 120, 180, 240, 300, 360, 390])
231 ax.set_xticklabels(labels=['30', '60', '120', '180', '240', '300', '0'
        , '30'])
232 ax.set_xlim([30, 390])
233
234 fig.savefig(f'Results/Common_years_results/Figures/Temp_monthly_bins.
        pdf', bbox_inches='tight', dpi=300)
235
```

```
236
237 ## SORT THE DATASET IN TERM OF MONTHLY COVERAGE ##
238 # This index sorts the monthly means dataset by time coverage per baz
        bin
239 _sort_by_time_coverage_ = monthly_means.app_vel.count('time').argsort
        ().data
240 # For example the baz with the best coverage is 187.5:
241 monthly_means.sel(baz=monthly_means.baz[_sort_by_time_coverage_[-1]])
242
243
244 ## TRACKING OF THE TEMPERATURE - POLYNOMIAL REGRESSION ##
245 # Figure settings
246 fig = plt.figure(figsize=(10,15))
247 gs = fig.add_gridspec(4, 2)
248 ax1 = fig.add_subplot(gs[0, 0])
249 ax2 = fig.add_subplot(gs[1, 0])
250 ax3 = fig.add_subplot(gs[0, 1])
251 ax4 = fig.add_subplot(gs[1, 1])
252 fig2 = plt.figure(figsize=(10,15))
253 gs2 = fig.add_gridspec(4, 2)
254 ax5 = fig2.add_subplot(gs2[2, 0])
255 ax6 = fig2.add_subplot(gs2[3, 0])
256 ax7 = fig2.add_subplot(gs2[2, 1])
257 ax8 = fig2.add_subplot(gs2[3, 1])
258
259 poly_deg = 2 # Degree of polynomial regression
260
261 axs1 = [ax1, ax3, ax5, ax7]
262 axs2 = [ax2, ax4, ax6, ax8]
263 baaz = -1
264
265 for axi, axj in zip(axs1, axs2):
266     mm = monthly_means.sel(baz=monthly_means.baz[
            _sort_by_time_coverage_[baaz]])
267     mm2 = xr.Dataset.dropna(mm, dim='time', how='any')  # Drop NaN
            otherwise fitting fails
268     mm2 = mm2.where(mm2.h > 0.1, drop=True)             # Remove zero
            values otherwise polyfit weighting fails
269
270     mm_shorter = monthly_means2.sel(baz=monthly_means2.baz[
            _sort_by_time_coverage_[baaz]])
271     mm2_shorter = xr.Dataset.dropna(mm_shorter, dim='time', how='any')
272     mm2_shorter = mm2_shorter.where(mm2_shorter.h > 0.1, drop=True)
273
274     # App vel
275     axi.errorbar(x=pd.to_datetime(mm2.time.data), y=mm2.app_vel, yerr=
            mm2.h, linestyle='None', color='k', marker='.', ms=3, ecolor='
            silver', capsize=2, capthick=1, elinewidth=1, rasterized=True)
276     params = np.polyfit(mm2.time.astype('datetime64[D]').astype(int),
            mm2.app_vel, poly_deg)
277     axi.plot(pd.to_datetime(mm2.time.data), np.polyval(params, mm2.
            time.astype('datetime64[D]').astype(int)), color='dodgerblue',
             linestyle='--', linewidth=2, rasterized=True)
278     params = np.polyfit(mm2.time.astype('datetime64[D]').astype(int),
            mm2.app_vel, poly_deg, w=1/mm2.h)
279     axi.plot(pd.to_datetime(mm2.time.data), np.polyval(params, mm2.
            time.astype('datetime64[D]').astype(int)), color='dodgerblue',
```

```
            linewidth=2, rasterized=True)
280
281         params = np.polyfit(mm2_shorter.time.astype('datetime64[D]').
                astype(int), mm2_shorter.app_vel, poly_deg)
282         axi.plot(pd.to_datetime(mm2_shorter.time.data), np.polyval(params,
                 mm2_shorter.time.astype('datetime64[D]').astype(int)), color=
                'limegreen', linestyle='--', linewidth=2, rasterized=True)
283         params = np.polyfit(mm2_shorter.time.astype('datetime64[D]').
                astype(int), mm2_shorter.app_vel, poly_deg, w=1/mm2_shorter.h)
284         axi.plot(pd.to_datetime(mm2_shorter.time.data), np.polyval(params,
                 mm2_shorter.time.astype('datetime64[D]').astype(int)), color=
                'limegreen', linewidth=2, rasterized=True)
285
286         # Temp
287         axj.errorbar(x=pd.to_datetime(mm2.time.data), y=mm2.T, yerr= mm2.
                h_T, linestyle='None', color='k', marker='.', ms=3, ecolor='
                silver', capsize=2, capthick=1, elinewidth=1, rasterized=True)
288         params = np.polyfit(mm2.time.astype('datetime64[D]').astype(int),
                mm2.T, poly_deg)
289         axj.plot(pd.to_datetime(mm2.time.data), np.polyval(params, mm2.
                time.astype('datetime64[D]').astype(int)), color='dodgerblue',
                 linestyle='--', linewidth=2, rasterized=True)
290         params = np.polyfit(mm2.time.astype('datetime64[D]').astype(int),
                mm2.T, poly_deg, w=1/mm2.h_T)
291         axj.plot(pd.to_datetime(mm2.time.data), np.polyval(params, mm2.
                time.astype('datetime64[D]').astype(int)), color='dodgerblue',
                 linewidth=2, rasterized=True)
292
293         params = np.polyfit(mm2_shorter.time.astype('datetime64[D]').
                astype(int), mm2_shorter.T, poly_deg)
294         axj.plot(pd.to_datetime(mm2_shorter.time.data), np.polyval(params,
                 mm2_shorter.time.astype('datetime64[D]').astype(int)), color=
                'limegreen', linestyle='--', linewidth=2, rasterized=True)
295         params = np.polyfit(mm2_shorter.time.astype('datetime64[D]').
                astype(int), mm2_shorter.T, poly_deg, w=1/mm2_shorter.h_T)
296         axj.plot(pd.to_datetime(mm2_shorter.time.data), np.polyval(params,
                 mm2_shorter.time.astype('datetime64[D]').astype(int)), color=
                'limegreen', linewidth=2, rasterized=True)
297
298         baaz -= 1
299
300 plt.rcParams['axes.linewidth'] = 1
301
302 for ax in [ax1, ax2, ax3, ax4, ax5, ax6, ax7, ax8]:
303     ax.set_xticklabels(labels=[''])
304     ax.set_yticklabels(labels=[''])
305     ax.tick_params(top=True, right=True, length=6, width=1)
306 for ax, subscript in zip([ax1, ax3, ax5, ax7], ['(a)', '(b)', '(c)', '
        (d)']):
307     ax.set_ylim([1472,1495])
308     ax.annotate(subscript, (-0.15,1.1), xycoords='axes fraction')
309 for ax, subscript in zip([ax2, ax4, ax6, ax8], ['(a)', '(b)', '(c)', '
        (d)']):
310     ax.set_ylim([2,7])
311     ax.annotate(subscript, (-0.15,1.05), xycoords='axes fraction')
312 for ax in [ax1, ax2, ax5, ax6]:
313     ax.yaxis.set_label_coords(-0.25, 0.5)
```

```
314 for ax in [ax2, ax4, ax6, ax8]:
315     ax.set_xticklabels(labels=['04', '2006', '08', '10', '12', '14', '
        16', '18'])  # Added also 2004 due to a bug
316
317 ax1.set_title('Backazimuth bin: 187.5   \n')
318 ax3.set_title('Backazimuth bin: 192.5   \n')
319 ax5.set_title('Backazimuth bin: 307.5   \n')
320 ax7.set_title('Backazimuth bin: 302.5   \n')
321 for ax in [ax1, ax5]:
322     ax.set_ylabel('Apparent velocity (m/s)')
323     ax.set_yticklabels(labels=['1470', '1475', '1480', '1485', '1490',
          '1495'])
324 for ax in [ax2, ax6]:
325     ax.set_ylabel('Temperature ( C )')
326     ax.set_yticklabels(labels=['2', '3', '4', '5', '6', '7'])
327
328 fig.savefig(f'Results/Common_years_results/Figures/
    Monthly_app_vel_and_temp_per_baz.pdf', bbox_inches='tight', dpi
    =300)
329 fig2.savefig(f'Results/Common_years_results/Figures/
    Monthly_app_vel_and_temp_per_baz2.pdf', bbox_inches='tight', dpi
    =300)
```