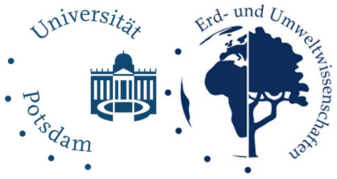
Institut fuer Erd- und Umweltwissenschaften
Geophysik

Automated seismic event location by waveform coherence analysis

Kumulative Dissertation
zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat)
in der Wissenschaftsdisziplin "Geophysik"

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultaet
der Universitaet Potsdam

von
Francesco Grigoli



Potsdam im November 2013

Published online at the
Institutional Repository of the University of Potsdam:
URL http://opus.kobv.de/ubp/volltexte/2014/7032/
URN urn:nbn:de:kobv:517-opus-70329
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-70329

Abstract

Automated location of seismic events is a very important task in microseismic monitoring operations as well for local and regional seismic monitoring. Since microseismic records are generally characterised by low signal-to-noise ratio, such methods are requested to be noise robust and sufficiently accurate. Most of the standard automated location routines are based on the automated picking, identification and association of the first arrivals of P and S waves and on the minimization of the residuals between theoretical and observed arrival times of the considered seismic phases. Although current methods can accurately pick P onsets, the automatic picking of the S onset is still problematic, especially when the P coda overlaps the S wave onset. In this thesis I developed a picking free automated method based on the Short-Term-Average/Long-Term-Average (STA/LTA) traces at different stations as observed data. I used the STA/LTA of several characteristic functions in order to increase the sensitiveness to the P wave and the S waves. For the P phases we use the STA/LTA traces of the vertical energy function, while for the S phases, we use the STA/LTA traces of the horizontal energy trace and then a more optimized characteristic function which is obtained using the principal component analysis technique. The orientation of the horizontal components can be retrieved by robust and linear approach of waveform comparison between stations within a network using seismic sources outside the network (chapter 2). To locate the seismic event, we scan the space of possible hypocentral locations and origin times, and stack the STA/LTA traces along the theoretical arrival time surface for both P and S phases. Iterating this procedure on a three-dimensional grid we retrieve a multidimensional matrix whose absolute maximum corresponds to the spatial and temporal coordinates of the seismic event. Location uncertainties are then estimated by perturbing the STA/LTA parameters (i.e the length of both long and short time windows) and relocating each event several times. In order to test the location method I firstly applied it to a set of 200 synthetic events. Then we applied it to two different real datasets. A first one related to mining induced microseismicity in a coal mine in the northern Germany (chapter 3). In this case we successfully located 391 microseismic event with magnitude range between 0.5 and 2.0 Ml. To further validate the location method I compared the retrieved locations with those obtained by manual picking procedure. The second dataset consist in a pilot application performed in the Campania-Lucania region (southern Italy) using a 33 stations seismic network (Irpinia Seismic Network) with an aperture of about

150 km (chapter 4). We located 196 crustal earthquakes (depth $<$ 20 km) with magnitude range $1.1 < Ml < 2.7$. A subset of these locations were compared with accurate locations retrieved by a manual location procedure based on the use of a double difference technique. In both cases results indicate good agreement with manual locations. Moreover, the waveform stacking location method results noise robust and performs better than classical location methods based on the automatic picking of the P and S waves first arrivals.

Zusammenfassung

Die automatische Lokalisierung von Mikroerdbeben durch die Wellenform Kohärenzanalyse
Die automatische Lokalisierung seismischer Ereignisse ist eine wichtige Aufgabe, sowohl im Bereich des Mikroseismischen Monitorings im Bergbau und von Untegrund Aktivitäten, wie auch für die lokale und regionale Überwachung von natürlichen Erdbeben.
Da mikroseismische Datensätze häufig ein schlechtes Signal-Rausch-Verhältnis haben müssen die Lokalisierungsmethoden robust gegen Rauschsignale und trotzdem hinreichend genau sein. Aufgrund der in der Regel sehr hochfrequent aufgezeichneten Messreihen und der dadurch sehr umfangreichen Datensätze sind automatische Auswertungen erstrebenswert. Solche Methoden benutzen in der Regel automatisch gepickte und den P und S Phasen zugeordnete Ersteinsätze und Minimieren die Summe der quadratischen Zeitdifferenz zwischen den beobachteten und theoretischen Einsatzzeiten. Obgleich das automatische Picken der P Phase in der Regel sehr genau möglich ist, hat man beim Picken der S Phasen häufig Probleme, z.B. wenn die Coda der P Phase sehr lang ist und in den Bereich der S Phase hineinreicht. In dieser Doktorarbeit wird eine Methode vollautomatische, Wellenform-basierte Lokalisierungsmethode entwickelt, die Funktionen des Verhältnisses "Short Term Average / Long Term Average" (STA/LTA) verwendet und keine Pickzeiten invertiert. Die STA/LTA charakteristische Funktion wurde für unterschiedliche Wellenform Attribute getestet, um die Empfindlichkeit für P und S Phasen zu erhöhen. Für die P Phase wird die STA/LTA Funktion für die Energie der Vertikalkomponente der Bodenbewegung benutzt, wohingegen für die S Phase entweder die Energie der horizontalen Partikelbewegung oder eine optimierte Funktion auf Basis der Eigenwertzerlegung benutzt wird. Um die Ereignisse zu lokalisieren wird eine Gittersuche über alle möglichen Untergrundlokalisierungen durchgeführt. Für jeden räumlichen und zeitlichen Gitterpunkt werden die charakteristischen Funktionen entlang der theoretischen Einsatzkurve aufsummiert. Als Ergebnis erhält man eine 4-dimensionale Matrix über Ort und Zeit des Ereignisses, deren Maxima die wahrscheinlichsten Lokalisierungen darstellen. Um die Unsicherheiten der Lokalisierung abzuschätzen wurden die Parameter der STA/LTA Funktionen willkürlich verändert und das Ereignis relokalisiert. Die Punktwolke aller möglichen Lokalisierungen gibt ein Maß für die Unsicherheit des Ergebnisses. Die neu entwickelte Methode wurde an einem synthetischen Datensatz von 200 Ereignissen getestet und für zwei beobachtete

Datensätze demonstriert. Der erste davon betrifft induzierte Seismizität in einem Kohlebergbau in Norddeutschland. Es wurden 391 Mikrobeben mit Magnituden zwischen Ml 0.5 und 2.0 erfolgreich lokalisiert und durch Vergleich mit manuell ausgewerteten Lokalisierungen verifziert.Der zweite Datensatz stammt von einem Anwednung auf des Regionale Überwachungsnetz in der Region Campania-Lucania (Süditalien) mit 33 seismischen Stationen und einer Apertur von etwa 150 km. Wir konnten 196 Erdbeben mit Tiefen ¡ 20 km und Magnituden zwischen Ml 1.1 und 2.7 lokalisieren. Eine Untergruppe der eigenen Lokalisierungen wurde mit den Lokalisierungen einer Standard Lokalisierung sowie einer hochgenauen Relativlokalisierung verglichen. In beiden Fällen ist die Übereinstimmung mit den manuellen Lokalisierungen gross. Außerdem finden wir, dass die Wellenform Summations Lokalisierung ronbust gegen Rauschen ist und bessere Ergebnisse liefert als die Standard Lokalisierung, die auf dem automatischen Picken von Ersteinsatzzeiten alleine basiert.

Acknowledgements

Now, at the end of this long path I would like to thank all people that helped me to grow scientifically but, more important, humanely. First of all I would like to thank my wife, for its patience and support. She is a fixed point in my chaotic life. A special thank is for my boss and friend Simone, for the very nice moments spent in working and laughing together, for his constant attitude to help me in everything (house moving included) and for his confidence in my capabilities. I would like to thank my supervisor Prof. Torsten Dahm, I always felt sure of his support, patience and confidence and his supervision could not have been better. I thank all my friend sharing my "adventure in Germany" with me: Eleonora, Sebastian, Lars (now in the kangaroo land) , Fabio (the new entry), Samira and Tolga. They created a pleasant working environment, where I felt like at home. Of course I cannot forget my friends and officemates (to be honest I squatted their office at GFZ) Francesco and Luigi, for the many dinners and the amazing moments spent together. Last but not least... (For Italian readers) ringrazio la mia famiglia (mamma, zia Mariolina, Franca, Enzo etc.) per essermi stata sempre vicina e per avermi spronato a dare sempre il massimo. Infine, ringrazio il mio scalcinato gruppo di amici Giuseppe, Emanule (ai quali sono legato da un rapporto fraterno), Nicola e Piero per la loro costante presenza nei momenti sia belli che brutti e per i nostri venerdì sera passati a giocare a Risiko. Grazie di cuore

Contents

Abstract **1**

Zusammenfassung **3**

Acknowledgements **5**

1 Introduction **14**

2 A complex linear least-squares method to derive relative and absolute orientations of seismic sensors **19**
2.1 Abstract . . . 20
2.2 Introduction . . . 20
2.3 Theory . . . 22
2.4 Synthetic tests . . . 26
2.5 Application to VSP data (Sudbury basin, Canada) . . . 28
2.6 Application to OBS data (Aegean sea, Greece) . . . 30
2.7 Application to a seismological array (Graefenberg, Germany) . . . 34
2.8 Discussion and Conclusions . . . 36
2.9 Acknowledgements . . . 39

3 Automated Seismic Event Location by Travel-Time Stacking: An Application to Mining Induced Seismicity **40**
3.1 Methodology . . . 42
3.2 Synthetic tests . . . 44
3.3 Application to real data . . . 49
3.4 Conclusions . . . 52
3.5 Acknowledgements . . . 55
3.6 Electronic supplement . . . 56

4 Automated seismic event location by waveform coherence analysis **59**
4.1 Abstract . . . 60
4.2 Introduction . . . 60

4.3 Methodology 62
4.4 Application to the ISNet data 65
4.5 Discussion and Conclusions 71
4.6 Acknowledgements 76
4.7 Appendix:Electronic supplement 77

5 Conclusions **85**

List of Figures

1.1 This sketch represents: a) The acquisition geometry and the source location (yellow star) 2) The recorded traces and 3) The STA/LTA traces of the recorded waveforms. 17
1.2 The waveform stacking location process for different source locations (indicated by yellow diamonds in panels a.X) and time steps (panels b, c, and d). 18
1.3 Coherence matrix $C(x, z)$ at the time $\hat{t}$, where the maximal coherence is detected. 18

2.1 Horizontal components of two non aligned sensors. A clockwise rotation of the sensor (j+1) by an angle $-\phi_j$ aligns it with the sensor j. 24
2.2 Sketch of a vertical borehole with a linear array of seismic sensors. Vertical components of all geophones are parallel to the case of the borehole, while the horizontal components are not aligned. Figure (a) represents the acquisition geometry referred to synthetic test while figure (b) the acquisition geometry referred to VSP data from Canada. 27
2.3 Synthetic data results. Hodograms of 4 sensors pairs: before (top) and after alignment (bottom). The red hodogram is related to the sensor we want align with the reference one (blue hodogram). 27
2.4 VSP data from Sudbury basin (Canada). The part of waveform used for alignment is highlighted in yellow. 29
2.5 VSP data from Sudbury basin (Canada). Average of the cross-correlation of the total energy traces between all adjacent sensors. The yellow time window corresponds with the part of the waveform containing the first arrival of the P-wave. 30
2.6 Field data results (VSP) showing hodograms for all sensors pairs. The red hodogram is related to the sensor $(j + 1)$ we want align with the reference one j (blue hodogram), with $j = 2, .., 9$. 31
2.7 Map of Santorini island (Greece) showing the network configuration. OBS stations are represented by black diamonds and land station by a white circle. 32
2.8 The Kuril island seismic event recorded by OBS 50, OBS 51 and land station SANT (time axis with respect to the event origin time). Yellow window highlight the part of the waveforms used to show alignment results (traces were filtered using a band-pass Butterworth filter within the band 0.03-0.07 Hz). 33

2.9 Results relative to the alignment of the OBSs and Land stations (Kuril event). Although we used the full waveform, for visualization purposes we show only the part of the waveforms relative to the yellow marked window in figure (2.8). Reference traces are blue while traces to rotate are red (traces were filtered using a band-pass Butterworth filter within the band 0.03-0.07 Hz). 34
2.10 Results relative to the alignment of the OBSs and Land stations (Vanuatu event). For visualization purposes we show only a part of the recorded waveform. Reference traces are blue while traces to rotate are red. Zero of the time axis corresponds to 01:20 GMT time (traces were filtered using a band-pass Butterworth filter within the band 0.03-0.07 Hz). 35
2.11 (a) Configuration of the network. (b) Complex traces (hodograms) related to the stations pairs ((blue) GRA3-GRA1 (red)) and ((blue) GRC3-GRC4 (red)), before (left) and after alignment (right). (c) Horizontal component traces of the same pairs of stations. Reference traces are blue while traces to rotate are red (traces were filtered using a low-pass Butterworth filter with corner frequency $f_c = 0.03$ Hz). 37

3.1 Flow diagram of the location algorithm. Coherence matrix XY is obtained by projecting, for each X-Y, its maximum along Z (coherence matrices XZ and YZ are obtained in a similar way). 45
3.2 Geometry of the network (a) and results (a,b,c) for synthetic data with different noise levels: True locations are represented by red dots, location results with data affected by noise are represented by blue (noise level at 10%), green (noise level at 30%) and yellow (noise level at 70%) spots. Subfigure (d) represents the layered velocity model we used to perform location. The reference point $(x, y) = (0, 0)$ corresponds to $x = 411617$ and $y = 5722111$ (in the UTM system) 46
3.3 Synthetic traces (Vertical component) with different noise levels: noise level at 10% of the maximum amplitude (top figure), noise level at 30% of the maximum amplitude (middle figure) and noise level at 70% of the maximum amplitude (bottom figure). 47
3.4 Coherence matrices related to one of the synthetic event shown previously: noise level at 10% (a), noise level at 30% (b) and noise level at 70% (c). Coherence matrix XY is obtained by projecting, for each X-Y, its maximum along Z (coherence matrices XZ and YZ are obtained in a similar way). Coherence values are represented in color scale. 48
3.5 Coherence values $C(\hat{x}, \hat{y}, \hat{z}, t)$ for different time steps t ($\hat{x}$, $\hat{y}$ and $\hat{z}$ are the estimated coordinates of the seismic event) considering noise levels at 10% (a), 30% (b) and 70% (c). Coherence maximum corresponding to the time step t_M is related to the origin time of the seismic event by $t_0 = t_M - \tau_{min}(\hat{x}, \hat{y}, \hat{z})$. 49
3.6 Three component velocity traces of a microseismic event occurred on October 4^{th}, 2006 at 13.50.30 GMT, recorded by four station of the network. 50

3.7 Vertical and horizontal STALTA traces (normalized) related to the waveforms shown in Figure 3.6. 50
3.8 Coherence matrices obtained by stacking the STA/LTA traces using: a) only P arrival times for an homogeneous velocity model, b) both P and S arrival times for an homogeneous velocity model, c) only P arrival times for a layered velocity model and d) both P and S arrival times for a layered velocity model. Coherence matrix XY is obtained by projecting, for each X-Y, its maximum along Z (coherence matrices XZ and YZ are obtained in a similar way). Coherence values are represented in color scale. (Coordinates are in the Gauss-Krieger system). 51
3.9 Geometry of the network (a) and location results (a,b,c) for all 391 seismic events (blue dots) and comparison with results obtained using a manual location procedure (red dots). Subfigure (d) represents a comparison between the homogenous velocity model used by the University of Bochum (red dashed lines) and the layered model we used to retrieve locations (blue lines). The reference point $(x, y) = (0, 0)$ corresponds to $x = 3411193$ and $y = 5723463$ (in the Gauss-Krieger system). 53
3.10 Solid line represents the percent of events located within a given distance from the reference locations (true locations for synthetics and manual ones for real data), while dashed line represents the percent of events with a given time shift (absolute value) respect to their origin times. Plot (a) shows the results for the real 391 seismic events, reference locations and origin times are based on the results obtained by the University of Bochum through a manual procedure. In analogy, plot (b) and (c) show the results related to the 200 synthetic events with noise level at 30% and 70% respectively. 54
3.11 his figure shows our location results (blue dots) for 200 synthetic events with noise levels at 30% (subfigures a.1, a.2 and a.3) and 70% (subigures b.1, b.2 and b.3). The true locations are represented by red dots. The reference point (x,y)=(0,0) corresponds to x=411617 and y=5722111 (in the UTM system). 57
3.12 Quality of the location results for the 200 sythetic events affected by a noise level at 30%. The plot on the top shows the percent of events located within a given distance from the true locations. On the bottom, an analogue plot shows the absolute time shift respect the true origin time (bottom figure) 58
3.13 Quality of the location results for the 200 sythetic events affected by a noise level at 70%. The plot on the top shows the percent of events located within a given distance from the true locations. On the bottom, an analogue plot shows the absolute time shift respect the origin time of manual locations (bottom figure) 58

4.1 Irpinia seismic network, short period stations are represented by triangles, broadband stations by squares and the blue circles represent three seismic events occurred in the region. Seismograms of these events recorded by the yellow marked stations are shown in figures 2-4. 66

4.2 Waveforms related to the Ml=2.7 seismic event occurred on May 27^{th} 2008 (figure 3.1) and recorded by the stations RSF3 (a) and CMP3 (b). We show the three component seismograms and the STA/LTA traces of three different characteristic functions: the horizontal energy trace (red line), the principal eigenvalue of the instantaneous covariance matrix (green line) and the P wave characteristic function based on the vertical energy trace (blue line). (for a more detailed figure see the electronic supplement) 68
4.3 Waveforms related to the Ml=1.7 seismic event occurred on October 23^{th} 2008 (figure 3.1) and recorded by the stations SCL3 (a) and CSG3 (b). We show the three component seismograms and the STA/LTA traces of three different characteristic functions: the horizontal energy trace (red line), the principal eigenvalue of the instantaneous covariance matrix (green line) and the P wave characteristic function based on the vertical energy trace (blue line).(for a more detailed figure see the electronic supplement) 69
4.4 Waveforms related to the Ml=1.1 seismic event occurred on April 24^{th} 2008 (figure 3.1) and recorded by the stations MNT3 (a) and COL3 (b). We show the three component seismograms and the STA/LTA traces of three different characteristic functions: the horizontal energy trace (red line), the principal eigenvalue of the instantaneous covariance matrix (green line) and the P wave characteristic function based on the vertical energy trace (blue line). (for a more detailed figure see the electronic supplement) 70
4.5 Three dimensional velocity model of the Campania-Lucania region (southern Italy) obtained by traveltime tomography. The figures on top show the P velocity model at different depths, in analogy figures at the bottom show depth slices of the S velocity model. Network stations are represented by triangles. 71
4.6 Coherence matrices related to the three seismic events shown in figure (4.1): (a) event 1 (b) event 2 (c) event 3. The coherence matrix XY is obtained by projection, for each X-Y, its maximum along Z (coherence matrices XZ and YZ are obtained in a similar way). Coherence values are represented in color scale. The reference point (X;Y)=(0;0) corresponds with the point (X;Y)=(493718;4458627) in the UTM coordinates system. 72
4.7 Comparison between our solutions (blue lines) and the locations obtained by automatic picking (red lines) with respect to manual locations (white circles) obtained using a double difference location algorithm. The reference point (X;Y)=(0;0) corresponds with the point (X;Y)=(510000;44480000) in the UTM coordinates system. The cross sections are perpendicular to the Northing and Easting axis. 73

4.8 The histogram plot (a) shows the number of events within a given distance from the manual locations. Blue bars are related to our automated locations, while red bars to the locations obtained by automatic picking. The cumulative plot (b) shows the percent of events located within a given distance from the manual locations. Also in this case, the blue line is related to the location obtained using our approach, while the red line is related to the location obtained using automated picking. For both plots, reference locations are based on the results obtained by [De Matteis et al., 2012; Stabile et al., 2012]. 74
4.9 Map view of about 196 seismic events located using our approach. Red lines represent the surface projections of three fault segments ruptured during the 1980 Irpinia earthquake. Cross sections are represented in the small plots. The reference point (X;Y)=(0;0) corresponds with the point (X;Y)=(475000;4455000) in the UTM coordinates system. 75
4.10 Zoomed version of the figure 2 (in the manuscript). Waveforms are related to the 2008-05-27 16:19:33 seismic event with Ml=2.7 and recorded by the stations RSF3 (a) and CMP3 (b). We show the three component seismograms and the STA/LTA traces of different characteristic functions. The red line represents the STA/LTA of the horizontal energy trace, while the green line represents the STA/LTA trace of the principal eigenvalue characteristic function. In the same plot the blue line is related to the STA/LTA trace of the P characteristic function based on the vertical energy trace. 77
4.11 Zoomed version of the figure 3 (in the manuscript). Waveforms related to the 2008-10-23 16:02:13 seismic event with Ml=1.7 and recorded by the stations SCL3 (a) and CSG3 (b). We show the three component seismograms and the STA/LTA traces of different characteristic functions. The red line represents the STA/LTA of the horizontal energy trace, while the green line represents the STA/LTA trace of the principal eigenvalue characteristic function. In the same plot the blue line is related to the STA/LTA trace of the P characteristic function based on the vertical energy trace. 78
4.12 Zoomed version of the figure 4 (in the manuscript). Waveforms related to the 2008-04-24 04:49:10 seismic event with Ml=1.1 and recorded by the stations MNT3 (a) and COL3 (b). We show the three component seismograms and the STA/LTA traces of different characteristic functions. The red line represents the STA/LTA of the horizontal energy trace, while the green line represents the STA/LTA trace of the principal eigenvalue characteristic function. In the same plot the blue line is related to the STA/LTA trace of the P characteristic function based on the vertical energy trace. 79

List of Tables

2.1 Synthetic test results: Comparison between true alignment angles and the estimated ones. Data+noise 2 refers to the results obtained using two explosions at different azimuths. Uncertainties were estimation using the bootstrap approach. 28

2.2 Synthetic test results: Comparison between true alignment angles and the estimated ones. Uncertainties were estimated via complex covariance matrix [Miller, 1973]. 28

2.3 Estimated alignment angles for VSP field data from Sudbury basin (Canada). 30

2.4 Estimated rotation angles for OBS data. OBS stations are aligned here with the land station SANT, Hensch [2009] used a different reference station (REF)). 32

2.5 Estimated rotation angles for Graefenberg array. Results in bold are related to the figure (2.11). 38

3.1 This table summarizes the location results related to the 4 synthetic events introduced in the manuscript. Here we show a comparison between the true locations and the estimated ones with different noise levels. Uncertainties were estimated through perturbations of the STALTA parameters and jack-knife method (Location coordinates are in UTM and in meters (m), while origin time is in seconds (s)). 56

4.1 Locations results related to 196 seismic events occurred in the Irpinia region between the February 2008 and March 2010. Since we used a grid spacing of 500 m, all uncertainties lower than this value are set to 500 m (Coordinates are provided in the UTM system). 84

Chapter 1

Introduction

Seismic networks designed for seismic monitoring purposes allows, nowadays, to detect microseismic events. A common definition of microseismicity in terms of magnitude ranges is not fully agreed and depends on different applications. The datasets used in this thesis are characterized by seismic events with magnitudes ranging from 0.5 to 2.7 Ml and the seismic networks consist of 15-33 stations and have an aperture ranging between 5 and 150 km. The analysis of the microsesmicity provide important informations on active processes in the subsurface. For example, the location of microseismic events can be use to estimate the distribution and orientation of active faults. A second important application concerns local scale tomography. Microseismic events can be either natural (e.g. tectonic or volcanic events) or induced. Human induced seismic events have been observed/evidenced for geomechanical operations close to oil and gas reservoirs, mines, water reservoirs and geothermal systems. In oil and gas applications microseimic events can be used to map fractures distribution inside reservoirs, in order to find areas characterized by higher permeability and enhance production. In mining environments the analyis of microseismicity is important for mines stability monitoring [Gharti et al., 2010]. The study of distribution of microsesmic event can also be used at larger scale to study the seismogenic structures of a particular area [Stabile et al., 2012]. According to the frequency magnitude distribution of earthquakes smaller earthquakes occur more often than larger ones. Since microseismic monitoring networks allow to detect very low magnitude earthquakes, the rate of detected events can become very large. For this reason, manual seismic event location procedures are time consuming or, in some cases, not feasible.
The automated location of seismic events is an important and challenging task in microseismic monitoring applications (e.g., to analyse induced seismicity following oil/geothermal field exploitation and mining operations), where we generally deal with a large number and weak seismic events characterized by low signal-to-noise ratios. Standard automated location routines require precise automated picking procedure, phases identification and association [Gharti et al., 2010]. Picking procedure consists in the determination of the correct arrival time of a particular seismic phase, the identification step consists in identifying the picked phase (first P onset, first S onset etc.). Picks are associated to each single station. A second

procedure is needed to associate the picks at each station to particular phases of a seismic event. These location methods are usually modified versions of the Geiger [1912] algorithm, based on the minimization of time residuals between theoretical and observed arrival times of bodywaves (mainly the first P and S onsets) by iterative inversion algorithms. In order to locate seismic event in a automatic way, in the last two decades a large number of picking algorithms have been developed: while P onsets can now be accurately picked, the automatic picking of later seismic phases (including S onsets) is still challenging. Their performance is limited in presence of noisy data (e.g when the P coda overlaps the S first onset), when picking and phases identification might be difficult.
The growing interest on microseismic monitoring operations, particularly for oil and gas applications, pushed the development of alternative techniques based on the *migration*[1] concept taken from reflection seismics. These methods do not require phase picking nor phase identification and directly exploit the waveform information contained in seismograms. Migration based location methods can be divided in two main categories. The first one makes use of time-reversed seismograms as virtual sources. The wavefield is then backpropagated from each virtual source to the original source which corresponds to the point where the maximum energy focuses [Gajewski and Tessmer, 2005]. These methods are computationally intensive (e.g. synthetic seismograms have to be calculated), and energy focusing can be ambiguous with noisy data and very heterogeneous models [Gharti et al., 2010]. The second category of such techniques are based on the idea of delay and sum of the measured seismic waveforms (i.e. synthetic seismograms are not required). Among these methods one of the most popular is the Source Scanning Algorithm (SSA) developed by Kao and Shan [2004, 2007]. The source location is performed using a britghtness function, which is obtained by stacking the absolute amplitudes of normalized seismograms recorded at different stations. Unlike the backpropagation location approach these methods results faster in terms of computing time, since they exploit the waveform information without the need to compute synthetic seismograms. The SSA method is designed to locate typical tremor events with emergent waveforms and is based on the stacking of the absolute amplitudes at all stations at their respective predicted arrival times.
In this thesis I discuss how the SSA method can be optimized and improved to locate microseismic events using both the P and S first onsets. From conventional locations methods is well known that the simultaneous inversion of P and S picks is very important to better constrain the earthquake hypocenter. Not surprising, the use of P and S phases improves the location performance of sparse networks. The method I developed is based on the use of the Short-Term-Average/Long-Term-Average (STA/LTA) traces at different stations as observed data. For different trial locations and origin times, observed STA/LTA traces are stacked along the

[1] with the term *migration* I refer to an inversion operation involving rearrangement of seismic information elements so that reflections and diffractions are plotted at their true locations. The need for this arises since variable velocities and dipping horizons cause these elements to be recorded at surface positions different from the subsurface positions [Sheriff, 2002].

theoretical P and S first arrival time surface. Iterating this procedure on a three dimensional grid we retrieve a multidimensional matrix whose absolute maximum corresponds to the spatio-temporal coordinates of the seismic event.
This is a cumulative thesis composed by three peer reviewed articles. In the first part of this thesis (chapter 2), I discuss the importance of the data quality control as a preliminary step before starting further data processing steps. Determining the orientation of the horizontal components of seismic sensors is one of such operations and is a common problem that, when not properly solved, limits data analysis and interpretation for several acquisition setups. These include linear arrays of geophones deployed in borehole installations, ocean bottom seismometers deployed at the sea-floor or misoriented seismic station at the earth surface. In fact the knowledge of the correct orientations of the horizontal components of three component seismic sensors is needed in different applications, from seismic event location (when we are dealing with single seismic station or a single vertical borehole array) to moment tensor inversion. In this chapter I introduce a novel technique to retrieve the orientation of seismic sensor using a complex linear least squares approach. This orientation method is based on waveform comparison between stations within a network using seismic sources outside the network. The main advantage of our methodology is that, in the complex domain, the relative orientations of seismic sensors can be viewed as a linear inverse problem, which ensures that the preferred solution corresponds to the global minimum of a misfit function. It is also possible to use simultaneously more than one independent dataset (other seismic events) to better constrain the solution of the inverse problem. I successfully applied this method to datasets resembling different acquisition geometries and environments; 1) a linear array of geophones deployed in a borehole, 2) Ocean bottom seismometers in the Aegean sea and 3) stations of the Graefenberg array (Germany).
In the second part of this thesis (chapter 3 and 4) I describe the waveform stacking based automated location method I have developed. The methodology I propose is based on the stacking of the STA/LTA traces along the P and S first arrival times. The main benefits of our method are: 1) phase identification and picking are not required, 2) robustness against noise and 3) high level of automatization. In chapter 3 I apply this location method to both synthetic and real data. The Synthetic dataset consists of 200 seismic events with random location and focal mechanism. All synthetic events have been located simulating two different noise levels, 30% and 70% of the maximum amplitude of each seismic trace . In both cases, even with a very high noise level, results confirmed the robustness of our method (I located the 90% of the events within 150 m from the true location with a noise level at 70%). Concerning the real data I successfully locate 391 microseismic events (with magnitudes Ml between 0.5 and 2.0) induced by coal mining activity in the Ruhr region (Germany). The network consisted of 15 stations (9 short-period stations and 6 broad-band stations) and an aperture of about 5 km. To further validate our method I compare our locations with those obtained by a manual location procedure.

In chapter 4 I apply this location approach to crustal earthquakes (with magnitudes Ml between 1.1 and 2.7) recorded by a regional seismic network in the Campania lucania region (southern Italy). The network consisted of 33 stations (28 short-period stations and 5 broad-band stations) and an aperture of about 150 km. Due to the complex geology of the area, seismic signals related to this dataset are characterized by a strong P coda overlapping the S wave onset making problematic the use of our location method. To overcome these problems I improved the location method introduced in chapter 3, by using the STA/LTA trace of a characteristic function more sensitive to the S waves. With the use of this characteristic function our waveform stacking location method results more robust and shows a better performance even when the identification of the S wave is difficult (overlapping of the P coda with the S phases, emergent S arrival, noise contaminated data, etc.). I compared our locations with accurate manual locations obtained by using double difference algorithm. Finally in the last part of the thesis I briefly resume the advantages and the limitations of our approach, discussing what are the main features of the method that can be improved in a future development version.

The waveform coherence analysis location method by sketches

In this section I will briefly describe step by step, through sketches, how the location method developed within this thesis works. A detailed and more rigorous description of the method is given in the chapters 3 and 4. In order to explain the idea behind the location method I will show a very simple example in 2D (the extension to 3D is trivial). Let us consider an halfspace with a linear array of receivers deployed at the surface (figure 1.1.a). Now let us suppose that a seismic events occur within the halfspace (the yellow star in figure 1.1.a) and that recorded traces are composed by direct P and S phases only (figure 1.1.b). The STA/LTA traces of the recorded waveforms (figure 1.1.c) are then used as the input data of our location method (for further details about the STA/LTA see the chapter 4). The location process starts with the scanning of all potential

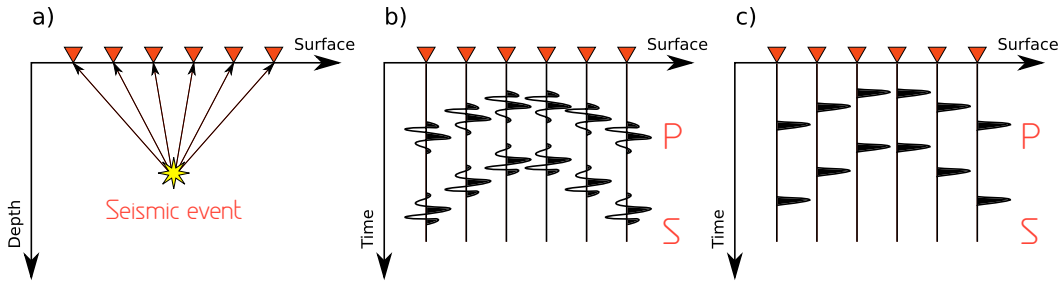


Figure 1.1: This sketch represents: a) The acquisition geometry and the source location (yellow star) 2) The recorded traces and 3) The STA/LTA traces of the recorded waveforms.

source location. Starting from the first trial location (figure 1.2.a1), we com-

pute the theoretical traveltimes for the P and the S phases (the blue and the red line respectively) and we stack the observed traces along these path. We iterate this process for different time steps (figures 1.2.b1,c1,d1) and locations. When we arrive to the correct location (figure 1.2.a2) and time step (figure 1.2.c2) we observe the maximum coherence along the P and S arrival times. Once scanned

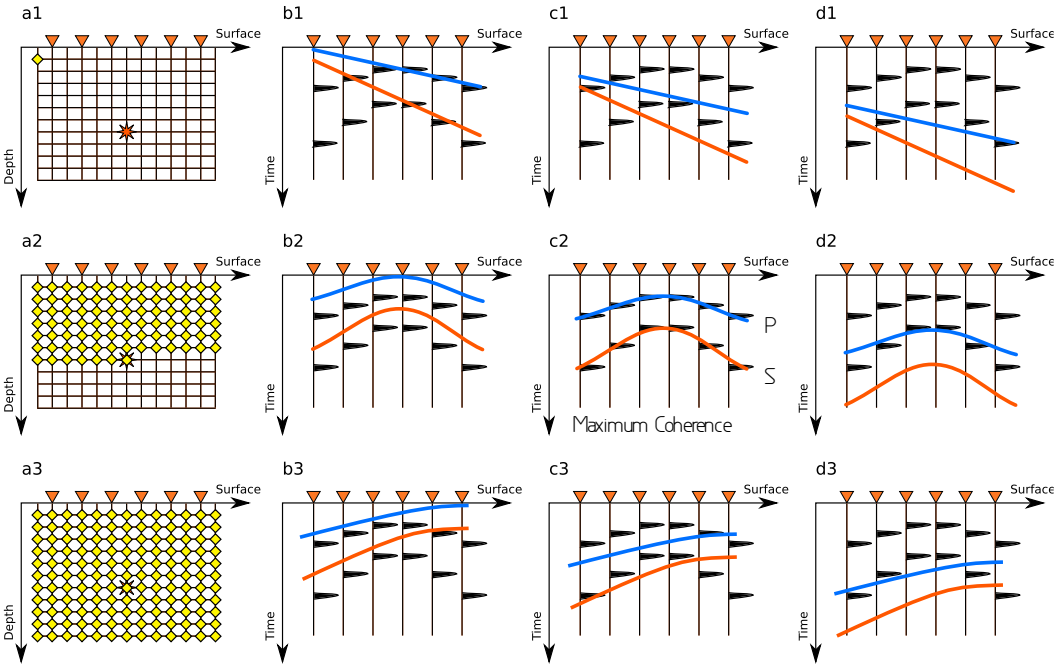


Figure 1.2: The waveform stacking location process for different source locations (indicated by yellow diamonds in panels a.X) and time steps (panels b, c, and d).

all the possible source locations and time steps (figures 1.2.a3,b3,c3,d3), we retrieve a coherence matrix $C(x, z, t)$ (figure 1.3). The maximum of $C(x, z, t)$ is related to the hypocentral coordinates $(\hat{x}, \hat{z})$ of the seismic event by the relation: $C(\hat{x}, \hat{z}, \hat{t}) = \max\{\mathrm{C(x, z, t)}\}$. Where the $\hat{t}$ is the time step where the maximal coherence is observed. The time $\hat{t}$ is also related to the origin time of the event (see chapter 3 and 4 for more details about the location process).

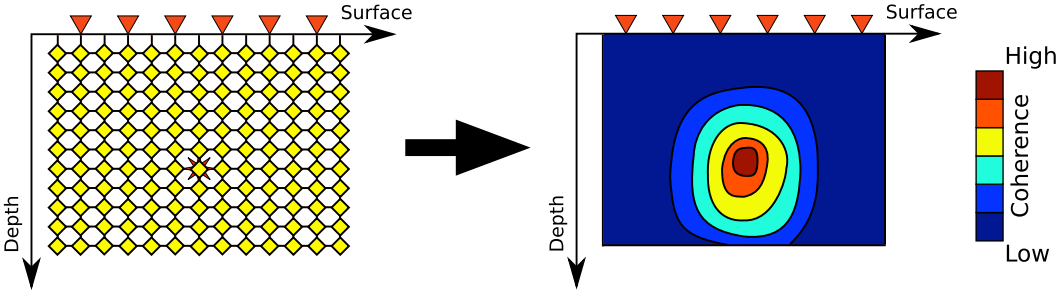


Figure 1.3: Coherence matrix $C(x, z)$ at the time $\hat{t}$, where the maximal coherence is detected.

Chapter 2

A complex linear least-squares method to derive relative and absolute orientations of seismic sensors

Authors:
Francesco Grigoli[a,d], Simone Cesca[a,b,d], Torsten Dahm,[a,b,d] Lars Krieger[c,d]
Journal:
Geophysical Journal International, 2012, 188: 1243–1254,
doi: 10.1111/j.1365-246X.2011.05316.x
Status:
Published

Authors affiliation:
a) Institute of Earth and Environmental Sciences,
University of Potsdam,
Potsdam, Germany
b) GFZ (German Research Centre for Geosciences) Potsdam,
Section 2.1, Physics of earthquakes and volcanoes,
Potsdam, Germany
c) Institute for Mineral and Energy Resources,
University of Adelaide,
Adelaide, Australia
d) Formerly at Institute of Geophysics,
University of Hamburg,
Hamburg, Germany

2.1 Abstract

Determining the relative orientation of the horizontal components of seismic sensors is a common problem that limits data analysis and interpretation for several acquisition setups, including linear arrays of geophones deployed in borehole installations or ocean bottom seismometers deployed at the sea-floor. To solve this problem we propose a new inversion method based on a complex linear algebra approach. Relative orientation angles are retrieved by minimizing, in a least-squares sense, the l_2-norm between the complex traces (hodograms) of adjacent pairs of sensors. This methodology can be applied without restrictions only if the wavefield recorded by each pair of sensors is very similar. In most cases, it is possible to satisfy this condition by low-pass filtering the recorded waveforms. The main advantage of our methodology is that, in the complex domain, the relative orientations of seismic sensors can be viewed as a linear inverse problem, which ensures that the preferred solution corresponds to the global minimum of a misfit function. It is also possible to use simultaneously more than one independent dataset (other seismic events) to better constrain the solution of the inverse problem. Furthermore, by a computational point of view, our method results faster than the relative orientation methods based on waveform cross-correlation. After several tests on synthetic datasets we applied successfully our methodology to different types of real data. These applications include the alignment of borehole sensors relative to a Vertical Seismic Profiling (VSP) acquisition and the orientation of Ocean Bottom Seismometers (OBS) relative to a neighbouring land station of known orientation. Using land stations, the absolute orientation of Ocean Bottom Seismometers can be retrieved. Finally, as a last application, we checked the correct orientation for land stations of a seismological array in Germany.

2.2 Introduction

Horizontal components of three-components seismic sensors (geophones or seismometers) deployed in borehole installations or at the sea bottom are usually in unknown, random orientation. In Vertical Seismic Profiling (VSP) acquisition, for instance, the sonde cables for current borehole systems cannot control the orientation of the horizontal elements [Di Siena et al., 1984]. For this reason the first step in VSP data processing consists in the alignment and absolute orientation of all borehole sensors. The standard techniques used to solve this problem are based on polarization analysis or cross-correlation of seismic signals. Di Siena et al. [1984] used a power maximization scheme in VSP processing to determine the azimuthal orientation of three-components geophones in a vertical borehole. This technique consists on maximizing the signal energy of the first P-wave arrival by orientation on a particular axis. Becquey and Dubesset [1990] applied polarization analysis based method to derive three-components sonde orientation in a deviated well. Michaels [2001] proposed a method based on principal component analysis to determine the tool orientation relative to source polarization direction using

SH-wave. Once rotation angles are found, the horizontal components of the sensors can be rotated to a radial/transversal coordinate system. Oye and Ellsworth [2005] applied a method based on polarization analysis to find the geophones orientation angles for a linear array installed in a borehole located near the San Andreas fault, California. Absolute orientations were derived comparing azimuths obtained from P-wave polarization analysis and theoretical azimuths estimated from ray tracing. Zeng and McMechan [2006] used traces cross-correlation to infer relative angles between adjacent geophone pairs in borehole arrays. The methods used to derive the orientation of broadband seismological arrays deployed on the sea-floor are similar to those encountered for borehole arrays. The main difficulty in the deployment of a free-fall OBS stations is that users do not have full control over how and where the station is landing [Dahm et al., 2002]. Horizontal components orientations of OBS stations are generally unknown and it is necessary to estimate them directly from recorded data. There are several existing methods to solve this problem and most of them are quite similar to those used in borehole applications. Nakamura et al. [1987], for instance, used air-gun shots to determine the location and orientation of OBS stations. They estimated OBS orientations using the amplitude ratio of the water wave arrival recorded by the horizontal components of the OBS. Li and Yuan [1999] used a method based on polarization analysis of the first arrival P-wave for determining geophone orientation of a multicomponent ocean bottom cable (OBC) deployed on sea-floor. Hensch [2009] used waveform cross-correlation to estimate alignment angle between OBSs and land seismic stations in the Aegean sea. In this paper we present a new method for determining relative seismic sensor orientations based on a linear least square inverse problem in the complex domain. This method can be applied if the distance between each pair of sensor is much smaller than the dominant wavelength of the seismic event (or explosion). However through low-pass filtering it is possible to satisfy this condition for most cases. Finding orientation of seismic sensors by complex linear least-squares approach has the advantage that we are dealing with a linear inverse problem. For this reason there are no complications with local minima and it is possible to add more independent data (other seismic events) to better constrain the solution of the inverse problem. Furthermore, our methods is faster than relative orientation methods based on waveform cross-correlation, and allows to estimate simultaneously all relative orientation angles of each sensors pair. An other advantage of our methods is that, unlike polarization analysis based methods, it can be applied to the full waveform and not only to highly linear polarized part of the waveform. To test the reliability of our methodology, we firstly applied it to synthetic data, simulating a test shot recorded by a vertical borehole array. Then, we applied it to three different real datasets. First, we used our methodology to derive geophones relative orientations in a borehole array relative to a set of VSP field data from Sudbury basin (Canada). Then we applied it to OBS data in the Aegean sea (Greece). In this application we further obtain absolute orientations by aligning OBS sensors with a reference land station. The last application is relative to the Graefenberg seismological array (Germany). In

this case we show that our methodology can be successfully used to check the correct orientations of stations in seismic arrays.

2.3 Theory

Our method relies on the hypothesis that for two aligned sensors the misfit between the respective waveforms is an absolute minimum, while misfit increases if the sensors are not aligned. Thus, we can find the alignment angle solving an inverse problem in the complex domain. At each pair of neighbouring sensors we assume the condition of plane wave approximation. This requirement must be satisfied to retrieve a correct estimation of the alignment angles. To ensure this condition, the relation $d \ll \lambda$ (where λ is the dominant wavelength of the recorded waveform and d is the linear distance between the two adjacent sensors to align) must be respected. Each pair of traces shows, after alignment, a high similarity. For this reason, we use a multistep processing approach in order to maximize signal similarity between waveforms related to each sensors pair. The first processing step consists in low-pass filtering recorded waveforms. The corner frequency of the filter is chosen to fulfil the relation $f_c \ll V_{min}/d$, where f_c is the cut-off frequency, V_{min} is the lowest seismic velocity of the medium and d is the distance between sensors. The next step consists in the time alignment of all traces with respect to a common reference time and in the normalization of each trace with respect to the maximum amplitude recorded by the horizontal components of the sensor. Time aligned traces are obtained by using the lag of the maximum cross-correlation value between total energy traces (sum of squares of horizontal components), which are invariant by rotation. In the last step of the data processing procedure we define the time window to be analysed. Depending on data quality, we can use the full recorded waveforms or only a part of the signal (for instance, the first P-wave arrivals or the surface wave trains). If similarity conditions are not satisfied for the full waveforms, in order to find the best time window, we use the normalized cross-correlation of the total energy trace starting from a fixed point (generally the first onset) and increasing, sample by sample, the length of the time window to the end of the selected waveforms. Since the total energy function is independent from sensor rotations (i.e. is an invariant), it can be used to evaluate the wavefield similarity at neighbouring sensors and therefore it can be used to define the best time window. A vector rotation in two dimensions can be described by a rotation matrix in $\mathbb{R}^2$. An alternative description of rotations makes use of complex numbers. In the complex field $\mathbb{C}$, rotations can be performed by simple multiplication of the complex number (i.e the two dimensional vector) with a complex exponential whose phase is the rotation angle.

$$w = u + iv \quad \text{with} \quad u, v \in \mathbb{R} \tag{2.1}$$

$$w = |w|e^{i\alpha} \quad \text{where} \quad |w| = \sqrt{ww^*} \tag{2.2}$$

Where α is the phase of w and w^* is its complex conjugate. Let be w' the result of the rotation of w by an angle β. Since $|w'| = |w|$ (rotation does not change

the modulus of the complex number), we can write:

$$w' = |w'|e^{i(\alpha+\beta)} = |w'|e^{i\alpha}e^{i\beta} = we^{i\beta} \tag{2.3}$$

Equation (2.3) shows that the mapping of the complex number w to w' is analogue to a counter clockwise rotation by an angle β of the vector w. We can now apply these concepts to rotate and align the horizontal components (that we suppose to be mutually orthogonal) of one or more sensor pairs. If we assume that vertical components of all seismic sensors are parallel to the casing of the vertical borehole (VSP data) or perfectly coupled with the sea-floor and levelled (OBS data), the alignment of the sensors requires only a rotation of the two horizontal components (X and Y) around the vertical axis. Our aim is to determine the relative orientation of a particular set of seismic sensors with respect to a reference one, finding the alignment angle between sensor pairs. We start defining the *complex trace* $\hat{S}(t)$ (equation 2.4) as a complex valued function (the symbol $\hat{}$ will denote complex variables) whose real part is the seismic trace $X(t)$ and the imaginary part is the seismic trace $Y(t)$:

$$\hat{S}(t) := X(t) + iY(t) = A(t)e^{i\theta(t)} \tag{2.4}$$

with $A(t) = |\hat{S}(t)|$ and $\theta(t) = \arctan(\mathrm{Im}[\hat{\mathrm{S}}(\mathrm{t})]/\mathrm{Re}[\hat{\mathrm{S}}(\mathrm{t})])$.
Since a digitally recorded seismic trace is a discrete time series, we write $\hat{S}(t)$ as the vector $\hat{\mathbf{S}}$, $X(t)$ as $\mathbf{X}$ and $Y(t)$ as $\mathbf{Y}$:

$$\hat{\mathbf{S}} = \mathbf{X} + i\mathbf{Y} \tag{2.5}$$

This reads in matrix form:

$$\begin{pmatrix} \hat{S}_1 \\ \hat{S}_2 \\ \vdots \\ \hat{S}_n \end{pmatrix} = \begin{pmatrix} X_1 + iY_1 \\ X_2 + iY_2 \\ \vdots \\ X_n + iY_n \end{pmatrix} = \begin{pmatrix} A_1 e^{i\theta_1} \\ A_2 e^{i\theta_2} \\ \vdots \\ A_n e^{i\theta_n} \end{pmatrix} \tag{2.6}$$

Where $A_m = |\hat{S}_m|$ with $m = 1, 2, ..., n$ and n the number of samples of the seismic trace. In analogy with the equation (2.3), we can write the rotation of $\hat{\mathbf{S}}$ by an angle ϕ as:

$$\hat{\mathbf{S}}' = \hat{\mathbf{S}}e^{i\phi} \tag{2.7}$$

or in matrix form:

$$\begin{pmatrix} \hat{S'}_1 \\ \hat{S'}_2 \\ \vdots \\ \hat{S'}_n \end{pmatrix} = \begin{pmatrix} A_1 e^{i\theta_1+\phi} \\ A_2 e^{i\theta_2+\phi} \\ \vdots \\ A_n e^{i\theta_n+\phi} \end{pmatrix} \tag{2.8}$$

Now let's suppose to have a set of k non aligned seismic sensors. The angle ϕ_j is the angle between the X-component of the j-th sensor and the X-component of

the $(j+1)$-th sensor (with $j = 1, 2, ..., (k-1)$) as shown in figure 2.1. In order to align this pair of sensors we need to rotate the $(j+1)$-th sensor by an angle $-\phi_j$ around the Z axis. Following an inverse problem approach we can think $e^{i\phi_j}$ as a model parameter which can be inverted to minimize the misfit between the complex traces of the j-th and $(j+1)$-th sensor. Using these assumptions, we define a *concatenated trace* as the vector **d** containing complex traces of the first $k-1$ sensors:

$$\mathbf{d} = \left(\hat{S}_1^1\ \hat{S}_2^1\ \ldots\ \hat{S}_n^1\ \hat{S}_1^2\ \hat{S}_2^2\ \ldots\ \hat{S}_n^2\ \ldots\ldots\ \hat{S}_1^{k-1}\ \hat{S}_2^{k-1}\ \ldots\ \hat{S}_n^{k-1}\right)^T \tag{2.9}$$

Where $\hat{S}_m^j$ is the m-th sample of the complex trace associated to the j-th sensor (with $m = 1, 2, ..., n$ and $j = 1, 2, ..., k$). Let n the number of samples of each complex trace and k the number of traces (i.e sensors), our concatenated trace will be a vector with $n(k-1)$ elements. In order to relate each sensor j with its

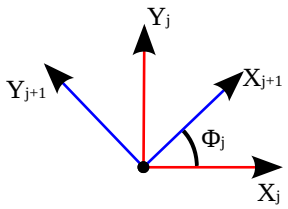


Figure 2.1: Horizontal components of two non aligned sensors. A clockwise rotation of the sensor (j+1) by an angle $-\phi_j$ aligns it with the sensor j.

adjacent one $(j+1)$, which is rotated by an angle ϕ_j, we define the sparse matrix **G**:

$$\mathbf{G} = \begin{pmatrix} \hat{S}_1^2 & 0 & \ldots & 0 \\ \hat{S}_2^2 & 0 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{S}_n^2 & 0 & & 0 \\ 0 & \hat{S}_1^3 & & 0 \\ 0 & \hat{S}_2^3 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \hat{S}_n^3 & & 0 \\ 0 & 0 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \hat{S}_1^k \\ 0 & 0 & & \hat{S}_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \hat{S}_n^k \end{pmatrix} \tag{2.10}$$

The relation between the traces of adjacent sensors is defined by the respective rotation angle $e^{i\phi_j}$. These are setting the model parameter vector **m**:

$$\mathbf{m} = \left(e^{i\phi_1}\ e^{i\phi_2}\ \ldots\ e^{i\phi_{k-1}}\right)^T \tag{2.11}$$

Solving the orientation problem is equivalent to solve the system of linear equations:

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \mathbf{r} \tag{2.12}$$

where **r** is the error vector. Since **G** is a complex valued and non square matrix and the inverse problem is overdetermined, we calculate the generalized inverse of **G** to find the vector **m**. For complex matrices the generalized inverse makes use of the hermitian transpose $\mathbf{G}^H$ [Miller, 1973]. Thus, if $(\mathbf{G}^H\mathbf{G})^{-1}$ is non singular, the estimated $\mathbf{m}_{l_2}$ equals to:

$$\mathbf{m}_{l_2} = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H\mathbf{d} \tag{2.13}$$

After the determination of **m**, all rotation angles ϕ_j are implicitly known. They can be calculated through Euler's formula as:

$$\phi_j = \begin{cases} \arctan\left(\dfrac{\sin\phi_j}{\cos\phi_j}\right) & \text{if } \cos\phi_j > 0 \\ \text{sgn}(\sin\phi_j)\dfrac{\pi}{2} & \text{if } \cos\phi_j = 0 \\ \arctan\left(\dfrac{\sin\phi_j}{\cos\phi_j}\right) + \pi & \text{if } \cos\phi_j < 0 \end{cases} \tag{2.14}$$

Error estimation is here obtained using a method similar to the one proposed by Knowlton and Spencer [1996] to estimate azimuth uncertainties. It consists in the perturbation of the initial time and the length of the window containing the part of waveform used to estimate rotation angles. For each perturbed window, a new set of model parameters is computed. Repeating this procedure several times we obtain different estimations for each model parameter (i.e rotation angle). Then, we use a weighted mean and standard deviation to compute the best estimation and the error of each model parameter. Let be $\hat{\mathbf{r}}_j$ the vector of residual between the j-th reference trace and the (j+1)-th trace rotated by ϕ_j:

$$\hat{\mathbf{r}}_j = \hat{\mathbf{S}}_j - e^{i\phi_j}\hat{\mathbf{S}}_{j+1} \tag{2.15}$$

the module of this vector is:

$$E_j = |\hat{\mathbf{r}}_j|^2 = (\hat{\mathbf{S}}_j - e^{i\phi_j}\hat{\mathbf{S}}_{j+1})^H(\hat{\mathbf{S}}_j - e^{i\phi_j}\hat{\mathbf{S}}_{j+1}) \tag{2.16}$$

and defining a weight as:

$$A_j = \frac{1}{E_j} \tag{2.17}$$

we can write, for m observation, the normalized weighting factor W_k^j relative to k-th observation as:

$$W_j^k = \frac{A_j^k}{\sum_{h=1}^{m} A_j^h} \tag{2.18}$$

Using the weighting factor from the equation (18) we can write the following equations:

$$P_j = \sum_{h=1}^{m} \cos\left(\phi_j^h\right) W_j^h \quad \text{and} \quad Q_j = \sum_{h=1}^{m} \sin\left(\phi_j^h\right) W_j^h \tag{2.19}$$

$$R_j = \sqrt{P_j^2 + Q_j^2} \tag{2.20}$$

from the latter equation we can compute the best estimation μ_j of the model parameter related to j-(j+1) geophone pair and its error σ_j by:

$$\mu_j = \arctan\left(\frac{Q_j}{P_j}\right) \quad \text{and} \quad \sigma_j = \sqrt{2\left(1 - R_j\right)} \tag{2.21}$$

Since we are dealing with angles, we used the definitions of mean and standard deviation from directional statistics [Mardia and Jupp, 2009]. Finally we can write our rotation angle as:

$$\phi_j = \mu_j \pm \sigma_j \tag{2.22}$$

Since the weighting factors are estimated from residuals, for each sampling we have to normalize the part of the waveform sampled. In this way we avoid that the misfit is dominated by large amplitude wavelets, giving a more homogeneous contribution on the residuals. Although Miller [1973] describes how to retrieve uncertainties using the complex covariance matrix, we prefer to use the bootstrap based method (previously described) because it is non parametric. This means that a priori assumptions about data statistics are not required. For comparison purposes, we will use both methods only on synthetic data.

2.4 Synthetic tests

To test the reliability of our methodology we first applied it to synthetic data generated using the finite difference code E3D [Larsen and Grieger, 1998]. We created synthetic seismograms for an homogeneous medium ($V_p = 4500$ m/s, $V_s = 2700$ m/s and $\rho = 2.5$ g/cm^3), using an explosive source (Ricker wavelet with central frequency of 10 Hz) placed at 1100 m depth and 950 m offset. Our acquisition geometry consists of five three-components geophones placed along a vertical borehole from 2000 to 2120 m depth, with a group interval of 30 m. The vertical components of the geophones are aligned along the Z-axis, while horizontal components are randomly oriented. The sampling period is 2 ms. The figure (2.2.a) shows a sketch of our acquisition geometry. After time alignment of all traces we did not filter the data, because the group interval of the

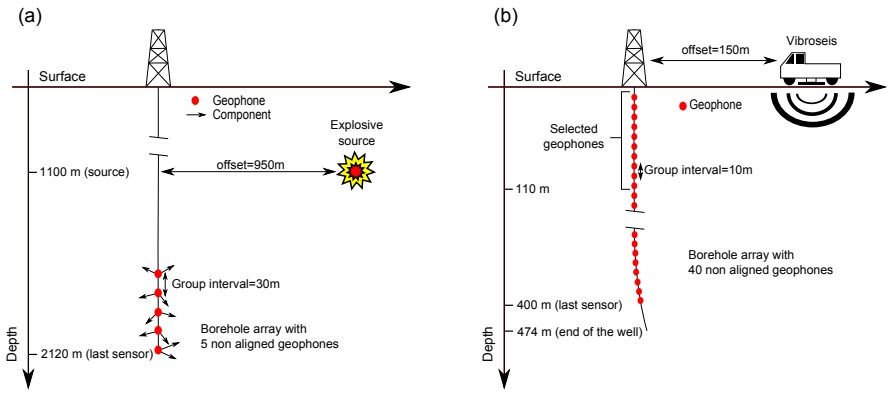


Figure 2.2: Sketch of a vertical borehole with a linear array of seismic sensors. Vertical components of all geophones are parallel to the case of the borehole, while the horizontal components are not aligned. Figure (a) represents the acquisition geometry referred to synthetic test while figure (b) the acquisition geometry referred to VSP data from Canada.

array is about ten time smaller than the dominant wavelength of the waveform. Our target is to find the alignment angles between the j-th geophone and the $(j+1)$-th geophone (with $j = 1, 2, 3, 4$). The true alignment angles of each pair of geophones are listed in the table (2.1). We have performed different tests using both noise free synthetic traces and including random noise (noise level 10%). Results related to noise free data are shown in figure (2.3). Table

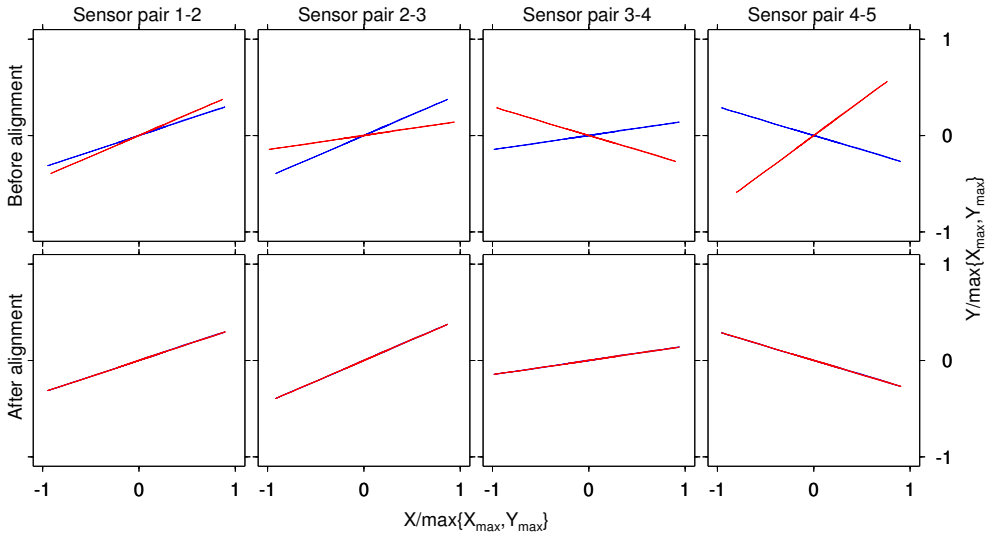


Figure 2.3: Synthetic data results. Hodograms of 4 sensors pairs: before (top) and after alignment (bottom). The red hodogram is related to the sensor we want align with the reference one (blue hodogram).

(2.1) shows satisfactory results obtained applying our technique. To test the improvement of the solution using a larger dataset, we add a second explosion

to our dataset. This explosion has the same depth of the previous one but a different azimuth and an offset of 565 m. Results listed in table (2.1) show also that the inversion including a second explosion leads to a better estimation of the alignment angles and smaller uncertainties. Uncertainties were estimated by random perturbations (1000 times) of the initial time and length of the window containing the signal. We have analysed the effect of a polarized noise on the estimation of alignment angles. We found that directional noise has a minor influence on the estimation of the alignment angles. Effects are larger for directions perpendicular to the signal direction (from 0.5° to 2.5°), and lower when noise and signal have the same direction (from 0.01° to 0.1°). However, it is possible to reduce the directional noise effect by using more events (earthquake or explosion) simultaneously (e.g. adding a second explosion with a different azimuth and we found that error decreased by a factor 2). In table (2.2) we show, only for comparison purposes, results (related to the first dataset) computing uncertainties with the alternative procedure described by Miller [1973].

True and estimated alignment angles				
Sensors pair	True angle	Noise free data	Data+noise	Data+noise 2
1-2	-5°	$-5.0^\circ \pm 0.3^\circ$	$-5.2^\circ \pm 0.7^\circ$	$-5.0^\circ \pm 0.1^\circ$
2-3	15°	$15.1^\circ \pm 0.7^\circ$	$15.3^\circ \pm 0.9^\circ$	$14.9^\circ \pm 0.2^\circ$
3-4	25°	$24.9^\circ \pm 0.3^\circ$	$24.7^\circ \pm 0.6^\circ$	$25.0^\circ \pm 0.1^\circ$
4-5	-53°	$-52.9^\circ \pm 0.6^\circ$	$-52.7^\circ \pm 0.9^\circ$	$-53.0^\circ \pm 0.2^\circ$

Table 2.1: Synthetic test results: Comparison between true alignment angles and the estimated ones. Data+noise 2 refers to the results obtained using two explosions at different azimuths. Uncertainties were estimation using the bootstrap approach.

True and estimated alignment angles			
Sensors pair	True angle	Noise free data	Data+noise
1-2	-5°	$-5.0^\circ \pm 0.1^\circ$	$-5.2^\circ \pm 0.3^\circ$
2-3	15°	$15.0^\circ \pm 0.2^\circ$	$15.2^\circ \pm 0.8^\circ$
3-4	25°	$24.9^\circ \pm 0.3^\circ$	$24.9^\circ \pm 1.3^\circ$
4-5	-53°	$-53.0^\circ \pm 0.5^\circ$	$-54.5^\circ \pm 2.4^\circ$

Table 2.2: Synthetic test results: Comparison between true alignment angles and the estimated ones. Uncertainties were estimated via complex covariance matrix [Miller, 1973].

2.5 Application to VSP data

(Sudbury basin, Canada)

We tested our methodology on VSP field data from the Sudbury Basin, Ontario, Canada. These VSP data are part of the LITHOPROBE Sudbury Transect and

have been provided by the Canadian Geological Survey (http://gsc.nrcan.gc.ca). The Chelmsford borehole is 447.1 m deep. It is vertical at the surface but deviates slightly along depth with a maximum deviation of 10° from vertical at a depth of 400 m [Miao et al., 1995]. The dataset we used is characterized by a minimum offset of 150 m, receivers intervals of 10 m and vibroseis source with sweep frequency from 30 to 140 Hz (figure 2.2.b). Before the alignment process we applied a band pass filter (30-140 Hz) and a notch filter with a band rejection between 58 and 66 Hz in order to eliminate the 62 Hz generator noise. Geophones are close enough to satisfy the condition $d \ll \lambda$ (the dominant wavelength of the recorded waveforms is about 50 m). Since the horizontal components of sensors placed at shallow depths seem to be not well aligned [Miao et al., 1995], we focused our attention on these stations. In order to show the reliability of the methodology described before, we find relative orientation angles for a subset of 10 geophones in the firsts 100 m of the borehole. Waveforms related to the first sensor of the borehole array were rejected because of their poor quality. The dataset used for this application (figure 2.4) is not an ideal case, the fast decrease of similarity after the first arrival of the compressional wave does not allow to use the full waveforms. However, it is possible to select the part of the waveforms that best fit the similarity conditions required by our method. From the mean cross-correlation trace (figure

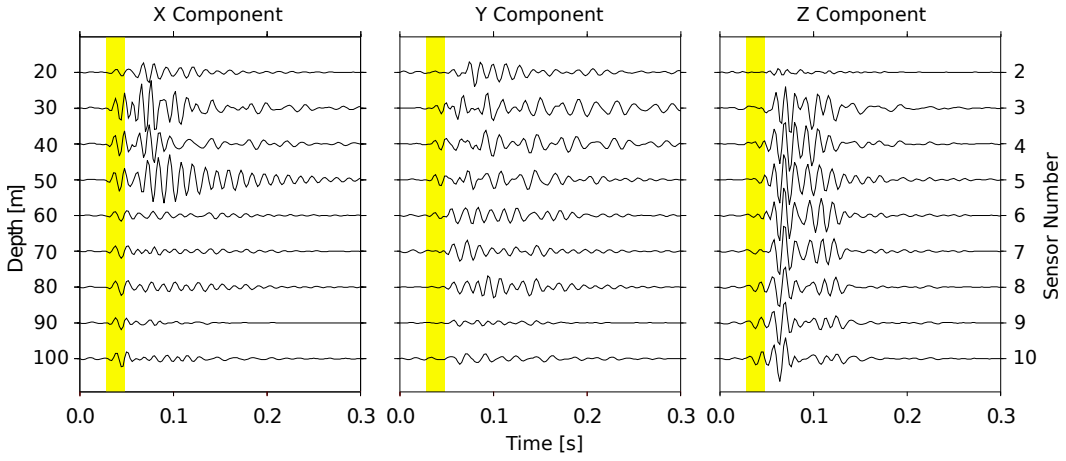


Figure 2.4: VSP data from Sudbury basin (Canada). The part of waveform used for alignment is highlighted in yellow.

2.5), we found that in the time window containing first P-onsets waveforms of adjacent sensors show the maximum similarity. Alignment results are shown in figure (2.6), while angles estimations and corresponding uncertainties are listed in table (2.3). Alignment angles and related uncertainties were estimated again by random perturbations (1000 times) of length and initial time of the selected time window. These perturbations allow to sample the part of waveforms between 20 and 60 ms. The large uncertainties related to sensors pairs 2-3 and 5-6 (2.3) are due to the poor similarity between waveforms.

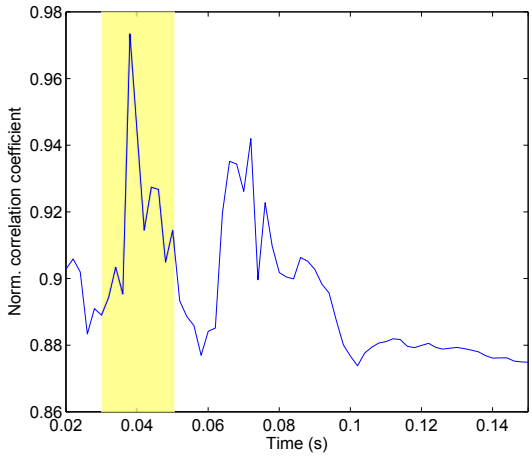


Figure 2.5: VSP data from Sudbury basin (Canada). Average of the cross-correlation of the total energy traces between all adjacent sensors. The yellow time window corresponds with the part of the waveform containing the first arrival of the P-wave.

Geophones pair	Estimated alignment angle
2-3	$-1^\circ \pm 11^\circ$
3-4	$-11^\circ \pm 2^\circ$
4-5	$-5^\circ \pm 3^\circ$
5-6	$1^\circ \pm 10^\circ$
6-7	$23^\circ \pm 3^\circ$
7-8	$9^\circ \pm 4^\circ$
8-9	$-2^\circ \pm 1^\circ$
9-10	$-3^\circ \pm 2^\circ$

Table 2.3: Estimated alignment angles for VSP field data from Sudbury basin (Canada).

2.6 Application to OBS data
(Aegean sea, Greece)

A second application to OBS stations deployed in the Aegean Sea shows the performance of our method in a different environment. In this case our aim is to align OBS stations with a permanent land station, obtaining the absolute orientations of each OBS. The dataset consists in two OBS broadband sensors and one permanent land station. The OBS systems of the Hamburg University are composed by a three component seismometer which has a sampling frequency of 50 Hz and provides flat response in the range 0.025-25 Hz, and an hydrophone, which covers the frequency range 0.5-25 Hz (sampling frequency 50 Hz). The land station SANT (Santorini Island, GEOFON network) is used as reference to estimate the absolute orientation of nearby OBSs. Figure (2.7) shows the configuration of our

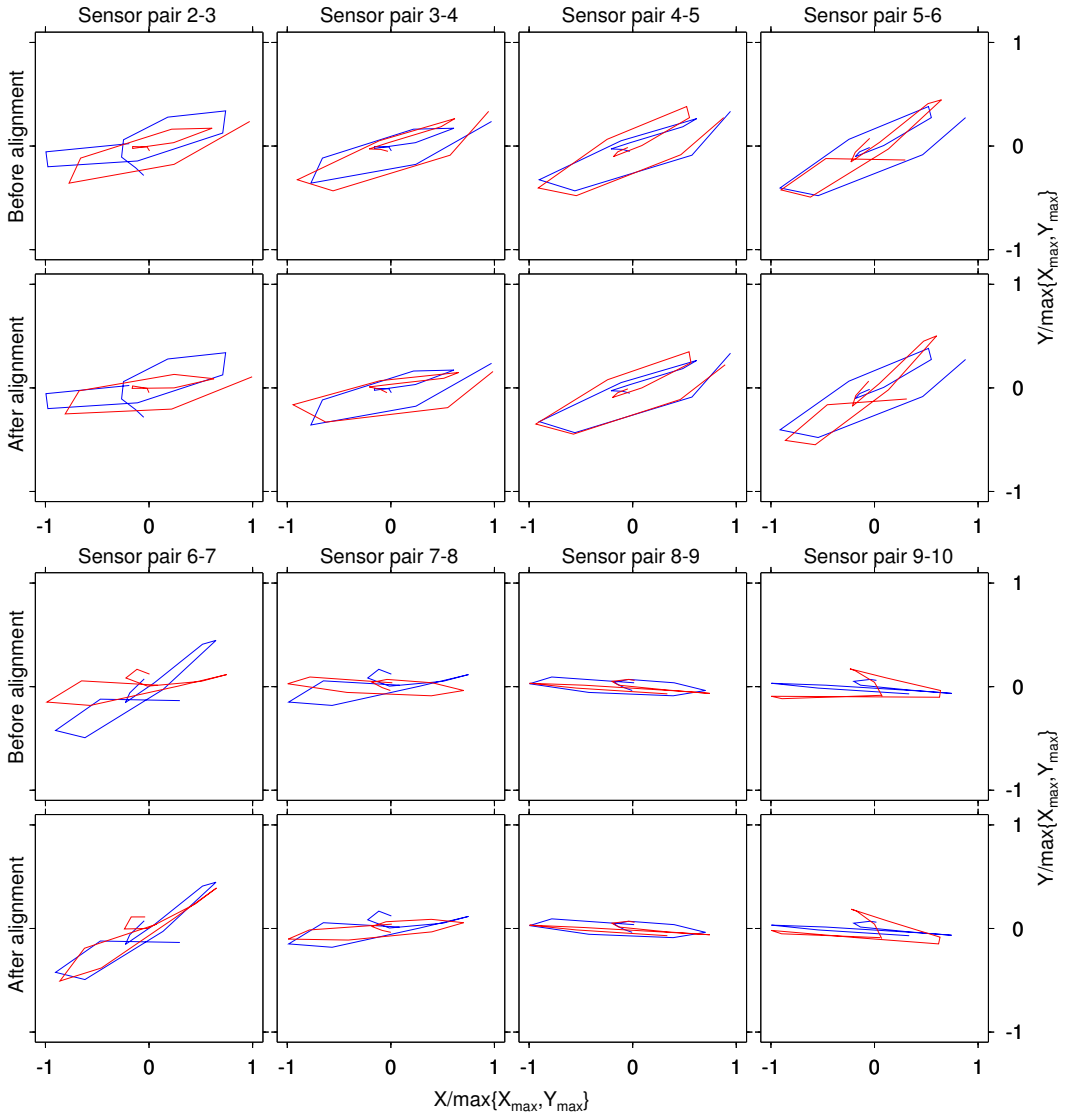


Figure 2.6: Field data results (VSP) showing hodograms for all sensors pairs. The red hodogram is related to the sensor $(j+1)$ we want align with the reference one j (blue hodogram), with $j = 2, .., 9$.

network. The largest distance among sensors is about 14 km, so that 0.07 Hz is an acceptable corner frequency to filter our data. To find orientation of OBS stations we used seismic signals from two teleseismic earthquakes: the 15 Nov 2006, $\mathrm{M_w}$ 8.3 Kuril Island event (Lat. $46.7°$, Lon. $153.2°$, Depth 27 km) and the 25 Mar 2007, $\mathrm{M_w}$ 7.2 Vanuatu Islands event (Lat. $-20.6°$, Lon. $169.4°$, Depth 35 km). For the Kuril island seismic event, stations azimuths are about $36°$ while epicentral distances are about $85°$. Stations azimuths of the Vanuatu island seismic event are about $55°$, while epicentral distances are about $145°$. We applied a band pass filter in the range 0.03-0.07 Hz (for both events), to optimize signal-to-noise ratio. We used the full waveforms for both events. Since the land station and OBSs

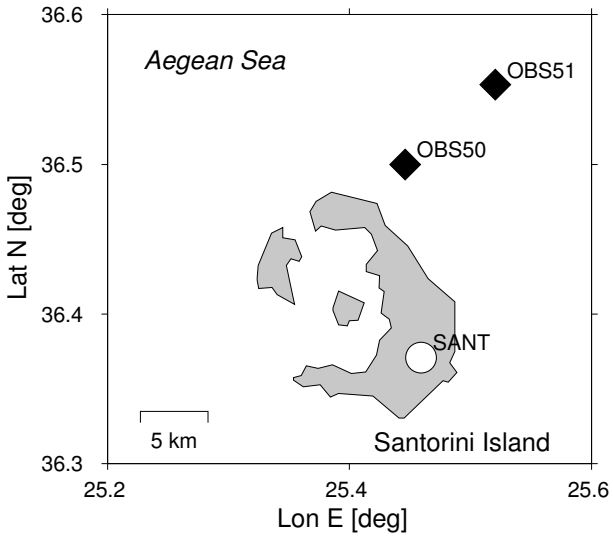


Figure 2.7: Map of Santorini island (Greece) showing the network configuration. OBS stations are represented by black diamonds and land station by a white circle.

do not have the same sampling frequency we downsampled our data to 10 Hz. Firstly we estimated rotation angles using the Kuril Island seismic event (Figure 2.8). Results are shown in figures (2.9) and table (2.4). Then, to check the consistency of our results we used the Vanuatu island seismic event as a second complementary dataset. Results are shown in figures (2.10) and table (2.4). We

Estimated alignment angles			
	SANT-OBS50	OBS50-OBS51	SANT-OBS51
Kuril Island ev.	$-30° \pm 3°$	$-121° \pm 4°$	$-151° \pm 6°$
Vanuatu Island ev.	$-31° \pm 4°$	$-124° \pm 4°$	$-150° \pm 8°$
Kuril ev. + Vanuatu ev.	$-31° \pm 3°$	$-122° \pm 3°$	$-152° \pm 5°$
	REF-OBS50	OBS50-OBS51	REF-OBS51
Hensch (2009)	$-39° \pm 4°$	$-130° \pm 8°$	$-169° \pm 4°$

Table 2.4: Estimated rotation angles for OBS data. OBS stations are aligned here with the land station SANT, Hensch [2009] used a different reference station (REF)).

also performed a joint inversion using both seismic events (table (2.4)). For all cases we found consistent solutions. Furthermore, these solutions were compared with those obtained by Hensch [2009] (table(2.4)), where the absolute alignment of OBS stations were performed using other neighbouring reference land stations. This may explain why results of the alignment of marine stations with land stations are not perfectly matching, while results of the relative OBSs orientations are consistent. We estimated uncertainties by random sampling of the waveforms with a time window of variable length. For the Kuril Island seismic event we perturbed (2000 times) the length and the position of the sampling window

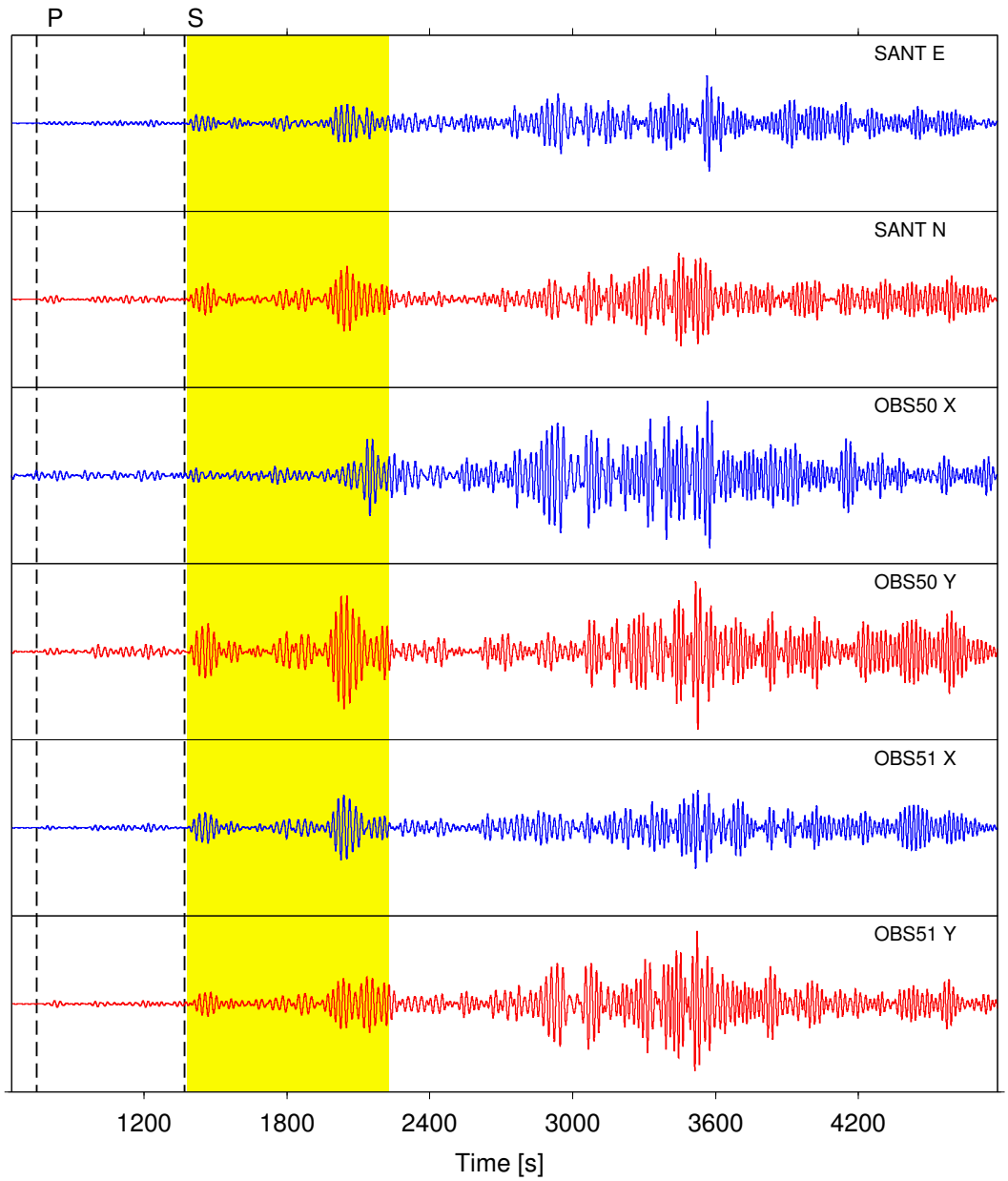


Figure 2.8: The Kuril island seismic event recorded by OBS 50, OBS 51 and land station SANT (time axis with respect to the event origin time). Yellow window highlight the part of the waveforms used to show alignment results (traces were filtered using a band-pass Butterworth filter within the band 0.03-0.07 Hz).

(window length in the range 500-1000 s). We used the same window parameters of the previous event to estimate the uncertainties for the Vanuatu Island seismic events and for the joint event.

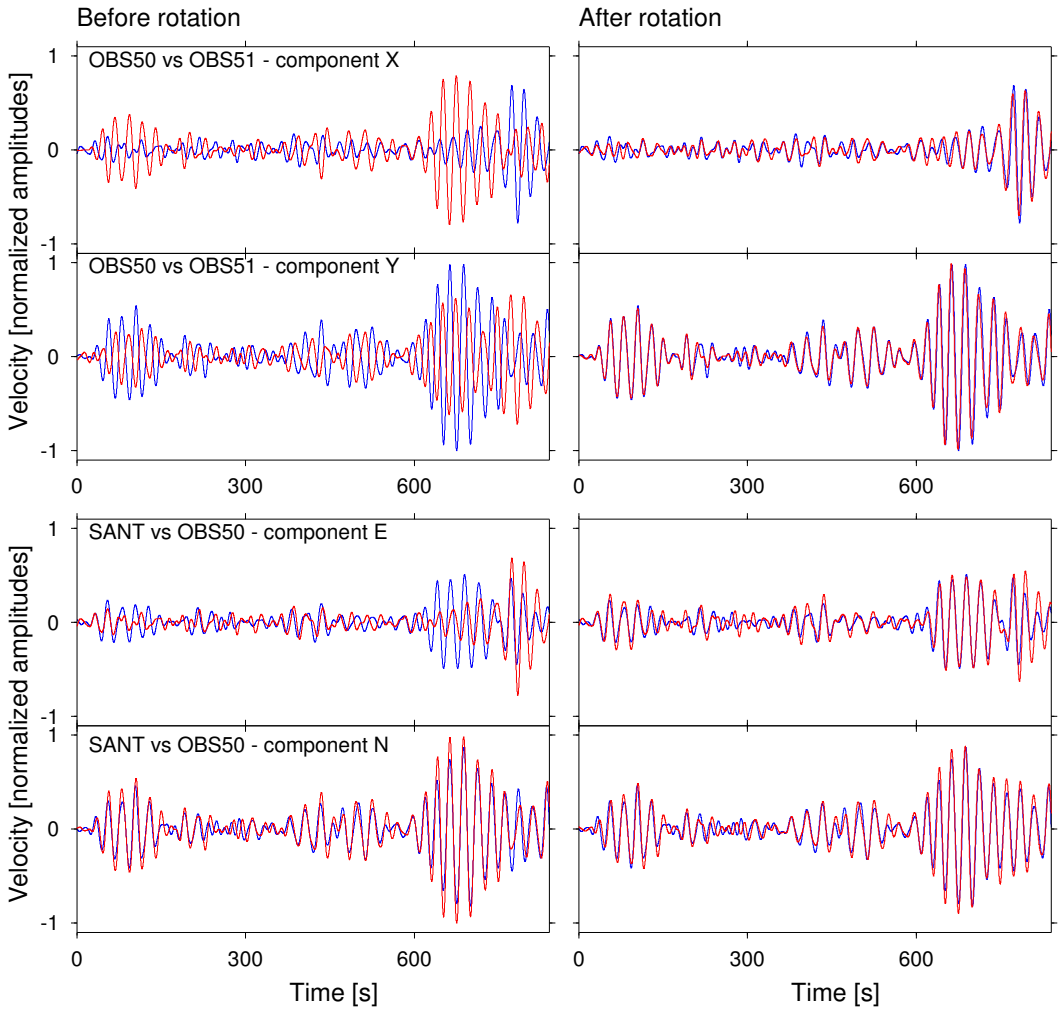


Figure 2.9: Results relative to the alignment of the OBSs and Land stations (Kuril event). Although we used the full waveform, for visualization purposes we show only the part of the waveforms relative to the yellow marked window in figure (2.8). Reference traces are blue while traces to rotate are red (traces were filtered using a band-pass Butterworth filter within the band 0.03-0.07 Hz).

2.7 Application to a seismological array
(Graefenberg, Germany)

In this part of the work we show that our method can be also applied for testing the correct orientation/alignment of multicomponent land stations in seismological arrays or temporary networks. The orientation of seismic stations may be affected by human errors or instrumental malfunction. Similar orientation problems can be encountered in presence of significant magnetic anomalies (e.g volcanic areas or building whose structure is composed by magnetic metals), where the orientation cannot be correctly performed using compass. We apply here our method to the Graefenberg array. In this case, almost all stations have been carefully oriented

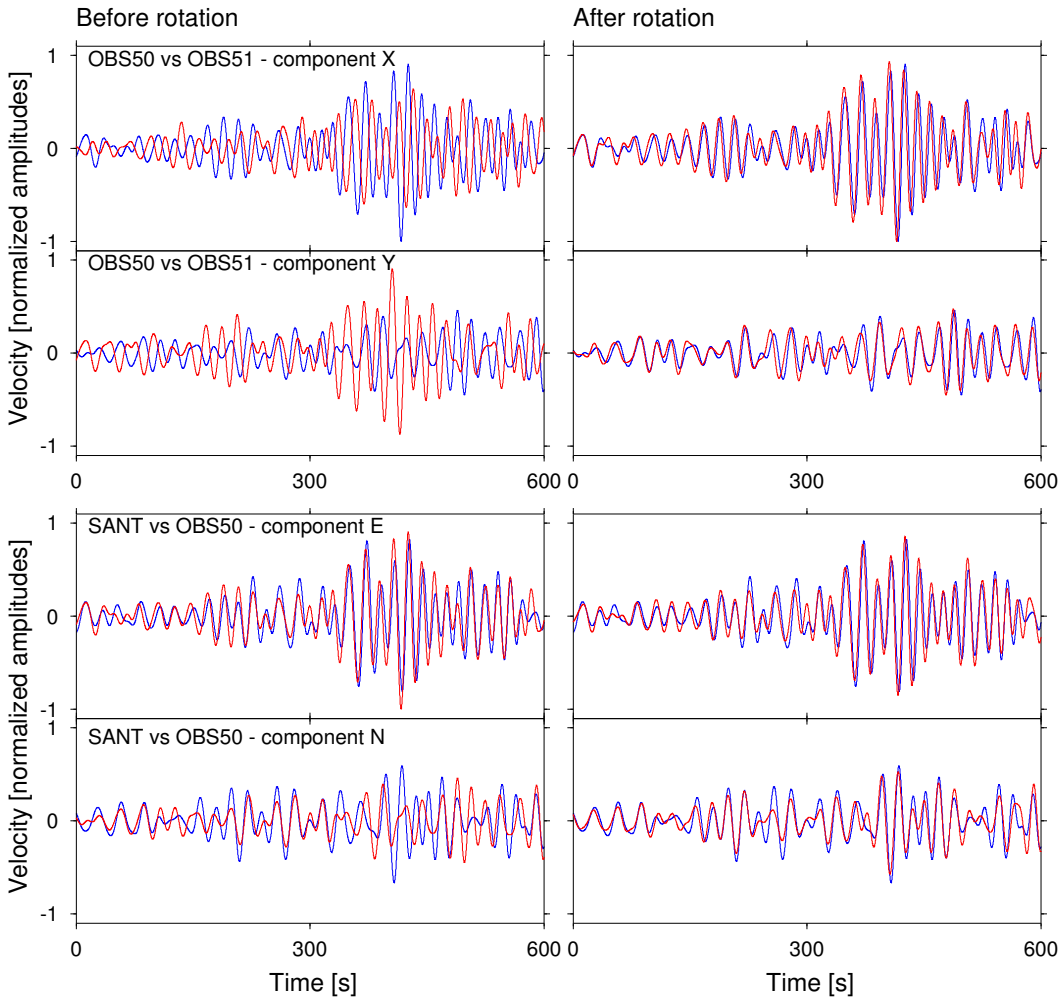


Figure 2.10: Results relative to the alignment of the OBSs and Land stations (Vanuatu event). For visualization purposes we show only a part of the recorded waveform. Reference traces are blue while traces to rotate are red. Zero of the time axis corresponds to 01:20 GMT time (traces were filtered using a band-pass Butterworth filter within the band 0.03-0.07 Hz).

[Greenhalgh and Mason, 1995]. The Graefenberg array is composed by 13 stations, located within an area of about 50 by 100km east of the city of Nuremberg (figure 2.11.a) Germany. The array became fully operational in April 1980 with ten single component stations and three stations equipped with three-components sensors (GRA1, GRB1 and GRC1). In autumn 2006 all stations were equipped with three-components very broad band (VBB) sensors [Plenefisch and Stammler, 2008]. The array is operated by the Seismologisches Zentralobservatorium (SZGRF), which is part of the Bundesanstalt fuer Geowissenschaften und Rohstoffe (BGR). The mean distance between neighbouring stations is about 15 km. Krüger and Weber [1992] studied the influence of local geological structures on the mislocation of telesesimic events using the Graefenberg array, founding that the mislocation can be reduced taking in account these structures. The effect of the geological

structures beneath the array is more evident on the whole array, rather than when considering only neighbouring stations. Since the distance between each pair of neighbouring stations is much smaller than the dominant wavelength of the wavefield, we can consider the effect of these geological structures at each pair of sensors negligible. This is confirmed by a very high wavefield similarity at each pair of stations. Our dataset consists in long period (LH channel) recordings of the June 24th, 2011, $M_w = 7.4$ Aleutian Islands earthquake (Lat. $52.0°$, Lon. $-171.8°$, Depth 46.0 km, Epicentral dist. $79°$ and Stations azimuths about $-2°$). The sampling frequency is 1 Hz. Waveforms are lowpass filtered below 0.03 Hz and then time aligned, using crosscorrelation of energy traces. We use time windows of about 8000 s, including the full waveforms of the event. Table (2.5) summarizes results for the all available stations of the array. These results show that stations GRA1, GRA2 and GRB3 are not correctly aligned with the other ones, whereas the remaining stations are correctly aligned (relative orientation angles below $5°$). In the worst case (GRA1 with respect to GRA3), the misalignment is about $18°$. Figures (2.11.b) and (2.11.c) show results of the alignment both for the case of sensors with consistent orientation (GRC3-GRC4) and for the case of two sensors which are not aligned (GRA3-GRA1). The improvement of the fit after alignment, for the staion pair GRA1-GRA3, can be seen (especially at time about 2880 s on the E component and about 2640 s on the N component)in figure (2.11.c). Uncertainties were estimated by random sampling of the full waveforms with a time window of variable length. We perturbed (5000 times) the length (500-3500 s)and beginning of the sampling window. Results were confirmed, when using seismic signals from others earthquakes: the 6th July, 2011, $M_w = 7.7$, Kermadec Islands event (Lat. $-29.3°$, Lon. $-176.2°$, Depth 1.0 km, Epicentral dist. $159°$ and Stations azimuths about $-14°$) and the 9th September, 2011, $M_w = 6.4$, Vancuver Island event (Lat. $49.5°$, Lon. $126.8°$, Depth 20.0 km, Epicentral dist. $67°$ and Stations azimuths about $-40°$). The misalignment of station GRA1 can be related to the station updating carried out by the BGR in September 2010, when the orientation mark was lost and the station was reoriented (K. Stammler, pers. comm.). Before this date, the orientation of the station GRA1 is assumed correct. We used a teleseimic event occurred before September 2010 (the Kuril Island seismic event of the previous section) and found that the misalignment of the station pair GRA3-GRA1 was smaller than $2°$.

2.8 Discussion and Conclusions

Based on a complex linear least-squares approach we developed a methodology to derive the relative orientation angles between seismic sensors. Our method has been tested using both synthetic and real datasets. Results obtained are satisfactory and show that our methodology can be successfully used to find relative, and in some case absolute (OBS data), orientations of seismic sensors for different acquisition geometries and environments. Furthermore, it can be used as a tool to test quickly the correct orientation of stations for seismic arrays or temporary

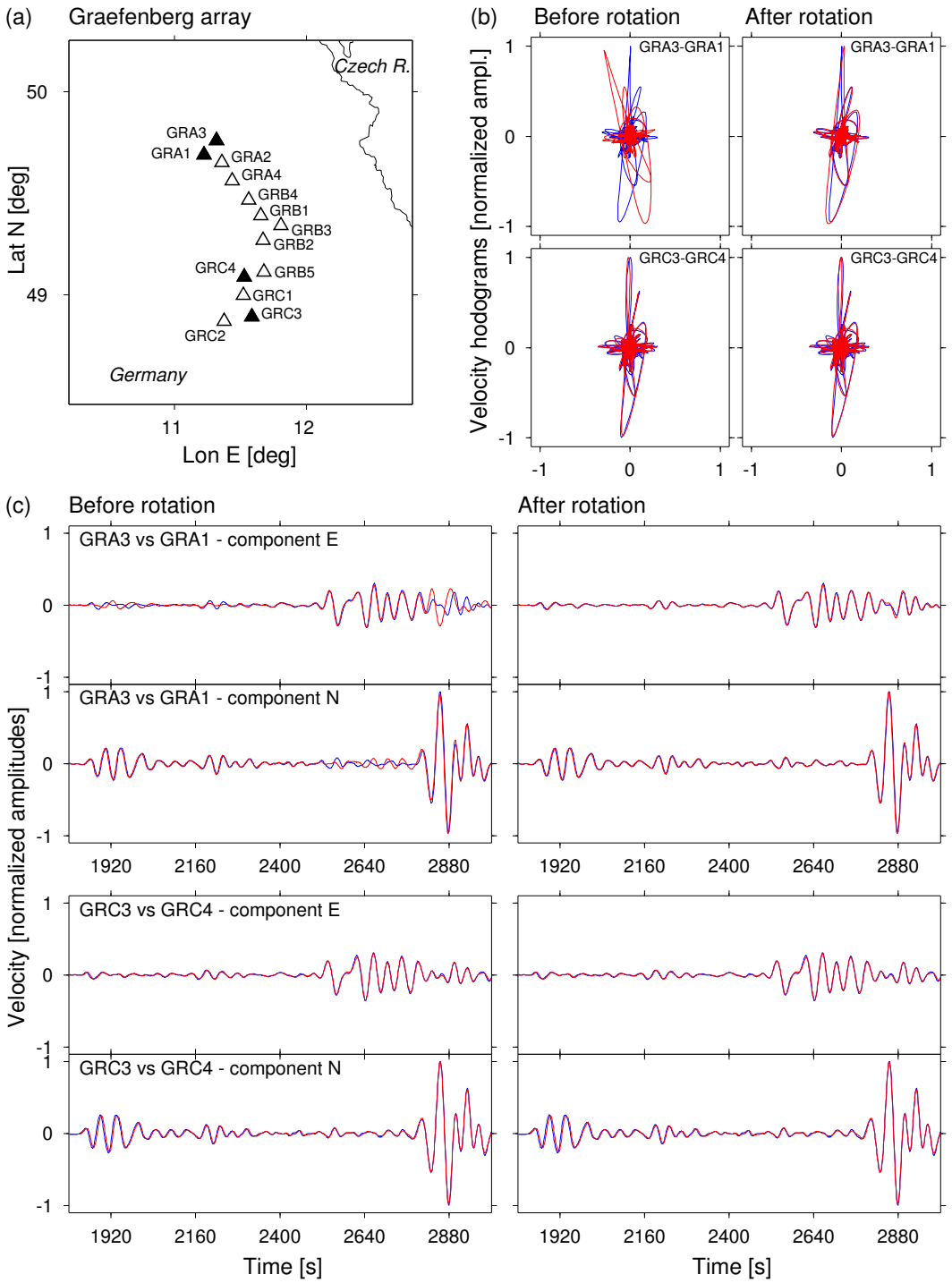


Figure 2.11: (a) Configuration of the network. (b) Complex traces (hodograms) related to the stations pairs ((blue) GRA3-GRA1 (red)) and ((blue) GRC3-GRC4 (red)), before (left) and after alignment (right). (c) Horizontal component traces of the same pairs of stations. Reference traces are blue while traces to rotate are red (traces were filtered using a low-pass Butterworth filter with corner frequency $f_c = 0.03$ Hz).

Pair of stations	Align. angles
GRA1-GRA2	$10.5^\circ \pm 0.6^\circ$
GRA2-GRA3	$7.7^\circ \pm 0.9^\circ$
GRA3-GRA4	$-2.4^\circ \pm 2.2^\circ$
GRA4-GRB1	$0.5^\circ \pm 2.3^\circ$
GRB1-GRB2	$-4.7^\circ \pm 1.1^\circ$
GRB2-GRB3	$-6.8^\circ \pm 0.4^\circ$
GRB3-GRB5	$11.7^\circ \pm 1.3^\circ$
GRB5-GRC1	$-3.2^\circ \pm 1.6^\circ$
GRC1-GRC3	$1.8^\circ \pm 0.7^\circ$
GRC3-GRC4	$-0.3^\circ \pm 2.2^\circ$
GRA3-GRA1	$-18.3^\circ \pm 0.8^\circ$

Table 2.5: Estimated rotation angles for Graefenberg array. Results in bold are related to the figure (2.11).

networks. There are several advantages in using our approach. (1) The method is not requiring a linearly polarized waveform and can be either applied to full waveform or to a shorter part of the signal (e.g. P-wave onset); while polarization based method can use only seismic phases characterized by an high rectilinearity (P or S waves), our method can be applied using both body and surface waves. (2) By solving a linear inverse problem in complex domain, our solution corresponds to the global minimum of the misfit function; furthermore, it is possible to use more seismic events (earthquakes or explosions) simultaneously to better constrain the final solution. (3) Computing relative orientation angles result faster than cross-correlation based methods, especially when dealing with a very large number of sensors. On the other hand, the main limitation of the method relies on the initial assumption we have done. Our method can be applied only when the plane wave approximation is valid and when the wavefield is very similar at each pair of sensors. Similarly to all other relative orientation methods, error propagation can have an heavy effect for sensors which are far from the reference one. If the energy traces show high similarity also between non adjacent sensors,errors can be reduced by considering more pair of sensors. When the distance between two neighbouring sensor is large (more than 50 km) there are problems to find seismic events with a suitable wavelengths. This problem can be solved using body waves of teleseismic events. In this way, since body waves related to these events arrive with a near vertical direction, the apparent wavelength should be large enough to satisfy the conditions which our method requires. The broad range of applications proofs that this methodology can be applied to different type of data and environments. Other possible applications of this method can be, for instance, the alignment of geophones in land acquisition or the sensor alignment in ocean bottom cables.

2.9 Acknowledgements

We thank the editor Dr. Eiichi Fukuyama and three anonymous reviewers for their comments and suggestions. We thank the Canadian Geological Survey for providing their data. We wish to thank Dr. Klaus Stammler and Dr. Thomas Plenefisch from BGR Hannover for the useful discussions, Dr. Martin Hensch from Nordic Volcanological Center (University of Iceland) for providing data. We further thank Dr. Eleonora Rivalta and Prof. Dario Luzio for the helpful advises during the writing of this paper. This work was realized within the research project MINE. The project MINE is part of the research and development programme GEOTECHNOLOGIEN and is funded by the German Ministry of Education and Research (BMBF). Grant of project BMBF03G0737A.

Chapter 3

Automated Seismic Event Location by Travel-Time Stacking: An Application to Mining Induced Seismicity

Authors:
Francesco Grigolii[a], Simone Cesca[a,b], Maurizio Vassallo[c], Torsten Dahm[b,a]
Journal:
Seismological Research Letters,84 (4), 666-677,
doi: 10.1785/0220120191
Status:
Published

Authors affiliations:
a) Institute of Earth and Environmental Sciences,
University of Potsdam,
Potsdam, Germany
b) GFZ (German Research Centre for Geosciences) Potsdam,
Section 2.1, Physics of earthquakes and volcanoes,
Potsdam, Germany
c) Istituto Nazionale di Geofisica e Vulcanologia (INGV),
Rome, Italy

Introduction

The automated location of seismic events is an important and challenging task in microseismic monitoring applications (e.g., to analyse induced seismicity following oil/geothermal field exploitation and mining operations), where we deal with a large number of seismic events and weak signals characterized by low signal-to-noise ratios. Given the large number of seismic events, manual location procedures are time consuming, when not feasible. Standard automated location routines require precise automated picking procedure and phases identification [Gharti et al., 2010]. These methods are, generally, modified versions of the Geiger [1910, 1912] algorithm, based on the minimization of time residuals between theoretical and observed arrival times of bodywaves (generally first P and S onsets) by iterative inversion algorithms. In the last two decades a large number of picking algorithms have been developed: while P onsets can now be accurately picked, the automatic picking of later seismic phases (including S onsets) is still problematic. Their performance is limited in presence of noisy data, when picking and phases identification might be difficult. The increasing interest on microseismic monitoring applications pushed the recent development of alternative techniques for automated seismic event location. These methods, similar to migration techniques used in reflection seismology, exploit the full waveforms and do not need any prior phase identification. Some methods are based on wavefield backpropagation, using time-reversed seismograms as sources [Gajewski and Tessmer, 2005; McMechan, 1982]: the seismic source is then located where the maximum energy focuses. These methods are computationally intensive, and energy focusing can be ambiguous with noisy data and very heterogeneous models [Gharti et al., 2010]. Other location methods use the coherence of a selected wave package recorded at different stations (e.g. Rubinstein and Beroza [2007]); this approach successfully located seismic events in volcanic environments [Wassermann, 1997]. Ekström [2006] applied a waveform stacking of low-pass filtered seismograms and surface waves to detect and locate teleseismic events; similar methods have been also used to map the rupture propagation of large earthquakes [Kao and Shan, 2007; Krüger and Ohrnberger, 2005a,b; Maercklin et al., 2012]. Finally, some techniques are based on the idea of delay and sum of seismic waveforms. Kao and Shan [2004, 2007] developed the Source Scanning Algorithm (SSA). The source location is performed using a britghtness function, which is obtained by stacking the absolute amplitudes of normalized seismograms recorded at different stations. Baker et al. [2005] developed a similar method, but stacking the seismogram envelope at first P onsets, instead of the normalized trace; the method was tested for real time seismic monitoring in southern California. For microseismic applications, Drew et al. [2005] stacked the product of P and S signal-to-noise ratios at computed arrival times to obtain a coalescence map in a 4D space: the event is detected when the coalescence function is higher than a fixed threshold, and then located through the polarization analysis of P waves. This method was applied to locate microseismic events recorded by three-components geophones deployed

in a vertical array. Gharti et al. [2010] proposed to rotate seismic traces to the ray coordinate, compute the envelope and finally perform stacking along P and S arrival times. All these stacking methods have the main advantages of not requiring phase picking nor their identification. Moreover, they exploit the waveform information without the need to compute synthetic seismograms. In this paper we present a modified version of the SSA developed by Kao and Shan [2004] and improved by Liao et al. [2012]. While the SSA method is designed to locate typical tremor events with emergent waveforms and is based on the stacking of the absolute amplitudes at all stations at their respective predicted arrival times, our approach is based on the stacking of the STA/LTA (Short Time Average to Long Time Average ratio) traces at all stations using both P and S phases. The chief benefits of our method are: 1) phase identification and picking are not required, and 2) it can an be successfully applied to noisy data. STA/LTA traces have been used for location purposes in the past, e.g. by Withers et al. [1999], who proposed a correlation based method to locate seismic events at a regional scale. To test the performance of our method, we first apply it to different synthetic datasets. Then, we discuss an application to real data, and successfully locate 391 microseismic events (with magnitudes MI between 0.5 and 2.0) induced by coal mining in Germany. To validate our method we compare our locations with those obtained by a manual location procedure.

3.1 Methodology

Location process

Let us suppose that a seismic event is recorded by n three-component seismic stations. In order to locate the event we first define a 3D cartesian space containing the whole seismogenic region. For each potential source location (x, y, z) we compute the theoretical arrival times for the first P ($\tau_i^p(x, y, z)$) and S ($\tau_i^s(x, y, z)$) onsets at all n stations (i is the station index). We then define τ_{min} and τ_{max} as:

$$\tau_{min}(x, y, z) = \min(\{\tau_i^p(x, y, z)\}_{i=1}^n) \ , \tag{3.1}$$

$$\tau_{max}(x, y, z) = \max(\{\tau_i^s(x, y, z)\}_{i=1}^n) \ , \tag{3.2}$$

which define for each source location the minimum P and maximum S arrival time in the network. Using the previous equations we introduce $\theta_i^p(x, y, z, t)$ and $\theta_i^s(x, y, z, t)$:

$$\theta_i^p(x, y, z, t) = \tau_i^p(x, y, z) - \tau_{min}(x, y, z) + t = \tau_i^p(x, y, z) + t_0(x, y, z, t) \ , \tag{3.3}$$

$$\theta_i^s(x, y, z, t) = \tau_i^s(x, y, z) - \tau_{min}(x, y, z) + t = \tau_i^s(x, y, z) + t_0(x, y, z, t) \ , \tag{3.4}$$

with $0 \leq t \leq t_M(x, y, z)$ and $t_M(x, y, z)$ defined as:

$$t_M(x, y, z) = t_{end} - (\tau_{max}(x, y, z) - \tau_{min}(x, y, z)) \ . \tag{3.5}$$

$\theta_i(x, y, z, t)$ are the computed arrival times at station i (for P and S first onsets) relative to τ_{min} and shifted by a delay t. t_{end} is the ending time of the observed traces. $t_0(x, y, z, t)$ is the origin time of the event which is equal to $t - \tau_{min}(x, y, z)$.

We first calculate the horizontal ($H(t) = n(t)^2 + e(t)^2$) and vertical ($V(t) = z(t)^2$) energy of the waveforms recorded at each three-component station ($n(t)$, $e(t)$, and $z(t)$ are the seismic traces related to the north, east and vertical components respectively), then we compute the recursive STA/LTA [Withers et al., 1998] traces using the H and V energy traces as characteristic functions separately. The recursive STA/LTA provides reduced memory requirement and is smoother than standard STA/LTA in absence of signal [Withers et al., 1998]. According to Withers et al. [1998], the characteristic decaying factor is set to $1/n_{short}$ for the short time window and to $1/n_{long}$ for the long one (n_{short} and n_{long} are, respectively, the length in samples of the short and long time windows). We evaluate the coherence functions C^p and C^s at each point (x, y, z), by using the following equations:

$$C^p(x, y, z, t) = \int \sum_{i=1}^{n} W_i^Z(\tau)\delta(\tau - \theta_i^p(x, y, z, t))d\tau \ , \tag{3.6}$$

$$C^s(x, y, z, t) = \int \sum_{i=1}^{n} W_i^H(\tau)\delta(\tau - \theta_i^s(x, y, z, t))d\tau \ , \tag{3.7}$$

where W_i^Z is the normalized STA/LTA trace of the vertical energy related to the i-th station (in analogy W_i^H is the normalized STA/LTA trace of the horizontal energy related to the i-th station) and δ is the Dirac's delta. To obtain smoother results (e.g. when the velocity model is poorly known) the Dirac's delta can be replaced by rectangular or triangular function centred at $\theta_i^p(x, y, z, t)$ and $\theta_i^s(x, y, z, t)$. Normalization is required to take care of propagation effect (geometrical spreading and intrinsic attenuation) in order to avoid that stations close to the source dominate the stacking. The coherence function is finally defined as:

$$C(x, y, z, t) = \frac{\sqrt{C^p(x, y, z, t)C^s(x, y, z, t)}}{n} \ . \tag{3.8}$$

From the previous equation it is clear that $C(x, y, z, t)$ is a bounded function whose theoretical bounds are 0 (no coherence) and 1 (perfect coherence for both P ans S first arrivals phases). The hypocentre coordinates $(\hat{x}, \hat{y}, \hat{z})$ and origin time $\hat{t}_0$ of the seismic events corresponds with the location of the coherence function maximum.

$$C(\hat{x}, \hat{y}, \hat{z}, \hat{t}) = \max\{C(x, y, z, t)\} \ , \tag{3.9}$$

from $\hat{t}$ we can derive the origin time by:

$$\hat{t}_0 = \hat{t} - \tau_{min}(\hat{x}, \hat{y}, \hat{z}) \ . \tag{3.10}$$

Figure 3.1 shows a flow-diagram which describes how the location algorithm works.

Uncertainty estimation

Uncertainties are estimated by (a) data perturbation and (b) a jack-knife method. In the first approach we iteratively repeat the location procedure after perturbing the lengths of both short and long time windows of the STA/LTA. In the second case, we repeat the location procedure after removing, each time, all traces related to a different station. In both cases, by repeating this procedure k times we obtain k estimations for each model parameter (i.e. the hypocentre location). Then, we use a weighted mean and standard deviation to compute the best estimation and its uncertainty for each model parameter. The largest value of the coherence function related to the h-th iteration is then used as weighting factor:

$$q_h = C_h(\hat{x}_h, \hat{y}_h, \hat{z}_h, \hat{t}_{0h}) \ , \tag{3.11}$$

where C_h, q_h and $(\hat{x}_h, \hat{y}_h, \hat{z}_h, \hat{t}_{0h})$ are respectively the coherence function, the weighting factor and the hypocentre estimation related to the h-th solution. After the k-th iteration we can compute the normalized weighting factor:

$$Q_h = \frac{q_h}{\sum_{m=1}^{k} q_m} \ . \tag{3.12}$$

The weighted average of all k solutions is the best estimation of the hypocentral coordinates:

$$\bar{x}_i = \sum_{m=1}^{k} Q_m \hat{x}_{im} \quad \text{with} \quad i = 1, 2, 3 \ , \tag{3.13}$$

where $\hat{x}_{1m} = \hat{x}_m$, $\hat{x}_{2m} = \hat{y}_m$, $\hat{x}_{3m} = \hat{z}_m$ and $\hat{x}_{4m} = \hat{t}_{0m}$, Uncertainties can be estimated through the weighted covariance matrix R defined as:

$$R_{ij} = \frac{\sum_{m=1}^{k} Q_m}{\left(\sum_{m=1}^{k} Q_m\right)^2 - \sum_{m=1}^{k} Q_m^2} \sum_{m=1}^{k} Q_m \left(\hat{x}_{im} - \bar{x}_i\right)\left(\hat{x}_{jm} - \bar{x}_j\right) \ . \tag{3.14}$$

with $i = (1, 2, 3, 4)$ and $j = (1, 2, 3, 4)$.

3.2 Synthetic tests

In order to test the performance of our method we first applied it to a synthetic dataset resembling an existing network geometry (Hammnet, Ruhr region, Germany; Bischoff et al. [2010]), composed by 15 three-components surface stations. We consider 4 synthetic events with different locations and normal faulting focal mechanisms (Figure 4.10) with different striking, which resemble the most common focal mechanisms in the area according to moment tensor inversion results [Sen et al., 2012]. All events have a depth of 1000 m, slightly above the mining level, as found by manual location. For all considered source models, we computed synthetic waveforms using a layered velocity model (Figure 4.10-d).

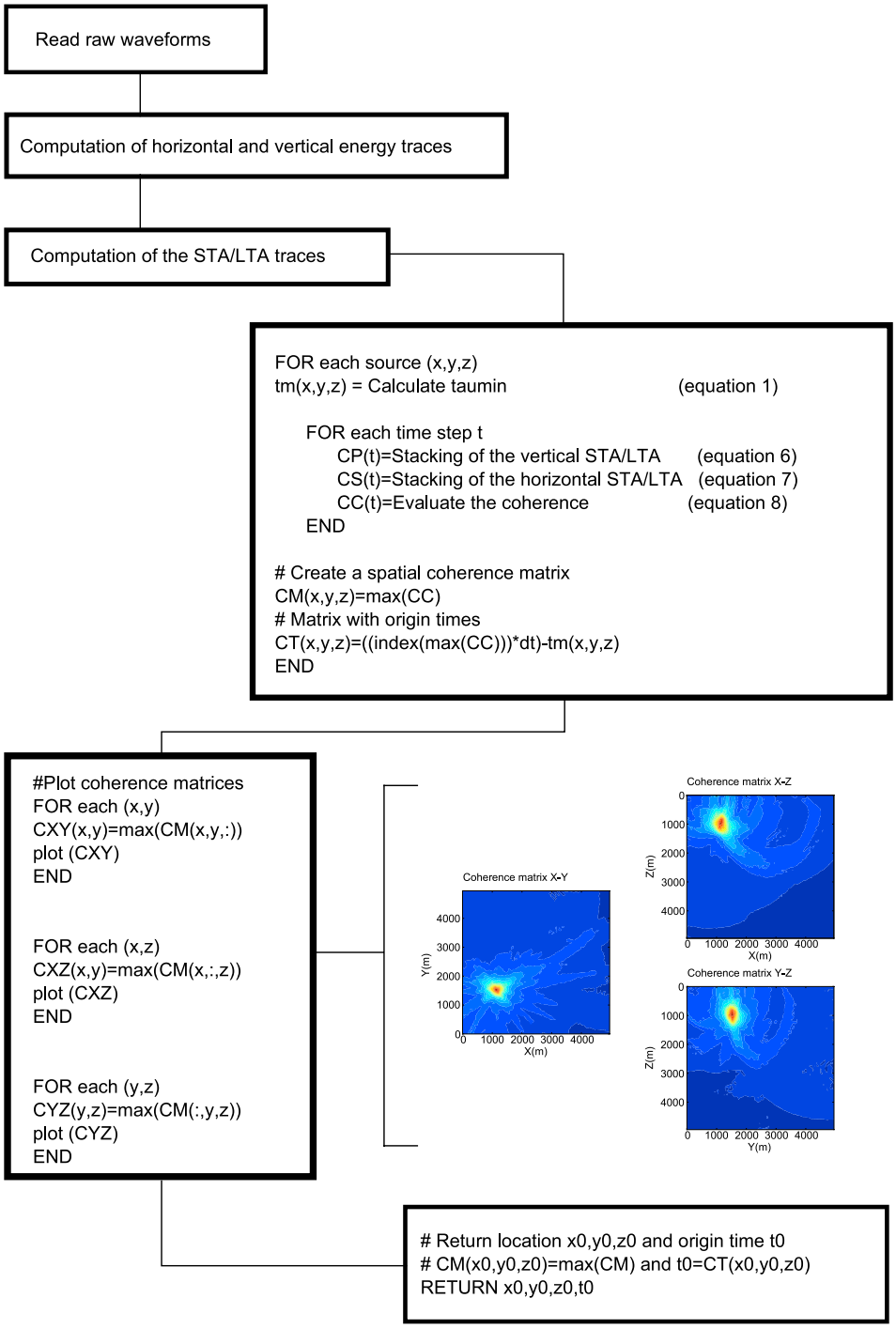


Figure 3.1: Flow diagram of the location algorithm. Coherence matrix XY is obtained by projecting, for each X-Y, its maximum along Z (coherence matrices XZ and YZ are obtained in a similar way).

Synthetic waveforms have been generated using the software Qseis [Wang, 1999], with a sampling period of 0.01 s. Theoretical travel times for both models have

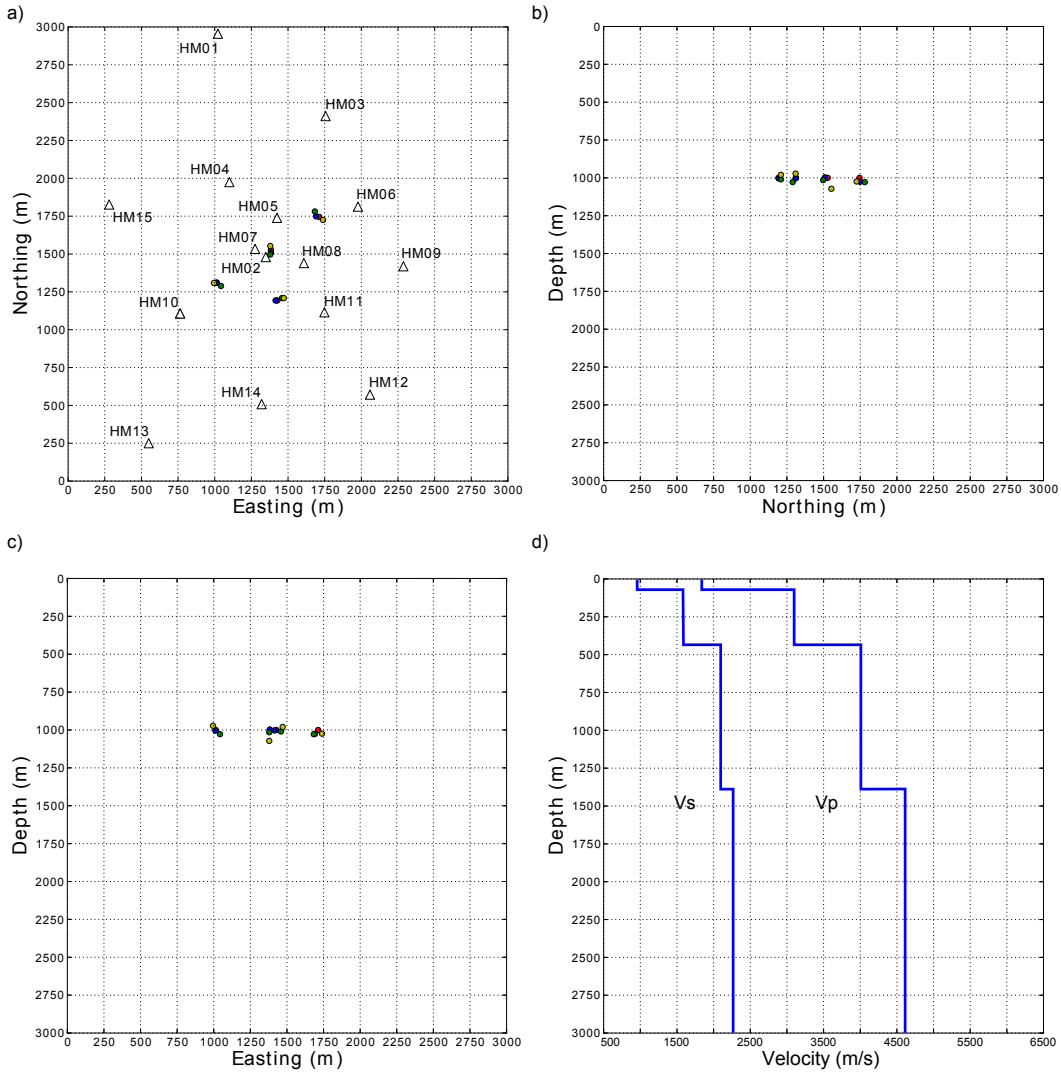


Figure 3.2: Geometry of the network (a) and results (a,b,c) for synthetic data with different noise levels: True locations are represented by red dots, location results with data affected by noise are represented by blue (noise level at 10%), green (noise level at 30%) and yellow (noise level at 70%) spots. Subfigure (d) represents the layered velocity model we used to perform location. The reference point $(x, y) = (0, 0)$ corresponds to $x = 411617$ and $y = 5722111$ (in the UTM system)

been computed using the finite difference code developed by Podvin and Lecomte [1991]. Finally, velocity synthetic waveforms were contaminated using different noise levels (10%, 30% and 70% of the maximum waveforms amplitude, Figure 4.11). Coherence matrices (related to the same data shown previously) show for all cases (even with noise level at 70%) a clear maximum within 1 grid point from the true location (Figure 4.12). Figure 3.5 shows the coherence values $C(\hat{x}, \hat{y}, \hat{z}, t)$ for different time steps t, where $\hat{x}$, $\hat{y}$ and $\hat{z}$ are the estimated coordinates of the seismic event considered in the previous figure. The scanned source region has an

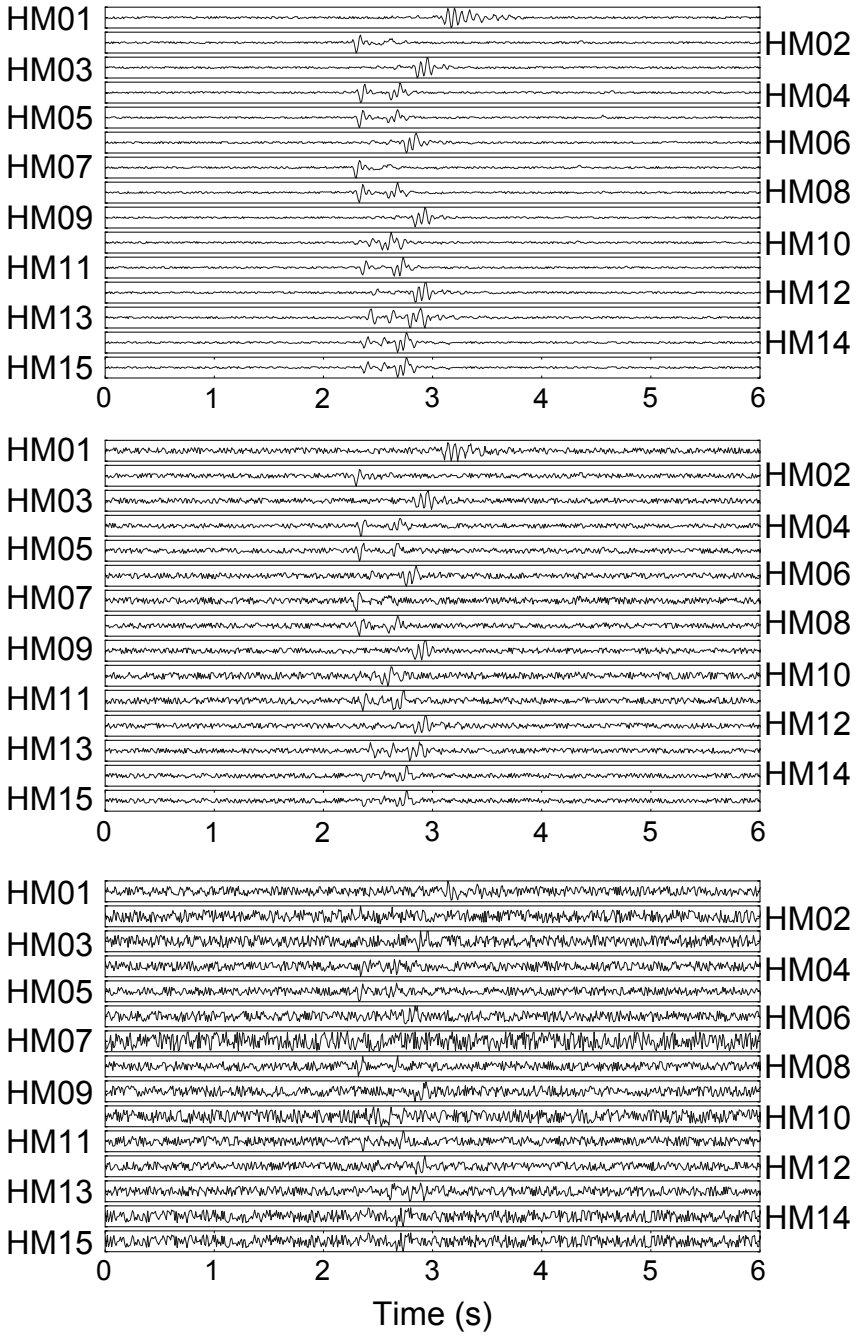


Figure 3.3: Synthetic traces (Vertical component) with different noise levels: noise level at 10% of the maximum amplitude (top figure), noise level at 30% of the maximum amplitude (middle figure) and noise level at 70% of the maximum amplitude (bottom figure).

extension of 5 x 5 x 5 km^3, and a 50 m grid spacing. In all cases results (Table enclosed in the electronic supplement to this manuscript and Figure 4.10) confirmed the robustness of our approach, even with very noisy traces. Uncertainties have been estimated by jack-knife and perturbing (100 times) the length of the short

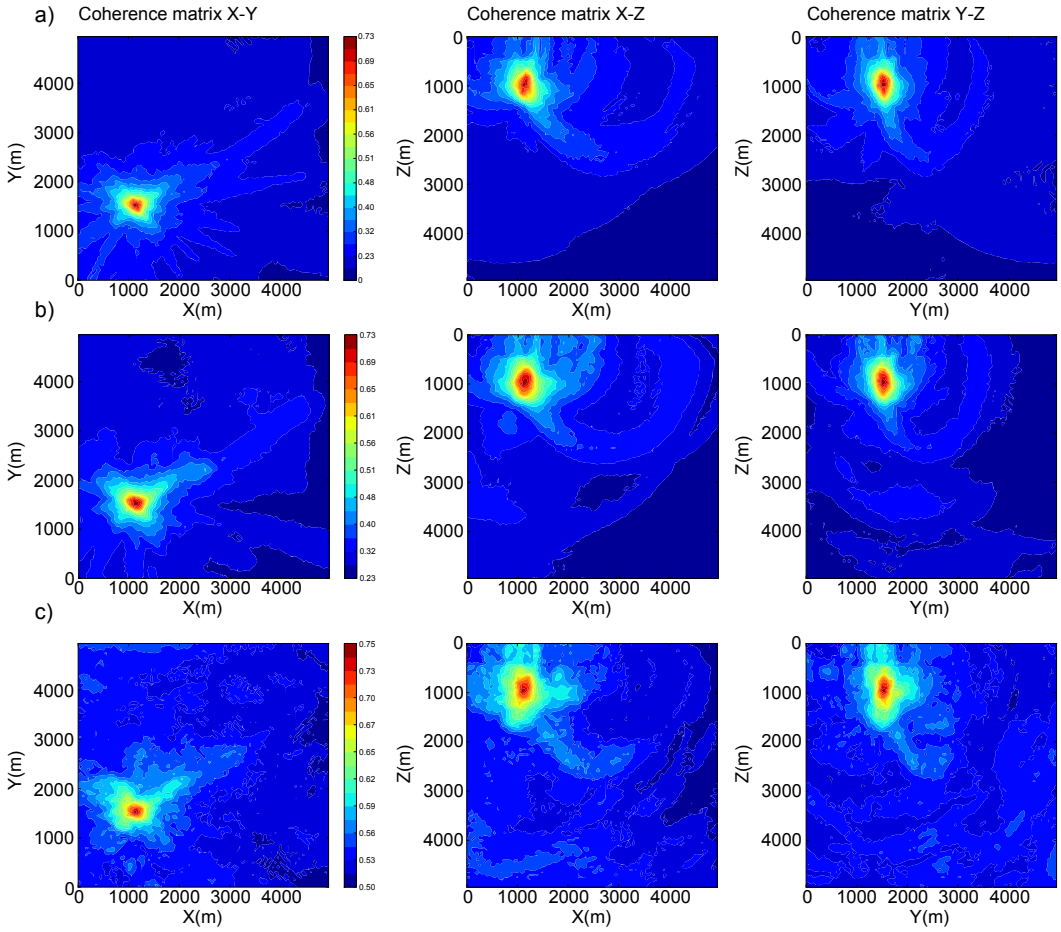


Figure 3.4: Coherence matrices related to one of the synthetic event shown previously: noise level at 10% (a), noise level at 30% (b) and noise level at 70% (c). Coherence matrix XY is obtained by projecting, for each X-Y, its maximum along Z (coherence matrices XZ and YZ are obtained in a similar way). Coherence values are represented in color scale.

and long time windows (the short time window has random length in the range 0.1-0.3 s, the long time window is 2 times longer). In order to test the location accuracy even using a wrong velocity model, we located synthetic events using an homogeneous model. We obtained the homogeneous model by weighted average of the layered one. While location results for epicentral coordinates are consistent with true locations (within 50 m), concerning depth estimation we observed larger errors (for all events we obtained results about 200 m deeper than true locations) and uncertainties (about 150 m). To further validate our method we extended our tests, locating 200 synthetic events, with random hypocentral locations and focal mechanisms. Synthetic waveforms have been contaminated by noise, considering two different noise levels (30% and 70% of the maximum signal amplitude). Again, we chose a short time window length within the range 0.1-0.3 s and a long time window 2 times longer. In both cases results confirmed the robustness of our approach: about 90% of the events are located within 80 and 150 m from their

true location, for data with noise level at 30% and at 70% respectively (Figures 3.10-b and 3.10-c). Concerning origin times, for most of the events (about 90%) the absolute time shifts with respect to the true origin time are within 0.035 and 0.060 s for data with noise level at 30% and 70% respectively. Figure S1 in the electronic supplement shows a comparison between the locations retrieved using our method and the true ones.

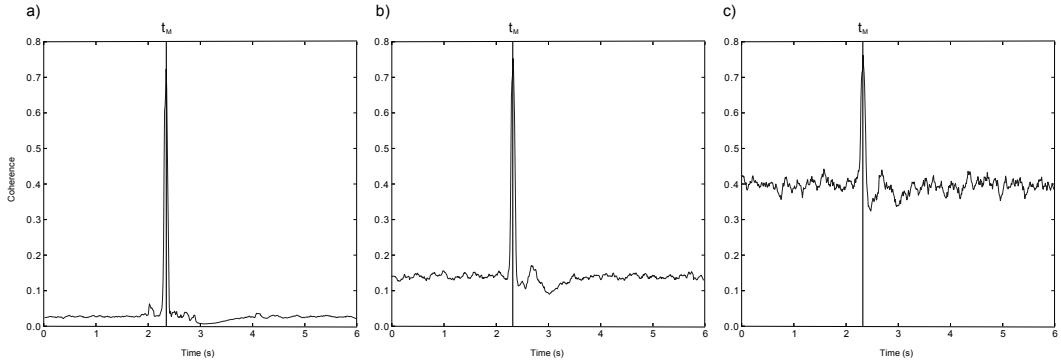


Figure 3.5: Coherence values $C(\hat{x}, \hat{y}, \hat{z}, t)$ for different time steps t ($\hat{x}$, $\hat{y}$ and $\hat{z}$ are the estimated coordinates of the seismic event) considering noise levels at 10% (a), 30% (b) and 70% (c). Coherence maximum corresponding to the time step t_M is related to the origin time of the seismic event by $t_0 = t_M - \tau_{min}(\hat{x}, \hat{y}, \hat{z})$.

3.3 Application to real data

Seismicity in the tectonically inactive Ruhr region is induced by deep coal mining. Since 1983 mining induced seismicity has continuously been monitored by the Ruhr University Bochum. A small scale network, HAMNET, was deployed in 2006 to monitor and analyse seismicity induced by the longwall mining of a single panel close to the town of Hamm. The network consisted of 9 short-period (Mark L-4C-3D, 1 Hz) and 6 broadband stations (5 Guralp CMG, 60 s and 1 Trillium 40, 40 s), installed at the surface, covering a region of about 3 x 2 km^2 (Figure 4.10-a). Data sampling rate was 200 Hz. With this setup, and using a standard STA/LTA detection algorithm, Bischoff et al. [2010] identified more than 7000 events along a time period of about 13 months in the years 2006-2007. Source location was performed by the inversion of manual picked P-wave first onsets, based on the assumption of a homogeneous halfspace velocity model ($V_p = 3880$ m/s). Estimated MI magnitudes range between -2.0 and 2.0. We focus here on a smaller dataset, composed by 391 events, all those with magnitudes above MI 0.5. An overview on the seismicity in the Hamm region is given by Bischoff et al. [2010], focal mechanisms for largest events are discussed in Sen et al. [2012]. Figure 3.6 shows a sample (M_l=1.1 occurred on October 4th, 2006 at 13.50.30 GMT) of the waveform recorded at few selected stations and their STA/LTA traces (Figure 3.7). We apply our automated location technique on cut velocity

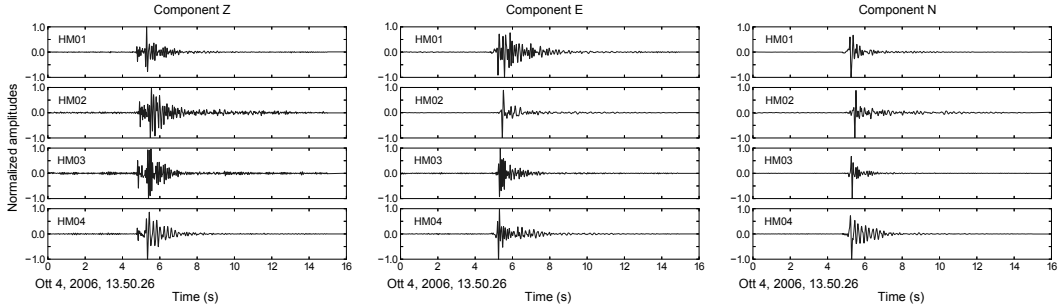


Figure 3.6: Three component velocity traces of a microseismic event occurred on October 4^{th}, 2006 at 13.50.30 GMT, recorded by four station of the network.

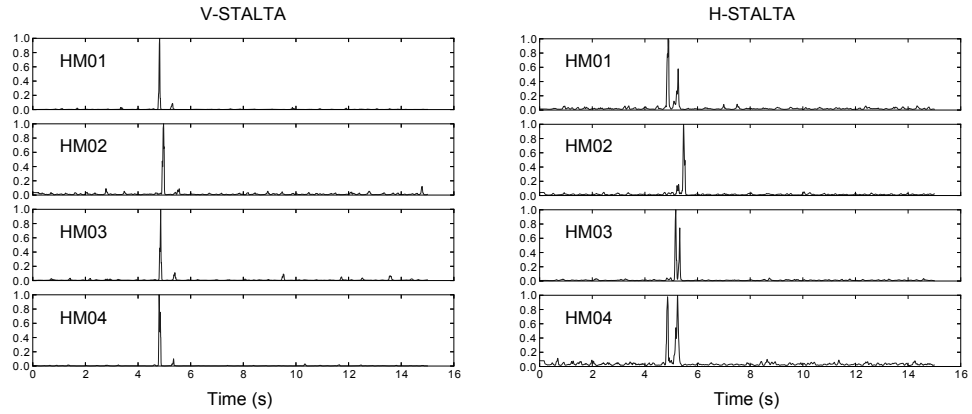


Figure 3.7: Vertical and horizontal STALTA traces (normalized) related to the waveforms shown in Figure 3.6.

waveforms. Traces start few seconds before the event trigger and have a length of about 16 seconds. We located 391 seismic events by stacking the chosen characteristic functions along both P and S arrival times. In order to show the quality of our results we compared the coherence matrices obtained using P arrival times only with those obtained using both P and S arrival times. Figure 3.8 shows a comparison between: (a) coherence matrices computed considering P arrival times only and an homogeneous velocity model, (b) coherence matrices computed considering both P and S arrival times and an homogeneous velocity model, (c) coherence matrices computed considering P arrival times only but using a layered velocity model and (d) coherence matrices computed considering both P and S arrival times and, also in this case, a layered velocity model. When we locate seismic events using P arrival time only the trade-off between origin time and depth heavily affects the quality of results (Figures 3.8-a and 3.8-c). Comparison between the figures 3.8-b and 3.8-d and the previous ones shows clearly how the location performance increases. We also compare source locations obtained both assuming the reference homogeneous velocity model and an improved 1D velocity model. The adoption of the layered model significantly improves the quality of our

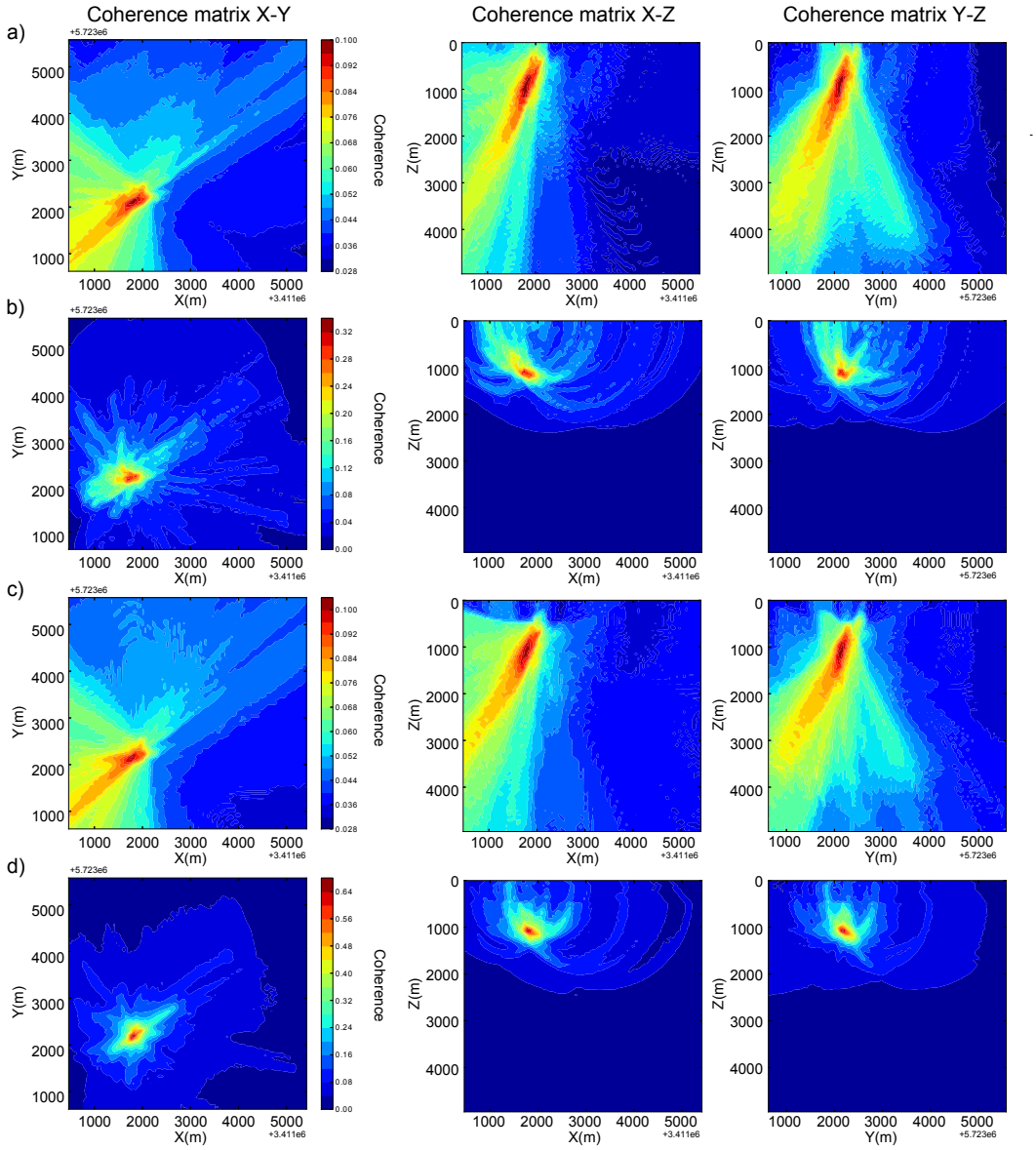


Figure 3.8: Coherence matrices obtained by stacking the STA/LTA traces using: a) only P arrival times for an homogeneous velocity model, b) both P and S arrival times for an homogeneous velocity model, c) only P arrival times for a layered velocity model and d) both P and S arrival times for a layered velocity model. Coherence matrix XY is obtained by projecting, for each X-Y, its maximum along Z (coherence matrices XZ and YZ are obtained in a similar way). Coherence values are represented in color scale. (Coordinates are in the Gauss-Krieger system).

results (this is confirmed by the analysis of the full dataset): the coherence values at the hypothetical hypocenters are larger for the layered models, and uncertainties are smaller (Figures 3.8-a and 3.8-b, Figures 3.8-c and 3.8-d and Figure 3.9-d). Events locations cluster in the region of active mining with distances up to 500 m from the panel (Figure 3.9). Only three events are located at further distance. The comparison of our results with those obtained through a manual location

procedure are summarized in Figure 3.10, where we show a cumulative plot of the events located within a given distance from the locations retrieved manually. Note that our solutions make use of both P and S arrival information, whereas reference solutions were derived based on P arrival times only. Results indicate that a general agreement between both locations, with more than 90% of the events within 200 m from the manually located sources (Figure 3.10-a). Seismic events results are more clustered in depth than locations retrieved manually. The comparison between real data results (Figure 3.10-a) and synthetic ones (Figures 3.10-b and 3.10-c) shows that event location/time error for the real dataset is comparable to the error related to the synthetic results with a noise level of 70%. For about 90% of the events we obtained an error within 200 m and 0.06 s for the real data against 150 m and 0.06 s for the synthetics (noise level 70%). Also in this case uncertainties have been estimated by jack-knife and perturbing (25 times) the length of the short and long time windows (the short time window has random length in the range 0.1-0.3 s, the long time window is 2 times longer), about 90% of the events have uncertainties of about 50 m for epicentral coordinates and about 100 m for depth.

3.4 Conclusions

We developed a methodology to locate seismic events based on waveform stacking. The proposed method is fully automatic and requires only few control parameters: the range of variation, in terms of length, for the short time window and for the long one. These two parameters can be chosen following a trial and error approach on a small subset of events, taking in account the dominant frequency of the recorded waveforms related to the seismic event we plan to locate (for all applications introduced in this study we used a short time window length in the range 0.1-0.3 s and a long time window 2 times longer). Our methodology is designed to work with dense seismic networks of 10 or more stations at very local scales and make use of cut seismic traces related to triggered events as input data. We have successfully tested our approach on both synthetic and real datasets. Results obtained with synthetics show that our methodology is reliable even with noisy traces. The location of real microseismic events was successfully determined and results are in agreement with those based on standard phase picking. We remind that our locations are based on both P and S phases, while the manual locations have been obtained by using only the P onsets time. Uncertainties provide informations about the quality and stability of the solutions (a larger uncertainties mean that location is unstable and for each perturbation the location might have large variations). However the lowest value that uncertainties can assume is constrained by the adopted grid spacing. The proposed approach has several advantages: (1) the method is completely automated, (2) it is picking free and does not require phase association, (3) it is robust and it has good performance also with low-quality data, (4) it consider P and S arrivals for location and therefore improves resolution in comparison to sole P arrival methods, (5) it exploits the

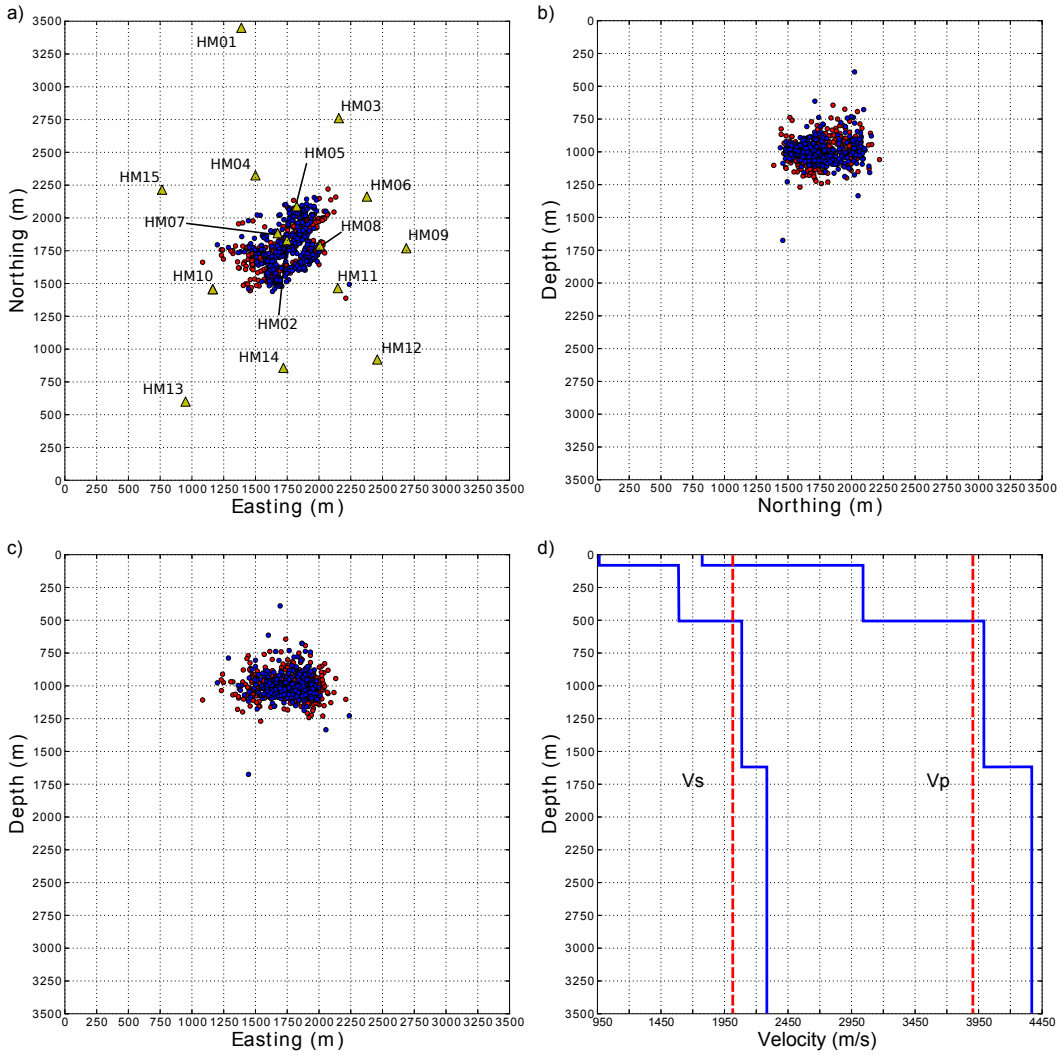


Figure 3.9: Geometry of the network (a) and location results (a,b,c) for all 391 seismic events (blue dots) and comparison with results obtained using a manual location procedure (red dots). Subfigure (d) represents a comparison between the homogenous velocity model used by the University of Bochum (red dashed lines) and the layered model we used to retrieve locations (blue lines). The reference point $(x, y) = (0, 0)$ corresponds to $x = 3411193$ and $y = 5723463$ (in the Gauss-Krieger system).

waveform information content without the need of synthetic seismograms computation. On the other hand, similarly to all other migration based methods it is computationally intensive and the location process require more computing time than standard approaches. The location process (without uncertainty estimation) of a seismic event recorded by 15 stations (each trace consists of 3000 samples) takes about 1 minute on a 2.6 Ghz Intel Xeon CPU. Since our method is based on the waveform stacking along theoretical traveltimes, inaccurate velocity models can affect the location performance. Our python location module, LOKI (LOcation of seismic events trough travetlime stacKIng) make use of the OBSPY library

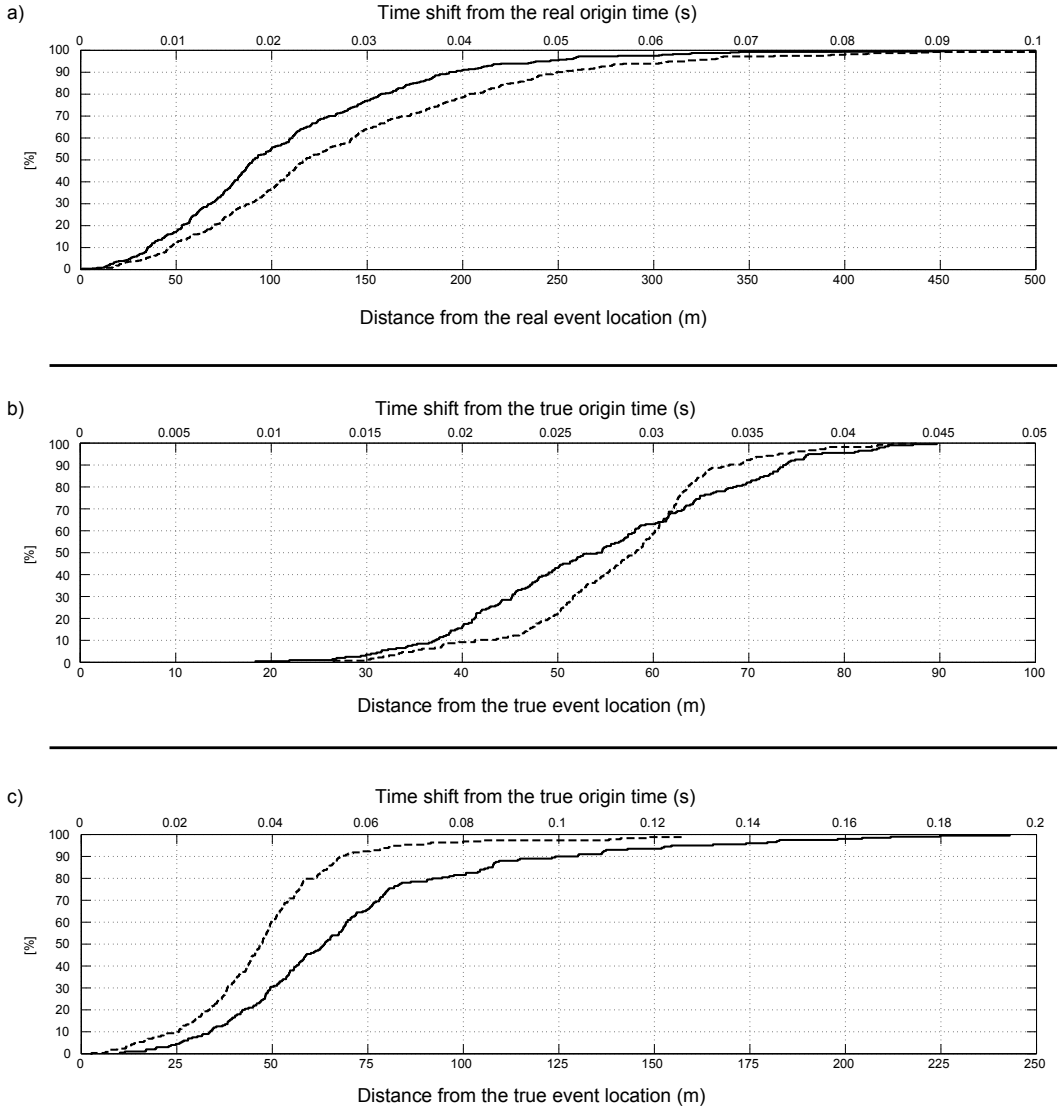


Figure 3.10: Solid line represents the percent of events located within a given distance from the reference locations (true locations for synthetics and manual ones for real data), while dashed line represents the percent of events with a given time shift (absolute value) respect to their origin times. Plot (a) shows the results for the real 391 seismic events, reference locations and origin times are based on the results obtained by the University of Bochum through a manual procedure. In analogy, plot (b) and (c) show the results related to the 200 synthetic events with noise level at 30% and 70% respectively.

[Beyreuther et al., 2010] for reading different data formats, is open source and freely available at the project MINE web page http://mine.zmaw.de.

3.5 Acknowledgements

We thank the editors Dr. J. M. Lees, the associate editor Dr. E. M. Thomson and an anonymous reviewer for their useful comments. We wish to thank Prof. Aldo Zollo, Dr. Antonio Emolo, Dr. Nils Maerklin, Dr. Luca Elia and the Rissclab team for their suggestions and for the useful discussions. We thank Dr. Leo Eisner from Czech Academy of Sciences for comments and suggestions and Dr. Sebastian Heimann for his support during the development of the code. This work has been realized within the FP7 EU research project NERA and the research project MINE. The project MINE is part of the research and development programme GEOTECHNOLOGIEN and is funded by the German Ministry of Education and Research (BMBF). Grant of project BMBF03G0737A.

3.6 Electronic supplement

Supplementary table include:
- A detailed location results for the 4 synthetic tests introduced in the manuscript (Table 3.1).
Supplementary figures show:
- Location results for 200 synthetic events (Figure 3.11)
- Cumulative plots showing the location misfit in space and time (Figures 3.12 and 3.13)

Results with noise level at 10%, 30% and 70%			
Event No. 1			
$x_{true} = 412631$	$y_{true} = 5723423$	$z_{true} = 1000$	$t0_{true} = 2.00$
$x_{10} = 412628 \pm 50$	$y_{10} = 5723420 \pm 50$	$z_{10} = 1004 \pm 50$	$t0_{10} = 2.01 \pm 0.01$
$x_{30} = 412660 \pm 50$	$y_{30} = 5723400 \pm 50$	$z_{30} = 1028 \pm 53$	$t0_{30} = 2.01 \pm 0.03$
$x_{70} = 412612 \pm 50$	$y_{70} = 5723420 \pm 50$	$z_{70} = 972 \pm 50$	$t0_{70} = 2.03 \pm 0.05$
Event No. 2			
$x_{true} = 413000$	$y_{true} = 5723639$	$z_{true} = 1000$	$t0_{true} = 2.00$
$x_{10} = 413000 \pm 50$	$y_{10} = 5723620 \pm 50$	$z_{10} = 997 \pm 50$	$t0_{10} = 2.00 \pm 0.02$
$x_{30} = 412996 \pm 50$	$y_{30} = 5723608 \pm 50$	$z_{30} = 1015 \pm 50$	$t0_{30} = 2.01 \pm 0.05$
$x_{70} = 412996 \pm 50$	$y_{70} = 5723664 \pm 50$	$z_{70} = 1072 \pm 75$	$t0_{70} = 1.98 \pm 0.04$
Event No. 3			
$x_{true} = 413044$	$y_{true} = 5723304$	$z_{true} = 1000$	$t0_{true} = 2.00$
$x_{10} = 413032 \pm 50$	$y_{10} = 5723304 \pm 50$	$z_{10} = 1003 \pm 50$	$t0_{10} = 2.00 \pm 0.01$
$x_{30} = 413076 \pm 50$	$y_{30} = 5723320 \pm 50$	$z_{30} = 1010 \pm 50$	$t0_{30} = 2.03 \pm 0.05$
$x_{70} = 413088 \pm 50$	$y_{70} = 5723320 \pm 50$	$z_{70} = 980 \pm 50$	$t0_{70} = 2.03 \pm 0.05$
Event No. 4			
$x_{true} = 413330$	$y_{true} = 5723856$	$z_{true} = 1000$	$t0_{true} = 2.00$
$x_{10} = 413308 \pm 50$	$y_{10} = 5723860 \pm 50$	$z_{10} = 1026 \pm 50$	$t0_{10} = 2.02 \pm 0.02$
$x_{30} = 413300 \pm 50$	$y_{30} = 5723892 \pm 50$	$z_{30} = 1028 \pm 50$	$t0_{30} = 2.02 \pm 0.03$
$x_{70} = 413356 \pm 50$	$y_{70} = 5723836 \pm 50$	$z_{70} = 1024 \pm 60$	$t0_{70} = 2.02 \pm 0.03$

Table 3.1: This table summarizes the location results related to the 4 synthetic events introduced in the manuscript. Here we show a comparison between the true locations and the estimated ones with different noise levels. Uncertainties were estimated through perturbations of the STALTA parameters and jack-knife method (Location coordinates are in UTM and in meters (m), while origin time is in seconds (s)).

We additionally show location results for a database of 200 synthetic events with random location and focal mechanism. All events have been located simulating two different noise levels, 30% and 70% of the maximum amplitude. In both cases, even with a very high noise level, results confirmed the robustness of our method (we located the 90% of the events within 150 m from the true location with a noise level at 70% of the maximum amplitude). For these tests we used a short time window length in the range 0.1-0.3 s and a long time window 2 times longer.

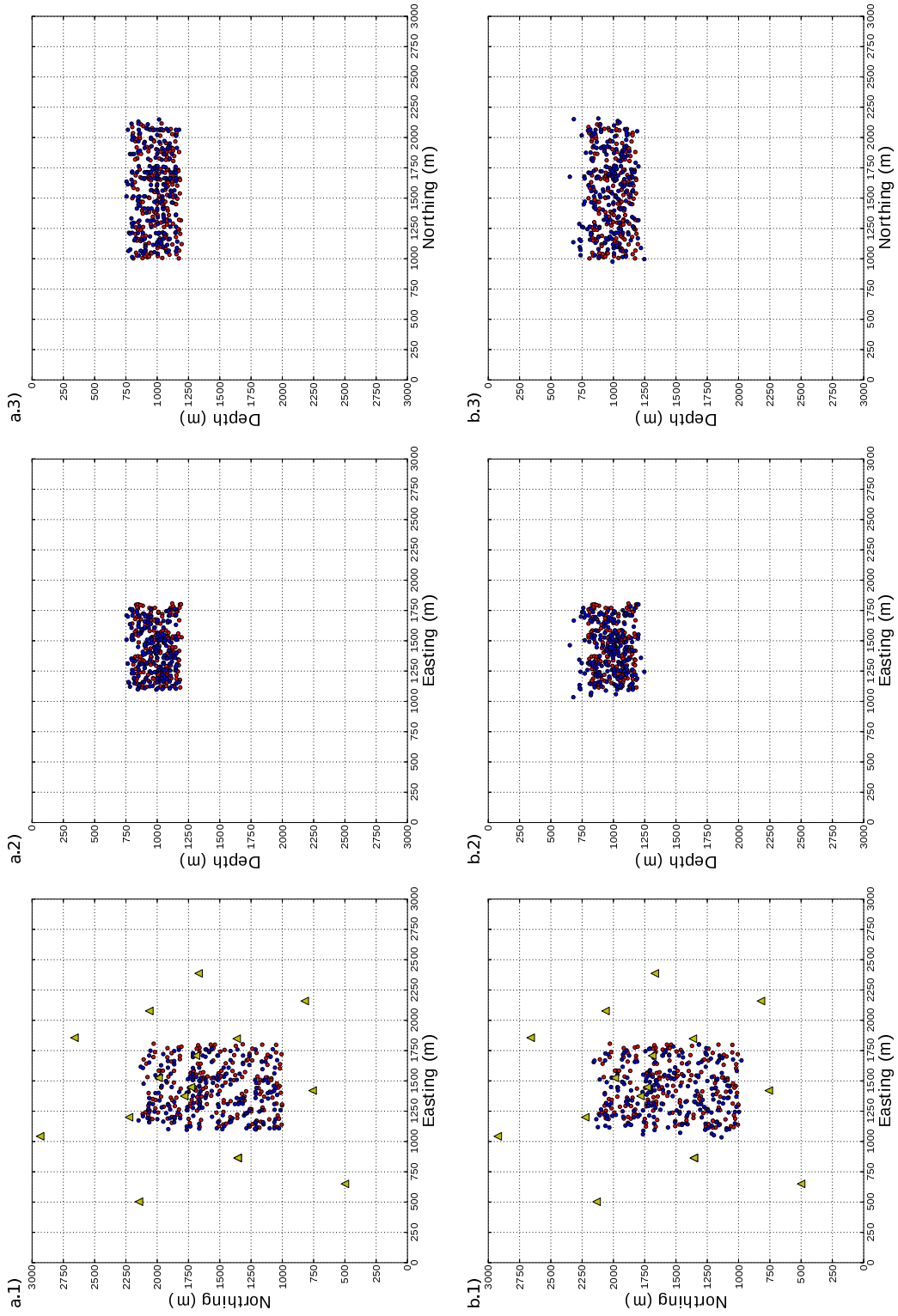


Figure 3.11: his figure shows our location results (blue dots) for 200 synthetic events with noise levels at 30% (subfigures a.1, a.2 and a.3) and 70% (subigures b.1, b.2 and b.3). The true locations are represented by red dots. The reference point (x,y)=(0,0) corresponds to x=411617 and y=5722111 (in the UTM system).

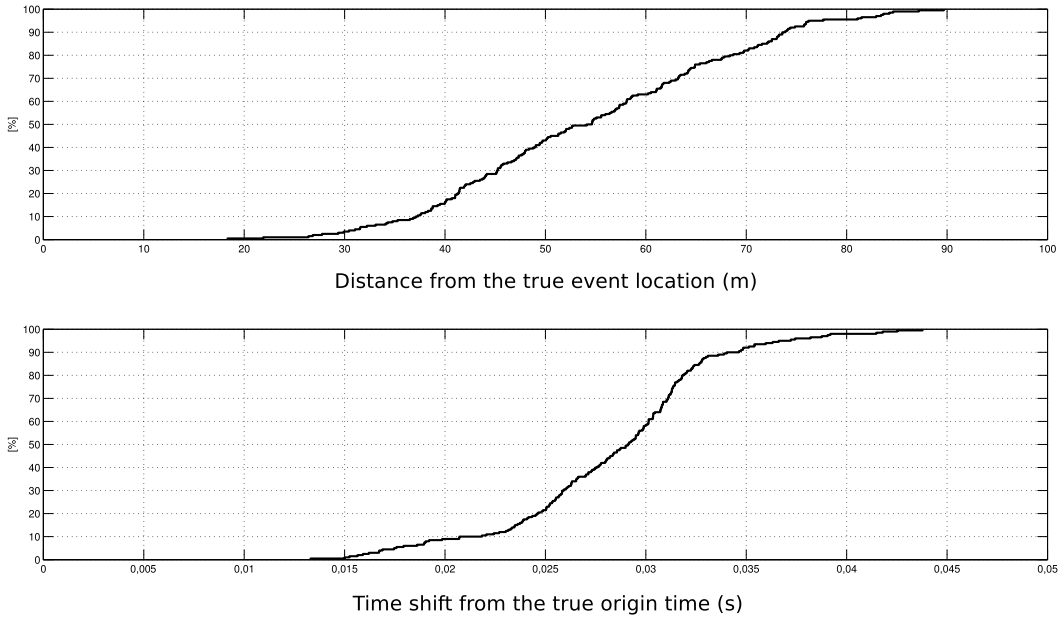


Figure 3.12: Quality of the location results for the 200 sythetic events affected by a noise level at 30%. The plot on the top shows the percent of events located within a given distance from the true locations. On the bottom, an analogue plot shows the absolute time shift respect the true origin time (bottom figure)

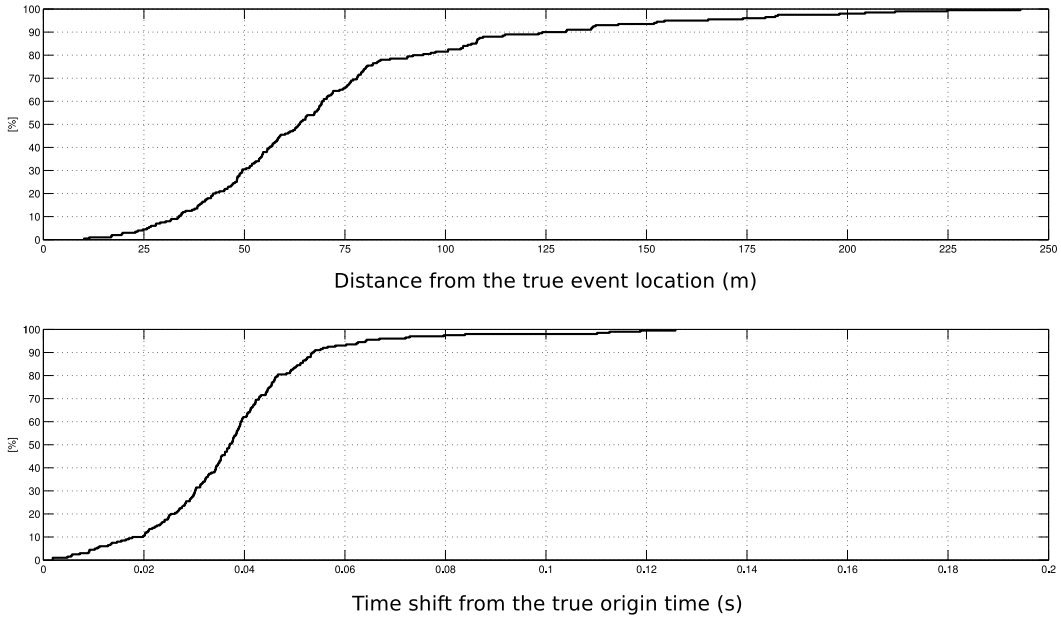


Figure 3.13: Quality of the location results for the 200 sythetic events affected by a noise level at 70%. The plot on the top shows the percent of events located within a given distance from the true locations. On the bottom, an analogue plot shows the absolute time shift respect the origin time of manual locations (bottom figure)

Chapter 4

Automated seismic event location by waveform coherence analysis

Authors:
Francesco Grigoli[a], Simone Cescaa[a,b], Ortensia Amoroso[c,d], Antonio Emolo[d], Aldo Zollo[d], Torsten Dahm[b,a]
Journal:
Geophysical Journal International,
Submitted on 19th November 2013,
in original form on 30th August 2013
Status:
Accepted for publication on 21th November 2013,
Doi:10.1093/gji/ggt477

Authors affiliation:
a) Institute of Earth and Environmental Sciences,
University of Potsdam, Potsdam, Germany
b) GFZ (German Research Centre for Geosciences) Potsdam,
Section 2.1, Physics of earthquakes and volcanoes,
Potsdam, Germany
c) Istituto Nazionale di Geofisica e Vulcanologia (INGV),
Osservatiorio Vesuviano,
Naples, Italy
d) Department of Physics,
University of Naples Federico II, Naples, Italy

4.1 Abstract

Automated location of seismic events is a very important task in microseismic monitoring operations as well for local and regional seismic monitoring. Since microseismic records are generally characterised by low signal-to-noise ratio, automated location methods are requested to be noise robust and sufficiently accurate. Most of the standard automated location routines are based on the automated picking, identification and association of the first arrivals of P and S waves and on the minimization of the residuals between theoretical and observed arrival times of the considered seismic phases. Although current methods can accurately pick P onsets, the automatic picking of the S onset is still problematic, especially when the P coda overlaps the S wave onset. In this paper we propose a picking free earthquake location method, based on the use of the Short-Term-Average/Long-Term-Average (STA/LTA) traces at different stations as observed data. For the P phases we use the STA/LTA traces of the vertical energy function, while for the S phases, we use the STA/LTA traces of a second characteristic function, which is obtained using the principal component analysis technique. In order to locate the seismic event, we scan the space of possible hypocentral locations and origin times, and stack the STA/LTA traces along the theoretical arrival time surface for both P ans S phases. Iterating this procedure on a three-dimensional grid we retrieve a multidimensional matrix whose absolute maximum corresponds to the spatial coordinates of the seismic event. A pilot application was performed in the Campania-Lucania region (southern Italy) using a seismic network (Irpinia Seismic Network) with an aperture of about 150 km. We located 196 crustal earthquakes (depth $<$ 20 km) with magnitude range $1.1 < Ml < 2.7$. A subset of these locations were compared with accurate manual locations refined by using a double difference technique. Our results indicate a good agreement with manual locations. Moreover, our method is noise robust and performs better than classical location methods based on the automatic picking of the P and S waves first arrivals.

4.2 Introduction

Automated seismic event location is nowadays a common practice for most seismological applications, from early warning systems to microseismic monitoring. The demand for automated location tools follows the increasing number of dense seismic networks and larger datasets to be processed. Since most datasets, as for the case of seismic events recorded by regional seismic networks, can be strongly contaminated by seismic noise, automated location methods are requested to be noise robust. Standard automated location routines are based on the automated picking and identification of the main seismic phases (generally P and S). Most of these methods are modified versions of the Geiger [1910] algorithm, based on the iterative minimization of the residuals between the theoretical and observed arrival times of the main seismic phases. A large number of algorithms has been

proposed for the automatic picking and the identification of the P and S phases, as for instance the Filterpicker algorithm [Lomax et al., 2012; Vassallo et al., 2012]. While P onsets can be accurately picked with current techniques, the automatic picking of the S onsets is still challenging. Usually the S waves are identified on the horizontal components of a seismogram through the variation in amplitude and frequency of the signal, with respect to the preceding P waves [Amoroso et al., 2012a]. Nevertheless, reliable picking of the S phase might be problematic for local events where the P coda overlaps the S wave. Another class of S-pickers make use of characteristic functions based on the polarization analysis of three-component seismic traces [Amoroso et al., 2012a; Cichowicz, 1993; Diehl et al., 2009]. However the performance of automatic pickers is limited in presence of noisy data, when picking and phases identification might be difficult. The growing interest on microseismic monitoring applications, particularly for oil and gas applications, has led to the recent development of alternative techniques, for automated seismic event location, similar to migration techniques used in reflection seismics. These methods are based on the concept of delay and sum of seismic waveforms and do not need prior phase picking nor phase identification. Among these techniques, the Source Scanning Algorithm (SSA) developed by Kao and Shan [2004, 2007] makes use of a brightness function to localize seismic tremors. The brightness function is obtained by stacking the absolute amplitudes of normalized seismograms recorded at different stations. A similar approach, based on the envelope stacking at first P arrival times, was introduced by Baker et al. [2005] and tested for real time seismic monitoring in southern California. Gharti et al. [2010] proposed to rotate seismic traces to the ray coordinates, compute the envelope and finally perform a stacking along P and S arrival times. The use of STA/LTA traces for seismic event location has been firstly introduced by Withers et al. [1999], who proposed a correlation based method to locate seismic events at regional scale. In the approach developed by Withers et al. [1999], processed waveforms are correlated with theoretical travel-time envelopes, then a grid search is performed to find the highest correlation value (i.e. the spatio-temporal coordinates of the seismic event). Grigoli et al. [2013] and Drew et al. [2013] used the stacking of the STA/LTA traces along P and S onsets times to locate microseismic events in mining and volcanic environments. All these migration based techniques do not require phase picking nor phase identification and they exploit the waveform information without the need to compute synthetic seismograms [Kao and Shan, 2007].

In this paper we present an improved version of the approach introduced by Grigoli et al. [2013], designed to be more stable when the S wave onsets are hardly picked. In this approach we make use of a new characteristic function, which is more sensitive to the S waves than the former one (horizontal energy trace). The new characteristic function maintains its performance even when the P coda overlaps the S wave onset (i.e. when P and S phases are not well separated). The main advantage of this function relies on its robustness: even with noisy data it shows a clear pick corresponding to the S wave arrival. To evaluate its performance we

test it with real data and compare its ability to recognize S phases with respect to the former characteristic function. Finally, we demonstrate the improved location method through an application to a dataset from a regional seismic network in the Campania-Lucania region (southern Italy). The dataset consists of 196 microseismic events with magnitude Ml between 1.2 and 2.7. Our location results are then compared with those obtained by an alternative automated location technique as well as with manual location results.

4.3 Methodology

Characteristic functions

The first step of the location process consists in the computation of characteristic functions. These should be generally sensitive to changes in energy, frequency content, polarization or other characteristics of the target signal relative to the background noise at each individual station [Lomax et al., 2012]. Here we calculate two characteristic functions, a first one sensitive to the first P phase arrival and a second one sensitive to the first S phase arrival. Three-component seismic traces can be viewed as discrete time series. We then denote the east component with $x(j)$, the north component with $y(j)$ and the vertical one with $z(j)$. The integer value $j = 1, n_{samples}$ is the time index of the series while $n_{samples}$ denotes the last sample of the trace. Following the same approach proposed by Grigoli et al. [2013] the P characteristic function CF^P is defined as the energy of the vertical component of the seismic trace:

$$CF^P(j) = z^2(j) \tag{4.1}$$

with $z(j)$ the vertical component seismic trace. To compute the S characteristic function CF^S, we firstly compute the analytic traces of both horizontal component traces defined as follow:

$$X(j) = x(j) + iH\{x(j)\} \tag{4.2}$$

$$Y(j) = y(j) + iH\{y(j)\} \tag{4.3}$$

where H is the Hilbert transform and $i^2 = -1$.Then, as proposed by Vidale [1986], we can compute the instantaneous covariance matrix $\mathbf{Q}(j)$ as:

$$\mathbf{Q}(j) = \begin{pmatrix} X(j)\hat{X}(j) & X(j)\hat{Y}(j) \\ Y(j)\hat{X}(j) & Y(j)\hat{Y}(j) \end{pmatrix} \tag{4.4}$$

where the hat ˆ denotes complex conjugation. Since the matrix $\mathbf{Q}(j)$ is Hermitian it has, for each sample j, two real positive eigenvalues λ_1 and λ_2 (with $\lambda_1 \geq \lambda_2$). At local and very local scale there is no guarantee that the incident S waves is

nearly vertical. In such cases, better results can be obtained by computing the instantaneous covariance matrix using the three component traces [Rowe et al., 2002]. Nevertheless, in this study we obtained better results by using a covariance matrix based on the horizontal components only. The S characteristic functions is then defined as:

$$CF^S(j) = \lambda_1(j)^2 + \epsilon \tag{4.5}$$

The ϵ term is a small positive number needed to overcome numerical issues related to the computation of the STA/LTA trace when $\lambda_1(j)$ tends to zero. Finally we compute the STA/LTA traces using the CF^P and CF^S characteristic functions separately. The original STA/LTA algorithm by Allen [1978, 1982] is here modified, through the adoption of a recursive one, which reduces memory requirements and results smoother than standard STA/LTA in absence of signal [Withers et al., 1998]. If we denote as n_s and n_l the number of samples of the short and long time windows respectively, a recursive STA/LTA algorithm is described by the following equations:

$$STA(j) = K_s(CF(j)) + (1 - K_s)STA(j-1) \tag{4.6}$$

$$LTA(j) = K_l(CF(j - n_s - 1)) + (1 - K_l)LTA(j-1) \tag{4.7}$$

where the index j varies in the range between $h = n_s + n_l$ and the last sample $n_{samples}$ of the characteristic function. According to Withers et al. [1999], the decaying constants K_s and K_l are set to $1/n_s$ and to $1/n_l$ respectively. Equations for the STA and LTA functions represent two single-pole low pass filters in the time domain with filter constants K_s and K_l respectively [Baer and Kradolfer, 1987]. Finally STA/LTA traces are computed in the following way:

$$W^P(j) = \frac{STA^P(j)}{LTA^P(j)} \tag{4.8}$$

$$W^S(j) = \frac{STA^S(j)}{LTA^S(j)} \tag{4.9}$$

where W^P and W^S denote the STA/LTA traces of the P ans S characteristic function respectively. Normalization of the STA/LTA traces is then required to take care of propagation effects, in order to avoid that stations close to the source dominate the stacking and also in order to balance the P and S contributions.

Location process

The STA/LTA traces for both P and S characteristic functions for all seismic stations of the network are now used as input data of the waveform based location

process. We now describe the waveform stacking location process as introduced by Grigoli et al. [2013]. Let us suppose that a seismic event is recorded by N three-component seismic stations. In order to locate the event we first define a 3D cartesian grid space containing the whole seismogenic region. For a grid with nx, ny and nz grid points along the x, y and z we can write:

$$x(l) = x_{ref} + l\delta x \quad \text{with} \quad l = 0, 1, ..., nx \tag{4.10}$$

$$y(m) = y_{ref} + m\delta y \quad \text{with} \quad m = 0, 1, ..., ny \tag{4.11}$$

$$z(n) = z_{ref} + n\delta z \quad \text{with} \quad n = 0, 1, ..., nz \tag{4.12}$$

where δx, δy and δz are the grid spacings along x, y and z directions (which correspond to East, North and vertical respectively) while x_{ref}, y_{ref} and z_{ref} are the cartesian coordinates of the reference point. Each grid point represents a potential source location. For each possible trial source location, we compute the theoretical arrival times for the first P and S onsets ($\tau_k^p(x, y, z)$ and $\tau_k^s(x, y, z)$ respectively) at all N stations of the recording network (k is the station index). We then define τ_{min} and τ_{max} as:

$$\tau_{min}(l, m, n) = \min(\{\tau_k^p(l, m, n)\}_{k=1}^N) \ , \tag{4.13}$$

$$\tau_{max}(l, m, n) = \max(\{\tau_k^s(l, m, n)\}_{k=1}^N) \ , \tag{4.14}$$

which denote the minimum P and maximum S arrival time in the network. Using the previous equations we introduce $T_k^p(l, m, n)$ and $T_k^s(l, m, n)$:

$$T_k^p(l, m, n) = \tau_k^p(l, m, n) - \tau_{min}(l, m, n) \ , \tag{4.15}$$

$$T_k^s(l, m, n) = \tau_k^s(l, m, n) - \tau_{min}(l, m, n) \ , \tag{4.16}$$

$T_k(l, m, n)$ are the computed arrival times at station k (for P and S first onsets) relative to τ_{min}. Considering the sampling rate δt of the recorded waveforms we can discretise the previous equations as follow:

$$\Delta T_k^p(l, m, n) = \text{round}\left\{\frac{T_k^p(l, m, n)}{\delta t}\right\} \ , \tag{4.17}$$

$$\Delta T_k^s(l, m, n) = \text{round}\left\{\frac{T_k^s(l, m, n)}{\delta t}\right\} \ , \tag{4.18}$$

Using the equations 17 and 18 we can evaluate the coherence functions C^p and C^s at each grid point and time sample, by using the following equations:

$$C^p(l, m, n, j) = \sum_{k=1}^{N} W_k^P(j + \Delta T_k^p(x, y, z)) \ , \tag{4.19}$$

$$C^s(l, m, n, j) = \sum_{k=1}^{N} W_k^S(j + \Delta T_k^s(x, y, z)) \ , \tag{4.20}$$

where W_k^P is the normalized STA/LTA of the P characteristic function related to the k-th station and, in analogy, W_k^S is the normalized STA/LTA of the S characteristic function related to the k-th station. According to Kao and Shan [2004], to obtain smoother results (e.g. when the velocity model is poorly known) the method can be modified stacking all samples within a chosen window centred around ΔT_k^p (or ΔT_k^s). The coherence matrix is finally defined as:

$$C(l,m,n,j) = \frac{\sqrt{C^p(l,m,n,j)C^s(l,m,n,j)}}{N} \ . \tag{4.21}$$

From the previous equation it is clear that $C(l,m,n,j)$ is a bounded function whose theoretical limits are 0 (no coherence) and 1 (perfect coherence for both P ans S first arrivals phases). The location is then obtained by taking the maximum of the matrix

$$C(\hat{l},\hat{m},\hat{n},\hat{j}) = \max\{C(l,m,n,j)\} \ , \tag{4.22}$$

Finally, the coordinates of the seismic event are $(\hat{x},\hat{y},\hat{z}) = (x(\hat{l}),y(\hat{m}),z(\hat{n}))$ while its origin time $\hat{t}$ can be retrieved by:

$$\hat{t} = \hat{j}\delta t - \tau_{min}(\hat{l},\hat{m},\hat{n}) \ . \tag{4.23}$$

In order to estimate uncertainties, we follow the approach proposed by Grigoli et al. [2013], based on the distribution of locations after processing the same event several times by perturbing the parameters of the STA/LTA traces. Location uncertainties are then estimated by perturbing the STA/LTA parameters (i.e the length of both long and short time windows) and relocating each event several times. From the location distribution we computed the weighted mean, using the coherence value as weight, and a weighted covariance matrix. The weighted mean provides the best estimation of the hypocentral location, while uncertainties information can be directly extracted from the covariance matrix.

4.4 Application to the ISNet data

On November 23th 1980, a M_s 6.9 earthquake struck the Irpinia region (southern Italy), causing almost 3,000 fatalities and more than 10,000 injured. Even more than 20 years after the main event, the region shows continued background seismic activity including moderate-size events with magnitude (Ml) up to 5.4 [Ameri et al., 2011; De Matteis et al., 2012]. A dense seismic network, the Irpinia Seismic Network (ISNet), has been deployed to study the active fault systems in this region and, because of the high density of population in that area, for early warning purposes (Iannacone et al., 2010). The network started to be operational in 2005; at the current state it consists of 28 short period (1 Hz) and 5 broadband (0.25-50 Hz) seismic stations deployed in an area of about 100 km x 70 km (figure 1) with an average inter-station distance of about 15 km. Each seismic station is equipped with a three-component velocity sensor and a three component strong motion

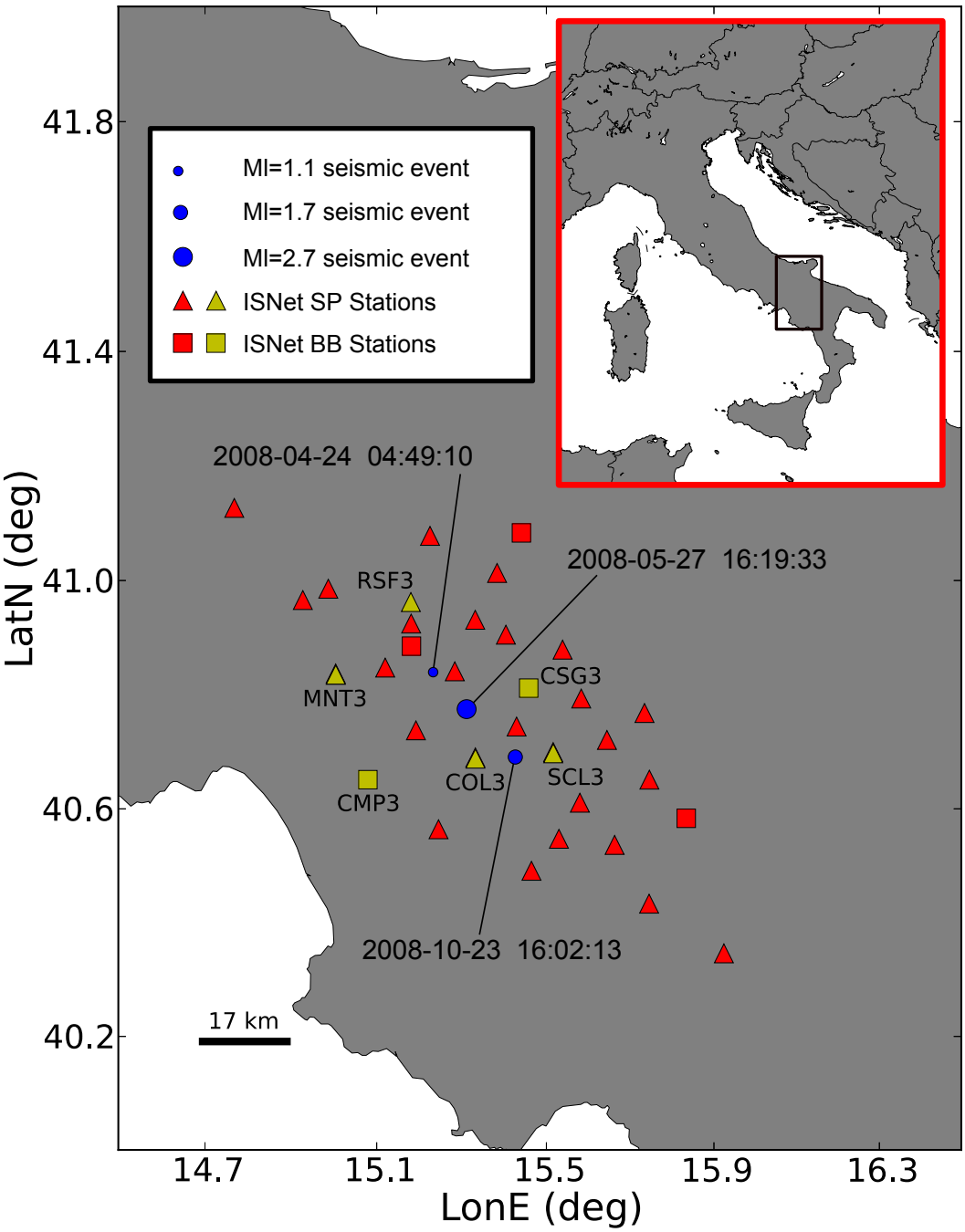


Figure 4.1: Irpinia seismic network, short period stations are represented by triangles, broadband stations by squares and the blue circles represent three seismic events occurred in the region. Seismograms of these events recorded by the yellow marked stations are shown in figures 2-4.

accelerometer [Weber et al., 2007]. In the period 2008-2012, ISNet recorded about 1200 in-network events with local magnitude up to Ml 3.4 [Stabile et al., 2012]. In this study we consider 196 microseismic events occurred between February 2008 and March 2010, with a magnitude Ml between 1.1 and 2.7 and detected at lest

by 8 seismic stations. The last condition is required since location methods based on waveform staking are designed to work with a large number of stations. In this application only velocity sensors have been used. All waveforms are characterized by a sampling frequency of 125 Hz and have been filtered with a bandpass filter within the frequency band 1-25 Hz. In order to show the performance of the new approach we first compare the STA/LTA of the horizontal energy trace with the STA/LTA of the new characteristic function (figures 4.2,4.3 and 4.4) related to three selected seismic events shown in figure (4.1). The STA/LTA traces in figures (4.2,4.3 and 4.4) were computed using a short time window length of 0.7 s and a long one of 1.0 s. From these figures it is clear that the STA/LTA of the characteristic function based on the principal eigenvalue (green line) shows, in all cases, a clear pick corresponding to the S phase, while the STA/LTA of the horizontal energy characteristic function (red line) is generally more noisy and prone to fail, showing in some cases a pick at the P phase. This result supports our choice to locate seismic events using both the STA/LTA of the vertical energy for the P phase and the STA/LTA of the principal eigenvalue characteristic function for the S phase. We located the seismic events by direct search within a grid, with size 102 x 102 x 36 km^3 and a 0.5 km grid spacing. To compute the P and S traveltimes we used the finite difference code developed by Podvin and Lecomte [1991] for a 3D velocity model (shown on figure 4.5) retrieved from traveltime tomography in the Irpinia region [Amoroso et al., 2012b; Matrullo et al., 2013]. Figure (4.6) shows the coherence matrices for the three selected seismic events (shown in figure 1) with magnitude between Ml 1.1 (figure 4.6.c) and 2.7 (figure 4.6.a). The coherence matrices show a clear absolute maximum localized within 2 grid points (i.e. 1 km) from the hypocentral location retrieved using manual picking. The coherence matrices related to the event with lower magnitude (figure 4.6.c) show a larger smearing in depth because of a lower number of stations used for the stacking. To verify the quality of our results, we compare 55 event locations with those obtained by a manual location procedure [De Matteis et al., 2012; Stabile et al., 2012]. All manual locations have been performed using a nonlinear global approach (NonLinLoc) [Lomax, 2008], subsequently these locations have been refined by applying a double-difference technique (HypoDD) [Waldhauser and Ellsworth, 2000]. This second step allow to reduce the location error due to the un-modeled velocity structures [Stabile et al., 2012]. We further compared the performance of our approach with a standard location method based on the RTLOC algorithm developed by Satriano et al. [2008]. The comparison among our locations, those retrieved by automatic picking and the reference ones is shown in figure (4.7). The better performance of our approach can be easily seen from the cumulative plot in figure (4.8). All the locations obtained using our approach (blue line) are within 4.5 km distance from the manual locations, while the locations obtained using standard automated method based on automated picking (red line) are affected by larger errors (up to 31 km). Using our approach more than 90% of the events have been located within 3.5 km from the reference location, against the 6.5 km distance from the reference locations related to the results obtained

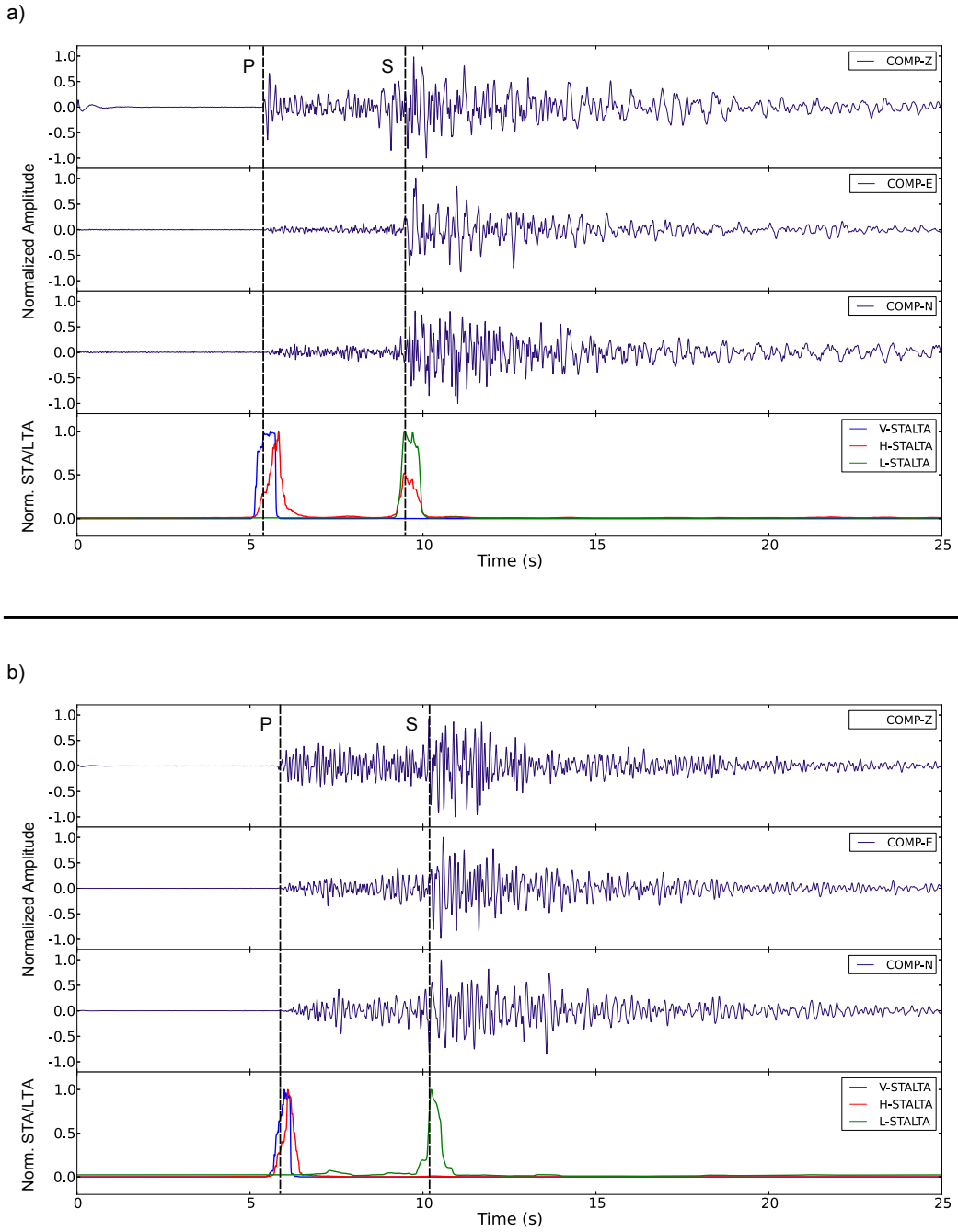


Figure 4.2: Waveforms related to the MI=2.7 seismic event occurred on May 27^{th} 2008 (figure 3.1) and recorded by the stations RSF3 (a) and CMP3 (b). We show the three component seismograms and the STA/LTA traces of three different characteristic functions: the horizontal energy trace (red line), the principal eigenvalue of the instantaneous covariance matrix (green line) and the P wave characteristic function based on the vertical energy trace (blue line). (for a more detailed figure see the electronic supplement)

by automatic picking. The better results achieved by our approach are confirmed both in epicentral and depth estimations. We located 196 seismic events (55 of

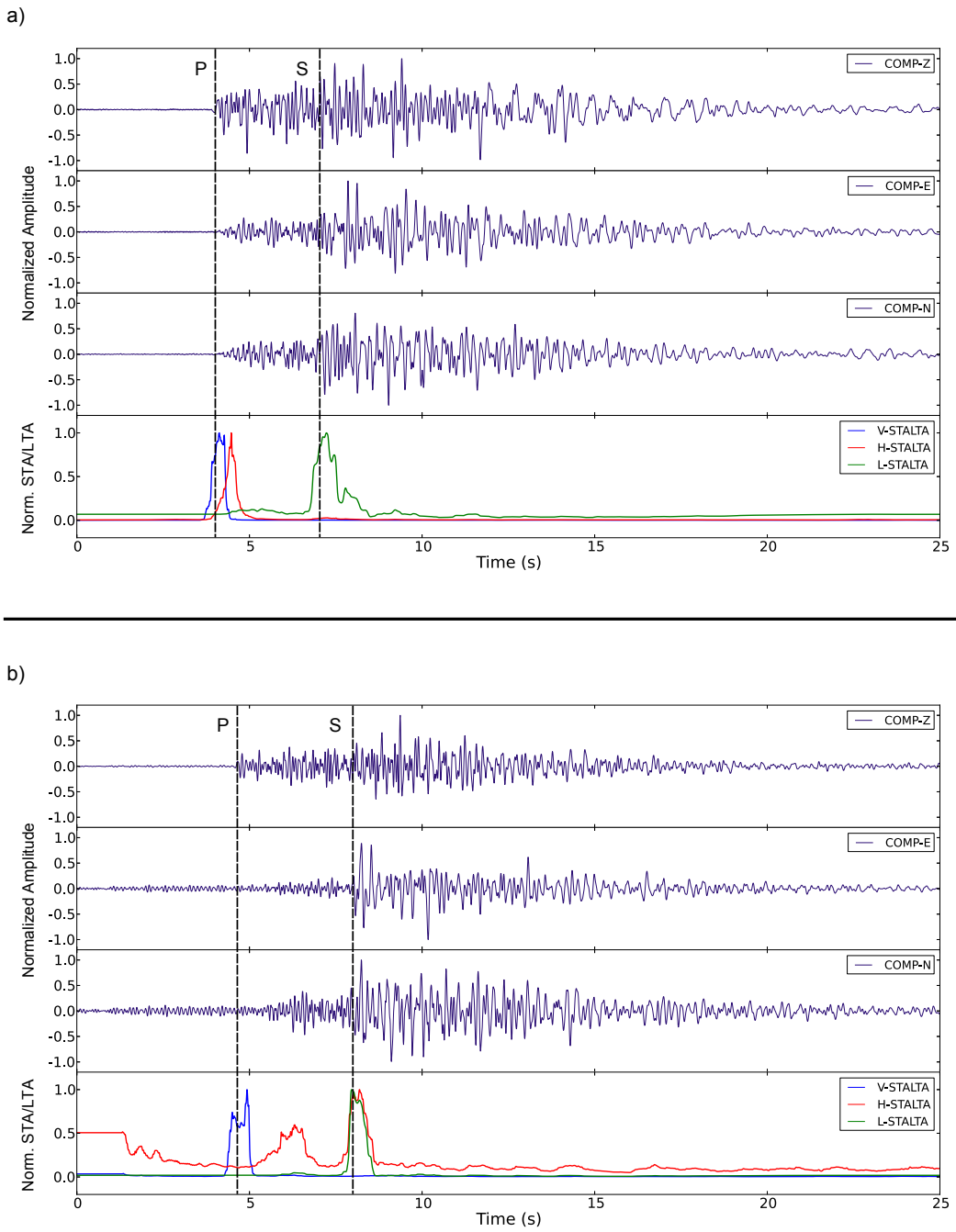


Figure 4.3: Waveforms related to the Ml=1.7 seismic event occurred on October 23^{th} 2008 (figure 3.1) and recorded by the stations SCL3 (a) and CSG3 (b). We show the three component seismograms and the STA/LTA traces of three different characteristic functions: the horizontal energy trace (red line), the principal eigenvalue of the instantaneous covariance matrix (green line) and the P wave characteristic function based on the vertical energy trace (blue line).(for a more detailed figure see the electronic supplement)

which were compared with the available manual locations as we shown previously) showing a good agreement between seismicity and tectonics structures. Indeed,

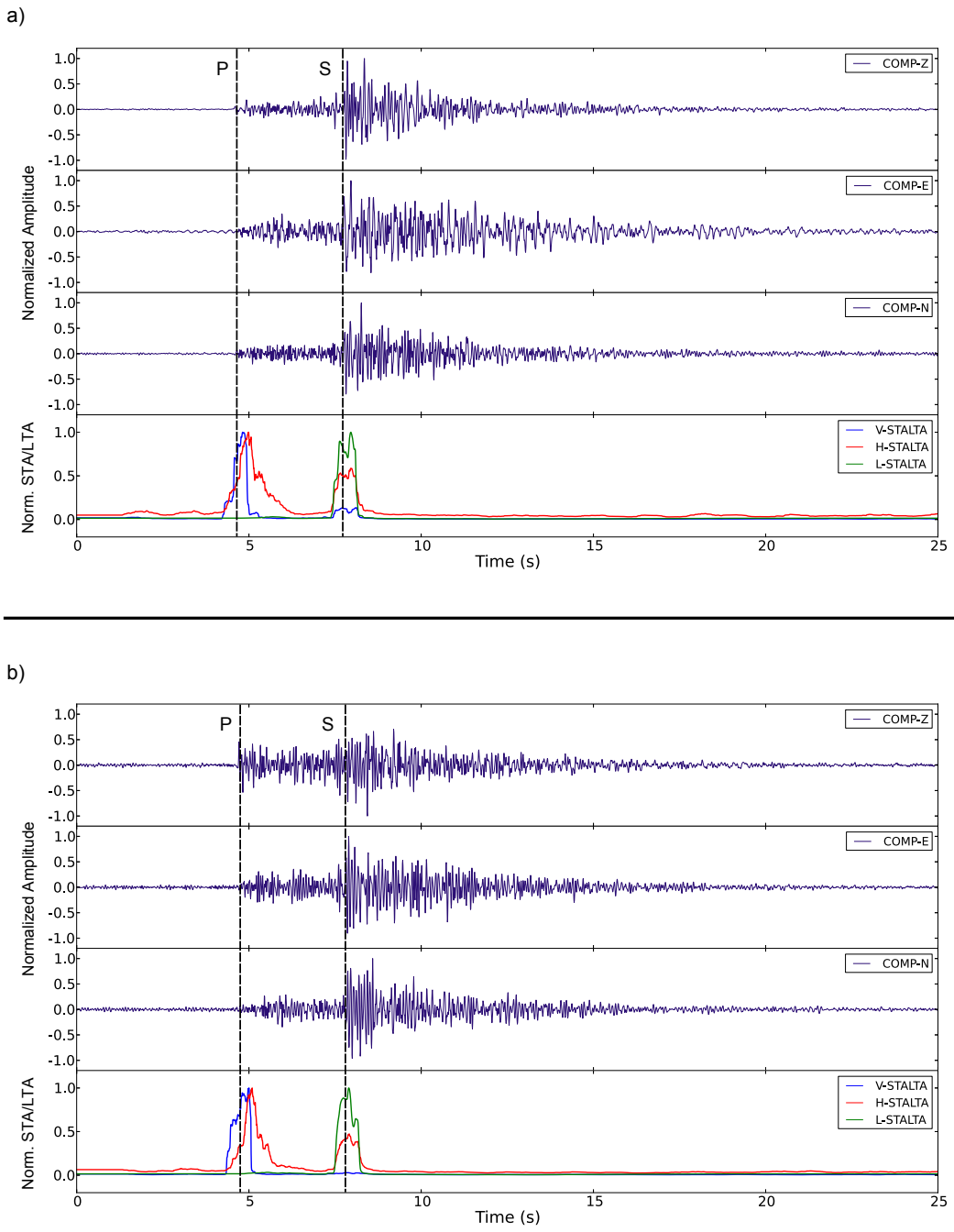


Figure 4.4: Waveforms related to the MI=1.1 seismic event occurred on April 24^{th} 2008 (figure 3.1) and recorded by the stations MNT3 (a) and COL3 (b). We show the three component seismograms and the STA/LTA traces of three different characteristic functions: the horizontal energy trace (red line), the principal eigenvalue of the instantaneous covariance matrix (green line) and the P wave characteristic function based on the vertical energy trace (blue line). (for a more detailed figure see the electronic supplement)

from figure (9), it is clear that most of the seismic events are located within an area surrounded by the main seismogenic structures of the region (red lines).

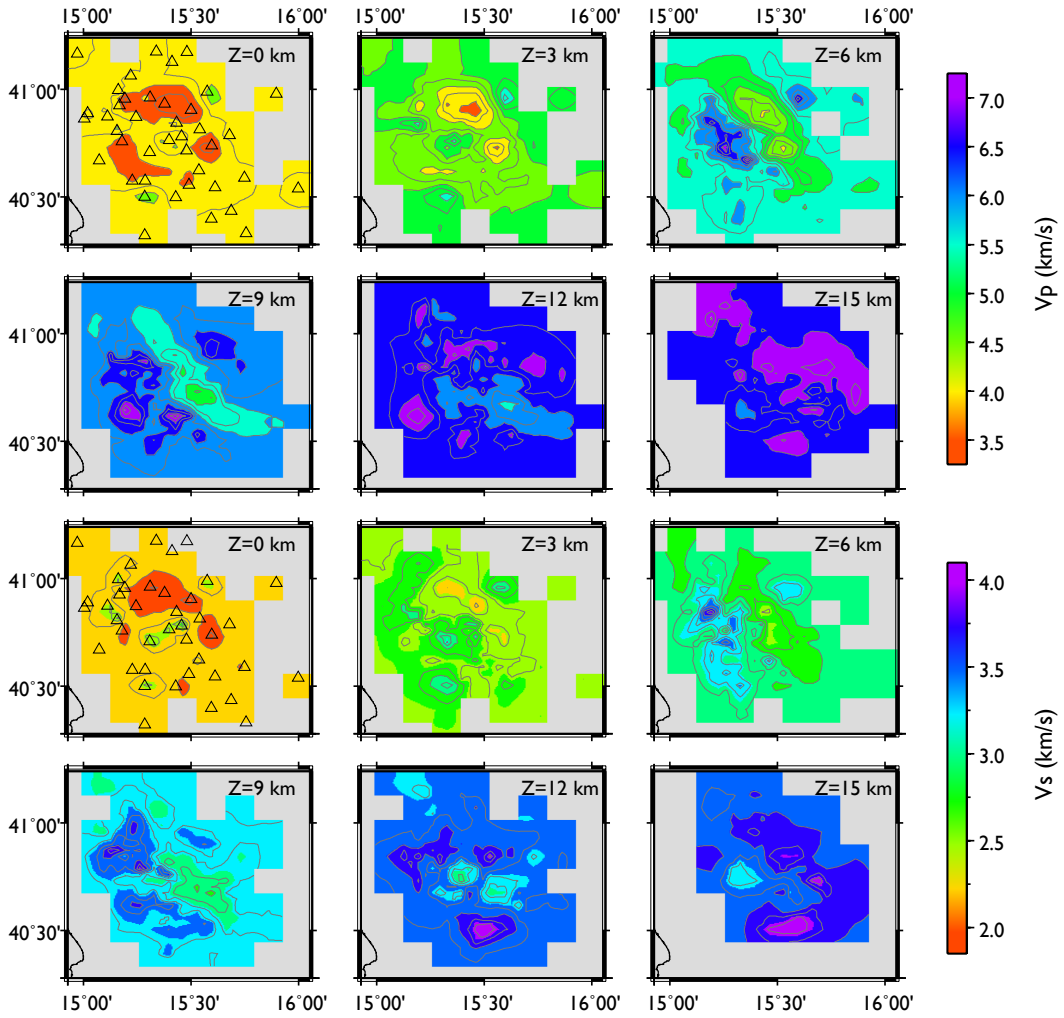


Figure 4.5: Three dimensional velocity model of the Campania-Lucania region (southern Italy) obtained by traveltime tomography. The figures on top show the P velocity model at different depths, in analogy figures at the bottom show depth slices of the S velocity model. Network stations are represented by triangles.

Location results and related uncertainties are summarized in table (1), enclosed in the electronic supplement of this manuscript. Uncertainties have been estimated by random perturbation (20 times) of the STA window length within the range 0.56-0.96 s and taking the LTA window length 1.5 times longer than the STA.

4.5 Discussion and Conclusions

We improved the methodology to locate seismic events introduced by Grigoli et al. [2013], where the S wave characteristic function is based on the horizontal energy trace. This automated methodology is applied, for the first time, to a regional seismic network in order to locate crustal earthquakes considering a 3D veloc-

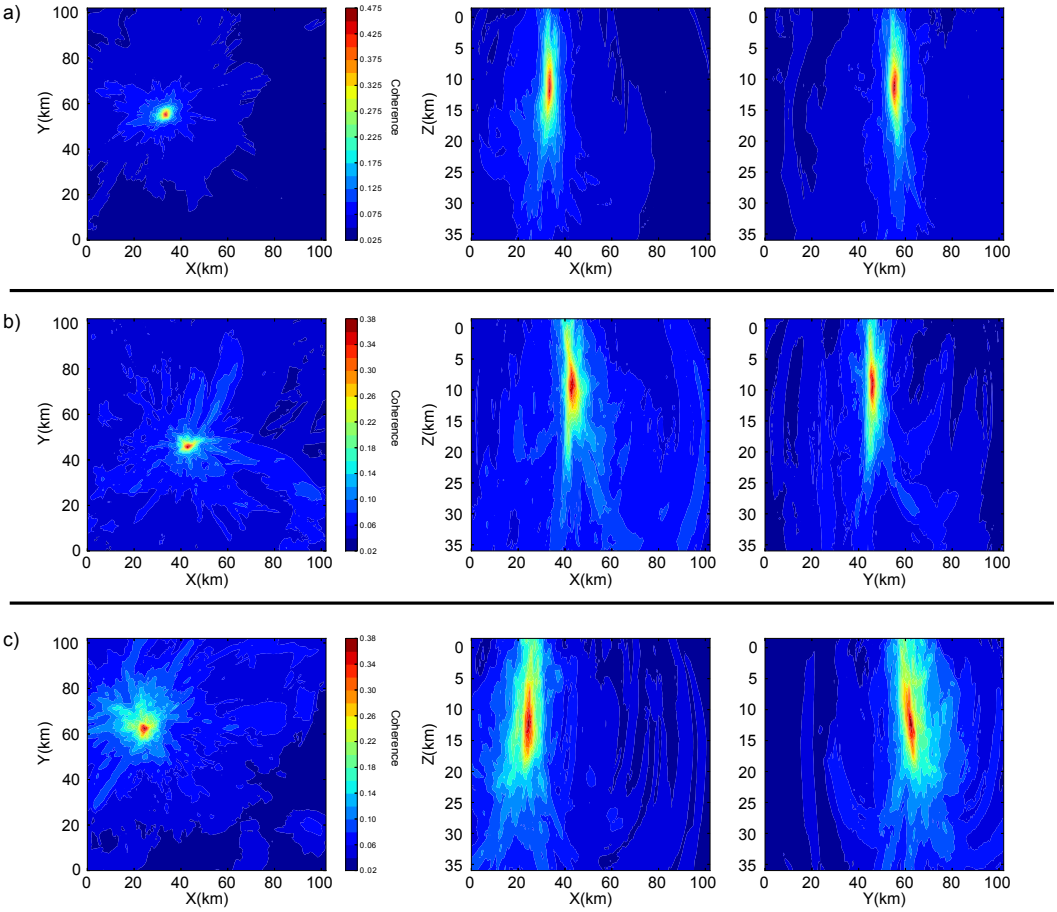


Figure 4.6: Coherence matrices related to the three seismic events shown in figure (4.1): (a) event 1 (b) event 2 (c) event 3. The coherence matrix XY is obtained by projection, for each X-Y, its maximum along Z (coherence matrices XZ and YZ are obtained in a similar way). Coherence values are represented in color scale. The reference point (X;Y)=(0;0) corresponds with the point (X;Y)=(493718;4458627) in the UTM coordinates system.

ity model obtained from traveltime tomography of the Campania-Lucania region [Amoroso et al., 2012b; Matrullo et al., 2013]. Since we are dealing with noisy seismograms contaminated by a strong P coda overlapping the S wave first onset, this approach results problematic. In such cases it is necessary to define a characteristic function which is more sensitive to the S waves. In this study we used a characteristic function based on the polarization analysis of horizontal component seismograms. Grigoli et al. [2013] extensively tested this waveform based location approach with synthetic data thus, in this work, the new characteristic function is directly tested on real data. Due to the complex geology of the area, seismic signals we used to test our approach are characterized by a strong P coda overlapping the S wave onset. The standard characteristic function often fails to correctly identify the S wave onset. Since the P coda is not linearly polarized the STA/LTA of the principal eigenvalue characteristic function results more sensitive to the S

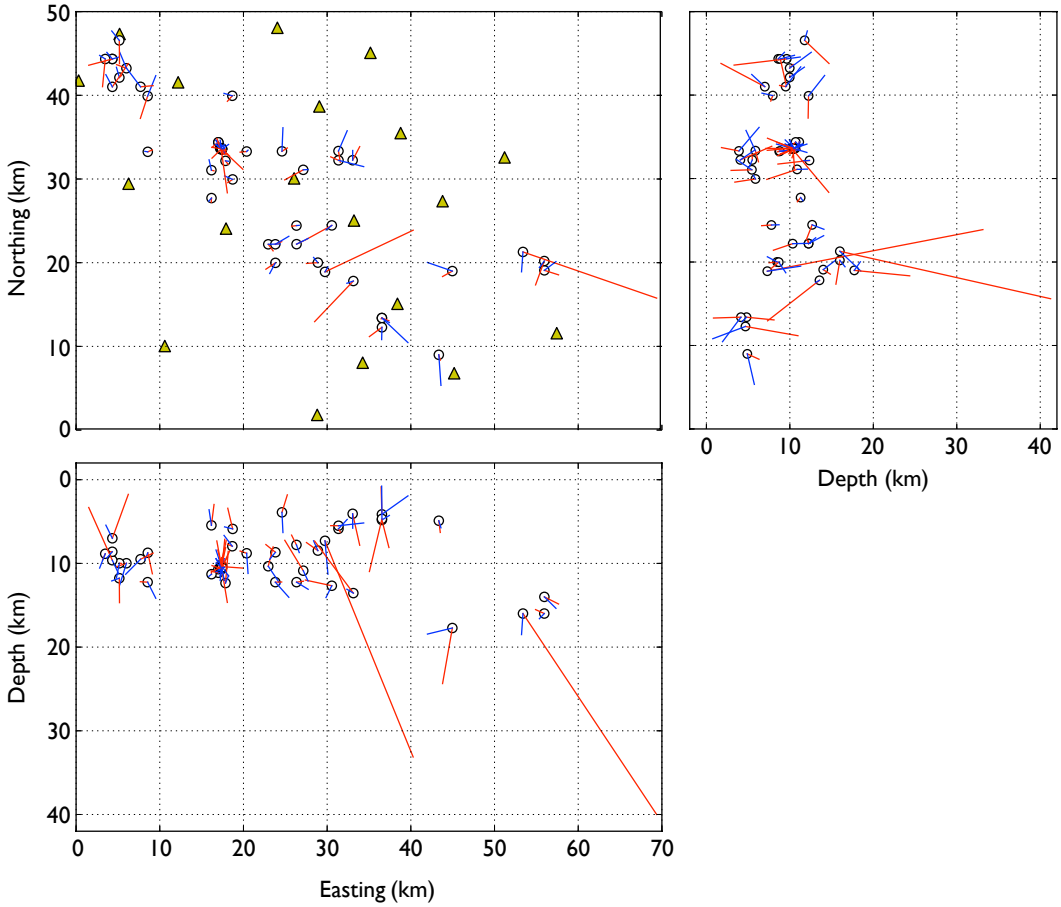


Figure 4.7: Comparison between our solutions (blue lines) and the locations obtained by automatic picking (red lines) with respect to manual locations (white circles) obtained using a double difference location algorithm. The reference point (X;Y)=(0;0) corresponds with the point (X;Y)=(510000;44480000) in the UTM coordinates system. The cross sections are perpendicular to the Northing and Easting axis.

wave and leads to a better performance during the location process. With the use of the new characteristic function our waveform stacking location method results more robust and shows a better performance even when the identification of the S wave is difficult (overlapping of the P coda with the S phases, emergent S arrival, noise contaminated data, etc.). The range of variation of both short and long time windows of the STA/LTA is determined on a empirical basis, considering the sampling period and the dominant frequency of the recorded waveforms. As a rule of thumb the length of the short time window should corresponds to half period of the P and S onsets, while the length of the long time window can be 1.5-2 times longer than the short one. To locate seismic events we make use of cut seismic traces related to triggered events as input data. We located 196 seismic event recorded from the Irpinia Seismic Network (ISNet) and we compared a subset of 55 events with the manual locations obtained by several studies [De Matteis et al., 2012; Stabile et al., 2012]. Our approach performs better than the standard

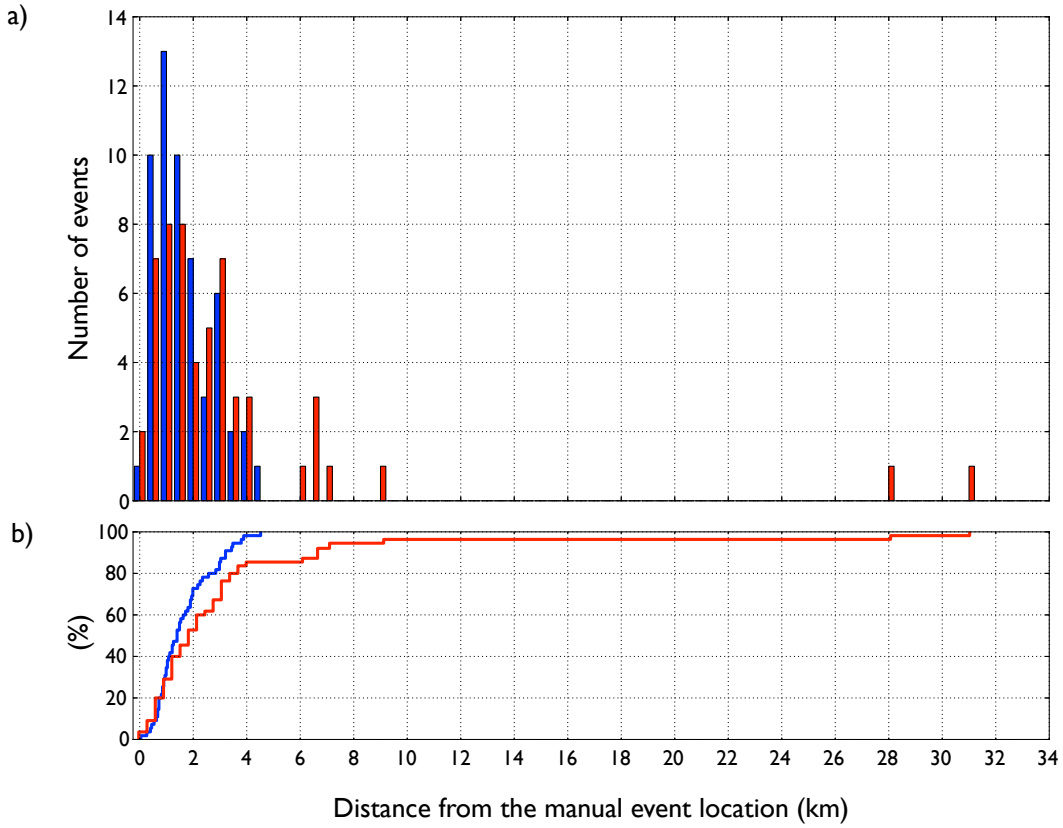


Figure 4.8: The histogram plot (a) shows the number of events within a given distance from the manual locations. Blue bars are related to our automated locations, while red bars to the locations obtained by automatic picking. The cumulative plot (b) shows the percent of events located within a given distance from the manual locations. Also in this case, the blue line is related to the location obtained using our approach, while the red line is related to the location obtained using automated picking. For both plots, reference locations are based on the results obtained by [De Matteis et al., 2012; Stabile et al., 2012].

automated location method. We located more than 90% of the events within 3.5 km from the reference location, against the 6.5 km distance obtained by location procedure based on the automatic picking. The location improvement is more evident with respect to the depth estimation (figure 7). This result depends on the fact that, unlike our approach, automatic picking algorithms fail to pick the S phases or they use a smaller number of S picks. Finally we successfully located a larger dataset composed by 196 seismic events. Our locations are distributed within a volume delimited by the major faults of the region. Uncertainties provide information about the quality and stability of the solutions (i.e. larger uncertainties are related to less stable solutions). The quality of the results is also related with the number of stations used for the stacking process as can be seen in figure (6). However, the lowest value that uncertainties can assume is constrained by the adopted grid spacing. For this reason It is important to note that our approach is not designed to obtain high precision locations (for instance as the double dif-

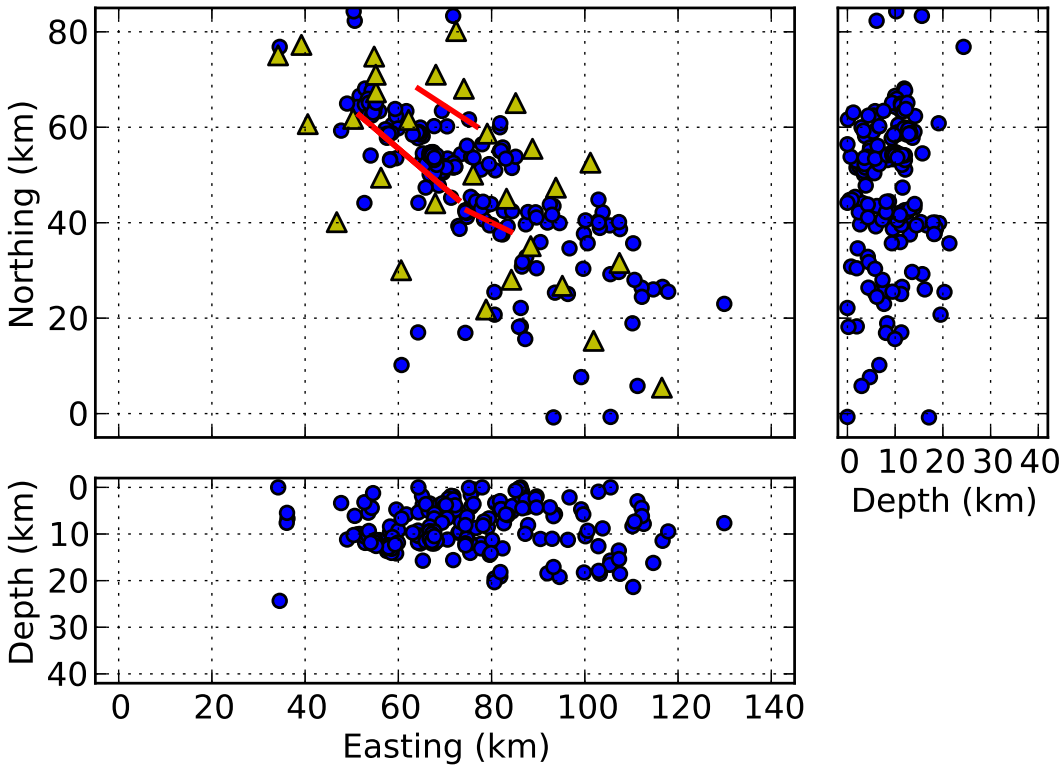


Figure 4.9: Map view of about 196 seismic events located using our approach. Red lines represent the surface projections of three fault segments ruptured during the 1980 Irpinia earthquake. Cross sections are represented in the small plots. The reference point (X;Y)=(0;0) corresponds with the point (X;Y)=(475000;4455000) in the UTM coordinates system.

ferences methods). Our aim was the development of a picking free, noise robust and full automated location method for processing large amount of data. Even if our approach performs better than the standard ones based on automatic picking, the results obtained by manual location procedures are the most reliable. Like the other waveform-based location methods, our approach is computationally intensive and the location process requires more computing time than those based on the iterative inversion of the arrival times. The use of such location approach can be useful to process (or reprocess) large amount of data in a completely automated fashion. Location accuracy depends on the choice of the grid spacing as well on the knowledge of the velocity model. In order to find the optimal grid spacing one should take in account that a larger number of grid points (fine grids) increase the computing time needed to locate a seismic event. If T_{stack} is the computing time needed to perform the trace stacking at one grid point, the total computing time for the full grid will be $T_{stack}(n_x n_y n_z)$, where n_x, n_y and n_z are, respectively, the total number of grid points along x, y, and z directions. However, the computing time can be strongly reduced using the parallel programming techniques in combination with a multicores workstation or a cluster. Concerning our application, a single location without error estimation needs less than 1 minute

on a 12 cores Intel Xeon workstation (each core has 2.4 MHz frequency clock) with 96 GB Ram. Our python location module, LOKI (LOcation of seismic events trough travetlime stacKIng) make use of the OBSPY library [Beyreuther et al., 2010] for reading different data formats, is open source and freely available, upon request, at the project MINE web page http://mine.zmaw.de.

4.6 Acknowledgements

We wish to thank the editor, Egill Hauksson and two anonymous reviewers for their comments and suggestion. We thank Claudio Satriano, Maurizio Vassallo, Luca Elia and the RiSSC lab team for the useful discussions we had during the development of the methodology. This work has been realized within the FP7 EU research project NERA under grant agreement n. 282862 and the research project MINE. The project MINE is part of the research and development programme GEOTECHNOLOGIEN and is funded by the German Ministry of Education and Research (BMBF). Grant of project BMBF03G0737A.

4.7 Appendix:Electronic supplement

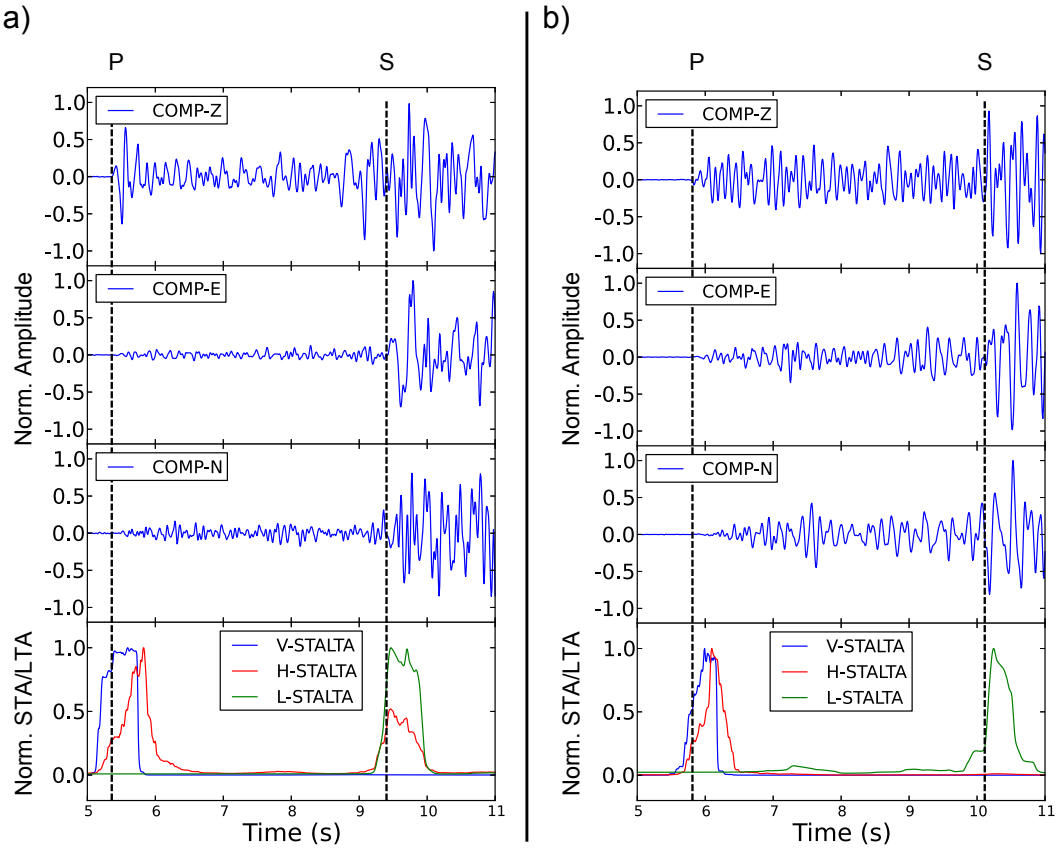


Figure 4.10: Zoomed version of the figure 2 (in the manuscript). Waveforms are related to the 2008-05-27 16:19:33 seismic event with MI=2.7 and recorded by the stations RSF3 (a) and CMP3 (b). We show the three component seismograms and the STA/LTA traces of different characteristic functions. The red line represents the STA/LTA of the horizontal energy trace, while the green line represents the STA/LTA trace of the principal eigenvalue characteristic function. In the same plot the blue line is related to the STA/LTA trace of the P characteristic function based on the vertical energy trace.

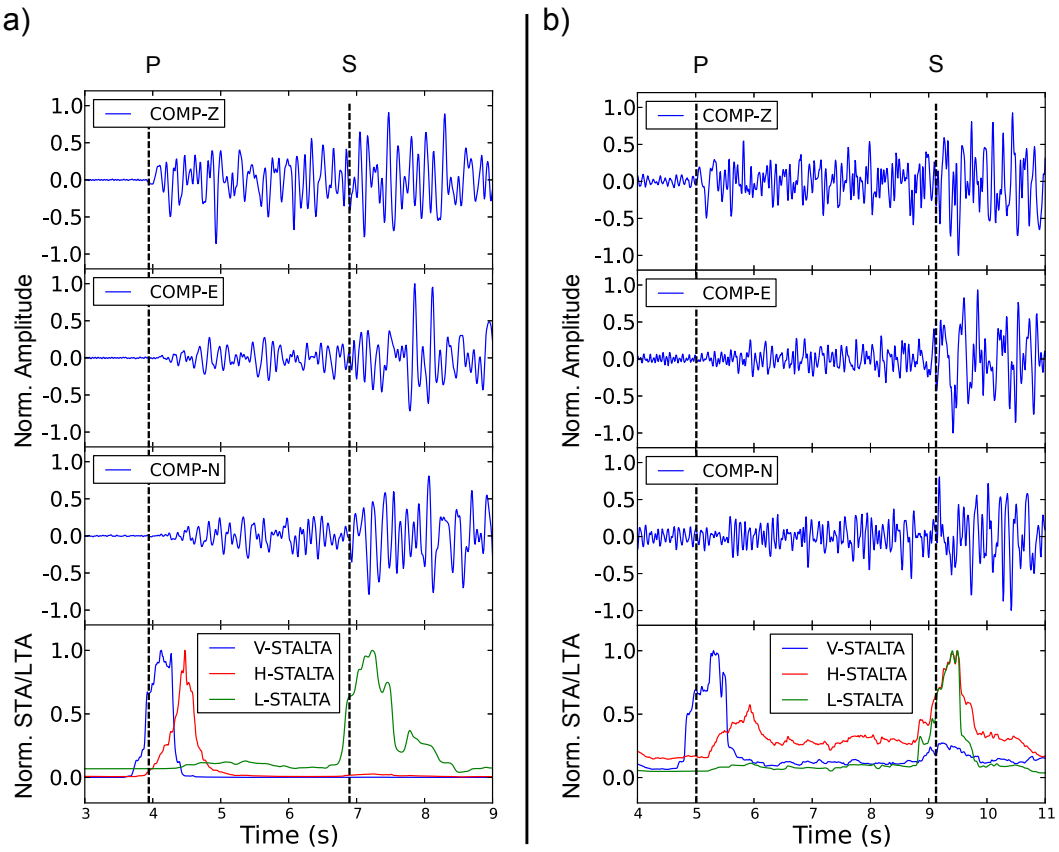


Figure 4.11: Zoomed version of the figure 3 (in the manuscript). Waveforms related to the 2008-10-23 16:02:13 seismic event with MI=1.7 and recorded by the stations SCL3 (a) and CSG3 (b). We show the three component seismograms and the STA/LTA traces of different characteristic functions. The red line represents the STA/LTA of the horizontal energy trace, while the green line represents the STA/LTA trace of the principal eigenvalue characteristic function. In the same plot the blue line is related to the STA/LTA trace of the P characteristic function based on the vertical energy trace.

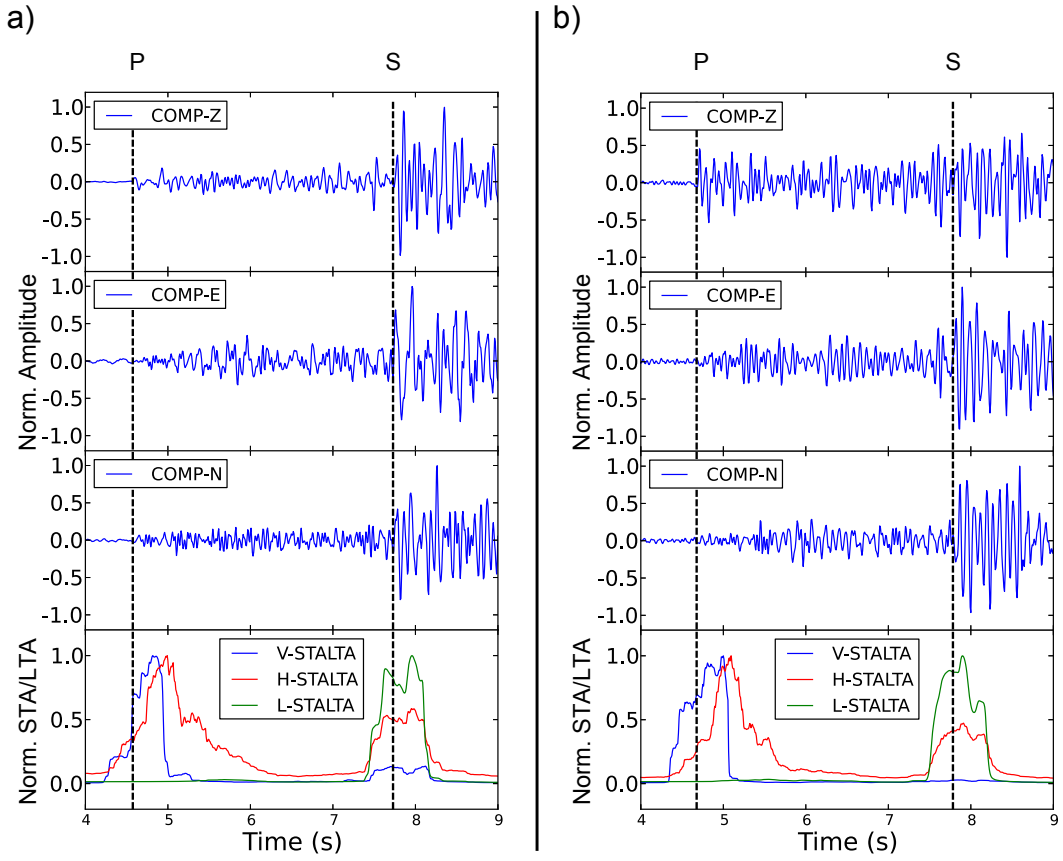


Figure 4.12: Zoomed version of the figure 4 (in the manuscript). Waveforms related to the 2008-04-24 04:49:10 seismic event with Ml=1.1 and recorded by the stations MNT3 (a) and COL3 (b). We show the three component seismograms and the STA/LTA traces of different characteristic functions. The red line represents the STA/LTA of the horizontal energy trace, while the green line represents the STA/LTA trace of the principal eigenvalue characteristic function. In the same plot the blue line is related to the STA/LTA trace of the P characteristic function based on the vertical energy trace.

Event date	Event time	Easting (m)	Northing (m)	Depth (m)
2008-02-18	06:29:49	524301 ± 500	4504171 ± 500	79 ± 500
2008-02-20	10:36:49	541889 ± 500	4497627 ± 500	18163 ± 500
2008-02-25	04:55:54	540602 ± 700	4485473 ± 550	20339 ± 1103
2008-02-25	04:56:17	540650 ± 702	4480733 ± 3399	19520 ± 1692
2008-02-25	05:41:23	562980 ± 500	4500026 ± 500	17909 ± 1032
2008-02-25	05:42:45	563227 ± 547	4498911 ± 979	18507 ± 1121
2008-03-19	01:49:36	514053 ± 500	4514072 ± 500	11868 ± 736
2008-03-21	08:00:25	528218 ± 500	4512122 ± 500	11435 ± 500
2008-04-13	23:02:12	531762 ± 514	4543338 ± 500	15598 ± 1274
2008-04-24	04:49:10	547218 ± 6417	4475627 ± 10149	9958 ± 762
2008-05-21	05:07:24	532469 ± 500	4505320 ± 658	12885 ± 500
2008-05-25	03:03:15	527079 ± 500	4514357 ± 500	9719 ± 1055
2008-05-25	03:21:55	526386 ± 500	4512967 ± 1178	6329 ± 2946
2008-05-25	02:56:36	527177 ± 500	4514279 ± 500	10458 ± 1137
2008-05-25	04:14:03	527458 ± 862	4514796 ± 500	10057 ± 1277
2008-05-25	05:23:09	527144 ± 500	4514792 ± 623	9393 ± 1839
2008-05-25	21:17:20	527449 ± 500	4513741 ± 500	11437 ± 672
2008-05-26	18:13:03	527611 ± 500	4514308 ± 500	9361 ± 623
2008-05-26	00:41:26	533877 ± 500	4506467 ± 500	1790 ± 500
2008-05-27	16:39:09	526860 ± 500	4514342 ± 500	10251 ± 500
2008-05-27	16:19:33	527218 ± 500	4513627 ± 500	11058 ± 572
2008-05-27	17:24:07	527356 ± 500	4514226 ± 500	10591 ± 526
2008-05-27	21:09:47	527032 ± 500	4514439 ± 500	10798 ± 820
2008-05-28	10:19:58	527392 ± 500	4514277 ± 526	11194 ± 556
2008-05-28	10:22:09	527636 ± 500	4513831 ± 501	11212 ± 522
2008-06-02	15:36:15	537982 ± 559	4516439 ± 986	81 ± 676
2008-06-02	22:51:38	538163 ± 8106	4504389 ± 1089	8155 ± 902
2008-06-02	23:11:03	539862 ± 500	4503921 ± 500	14081 ± 591
2008-06-09	21:39:06	562922 ± 3325	4504830 ± 4142	927 ± 1119
2008-06-15	21:26:12	529349 ± 1840	4523422 ± 2054	7568 ± 6495
2008-06-16	04:12:34	519312 ± 500	4523824 ± 500	12243 ± 625
2008-06-19	03:35:24	534369 ± 1117	4476910 ± 769	8084 ± 500
2008-06-27	05:32:22	515218 ± 500	4525127 ± 500	12562 ± 500
2008-06-29	01:38:36	543055 ± 500	4513408 ± 500	5792 ± 601
2008-07-06	22:28:52	494218 ± 500	4555156 ± 745	60 ± 500
2008-07-10	23:17:18	552985 ± 638	4501671 ± 500	11026 ± 877
2008-07-11	08:42:36	549607 ± 1711	4501096 ± 745	4315 ± 976
2008-07-14	00:11:15	559668 ± 1713	4490348 ± 968	5782 ± 582
2008-07-29	23:06:20	532383 ± 500	4516463 ± 500	11606 ± 555
2008-07-31	09:18:55	552496 ± 500	4503816 ± 500	4271 ± 1024
2008-07-31	09:19:39	549449 ± 4808	4502180 ± 2628	2390 ± 1857
2008-08-03	07:49:00	546602 ± 1239	4491746 ± 855	4477 ± 602
2008-08-06	19:07:04	545158 ± 500	4513821 ± 500	668 ± 1333

2008-08-07	01:00:10	510129 ± 500	4542921 ± 500	7875 ± 500
2008-08-07	01:23:51	510414 ± 616	4544287 ± 567	10223 ± 3195
2008-08-07	01:41:51	510648 ± 681	4542312 ± 500	6113 ± 1947
2008-08-10	03:37:41	514574 ± 500	4523079 ± 689	1234 ± 1134
2008-08-13	02:56:17	523099 ± 500	4518428 ± 1010	9711 ± 500
2008-09-04	02:31:52	525856 ± 500	4507354 ± 589	11552 ± 1220
2008-09-05	07:17:40	547818 ± 500	4502228 ± 500	8086 ± 908
2008-09-06	01:13:24	546363 ± 1430	4499724 ± 649	12627 ± 4294
2008-09-10	04:12:04	539585 ± 500	4499444 ± 500	14459 ± 501
2008-09-10	08:19:09	514819 ± 500	4523346 ± 500	11736 ± 514
2008-09-12	07:16:58	560608 ± 12016	4495694 ± 9210	9308 ± 4311
2008-09-15	00:44:20	545886 ± 500	4478165 ± 500	218 ± 853
2008-09-18	17:37:38	563818 ± 1860	4502176 ± 860	8799 ± 1175
2008-09-23	01:52:31	567339 ± 500	4500127 ± 500	15367 ± 500
2008-10-06	02:27:26	515088 ± 500	4524559 ± 500	11245 ± 500
2008-10-10	21:25:45	560147 ± 583	4500481 ± 500	10482 ± 1308
2008-10-11	11:02:57	541864 ± 600	4520840 ± 1029	19120 ± 1788
2008-10-14	18:05:16	537729 ± 500	4502901 ± 500	13092 ± 500
2008-10-20	02:22:13	509006 ± 500	4524964 ± 500	11184 ± 500
2008-10-20	03:28:24	507724 ± 1856	4519293 ± 1592	3360 ± 1659
2008-10-21	07:56:04	547343 ± 3628	4499652 ± 2200	2644 ± 871
2008-10-23	16:02:13	536667 ± 500	4504464 ± 500	8705 ± 500
2008-11-03	05:24:09	538382 ± 500	4500627 ± 500	7303 ± 585
2008-11-07	08:36:57	553279 ± 500	4503499 ± 500	6173 ± 1159
2008-11-08	09:24:22	547434 ± 500	4492843 ± 500	4270 ± 658
2008-11-08	09:46:42	556716 ± 2468	4494626 ± 560	2146 ± 1975
2008-11-12	19:31:39	570658 ± 1133	4488013 ± 1455	7341 ± 3784
2008-11-14	01:59:21	572246 ± 588	4484470 ± 906	6159 ± 1900
2008-11-14	20:44:51	565360 ± 500	4499430 ± 500	16611 ± 800
2008-11-14	21:04:26	567545 ± 722	4498660 ± 500	18534 ± 1558
2008-11-17	00:13:01	572180 ± 500	4486389 ± 920	4399 ± 4840
2008-11-18	19:54:00	574684 ± 745	4486018 ± 500	16208 ± 1354
2008-11-18	20:05:29	577881 ± 776	4485546 ± 1124	9451 ± 2410
2008-11-18	22:14:27	572639 ± 701	4485859 ± 537	7829 ± 1890
2008-11-19	05:00:28	534718 ± 643	4516127 ± 500	6312 ± 1082
2008-11-19	12:45:38	535155 ± 593	4521646 ± 549	102 ± 500
2008-11-19	19:17:53	549654 ± 2919	4490454 ± 1071	1933 ± 3164
2008-11-22	14:53:13	540094 ± 500	4499503 ± 500	11293 ± 514
2008-11-25	07:44:00	536123 ± 500	4513596 ± 654	3579 ± 527
2008-12-25	11:56:28	527679 ± 500	4520233 ± 593	6656 ± 944
2008-12-25	18:55:58	571302 ± 782	4465794 ± 1067	2970 ± 2415
2009-01-15	23:08:34	550479 ± 19249	4495937 ± 5546	11063 ± 11393
2009-02-12	17:26:59	556363 ± 593	4485068 ± 610	11253 ± 500
2009-02-26	21:26:26	532718 ± 500	4517127 ± 500	7875 ± 500

2009-02-28	04:44:13	565528 ± 1153	4459305 ± 500	125 ± 596
2009-02-28	15:24:50	546270 ± 634	4478272 ± 923	1955 ± 855
2009-03-02	18:34:12	546253 ± 500	4482134 ± 500	79 ± 500
2009-04-06	13:13:09	512718 ± 5513	4504127 ± 2842	3187 ± 548
2009-04-16	08:15:14	553283 ± 513	4459203 ± 500	17081 ± 3998
2009-05-02	21:57:12	539436 ± 9831	4512268 ± 1841	6655 ± 610
2009-05-08	21:48:12	543452 ± 557	4499231 ± 1032	5937 ± 5791
2009-05-09	20:50:14	544405 ± 500	4502406 ± 500	4001 ± 539
2009-05-13	03:13:23	513677 ± 500	4525127 ± 500	9349 ± 500
2009-05-15	08:51:16	534377 ± 500	4502239 ± 500	12446 ± 500
2009-05-17	20:17:03	534381 ± 500	4502617 ± 500	11990 ± 500
2009-05-17	20:20:20	534651 ± 500	4502060 ± 500	10926 ± 572
2009-05-17	21:41:50	514292 ± 500	4522440 ± 500	4391 ± 500
2009-05-18	00:01:10	527742 ± 500	4510373 ± 500	5635 ± 500
2009-05-18	02:02:15	513558 ± 500	4522467 ± 500	5432 ± 649
2009-05-18	16:26:03	525890 ± 500	4512289 ± 500	3590 ± 770
2009-05-18	17:15:02	526477 ± 500	4511681 ± 500	5042 ± 616
2009-05-18	18:23:48	513467 ± 500	4525960 ± 500	10305 ± 500
2009-05-19	00:45:19	535609 ± 500	4513838 ± 500	4191 ± 1711
2009-05-19	10:44:09	562294 ± 1923	4500530 ± 1396	15780 ± 3796
2009-05-19	22:03:36	530444 ± 500	4520188 ± 500	6611 ± 1806
2009-05-22	08:59:18	528969 ± 500	4510000 ± 500	5798 ± 637
2009-05-22	12:30:01	542641 ± 691	4502467 ± 873	7704 ± 1236
2009-05-22	19:25:53	529922 ± 500	4510439 ± 500	5057 ± 500
2009-05-29	02:52:40	534718 ± 612	4501217 ± 1381	10690 ± 1418
2009-05-31	04:13:40	570375 ± 1315	4495680 ± 541	21361 ± 606
2009-06-10	01:24:52	530211 ± 500	4511535 ± 583	3751 ± 635
2009-06-10	09:31:09	529848 ± 500	4511856 ± 500	4041 ± 500
2009-06-12	02:54:34	570219 ± 2028	4478920 ± 3034	8333 ± 2311
2009-06-12	17:58:42	525253 ± 522	4514520 ± 500	15724 ± 543
2009-06-13	05:28:14	529476 ± 500	4510820 ± 500	3390 ± 500
2009-06-15	18:23:40	520684 ± 500	4470179 ± 1141	6676 ± 972
2009-06-15	23:56:47	520718 ± 500	4470287 ± 1178	6145 ± 1227
2009-06-17	10:11:11	525077 ± 739	4512968 ± 500	1847 ± 2000
2009-06-17	22:36:02	532170 ± 500	4511734 ± 500	3848 ± 738
2009-06-18	07:57:15	529556 ± 500	4510472 ± 500	5028 ± 633
2009-06-19	20:34:57	531820 ± 500	4512525 ± 500	2560 ± 500
2009-06-23	00:19:19	528683 ± 500	4507782 ± 500	3834 ± 500
2009-06-23	04:15:24	526718 ± 500	4510127 ± 500	3447 ± 500
2009-06-23	18:49:11	532219 ± 500	4511684 ± 500	4804 ± 500
2009-06-24	21:56:27	530718 ± 500	4511874 ± 500	2408 ± 500
2009-06-25	23:04:19	512813 ± 500	4524766 ± 500	10555 ± 579
2009-06-28	07:32:13	567293 ± 1255	4489679 ± 500	13551 ± 2870
2009-06-28	23:48:06	554581 ± 869	4499918 ± 600	19210 ± 2700

2009-06-29	04:05:23	517934 ± 500	4518627 ± 500	12900 ± 500
2009-06-29	04:51:19	517490 ± 616	4518627 ± 500	13003 ± 682
2009-06-29	06:09:22	518565 ± 500	4518779 ± 500	13181 ± 500
2009-06-29	09:46:37	517213 ± 500	4519576 ± 627	12745 ± 808
2009-06-29	10:54:23	517988 ± 500	4518978 ± 500	12555 ± 601
2009-06-29	12:07:40	519912 ± 3380	4519749 ± 2771	11849 ± 2567
2009-06-30	21:07:33	551984 ± 5991	4500026 ± 618	18444 ± 5158
2009-07-06	06:30:51	546513 ± 500	4490810 ± 1833	767 ± 2489
2009-07-07	19:27:09	535534 ± 500	4505408 ± 500	1605 ± 1060
2009-07-12	00:34:22	531321 ± 500	4505231 ± 500	1812 ± 500
2009-07-14	05:19:45	518546 ± 500	4519067 ± 500	14048 ± 500
2009-07-14	06:00:31	517718 ± 500	4518473 ± 500	12399 ± 500
2009-07-15	18:25:36	519523 ± 692	4522349 ± 500	14165 ± 1127
2009-07-25	00:15:11	565485 ± 1768	4489217 ± 1804	15703 ± 4466
2009-07-25	06:56:22	514259 ± 500	4527684 ± 820	12093 ± 543
2009-07-25	07:46:39	512869 ± 892	4528116 ± 500	11902 ± 567
2009-07-27	21:29:50	542005 ± 500	4515088 ± 836	4664 ± 607
2009-07-27	21:56:03	542394 ± 518	4515764 ± 1298	4745 ± 984
2009-07-27	22:40:42	541765 ± 500	4515013 ± 1580	4815 ± 1093
2009-07-28	08:06:08	541798 ± 2124	4519997 ± 869	2952 ± 1651
2009-07-28	11:46:19	536218 ± 500	4504765 ± 500	462 ± 628
2009-08-05	12:08:03	496117 ± 572	4552352 ± 500	5465 ± 1282
2009-08-08	04:41:51	540780 ± 500	4510963 ± 500	3362 ± 950
2009-08-23	13:32:02	535421 ± 500	4502397 ± 500	11933 ± 500
2009-09-01	00:06:04	524071 ± 500	4518830 ± 500	10271 ± 877
2009-09-01	00:44:32	524404 ± 500	4518905 ± 500	10447 ± 751
2009-09-01	04:46:15	524218 ± 500	4518943 ± 500	11823 ± 1227
2009-09-01	05:39:18	524312 ± 500	4519053 ± 763	5351 ± 3195
2009-09-01	06:53:11	524718 ± 500	4519932 ± 500	12038 ± 500
2009-09-01	11:41:33	524718 ± 500	4518472 ± 500	10330 ± 625
2009-09-01	16:26:33	524218 ± 500	4517712 ± 500	9266 ± 1070
2009-09-01	18:19:24	531802 ± 500	4511535 ± 500	1896 ± 500
2009-09-01	18:43:45	525218 ± 500	4518127 ± 500	3708 ± 500
2009-09-02	11:37:10	494547 ± 574	4536839 ± 746	24340 ± 1002
2009-09-07	13:10:35	496053 ± 500	4553306 ± 500	7604 ± 966
2009-09-08	15:33:36	533422 ± 500	4514356 ± 500	7459 ± 792
2009-09-13	01:20:04	518718 ± 500	4520533 ± 500	10649 ± 1590
2009-09-13	04:11:28	496215 ± 500	4553616 ± 655	6671 ± 860
2009-09-23	01:30:27	519995 ± 1943	4560627 ± 500	9242 ± 11947
2009-09-27	16:43:50	521718 ± 500	4523379 ± 500	5791 ± 500
2009-09-28	16:01:25	518218 ± 500	4513127 ± 500	8395 ± 500
2009-09-28	22:57:07	518218 ± 500	4513299 ± 500	9594 ± 500
2009-10-04	17:25:29	562907 ± 1309	4500320 ± 927	12615 ± 2107
2009-10-05	23:36:18	535421 ± 500	4503154 ± 500	14049 ± 500

2009-10-06	02:11:32	524228 ± 2671	4477006 ± 4178	11287 ± 2668
2009-10-06	18:16:10	511144 ± 500	4524250 ± 500	10082 ± 500
2009-10-24	23:06:39	569218 ± 500	4473127 ± 500	5791 ± 500
2009-10-26	18:05:09	533545 ± 500	4521281 ± 500	12240 ± 982
2009-11-01	17:28:47	589914 ± 500	4482988 ± 6356	7667 ± 630
2009-11-06	02:52:10	519564 ± 739	4513510 ± 500	4749 ± 500
2009-11-16	09:04:20	533139 ± 606	4498712 ± 731	9512 ± 1423
2009-11-20	18:24:33	511684 ± 687	4526561 ± 500	9991 ± 543
2009-12-03	05:06:46	530547 ± 500	4513448 ± 500	11198 ± 556
2009-12-03	08:19:11	559218 ± 500	4467646 ± 865	4712 ± 500
2009-12-04	00:08:01	553599 ± 500	4485341 ± 2537	5744 ± 500
2009-12-18	18:00:57	532869 ± 500	4499312 ± 931	9725 ± 1053
2009-12-20	03:07:32	539191 ± 500	4503729 ± 942	8514 ± 2771
2009-12-21	21:08:16	517545 ± 500	4517793 ± 500	13645 ± 500
2010-01-14	00:53:34	537718 ± 500	4511127 ± 500	12099 ± 500
2010-01-24	00:20:19	576685 ± 1348	4486550 ± 2179	11453 ± 2067
2010-02-01	01:42:00	565113 ± 562	4490496 ± 500	30077 ± 773
2010-02-27	19:32:48	544365 ± 3100	4511479 ± 2068	15168 ± 7256
2010-03-10	02:52:14	559812 ± 1273	4497678 ± 788	18257 ± 2294
2010-03-12	18:08:22	542357 ± 613	4497539 ± 613	13083 ± 500
2010-03-17	16:59:42	515921 ± 500	4523377 ± 500	11300 ± 500

Table 4.1: Locations results related to 196 seismic events occurred in the Irpinia region between the February 2008 and March 2010. Since we used a grid spacing of 500 m, all uncertainties lower than this value are set to 500 m (Coordinates are provided in the UTM system).

Chapter 5

Conclusions

The main goal of this thesis was the development of an automated seismic event location method for microseismic monitoring purposes. Since data quality control is the first step, before starting other kind of applications (e.g seismic event location, moment tensor inversion etc.) I firstly developed a novel technique to retrieve the orientation of seismic sensor based on the comparison of recordings at different stations. In contrast to other procedures the method is based on a complex linear least-squares inversion approach and makes use of the full waveform, resulting more stable than standard approaches. The first paper (chapter 1) demonstrates the usefulness of this method for different acquisition geometries and has been tested using both synthetic and real datasets. The good quality of our results shows that our methodology can be successfully applied to different acquisition geometries, ranging from local to regional scales and environments. For example, it can be used as a tool to remotely validate the correct orientation of stations for seismic arrays or temporary networks during and after deployment. The procedure is very appealing towards seismic station deployments, where the sensor orientation is poorly known. Among the ocean bottom seismometers other examples include borehole installations and seismic station in volcanic areas or polar regions (where compass orientation can be affected by large errors). There are several advantages in using our approach. (1) The method is not requiring a linearly polarized waveform and can be either applied to full waveform or to a shorter part of the signal (e.g. P-wave onset); while polarization based method can use only seismic phases characterized by an high rectilinearity (P or S waves), our method can be applied using both body and surface waves. (2) By solving a linear inverse problem in complex domain, our solution corresponds to the global minimum of the misfit function; furthermore, it is possible to use more seismic events (earthquakes or explosions) simultaneously to better constrain the final solution. In consequence, one may use continuous seismic noise originated outside the network to retrieve the sensor orientations. (3) Computing relative orientation angles result faster than cross-correlation based methods, especially when dealing with a very large number of sensors.

In the second part of the thesis (chapter 3 and 4) I present a new location method. A large number of seismological tools for the automatic location of seis-

mic events have been proposed in the past, most of these algorithms are based on the automatic picking of the main seismic phases (generally P and S first arrivals). With the current available techniques it is possible to accurately pick the P onsets but reliable picking of the S onsets is still problematic, especially for local events where the P coda overlaps the S wave. The performance of automatic pickers is limited in presence of noisy data, when picking and phases identification might be difficult. In addition, the correct association of picks to P, S or other phases may be difficult in practice. In order to overcome these problems I developed an automatic seismic event location procedure that is based on a waveform stacking approach. The location method I propose is picking free, noise robust and completely automated. I have successfully tested our approach on both synthetic and real datasets. Results obtained with synthetics show that this methodology is reliable even with sparse networks and noisy traces, while real data results have been compared with accurate manual locations, when available. The application of this location approach to two different dataset is presented in two publications (one published and another one accepted for publication on an international peer reviewed journal). In the first paper (chapter 3), published on the Seismological Research Letters, I located microseismic events induced by mining activity in a coal mine in the Ruhr region (Germany). Location results were in agreement (with 90% of the event within 200 m from the manual locations) with manual location obtained by an other research group. In the second paper (chapter 4), accepted for publication to the Geophysical Journal International, I applied this method to locate crustal earthquakes recorderd by a regional network in the Campania-Lucania region (Southern Italy). Since this dataset is characterized by noisy seismograms contaminated by a strong P coda overlapping the S wave first onset, the use of the former location approach resulted problematic. Thus, it was necessary an additional technical improvement to the algorithm, making use of the principal eigenvalue characteristic function which is more sensitive to the S wave. Results confirmed the better performance of the improved location method. An amount of 196 seismic event recorded from the Irpinia Seismic Network (ISNet) were successfully located. A subset of these events has been compared with accurate manual locations obtained by using a double differences algorithm. In the 90% of cases the proposed location method preformed better than the standard method based on automatic picking, with an error reduction of about 3 km. In chapters 3 and 4 it has been shown that the location algorithm and its implementation are fully functional. For both datasets (induced microseismic events and crustal earthquakes) the location algorithm performs very well even in presence of noisy waveforms. The proposed approach has several advantages: (1) the method is completely automated, (2) it is robust and it has good performance also with low-quality data, (3) it consider P and S arrivals for location (4) it exploits the waveform information content without the need of synthetic seismograms computation. On the other hand, similarly to all other migration based methods it is computationally intensive and the location process require more computing time than standard approaches. The most critical point is the computing time needed

to estimate uncertainties. In this step, each event has to be located several time by using a bootstrap approach. Furthermore, the performance of the algorithm also depends on the number of grid points, on the number of stations and on the number of samples of each trace. Although the current version of the location algorithm is coded in parallel using a shared memory paradigm, the computing time is still larger than the standard location approaches.

As future development It would be useful to improve the grid search strategy moving to stochastic optimization methods (e.g. Montecarlo based methods) which also allow to evaluate uncertainties without the need of the bootstrap approach. I believe this improvement will strongly reduce the computing time needed to perform a location. Another important improvement (which is currently a work in progress) is to extend the algorithm to work with continuous data streams in order to simultaneously detect and locate seismic events. Supported by very good results I conclude that both methods discussed in this thesis (orientation method and location method) can contribute considerably to seismic monitoring operations and to process or reprocess large datasets. Thus, they are extremely valuable for a variety of different applications, such as microseismic monitoring of mining environments and oil reservoir or the automated location of crustal earthquakes recorded by regional networks. A third development at hand is to define proxies for the radiation pattern of the seismic sources using the final location of the stacking method. The additional computing time would be negligible, and the interest in such estimate would be enormous even if the uncertainties would be large. Another possible development is to stack the coherence matrix maps to visualize the seismicity rate of a cluster of events. Such a type of seismicity map would be very different to the conventional visualizations, where individual events are plotted, since the uncertainties from the stacks are directly considered. It would be possibly give a much better objective view of what is going on than possible with conventional seismicity maps.

Bibliography

R. Allen. Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, 68:1521–1532, 1978. ISSN 00371106.

R. Allen. Automatic phase pickers: their present use and future prospects. *Bulletin of the Seismological Society of America*, 72:S225–S242, 1982. ISSN 00371106.

G. Ameri, A. Emolo, F. Pacor, and F. Gallovic. Ground-Motion Simulations for the 1980 M 6.9 Irpinia Earthquake (Southern Italy) and Scenario Events. *Bulletin of the Seismological Society of America*, 101:1136–1151, 2011. ISSN 00371106. doi: 10.1785/0120100231.

O. Amoroso, N. Maercklin, and A. Zollo. S-Wave Identification by Polarization Filtering and Waveform Coherence Analyses. *Bulletin of the Seismological Society of America*, 102:854–861, 2012a. ISSN 00371106. doi: 10.1785/0120110140.

O. Amoroso, A. Zollo, and J. Virieux. 3D seismic imaging of an active, normal fault zone in southern Apennines (Italy): Clues on fluid driven microearthquake fracturing. *Abstract S31C-07 presented at 2012 Fall Meeting, AGU, San Francisco, Calif., 3-7 Dec.*, 2012b.

M. Baer and U. Kradolfer. An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, 77:1437–1445, 1987. ISSN 00371106.

T. Baker, R. Granat, and R. W. Clayton. Real-time Earthquake Location Using Kirchhoff Reconstruction. *Bulletin of the Seismological Society of America*, 95: 699–707, 2005. ISSN 00371106. doi: 10.1785/0120040123.

M. Becquey and M. Dubesset. Three-component sonde orientation in a deviated well. *Geophysics*, 55(10):1386–1388, 1990.

M. Beyreuther, R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann. ObsPy: A Python Toolbox for Seismology. *Seismological Research Letters*, 81: 530, 2010.

M. Bischoff, A. Cete, R. Fritschen, and T. Meier. Coal mining induced seismicity in the ruhr area, germany. *Pure and Applied Geophysics*, 167(1-2):63–75, 2010.

A. Cichowicz. An automatic S-phase picker. *Bulletin of the Seismological Society of America*, 83:180–189, 1993. ISSN 00371106.

T. Dahm, M. Thorwart, E. R. Flueh, T. Braun, R. Herber, P. Favali, L. Beranzoli, G. D'Anna, F. Frugoni, and G. Smriglio. Ocean bottom seismometers deployed in tyrrhenian sea. *Eos, Transactions American Geophysical Union*, 83(29):309–315, 2002.

R. De Matteis, E. Matrullo, L. Rivera, T. A. Stabile, G. Pasquale, and A. Zollo. Fault Delineation and Regional Stress Direction from the Analysis of Background Microseismicity in the southern Apennines, Italy. *Bulletin of the Seismological Society of America*, 102:1899–1907, 2012. ISSN 00371106. doi: 10.1785/0120110225.

J. P. Di Siena, J. E. Gaiser, and D. Corrigan. Horizontal components and shear wave analysis of three component vsp data. In Stewart R. R. Toksoz M.N., editor, *Vertical Seismic Profiling Part B: Advanced Concepts*. Geophysical Press, London, 1984.

T. Diehl, N. Deichmann, E. Kissling, and S. Husen. Automatic S-Wave Picker for Local Earthquake Tomography. *Bulletin of the Seismological Society of America*, 99:1906–1920, 2009. ISSN 00371106. doi: 10.1785/0120080019.

J. Drew, H. Leslie, P. Armstrong, and G. Michard. Automated microseismic event detection and location by continuous spatial mapping. In *SPE Annual Technical Conference and Exhibition*, 2005.

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

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

D. Gajewski and E. Tessmer. Reverse modelling for seismic event characterization. *Geophysical Journal International*, 163(1):276–284, 2005.

L. Geiger. Herdbestimmung bei erdbeben aus den ankunftzeiten. *Nachrichten von der Koniglicher Gesellschaft der Wissenschaften zu Gottingen Mathematisch Physikalische Klasse*, (4,331-349), 1910.

L. Geiger. Probability method for the determination of earthquake epicenters from the arrival time only (translated from geiger's 1910 german article). *Bulletin of St. Louis University*, 8(1):56–71, 1912.

H. N. Gharti, V. Oye, M. Roth, and D. Kühn. Automated microearthquake location using envelope stacking and robust global optimization. *Geophysics*, 75(4): MA27–46, 2010. ISSN 00168033. doi: 10.1190/1.3432784.

S. A. Greenhalgh and I. M. Mason. Orientation of a downhole triaxial geophone. *Geophysics*, 60(4):1234–1237, 1995.

F. Grigoli, S. Cesca, M. Vassallo, and T. Dahm. Automated Seismic Event Location by Travel-Time Stacking: An Application to Mining Induced Seismicity. *Seismological Research Letters*, 84(4):666–677, July 2013. ISSN 0895-0695. doi: 10.1785/0220120191.

M. Hensch. On the interrelation of fluid-induced seismicity and crustal deformation at the columbo submarine volcano (aegean sea, greece). *Ph. D thesis, University of Hamburg*, 2009.

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

H. Kao and S. J. Shan. Rapid identification of earthquake rupture plane using Source-Scanning Algorithm. *Geophysical Journal International*, 168:1011–1020, 2007. ISSN 0956540X. doi: 10.1111/j.1365-246X.2006.03271.x.

K. B. Knowlton and T. W. Spencer. Polarization measurement uncertainty on three-component vsp. *Geophysics*, 61(2):594–599, 1996.

F. Krüger and M. Ohrnberger. Tracking the rupture of the mw = 9.3 sumatra earthquake over 1,150 km at teleseismic distance. *Nature*, 435(7044): 937–939, 2005a.

F. Krüger and M. Ohrnberger. Spatio-temporal source characteristics of the 26 december 2004 sumatra earthquake as imaged by teleseismic broadband arrays. *Geophysical research letters*, 32(24), 2005b.

F. Krüger and M. Weber. The effect of low-velocity sediments on the mislocation vectors of the grf array. *Geophysical journal international*, 108(1):387–393, 1992.

S. Larsen and J. Grieger. Elastic modeling initiative, part iii: 3-d computational modeling. In *1998 SEG Annual Meeting*, 1998.

X. Y. Li and J. Yuan. Geophone orientation and coupling in three-component sea-floor data: a case study. *Geophysical Prospecting*, 47(6):995–1013, 1999.

Y. C. Liao, H. Kao, A. Rosenberger, S. K. Hsu, and B. S. Huang. Delineating complex spatiotemporal distribution of earthquake aftershocks: an improved source-scanning algorithm. *Geophysical Journal International*, 189(3):1753–1770, 2012.

A. Lomax. The nonlinloc software guide. *ALomax Scientific, Mouans-Sartoux, France*, 2008.

A. Lomax, C. Satriano, and M. Vassallo. Automatic Picker Developments and Optimization: FilterPicker–a Robust, Broadband Picker for Real-Time Seismic Monitoring and Earthquake Early Warning. *Seismological Research Letters*, 83 (3):531–540, May 2012. ISSN 0895-0695. doi: 10.1785/gssrl.83.3.531.

N. Maercklin, G. Festa, S. Colombelli, and A. Zollo. Twin ruptures grew to build up the giant 2011 tohoku, japan, earthquake. *Scientific reports*, 2, 2012.

K. V. Mardia and P. E. Jupp. *Directional statistics*, volume 494. Wiley. com, 2009.

E. Matrullo, R. De Matteis, C. Satriano, O. Amoroso, and A. Zollo. An improved 1-D seismic velocity model for seismological studies in the Campania-Lucania region (Southern Italy). *Geophysical Journal International*, July 2013. ISSN 0956-540X. doi: 10.1093/gji/ggt224.

G. A. McMechan. Determination of source parameters by wavefield extrapolation. *Geophysical Journal International*, 71(3):613–628, 1982.

X. G. Miao, W. M. Moon, and B. Milkereit. A multioffset, three-component vsp study in the sudbury basin. *Geophysics*, 60(2):341–353, 1995.

P. Michaels. Use of principal component analysis to determine down-hole tool orientation and enhance sh-waves. *Journal of Environmental & Engineering Geophysics*, 6(4):175–183, 2001.

K. S. Miller. Complex linear least squares. *Siam Review*, 15(4):706–726, 1973.

Y. Nakamura, P. L Donoho, P. H. Roper, and P. M. McPherson. Large-offset seismic surveying using ocean-bottom seismographs and air guns: Instrumentation and field technique. *Geophysics*, 52(12):1601–1611, 1987.

V. Oye and W. L. Ellsworth. Orientation of three-component geophones in the san andreas fault observatory at depth pilot hole, parkfield, california. *Bulletin of the Seismological Society of America*, 95(2):751–758, 2005.

T. Plenefisch and K. Stammler. Investigations of upper mantle anisotropy beneath the graefenberg array—first usage of the new three component broadband stations in the sks splitting analysis. In *Geophys. Res. Abstr*, volume 10, 2008.

P. Podvin and I. Lecomte. Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools. *Geophysical Journal International*, 105:271–284, 1991. ISSN 0956540X. doi: 10.1111/j.1365-246X.1991.tb03461.x.

C. A. Rowe, R. C. Aster, B. Borchers, and C. J. Young. An automatic, adaptive algorithm for refining phase picks in large seismic data sets. *Bulletin of the Seismological Society of America*, 92(5):1660–1674, 2002.

J. L. Rubinstein and G. C. Beroza. Full waveform earthquake location: Application to seismic streaks on the calaveras fault, california. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 112(B5), 2007.

C. Satriano, A. Lomax, and A. Zollo. Real-Time Evolutionary Earthquake Location for Seismic Early Warning. *Bulletin of the Seismological Society of America*, 98:1482–1494, 2008. ISSN 00371106. doi: 10.1785/0120060159.

A.T. Sen, S. Cesca, M. Bischoff, Meier T., T. Dahm, and D. Kühn. Full moment tensor inversion and rupture modelling of mining induced seismicity. In *EGU General Assembly Conference Abstracts*, volume 14, page 2775, 2012.

R. E. Sheriff. Encyclopedic dictionary of applied geophysics, 2002.

T. A. Stabile, C. Satriano, A. Orefice, G. Festa, and A. Zollo. Anatomy of a microearthquake sequence on an active normal fault. *Scientific Reports*, 2: submitted, 2nd revision, 2012. ISSN 20452322. doi: 10.1038/srep00410.

M. Vassallo, C. Satriano, and A. Lomax. Automatic Picker Developments and Optimization: A Strategy for Improving the Performances of Automatic Phase Pickers. *Seismological Research Letters*, 83(3):541–554, May 2012. ISSN 0895-0695. doi: 10.1785/gssrl.83.3.541.

J. E. Vidale. Complex polarization analysis of particle motion. *Bulletin of the Seismological Society of America*, 76:1393–1405, 1986. ISSN 00371106.

F. Waldhauser and W. L. Ellsworth. A double-difference earthquake location algorithm: Method and application to the northern hayward fault, california. *Bulletin of the Seismological Society of America*, 90(6):1353–1368, 2000.

R. Wang. A simple orthonormalization method for stable and efficient computation of green's functions. *Bulletin of the Seismological Society of America*, 89(3): 733–741, 1999.

J. Wassermann. Locating the sources of volcanic explosions and volcanic tremor at stromboli volcano (italy) using beam-forming on diffraction hyperboloids. *Physics of the earth and planetary interiors*, 104(1):271–281, 1997.

E. Weber, V. Convertito, G. Iannaccone, A. Zollo, A. Bobbio, L. Cantore, M. Corciulo, M. Di Crosta, L. Elia, C. Martino, A. Romeo, and C. Satriano. An advanced seismic network in the Southern Apennines (Italy) for seismicity investigations and experimentation with earthquake early warning. *Seismological Research Letters*, 78:622–634, 2007. ISSN 08950695. doi: 10.1785/gssrl.78.6.622.

M. Withers, R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, and J. Trujillo. A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bulletin of the Seismological Society of America*, 88:95–106, 1998. ISSN 00371106.

M. Withers, R. Aster, and C. Young. An Automated Local and Regional Seismic Event Detection and Location System Using Waveform Correlation. *Bulletin of the Seismological Society of America*, 89:657–669, 1999.

X. Zeng and G. A. McMechan. Two methods for determining geophone orientations from vsp data. *Geophysics*, 71(4):V87–V97, 2006.