GFZ Helmholtz Centre Potsdam

Institute of Geosciences

Universität Potsdam

Earthquake source and receiver array optimal configuration

Nasim Karamzadeh Toularoud

Date of defense: 10.02.2020

Cumulative dissertation
to obtain the academic degree
"doctor rerum naturalium" (Dr. rer. nat.)
in the scientific discipline Geophysics

Submitted to the
Faculty of Mathematics and Natural Sciences
at the University of Potsdam, Germany

Published online at the
Institutional Repository of the University of Potsdam:
https://doi.org/10.25932/publishup-45982
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-459828

Statement of originality

I hereby declare that this thesis is the product of my own work. All the assistance received in preparing this thesis and the sources used have been acknowledged.

Nasim Karamzadeh Toularoud
April 2020

Abstract

Seismic receiver arrays have a variety of applications in seismology, particularly when the signal enhancement is a prerequisite to detect seismic events, and in situations where installing and maintaining sparse networks are impractical. This thesis has mainly focused on the development of a new approach for seismological source and receiver array design. The proposed approach deals with the array design task as an optimization problem. The criteria and prerequisite constraints in array design task are integrated in objective function definition and evaluation of a optimization process. Three cases are covered in this thesis: (1) a 2-D receiver array geometry optimization, (2) a 3-D source array optimization, and (3) an array application to monitor microseismic data, where the effect of different types of noise are evaluated.

A flexible receiver array design framework implements a customizable scenario modelling and optimization scheme by making use of synthetic seismograms. Using synthetic seismograms to evaluate array performance makes it possible to consider additional constraints, e.g. land ownership, site-specific noise levels or characteristics of the seismic sources under investigation. The use of synthetic array beamforming as an array design criteria is suggested. The framework is customized by designing a 2-D small scale receiver array to monitor earthquake swarm activity in northwest Bohemia/ Vogtland in central Europe. Two sub-functions are defined to verify the accuracy of horizontal slowness estimation; one to suppress aliasing effects due to possible secondary lobes of synthetic array beamforming calculated in horizontal slowness space, and the other to reduce the event's mislocation caused by miscalculation of the horizontal slowness vector. Subsequently, a weighting technique is applied to combine the sub-functions into one single scalar objective function to use in the optimization process.

The idea of optimal array is employed to design a 3-D source array, given a well-located catalog of events. The conditions to make source arrays are formulated in four objective functions and a weighted sum technique is used to combine them in one single scalar function. The criteria are: (1) accurate slowness vector estimation, (2) high waveform coherency, (3) low location error and (4) high energy of coda phases. The method is evaluated by two experiments, (1) a synthetic test using realistic synthetic seismograms, (2) using real seismograms, and for each case optimized SA elements are configured using the data from the Vogtland area.

The location of a possible scatterer in the velocity model, that makes the converted/reflected phases, e.g. sp-phases, is retrieved by a grid search method using the optimized SA. The accuracy of the approach and the obtained results demonstrated that the method is applicable to study the crustal structure and the location of crustal scatterers when the strong converted phases are observed in the data and a well-located catalog is available.

Small aperture arrays are employed in seismology for a variety of applications, ranging from pure event detection to monitor and study of microcosmic activities. The monitoring of microseismicity during temporary human activities is often difficult, as the signal-to-noise ratio is very low and noise is strongly increased during the operation. The combination of small aperture seismic arrays with shallow borehole sensors offers a solution. We tested this monitoring approach at two different sites, (1) accompanying a fracking experiment in sedimentary shale at 4 km depth, and (2) above a gas field under depletion. Arrays recordings are compared with recordings available from shallow borehole sensors and examples of detection and location performance of the array are given. The effect of different types of noise at array and borehole stations are compared and discussed.

Zusammenfassung

Seismische Arrays haben eine Vielzahl von Anwendungen in der Seismologie, insbesondere wenn die Signalverbesserung eine Voraussetzung ist für die seismische Ereignisse erkennen, und in Situationen, in denen die Installation und Wartung spärlicher Netzwerke ist unpraktisch. Diese Arbeit hat sich vor allem auf die Entwicklung eines neuen Ansatzes für seismologische Quellen- und Empfänger-Array-Design konzentriert. Der vorgeschlagene Ansatz beschäftigt sich mit der Array-Design-Aufgabe als Optimierungsproblem. Die notwendigen Kriterien und Randbedingungen, die für die seismologische Array-Design-Aufgabe wichtig sind, werden in die objektive Funktionsdefinition und Bewertung eines Optimierungsprozesses integriert. In dieser Arbeit werden drei Fälle behandelt. (1) eine 2-D-Empfänger-Array-Geometrieoptimierung, (2) eine 3D-Quellfeldoptimierung, und (3) eine Array-Anwendung zum Überwachen mikroseismischer Daten, wobei die Auswirkungen verschiedener Arten von Lärm werden bewertet. Ein flexibles Empfänger-Array-Design-Framework wird eingeführt, das ein anpassbares Szenario-Modellierungs- und Optimierungsschema unter Verwendung synthetischer Seismogramme implementiert. Die Verwendung synthetischer Seismogramme zur Bewertung der Array-Leistung ermöglicht es, zusätzliche Einschränkungen, wie z.B. Landbesitz, zu berücksichtigen, standortspezifische Lärmpegel oder Eigenschaften der seismischen Quellen unter Berücksichtigung von Untersuchung. Die Verwendung von synthetischem Array-Strahlformung als Array-Design-Kriterium wird vorgeschlagen. Das Array-Design-Framework wird durch die Entwicklung eines 2-D-Kleinempfänger-Arrays zur Überwachung der Erdbebenschwarmaktivität im Vogtland in Mitteleuropa angepasst.

Es werden zwei Teilfunktionen definiert, um die Genauigkeit der horizontalen Langsamkeitsschätzung zu überprüfen. Eine zur Unterdrückung von Aliasing-Effekten aufgrund möglicher Nebenkeulen der synthetischen Strahlformung, berechnet im horizontalen Langsamkeitsraum, und zum anderen, um die Fehlstellung des Ereignisses durch eine Fehlberechnung des horizontalen Langsamkeitsvektors zu reduzieren. Anschliessend wird eine Gewichtungstechnik angewendet, um die Kombination von die Unterfunktionen zu einer einzigen skalaren Zielfunktion zusammenfassen, die in der Optimierungsprozess verwendet werden kann.

Die Idee des Array Optimal Design wird verwendet, um ein 3-D Source Array zu entwerfen, das einen gut lokalisierten Katalog von Erdbebenereignissen enthält. Die Bedingungen für die Herstellung von Quellarrays werden in vier Zielfunktionen formuliert, und eine gewichtete Summentechnik wird verwendet, um sie in einer einzigen skalaren Funktion zu kombinieren. Die Kriterien sind: (1) genaue Langsamkeitsvektorschätzung, (2) hohe Wellenform-Kohärenz, (3) niedriger Ortsfehler und (4) bis hohe Energie der Coda-Phasen. Die Methode wird durch zwei Experimente bewertet, (1) ein synthetischer Test mit realistischen synthetischen Seismogrammen, (2) mit realen Seismogrammen und optimierte SA-Elemente werden für jeden Fall unter Verwendung der Daten aus dem Vogtland gefunden. Die Position eines möglichen Streuers im Geschwindigkeitsmodell, der die konvertierten/reflektierten Phasen, z.B. sp-Phasen, erzeugt, wird durch ein Rastersuchverfahren mit dem optimierten SA ermittelt.

Die Genauigkeit des Ansatzes und die erhaltenen Ergebnisse sind überzeugend, dass die Methode anwendbar ist, um die Krustenstruktur und die Position von Krustalstreuern zu untersuchen, wenn die stark konvertierten Phasen in den Daten beobachtet werden und ein gut lokalisierter Katalog verfügbar ist.

Die Überwachung der Mikroseismizität bei solchen temporären menschlichen Aktivitäten ist

oft schwierig, da der Signal-Rausch-Verhältnis sehr niedrig ist und das Rauschen während des Betriebs stark erhöht wird. Die Kombination von seismischen Arrays mit kleiner Apertur und flachen Bohrlochsensoren bietet eine Lösung. Wir haben diesen Überwachungsansatz an zwei verschiedenen Standorten getestet. (1) Begleiten eines Fracking-Experiments in sedimentärem Schiefer in 4 km Tiefe und (2) über einem Gasfeld unter Erschöpfung. Die Aufzeichnungen von Arrays werden mit den Aufzeichnungen von flachen Bohrlochsensoren verglichen, und es werden Beispiele für die Detektions- und Standortleistung der Arrays gegeben. Die Auswirkungen verschiedener Arten von Lärm an Array- und Bohrlochstationen werden verglichen und diskutiert.

Contents

Chapter 1

Introduction

In seismology the word "array" mainly refers to the receiver arrays, which are a number of seismic sensors deployed in a special geometry with common precise timing, acquisition parameters and instrument type. Seismic arrays have gained magnificent interest from the early days of application and have been installed all around the globe, while their application extended from the initial mission, i.e monitoring compliance with Comprehensive Test Ban Treaty (CTBT) to various seismological research (Havskov and Alguacil, 2004; Havskov and Ottemöller, 2010; Schweitzer et al., 2012; Rost and Thomas, 2009, 2002).

Array recordings are processed as an ensemble, and array signal processing methods rely on high signal coherency across all stations. Thus the sensor deployment area should be small enough compared to wavelength of the desired signals, that the interested signals are correlated, while the background seismic noise at the frequency range of signals are uncorrelated, so the signal-to-noise ratio (SNR) is improved. Particular utility of an array in a seismic monitoring system is to estimate the direction and the horizontal slowness $\mathbf{S}$, or horizontal wavenumber vector $\mathbf{K} = \omega.\mathbf{S}$, of incoming waves, assuming ω is the frequency. While the number of array stations controls the SNR gain achievable by the array, the array geometry defines the limits for resolvable wavenumbers. For instance, small aperture arrays can not distinguish between waves with small wavenumber differences, and for crossing waves with long horizontal wavelengths compared to the array aperture, such arrays act like a single station. So theoretically, the upper limit for the longest horizontal wavelength that can meaningfully be analysed by array techniques is about the aperture of the array. In addition, a wave crossing the array should be sampled by at least two stations, i.e. the smallest recordable wavelength (Havskov and Alguacil, 2004; Havskov and Ottemöller, 2010; Schweitzer et al., 2012; Rost and Thomas, 2009, 2002). Array geometry, spatial dimensions and data quality play important roles in array analysis. In practice, to increase the accuracy of array processing results, the parameters that can be adjusted to the specific purpose and noise condition are the array geometry, sensor placements and the processing method.

Arrays are often designed for a specific purpose. Examples include the monitoring of seismicity related to specific faults, volcanoes or earthquake swarms, nuclear explosions and specific local sites studies of micro-earthquakes after nuclear explosions, the tracking of earthquake ruptures, or structural analyses to image reflections and diffractions. Array response function (ARF) is often used as array design criteria, that depends on array geometry and frequency content of signal of interest (Schweitzer et al., 2012; Rost and Thomas, 2009, 2002), and has been employed and discussed for many applications such as surface wave studies (Marano et al., 2014; Wathelet et al., 2008). Based on the ARF properties, Kennett et al. (2015) suggested the use of spiral-arm shaped arrays, which allows deployment of relatively large aperture arrays with a limited number of stations to achieve a high slowness resolution. In early years of array seismology developments, Haubrich (1968) investigated a special class of linear and planar arrays involving uniform patterns of stations, proposed a design method based on the coarray's properties and showed how the beam pattern can be optimized for a given number of sensors.

In many applications, information about deployment site and the experiment, such as site-specific noise level, properties of the target events, land-use classes (Büttner et al., 2004) and other constraints on the logistics are available. For instance, to ensure high signal coherence and low signal loss across the array stations, it is advantageous to take focal properties of the target events and the site-specific noise level into consideration. However, considering such information is not standard, a quantitative integration method is required.

In this research we realize the array design task as an optimization problem, which is finding the best array geometry (an optimized model) from all feasible geometries. Such an approach has been successfully used in sparse network geometry design and improvement, for example by Hardt and Scherbaum (1994) and Kraft et al. (2013), however not involved synthetic full waveform. An optimization algorithm (i.e. simulated annealing, genetic algorithm, ...) applies a procedure which is evaluating a cost function (objective function) iteratively by comparing different solutions until an optimum solution is found. Then, to employ the optimization approach in the optimal array design framework, an appropriate objective function definition is very important. We suggest that the objective function should have two special characteristics:

(1) It should be flexibly defined based on the general array application as well as monitoring goals of each case study. For example, as a general goal of any array application, attaining the highest accuracy of horizontal slowness vector estimation can be formulated as one objective function. Extending a single objective optimization to a multi objective optimization approach allows us to define and formulate other customized criteria as objective functions and include them in optimization procedure.

(2) It should be capable to include other information and general requirements of array station selection such as local noise properties and site specific conditions. In other words, the objective function should accept such information as input parameters.

To fulfill these requirements, in this thesis, it is suggested to make use of synthetic seismograms to model realistic conditions and scenarios of the experiment and to use synthetic array beamforming as main design criteria. The use of realistic synthetic seismograms makes it possible to take into account ranges of possible source mechanisms, geometry of the seismogenic zone, propagation velocities and noise levels at potential deployment sites. It is also possible to use segments of real noise (if it is available for deployment site) as complementing synthetic seismograms to make more realistic estimation of seismic monitoring system (López-Comino et al., 2017). The synthetic array data is then used to evaluate one or more objective functions to be minimized during the optimization process.

The idea of array elements optimal design is employed for the source array elements optimal design, as well. Based on the reciprocity theorem of Green's function and under particular conditions, a number of seismic sources recorded at a single station are named a "source array", in an analogy of array of sensors, then the same array processing methods are applicable on recordings of a single station (Spudich and Bostwick, 1987; Scherbaum et al., 1991; Krüger et al., 1996). The conditions to make source arrays are that the coordinates, the origin times, and the depths of the sources are precisely known. In addition, akin to the waveform coherency requirement for a receiver array, waveforms of a source array should be coherent. This criterion implicitly indicates that the travel path and source mechanisms of events are similar. If they are not identical, it is necessary to correct the effect of source mechanisms from the recorded waveforms. This basic idea provides a unique tool to study the seismic structure of the areas with as few as a single recording instrument using array processing methods. In addition, by simultaneous use of a receiver and source array, the double array method (Krüger et al.,

1993, 1995; Scherbaum et al., 1997) is applicable to achieve a further SNR improvement, which makes it possible to study low amplitude scattered phases and to image scatterers locations more precisely. A scatterer is a small scale heterogeneity in the wave propagating media, that can cause seismic phases to be converted or reflected, for example the waves which are emitted from the seismic source as S- or P-phases are converted to P- or S-phases, respectively when they hit a scatterer. Given a well-located catalog, which can be available by advanced relocation procedures, the main challenge to make use of a source array is to eliminate the effect of waveforms dissimilarity. If events are big enough to calculate the related focal mechanisms and source time functions, one can deconvolve source effects from the signals. However, this is often not possible for smaller events. Our approach to solve this problem is to systematically explore the available data to find a collection of sources with similar waveforms, so that we can interpret them as a source array. As the array data are processed together, a single non-fitting event can destroy the coherency of particular phase beam trace. Finding a collection of similar events, is feasible when a large number of well-located events in a relatively small volume exist in the catalog. In analogy to receiver array the distribution of array elements is imperative aside from the coherent signals to get the precise slowness vector components (Schweitzer et al., 2012; Rost and Thomas, 2009). To fulfill all these criteria, our solution is to employ the optimization technique, while all requisite conditions are formulated to define a single objective function.

Three cases are covered in this thesis: (1) a 2-D receiver array optimization, (2) a 3-D source array optimization, and (3) an array application to monitor microseismic data where the effect of different types of noise are evaluated.
In chapter 2 of this thesis, I addressed the problem of seismic receiver array optimal design. The research questions answered in that chapter are: "How to define the geometry of an array to monitor a specific target with highest possible quality, knowing the source properties, SNR and other conditions and constraints related to the deployment site?" and "How to quantitatively integrate all available information into an applicable array design criterion?"

Such questions are important to answer, for instant when a small aperture array is planed to monitor weak events in local distances, and the accessible places to deploy stations are limited and regular geometries are not practical.

In chapter 3, details of my study about source array optimal configuration is presented. The research questions covered in that study are "How a well-located (swarm) catalog can be used to employ source array processing methods?" and "How an optimal source array can be used to localize the scatterers in the velocity structure?"

In chapter 4, examples of small aperture high frequency arrays are discussed. The example arrays were operating in combination with shallow borehole sensors to monitor induced seismicity during industrial operations at two different sites, (1) accompanying a fracking experiment in sedimentary shale at 4 km depth, (in Wysin Poland) and (2) above a gas field under depletion, (Groningen gas field The Netherlands). Arrays recordings are compared with recordings available from shallow borehole sensors and examples of detection and location performance are given. The array characteristics and transfer (response) functions are discussed in the context of micro-earthquake detection at depth. Concluding, recommendations on the design of microseismic monitoring networks involving seismological surface arrays are provided.

Details on more specific contributions of this dissertation conclude this chapter.

1.1 Technical contributions and innovations

The innovations and technical contributions in this thesis are listed and briefly introduced:

- Addressing the array design task as an optimization problem, which allows to include the design criteria as mathematically formulated objective functions, that should be eval-

uated during the optimization process. Using multi objective optimization method for source and receiver array design allows the user to include more than one criteria in the procedure, based on customized preferences. Then using techniques such as weighted sum method, the objective functions are combined to one single scalar function. (Chapter 2 and 3)

- Using synthetic waveform data and scenario modelling in optimal array design framework that make it possible to integrate seismic source effects such as magnitude, focal mechanisms, radiation pattern, noise condition and other information about deployment site into the design criteria. (chapters 2 and 3)

- Objective function definitions:
  In the example application of receiver array optimization, two sub-functions are introduced to verify the accuracy of horizontal slowness estimation; (1) to suppress aliasing effects due to possible secondary lobes of synthetic array beamforming calculated in horizontal slowness space, and (2) to reduce the event's mislocation caused by miscalculation of the horizontal slowness vector. (chapter 2)
  In the example application of source array optimization four quantitative criteria as sub-functions are formulated and a the weighted sum technique is used to combine them in one single scalar function. The criteria are: (1) to control the accuracy of the slowness vector estimation using time domain beamforming method, (2) to measure the waveform coherency of the array elements, (3) to select events with lower location error and (4) to select events with high energy of specific phases, i.g, sp- or ps-phases. (chapter 3)

- Optimization algorithm: The simulated annealing optimization procedure is modified so that it could converge to a minimum value in less number of evaluations and decrease the calculation time. The modified version benefits from a sampling technique based on the neighbourhood concept (Sambridge, 1999; Wathelet, 2008) and objective function approximation using Voronoi cells (Okabe et al., 1992). (chapters 2 and 3)

- Modelling 3-D seismic noise with spatial correlation: Localized noise sources are modelled by assuming a cluster of point sources and explosion sources with random mechanisms, origin times and locations at the surface of the search area. Accordingly, a stochastic 3-component noise trace is produced that can be tuned to have the same PSD level as that calculated from real samples. (chapter 2)

- Developing computer program for determination of scatterers in the velocity model. Optimized source array elements are selected form a given catalog. The related codes in Python are provided to process both real and synthetic waveforms. Study the location of velocity interface. Algorithms are provided to search for the location of reflector using converted phases sp and optimized source arrays, both using synthetic and real data. (chapter 3)

- Array geometry assessment. Event location and detection using array beamforming. (chapter 4)

1.2 Author's publications

The chapters 2, 3 and Chapter 4 of this thesis are each a peer-reviewed article in a scientific journal. These articles and the author's contributions therein are listed below:

- Chapter 2:

  Karamzadeh, N., Heimann, S., Dahm, T. and Krüger, F., (2019). Application based seismological array design by seismicity scenario modelling. Geophysical Journal International, 216(3), pp.1711-1727; doi:10.1093/gji/ggy523.

  - The author, NK, developed and implemented the array design framework in this article and prepared the as associated Python codes and wrote up the manuscript, implementing comments from her Ph.D. advisers and co-authors

- Chapter 3:

  Karamzadeh, N., Heimann, S., Dahm, T. and Krüger, F., (2020) Earthquake source arrays: optimal configuration and applications in crustal structure studies, Geophysical Journal International, 221(1), pp.352-370; doi:10.1093/gji/ggaa002.

  - The author, NK, developed and implemented the source array design framework in this article and prepared the associated Python codes and wrote up the manuscript, implementing comments from her Ph.D. advisers and co-authors

- Chapter 4:

  Karamzadeh, N., Kühn, D., Kriegerowski, M., López-Comino, J.A., Cesca, S. and Dahm, T., (2019). Small-aperture array as a tool to monitor fluid injection-and extraction-induced microseismicity: applications and recommendations. Acta Geophysica, 67(1), pp.311-326; doi:0.1007/s11600-018-0231-1.

  - The author, NK, did the array geometry assessment and signal spectral analysis and comparisons as well as array based event location in Groningen data, and she wrote up some part of the manuscript, and merged the parts provided by other co-authors, implementing comments from her Ph.D. adviser and co-authors

Chapter 2

Application Based Seismological Array Design by Seismicity Scenario Modelling

This chapter is published in Karamzadeh et al. (2019a).

Abstract

The design of an array configuration is an important task in array seismology during experiment planning. Often the array response function (ARF), which depends on the relative position of array stations and frequency content of the incoming signals, is used as the array design criterion. In practice, additional constraints and parameters have to be taken into account, e.g. land ownership, site-specific noise levels or characteristics of the seismic sources under investigation. In this study a flexible array design framework is introduced which implements a customizable scenario modelling and optimization scheme by making use of synthetic seismograms. Using synthetic seismograms to evaluate array performance makes it possible to consider additional constraints. We suggest to use synthetic array beamforming as an array design criteria instead of ARF. The objective function of the optimization scheme is defined according to the monitoring goals, and may consist of a number of sub-functions. The array design framework is exemplified by designing a 7-station small scale array to monitor earthquake swarm activity in Northwest Bohemia/Vogtland in central Europe. Two sub-functions are introduced to verify the accuracy of horizontal slowness estimation; one to suppress aliasing effects due to possible secondary lobes of synthetic array beamforming calculated in horizontal slowness space, and the other to reduce the event's mislocation caused by miscalculation of the horizontal slowness vector. Subsequently, a weighting technique is applied to combine the sub-functions into one single scalar objective function to use in the optimization process.

2.1 Introduction

A seismic array consists of a number of sensors deployed in a special geometry with common precise timing, acquisition parameters and instrument type. Arrays have a variety of applications, particularly when the signal enhancement is a prerequisite to detect seismic events, and in situations where installing and maintaining sparse networks around the source is impractical. Study and monitoring of nuclear and chemical explosions (Baumgardt and Der, 1998; Kim et al., 1998), volcanic and non-volcanic tremors (La Rocca et al., 2008; Saccorotti and Del Pezzo, 2000; Ghosh et al., 2009), earthquake swarms (Hiemer et al., 2012) and real time monitoring and early warning systems of active faults (Meng et al., 2014) are examples of array applications.

A remarkable utility of an array is to estimate the horizontal slowness vector of the incoming

wave, which yields to phase identification and further seismological findings, such as event location and rupture front tracking (Krüger and Ohrnberger, 2005; Ishii, 2011). The accuracy of the array processing results relies on the coherency of the signal across the array. Coherency distortion may result from factors such as strong near surface heterogeneity, attenuation, strong local noise sources and radiation pattern of seismic waves. Near surface heterogeneities cause violation of the plane wave approximation that increases uncertainties of the estimated horizontal slownesses.

There are other sorts of uncertainties and aliasing in slowness vector estimation which are related to the array geometry. To eliminate them, some considerations in array configuration are commonly applied. For instances, each apparent wavelength of the signal under the study, λ_{app}, should be sampled in at least two discrete sampling locations so that the minimum interstation distance should be at least $\lambda_{app}/2$. In addition, to increase the resolution for the small wavenumbers, array aperture should be comparable with the highest apparent wavelength of the incoming signals.

Array geometry design has been systematically studied for seismic surface wave studies based on the Array Response Function (ARF) characteristics (Marano et al., 2014; Wathelet et al., 2008). Kennett et al. (2015) suggested the use of spiral-arm shaped arrays, which allows deployment of relatively large aperture arrays with a limited number of stations to achieve a sharp central peak in the ARF and consequently a high slowness resolution. Haubrich (1968) investigated a special class of linear and planar arrays involving uniform patterns of stations, proposed a design method based on the coarray's properties and showed how the beam pattern can be optimized for a given number of sensors. However, idealized array configurations are difficult to realize in practice, especially on small scales, due to geographical and geological limitations.

The ARF depends on the relative position of array stations and the frequency content of the interesting signals, and for a specific frequency ω, it is defined as:

$$\left| \frac{1}{N} \Sigma_{j=1}^{N} e^{2\pi i \omega (\mathbf{S}-\mathbf{S_0}).\mathbf{r}_j} \right|^2, \tag{2.1}$$

where $\mathbf{S} = (S_x, S_y)$ is the horizontal slowness vector, S_0 is the correct horizontal slowness vector of the incoming waves, $\mathbf{r}_j = (\delta x_j, \delta y_j)$ is the horizontal location vector of the jth station relative to the array reference point and N is the number of stations (Schweitzer et al., 2012; Rost and Thomas, 2002). In practice, ARF is evaluated for a monochromatic wave with a fixed frequency, e.g. 1 Hz. Assuming the power spectral density of the analysed waves is normalized to 1, then the ARF is considered to be equivalent of the total energy recorded by the array (Rost and Thomas, 2002).

In many applications, information about deployment site and the experiment, such as site-specific noise level, properties of the target events, land-use classes (Büttner et al., 2004) and other constraints on the logistics are available. Such information needs to be integrated quantitatively to be used in the array geometry design and not just collected and qualitatively compared. For instance, to ensure high signal coherence and measure loss of signal across the array stations, it is advantageous to take focal properties of the target events and the site-specific noise level into consideration. However, considering such information is not standard, a quantitative integration method is required.

In this study a flexible array design framework is introduced in order to perform an automatic search for an optimized array geometry considering additional information and constraints. This framework operates by modelling realistic conditions and scenarios and making use of synthetic seismograms to evaluate array beamforming as the design criteria instead of using the ARF (equation 4.1). While the ARF is unique for a given array geometry regardless of seismological background and deployment site conditions, array synthetic beamforming is

customizable for individual experiments. It is not straightforward to include station elevation differences in the ARF, but this can be achieved effortlessly using synthetic array beamforming. In addition, the plane wave approximation is an implicit assumption while computing the ARF, whereas the time correction due to deviation from the plane wave approximation at short source-receiver distances is easy to handle using realistic synthetic seismograms. Furthermore, the approach also allows to consider the joined effect of the newly created array and pre-existing network stations or arrays. One or more objective functions are defined depending on the purpose of the survey and specific boundary constraints. For instance, objective functions can ensure the precise slowness vector estimation of the earthquakes for a site-specific signal-to-noise ratio (SNR).
To describe the framework in detail, and to show how it works, we customized it with specified realistic constraints. The problem is to design a 7-station small scale array to monitor earthquake swarm activity in the region of NW Bohemia/Vogtland. According to the survey goal, which is to obtain precise event locations derived by array method, two objective functions are defined. Then the weighting sum method is used to combine them in one scalar function. Both functions are used to control the accuracy of horizontal slowness vectors of P- and S-phases estimated using the time domain beamforming of (noisy) seismograms from the hypothetic events in the target source region. One objective is to minimize the array beam power in a specified slowness range. The other one is to minimize the possible miscalculation of the main peak in horizontal slowness space. Details of the objective functions and the weighting technique are described in the section 2.

2.2 Theoretical approach

We define a model as a specific geometrical configuration of the array for a given number of stations and specified deployment site, under optional boundary conditions as pre-existing stations. The model space covers all possible combinations of station positions. The aim is to find the best geometry, i.e. a point in model space, at the minimum of a well-defined scalar objective function.
A key point of our approach is to solve the problem with the help of synthetic, full waveform seismograms. Although this is computationally demanding, it has the big advantage of flexibility when considering realistic source and site configurations and different types of noise. Additionally, user-specific processing schemes can be easily implemented and tested. We overcome the technical challenge of quickly calculating seismograms for many station-source configurations by means of pre-calculated Green's functions databases and efficient storage and accessing tools (Heimann, 2017).
To generate synthetic seismograms in the context of our scenario testing procedure, several pieces of information are needed: a) A seismic source model including magnitude range, seismogenic zone and source mechanisms; b) Propagation velocity and attenuation model (if it is available), ideally with error bounds; c) Site specific characteristics such as noise level and amplification factors. Given a specific model has been generated during one iteration of the optimization process, the corresponding synthetic seismograms are re-calculated for a given ensemble of seismic sources, and the scalar value of the objective function is calculated considering the complete ensemble of sources.
The random models are in principle drawn from a uniform distribution of virtual station points within the given boundaries of the deployment site. However, in order to consider constraints for the selection of sites, we work with 2D probability density functions for selecting new models during each iteration. The probability density function may consider surface geology, boundaries of cities or countries, land use, accessibility as well as the natural and man made sources of noise. For example, to avoid noisy station locations, land-use classes defined by the EU-project

CORINE (Büttner et al., 2004) and open GIS data can be used to define such a 2D probability density function. In the following subsections we describe the selection of objective functions and the implemented search algorithm.

2.2.1 Objective functions and their scaling

As an example of application, we want to design a small aperture array for a given number of stations to detect, absolutely locate and map the migration pattern of clustered, weak events of earthquake swarms or aftershock activity. Therefore, the aim is to ensure a high resolution of the horizontal slowness vectors estimation of P- and S-phases for individual events occurring within a confined source volume with given, pre-dominant mechanism and magnitudes. In this example, we define two objective functions using time domain beamforming of P- and S-phases in the (noisy) seismograms from hypothetic events in the target source region to verify the accuracy of horizontal slowness estimation, without using the idealized concept of ARFs. We measure the total amount of relative power of array beam in slowness space, for a given source and array geometry, that ideally should contain a concentrated main peak and no other localized secondary peak. The array beam trace is calculated as the sum of all recorded, time shifted traces:

$$B(t) = \frac{1}{n}\Sigma_{j=1}^{n} \; Y_j(t + dT_j) \qquad \text{with} \qquad dT_j = S_x\delta x_j + S_y\delta y_j. \tag{2.2}$$

where, Y_j is the normalized trace recorded at the station j and n is the number of stations. Assuming the plane wave approximation is valid, the time shift, dT_j, for the station j, depends on the horizontal slowness components of the incoming wavefront, S_x and S_y, and the relative distance to the array reference point, δx_j and δy_j. We assumed that there is not any overlap of phases in the processing time window, so that using the normalized traces eliminates the effect of radiation pattern on amplitudes. A common strategy to find the correct value of time shifts is described in e.g. Schweitzer et al. (2012) and Rost and Thomas (2002). $B(t)$ is calculated for an appropriate time window of P- and S-phases.
The first objective function, f_1, is defined by assuming that the average relative beam power of the P- and S-phases in the slowness space is minimal. For P-phase that is calculated by:

$$f_p = \frac{1}{n_x n_y}\Sigma_k^{n_y}\Sigma_l^{n_x} A^2_{S_{x_l},S_{y_k}}, \tag{2.3}$$

where, n_x and n_y are the number of grid points in the predefined slowness ranges S_x and S_y, index p of f_p indicates the P-phase and A^2 is the value of beam power which is normalized to the global maximum of all tested grid points, so it is called relative beam power. In the same way the average relative beam power for the S-phase is written as f_s, and finally both phases are combined in a single formula:

$$f_b = \sqrt{{f_p}^2 + {f_s}^2}. \tag{2.4}$$

f_b is calculated using synthetic waveforms of a given specific source from the ensemble of target sources and can also be calculated from the whole ensemble to include a realistic model of seismicity, leading to f_1 defined by:

$$f_1 = \frac{1}{n}\Sigma^n f_b, \tag{2.5}$$

where, n is the number of events. A larger value of f_1, indicates higher value of average beam power in the slowness space for the whole set of earthquakes, which can happen either because of high or numerous side lobes or a wide main lobe in the slowness map, for a number of

individual earthquakes, both of which are undesirable.
The second objective function, f_2, is defined to minimize the possible miscalculation of the horizontal slowness vector, i.e. the main peak in the horizontal slowness space. Such a misplaced beam power peak in the slowness space can occur if for instance the wavefront is affected by a nodal plane of the source radiation pattern, if the waveform becomes non-coherent close to the nodal plane, and if the plane wave approximation is not valid. Other causes may be related to the heterogeneous distribution of ambient noise or noise sources across the array, or if local site effects and amplification factors distort the waveform coherency. Nevertheless, modelling 3D propagation effects such as lateral heterogeneities is hard to realise before the experiment, and in practice, the measurement bias of the horizontal slowness vector due to such incoherent sources is usually corrected by employing station azimuthal correction terms during the operation of an array (Bondár et al., 1999). In our modelling, we try to minimize errors caused by source radiation pattern and local site conditions such as the noise level.

A miscalculation of slowness vector would lead to an error in the location of the earthquake source, i.e. a deviation between calculated and true event location. We prefer to include mislocation in the objective function rather than the miscalculation of slowness, because mislocation is a quantity that can be shared between array and network processing if network stations are included. We implement a location procedure which relies on ray tracing of P- and S-phases given the measured slowness values and the velocity model of the propagation media. We use a common layered model in our example, but it would be possible to consider local site effect models for every station. A unique ray geometry, $R(X, Z)$, and travel time function of ray position, $X(t)$ and $Z(t)$, are determined for both the P- and S-phases. $X(t)$ and $Z(t)$ are time sequences of distance from the source and depth, respectively. Since the aim is to calculate the location error caused by miscalculation of slowness, we employed theoretical phase arrival times using the velocity model and source location without assuming any uncertainty. Epicentral distance and depth are then estimated from the crossing point of P- and S-phase ray trajectory, given the arrival times. First, using arrival time differences, epicentral distance is calculated. Then, corresponding depth, z is derived from the average value of Z(t) for P- and S-phases. Together with the backazimuth angle estimated from the horizontal slowness vector of the P-phase, the latitude and longitude of the epicentre are determined. Horizontal mislocation, δ_h, and depth error, δ_d, are obtained by comparing calculated and true location. The total mislocation is defined by:

$$f_l = \sqrt{\delta_h^2 + \delta_d^2} \tag{2.6}$$

After calculating f_l for n events of ensemble of target events, the average value of total mislocations is defined as the second objective function, f_2, by:

$$f_2 = \frac{1}{n}\Sigma^n f_l \tag{2.7}$$

Having defined two objective functions, we use a weighting technique to combine them into one scalar function. In general one can customize the problem by defining any number of objective functions, so we describe the weighting technique in a general form.
A multi-objective optimization problem is defined in mathematical terms as:

$$min\ [f_n(\mathbf{m})]; \quad \mathbf{m} \in S \tag{2.8}$$

where $f_n(\mathbf{m})$ is a set of objective functions with n member, $\mathbf{m}$ is a vector of design variables, and S is a feasible search region defined by a set of constraints. Single objective optimization is formulated by $n = 1$. Without loss of generality, the number of objective functions in our problem is $n = 2$ and f_1 and f_2 are the two independent objective functions for two criteria

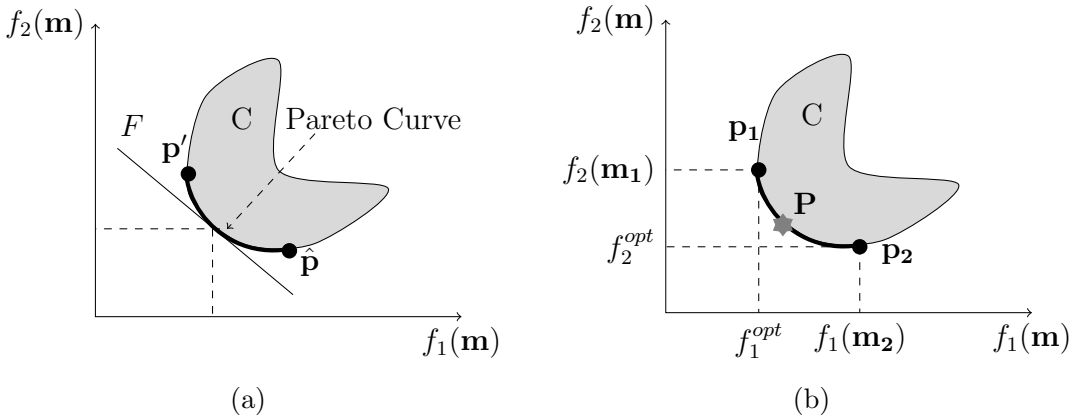


Figure 2.1: (a) Illustration of the Pareto curve for a two dimensional objective function space. All points between $\mathbf{p}'$ and $\hat{\mathbf{p}}$ define the Pareto curve, the line F is a tangent to the curve at an arbitrary point and the region C is the feasible objective space. (b) Geometrical representation of the scalarization approach used to approximately define F as a weighted sum of the individual objective functions f_1 and f_2. Points $\mathbf{p_1}$ and $\mathbf{p_2}$ on the assumed Pareto curve are approximated by optimization of the single objective functions f_1 and f_2, respectively and are used to define the weighting factor. The point $\mathbf{P}$ indicates a touching point on the Pareto curve that can be reached using the scalar objective function defined by the line F.

defined in equations (2.5) and (2.7).

Contrary to a single objective problem, a solution to a multi objective problem is not a single global solution, and it is required to specify a set of solutions that all satisfy a predetermined definition of optimum solution (Marler and Arora, 2004). "Optimal" solution for the non-unique multi-objective optimization is defined using the concept of the "Pareto optimality" (Marler and Arora, 2004; Laponce, 1972). A vector $\mathbf{m}^* \in S$ is named Pareto optimal if all other vectors $\mathbf{m} \in S$ have higher values for at least one objective function, or have the same value for all the objective functions (Laponce, 1972; Marler and Arora, 2004; Caramia and Dell' Olmo, 2008). Let's assume a region, C in Fig. 2.1, in the two dimensional objective function space, $f_1 - f_2$, is covered by feasible values of functions f_1 and f_2 for all possible models $\mathbf{m} \in S$. All points between $\mathbf{p}'$ and $\hat{\mathbf{p}}$, indicated by the thick line on Fig. 2.1a., define the Pareto curve or Pareto front which is the image of all the Pareto optimal solutions in the objective function space. The shape of the Pareto curve indicates the nature of the trade-off between the different objective functions, and there is mathematically no better solution for the optimization problem than any point on the Pareto curve. Finding the Pareto curve from random sampling in the model space is computationally often not possible, so approximation methods are frequently used (Caramia and Dell' Olmo, 2008). In our example of a two-dimensional objective space the final objective function can be written as a line (dimensionless) tangent to the Pareto curve:

$$F = \gamma f_1(\mathbf{m}) + (1 - \gamma) f_2(\mathbf{m}), \tag{2.9}$$

where γ $(0 =< \gamma <= 1)$ is a weighting or normalization factor between f_1 and f_2. Obviously, changing the weight γ leads to possibly different touching points of F (Fig. 2.1a). There is not a deterministic correspondence between the weighting factor and the solution of the optimization, so the user should determine the appropriate weight. However, the weighting factor does not necessarily correspond directly to the relative importance of the objective functions (Caramia and Dell' Olmo, 2008). Equation (2.9) can be re-written as:

$$f_2(\mathbf{m}) = \frac{-\gamma f_1(\mathbf{m})}{1 - \gamma} + \frac{F}{1 - \gamma} \tag{2.10}$$

So, the minimization in the scalarization approach can be interpreted as the attempt to find the

touching point for a tangent with slope $-\frac{\gamma}{1-\gamma}$ (Fig. 2.1a). By determining γ, F is formulated as a joint scalar objective function, which can lead the optimization process to one Pareto optimal solution, e.g. point P on Fig. 2.1b. Our strategy to estimate γ is to use points $\mathbf{p}_1$ and $\mathbf{p}_2$ (Fig. 2.1b) to estimate the slope in equation (2.10). We assume that $\mathbf{p}_1$ and $\mathbf{p}_2$ represent the optimized solution of individual objective functions f_1 and f_2 respectively. Accordingly, the optimized models, $\mathbf{m}_1$ and $\mathbf{m}_2$, and related minimum values, $f_1(\mathbf{m}_1)$ and $f_2(\mathbf{m}_2)$ are obtained through the optimization process. The associated values of $f_2(\mathbf{m}_1)$ and $f_1(\mathbf{m}_2)$ are obtained by simple forward simulation. Finally, the slope of an assumed passing line from $\mathbf{m}_1$ and $\mathbf{m}_2$ is calculated and the value of γ is estimated from $-\frac{\gamma}{1-\gamma}$. Knowing γ the final solution can be obtained by minimizing F for the joint problem.
In general, the procedure is extendible for any number of objective functions. For instance if $n = 3$, two normalization factors γ_1 and γ_2 are used to modify equation (2.9).

2.2.2 Optimization method

The aim is to search for an array geometry that, given a set of synthetic events, gives the best array performance according to the scalar objective function described in section 2.1. For a similar problem of optimized network design, Hardt and Scherbaum (1994) applied a simulated annealing (SA) technique. The SA algorithm operates by sequence of steps, controlled by a predefined cooling scheme, and starts with an initial model estimation and the objective function evaluation. The next steps are new model generation by perturbation of a given model and making decision about the acceptance or rejection of that model based on specified criteria (Sen and Stoffa, 2013). The main drawback of SA technique is the large number of function evaluations needed to generate converged statistics. For instance, Hardt and Scherbaum (1994), suggested a cooling scheme based on which 14×10^5 function evaluations are needed to find an optimum configuration for a 7-station network. The objective function calculation in our approach is computationally expensive as it takes about 2.5 minutes for one function evaluation using a 8-core processor. Thus we tried to modify the SA procedure so that it converges to a minimum value in less number of evaluations. Accordingly, we benefit from a sampling technique based on the neighbourhood concept (Sambridge, 1999; Wathelet, 2008) and objective function approximation using Voronoi cells (Okabe et al., 1992).
A model is determined by the position vectors of the array, while the array stations are ordered based on the distance from the lower left corner of the deployment site. The outline of the algorithm is shown in the flowchart in Fig. 2.2 and works according to the following steps:
I: A number of initial models, n_{start}, with known number of stations, $n_{station}$, are randomly drawn from a given distribution defined based on the search area and possible one or more forbidden zones, so the model population, M, is generated. II: For every element of the model population, $\mathbf{m_j}$, a forward calculation is done, so the objective function population, $F(\mathbf{m})$, is created. III: In every iteration, a number of n_{select} models with lower value of $F(\mathbf{m})$ are chosen from the model population as guiding models. IV: New models are generated by random perturbation of the guiding models so an ensemble of candidate models, $n_{candidate}$ models, is created. The radius of random perturbation is reduced by increasing the iteration steps using a cooling scheme. V: The neighbourhood check is done to determine if the candidate models fall into the same Voronoi cell as the corresponding guiding model or not. $n_{new}(< n_{candidate})$ models are accepted from each ensemble and are added to the model population. If in the first try n_{new} models are not found for a guiding model, this step is repeated several times until n_{new} models are found, but the number of tries are limited to n_{try}.
In every iteration, new accepted models are added to the M and $F(\mathbf{m})$ is updated. So the iteration is done over steps II to V. After the last iteration, the model with lowest $F(\mathbf{m})$ is reported as the final model.

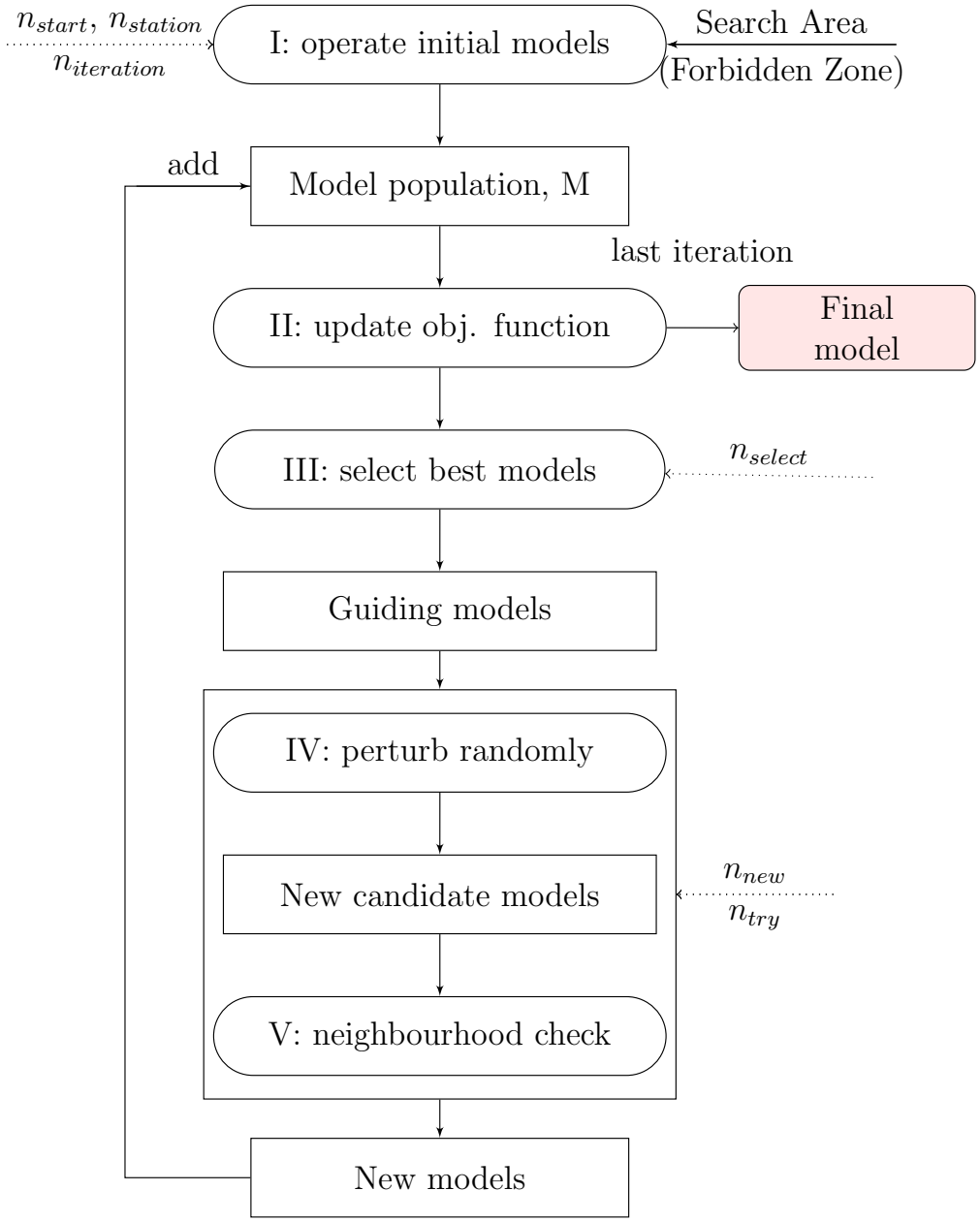


Figure 2.2: Optimization algorithm flowchart: The initial "Model population, M" is created by drawing models from a given distribution, based on geometrical settings, the desired number of stations ($n_{station}$), and the number of initial models (n_{start}). $n_{iteration}$ is the number of iterations to create new models and to update M and the objective function population. n_{select} is the number of selected models from M with lowest value of the objective function used as "Guiding models" to generate "New candidate models" by random perturbations. n_{new} is the number of new models to be accepted in the neighbourhood of each of the n_{select} guiding models, n_{try} is the number of tries to find n_{new} accepted models to be added as "New models" to M. After the last iteration, "Final model" is the model in M yielding the lowest value in the objective function.

2.3 Application to design of a monitoring array for midcrustal earthquake swarms

2.3.1 Model setup

The target area, Northwest Bohemia/Vogtland, is one of the most active intraplate earthquake swarm regions in Europe. It is situated in the border region between Germany and the Czech Republic. In 2010 for 6 months a seismic array of 7 stations was installed, by University of Potsdam, near the German-Czech border to monitor swarm activities. Position of the array is shown by a black triangle in Fig. 2.3a, and array geometry and ARF are depicted in Fig. 2.3b and Fig. 2.3c, respectively.
Earthquake swarms encompass an immense number of weak events occurring in a spatio-temporal cluster over a period of weeks or months, and are not associated with typical mainshock-aftershock sequences of an earthquake. They are considered to result either from magmatic activity (Dahm et al., 2008; Morita et al., 2006), from fluid-migration (Hensch et al., 2008; Hainzl et al., 2016), or from aseismic creep on faults (Neunhöfer and Hemmann, 2005; Passarelli et al., 2015). Seismic swarms in Northwest Bohemia/Vogtland often comprise more than $10^3 - 10^4$ individual events with magnitudes between M1 and M4 (Ružek and Horálek, 2013). Typically, the events are clustered in several focal zones (Novy Kostel, Kraslice-Klingenthal, Plesna, etc.) and the hypocentral depths of events in the whole region vary between 5 and 20 km (Ružek and Horálek, 2013). However, during the last 25 years, most of the seismic activity took place near the village of Novy Kostel, where the hypocentres are located in a depth range between 7 to 10 km (Ružek and Horálek, 2013). The Novy Kostel zone shows a fault plane at depths between 6 and 11 km oriented nearly $S - N$ (strike 169°) and steeply dipping (dip 80°) westward. The epicenters of earthquake swarms that occurred between 1991-2001 and in 2008, are distributed along the 12.5 km long section of the fault line (Fischer and Horálek, 2003; Fischer et al., 2010).

According to Horálek and Šílený (2013), the focal mechanisms of the swarm activity which occurred close to Novy Kostel in 2000 are mostly oblique-normal and oblique-thrust but the oblique-normal faulting predominates. The mechanisms of the individual types are fairly consistent, where the oblique-normal events have predominant strikes of 160° to 170°, dips of 72° to 80° and rakes of −28° to −38°, whereas the oblique- thrust events show mainly strikes of 355° to 360°, dips of 80° to 85° and rakes of 35° to 40°. This area has been of a great interest for seismologists and has been studied for different subjects such as location of hypocentres (Fischer et al., 2010; Hiemer et al., 2012), magnitude estimation (Horálek et al., 2009), determination of focal mechanisms (Dahm et al., 2000; Horálek and Šílený, 2013) and determination of optimum velocity models (Málek et al., 2004, 2005). In this region during a swarm event in 2008, a seismic array helped in the detection of small magnitude events (Hiemer et al., 2012) and the region is also selected for developing an interdisciplinary observatory using shallow drilling and small aperture seismic arrays (Larsen, 2012).
We use this region as an example and ask the question of how the design of a temporary array installation could be improved by re-arranging its stations. However, by presenting this example the aim is to show how the array design framework can be customized given a specific problem.

2.3.2 Simulating synthetic waveforms

The essential information needed to set up the array design framework and to simulate synthetic waveforms are summarized in Table 3.1. The velocity model which is used to generate synthetic

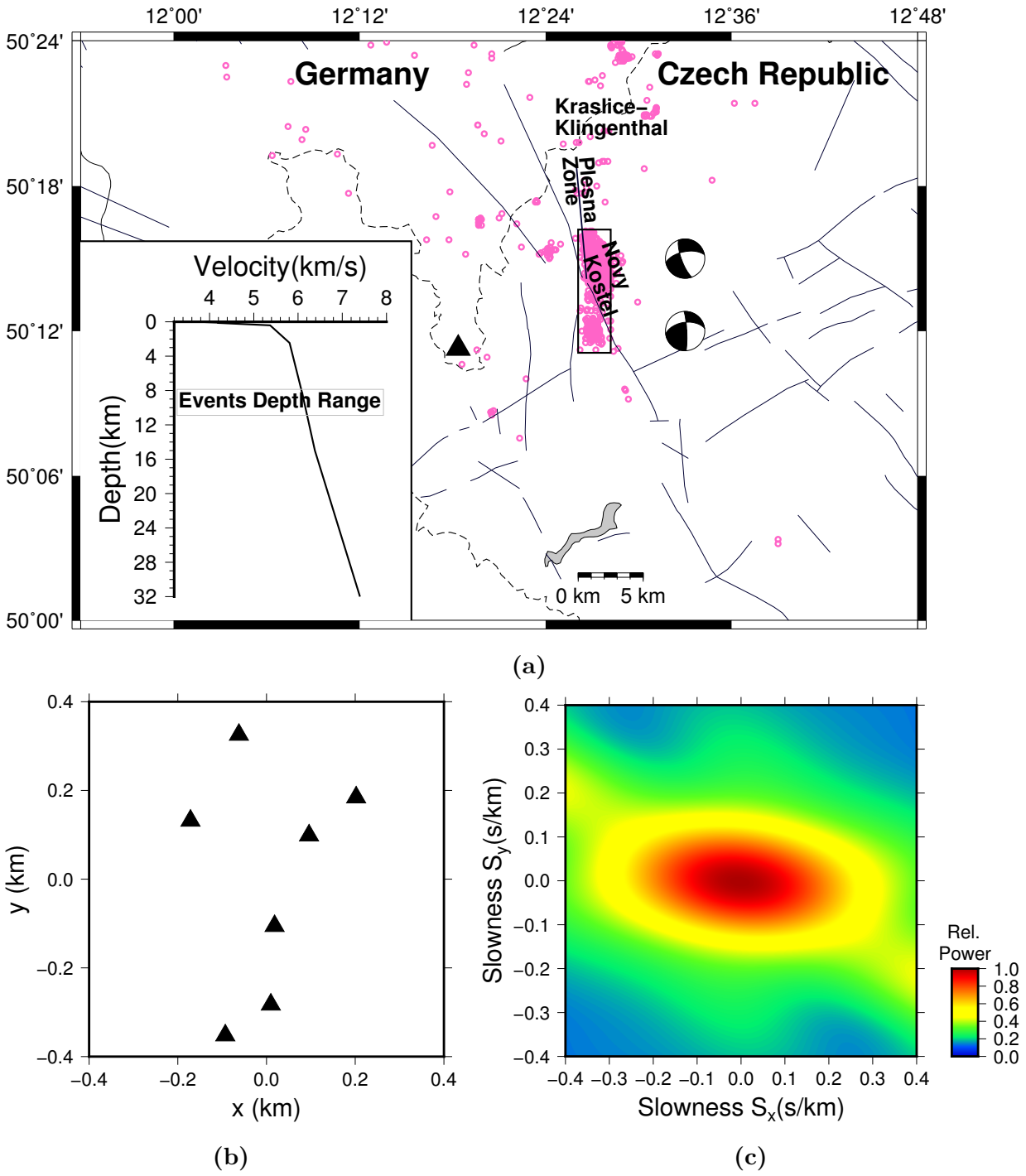


Figure 2.3: (a) Test area, which is an interaplate swarm region, pink circles are epicentres of earthquakes which occurred during 2013, the black triangle represents a temporary small aperture array, the Vogtland array, which operated from 29 August to 2 November 2011. The rectangle shows the Novy Kostle swarm zone, which we used as target area. Two predominant focal mechanisms are shown. The inlay depicts the velocity model used to generate synthetic seismograms. (b) Geometry of the Vogtland array. (c) Array response function in slowness space.

Table 2.1: Input parameters used in the scenario modelling for swarm monitoring in NW Bohemia/Vogtland.

Parameter	Value
Velocity model	1-D model based on (Málek et al., 2004)
Noise level (PSD)	Calculated from recorded noise samples
Focal mechanisms	1) Oblique-normal; strike: $160° \pm 5°$; dip: $75° \pm 5°$; rake:$-30° \pm 10°$
	2) Oblique-thrust; strike: $357° \pm 5°$; dip: $83° \pm 5°$; rake: $37° \pm 5°$
Fault plane	Novey Kastel fault geometry; strike: 169°, dip: 80°
Hypocentral distribution	Fault plane $\pm$ 0.1 km
Sampling rate	originally 100 sps, upsampled to 400
Magnitude	M_w : 1
Frequency band	2-8 Hz
Slowness range	$\pm 0.3 s/km$ for both P- and S-phase
Number of slowness grids	200
Number of simulated events	100

signals as well as to predict phase arrival times is based on Málek et al. (2004) (see Fig. 2.3a).

Hundred simulated events are assumed to be distributed uniformly within 100 m bands around the main fault plane near Novy Kostel. Consequently, the depth of events are between 8 and 11 km. The source mechanism of each particular event is assigned randomly from valid ranges of strike, rake and dip values of oblique-normal and oblique-thrust type events (50 mechanisms of each type, see Table 3.1 for valid ranges). We decided to use magnitude $M_w 1$ for all of the simulated events, so the variation of SNR in the analysed signals results from the change in noise amplitude and signal amplitude due to the radiation pattern.

To simulate realistic background noise traces, noise power spectral density (PSD) is calculated using real samples recorded by the array that was operated in the same zone (Fig. 2.3a). In reality, noise PSD can vary depending on the measurement location even in 1 km distances, due to localized noise sources. Localized noise sources are modelled by assuming a cluster of point sources and explosion sources with random mechanisms, origin times and locations at the surface of the search area. Accordingly, a stochastic 3-component noise trace is produced that can be tuned to have the same PSD level as that calculated from real samples. In this way, synthetic noise samples show realistic variation in PSD in different random stations, while properties of the noise signals such as high degree of coherency at short distances and within three components are simulated realistically. Fig. 2.4 illustrates a distribution of noise amplitude which are simulated by the noise generator used in this study.
Examples of real and synthetic waveforms are shown in Fig. 2.5 where the array geometry was introduced in Fig. 2.3. Fig. 2.6 shows noise signal spatial correlation derived from 10-second duration samples of real and simulated noise. For both cases array station locations are the same. Comparing the overall trends of synthetic and real noise data indicates that the distance dependent coherency of synthetic noise signals is realistic.
Fig. 2.7 shows average radiation pattern of P- and S-phases for simulated events in an area including the search box (black rectangle). The oblique-normal and oblique-thrust events are analysed separately, and the main focal mechanisms are shown in the location of assumed sources. These plots show that within the small search area, signal amplitudes vary slightly. According to Fig. 2.7a and Fig. 2.7c P-phases average amplitudes are in the same range for both type of events, while the amplitude of S-phases are higher for oblique-normal events (Fig. 2.7b) compared to the oblique-trust events (Fig. 2.7c). Using the noise model illustrated in Fig. 2.4, the average signal-to-noise ratio (SNR) for phases recorded at the deployment site is calculated

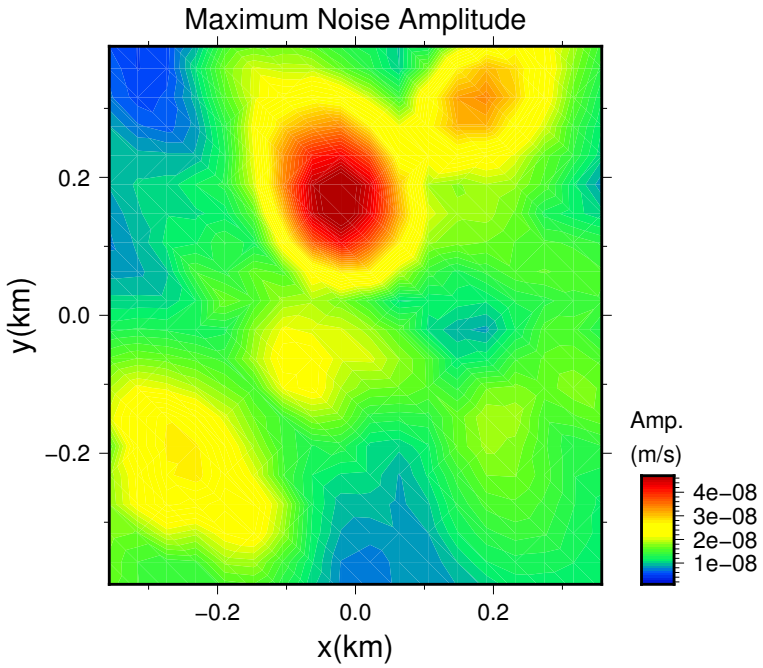


Figure 2.4: Simulated noise signal amplitude by assuming a cluster of point sources and explosion sources at the surface of the deployment site.

Table 2.2: Input parameters used to set up the optimization algorithm.

Parameter	$n_{station}$	$n_{iteration}$	n_{start}	n_{select} [1]	n_{new}	n_{try}
Value	7	12	500	50; 40; 30; 25; 25; 20; 20; 10; 10; 10; 10; 10	15	5

[2] Values of n_{select} are specified for each iteration.

(Fig. 2.8). We express the SNR in decibel as:

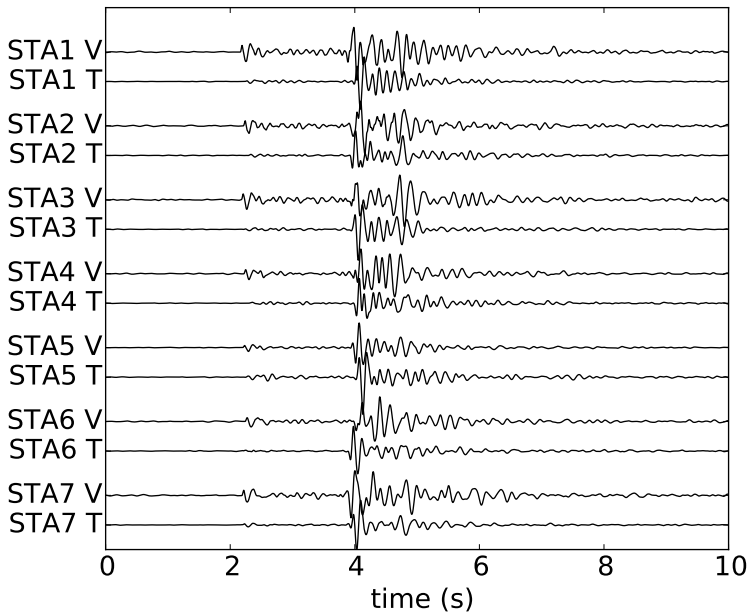
$$SNR = 10 \log_{10} \frac{P_S}{P_N}, \tag{2.11}$$

where P_S is the power of the signal, P- or S-phase wave, and P_N is the power of the noise. P_S and P_N are computed in the time domain as an average power over the duration of 1 second. All synthetic traces are filtered before beamforming using the bandpass Butterworth filter of order 4 and corner frequencies of 2 and 8 Hz. According to the velocity model and the hypocentral distribution of events, realistic slowness ranges for P- and S-phases are about 0.12-0.16 s/km and 0.17-0.24 s/km, respectively. Thus, the wavelength (λ) range of the waves, 0.5-4 km, is much smaller than the distance travelled by the wavefront, $r \approx 13$ km, and the plane wave approximation condition, $r \gg \frac{\lambda}{2\pi}$ (Ben-Menahem and Beydoun, 1985), is valid. The highest value for the slowness is set to 0.3 s/km for both phases, and the number of grid points for each slowness component, i.e. n_x and n_y, is equal to 200. Required parameters to set up the optimization algorithm are summarized in the Table 3.2.

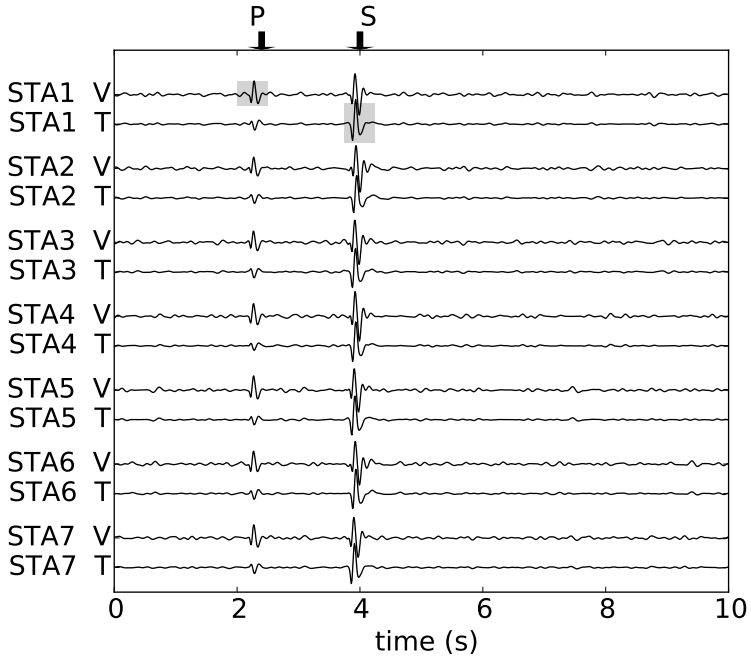
2.3.3 Results and discussions

2.3.3.1 Estimation of weighting factor: γ

To estimate the weighting factor in equation (2.10), according to the procedure described in section 2.1, the optimization program is initiated using each objective function f_1 (equation 2.5) and f_2 (equation 2.7) individually. For each case the same scenario and boundary conditions are applied that are described in detail in section 3.1. The optimized models for each objective



(a) Real waveforms.



(b) Synthetic waveforms.

Figure 2.5: An example of real and synthetic waveforms recorded on vertical (V) and transverse (T) components. (a) Real signals recorded at the 7-station Vogtland array shown in Fig. 2.3b, magnitude of the event is $Ml = 0.9$ and is located in the Novey-Kostel zone. Traces are normalized to one, bandpass filtered 2-8 Hz. (b) Synthetic waveforms after adding realistic noise and bandpass filtered 2-8 Hz. P-phase analysis is done on the phase window on V, to estimate slowness and backazimuth angle, then horizontal components are rotated according to the backazimuth and the T components are computed. The S-phase processing is done on the phase window on T.

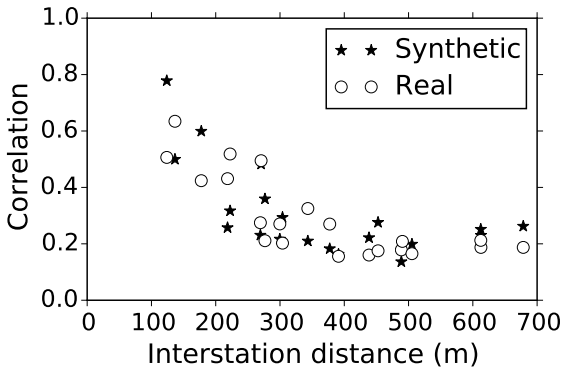


Figure 2.6: Cross-correlation of real (circles) and synthetic (stars) noise samples for the 7-station Vogtland array shown in Fig. 2.3b.

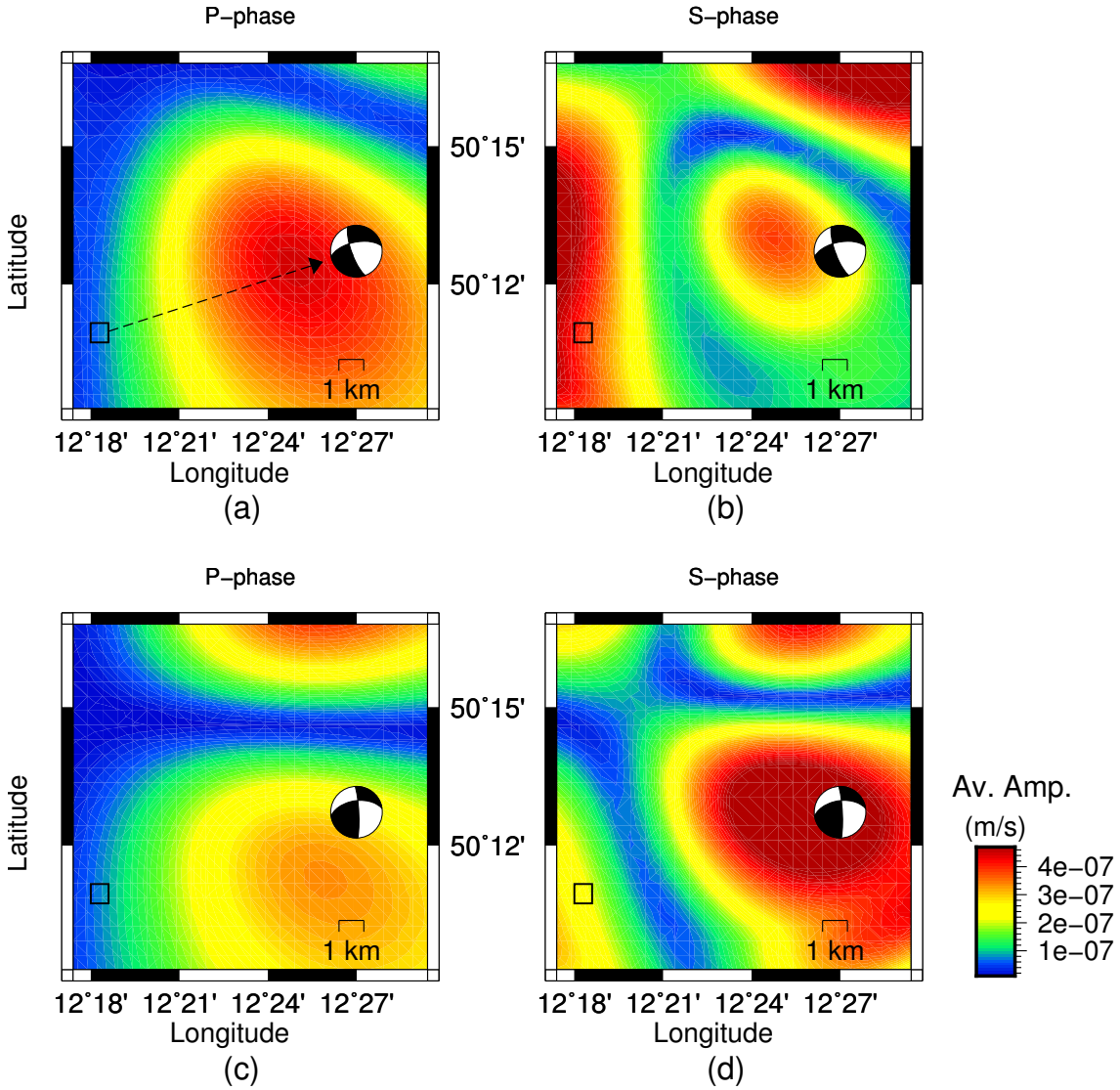


Figure 2.7: Radiation pattern of body waves for two types of events characteristic of the seismicity of the target area. (a) and (b) show the average value of P- and S-phase amplitudes for oblique-normal events and (c) and (d) show the average value of P- and S-phase amplitudes for oblique-thrust events. The small area shown by the rectangle is the area where the array is supposed to be deployed and the arrow on plot (a) indicates the backazimuth direction to the target events.

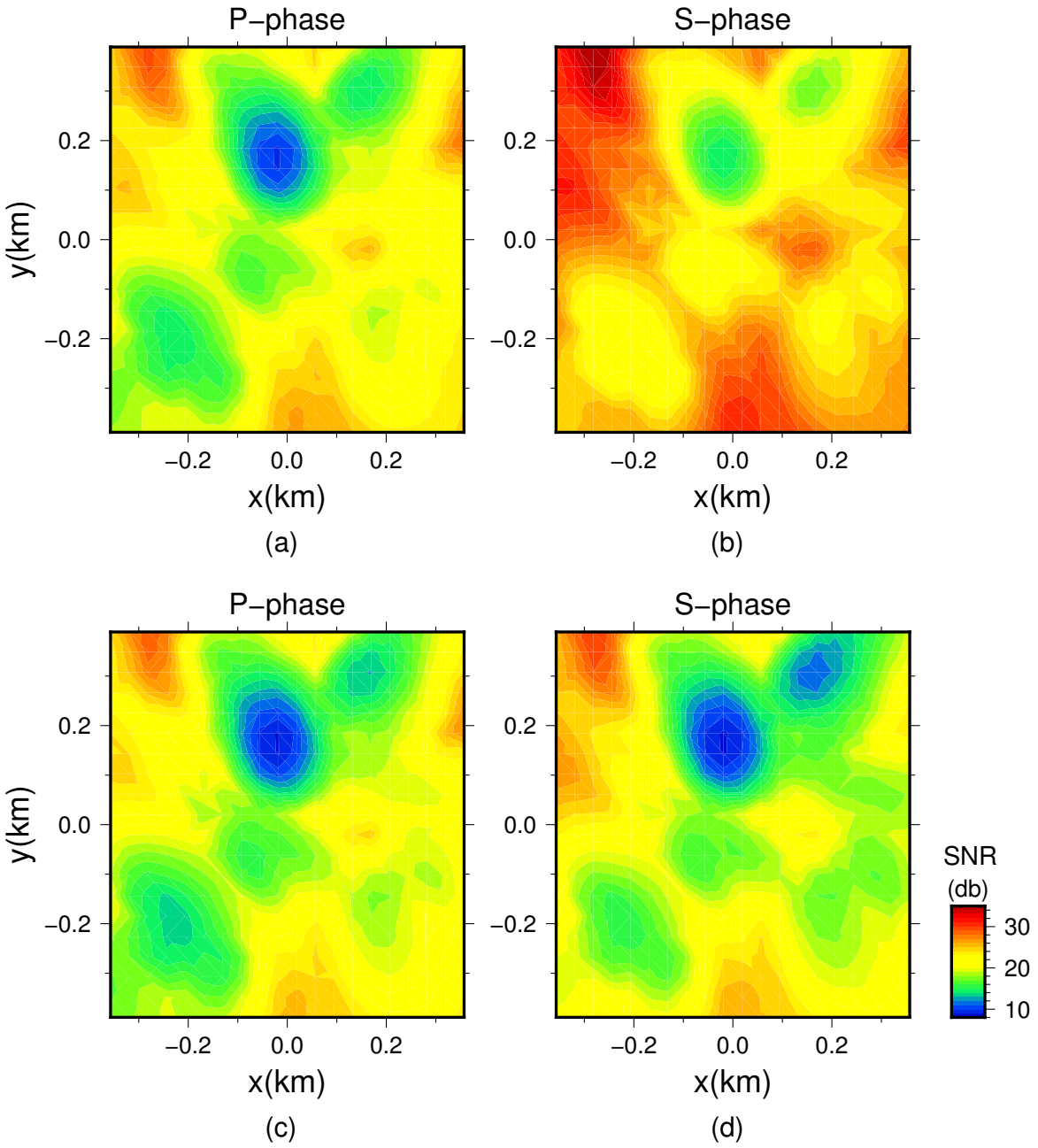


Figure 2.8: Average SNR of P- and S-phases for the synthetic events modelled in the Vogtland area (small rectangle in Fig. 2.7). The noise model is shown in Fig. 2.4 and signal amplitudes are shown in Fig. 2.7. (a) and (b) are related to the oblique-normal events and (c) and (d) are related to the oblique-thrust events.

function are shown in Fig. 2.9a and Fig. 2.9b, where the assumed noise maximum amplitude distribution is depicted in the background. The values of the two objective functions for the geometry shown in Fig. 2.9a are $f_1 = 1.0$ and $f_2 = 2.18$, and for the geometry shown in Fig. 2.9b are $f_1 = 1.26$ and $f_2 = 1.0$, and are marked in Fig. 2.10 by a circle and filled square, respectively. By substituting those values in equation (2.10), γ is calculated equal to 0.88. Accordingly, the final scalar objective function is:

$$F = 0.88 f_1 + 0.12 f_2 \tag{2.12}$$

2.3.3.2 Optimized array

In the ideal case, the final objective function should lead to a solution that minimizes both functions f_1 and f_2 simultaneously. However, in practice there might be an inherent trade off between the two functions, and such an ideal solution might not exist. In general, the actual shape of the Pareto front is unknown in this problem. Using the final objective function (equation 2.12), which is a weighted sum of f_1 and f_2, we try to get as close as possible to an ideal solution.

The minimum value of the final objective function gives the solution which is shown in Fig. 2.9c, and the corresponding values of the individual objective functions $f_1 = 1.05$ and $f_2 = 1.4$ are shown in Fig. 2.10 by a white star. To allow for small variations in each station location to take into account unforeseen restrictions in the field, we suggest to consider the family of best solutions as well. The family of solutions is determined by applying a threshold on the final objective function. For instance, for the solution shown in Fig. 2.9c, the minimum of the objective function augmented by a factor of 0.02 is assumed as threshold value. Accordingly, the spatial distribution of all models whose objective functions fall below the threshold, gives an indication for possible adjustments of the final deployment, while for each station, the average location of the clustered solutions reveals an alternative station location. Figure 2.11 shows the family of solutions in terms of histograms of the relative distances of the stations that are measured in east-west (red) and north-south (green) direction from the average model, where the station numbers correspond to the numbers shown in Fig. 2.9c.

A comparison between the three solutions, shown in Fig. 2.9a-c, is performed by evaluating the coarray of each array geometry, which illustrates the set of all interstation distances. (Haubrich, 1968). According to the coarrays depicted in Fig. 2.12, it is obvious that the objective function which is related to the array beam power, i.e. f_1, has a tendency to distribute stations regularly with large interstation distances, such that the smallest interstation distance for the related optimized model (Fig. 2.9a) is bigger than those in the other two arrays. In addition, the maximum interstation distance for this case is comparable with the dimension of the deployment site. On the other hand, the objective function f_2, which is related to the precision of the slowness vector, results in a model (Fig. 2.9b) that has stations at closer interstation distances to achieve a better azimuthal resolution of the target swarm area (the backazimuth to the swarm region is shown in Fig .2.7a). Nevertheless, the final optimized solution (Fig. 2.9c) keeps properties of two other models as much as possible to achieve optimization in terms of both desired properties measured by the contributing functions.

In all cases, the optimization algorithm avoids the high noise area (red colors in Fig. 2.9) for deployment of any station, although to achieve optimal interstation distances and optimal azimuthal resolution, some stations are selected at the margin of the high noise area (yellow colors in Fig. 2.9). Although the ARF has not been considered directly as a design criteria, it can be used to compare the array geometries as a standard tool, so the ARFs of the models are plotted in Fig. 2.13. A circular shape of the ARF's main lobe in Fig. 2.13a reflects a uniform distribution and azimuthal symmetry of the related interstation distance vectors (Figs. 2.9a and 2.12a), while non-uniform distribution of the other two geometries (see Figs. 2.9b,c and

Table 2.3: Input parameters to set up ARF as an objective function.

Parameter	Value
Frequency band	2-8 Hz
Slowness range	$\pm 0.3 s/km$
Number of slowness grids	200
Number of stations	N=7

2.12b,c) causes the elliptical main lobes with the minor axes, i.e. highest array resolution, in the azimuthal direction of the events.
Convergence rate of the final objective function is shown in Fig. 2.14. According to the initial setup parameters (Table 3.2), the first 500 models are randomly chosen, while the later models are generated by perturbing the selected guiding models to reduce the value of the objective function. A relatively large variation in the early iterations is due to the large dimension of the explored model space, while at the final iterations the selected guiding models change little. The locations of the individual tested models in the objective function space are shown in Fig. 2.15, where the white star indicates the final optimized solution. In this figure the value of the final objective function is represented by the color bar.

2.3.3.3 More constraints: Forbidden zone

We imposed a hypothetical condition to the scenario, that is the existence of a forbidden zone in the deployment site that should be excluded in the model sampling. Such a forbidden zone is realized for instance because of ownership issues. The assumed forbidden zone is marked with a rectangle in Fig. 2.9d. The final optimum geometry (Fig. 2.9c) has already two stations situated in the forbidden zone. So the optimization algorithm has to find another distribution of stations considering this geometrical limitation. The new solution is depicted in Fig. 2.9d, and the corresponding point in the objective function space is shown in Fig. 2.15. The corresponding coarray and ARF of this geometry are shown in Fig. 2.12d and Fig. 2.13d. For the spatial distribution of the family of solutions for each station see Fig. 2.16. The performance of two final geometries (black and white stars in the Fig. 2.15) are quite the same in terms of the objective function related to the array beam power, f_1, and the final geometry with forbidden zone is slightly better than the other one in terms of the objective function related to the accuracy of slowness vector, f_2.
Although in this example we used a rectangular shape as a forbidden zone, the algorithm is flexible to consider an arbitrary shape and a number of forbidden zones.

2.3.3.4 Comparing with regular geometries

We evaluated the objective functions for some well-known regular arrays to compare with the final solutions suggested in this study. Regular shape arrays are depicted in Fig. 2.17 and the values of the objective functions are shown in Fig. 2.18. In addition the Vogtland array (Fig. 2.3b) is included in this comparison.
We concluded that, the final arrays which are suggested for the region using introduced scenario and boundary conditions perform better than the other arrays in terms of f_2 (precise slowness vector). However, a uniform array geometry, shown in Fig. 2.17d, can perform better than the others in terms of the objective function f_1 (array beam power). Such an array gives approximately the same value of the objective function f_2 as both circular arrays with and without a station in the center. The spiral arm shaped array and the Vogtland array operate rather similarly for both objective functions.

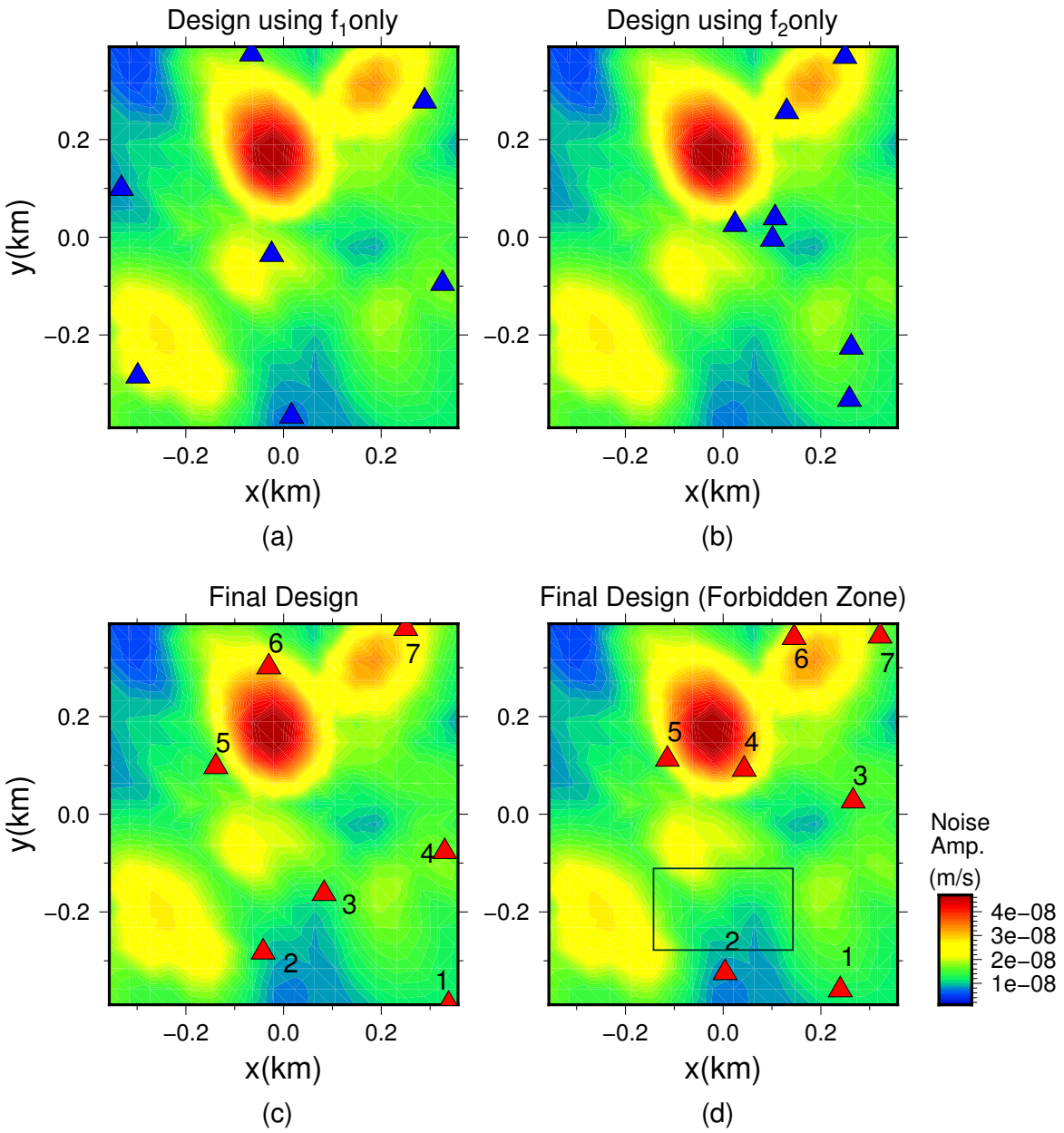


Figure 2.9: Optimized geometries found by using (a) f_1 (array beam power) and (b) f_2 (precise slowness vector) as objective function. Corresponding values of the objective functions are used to estimate γ in equation (2.10). (c) Final optimized geometry suggested for the test area using the final combined objective function (equation 2.12). (d) The geometry suggested assuming an extra condition which is a forbidden zone in the area is indicated by a black rectangle. The simulated noise amplitude pattern is depicted in the background of all figures. The related coarrays and ARFs are plotted in Figs. 2.12 and 2.13, respectively. In plots (c) and (d), station numbers are correspond to the histograms depicted in Figs. 2.11 and 2.16, respectively, which indicate potential adjustments of each station calculated from the family of solutions.

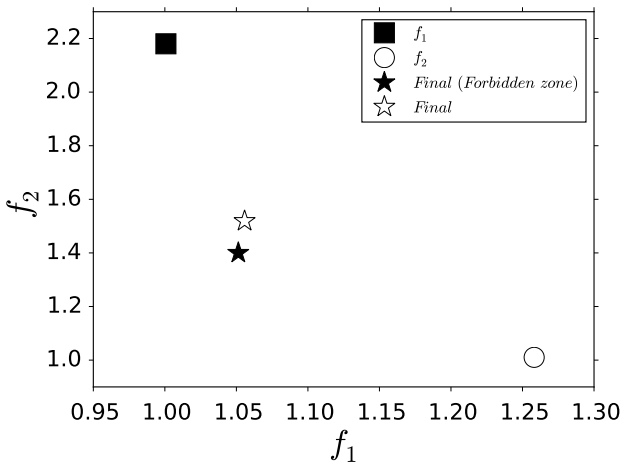


Figure 2.10: The value of sub-functions for the geometries introduced in Fig. 2.9. The square and the circle represent geometries which are outcomes of a single objective optimization using f_1 and f_2 individually to approximate the Pareto curve and to estimate the coefficient γ in equation (2.9). The white star represents the final optimized model and the black star corresponds to the optimized geometry under the additional constraint of a forbidden zone in the search area; both solutions are obtained from the minimum value of the objective function.

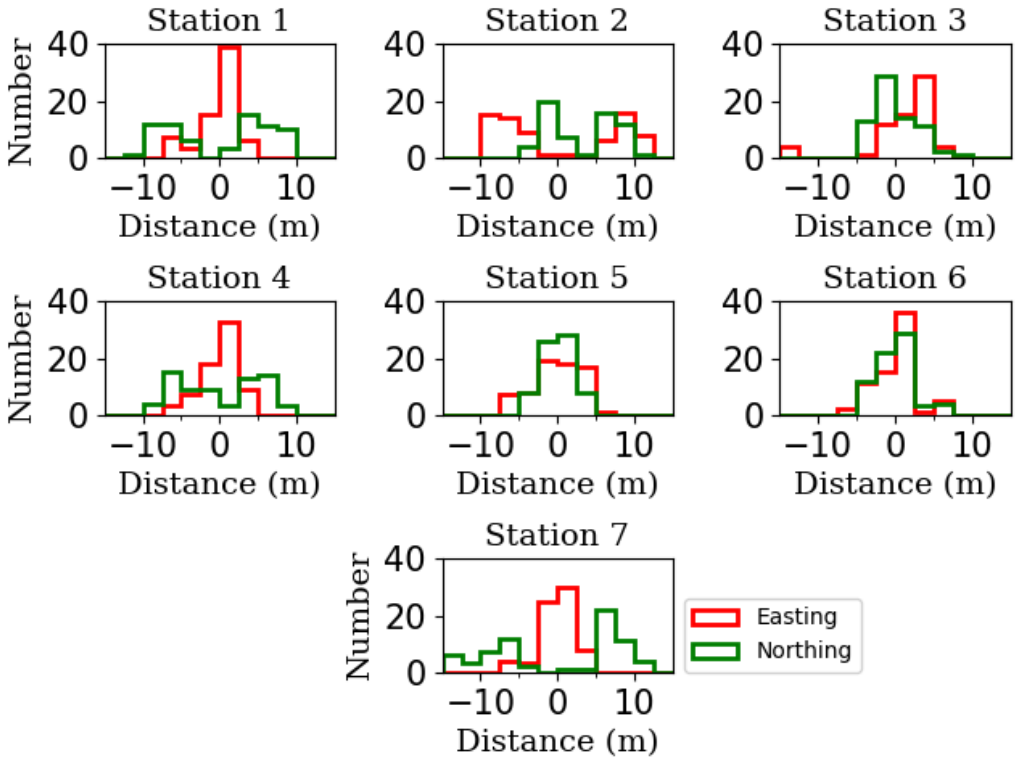


Figure 2.11: Histograms of the spatial distribution of family of solutions for the geometry shown in Fig. 2.9c. For each station the red line shows the distribution of distances in east-west direction relative to the average location and the green line illustrates the same values in the north-south direction. Station numbers are in agreement with the station order in Fig. 2.9c.

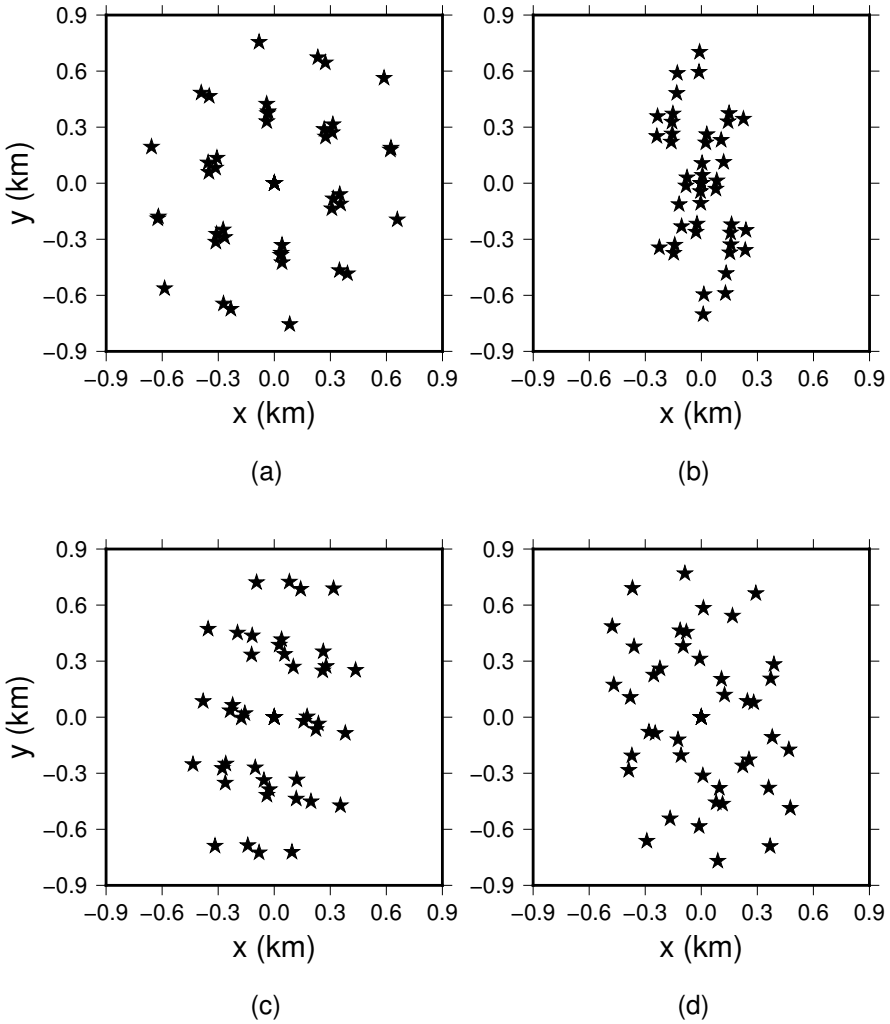


Figure 2.12: Coarrays of the arrays shown in Fig. 2.9.

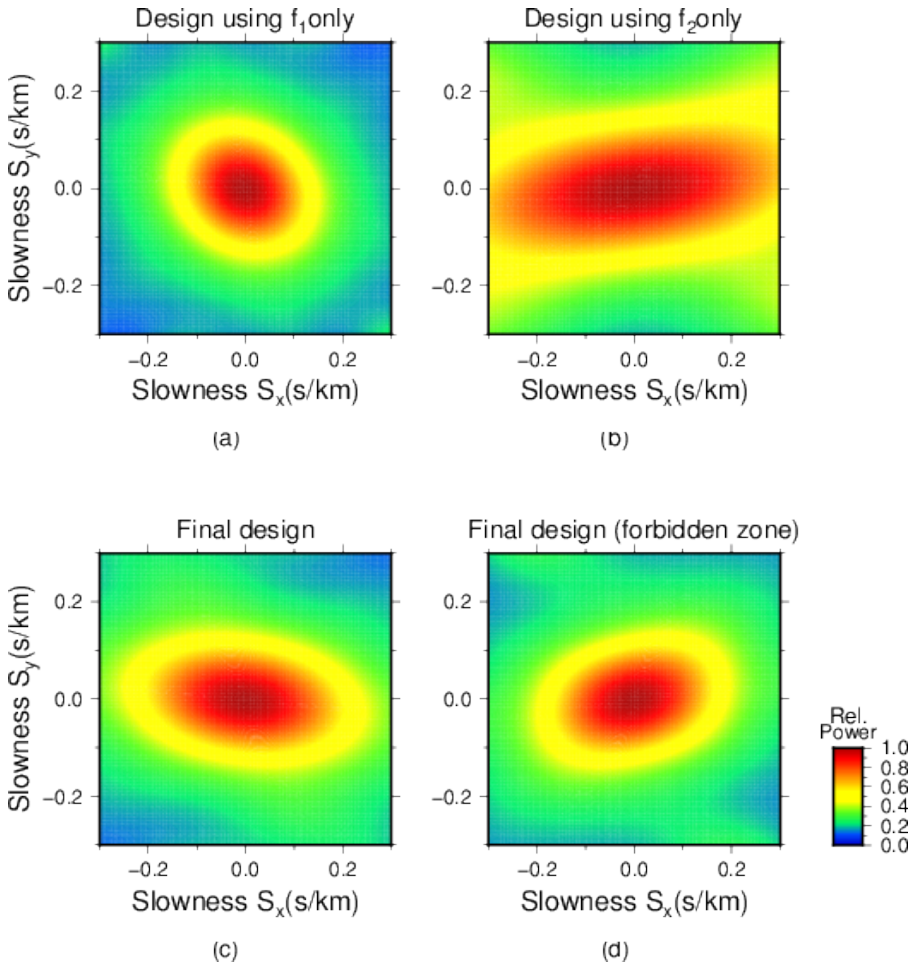


Figure 2.13: ARFs of the arrays shown in Fig. 2.9

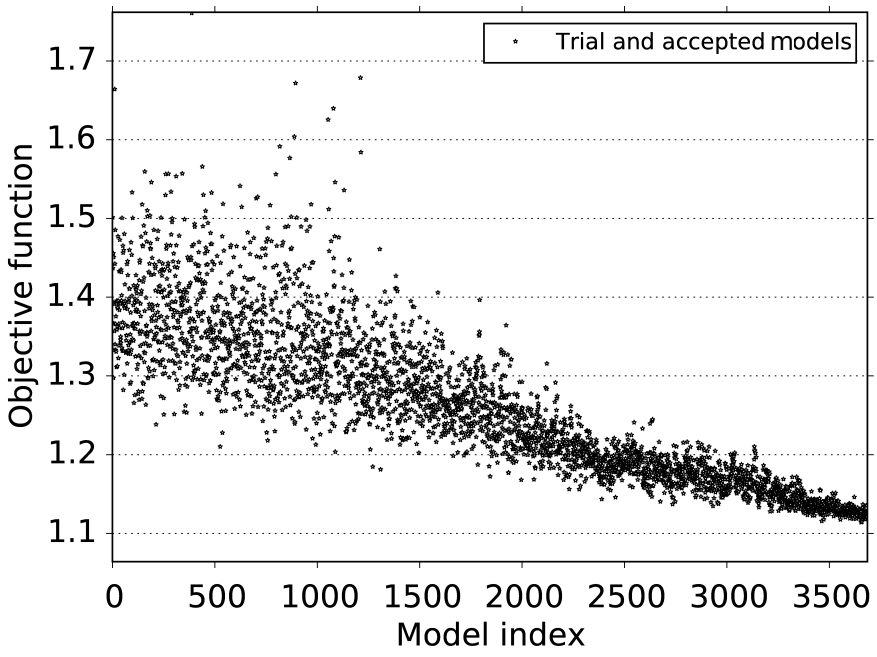


Figure 2.14: Variation of the final objective function for random models generated and tested in the optimization process. According to the initial setup of the optimization process, the first 500 models are randomly chosen, while the later models are generated by perturbing some guiding models. The amplitude of model perturbation is reduced in each iteration.

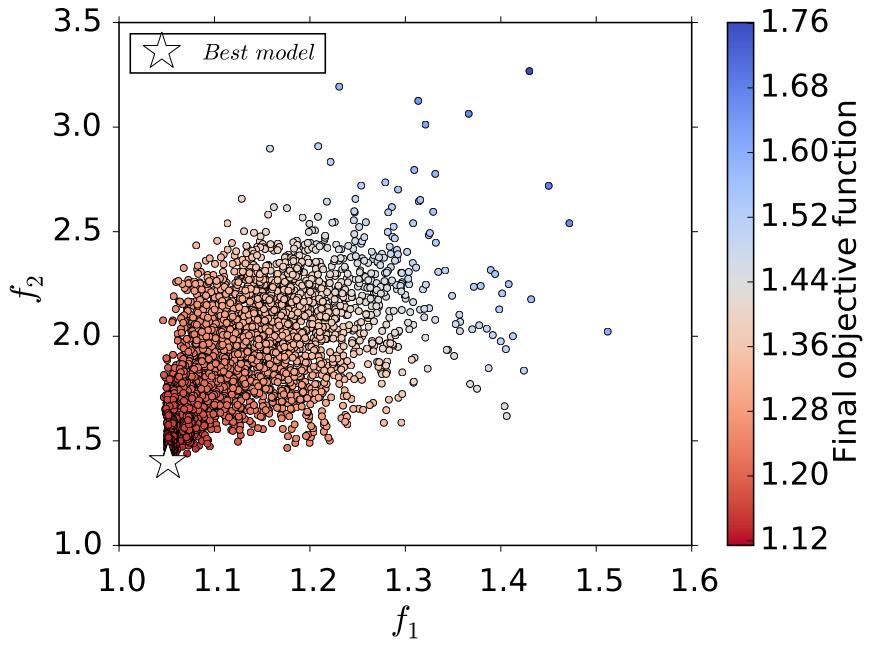


Figure 2.15: Values of individual tested models in the objective function space, for which the convergence pattern is shown in Fig. 2.14. The white star indicates to the final model. The colorbar represents the value of the final objective function.

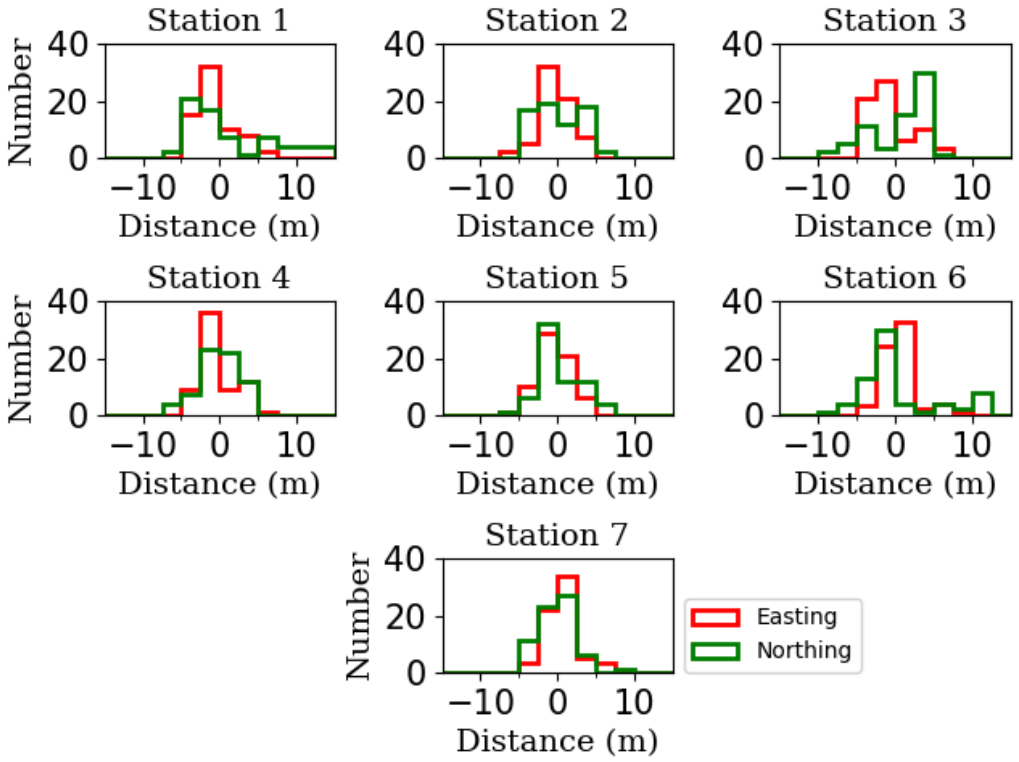


Figure 2.16: Histograms of the spatial distribution of family of solutions for the geometry shown in Fig. 2.9d. For each station the red line shows the distribution of distances in east-west direction relative to the average location and the green line illustrates the same values in the north-south direction. Station numbers are in agreement with the station order in Fig. 2.9d.

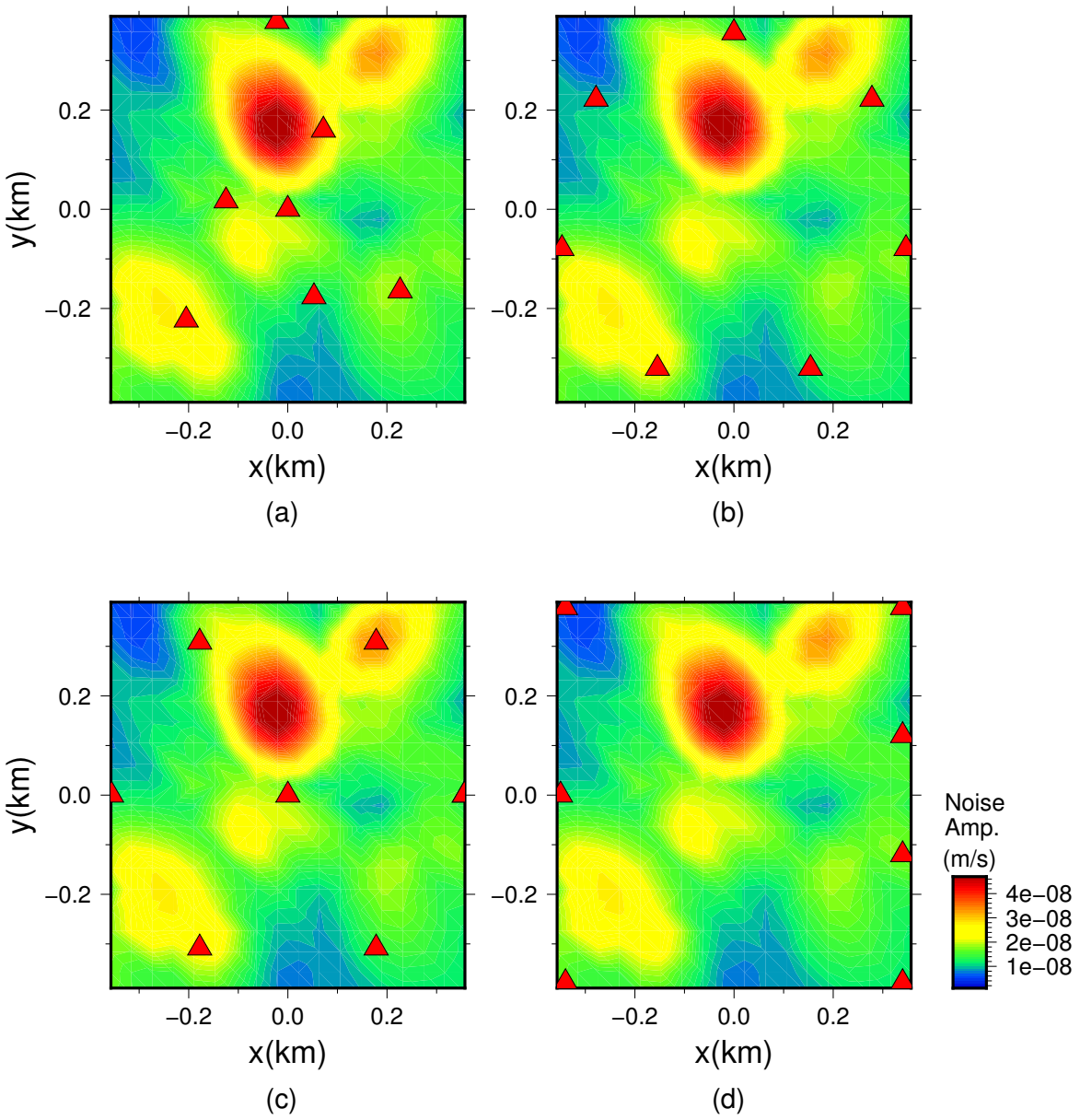


Figure 2.17: Examples of 7-station regular arrays, spiral arms (a), circular (b), circular with center (c) and two uniform lines (d), which are used for comparison with the final optimized array. Corresponding values of the objective functions are shown in Fig. 2.18.

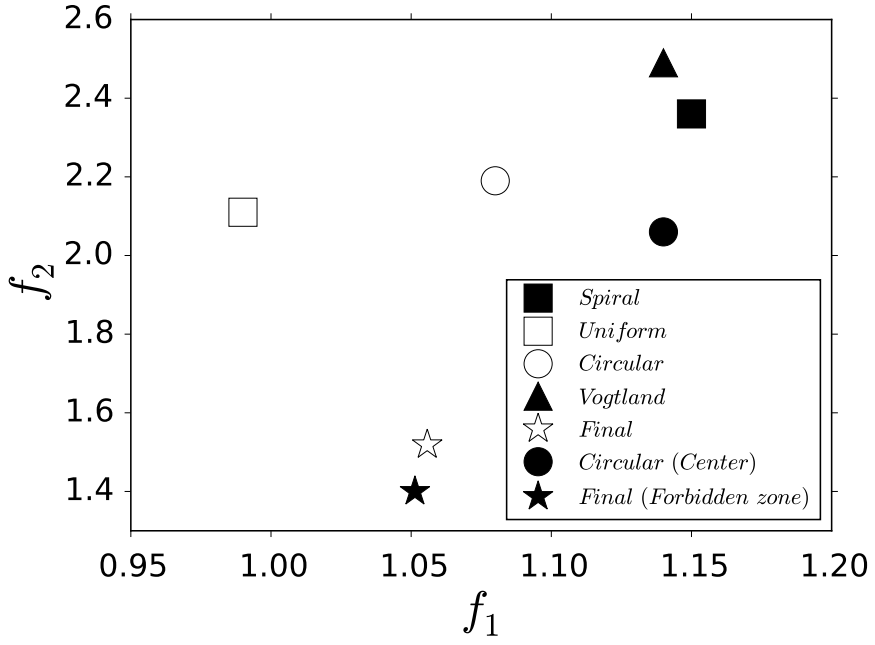


Figure 2.18: Comparing the performance of the final suggested arrays (white and black stars) with regular geometries depicted in Fig. 2.17 and the Vogtland array shown in Fig. 2.3b.

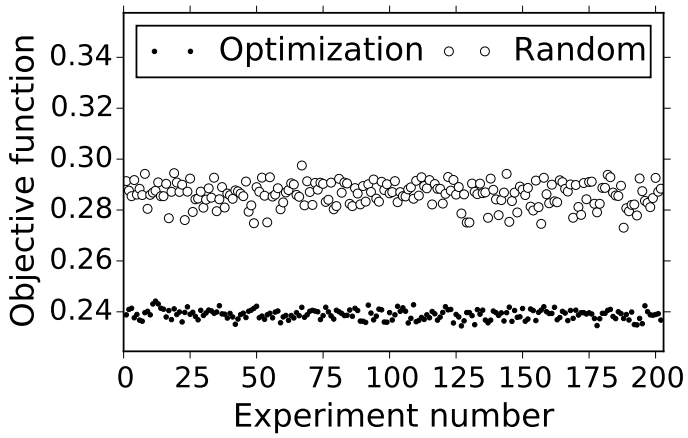


Figure 2.19: Results of repeating the optimization process using the same initial conditions. The ARF is used as objective function. Circles mark the minimum value of the initial populations that are randomly drawn, and black dots are the minimum value of the objective function obtained during the optimization process.

2.3.3.5 Robustness of the optimization method

We did an experiment to test the robustness of the optimization process by running it 200 times and analysing the minimum value of the objective function obtained in every run. In this test, to decrease the calculation time, we used the theoretical ARF (equation 4.1) as the objective function with the set up parameters as summarized in Table 2.3. The setup parameters for the optimization program are the same as the optimizations using the other objective functions discussed in this paper and are already introduced in Table 3.2.
In Fig. 2.19, for each experiment, the minimum values of the objective function calculated from the initial random population (circles) and the optimization process (dots) are plotted. From the results we conclude that, as we repeat the optimization process, while the final solutions are not always the same, the minimum value of the objective function remains in the same level. In addition, this figure indicates that, since the improvement of the minimum value of the objective function from a random population to the final optimized solution is almost consistent, the general trend of the convergence remains the same if we repeat the experiment. In practice, it is a useful property of the algorithm, that it can suggest several geometries with approximately the same performance. Accordingly, the user can decide to deploy an array geometry based on the unforeseen circumstances at the deployment site.

2.3.4 Conclusion

In this study, we have developed an array design framework which provides a flexible tool to perform an automatic search for an optimized array geometry. The framework implements a customizable scenario modelling and optimization scheme. The main point of this framework is to use realistic synthetic seismograms, which makes it possible to take into account ranges of possible source mechanisms, geometry of the seismogenic zone, propagation velocities and noise levels at potential deployment sites. The synthetic array data is then used to evaluate one or more objective functions to be minimized during the optimization process. The objective functions can be defined by user preference based on the monitoring goals.
We showed an example of a 7-station array design in a swarm activity zone in northwest Bohemia/Vogtland area. By presenting this example, our aim is to show how the array design framework can be initialized and customized for a specific problem. The parameters needed for the synthetic array waveform modelling are derived from previous seismological studies of this region added by realistic estimation of synthetic noise. The synthetic noise level is estimated from real samples of waveforms recorded by an array that was installed temporarily in the same area (Fig. 2.3b).
Two objective functions are introduced to ensure the accuracy of the estimated P- and S-phase slowness of synthetic signals. Using the scalarization technique, two objective functions are combined in one single function and a final optimized geometry is introduced.
By a statistical robustness test of applied optimization process, we conclude that if we rerun the algorithm with the same initial set up, while the final model is not unique, the related value of the objective function remains almost the same. In other words, it is possible to make more than one suggestion with similar performance in terms of the defined objective function.

Chapter 3

Earthquake source arrays: configuration and applications in crustal structure studies

This chapter is published in Karamzadeh et al. (2020).

Abstract

A collection of earthquake sources recorded at a single station, under specific conditions, are considered as a source array (SA), that is interpreted as if earthquake sources originate at the station location and are recorded at the source location. Then, array processing methods, i.e array beamforming, are applicable to analyze the recorded signals. A possible application is to use source array multiple event techniques to locate and characterize near-source scatterers and structural interfaces. In this work the aim is to facilitate the use of earthquake source arrays by presenting an automatic search algorithm to configure the source array elements. We developed a procedure to search for an optimal source array element distribution given an earthquake catalog including accurate origin time and hypocenter locations. The objective function of the optimization process can be flexibly defined for each application to ensure the prerequisites (criteria) of making a source array. We formulated four quantitative criteria as sub-functions and used the weighted sum technique to combine them in one single scalar function. The criteria are: (1) to control the accuracy of the slowness vector estimation using the time domain beamforming method, (2) to measure the waveform coherency of the array elements, (3) to select events with lower location error and (4) to select traces with high energy of specific phases, i.g, sp- or ps-phases. The proposed procedure is verified using synthetic data as well as real examples for the Vogtland region in northwest Bohemia. We discussed the possible application of the optimized source arrays to identify the location of scatterers in the velocity model by presenting a synthetic test and an example using real waveforms.

3.1 Introduction

In seismology the word “array” implicitly refers to a “ receiver array”, that is a number of sensors deployed in a special geometry with common precise timing, acquisition parameters and instrument types. A receiver array signal processing provides signal-to-noise ratio (SNR) improvement by enhancing coherent signal arrivals while suppressing the incoherent background, using a time-domain or frequency-domain stacking procedure (Rost and Thomas, 2002; Schweitzer et al., 2012; Rost and Thomas, 2009). Using the so called “array beamforming” method, slowness vectors of the incoming waves are estimated, which yields to seismic phase identification and further seismological findings, i.e. event location and rupture front tracking (Krüger and

Ohrnberger, 2005; Ishii, 2011). Based on the reciprocity theorem of the Green's function, under particular conditions, a number of seismic sources recorded at a single station are called a " source array" (SA), in an analogy of an array of sensors (Spudich and Bostwick, 1987; Scherbaum et al., 1991; Krüger et al., 1996). The reciprocity theorem implies that for a vectorial force and receiver, if the positions of source and receiver in a seismic experiment are exchanged, the observed seismograms remain identical. Thus, after corrections due to different origin times and radiation pattern, signals of an SA received at a surface station are interpreted as if they originated at the location of the station and are recorded at the location of sources. This basic idea provides a unique tool to study the seismic structure of the areas with as few as a single recording instrument using array processing methods. In addition, by simultaneous use of an RA and an SA, double array method (Krüger et al., 1993, 1995; Scherbaum et al., 1997) is applicable to achieve a further SNR improvement which makes it possible to study low amplitude scattered phases and to locate scatterers out of the great circle path more precisely. To benefit from SA requires that the coordinate, origin time, and depth of each source are precisely known. In addition, akin to the waveform coherency requirement for an RA, waveforms of an SA should be coherent. This criterion implicitly indicates that the travel path and source mechanisms of events are similar. If they are not identical, it is necessary to remove the source mechanisms from the recorded waveforms. Due to these pre-requisites, either nuclear explosions or big earthquakes are most often used to define an SA. The former due to more simple waveforms and well controlled source parameters, and the latter due to the possibility to remover a well-known source mechanism. The use of SA technique and the double array method are mostly focused on the study of properties of the velocity interfaces in different scales using converted or reflected phases, such as crust-upper mantle boundary structure (Niazi, 1969; Goldstein et al., 1992), the distribution of heterogeneities in the Earth's mantle and subduction zones (Weber and Wicks, 1996), S-wave coda composition of microearthquakes (Scherbaum et al., 1991), origin of S-phase coda from earthquakes (Dodge and Beroza, 1997), and the imaging of crust and upper mantle structure using scattered energy within the coda of the teleseismic P-phases (Revenaugh, 1995). Inhomogeneities in the lowermost mantle was studied by reflected phases using double beam forming techniques (Krüger et al., 1993, 1995; Scherbaum et al., 1997; Krüger et al., 2001; Rietbrock and Scherbaum, 1999). In addition, recently the double array method is used to extract body and surface waves in volcano ambient noise processing (Nakata et al., 2016). The double array technique was also used to identify and separate low-amplitude body waves from high-amplitude dispersive surface waves at the exploration geophysics scale (Boué et al., 2013) and for acoustic tomography in a shallow ultrasonic waveguide (Roux et al., 2008).

SA allows for full slowness field calculation of the phases emerging from the source region. Thus, the positive attribute of using an SA is due to the possibility of constructing a 3-D array to study the near source structure and the travel path of the wavefield. In addition, providing the prerequisite conditions are fulfilled, without bearing the extra cost of sensor installation array processing methods are applicable and higher SNRs are achievable. Accordingly the number of SA elements can be increased, without paying extra cost. On the other hand, the negative attributes are due to the prerequisite conditions, which can be hard to achieve, such as similarity of sources, similarity of near source structure, and earthquake location and origin time determination with high precision.

Given a well-located catalog, which can be available by advanced relocation procedures and a well-configured monitoring system, the main challenge to make use of an SA is to eliminate the effect of waveforms dissimilarity and search for coherent arrivals. If events are big enough to calculate the related focal mechanisms and source time functions, one can deconvolve source effects from the signals. However, this is often not possible for smaller events. But smaller events are more numerous and we can alternatively search for similar waveforms indicating

similar source mechanisms and travel paths. This can be feasible when a large number of well-located events in a relatively small volume exist in the catalog. In analogy to an array of receivers, the distribution of source array elements is also imperative aside from the signals coherency, in order to get the precise slowness vector components (Schweitzer et al., 2012; Rost and Thomas, 2009). To fulfill all these criteria, our solution is to search for similar events in a specific geometry that imposes enough time delays in signal arrival times to estimate the slowness vector precisely based on coherent waveforms of interested phases.
In the present study, the aim is to facilitate the use of SAs by automatic search of array elements, given a catalog of events and associated waveforms. We propose the use of an optimization technique to find the best source array configuration. Using an optimization technique allows to define one or more objective functions to handle necessary criteria of setting the SA elements. We suggest four sub-functions, but in general the number of sub-functions can be adapted according to the specific goals of the studies: (1) to ensure high resolution of SA beamforming in 3-D slowness space, that can be evaluated using either synthetic or real data. This criterion controls the SA geometry, (2) to check the P-phase waveforms similarity over the array elements, (3) to select events with low location error and (4) to check the presence of reflected/converted phases. Then the weighting sum technique (Caramia and Dell' Olmo, 2008) is used to linearly combine the defined sub-functions in one scalar objective function. The proposed procedure can promote a variety of interesting applications of SA techniques related to the study of the near source structural complexity, e.g. to localize the position of a velocity or density contrast within the source region of an earthquake swarm. The double array technique further allows to study the velocity interfaces in different scales.

3.2 Theoretical approach

We define a "model", m, as a source array with a specified number of elements, K. Each element of m, i.e. m_k is the hypocenter of an earthquake which occurred within a confined epicentral area. A random model has a 3-D spatial configuration, every element of which is drawn from a given event's catalog E with J events. It is assumed that the given catalog, includes precise event locations and origin times as well as related errors, but the source mechanisms of the events are not necessarily available. The "model space" accommodates all possible K-element combinations of the J given events. The aim is to find the best model, i.e. a member of model space, that optimizes (minimizes) a well-defined scalar objective function.

In this section the objective function formation as well as the implemented optimization algorithm are introduced. The objective function is evaluated during the optimization process employing the data associated to a trial model, such as the waveforms and location error in the catalog. A key point of our approach is to automatize the elaborate task of the SA set-up by utilizing an optimization process.

As an example application of the source arrays in crustal structure studies, a grid search algorithm to locate a scatterer is described in this section as well.

3.2.1 Objective function formation

We introduced and evaluated four criteria to search for the best model. Each criterion is to target one specific property of the desired SA, and is formulated in one individual sub-function. The weighted sum technique is used to estimate the relative weights of the sub-functions and to combine them in one single scalar function. The single final objective function is used in the final optimization process, which results in the optimized SA. In the following subsections, first, individual sub-functions are introduced, then the final objective function formation is presented.

3.2.1.1 3-D array slowness resolution

Assuming the signal coherency is valid and the sources have similar mechanisms, the 3-D source array beam trace is calculated as the sum of all time shifted traces recorded at one single station:

$$B(t) = \frac{1}{K}\Sigma_{j=1}^{K} \; Y_j(t - \tau_j + dT_j) \qquad \text{with} \qquad dT_j = s_x\delta x_j + s_y\delta y_j + s_z\delta z_j. \tag{3.1}$$

where, Y_j is the normalized trace of source j and K is the number of sources. τ_j is the origin time of the source j. Assuming the plane wave approximation is valid, the time shift dT_j, for the source j, depends on the slowness components of the leaving wavefront s_x, s_y and s_z, and the relative distance to the SA reference point (center) δx_j, δy_j and δz_j. Accordingly, the time shift needed for array beamforming depends on the slowness vector components and source array elements distance vector from the array center. The precision of the estimated slowness vector using array beamforming depends on the aperture and the 3-D inter-source distances, i.e. 3-D geometry of the elements. Thus, by defining an objective function based on the source array beam power, we can take into account the SA geometry. We measure the average amount of relative power of the 3-D array beam which is calculated in 3-D slowness space using the given sources and one station location. Ideally, array beam power in slowness space contains a concentrated sharp main peak, like a delta function, and no other localized secondary peaks. The maximum value of the array beam relates to the resolved slowness vector. The sharpness of the main peak shows the resolution of the estimated slowness vector. A sharp main peak causes a relative low plateau surrounding the peak, resulting in a small average of the beam power in s_x-s_y-s_z space. To control the accuracy of slowness vector estimation and achieve the highest possible slowness resolution using the SA, we set our objective to minimize the average value of the array beam power. An SA beam is calculated for all 3-D grid points of the slowness space, considering appropriate time windows of the traces around arrival times of the selected phases. Then, three 2-D cross-sections of the 3-D beam pattern, crossing the beam power maximum value, are extracted from the full 3-D pattern and the related average beam power is calculated as:

$$f_{xy} = \frac{1}{n_x n_y}\Sigma_k^{n_y}\Sigma_l^{n_x} A^2_{x_l,y_k}, \tag{3.2}$$

$$f_{xz} = \frac{1}{n_x n_z}\Sigma_k^{n_z}\Sigma_l^{n_x} A^2_{x_l,z_k}, \tag{3.3}$$

$$f_{zy} = \frac{1}{n_z n_y}\Sigma_k^{n_y}\Sigma_l^{n_z} A^2_{z_l,y_k}, \tag{3.4}$$

where, n_x, n_y and n_z are number of grid points in the predefined slowness ranges for s_x, s_y and s_z; and A^2 is the value of beam power which is normalized to the global maximum of all tested grid points, so it is called relative beam power. The average relative beam power, as a measure of 3-D array resolution, is defined as f_1 to combine above equations in one formula and to define the first objective function as:

$$f_1 = \frac{1}{3}(f_{xy} + f_{xz} + f_{zy}). \tag{3.5}$$

A larger value of f_1, indicates higher value of the average beam power in the slowness space for the SA, which can happen either because of high or numerous side lobes or a wide main lobe in each of the 2-D slowness maps, both of which are undesirable.

3.2.1.2 Waveform similarity

Similar to the RA, waveforms of an ideal SA should be identical in full measure. This condition is fulfilled if the source parameters of all events are identical and the travel paths between the sources and the station are similar. The given event catalog can include heterogeneous source mechanisms, accordingly for a randomly selected model, including a number of random sources, dissimilarity of the related waveforms is likely to happen. Nevertheless, the desire is to have maximum waveform coherency of the direct and the reflected/converted phases for the events of the optimized SA. Waveform similarity over the full length of the recording is not needed and cannot be ensured, because due to different sources-receiver and sources-reflector paths the time delay of the secondary phases from the first arrival P-phases may not be identical for all sources. In addition, we presume that events of similar P-phase waveform, aside from similar source mechanisms, might travel along similar source-receiver paths, consequently the waveforms of the secondary phases, i.e P to S converted phases ps, may be similar as well. Thus, an objective function f_2, is introduced to measure P-phases waveform similarity of a given SA, by calculating the waveform cross-correlation matrix, where each element is the cross-correlation of a pair of P-phase waveforms measured in time domain. Accordingly, a K^2 element symmetric the cross-correlation matrix, is associated to each model.

Given two signals x_k and $x_{k'}$ are related to the sources k and k', respectively, we can delay x_k by m samples and then calculate the cross-covariance between the pair of signals, that is:

$$\sigma_{kk'}(m) = \frac{1}{N-1}\Sigma_{n=1}^{N}(x_k(n-m) - \mu_{x_k})(x_{k'}(n) - \mu_{x_{k'}}), \tag{3.6}$$

where μ_{x_k} and $\mu_{x_{k'}}$ are the means of each signal and there are N samples in each. n is the sample's number; assuming t is the time, $n = \frac{t}{\delta t}$ and δt is the sampling rate of the signal. The function $\sigma_{kk'}(m)$is the cross-covariance function. The cross-correlation function is a normalized version of the cross-covariance function:

$$\sigma'_{kk'}(m) = \frac{\sigma_{kk'}(m)}{\sqrt{\sigma_{kk}(0)\sigma_{k'k'}(0)}}. \tag{3.7}$$

The $\sigma'_{kk'}(m)$ is calculated for a possible range of valid m, which is defined based on the signal's length. Then the maximum of $\sigma_{kk'}(m)$, $\hat{\sigma}$, is used to define the cross-correlation matrix:

$$C = [\hat{\sigma}_{kk'}], k = 1:K, k' = 1:K, \tag{3.8}$$

where, C is the cross-correlation matrix. Assuming the given signals are P-phase waveforms, then f_2, as a second sub-objective function is defined as:

$$f_2 = \left(\Sigma_{k=1}^{K-1}\Sigma_{k'=k+1}^{K}(1 - C_{kk'})^2\right)^{\frac{1}{2}}. \tag{3.9}$$

3.2.1.3 Location error

Although we assumed that the given catalog includes well-located events, any event location solution comes with a specific location error, regardless of the type of implemented location method. We defined another criterion, f_3, to ensure the selection of the events with the lowest location error. The total location error for one event is measured as:

$$e = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}, \tag{3.10}$$

where σ_x, σ_y and σ_z are the errors in latitude, longitude and depth of each earthquake, respectively. f_3 for an assumed SA is defined as:

$$f_3 = \sqrt{\Sigma_{k=1}^{K} e_k^2}, \tag{3.11}$$

3.2.1.4 Presence of scattered phases

Arrival time and amplitude of the near-source scattered waves on seismograms, depends on the geometry and position of source, receiver and the scatterer, while their amplitude is strongly dependent on the focal mechanisms and the radiation pattern (e.g. Hrubcová et al., 2016). The scatterer, i.e velocity interface, can be localized or laterally extended in part of the propagation media. Thus, depending on the incident angle of the phases impinging on the interface, the relative distances between the sources and the interface, and the velocity contrast, the scattered phases can be strong or weak or not existing at all. Furthermore, if the media is highly scattering or more than one scatterer exists, deconstructive interference of the scattered phases is likely to happen. Accordingly, the waveforms are inspected to verify if such phases are present with considerable energy.
The ideal is to select events with the strongest reflected/converted phases produced from the same reflector. Accordingly, we suggest to measure the average kinetic energy of the waveforms, in a specified time window between the P- and S-phase arrivals, for each element of model after rectifying the magnitude differences between events. To rectify the magnitude differences, the traces are scaled by the maximum/minimum amplitude of P-phases, so that the maximum/minimum amplitude of P-phases are normalized to 1 or -1. So, for P to S and S to P phase scatterings, the fourth objective function is defined as:

$$f_4 = \frac{w}{\Sigma_{n_i}^{n_{i+w}} x_n^2}, \tag{3.12}$$

where w is the number of samples in the specified time window and n_i is its beginning sample. f_4 is defined as the inverse of the mean energy of the normalized traces to contribute in the overall objective function. In our tested example we assumed the segment of the trace after the P- phase and before the S-phase arrival time. However, in application, the time window can be adjusted according to the theoretical travel time of the interested secondary phases.

3.2.1.5 Final objective function

Each sub-function introduced before, has different unit and range of values, so before combining, they should be scaled and made dimensionless. The range of the sub-functions are normalized using the general formula given by:

$$\hat{P} = \frac{P - P_{min}}{P_{max} - P_{min}}, \tag{3.13}$$

where, P is the original value and $\hat{P}$ is the normalized value, assuming P has the limited range between P_{min} and P_{max} which are estimated by running optimization algorithm using each sub-function individually. Having normalized each of the sub-functions, the weighting coefficients are calculated to combine them in one single scalar function. In general there is not a unique way to weight the sub-functions and different weights lead to the different final solutions, which all are mathematically correct. We employed a procedure to define weighting coefficients between two sub-functions so that both sub-functions could contribute almost equally into the final solution (Karamzadeh et al., 2019a). The method is extendable to define the weighting coefficients of more sub-functions (Appendix A). To combine four sub-

Inputs: Optimized source array (m_0 and Waveforms); Velocity model; Phase name (i.e: sp or ps)
Output: Location of the scatterer
procedure SEMBLANCE
  Grid the 3-D travel path $\rightarrow N_g$ Grids.
  for each grid point g in Grids **do**: ▷ i=1:N_g
    for each source e in m_o **do**: ▷ k=1:K
      Estimate the travel time of the scattered phase using the velocity model assuming g is the scatterer point (eq:15).
      Select the phase waveform using the estimated arrival time.
    Calculate the semblance (Eq. 17) for the selected waveforms and assign to grid point g_i.
  Maximum value of the semblance $\rightarrow$ location of the scatterer g_f.

Figure 3.1: Source array optimization algorithm. A model m is defined as a source array with K events. The value of input parameters in the test examples in this study are selected as: E the given catalog of 570 events; $K = 20$ number of events in each model; $n_{start} = 200$ number of initial random models; $n_{select} = 20$ number of guiding models in each iteration; $n_{new} = 20$ number of new accepted models per each guiding model; $n_{try} = 10$ number of attempts to make n_{new} accepted models per each guiding model; $n_{iteration} = 15$ number of iterations. After the last iteration a model m_o with the least value of objective function is selected as the final optimized source array.

functions, first the weighting coefficients between $\hat{f}_1$ and $\hat{f}_2$ are estimated and a combined function, f_{12} is made, then following the same procedure the weighting coefficients between $\hat{f}_3$ and f_{12} are calculated and a combined function is named f_{123}. Finally, the coefficients between $\hat{f}_4$ and f_{123} are calculated and the final scalar objective function is defined as:

$$F = \alpha\hat{f}_1 + \beta\hat{f}_2 + \gamma\hat{f}_3 + \lambda\hat{f}_4, \tag{3.14}$$

where, $\hat{f}_i$ are scaled sub-functions, and weighting coefficients α, β, γ and λ are defined so that: $\alpha + \beta + \gamma + \lambda = 1$.

3.2.2 Optimization algorithm

The aim of the optimization process is to find the best possible model, m_o, from the model space. Assuming a specific receiver, it gives the best performance of the defined scalar objective function (eq: 3.14) as described in section 3.2.1.5. Since the objective function calculation can be computationally expensive (it takes about 1 minute for one function evaluation using a 8-core processor), we used a fast converging procedure (Karamzadeh et al., 2019a) which is a modified version of the simulated annealing procedure that benefits from a sampling technique based on the neighbourhood concept (Sambridge, 1999; Wathelet, 2008) and an objective function approximation using Voronoi cells (Okabe et al., 1992).
The optimization process (see Fig. 3.1) works according to the following steps: (1) Operate initial models: The n_{start} (input parameter) initial models are generated as the model population, M, which is a subset of the model space. Each model is created by a random selection of K (input parameter) events from the given catalog. However, a minimum inter-event distance threshold of 100 m is considered to avoid spatial overlap of array elements. (2) Update objective function: The objective function is evaluated for models in M to make, at the first iteration, and update, at the later iterations, the objective function population, O. (3) Select guiding models: M is sorted according to the corresponding values in O, and n_{select} (input parameter) models with the lowest value in O are selected as the "best models" to produce the "guiding models" population, G. The guiding model selection is repeated in the subsequent iterations since M and O are updated. (4) Perturb randomly (update models): n_{new} (input parameter) new models are generated using each of the models g, out of the G. All events in the input catalog are divided into the K clusters using the euclidean distances of hypocen-

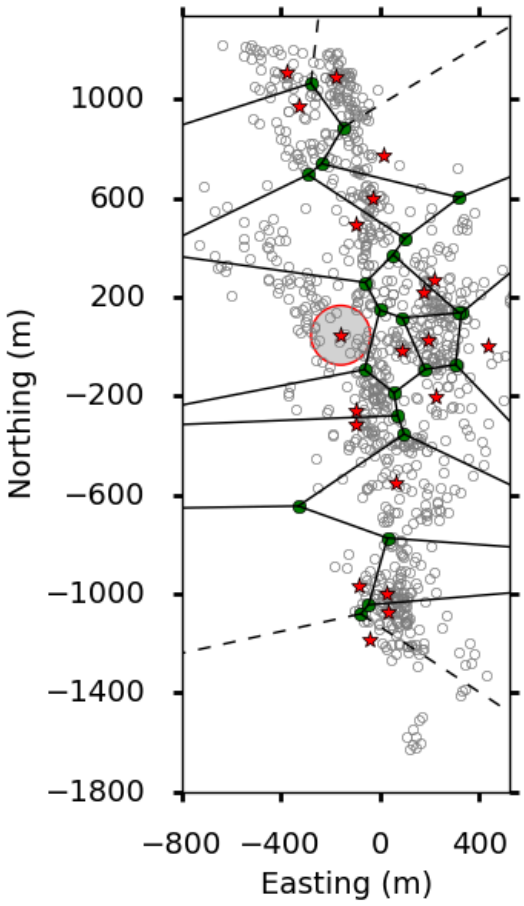


Figure 3.2: Example of Voronoi diagrams (black) used to partition a 2-D distribution of earthquake locations (gray circles) by a set of irregular cells. Each cell includes events closest to its center (red star) and so the shape of the cells is entirely defined by the location of the center. Arbitrary source array (SA) elements are used as central points of Voronoi cells. The radius of the red circle drawn around one of the SA elements indicates the maximum distance for random walk and the events inside the circle (in gray area) can be replaced by the Voronoi center. For each iteration a new set of Voronoi centers is selected.

Inputs: Catalog, E; K; n_{start}; n_{select}; n_{new}; $n_{iteration}$; n_{try}
Output: Optimized source array (m_o)

procedure SOURCE ARRAY OPTIMIZATION
  Operate initial models $\rightarrow$ model population $M = \{m_i, i : 1, n_{start}\}$.
  Create objective function population $O = \{o_i, i : 1, n_{start}\}$.
  Create guiding models population $G = \{g_i \in M, i : 1, n_{select}\}$.
  for each iteration **do**:
    for each model g in G **do**: ▷ This is to update models.
      Divide the E into K clusters, C, by Voronoi tessellation algorithm (Fig. 2).
      for n_{new} times **do**: ▷ n_{new}: No. of new models per each g.
        for each e in g **do**: ▷ each element (event location) of g is updated.
          Update e by random walk in the related C.
        Accept or Reject the updated models based on inter-event distance comparison with
          models in G.
        Add new accepted models into the M
          ▷ The procedure to find n_{new} accepted models is tried n_{try} times.
    Update objective function population O.
    Update guiding model population G. ▷ The no. of models in G: n_{select}.
  Model with minimum value in $O \rightarrow m_o$

Figure 3.3: Algorithm to find the location of the scatterer.

ters using the Voronoi clustering algorithm implemented by SciPy-Spatial algorithms and data structures package (Virtanen et al., 2019)(see Fig. 3.2). Each cluster is associated to one event $g_l \in g$, and all events in that cluster are closer to the g_l compared to the other $g_j, j \neq l$. New models are generated by replacement of each element g_l with a randomly selected event from the same cluster. The maximum possible distance between two events for random replacement is reduced by increasing the iteration steps using a cooling scheme. In this way, an ensemble of new candidate models are created using each guiding model. (5) Accept or reject the updated models: Before a candidate model is accepted as a new model, the inter-event distances of that is compared with the corresponding values of all models in M. For each model, the inter-event distance matrix I is evaluated, that is a squared matrix of K^2 elements. All values above the main diagonal of I are sorted to make a quantitative accept or reject criterion. A candidate model is accepted as a new model, if it is in the closest neighborhood of one of the best models in M, i.e. models in$G$. By applying this condition, for instance, models with small aperture (largest inter-event distance) compared to the aperture of the guiding models and models that are not well distributed in space are rejected. In every iteration, accepted models are added to the M, and O and subsequently G are updated. After the last iteration, the model with lowest value in O is reported as the final optimized model.
In this algorithm, the reason to employ the Voronoi tessellation algorithm is to cluster the given catalog into a fix number of cells, where the center of each cell is predefined by the guiding model's elements (event's location in the guiding model). Any other clustering algorithm that can do such a division is applicable as well.

3.2.3 Location of the scatterer

A possible application of an SA is to locate the scatterer which generates strong coherent scattered waves. We investigated the SA application by applying a grid search procedure to find the location of the scatterer. The procedure (See Fig. 3.3) requires an estimation of the travel path and the arrival time of scattered phases such as sp-phase and ps-phases for a single receiver and works according to the following steps: (1) A 3-D volume which includes the travel path of the scattered phases originated at the source array and recorded at the single receiver, is divided into small size 3-D grids. (2) Assuming each grid point, g_i is the trial position of the scatterer, for all sources, trial travel times of the phases are calculated by summing the s- and p- phase segments:

$$\tau_{sp}(i,k) = \tau_s(i,k) + \tau_p(i,k), k = 1 : K, \tag{3.15}$$

where K in the size of the source array. (3) The trial sp-phase waveforms are extracted from the related waveforms, using the trial travel-time values:

$$z_k^{sp}(i) = z_k(\tau_{sp}(i,k) + t_{origin}(k) : \tau_{sp}(i,k) + w + t_{origin}(k)) \tag{3.16}$$

where w is the length of the time window, and t_{origin} is the origin time of each source. Then the semblance coefficient is calculated for the selected waveforms. The grid point g_f, where the related semblance value is maximum, is assumed to be the location of the scatterer. It should be noted that for grid points, whose related $\tau_{sp}(i,k)$ were too large, and close to the $\tau_{ps}(i,k)$ (ps- phase arrival time), the semblance is not calculated and the value set to a constant value calculated from the background noise semblance. Semblance is calculated from the following formula and assigned to each g_i.

$$S = \frac{\sum_{t=0}^{t=w}(\sum_{k=1}^{k=K} z_k(t))^2}{K \sum_{t=0}^{t=w} z_k(t)^2}, \tag{3.17}$$

This approach does not require the plane wave approximation condition to be hold, so it can be applied for any source-receiver distances.

3.3 Data and experiments

The presented method is evaluated by performing two experiments, (1) a synthetic test using realistic synthetic seismograms based on a real catalog, (2) using real seismograms. In both experiments, subject of the study is northwest Bohemia/Vogtland, in the border region between Germany and Czech Republic, that is well known for the repeated occurrence of earthquake swarms. Earthquake swarms are a vast number of weak events occurring in a spatio-temporal cluster over a period of weeks or months and are not generally associated with a typical mainshock-aftershock sequence of earthquakes. The causes of earthquake swarms are considered to be either magmatic activities (Dahm et al., 2008; Morita et al., 2006), fluid-migration (Hensch et al., 2008; Hainzl et al., 2016) or aseismic creep on faults (Neunhöfer and Hemmann, 2005; Passarelli et al., 2015).

The Vogtland area has experienced various swarm activities during 1985-1986, 1997, 2000, 2008, 2011, 2014, 2017 and 2018 (Fischer et al., 2014; Hainzl et al., 2016; Krentz, 2019). The region has been subject of many seismological studies, mainly based on the observations of the West Bohemia seismic network (WEBNET) operating in Czechia territory, and is selected for developing an interdisciplinary observatory using shallow drilling and small aperture seismic arrays (Dahm et al., 2013)

Our application is focused on a swarm in 2008 which occurred in the Nový Kostel (NK) fault zone in Vogtland (see Fig. 3.4), a planar structure steeply dipping westward (Fischer and Horálek, 2003; Fischer et al., 2010), oriented nearly S-N, and reactivated at a depth between 6 and 11 km. According to (Horálek and Šílený, 2013), the focal mechanisms of the swarm activity which occurred close to Nový Kostel are mostly oblique-normal and oblique-thrust types but the oblique-normal faulting predominates. The oblique-normal events have predominant strikes of 160° to 170°, dips of 72° to 80° and rakes of $-28°$ to $-38°$ whereas the oblique- thrust events show mainly strikes of 355° to 360°, dips of 80° to 85° and rakes of 35° to 40°.

The waveforms of swarm earthquakes in Vogtland typically display distinct direct P- and S-waves followed by high frequency coda waves generated at crustal interfaces and at small-scale inhomogeneities. The secondary phases are proved to be sensitive to the focal mechanisms and are different for each station of the local network and can be used to identify and image the prominent crustal discontinuities (Hrubcová et al., 2013, 2016) (See Fig. 3.17a for examples of waveforms). The 2008 swarm activity was also recorded by a small-aperture seismic array (Rohrbach array) operated by the University of Potsdam from 19 October 2008 until 18 March 2009 (see Fig. 3.4) (Hiemer et al., 2012; Rößler et al., 2008). The analysis of high-quality data recorded in Rohrbach array indicated near-vertical ray incidence of P- and S-phases, while calculated back azimuths showed 30° deviation from theoretical values, due to the structural inhomogeneities in the propagating zone (Hiemer et al., 2012). A well-located catalog based on relative master event location using precise arrival-time picking of WEBNET seismic network is used (Fischer et al., 2010) for both experiments. The original catalog includes 5679 events in the magnitude range of $M = -1$ up to $M = 3.5$. For our test 570 events of the total catalog are selected based on the availability of waveforms in station $V02$ of the Rohrbach array. Figure 3.5 shows the relative location of hypocentral parameters of those selected events. The overall reported location error in the catalog versus magnitude of events is plotted in Fig. 3.6. The overall error is calculated from root mean square of errors in latitude, longitude and depth estimation. To calculate synthetic waveforms, we assigned realistic hypothetical focal mechanisms to each event in the catalog. The hypothetical mechanisms are generated by assuming 5° variation in each of the rake, strike and dip values, reported for two possible

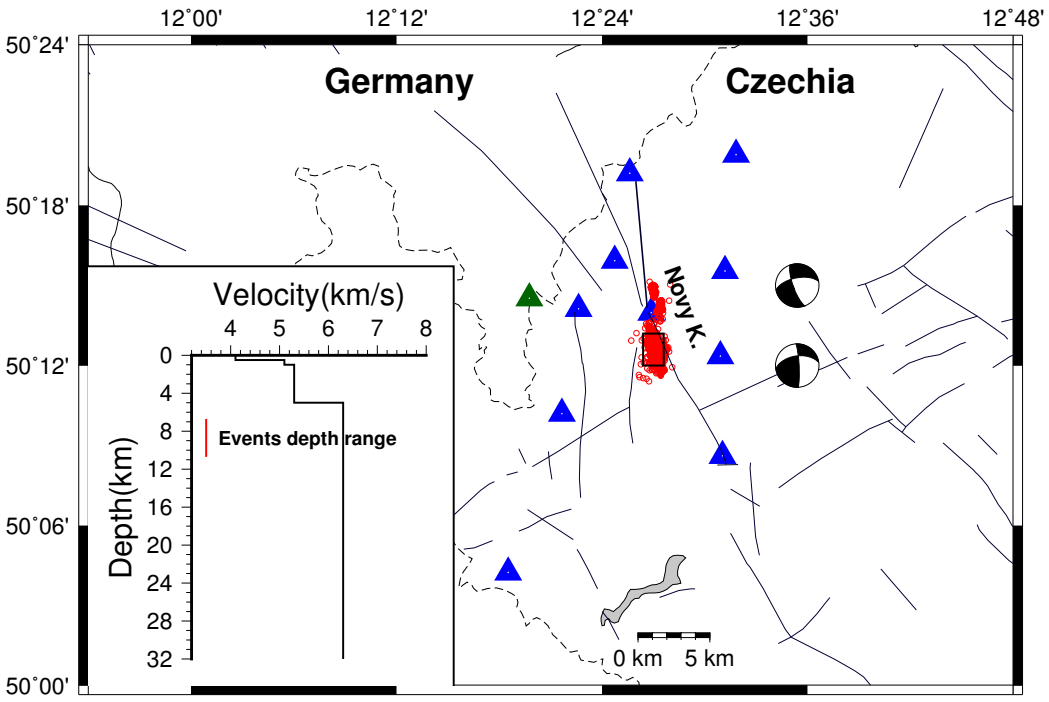


Figure 3.4: The Vogtland area which is an intraplate swarm region in German-Czechia border is shown. The green triangle shows the location of Rohrbach receiver array, and blue triangles represent the WEBNET stations. Red circles are epicenter of 2008 Nový Kostel swarm activity, and the black rectangle shows the search area for an optimized source array. Two predominant focal mechanisms are shown. The inlay depicts the P-phase velocity model used to generate synthetic seismograms.

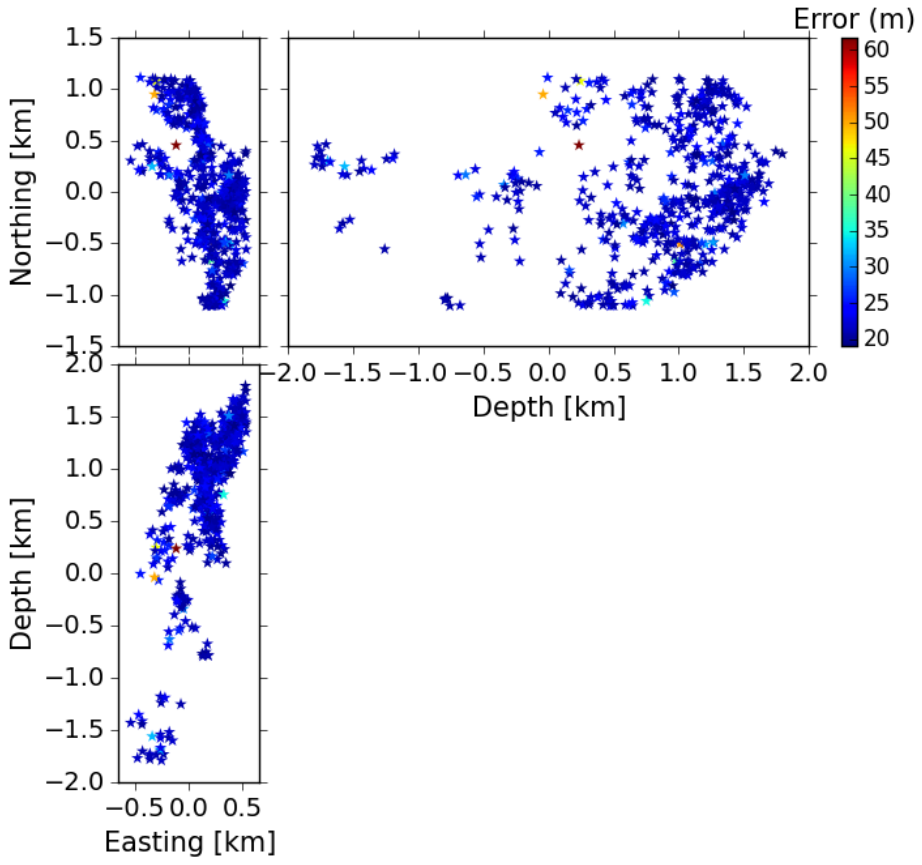


Figure 3.5: Distribution of the events relative locations (stars) and their location errors (color). The geometrical center is set to zero in all planes, and the positive direction of the depth axis directs to the shallower events.

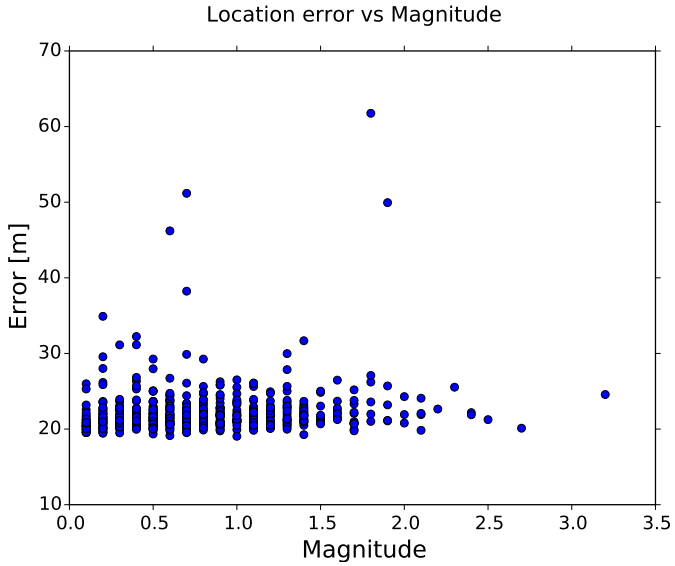


Figure 3.6: Location precision for different magnitudes. The total error is calculated assuming the latitude, longitude and depth error reported for each individual event in the catalog.

predominant clusters in NK zone. We used a velocity model with a strong contrast at depth of 5 km, to produce converted phases (such as ps- and sp-phases). The velocity model is shown in Fig. 3.4. Random uniform noise is also considered for the individual synthetic traces. Both synthetic and real waveforms are band-pass filtered in a frequency range which is selected to ensure good resolution and reasonable small aliasing of final results, given the magnitude and epicentral distance range of events and the visibility of the scattered phases on signals.

We use this region as an example and ask the question if the swarm catalog can be used to systematically search for the source array elements to be optimally used for further seismological studies. Using the WEBNET stations distributed at different backazimuths to the sources can help to study lateral variations of the crustal discontinuities, while Rohrbach array can help to apply the double array method.

3.4 Synthetic test

Using the real catalog of swarm events (including 570 events), assuming realistic random focal mechanisms and an appropriate velocity model, realistic synthetic seismograms at one station of the Rohrbach array are simulated by means of pre-calculated Green's function databases and efficient storage and accessing tools (Heimann, 2017). The generated synthetic signals are used during the optimization process to evaluate the SA beamforming, P-phase similarity and the energy of coda phases. The search algorithm is initiated to find a model, i.e. an SA including 20 sources. The process starts by defining the weighting coefficients of the sub-objective functions (eq. 3.14) and formulating the final scalar function. Following the procedure described in (Karamzadeh et al., 2019a), we have to run the optimization algorithm using the individual objective functions to calculate the normalization and relative weighting coefficients. So, the optimization algorithm is operated given the defined objective functions, f_1, f_2, f_3 and f_4, while in each operation for the tested models, the values of the other functions are evaluated as well. The maximum and minimum value of each objective function used for normalization are listed in Table 3.1, and the values of objective functions to define the weighting coefficients are given in Table 3.2. For instance, p_1 is a model obtained by optimizing f_1, and the optimal/minimal value of f_1 is 0.33, the values of other three objective functions for p_1 are evaluated and listed in the first row of the table. Using values of Table 3.1, values of Table 3.2 are scaled and normalized,

Table 3.1: The maximum and minimum value of each sub-objective function for the given catalog (including 570 events) used to define the normalization coefficients

Type	f_1	f_2	f_3	f_4
Synthetic	0.33-0.5	0.02-0.1	12.-32.6	1.89 - 5.83
Real	0.36 - 0.65	0.15-0.61	12.-31.3	0.36 - 2.7

Table 3.2: The value of sub-functions used to define weighting factors

	Type	f_1	f_2	f_3	f_4
p_1	Synthetic	0.33	0.11	14.8	4.5
p_2	Synthetic	0.51	0.02	14.1	4.5
p_3	Synthetic	0.43	0.121	12.1	4.7
p_4	Synthetic	0.5	0.2	14.6	1.9
p_1	Real	0.36	0.39	15.2	1.15
p_2	Real	0.51	0.15	15.7	2.65
p_3	Real	0.63	0.56	12.0	1.33
p_4	Real	0.53	0.45	14.2	0.36

then following the procedure described in Karamzadeh et al. (2019a) (See Appendix A.1) the final objective function is formulated as:

$$F = 0.14\hat{f}_1 + 0.2\hat{f}_2 + 0.18\hat{f}_3 + 0.48\hat{f}_4, \tag{3.18}$$

Where $\hat{f}_i$ indicates the normalized and scaled version of f_i.

3.4.1 Results of synthetic data test

Using the defined final objective function (eq. 3.18), the optimization process gives us the final optimized SA which is a model that is supposed to fulfill all 4 defined criteria. In the ideal case, it is expected that the final solution minimizes all sub-functions f_1-f_4 simultaneously. However, in practice there might be an inherent trade-off between sub-functions, and such an ideal solution might not exist. For example, a model with minimum overall location error is not necessarily the model with highest waveform similarity and a model with highest waveform similarity might not have well distributed elements in 3-D space to ensure the highest possible slowness vector estimation resolution. However, by using the final objective function (eq. 3.18), which is a weighted sum of all sub-functions, we try to find a model as close as possible to an ideal solution for each sub-function.

Fig. 3.7 shows variation of individual sub-functions (a)-(d) as well as the final function (e) during the optimization process. In this example, the first 300 models are selected in random and show higher variation range in all 5 plots. The sub-function defined based on the array beam power f_1 (Fig. 3.7a) shows steady improvement by increasing the iteration number, as it depends on the geometry and relative location of a tested model's element. In the applied optimization method, the radius of the random walk decreases by the iteration, so variation of the geometry gets steadily smaller during the process. The other sub-functions experience local fluctuations as the corresponding properties can even change from one event to the adjacent event. For example, according to Fig. 3.5 the location error is distributed heterogeneously. Most of events have similar location error (20-25 m), but events with large errors up to 60 m are existing in the catalog and can be included in the trial models during the optimization process. Fig. 3.8 shows the geometry of the final optimized SA. The waveforms of the SA are depicted in Fig. 3.9a, where the theoretical ps- and sp-phases onset times are marked on individual traces.

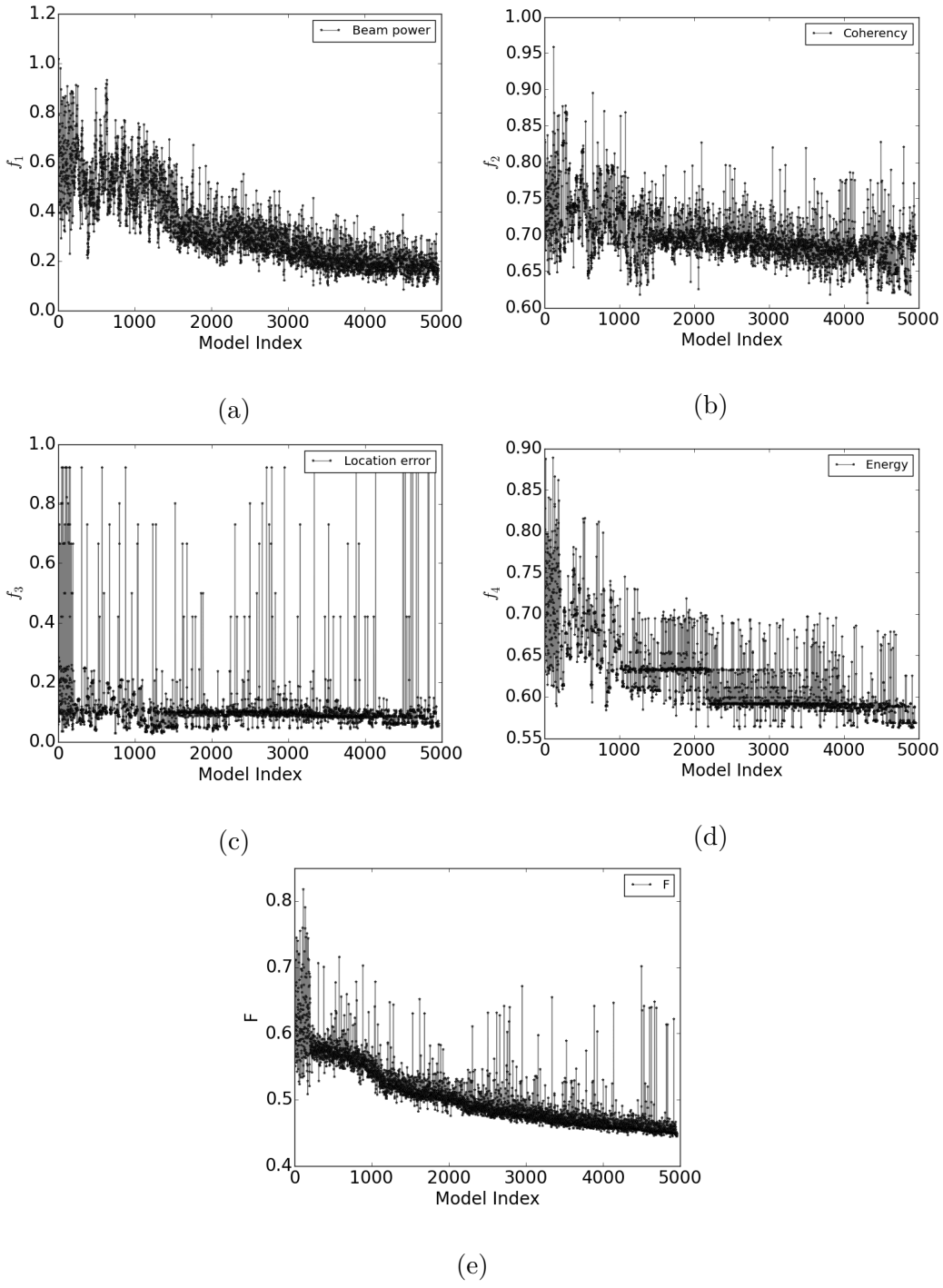


Figure 3.7: Variation of sub-functions and final objective function, for random models (20-element source arrays) selected from the given catalog and tested during the optimization process in synthetic test. (a)-(d) are related to the f_1 array beam power, f_2 waveform coherency, f_3 location error, and f_4 energy of the coda phases, respectively, and panel (e) shows the final objective function defined by (3.18). The first 300 models are randomly chosen, while the later models are generated by random walk in the neighborhood of the guiding models as introduced in section (3.2.2).

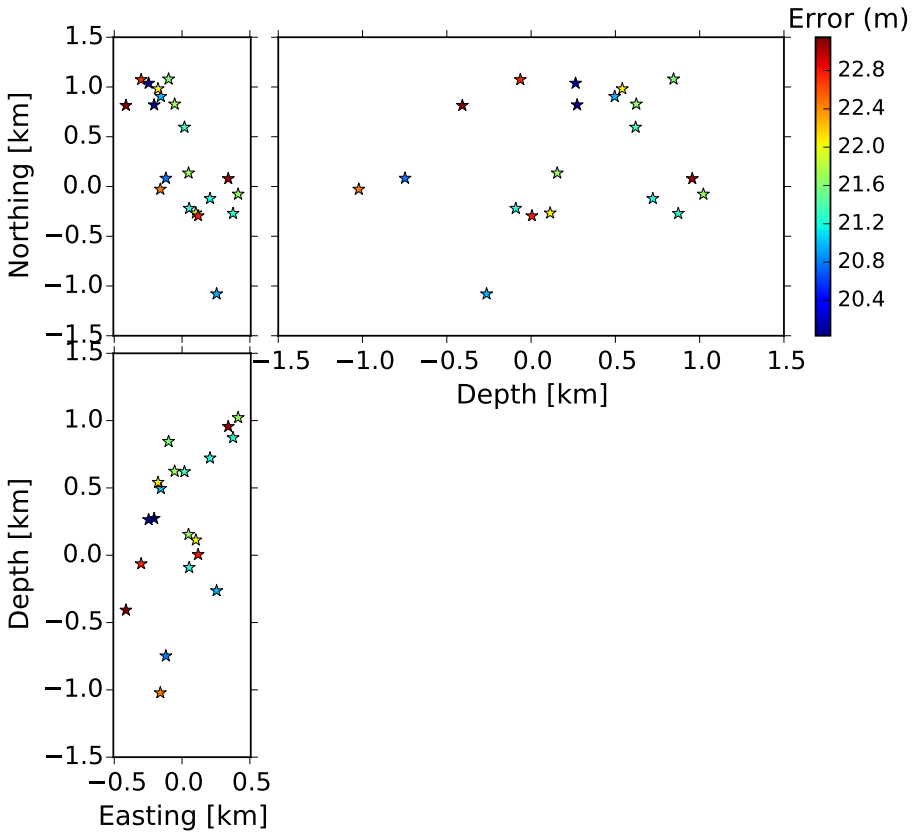
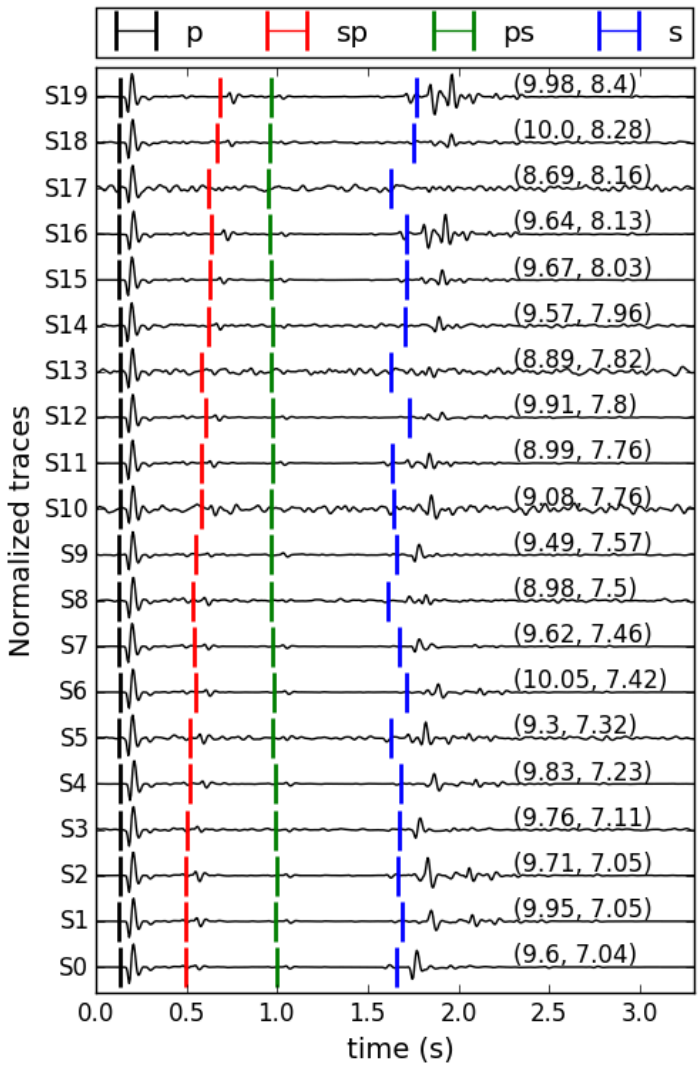
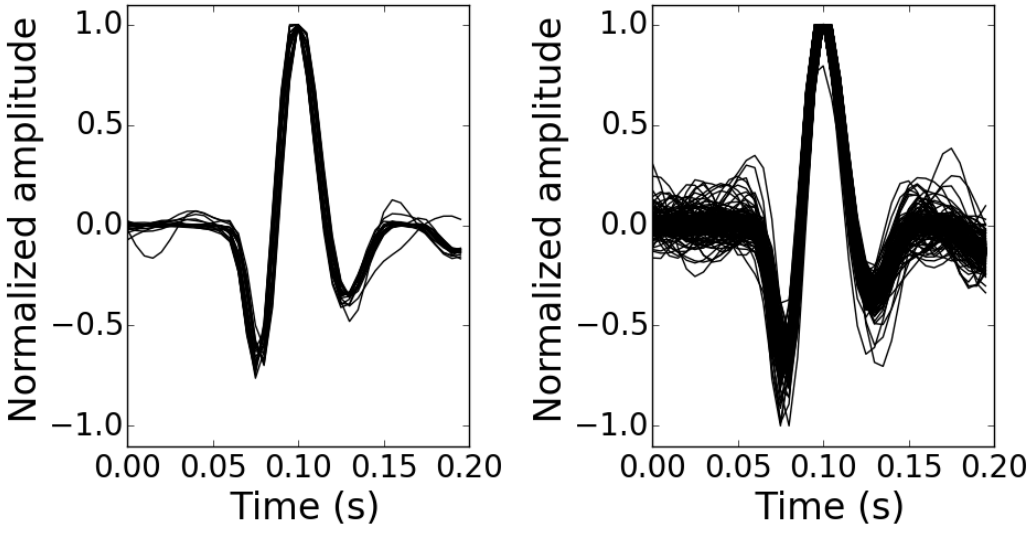


Figure 3.8: Geometry of an optimized 20-element source array obtained from synthetic test. See the Fig.(3.5) caption for further explanations.



(a)



(b) (c)

Figure 3.9: (a) Full synthetic waveforms of the optimized source array (see Fig. 3.8) aligned according to the P-phases arrival times. Arrival times of directed p- and s-phases and sp- and ps- converted (scattered) phases are indicated by markers. The pair of numbers on each trace are epicentral distance and depth given in km. (b) P-phase waveforms of the optimized source array. (c) P-phase waveforms of all events in the used catalog. All traces are band-pass filtered (2-20 Hz).

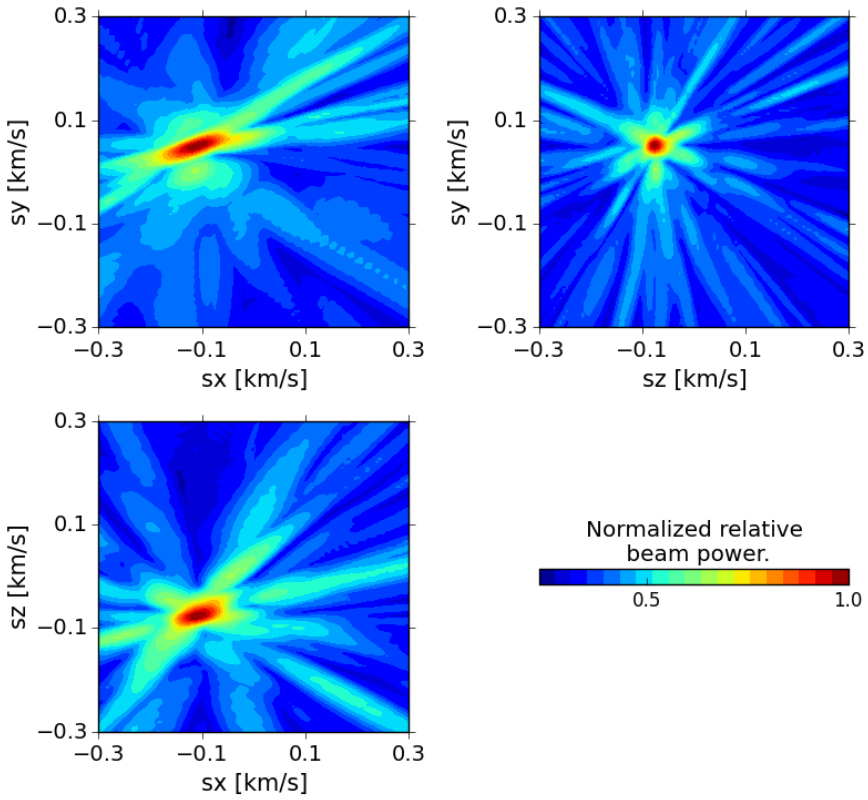
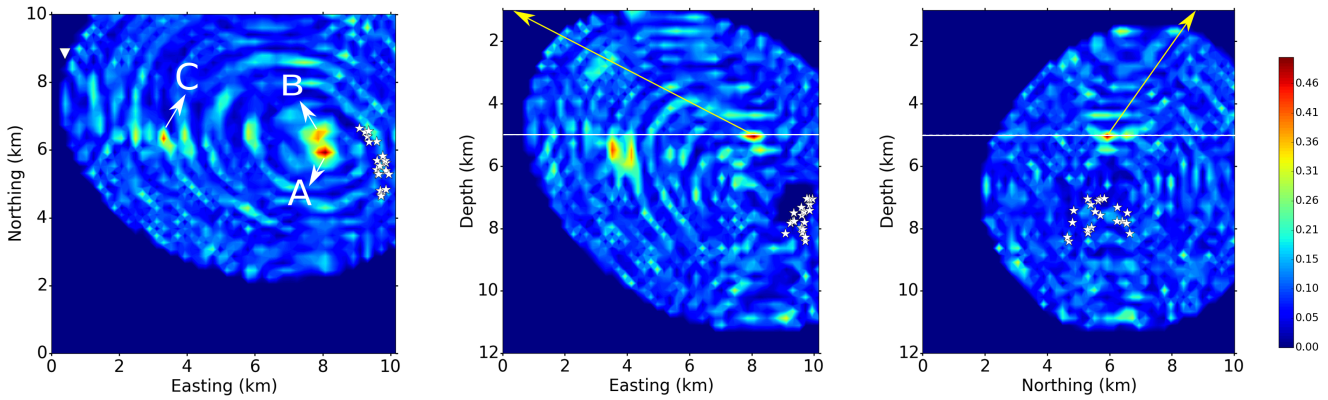
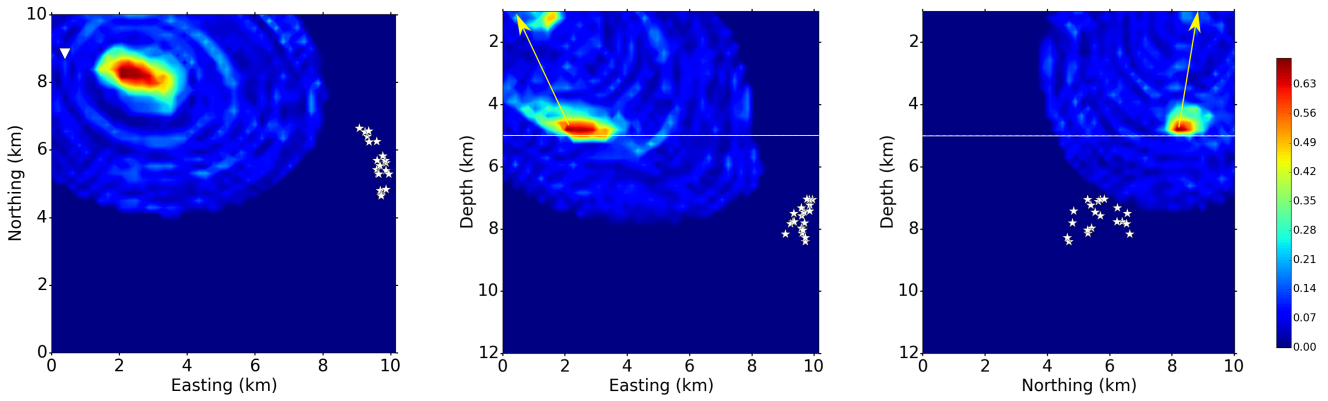


Figure 3.10: The 3-D array beam pattern for the optimized SA calculated using synthetic waveforms (see Fig. 3.8 for array geometry).



(a) (b) (c)

Figure 3.11: The 3-D cross-sections of the semblance values including its maximum value. Semblance is calculated using the time windows of 0.1 s length, extracted from the source array synthetic recordings, according to the sp-phase theoretical arrival time. White stars show the location of the sources and the white triangle on (a) is the projection of the receiver at the surface. The horizontal white lines on depth axes (b and c) indicate the depth of the velocity interface at 5 km. The yellow arrow in (b) and (c) plotted from the location of the maximum semblance towards the location of the station at the surface. The traces are bandpass filtered: 2-20 Hz.



(a) (b) (c)

Figure 3.12: The 3-D cross-section of the semblance values including its maximum value, calculated using synthetic traces and according to the ps-phase travel time, see the Fig. 3.11 caption for more explanations.

The waveforms are sorted according to the depth. The values of depth and source-receiver horizontal distances are written on each trace as well. In Fig. 3.9b P-phases waveforms are plotted in the shared frame to compare the similarity and quality of the waveforms visually. Fig. 3.9c shows the P-phases waveforms of all events in the catalog, including the waveforms of the optimized SA. Comparing Fig. 3.9b and Fig. 3.9c indicates that the optimization algorithm could successfully select a similar subset of events. In Fig. 3.10 the related array beam power is depicted. Fault geometry imposes a narrower distribution of events in east-west direction, accordingly the maximum achievable source array apertures in east-north and east-depth plane are smaller than the maximum achievable aperture in north-depth plane, which results in the lower slowness resolution in both planes compared to the east-depth plane.

Using the final optimized source array, the procedure described in section 3.2.2 is employed to locate possible scatterers in the wave propagation media. The calculated semblance value for all trial grid points, using the sp- and ps-phase arrival times, are plotted in the cross-section plots shown in Figs. 3.11 and 3.12 respectively. In these plots, white stars represent the location of the source array, the white triangle in the horizontal northing-easting plane (plot a) shows the relative location of the receiver at the surface. In plots (b) and (c) the horizontal white lines represent the location of the interface at 5 km depth in the velocity model, the yellow arrows are plotted from the grid point showing the maximum semblance and direct to the relative position of the receiver at the surface. In each cross-section plot, the grid point with maximum value of the semblance shows the location of the point scatterer related to the converted phase origin.

The calculated depth and location of the scatterers in horizontal plane are in accordance with the theoretical ray paths of sp- and ps-phases for this geometry. Fig. 3.13 illustrates the ray path of p to s and s to p converted phases at the velocity interface at 5 km depth for the source-receiver geometry similar to our example, that is a source located at 8 km depth and a receiver at 10 km epicentral distance. For this setup, the impinging points of the sp-phase ray on the interface is closer to the source than to the receiver, while the impinging point of ps-phase ray is closer to the receiver than to the source. This point is visible in the semblance patterns shown in Figs. 3.11 and 3.12, while the depth of the maximum semblance is very close to the 5 km depth, that is 4.95 km for the sp-phase and 4.85 km for the ps-phase. For the sp-phase, in the horizontal cross-section plot, i.e. Fig. 3.11a, another peak (label B) is appeared very close to the main peak (label A). It is due to the fact that the ray path of each event can be slightly different from the others and the impinging points of all rays to the interface surface may not be at the same location. The reason for another localized high semblance in this plane (label C) can be due to the existence of other low energy coherent waves (e.g. ps-phase) with

similar slowness vector as the sp-phase which is targeted in this example.
The sp- and ps-phase beam traces are plotted in Fig. 3.14, labeled as "Beam 1" and "Beam 2", respectively.

3.5 Application on real data

Using the given catalog and the related waveforms, the algorithm is initiated to search for an optimized 20-element SA at one station of the Rohrbach array (Rößler et al., 2008), the same as used for the synthetic test. Similar to the experience with the synthetic data in section 3.4, we use the values reported in Table 3.1 and 3.2, and define the final objective function as:

$$F = 0.17\hat{f}_1 + 0.175\hat{f}_2 + 0.34\hat{f}_3 + 0.315\hat{f}_4, \tag{3.19}$$

Where $\hat{f}_i$ indicates the normalized and scaled version of f_i. The dynamic range of the array beam power f_1, is comparable for both experiments (Table 3.1). However, the waveform similarity f_2 is scaled up to 7.5 and 6.1 times for the lower and upper bands of real data compared to the synthetic data, which indicates that the P-phase waveforms of the real test are more diverse compared to the synthetic test. The fact that a uniform magnitude (ML=0-3.2) has been used in the synthetic tests to generate the earthquake catalog might partially explain this observation. Additionally, the noise level in the synthetic test has less variation for different events compered to the real waveforms. We have not excluded larger magnitude events from the catalog to minimize the effect of later source extension on waveform dissimilarity, but preferred to keep them in the catalog, and let the optimization algorithm to avoid automatically the larger events from the optimized SA.
As we expect, because the location error for both cases are identical, the dynamic range of f_3 is almost the same as well (Table 3.1). In particular the lower band of the dynamic range is identical for both tests, indicating that the optimization algorithm successfully could select the best event ensembles in terms of the location error.
In the case of the f_4 criterion, which is to measure the energy of the coda waves, the values of both limits of dynamic range are measured lower for the real test compared to the synthetic test. This indicates that the signal-to-noise ratio of coda phases in the synthetic test is higher than for the real data, whereas the ratio of upper-to-lower value in real case is 2.5 times higher than the same value measured in the synthetic test. This may be explained by pairs of time overlapping events, for which the P-phase of the second event could be misinterpreted as a secondary phase of the first event in each pair. In synthetic traces, such pairs of events are stored in separated waveform files, and the related waveforms are not overlapping in time.
It is noteworthy that the effect of some different dynamic ranges for real and synthetic tests are eliminated by the normalization process, as each criterion is normalized regarding to its individual measured dynamic range.

3.5.1 Results of real data application

Having defined the final objective function for the real waveforms recorded at one single station of the Rohrbach array (see Fig. 3.4), the optimized real SA is specified. Fig. 3.15 shows the values of all sub-functions in panels (a)-(d) and the final function in (e), which are evaluated during the optimization process. Fig. 3.16 shows the geometry of the final optimized SA. The waveforms of the optimized SA are depicted in Fig. 3.17a and are sorted according to the depth. The values of depth and source-receiver horizontal distances are indicated on each traces. In Fig. 3.17b the P-phases are plotted in a shared frame to compare the similarity and quality of the waveforms. The P-phase waveforms of all events in the catalog are depicted in Fig. 3.17c.

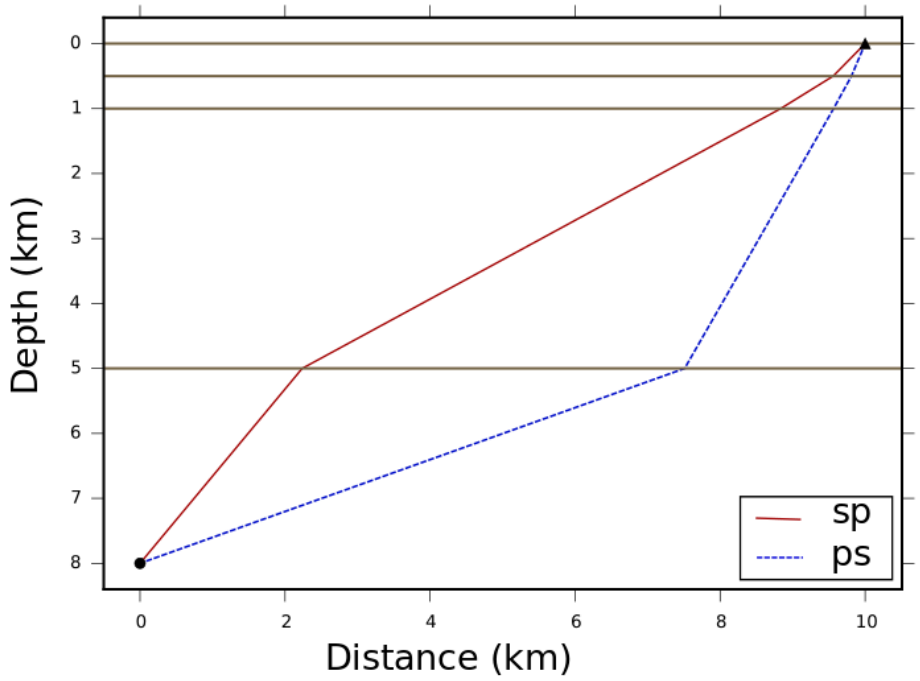


Figure 3.13: Schematic ray path of sp- and ps-phases calculated using the given 1-D velocity model used for the synthetic test in this study. The ray paths are calculated by using the Cake tool implemented in Pyrocko package (Heimann, 2017). The horizontal lines represent the velocity layers. The source is at 8 km depth and epicentral distance is 10 km.

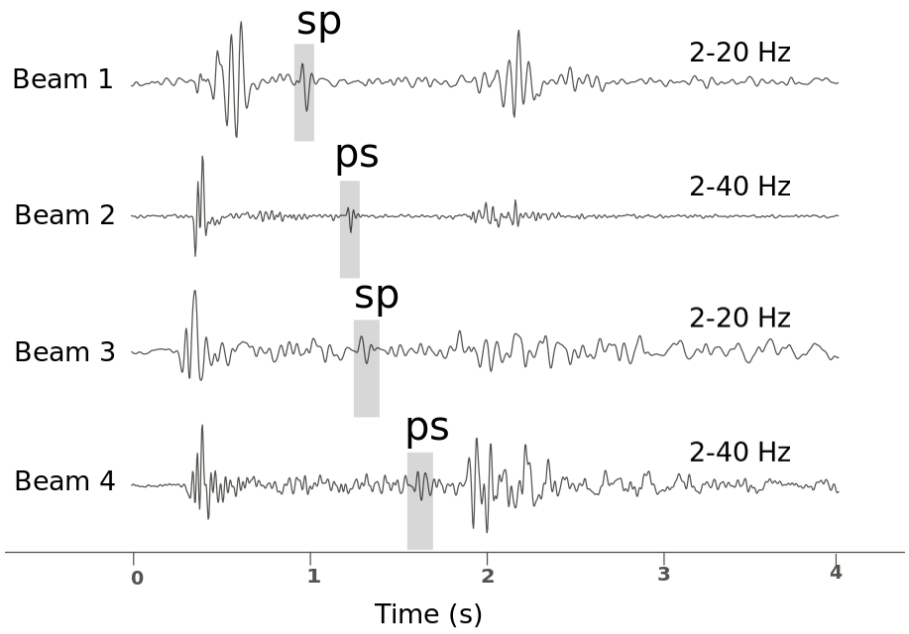


Figure 3.14: Source array beam traces calculated using time delays according to the grid points with maximum semblance values. Beam 1 and Beam 2 are computed using synthetic traces and Beam 3 and Beam 4 are related to the real data.

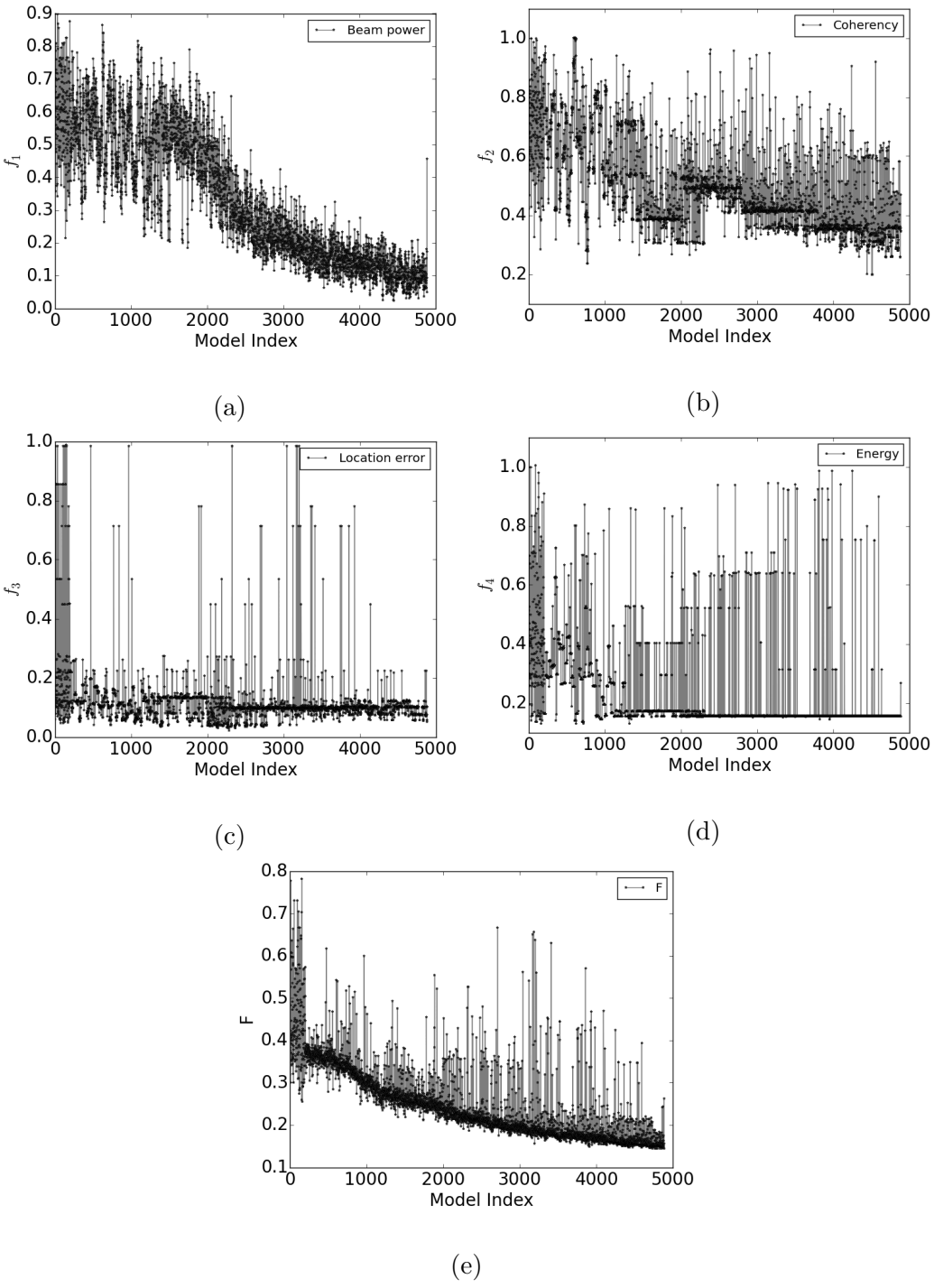


Figure 3.15: Variation of the sub-functions $f_1 - f_4$ and the final objective function (F), for random models selected from the given catalog and tested during the optimization process using the real waveforms. See the Fig. 3.7 caption for more explanations.

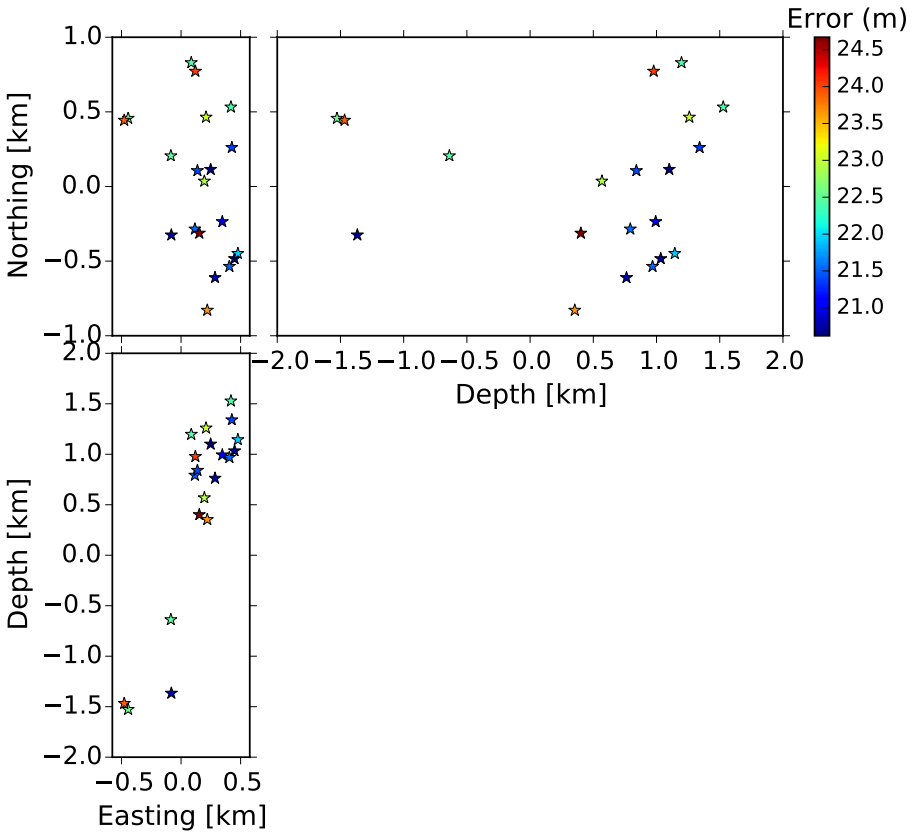


Figure 3.16: Geometry of an optimized 20-element source array obtained from real data. See the Fig. 3.5 caption for further explanations.

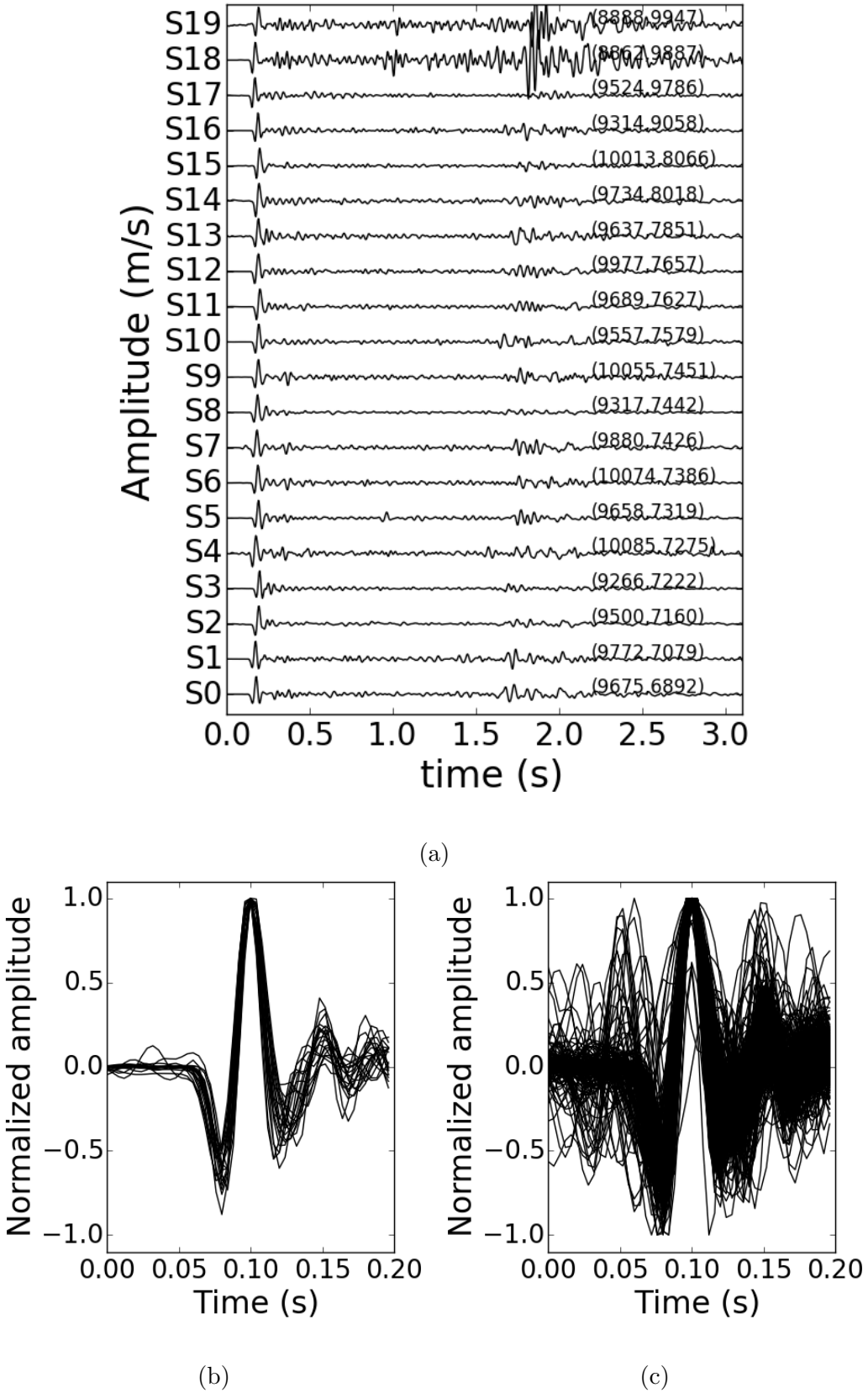


Figure 3.17: (a) Full real waveforms of the optimized source array (3.16) aligned according to P-phases arrival times. The pair of numbers on each trace are epicentral distance and depth. (b) P-phase waveforms of the optimized source array. (c) P-phase waveforms of all events in the used catalog. All traces are band-pass filtered (2-20 Hz).

The array beam power in 3-D slowness space for the SA geometry is shown in Fig. 3.18. Fig. 3.19a shows waveforms collected using a single source and 10 stations of the receiver array (Rohrbach Array), and the P-phases are extracted and plotted in a single frame in Fig. 3.19b. Comparing the P-phase similarity of a source and receiver array (Fig. 3.17b and Fig. 3.19b) we can conclude that waveforms of an SA recorded in a single station can show higher similarity compared to the waveforms of a single source recorded in a receiver array. In other words near receiver structure differences are in this case stronger than the near source structure differences. The value of P-phase cross-correlation for the receiver array is 0.36, whereas for the source array waveforms shown in Fig. 3.17c this value is 0.87.

Using the final optimized source array, according to the procedure described in section 3.2.2, similar to the synthetic test case, we tried to locate possible scatterers existing in the wave propagation media. The same velocity model as was used in the synthetic test case, employed in this test to predict arrival times of ps- and sp-phases assuming trial point scatterers in the velocity model. The calculated semblance value for all trial grid points are plotted in the cross-section plots shown in Figs. 3.20 and 3.21 for sp- and ps-phases, respectively. In these plots, white stars are locations of source array elements. The white triangle on horizontal Northing-Easting plane, i.e. plot (a), shows the relative location of the receiver at the surface. In plots (b) and (c) the yellow arrows are plotted from the grid points showing the maximum semblance and direct to the position of the receiver at the surface. The maximum value of the semblance in each cross-section plot, indicates the location of the scatterer. In both cases, Figs. 3.20 and 3.21, clear peaks in semblance map are visible, however the indicated scatterers are localized in different location. The sp- and ps- beam traces are plotted in Fig. 3.14, labeled as “Beam 3” and “Beam 4”, respectively.

3.6 Discussion and conclusions

In this paper we presented a method to search for seismic source array elements, given a well-located earthquake swarm catalog. The proposed method benefits from an optimization scheme, which allows to evaluate a number of customizable objective functions quantitatively. We formulated the preconditions to make an SA, by defining 4 objective functions, f_1 - f_4. The synthetic test and the test with real data proved that the proposed method by using the suggested optimization approach is capable to find an optimized SA according to the defined objective functions. In addition, the values of the individual sub-functions at the final optimized SA, indicates that our weighting strategy performs well.

One of the key conditions to define an SA is the accuracy of source locations, which is formulated in f_3. A test is performed to see how the location error propagates into the final outcomes of the source array beamforming analysis, i.e slowness vector estimation. First, 3-D slowness vector elements s_x, s_y, s_z, for a 20-element source array are estimated using real waveforms according to equation (4.2) assuming the exact locations of the sources are given. The corresponding 3-D array beam patterns are shown in Fig. 3.22 and the slowness vector is indicated by red stars on 2-D slowness planes. Then, assuming arbitrary error vectors for elements of SA, they are shifted to new locations. Afterwards, the time shifts (equation 4.2) are recalculated for the new locations and the slowness vector is updated and compared with the value obtained initially without considering the location errors. The circles shown in Fig. 3.22 indicate the updated slowness values for the perturbed SAs. Components of the tested error vectors are determined randomly, assuming an individual Gaussian distribution for each component, i.e. latitude, longitude and depth, with a specific standard deviation. The colorbar in Fig. 3.22 shows the maximum value of the assumed standard deviation for each perturbed model. Histograms depicted in Fig. 3.23 illustrate the values of deviations obtained for each component. According to the results, for the tested source-receiver geometry and SA configuration, as it

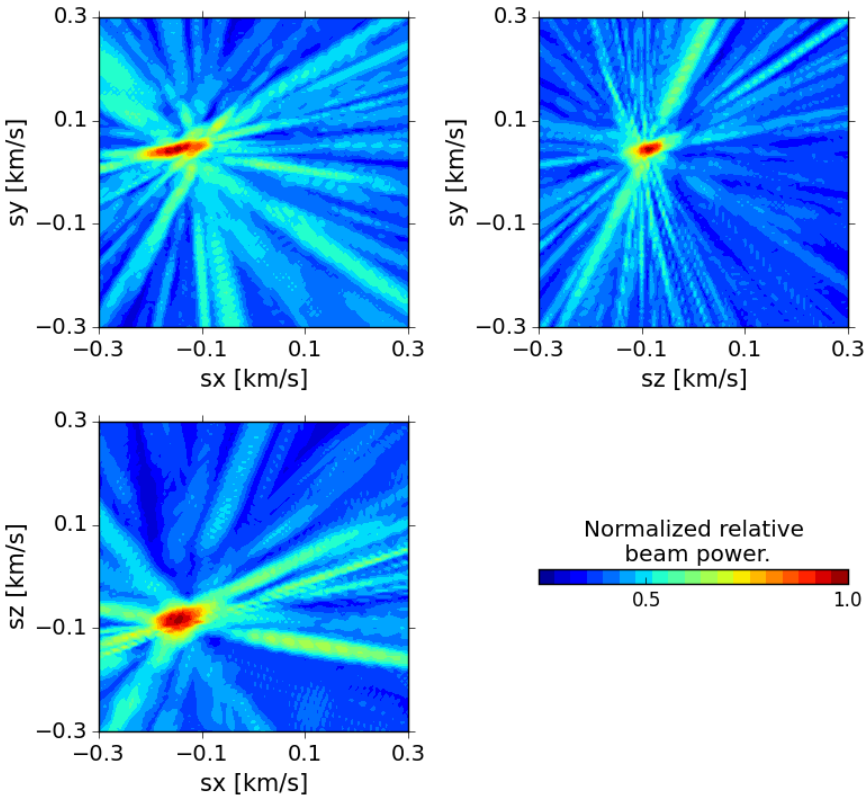
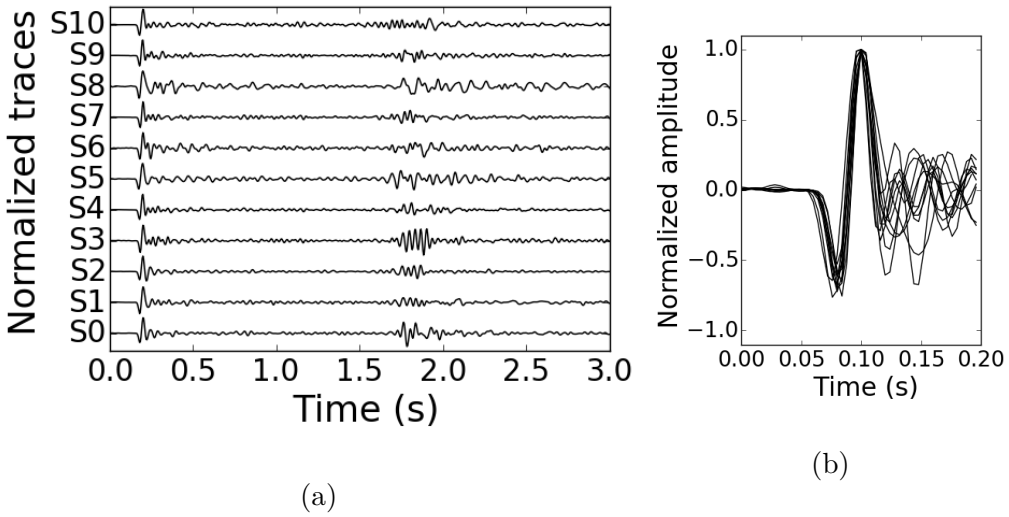
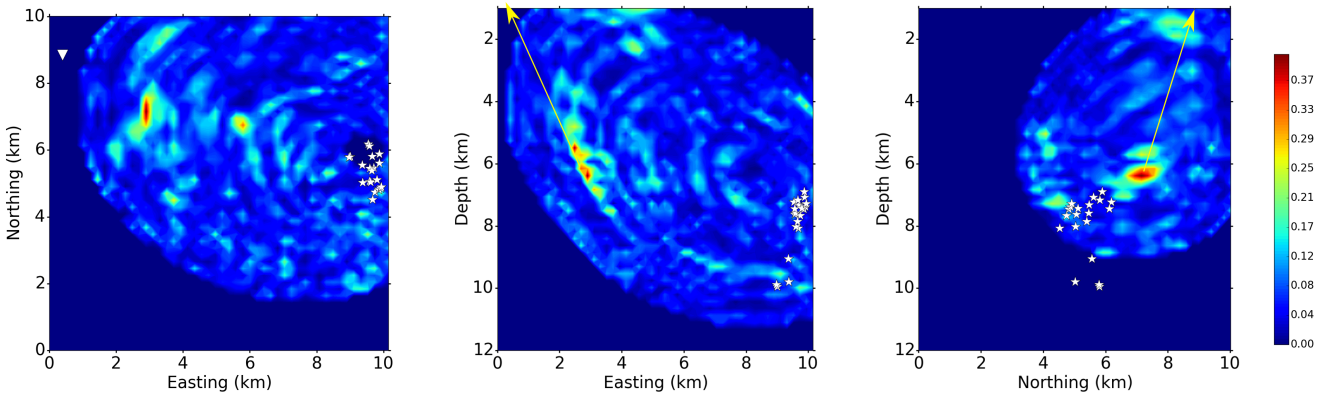


Figure 3.18: Array beam pattern for an example SA using real traces (see Fig. 3.16 for the array geometry)



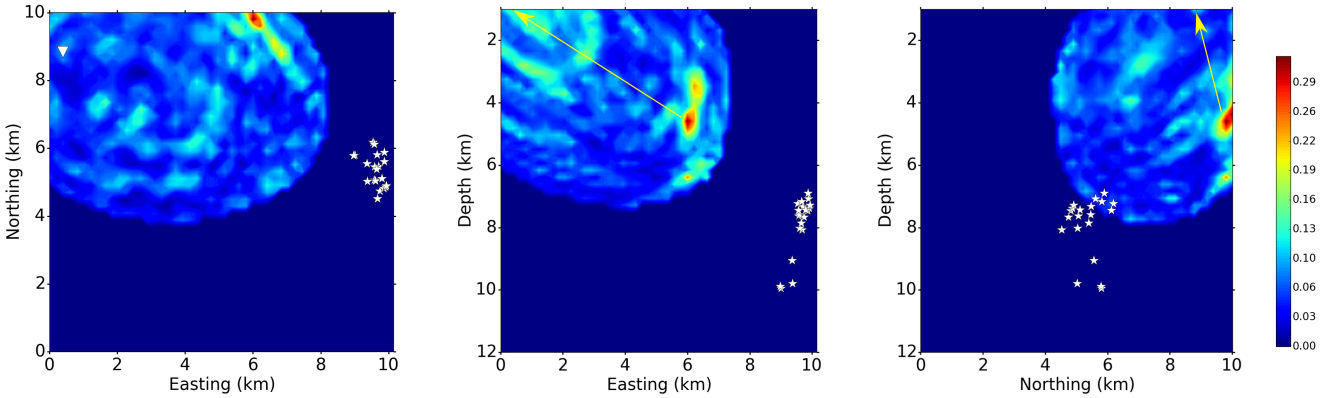
(a) (b)

Figure 3.19: (a) Waveforms recorded from source $S11$ (Fig. 3.17a) at all 10 stations of the Rohrbach array. (b) P-phase waveforms.



(a) (b) (c)

Figure 3.20: Semblance patterns calculated using the time windows extracted from the real source array recordings, according to the sp-phase theoretical arrival times. The location of maximum semblance in each plot indicates the location of the scatterer (see the Fig. 3.11 caption for more explanations.). Array beam trace for the time shifts related to the maximum semblance location is plotted in Fig. 3.14 and the related phase is highlighted on the trace named "Beam 3".



(a) (b) (c)

Figure 3.21: Semblance patterns calculated using the time windows extracted from the real source array recordings, according to the ps-phase theoretical arrival times. The location of maximum semblance in each plot indicates the location of the scatterer (see the Fig. 3.11 caption for more explanations.). Array beam trace for the time shifts related to the maximum semblance location is plotted in Fig. 3.14, and the related phase is highlighted on the trace named "Beam 4".

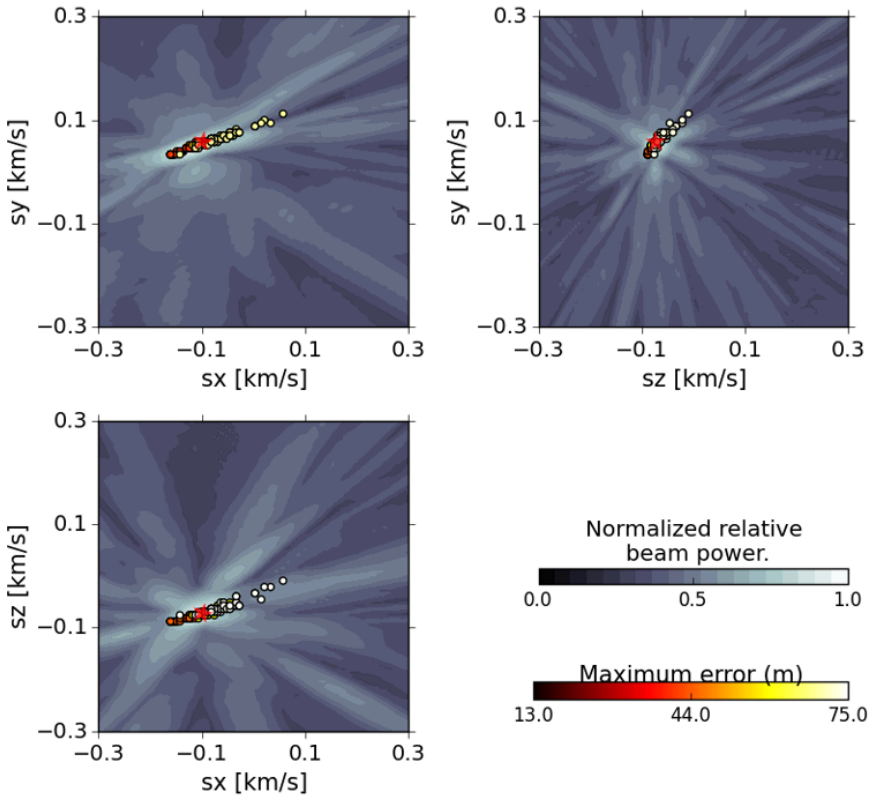


Figure 3.22: Location error propagation in slowness vector components. The array beam power depicted in the background is calculated assuming zero error. Each circle represents a slowness vector assuming a specific error in source array elements location. The colorbar shows the maximum error in each tested case.

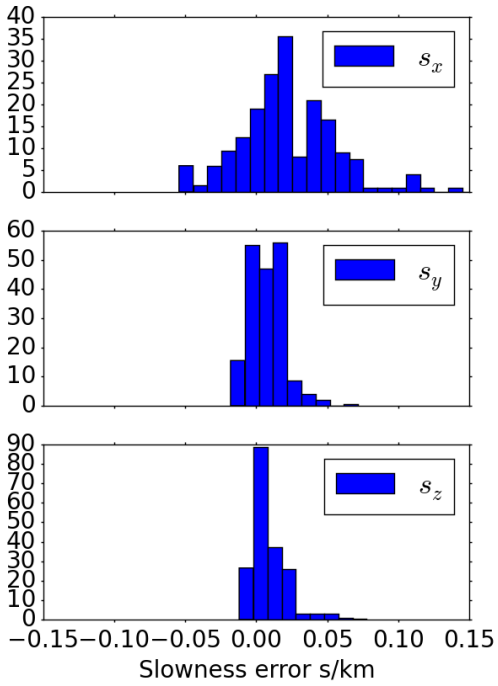


Figure 3.23: Deviation of slowness vector components calculated after perturbation of the array elements. Values of the maximum errors are shown in Fig. 3.22.

is expected, by increasing the value of location error, the error of estimated slowness vector increases, so that assuming a maximum 75 m location error, up to 0.15 s/km deviation in S_x component of slowness vector is likely to happen, while for the other components, the value of deviation is less than 0.07 s/km. It is interesting that the deviating slowness values are distributed around the main peaks and show correlation with the array beam power patterns of the original error-less SA. Accordingly, in the $s_x - s_z$ plane, where we find the sharpest central peak, the lowest deviation of slowness vector component is observed. In other words, for the smaller aperture source arrays, more precise locations are necessary to produce reliable final results. Larger aperture source arrays are more robust against location errors.

The results of the performed synthetic test to localize the point scatterers related to the observed sp- and ps-phases, demonstrated that if the coherency of the scattered phases are persistent over the SA elements, it is possible to image the scatterer location by measuring the semblance of the phases. However, even in a synthetic test, resolution of the resolved scatteres depends on the geometry of the experiment, that is the source array and receiver relative location and ray path of the phase. For instant, the semblance pattern calculated using ps-phase (Fig. 3.12) shows one distinct peak, but the peak is wider compared to the main peak observed for sp-phase (Fig. 3.11), that implies to a lower resolution in determining the scatterer location using ps-phase which is closer to the receiver than to the source region. This simple experiment demonstrated that to image different parts of an extended scatterer, or if there are many local scatterers, different phases and variable SA and receiver set ups are required. However in such cases, phase interference can violate the coherency of the phases which arrive in a specific time window.

In our experiments we did not include the error in origin time, as it was not available in our catalog. However, in case of origin time error, aligning the waveforms based on P-phase beamforming will eliminate the timing errors. In this case, usually the value of the relative slowness of the reflected phases are measurable.

According to the test using real data, there are indications for scatterers in the velocity structure in the tested area, northwest Bohemia/ Vogtland. The scatterer location revealed by using ps-phases in this study (Fig. 3.21) is in agreement with findings of Rüssler et al. (2009). Using the receiver array analysis (Rohrbach array) Rüssler et al. (2009) concluded that backazimuth angle of the converted phases can be deviated towards the north direction compared to the backazimuth angle of the direct phases, indicating the heterogeneity and inclined discontinuities along the ray path. Nevertheless, to get reliable image of the scatterers for such an application, we suggest using a collection of the optimal source arrays, by repeating the SA optimization algorithm several times. Considering other single receivers is also useful, to take into account variable SA to receiver geometries. Then, the search algorithm to discover the strong scatterers, using different SAs, should reveal consistent results in overall, to make reliable interpretation of the velocity structure.

For the examples given in this paper, the number of elements in SA is chosen to be 20 elements. In general more high quality array elements can increase the signal-to-noise ratio of the coherent scattered phases, and subsequently increase the resolution of the scatterer location. Larger aperture helps in slowness resolution, however the coherency of the scattered phase signals might decrease because of heterogeneous model and different ray path. The optimal number therefore depends on the quality of the available data, the wavelengths the study wants to resolve and the heterogeneity in and around the source volume. To define the optimal number of a source array for the specific catalog and region, the algorithm can be run with different numbers and the misfit functions can be evaluated and compared to find an optimal number.

SA technique can be applied whenever earthquake clusters are accurately located and contain events with similar waveforms. Often relative localizations are already available by applying

a waveform cross-correlation and double difference location techniques. Effects due to varying source mechanisms can possibly be taken into account by deconvolving the source mechanism from measured waveforms. Examples for potential datasets include, beside midcrustal swarms as in NW Bohemia, deep seismic nests, as often observed in slabs beneath orogens as in Buccaramanga in Colombia (e.g. Prieto et al., 2012) or the Hindu Kush (e.g. Kufner et al., 2017). Near-surface swarms of earthquakes are often observed for induced seismicity (e.g. Cesca et al., 2014) and offer another possible future application of SA. Volcanic intrusions induced seismicity (e.g. Cesca, 2019); is a third target of interest to apply the SA technique. The SA application examined here was aimed at the detection of scattered bodies near the source. However, SA is also interesting for illuminating structures at a greater distance from the source, e.g. using the so-called double beam technique, in which receiver and source arrays are processed simultaneously. For example, the slab location in a subduction zone with S-to-P phases (Kaneshima, 2019) could be investigated, the Moho depth determination with PmP phases (Hrubcová et al., 2013), or low shear wave provinces with P (Pdiff) phases (Frost and Rost, 2014).

The method may be further developed for future double array applications from vertical seismic profiles (Boué et al., 2013) using passive sources, or to define discrete, optimal source arrays in linear continuous seismic sensors as will be available for Distributed Acoustic Sensors (e.g. Jousset et al., 2018) and Ocean-Bottom Distributed Acoustic Sensing arrays (e.g. Williams et al., 2019).

3.7 Acknowledgements

We acknowledge the University of Potsdam especially Daniel Vollmer and Matthias Ohrnberger for the data at the temporary array installed in Vogtland/Rohrbach. We also acknowledge Potsdam Graduate School (PoGS) that supported the first author to finish this work.

A.1 Weighting coefficients calculation

A multi-objective optimization problem is defined in mathematical terms as:

$$min\ [f_n(\mathbf{m})]; \quad \mathbf{m} \in S \tag{A.1}$$

where $\mathbf{m}$ is the model space, $f_n(\mathbf{m})$ is a set of objective functions with index n in the number, and S is a set of constraints. One strategy to solve a multi-objective problem is combine all objective functions in a single scalar function using the weighting sum technique. To define the weighting coefficients, we can start by evaluating them between two functions according to (Karamzadeh et al., 2019a), and then increasing the number of functions step by step:

$$F_{1,2} = \gamma_1 f_1 + (1 - \gamma_1) f_2, \tag{A.2}$$

where, $0 < \gamma_1 < 1$ is the weighting factor, $F_{1,2}$ is the combined scalar function for f_1 and f_2. We can proceed the procedure using $F_{1,2}$ and f_3, to make F_{123}:

$$F_{1,2,3} = \gamma_2 F_{1,2} + (1 - \gamma_2) f_3, \tag{A.3}$$

and, to include f_4:

$$F_{1,2,3,4} = \gamma_3 F_{1,2,3} + (1 - \gamma_3) f_4, \tag{A.4}$$

$F_{1,2,34,}$ can be written as:

$$F_{1,2,3,4} = \gamma_3\gamma_2\gamma_1 f_1 + \gamma_3\gamma_2(1 - \gamma_1) f_2 + \gamma_3(1 - \gamma_2) f_3 + (1 - \gamma_3) f_4. \tag{A.5}$$

Finally, for n objective functions the final combined scalar function is formulated as:

$$F_{1,2,...,n} = \gamma_{n-1}\gamma_{n-2}...\gamma_1 f_1 + \gamma_{n-2}\gamma_{n-3}...\gamma_2(1-\gamma_1)f_2 + ... \tag{A.6}$$

$$+\gamma_{n-1}(1-\gamma_{n-2})f_{n-1} + (1-\gamma_{n-1})f_n.$$

Chapter 4

Small aperture arrays as a tool to monitor fluid injection- and extraction-induced microseismicity: applications and recommendations

This chapter is published in Karamzadeh et al. (2019b).

Abstract

The monitoring of microseismicity during temporary, human activities such as fluid injections for hydrofracturing, hydrothermal stimulations or waste water disposal is a difficult task. The seismic stations often cannot be installed on hard rock and at quiet places, noise is strongly increased during the operation itself and the installation of sensors in deep wells is costly and often not feasible. The combination of small aperture seismic arrays with shallow borehole sensors offers a solution. We tested this monitoring approach at two different sites, (1) accompanying a fracking experiment in sedimentary shale at 4 km depth, and (2) above a gas field under depletion. The small aperture arrays were planned according to theoretical wavenumber studies combined with simulations considering the local noise conditions. We compared arrays recordings with recordings available from shallow borehole sensors and give examples of detection and location performance. Although the high frequency noise on the 50 m deep borehole sensors was smaller compared to the surface noise before the injection experiment, the signals were highly contaminated during injection by the pumping activities. Therefore, a set of three small aperture arrays at different azimuths was more suited to detect small events, since noise recorded on these arrays is uncorrelated with each other. Further, we developed recommendations for the adaptation of the monitoring concept to other sites experiencing induced seismicity.

4.1 Introduction

Fluid injection and extraction operations, including those related to hydraulic fracturing, can trigger and induce seismicity through different physical processes, favouring shear failure along pre-existing faults or creating new fractures (Grigoli et al., 2017). Since the first documented cases of earthquakes triggered by fluid injections in the 1970ies (Healy et al., 1968), the number and types of industrial crustal fluid injections or extractions have steadily increased. In recent years, such types of operations were discussed in relation with the occurrence of significant earthquakes, which may lead to damage or change the seismic hazard with a possible feedback to the planning and development of injection projects. Examples include the geothermal

stimulation activities in deep hot dry rock environments (Grigoli et al., 2018; Deichmann and Giardini, 2009; Brodsky and Lajoie, 2013), the development of gas storage facilities (Cesca et al., 2014), waste water injections (Ellsworth, 2018; Tadokoro et al., 2000; Horton, 2012; Rubinstein et al., 2014; Hincks et al., 2018), or hydraulic fracturing operation in shale gas (Kim, 2013; Sasaki, 1998).

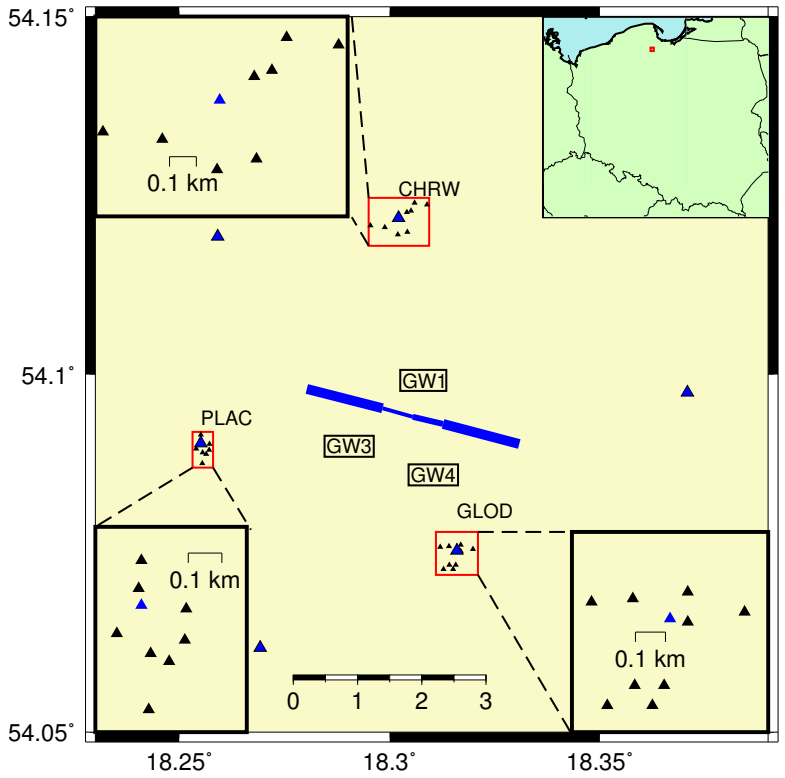
As a reaction, authorities in different countries have started to define regulations, which often specify criteria for the performance of a monitoring network and the magnitude of completeness. The aim is to be able to detect and locate micro-earthquakes before, during and after injection operations, in order to better understand changes in the seismic hazard and to develop traffic light systems for mitigating the consequences of induced seismicity (e.g. Green et al., 2012).

Monitoring of injection-induced micro-earthquakes in sedimentary basins is challenging due to high background noise level. Detections and locations of such microseismic events are key to judge the effectiveness of geomechanical operations, track the migration of the fracturing processes and ensure the preservation of reservoirs and the integrity of wells. A monitoring system should allow to detect, locate and characterize (1) microseismicity ($Mw < 0.5$) taking place in the vicinity (max 500 m distance) of the operational well, and (2) weak to moderate seismicity ($Mw > 0.5$) taking place at least up to 10 km distance from the operational well. The Mw 0.5 magnitude threshold, as the distance threshold, is indicative and chosen upon our current experience and guidelines of several European states. Specific accuracy in the detection and location of weak events down to a specific minimum magnitude threshold may be needed to track the migration of the fracturing processes, e.g. to ensure the preservation of local underground water reservoirs and the integrity of wells. Similarly, the monitoring should be tuned to allow the prompt detection and characterization of moderate events at further distances, if specific seismogenic faults are recognized in the local surrounding of the operation site.

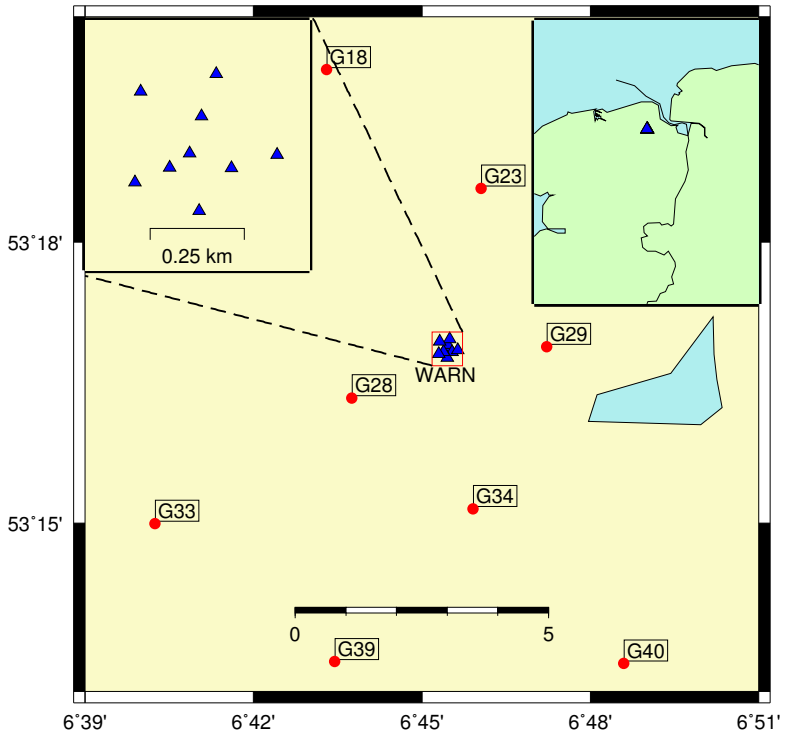
Often the signal to noise ratios (SNR) are poor, the urban and industrial activities are ongoing during the operations, and the sites may be subject to logistical and environmental restrictions. A seismic monitoring network therefore needs to be not only sufficiently sensitive to detect smallest earthquakes at depth, but also flexible in order to adapt to changing conditions and activities at the surface. Borehole seismometers located in deep monitoring wells reaching basement rocks are usually of high sensitivity and improved SNR. However, they are expensive and cannot be adapted to changing conditions. Seismic monitoring approaches employing a network of shallow boreholes may be an alternative, although the SNR improvement from shallow borehole stations is potentially not very large if the sensors are placed in unconsolidated quaternary layers. A combined network of shallow borehole sensors and small aperture arrays of surface sensors can be interesting, since such installations improve the SNR by stacking and at the same time allow to apply beamforming filter techniques to detect waves with specific slowness.

Small aperture arrays have been used in seismology for a variety of applications, ranging from pure detection arrays for regional seismicity and the study of earthquake swarms associated with natural fluid migration (Hiemer et al., 2012) to studies of induced seismicity in relation to fracking experiments (e.g., López-Comino et al., 2017).

In the present paper, we show and discuss examples of small aperture high frequency arrays combined with shallow borehole sensors to monitor induced seismicity during industrial operations. The array characteristics and transfer functions are discussed in the context of micro-earthquake detection at depth. Field tests have been performed above a gas field under production in the Netherlands and during hydraulic fracturing operations at a depth of about 4 km in Poland. We evaluate the fidelity and SNR of the arrays in comparison to shallow borehole sensors under field conditions. A waveform attribute stacking and beamforming method is applied to detect and partially locate events. We test the arrays' event detection capabil-



(a)



(b)

Figure 4.1: a) Overview of the Wysin hydrofracturing experiment site in Poland and the locations of the installed seismic stations to monitor related induced seismicity. b) Wittewierum site above the Groningen gas field, the Netherlands. Blue line in (a) shows the location of the fluid injection at the depth of 4 km. In (a) and (b): red dots show the location of borehole stations. Blue triangles represent broadband stations and black triangles mark short-period stations. Red rectangles show location of inset maps.

ity by beamforming and compare instances of noise levels between array and shallow borehole stations at different depth levels. In addition, we compare the location ability of one array to network based locations and discussed the benefit of using multiple arrays for event location. Concluding, we provide recommendations on the design of microseismic monitoring networks involving seismological surface arrays.

4.2 Data

We employ data recorded at a hydraulic fracturing operation in Wysin (Poland), where a seismic monitoring system was installed consisting of surface broadband stations, small-scale arrays and shallow borehole stations. Additionally, we analyse data recorded on a small-aperture seismic array deployed temporarily in Wittewierum above the Groningen gas field (The Netherlands). Both installations were part of the SHEER project (SHale gas Exploration and Exploitation induced Risks, www.sheerproject.eu). In the following, we describe the instrumentation at both sites in more detail.

4.2.1 The Wysin seismic monitoring system

A dedicated seismic network was installed at a shale gas play close to the village of Wysin in the central-western part of the Peribaltic syncline at Pomerania (Poland). In this area, a Polish oil and gas company drilled two horizontal boreholes designed for fracturing for prospecting and exploration of oil and natural gas. Hydrofracking operations were performed along two horizontal wells at 3955 m and 3865 m depth with an approximate horizontal length of 1.7 km each, in the time periods 9-18 June and 20-29 July, 2016 (López-Comino et al., 2017, 2018).

A hybrid installation, including a distributed network of six broadband stations, three borehole geophones and three small-scale arrays (Fig. 4.1a) to account for both triggered and induced seismicity in the vicinity of the operational wells, was installed in summer 2015 and fully operational from November 2015 until January 2017. All stations operated in continuous mode. The six broadband stations surround the drilling site at distances between 2.1 and 4.3 km with a good azimuthal coverage (maximal gap 90°). Broadband stations were equipped with GÜRALP CMG-3ESP sensors recording with a sampling rate of 200 Hz. In addition, short-period stations were arranged in three small-scale arrays with apertures between 450 and 950 m. Short period stations were equipped with MARK L-4C-3D sensors (GLOD array) and GeoSIG VE-53-BB sensors (CHRW and PLAC arrays) with sampling rates of 500 Hz. The shallow underground installation is composed of three seismometers installed at 50 m depth (initially Geotech Instruments KS-2000). Seismometers at two borehole stations (GW3 and GW4) were replaced by Nanometrics Trillium Compact Posthole 120s sensors at the end of April 2016 due to technical problems. The sampling rate of all downhole instruments was 500 Hz.

4.2.2 The Wittewierum array

The objective of the temporary array deployed above the Groningen gas field was to test the usage of a conventional array layout for detection of microseismicity. The region of the Groningen gas field is an excellent test ground, since the operating company NAM (Nederlandse Aardolie Maatschappij) installed a multitude of shallow borehole stations from 2014 to 2017, of which 65, in addition to the already existing shallow borehole stations installed by KNMI (Koninklijk Nederlands Meteorologisch Instituut), were already online during the time of measurement, thus ensuring an earthquake catalogue that is complete down to M_L 0.5 during the time of array installation (Dost et al., 2017).

Table 4.1: KNMI catalogue. See Fig 4.10 to compare the locations with the locations obtained from single array beamforming (BF).

No	Date	Time	Lat	Lon	Depth	Ml	Vis. by eye	Event BF	Phase BF
1	2016/07/17	12:01:18.89	53.182	6.887	3	0.5			
2	2016/07/18	08:58:11.50	53.378	6.709	3	1.7	x	x	x
3	2016/07/22	10:55:15.30	53.280	6.855	3	0.3			
4	2016/07/23	17:59:45.00	53.219	6.898	3	0.1			
5	2016/07/26	14:02:10.40	53.277	6.907	3	0.9			
6	2016/07/28	05:32:13.09	53.281	6.860	3	0.2		x	x
7	2016/07/28	15:57:28.10	53.250	6.824	3	0.8	x	x	x
8	2016/08/07	20:40:22.00	53.374	6.644	3	1.3	x	x	x
9	2016/08/08	00:03:39.39	53.170	6.892	3	0.4		x	
10	2016/08/10	18:16:25.30	53.312	6.669	3	0.5	x	x	x
11	2016/08/14	01:50:44.89	53.234	7.019	3	0.7	x	x	x
12	2016/08/14	04:07:50.70	53.220	6.678	3	0.2	x	x	x
13	2016/08/23	02:11:16.10	53.224	7.027	3	0.6			
14	2016/08/23	03:53:30.30	53.223	7.036	3	1.0	x	x	x
15	2016/08/24	13:09:08.40	53.305	6.903	3	0.8	x	x	x
16	2016/08/24	18:44:23.19	53.372	6.724	3	0.6	x	x	x
17	2016/08/24	23:44:03.00	53.354	6.950	3	1.1	x	x	x
18	2016/08/28	03:27:53.10	53.401	6.636	3	1.3	x	x	x

The site for the installation was agreed on with local parties involved in the seismicity monitoring, i.e. KNMI and NAM. Stations were installed from July 12 to August 29, 2016 for a period of almost 50 days. Fig. 4.1b displays the location of the Groningen gas field with the placement of the array stations shown as blue triangles, and the locations of borehole stations in the vicinity of the array displayed as red circles.

IMS (International Monitoring System) modern small aperture arrays usually consist of a central station plus further stations placed on concentric rings, each with an odd number of sites, spaced at log-periodic intervals (Schweitzer et al., 2012). We based the geometry of the Wittewierum array on this construction, but were not entirely free in choosing the ring diameters and station sites. The array was composed of 9 seismometers and constructed as three concentric rings of 75 m, 150 m and 225 m radius including a central station. Each station consisted of a broadband sensor (Trillium 120 s), an acquisition system (CUBE datalogger), a battery and a GPS antenna. Sensors were installed at about 1 m depth. All array stations recorded continuously with little outages (Cesca et al., 2016).

During the installation time, KNMI registered 18 events, which are listed in Table 4.1. (`https://data.knmi.nl/datasets/aardbevingen_catalogus/1`), the largest of which had a local magnitude of 1.7 and occurred on July 18, 2016, at a distance of about 11m km to the array. The event closest to the array occurred at a distance of about 5.5 km on July 26, 2016, and had a local magnitude of 0.9.

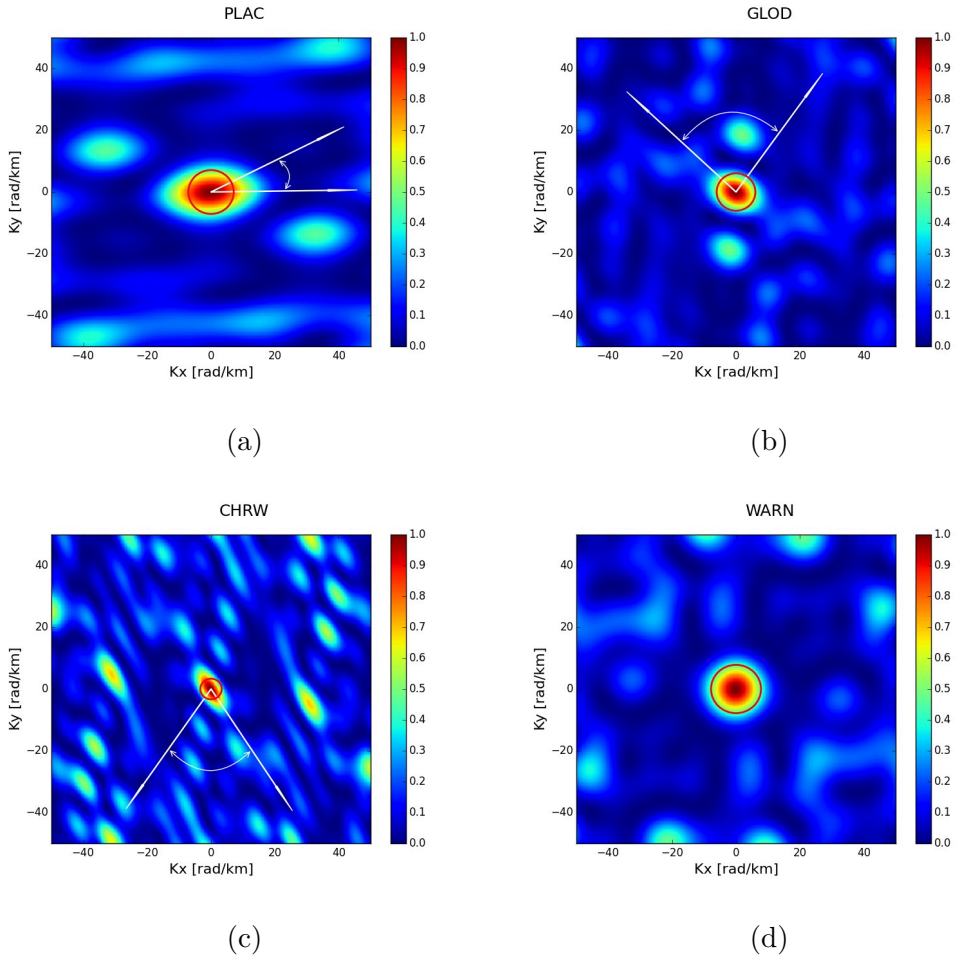


Figure 4.2: Array transfer functions (ATFs) for the three arrays at Wysin (a)-(c) and Wittewierum (d). Red circles show the width of the main lobe which is equivalent to the array resolution, i.e. K_{min}. The arrows in (a)-(c) indicate the expected backazimuth range for incoming signals as inferred from Fig. 4.1.

4.3 Methodology

4.3.1 Array assessment

Arrays have a special ability to distinguish between signals with different wavenumbers (slownesses) crossing the array simultaneously. Array signal processing methods are based on improving the SNR by highlighting the arriving seismic waves with a specific wavenumber (slowness) and suppressing the background signals travelling with different wavenumbers (slownesses). The theoretical value of SNR improvement by an array with n stations, is $\sqrt{n}$ (Schweitzer et al., 2012, Eq. 9.7). While the number of array stations controls the SNR gain achievable by the array, the array geometry defines the limits for the resolvable wavenumbers. For instance, small aperture arrays can not distinguish between waves with small wavenumber differences, and for crossing waves with long horizontal wavelengths (λ) compared to the array aperture (a), such arrays act like a single station. So theoretically, the upper limit for the longest horizontal wavelength that can meaningfully be analysed by array techniques is about the aperture of the array: $\lambda_{max} \simeq a$, so the lower band of resolvable wavenumber, K_{min}, or array resolution is equal to $\frac{2\pi}{a}$. In addition, a wave crossing the array should be sampled by at least two stations, i.e. the smallest recordable wavelength is $\lambda_{min} = 2d_{min}$ assuming d_{min} is the minimum interstation distance, and thus, the maximum resolvable wavenumber is $K_{max} = \frac{\pi}{d_{min}}$.

The array transfer function (ATF) is a standard tool to quantitatively analyse the array performance and to study the capability of the array as seismic monitoring system. The ATF

depends on the relative position of array stations and the frequency content of the signals of interest and, for a specific frequency ω, is defined as:

$$\left|\frac{1}{n}\Sigma_{j=1}^{n} e^{i(\mathbf{K}-\mathbf{K_0}).\mathbf{r}_j}\right|^2, \tag{4.1}$$

where $\mathbf{K} = [k_x, k_y] = \omega[s_x, s_y]$ is the horizontal wavenumber vector, $\mathbf{r}_j = (\delta x_j, \delta y_j)$ is the horizontal location vector of the jth station relative to the array reference point and n is the number of stations (Rost and Thomas, 2002).
The characteristics of the array transfer function such as the presence of side lobes and the shape and sharpness of the main lobe are related to the array layout. For instance, a circular shape of the main lobe implies a symmetric distribution of the array stations ensuring a similar resolution of signals arriving from different backazimuth angles. The width of the main lobe depends on the aperture of the array and defines the array resolution, K_{min}. The larger the aperture of the array, the sharper the main lope and higher the resolution of the array.
The presence and distribution of side lobes depends mainly on the interstation distances and the frequency of the incoming signals. The larger the interstation distances, the closer the side lobes are to the main lobe, which threatens the accurate slowness (wavenumber) determination by increasing the danger of slowness (wavenumber) aliasing depending on the relative beampower ratio between main and side lobes.
The estimation of K_{min} and K_{max} are theoretically valid if the array geometry is regular with uniform interstation distances. Nevertheless, given a potential irregular geometry in the two spatial directions, the true resolution is azimuth dependent (Zywicki, 1999). Experience from ambient vibration studies with synthetic and ground truth data show that the resolution capability of an array lies approximately between $K_{min}/2$ and K_{max} (Wathelet, 2008). Additionally, considering the energy content of the signal, even under the best experimental conditions and inside the resolution limits, if the wave energy is too low, identifying the correct wavenumber is difficult and maxima are hardly visible in the slowness-âĂŞazimuth plane. In practice, the level of incoherent noise rather than the array geometry is the main factor controlling this lower bound (Poggi and Fäh, 2010).

4.3.2 Array beamforming and event detection

The array beam trace is calculated as the sum of all recorded, time shifted traces:

$$B(t) = \frac{1}{n}\Sigma_j^n \;\; Y_j(t + dT_j) \qquad \text{with} \qquad dT_j = s_x\delta x_j + s_y\delta y_j. \tag{4.2}$$

where, Y_j is the trace recorded at the array station j, and n is the number of stations. Assuming the plane wave approximation is valid, the time shift, dT_j, for station j depends on the horizontal slowness components of the incoming wavefront, s_x and s_y, and the relative distance to the array reference point, δx_j and δy_j. The common strategies to find horizontal slowness vector components by estimating the correct values of time shifts and computing the array beam are described in e.g. Schweitzer et al. (2012) and Rost and Thomas (2002).
In the present study, we apply *Lassie*, a recently developed automated full waveform event detection algorithm based on systematic shifting and stacking of smooth characteristic functions and subsequent identification of instances of high coherence in signals recorded at different stations (Lassie, `https://gitext.gfz-potsdam.de/heimann/lassie`; Matos et al., 2018, Heimann et al., in preparation). *Lassie* was initially developed to be applied to data recorded on monitoring networks. We extended *Lassie* using a standard delay-and-sum beamforming approach to shift characteristic functions on a predefined slowness and backazimuth grid. The characteristic functions of traces implement bandpass filtering, taking the absolute, Hanning window

convolution, downsampling and final continuous normalization (in that order) to produce a smooth representation of energy contained in the signal. Thanks to efficient implementation and parallelization, the algorithm applies a dense grid search to full waveforms and produces event detections at occurrences of coherent energy crossing the array along with estimates of backazimuth and apparent horizontal slownesses. These information can be employed for signal classification and subsequent event location (Schweitzer et al., 2012).

4.4 Application

We investigate the theoretical capabilities of the three installed arrays at the Wysin site in Poland (PLAC, GLOD, and CHRW) and at the Wittewierum site above the Groningen gas field in the Netherlands (WARN) with respect to their ability to detect expected target events. The transfer functions of the arrays are plotted in Fig. 4.2. In this figure, white arrows displayed in (a)-(c) indicate the direction to expected target events, considering the location of the array and the fracturing experiment (Fig. 4.1). Red circles show the array resolution (K_{min}). Due to the irregular shape of the arrays at the Wysin site, the array resolutions are not uniform for all backazimuth directions. In contrast, the WARN array's stations are regularly spaced and thus, it is expected to have uniform azimuthal resolution.

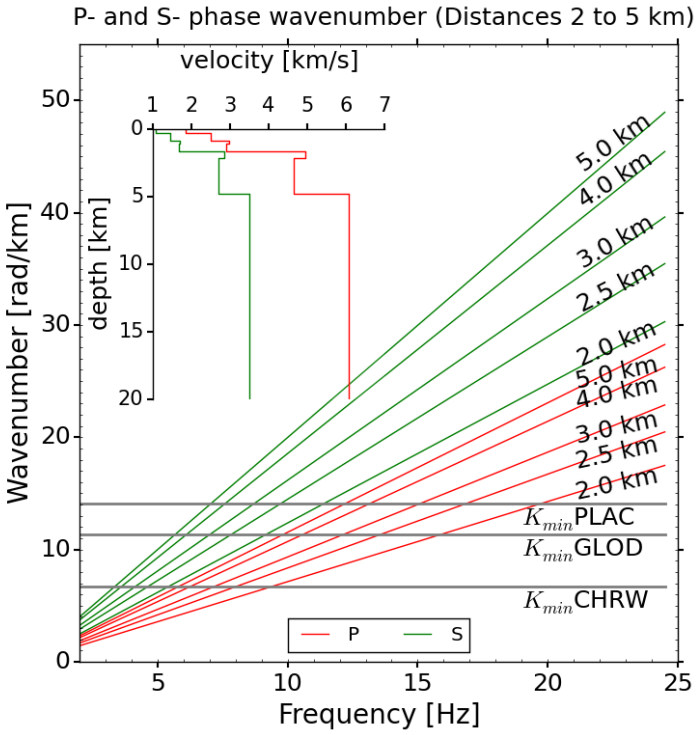
Theoretical frequency-wavenumber curves of P- and S-phases resulting from target events are depicted in Fig. 4.3 providing information on the capabilities of the arrays in terms of resolution and expected aliasing features. The depths of events are assumed to be 4 km and 3 km for the Wysin and Wittewierum area, respectively. The distance dependent wavenumber lines are estimated using the theoretical slowness values depicted in Fig. 4.4. The distance range in each case is selected according to the expected event distances. The velocity models for two the sites are shown in Fig. 4.3. In practice, waves may travel with higher slowness values. Especially for the Groningen field, the seismic velocities in the uppermost layers derived recently (Hofman et al., 2017; Kruiver et al., 2017) are much lower than defined in the velocity model depicted in Fig. 4.3, which was derived from the average velocity model employed by KNMI for event location in the Northern parts of the Netherlands including, but not being limited to the Groningen field (Spetzler and Dost, 2017). For the computation of slownesses, it was combined with the CRUST2.0 model (Bassin, 2000) for depths larger than reservoir depth, since the velocity structure of the deeper part of the Carboniferous layer is not well known (Dost et al., 2017). In addition, S-wave velocities for the sediments down to 3000 m depth were estimated from P-wave velocities using Castagna's relation (Castagna et al., 1985).
In Fig. 4.3, the value of K_{min} for individual arrays is indicated by the horizontal lines in order to ease discussion and comparison of the expected performance of the three arrays for different frequency content of P- and S- phases for events at different locations in the fracturing zone. The value of the K_{max} for the WARN array is also depicted. For the other arrays, K_{max} is larger than wavenumber range plotted in the figures and therefore is stated only in the figure caption.

4.4.1 Assessment of the arrays installed at Wysin

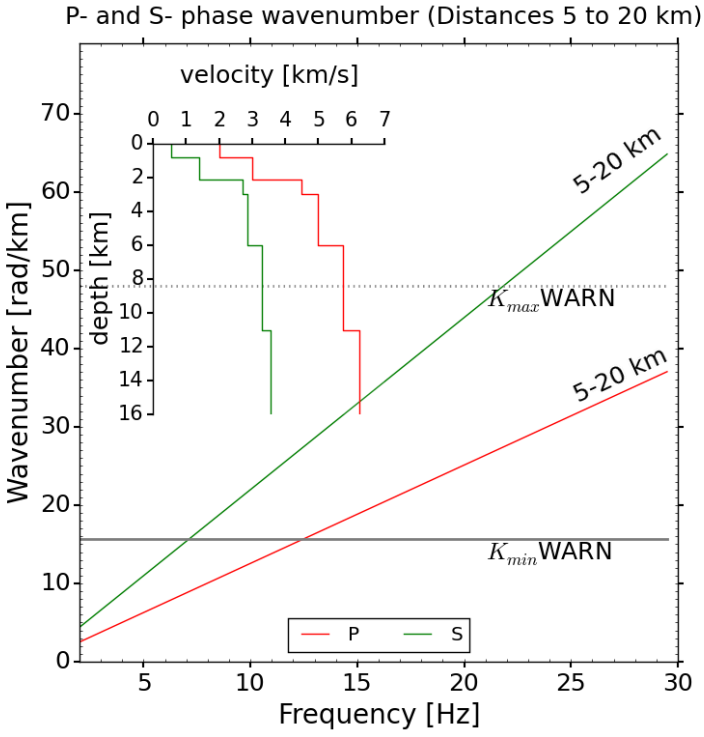
In the following, the assessment of the theoretical capability of individual arrays is described in detail:
PLAC: The PLAC array is expected to record events from a distance range of about 2 - 5 km, from the direction shown in Fig. 4.2a. The ATF of this array shows relatively strong side lobes at about 25 rad/km distance from the main lobe, with relative power as high as 50% of the main lobe. However, these side lobes are not oriented in the expected direction of incoming

events and thus, may not cause a problem in estimation of the slowness vector.

Due to the small aperture of this array, the array is not expected to be sensitive to wavenumbers below $K_{min} = 14$ rad/km. According to the Fig. 4.3, this limiting value of the wavenumber is related to P- and S-phases with certain frequencies and slownesses. Since the slowness in the distance range of 2 - 5 km is increasing with distance (see Fig. 4.4), the minimum wavenumber is related to the higher frequencies at closer distances and shifts towards the lower frequencies by increasing the epicentral distances. In other words, events at closer distances ($\leq$ 2 - 3 km) with frequency content of P-phases of less than 15 - 20 Hz are not detectable by the array, while for the S-phases, the frequency limits shift towards the lower frequencies (since the slowness of S-phases is higher than that of P-phases), so the related lower frequency bands shift from 7 to 12 Hz for distances 5 to 2 km.
Considering this frequency limitation, P-phases will be difficult to detect in array beams and more likely, events at all distances will be detected once the S-phase energy is exceeding the noise level. The difference in wavenumbers from waves arriving from the edges of the fracturing zone, i.e., at 2 km to 5 km distance, is a fraction of the resolution limit at lower frequencies and is the same as at higher frequencies, which implies that distinguishing events that arrive simultaneously at the array will probably not be possible. The value of K_{max} is about to 62 rad/km, and frequencies up to about 30 Hz are expected to be resolved safely by this array.
GLOD: The GLOD array is situated at a distance of 2 - 3.5 km of the hydrofracturing experiment. According to the ATF shown in Fig. 4.2b, a secondary lobe is situated in direction of the backazimuth of interest, which is about 20 rad/km away from the main lobe, with the relative power as high as 50% of the main lobe. The wavenumber limit K_{min} for this array is 11.3 rad/km. So the frequency limit for this array is shifted to lower frequencies compared to the PLAC array. This means that P-phases originating from events at distances of 2 km to 3.5 km with frequencies less than 14 to 16 Hz cannot be detected by the array, while for S-phases, the frequency limit is 6 to 9 Hz.
The side lobes at 20 rad/km imposes another limitation on the resolvable signals, since they are situated in the backazimuth range of expected signals. If they lead to spatial aliasing depends on the relative power between main lobe and side lobes. Accordingly, the highest frequency for arriving phases should be considered above which such spatial aliasing would occur, which is 20 Hz and 10 Hz, respectively, for P- and S-phases emanating at distances of 3.5 km.
CHRW: The fracturing operation occurs about 3 to 4 km away from the CHRW array. The ATF of this array is depicted in the Fig. 4.2c. Although a number of side lobes are present, the expected azimuth direction does not contain any high amplitude secondary lobe. However, some small amplitude side lobes (30% of main lobe amplitude) are visible in those directions. The width of the main lobe is smaller compared to the other arrays as the aperture of the array is larger, allowing this array to be sensitive to lower wavenumbers. The resolution is not uniform in all backazimuth directions, as the array itself is elongated in approximately SW-NE direction. Thus, in this direction, K_{min} is lowest corresponding to the best resolution for small wavenumber differences. Contrary, the resolution is poorest in the SE-NW direction. Therefore, events from the western edge of the hydrofrack will be easier to observe than from the eastern edge. According to Fig. 4.3, P- and S-phases will be detectable at frequencies above 7 Hz and 2.5 Hz, respectively, at about 4 km epicentral distance. Compared to the other arrays, CHRW array has a better chance to detect P- and S-phase arrivals, however, similar to the other arrays, the resolution of the array to separate between simultaneously arriving waves is insufficient.



(a)



(b)

Figure 4.3: Theoretical wavenumbers of P- and S-phases for the frequency range of incoming waves and expected epicentral distances and depths for (a) arrays at the Wysin sites PLAC, GLOD and CHRW and (b) the array at the Wittewierum site (WARN). The velocity models that are used to estimate slownesses are shown, and depths of events are 4 km and 3 km in (a) and (b), respectively. The value of K_{min} is indicated for each array by a horizontal solid black line. The values of K_{max} for the PLAC, CHRW and GLOD arrays are 61, 45 and 52.5 rad/km, respectively. For the WARN array, it is about 48 rad/km, which is indicated in (b) by the horizontal dotted line.

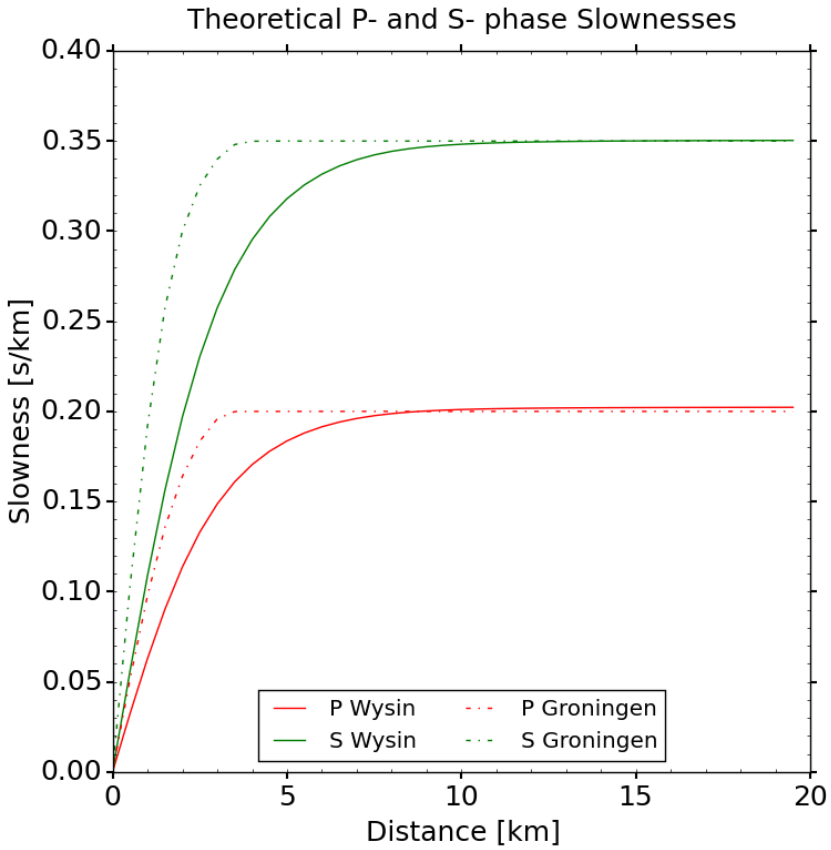


Figure 4.4: Theoretical slownesses of P- and S-phases for the Wittewierum (dashed lines) and Wysin (solid lines) arrays for an assumed event at 3 km and 4 km depth respectively; the velocity models are shown in Fig. 4.3.

4.4.2 Assessment of the arrays installed at Wittewierum (WARN)

The source-array distance is expected to be about 5 to 20 km and the array is supposed to detect seismic waves originating from all directions.
According to Fig. 4.2d, the width of main lobe is circular, so the resolution is uniform for all directions. In addition, some relatively strong secondary lobes exist, but only at 40 rad/km from the main lobe. So the array is capable to resolve larger wavenumber ranges.
The array aperture is 0.4 km, which means that K_{min} is about 16 rad/km. According to Fig. 4.4, for the distance range of 5 to 20 km, P- and S-waves possess constant slowness values of 0.2 and 0.35 s/km for the velocity model assumed for this region, so the frequency limits are not distance-dependent. According to Fig. 4.3b, the lower frequency limits for P- and S-phases are about 12 Hz and 7 Hz, respectively. In contrast to the arrays installed in Wysin, WARN exhibits an upper frequency limit for P-waves for the epicentral distances of interest at about 22 Hz.

4.5 Results

4.5.1 Event detection on the Wysin arrays

We applied the modified version of *Lassie* using a frequency pass band between 9 and 20 Hz. Backazimuths were scanned between -30° to +30° in 0.5° steps and slownesses between 0.05 s/km and 0.3 s/km in 0.01 s/km steps. After manual revision of the detections, we could verify that

none of them were of seismic origin from the nearby fracturing site. Most of the detections correspond to local noise sources (López-Comino et al., 2018). An example of waveforms and a detection at array GLOD is depicted in Fig. 4.5.

4.5.2 Event detection on the Wittewierum array

In order to process data recorded on the Wittewierum array, we applied a bandpass filter between 9 and 30 Hz following a spectrogram analysis and employed a full backazimuth grid search (from 0° to 360° with a grid step of 5°). Slownesses were scanned between 0 s/km and 0.5 s/km (corresponding to horizontal apparent velocities from 2 km/s to infinity). At first, the detection algorithm was tested on eleven events from the KNMI catalogue that where visible by eye (See column 8 in Table 4.1) in the data in order to evaluate the detection threshold. Subsequently, the complete data set was processed. An example detection is shown in Fig. 4.6 for an event that occurred on July 18, 2016 (08:58:11h). The waveforms of this event recorded on the WARN array as well as the KNMI shallow borehole station G28 are depicted in Fig. 4.7. The application to the complete data set results in more than 65 000 detections, albeit half of which with a detector strength lower than 16.5. When plotting backazimuth estimates versus slowness for different detector strengths (Fig. 4.8), there is neither a preferred slowness range nor orientation recognizable. Since seismic events at Groningen are supposed to originate at reservoir depth (Dost et al., 2017), differences in slowness mainly imply changes in the distance to the events. However, with the exception of some events detected with zero slowness, the backazimuth-slowness pattern is similar for all detector strengths, which strengthens the assumption that at least a part of the detections constitutes real events. Detection performance is stable over time, but decreases in the period from the August 20 to August 22, when two stations were malfunctioning.
In addition, *Lassie* detects all events which served for parameter tuning, as well as two additional events catalogued by KNMI that are less obvious in the single seismic traces (See column 9 in Table 4.1). However, two of those thirteen events exhibit a large difference in backazimuth compared to the KNMI event location (Events number 8 and 15 in Table 1 which are marked by grey stars in Fig. 4.8). Five other events listed by KNMI where not detected. In general, for the KNMI catalogue events, detection levels correlate with event magnitude and anti-correlate with distance.
The apparent velocities vary between 2 and 6.6 km/s and thus are slightly more variable than what is expected from 1-D raytracing (2.8 km/s to 5 km/s, Fig. 4.4). In fact, the average velocity model for the Northern Netherlands used to derive slownesses is not describing the complex structure of the Groningen gas field very well. The Rotliegend gas reservoir (average P-wave velocity v_P=3.8 km/s) is overlain by anhydrite with a much higher velocity of v_P=5.9 km/s and underlain by the Carboniferous with v_P=4.25 km/s (Willacy et al., 2018). These high-impedance contrasts channel earthquake energy within the reservoir and result in significant mode conversions (Willacy et al., 2018). In addition, there are strong impedance contrasts between the Zechstein reservoir seal and the overburden as well as within the overburden itself, further complicating the propagation of seismic waves, such that seismograms recorded at the surface contain considerable P-to-S and S-to-P conversions (Willacy et al., 2018). Including single and multiple reflections seismograms are difficult to interpret (Willacy et al., 2018) and beamforming may stack converted instead of direct arrivals due to a wrong phase association. In addition, as mentioned above, seismic velocities, especially S-wave velocities, in the uppermost layers derived recently (Hofman et al., 2017; Kruiver et al., 2017) are much lower than defined in the average velocity model.

It is difficult to distinguish automatically between noise and earthquake signals. One indica-

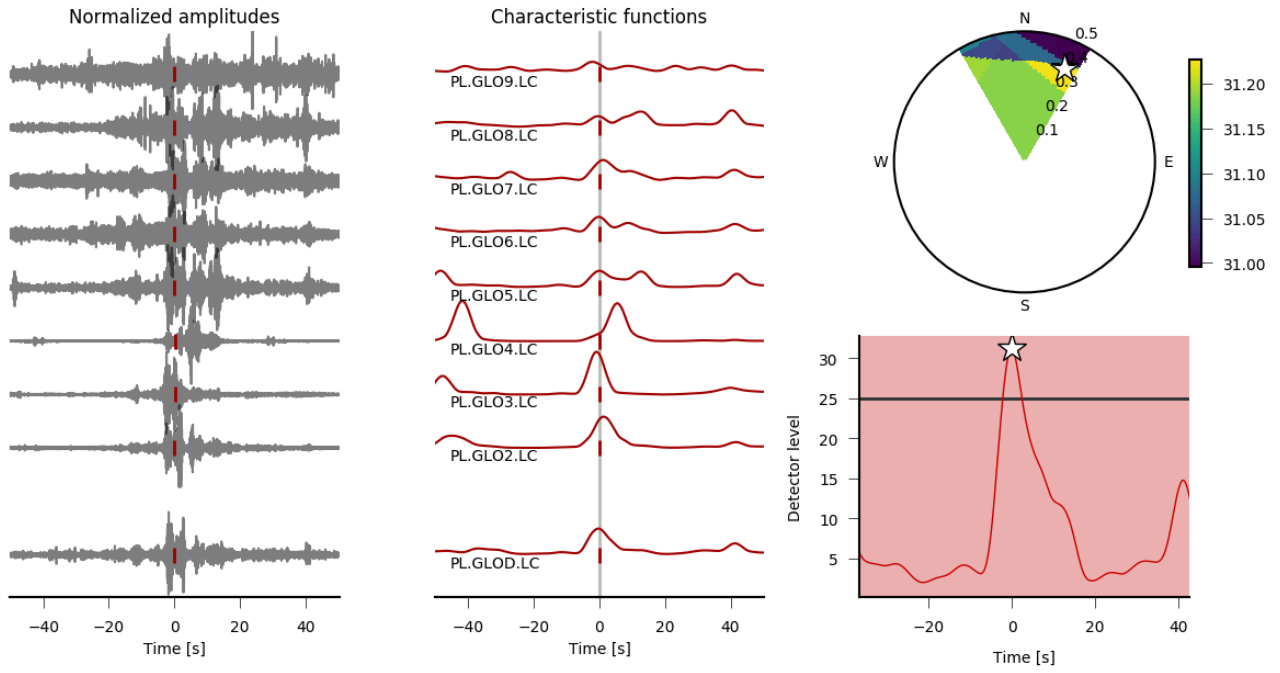


Figure 4.5: A detection on the GLOD array shortly after the injection was stopped (see white star no. 1 in Fig. 4.11). This detection is not confirmed to be an event related to the hydrofracturing experiment. a) Left: filtered seismic traces; centre: characteristic functions, vertical dashes mark applied shifts according to the maximum in the slowness-backazimuth domain; top right: slowness-backazimuth slice coloured by amplitude of stacked characteristic functions, white star denotes the maximum coherence; bottom right: detector level, the white star represents the local maximum detected once the coherence exceeds the detector threshold indicated by the black horizontal line.

tion is the distribution of events with time of day (Fig. 4.9). Clearly, the detection distribution with time of day is not even. Additionally, the temporal behaviour varies for different detector strengths; events with a detector strength below 16 occur more often between 21:00 (9:00 P.M.) and 4:00 A.M., whereas events with a detector strength above 18 have a pronounced peak between 21:00 (9:00 P.M.) and 10:30 P.M. and a second between 13:30 and 14:30 (1:30 P.M. and 2:30 P.M.). Most of the detections occur during night time, indicating that the day time noise from superficial sources increases the magnitude of completeness.
Surprisingly, the distribution of events with time seems to be relatively independent of apparent velocity, although the absolute number of detections with apparent velocity lower than 5 km/s is ten times higher than the number of events with apparent velocity between 5 and 10 km/s. That means that a low apparent velocity cannot be used to distinguish between shallow artificial sources close to the array and natural sources at larger distances and depths.

4.5.3 Event location capability of a single array

In order to locate events using a single array, it would be necessary to form separate beams for P- and S-wave onsets searching different slowness ranges and being filtered in different frequency bands, ensure that direct arrivals are detected and associate phases belonging to the same event prior to locating it based on S-P travel time differences and backazimuth estimate (Mykkeltveit and Bungum, 1984). However, the event location precision of single small aperture arrays is limited due to scatter in the backazimuth and uncertainties in automatically measuring the travel time differences (Schweitzer et al., 2012), such that at seismological observatories, observations from several small aperture arrays as installed at the Wysin site are usually interpreted jointly, employing for example the generalized beamforming location algorithm (Ringdal and

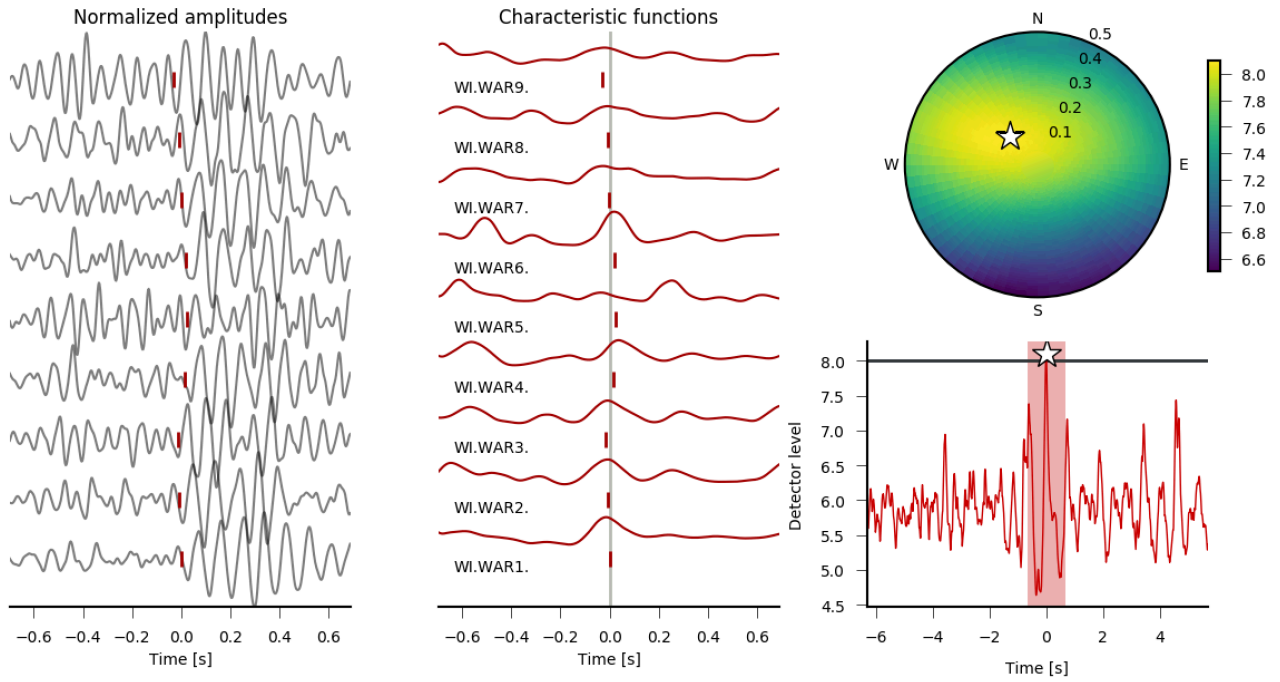


Figure 4.6: Example of event detected by *Lassie* on the WARN array on July 18, 2016. For a description of the plot see caption of Fig. 4.5.

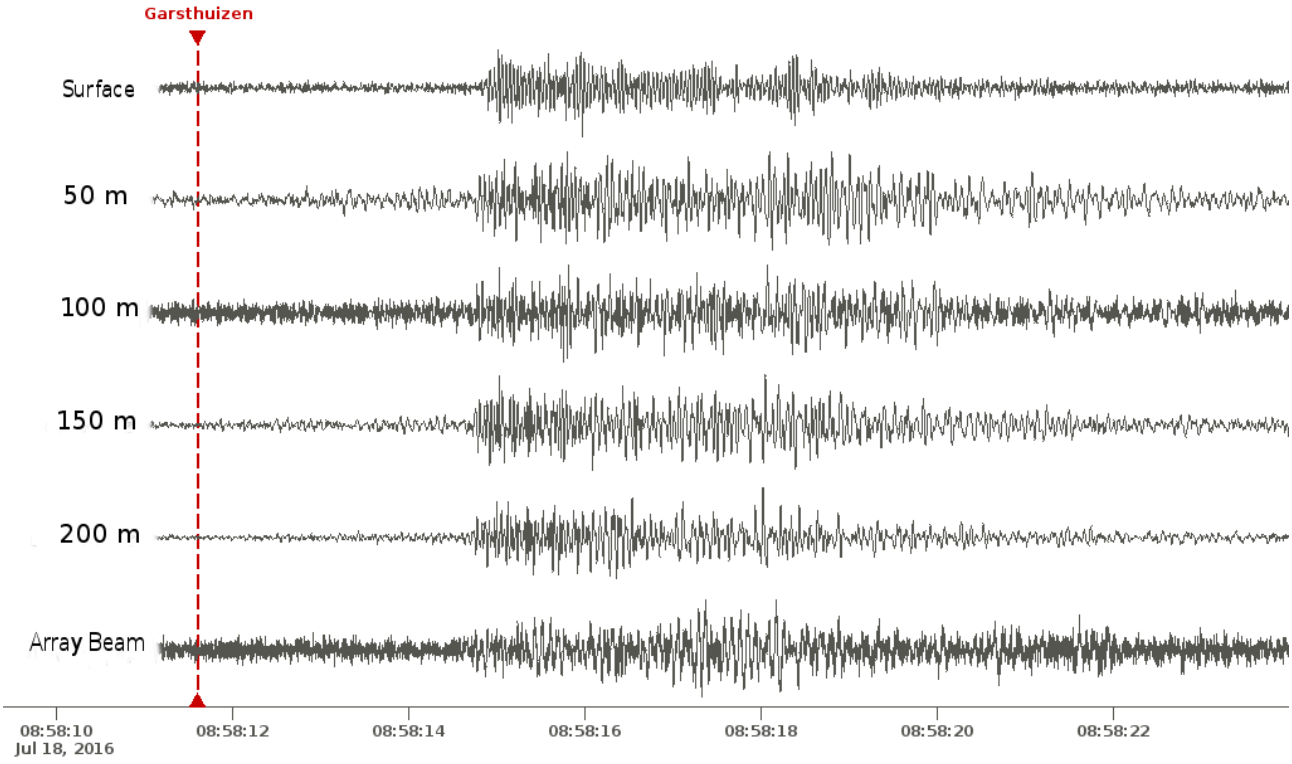


Figure 4.7: Waveforms of the event on July 18, 2016 (08:58:11). The five top traces are recorded at the G28 shallow borehole station. The top trace stems from the surface accelerometer, the following four traces from different levels within the borehole. The last trace shows the WARN array's beam (according to the P-phase horizontal slowness vector).

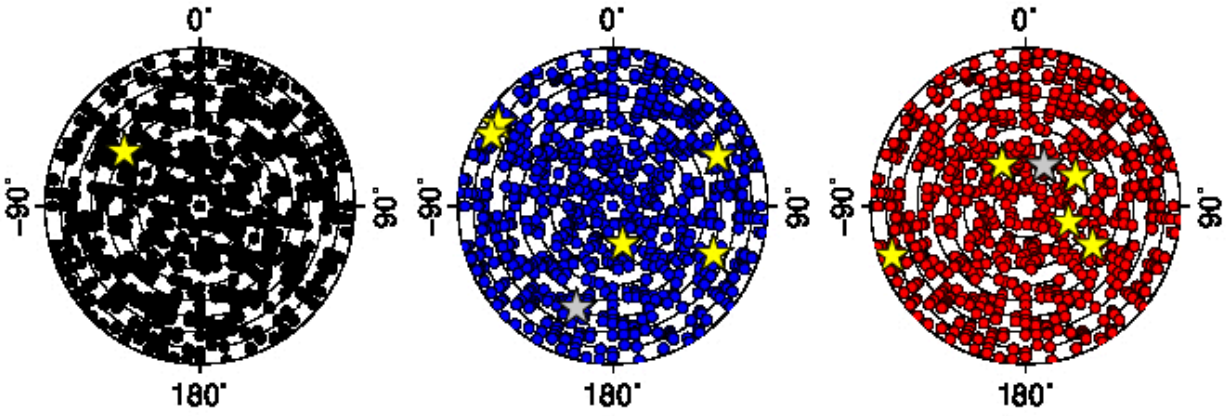


Figure 4.8: Polar plots showing the distribution of measured backazimuths and slownesses for different detector strengths (slowness varies from 0 s/km in the centre to 0.5 s/km at the outer rim, intergrid line distances correspond to 0.05 s/km). Left: detector strength < 16.5, middle: detector strength between 16.5 and 18, right: detector strength > 18. Stars indicate events registered in the KNMI catalogue.

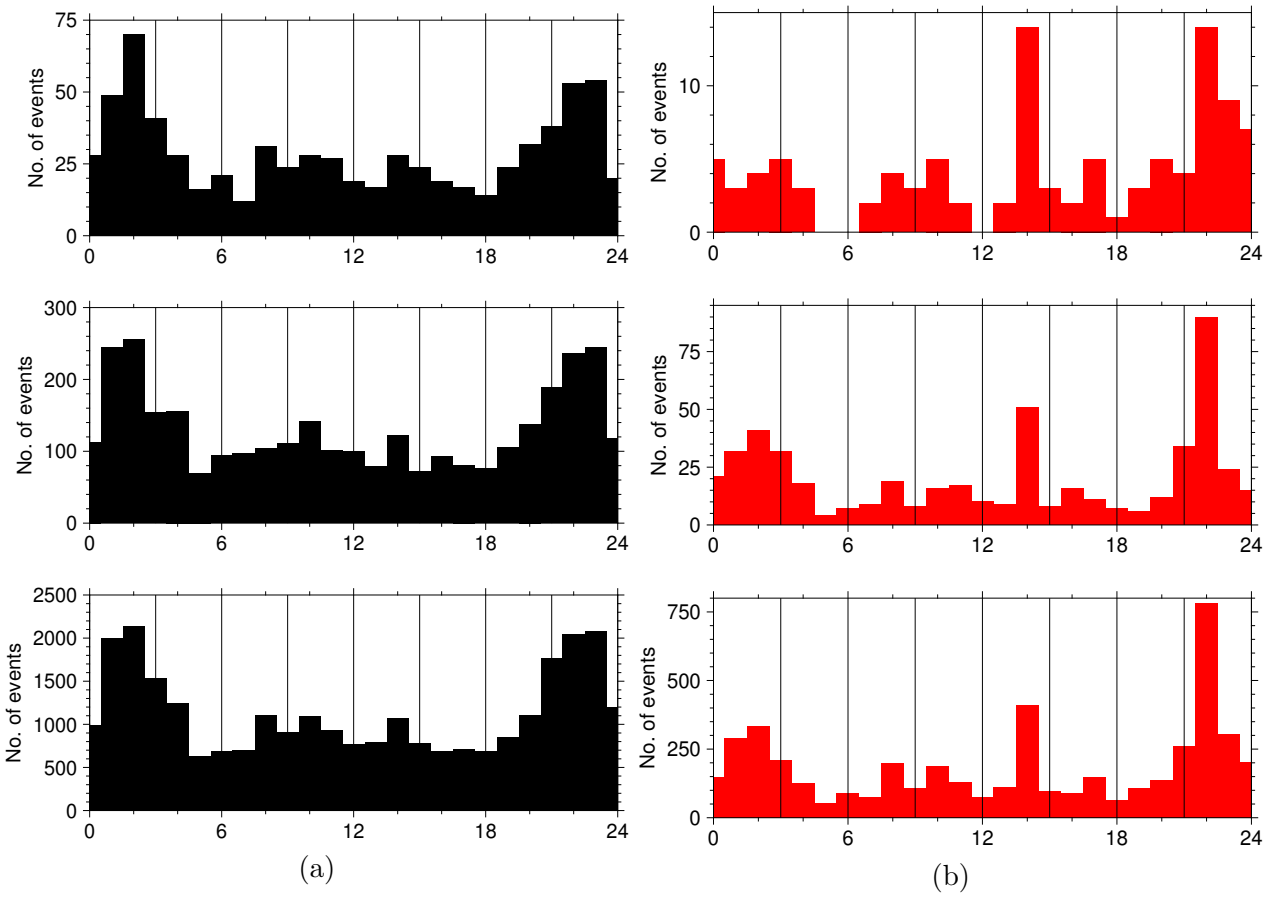


Figure 4.9: Distribution of the number of events with time of day for three different ranges of apparent velocity (top: apparent velocity > 10 km/s, middle: apparent velocity between 5 and 10 km/s, bottom: apparent velocity < 5 km/s); left: detector strength < 16, right: detector strength > 18.

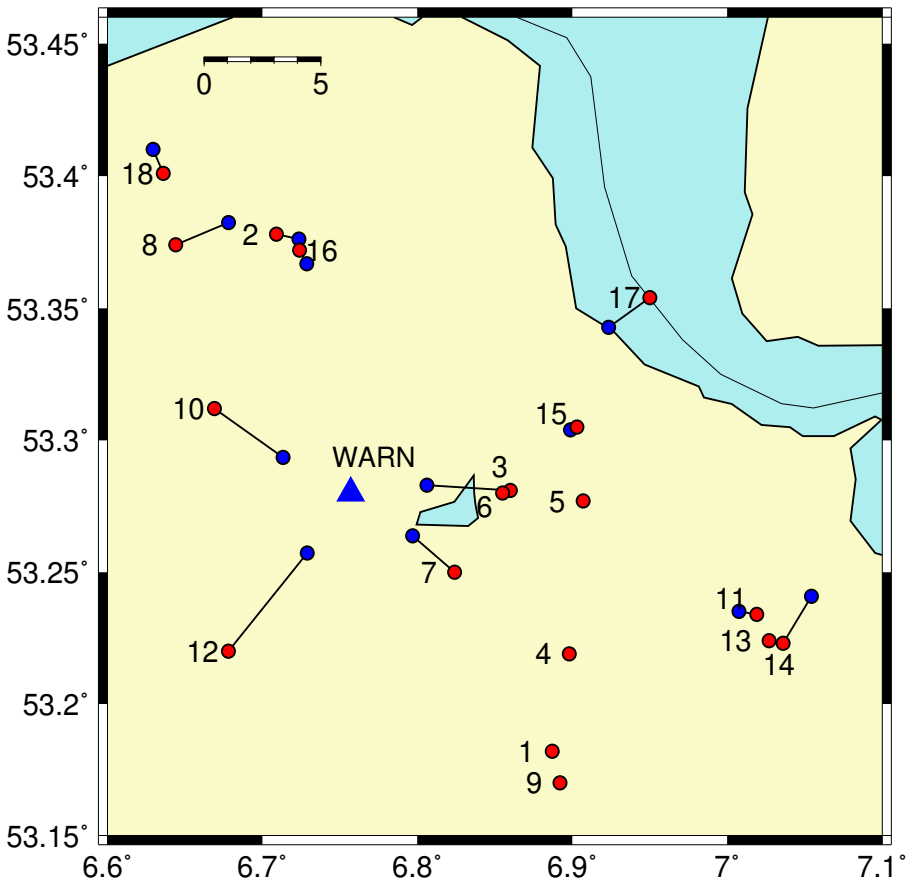


Figure 4.10: Blue triangle shows the WARN array. Red circles are locations reported in KNMI catalogue (Table 4.1) and blue circles are location calculated from single array beamforming method (phase detection module). Numbers are in accordance with the numbers in Table 4.1.

Kværna, 1989). Recently, techniques have been developed to integrate array recordings with network recordings for event location (Sick and Joswig, 2016; López-Comino et al., 2017). In addition, the use of multiple use for event location has been picked up (Stipčević et al., 2017).

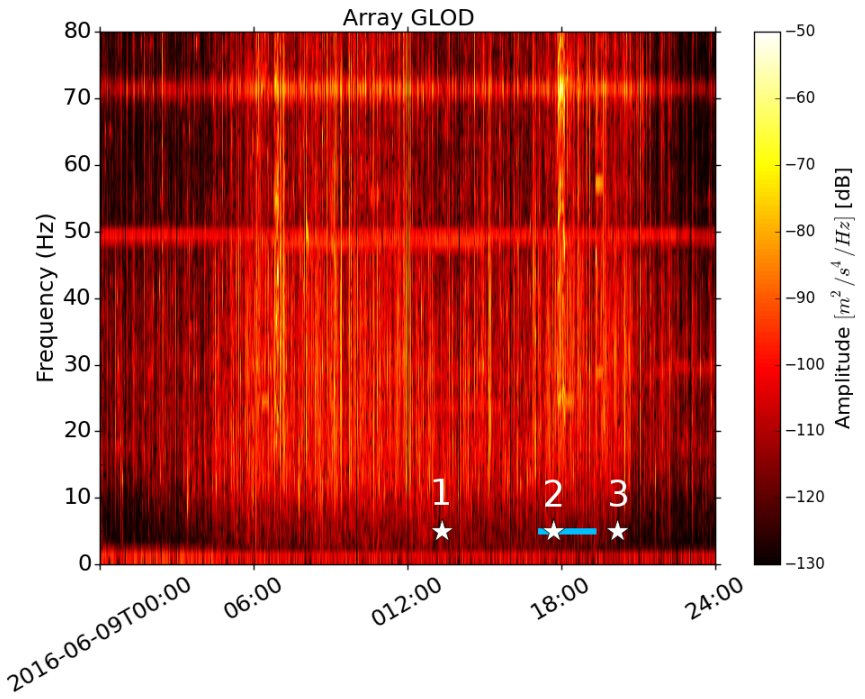
In the following, we evaluate the location capability of the WARN array. We use a phase detection module developed in-house, which will be integrated into *Lassie* in near future. This module performs a semblance analysis in short moving time windows to measure backazimuths and slowness of arrivals with higher precision than feasible during the *Lassie* automatic beamforming. While events are detected automatically applying a Short Time Average over Long Time Average (STA/LTA) detector on the semblance traces, accurate $P-$ and $S-$ phases are picked manually based on slowness values. We analyzed time segments of data containing the KNMI reported events. In case of event detection (See the last column in Table 4.1), the event location is estimated using the S-P arrival time difference as well as the estimated backazimuth employing the velocity model presented in Fig. 4.3b. The obtained twelve event locations are compared to the KNMI catalogue event locations in Fig. 4.10. For almost all events, backazimuth estimates agree very well with the KNMI catalogue event locations. The mean deviation is 3°, while the largest is only 11°, whereas the average backazimuth deviation for the detected events using *Lassie* automatic beamforming is 38°, excluding two very large values. The mean value of epicentral mislocation is 2.1 km, whereas its maximum is 5.3 km. However, the error is largest for events closest to the array. Therefore, we suspect that this deviation does not originate from errors in arrival time measurements of P- and S-phases, which would presumably be more randomly distributed, but is caused more likely by an erroneous V_p/V_s ratio. Hofman et al. (2017) demonstrate that especially the shallow S-wave velocities vary significantly, which leads to a laterally fluctuating V_p/V_s ratio. There is no reason to assume that such lateral

variations cannot be present in the deeper sedimentary layers as well, which could explain the systematic distribution of error in distance. Such errors in event location can be avoided by employing multiple arrays. Nevertheless, we think that this comparison shows the inherent capability of arrays to measure slowness vectors of incoming waves with high precision.

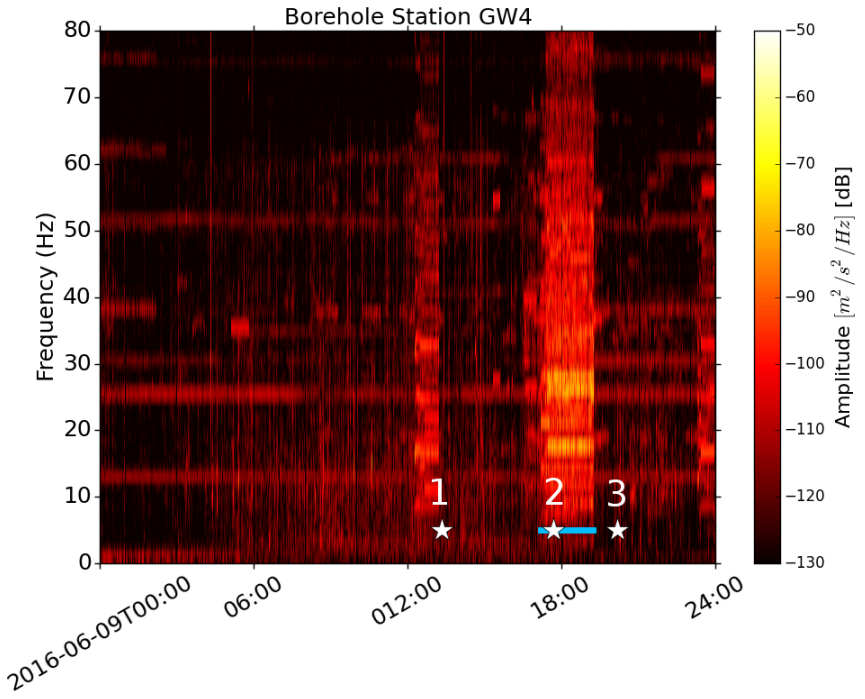
4.6 Discussion

In this section. we discuss the special ability of a surface array to reduce the background noise level during fluid injection and extraction experiments. Since at both sites, arrays were operating as a complementary element to surface or borehole sensor installations, we can compare noise levels of surface arrays with nearby borehole instruments.

At the Wysin site, the noise levels at the borehole GW4 and the nearest array, GLOD, both of them in about 2 km distance to the injection well (see Fig. 4.1), are compared using spectral analysis. Fig. 4.11 shows the variation of the noise spectral content measured on the starting day of the injection (June 9, 2016) at the array (Fig. 4.11a) and the borehole (Fig. 4.11b). The spectral content of the recorded signal at the borehole shows an increase in the two-hour interval between 17:10h and 19:10h (indicated by a blue line on each panel), which is in the agreement with the injection time, whereas at the surface array, such a correlation is not visible. The source of the noise was most likely related to the pumping activity at the surface close to the injection well (López-Comino et al., 2018). More information about the timing of the injection activity is given by (López-Comino et al., 2017).
Fig. 4.12 shows the noise power spectral density (PSD) during three periods before, during and after fluid injection for the surface array and the borehole station. Ten-minute time windows were analysed and the start time of each period is indicated by white stars on Fig. 4.11. For the array stations, the PSD is calculated from the array beam, which is formed to detect P-waves generated at the location of the injection at a depth of 4 km. According to the graphs, while the borehole station shows an increase of noise level during the fluid injection for frequencies above 4 Hz, after the injection, the noise level falls to almost the same level as before the experiment, with 10 db fluctuation. However, for frequencies above 60 Hz, the noise is still slightly increased. On the contrary, the noise levels of the array beam before and during the experiment are almost identical, except for the narrow frequency band between 10 Hz to 18 Hz, where the noise level increases about 5 db during the injection. Furthermore, comparing the surface array and borehole analysis, it is concluded that below 6 Hz, the noise level at the surface array is lower than at the borehole station and this pattern is visible for all three periods. However, for frequencies above 6 Hz before the injection, the noise at the array is larger than in the borehole. During the injection experiment, the noise level for frequencies above 60 Hz in the borehole reaches the noise level of the surface array. Seismic noise at the surface array shows a strong variation before and after hydraulic fracturing, in contrast to borehole stations, where the variation is not significant. This observation can be explained with the higher sensitivity of surface installations to daily variation of human activity producing higher seismic noise during daytime and lower seismic noise during night hours.
Fig. 4.12b shows the noise level comparison between the WARN surface array at Wittewierum and at different depth levels of the close-by KNMI station G28. This station consists of an accelerometer placed at the surface and 4.5 Hz geophones placed at depths of 50, 100, 150 and 200 m. The horizontal distance between both locations is about 1 km (Fig. 4.1). The time segments employed for this comparison are 10 minutes long and are extracted before the detection of the largest event with magnitude Ml 1.7 on July 18, 2016. According to the figure, the noise level of the array beam is in general smaller than the noise at a single surface station (see blue dotted and dashed curves). The noise level reduction from a surface measurement

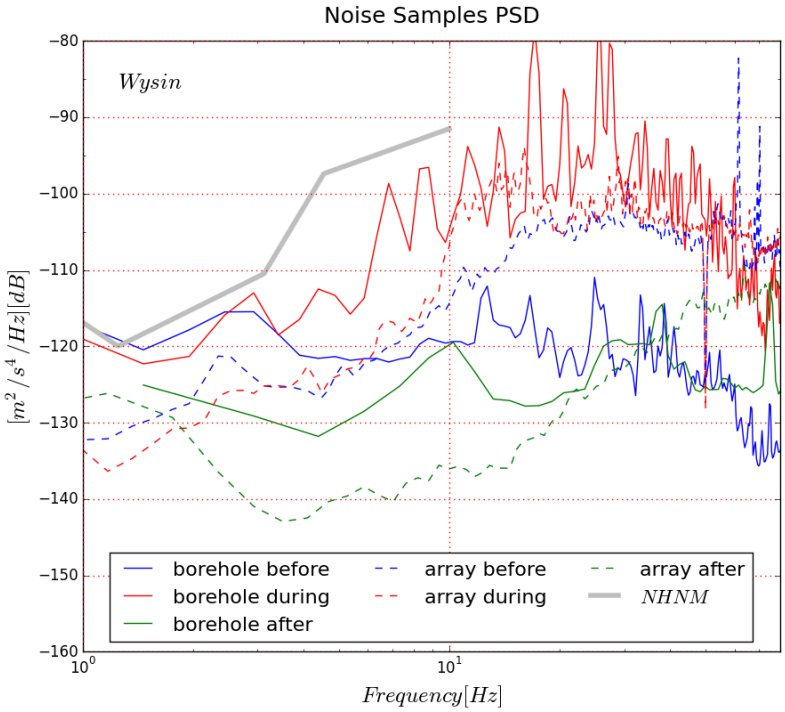


(a)

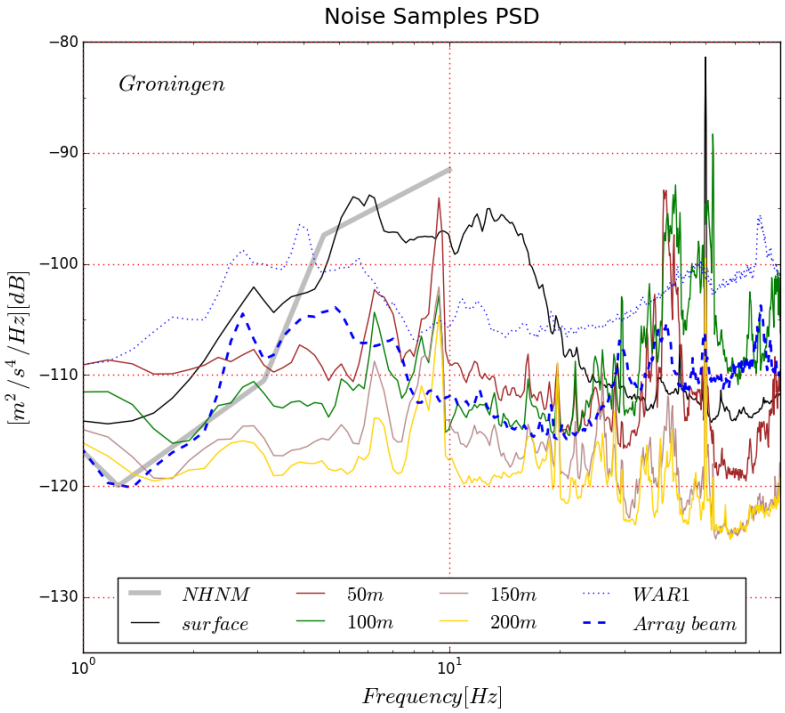


(b)

Figure 4.11: Noise spectral variation during the hydrofracturing operation at about 2 km distance from the injection point measured on a) a surface array beam (GLOD) and b) at a borehole station (GW4) (see Fig. 4.1 for the locations of instruments). The blue lines on each plot show the injection time period and the three white stars show the start time of the two-minute time segments used to compare the PSD (Fig. 4.12) for three periods before, during and after the injection.



(a)



(b)

Figure 4.12: PSD of the noise samples recorded during a) three periods before (blue lines), during (red lines) and after (green lines) the hydrofracturing experiment in Wysin and b) in Wittewierum The thick grey line displays the new high noise model according to Peterson (1993). In a), solid lines mark PSDs of recordings in the borehole, wheras dashed lines display PSDs of recordings on the surface array. In b), solid lines mark PSDs of recordings from the instruments of the shallow borehole station G28, the dashed blue line indicates the PSD of the WARN array's beam and the dotted blue line displays the PSD of a record from the central array station, WAR1.

(accelerometer) to the 50 m deep instrument (velocity meter) can reach 10 to 15 db in the 3-20 Hz frequency band. By means of array beam forming we can achieve a 5 to 10 db reduction in the noise level for all frequency bands. At frequencies below 3 Hz, the noise level of the array is similar to the noise level on the borehole stations at 200 m depth. For the frequency range of 6-20 Hz, the array beam shows a lower noise level than achieved at 50 m depth in the borehole and the same level as reached at 100 m depth in the borehole. As expected, the high frequency noise on the borehole sensors decreases the deeper the sensors are placed. Below 50 Hz, the incremental decrease with depth is larger the higher the frequencies are. However, the borehole sensor at 100 m depth is an exception, since it shows high noise at frequencies above 30 Hz, increasing even to the noise level of the surface stations (measured on the accelerometer of G28 and the central station of the WARN array). The reason is unclear, but may be related to the local geology and potentially, waveguides at depth.

In general, the noise level at the Wysin site is lower than at the Wittewierum site. Especially, an instrument placed at 50 m depth at Wysin experiences lower noise levels than an instrument placed at 200 m depth above the Groningen field. This is a result of a higher level of cultural noise in the Netherlands compared to Poland (Kraft, 2016) and represents another aspect that should be included when planning a monitoring network. Unfortunately, so far the only source of information are up-front test measurements, since no general database for a comparison of noise levels at different locations is available yet.

4.7 Conclusions

In this paper, we evaluated the performance of small aperture arrays with respect to their ability to detect target events. For the purpose of planning array measurements before injection experiments, we recommend:

1. In order to design a small aperture array for the specific target to monitor weak induced seismicity at shallow depths, the array transfer function should be analysed and different array geometries should be evaluated and compared, specifying the expected source-receiver distances and expected slowness range of incoming P and S waves, the anticipated magnitudes to be monitored and the estimated frequency range and expected horizontal wavenumber ranges. Figs. 4.1 and 4.2 give examples for two real case studies.

2. Plots as shown in Fig. 4.3 and Fig. 4.4 are helpful to support the planning of the array design. The velocity models and targeted event depths are used to derive slowness ranges, and from the expected frequencies the range of horizontal wavenumbers. The theoretical wavenumber analysis showed that increasing the aperture of the array leads to a decrease in K_{min} and thus, the crossing point with the wavenumber-distance lines is shifted to lower frequencies. This means that a lower frequency band can be included in the beamforming analysis. Especially for monitoring of nearby microseismicity, increasing the aperture has a limitation, though, since the plane wave approximation may be violated if the aperture of the array is in the same order as the source-array distance.

3. Planning a microseismic monitoring array often is subject to restrictions such as land use, accessibility of the stations, and other logistics. The local noise level at individual stations poses constraints as well, since high noise sites should be definitely avoided. Additionally, the source mechanisms of the individual earthquake events can influence the performance of the array. We suggest to apply synthetic simulations and design the array geometry based on an optimization approach considering all seismological and logistical information about the targeted site and sources. An example is provided in the study by **?**.

From the specific experience we gained by analysing data recorded by small scale arrays at Wysin and Wittewierum, we conclude:

1. Borehole installations should be combined with surface arrays during hydrofracturing operations. Although no injection-induced event occurred at Wysin above the magnitude of completeness of Mw 0.5, we could demonstrate how hydrofracturing operations impact the SNR at shallow boreholes, while small aperture surface arrays, located at larger distances to the injection well, are less affected. For instance, the shallow borehole installations suffered from very high noise related to the pumping activitites during the injection itself. Therefore, combining boreholes close to the injection site with small aperture arrays at larger distances is beneficial to ensure a constant magnitude of completeness over the full period of the experiment.

2. It is preferrable to employ multiple surface array installations as an alternative to a dense network of borehole sensors, especially in areas experiencing high levels of noise. In case of the Groningen gas field, we could detect a multitude of potential events below the magnitude of completeness of the KNMI catalogue and locate events comprised in the catalogue with $Ml > 0.2$. According to the comparison between KNMI network and single array event locations, the WARN array was capable to determine the backazimuth and arrival time differences of P- and S-phases with high precision. Using more than one array, will decrease the location errors caused by an improper velocity model (Stipčević et al., 2017).

Chapter 5

Concluding remarks

This chapter concludes the thesis. A summary of the main outcomes is presented in Section 5.1., while Section 5.2 contains outlooks and suggestions for future research.

5.1 Summary of main conclusions

This thesis has mainly focused on the development of a new approach for seismological array design. The approach see the array design as an optimization problem, in which all considerations and criteria which are important in array design task, can be defined as individual objective functions or can be integrated in one relevant objective function. To mathematically engage all the available information, the expected seismic data at assumed array stations are modelled based on realistic facts, and the realistic synthetic earthquake waveforms are employed to evaluate the objective functions. This section summarizes the main conclusions of the thesis chapters, excluding the introductory chapter 1, that does not contain research material.

In chapter 2, the developed receiver array design framework is introduced in details, and an example application is presented and discussed. Receiver arrays have a variety of applications in seismology, particularly when the signal enhancement is a prerequisite to detect seismic events, and in situations where installing and maintaining sparse networks is impractical. Study and monitoring of nuclear and chemical explosions (Baumgardt and Der, 1998; Kim et al., 1998), volcanic and non-volcanic tremors (La Rocca et al., 2008; Saccorotti and Del Pezzo, 2000; Ghosh et al., 2009), earthquake swarms (Hiemer et al., 2012) and real time monitoring and early warning systems of active faults (Meng et al., 2014) are examples of array applications. A remarkable utility of a receiver array is to estimate the horizontal slowness vector of the incoming wave, which yields to phase identification and further seismological findings, such as event location and rupture front tracking (Krüger and Ohrnberger, 2005; Ishii, 2011). I showed an example of a 7-station array design in a swarm activity zone in northwest Bohemia/Vogtland area. By presenting this example, the aim was to show how the array design framework can be initialized and customized for a specific problem. The parameters needed for the synthetic array waveform modelling are derived from previous seismological studies of this region added by realistic estimation of synthetic noise. The synthetic noise level is estimated from real samples of waveforms recorded by an array that was installed temporarily in the same area.

To simulate realistic background noise traces, noise power spectral density (PSD) is calculated using real samples recorded by the array. Localized noise sources are modelled by assuming a cluster of point sources and explosion sources with random mechanisms, origin times and locations at the surface of the search area. Accordingly, a stochastic 3-component noise trace is produced that is tuned to have the same PSD level as that calculated from real samples. In this way, synthetic noise samples show realistic variation of PSD at different random stations, while properties of the noise signals such as high degree of coherency at short distances and within three components are simulated realistically. Comparing spatial correlation of synthetic

and real noise data indicates that the distance dependent coherency of synthetic noise signals is realistic.

Two objective functions are introduced to ensure the accuracy of the estimated P- and S-phase slowness of synthetic signals. Using the scalarization technique, two objective functions are combined in one single function and a final optimized geometry is introduced.
A comparison between the three solutions, obtained using individual and final combined objective function, is performed by evaluating the coarray of each array geometry, which illustrates the set of all inter-station distances (Haubrich, 1968). I concluded that the objective function which is related to the array beam power, i.e. f_1, has a tendency to distribute stations regularly with large inter-station distances, such that the smallest inter-station distance for the related optimized model is bigger than those in the other two arrays. In addition, the maximum inter-station distance for this case is comparable with the dimension of the deployment site. On the other hand, the objective function f_2, which is related to the precision of the slowness vector, results in a model that has stations at closer inter-station distances to achieve a better azimuthal resolution of the target swarm area. Nevertheless, the final optimized solution keeps properties of two other models as much as possible to achieve optimization in terms of both desired properties measured by the contributing functions.
I compared the final optimized array geometry with some well-known regular arrays. I concluded that, the final arrays which are suggested for the region using introduced scenario and boundary conditions perform better than the other arrays in terms of objective fucntion f_2 (precise slowness vector). However, a uniform array geometry, can perform better than the others in terms of the objective function f_1 (array beam power).
The robustness of the applied optimization process is evaluated by doing a statistical test, by running it 200 times and analyzing the minimum value of the objective function obtained in every run. From the results we concluded that, as we repeat the optimization process, while the final solutions are not always the same, the minimum value of the objective function remains in the same level. In other words, it is possible to make more than one suggestion with similar performance in terms of the defined objective function.
To allow for small variations in each station location to take into account unforeseen restrictions in the field, I suggest to consider the family of best solutions as well. The family of solutions is determined by applying a threshold on the final objective function. Accordingly, the spatial distribution of all models whose objective functions fall below the threshold, gives an indication for possible adjustments of the final deployment, while for each station, the average location of the clustered solutions reveals an alternative station location.

Another contribution of this research is to use the optimization process to make source arrays using a well located catalog of events which are occurred in a spatial cluster. The automatize procedure which is applied and tested both using synthetic and real data, is capable to find optimal set up of sources that can be processed in combination as array recordings. The application of SA technique are mostly focused on the study of properties of the velocity interfaces and heterogeneities in different scales using converted or reflected phases, (Niazi, 1969; Goldstein et al., 1992; Weber and Wicks, 1996; Revenaugh, 1995; Krüger et al., 1993, 1995; Scherbaum et al., 1997; Krüger et al., 2001; Rietbrock and Scherbaum, 1999), and coda composition and microearthquakes (Scherbaum et al., 1991; Dodge and Beroza, 1997). In chapter 3, the idea of receiver array optimization is extended to include source arrays. I presented a method to search for seismic source array elements, given a well-located earthquake swarm catalog. I formulated the preconditions to make an SA, by defining 4 objective functions. One of the key conditions to define an SA is the accuracy of source locations, I performed a test to see how the location error propagates into the final outcomes of the source array beamforming analysis, i.e slowness vector estimation.

According to the results, for the tested source-receiver geometry and SA configuration, as it is expected, by increasing the value of location error, the error in slowness vector estimation increases, so that for assuming a maximum 75 m location error, up to 0.15 s/km deviation in the calculated slowness vector is likely to happen for S_x component, while for the other components, the range of deviation is less than 0.07 s/km. It is interesting that the deviated slowness values are distributed around the main peaks and show correlation with the array beam power patterns of the original error-less SA. Accordingly, in the $s_x - s_z$ plane, where we find the sharpest central peak, the lowest deviation of slowness vector component is observed. In other words, for the smaller aperture source arrays, more precise locations are necessary to produce reliable final results. Larger aperture source arrays are more robust against location errors.

We did not include the events origin time errors in our calculations, as such data were not available in our catalog. To eliminate the timing error, aligning the waveforms based on P-phase arrival times, that can be calculated using array beamforming method, offers a solution. In this case relative slowness value of the reflected phases are measurable.

The synthetic test and test with real data, proved that the proposed method by using the suggested optimization approach is capable to find an optimized SA according to the defined objective functions. In addition, the values of the individual sub-functions at the final optimized SA, indicates that our weighting strategy performs well.

Results of the synthetic test to localize the point scatters related to the observed sp- and ps-phases, demonstrated that if the coherency of the scattered phases are persistent over the SA elements, it is possible to image the scatterer location by measuring the semblance of the phases. However, even in a synthetic test, the accuracy of the resolved scatter depends on the geometry of the experiment, that is the source array and receiver relative location and ray path of the phases. For instant, the semblance pattern calculated using ps-phase shows one distinct but wider peak compared to the main peak obtained using sp-phase. This point implies to the lower resolution of calculated scatterer location which is closer to the receiver than to the source region using ps-phase. This simple experiment also demonstrated that to image different part of an extended scatterer, or if there are many local scatterers, different phases and variable SA receiver set up are required. However, in such cases, phase interference can violate the coherency of the phases arrive in specific time window. According to the test using real data, there are indications for scatterers in the velocity structure in the tested area. The scatterer location revealed by using ps-phases in this study is in agreement with findings of Rüssler et al. (2009). Using the receiver array analysis (Rohrbach array) Rüssler et al. (2009) concluded that backazimuth angle of the converted phases can be deviated towards the north direction compared to the backazimuth angle of the direct phases, indicating the heterogeneity and inclined discontinuities along the ray path.

P-phase waveform similarity of source array elements recorded in a single station are higher than the P-waves waveform similarity measured at one single source and a receiver array. This implies that near surface structure is more pronounced on receiver array recordings, than the near source structure on source array recordings. In our experiment, the dimension of source array and receiver array are comparable.

In chapter 4, the performance of small aperture arrays with respect to their ability to detect target events, at various noise condition is evaluated. For the purpose of planning array measurements before injection experiments, some recommendations are concluded. From the specific experience we gained by analysing data recorded by small scale arrays at Wysin and Wittewierum, I conclude: Borehole installations should be combined with surface arrays during hydrofracturing operations. Although no injection-induced event occurred at Wysin above the magnitude of completeness of Mw 0.5, we could demonstrate how hydrofracturing operations impact the SNR at shallow boreholes, while small aperture surface arrays, located at larger

distances to the injection well, are less affected. For instance, the shallow borehole installations suffered from very high noise related to the pumping activities during the injection itself. Therefore, combining boreholes close to the injection site with small aperture arrays at larger distances is beneficial to ensure a constant magnitude of completeness over the full period of the experiment.
It is preferrable to employ multiple surface array installations as an alternative to a dense network of borehole sensors, especially in areas experiencing high levels of noise. In case of the Groningen gas field, we could detect a multitude of potential events below the magnitude of completeness of the KNMI catalogue and locate events comprised in the catalogue with $Ml > 0.2$. According to the comparison between KNMI network and single array event locations, the installed small aperture array was capable to determine the backazimuth and arrival time differences of P- and S-phases with high precision. Using more than one array, will decrease the location errors caused by an improper velocity model (Stipčević et al., 2017).

5.2 Outlook

The receiver array design framework introduced in this dissertation, can be used to plan seismological arrays in various applications and scales. As long as we can provide the required information to generate the realistic synthetic full waveforms, the framework is applicable to suggest the optimal array configuration. In addition, as the framework suggests a family of optimal solutions, small variations in each station location is possible to take into account unforeseen restrictions in the field while array implementation. Since the objective function is adaptable based on the specific application, the framework is applicable to design multiple arrays and to design arrays as completing elements of an existing seismic network, for instance to design and improve a combined array-network observatory system. With the increasing application of the arrays, specially small aperture arrays, employing the presented framework is very practical and beneficial in array seismology.

The optimized source array processing is applicable to study the ray path and origin of coda waves, which leads to locate the strong scatterers in the wave propagation medium. As the location of the sources and receiver(s) are assumed to be known, the interesting problem to be solved, for instance is to locate a discontinuity in the velocity model which makes strong converted phases such as p to s or s to p. To get reliable image of the scatterers for such an application, we suggest using a collection of the optimal source arrays, by repeating the SA optimization algorithm a number of times. In this case, different time segments of the seismograms are explored to extract coherent scattered phases. Including more single receivers is useful as well, to take into account various SA to receiver geometries to increase the ray coverage of propagation medium. Then, the search algorithm to discover the strong scatterers, using different SAs, should reveal consistent results in overall, to make reliable interpretation of the velocity structure. Simultaneous use of source arrays and receiver array data, allows application of double beamforming method to achieve higher signal-to-noise ratio and to include more constraints on the location procedure.
With recent development in earthquake location techniques and availability of catalogs with high location precision, particularly swarm catalog, the approach proposed in this dissertation to search for optimal source arrays promotes the applications of source array processing and double beam method.

Bibliography

Bassin, C. (2000). The current limits of resolution for surface wave tomography in North America. *EOS Trans. AGU. 81: Fall Meet. Suppl., Abstract*.

Baumgardt, D. R. and Der, Z. (1998). Identification of presumed shallow underwater chemical explosions using land-based regional arrays. *Bulletin of the Seismological Society of America*, 88(2):581–595.

Ben-Menahem, A. and Beydoun, W. B. (1985). Range of validity of seismic ray and beam methods in general inhomogeneous media–i. general theory. *Geophysical Journal International*, 82(2):207–234.

Bondár, I., North, R. G., and Beall, G. (1999). Teleseismic slowness-azimuth station corrections for the international monitoring system seismic network. *Bulletin of the Seismological Society of America*, 89(4):989–1003.

Boué, P., Roux, P., Campillo, M., and de Cacqueray, B. (2013). Double beamforming processing in a seismic prospecting context. *Geophysics*, 78(3):V101–V108.

Brodsky, E. and Lajoie, L. (2013). Anthropogenic seismicity rates and operational parameters at the Salton sea geothermal field. *Science*, 341(6145):543–546, 10.1126/science.1239213.

Büttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G., and Soukup, T. (2004). The Corine Land Cover 2000 Project. *EARSeL eProceedings*, 3(3):331–346.

Caramia, M. and Dell' Olmo, P. (2008). Multi-objective optimization. *Multi-objective Management in Freight Logistics: Increasing Capacity, Service Level and Safety with Optimization Algorithms*, pages 11–36.

Castagna, J., Batzle, M., and Eastwood, R. (1985). Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks. *Geophysics*, 50(4):571–581.

Cesca, S. (2019). Drainage of a deep magma reservoir near mayotte inferred from seismicity and deformation. *Nature geoscience*, 198(2):941–953.

Cesca, S., Grigoli, Francesco and, S., González, A., Buforn, E., Maghsoudi, S., Blanch, E., and Dahm, T. (2014). The 2013 september–october seismic sequence offshore spain: a case of seismicity triggered by gas injection? *Geophysical Journal International*, 198(2):941–953.

Cesca, S., López-Comino, J., Kühn, D., and Dahm, T. (2016). Array in wittewierum, netherlands. deutsches geoforschungszentrum gfz. other/seismic network.

Dahm, T., Fischer, T., and Hainzl, S. (2008). Mechanical intrusion models and their implications for the possibility of magma-driven swarms in NW Bohemia Region. *Studia Geophysica et Geodaetica*, 52(4):529–548.

Dahm, T., Horálek, J., and Šílený, J. (2000). Comparison of absolute and relative moment tensor solutions for the january 1997 west bohemia earthquake swarm. *Studia Geophysica et Geodaetica*, 44(2):233–250.

Dahm, T., Hrubcova, P., Fischer, T., Horalek, J., Korn, M., Buske, S., and Wagner, D. (2013). Eger Rift ICDP: an observatory for study of non-volcanic, mid-crustal earthquake swarms and accompanying phenomena. *Scientific drilling : reports on deep earth sampling and monitoring*, 16:93–99.

Deichmann, N. and Giardini, D. (2009). Earthquakes induced by the stimulation of an enhanced geothermal system below Basel (Switzerland). *Seismological Research Letters*, 80(5):784–798, 10.1875/gssrl.80.5.784.

Dodge, D. A. and Beroza, G. C. (1997). Source array analysis of coda waves near the 1989 loma prieta, california, mainshock: Implications for the mechanism of coseismic velocity changes. *Journal of Geophysical Research: Solid Earth*, 102(B11):24437–24458.

Dost, B., Ruigrok, E., and Spetzler, J. (2017). Development of seismicity and probabilistic hazard assessment for the Groningen gas field. *Netherlands Journal of Geosciences*, 96(5):s235–s245.

Ellsworth, W. (2018). Injection-induced earthquakes. *Science*, 341(6142):10.1126/science.1225942.

Fischer, T. and Horálek, J. (2003). Space-time distribution of earthquake swarms in the principal focal zone of the NW Bohemia/Vogtland seismoactive region: Period 1985-2001. *Journal of Geodynamics*, 35(1-2):125–144.

Fischer, T., Horálek, J., Hrubcová, P., Vavryčuk, V., Bräuer, K., and Kämpf, H. (2014). Intracontinental earthquake swarms in west-bohemia and vogtland: a review. *Tectonophysics*, 611:1–27.

Fischer, T., Horálek, J., Michálek, J., and Boušková, A. (2010). The 2008 West Bohemia earthquake swarm in the light of the WEBNET network. *Journal of Seismology*, 14(4):665–682.

Frost, D. A. and Rost, S. (2014). The p-wave boundary of the large-low shear velocity province beneath the pacific. *Earth and Planetary Science Letters*, 403:380–392.

Ghosh, A., Vidale, J. E., Sweet, J. R., Creager, K. C., and Wech, A. G. (2009). Tremor patches in Cascadia revealed by seismic array analysis. *Geophysical Research Letters*, 36(June):1–5.

Goldstein, P., Walter, W. R., and Zandt, G. (1992). Upper mantle structure beneath central eurasia using a source array of nuclear explosions and waveforms at regional distances. *Journal of Geophysical Research: Solid Earth*, 97(B10):14097–14113.

Green, C., Styles, P., and Baptie, J. (2012). Review and recommendations for induced seismicity mitigation. *Preese Hall Shale Gas Fracturing - Induced Seismicity Report*, pages 1–22.

Grigoli, F., Cesca, S., Priolo, E., Rinaldi, A. P., Clinton, J. F., Stabile, T. A., Dost, B., Fernandez, M. G., Wiemer, S., and Dahm, T. (2017). Current challenges in monitoring, discrimination, and management of induced seismicity related to underground industrial activities: A European perspective. *Reviews of Geophysics*, 55(2):310–340.

Grigoli, F., Cesca, S., Rinaldi, A., Malconi, A., López-Comino, J., Westaway, R., Cauzzi, C., Dahm, T., and Wiemer, S. (2018). The November 2017 Mw 5.5 Pohang earthquake: A possible case of induced seismicity in South Korea. *Science*, page 10.1126/science.aat2010.

Hainzl, S., Fischer, T., Čermáková, H., Bachura, M., and Vlček, J. (2016). Aftershocks triggered by fluid intrusion: Evidence for the aftershock sequence occurred 2014 in west bohemia/vogtland. *Journal of Geophysical Research: Solid Earth*, 121(4):2575–2590. 2015JB012582.

Hardt, M. and Scherbaum, F. (1994). The design of optimum networks for aftershock recordings. *Geophysical Journal International*, 117(3):716–726.

Haubrich, R. A. (1968). Array design. *Bulletin of the Seismological Society of America*, 58(3):977.

Havskov, J. and Alguacil, G. (2004). *Instrumentation in earthquake seismology*, volume 358. Springer.

Havskov, J. and Ottemöller, L. (2010). Array processing. In *Routine Data Processing in Earthquake Seismology*, pages 283–297. Springer.

Healy, J., Rubey, W., Griggs, D., and Raleigh, C. (1968). The Denver earthquakes. *Science*, 161:1301–1310–351.

Heimann, Sebastian; Kriegerowski, M. I. M. (2017). Pyrocko - An open-source seismology toolbox and library. V. 0.3. GFZ Data Services.

Hensch, M., Riedel, C., Reinhardt, J., and Dahm, T. (2008). Hypocenter migration of fluid-induced earthquake swarms in the Tjörnes Fracture Zone (North Iceland). *Tectonophysics*, 447(1-4):80–94.

Hiemer, S., Rüssler, D., and Scherbaum, F. (2012). Monitoring the West Bohemian earthquake swarm in 2008/2009 by a temporary small-aperture seismic array. *Journal of Seismology*, 16:169–182.

Hincks, T., Aspinall, W., Cooke, R., and Gernon, T. (2018). OklahomaâĂŹs induced seismicity strongly linked to wastewater injection depth. *Science*, 161:10.1126/science.aap7911.

Hofman, L., Ruigrok, E., Dost, B., and Paulssen, H. (2017). A shallow seismic velocity model for the Groningen area in the Netherlands. *Journal of Geophysical Research: Solid Earth*, 122(10):8035–8050.

Horálek, J., Fischer, T., Bouškova, A., Michálek, J., and Hrubcová, P. (2009). The West Bohemian 2008-earthquake swarm: When, where, what size and data. *Studia Geophysica et Geodaetica*, 53(3):351–358.

Horálek, J. and Šílený, J. (2013). Source mechanisms of the 2000 earthquake swarm in the West Bohemia/Vogtland region (Central Europe). *Geophysical Journal International*, 194(2):979–999.

Horton (2012). Disposal of hydrofracking waste fluid by injection into subsurface aquifers triggers earthquake swarm in Central Arkansas with potential for damaging earthquake. *Seismological Research Letters*, 83(2):doi.org/10.1785/gssrl.83.2.250.

Hrubcová, P., Vavryčuk, V., Bouškova, A., and Bohnhoff, M. (2016). Shallow crustal discontinuities inferred from waveforms of microearthquakes: Method and application to ktb drill site and west bohemia swarm area. *Journal of Geophysical Research: Solid Earth*, 121(2):881–902.

Hrubcová, P., Vavryčuk, V., Boušková, A., and Horálek, J. (2013). Moho depth determination from waveforms of microearthquakes in the west bohemia/vogtland swarm area. *Journal of Geophysical Research: Solid Earth*.

Ishii, M. (2011). High-frequency rupture properties of the Mw 9.0 off the Pacific coast of Tohoku Earthquake. *Earth, Planets and Space*, 63:609–614.

Jousset, P. et al. (2018). Dynamic strain determination using fibre-optic cables allows imaging of seismological and structural features. nat. commun.

Kaneshima, S. (2019). Seismic scatterers in the lower mantle near subduction zones. *Geophysical Journal International*, 218(3):1873–1891.

Karamzadeh, N., Heimann, S., Dahm, T., and Krüger, F. (2019a). Application based seismological array design by seismicity scenario modelling. *Geophysical Journal International*, 216(3):1711–1727.

Karamzadeh, N., Heimann, S., Dahm, T., and Krüger, F. (2020). Earthquake source arrays: optimal configuration and applications in crustal structure studies. *Geophysical Journal International*, 221(1):352–370.

Karamzadeh, N., Kühn, D., Kriegerowski, M., López-Comino, J. Á., Cesca, S., and Dahm, T. (2019b). Small-aperture array as a tool to monitor fluid injection-and extraction-induced microseismicity: applications and recommendations. *Acta Geophysica*, 67(1):311–326.

Kennett, B. L. N., Stipčević, J., and Gorbatov, A. (2015). SpiralâĂŘArm Seismic Arrays. *Bulletin of the Seismological Society of America*, 105(4):2109–2116.

Kim, S. G., Park, Y. C., and Kim, W. Y. (1998). Discrimination of small earthquakes and artificial explosions in the Korean Peninsula using Pg/Lg ratios. *Geophysical Journal International*, 134(1):267–276.

Kim, W.-Y. (2013). Induced seismicity associated with fluid injections into a deep well in Youngstown, Ohio. *Journal of Geophysical Research*, 118(7):3506–3518.

Kraft, T. (2016). A high-resolution and calibrated model of man-made seismic noise for europe. 76th Annual Meeting of the DGG (German Geophysical Society), 14-17.03.2016, Münster, Germany.

Kraft, T., Mignan, A., and Giardini, D. (2013). Optimization of a large-scale microseismic monitoring network in northern switzerland. *Geophysical Journal International*, 195(1):474–490.

Krentz, O. (2019). Saxon state ministry of environment and agriculture. http://www.geologie.sachsen.de/schwarmbeben-im-vogtland-12796.html#a-26016. Accessed: 2019-07-29.

Krüger, F., Baumann, M., Scherbaum, F., and Weber, M. (2001). Mid mantle scatterers near the mariana slab detected with a double array method. *Geophysical research letters*, 28(4):667–670.

Krüger, F. and Ohrnberger, M. (2005). Tracking the rupture of the Mw = 9.3 Sumatra earthquake over 1,150 km at teleseismic distance. *Nature*.

Krüger, F., Scherbaum, F., Weber, M., and Schlittenhardt, J. (1996). Analysis of asymmetric multipathing with a generalization of the double-beam method. *Bulletin of the Seismological Society of America*, 86(3):737–749.

Krüger, F., Weber, M., Scherbaum, F., and Schlittenhardt, J. (1993). Double beam analysis of anomalies in the core-mantle boundary region. *Geophysical research letters*, 20(14):1475–1478.

Krüger, F., Weber, M., Scherbaum, F., and Schlittenhardt, J. (1995). Evidence for normal and inhomogeneous lowermost mantle and core-mantle boundary structure under the arctic and northern canada. *Geophysical Journal International*, 122(2):637–657.

Kruiver, P., van Dedem, E., Romijn, R., de Lange, G., Korff, M., Stafleu, J., Gunnink, J., Rodriguez-Marek, A., Bommer, J., van Elk, J., and Doornhof, D. (2017). An integrated shear-wave velocity model for the Groningen gas field, The Netherlands. *Bulletin of Earthquake Engineering*, 15(9):3555–3580.

Kufner, S.-K., Schurr, B., Haberland, C., Zhang, Y., Saul, J., Ischuk, A., and Oimahmadov, I. (2017). Zooming into the hindu kush slab break-off: A rare glimpse on the terminal stage of subduction. *Earth and Planetary Science Letters*, 461:127–140.

La Rocca, M., Galluzzo, D., Malone, S., McCausland, W., Saccorotti, G., and Del Pezzo, E. (2008). Testing small-aperture array analysis on well-located earthquakes, and application to the location of deep tremor. *Bulletin of the Seismological Society of America*, 98(2):620–635.

Laponce, J. (1972). Vilfredo Pareto, Manual of Political Economy. Trans. Ann S. Schwier, ed. Ann S. Schwier and Alfred N. Page. New York: Augustus M. Kelley, 1971, pp. xii, 504. *Canadian Journal of Political Science*, 5(04):599–600.

Larsen, C. H. (2012). Scientific Drilling. *Reports on Deep Earth Sampling and Monitoring*, (June):80.

López-Comino, J., Cesca, S., Jarosławski, J., Montcoudiol, N., Heimann, H., Dahm, T., Lasocki, S., Gunning, A., Capuano, P., and Ellsworth, W. (2018). Induced seismicity response of hydraulic fracturing: results of a multidisciplinary monitoring at the Wysin site, Poland. *Scientific Reports*, in press:251–267.

López-Comino, J., Cesca, S., Kriegerowski, M., Heimann, S., Dahm, T .and Mirek, J., and Lasocki, S. (2017). Monitoring performance using synthetic data for induced microseismicity by hydrofracking at the Wysin site (Poland). *Geophysical Journal International*, 210(1):42–55.

Málek, J., Horálek, J., and Jansky, J. (2005). One-dimensional qP-wave velocity model of the upper crust for the West Bohemia/Vogtland Earthquake swarm region. *Studia Geophysica et Geodaetica*, 49(4):501–524.

Málek, J., Jansky, J., Novotný, O., and Rüssler, D. (2004). Vertically inhomogeneous models of the upper crustal structure in the West-Bohemian seismoactive region inferred from the celebration 2000 refraction data. *Studia Geophysica et Geodaetica*, 48(4):709–730.

Marano, S., Fah, D., and Lu, Y. M. (2014). Sensor placement for the analysis of seismic surface waves: sources of error, design criterion and array design algorithms. *Geophysical Journal International*, 197(3):1566–1581.

Marler, R. T. and Arora, J. S. (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6):369–395.

Matos, C., Custódio, S., Batló, J., Zahradník, J., Arroucau, P., Silveira, G., and Heimann, S. (2018). An active seismic zone in intraplate West Iberia inferred from high-resolution geophysical data. *Journal of Geophysical Research: Solid Earth*, 123(4):2885–2907.

Meng, L., Allen, R. M., and Ampuero, J.-P. (2014). Application of Seismic Array Processing to Earthquake Early Warning. *Bulletin of the Seismological Society of America*, 104(5):2553–2561.

Morita, Y., Nakao, S., and Hayashi, Y. (2006). A quantitative approach to the dike intrusion process inferred from a joint analysis of geodetic and seismological data for the 1998 earthquake swarm off the east coast of Izu Peninsula, central Japan. *Journal of Geophysical Research: Solid Earth*, 111(6):1–18.

Mykkeltveit, S. and Bungum, H. (1984). Processing of regional seismic events using data from small-aperture arrays. *Bulletin of the Seismological Society of America*, 74(6):2313–2333.

Nakata, N., Boué, P., Brenguier, F., Roux, P., Ferrazzini, V., and Campillo, M. (2016). Body and surface wave reconstruction from seismic noise correlations between arrays at piton de la fournaise volcano. *Geophysical Research Letters*, 43(3):1047–1054.

Neunhöfer, H. and Hemmann, A. (2005). Earthquake swarms in the Vogtland/Western Bohemia region: Spatial distribution and magnitude-frequency distribution as an indication of the genesis of swarms? *Journal of Geodynamics*, 39(4):361–385.

Niazi, M. (1969). Use of source arrays in studies of regional structure. *Bulletin of the Seismological Society of America*, 59(4):1631.

Okabe, A., Boots, B., and Sugihara, K. (1992). *Spatial Tesselation: Concepts and Applications of Voronoi Diagrams*. John Wiley and Sons.

Passarelli, L., Hainzl, S., Cesca, S., Maccaferri, F., Mucciarelli, M., Roessler, D., Corbi, F., Dahm, T., and Rivalta, E. (2015). Aseismic transient driving the swarm-like seismic sequence in the Pollino range, Southern Italy. *Geophysical Journal International*, 201(3):1553–1567.

Peterson, J. (1993). Observations and modeling of seismic background noise. Technical report, US Geological Survey Albuquerque, New Mexico. Open-file report 93-322.

Poggi, V. and Fäh, D. (2010). Estimating Rayleigh wave particle motion from three-component array analysis of ambient vibrations. *Geophysical Journal International*, 180(1):251–267.

Prieto, G. A., Beroza, G. C., Barrett, S. A., López, G. A., and Florez, M. (2012). Earthquake nests as natural laboratories for the study of intermediate-depth earthquake mechanics. *Tectonophysics*, 570:42–56.

Revenaugh, J. (1995). A scattered-wave image of subduction beneath the transverse ranges. *Science*, 268(5219):1888–1892.

Rietbrock, A. and Scherbaum, F. (1999). Crustal scattering at the ktb from a combined microearthquake and receiver analysis. *Geophysical Journal International*, 136(1):57–67.

Ringdal, F. and Kværna, T. (1989). A multi-channel processing approach to real time network detection, phase association, and threshold monitoring. *Bulletin of the Seismological Society of America*, 79(6):1927–1940.

Rößler, D., Hiemer, S., Bach, C., Delavaud, E., Krueger, F., Ohrnberger, M., Sauer, D., Scherbaum, F., and Vollmer, D. (2008). Rohrbach/vogtland seismic array 2008, univ. potsdam.

Rost, S. and Thomas, C. (2002). Array seismology: methods and applications. *Reviews of geophysics*, 40(3):1–2.

Rost, S. and Thomas, C. (2009). Improving seismic resolution through array processing techniques. *Surveys in Geophysics*, 30(4-5):271–299.

Roux, P., Cornuelle, B. D., Kuperman, W., and Hodgkiss, W. (2008). The structure of raylike arrivals in a shallow-water waveguide. *The Journal of the Acoustical Society of America*, 124(6):3430–3439.

Rubinstein, J., Ellsworth, W., McGarr, A., and Benz, H. (2014). The 2001-present induced earthquake sequence in the Raton Basin of Northern New Mexico and Southern Colorado. *Bulletin of the Seismological Society of America*, 104(5):10.1785/012014.

Rüssler, D., Hiemer, S., Bach, C., Delavaud, E., Krüger, F., Ohrnberger, M., Sauer, D., Scherbaum, F., and Vollmer, D. (2009). Small-aperture seismic array monitors vogtland earthquake swarm in 2008/09.

Ružek, B. and Horálek, J. (2013). Three-dimensional seismic velocity model of the west Bohemia/Vogtland seismoactive region. *Geophysical Journal International*, 195(2):1251–1266.

Saccorotti, G. and Del Pezzo, E. (2000). A probabilistic approach to the inversion of data from a seismic array and its application to volcanic signals. *Geophysical Journal International*, 143(1):249–261.

Sambridge, M. (1999). Geophysical inversion with a neighbourhood algorithm-I. Searching a parameter space. *Geophysical Journal International*, 138(2):479–494.

Sasaki, S. (1998). Characteristics of microseismic events induced during hydraulic fracturing experiments at the Hijori hot dry rock geothermal site. *Tectonophysics*, 289:171–188.

Scherbaum, F., Gillard, D., and Deichmann, N. (1991). Slowness power spectrum analysis of the coda composition of two microearthquake clusters in northern switzerland. *Physics of the earth and planetary interiors*, 67(1-2):137–161.

Scherbaum, F., Krüger, F., and Weber, M. (1997). Double beam imaging: mapping lower mantle heterogeneities using combinations of source and receiver arrays. *Journal of Geophysical Research: Solid Earth*, 102(B1):507–522.

Schweitzer, J., Fyen, J., Mykkeltveit, S., and Kværna, T. (2012). Seismic Arrays. *New Manual of Seismological Observatory Practice 2 (NMSOP-2)*, (December):1–80.

Sen, M. K. and Stoffa, P. L. (2013). *Global optimization methods in geophysical inversion*. Cambridge University Press.

Sick, B. and Joswig, M. (2016). Combining network and array waveform coherence for automatic location: examples from induced seismicity monitoring. *Geophysical Journal International*, 208(3):1373–1388.

Spetzler, J. and Dost, B. (2017). Hypocentre estimation of induced earthquakes in Groningen. *Geophysical Journal International*, 209(1):453–465.

Spudich, P. and Bostwick, T. (1987). Studies of the seismic coda using an earthquake cluster as a deeply buried seismograph array. *Journal of Geophysical Research: Solid Earth*, 92(B10):10526–10546.

Stipčević, J., Kennett, B. L., and Tkalčić, H. (2017). Simultaneous use of multiple seismic arrays. *Geophysical Journal International*, 209(2):770–783.

Tadokoro, K., Ando, M., and Nishigami, K. (2000). Induced earthquakes accompanying the water injection experiment at the Nojima fault zone, Japan: seismicity and its migration. *Journal of Geophysical Research*, 105(B3):6089–6104.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors, S. . . (2019). SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python. *arXiv e-prints*, page arXiv:1907.10121.

Wathelet, M. (2008). An improved neighborhood algorithm: Parameter conditions and dynamic scaling. *Geophysical Research Letters*, 35(9):1–5.

Wathelet, M., Jongmans, D., Ohrnberger, M., and Bonnefoy-Claudet, S. (2008). Array performances for ambient vibrations on a shallow structure and consequences over Vs inversion. *Journal of Seismology*, 12(1):1–19.

Weber, M. and Wicks, C. W. (1996). Reflections from a distant subduction zone. *Geophysical Research Letters*, 23(12):1453–1456.

Willacy, C., van Dedem, E., Minisini, S., Li, J., Blokland, J., Das, I., and Droujinine, A. (2018). Application of full-waveform event location and moment-tensor inversion for Groningen induced seismicity. *The Leading Edge*, 37(2):92–99.

Williams, E., Fernandez-Ruiz, M. R., Magalhaes, R., Vanthillo, R., Zhan, Z., Gonzalez-Herraez, M., and Martins, H. F. (2019). Teleseisms and microseisms on an ocean-bottom distributed acoustic sensing array.

Zywicki, D. J. (1999). *Advanced Signal Processing Methods Applied to Engineering Analysis of Seismic Surface Waves*. PhD thesis, Georgia Institute of Technology.