UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Fisica "E.R. Caianiello"

XXXIII Ciclo di Dottorato in Matematica, Fisica ed Applicazioni

Tesi di Dottorato in

Ambient seismic noise to image and monitor the propagation medium elastic properties in georesource exploration and exploitation areas

Tutor:
Prof.re **Paolo Capuano**
Co-Tutor:
Dott. **Maurizio Vassallo**
Coordinatore:
Prof.re **Carmine Attanasio**

Dottorando:
Paola Forlenza
Matr.8800200030

Anno Accademico 2020/2021

Contents

Introduction

Traditional seismological methods for the study of the Earth structure are based on the information that can be extracted from direct waves emitted by seismic sources such as earthquakes or explosions. Useful information about the Earth interior, that is not directly available, can also be contained in the ambient seismic noise sources. In the last 15 years the interest in these alternative methods has grown because they are particularly advantageous both for economic reasons and for overcoming some limitations of the traditional techniques. Indeed, they offer the opportunity for studying areas without a localized natural or artificial sources and can be used to assess the aseismic response of the subsurface. Correlations of ambient seismic noise traces are used to reconstruct the Earth impulse responses (Green's function) between two receivers as if a source was placed at one of them. From these information it is possible to reconstruct the elastic properties of the propagating medium in terms of spatial and temporal velocity variations of the seismic waves that travel through it, performing an *ambient noise tomography*. Velocity changes in the propagation medium are closely related to the physical mechanisms triggering natural or inducing earthquakes processes.

The aim of this thesis is the application of methods based on ambient noise analysis to the areas where industrial operations for the geo-resources exploration and exploitation take place. Any energy technologies can originate several environmental risks among which seismic risk, therefore application of efficient methods can enhance the ability to monitor and reduce seismic risks in specific areas.

This work was carried out within the three-year project *S4CE* (*Science For Clean Energy*) started in 2017 and funded by the European Union's Horizon 2020 research and innovation program (https://science4cleanenergy.eu/). This multi-disciplinary consortium aimed to improve the knowledge on the mechanisms concerning the sub-surface geo-energy operations in order to better monitor, quantify and reduce the environmental risks that may derive from it. The Università degli Studi di Salerno was one of the main partners and dealt with the management of data on microseismicity and the study of induced seismicity. In the 2019, the Università degli Studi di Salerno started to coordinate the three-year project *MATISSE* (*Methodologies for the AssessmenT of anthropogenic environmental hazard: Induced Seismicity by Sub-surface geo-resources Exploitation*) financed by the Italian Ministry of Education, University and Research, PRIN 2017. This thesis work is also part of this project, developed for detecting and quantifying the potential environmental hazards related to sub-surface geo-energy exploitation.

It was analysed the continuous noise recorded of two sites where different energy technologies was implemented for the exploitation of their geo-resources. This enabled testing of the methodology in different types of areas and industrial operations, and to assess differences and limitations. One of the study areas is the shale gas site of Wysin, which belongs to the Stara Kiszewa concession of Polish Oil and Gas Company (PGNiG) in Pomerania region of the northern Poland. In this region, the operations for the exploration and exploitation of the hydrocarbon was conducted in 2016 by using the hydrofracturing technique. The second investigated site is one of the test fields of the European project S4CE, located in St. Gallen, northern Switzerland. This area was involved in a deep geothermal energy project for power production using water in aquifers at a depth of over 4 km. The project was realised in 2013 by the St. Gallen public utilities company (Sankt Galler Stadtwerke) in the proximity of an existing fault zone that offers greater permeability in favour of industrial activities.

When studying areas where anthropogenic activity related to the exploitation of subsoil resources can generate induced seismicity, it is essential to increase the detection of weak events to improve knowledge about the seismicity of the area and its related consequences. In seismology, when dealing with low signal-to-noise recordings, traditional event detection methods are often unable to recognise all the weak events hidden within the seismic noise. Machine learning techniques can be a useful tool to improve automatic event detection by recognising the similarity between events. In this study, the implementation of an automatic waveform-based detection system for recordings of the St.Gallen deep geothermal site is reported.

The first chapter of the thesis provides an overview of the most common used energy technologies highlighting their induced seismicity potential and the difficulties in dealing with this issue. Then, the state of the art in using ambient noise for propagation medium properties investigation is reported, followed by a brief description of seismic noise sources. The second chapter explains the detailed methodology used to conduct a tomographic study with long ambient-noise time series. In the third and fourth chapters there are the applications of the exposed methods to the areas of interest Wysin and St. Gallen, respectively. The final chapter contains the final conclusions of the thesis.

Chapter 1

Ambient seismic noise to study medium properties in geo-resource exploration and exploitation areas

From methods based on ambient noise analysis it is possible to extract the Green's function from the cross-correlation of random ambient noise recorded at couples of stations (Weaver and Lobkins, 2001). The information obtained from the cross-correlation analysis can be used in two main applications: imaging and monitoring. The former allows to construct group and phase velocity maps of the surface waves at different scales, and the latter allows to asses the temporal seismic velocity variations (Vassallo et al., 2016). The use of ambient noise in seismic areas has several advantages compared to traditional methods. It solves the problem of uncertainties in velocity models deriving from earthquake location and offers more data availability for the specific area. Another advantage is the enhanced resolution because measurements are made between regularly spaced receivers, which may lie much closer to one another than to earthquakes (Shapiro et al., 2005). It is also possible to investigate the properties of the medium in an area at any time without the occurence of earthquakes, and in any world area even where seismicity is low or absent. This emerging approach avoids the need for high cost and environmentally harmful active seismic sources, therefore also allows surveys in sensitive urban areas and in protected nature reserves. Because localized natural and artificial sources are not needed, ambient seismic noise analysis can be used to assess the aseismic response of the subsurface. Methods based on ambient noise analysis are simple to implement and can be performed continuously to follow the evolution of the

reservoir at low cost, avoiding the traditional use of expensive artificial impulsive sources (e.g. explosions). However, these methods are based on the assumption of spatial and temporal uniformity of seismic noise sources. If this condition does not occur, the interpretation of seismic velocities cannot be done properly due to privileged directions of noise propagation.

These methods are particularly useful to examine geo-resource exploration and exploitation areas in the phase that precedes the operations. In fact, it is possible to probe the targeted area in order to characterize the reservoir and its relation to pre-existing geological formations, and better reconstruct any variations in the elastic properties of the propagation medium during or after the operations. This leads to improved knowledge of the basic physics of triggering natural or induced earthquakes processes and, consequently, to improved competences in monitoring and reducing seismic risks.

This chapter describes the most commonly used technological energies and the corresponding physical conditions that can cause seismicity.

Induced seismicity associated to underground industrial activities is currently a relevant topic of interest for scientific, political, and social community. In recent years, the development of the industrial activities in the vicinity of highly populated regions combined with many cases of earthquakes felt by locals increased the public concern about this topic. In particular, a significant growth of the interest of scientific research in induced seismicity occurred after the November 2011 induced earthquake in Oklahoma, indeed an increasing yearly number of scientific papers on the topic was recorded (Grigoli et al., 2017). At the same time, also non-scientific community increased its attention in the theme. Analysis performed using Google trends show the global web queries in that period containing the non-technical term "fracking earthquakes". The observed seismic activity associated with energy technologies typically has low energy ($M \leq 3$). Earthquakes can be felt by the population but are small enough to cause damage. However, there are well documented cases in which human activity has been associated with the generation of even high magnitude earthquakes which will be presented in this chapter.

Induced seismicity is a complex problem to address, the state of the art of works that use the ambient noise as a tool of propagation medium investigation is reported in the following sections, in particular the applications on the type of areas studied in this thesis.

The chapter concludes with a description of seismic noise sources and some studies that revealed a variation in its recordings due to measures on human activities to manage the Covid-19 pandemic in the first part of 2020.

1.1 Energy technologies and their induced seismicity potential

The planning and implementation of operations that are performed to obtain energy from the subsoil are essential for our society. Within the framework of energy supply a variety of energy technologies have been developed to satisfy the increasing demand for energy and better exploit the different geo-resources that the subsoil offers.

In the last decades, in addition to traditional (*conventional*) technologies, new energy technologies, called *unconventional,* have been developed with the aim to perform or improve production in reservoirs where there are no advantageous natural conditions for geo-resources extraction. These conditions, unlike the conventional case, include the low permeability of the rocks, the absence of a pressure gradient that is sufficient for fluids to migrate upwards, the absence of connection pathways (e.g. fractures) between the formation and the well, and the high depths location of the geo-resource.

All these energy activities have some potential risks such as emission of gases, faulty wells casings, contamination of the water table, and seismicity that are referred to as *TechNa* (Technological-Natural) accidents. Conversely, it is also possible that natural events have an impact on industrial facilities causing technological accidents. In this case, it is talking about *NaTech* events such as the release of toxic substances, fires, and explosions (Krausmann et al., 2019). In both cases, the possibility of cascading events can occurs making industrial risks a complex issue. Therefore, the implementation of any energy technology is not immune to risks (No risks, no energy!).

In this work we focus our attention on the risk of seismicity that can be generated by human activities.

Seismicity attributable to anthropogenic activities can be distinguished into two categories:

- *induced seismicity*, when human activities can cause variations in stress field that is large enough to produce a seismic event in area of quiescent pre-existing faults;

- *triggered seismicity*, for which a small perturbation generated by human operations is sufficient to move the system from a quasi-critical to an unstable state. In this case the event would have occurred anyway, but probably in different geological times. The operations only accelerate the process of tectonic stress release.

For the sake of simplicity, only the term "induced seismicity" will be used to cover both categories and make a distinction from natural seismicity result-

ing from tectonic activity.

Energy technologies have potential for induced seismicity because they involve activities that can alterate chemical or thermal properties, and/or change stress field in the shallow Earth's crust. The physical conditions for shallow earthquakes result from slip along a pre-existing fault are described by the Coulomb criterion. Considering the normal σ and shear τ stresses acting across a fault plane, and the pore pressure ρ assuming the presence of fluids (Fig. 1.1.1 a), they are related by the equation

$$\tau = \tau_0 + \mu\,(\sigma - \rho)$$

where τ_0 is the cohesion, μ is the friction coefficient, and $(\sigma - \rho)$ is the effective stress (the pore pressure ρ counterbalances the effect of the normal stress σ).

Graphically, the rupture threshold is identified in the plane $(\sigma - \rho, \tau)$ by the line with slope μ (Fig. 1.1.1 b).

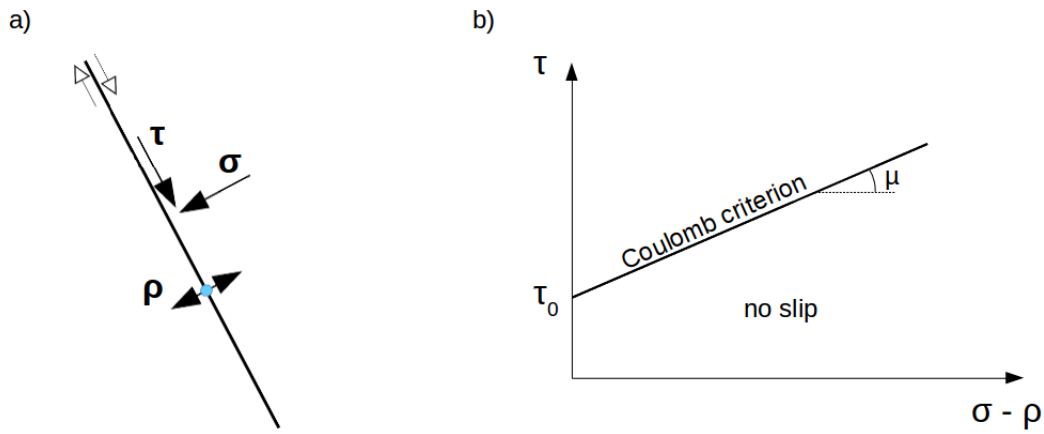


Figure 1.1.1: *a) Schematisation of the normal stress σ, the shear stress τ , and the pore pressure ρ acting across a fault plane. b) Graphical representation of the Coulomb criterion. No slip occurs for points below the critical line defined by the slope μ.*

According to the Coulomb criterion, the slip condition can be achieved by a decrease of the normal stress, an increase of the pore pressure, and/or an increase of the shear stress. In the case of an injection well in contact with pre-existing faults (Fig. 1.1.2 a) through porous and permeable layers, the fluid injection activities increase the pore pressure reducing the effective normal stress acting on the active fault, and thus promoting slip on the fault. Instead, in the case with no hydrologic connection (Fig. 1.1.2 b),

the industrial operations can alter the stress conditions around faults by involving mass/volume changes. Seismicity can be generated by industrial operations even without the presence of faults (Fig. 1.1.2 c). In this case, the rupture is attributable to high pressure fluid injection in impermeable layers. The rupture process takes place when the fluid pressure exceeds the minimum principal stress of the in situ stress field.

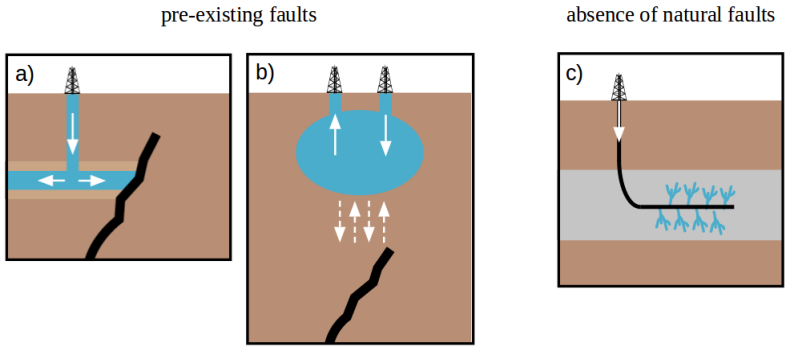


Figure 1.1.2: *Possible physical mechanisms that can cause induced seismicity in b), c) presence or (d) absence of faults. Figure modified from Grigoli et al. (2017).*

The most common energy technologies concern geothermal energy, hydrocarbons extraction, fluids injection, and reservoires creation (Fig. 1.1.3).

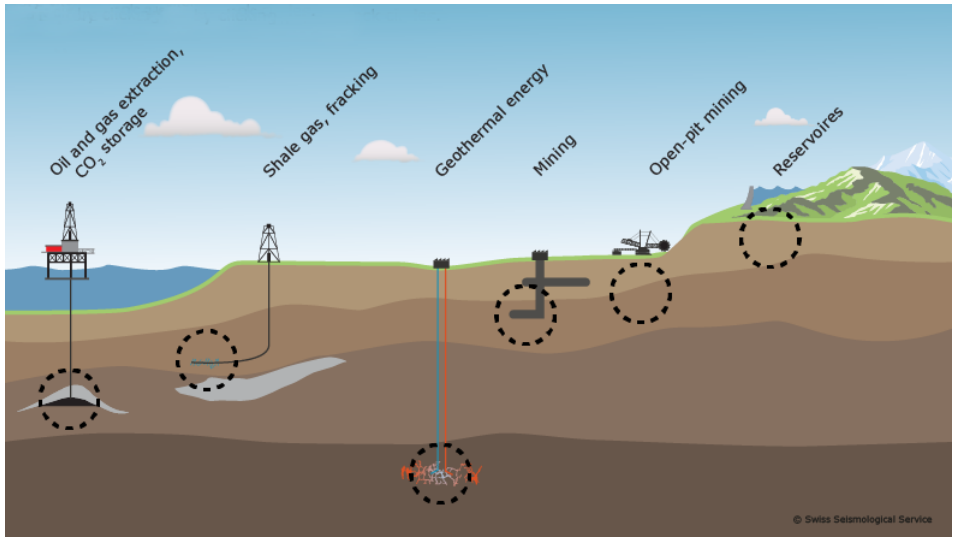


Figure 1.1.3: *Geotechnical applications that can cause induced seismicity. Source: Swiss Seismological Service (SED), http://seismo.ethz.ch.*

Geothermal energy production exploits the natural heat of the Earth as an energy resource. Depending on the depth and heat used, two forms of geothermal energy can be identified: shallow geothermal energy and deep geothermal energy (Fig. 1.1.4), which are used for different purposes based on the subsoil temperatures. Shallow geothermal energy penetrates no more than 400 meters into the subsoil and makes use of layers with temperatures between 8°C and 20°C. This traditional geothermal energy includes geothermal probes or geothermal loop systems and the use of groundwater heat, energy piles and geostructures (Fig. 1.1.4 a). Deep geothermal energy covers depths ranging from 400 meters. At a depth of 400-2000 meters, temperatures between 20 and 70 °C can be found, whereas at a depth of 4000 meters, temperatures between 150 and 200 °C can be reached. Deep geothermal energy can involve enhanced geothermal systems (EGS) or hydrothermal systems (Fig. 1.1.4 b).

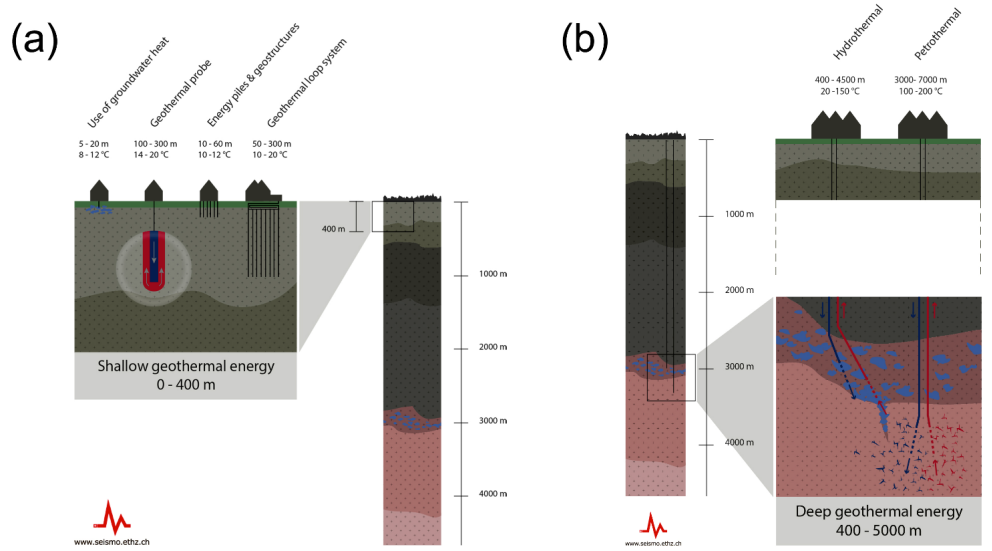


Figure 1.1.4: *Schematisation of (a) shallow geothermal energy and (b) deep geothermal energy. Source: Swiss Seismological Service (SED), http://seismo.ethz.ch.*

EGS exploit the thermal energy stored typically into crystalline basement rocks characterized by low permeability. These systems use the unconventional technique of high-pressure fluids injection (hydraulic stimulation) to create or reactivate fractures in the future reservoir, increasing the permeability of the rocks, and subsequently allowing fluids heating and pumping on the surface. Hydrothermal systems use existing aquifers in sediments and are only implemented at certain sites with high permeability, e.g. in

fault zones or volcanic areas. In this case, the fluids injection is necessary to have a vector to transport thermal energy from deep hot formations to the surface through naturally occurring fractures in the permeable rocks. In each of these deep geothermal development, produced fluid is replaced by injected fluid in order to maintain fairly constant reservoir pressure (close to the pre-production value) enhancing the exploitation of the energy resource. The injection or extraction of fluid has the potential to induce seismic activity. The number of earthquakes and their magnitude may depend on the rock extraction volume and the fluid flow rate. Also a large temperature difference between the injected fluid and the geothermal reservoir can lead to seismicity. Significant cooling of the reservoir rocks can cause contraction, reducing the confining pressure (vertical stress) and thus the release of local stress. According to current understanding, deep geothermal energy is the only form that can lead to induced earthquakes because shallow geothermal energy involves shallow depth and closed systems.

Oil and gas production involves pumping hydrocarbon fluids: petroleum and natural gas. Fluids in a conventional oil or gas reservoir are usually under significant natural pressure thus, when a well is drilled, they flow to the surface, generally aided by pumping. Reservoirs often can reach an insufficient pressure value for hydrocarbon recovery and some technologies can be used to increase production such as *secondary* and *tertiary recovery*. For this purpose both operations involve fluids injection into the subsurface and maintain reservoir pore pressure. Secondary recovery often uses water injection (waterflooding) whereas tertiary recovery, known as *enhanced oil recovery (EOR)*, often injects carbon dioxide (CO_2). Hydrocarbons often are extracted together with aqueous fluids (groundwater) containing high amounts of dissolved solids and salts (brine). Commonly these fluids are reused in the recovery process or disposed of by injection into the deep subsurface in wells located at some distance from the oil or gas extraction ones.

A recent technology in the oil and gas industry is the horizontal drilling technique together with the use of *hydraulic fracturing* (HF), or colloquially *fracking* (Fig. 1.1.5). It is an unconventional technique that optimises the production of reservoir in which extremely low-permeability rocks, such as shale formations, prevent the hydrocarbons from easily flowing into a production well without any stimulation. Hydraulic fracturing consists in fluid injection at high pressure into the reservoir rocks such that the fluid pressure creates new fractures or stimulates slip across pre-existing faults or fractures. The fluid consists of water with chemical additives and proppants (3-4%) which maintain the fractures open for the needed time. After hydraulic fracturing and before the start of hydrocarbon extraction another

fluids movement is carried out. These are the flow-back operations that allow the hydraulic fracturing fluids to flow back to the surface for reuse or disposal. The injection of fluid to facilitate oil and gas withdrawal or for permanent disposal and fluids related to hydraulic fracturing, similar to fluid injection for geothermal systems, may cause induced seismic events. As already mentioned, energy production activities can also produce by-products such as wastewater which are then reused for other purposes or returned to the subsoil.

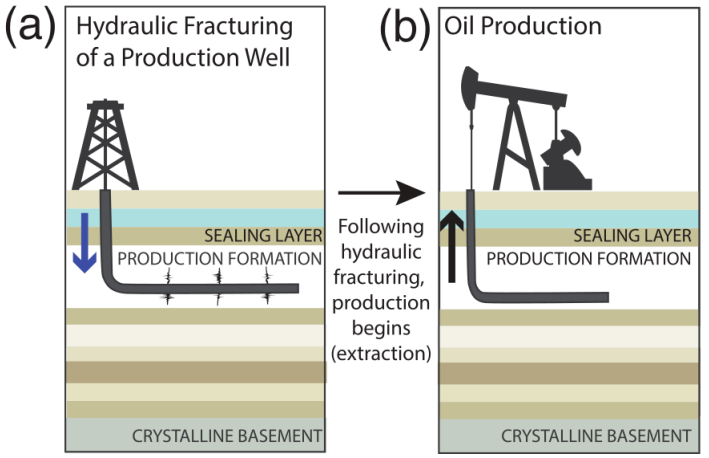


Figure 1.1.5: *schematisation of (a) HF of production well and (b) the following oil production (Rubinstein and Mahani, 2015).*

Carbon Capture and Storage (CCS) technologies allow capturing and sequestering CO2 in suitable deep formations. Carbon dioxide is another by-product that can be injected into oil and gas reservoirs to enhance oil recovery or stored permanently underground reducing carbon emissions to the atmosphere. As with other technologies, involving underground fluid migration, CCS may also cause changes to subsoil properties that could generate seismicity.

Reservoires creation using dams to store water permits the process for generating electricity from hydropower. Filling of reservoir at sites causes stress changes due to load of water and thus can trigger seismicity where the underlying rocks are in a state close to rupture. This is because the additional water load is too small to create new fractures in the earth's crust, thus reservoirs might only advance earthquakes which would otherwise have occurred at a later time.

1.1.1 Critical cases of induced seismicity

Induced seismicity has been observed worldwide related with several industrial activities (Fig. 1.1.6). The connection between human activities and seismicity is an important issue in areas especially where natural seismicity is low or non-existing. Many of these relevant cases were documented in different parts of the world, some of which led to the suspension or reduction of the related industrial activities especially because occurred nearby urbanized areas.

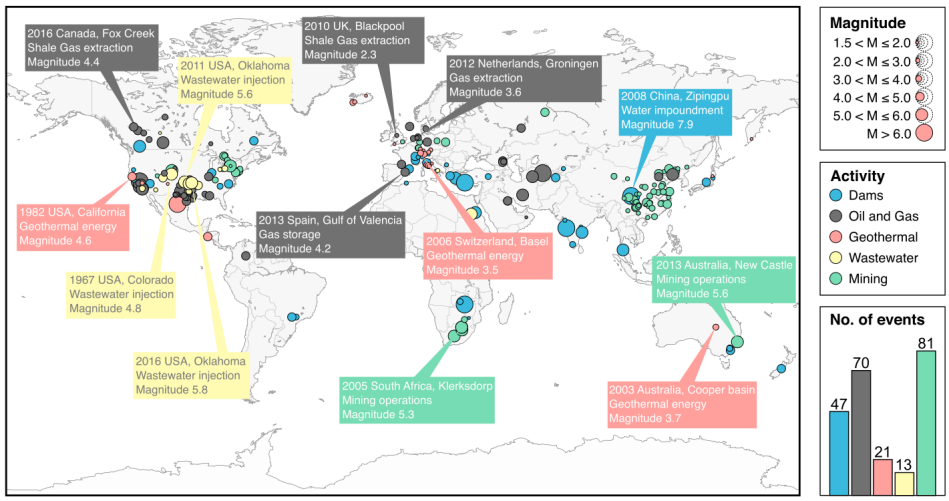


Figure 1.1.6: *Global distribution of anthropogenic seismicity until August 2016. The maximum magnitude is reported at each site (Grigoli et al., 2017).*

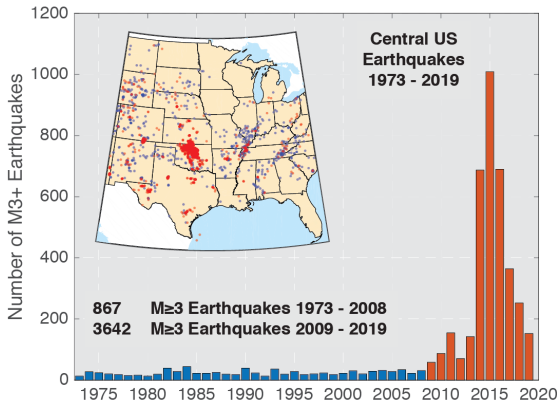


Figure 1.1.7: *Number of earthquakes with M>= 3.0 in the central and eastern United States from 1973 to 2019. Source: U.S. Geological Survey (USGS), https://www.usgs.gov.*

A relevant phenomenon is the fact that the seismicity rate in central and eastern United States dramatically increased since 2009 (Ellsworth, 2013). The figure 1.1.7 shows the annual number of earthquakes of magnitude three and larger in the central and eastern United States in the period 1973-2019. Since 2009, United States experiencing at least 58 events/year, and at least 100 events/year since 2013, compared with an average rate of 25 events/year observed from 1973–2008. In 2015 a peak was recorded with 1010 earthquakes.

The rapid increase in earthquake activity mainly concerned the ***Oklahoma***, a relatively low seismicity area that from 2009 experienced a number of earthquake greater than California, known for its tectonic seismicity (Fig. 1.1.8). The strongest earthquake recorded in Oklahoma occurred on September 3, 2016. An event of 5.8 Mw which caused damage to various buildings, especially in the Pawnee city. This rapid change in Oklahoma and surrounding areas is highly correlated with oil and gas exploration, specifically the deep injection of wastewater arising from the hydraulic fracturing process (Keranen et al., 2013). The 2016 Oklahoma sequence is the largest induced ones associated to these kind of operations.

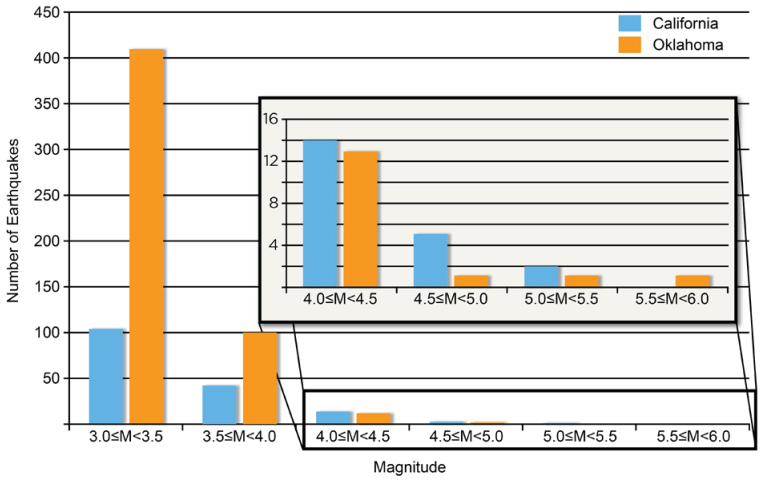


Figure 1.1.8: *Comparison of the number of earthquakes $M \geqq 3$ between California and Oklahoma in 2016. Source: U.S. Geological Survey (USGS), http://earthquake.usgs.gov.*

The most important case of induced seismicity in Europe is the ***Groningen field***, a natural gas production site in the Netherlands. Since the 1960s, natural gas drilling has converted a tectonically stable area into a seismic one (van Thienen-Visser and Breunese, 2015). The Mw 3.6 Huizinge largest event occurred on 16 August 2012. This episode reactivated the interest

of the population in induced seismicity, pushing the government and the company to build a large microseismic monitoring network composed of more than 60 borehole stations.

Two important European cases of induced seismicity related to deep geothermal energy projects developed nearby urban areas occurred in ***Basel*** and ***St. Gallen***, both cities in Switzerland. The Basilea project had the aim of developing a geothermal power plant from an EGS systems. The stimulation experiments started on December 2006 and during the injection phase, an intense seismic activity was recorded reaching an event of Ml 2.6 at the peak of the injection surpassing the threshold of the safe protocol (Haering et al., 2008). Then the well was shut-in, but on December 2006 a Ml 3.4 earthquake occurred prompting a strong public debate and a considerable number of damage claims. Between 2006 and 2008 seismic events were felt by local residents and considered connected to geothermal energy development. A conducted seismic-hazard evaluation led to the cancellation of the project in December 2009. The St. Gallen project in 2013 aimed to make use of naturally circulating water in existing aquifers at a depth of over 4000 meters for power production. The drilling work for an hydrothermal system took place in a fault zone whose seismic potential was rated as low. In July 2013, after an injection test and two acid injections a sequence of microearthquakes was induced. Then an uncontrolled gas release (gas kick) from the formation led to a Ml 3.5 earthquake on July 20, 2013 (Obermann et al., 2015). While the earthquake was noticed extensively by the public, very few reports of damage were received in comparison with Basel. The companies concluded the first phase of its geothermal energy project following the production tests, and the project was terminated permanently in spring 2014 following further investigation.

The first documented example in Europe of seismicity related to hydraulic fracturing activities is the case of ***Blackpool***, a shale gas site in UK. On 1 April 2011, a felt seismic event with a 2.3 Ml occurred several hours after the injection operations was terminated, preceded by a low seismic activity. The seismicity resulted from the interaction of hydraulic fracturing and a pre-existing natural fault located some 300 m from the injection well (Clarke et al., 2014). Fracking fluids penetrated into the fault changing its stress field reactivating it. After the crises, the industrial activity was suspended and a local monitoring network was installed.

In Italy, the ***Emilia Romagna*** region is an important focus area where the difficulty and importance of discriminating between natural and induced seismicity close to industrial sites has been highlighted (Grigoli et al., 2017). The Emilia plays an historical fundamental role in the national exploration and

production of hydrocarbons. Being tectonically active region, the reverse and strike-slip fault systems facilitate vertical communication between the different levels. Starting from the late May 2012, the region was shaken by an intense seismic sequence characterized by two mainshocks on 20 and 29 May with Ml 5.9 and 5.8 respectively, occurred in the proximity of the Cavone oil field. The sequence caused tens of victims, the destruction of public and private buildings and serious damage to the economy. A technical-scientific commission (ICHESE, 2014) was appointed for evaluating the possible relationships between anthropogenic actions and increase of seismicity observed in the area. The commission did not exclude that the oil extraction operations performed in the Cavone oil field had its contribution, although there was no clear link with the May 2012 earthquake sequence (ICHESE, 2014). This complex case is still affected by numerous studies, Pezzo et al. (2018) work is the most recent which states that the seismic sequence was driven by an increase pore pressure at the base of a pre-existing fault system.

Two simultaneous studies (Grigoli et al., 2018; Kwang-Hee et al., 2018) published in Science analysed the possible connection between the 2017 Mw 5.5 ***Pohang*** (South Korea) earthquake and a nearby enhanced geothermal system site. The Pohang earthquake was the largest and most damaging event to have occurred in South Korea and at any EGS site worldwide. Combined seismological and geodetic investigations aided to understand the nature of the event that is suggested by taking into account the location of earthquakes, their depths, the temporal relationship between seismicity and fluid injection, and the lack of seismicity in the area before the EGS was established. The mainshock occurred about two months after the last stimulation activities within 2 km distance or less of the geothermal site at shallow depths of 3 to 7 kilometres. All these factors suggest that the Pohang earthquake was induced. Similar to Italy, the Korean government has also appointed an independent expert commission to evaluate if the event was triggered or induced by the nearby stimulation activities.

1.1.2 Challenges in monitoring, discrimination, and management of induced seismicity

Induced seismicity is not an easy issue to face and the related phenomena are intricate and variable from case to case, and the correlation with the process parameters is far from being completely understood. Cases of seismicity that are thought to be associated with industrial activities have been observed even at a distance of several tens of kilometres from the systems and after tens of years (Keranen et al., 2013; Mulargia and Bizzarri, 2014). The

complexity of the problem is also due to the deep socio-economic impact on industrial companies that manage underground activities and on local communities. For a safe and sustainable future exploitation of geo-resources, a proper discrimination between natural and anthropogenic seismic events is one of the primary objectives in current research areas.

Microseismic monitoring plays a key role in better understanding the physical mechanisms governing induced seismicity. Characterization of microseismicity and its spatio-temporal evolution might be used to track fluid migration and identify potential interactions with known pre-existing faults. To ensure optimal monitoring of seismic activity, a dense network with adequate characteristics is necessary to achieve the objectives. Monitoring infrastructure performance depends on many factors including the number of stations and its geometric distribution, the type of instrumentation, the noise level, and the installation conditions. An efficient seismic monitoring network has the ability to detect weak seismic signals, obtain high precision locations, and reliable magnitude and source parameters. This enables an early and accurate microseismicity analysis solving many problems related to the interpretation of the results in particular for discriminating between induced and natural seismicity. The design of a monitoring network can be optimised by evaluating its performance prior to its development or integration of an existing ones, before the start of extraction operations. A criterion for evaluation, applied to the Wysin site in Poland, is described in the work of López Comino et al. (2017).

A high quality monitoring network must be combined with advanced data analysis methods to provide reliable results for interpretation. These procedures must be fast, robust, and automated for a good management of big datasets. Monitoring is also an important tool which affects the management of industrial activities and aid the related decisional protocols in case of critical situations. Several scientific and sociological problems are still completely unsolved. Production and/or monitoring data are often not accessible to the scientific community, this restriction does not allow the reproducibility and reliability of results. Correct communication campaigns are strongly necessary because a misinformation can alter the perception of the general public and have a negative socioeconomic impact (a clear example is given by the 2012 Emilia sequence). Adequately open data policy would give valuable help not only to improve the scientific knowledge about the physical processes governing induced seismicity but also to increase the social acceptance of the related industrial activities (Grigoli et al., 2017).

1.2 State of the art of ambient seismic noise as a tool of propagation medium investigation

The work of Shapiro et al., published in Science in 2005, was the precursor in application of methods using ambient seismic noise. Starting from one month of seismic noise recordings, a high-resolution tomographic image of the Californian crust was produced allowing to well discriminate the different compositions of the area. In the last 15 years, seismic interferometry methods have grown considerably involving studies at local (Roux, 2009; Picozzi et al., 2009; Vassallo et al. 2019), regional (Yao et al. 2006; Vassallo et al. 2016) and continental (Yang et al., 2007) scales. Applications have developed in different geological environments, ranging from fault zones (Brenguier et al., 2008; Roux, 2009; Zigone et al. 2015; Vassallo et al. 2016), volcanic areas (Brenguier et al., 2016; Cabrera-Pérez et al. 2020), urban areas (Vassallo et al. 2019), marine environment (Lin et al., 2006; de Ridder & Dellinger, 2011; Mordret et al., 2013), to underground mines (Olivier et al., 2015). In recent years, these methods have become useful for imaging and monitoring geo-resource exploitation sites. These applications are the main research reference for the purpose of this thesis.

Obermann et al. (2015) highlighted the potential of seismic noise techniques to monitor the St. Gallen deep geothermal site in Switzerland. The study of the temporal evolution of velocity and waveform coherence in the coda of the cross-correlations, had permitted to observe a significant loss of waveform coherence parallel to the injection tests and prior to the Ml 3.5 event. The ability of the ambient noise analysis to observe the aseismic response of the reservoir to well operations, led to attribute medium changes to a local perturbation of the medium escluding local source changes.

The application of the ambient noise correlation technique in the context of a geothermal reservoir was also performed by Lehujeur et al. (2015), characterising more than 4 years noise data recorded near two deep geothermal sites (Soultz-sous-Forêts and Rittershoffen) in northern Alsace, France. They analysed properties of the computed cross-correlation functions that are crucial to perform the imaging and monitoring of the site. In particular, they analysed the seismic noise distribution in the 0.2 - 5 s period range showing ranges in which Green's functions cannot be reliably reproduced, making dispersion measurements difficult. Secondly, they evaluated the time stability of the correlation functions coda, promoting the use of nocturnal noise for future monitoring temporal variations induced by the geothermal activities inside the reservoir.

Later, Lehujeur et al. (2018) shows the potential of a dedicated dense net-

works to image a deep geothermal reservoir at several tens of kilometers. For the same area of their previous studies (Lehujeur et al., 2015), the dense network of 288 stations enabled to analyse only one-month ambient noise records, deriving correlation functions with quality comparable to those extracted with records of more than one year from a network of 34 stations. The obtained Rayleigh wave group velocity maps and 3-D S-wave velocity model of the considered area exhibit significant velocity anomalies that are consistent with previous geological and geophysical observations.

Studies on hydrocarbon reservoirs promote ambient seismic noise tomography as a valid alternative to traditional methods that are used in modelling S-wave velocity. Chmiel et al., 2019 analysed one of the largest gas fields in Europe, the Groningen site in the northern of the Netherlands. Performing depth inversion of surface waves from cross-correlation of 1 month of ambient noise data, a detailed S-wave models up to 1 km depth were estimated. These models reflect the geological structures, and are consistent and complementary with the existing S-wave models of the area, therefore they can be used in seismic risk modelling.

Mordret et al. (2013) proved the applicability of the ambient noise surface wave tomography for a wide frequency range to characterise the seabed structure at the Valhall oil field, in the Norwegian North Sea at a 70-m water depth. They used continuous seismic noise records from the permanent Life of Field Seismic network (LoFS) to reconstruct the features of the complicated geological structure of the offshore oil reservoir. In this study, the limitations of the methodology relying on assumptions that noise sources should be homogeneously distributed were highlighted. In fact, it was performed an ambient noise surface wave tomography at frequencies below 2 Hz, since at higher frequencies the dominant noise source was the oil platform at the centre of the network.

All the works show that processing seismic ambient noise data to study sites where industrial activities take place is a very useful tool because it allows to improve and integrate the limited knowledge obtained with the standard monitoring of microseismicity. Studying the aseismic response of the reservoir prior to the start of operations permits to better understand and interpret the changes of the medium properties due to the activities. This also enables to recognize the unexpected reservoir dynamics at an earlier stage, which is crucial for the management of activities and the mitigation of environmental risks. Application of seismic interferometry method on geo-resource sites are not sufficiently developed so far. Further works are necessary to validate and improve the techniques, and to understand the possible limits and differences between the different types of industrial site.

1.3 Seismic noise sources

Ambient seismic noise represents a continuous ground oscillation that can be recorded anywhere on the planet. It is produced, in fact, from all sources that are able to stimulate the Earth surface by generating vibrations in it. Vibrations are the result of the interaction between the solid Earth, the atmosphere and the hydrosphere.

Seismic noise sources are classified into two main categories according to their frequency content: natural sources operating at low frequencies ($<$ 1 Hz) and anthropogenic sources operating at high frequencies ($>$ 1 Hz) (Bonnefoy-Claudet et al., 2006).

Vibrations at low frequencies ($<$ 1 Hz) are called *microseism*, whereas those at high frequencies ($>$ 1 Hz) are called *microtremors*.

The ***anthropogenic activity*** carried out on the Earth surface or underground (e.g. vehicular traffic and industrial machinery operation) causes a signal propagating mainly through surface waves at frequencies generally higher than 1 Hz and attenuating at a distance of a few kilometres from the source. The major noise sources of low frequency ($<$ 1 Hz) are ***large-scale oceanic and meteorological phenomena***. Thanks to pressure gradients generated in conjunction with cyclonic events above the oceanic masses, marine wave phenomena are created. Interacting with the sea coasts, they induce the propagation of elastic waves inside the continental crusts. The intensity of these phenomena varies on average with seasonal cycles. ***Wind and local weather conditions*** effects produce a signal with a frequency close to 1 Hz. The wind interacts with objects connected to the ground and causes the vibration which is transmitted into the ground. The generated signals have frequencies mainly related to objects height. For example, the interaction between the wind and the topographical irregularities or trees causes high frequency signals, whereas the interaction of the wind with large objects, such as towers or buildings with own period of oscillation, inputs lower frequency energy into the ground. Other natural sources are ***watercourses***, ***tremors from volcanic activity***, and ***underground fluids migration***. These act in frequency ranges overlapped with those generated by human activity.

Seismic noise is the result of the temporal combination of signals from different type, spatially distributed, often unrelated and continuous sources. Therefore, it can be compared to a stationary stochastic process, which does not have a well-defined phase spectrum (Bormann, 2002).

Ambient seismic noise is not like white noise having a constant amplitude spectrum, but has peaks around 0.07 Hz (*primary microseismic peak*) and

0.2 Hz (*secondary microseismic peak*), both due to ocean waves (Aki and Richards, 2002). For periods longer than 30 s an additional peak, denoted as *Earth hum*, is detectable.

Chapter 2

Seismic interferometry method based on cross-correlation analysis

The procedure to performe an ambient noise surface-wave tomography is divided into several steps. A preliminary analysis of the distribution of the noise source must be conducted because seismic interferometry methods based on cross-correlation analysis work only if the noise sources have an isotropy distribution. In this thesis, this hypothesis was verified with the beamforming analysis based on frequency-wave number method.

The first step of the analysis is the data processing in which interstations Green's functions are reconstructed by the cross-correlation of continuous ambient noise data. Initially, a single station data preparation phase is performed to improve the quality of the signal to be analysed. In this pre-processing phase, operations (such as de-meaning, de-trending, filtering, normalization in time and frequancy domain, etc.) are carried out in aim of removing earthquake signals and instrumental irregularities that tend to obscure ambient noise (Bensen et al., 2007). Afterwords, daily cross-correlations are computed between all possible pairs of stations and temporal stacked to correspond to longer time-series improving the signal-to-noise ratio and emphasizing the coherent signals that correspond to the Green's functions.

The second step consists in the measurement of dispersion curves from cross-correlations which can be achieved with different methods. For example, the group and phase velocities as a function of period can be measured by using traditional frequency-time analysis (FTAN), or the traveltimes of surface waves can be picked directly from the stacked cross-correlation functions.

The choice of method depends on the data availability and how clear the signals are.

The third step of the analysis concerns the tomographic inversion in which the obtained dispersion curves are used as input for solving an inverse problem, to reconstruct 2-D S-waves velocity model of the subsoil. Each map refers to a single period and allows to determine the structure of the model at a given depth. In the tomographic inversion phase, images are obtained as a function of frequency.

The methodology for each step is explained in detail in this chapter.

2.1 Beamforming (f-k analysis)

Application of seismic interferometry methods based on cross-correlation analysis requires a preliminary noise analysis to verify the hypothesis of isotropy distribution of noise sources. For this aim, one of the methods that can be employed is the classical beamforming technique (Rost and Thomas, 2002). The beamforming (or frequency-wave number analysis) is an array methods that allows to derive the directional properties of a signal coming along an array assuming that the seismic wavefield is composed of a finite number of plane waves.

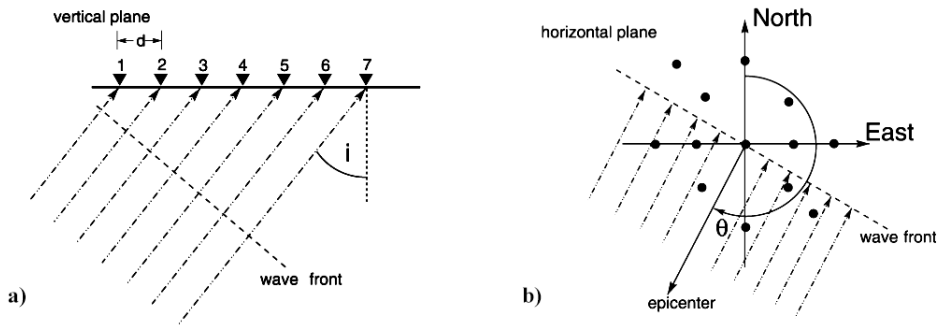


Figure 2.1.1: *(a) The vertical plane of an incident wave front crossing an array at an angle of incidence i. (b) Sketch of the horizontal plane of an incident plane wave arriving with a backazimuth θ (Rost and Thomas, 2002).*

A plane wave can be described by two parameters:

- *Backazimuth* θ that is the angle of the wavefront arriving at the array measured clockwise from the North to the direction towards the epicenter in degrees (Fig. 2.1.1 b).

- *Slowness* s defined as the inverse of the *apparent velocity* V_{app}

$$s = \frac{1}{V_{app}} \tag{2.1.1}$$

The apparent velocity is the velocity at which the wavefront moves and it is related to the effective mean crustal velocity V by

$$V_{app} = \frac{V}{\sin i} \tag{2.1.2}$$

where i is the vertical incident angle.

These two parameters can be combined in the slowness vector $\mathbf{s}$, a vector oriented in the propagation direction (Fig. 2.1.1) with modulus as the reciprocal of wave velocity

$$\mathbf{s} = (s_x,\, s_y,\, s_z) = \left(\frac{\sin\theta}{V_{app}},\, \frac{\cos\theta}{V_{app}},\, \frac{1}{\tan i}\right) \tag{2.1.3}$$

We assume that all array sites are in the same horizontal plane of the wavefront propagation. In this case the vertical apparent velocity component is infinite $\left(V_{app,z} = \frac{V}{\sin 0} = \infty\right)$ and the corresponding slowness component becomes zero $\left(s_z = \frac{1}{V_{app,z}} = 0\right)$

$$\mathbf{s} = (s_x,\, s_y) = \left(\frac{\sin\theta}{V_{app}},\, \frac{\cos\theta}{V_{app}}\right) \tag{2.1.4}$$

When a plane wave arrives at an array, each station records a signal with a certain delay time depending on the slowness vector s of the wavefront and the position of the station in the array. For example, the delay time for the j-th station in vector sintax is

$$t_j = \mathbf{r}_j \cdot \mathbf{s} \tag{2.1.5}$$

where $\mathbf{r}_j$ is the location vector of the j-th station relative to the array reference point.

By bringing the signals into phase, the sum of their coherent part creates constructive interference and then increase the signal-to-noise ratio (SNR). Since the delay times depend on a specific backazimuth and apparent velocity combination, from them it is possible to derive the directional properties of the wavefront crossing the array.

The method operates in the frequency domain. The relation between the parameters of a plane wave and the actual seismic signal is given by the wave number vector $\mathbf{k}$ defined as

$$\mathbf{k} = \omega \cdot \mathbf{s} \tag{2.1.6}$$

Beamforming is essentially a "delay and sum" process to calculate the array beam. First the single-station recordings are appropritely shifted in time, then all signals are summed and divided by the number of traces.

Let $x_j(\mathbf{r}_j, t)$ be the signal recorded at the j-th station. Since a plane wave is defined by its propagation direction and apparent velocity (or slowness vector $\mathbf{s}_0$), we can write

$$x_j(\mathbf{r}_j, t) = x(t - \mathbf{r}_j \cdot \mathbf{s}_0) \tag{2.1.7}$$

The seismic beam of an array with N sensors for a seismic signal for the slowness $\mathbf{s}_0$ is defined as

$$b(t) = x(t) = \frac{1}{N} \sum_{j=1}^{N} x_j(t + \mathbf{r}_j \cdot \mathbf{s}_0) \tag{2.1.8}$$

For a signal with any other slowness vector $\mathbf{s}$ the array beam is given by

$$b(t) = \frac{1}{N} \sum_{j=1}^{N} x_j(t + \mathbf{r}_j \cdot \mathbf{s}) \tag{2.1.9}$$

If we calculate all time shifts for a signal with the correct slowness $\mathbf{s}_0$ with respect to any other slowness $\mathbf{s}$ we get the difference for the signal $x(t + \mathbf{r}_j \cdot \mathbf{s}_0 - \mathbf{r}_j \cdot \mathbf{s})$ and the array beam can be written as

$$b(t) = \frac{1}{N} \sum_{j=1}^{N} x_j[t + \mathbf{r}_j \cdot (\mathbf{s}_0 - \mathbf{s})] \tag{2.1.10}$$

The seismic energy of the beam can be calculated by integrating over the squared amplitudes

$$E(t) = \int_{-\infty}^{\infty} b^2(t)\ dt = \int_{-\infty}^{\infty} \left\{ \frac{1}{N} \sum_{j=1}^{N} x_j[t + \mathbf{r}_j \cdot (\mathbf{s}_0 - \mathbf{s})] \right\}^2 dt \tag{2.1.11}$$

Parseval's theorem permits to convert into frequency domain

$$E\left(\omega,\, \mathbf{k}_0 - \mathbf{k}\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|X\left(\omega\right)\right|^2 \left| \frac{1}{N} \sum_{j=1}^{N} e^{i\omega \mathbf{r}_j \cdot (\mathbf{k}_0 - \mathbf{k})} \right|^2 d\omega \qquad (2.1.12)$$

where $\mathbf{k}$ is the wave number vector and $X\left(\omega\right)$ the Fourier transform of the original signal $x(t)$.

Replacing $\left|A\left(\mathbf{k}_0 - \mathbf{k}\right)\right|^2 = \left| \frac{1}{N} \sum_{j=1}^{N} e^{2\pi i \mathbf{r}_j \cdot (\mathbf{k}_0 - \mathbf{k})} \right|^2$ in the eq. 2.1.12 we obtain

$$E\left(\omega,\, \mathbf{k}_0 - \mathbf{k}\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|X\left(\omega\right)\right|^2 \left|A\left(\,\mathbf{k}_0 - \mathbf{k}\right)\right|^2 d\omega \qquad (2.1.13)$$

The energy of an array beam is determined by the power spectral density $\left|X\left(\omega\right)\right|^2$ and the array transfer function $\left|A\left(\,\mathbf{k}_0 - \mathbf{k}\right)\right|^2$ that depends on the wave number of the observed signal and the array geometry.

The (apparent horizontal) slowness s can be calculated from the module of $\mathbf{k}$

$$s = \frac{|\mathbf{k}|}{\omega} \qquad (2.1.14)$$

and the backazimuth θ by

$$\theta = \arctan\left(\frac{k_x}{k_y}\right) \qquad (2.1.15)$$

2.2 Cross-Correlation of diffuse fields

It has been shown (Weaver and Lobkins, 2001; Shapiro and Campillo, 2004) that the response of the Earth after the action of an impulsive source, the so-called Green's function, can be estimated by computing cross-correlation (c-c) function of continuous ambient noise data recorded at couples of synchronized stations. The method rely on assumption that noise sources should be uniformly distributed with respect to the seismic network. The noise distribution generally is not homogeneous due to the presence of localized noise sources. The use of sufficiently long time series and the scattering from heterogeneities within Earth allow to randomize the distribution of the ambient sources (Shapiro et al., 2005).

From a mathematical point of view the c-c is a measure of similarity between two waveforms as a function of a time lag applied to one of them. It allows to obtain coherent information from seismic noise recorded by a station pair. Given two continuous signals x(t) and y(t), the c-c is defined as the integral

of the product of the two functions after one is shifted by amount τ (it is equivalent to the convolution with no time reversal action)

$$C_{xy}(\tau) = (x \star y)(\tau) \stackrel{def}{=} \int_{-\infty}^{\infty} x^*(t)\, y(t+\tau)\, dt \tag{2.2.1}$$

Similarly, for discrete signals

$$C_{xy}[n] = (x \star y)[n] \stackrel{def}{=} \sum_{m=-\infty}^{\infty} x^*[m]\, y[m+n] \tag{2.2.2}$$

To demonstrate the equivalence between c-c and Green's function we follow the approach of Weaver and Lobkins (2001) based simply on the definition of a diffuse field and its properties within an elastic medium (the Earth in our case).

A diffuse field in a finite body may be expressed in modal form by

$$\phi(x,t) = \sum_{n=1}^{\infty} a_n u_n(x)\, e^{iw_n t} \tag{2.2.3}$$

where x is location, t is time, a_n are modal amplitudes, u_n and w_n are eigenfunctions and eigenfrequencies.

A property of diffuse elastic waves is that the modal amplitudes are uncorrelated random variables

$$\langle a_n a_m^* \rangle = \delta_{nm} F(\omega_n) \tag{2.2.4}$$

where $F(\omega)$ is the spectral energy density.

Therefore, the cross-correlation between the fields at x and y results

$$C_{xy}(\tau) = \sum_{n=1}^{\infty} F(\omega_n)\, u_n(x)\, u_n(y)\, e^{-iw_n \tau} \tag{2.2.5}$$

By comparing the 2.2.5 with the expression of the actual Green's function between x and y

$$G_{xy}(\tau) = \sum_{n=1}^{\infty} u_n(x)\, u_n(y)\, \frac{\sin \omega_n \tau}{\omega_n} \qquad \tau > 0 \tag{2.2.6}$$

we find that the two functions are proportional up to an amplitude factor depending on frequency.

From a physical point of view the c-c represents the seismic signal that would be recorded at one of the receivers if an impulsive source were acting at the other receiver.

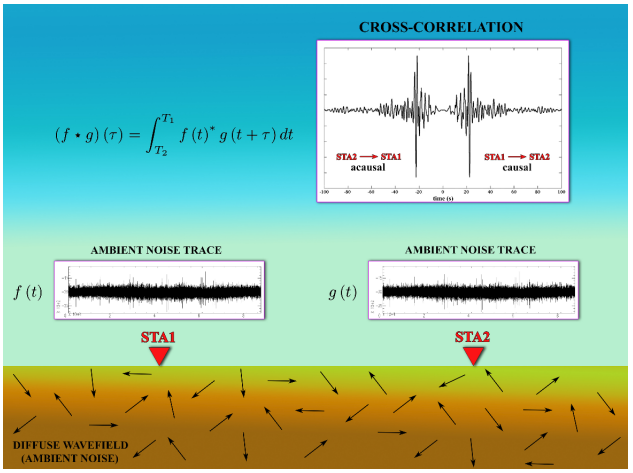


Figure 2.2.1: *Principles of cross-correlation analysis.*

The figure 2.2.1 shows a schematisation of the method based on c-c analysis. Computing c-c function (correlogram) of the noise traces recorded at two stations, sta1 and sta2, it is possible to obtain information on the propagation medium between these stations, as in the case with a seismic signal that propagates from a source to a receiver (seismogram). The c-c function is composed of a causal part (positive time) which represents the Green's function that propagates from the source sta1 to the receiver sta2, and an acausal part (negative time) which represents the Green's function that propagates from the source sta2 to the receiver sta1.

Considering a seismic network, for each station pairs a c-c function is calculated over several days. The total number of computed c-c function is equal to n(n-1)/2, where n is the number of stations. Then, the daily c-c functions are temporal stacked to correspond to longer time-series improving the signal-to-noise ratio and emphasizing the coherent signals that correspond to the Green's functions.

Theoretically, it is expected to observe the propagation of waves at all pairs of stations symmetrically with respect to zero time because of the noise isotropy with respect to the seismic network. For each station pairs the presence of a coherent signal and its stability with time in different ranges of frequency is analysed.

The computed Green's functions from the correlation of noise data are dominated by surface waves because most noise sources occur close to the Earth's

surface (Shapiro et al., 2005), but also body waves can be extracted (Poli et al., 2012).

To compute c-c functions, the noise signals are considered into the ZRT (Vertical, Radial, Transverse) system. The North (N) and East (E) components of the data original horizontal orientations are rotated into components that are radial (R) and transverse (T) to the propagation direction of the signals (Fig. 2.2.2), through the rotation matrix

$$M = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \qquad \begin{bmatrix} R \\ T \end{bmatrix} = M \begin{bmatrix} N \\ E \end{bmatrix}$$

where α is the azimuth measured clockwise from north.

The c-c functions between the radial (R), transverse (T), and vertical (Z) components of the noise signals of two stations are represented by a nine-component tensor corresponding to the elastic Green's tensor. Rayleigh waves are reconstructed on the ZZ, RR, RZ, and ZR components because their elliptical particle motion is confined to the vertical-radial plane. Love waves, for which the particle motion is only in the transverse plane, are emphasised by the TT component (Lin et al., 2008). The remaining transverse components (RT, TR, TZ, ZT) are characterized by weak diffuse phases and can be used as support for analysis (Zigone et al., 2015).

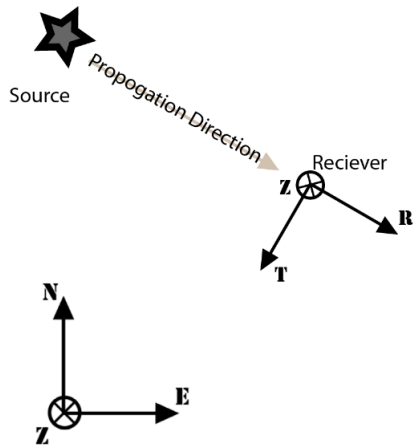


Figure 2.2.2: *Map view of the (down) ZNE system used when instruments have horizontal components aligned with North and East, and (top) ZRT system used when signals horizontal components are radial and transverse to the propagation direction. The vertical component Z is directed in of the page. Source: IRIS Web Service, http://www.iris.edu.*

The classical procedure to perform ambient noise data processing and cross-

correlation following the approach described in Bensen et al. (2007) will be developed in chapter 3 where the steps conducted for the analysis are described in detail.

2.3 Dispersion measurement - FTAN method

An important characteristic of surface waves propagating in stratified media is the dispersion phenomenon. Therefore, their velocity depends on frequency and consequently on penetration depth. It is possible to determine a phase velocity for each wave frequency and a group velocity of the wave packet moving between source and receiver (or between two receivers, in the case of seismic noise). The value of these quantities strongly depends on the characteristics of the subsoil (density, and velocity P and S), therefore allows to obtain information on the structure of the propagation medium. Short period waves investigate characteristics of shallow layers and long period waves explore layers at greater depth.

The study of Rayleigh surface wave dispersion can be performed by applying the *Frequency-Time ANalysis* (FTAN) (Levshin et al., 1972; 1992) to noise cross-correlation. The analysis consists in a multiple filtering in the frequency domain that permits an easy identification and isolation of the fundamental mode.

Given $s\,(t)$ the correlogram from which to extract the Rayleigh group-velocity dispersion curve, the Fourier transform is computed to obtain the spectum

$$S\,(\omega) = \int_{-\infty}^{\infty} s\,(t)\, e^{-i\omega t} dt \tag{2.3.1}$$

Then, to obtain a frequency-time representation of the signal, the method applies a set of narrow-band filters $G\,(\omega - \omega_i)$

$$Y\,(\omega_i,\, t) = \int_{-\infty}^{\infty} S\,(\omega)\, G\,(\omega - \omega_i)\, e^{i\omega t} d\omega \qquad\qquad i = 1, ..., N \tag{2.3.2}$$

In the case of surface waves, the best resolution in the frequency-time domain with no phase distortion is achieved using a Gaussian filter with variable central frequency ω_i (Levshin et al., 1972).

$$G\,(\omega - \omega_i) = e^{-\alpha\left(\frac{\omega - \omega_i}{\omega_i}\right)^2} \tag{2.3.3}$$

The α parameter governs the frequency bandwidth of the filter and depends on the source-receiver distance. As the distance increases, the bandwidth

decreases. In other words α controls the resolving power of the filter, i.e. the ability to identify the dispersion curves.

$|Y(\omega_i, t)|$ represents the signal envelope and depends on the characteristics of the chosen filter.

The group arrival time $\tau(\omega_i)$ is determined from the peak of the signal envelope so that it is possible to calculate the group velocity defined as the ratio between the distance source-receiver r and the arrival time of the envelope $\tau(\omega_i)$

$$U(\omega_i) = \frac{r}{\tau(\omega_i)} \tag{2.3.4}$$

2.4 Surface wave velocity model

S-wave velocities control changes in Rayleigh wave phase and group velocities for a layered Earth model. Therefore, S-wave phase and group velocities can be estimated from the inversion of Rayleigh wave dispersion curves obtained in previous analysis.

The inversion procedure consists of determining the Earth model parameters (S-wave velocities) that are consistent with the measurements (dispersion curves), optimizing in the parameter space an appropriate cost function (misfit) that measures the discrepancy between data and model. The inversion method is selected depending on the type of relationship between data and model parameters. For example, linear or linearized approach is used to solve linear or quasi-linear problems, respectively. When the inverse problem is highly non-linear, robust and efficient global optimization strategies are most appropriate to find the global minimum. To provide stable and robust noise tomographic images, a combination of the *Fast Marching Method* (FMM) (Sethian, 1996) and the *subspace inversion method* (Kennett et al., 1988) is widely used.

FMM is a numerical algorithm which solves the forward problem of traveltime prediction simulating the evolution of the first arrival wavefront by employing a narrow band approach (Fig. 2.4.1). The algorithm considers a gridded velocity field with a traveltime associated to each node and calculated by solving the eikonal equation via finite-difference solution. FMM follows an upwind structure that means the wavefront propagates from smaller values of T to larger ones. The iteration stops when each point has been analysed.

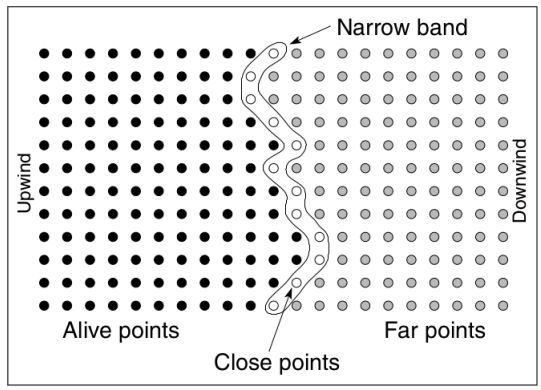


Figure 2.4.1: *Principle of narrow band method. The shape of the narrow band approximates the shape of the first arrival wavefront, the alive nodes represent the region that has already been reached from the wavefront, and for the far nodes the eikonal equation has not yet been solved (Rawlinson and Sambridge, 2005).*

The eikonal equation which governs the propagation of seismic waves in the high-frequency limit is

$$|\nabla_{\mathbf{x}} T| = s\left(\mathbf{x}\right) \tag{2.4.1}$$

where $\nabla_{\mathbf{x}}$ is the gradient operator, T is traveltime, and $s\left(\mathbf{x}\right)$ is slowness as a function of position $\mathbf{x}$.

The equation to solve for satisfying the upwind scheme is

$$\left[\begin{array}{l} max\left(D_a^{-x}T, -D_b^{+x}T, 0\right)^2 \\ +max\left(D_c^{-y}T, -D_d^{+y}T, 0\right)^2 \\ +max\left(D_e^{-z}T, -D_f^{+z}T, 0\right)^2 \end{array}\right]_{i,j,k}^{\frac{1}{2}} = s_{i,j,k} \tag{2.4.2}$$

where (i, j, k) are Cartesian grid increment variables in (x, y, z) , and the variables a, b, c, d, e, f are the order of accuracy of the upwind finite-difference operators.

For example, making $D^{-x}T_i$ explicit, the two first operators are

$$D_1^{-x}T_i = \frac{T_i - T_{i-1}}{\delta x}\,, \tag{2.4.3}$$

$$D_2^{-x}T_i = \frac{3T_i - 4T_{i-1} + T_{i-2}}{2\delta x}$$

where δx is the grid spacing in x.

The operators in 2.4.2 depend on the availability of upwind traveltimes and the maximum order allowed.

When this algorithm is applied to a 3-D teleseismic tomography, the operators 2.4.3 must be convert from Cartesian to spherical coordinates because the Earth is assumed to be spherically symmetric. For example, the first-order operators in spherical coordinates are

$$D_1^{-r}T_i = \frac{T_i - T_{i-1}}{\delta r}\,, \tag{2.4.4}$$

$$D_1^{-\theta}T_j = \frac{T_j - T_{j-1}}{r\delta\theta}\,,$$

$$D_1^{-\phi}T_k = \frac{T_k - T_{k-1}}{rsin\theta\delta\phi}$$

where δr, $\delta\theta$, $\delta\phi$ are the grid spacing in depth, co-latitude and longitude, respectively.

The marching of this method is considered fast because of the heapsort algorithm that require a total computational cost of O(NlogN), where N is the total number of grid points. The FMM efficiency is due to the fact that calculation scales with grid size.

In a non-linear inverse problem with different type model parameters (different character and physical dimensionality), methods based on gradient techniques that minimize a misfit function may not converge and can depend strongly on the scaling of the different parameters types. To overcome these obstacles the subspace inversion method is used to solve large inverse problems with a iterative procedure that represents a cross between a gradient and a matrix approach. The perturbation to the current model is construct by performing a least-squares inversion within a small dimension subspace spanned by vectors which reflect the dependence on all the parameter classes.

2.5 Ambient noise surface-wave tomography

Seismic tomography is a data inference technique used to study the internal structure of the Earth. Indeed, it allows reconstruction of 2-D or 3-D models of the subsurface extracting information from a set of observable seismic data. In particular, the obtained static images are the representation of spatial velocity variations of P and S or surface seismic waves. The analysis of

velocity variations is directly related to the knowledge of physical properties of the propagation medium, and therefore it can improve the insight of the phenomena that originate a seismic event in a specific area. In the tomographic method, the investigation of the medium is carried out by studying the behaviour of seismic waves crossing it when propagating from seismic sources to receivers located at the surface (Fig. 2.5.1). From the recordings of the signals (seismograms) at the receivers, it is possible to observe the waves traveltimes. These values, and consequently their velocities, depend on the physical properties of the crossed medium. The use of tomographic method requires knowledge, prior to the tomographic survey, of a reference velocity model of the study area, obtained using information from other geological or geophysical studies. Using this model, theoretical traveltimes are calculated and compared with observed traveltimes. The difference between the observed and theoretical times is called residual and indicates the presence of a velocity anomalies zone. Seismic tomography, thus reports the distribution of velocity anomalies that can be calculated at different depth levels (Fig. 2.5.1), depending on the considered frequency of the signal.

Performing a tomography consists of solving an inverse problem. Indeed, given the arrival times of the waves that propagated from the sources to the receivers, a heterogeneous seismic model that is consistent with observations is determined. Since the relationship between the arrival times and the velocity field is non-linear, the implementation of non-linear inverse methods are required.

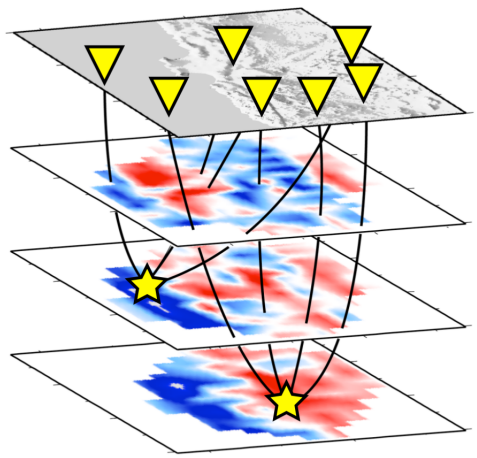


Figure 2.5.1: *Principle of the seismic tomographic method. Sources (stars), receivers (triangles), and raypath (black lines) within the study volume are shown. Source: S. Husen (2011), from lecture notes on seismic tomography - ETH Zurich.*

Traditional procedures extract information about the structure of the Earth by analysing seismic waves generated by earthquakes (active seismic tomography) or explosions (passive seismic tomography). In recent years, it is also possible to perform imaging considering ambient seismic noise as an alternative to earthquakes or artificial sources. In this case we talk about *ambient noise tomography*, a new revolutionary approach for imaging Earth structure at a variety of scales. Ambient noise tomography is based on the reconstruction of the Green's function between different receivers from the cross-correlation of ambient noise recordings made the Earth's surface. This result theoretically synthesises the seismic energy that would be recorded at one of the receiver locations if there had been an impulsive seismic source at the other. The most common approach is to extract Rayleigh wave group traveltimes from the cross-correlated waveforms and invert for group velocity at different periods. Using ambient noise tomography is advantageous because the study of the aseismic response of the propagation medium can be performed, the absence of earthquakes guarantees the absence of induced effects in the medium.

The resolution of a tomographic image (*physical resolution*) indicates what is the smallest spatial dimension of the obtained tomographic result that can be distinguished. Image resolution reflects *model resolution*, therefore it allows to establish the quality of the solution through the knowledge of the best solved model parameters. A qualitative estimate is provided by the ray coverage, i.e. the number of rays and their path in each block of the investigated propagation medium. In addition, several synthetic tests can be performed to evaluate the reliability of the reconstructed shape and amplitude of anomalous areas obtained in tomographic results. The most common are *checkerboard test*, *spike test*, and *shape test*.

In the checkerboard test, the final model obtained from the tomographic inversion is perturbed by adding a distribution of checkerboard anomalies with equal amplitude with alternating sign. The value of the perturbation must be sufficiently small (usually 10% of the velocity) in order to not alter significantly the raypaths obtained in the final tomographic model. The size of the checkerboard cells depends on the minimum dimension of the anomalies to be highlighted in the resolution test. Considering the perturbed model, synthetic data corresponding to the arrival times of the seismic waves are calculated. Then they are used as input data for a tomographic inversion with the same configuration as the one with the real data. The difference between the final tomographic models obtained with the synthetic data and the real data is calculated and should reproduce the form and amplitude of the checkerboard perturbation (Fig. 2.5.2). Therefore, the regions where the checkerboard perturbation is well reconstructed are those where the max-

imum resolution of the data is expected, then a higher reliability of the tomographic model.

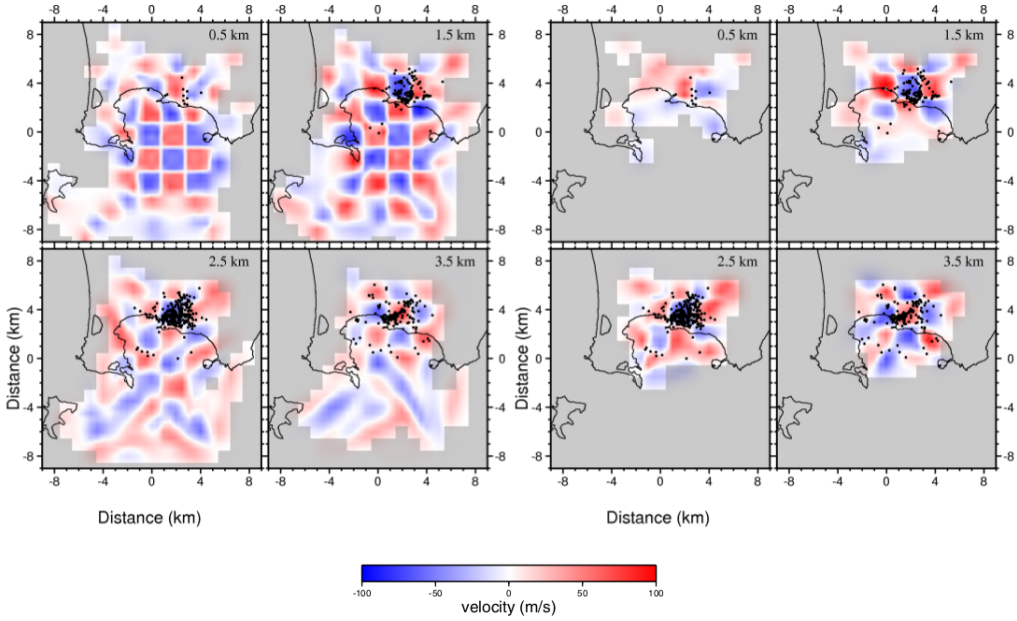


Figure 2.5.2: *Example of a checkerboard test results for with 2*2*2 km cells with ±50 m/s anomalies added to P (left) and S (right) velocities in the area of Campi Flegrei (Battaglia et al., 2008).*

The checkerboard test strongly depends on the extent, shape and amplitude of the anomalies, and on the heterogeneity of the final model.

The spike test follows the same principle as the checkerboard test, but adds an anomaly at only one point on the tomographic grid. The test can be performed several times for different grid positions. Finally, the shape test consists in considering a velocity model that reproduces, with a simplified form, a specific anomaly obtained in the final tomographic model, and then resolving the tomographic problem again maintaining its configuration unchanged. The ability of the inversion method to reconstruct the location, the shape, and the amplitude of the specific anomaly is evaluated. By comparing the final and the retrieved models it can be extracted information on the image resolution of specific regions and also highlight any smearing effect due to non optimal ray sampling.

Chapter 3

A shale gas extraction site: Wysin (Poland)

Over the last decades, the application of hydraulic fracturing (HF) has increased interest and expectations about production of hydrocarbons in many European countries. At the same time, also has raised concerns about the entity of the associated environmental impact. In UK, HF operations for shale gas exploration were suspended on April 1 st , 2011 at Blackpool after a 2.3 Ml induced earthquake; while in Poland, this type of activity were halted in 2017 due to unfavourable oil market conditions and unsatisfying results from the exploration phase due to the geology. The exploitation of unconventional hydrocarbon resources is therefore still a controversial issue.

One of the last operating wells in Europe was at shale gas Wysin site located in northern Poland, which also represents one of the first real-scale HF experiment in Europe. It is a test site of *SHEER* project (*Shale gas Exploration and Exploitation Induced Risks*), a 3-year plan of the European Union H2020 programme that concluded in 2018. The aim of the project was to develop best practices for assessing and mitigating the environmental impacts and risks of shale gas exploration and exploitation (www.sheerproject.eu). A main activity of the SHEER consortium was the installation and maintenance of a dence dedicated multidisciplinary monitoring system at HF operational site at Wysin. Such multi-monitoring activity was the first in Europe and aimed of jointly assess the short- and long-term risk connected to the most relevant potential hazards of HF operations such as induced seismicity, air pollution, and groundwater contamination.

The first part of this chapter introduces a brief description of geological features of the study area, industrial operations carried out in 2016, and previous seismic studies that provided information about the magnitude of

completeness, local microseismicity and attenuation properties of the area. Afterwards, the cross-correlation analysis, the study of the temporal coherence and possible seasonal variability of their energetic phases are reported. Finally, two different approch for performing tomographic inversion was used. The first is the FMST, an iterative non-linear inversion method that uses the fast marching method (FMM) for the forward prediction step and the subspace inversion method to performe each local inversion assumed linear. Iterative application of FMM and subspace inversion permits to account for the non-linear of the problem. The second approch is the MANgOSTA method (Cabrera-Pérez et al., 2020) based on non-linear multiscale inversions which progressively upgrade the model parameterisation to reduce the non-linearity of the inverse problem. Finally, the results of the two approaches are analysed, compared and discussed.

3.1 Geological setting, drilling site, and HF operations

The study area of this thesis work is located in Pomerania region, northern Poland about 45 km from Gdańsk city. The geological structure belongs to the Baltic Basin, a large marginal synclinal structure in the southwestern part of the East European Craton, and it is relatively tectonically undeformed. It is a sedimentary basin as the result of a continuos dynamic geological evolution in time of the European continent. The Baltic Syneclise is a stratification of deposits belonging to different geological eras, from the Lower Palaeozoic through the Mesozoic to the Cenozoic. In particular at about 4 km depth there are thin Cambrian and Ordovician layers overlain by 2 km of Silurian sediments. In this era, marine organic matter-rich fossils were accumulated and led to the formation of the unconventional shale oil and gas reserves, thanks to the biogeochemical conditions of the region.

The Wysin site was a candidate area for shale gas extraction belonging to the Stara Kiszewa concession of Polish Oil and Gas Company (PGNiG), currently inoperative. The exploration and exploitation of the shale gas was carried out by the company through hydro-fracturing (HF) operations. A pre-drilling analysis were carried out by the Polish Geological Institute (PIG-PIB) through the research borehole Koscierzyna IG-1 located 8.25 km from Wysin site (Fig. 3.1.1), obtaining detailed insight on the local lithology and stratigraphy. The drilling pad had three boreholes located at about 4 km depth crossing the stratigraphy of the various geological formations (Fig. 3.1.2).

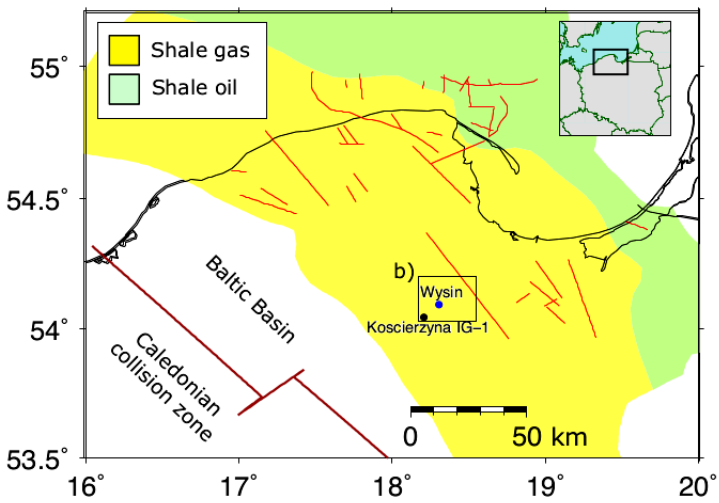


Figure 3.1.1: *Structural-tectonic map of the Polish Baltic region with an indication of resources area of shale gas (yellow area) and shale oil (green area) from a study conducted by the PIG-PIB. Fault structures (red lines), wellhead of the Wysin site (blue dot) and Koscierzyna IG-1 research borehole (black dot) are shown (López Comino et al., 2018).*

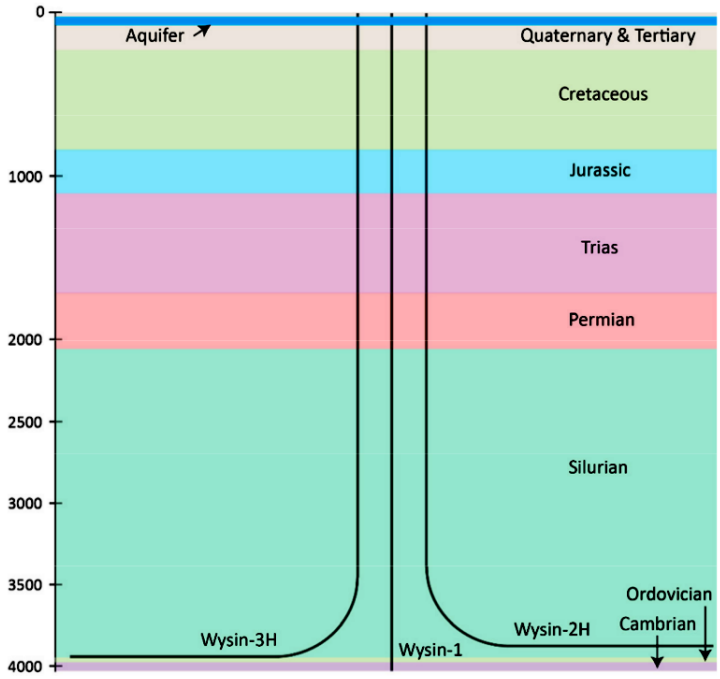


Figure 3.1.2: *Schematisation of the drilling pad and geological composition stratigraphy in which the wells are located (Montcoudiol et al., 2017).*

A vertical well (Wysin-1) was drilled in 2013 and two horizontal wells (Wysin-2H and Wysin-3H) were completed in summer and autumn 2015, respectively. The latter have a horizontal lenghts of 1.7 km each, oriented WNW-ESE (Fig. 3.1.3), and quite parallel to the NW-SE trend of the region faults (Fig. 3.1.1).

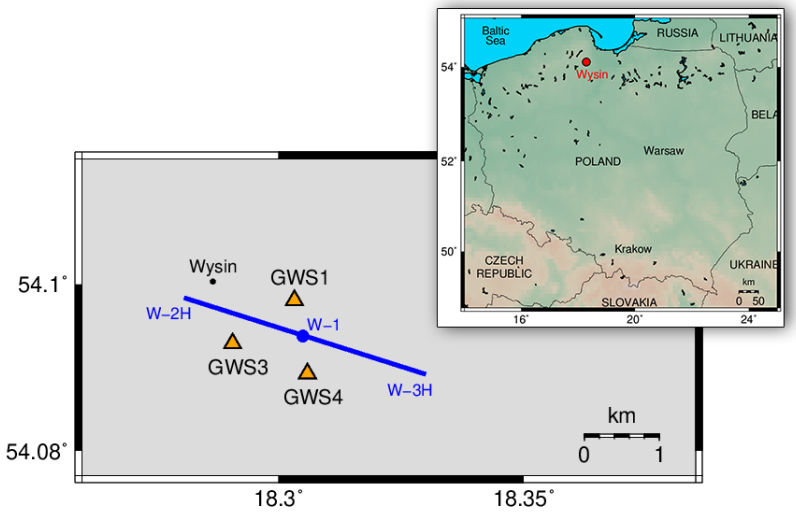


Figure 3.1.3: *Location of the wellhead Wysin-1 (blue dot) and horizontal boreholes Wysin-2H and Wysin-3H (blue lines). The orange triangles represent the three shallow borehole stations located closest to the wells.*

HF operations were performed in 2016 from 9^{th} June to 18^{th} June along Wysin-2H well and from 20^{th} July to 29^{th} July along Wysin-3H well, each requiring eleven stimulation stages. The injection operations involved a total volume of 18812 m^3 and 17230 m^3 for Wysin-2H and Wysin-3H respectively, and maximum pressures at the well head between 84.3 and 90.5 MPa (López Comino et al., 2018).

3.2 Previous seismic works on the study area

The seismic response of the HF experiment at the Wysin site has been assessed by some studies analysing seismic records before, during and after the HF operations.

López Comino et al. (2017) evaluated the performance of the monitoring network in detecting induced microseismicity around the fracking area in terms of the magnitude of completeness (M_c) using information of the

pre-operational phase. In this study a synthetic catalogue was generated considering the characteristics of different seismogenic sources and the type of seismic activity expected from the HF activities. A realistic synthetic dataset at all the network stations were also generated by taking into account the instrumental response and adding the real noise recorded prior to the HF operations. For the computation, a 1-D velocity model of the site (Fig. 3.2.1) was extracted from a high-resolution 3-D seismic model for Poland provided by Grad et al. (2015).

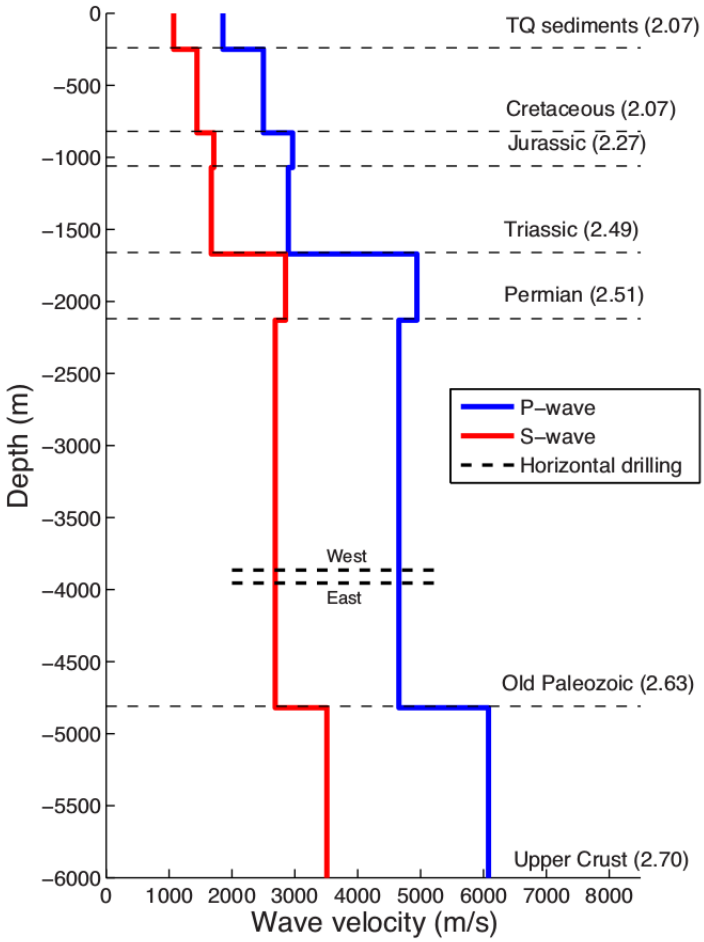


Figure 3.2.1: *1-D crustal velocity model for the Wysin site. The P-wave velocity profile (blue line) at the location of the drilling site is an extraction of the high-resolution 3-D seismic model for Poland from Grad et al. (2015), and the S-wave velocity profile (red line) comes from the assumption Vp = 1.73 Vs. For each layer the density value ($g\,cm^{-3}$) is reported. The depth of the west and east horizontal drilling (thick black dashed lines) is also shown (López Comino et al., 2017).*

The M_c was estimated using two different techniques. The first use an amplitude threshold that takes into account the noise recorded at the stations and the maximum amplitude of the synthetic waveforms, for different values of signal-to-noise ratio (SNR). Considering an SNR = 2, a M_c of about 0.55 was estimated with an increase of 0.05 during day hours, whereas M_c decreases to 0.45 using arrays techniques. The second method involves the application of an automatic detection algorithm based on the evaluation of the coherence of the signal arrival times at the various stations. In this case the detection performance is improved with a lower M_c of about 0.10 during night hours, but an increase of false detections can occurs. In this study it is also show that M_c differs for different source processes. The double-couple sources are better detected than tensile cracks resulting from fracking.

López Comino et al. (2018) continued the seismic noise characterisation for the period during all HF stages finding additional shallow artificial seismic noise sources at the wellhead. In day hours, the noise conditions change due to HF operations temporally reduce the SNR of the three borehole stations, increasing the M_c from 0.55 to 0.80. Local microseismicity between June and September 2016 was also detected and characterised. For this aim, an automated full waveform detection and location algorithm based on waveform stacking and coherence analysis was deployed.

The results show that the fault system around the HF area remaine inactive in the analysed period. Most of the detected events are weak ($M_w < 1$) and their signals are visible only at the three borehole stations, often with a low SNR. Only two weak events were recorded at all stations with an estimated M_w of 1.0 and 0.5. They occurred days after the stimulation and close to the HF site, but they are several kilometres shallower than the horizontal wells (Fig. 3.2.2). The authors supposed a link with industrial activities due to the spatial vicinity between the epicentres and wellhead, and the temporal correlation between HF injections and seismicity occurrence.

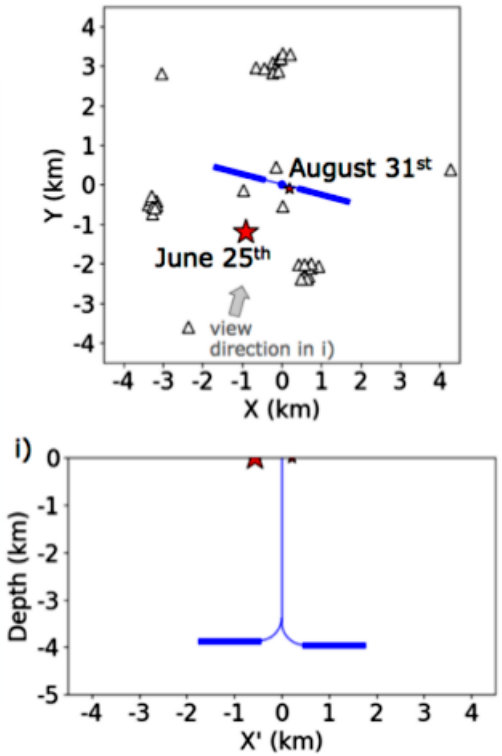


Figure 3.2.2: *(Top panel) Map view and (down panel) depth section depicting the location estimated for the two main events (red stars) at Wysin site. The 1.0 M_w event occurred on 25^{th} June 2016 is located 1500 m SSW of the wellhead (blue dot), while the 0.5 M_w event on 31^{st} August 2016 results closer, i.e. 220 m ESE of the wellhead. Both events have very shallow depths (< 150 m). Seismic stations (triangles) and horizontal boreholes (blue lines) are also shown (López Comino et al., 2018).*

Information on the attenuation properties of the study area of this thesis is provided by the work of Wandycz et al. (2019). They estimated the P- and S-wave quality factors (Q_P and Q_S) analysing microseismicity records from two indipendent gas reservoirs of the northern Poland, in particular from the Wysin-2H and Lubocino-2H horizontal wells. In both cases datasets was recorded during each stage of the hydraulic stimulation through a vertical array of geophones deployed in the respective wells. The Q factor was calculated using the peak frequency method, thus it is proportional to the peak frequency and traveltime of the corresponding seismic wave. The results show an homogeneous and isotropic attenuation in both reservoirs (Fig. 3.2.3).

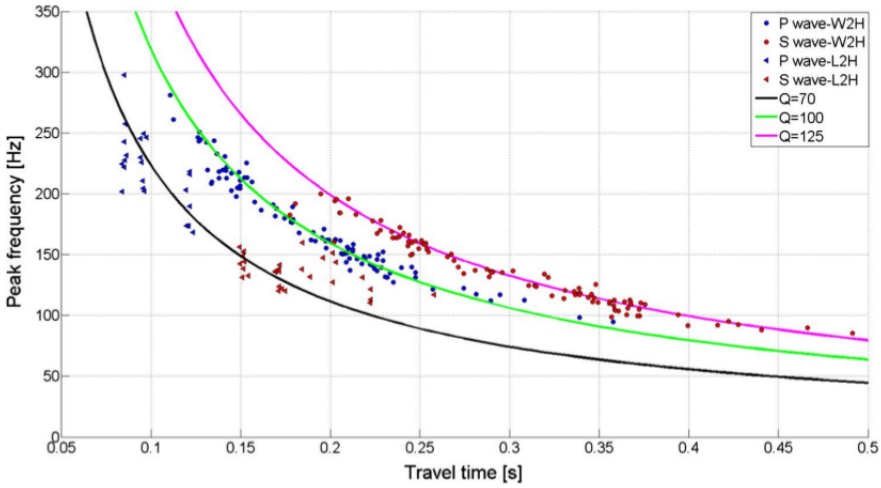


Figure 3.2.3: *Measured peak frequency as a function of computed traveltime for P and S waves for Wysin-2H (circle) and Lubocino-2H (triangles). The lines represent theoretical peak frequencies calculated with (black) Q = 70, (green) Q = 50 and (magenta) Q = 125 (Wandycz et al., 2019).*

In the case of Wysin site, the P-wave and S-wave quality factors have average values across all stages of approximately 100 and 120, respectively. For Lubocino site, the observed Q_P and Q_S factors range from 60 to 90. In both cases, S-wave attenuation factors is higher than P-wave which is expected for highly compacted and saturated sedimentary reservoirs. Moreover, despite the same lithological formations, the effective attenuation is lower in the deeper reservoir Wysin consistent with an increased pressure and compactness. The low attenuation affects events detection improving the performance. Indeed, for Wysin almost 100 events were located compared to a few for Lubocino over the same distance. This proves that the attenuation study is another fundamental step for evaluating monitoring networks and, consequently, for assuming reliable results of a study area.

3.3 Seismic network and data

The local seismic network was installed in the Wysin site in the framework of *SHEER* project. The seismic monitoring was part of a wider investigation which also included the indipendent monitoring of groundwater and air conditions. It was performed to assess the sensitivity of microseismicity to industrial operations.

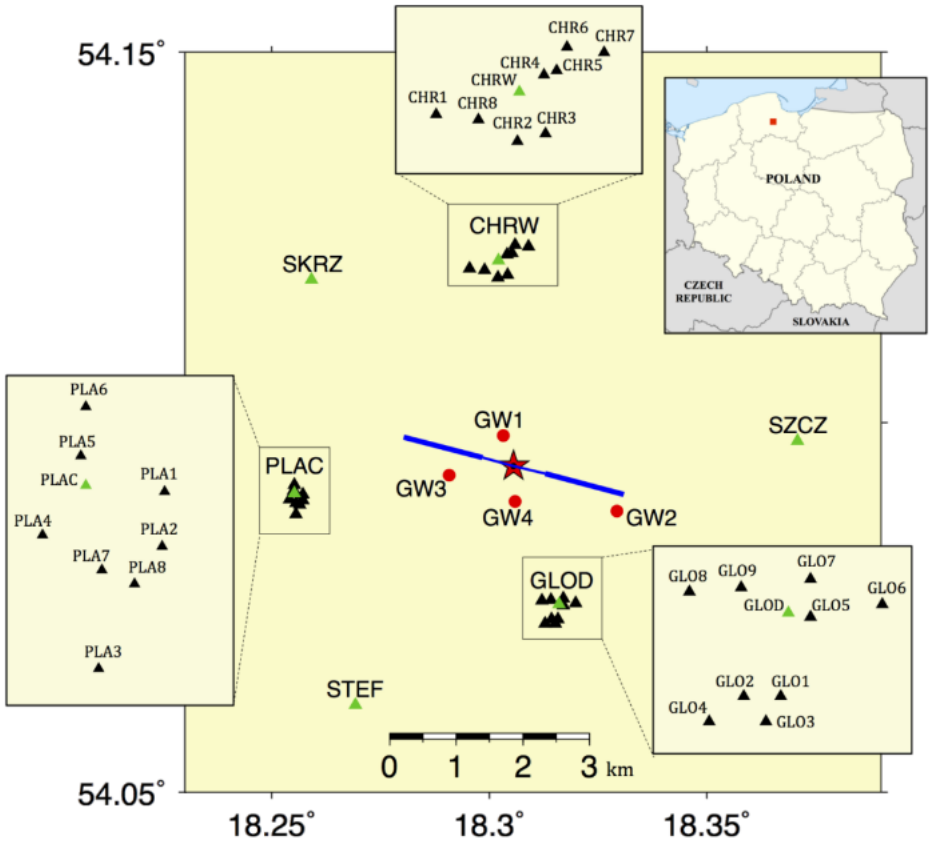


Figure 3.3.1: *The entire network geometry of the Wysin site composed of 6 surface BB stations (green triangles), three small-scale arrays (inset boxes) with 8–9 surface SP stations (black triangles) around a BB station, and 3 shallow BBB stations (red circles). Fracking area (red star), projection of the vertical (blue dot) and horizontal wells (blue lines) are shown. (Lòpez Comino et al., 2017).*

The implemented seismic network (Fig. 3.3.1) consists of 6 surface broadband (BB) seismometers, 25 surface short-period (SP) seismometers, and 3 borehole broadband (BBB) seismometers. The SP devices are distributed into three small-aperture arrays of about 300-400 m, each composed of 8 to 9 SP stations and one BB station (GLO1-9 (SP) and GLOD (BB), PLA1-9 (SP) and PLAC (BB), CHR1-8 (SP) and CHRW (BB)) surrounding the operational wells at distances of about 2-4 km. The remaining BB stations (SZCZ, STEF, SKRZ) are located at a longer distance, up to about 5 km, reducing the azimuthal gap of the network geometry. All the BB sensors are characterized by a sampling rate of 200 Hz in order to detect signal in a broad range of frequencies. The shallow BBB seismometers (GWS1, GW3S, GW4S) was installed at depths of about 50-60 m near the injection well

within a maximum distance of 500 m. The sampling rate of all short-period stations is 500 Hz. In this thesis work the data acquired from the 9 broad-band stations were analysed (6 surface and 3 borehole seismometers). The table 3.1 summarizes the main characteristics of the employed BB stations.

Station code	Sensor type	Sample rate (Hz)	Installation date	Coordinates
GLOD	GURALP CMG-3ESP / Nanometrix Trillium Compact PH TC120PH2	200	9th July 2015 / 18th January 2017	54.07534 N 18.31600 E
CHRW	GURALP CMG-3ESP / Nanometrix Trillium Compact PH TC120PH2	200	27th October 2015/ 18th January 2017	54.12198 N 18.30217 E
PLAC	GURALP CMG-3ESP / Nanometrix Trillium Compact PH TC120PH2	200	27th October 2015/ 18th January 2017	54.09038 N 18.25504 E
SKRZ	GURALP CMG-3ESP	200	27th October 2015	54.11930 N 18.25904 E
STEF	GURALP CMG-3ESP	200	27th October 2015	54.06176 N 18.26926 E
SZCZ	GURALP CMG-3ESP	200	27th October 2015	54.09750 N 18.37083 E
GWS1	Geotech KS2000BH/120 / Nanometrix Trillium Compact PH TC 120PH2	200	17th December 2015/ 17th January 2017	54.09805 N 18.30327 E
GW3S	Geotech KS2000BH/120 / Nanometrix Trillium Compact PH TC 120PH2	500	15th December 2015/ 26th April 2016	54.09287 N 18.29067 E
GW4S	Geotech KS2000BH/120 / Nanometrix Trillium Compact PH TC 120PH2	500	17th December 2015/ 26th April 2016	54.08927 N 18.30592 E

Table 3.1: *Details of the nine BB stations considered for the analysis.*

The network geometry of selected BB stations (Fig. 3.3.2) is suitable for the purpose of using a seismic interferometry method, in fact it has a good azimuthal coverage (maximal gap 90°) of the stations whose interdistances vary from few hundred meters to almost eight kilometers. BB stations allow to start studying ambient seismic noise over a wide frequency range. The figure 3.3.2 also presents the connections between the stations (black line) revealing the combinations of station pairs used for the cross-correlation calculation as well the areas of the propagation medium most constrained by the interferometric analysis (where there are many rays and where the azimuthal distribution of the rays is greater). The choice of using only BB stations is due to the fact that the sub-networks are very concentrated. The SP stations are very close to the BB station belonging to the same mini-array, therefore no particular variations in terms of azimuthal coverage and interstation distances are expected that improve results.

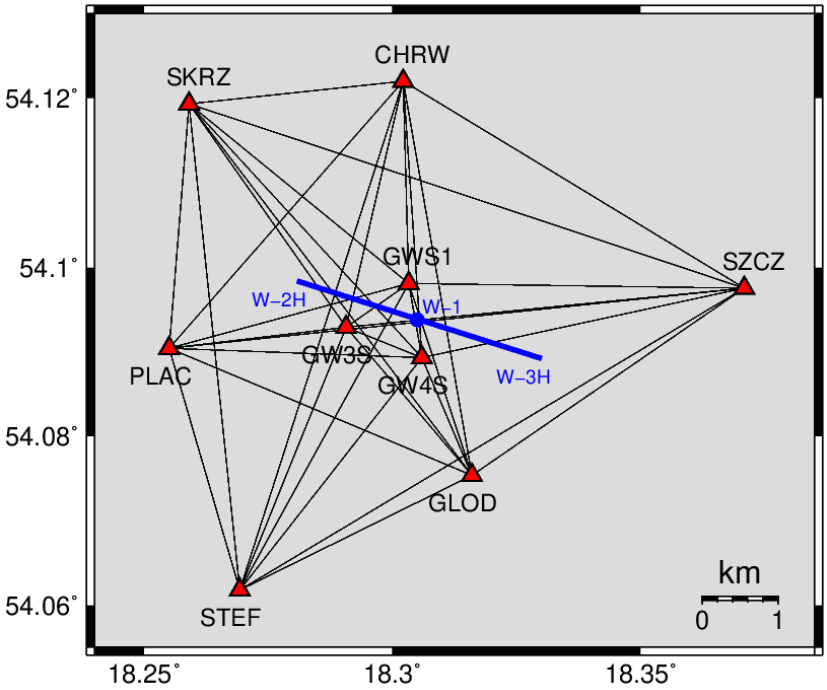


Figure 3.3.2: *The considered network geometry and ray coverage of the investigation area. Wellhead (blue dot) and horizontal boreholes (blue lines) are shown.*

The continuous seismic recordings were downloaded from the SHEERWER platform of the SHEER project. The data are available from 27^{th} October 2015 to 19^{th} July 2017 but the acquisition of the stations presents interruptions in some periods or days. In addition, the sensors of the two borehole stations GW3S and GW4S were replaced with different ones after technical problems at the end of 2016 April. For all these reasons, the datatset influenced the choice of the analysed periods in this work. The graphs in figure 3.3.3 show the availability of the data selected per year, highlighting also the dates of HF operations.

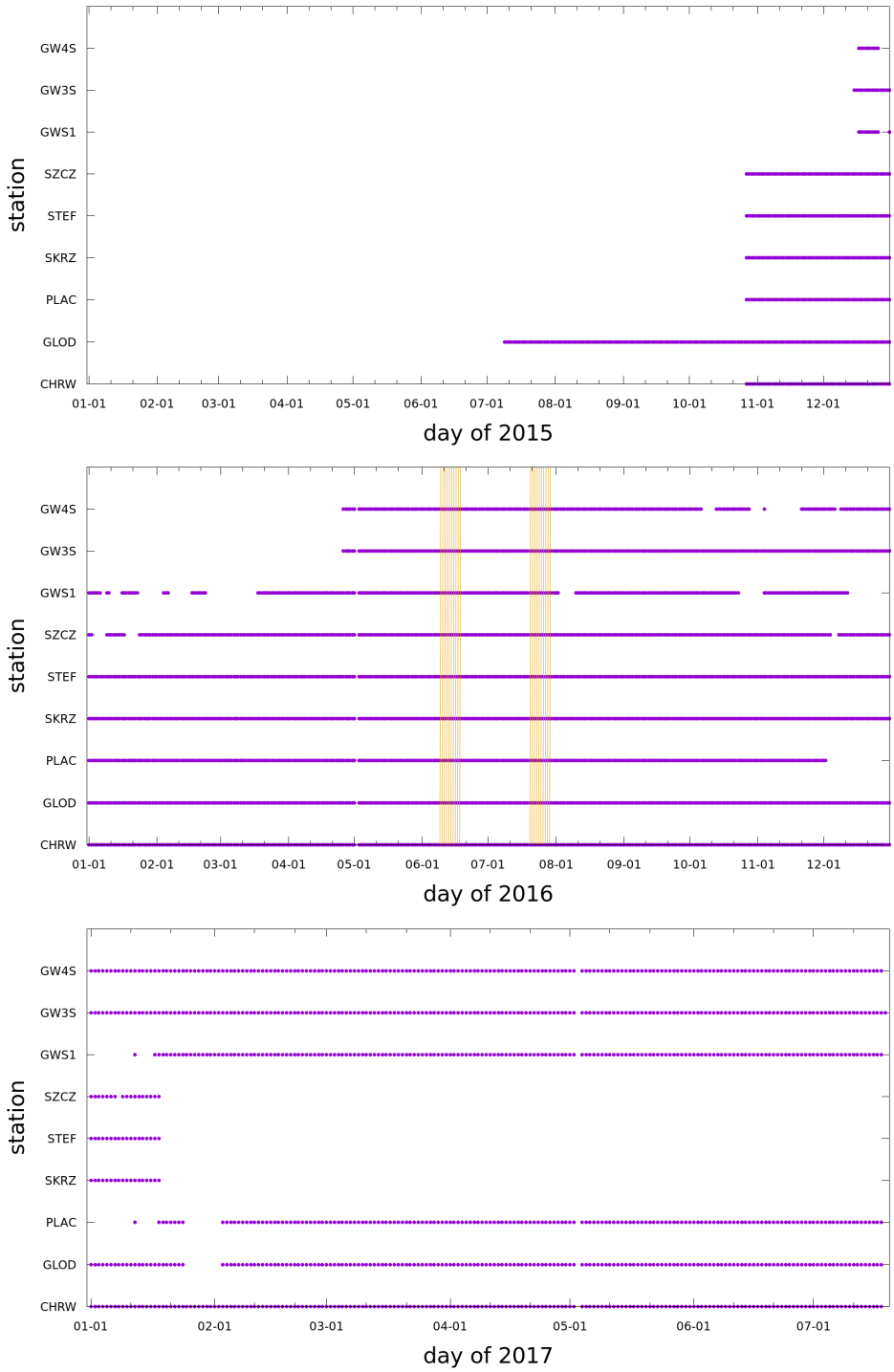


Figure 3.3.3: *Continuous seismic data available for the 2015 (top panel), 2016 (middle panel), and 2017 (bottom panel). The vertical orange lines mark the date of the HF operations.*

3.4 Preliminary noise analysis

In this work a beamforming analysis based on f-k method was performed to test the hypothesis of isotropy distribution of noise sources with respect to the considered seismic network. The analysis was performed using the dedicated script of the open-source project ObsPy (a Python toolbox for seismology/seismological observatories) available on the GitHub platform (https://github.com). For the entire investigated period, the daily vertical signals recorded by the network stations was divided in time-windows and bandpass filtered. Afterwards, the relative and absolute power, the backazimuth, and the slowness of the signals were computed. Different frequency bands was investigated within the range 0.1-3.0 Hz as during the cross-correlation analysis. The beamforming results for one data month and two frequency bands are reported and presented in two different representations. The figures 3.4.1, 3.4.2, 3.4.5 and 3.4.6 show the relative and absolute power, backazimuth and slowness as a function of time for each day of June 2016 in the 0.1-3.0 and 1.5-2.0 Hz frequency bands, respectively. For the same data, the figures 3.4.3, 3.4.4, 3.4.7 and 3.4.8 show the relative power in the slowness space as a function of the angle. In the wide range 0.1-3.0 Hz signal characteristics change over time and are dominated by a northern source. In the range 1.5-2.0 Hz, instead, the power level is comparable in all time windows and the scattered backazimuth values show that there is no privileged direction of the signal. As will be seen in the next paragraphs, this range is where significant results are found in the cross-correlation analysis. In figure 3.4.9 is reported an example of the beamforming results obtained for the day 2016/06/14 for the four frequency sub-bands investigated in this analysis (0.1-0.2, 0.2-0.5, 0.5-0.9, and 1.5-2.0 Hz). As can be seen, the beamforming reveals that interesting results can be expected in the band 1.5-2.0 Hz.

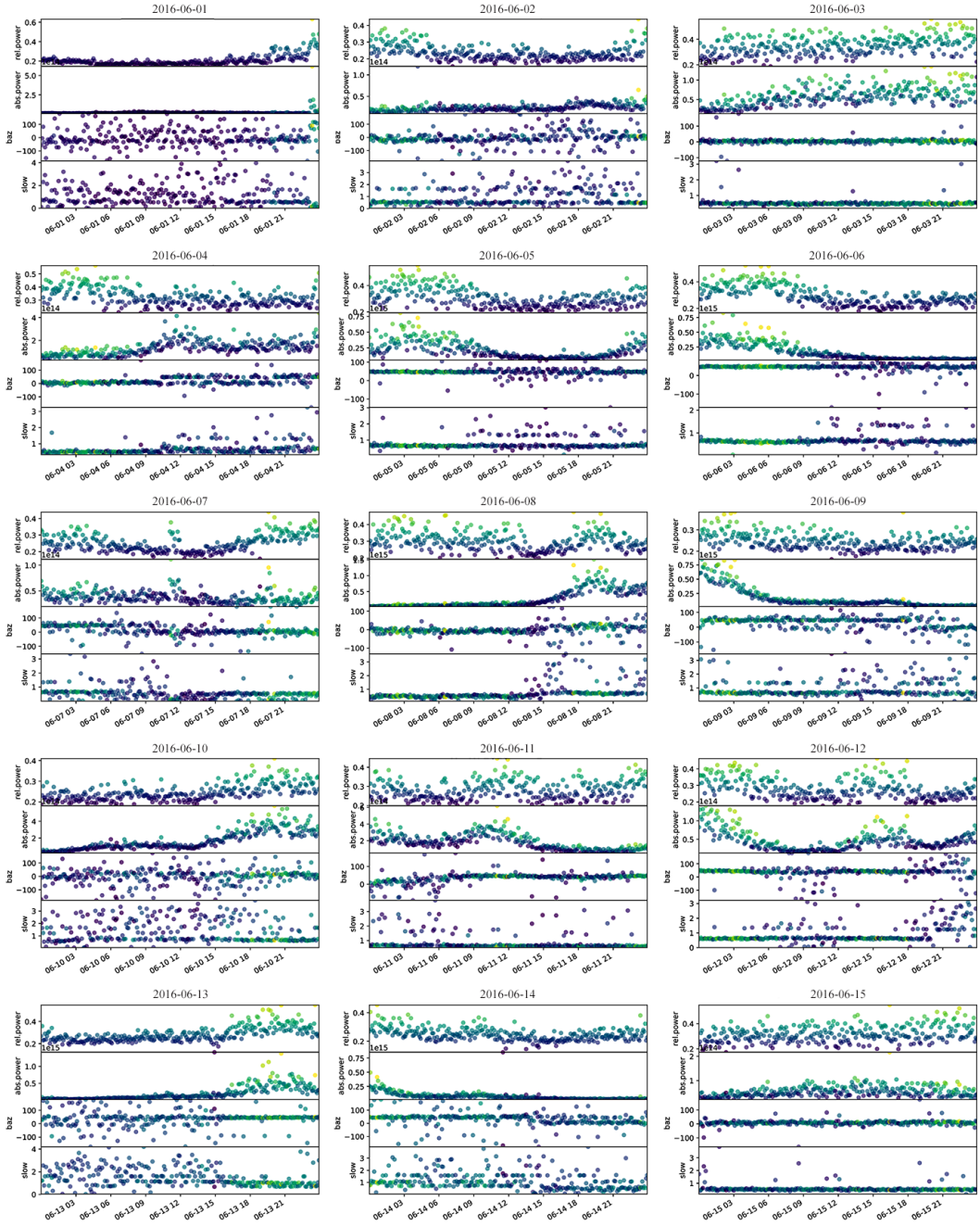


Figure 3.4.1: *Time evolution of relative and absolute power, backazimuth, and slowness for 1st- 15th June 2016 in the 0.1-3.0 Hz range.*

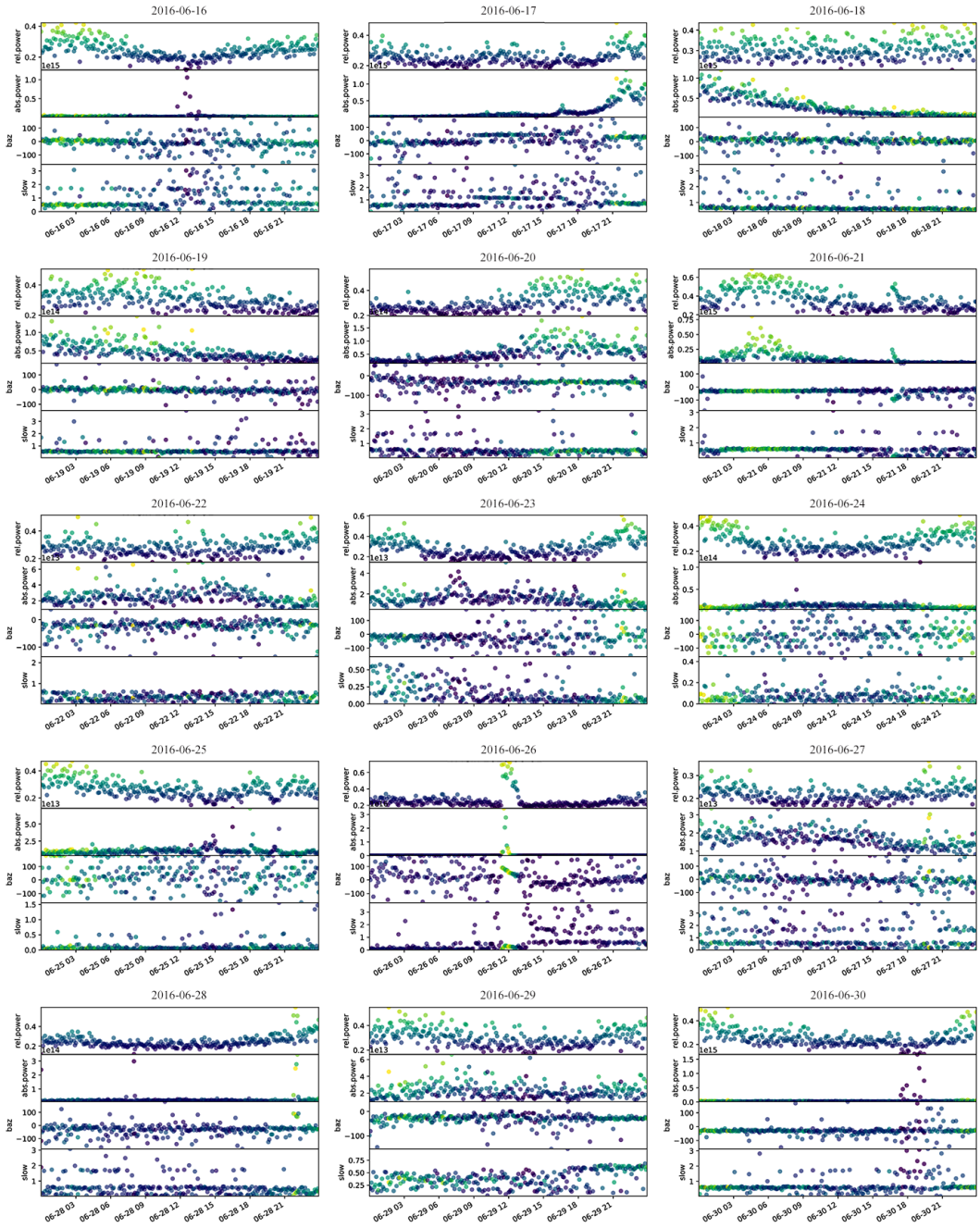


Figure 3.4.2: *Time evolution of relative and absolute power, backazimuth, and slowness for 16^{th}- 30^{th} June 2016 in the 0.1-3.0 Hz range.*

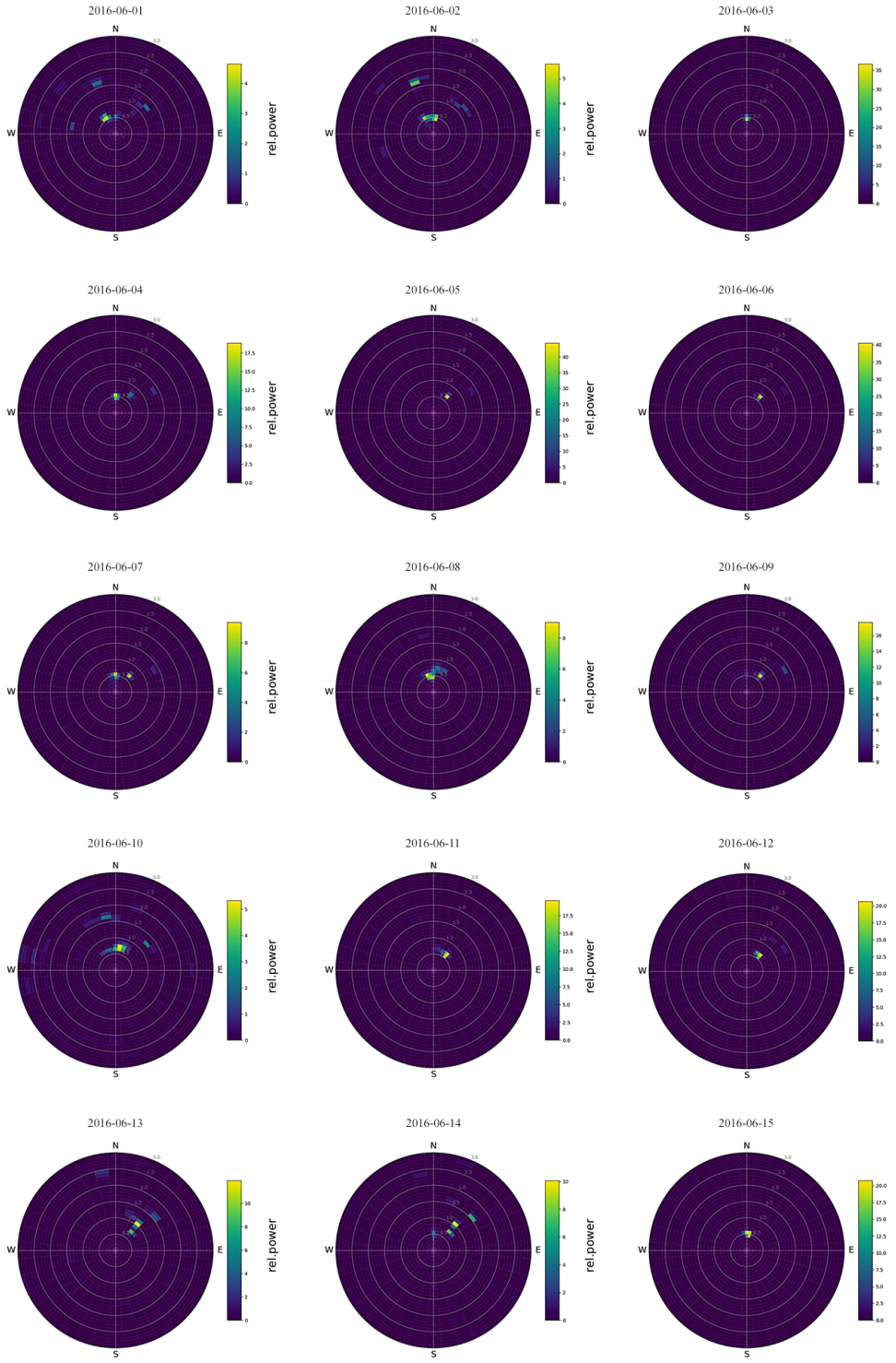


Figure 3.4.3: *Polar plot of relative power for 1st- 15th June 2016 in the 0.1-3.0 Hz range (angular step=10°; slowness step=0.5 s m^{-1}).*

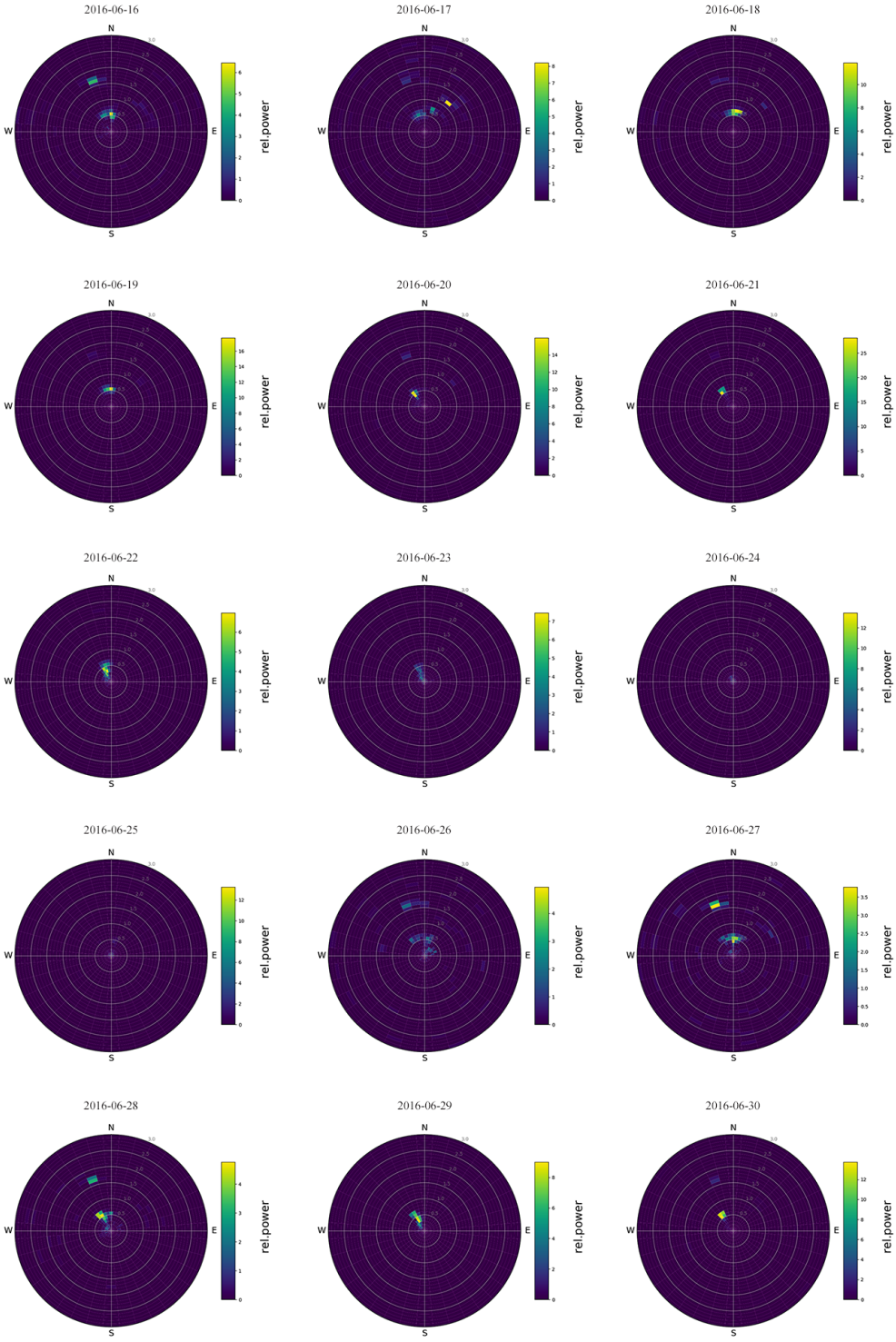


Figure 3.4.4: *Polar plot of relative power for 16^{th}- 30^{th} June 2016 in the 0.1-3.0 Hz range (angular step=10°; slowness step=0.5 s m^{-1}).*

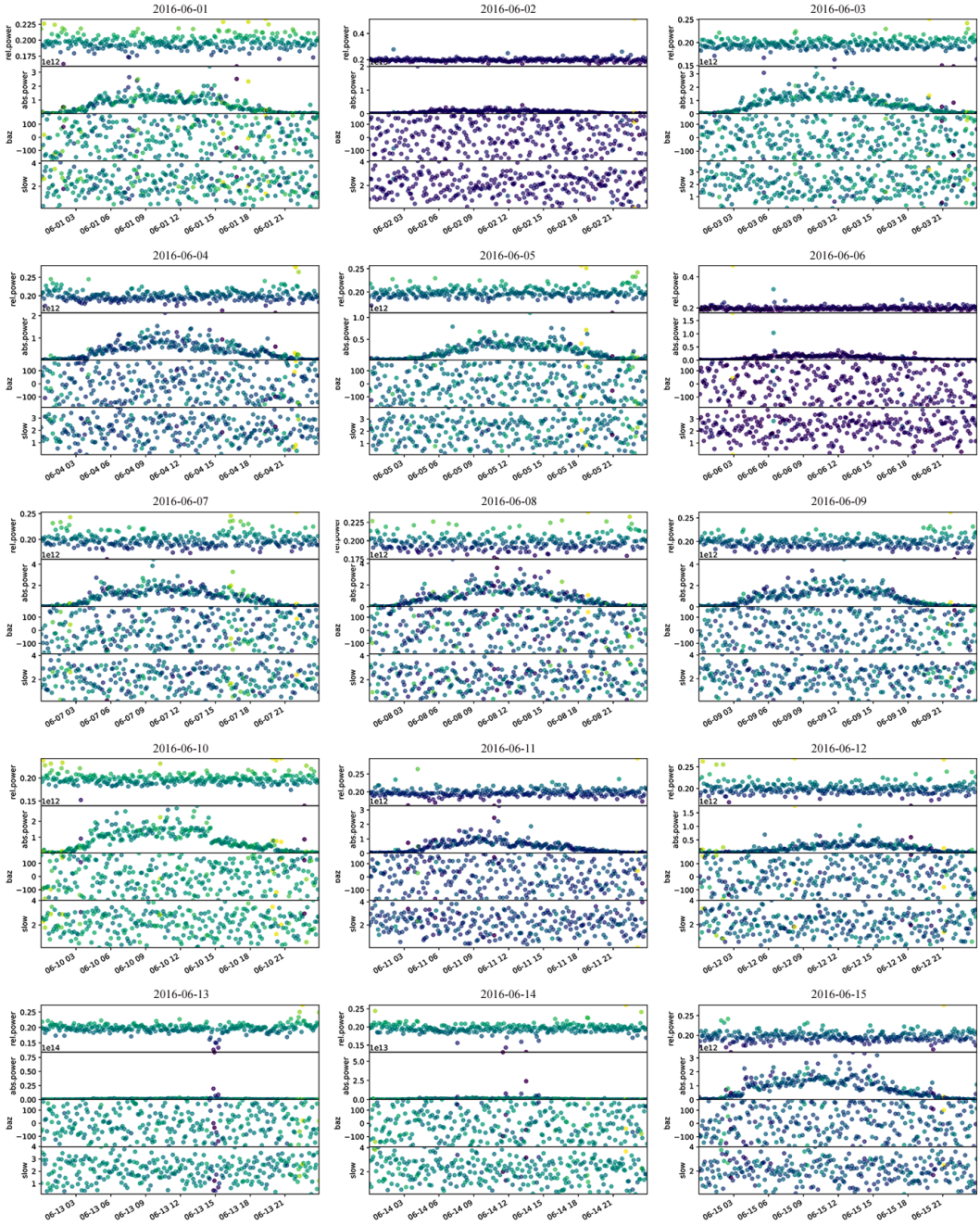


Figure 3.4.5: *Time evolution of relative and absolute power, backazimuth, and slowness for 1st- 15th June 2016 in the 1.5-2.0 Hz range.*

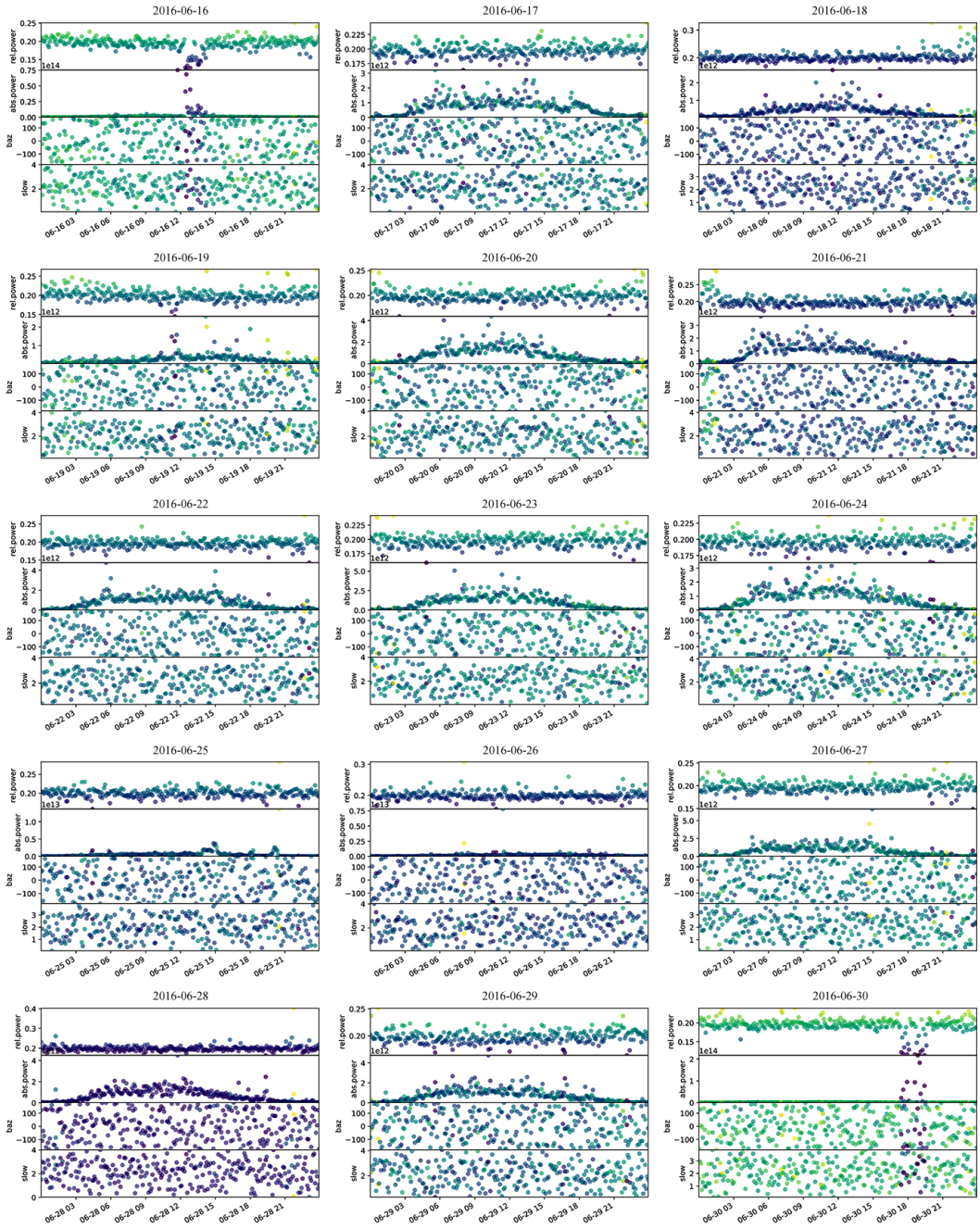


Figure 3.4.6: *Time evolution of relative and absolute power, backazimuth, and slowness for 16th- 30th June 2016 in the 1.5-2.0 Hz range.*

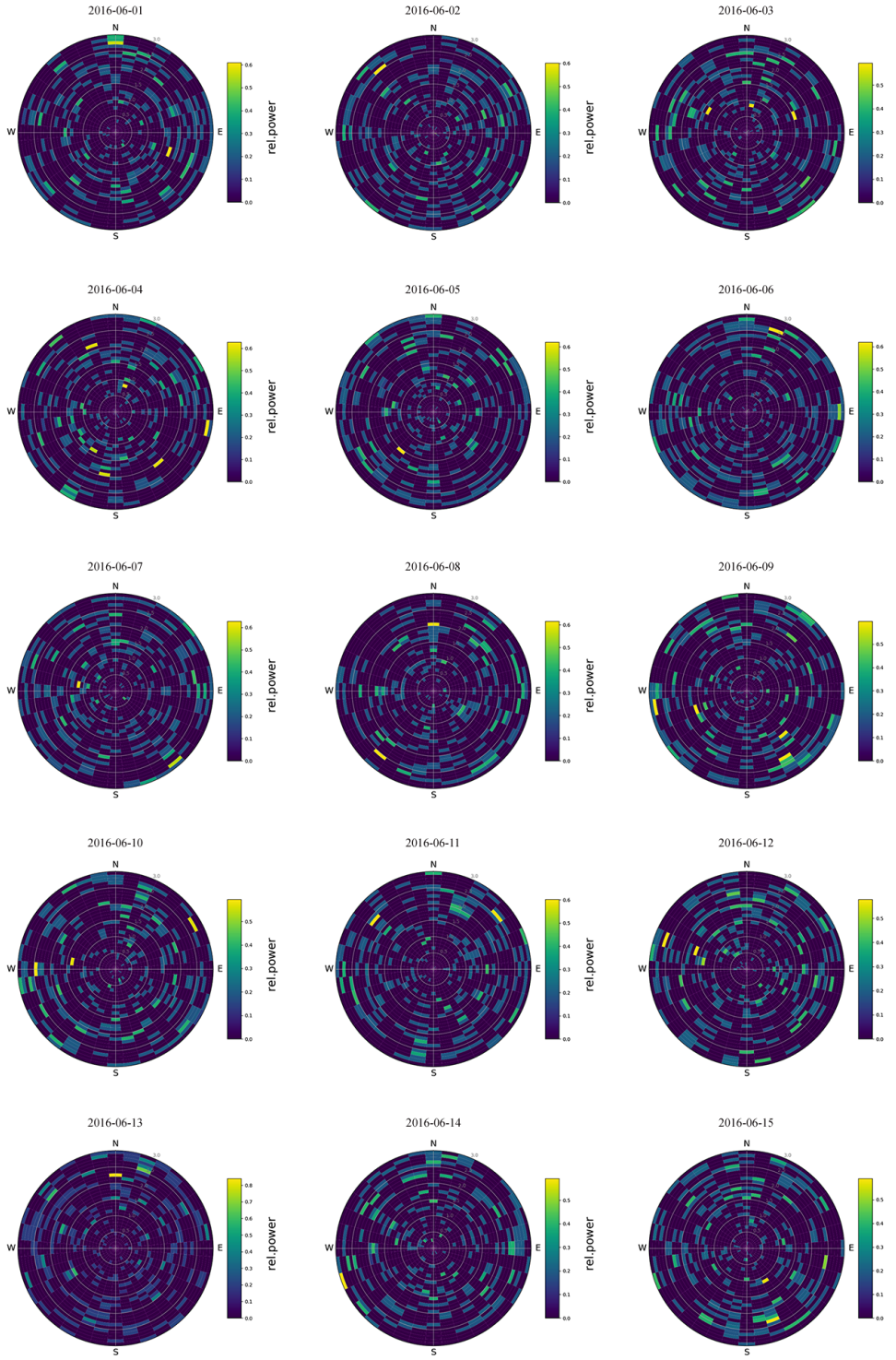


Figure 3.4.7: *Polar plot of relative power for 1st- 15th June 2016 in the 1.5-2.0 Hz range (angular step=10°; slowness step=0.5 s m^{-1}).*

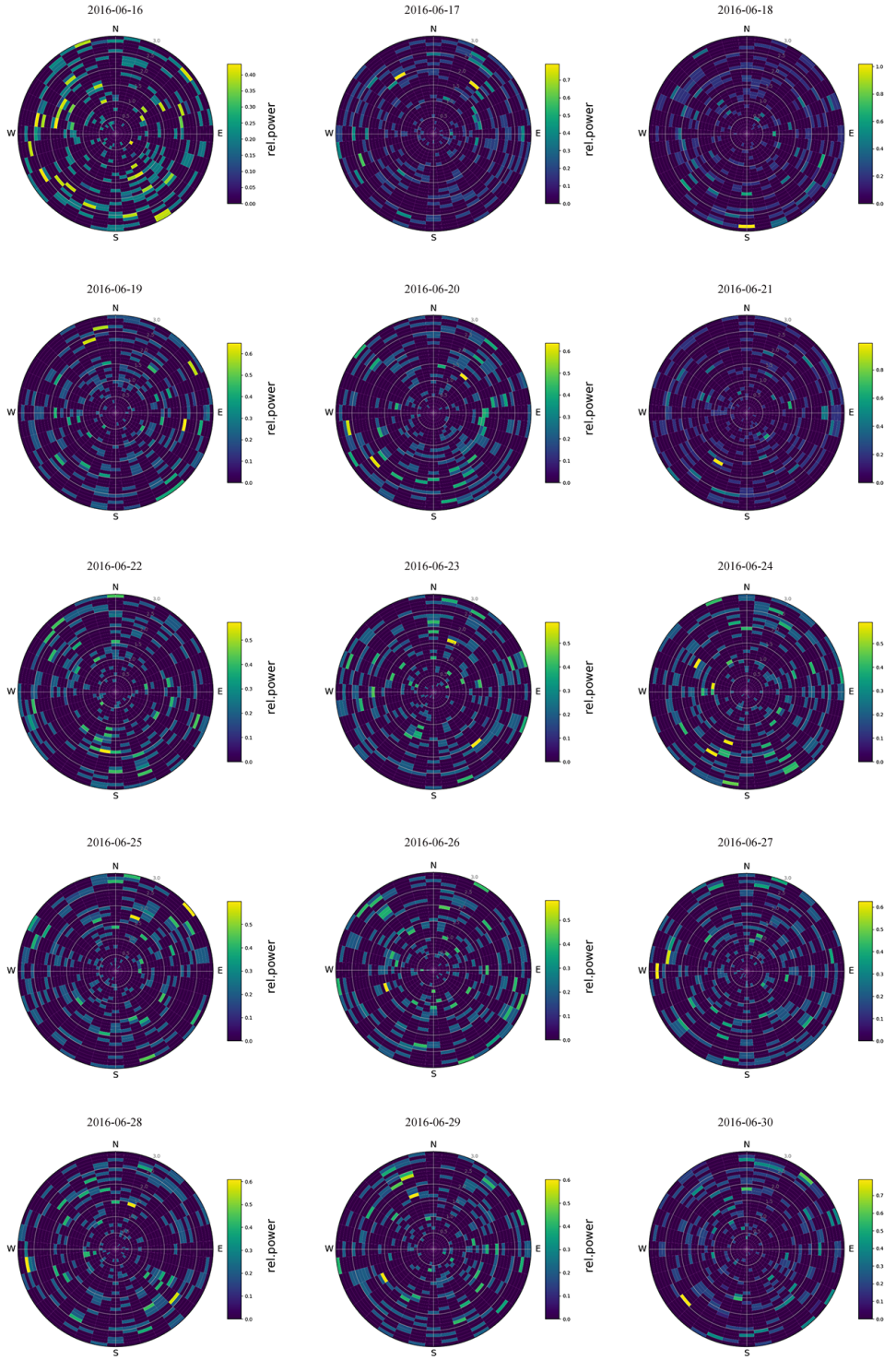


Figure 3.4.8: *Polar plot of relative power for 16th- 30th June 2016 in the 1.5-2.0 Hz range (angular step=10°; slowness step=0.5 s m^{-1}).*

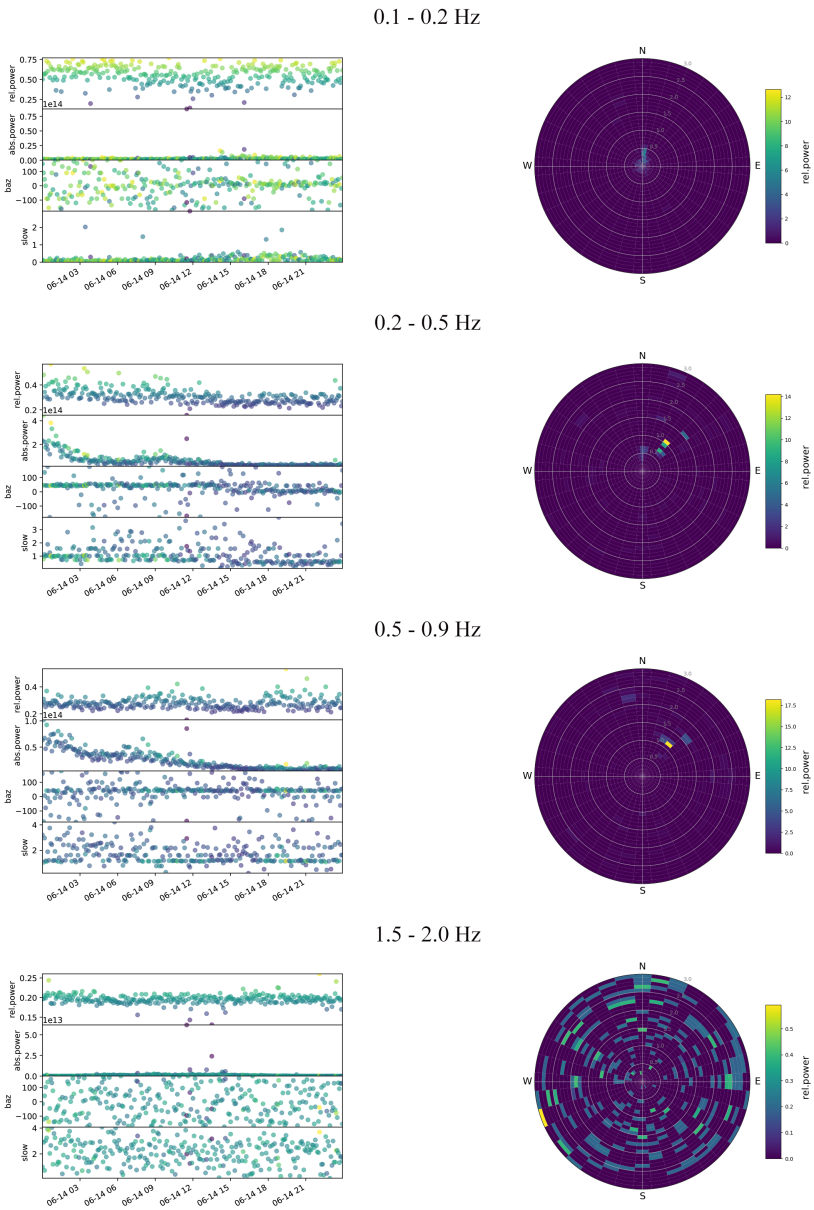


Figure 3.4.9: *(Left panels) Time evolution of relative and absolute power, backazimuth, and slowness, (right panels) with corresponding polar plot of relative power (angular step=10°; slowness step=0.5 s* m^{-1}*) using the data acquired on 2016/06/14 in the frequency band 0.1-0.2, 0.2-0.5, 0.5-0.9 and 1.5-2.0 Hz.*

3.5 Fast Marching Surface-wave Tomography (FMST)

Computation of cross-correlation functions

The present procedure for extracting Green's functions from cross-correlation analysis follows the approach described in Bensen et al. (2007). It is used by Vassallo et al. (2016) on the study of the role of fluids in nucleation and propagation of large earthquakes in Irpinia area, and Vassallo et al. (2019) on an application to the urban area of Benevento city (Southern Italy). The method was implemented by the same authors in a Matlab code that is also considered for this work. The Matlab script allows data processing operations that can be divided into two distinct phases (Fig. 3.5.1).

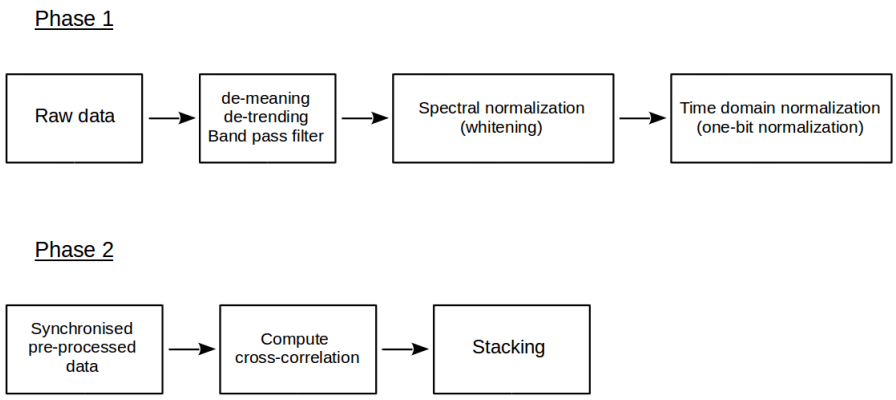


Figure 3.5.1: *schematisation of data processing steps.*

The first phase concerns the data pre-processing performed on single-station data, and the second phase executes the cross-correlation procedure for couples of stations (Weaver and Lobkins, 2001; Shapiro and Campillo, 2004). The first phase of data processing aims to prepare waveform data recorded at each station in an optimal way for cross-correlation, removing any signal that obscures the seismic noise and does not provide information. This procedure was applied to a one-day traces previously converted into the SAC format (Goldstein et al., 2003). The data were decimated at 50 Hz in order to reduce the computation time. The mean and the trend were removed and a band pass filter was applied between 0.1 and 10 Hz. The selection of this range is the result of repeated tests with a data sample. For the first

attempt, the minimum frequency f_{min} can be estimated from

$$\lambda_{min} = \frac{V_S}{f_{min}} \tag{3.5.1}$$

where V_S is an estimate of the S-velocity from an average of the velocities involving the surface wave propagation layers of the site known from other studies; and λ_{min} is the corresponding wavelength. In this case it can be replaced by the value of the maximum interstation distance (array aperture). On the contrary, the estimation of the maximum frequency f_max can not be achieved with the same procedure. A sufficiently large value is considered to ensure that the cross-correlation results belong within the defined frequency range. Finally, the one-bit normalization that consist in a time domain normalization by replacing all positive amplitudes with a 1 and all negative amplitudes with a -1, was used in order to consider only the phases of the signals. This step also reduces the effects of transient signals (as spikes, earthquakes and anthropogenic events). After this phase, for all the 36 station pairs, daily cross-correlation functions was computed over the two selected months and added together resulting stacked cross-correlation funtions. Stacking permits to improve the signal-to-noise ratio and emphasise the coherent signals that correspond to the Green's functions. In this work, the depths of the borehole stations have not been taken into account in the calculation of the interstation distances, because they are located at 50 m depth. This distance can be negligible compared to the distances between their surface projection and other surface stations.

The cross-correlation functions between each station pairs were computed for two months of data (1^{st} June – 31^{th} July 2016) in which the HF operations took place. Different frequency bands was investigated to search for the presence of a coherent signal. First, the 0.1-3.0 Hz frequency bands was examined where the main signal from the cross-correlation was expected to occur, then the analysis was refined for several sub-bands within this range. The figure 3.5.2 shows the resultant 36 cross-correlation stacks arranged by increasing interstation distance obtained in the some of the reviewed frequency bands, in particular 0.1-0.2, 0.2-0.5, 0.5-0.9, and 1.5-2.0 Hz. In the case of the 0.1-0.2 and 0.2-0.5 Hz bands, there is a variable signal with also an overlap of the causal and acausal part which does not allow to evaluate the useful signal part. Between 0.5-0.9 Hz a coherent signal is visible more in the causal part. A coherent phase is well identified in the band 1.5-2.0 Hz (Fig. 3.5.2 d) where it is present in almost all couple of stations at both causal and acausal part and propagates with an estimated velocity of about 400 m/s, consistent with expected near-surface shear-wave velocities. This is proposed to be the band of interest for the whole analysis. For frequencies

higher than 2 Hz there are no emerging signals that can be useful for a tomographic study.

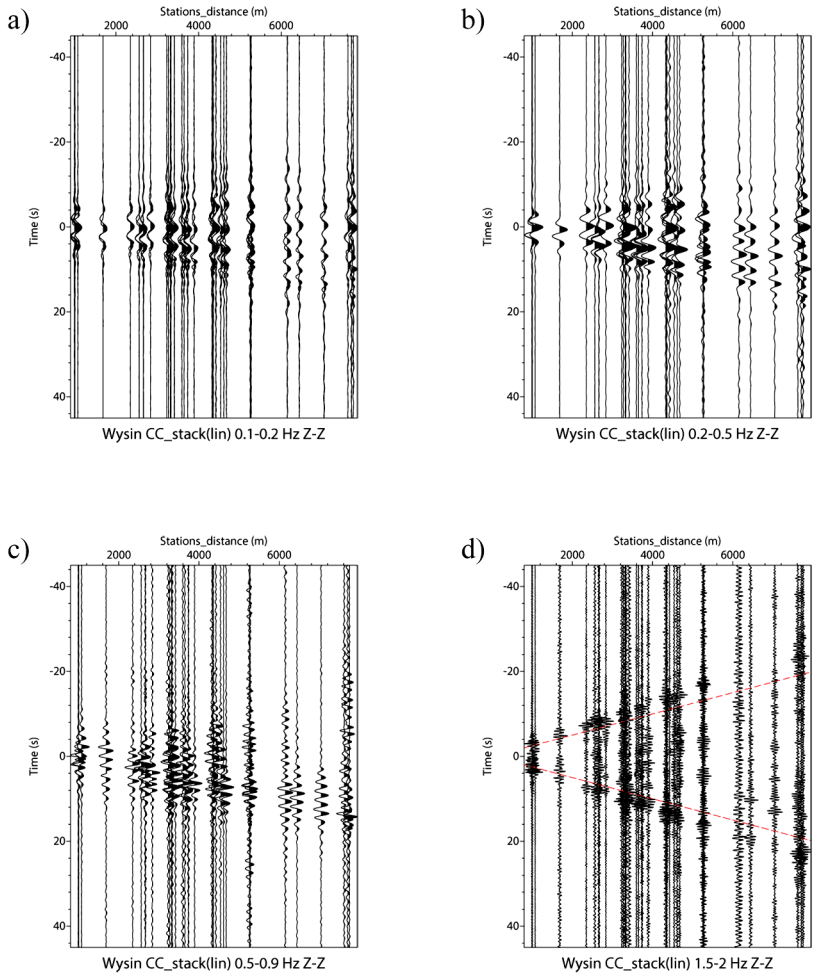


Figure 3.5.2: *The two-month vertical stacked traces plotted with respect to interstation distance for the a) 0.1-0.2, b) 0.2-0.5, c) 0.5-0.9, and d) 1.5-2.0 Hz ranges, respectively. The dashed red lines in d) indicate the theoretical arrival times of a surface wave propagating at a velocity of 400 m/s.*

To further reduce the incoherent noise, the phase-weighted stack (PWS) method proposed by Schimmel and Paulssen (1997) was tested in the computation of the cross-correlation stacks. The PWS is a non-linear stacking technique in which every sample of the linear stack are weight by the coherency of its instantaneous phases

$$g(t) = \frac{1}{N} \sum_{j=1}^{N} s_j(t) \left| \frac{1}{N} \sum_{k=1}^{N} e^{i\phi_k(t)} \right|^{\nu}$$

where

$s_j(t)$= j-th seismic trace;

$\varphi_k(t)$= k-th instantaneous phase of a seismic trace.

In this way stacked traces which are not coherent are suppressed enhancing weak coherent signals. The attenuation of the incoherent signals is controlled by the power ν. The linear stack is retrieved with $\nu = 0$. Phase-weighted cross-correlations stacks were calculated testing two small power values ($\nu = 2$ and $\nu = 5$) as the PWS is a non-linear transformation and therefore a waveform distortion is expected. The results for the four frequency sub-bands studied were compared with the linear cross-correlation stacks and show that this new technique does not lead to any particular improvement, therefore the linear stack method can be considered sufficient for the purpose of these analysis. The figure 3.5.3 shows the results in the 1.5-2.0 Hz range.

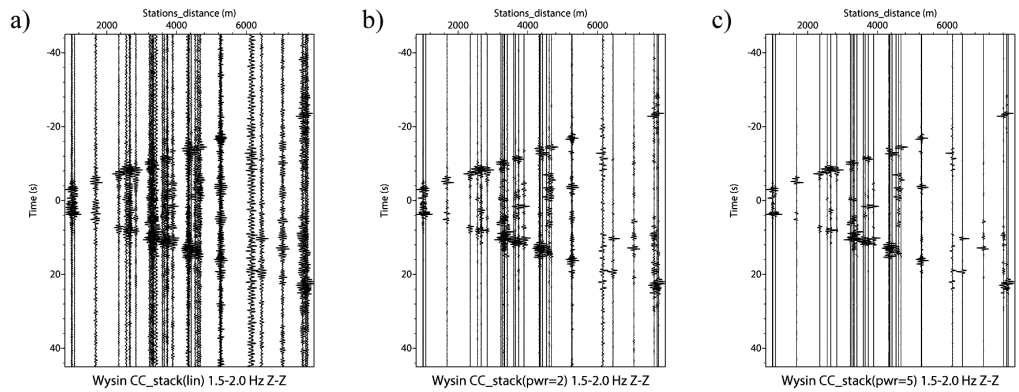


Figure 3.5.3: *a) Linear and b)($\nu = 2$), c) ($\nu = 5$) phase-weighted cross-correlations stacks for 1.5-2.0 Hz frequency band.*

A study of the temporal coherence of the energetic phases of the cross-correlation funtions was carried out in the same considered sub-ranges of frequency. For each station pairs the presence of a coherent signal and its stability with time was analysed over the entire investigated period. In the figures 3.5.4, 3.5.5, 3.5.6, and 3.5.7 are reported the results reconstructed in the band of interest 1.5-2.0 Hz for the 36 couple of stations. Each panel refers to a specific pair and represents the corrisponding time lag of the cross-correlation traces computed for each day of recorded data (horizontal axis).

Each function is normalised to its maximum amplitude value. The blue and red colour scale indicates negative and positive values of the amplitude, respectively. The stack of all the cross-correlation functions is also reported at the top of each panels. For some pairs the cross-correlation functions are symmetric with respect to zero although the acausal part has larger amplitude (ex. GWS1-PLAC, GWS1-SZCZ, SKRZ-CHRW, GW3S-GLOD) or vice versa (ex. GW4S-PLAC, GW4S-CHRW, GWS1-SKRZ, PLAC-GLOD). In other cases, the cross-correlation functions are asymmetric around zero with only the presence of the causal signal (ex. GW4S-SZCZ, GW3S-SKRZ, STEF-GLOD, SZCZ-GLOD). There are cases where the signal is centred around zero (ex. GW4S-GWS1, GW4S-GW3S, GW3S-GWS1). This occurs for pairs involving the borehole stations which have a minor interstation distance (about 1 km) and therefore can be attributable to the fact that signals cross these stations almost simultaneously. Finally, for some pairs is difficult to distinguish an emergent signal and evaluate its coherence (ex. SZCZ-PLAC, CHRW-GLOD, SKRZ-GLOD, STEF-CHRW). For pairs with a predominant signal a good coherence is observed. Indeed, the most energetic signal is present on many time windows analysed and can be considered reasonably stable with time although a loss of coherence occurs at almost regular time intervals.

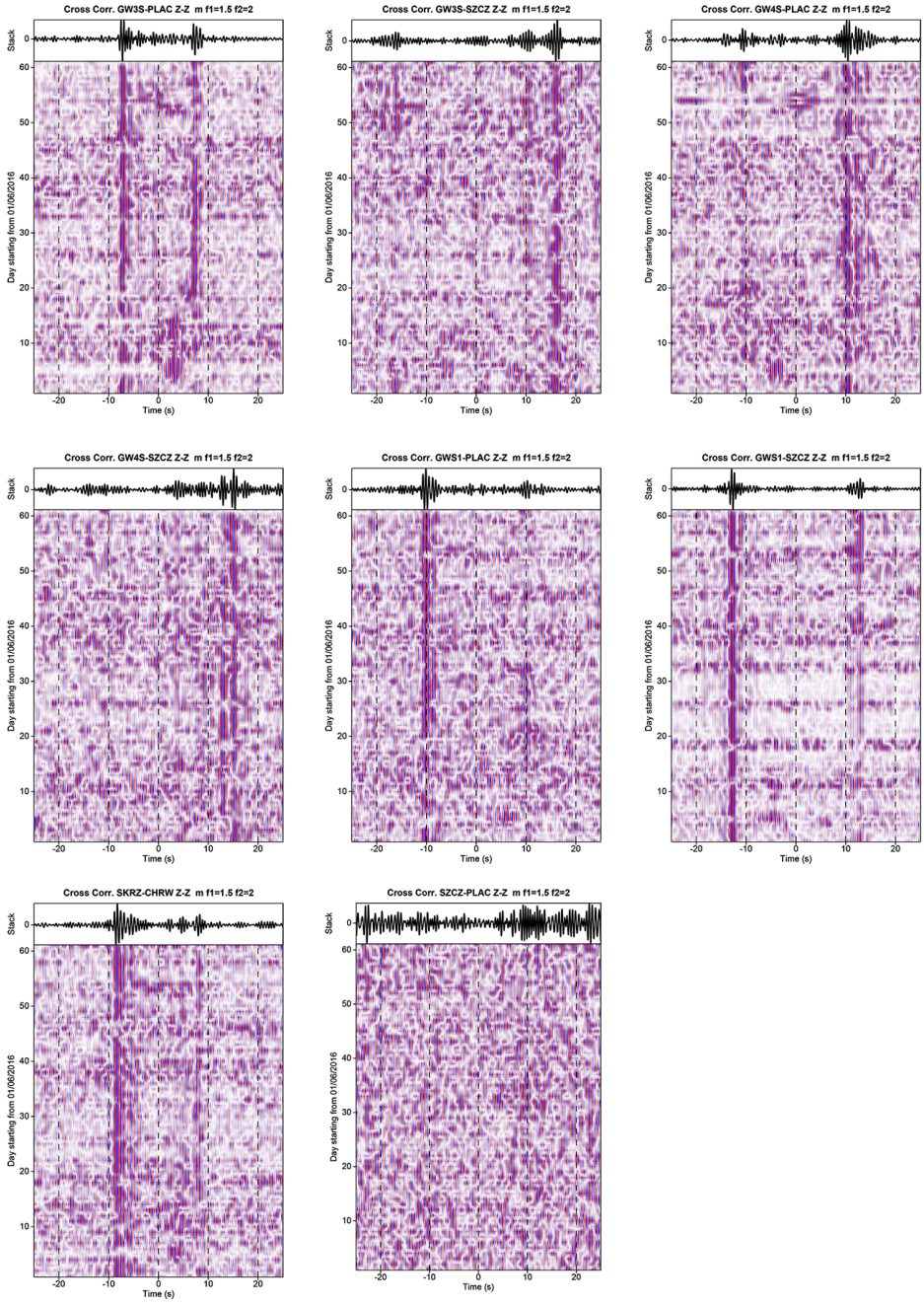


Figure 3.5.4: *Daily cross-correlations as a function of the ambient noise acquisition date filtered in the 1.5-2.0 Hz range for the couple of stations E-W oriented. The stack of the all cross-correlation functions is also depicted (top of panels).*

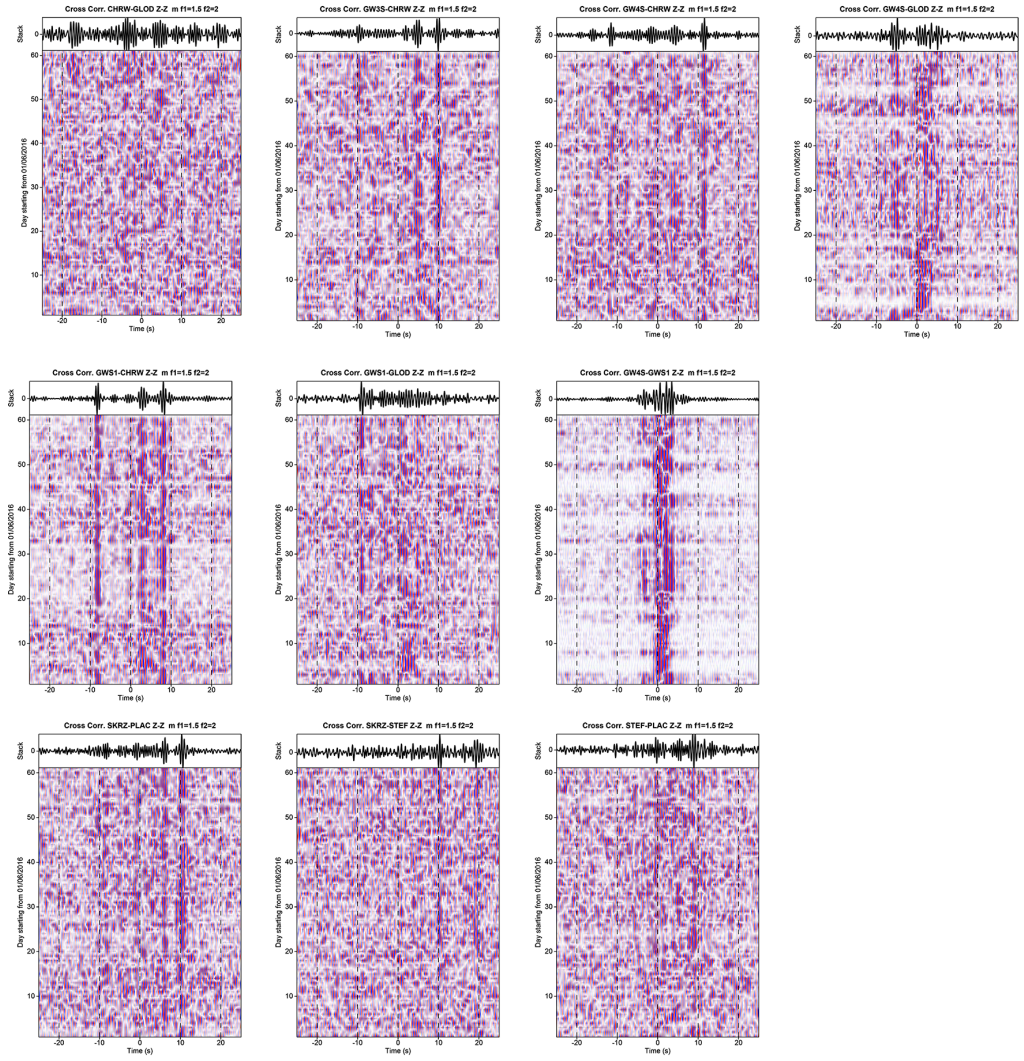


Figure 3.5.5: *Daily cross-correlations as a function of the ambient noise acquisition date filtered in the 1.5-2.0 Hz range for the couple of stations N-S oriented. The stack of the all cross-correlation functions is also depicted (top of panels).*

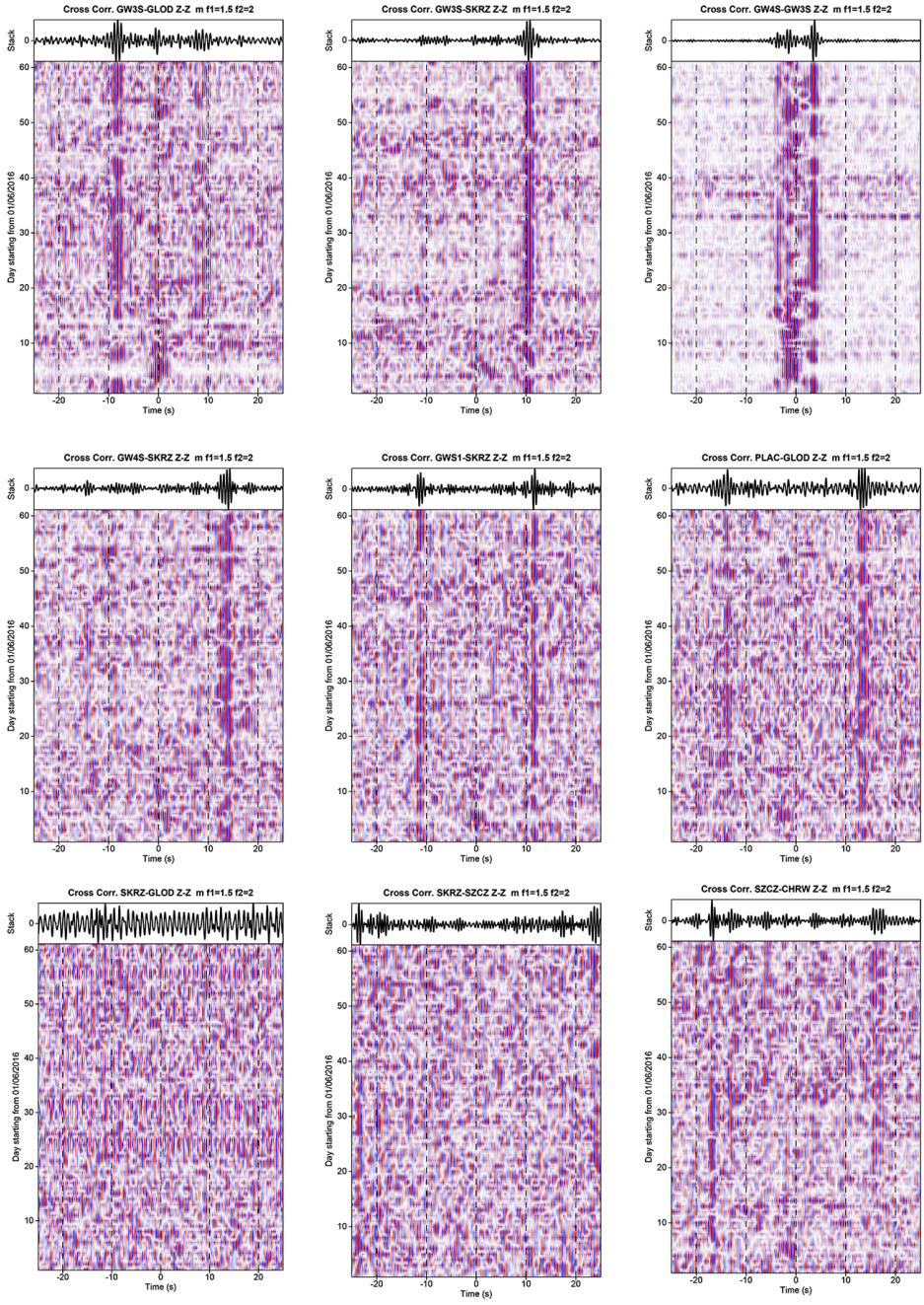


Figure 3.5.6: *Daily cross-correlations as a function of the ambient noise acquisition date filtered in the 1.5-2.0 Hz range for the couple of stations N-SE oriented. The stack of the all cross-correlation functions is also depicted (top of panels).*

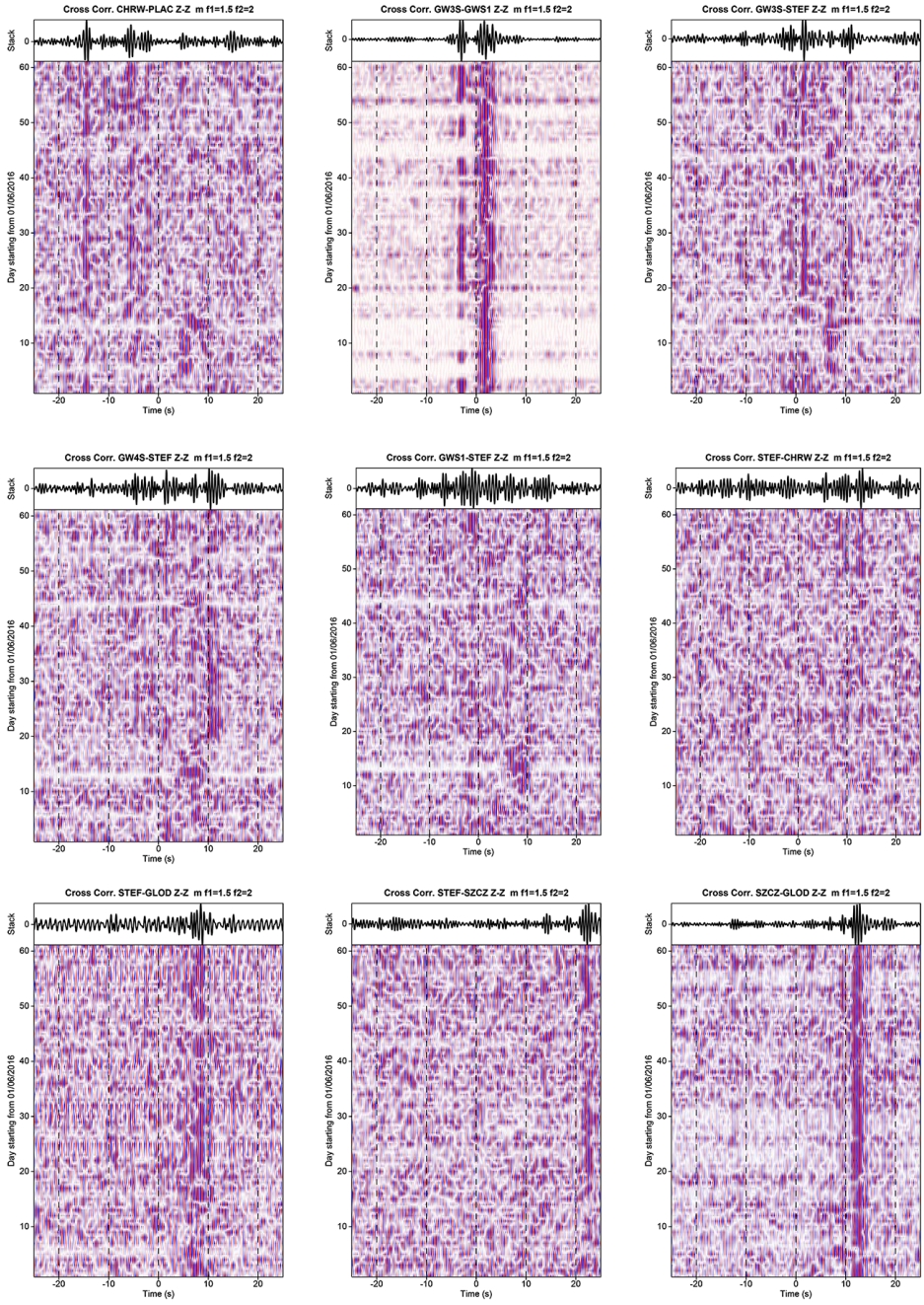


Figure 3.5.7: *Daily cross-correlations as a function of the ambient noise acquisition date filtered in the 1.5-2.0 Hz range for the couple of stations N-SW oriented. The stack of the all cross-correlation functions is also depicted (top of panels).*

Analysing the loss of coherence taking into account of when the HF operations took place (examples in figure 3.5.8), it is observed that the interruptions of signals are present before and after the human activities. This leads to the consideration that they are not be related with the HF operations.

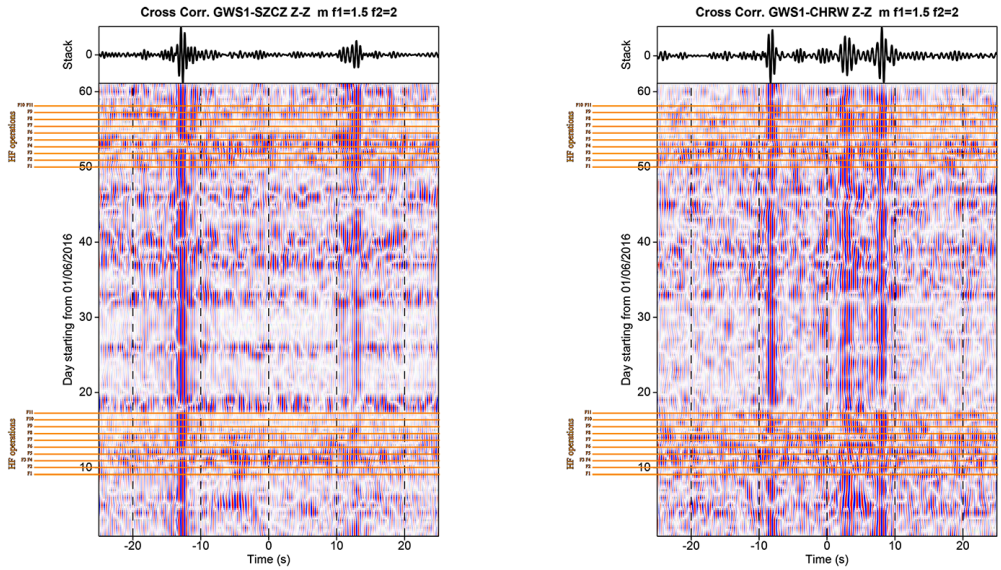


Figure 3.5.8: *Two examples of the temporal analysis in the frequency band 1.5-2.0 Hz of the cross-correlations for the couple GWS1-SZCZ and GWS1-CHRW in which the date of the HF operations are marked with orange lines.*

In this analysis the effects of spectral whitening were also tested in the attempt of improving the results by adjusting the single-station data. It was applied before one-bit normalization during the data processing steps (Fig. 3.5.1). This operation on waveforms makes the magnitude spectrum uniform at all frequencies preserving the phase spectrum of the original seismic traces. The order of the spectral and temporal normalization is important due to non-linearity of the operations on waveforms. The results in figure 3.5.9 show that the spectral amplitude equalization leads to a modest improvement of the cross-correlations only for few station pairs, which is not sufficient to provide additional information for the analysis. In addition, only for some pairs is visible an improvement of the daily cross-correlation functions that confirm the stability of signals over time. Examples for the couple GW3S-GWS1 and PLAC-GLOD are reported in figures 3.5.10 and 3.5.11.

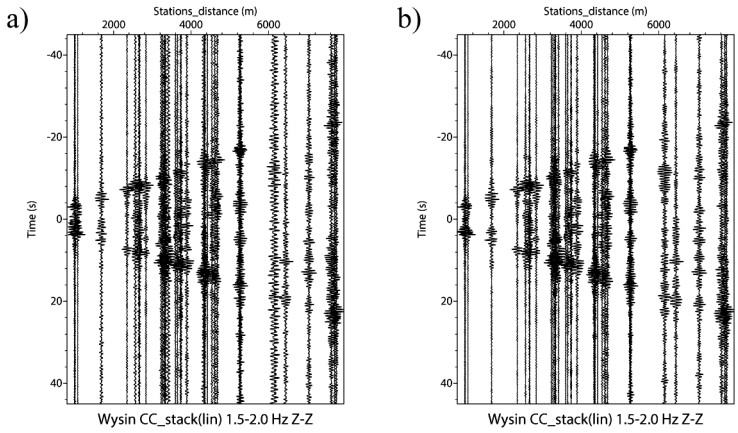


Figure 3.5.9: *Cross-correlation stacks filtered in the 1.5-2.0 Hz range resulting from a data pre-processing a) without and b) with spectral whitening.*

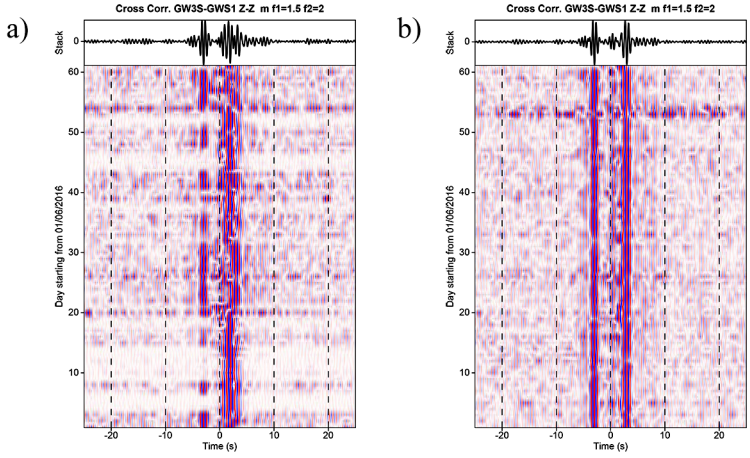


Figure 3.5.10: *Cross-correlations for the couple GW3S-GWS1 filtered in the 1.5-2.0 Hz range obtained with a data pre-processing a) without and b) with spectral whitening.*

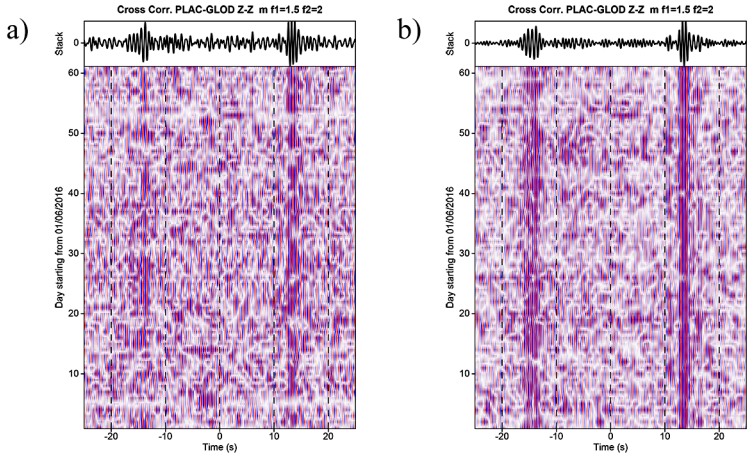


Figure 3.5.11: *Cross-correlations for the couple PLAC-GLOD filtered in the 1.5-2.0 Hz range obtained with a data pre-processing a) without and b) with spectral whitening.*

In order to assess whether the noise source of the study area could have seasonal characteristics in the band of interest 1.5-2.0 Hz, daily cross-correlation functions and stacks were also computed considering the seismic noise acquired in different time windows. For this purpose, a period before (PRE), during (DURING), and after (POST, POST2, POST3) the HF operations were selected (Table 3.2), depending also on the availability of data. In all periods analysed, not all the stations were simultaneously operating. Therefore, the analysis in each period was made with the stations reported in table 3.2.

	Analysed period	Stations in operation
PRE	3^{rd}May - 6^{th} June 2016	GLOD, CHRW, SZCZ, STEF, SKRZ, GWS1, GW3S.
DURING	7^{th}- 30^{th} June 2016	GLOD, CHRW, SZCZ, STEF, SKRZ, GWS1, GW3S.
POST	4^{th}- 30^{th} November 2016	GLOD, CHRW, SZCZ, STEF, SKRZ, GWS1, GW3S.
POST2	4^{th}May - 6^{th} June 2017	GLOD, CHRW, PLAC, GWS1, GW3S, GW4S.
POST3	7^{th}- 30^{th} June 2017	GLOD, CHRW, PLAC, GWS1, GW3S, GW4S.

Table 3.2: *Date of the considered periods for the seasonality analysis and relative operating stations.*

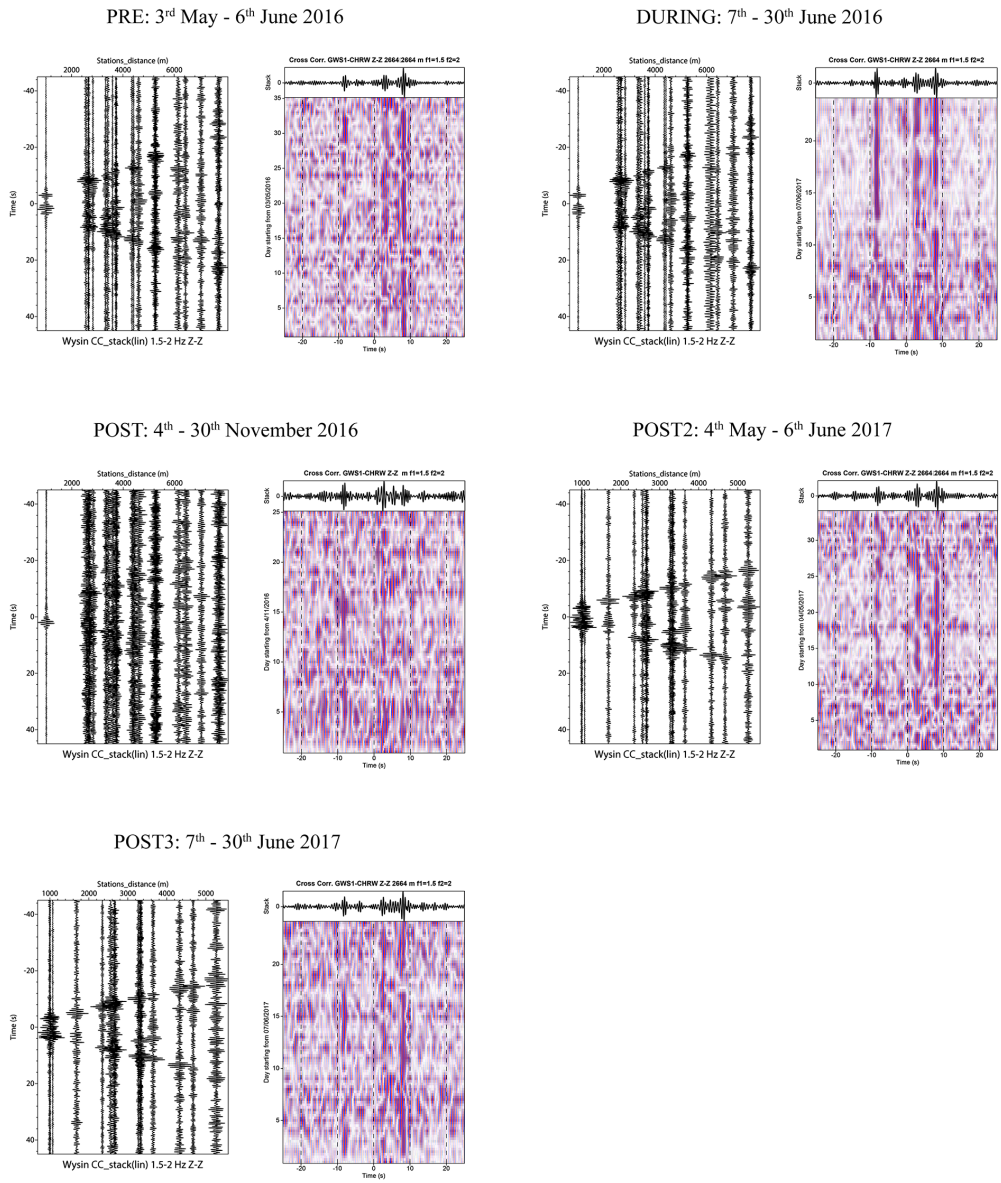


Figure 3.5.12: *Seasonality analysis for five analysed periods (see top label for each panel) in the frequency band 1.5-2.0 Hz.*

A beamfoarming analysis was also carried out in these different selected periods to study the noise source. Also in these cases, an isotropically distributed noise source is found by filtering between 1.5 and 2 Hz. For each period, the results of the cross-correlation stacks and an example of the daily cross-correlation for the couple GWS1-CHRW are reported in the figure 3.5.12. The presence of a coherent signal propagating across all pairs is confirmed

in each period except in the winter time (POST), where no energetic signal emerges from the analysis. Therefore, it is supposed that there are seasonal factors influencing the ambient seismic noise in the Wysin site.

So far only the vertical ZZ components of continuous ambient noise records have been evaluated, but it is possible to calculate all nine components of the cross-correlation tensor (ZZ, RR, TT, RZ, ZR, RT, TR, TZ, ZT). To integrate and compare the results inferred from the vertical components, the cross-correlation analysis was also performed for the RT, ZR, ZT components. All data processing was the same as for ZZ components but in this case, before the pre-processing phase, the raw data of the horizontal components was rotated in the radial (R) and transverse (T) directions with respect to the connections between the positions of the stations. This operation is necessary because the Rayleigh and Love surface waves can be emphasized and reconstructed along these directions. In addition, the ZRT system results best suitable for studying the wave field produced by one receiver (which acts as a virtual source) and recorded to another receiver. The analysis was carried out in the same frequency bands investigated for the ZZ components. The comparison of the results between ZZ, RT, ZR, and ZT components is displayed in the figures 3.5.13, 3.5.14, 3.5.15, and 3.5.16. For all the frequency ranges, the results of the mixed components mostly confirm the information obtained by investigating only the vertical components.

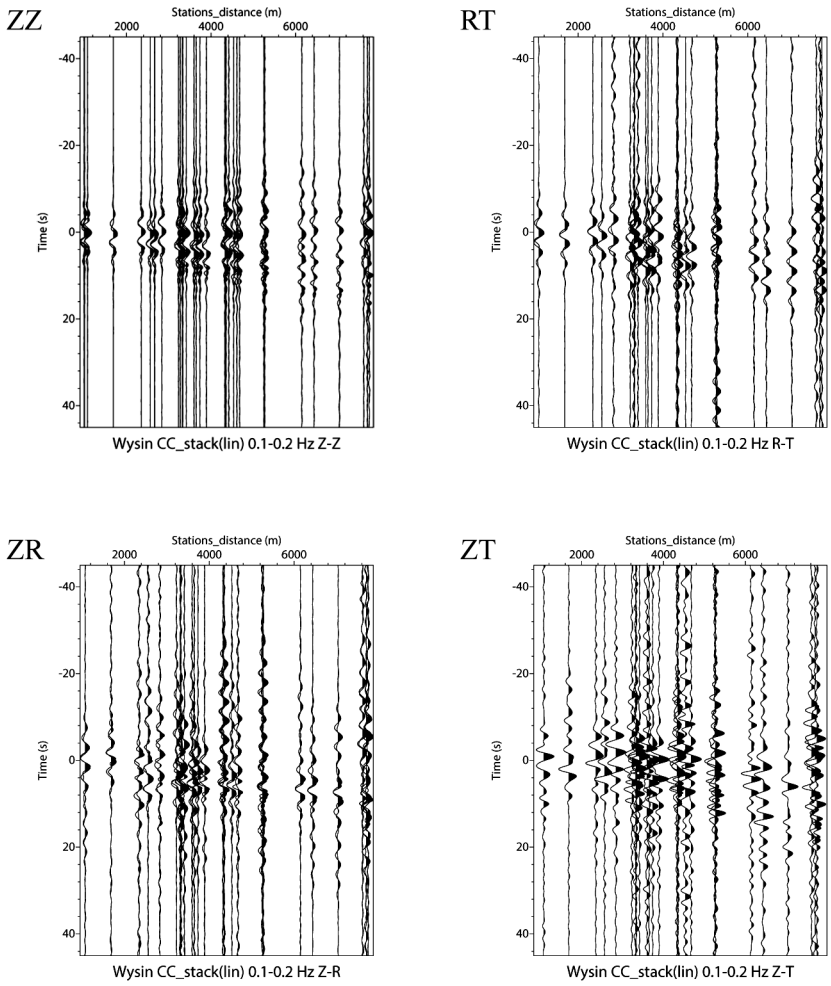


Figure 3.5.13: *Cross-correlation stacks as a function of the interstation distance for ZZ, RT, ZR, and ZT components in the 0.1-0.2 Hz range.*

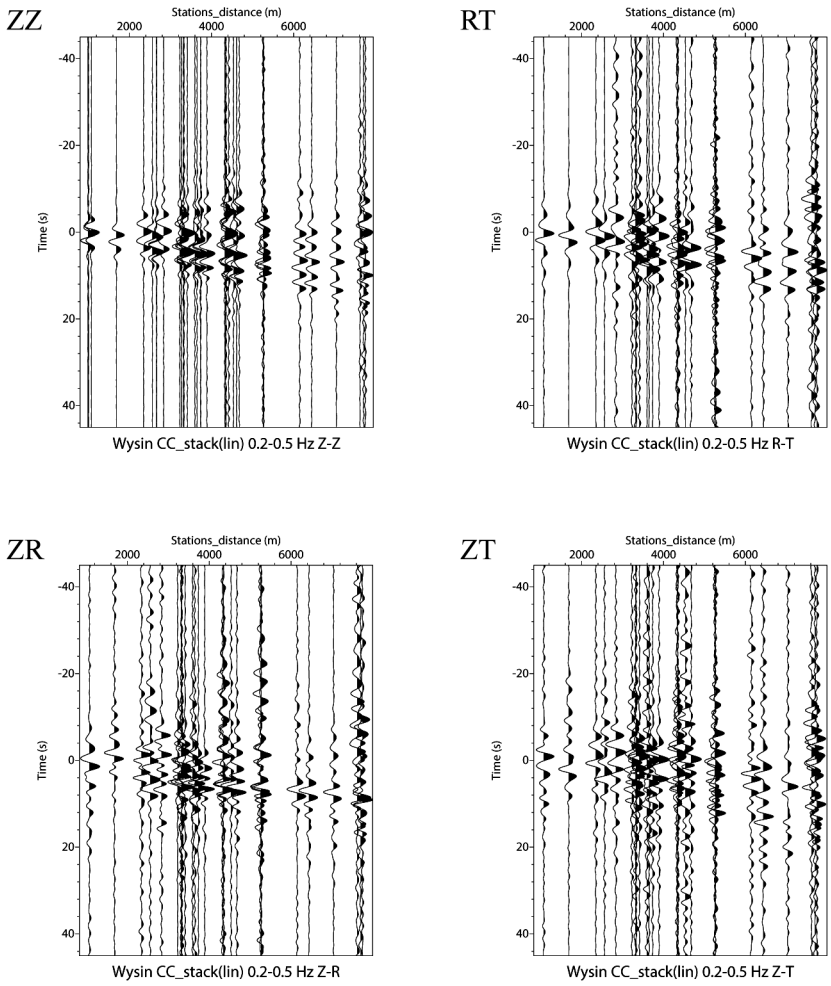


Figure 3.5.14: *Cross-correlation stacks as a function of the interstation distance for the ZZ, RT, ZR, and ZT components in the 0.2-0.5 Hz range.*

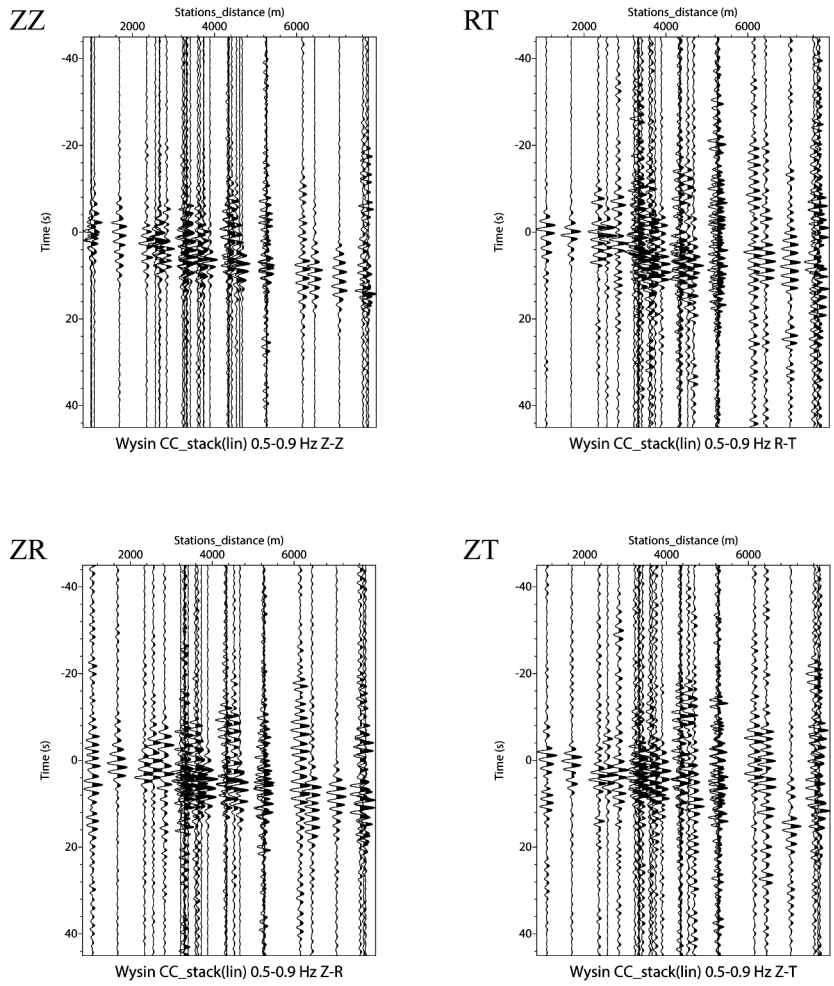


Figure 3.5.15: *Cross-correlation stacks as a function of the interstation distance for the ZZ, RT, ZR, and ZT components in the 0.5-0.9 Hz range.*

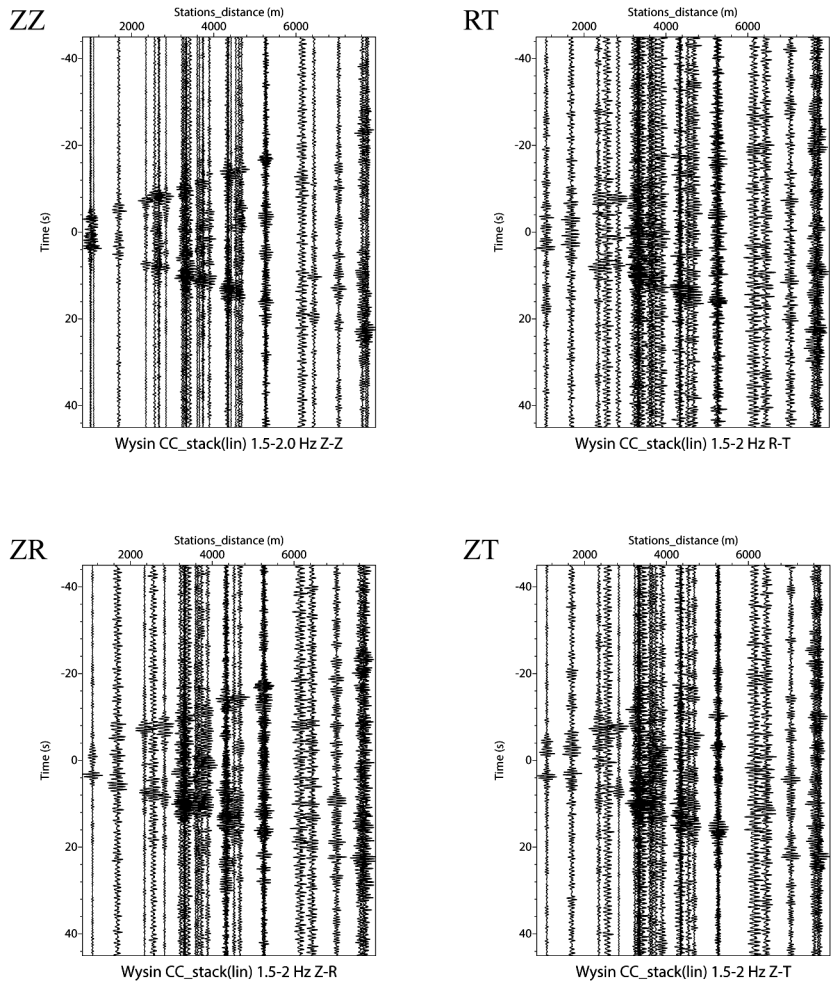


Figure 3.5.16: *Cross-correlation stacks as a function of the interstation distance for the ZZ, RT, ZR, and ZT components in the 1.5-2.0 Hz range.*

Tomographic inversion

In this work tomographic images of the investigated area was reconstructed using the Fast Marching Surface-wave Tomography (FMST), a Fortran 90 software package for performing traveltime tomography (Rawlinson N., http://rses.anu.edu.au/~nick/surftomo.html). It is an iterative non-linear inversion method in which traveltimes from point sources to receivers are computed in 2-D spherical shell coordinates (θ, φ). FMST is based on the FMM algorithm (Sethian, 1996; Rowlinson and Sambridge, 2005) to solve the forward problem of traveltime prediction, and the subspace inversion method (Kennett et al., 1988) to solve the inverse problem.

The tomography was performed by the inversion of the surface wave arrival times extracted from the previous cross-correlation analysis in the 1.5-2.0 Hz frequency band where the most energetic signal emerged. This implies that a fixed-frequency tomography was performed. In the case of Rayleigh surface waves, a skin depth given by 1/3 of the wavelength of the analysed signal can be considered a good approximation for the estimation of the maximum depth of the medium investigation. In this study the investigation depth reaches a maximum of about 70 m. Arrival times for tomographic inversion were obtained by performing manual picking on the stacked cc functions in the time domain. The phase-weighted stacks ($\nu = 2$) were considered because of improved signal-to-noise ratio compared to the linear stack, as shown in the histogram in figure 3.5.17. For each traces, the symmetric component was derived by summing the causal part and the reversed anticausal part, and the envelope was calculated. The arrival times were picked on the maximum of the envelope of all 36 phases sorted by the offset, following the lateral continuity of patterns associated with the surface wave in the time domain (Fig. 3.5.18 a). The error associated with arrival time measurements is 0.5 s, given by the amplitude at half height of the maximum of the envelope function in the time domain.

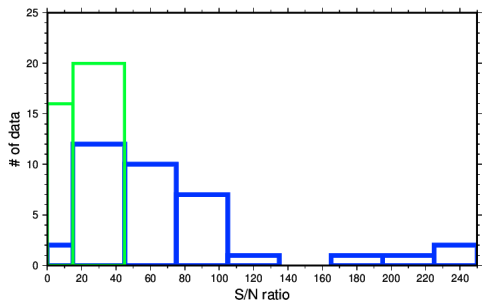


Figure 3.5.17: *Frequency histogram of the signal-to-noise ratio computed for the (green) linear and (blue) phase-weighted ($\nu = 2$) cross-correlation stacks.*

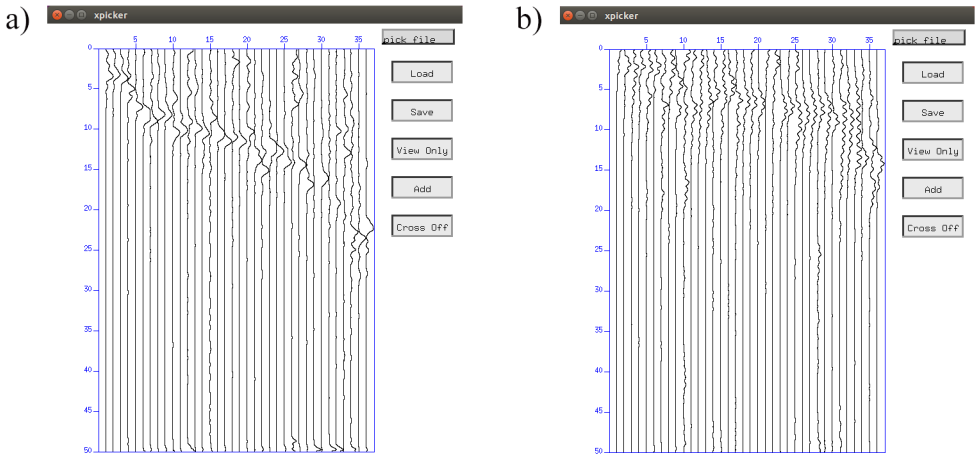


Figure 3.5.18: *The 36 envelope functions in the time domain sorted by the offset for manual picking in the a) 1.5-2.0 Hz and b) 0.5-0.9 Hz ranges.*

The 0.5-0.9 Hz frequency range was not considered for tomography because these frequencies are not easily detectable for performing manual picking in the time domain (Fig. 3.5.18 b).

To perform the tomographic inversion, an initial uniform velocity model of 350 m/s was estimated by averaging the values of the 36 velocities evaluated between each path joining the stations (Fig. 3.5.19).

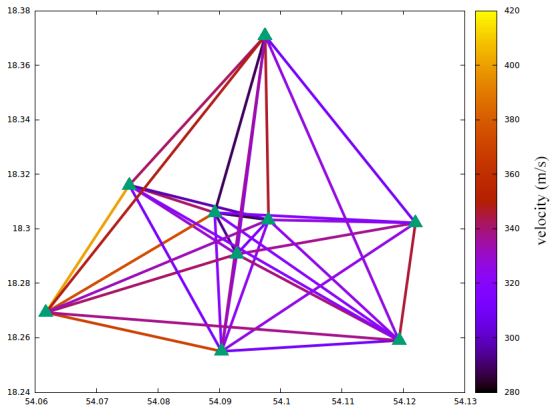


Figure 3.5.19: *Velocity values related to lines joining the stations (green triangles) computed considering the arrival times picked on the maximum of the phase-weighted stack envelopes and interstation distances.*

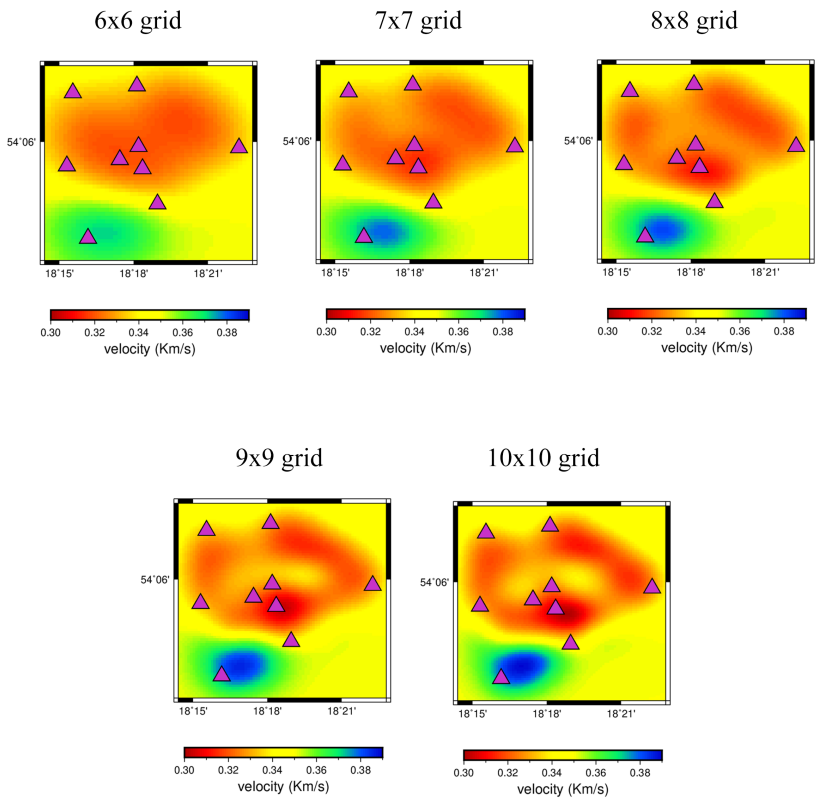


Figure 3.5.20: *Solution models resulting from tomographic inversions by varying only the inversion grid size (see top label for each panel).*

Several tests were made in order to select the optimal size of a 2-D inversion grid. In the figure 3.5.20 the tomography results for the tested grid sizes are reported. The criterion for selecting a first size value of inversion grid was based on the RMS trend related to iteration number of each tomographic inversions (Fig. 3.5.21). Although the RMS still decreases as the number of iterations increases, 6 is the maximum number of iterations chosen for tomographic inversion. This is because by increasing the number of iterations the value of RMS becomes smaller than the uncertainty of the data (0.5 s), so that the solution model is not acceptable.

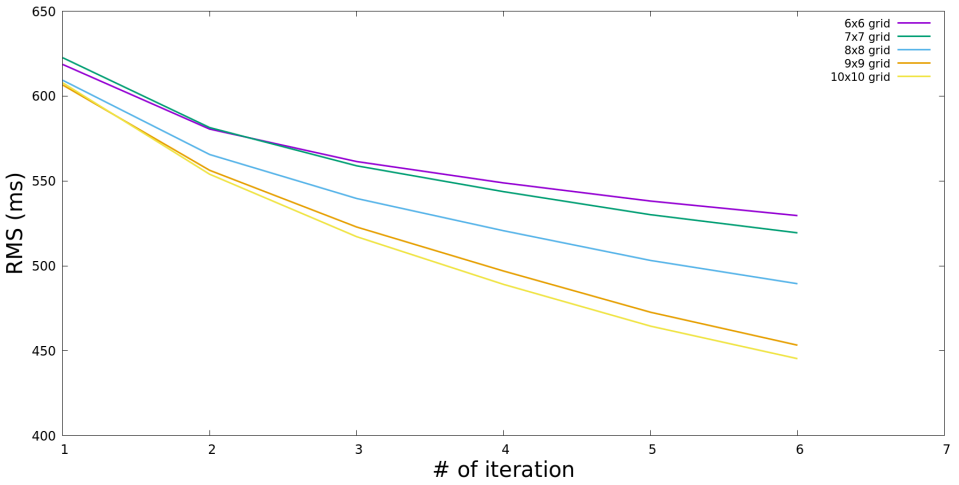


Figure 3.5.21: *RMS behaviour as a function of the iteration number of the tomographic inversion performed for different grid sizes.*

Follow the principle of parsimony, the 2-D grid of 9x9 velocity node represented a good compromise between the RMS value and model semplicity. The nodes are not equally spaced along the longitude and latitude, with a relative distance of 0.140° and 0.077°, respectively. Once the inversion grid was defined, the tuning of the regularisation parameters was carried out by constructing the trade-off curve between data variance after inversion and the model variance for different damping parameters, and the trade-off curve between data variance after inversion and the model roughness for different smoothing parameters (Fig. 3.5.22). Analysing the trade-off curves, the optimal choice of damping and smoothing factors is given by the values at the knee of the curves because they represent the best compromise between the data fit and the simplicity of the model, addressing the problem of nonuniqueness of the solution. The damping value 2.5 and the smoothing value 5 were selected for the final inversion.

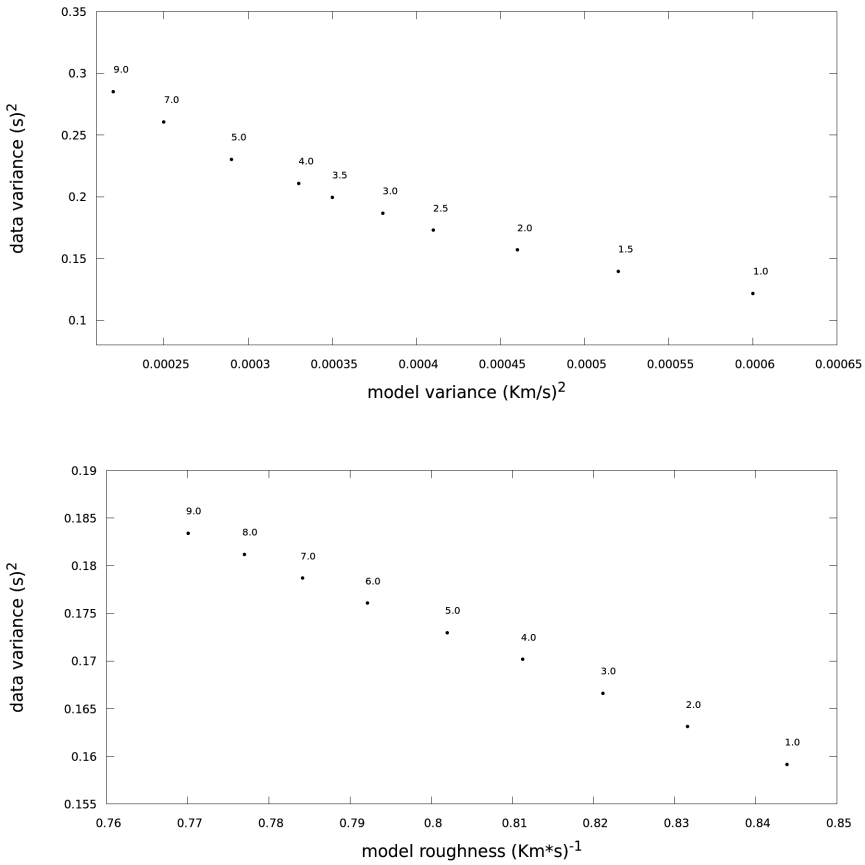


Figure 3.5.22: *Trade-off curves for the selection of the best (top) damping and (bottom) smoothing factors. Each point on the curves represents one inversion carried out with the indicated value of the these two parameters.*

The tomographic inversion strategy allowed the parameters setting to solve the inverse problem, obtaining the optimal tomographic model solution that minimises the traveltime residuals. The histogram in the figure 3.5.23 compares the traveltime residuals between the initial and solution model. The final solution can be considered reliable as the residuals are zero-centred and have a narrower dispersion around the zero value. The image in the figure 3.5.23 represents the tomographic model solution indicating the velocity variations for the Wysin site within a penetration depth of approximately 70 m. The velocities ranges between 300 m/s and 380 m/s. A low-velocity zone is present in the central part of the map in the proximity of wells and is surrounded by higher velocities, whereas a high-velocity anomaly is located in the south-west part.

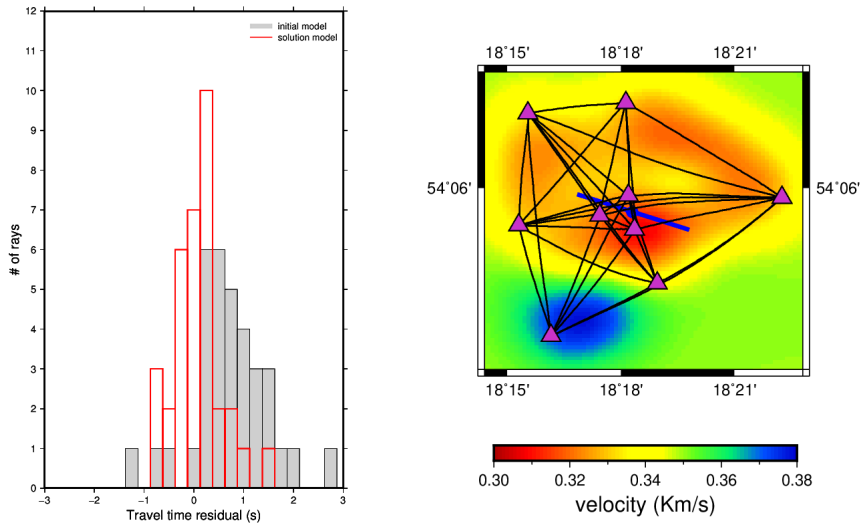


Figure 3.5.23: *(left panel) Frequency histogram of the (grey) initial and (red) final data fit; (right panel) solution of the tomographic model in the frequency band 1.5-2.0 Hz, with raypaths, stations and wells locations superimposed.*

A resolution study of the tomographic results was performed using a synthetic checkerboard test. Different synthetic models were generated adding three different values of velocity perturbations to the tomographic solution model, for two choices of checkerboard grid. The inversion of the synthetic data was performed using the resulting model of the real data inversion as the starting model and considering the same parameters setting in the real data inversion. Figure 3.5.24 shows the checkerboard perturbations added to final tomographic solution model using anomalies with a size of about 4.59 × 4.27 km with 0.03, 0.01 and 0.005 km/s maximum perturbation at the grid vertices, compared with the results of the resolution test. Figure 3.5.25 shows the same graphs but with anomalies having a size of about 1.84 × 1.71 km. The areas where the perturbation reconstructed by the checkerboard test has the same sign of the input velocity anomaly can be considered resolved. The results of the resolution test show that the checkerboard pattern is well reconstructed in the case of the 4.59 × 4.27 km checkerboard anomalies size for the three considered maximum velocity perturbation. Instead, for the smallest size checkerboard the shape of the anomalies are not correctly reconstructed. However, for the velocity model obtained from the tomographic inversion it is not necessary to investigate minor anomalies.

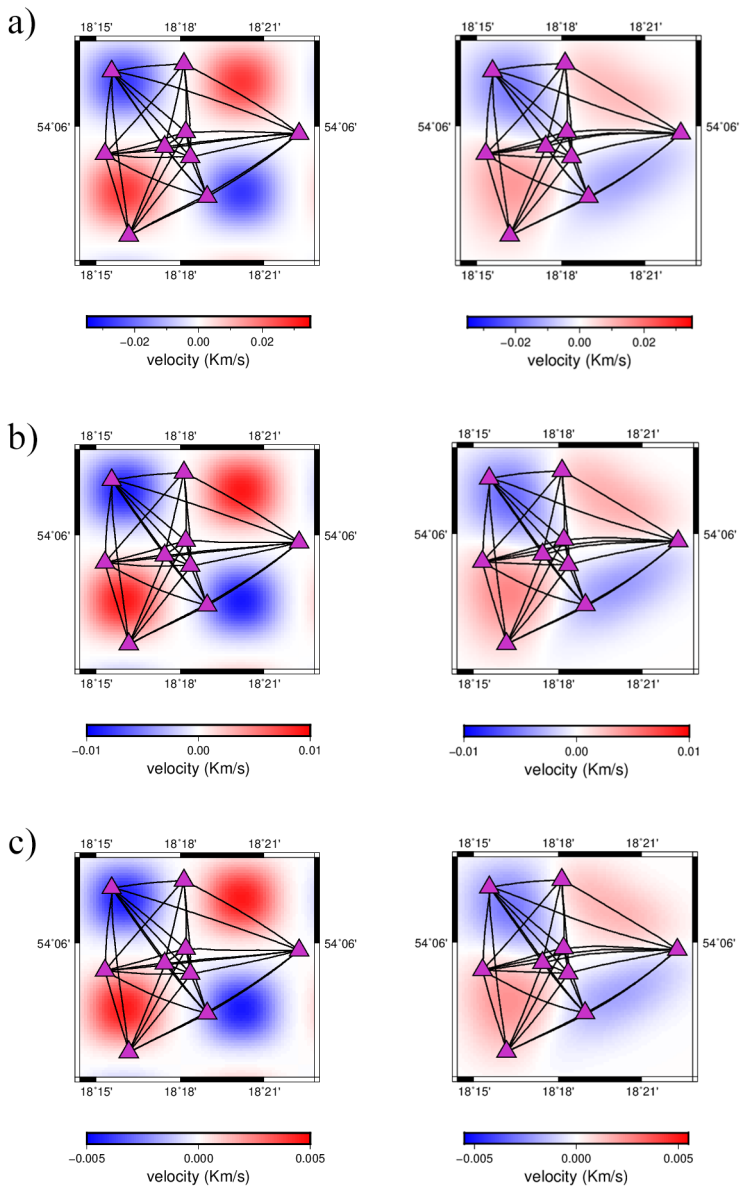


Figure 3.5.24: *(left panel) Checkerboard perturbation obtained with a 4.59 × 4.27 km checkerboard anomalies size and a) 0.03 b) 0.01 and c) 0.005 km/s maximum perturbation at the grid vertices compared with (right panel) checkerboard perturbation retrieved from the resolution test.*

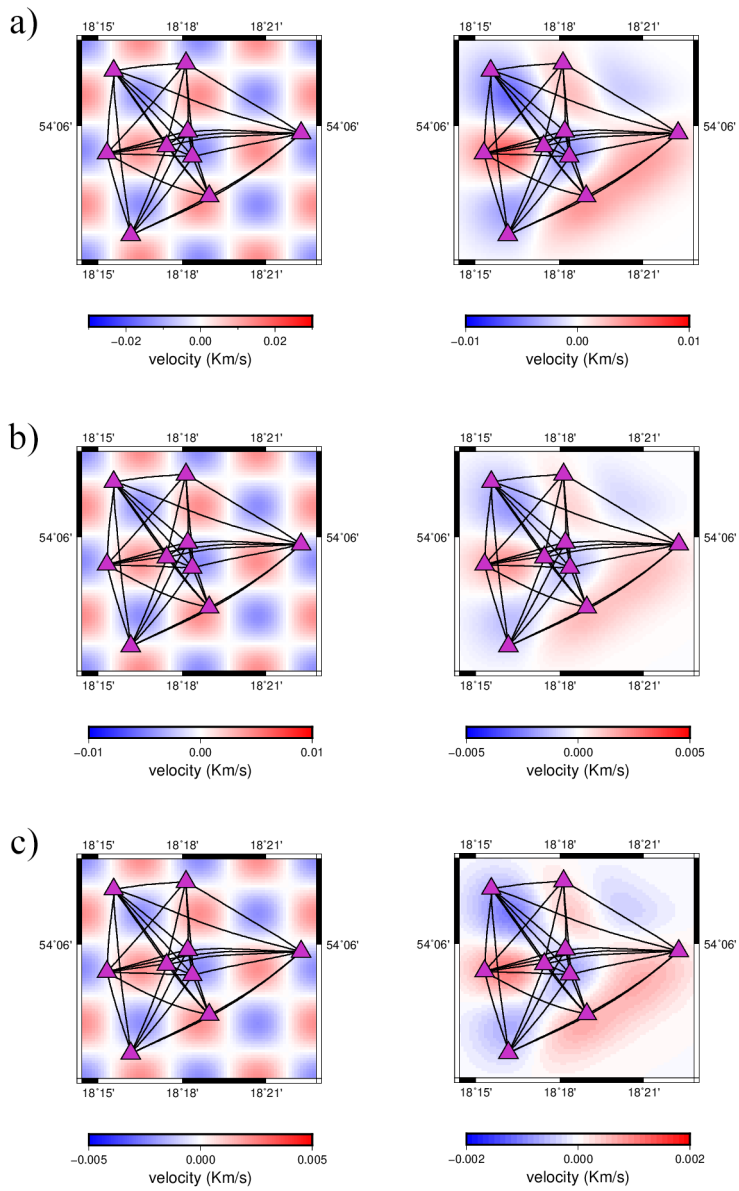


Figure 3.5.25: *(left panel) Checkerboard perturbation obtained with a 1.84 × 1.71 km checkerboard anomalies size and a) 0.03 b) 0.01 and c) 0.005 km/s maximum perturbation at the grid vertices compared with (right panel) checkerboard perturbation retrieved from the resolution test. In this case the color scales on the resulting maps were adapted in order to make the recovered anomaly amplitude more visible.*

3.6 Multiscale Ambient NOiSe TomogrAphy (MANgOSTA)

For performing ambient noise tomography from cross-correlations it was also used the innovative technique developed by Cabrera-Pérez et al. (2020), tested for a geothermal exploration study at Cumbre Vieja volcano (Canary Islands). The method is based on non-linear multiscale inversions which progressively upgrade the model parameterisation to reduce the non-linearity of the inverse problem. The methodology was implemented by the authors in the Python package MANgOSTA (Multiscale Ambient NOiSe TomogrAphy), divided into modules which perform all the steps required for ambient noise tomography.

The same dataset (1^{st} June – 31^{th} July 2016) used in the previous analysis was considered and pre-processed prior the cross-correlation procedure. The raw single-station data, decimated at 50 Hz, was filtered between 0.1 and 10 Hz, and the spectral whitening was applied. The daily vertical cross-correlation of signals from the different station pairs were computed and then stacked. Also in this case, starting from a wider frequency range, the same frequency sub-bands were investigated. The figure 3.6.1 shows the cross-correlation stacks ranked by increasing interstation distance in the wider frequency band 0.1-10 Hz, and the figure 3.6.2 shows the results obtained for the narrower bands 0.1-0.2, 0.2-0.5, 0.5-0.9, and 1.5-2.0 Hz.

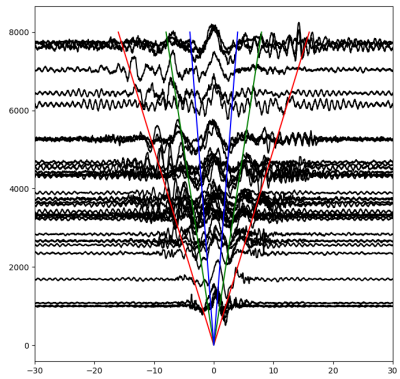


Figure 3.6.1: *The two-month vertical stacked functions filtered in the wider frequency band 0.1-10 Hz, sorted according to the interstation distance. The red, green and blue lines indicate the theoretical arrival times due to a propagation velocity of 0.5, 1, and 2 km/s, respectively.*

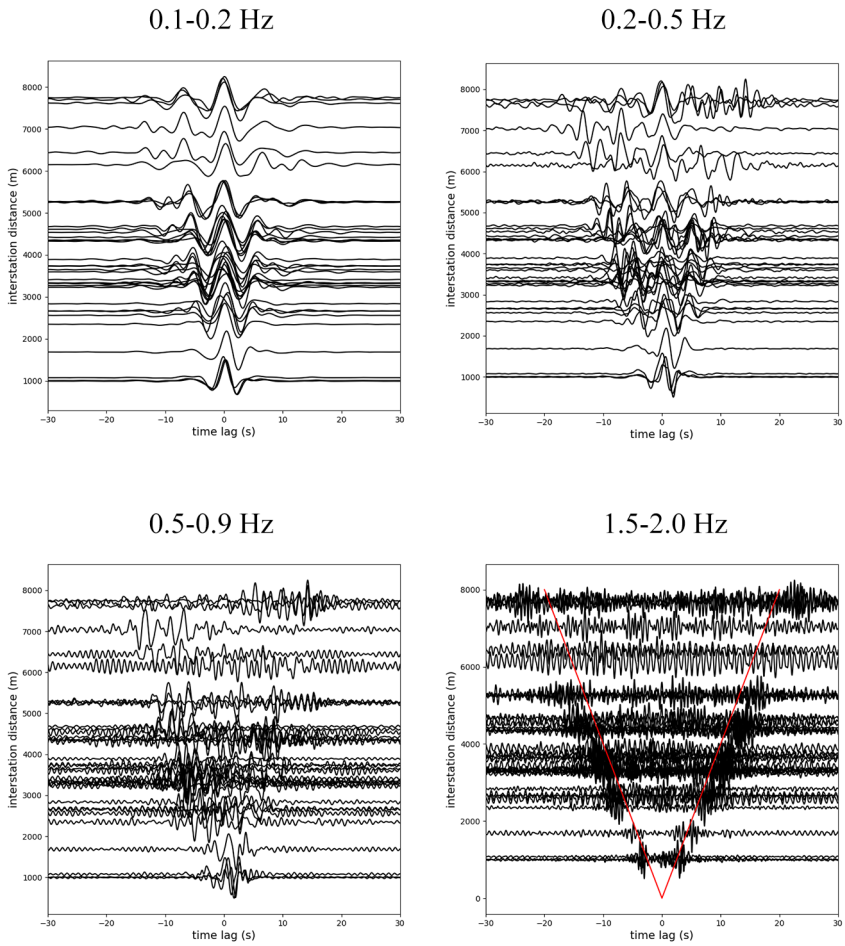


Figure 3.6.2: *The two-month vertical stacked traces (see top label for each panel) filtered in the 0.1-0.2, 0.2-0.5, 0.5-0.9 and 1.5-2.0 Hz ranges, sorted according to the interstation distance. The red line in the 1.5-2.0 Hz panel indicates the theoretical arrival times due to a propagation velocity of 400 m/s.*

Also in this case, refining the study in narrower frequency bands, a well-defined coherent wavetrains of dispersive Rayleigh waves emerges from vertical stacked functions filtered in the 1.5-2.0 Hz range. It is present in almost all couple of stations symmetrically in the causal and acausal parts, and a propagation velocity of about 400 m/s can be attributed. In the other sub-bands the signal is variable and there is no distinct separation between the causal and acausal parts. From the estimated cross-correlations functions, the group velocities as a function of period was measured by using the standard frequency-time analysis (FTAN; Levshin et al., 1972). The measurement of dispersion curves was carried out manually using the graphical

interface in MANgOSTA. The software allows to select cross-correlation signals individually and display them (Fig. 3.6.3 down).

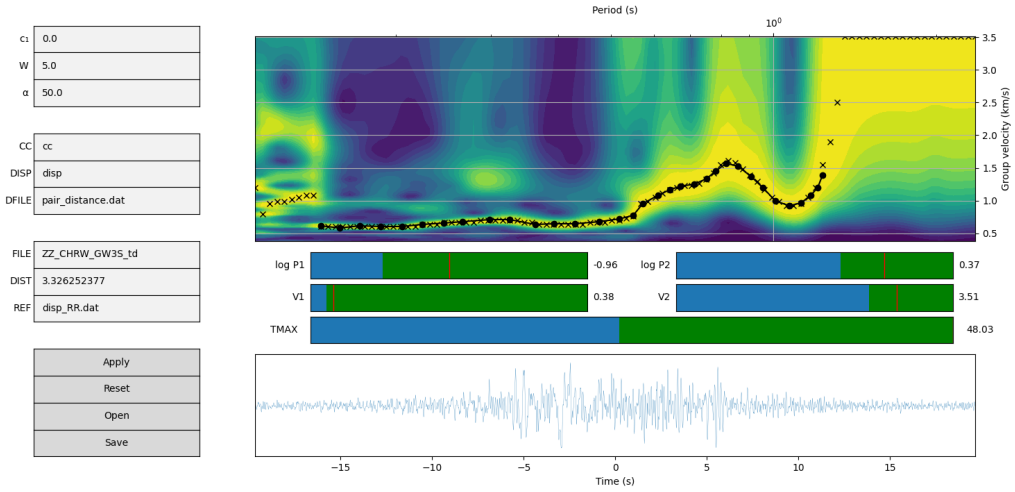


Figure 3.6.3: *Graphical interface for measuring dispersion curves. The correlogram (down) and the relative dispersion curve (top) are shown. The black dots represent the group velocities from the picking and the black line represents the best-fits curve for the data.*

Before applying the FTAN, the code sums the causal and anticausal part of the stacked function. The corresponding dispersion curve is displayed (Fig. 3.6.3 top) and it is possible to select the points of the curve that constitute the group velocity corresponding to a different period. It is possible to set parameters that improve the graphical display of data or facilitate the choice of group velocity measurement. For the FTAN diagram, the upper and lower limits along the x-axis (period) are adjusted via the log P1 and log P2 parameters, respectively. Similarly, along the y-axis (group velocity) the V1 and V2 parameters are varied. The α parameter, instead, controls the resolving power of the filter, and then the ability to measure the dispersion curves. A value 50 for α resulted in a good choice. The Rayleigh wave group velocity dispersion curves of all the 36 noise cross-correlation stacked function were estimated. They are depicted in the figure 3.6.4, showing the limits of the measured group velocities and the error estimation made during the data analysis. The retrieved Rayleigh wave group velocities range between 0.50 km/s and 2 km/s.

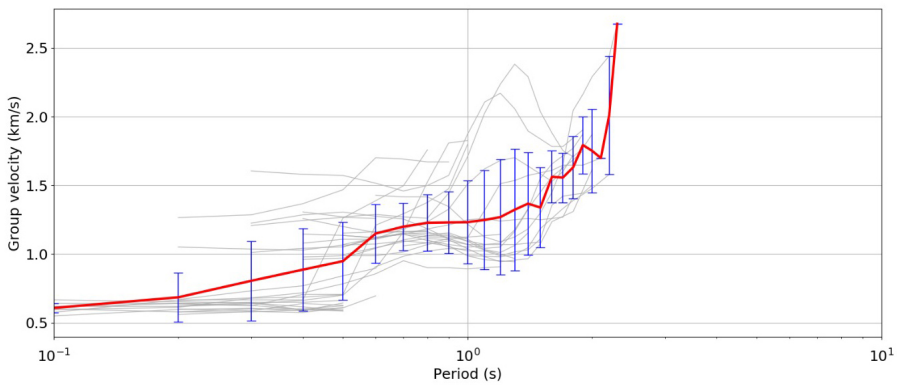


Figure 3.6.4: *Measured dispersion curves (gray curves) from FTAN analysis. The mean (red curve) and standard deviation at each period (blue vertical bars) are also shown.*

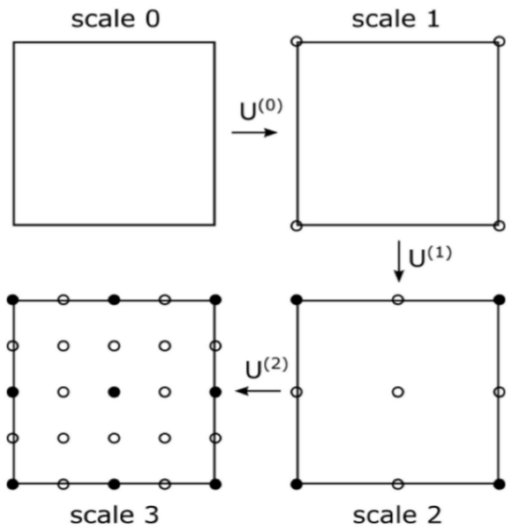


Figure 3.6.5: *Schematisation of the multiscale approach. New grid nodes (blank circles) are added by upgrading to the next scale. Filled circles represents the nodes from the previous scale (Cabrera-Pérez et al., 2020).*

The group velocity dispersion curves constitute the input data for tomographic inversion to obtain the 2-D group velocity maps. The tomographic images were obtained through a multiscale inversion approach. It is an iterative process that consists in analysing models with different parametrisation scales. Starting from a simple model, at each iteration an upgrade operator

U(s) switches from a model with scale s to a model with scale s + 1 changing the number of model parameters. MANgOSTA allows to implemented this operator in different ways, e.g. interpolation, Fourier series, or Discrete Wavelets.

In this work, a linear interpolation strategy that increases the number of nodes arranged on a regular grid (Fig. 3.6.5) was used.

The number of parameters for a model in the scale s results

$$N(s) = \begin{cases} 1 & if\ s = 0 \\ (2^{s-1} + 1)^2 & if\ s > 0 \end{cases} \tag{3.6.1}$$

At each iteration, the best-fit model at scale s can be determined using linear o non-linear inversion. Then, it is upgraded and used as a starting reference model for the new inversion at scale s+1, and so forth.

In this work, the inversion of data was performed using the multiscale inversion method with a Tikhonov regularization and $\vartheta = 0.5$. A linear inversion with an homogeneous model was performed at scale 0 using the damped least squares approach, and non-linear inversions was performed at superior scales using the Levemberg-Marquadt like approach (Aster et al. 2018). For the inversion, the group velocity for periods between 0.30 s and 1.20 s were considered selecting only periods having a minimum number of 10 dispersion curves.

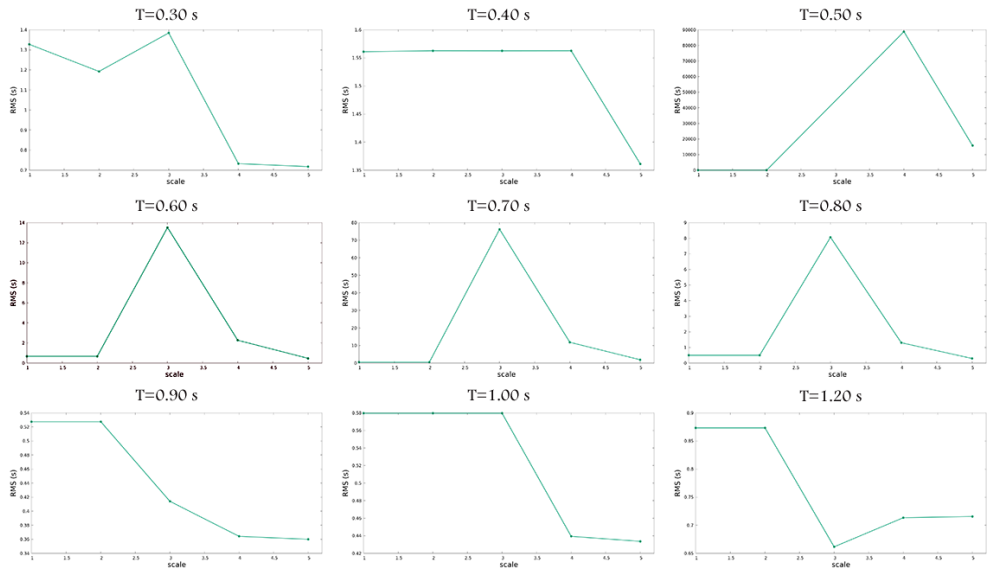


Figure 3.6.6: *RMS as a function of scale, for periods from T = 0.30 s to T = 1.20 s (see top label for each panel).*

The maximum scale considered in the inversion for each period was chosen corresponding to the minimum RMS value calculated from scale = 1 to scale = 5 (Fig. 3.6.6). The figure 3.6.7 shows the 2-D maps of group velocity for the Wysin site obtained by the tomographic inversion in the 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 ,and 1.2 s period, allowing the investigation of a maximum depth of about 150 m. The retrieved values range between 0.10 km/s and 2.5 km/s. For period 0.4 and 1.0 s, large spatial variations in velocity are observed with two high-velocity anomaly located in the south-west part of the investigated area and at the north-east in the proximity of wells. In addition, for some periods such as 0.50, 0.60, 0.70 and 0.90 s, because there are no spatial velocity variations a uniform model can be estimated for the whole investigated area. The reliability of the tomographic results can be inferred from the histograms in figure 3.6.8 that show the traveltime residuals of the solution models compared with model solutions at first non-inversion (scale = 1). For some periods the scale = 1 model coincides with the final model solution.

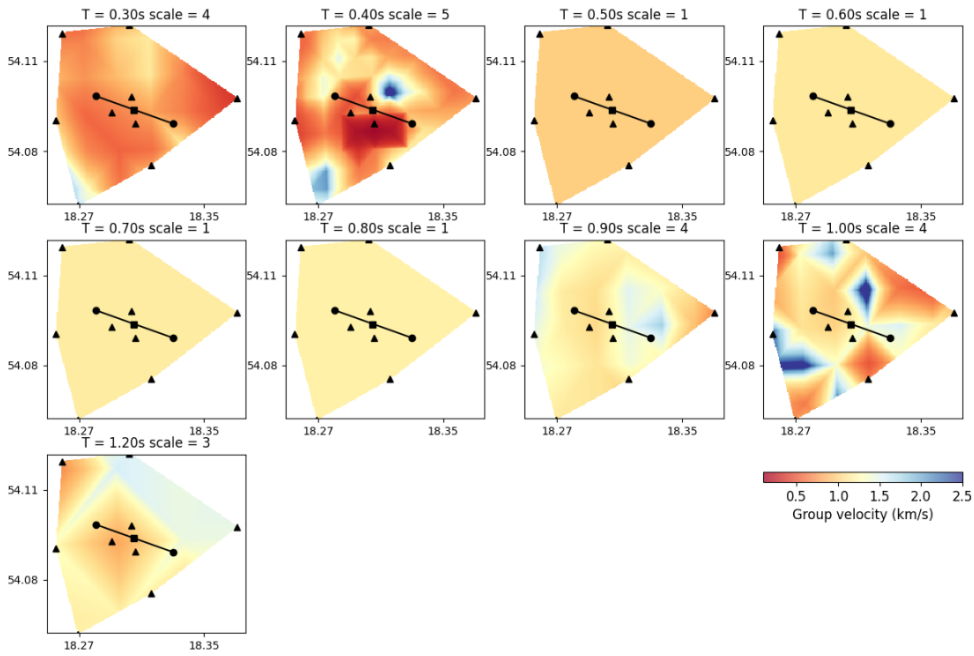


Figure 3.6.7: *Solution models resulting from tomographic inversions for different periods. The maximum scale chosen for each period is indicated on top of each panel. The stations and wells locations are also reported.*

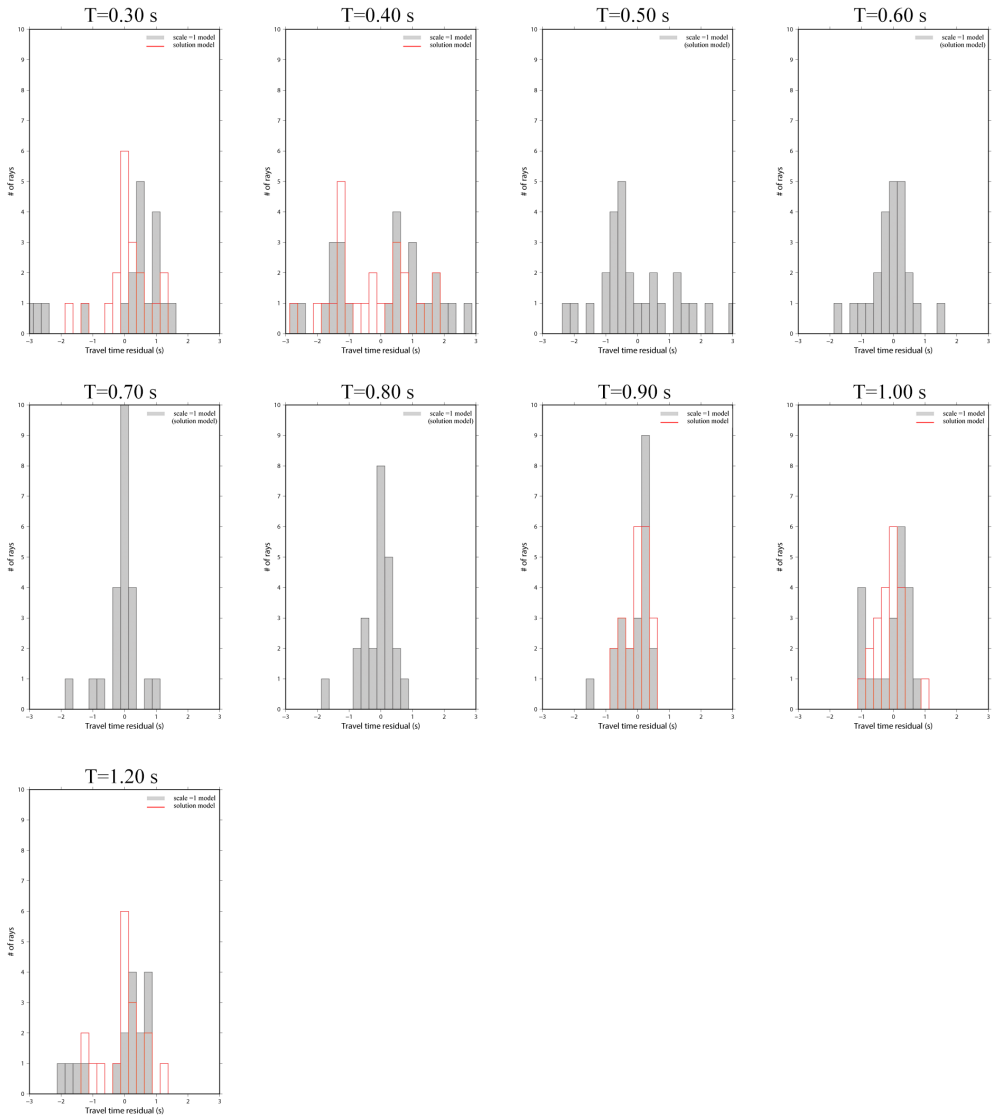


Figure 3.6.8: *Frequency histogram of the (grey) solution model using scale = 1 and (red) final data fit, for periods from $T = 0.30$ s to $T = 1.20$ s (see top label for each panel).*

3.7 Comparison of results and discussion

Daily cross-correlation stacks between 36 station pairs of the seismic network at the Wysin site were analysed using two algorithms that allow the traditional computation of cross-correlation. It was used a Matlab code where the computation is performed in the time domain whereas in Mangosta in the frequency domain. In both cross-correlation analysis the results show a well-defined coherent phase emerging from two-month vertical stacked functions filtered in the 1.5-2.0 Hz range. At these frequencies the emerging signal is present in almost all couple of stations symmetrically in the causal and acausal parts. The signal propagates with a velocity of about 400 m/s estimated in both analysis, consistent with expected near-surface shear-wave velocities. The obtained velocity value is retrived in analysis of López Comino et al. (2018).

The fixed-frequency tomography performed with FMST involved the arrival times of the weighted cross-correlation stacks in the 1.5-2.0 Hz frequency band, enabling the investigation of a maximum depth of about 70 m of the medium. The tomographic solution model shows a velocity range between 300 m/s and 380 m/s consistent with the expected values from the cross-correlation analysis. A decrease in velocity is present in the central part of the investigated area (Fig. 3.5.23), near the wells, therefore a possible correlation with the HF operations can be assessed. A high-velocity anomaly is also present and located in the south-west part of the map.

Whereas, in the case of tomographic analysis performed with MANgOSTA, nine periods between 0.3 and 1.20 s were examined. In this case, the depth of investigation reaches about 150 m and the retrieved velocities vary from 0.40 km/s to 2.5 km/s, except for the period T = 0.4 s where the minimum velocity is 0.1 km/s. The results reveal that the velocity does not always increase as the period increases, leading to the assumption that the propagation could be not normal dispersive. Comparison with the results of the first approach can be made with the images obtained for the 0.5 and 0.6 s period. The velocity values are higher in the second approch. However, in this analysis for the periods 0.5, 0.6, 0.7 and 0.8 s, the tomographic inversion does not retrive any improvement as the scale increases. The issue could be related to the low number of input data with the presence of some outliers that could have a relevant weight in the inversion process.

Chapter 4

A deep geothermal energy site: St. Gallen (Switzerland)

In Europe, the use of energy from deep geothermal reservoirs is increasing because is a valid alternative energy source. It allows to reduce dependence on imported fossil fuels contributing to the European Green Deal objectives on the use of energy from sustainable sources. The temperature and permeability conditions required to generate electricity are reached at depths of at least 3 km in Europe, and usually the produced energy is used by neighbouring cities to limit energy loss through long distance transport.

The deep geothermal energy project in St. Gallen was designed to produce energy for the local community by exploiting hot water from a deep natural aquifer. The project followed in the footsteps of several deep geothermal projects in Southern Germany successfully developed in similar geological structures. Due to complications that arose during the project, St. Gallen represents an important study field for improving knowledge about this type of activity, which can only be successfully developed and operated if the risk of potential induced seismic events can be controlled.

In the first section of this chapter the geothermal energy project is described with also a brief introduction of the geological setting of the study area. Previous studies of the St. Gallen site concerning a seismic and aseismic study are illustrated next, highlighting the complexity of investigating the induced seismicity and related seismic hazard. Afterwards, the application of the seismic interferometry method and a tomographic study with the FMST method are reported.

The study of the phenomena of induced seismicity in these particular areas can also be improved by enhancing the ability to detect weak events. Since the recordings from the St. Gallen site are characterised by a low signal-

to-noise ratio, and due to the availability of a event dataset (although reduced), an application of automatic event detection based on neural network was performed. In a dedicated section of the chapter, the pre-processing and clustering analysis using an unsupervised network is reported. In addition, the results were also evaluated in terms of magnitude, events location, hypocentral distance, and origin time.

4.1 Geological setting and geothermal energy project

The study area is located in the eastern part of the North Alpine Foreland Basin which is the Swiss part of the Molasse basin, north of the Alps (Fig. 4.1.1). The Molasse basin was formed during the Oligocene and Miocene epochs due to the convergent movement of the European and Apulian plates. The basin formed as a result of the flexure of the European plate under the weight of the orogenic wedge of the Alps that was forming to the south. The Molasse basin is stratified into formations that are distinguished by the presence of continental or marine sediments deposited during the succession of epochs (Upper Freshwater Molasse; Upper Marine Molasse; Lower Freshwater Molasse). Under this stratification there are Mesozioic sediments from the period before the development of the Molasse basin when the region was covered by a shallow sea (Fig. 4.1.2).

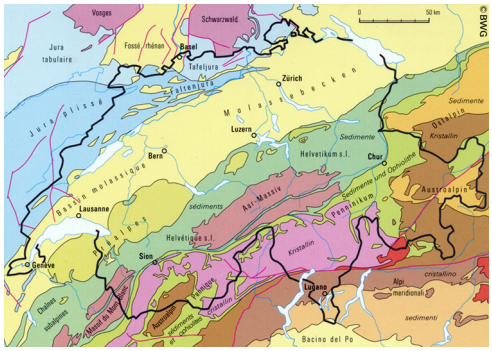


Figure 4.1.1: *Geological map of Switzerland. Source: swisstopo (https://www.swisstopo.admin.ch/).*

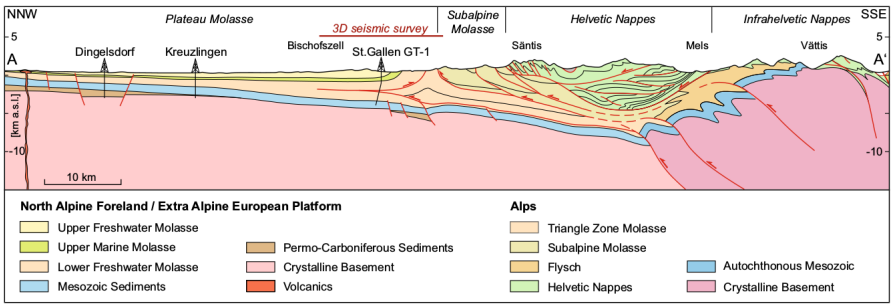


Figure 4.1.2: *Tectonic cross-section across northeastern Switzerland (Heuberger et al., 2016).*

The deep geothermal energy project in St. Gallen aimed to produce power using water in aquifers hosted in the Mesozoic strata at a depth of over 4 km (Fig. 4.1.2). The project was inspired by the success of other deep geothermal projects in southern Germany characterised by the same geological structures. Furthermore, it was strongly promoted by the citizens of St. Gallen through a public poll in supporting of green energies.

A large-scale 3-D seismic survey was conducted in 2010 (Heuberger et al., 2016) to investigate the target area prior to the drilling of the exploration well to set the optimal implementation. The survey identified a fault zone in the Mesozoic sediments and the pre-Mesozoic basement with a major system of NNE-SSW striking normal faults forming the St. Gallen Fault Zone (SFZ). This area was selected as a suitable location for drilling activity because its orientation promises naturally enhanced permeability. Furthermore, based on past seismic activity, the operators considered the area to be of low seismic risk.

The project was realised by the St. Gallen public utilities company (Sankt Galler Stadtwerke) and a dedicated seismic monitoring network was installed and managed by the Swiss Seismological Service (SED).

The project started the drilling phase in March 2013 and the well GT-1 was completed to a depth of 4450 m. On 14 July 2013 a first hydraulic stimulation test was performed involving a total of 175 m^3 of cold water. On 17 July, two acid stimulations were performed, each injecting $145m^3$ of diluted hydrochloric acid into the reservoir. A low-level of microseismicity strongly correlated with these activities was observed in the vicinity of the injection point. On 19 July, during the production test setup, gas (90% methane) from an unknown source at depth unexpectedly entered the well. The borehole was closed immediately but the wellhead pressure gradually increased with a possible risk to damage the well integrity. Therefore, the

operators decided to deal with the gas kick by pumping drilling mud into the well to push the gas back into the formation reducing the wellhead pressure. This well control procedure was accompanied by a new sequence of induced earthquakes that culminated in the largest M_L 3.5 earthquake at 2:40 A.M. local time on 20 July, which was felt within 10–15 km from the epicenter (Diehl et al., 2017).

Then, the project was temporarily stopped and resumed in October 2013 with a production tests extracting gas and water from the subsoil. The St. Gallen public utilities company decided to halt the geothermal project in May 2014 due to a "combination of insufficient water availability, increased earthquake risk and unexpected gas flow in the layers of rock accessed".

4.2 Previous works on the site

Several studies have investigated both seismic and aseismic response of the St. Gallen site to activities related to the geothermal project.

Edwards et al. (2015) focused on crucial aspects of the seismic hazard associated to the project, such as the robust and consistent computation of earthquake magnitude and ground motion and macroseismic intensity. They defined a specific magnitude scale M_L^{corr} for events occurring in the St. Gallen area to overcome the overestimation of the magnitudes of the smaller events ($M_L < 1.0$). Indeed, the smaller events was recorded only by the local network with station spacing on the order of a few kilometres. Relying on the almost constant source-station distance for events, the M_L^{corr} was defined by adding station specific correction factors that account for the bias of the attenuation correction in the near field (R $<$ 20 km) and potential site amplification effects. The M_L^{corr} magnitude was calibrated by using a set of reference events large enough to be recorded and assigned M_L using only broad-band stations at R $>$ 20 km of the Swiss Digital Seismic Network (SDSNet). The authors analysed the differences in ground-motion amplitudes and macroseismic observations between the $M_L = 3.5$ ($M_w = 3.4$) St. Gallen main shock and $M_L = 3.4$ ($M_w = 3.2$) Basel main shock occurred in 2006 at the respective EGS systems in Switzerland. In the long period motions ($\geq$ 1 s) it is observed, as expected, that the higher M_w St. Gallen event produced higher ground-motion, whereas for shorter period the ground motions of the two events are comparable. In this case, the dependence of ground-motion not only on fault area and slip (i.e. seismic moment), but also radiated energy. Indeed for the Basel event the source corner frequency (related to the source rupture velocity and stress-drop) is higher. In addition, a difference in macroseismic intensity was also found

analysing the effects on objects depending on the signals frequency content. Effects requiring high-frequency excitation is clearly predominant in Basel which thus is the event more strongly perceived.

Obermann et al. (2015) demonstrate for the St. Gallen case that using information from ambient seismic noise provides new possibilities to monitor well changes being able to observe the aseismic response of the reservoir to operator activities. They analysed the temporal evolution of waveform coherence and velocity in the coda of ambient seismic noise cross-correlations using 17 months data. Any important change of apparent velocity ε was revealed, whereas a significant loss of waveform coherence parallel to the injection tests was observed (Fig. 4.2.1).

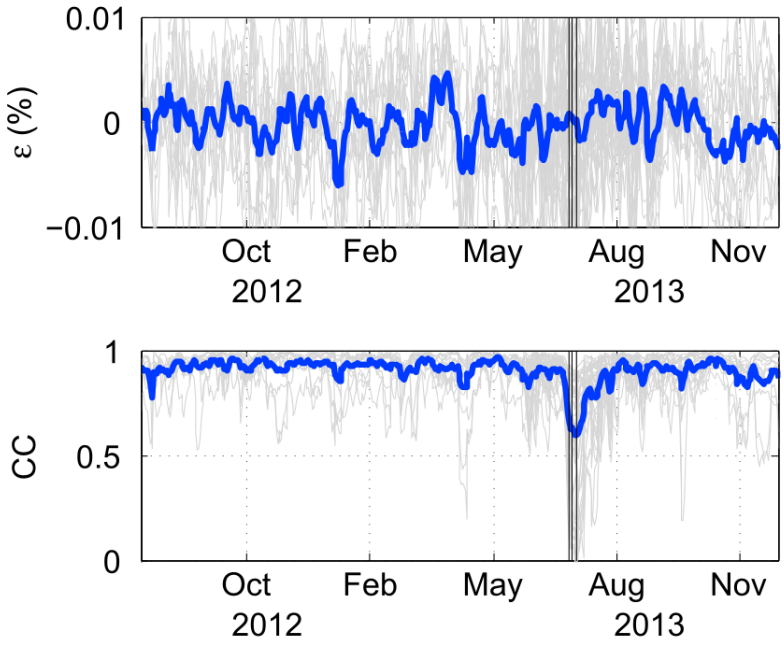


Figure 4.2.1: *(Top panel) Apparent velocity changes ϵ and (down panel) waveform coherence CC with their averages over all station pairs (blue lines). The injection tests and M_L 3.5 earthquake are indicated with the vertical lines (Obermann et al., 2015).*

The authors interpreted the loss of coherence as due to a perturbation in the subsurface and estimated its spatial distribution horizontally with an inversion procedure and vertically through a spectral analysis. The changes are spatially constrained within 3–6 km of the injection location suggesting that they are related to the well operations (Fig. 4.2.2).

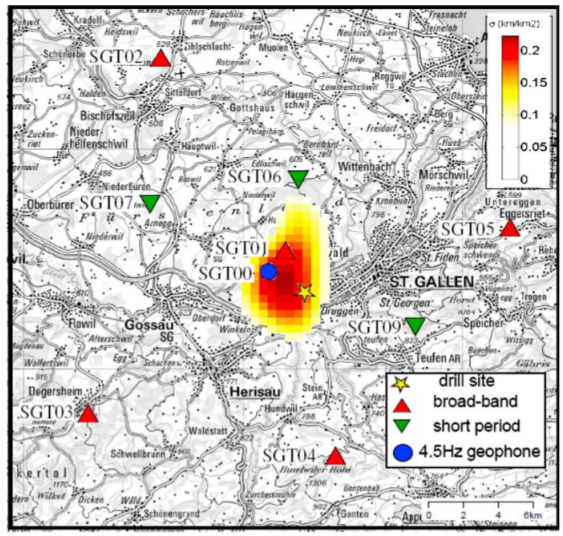


Figure 4.2.2: *Map of the scattering cross-section density σ associated with the strength and size of the medium change, averaged over July 2013. (Obermann et al., 2015).*

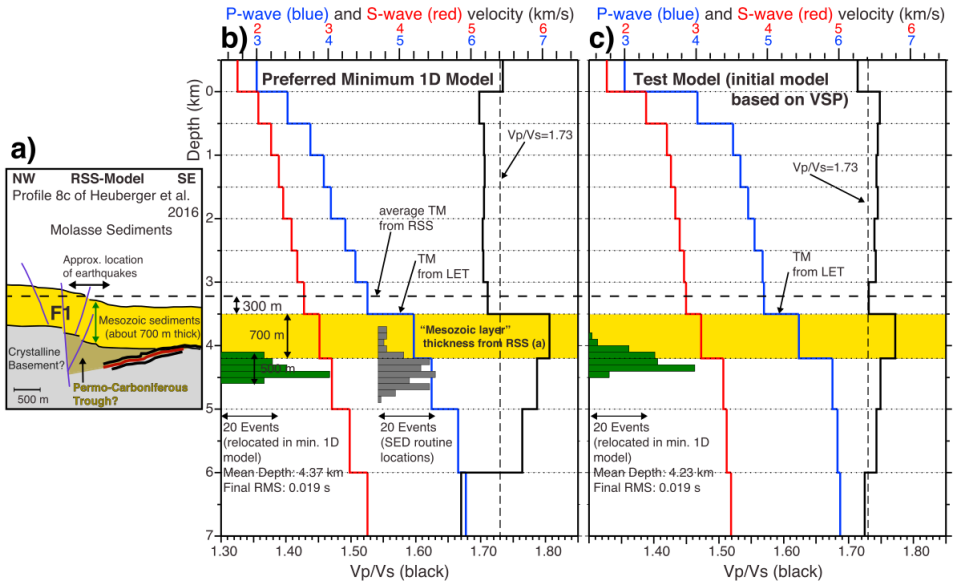


Figure 4.2.3: *a) Structural model of the source region from the reflection seismic survey of Heuberger et al. (2016). The yellow band indicates the interpreted Mesozoic sediments of 700 m thickness. b) Preferred minimum 1-D VP and VS models derived from simultaneous inversion of P and S wave traveltimes of earthquakes. c) Alternative 1-D model to assess effects of higher initial VP and VS values in the Molasse sediments according to the vertical seismic profiling (VSP) in Heuberger et al. (2016). (Diehl et al., 2017).*

A study of the St. Gallen induced sequence was carried out by Diehl et al. (2017) providing a high-precision earthquake catalog consisting of 347 earthquakes with M_L^{corr} ranging from -1.2 to 3.5. The events relocation was performed by solving a coupled hypocenter-velocity inversion and the absolute location was constrained combining results of a 1-D seismic velocity structure obtained by inversion of earthquake data (Fig. 4.2.3 b,c) with information from the 3-D reflection seismic survey of Heuberger et al., (2016). The majority of the induced seismicity is observed in the pre-Mesozoic basement (Fig. 4.2.3 a).

A double-difference techniques also were applied to further refine the hypocenter locations and image the fine-scale structures of the reactivated fault (Fig. 4.2.4).

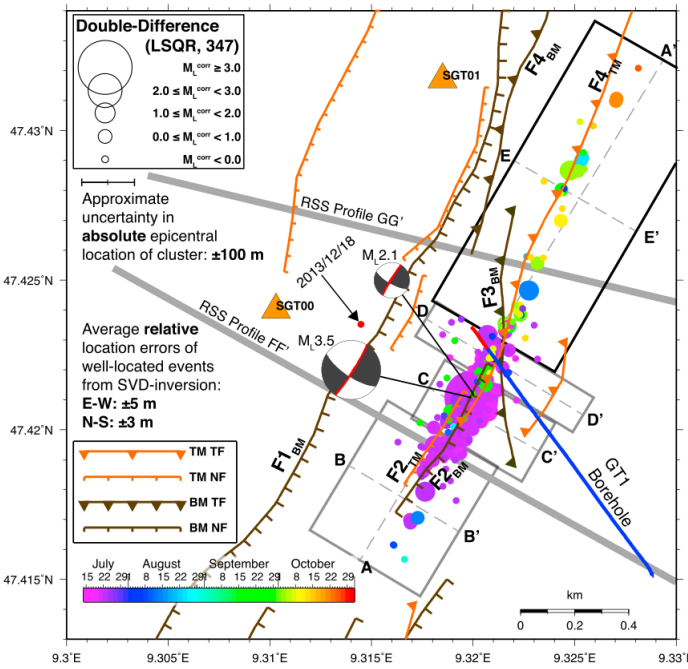


Figure 4.2.4: *The map of St. Gallen sequence resulting from the double-difference method. The size of the circles is proportional to magnitude (M_L^{corr}) and colour indicate the origin time of events. Beach ball symbols indicate the focal mechanisms of the two largest events of the sequence. The surface projection of the borehole GT-1 (blue bold line) and the faults from Heuberger et al. (2016) are shown. TM: top-Mesozoic; BM: bottom-Mesozoic; NF: normal Fault; TF: thrust Fault. (Diehl et al., 2017).*

Diehl et al. (2017) investigated a possible correlation beetween the spatiotemporal evolution of the sequence and borehole operations. They propose the existence of a hydraulic connection between the reactivated fault

and the borehole, likely through faults mapped by seismic data. Despite the excellent quality of the seismic data, this association remains ambiguous. The crustal fault zone can have high complexity in terms of structure, hydraulic condition, and state of stress. These complexities and incomplete knowledge complicate the study of induced seismicity caused by fluid injection, consequently also the assessment and mitigation of the associated seismic hazard.

4.3 Seismic network and data

The seismic monitoring of the St. Gallen geothermal project was managed by the Swiss Seismic Service (SED) within the GeoBest research project (http://www.seismo.ethz.ch/en/ research-and-teaching/ongoing-projects), funded by the Swiss Federal office of Energy (SFOE). This monitoring aimed to detect microseismicity near deep boreholes and study whether its cause was natural or related by the geothermal power project.

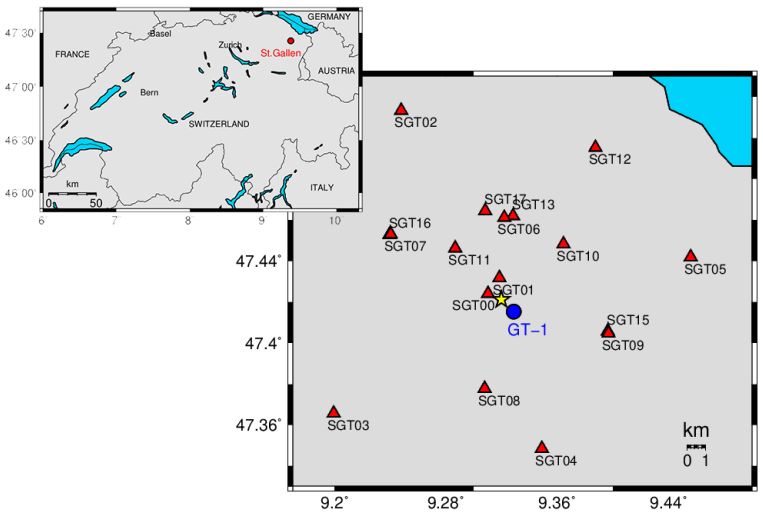


Figure 4.3.1: *Station array for monitoring seismicity of the St. Gallen geothermal project. The drilling site GT-1 (blue dot) and 3.5 Ml earthquake (yellow star) are shown.*

Station code	Sensor type	Sample rate (Hz)	Installation date	Coordinates
SGT00	IESE-S21g-4.5 Hz	200	13th March 2012	47.4240 N 9.3103 E
SGT01	Nanometrix Trillium compact	200	10th February 2012	47.4317 N 9.3185 E
SGT02	Nanometrix Trillium compact	200	23th January 2012	47.5133 N 9.2479 E
SGT03	Nanometrix Trillium compact	200	23th January 2012	47.3656 N 9.1994 E
SGT04	Nanometrix Trillium compact	200	10th February 2012	47.3484 N 9.3490 E
SGT05	Nanometrix Trillium compact	200	2nd February 2012	47.4419 N 9.4560 E
SGT06	Lennartz LE3D-1Hz	200	4th June 2013	47.4611 N 9.3220 E
SGT07	Lennartz LE3D-1Hz	200	4th June 2013	47.4530 N 9.2404 E
SGT08	Lennartz LE3D-1Hz	200	4th June 2013	47.3777 N 9.3079 E
SGT09	Lennartz LE3D-1Hz	200	4th June 2013	47.4046 N 9.3969 E
SGT10	Lennartz LE3D-1Hz	200	25th July 2013	47.4482 N 9.3645 E
SGT11	Lennartz LE3D-1Hz	200	25th July 2013	47.4462 N 9.2867 E
SGT12	Lennartz LE3D-1Hz	200	25th July 2013	47.4952 N 9.3875 E
SGT13	Lennartz LE3D-1Hz	200	9th October 2013	47.4619 N 9.3283 E
SGT15	Lennartz LE3D-1Hz	200	9th October 2013	47.4057 N 9.3964 E
SGT16	Lennartz LE3D-1Hz	200	7th November 2013	47.4528 N 9.2395 E
SGT17	Lennartz LE3D-1Hz	200	7th November 2013	47.4644 N 9.3083 E

Table 4.1: *Main characteristics of the seismic stations in St. Gallen.*

Before starting stimulation tests the SED, in cooperation with the St. Gallen public utilities company (Sankt Galler Stadtwerke), densified the monitoring network to improve the sensitivity and precision of the location of natural and induced seismicity in the target area. In February 2012 one borehole short-period station at 205 m depth (SGT00) and five surface broadband stations (SGT01–SGT05) were installed within 12 km from the drilling site GT-1 (Fig. 4.3.1). In July 2013 four additional surface short-period stations

(SGT06–SGT9) were added at a distance of about 5 km around the geothermal well. After the induced 3.5 Ml earthquake on July 20, the network was further densified by three surface short-period stations (SGT10–SG12). In the following months some of the stations were slightly relocated and renamed (SGT06→SGT13→ SGT17; SGT09→SGT15; SGT07→SGT16) because of hardware upgrades. In practice the network occupied 11 different sites, with a maximum of seven stations running at any one time (Edwards et al. 2015). The main characteristics of the stations are reported in the table 4.1.

The continuous waveforms are provided by the geomecon GmbH company and recorded in the framework of GeoBest project, while realizing well control measures after drilling and acidizing the GT-1 well. The available data range from 1^{st} July to 31^{th} December 2013, but not all stations were operating at the same time as can be seen in the figure 4.3.2.

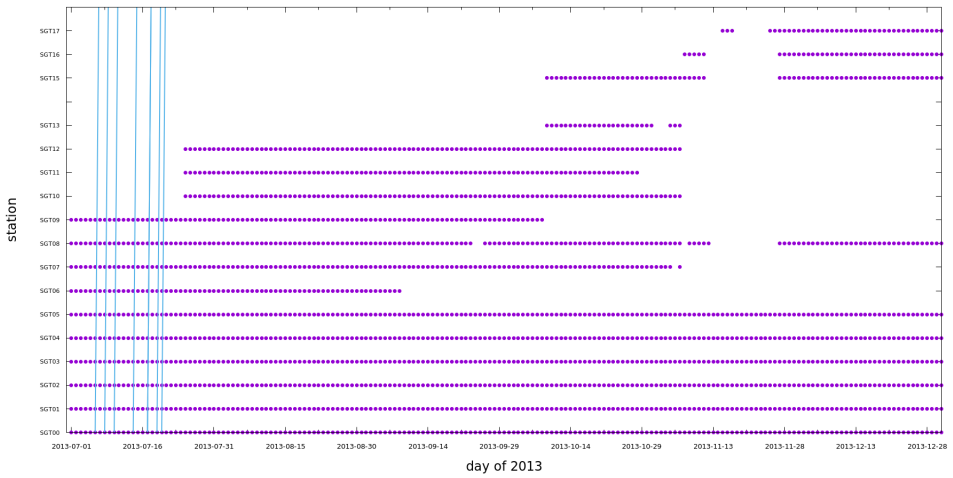


Figure 4.3.2: *Available continuous seismic data for the St. Gallen site recorded in 2013. The vertical blue lines mark the date of the main phases of the project and of the seismic event of Ml 3.5.*

The analysed dataset was chosen according to both the availability of the recordings and the period of well stimulation and testing followed by complications such as gas kick and the occurrence of a Ml 3.5 seismic event. The stations that operated during the analysed period are displayed with their connections in figure 4.3.3.

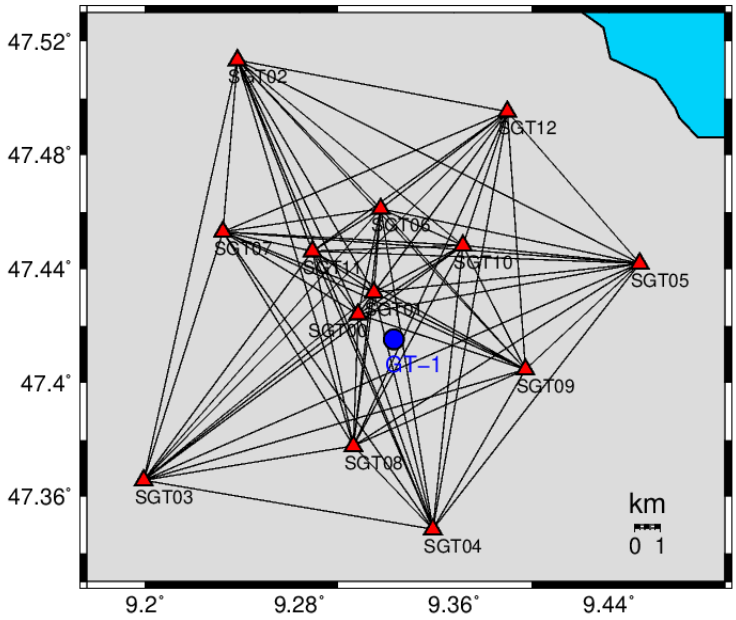


Figure 4.3.3: *Raypaths between the stations used in this study.*

An high-precision earthquake catalog of the induced St. Gallen sequence and the corrisponding waveforms are provided by Diehl et al. (2017). The sequence between 14^{th} July and the end of October 2013 consists of 347 locatable earthquakes with M_L^{corr} defined in Edwards et al. (2015) ranging from -1.2 to 3.5.

4.4 Fast Marching Surface-wave Tomography (FMST)

Computation of cross-correlation functions

The procedure for extracting Green's functions from cross-correlation analysis in the case study of the St. Gallen site follows the same approach implemented in the same Matlab code as described for the Wysin site. The hypothesis of isotropy distribution of noise sources with respect to the considered seismic network, for applying the methodology, was verified by a beamforming analysis. An example of the beamforming results obtained for the day 2013/08/15 is reported in figure 4.4.1 for 0.5-0.9 Hz, range of interest for this analysis.

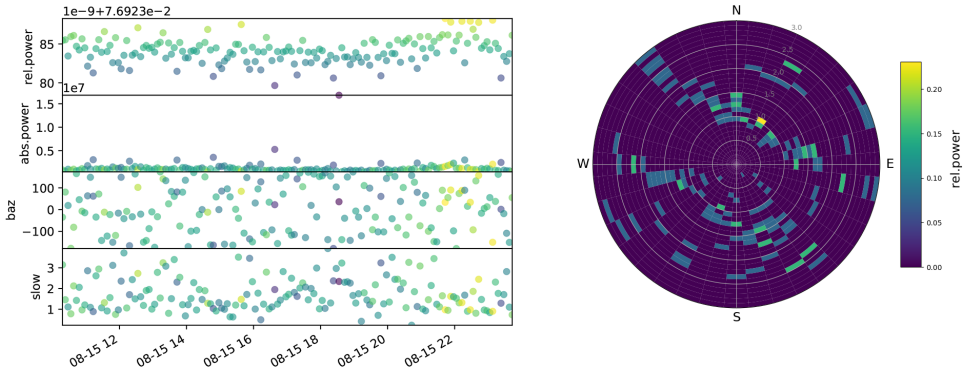


Figure 4.4.1: *Beamforming results obtained using the data acquired on 2013/08/15 in the frequency band 0.5-0.9 Hz.*

In the pre-processing step, the single-station data were transformed into day-long traces in SAC format, and then decimated at 50 Hz. The mean and the trend were removed, a band pass filter between 0.1 and 10 Hz and the one-bit normalization were applied. The daily stacked cross-correlation functions between the 78 station pairs were computed for three months of data (1^{st} July – 30^{th} September 2013) that include the period of well stimulation and the occurence of the gas kick and the Ml 3.5 seismic event. Also in this analysis, the 205 m depth of the borehole station SGT00 was considered negligible for calculating the interstation distances. The presence of a coherent signal was investigated for different frequency bands for four components (ZZ, RT, ZR, ZT) of the cross-correlation tensor. A coherent phase, visible in almost all couple of stations at both causal and acausal part, is retrieved in the band 0.5-0.9 Hz for all four components and travels with an estimated propagation velocity of about 1750 m/s. The figure 4.4.2 shows the 78 cross-correlation stacks organized for increasing interstation distance in this frequency band obtained for the ZZ, RT, ZR, ZT components.

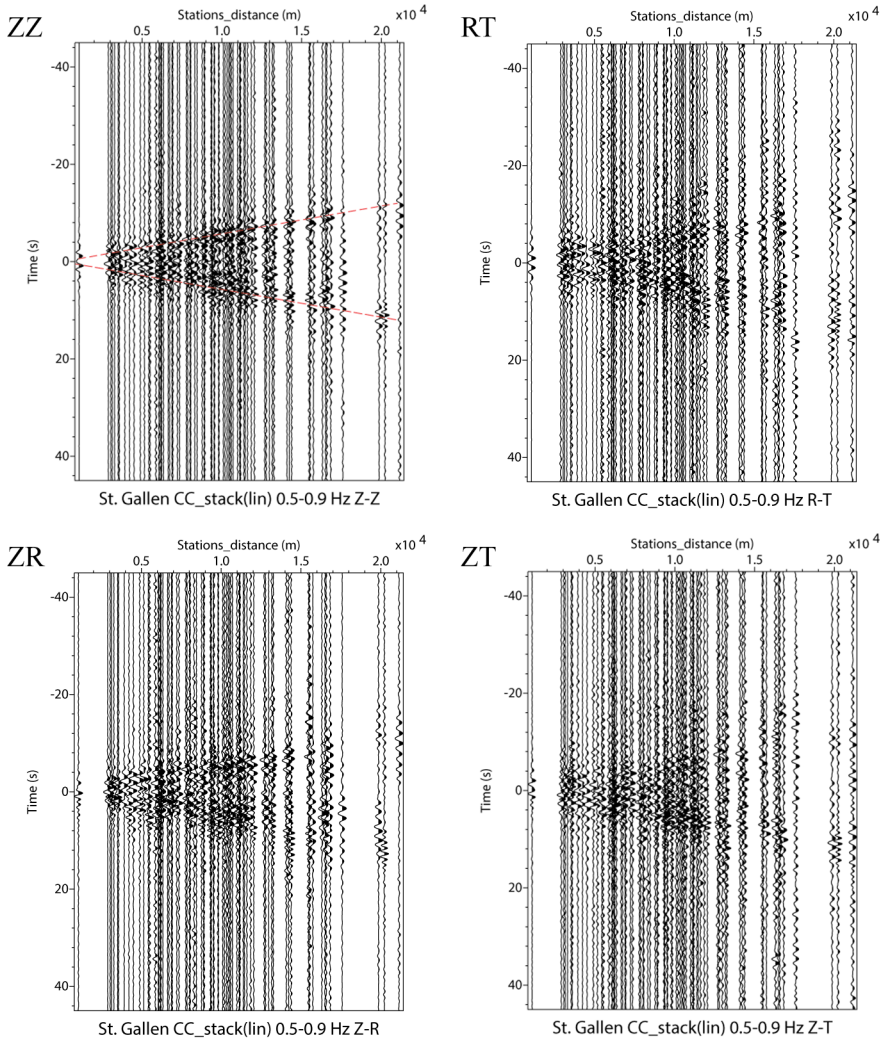


Figure 4.4.2: *The three-month cross-correlation stacks as a function of the interstation distance for the ZZ, RT, ZR, and ZT components in the 0.5-0.9 Hz range. The dashed red lines in ZZ graph indicate the theoretical arrival times of a surface wave propagating at a velocity of 1750 m/s.*

In the band of interest 0.5-0.9 Hz and for the vertical component, the presence of a coherent signal and its stability with time was analysed for all the 78 station pairs. The figure 4.4.3 shows the results for some examples of station pairs.

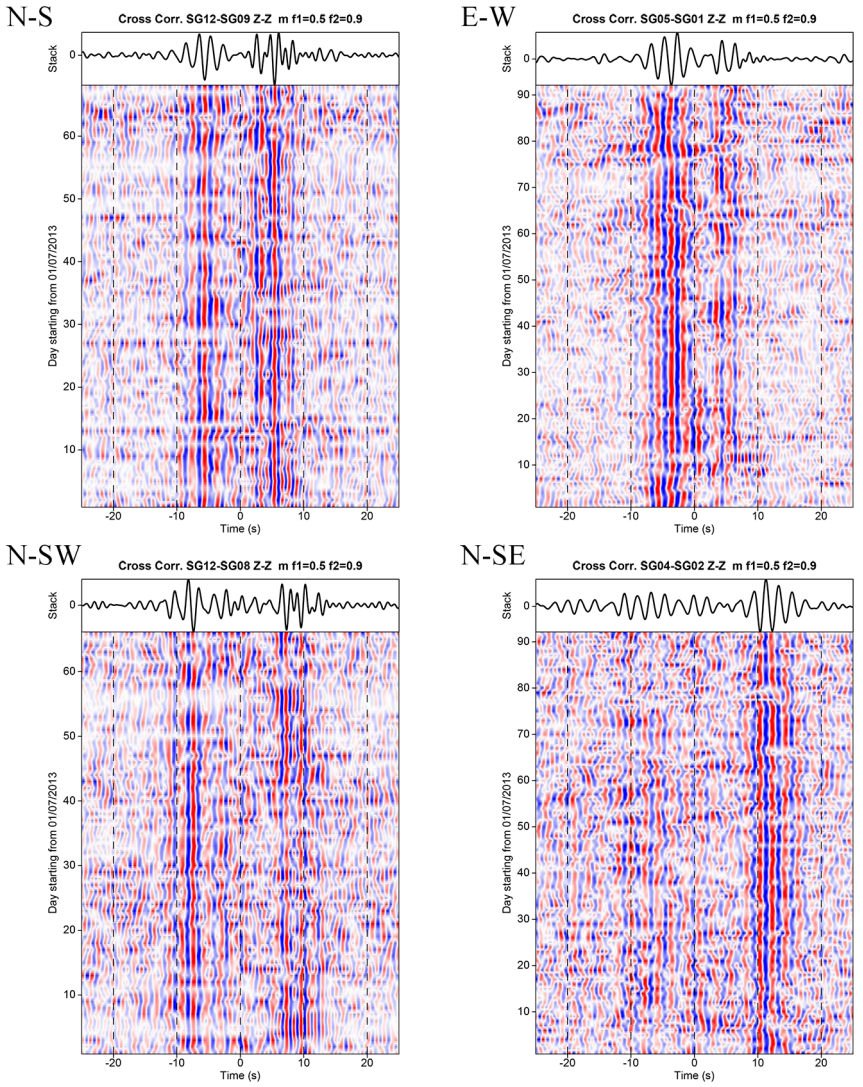


Figure 4.4.3: *Daily vertical cross-correlations as a function of the date of data acquisition filtered in the 0.5-0.9 Hz range. Each panel refers to four examples of station pairs with different orientation (see top label for each panel). The stack of the all cross-correlation functions is also depicted on the top of panels.*

Also in this case, there are station pairs for which the cross-correlation functions have different patterns in the causal and acausal part or a signal is centred around zero, depending on the orientation and location of the stations. The energetic signal retrived for most station pairs is reasonably stable over time with a loss of coherence observed before, during, and after well stimulation tests.

Tomographic inversion

Tomographic images of the investigated area of the St. Gallen site was reconstructed using the FMST algorithm, following the same procedure as described in the case of the Wysin site. The tomographic inversion was performed using the surface wave arrival times of the stacked cross-correlation functions derived from the previous analysis. Inversions was carried out for the frequency sub-bands 0.5-0.7, 0.6-0.8, 0.7-0.9, 0.8-1.0, and 0.9-1.1 Hz within the range in which the most energetic signal emerged. Thus, in this case the maximum investigated depth of the medium is about 1.2 km.

For each narrow frequency band, the arrival times were obtained by picking manually the maximum envelope of the weighted ($\nu = 2$) cross-correlation stacks of all the 78 station pairs, ordered by the offset in the time domain. The error assigned to the picked arrival times is 0.4 s, considering the amplitude of the maximum of the envelope function in the time domain. Not all 78 measures were considered good for inversion. For each of them, the propagation velocity between the corresponding station pair was calculated and considered only velocities close to that found in the cross-correlation analysis, in particular in the range 1750 m/s +- 500 m/s.

For each analysed range, initial uniform velocity model for the inversion was estimated by averaging the values of the propagation velocities estimated between station pairs. The 2-D inversion grid size was selected according to the minimum RMS value as a function of inversion iteration number. Finally, the damping and smoothing parameters were selected by constructing the relative trade-off curves.

The figures 4.4.4, 4.4.5, 4.4.6, 4.4.7, and 4.4.8 show the parameters setting for the tomographic inversion and results obtained with the corresponding histograms of residuals for the frequency bands 0.5-0.7, 0.6-0.8, 0.7-0.9, 0.8-1.0, and 0.9-1.1 Hz, respectively.

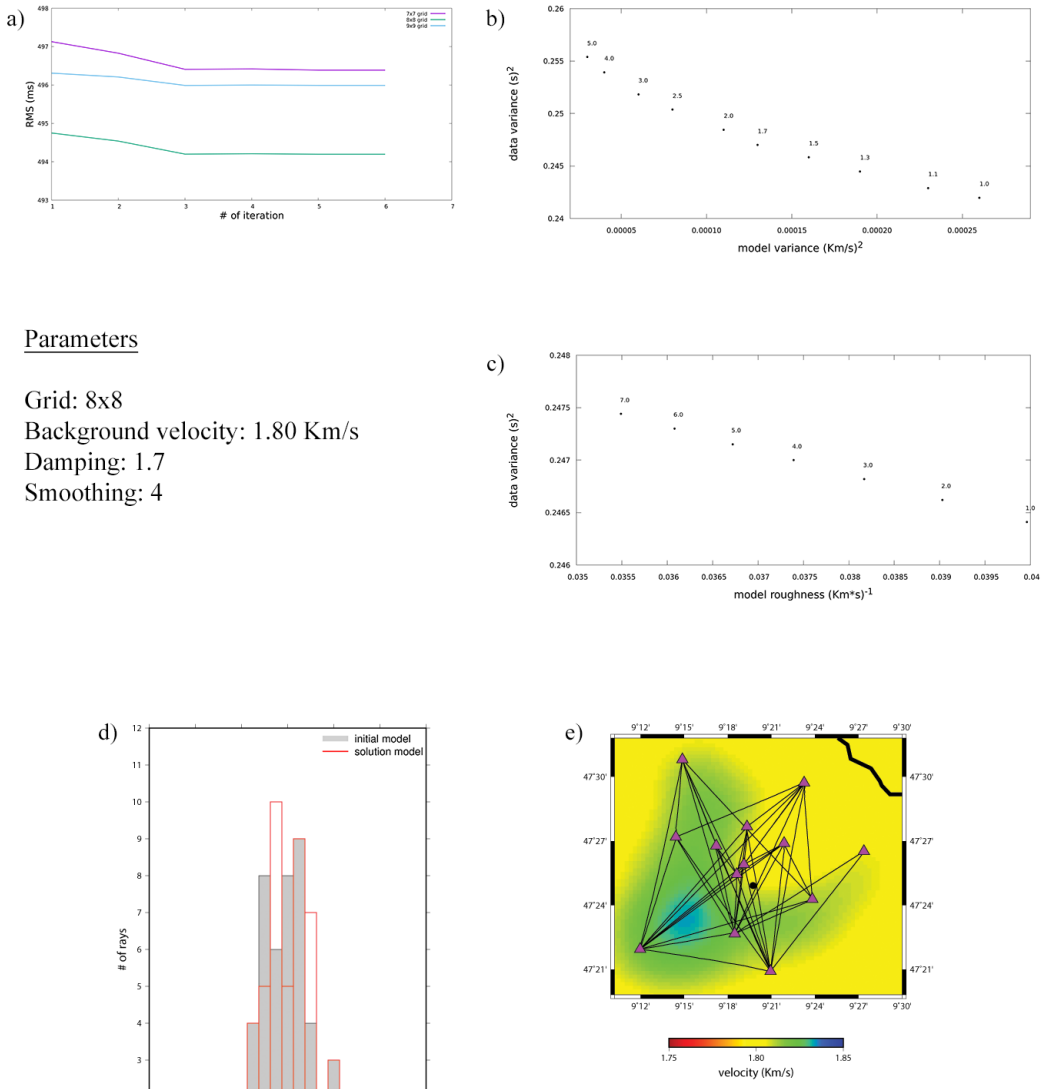


Figure 4.4.4: *Tomographic inversion strategy and solution model in the frequency band 0.5-0.7 Hz. a) RMS as a function of the iteration number of the tomographic inversion performed for different grid sizes. b) Damping and c) Smoothing factor trade-off curves. d) Frequency histogram of the (grey) initial and (red) solution model. e) Solution of the tomographic model with raypaths and stations used for the analysis in the specific range. The GT-1 well is represented with a black dot.*

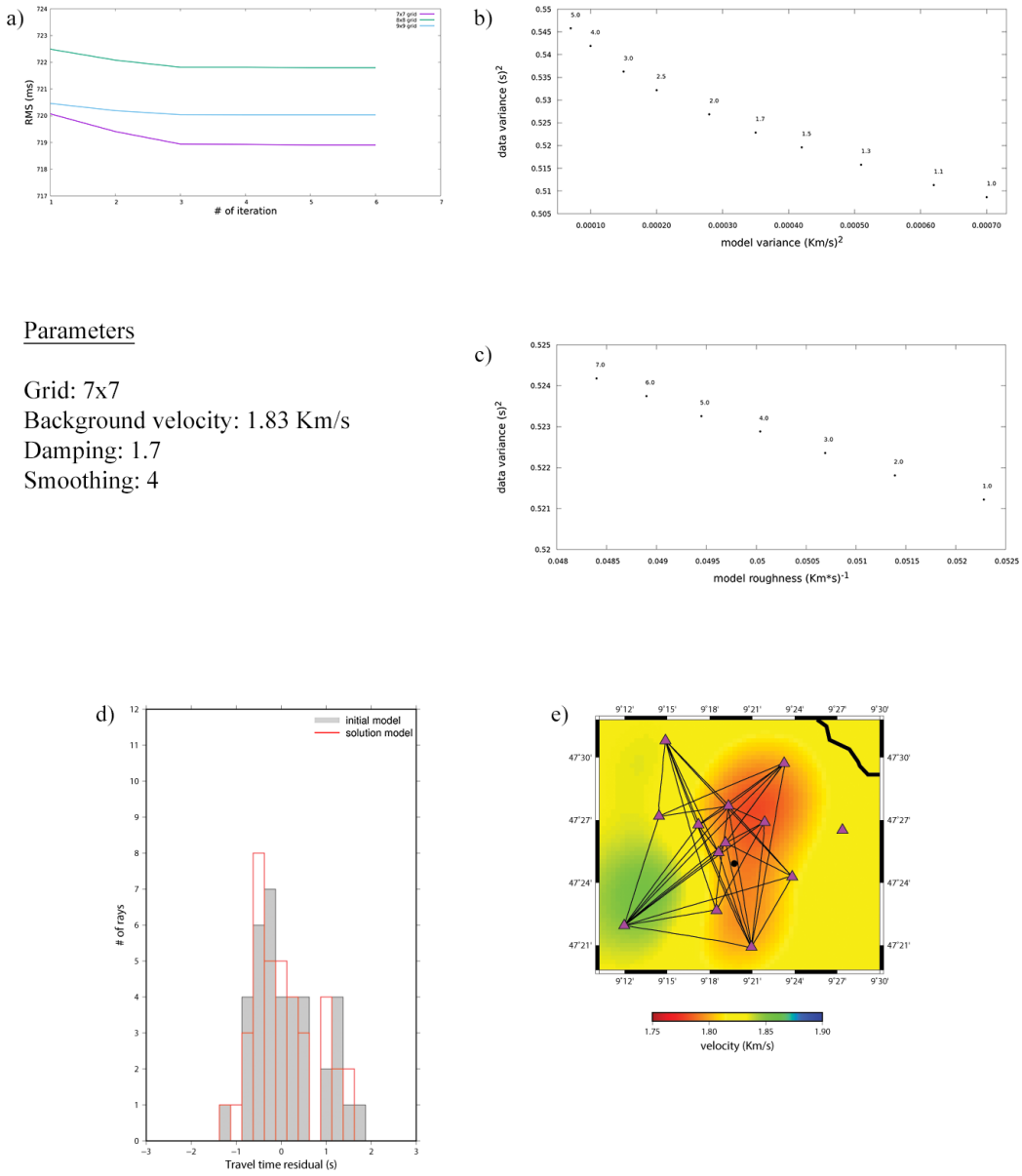


Figure 4.4.5: *Tomographic inversion strategy and solution model in the frequency band 0.6-0.8 Hz. a) RMS as a function of the iteration number of the tomographic inversion performed for different grid sizes. b) Damping and c) Smoothing factor trade-off curves. d) Frequency histogram of the (grey) initial and (red) solution model. e) Solution of the tomographic model with raypaths and stations used for the analysis in the specific range. The GT-1 well is represented with a black dot.*

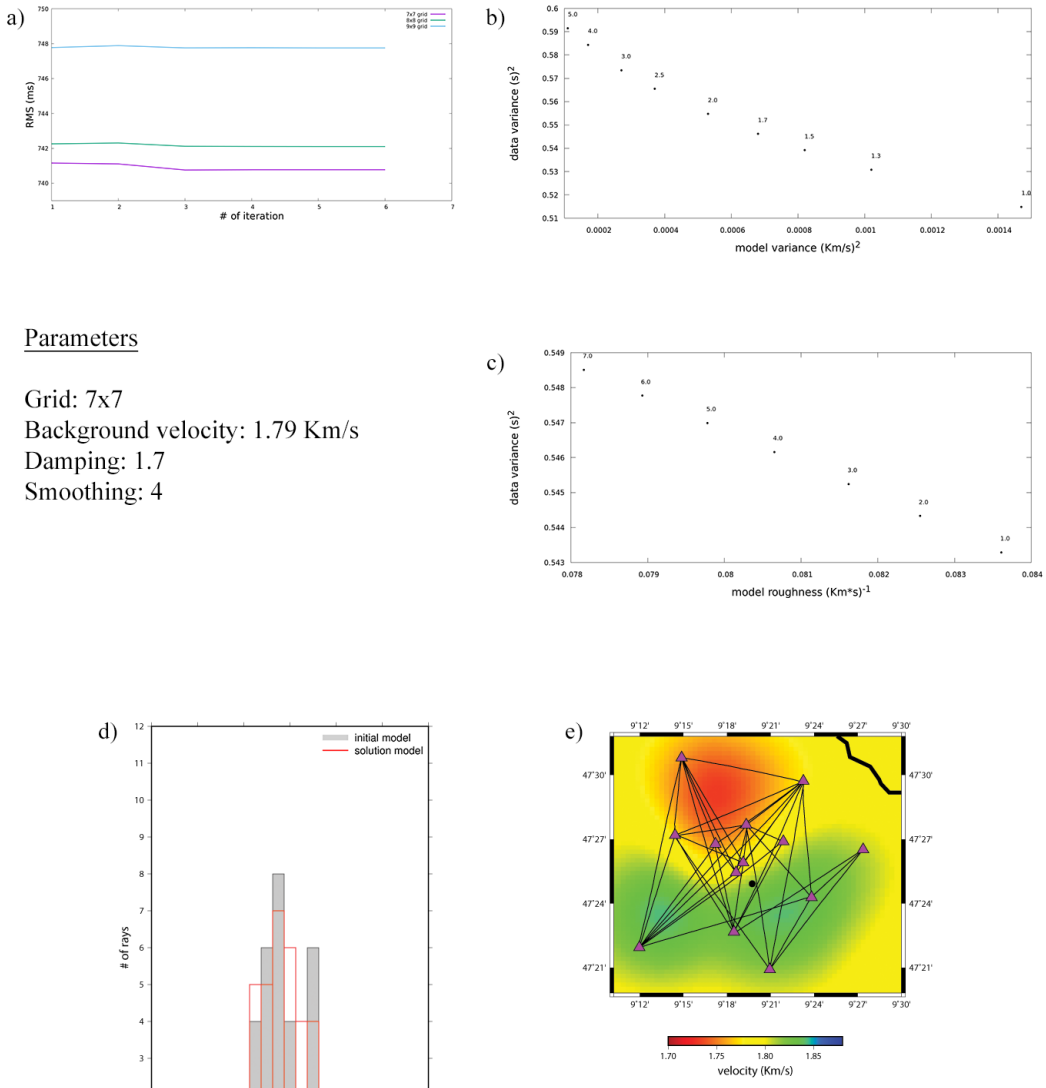
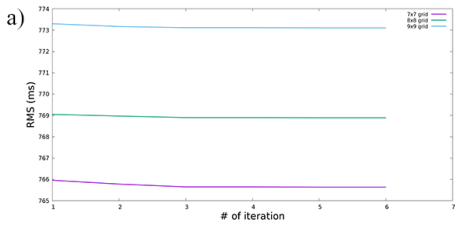
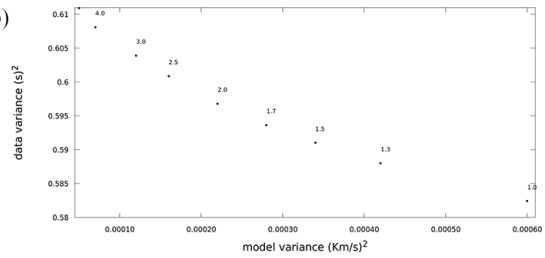


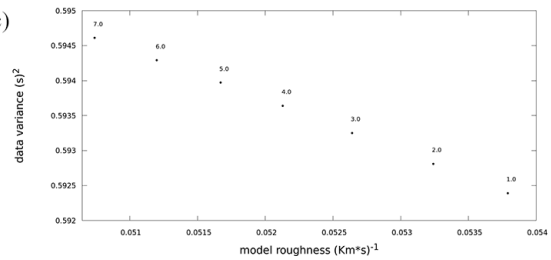
Figure 4.4.6: *Tomographic inversion strategy and solution model in the frequency band 0.7-0.9 Hz. a) RMS as a function of the iteration number of the tomographic inversion performed for different grid sizes. b) Damping and c) Smoothing factor trade-off curves. d) Frequency histogram of the (grey) initial and (red) solution model. e) Solution of the tomographic model with raypaths and stations used for the analysis in the specific range. The GT-1 well is represented with a black dot.*

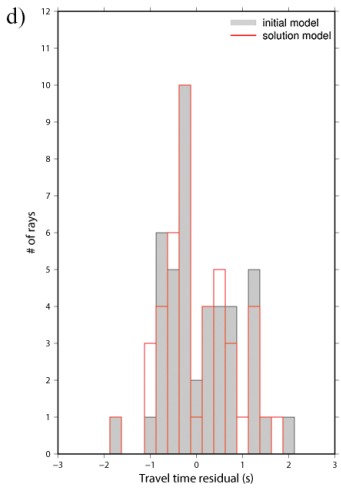




Parameters

Grid: 7x7
Background velocity: 1.77 Km/s
Damping: 1.7
Smoothing: 4





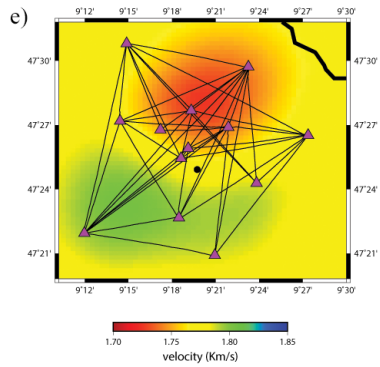


Figure 4.4.7: *Tomographic inversion strategy and solution model in the frequency band 0.8-1.0 Hz. a) RMS as a function of the iteration number of the tomographic inversion performed for different grid sizes. b) Damping and c) Smoothing factor trade-off curves. d) Frequency histogram of the (grey) initial and (red) solution model. e) Solution of the tomographic model with raypaths and stations used for the analysis in the specific range. The GT-1 well is represented with a black dot.*

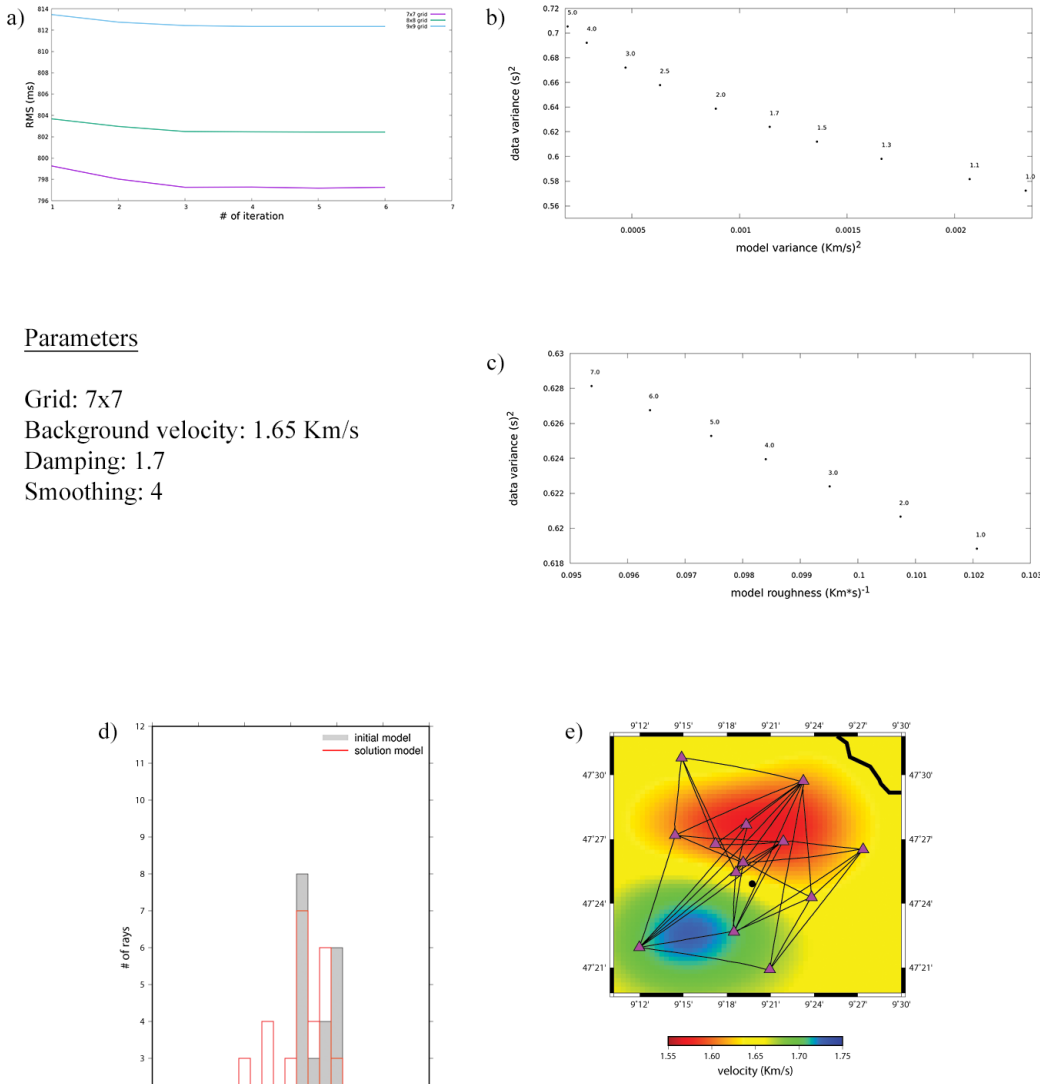


Figure 4.4.8: *Tomographic inversion strategy and solution model in the frequency band 0.9-1.1 Hz. a) RMS as a function of the iteration number of the tomographic inversion performed for different grid sizes. b) Damping and c) Smoothing factor trade-off curves. d) Frequency histogram of the (grey) initial and (red) solution model. e) Solution of the tomographic model with raypaths and stations used for the analysis in the specific range. The GT-1 well is represented with a black dot.*

4.5 An application of automatic event detection using Neural Network

This analysis aimed to exploit the potential of machine learning techniques to detect small magnitude events in low signal-to-noise recordings, such as those collected in induced seismicity. The considered dataset concerns the recordings of St. Gallen deep geothermal site and is composed of 388 records of seismic noise and 347 earthquakes with magnitude (M_L^{corr}) between -1.2 and 3.5 (Fig. 4.2.4). Also 20 synthetic traces generated by hiding selected events in the noise traces was jointly tested. It was performed a single-station analysis using the records from the only borehole station of the local seismic network SGT00. In addition, the three components EH1, EH2 and EHZ of each signal were analysed simultaneously.

Different approaches to discriminating automatically events have been developed over years, in this analysis an unsupervised network based on the Self-organizing map (SOM) algorithm was considered for its promising approach in data clustering (Kohonen, 2008; Vellido, 2019). In addition, the availability of a large unlabelled dataset but a limited number of available labelled event data makes an unsupervised algorithm suitable. A pre-processing method is required to identify the features of the signals that are useful for the SOM mapping. The features extraction was done by using the Linear Predictive Coding (LPC) technique for coding the spectrograms and a waveform parameterization for characterizing amplitudes in the time domain.

Pre-processing Stage

In the pre-processing, spectral and amplitude-versus-time features were extracted from the data, as they approximate the waveform characteristics considered in visual classification of signals. The traces are previously normalised with zero mean and unit variance.

To obtain the main features of the spectrum, the Linear Predictive Coding (LPC) technique (Makhoul, 1975) was used. It is widely used in speech processing for representing the spectral envelope of a digital signal of speech in compressed form. This method is preferred to the traditional Fourier spectral analysis because it is able to extract robust spectral information from the recordings in a compressed form, improving the performance efficiency of the discrimination process. Using signals Fourier transform, a number of useful frequencies equal to half the number of data points for each spectrum is required.

LPC predicts a signal sample at the time n, $S^*(n)$, making a linear combination of k previous signal samples $S(n)$

$$S^{*}(n) = \sum_{l=1}^{k} c_l \, S(n-l) \tag{4.5.1}$$

where c_l are the predictor coefficients (model parameters) and k is the number of the prediction coefficients (model order). In the frequency domain this is equivalent to modeling the signal spectrum by an all-pole filter. The estimate of the vector of the predictor coefficients $\mathbf{c}\,(c_1..c_l..)$ is performed by an optimization procedure that minimizes the error between the real signal at time n and its LPC prediction, i. e. the misfit function $E(\mathbf{c})$

$$E(\mathbf{c}) = \sum_{n} \left(S(n) - S^{*}(n)\right)^2 \tag{4.5.2}$$

The number k of predictive coefficients is a trade-off between the loss of information and the compactness of the representation. It is evaluated from the residual error trend as a function of the k.

In this study, each signal segment of 2.75 s (550 samples) was divided into N=10 Hanning windows of 0.5 s (100 sample) sliding along the signal with a half-window overlap to extract features over time. In each of the Hanning window, it was sufficient extract 6 coefficients (low residual error), obtaining a 60-component vector (6 coefficients x 10 windows) that encodes the spectral content over time for each waveform.

The information of amplitude in time-domain were extracted computing the function f_m as the difference between the maximum and minimum signal amplitudes s_i in the same Hanning window

$$f_m = \frac{\left(max_{i\varepsilon W_m}[s_i] - min_{i\varepsilon W_m}[s_i]\right) \times N}{\sum_{n=1}^{N}\left(max_{i\varepsilon W_n}[s_i] - min_{i\varepsilon W_n}[s_i]\right)} \qquad m = 1, ..., N \tag{4.5.3}$$

where N is the number of Hanning windows W_m.

For N=10 a vector of 10 amplitude features is obtained for each seismogram. Therefore, each of the 755 analysed signal was encoded with a feature vector of size 210 = [(60 frequency features + 10 amplitude features)x 3 components] and represents the input for the network training and discrimination process.

Self-Organizing Map Algorithm

To perform an automatic event detection, the neural network based on the Kohonen *Self-organizing map* (*SOM*) algorithm (Kohonen et al., 1996; Kohonen, 1997) was employed. It is an unsupervised machine learning approach

that is widely used for clustering and visualization of data sets into a topological map. In contrast to supervised feed-forward nets, SOMs learn in an unsupervised manner, without a priori assumptions, guaranteeing minimal bias from the investigator. Analysing the features extracted in the pre-processing stage, the SOM performs a non-linear mapping of input data onto a two-dimensional regular grid of processing units known *nodes* (Fig. 4.5.1). The mapping preserves the most important topological and metric relationships of the data.

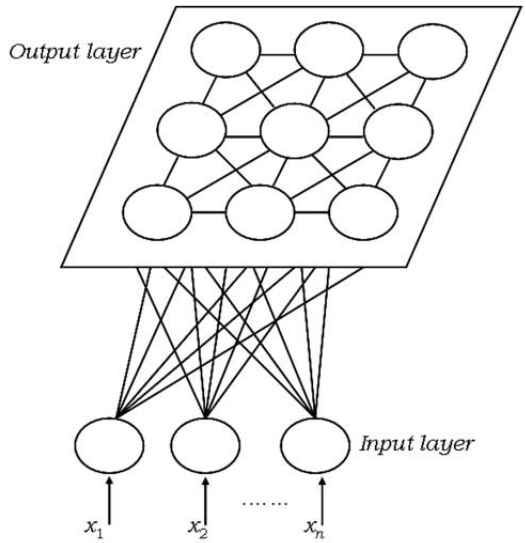


Figure 4.5.1: *Schematization of a Self-organizing map in which each input unit is connected to all units of the output layer (Esposito et al., 2006).*

The SOM clasterizes the input data associating a prototype with each node and placing similar input features on neighbouring nodes in the grid. Considering the feature input vector of the signal, $\mathbf{x}(t)$, extracted at step t and the prototype of $i-th$ node, $\mathbf{m}_i(t)$, a new prototype is assigned to node c (winner index) by the condition

$$\|\mathbf{x}(t) - \mathbf{m}_c(t)\| \leq \|\mathbf{x}(t) - \mathbf{m}_i(t)\| \qquad \forall\, i \tag{4.5.4}$$

All prototypes in the neighborhood are updated by the equation

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{ci}(\mathbf{x}(t) - \mathbf{m}_i(t)) \tag{4.5.5}$$

Where h_{ci} is the neighborhood function, a Gaussian function that decrease with the distance between the $i-th$ and $c-th$ nodes on the map grid

$$h_{ci} = e^{-\frac{d_{ci}^2}{\sigma_t^2}} \tag{4.5.6}$$

where σ_t is the neighborhood radius at step t and d_{ci} is the distance between map units c and i on the map grid.

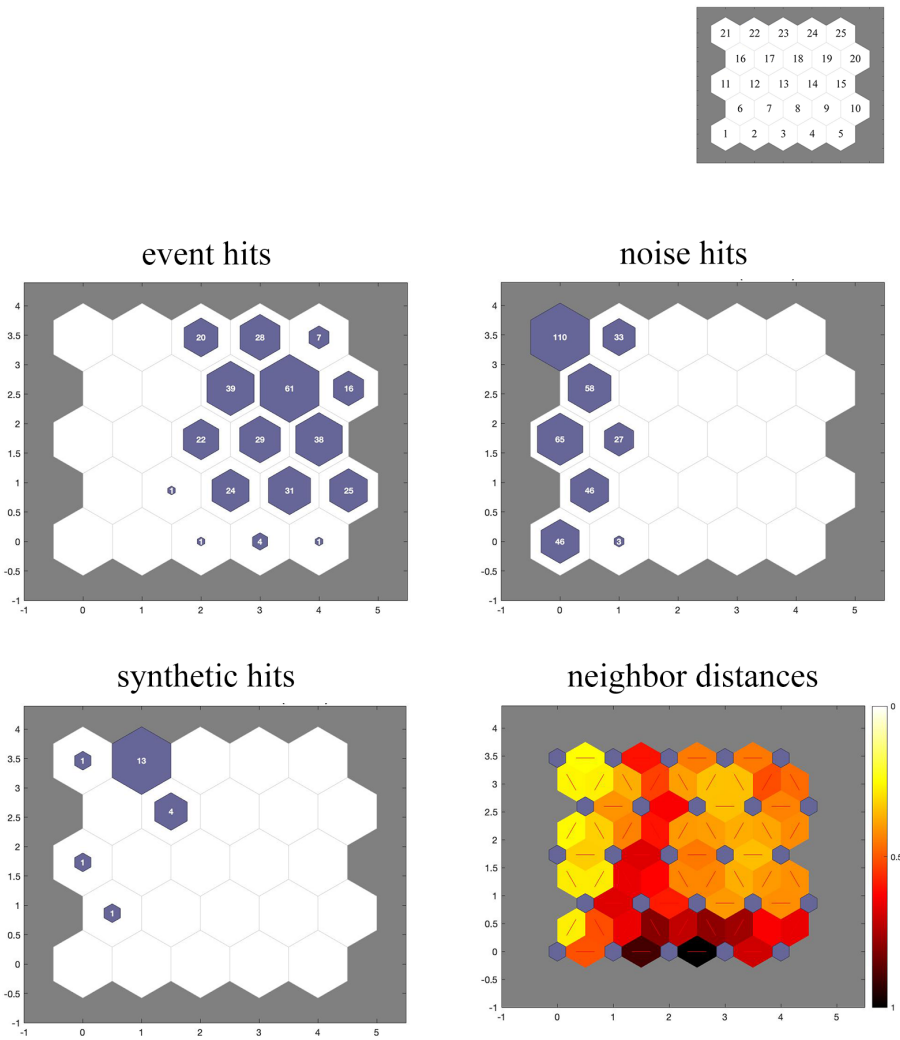


Figure 4.5.2: *Results of the SOM map with 5x5= 25 nodes representing (see top label for each panel) the event hits, noise hits, synthetic hits, respectively. The neighbour distances graph denotes the distances between the prototypes, the lighter the colour, the more similar the features between neighbouring nodes. The upper panel shows the numeration of the 25 nodes.*

The figure 4.5.2 shows the results obtained for the SOM map with $5 \times 5 = 25$ nodes. Each node is represented as a hexagon where the number of data associated with that node is shown. For simplicity, three graphs of the same SOM map were generated to highlight the position of the nodes containing the event, noise, and synthetic prototypes, respectively. The topological relationship between the nodes is reported in the neighbour distances graph with a color map in which the lighter colour indicates more similar characteristics between the prototypes.

As can be seen in the results, the network clearly distinguishes noise traces from events, their graphs are indeed complementary. Concerning synthetic traces, only 4 traces in the node 17 are distinct from noise nodes. This depends on the signal-to-noise ratio of their 3 components. Figure 4.5.3 compares the signal-to-noise ratios of examples of traces used to build the synthetic traces for which the analysis leads to a node noise and the node 17 hit, respectively.

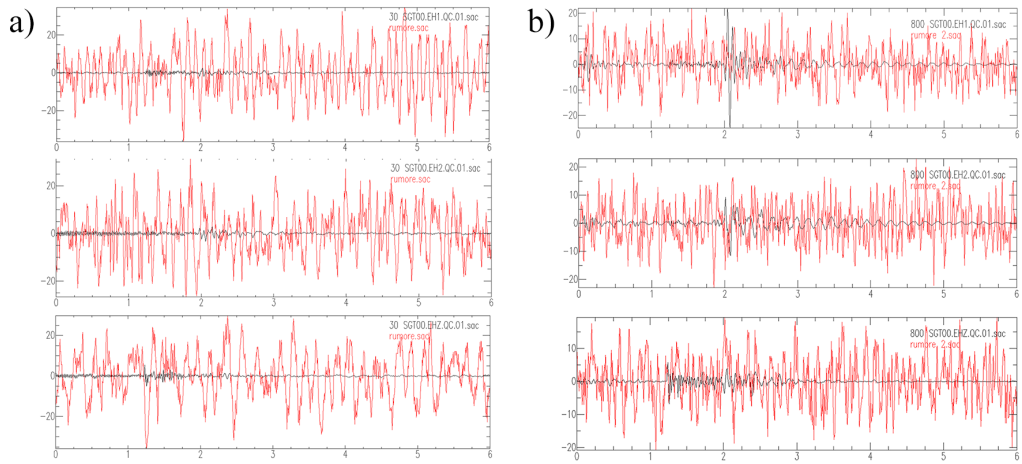


Figure 4.5.3: *Example of the event (black) and noise (red) traces overlapping, for the three components (see labels), used to generate the three synthetic traces analysed simultaneously for a specific hidden event that leding a) to the node noise 6 and b) to the node 17 hit.*

The results concerning the clustering of events were also analysed in terms of the main characteristics of an event, such as magnitude, events location, hypocentral distance, and origin time. Therefore, the similarity of these characteristics was assessed at each node of the SOM map.

The figure 4.5.4 shows the magnitude values of the events and the average at each node. It is observed that the average magnitude has an increasing

trend for adjacent nodes from left to right in the SOM map (e.g. for nodes $8 \rightarrow 9 \rightarrow 10$, nodes $13 \rightarrow 14 \rightarrow 15$, and so on).

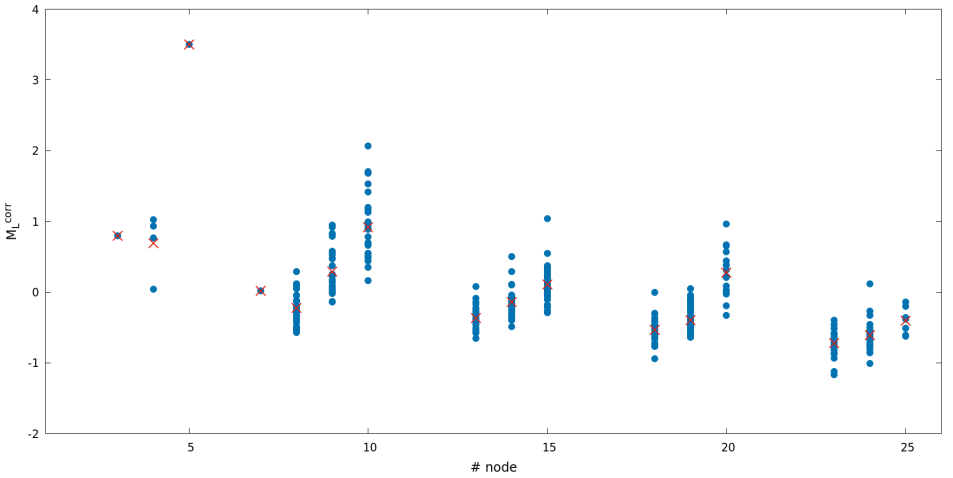


Figure 4.5.4: *Values of event magnitude as a function of node. The average value (red x) for each node is also shown.*

The figure 4.5.5 shows the location of events coloured according to the corresponding node. For a better analysis of the locations, the number of nodes was split into three graphs. The colour of the nodes in each graph is independent to the others. In particular, when analysing the events in the deepest part of the location cluster (see the red arrows), it is observed that some of them belong to the same node (same colour) or neighbouring nodes in the SOM map. Specifically, the events with deepest locations in map 1 are blue and refer to node 8; in map 2 they are burgundy and belong to node 9; while in map 3 they are green concerning the node 23. To complete the cluster of deeper events there are also those at node 14 (green in map 2) which is adjacent to nodes 8 and 9 in the SOM map and node 24 (blue in map 3) which is adjacent to node 23. In these two nodes, however, the events show more scattered depth values.

Therefore, despite the use of a completely unsupervised technique, the network also partially recognises some main events characteristics such as their location and magnitude.

The evaluations carried out on the event origin time and hypocentral distance do not provide any information on a possible correlation between the clustering of events and these characteristics.

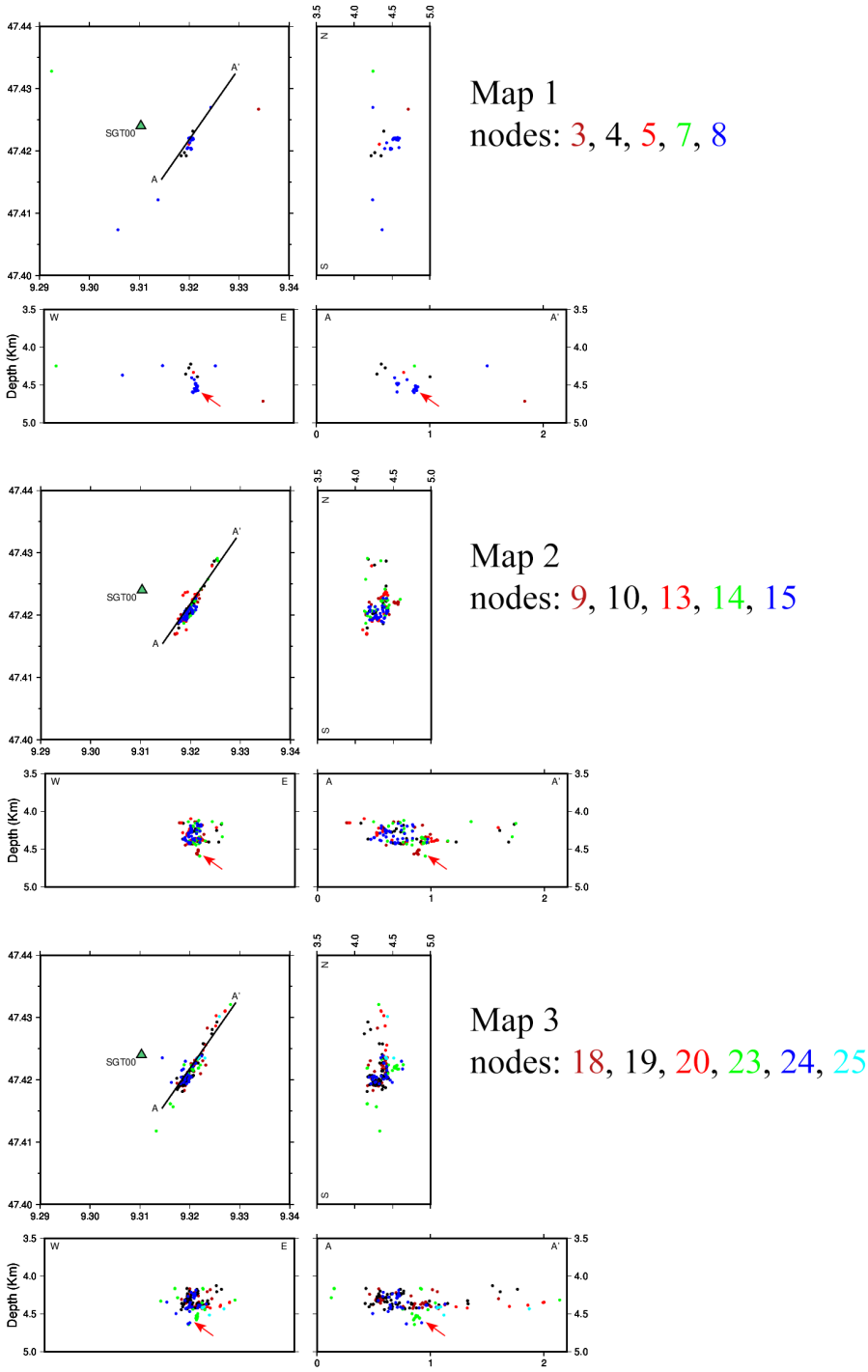


Figure 4.5.5: *Location of events with colour map as a function of node. The four panel represent the seismicity in the 1) map view, 2) West-East and 3) North-South vertical section, and 4) cross-section along the profiles AA' indicated in the map view.*

4.6 Discussion

Daily stacked cross-correlation functions between 78 station pairs were computed using three-month continuous data recorded by the seismic network at the St. Gallen site. Results show the presence of a coherent signal in almost all couple of stations at both causal and acausal part in the band 0.5-0.9 Hz. Its estimated propagation velocity is about 1750 m/s, consistent with the values of surface wave velocities retrived in Heuberger et al. (2016). The analysis of the temporal evolution of waveform coherence in the band of interest 0.5-0.9 Hz shows an energetic signal that is time stable. For some station pairs, a loss of coherence occurs on days before the first well stimulation on 14^{th} July 2013 (Fig. 4.4.3), and is present on some days between 14^{th} and 20^{th} July during all stimulation tests and the occurrence of the gas release and the M_L 3.5 earthquake. The loss of coherence is also observed after many days following stimulation (e.g. 20 to 40 days). Obermann et al. (2015) also analysed ambient seismic noise cross-correlations using 17 months data from nine stations of the local network of St. Gallen. They retrived an emerging wavetrain in the 0.1-1.0 Hz range and analysed the temporal evolution of waveform coherence in the cross-correlation functions coda. They observed a loss of waveform coherence starting after the first injection test that was interpret as a local pertubation of the medium caused by the injection of fluids. The coherence loss is maximum on average on 19^{th} and 20^{th} July 2013 with the gas kick and the M_L 3.5 earthquake, and recovers to the initial values after roughly 30 days (Fig. 4.2.1).

Tomographic images were retrived from the inversion of the surface wave arrival times of the cross-correlation stacks in the narrow frequency bands 0.5-0.7, 0.6-0.8, 0.7-0.9, 0.8-1.0, and 0.9-1.1 Hz. The maximum depth of investigation is about 1.2 km. The RMS and residuals histograms (Fig. 4.4.4 - 4.4.8 a and d) show that there is no significant improvement between the initial and final solutions in each analysed frequency band. The final model is strongly dependent on the number of considered input data which is equal, on average, to 40 observables per band (see raypaths in figures 4.4.4 - 4.4.8 e). The inversion carried out on a low number of data could also be influenced by any outliers present in the dataset.

The St. Gallen recordings resulted suitable for an application of automatic event detection using an unsupervised network. The SOM technique resulted a promising approach for visualization and comprehension of the inherent structure of the data without a priori assumptions, as also suggests by Esposito et al. (2006). First results show a clearly discrimination of events from noise traces, whereas hidden events in noise are not detected if the signal-to-noise ratio is very low. However, this type of network only clus-

tering data according to their similarities. The analysis of the cluster of deeper events is of particular interest since its presence is debated in Diehl et al. (2017) and verified performing a sequence of tests using different data types and velocity models for a double-difference inversion. The alternative technique of using neural networks could contribute to discriminate the real presence of deeper events belonging to the cluster. However, using data from only one station have limitations that affect the results.

Chapter 5

Conclusion

Cross-correlation analysis of two-month of continuous ambient seismic noise was performed to investigate the propagation medium of the shale gas site of Wysin. Coherent Rayleigh waves extracted from vertical cross-correlations at couples of synchronized stations exhibit a significant energy in the 1.5-2.0 Hz frequency band, information confirmed also by investigating other components in the radial and transverse directions of cross-correlation tensor. The predominant signal propagates with an estimated velocity of about 400 m/s and can be considered reasonably time stable with the occurrence of a few losses of coherence that seems not be related with the HF operations. A seasonal analysis in the band of interest 1.5-2.0 Hz, reveals that the characteristics of the noise source of the study area change in the winter time. The seasonal variability of the cross-correlation signals should be studied in more detail, perhaps using additional information such as meteorological information (wind, rainfall, etc.) which could explain the loss of the signal from the cross-correlation analysis.

The information obtained from the cross-correlation analysis were used to perform tomographic inversions with two different approch, to reconstruct 2-D surface waves velocity model of the medium. From the method using the surface wave arrival times picked on the stacked cross-correlation functions, a fixed-frequency tomography in the narrow range 1.5-2.0 Hz was obtained allowing an investigation of the medium up to a maximum depth of about 70 m. The tomographic model shows a velocity range between 300 m/s and 380 m/s. A low-velocity zone around the wells suggests a possible correlation with HF operations that could have changed the fluids state in the reservoir. Future work will explore the possibility to reconstruct a 3-D tomographic model to better investigate these findings. A study by selecting cross-correlations based on an established threshold of their signal-to-noise ratio could be carried out for a possible improvement of the emerging coher-

ent signal in the other frequency ranges, in order to reach and investigate depths greater than those currently obtained. A high-velocity anomaly is also present and located in the south-west part of the tomographic image. For data inversion it was also used a non-linear multiscale approach. In this case the data are the dispersion curves from cross-correlations obtained by frequency-time analysis (FTAN). Tomographic images were retrieved for nine periods within the range 0.3-1.20 s, permitting to reach an investigation depth of about 150 m. In this case, as the first inversion tests were performed during the last period of the thesis work, the results are considered to be entirely preliminary and the performed tomographic analysis is still under improvement. As a further development, to obtain a reliable 3-D Vs model, the direct surface wave tomography method (Fang et al., 2015; Li et al., 2016; Gu et al., 2019) can be used, where all the measured dispersion curves can be directly inverted for 3-D Vs model of the near surface.

Also for the second study area, the deep geothermal energy site of St. Gallen, a seismic interferometry method based on cross-correlation analysis was carried out. Vertical and mixed components cross-correlation analysis of three-month of continuous noise recording show a well identified coherent phases in the 0.5-0.9 Hz frequency band. The energetic signal propagates with an estimated velocity of about 1750 m/s. It results stable over time with a loss of coherence observed a few days before the well stimulation tests and in the time interval when these tests were followed by the gas kick and M_L 3.5 earthquake. Moreover, this loss of coherence continues to occur on subsequent days, even after about 40 days.

The tomographic inversion was performed using traveltimes of surface waves derived from the stacked cross-correlation functions. Inversions was carried out for five narrow frequency band in the range 0.5-1.1 Hz where the most energetic signal emerged, exploring the medium up to about 1.2 km. The results show a strong dependence of the input data both in terms of the number of available observables and the presence of outliers that do not allow to solve the tomographic problem in an optimal way.

Application of an automatic event detection using an unsupervised neural network algorithm was performed testing one-station low signal-to-noise recordings of St. Gallen site. The results are promising in discriminating events from noise and in clustering events by some of their main characteristics such as magnitude and location. Future developments on this work will include a joint analysis of recordings from the other stations of the seismic network and the implementation of a supervised strategy. Supervised strategies rely on a labelled dataset, i.e. set of data classified by analystes that are used to train and test the networks. The larger is the labelled dataset, the

better is the generalization ability of the net. The available event traces are not sufficient for this purpose, therefore, with a dataset augmentation technique, a large enough dataset of synthetic traces will be create from existing data. It is expected that the discrimination performances using a supervised strategy will be good even in the presence of low signal-to-noise ratio, as it was successfully achieved in many volcanic contexts (Del Pezzo et al., 2003; Scarpetta et al. 2005; Esposito et al., 2006). Volcanic data analysis can be useful to make comparisons with results obtained using recordings from geo-resource exploration areas for the small magnitudes and involved fluids in both cases.

Bibliography

Aki K. and Richards P.G.; 2002: Quantitative Seismology. Sausalito, University Science Books.

Aki K. and Lee W.H.K.; 1976: *Determination of three-dimensional velocity anomalies under a seismic array using first P arrival time from local earthquakes: 1. homogeneous initial model.* J. Geophys. Res. 81, 4381-4399.

Aster R. C., Borchers B. and Thurber C. H.; 2018: *Parameter estimation and inverse problems.* Elsevier.

Battaglia J., Zollo A., Virieux J. and Dello Iacono D.; 2008: *Merging active and passive data sets in travel time tomography: the case study of Campi Flegrei caldera (southern Italy).* Geophysical Prospecting 56, 555-573.

Bensen G. D., Ritzwoller M. H., Barmin M. P., Levshin A. L., Lin F., Moschetti M. P., Shapiro N. M. and Yang Y.; 2007: *Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurement.* Geophys. J. Int., 169, 1239–1260.

Bonnefoy-Claudet S.; Cotton F. and Bard, P.-Y.; 2006: *The nature of noise wavefield and its applications for site effects studies. A literature review.* Earth Science Review. 79 (3–4): 205 227. doi:10.1016/j.earscirev.2006.07.004.

Bormann P.; 2002: *NMSOP – New Manual of Seismological Observatory Practice.* IASPEI, GeoForschungsZentrum Potsdam, Germany.

Brenguier F., Kowalski P., Ackerley N., Nakata N., Boué P., Campillo M., Larose E., Rambaud S., Pequegnat C., Lecocq T., Roux P., Ferrazzini V., Villeneuve N., Shapiro N. M. and Chaput J.; 2016: *Toward 4D Noise-Based Seismic Probing of Volcanoes: Perspectives from a Large-N Experiment on Piton de la Fournaise Volcano.* Seismological Research Letters, 87, 15–25. https://doi.org/10.1785/0220150173.

Brenguier F., Campillo M., Hadziioannou C., Shapiro N. M., Nadeau R. M. and Larose E.; 2008: *Postseismic relaxation along the San Andreas fault at Parkfield from continuous seismological observations.* Science, 321(5895), 1478–1481. https://doi.org/10.1126/science.1160943.

Cabrera-Pérez I., D'Auria L., Soubestre J., Barrancos J., Padilla G. D., Pérez N. M., 2020: *A nonlinear multiscale inversion approach for ambient noise tomography.* Geophysical Journal International, Vol. 225, Issue 2, p 1158–1173, https://doi.org/10.1093/gji/ggaa574.

Chmiel M., Mordret A., Boué P., Brenguier F., Lecocq T., Courbis R., Hollis D., Campman X., Romijn R. and Van der Veen W.; 2019: *Ambient noise multimode Rayleigh and Love wave tomography to determine the shear velocity structure above the Groningen gas field.* Geophys. J. Int. (2019) 218, 1781–1795, doi: 10.1093/gji/ggz237.

Clarke H., Eisner L., Styles P. and Turner P.; 2014: Felt seismicity associated with shale gas hydraulic fracturing: The first documented example in Europe. Geophys. Res. Lett., 41, 8308–8314, doi:10.1002/2014GL062047.

Del Pezzo E., Esposito A., Giudicepietro F., Marinaro M., Martini M. and Scarpetta S.; 2003: *Discrimination of earthquakes and underwater explosions using neural networks.* Bull. Seism. Soc. Am. 93, no. 1, 215–223.

Diehl T., Kraft T., Kissling E. and Wiemer S.; 2017: *The induced earthquake sequence related to the St. Gallen deep geothermal project (Switzerland): Fault reactivation and fluid interactions imaged by microseismicity.* J. Geophys. Res. Solid Earth, 122, 7272–7290, doi:10.1002/2017JB014473.

Edwards B., Kraft T., Cauzzi C., Kastli P. and Wiemer S.; 2015: *Seismic monitoring and analysis of deep geothermal projects in St Gallen and Basel, Switzerland.* Geophysical Journal International, 201, 1020–1037.
doi: 10.1093/gji/ggv059.

Ellsworth W. L; 2013: *Injection-induced earthquakes.* Science, 341(6142), 1225942.

Esposito A. M., Giudicepietro F., Scarpetta S., D'Auria L., Marinaro M. and Martini M.; 2006: *Automatic Discrimination among Landslide, Explosion-Quake, and Microtremor Seismic Signals at Stromboli Volcano Using Neural Networks.* Bulletin of the Seismological Society of America, Vol. 96, No. 4A, pp. 1230–1240. doi: 10.1785/0120050097.

Fang H., Yao H., Zhang H, Huang YC and van der Hilst R.D.; 2015: *Direct inversion of surface wave dispersion for three-dimensional shallow crustal structure based on ray tracing: methodology and application.* Geophys.J. Int.,201, 1251-1263.

Goldstein P., Dodge D., Firpo M., Minner L., Lee W.H.K., Kanamori H., Jennings P.C. and Kisslinger C.; 2003: *SAC2000: Signal processing and*

analysis tools for seismologists and engineers. The IASPEI International Handbook of Earthquake and Engineering Seismology. 81, 1613-1620.

Grad M., Polkowski M. and Ostaficzuk S.R.; 2015: *High-resolution 3D seismic model of the crustal and uppermost mantle structure in Poland.* Tectonophysics, 666, 188–210.

Grigoli, F., S. Cesca, A. P. Rinaldi, Manconi A., López-Comino J. A., Clinton J. F., 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, 360 (6392), 1003-1006.

Grigoli, F., Cesca S., Priolo E., Rinaldi A. P., Clinton J. F., Stabile T. A., Dost B., Fernandez M. G., WiemerS. and Dahm T.; 2017: *Current challenges in monitoring, discrimination, and management of induced seismicity related to underground industrial activities: A European perspective.* Rev. Geophys., 55, doi:10.1002/2016RG000542.

Gu N., Wang K., Gao J., Ding N., Yao H., Zhang H. 2019. *Shallow crustal structure of the Tanlu fault zone near Chao Lake in eastern China by direct surface wave tomography from local dense array ambient noise analysis.* Pure and Applied Geophysics, 176(3), 1193–1206.

Haering M., Schanz U., Ladner F. and Dyer B. C.; 2008: *Characterization of the Basel 1 geothermal system.* Geothermics, 37, 469–495, doi:10.1016 /j.geothermics.2008.06.002.

Heuberger S., Roth P., Zingg O., Naef H. and Meier B. P.; 2016: *The St. Gallen Fault Zone: a long-lived, multiphase structure in the North Alpine Foreland Basin revealed by 3D seismic data.* Swiss J Geosci, 109:83–102. doi: 10.1007/s00015-016-0208-5.

ICHESE (International Commission on Hydrocarbon Exploration and Seismicity in the Emilia-Romagna region) 2014: *Report on the hydrocarbon exploration and seismicity in Emilia Region.* Regione Emilia-Romagna, E-R Ambiente, Geologia, sismica e suoli, Bologna, Italy, 213 pp.

Induced Seismicity Potential in Energy Technologies; 2013. The National Academies Press.

Kennett B. L. N., Sambridge M. S. and Williamson P. R.; 1988: *Subspace methods for large scale inverse problems involving multiple parameter classes.* Geophysical Journal. 94, 237–247.

Keranen K. M., Savage H. M., Abers G. A. and Cochran E. S.; 2013: *Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence.* Geology, 41(6), 699–702, doi:10.1130/G34045.1.

Kohonen T.; 2008: *Data Management by Self-Organizing Maps.* In: Zurada J.M., Yen G.G., Wang J. (eds) Lecture Notes in Computer Science, vol 5050. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-68860-0_15

Kohonen T., 1997: *Self-Organizing Maps.* Second Ed., Series in Infor- mation Sciences, Vol. 30, Springer, Heidelberg.

Kohonen T., Hynninen J., Kangas J. and Laaksonen J.; 1996: *SOM_PAK: the self-organizing map program package.* Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland. Also available at www.cis.hut.fi/research/ som_lvq_pak.shtml.

Krausmann E., Girgin S. and Necci A.; 2019: *Natural hazard impacts on industry and critical infrastructure: Natech risk drivers and risk management performance indicators.* International Journal of Disaster Risk Reduction, 40, 1-9, 101163, https://doi.org/10.1016/j.ijdrr.2019.101163.

Kwang-Hee K., Jin-Han R., YoungHee K., Sungshil K., Su Young K. and Wooseok S.; 2018: *Assessing whether the 2017 Mw 5.4 Pohang earthquake in South Korea was an induced event.* Science, 360 (6392), 1007-1009.

Lecocq T. et al.; 2020: *Global quieting of high-frequency seismic noise due to COVID-19 pandemic lockdown measures.* Science, 10.1126/science.abd2438.

Lehujeur M., Vergne J., Schmittbuhl J., Zigone D., Le Chenadec A. and EstOF Team; 2018: *Reservoir imaging using ambient noise correlation from a dense seismic network.* Journal of Geophysical Research: Solid Earth, 123, 6671–6686. https://doi.org/10.1029/2018JB015440.

Lehujeur M., Vergne J., Schmittbuhl J. and Maggi A.; 2015: *Characterization of ambient seismic noise near a deep geothermal reservoir and implications for interferometric methods: a case study in northern Alsace, France.* Geothermal Energy, 3(1), 3. DOI: 10.1186/s40517-014-0020-2.

Levshin A. L., Ratnikova L. and Berger J.; 1992: *Peculiarities of surface-wave propagation across Central Eurasia.* Bull. Seism. Soc. Am., 82, 2464-2493.

Levshin A. L., Pisarenko V. F. and Pogrebinsky G. A.; 1972: *On a frequency-time analysis of oscillations.* Ann. Geophys., 28, 211-218.

Li C., Yao H., Fang H., Huang X., Wan K., Zhang H. and Wang K.; 2016: *3D near-surface shear-wave velocity structure from ambient noise tomography and borehole data in the Hefei urban area, China.* Seismological Research Letters, 87(4), 882-892,doi: 10.1785/0220150257.

Lin F.-C., Moschetti M. P. and Ritzwoller M. H.; 2008: *Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps.* Geophys. J. Int., 173(1), 281–298, doi:10.1111/j1365-246X.2008.03720.x.

Lin F.-C., Ritzwoller M. H. and Shapiro N. M.; 2006: *Is ambient noise tomography across ocean basins possible?.* Geophys. Res. Lett., 33, L14304, doi:10.1029/2006GL026610.

López Comino J. A., Cesca S., Jarosławski J., Montcoudiol N., Heimann S., Dahm T., Lasocki S., Gunning A., Capuano P. and Ellsworth W. L.; 2018: *Induced seismicity response of hydraulic fracturing: results of a multidisciplinary monitoring at the Wysin site, Poland.* Scientific Reports, DOI:10.1038/*s41598-018-26970-9*.

López Comino J. A., Cesca S., Kriegerowski M., Heimann S., Dahm T., Mirek J. and Lasocki S.; 2017: *Monitoring performance using synthetic data for induced microseismicity by hydrofracking at the Wysin site (Poland).* Geophy. J. Int., 210, 42-55, DOI: 10.1093/gji/ggx148.

Makhoul J.; 1975: *Linear prediction: a tutorial review.* Proc. IEEE 63, 561–580.

Montcoudiol N., Isherwood, C., Gunning, A., Kelly, T. and Younger, P. L.; 2017: *Shale gas impacts on groundwater resources: Understanding the behavior of a shallow aquifer around a fracking site in Poland.* Energy Procedia 125, 106–115.

Mordret A., Landés M., Shapiro N. M., Singh S. C., Roux P. and Barkved O. I.; 2013: *Near-surface study at the Valhall oil field from ambient noise surface wave tomography.* Geophysical Journal International, 193(3), 1627–1643. https://doi.org/10.1093/gji/ggt061.

Mulargia F. and Bizzarri A.; 2014: Anthropogenic triggering of large earthquakes. Sci. Rep., 4, 6100–6107, doi:10.1038/srep06100.

Obermann A., Kraft T., Larose E. and Wiemer S.; 2015: *Potential of ambient seismic noise techniques to monitor the St. Gallen geothermal site (Switzerland).* J. Geophys. Res. Solid Earth, 120, 4301–4316, DOI: 10.1002/2014JB011817.

Olivier G., Brenguier F., Campillo M., Roux P., Shapiro N. M. and Lynch R.; 2015: *Investigation of coseismic and postseismic processes using in situ measurements of seismic velocity variations in an underground mine.* Geophys. Res. Lett., 42, 9261–9269, doi:10.1002/2015GL065975.

Pezzo, G., De Gori, P., Lucente, F. P., & Chiarabba, C.; 2018: *Pore pressure pulse drove the 2012 Emilia (Italy) series of earthquakes.* Geophysical Research Letters, 45, 682–690. https://doi.org/10.1002/2017GL076110.

Picozzi M., Parolai S., Bindi D. and Strollo A.; 2009: *Characterization of shallow geology by high-frequency seismic noise tomography.* Geophys. J. Int., 176, 164–174.

Poli P., Pedersen H. A., Campillo M. and the POLENET/LAPNET Working Group; 2012: *Emergence of body waves from cross-correlation of short period seismic noise.* Geophys. J. Int., 188, 549–558.

Rawlinson N., Pozgay S. and Fishwick S.; 2010: *Seismic tomography: a window into deep Earth. Physics of the Earth and Planetary Interiors.* 178(3-4). 101-135. DOI: 10.1016/j.pepi.2009.10.002.

Rawlinson N. and Sambridge M.; 2005: *The fast marching method: An effective tool for tomographic imaging and tracking multiple phases in complex layered media.* Exploration Geophysics. 36, 341-350.

Rawlinson N.; FMST: *Fast Marching Surface Tomography Package – Instructions.* http://rses.anu.edu.au/~nick/surftomo.html.

Rost S. and Thomas C.; 2002: *Array seismology: methods and applications.* Reviews of Geophysics, 40:1008. doi:10.1029/2000RG000100.

Roux P.; 2009: *Passive seismic imaging with direct ambient noise: application to surface waves and the San Andreas Fault in Parkfield, CA.* Geophys. J. Int., 179, 367–373.

Rubinstein J. L. and Mahani A. B.; 2015: *Myths and facts on wastewater injection, hydraulic fracturing, enhanced oil recovery, and induced seismicity.* Seismol. Res. Lett., 86(4), 1060–1067.

Scarpetta S., Giudicepietro F., Ezin E. C., Petrosino S., Del Pezzo E., Martini M. and Marinaro M.; 2005: *Automatic Classification of seismic signals at Mt. Vesuvius Volcano, Italy using Neural Networks*, Bull. Seism. Soc. Am. 95, 185–196.

Schimmel M. and Paulssen H.; 1997: *Noise reduction and detection of weak, coherent signals through phase-weighted stacks.* Geophys. J . Int., 130, 497-505.

Sethian J. A.; 1996: *A fast marching level set method for monotonically advancing fronts.* Proc. Natl. Acad. Sci. USA. Vol. 93, No. 4, pp. 1591-1595.

Shapiro N. M., Campillo M., Stehly L. and Ritzwoller M.H.; 2005: *High-Resolution surface wave tomography from ambient seismic noise.* Science, 307, 1615-1618.

Shapiro N. M. and Campillo M.; 2004: *Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise.* Geophys. Res. Lett., 31, L07614, doi:10.1029/2004GL019491.

van Thienen-Visser K. and Breunese J. N.; 2015: *Induced seismicity of the Groningen gas field: History and recent developments.* Leading Edge, 34(6), 664–671.

Vassallo M., De Matteis R., Bobbio A., Di Giulio G., Adinolfi G. M., Cantore L., Cogliano R., Fodarella A., Maresca R., Pucillo S. and Riccio G.; 2019: *Seismic noise cross-correlation in the urban area of Benevento city (Southern Italy).* Geophys. J. Int., 217, 1524–1542, DOI: 10.1093/gji/ggz101.

Vassallo M., Festa G., Bobbio A. and M. Serra; 2016: *Low shear velocity in a normal fault system imaged by ambient noise cross correlation: The case of the Irpinia fault zone, Southern Italy.* J. Geophys. Res. Solid Earth, 121, 4290–4305, DOI: 10.1002/2015JB012410.

Vellido A.; 2019: *Advances in self-organizing maps, learning vector quantization, clustering and data visualization: WSOM 2019.* Springer.

Wandycz P., Święch E., Eisner L., Pasternacki A., Wcisło M. and Maćkowski T.; 2019: *Estimation of the quality factor based on the microseismicity recordings from Northern Poland.* Acta Geophysica, 67:2005–2014. https://doi.org/10.1007/s11600-019-00362-7.

Weaver R. L. and Lobkis O. I.; 2001: *On the emergence of the Green's function in the correlations of a diffuse field.* J. Acoust. Soc. Am., 110, 3011 – 3017.

Yang Y., Ritzwoller M.H., Levshin A.L. and Shapiro N.M.; 2007: *Ambient noise Rayleigh wave tomography across Europe.* Geophys. J. Int., 168, 259–274.

Yao H., van der Hilst R.D. and de Hoop M.V.; 2006: *Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis – I. Phase velocity maps.* Geophys. J. Int., 166, 732–744, doi: 10.1111/j.1365–246X.2006.03028.x.

Zigone D., Ben-Zion Y., Campillo M. and Roux P.; 2015: *Seismic Tomography of the Southern California Plate Boundary Region from Noise-Based Rayleigh and Love Waves.* Pure Appl. Geophys. 172 (2015), 1007–1032, doi 10.1007/s00024-014-0872-1.