Evaluation of Ground Motion and Optimum Seismic Monitoring at an Inner-City Deep Geothermal Power Plant

Sabrina Keil

SIGILL·UNIVERSITAT·LUDOVIC·MAXIM

München 2023

Evaluation of Ground Motion and Optimum Seismic Monitoring at an Inner-City Deep Geothermal Power Plant

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Geowissenschaften
der Ludwig-Maximilians-Universität München

vorgelegt von
Sabrina Keil

München, den 29.03.2023

Erstgutachter: Prof. Dr. Heiner Igel
Zweitgutachterin: Prof. Dr. Eva Eibl
Tag der mündlichen Prüfung: 11.07.2023

Abstract

Experience shows that even non-pressure-stimulated geothermal power plants can generate induced seismicity even in low seismic hazard settings. Due to the shallow hypocentral depth of a few kilometers, already relatively small earthquakes can lead to ground shaking, which is felt by the population. As most of the geothermal power plants are located within or close to densely inhabited areas, monitoring and estimating the maximum ground motion and its spatial distribution are of crucial importance for the authorities and the public. However, this is challenging as in most cases the monitoring network is weak, due to high noise levels and limited station numbers, and only few information about local site effects are available. Therefore, several open questions for the development of proper risk governance strategies remain: How does the shallow subsurface structure look like and is there the risk of seismic wave amplification? How large is the expected ground motion and what areas are affected? How can the seismic monitoring within urban environments be improved?

These questions are going to be answered by the example of Munich, Germany, where Europe's largest inner-city geothermal project was carried out with a total of six deep wells, which increases the risk of induced seismicity in the area.

To answer the question of local site effects a microzonation study is conducted in Munich´s inner city. The common approach of array measurements is challenging in urban environments due to the space requirements. Therefore, the recently developed approach of single-station six-component (6C) measurements is applied, combining three translational and three rotational motions. This new method is much simpler in terms of logistics and therefore allows an easier and faster estimation of the local velocity structure.

Another problem that is encountered in microzonation studies is the existing ambiguity of the inversion results. Since conventional inversion methods suffer from different limitations, a machine learning algorithm is trained, which selects the appropriate number of subsurface layers and returns a complete probability distribution of the S-wave velocity structure.

The shallow velocity structure is important, as it can amplify seismic waves. However, the maximum induced ground motion is often not recorded because the station network is sparse at most geothermal sites. A dense network is neither feasible nor required by (Bavarian) law, which impedes a spatial interpretation of the ground motion after an earthquake. Because of that, large uncertainties remain in the determination of affected areas according to the German mining laws. This is a major problem when it comes to compensating for damage claims brought forward by local population and companies. I show that 3D numerical simulations are able to close observational gaps and can be used to estimate the maximum ground motion and its spatial distribution. This way public authorities can make fast and precise decisions in case of damaging events.

Not only the sparse number of stations, but also the high noise levels pose a big problem for microseismic monitoring within urban environments, which make it difficult to reach the required magnitude threshold and location accuracy. Therefore, a network optimization method is applied, which calculates the optimum number and location of seismic stations even in environments with heterogeneous noise conditions.

The dissertation results have implications for future inner-city geothermal projects, as they will facilitate the seismic risk assessment during the planning stage, the seismic monitoring during operation and the evaluation of shaking effects after an event.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Geothermal systems are a promising source of energy that is clean and provides a sustainable baseload for heat and electricity. However, induced seismicity is an inevitable and well-known, yet poorly understood by-product of this technology and has caused serious public concern leading to the shutdown of several geothermal projects in the past. One example is the Deep Heat Mining Project in Basel, Switzerland, where several M_L 3.0 events occurred following fluid injection, which led the operator´s insurance pay out property damages, attributed to the induced earthquakes, of about 7 million CHF (Häring et al. 2008, Baisch et al. 2009). The negative public reaction and the subsequent update of the risk assessment led to the cancellation of the project (Baisch et al. 2009). Another prominent example is the 2017 M_W 5.4 earthquake in Pohang, South Korea, where strong evidence indicates that geothermal operations triggered the event (Kim et al. 2018, Ellsworth et al. 2019). This earthquake was the second largest instrumentally recorded earthquake along southeast Korea, causing property damage of about US$ 75.8 million and injuring more than 90 people (Naik et al. 2020). Although these are examples for enhanced geothermal systems, induced seismicity was also observed at non-pressure stimulated geothermal power plants. In the Molasse Basin induced events of M_L 2.4 and M_L 2.1 were recorded in Unterhaching (Megies & Wassermann 2014) and Poing (Seithel et al. 2019). Due to the shallow hypocentral depth of a few kilometers and the proximity to urbanized areas, even such relatively small earthquakes can lead to ground shaking, which is felt by the population. This can severely reduce the level of acceptance of residents close to the plant. Managing the induced seismicity risk is therefore crucial for the development and further exploitation of geothermal technology for power and heat supply in urban environments. Hereby, the prediction of the ground motion already during the planning stages of a geothermal project, as well as a good seismic monitoring system are essential tools to mitigate earthquake damage.

The ground motion of an earthquake can be strongly influenced by local site effects, where low-velocity sediments have a large amplification potential (Field et al. 1997, Hartzell et al. 2001). Therefore, the shallow 1D velocity structure is estimated as part of microzonation studies typically by the inversion of surface wave dispersion curves, which are extracted from ambient vibration recordings (e.g. Fäh et al. 2008, Scherbaum et al. 2003). Several strategies exist, including array methods, such as frequency-wavenumber analysis (Capon 1969) and spatial autocorrelation (Aki 1957), as well as single-station approaches, such as horizontal-to-vertical (H/V) spectral ratios (Nogoshi & Igarashi 1971). These methods are already commonly applied, however, they suffer from different shortcomings. An array set-up can be quite challenging as it requires a lot of space, while the H/V method can give ambiguous results. This raises the requirement for a new single-station method for the estimation of the shallow velocity structure specifically in urban areas.

In addition, the choice of inversion process can strongly influence the resulting velocity profile. Classic inversion methods might be computationally demanding, become trapped in local minima or depend on the choice of starting model and tuning parameters (e.g. Gosselin et al. 2022, Tarantola 2005). Therefore, machine learning methods have been proposed to overcome these limitations (e.g. Chen et al. 2022, Luo et al. 2022). However, these studies implement classical neural network structures, which only provide deterministic velocity estimates and/or constrain either velocity or layer thickness to a fixed value. A fast and accurate inversion method, which outputs fully probabilistic solutions without any velocity or depth constraints is required.

Once the shallow velocity structure is computed it can be implemented in seismic simulations in order to estimate the maximum ground motion and its distribution in case of an earthquake. This is useful to estimate the seismic risk already during the planning stages of a geothermal project, as well as evaluate the shaking effects after an event. Since in most cases the station network is sparse, a spatial interpretation of the ground motion is not possible and the maximum ground motion is often not recorded. Therefore, numerical simulations should be used to supplement recorded and macroseismic data in order to identify affected areas.

The sparse station network in combination with high noise levels within urban areas make microseismic monitoring challenging. This is a large problem since well-designed monitoring networks are fundamental to improve the detection of weak seismic signals to obtain accurate locations, magnitudes and source parameters, both for natural and induced seismicity. Hereby, the quality of a monitoring network depends on several factors, such as sensor type, number of stations and network geometry (e.g. Trnkoczy et al. 2009). Nevertheless, network planning is still mainly performed as a manual task based on simple design rules, which may fail in complex settings. Therefore, this problem should be solved computationally in the framework of D-optimal network design (e.g. Steinberg & Rabinowitz 2003, Kijko 1977).

In this thesis the aforementioned challenges are tackled on the example of the greater Munich area, which is situated in the North Alpine Foreland Basin, where several geothermal power plants exploit a deep hydrothermal reservoir. This includes the geothermal project at Schäftlarnstraße (SLS) located in Munich´s city center, which is up to date Europe´s largest inner-city deep geothermal power plant with a total of six deep wells (Lentsch & Schweingruber 2022). Induced seismicity was observed at surrounding geothermal power plants with magnitudes up to 2.4 (Megies & Wassermann 2014, Seithel et al. 2019), which raises concern about the seismic risk within the densely inhabited city center. Therefore, the ground motion of induced earthquakes in the Munich area is evaluated and strategies for optimizing the seismic monitoring network are presented in the following.

This thesis is structured in three main chapters which deal with the topics of seismic microzonation, ground motion simulations and network optimization. Chapter 2 is subdivided into three parts and consists of three published papers. The first paper "Single-station Seismic Microzonation using 6C Measurements" is published in Journal of Seismology (Keil et al. 2021) and presents the first field test of the blueSeis-3A rotational motion sensor for microzonation purposes. Estimated velocity profiles from this novel six-component method are shown for two locations in Munich, Germany, and directly compared to borehole-derived lithologic profiles.

The second paper "Comparing Single-station 6C Measurements and Array Measurements for Seismic Microzonation in Munich, Germany" has been published in Geophysical Journal International (Keil et al. 2022a) and presents a direct comparison of the novel six-component measurements and established array methods for the estimation of the local 1D velocity structure. For this purpose, ambient noise measurements are conducted in Munich, Germany, using a rotational motion sensor in combination with a seismometer and a geophone array. The advantages and disadvantages of the different methods are discussed.

The third paper "Surface Wave Dispersion Curve Inversion using Mixture Density Networks" published in Geophysical Journal International (Keil & Wassermann 2023) presents a machine learning approach for the inversion of surface wave dispersion curves. Hereby, the shallow 1D velocity structure is predicted in a two-step approach, where 1) a regular neural network classifies the optimum number of layers within the upper 100 m of the subsurface and 2) a mixture density network outputs the layer depth estimates together with a fully probabilistic solution of the S-wave velocity structure. The machine learning approach is applied to recorded noise data in Munich and compared to a neighborhood algorithm and Markov chain Monte Carlo inversion.

Chapter 3 contains the paper "Estimation of Ground Motion due to Induced Seismicity at a Geothermal Power Plant near Munich, Germany, using Numerical Simulations" published in Geothermics (Keil et al. 2022b). The shallow velocity structure estimated in chapter 2 is implemented with additional geophysical and geological data into a 3D subsurface model and is used for seismic simulations. The paper presents simulation results of two induced events that occurred at the geothermal power plant in Poing, Germany, which is located 15 km east of Munich. The simulated waveforms are compared to recorded data and complete shakemaps are computed.

Chapter 4 contains the paper "Optimal Network Design for Microseismic Monitoring in Urban Areas - A Case Study in Munich, Germany" submitted to Seismica and presents a network optimization approach based on simulated annealing in order to find the optimum number and distribution of seismic stations in Munich. First of all, a detailed noise model of the Munich area is constructed in order to capture the heterogeneous noise conditions within the city. This noise model is implemented into the optimization program to compute signal-to-noise ratios at the possible new station locations. After that, the installation of surface and borehole stations is tested to reach the recommended magnitude threshold and location accuracy of the induced events.

Chapter 2

Seismic Microzonation

2.1 Single-station seismic microzonation using 6C measurements

by Sabrina Keil, Joachim Wassermann and Heiner Igel

Published in Journal of Seismology 25.1 (2021): 103-114

Abstract

Microzonation is one of the essential tools in seismology to mitigate earthquake damage by estimating the near surface velocity structure and developing land usage plans and intelligent building design. The number of microzonation studies increased in the last few years as induced seismicity becomes more relevant, even in low risk areas. While of vital importance, especially in densely populated cities, most of the traditional techniques suffer from different shortcomings. The microzonation technique presented here tries to reduce the existing ambiguity of the inversion results by the combination of single-station six-component (6C) measurements, including three translational and three rotational motions, and more traditional H/V techniques. By applying this new technique to a microzonation study in the downtown area of Munich (Germany) using an iXblue blueSeis-3A rotational motion sensor together with a Nanometrics Trillium Compact seismometer, we were able to estimate Love and Rayleigh wave dispersion curves. These curves together with H/V spectral ratios are then inverted to obtain P- and S-wave velocity profiles of the upper 100 m. In addition, there is a good correlation between the estimated velocity models and borehole-derived lithology, indicating the potential of this single-station microzonation approach.

2.1.1 Introduction

In seismic microzonation the velocity structure of the upper few 100 m is estimated in order to characetrize the local earthquake shaking characteristics. There are two common methods that allow the determination of 1D subsurface wave velocity structure: 1) Array-based methods, such as Spatial AutoCorrelation (SPAC) (Aki 1957) and Frequency-Wavenumber (FK) analysis (Capon 1969); 2) single-station approaches, including Horizontal-to-Vertical (H/V) spectral ratios (Nogoshi & Igarashi 1971). In general, array-based methods are well understood and give reliable results (e.g., Marano & Fäh (2014)). However, as a severe limitation, the installation and maintenance in an urban area is very complex. Due to its simplicity the single-station H/V method is commonly applied in microzonation studies, but its theoretical foundation is still not completely understood and the results highly depend on the quality of the noise (e.g., Malischewsky & Scherbaum (2004)).

Wassermann et al. (2016) demonstrated that a single-station six-component approach, combining rotational motion measurements (which are related to the gradient of a seismic wave field) with traditional translational recordings (i.e., recordings of ground velocity) may give comparable results to array techniques for the estimation of the 1D local velocity structure and the dominant direction of the incident wave field.

In addition, Bernauer et al. (2018) showed that a portable and reliable broadband rotational sensor, the blueSeis-3A (iXblue 2017), is now available. In order to test the performance of this six-component approach in a real world application, we conduct an experiment using a Trillium Compact seismometer and the blueSeis-3A rotational sensor within Munich, where the largest inner-city geothermal power plant gives rise to concern about induced seismicity in a densely inhabited area. Using noise measurements, the relation between rotational and translational motion is exploited to estimate Love and Rayleigh wave dispersion curves, which can then be inverted for the local 1D velocity structure (Wassermann et al. 2016). To get as much information on the subsurface as possible, the H/V method is used to complement the data in the lower frequency range (< 5 Hz). Finally, the inverted velocity models are compared to lithologic profiles, derived during the GeoPot (Geo-potentials of the tertiary subsurface) project (Technical-University-Munich 2016-2019) of the Technical University of Munich (TUM), to identify any correlations between wave velocity and geology.

2.1.2 Data acquisition

The six-component measurements require the simultaneous recording of translational and rotational motion. In this study two instruments are used, a Nanometrics Trillium Compact 120s seismometer and an iXblue blueSeis-3A rotational motion sensor (iXblue, 2017). In order to record the same movement, the seismometer and the rotational motion instrument have to be installed on a single rigid base. Nevertheless, there also exist six-component instruments, so-called Rotaphones, which measure three rotational and three translational components in a fixed frame (Brokešová et al. (2012)). The rotational seismometer blueSeis-3A (iXblue, 2017), is based on an interferometric fibre-optic gyroscope (FOG), which is strictly insensitive to translational motions (Bernauer et al. 2018). A more detailed description of the working principle of FOGs is given in Lefevre (2014). Bernauer et al. (2018) conducted several laboratory tests in order to determine the performance of the blueSeis-3A. They found that the sensor has a flat self-noise level lower than $30nrads^{-1}Hz^{-1/2}$ over a wide frequency range of 0.001-50 Hz, as can be seen in Fig. 2.1. In addition, the sensor is very stable in changing ambient conditions, such as temperature and magnetic field, which makes it suitable for field installations (Bernauer et al. 2018).

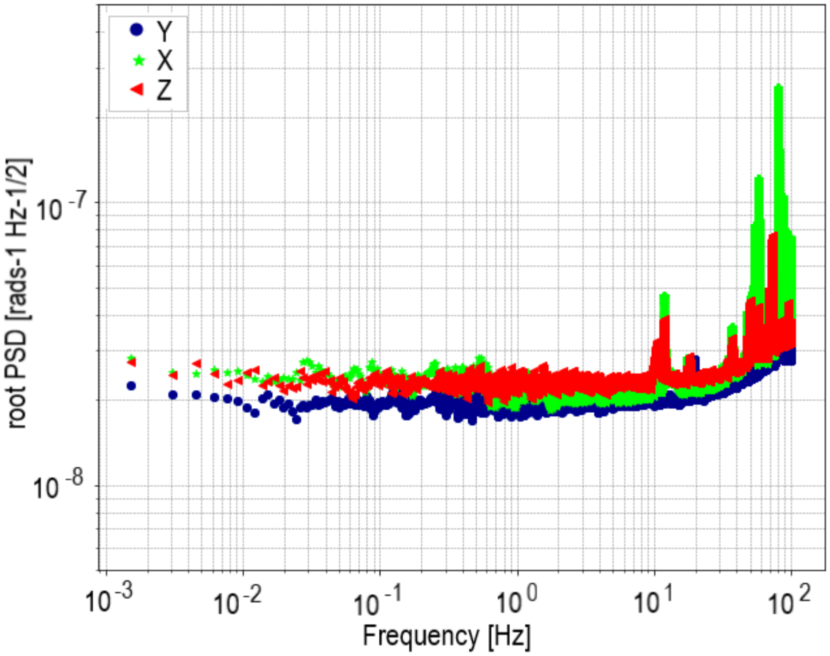

Figure 2.1: Self-noise power spectra of the three components of the blueSeis-3A (dots, triangles, and stars) determined during a laboratory experiment at the Geophysical Observatory in Fürstenfeldbruck, Germany. The sensor was placed on an auxiliary monument in a quiet location. All three components show flat self-noise levels over a frequency range of 0.001 Hz to 50 Hz. The X-component exhibits significant peaks at 58 and 83 Hz, and the Z-component at 74 Hz, probably due to seismic ambient signals (cf. Bernauer et al. (2018)).

The instrument set-up used in this study is shown in Fig. 2.2. For the installation of the Trillium Compact seismometer and the blueSeis-3A rotational sensor, a thin layer of the top soil has to be removed in order to establish better coupling to the ground. The two instruments are then placed next to each other on a concrete slab to guarantee a stable position and are connected to power and GPS. The seismometer is additionally connected to a REFTEK digitizer. The distance between the two instruments is about 10cm, which is negligible compared to the smallest wavelength of a few meters. Therefore, the set-up can be considered a single-station measurement, where both instruments record the same movement. To shield the devices against wind and direct sun radiation, a styrofoam insulation box is placed over them. Ambient noise is recorded during daytime with sampling rates of 200 Hz for approximately two hours, which is enough to get a good representation of the wavefield in the recorded frequency range. In addition, increasing the measurement time to several days would make installation and maintanance more complicated and thus reduce the impact of the method. The desired frequency band for the estimation of the velocity profiles lies between 1-20 Hz, which corresponds to urban noise, as was shown by numerous authors (e.g. Asten (1984), Gutenberg (1958)). Measurements are performed at several locations in the city of Munich in the vicinity of the landing points of the geothermal wells, since these are the most likely regions for the nucleation of induced earthquakes. The study area is marked in Fig. 2.3.

Figure 2.2: Set-up for the single-station six-component noise measurements. The two instruments, the blueSeis-3A rotational sensor on the left and the Trillium Compact 120s seismometer on the right, are placed about 10cm apart on a concrete slab. The instruments are connected to GPS, power and the seismometer to a REFTEK digitizer (inside of metal box). Before starting the recording of the data, a styrofoam insulation box is placed over the instruments

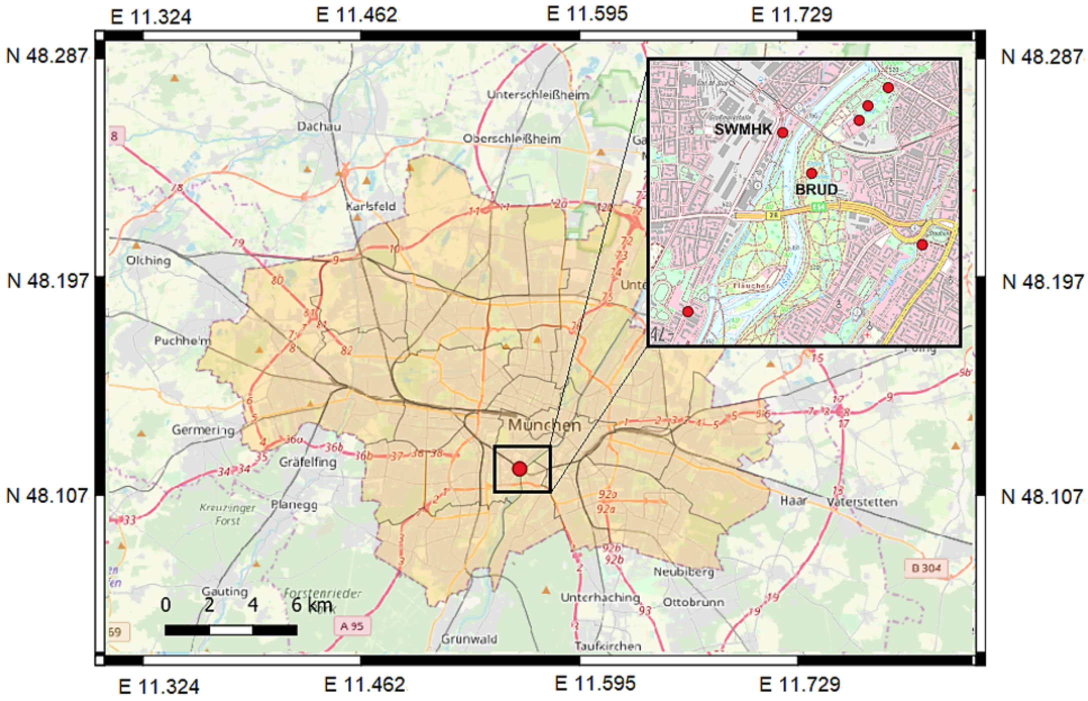

Figure 2.3: Map of Munich. The location of the geothermal power plant is marked by the dot. The study area is indicated by the rectangle. The magnified map in the upper right corner shows a part of the study area. The dots represent the measurement locations, including stations SWMHK and BRUD, which are discussed in more detail. Station SWMHK is located at the geothermal power plant.

2.1.3 Methods

2.1.3.1 Love and Rayleigh wave dispersion estimation

Six-component measurements provide a new way of computing Love and Rayleigh wave dispersion curves. Under the premise that we only have to deal with plane, fundamental mode surface waves, it was shown by several authors (e.g. Ferreira & Igel (2009), Kurrle et al. (2010)) that simple relations between the rotational motion and the translational acceleration of a seismic signal exist. In case of a fundamental mode Love wave the relation is:

$$c_L(f) = -\frac{a_T(f)}{2\dot{\omega}_Z(f)} \tag{2.1}$$

where $c_L(f)$ represents the frequency-dependent phase velocity, a_T the transversal component of acceleration, and $\dot{\omega_Z}$ the vertical rotation rate.

The transversal acceleration can be further defined as:

$$a_T(f) = sin(\phi_L)a_N(f) - cos(\phi_L)a_E(f) \tag{2.2}$$

where ϕ_L is the back azimuth of the wavefield, a_N the N-S component of acceleration, and a_E the E-W component, respectively. This allows equation (1) to be rewritten as:

$$2c_L(f)\dot{\omega}_Z(f) = sin(\phi_L)a_N(f) - cos(\phi_L)a_E(f) \tag{2.3}$$

The phase velocity $c_L(f)$ and the back azimuth ϕ_L are both unknown properties, which have to be estimated in the following steps.

Accordingly, the relation for the phase velocity of fundamental mode Rayleigh waves can be derived by taking the ratio of the vertical component of acceleration a_Z and the transverse rotation rate $\dot{\omega}_T$ (e.g. Lin et al. (2011)).

$$c_R(f) = \frac{a_Z(f)}{\dot{\omega}_T(f)} \tag{2.4}$$

By substituting $\dot{\omega}_T$ with its N-S and E-W rotation rate components ($\dot{\omega}_N$ and $\dot{\omega}_E$), equation (4) can be rewritten as:

$$\frac{a_Z(f)}{c_R(f)} = sin(\phi_R)\dot{\omega}_N(f) - cos(\phi_R)\dot{\omega}_E(f) \tag{2.5}$$

To solve these equations and estimate the Love and Rayleigh wave dispersion curves, an updated version of the python package ROLODE (ROtational LOve Dispersion Estimation; Wassermann et al. (2016)) is used. This program simultaneously estimates the direction and the velocity employing the principle of orthogonal distance regression (ODR). To fulfil the additional assumption of a single source active at a time, the data are analysed at each frequency band in short sliding time windows. Each time window gives an estimate of phase velocity and back azimuth. A kernel density estimation (kde) is used to bin the estimates in error weighted histograms and model these histograms with Gaussian functions. The quality of the fit can be improved by introducing a weighting scheme to the computation of the histograms, which accounts for the goodness of the straight line fit by the ODR (i.e. the correlation between the two quantities). The estimated phase velocities from the ODR are weighted according to:

$$w^x_{norm} = (1 - \frac{1}{w})^x \tag{2.6}$$

where

$$w = \frac{\sum_i^I (\dot{\omega}_Z(f)[i])^2}{\sum_i^i (\epsilon_i^2 + \delta_i^2)} \tag{2.7}$$

with I the number of processed time windows, $x > 0$, and δ and ϵ the errors of the dependent and observed value in the ODR.

The mode of the kde-function and its variance are then used to determine the phase velocity and the error. From the velocity estimates at each frequency band the dispersion curve is derived. The procedure is described in more detail in Wassermann et al. (2016).

2.1.3.2 Horizontal-to-Vertical spectral ratios

The method of Horizontal-to-Vertical spectral ratios (H/V) was first introduced by Nogoshi and Igarashi (1971), who described it as the ratio between the Fourier amplitude spectra of the horizontal and the vertical component of microtremors. Several more recent studies interpret the H/V ratio as the ellipticity χ of Rayleigh waves, which can be computed by dividing the horizontal over the vertical component of particle motion (e.g. Malischewsky & Scherbaum (2004)).

$$\chi = \mid \frac{H}{V} \mid \tag{2.8}$$

Numerous authors (e.g. Sylvette et al. (2006), Malischewsky & Scherbaum (2004)) have shown that the H/V ratio exhibits a pronounced peak close to the fundamental S-wave resonance frequency of the site, when the surface layer exhibits a sharp impedance contrast with the underlying stiffer formations. This makes the ellipticity an important parameter to reflect properties of the underground structure (Sylvette et al. 2006) and gives additional data, especially in the lower frequency band.

The H/V curves, computed with the Geophysical Signal Database for Noise Array Processing (GEOPSY) software package (Wathelet et al. 2004, Wathelet 2008), are used to complement the dispersion curves for the surface wave inversion. The 1D velocity profiles are then derived with the DINVER module in the GEOPSY software package, which implements a neighbourhood algorithm (Wathelet 2008).

2.1.4 Results and discussion

The power spectral densities (PSD) of the recorded rotational and translational data for station SWMHK, which is located right next to the geothermal power plant (Fig. 2.3), are presented in Fig. 2.4. It is easy to notice that the PSD for the translational components is about two orders of magnitude larger than for the rotational components. Additionally, an energy decrease towards lower frequencies can be observed in both cases. At about 5 Hz the self-noise level of the rotational sensor is reached, which lies at $30 nrads^{-1}Hz^{-1/2}$ (Bernauer et al. 2018), explaining the flat trend of the curve in Fig. 2.4a).

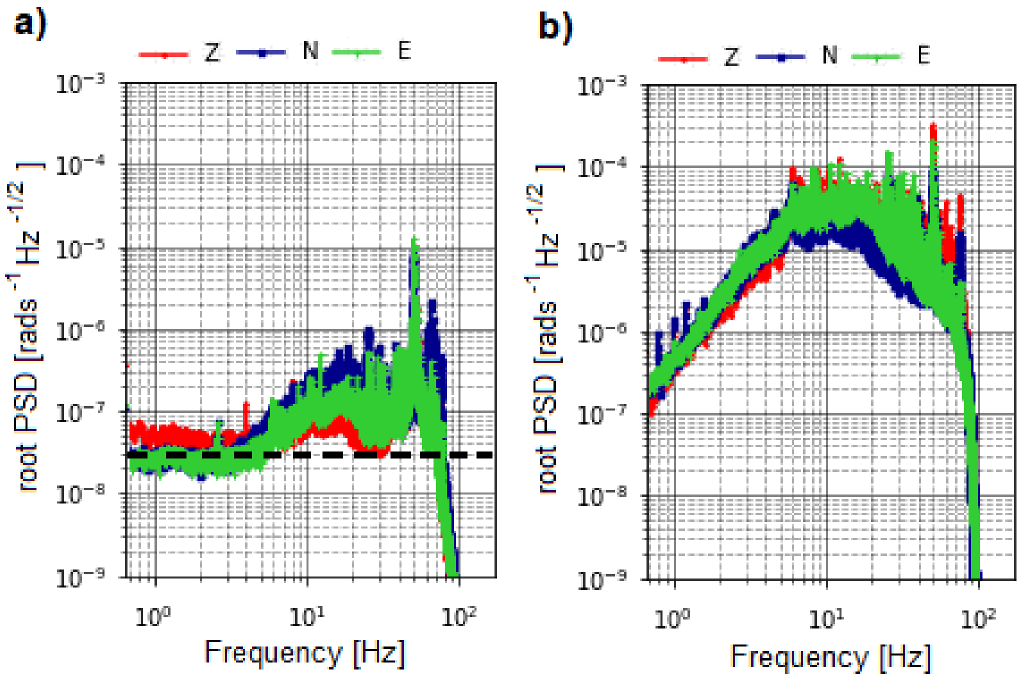

Figure 2.4: Power spectral density of the three components of the a) blueSeis-3A rotational sensor and b) Trillium Compact seismometer for station SWMHK. The dashed line in a) marks the self-noise level of the rotational sensor, which lies at about $30nrads^{-1}Hz^{-1/2}$

Applying the ROLODE method to the recorded data, the phase velocities for each frequency band can be estimated and are arranged in error weighted histograms. Fig. 2.5 displays the distribution for four different frequencies. The data are modelled with Gaussian basis functions, where the mode is marked by a vertical dotted line. For each frequency band under consideration, using the mode as phase velocity estimate and the corresponding standard deviation of the kde-function as the associated error, a complete dispersion curve can be derived. Fig. 2.6 shows the resulting Love and Rayleigh wave dispersion curves for station SWMHK. Between 5-20 Hz the data show a normally dispersive trend in which the phase velocity increases with decreasing frequency. However, below 5 Hz the curves drop, indicating a limitation of this method in the lower frequency range. This is observed in the dispersion curves of all the measurements.

This limitation in the lower frequency range can be explained by re-visiting the PSD of the rotational sensor in Fig. 2.4a). At about 5 Hz the self-noise level of the blueSeis-3A is reached and therefore no rotations can be recorded below that. This causes the drop in the estimated dispersion curves. Reasons could include either the lack of these lower frequencies in the noise spectrum of the city and/or the rotations are too small to be recorded by the rotational sensor. In this context one has to keep in mind that the rotational motion amplitudes are equal to the ground acceleration values scaled by the ground velocity (cf. Eq. 2.1). Exploiting this relation a rough estimate of the expected rotation rates can be calculated (Fig. 2.7). Herefore, a forward computation for the Love wave dispersion curve using the derived velocity profile at station SWMHK (Fig. 2.10) was performed. The phase velocity values together with the actual recorded translational data at this station were inserted into Eq. 2.1 to obtain the theoretical rotation rates. It can be seen that the rotation rates below 5 Hz are in the order of 10^{-8} to 10^{-9} and are therefore too small given the self noise level of the sensor.

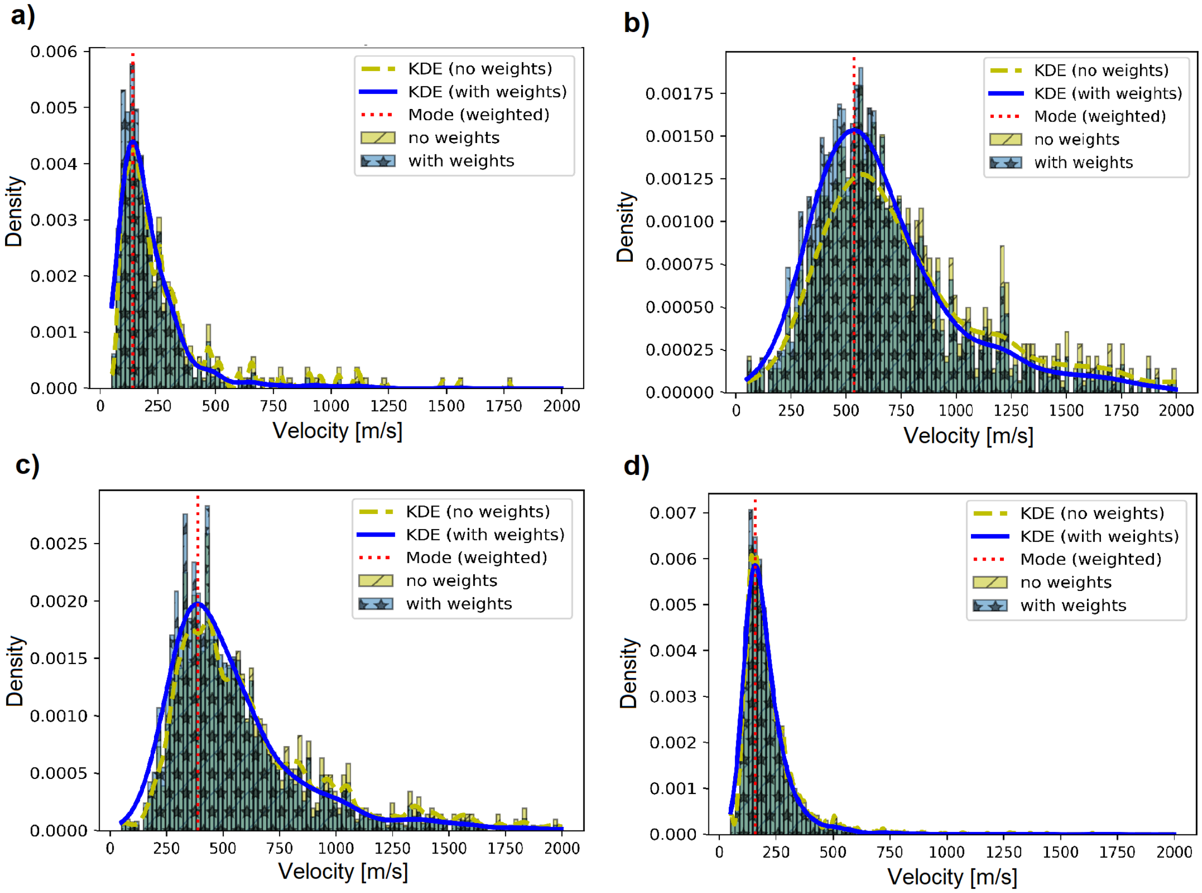

Figure 2.5: Histograms of estimated Love wave phase velocities using the SciPy ODR package for selected frequency bands a) 1.79 Hz, b) 5.08 Hz, c) 7.18 Hz, and d) 14.35 Hz at station SWMHK. The bright bars represent the histogram of the non-weighted velocity estimates and the darker bars indicate the weighted estimates. The dashed and solid curves give the best fitting Gaussian. The mode of the weighted distribution is shown as vertical dotted line.

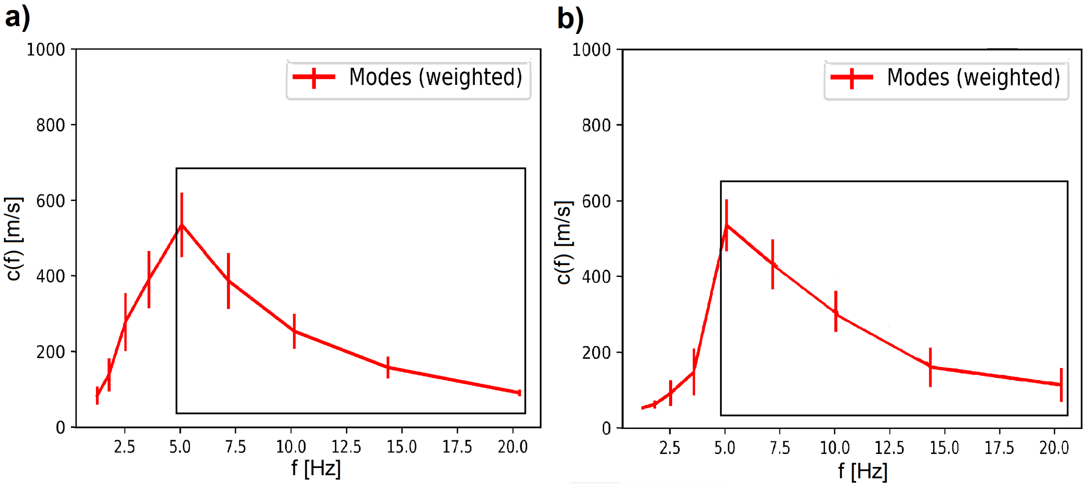

Figure 2.6: Estimated a) Love and b) Rayleigh wave dispersion curves using the ROLODE method. The error bars represent the standard deviations of each frequency band. The rectangular boxes give the frequency region where the inversion is applied.

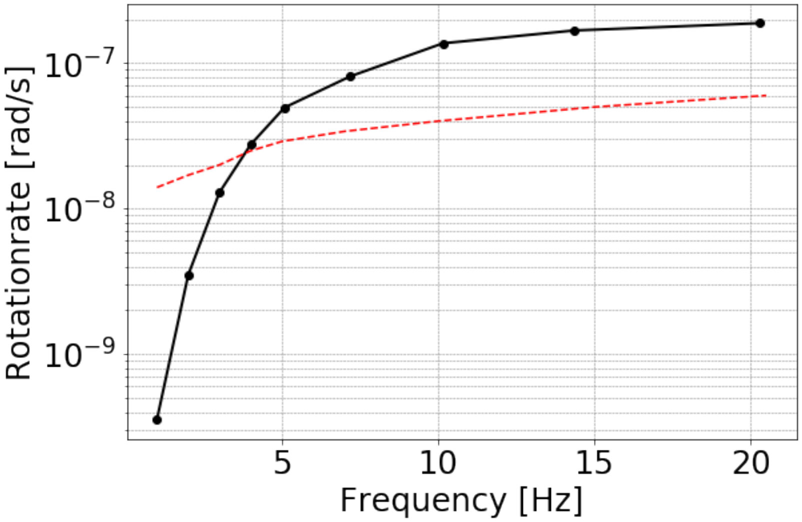

Figure 2.7: Expected rotation rates at station SWMHK calculated from a forward modelled Love wave dispersion curve and the measured translational data using Eq. 2.1. The dashed line marks the smallest rotation rates that can be recorded by the sensor. These values were extracted from an operating range diagram (ORD)

To complement the data in the lower frequency range and to enable an inversion to greater depth, the H/V curves are computed from the three translational components using the GEOPSY (Wathelet et al. 2004, Wathelet 2008) software package. The program divides the data into small time windows, for which the H/V ratio is computed separately. For the windows selection an anti-trigger algorithm is implemented with the objective to keep the most stationary parts of ambient vibrations and to avoid the transients, as explained in more detail by Bard et al. (2008). As a final step, an average over all single H/V ratios is computed. Fig. 2.8 shows the ellipticity curve for station SWMHK, which exhibits a pronounced peak at about 2.5 Hz.

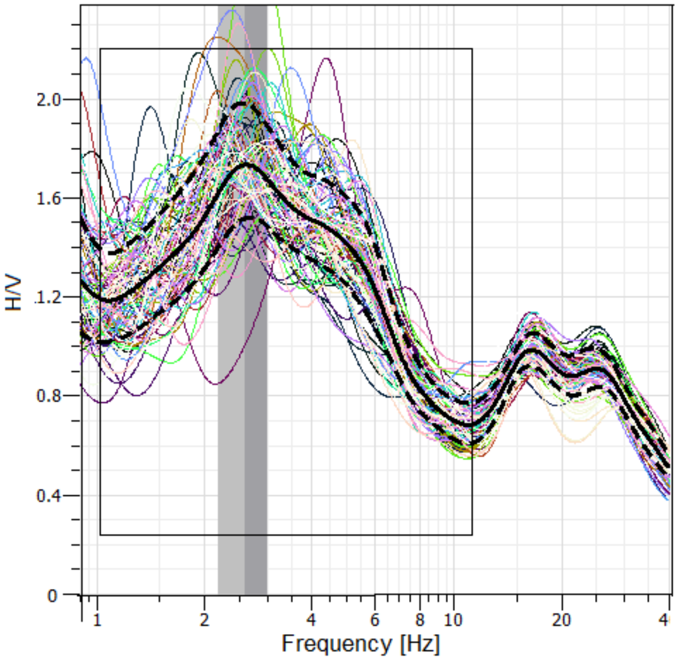

Figure 2.8: Ellipticity curve for station SWMHK computed with GEOPSY (Wathelet et al. 2004, Wathelet 2008). Each individual thin curve represents the computed H/V ratio for a single selected time window. The solid black curve indicates the geometrically averaged ellipticity curve over all individual H/V ratios. The two dashed lines represent the H/V standard deviation. The grey vertical bars mark the estimated ellipticity peak and its error. The rectangular box gives the frequency region where the inversion is applied.

For the following joint inversion of a 1D ground velocity profile, the appropriate frequency range of the input data has to be selected, marked by the black boxes in Fig. 2.6 and 2.8. For the inversion at station SWMHK the input consists of the H/V ellipticity estimates between 1-10 Hz, as well as the derived Love and Rayleigh wave dispersion curves. In case of the dispersion curves, normal dispersion is assumed, restricting us to use the data in the frequency range of 5-20 Hz. In the next step the number of layers to be inverted for has to be defined. The neighbourhood algorithm of the program then generates different ground models and computes the corresponding dispersion and H/V curves for each of those models. The comparison of the computation results with the measured dispersion and ellipticity curves provides a misfit value that indicates how far the generated model is from the data fit (Wathelet et al. 2004). In general, a velocity model with a low misfit value is desired, however, over-fitting of the data has to be avoided. Increasing the number of parameters for the inversion most likely decreases the misfit, because of the higher degree of freedom (DOF). Several tests with a different number of layers showed that the most significant reduction in the misfit can be achieved when increasing the number of model layers from 2 to 3. Because of that and in order to avoid over-fitting of the data, a three layer model for the inversion was chosen.

As part of the inversion, the linkage between the different free parameters (v_p, v_s, density and Poisson´s ratio) has to be chosen. While Love waves consist of multiple reflected SH waves only, Rayleigh waves are a combination of P and SV waves. Because of that, the analysis of these surface waves provides more information about the S- rather than the P-wave. Therefore, the P-wave velocity model was linked to the S-wave model during the inversion by Poisson´s ratios between 0.2 - 0.5, which includes the typical values for soil and sedimentary rocks (Gercek 2007).

The results for a three-layer inversion at station SWMHK, using only Love and Rayleigh wave dispersion curves as an input, are shown in Fig. 2.9. Due to the limitation of rotational motion in the lower frequency range, the inversion results are only sensitive to about 30 m depth. The velocity steps in the S-wave profile are well constrained, with the upper one at about 4 m depth and the second one at 14-16 m. Compared to this, the P-

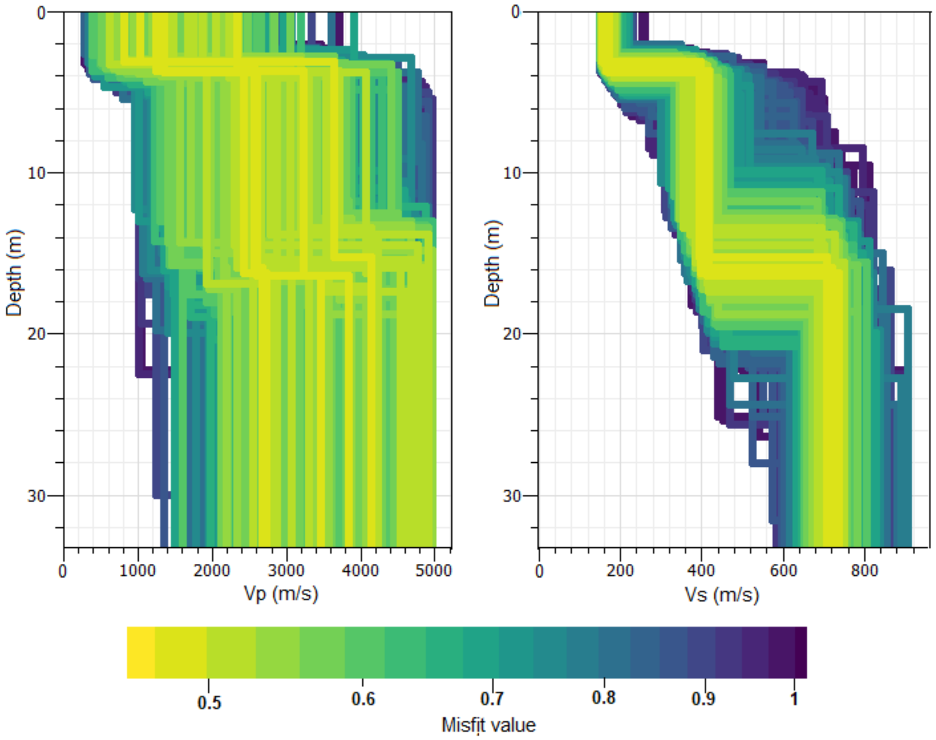

Figure 2.9: P- and S-wave velocity profiles for a three-layer inversion using DINVER of GEOPSY (Wathelet 2008). Only Love and Rayleigh wave dispersion curves were used as an input. The shading gives the misfit of the computed models.

wave velocities are widely spread, even though the P-wave model was linked to the S-wave model. The reason for this is the large variance of the other free parameters due to the fact that there are less information on the P-wave.
In order to get more information about deeper structures, the dispersion curves are complemented with the H/V ratios, which provide data to a lower frequency range. All input data are inverted with a weight of 1. Fig. 2.10a) gives the preferred P- and S-wave velocity model at station SWMHK for a three-layer inversion. There are two apparent velocity steps visible. Both velocity steps are shifted to greater depth, compared to the model in Fig. 2.9. Furthermore, the resolution of the P-wave profile increased due to the additional information from the Rayleigh wave ellipticity curve. Adding more layers to the inversion results in thin upper layers, however, the main velocity contrasts remain at the same depth range.
In the next step the velocity profiles are compared to lithologic profiles to identify any correlation between the wave velocity and the geology. The lithologic profiles presented in Fig. 2.10b) and Fig. 2.11b) were extracted at the exact measurement locations from the

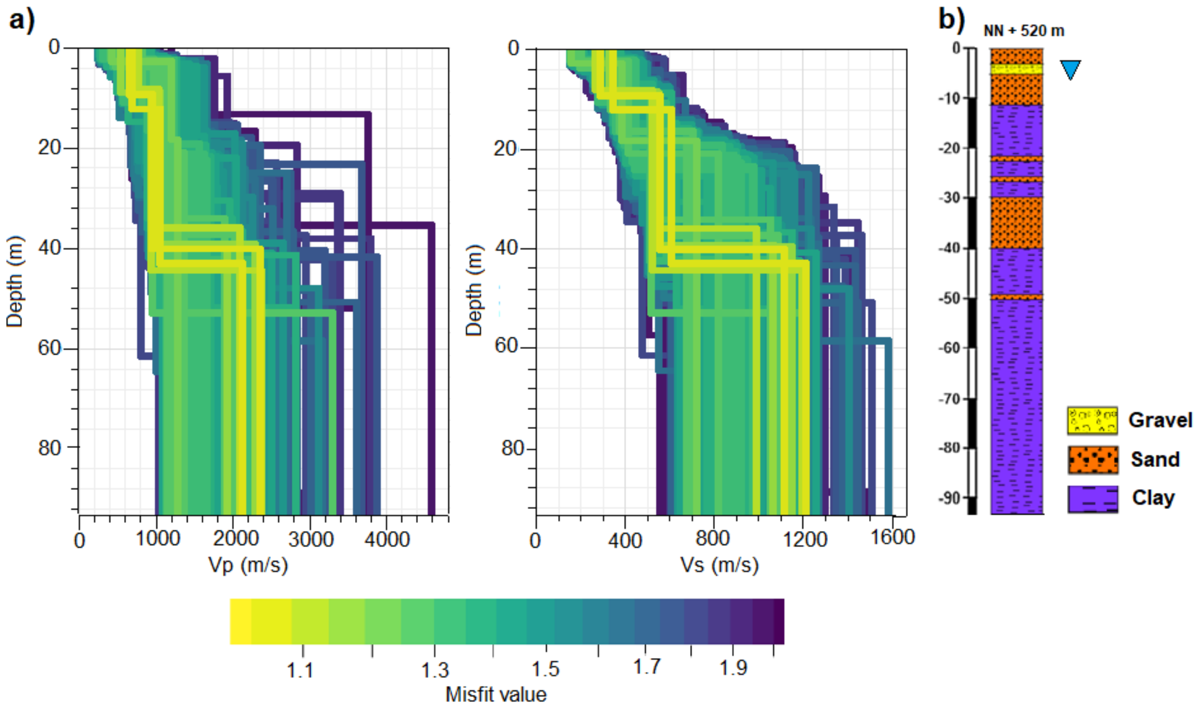

Figure 2.10: a) Resulting three-layer P- and S-wave velocity profiles at station SWMHK using DINVER of GEOPSY (Wathelet 2008). For the inversion Love and Rayleigh wave dispersion curves, together with H/V ratios, were used with a weight of 1. The shading gives the misfit of the computed models. b) Lithologic profile estimated from borehole data during the GeoPot project of TUM (Technical-University-Munich 2016-2019). The groundwater table is marked by the triangle.

geologic 3D model of Munich, derived during the GeoPot project (Technical-University-Munich 2016-2019). The 3D model was constructed through interpolation of borehole data. Therefore, the lithologic profiles presented here are interpolated and could slightly deviate from the real structure. However, we assume that the deviation is small, since the borehole data are very dense within Munich. The distance between the closest boreholes and the stations SWMHK and BRUD is less than 100 m. Comparing the velocity profile of station SWMHK with the geology (Fig. 2.10), it is apparent that the upper velocity step at 8-10 m depth coincides with the change in lithology from sand to clay. In addition, the groundwater table also occurs at this depth range, which could influence the wave velocity. The second velocity contrast at 35-45 m depth coincides with a 10 m thick sandstone layer. This indicates a correlation between the change in lithology and the increase in wave speed, while the very thin sandstone layers cannot be detected because of the decreasing resolution with depth.

Similar correlations can be found for the other stations. As a different example, the inversion results of station BRUD are shown in Fig. 2.11. Again, the P- and S-wave velocity profiles for the three-layer inversion compared to the 100 m deep lithologic profile estimated during the GeoPot project is displayed. The upper velocity increase occurs at 4 - 6 m depth, which coincides with the groundwater level at 5 m depth and the transition from gravel to sand. Also for location BRUD, as it is for station SWMHK, the velocity contrast might either reflect the presence of groundwater and/or the change in lithology. The second increase in velocity at 25 - 30 m depth lies in the range of the upper edge of the sandstone layer at 30 m depth. Therefore, the material change from clay to sand influences the wave speed. The sandstone lens at 50 m depth cannot be detected even when the number of layers in the inversion is increased, because it is too thin for the resolution at this depth.

2.1.5 Conclusion

The objective of this study was to test a new single-station technique for seismic microzonation in order to improve the resolution of the resulting 1D velocity models. The single-station approach using a Trillium Compact 120s seismometer and the blueSeis-3A rotational sensor makes measurements very simple in terms of logistics compared to an array set-up, especially when working in an urban area. The six-component data allow the computation of Love and Rayleigh wave dispersion curves. The limitation of rotational motion in the lower frequency range (< 5 Hz), which appears to be a combination of sensor self-noise and reduced rotational noise amplitudes, can be resolved by combining the dispersion curves with H/V ratios. The most reliable P- and S-wave velocity profiles are obtained by constraining the inversion to a three-layer model. Increasing the number of layers resolves more velocity steps close to the surface, however, the misfit value does not significantly decrease, which could be an indication for over-fitting the data. As an application, the resulting 1D velocity profiles will be used in future studies to estimate the local shaking characteristics in Munich.

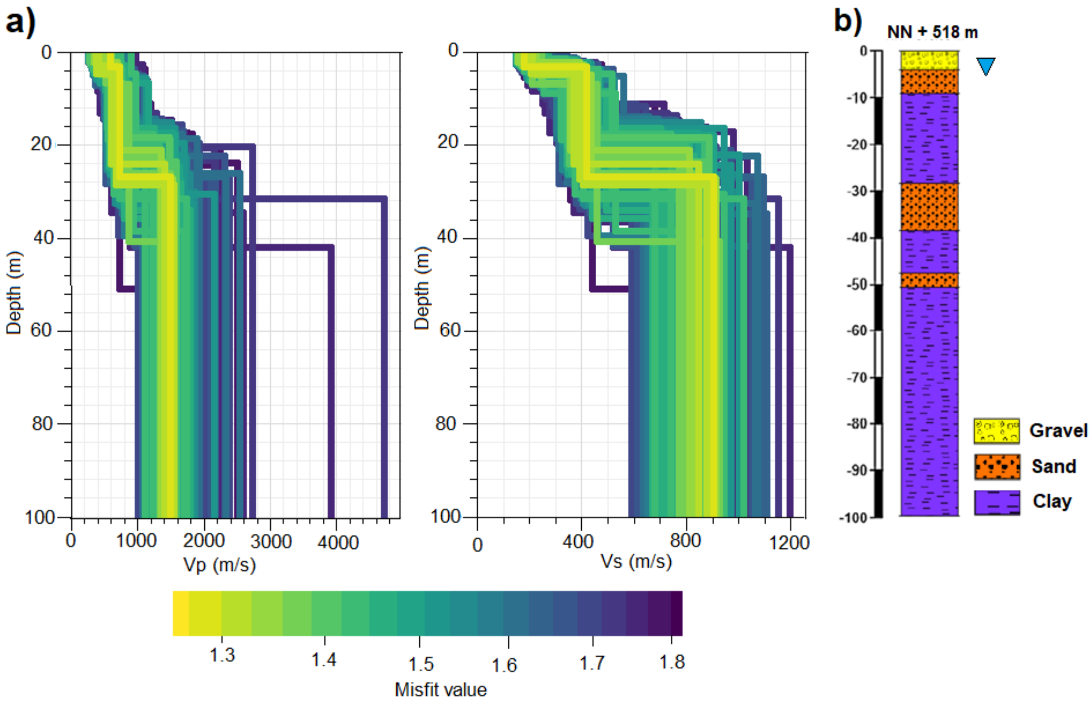

Figure 2.11: a) Resulting three-layer P- and S-wave velocity profiles at station BRUD using DINVER of GEOPSY (Wathelet 2008). For the inversion Love and Rayleigh wave dispersion curves, together with H/V ratios, were used with a weight of 1. The shading gives the misfit of the computed models. b) Lithologic profile estimated from borehole data during the GeoPot project of TUM (Technical-University-Munich 2016-2019). The groundwater table is marked by the triangle

In general, the velocity models show a correlation to the lithologic profiles that were derived during the GeoPot project. Especially layers close to the surface and the upper groundwater table could be identified.

The results have implications for any situations in which 1) near-surface velocity structure are sought and 2) multi-station networks are difficult or impossible to implement. This may happen in urban environments, at volcanoes, at the ocean bottom, or on planetary objects.

Acknowledgements

The authors like to thank the company iXblue for the cooperation in developing a portable, broadband rotation sensor. Thanks to SWM (Stadtwerke München) for providing access to the site of the geothermal power plant in order to conduct our measurements. Thanks to the support from the research project SEIGER (Project. no. 03EE4003G), funded by the German Federal Ministry of Economics and Energy (BMWi) and the research institute Jülich (PTJ). HI is grateful for support from the ERC-Adv Grant ROMY (Project. no. 339991). The authors thank the reviewer Johana Brokesova, who helped to improve this paper with useful comments.

2.2 Comparing single-station 6C measurements and array measurements for seismic microzonation in Munich, Germany

by Sabrina Keil, Alexander Wilczek, Joachim Wassermann and Simon Kremers

Published in Geophysical Journal International 231.3 (2022): 1634-1652

Abstract

The essential goal of seismic microzonation is the estimation of the shallow velocity structure in order to characterise the local earthquake shaking characteristics. This is of special importance in densely inhabited areas with unfavorable soil conditions. The common approach is the analysis of ambient noise array data using frequency-wavenumber (FK) or spatial autocorrelation (SPAC) techniques. However, the installation of arrays is difficult, especially within urban environments, making single-station approaches more desirable. In this study, we directly compare the recently developed approach of velocity estimation using single-station six-component (6C) measurements, combining three translational and three rotational motions, with the established methods of FK and SPAC analysis. We conduct measurements in Munich's inner city using a geophone array and an iXblue blueSeis-3A rotational motion sensor together with a Nanometrics Trillium Compact Seismometer, respectively. From the array data, as well as from the 6C data, Love and Rayleigh dispersion curves are estimated and further inverted for 1D P- and S-wave velocity profiles. We find that all methods give similar results, indicating the potential of the novel 6C approach. Furthermore, adding horizontal-to-vertical spectral ratios enables the inversion for structures at greater depth and increases the resolution of the velocity structure. In addition, we test different array geometries to evaluate the influence of the sensor configuration on the results. As a last step, we compare the estimated velocity models to lithologic profiles and find an overall positive correlation, which supports our inversion results.

2.2.1 Introduction

The assessment of local site effects is an important part of seismic hazard and engineering seismology studies. One of the controlling parameters on earthquake ground motion is the shear-wave velocity structure of the shallow subsurface. Hereby, the local geology has a strong influence, as low-velocity sediments have a large amplification potential and thus pose a great risk especially in densely populated areas (e.g. Hartzell et al. 2001, Field et al. 1997). This highlights the importance of seismic microzonation studies, which provide an estimate of the local ground motion in areas of unfavourable soil conditions, map potential areas of amplification and provide estimates for building codes. Information on the shallow

subsurface structure can be extracted from ambient vibration recordings at low costs and with good lateral coverage. Several strategies exist, including array measurements and single-station approaches. Common array methods are the spatial autocorrelation (SPAC) (Aki 1957) and frequency-wavenumber analysis (FK) (Capon 1969), which yield reliable results on the dispersion characteristics of surface waves (e.g. Ohrnberger et al. 2004, Fäh et al. 2008). However, it is difficult to apply these methods in areas, where an array-setup is challenging, such as in urban environments. In these cases, a single-station approach is desirable. The method of horizontal-to-vertical (H/V) spectral ratios (Nogoshi & Igarashi 1971) is already commonly applied, however, its theoretical foundation is still not completely understood and the results suffer from non-uniqueness (e.g. Hobiger et al. 2013).
Meanwhile, recent advances in sensor technology make it possible to record three components of rotational motion, in addition to the three components of translational motion. The development of portable rotational sensors, including the fibre-optic gyroscope based blueSeis-3A (Bernauer et al. 2018) and the mechanical six-degree-of-freedom Rotaphone (Brokešová et al. 2012), allow the application of such single-station six-component (6C) measurements for microzonation purposes. It could be shown that Love and Rayleigh wave dispersion curves can be computed from 6C data (Wassermann et al. 2016, Keil et al. 2021). The study presented here directly compares this novel 6C approach with the well established methods of FK and SPAC analysis, as well as the H/V spectral ratio method and highlights the advantages and disadvantages of the different approaches.
Therefore, several measurements have been performed in Munich´s (Germany) inner city, which is situated in the North Alpine foreland basin, that is filled with up to 3 km thick layers of Jurassic, Cretaceous and Tertiary sediments in the surroundings of Munich. A new geothermal power plant in the inner-city gives rise to concern regarding induced seismicity and raises the necessity to study the shallow subsurface in more detail, as the sedimentary layers could amplify the seismic waves. Urban noise was recorded using three-component 4.5 Hz geophones for the array set-up and a Trillium Compact 120s Seismometer together with the blueSeis-3A rotational motion sensor for the single-station 6C measurements. Different array geometries were tested in order to evaluate the influence of the instrument configuration on the results and find a trade-off between data quality and feasibility in urban settings. From the array data, as well as the 6C data, Love and Rayleigh wave dispersion curves are computed, which are then further inverted together with H/V ratios for the 1D velocity structure. Finally, the velocity profiles are compared to lithologic profiles that were extracted from the hydro-geologic 3D model of Munich, which was derived from borehole data during the GeoPot (Geo-potentials of the tertiary subsurface) project (Technical-University-Munich 2016-2019) of the Technical University of Munich (TUM).

2.2.2 Methodology

2.2.2.1 Passive array methods

We analyze the array data using the high-resolution frequency wavenumber (HRFK) approach after Capon (1969), which is an enhancement of the conventional beamforming FK method. FK analysis is a standard array technique which simultaneously calculates the power distributed among different slownesses and directions of a plane wavefield propagating across the sensor array. A grid search for all slowness and backazimuth combinations is applied in the spectral domain to find the best parameter combination. The conventional method uses a fixed wavenumber window, while the HRFK approach implements a weighting scheme, which results in a higher wavenumber resolution (e.g. Capon 1969, Woods & Lintz 1973, Asten & Henstridge 1984). The estimator can be written as:

$$P(\omega,\vec{k}) = \frac{1}{A(\omega,\vec{k})^{\mathrm{H}} R(\omega)^{-1} A(\omega,\vec{k})} \tag{2.9}$$

where $R(\omega)^{-1}$ is the inverse of the cross spectral matrix estimate and $A(\omega,\vec{k})$ are the steering vectors for wave vector $\vec{k}$ at frequency ω (e.g. Ohrnberger et al. 2004).

The resolution capability is specified by the ability of separating two waves with closely neighboured wave vectors and depends on the array geometry, as well as on the wave field characteristics. For regular array geometries the corresponding resolved wavelength range can be approximated by $\lambda_{\mathrm{max}} = 3D_{\mathrm{max}}$ and $\lambda_{\mathrm{min}} = 2D_{\mathrm{min}}$, where D represents the inter-sensor distance and λ the wavelength (Tokimatsu 1997). For irregular arrays the resolution limits depend on the spatial distribution of sensors and should be defined from the theoretical array response, which takes into account the real capabilities of the array geometry (Wathelet et al. 2008).

Love and Rayleigh dispersion curves are computed using the HRFK implementation of the GEOPSY software for three component data (Wathelet et al. 2018).

As a second analysis method for the array data we use the modified spatial autocorrelation (MSPAC) approach proposed by Bettig et al. (2001). The SPAC method assumes a stochastic wavefield, which is stationary in space and time. With this assumption spatial autocorrelation functions can be computed between station pairs, which depend directly on the unknown velocity of the seismic wave (Aki 1957). This conventional SPAC technique requires perfect shaped array geometries, which may be difficult to apply in urban environments. The MSPAC method can be applied to less ideal array configurations, as the autocorrelation values are computed for co-arrays, which are defined as the set of all possible combinations of two sensors. The averaged spatial autocorrelation coefficient $\overline{\rho_{r_1,r_2}}(\omega)$ can then be computed by taking into account the radius of the inner (r_1) and outer (r_2) ring of the sub-array, instead of using a fixed radius (Bettig et al. 2001):

$$\overline{\rho_{r_1,r_2}}(\omega) = \frac{2}{r_2^2 - r_1^2} \int_{r_1}^{r_2} r J_0\left(\frac{\omega r}{c(\omega)}\right) dr \tag{2.10}$$

J_0 denotes the Bessel function of the first kind and zero-th order and $c(\omega)$ is the phase velocity at a given frequency.

Love and Rayleigh dispersion curves are computed using the MSPAC implementation of GEOPSY, which extends this method to three-component data (Köhler et al. 2007).

2.2.2.2 Single-station methods

Among the single-station approaches the H/V spectral ratio method is most popular, even though it still lacks a clear theoretical explanation. The method was first introduced by Nogoshi & Igarashi (1971), who described it as the ratio between the Fourier amplitude spectra of the horizontal and the vertical component of microtremors:

$$H/V(f) = \frac{\sqrt{|N(f)|^2 + |E(f)|^2}}{|Z(f)|} \tag{2.11}$$

where N(f), E(f) and Z(f) are the north, east and vertical components of the noise signal. When the surface layer has a sharp impedance contrast with underlying stiffer formations the H/V curve often exhibits a distinct peak, which was found to be close to the fundamental S-wave resonance frequency of the site (e.g. Nakamura 1989, Lachet & Bard 1994). Other studies interpret the H/V ratio as the ellipticity of Rayleigh waves at the surface of a layered medium and account the peak of the curve to the fundamental Rayleigh wave mode. Therefore, the shape of the H/V curves can be used to determine the shear wave velocity profile (e.g. Malischewsky & Scherbaum 2004, Fäh et al. 2001). If the wavefield is only composed of single mode Rayleigh waves, the H/V ratio and ellipticity should be identical. However, in general Love waves are also present on the horizontal component. Therefore, the H/V method overestimates the actual Rayleigh wave ellipticity (e.g. Poggi et al. 2012). As a consequence, the H/V curve has to be corrected for the Love wave contribution, in order to be used for the inversion. In this study the RayDec (Hobiger et al. 2009) method is used to estimate the Rayleigh wave ellipticity, which is based on the random decrement technique. This approach is explained in more detail in Hobiger et al. (2009). As a comparison, classical H/V ratios are computed from the three-component data using the GEOPSY software (Wathelet et al. 2020).

Recent advances in sensor technology have made it possible to measure all three components of rotational ground motion, in addition to the traditionally recorded three components of translational motion. This gives rise to a new single-station method for microzonation studies, that makes use of such 6C data. The rotational motions of a wavefield (ω_x, ω_y, ω_z) are related to the curl of the translational motions according to equation (2.12) (u_x, u_y, u_z) (e.g. Cochard et al. 2006).

$$\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} = \frac{1}{2}\nabla \times \vec{u} = \frac{1}{2}\begin{pmatrix} \partial_y u_z - \partial_z u_y \\ \partial_z u_x - \partial_x u_z \\ \partial_x u_y - \partial_y u_x \end{pmatrix} \tag{2.12}$$

We now assume a transversely polarized plane wave propagating in x-direction with a displacement vector

$$\vec{u} = (0,\, A\sin(kx - kct),\, 0)^T \tag{2.13}$$

where A is the amplitude, k the wavenumber, and c the phase velocity. Applying equation (2.12) to equation (2.13), gives us the rotation angle around a vertical axis, which can be differentiated to obtain the rotation rate $\dot{\omega}_z$:

$$\dot{\omega}_z = \frac{1}{2}\, k^2\, c\, A\, \sin(kx - kct) \tag{2.14}$$

This rotation rate can be directly measured using a rotational motion sensor.
The transverse acceleration is further defined as:

$$a_\mathrm{T} = \ddot{u}_\mathrm{T} = -k^2\, c^2\, A \sin(kx - kct) \tag{2.15}$$

Dividing equation (2.15) by (2.14) results in a simple relation between rotational and translational motion for the estimation of the phase velocity:

$$\frac{a_\mathrm{T}}{\dot{\omega}_z} = -2c_\mathrm{L} \tag{2.16}$$

This relation holds for plane, fundamental mode Love waves, as was shown by several authors (e.g. Ferreira & Igel 2009, Kurrle et al. 2010). A similar relation can also be derived for fundamental mode Rayleigh waves by taking the ratio of the vertical acceleration a_z and the transverse rotation rate $\dot{\omega}_\mathrm{T}$ (e.g. Lin et al. 2011).

$$c_\mathrm{R} = \frac{a_z}{\dot{\omega}_\mathrm{T}} \tag{2.17}$$

Since surface waves are dispersive, the phase velocities c_L and c_R are frequency dependent and thus, Love and Rayleigh wave dispersion curves can be computed from six-component data using equations (2.16) and (2.17) (e.g. Wassermann et al. 2016, Keil et al. 2021). The data are analysed with the python package ROLODE (ROtational LOve Dispersion

Estimation), which is described in more detail in Wassermann et al. (2016). This program simultaneously estimates the backazimuth and the velocity of the wavefield for short sliding time windows using the principle of orthogonal distance regression.

2.2.2.3 Inversion for the velocity profile

The derivation of 1D velocity profiles from surface wave data is a classical inversion problem in geophysics. We perform the inversion using the DINVER module in the GEOPSY software package, which implements a neighbourhood algorithm (Wathelet 2008). The main parameters are the P-wave velocity (v_P), the S-wave velocity (v_S), the density, and the thickness of each layer. We constrain the inversion to a maximum depth of 100 m according to the frequency range of the input data. Furthermore, the density is fixed to a constant value of 2000 kg/m^3 since it has a small influence on the dispersion curves (Wathelet et al. 2004). The velocities v_P and v_S are linked together in the inversion by Poisson´s ratios between 0.2 and 0.5, which are typical values for sedimentary rocks (Reynolds 2011). A critical step in the application of DINVER is the choice of the number of layers to be inverted for. On the one hand, the parameterized model should be complex enough to match the measured input data, but on the other hand over-fitting has to be avoided. Therefore, we perform separate inversions for varying layer numbers to find the best match. We also test different combinations of input data, such as dispersion curves only, H/V curves only and both together. The computed results are compared with the measured curves, which then provides a misfit value given by (Wathelet et al. 2004):

$$\text{misfit} = \sqrt{\sum_{i=0}^{n_\mathrm{F}} \frac{(x_{di} - x_{ci})^2}{\sigma_i^2 n_\mathrm{F}}} \tag{2.18}$$

Where x_{di} is the apparent phase velocity of the data curve at frequency f_i, x_{ci} is the velocity of the calculated curve at frequency f_i, σ_i is the uncertainty of the frequency samples considered and n_F is the number of frequency samples considered.

2.2.3 Study area and measurement set-up

The study area is the inner city of Munich (Germany). In this work, measurements that were performed using the single-station 6C set-up together with a geophone array are discussed at two different sites. The aim was to select sites in the vicinity of the geothermal power plant, which is located in the southern part of Munich, since this will be the most likely area for induced earthquakes. However, due to the spatial requirements of the array set-up, measurements could only be performed in parks and on sports fields, which resulted in a larger spread of the selected locations (Fig. 2.12).

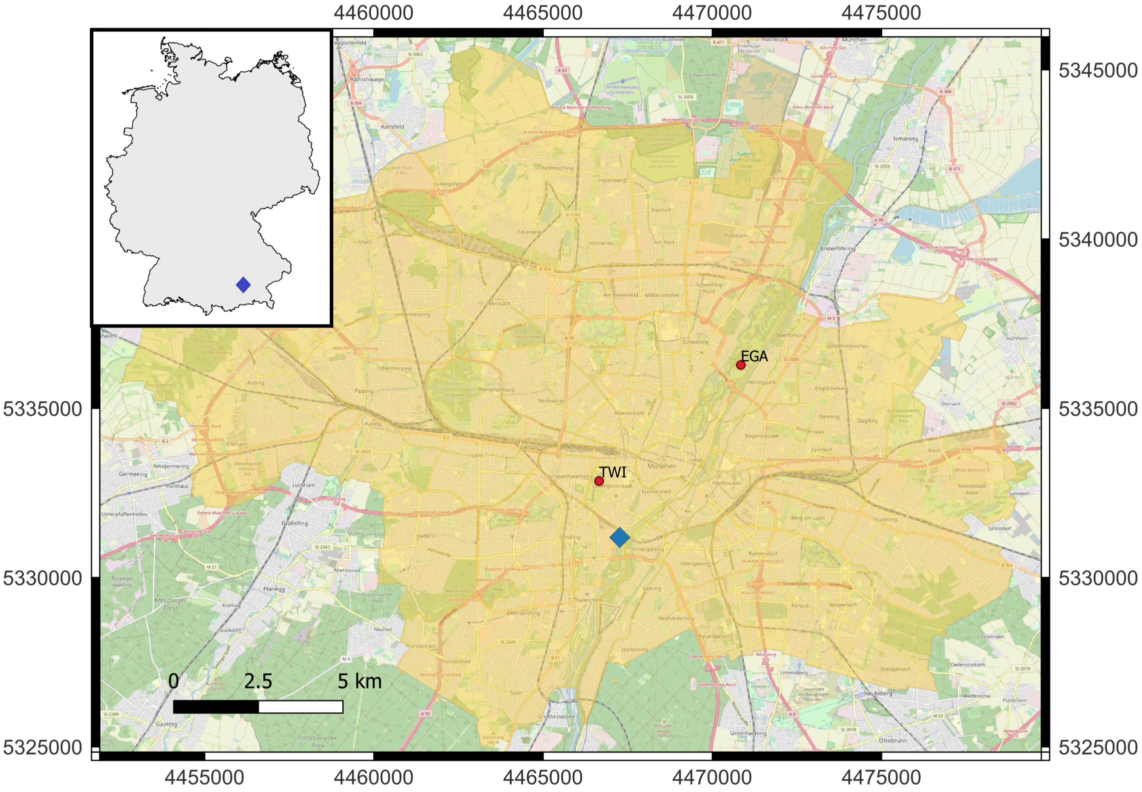

Figure 2.12: Map of Munich with measurement locations marked by the dots and corresponding station names. At each station 6C and array measurements were performed. The location of the geothermal power plant is marked by the diamond. The map is referenced by the Gauss-Krüger coordinate system (GK zone 4). In the upper left corner an overview map of Germany is shown, where the diamond marks the location of Munich.

Two geometries were tested for the array measurements: 1) a circular instrument configuration and 2) an L-shaped set-up. For the circular array, three rings consisting of 9 sensors in addition to one sensor in the center, were installed consecutively with different radii of 6 m, 18 m and 50 m in order to record data over a broad range in the f-k plane. Small array apertures are able to resolve higher frequencies and wavenumbers and can therefore resolve the shallow subsurface structure, while large apertures resolve low frequencies and wavenumbers and are thus used to investigate deeper structures. The 50 m array was set-up for 2 hours, the 18 m array for 1 hour and the 6 m array for 30 minutes, to record enough data in the corresponding frequency range. The L-shaped array consisted of 11 instruments with increasing sensor spacing from 5 m to 50 m and a side length of 100 m. The recording time for this configuration was 2 hours. The schematic array geometries are shown in Figs. 2.13a) and b). The positions of the instruments were measured with a differential GNSS receiver with a precision in the centimeter-scale. At the sites three-component 4.5 Hz geophones with a sampling frequency of 1000 Hz were installed on a slab, as shown in Fig. 2.13d). The instruments were aligned towards north, levelled and covered by buckets in order to shield them against wind. After the recording the geophone data were corrected by their individual frequency response function. The reason for this step was to recover also signals below the natural eigenfrequency of 4.5 Hz and thus a uniform sensor characteristic was simulated with a corner frequency of 1 Hz.

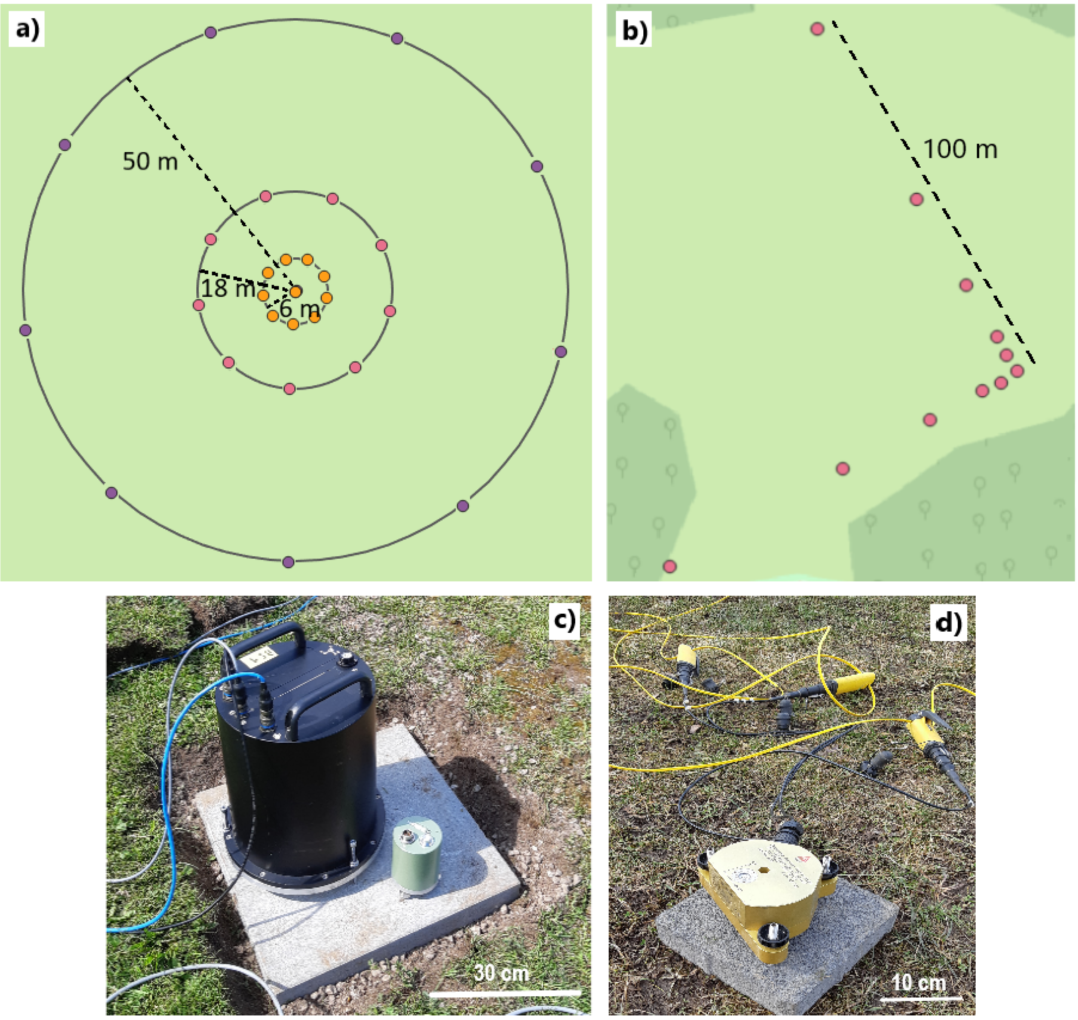

Figure 2.13: Shown is a) the schematic circular array geometry including the rings with radii of 6 m, 18 m, and 50 m, and b) the schematic L-shaped array with varying sensor spacing of 5 m to 50 m and a side length of 100 m. The 6C station was always installed close to the array. In c) the 6C set-up is shown. The larger black instrument is the blueSeis-3A rotational motion sensor. The Trillium Compact 120s seismometer is placed a few centimeters next to it on the same plate. Before the recording starts, the instruments are covered by a styrofoam insulation box. In d) the installation of the three-component geophone is shown. It is placed on a concrete plate, levelled, aligned to north and covered by a bucket before the recording starts.

Simultaneously, the 6C measurements with a sampling rate of 200 Hz were performed next to the array set-up using a Nanometrics Trillium Compact 120s seismometer, which was connected to a RT130 digitizer, and an iXblue blueSeis-3A rotational motion sensor. The blueSeis-3A is based on an interferometric fiber-optic gyroscope (FOG) and is strictly insensitive to translational motions. In addition, the rotational sensor has a flat self-noise level lower than 30 $\text{nrads}^{-1}\text{Hz}^{-1/2}$ over a wide frequency range of 0.001 Hz to 50 Hz and is suitable for field installations (Bernauer et al. 2018). In order to record exactly the same movement, the two instruments, seismometer and FOG, were installed on the same concrete base plate only a few centimetres apart from each other. This way it can be interpreted as a single-site measurement. To guarantee a better coupling to the ground, 10 cm of the topsoil has been removed. A styrofoam insulation box was placed over the instruments to shield them against wind and direct sun radiation. Past experiments have shown that a recording time of 2 hours is sufficient to acquire enough data in the desired frequency range of 1 Hz to 20 Hz. The 6C installation is shown in Fig. 2.13c).
Ambient noise data were recorded by the different instruments according to their sensitivity and analyzed in a frequency range of 1 Hz to 20 Hz, which corresponds to the frequency band of urban noise. The measurements were performed during day time, throughout which the highest noise levels can be expected. The site TWI is surrounded by busy roads and train tracks, in addition to a nearby subway station. Therefore, it can be assumed that the main sources of the recorded noise are cars and trains. The main frequency content of the recorded noise at this site lies between 3-25 Hz with a peak at 12 Hz. The site EGA is situated inside a meadow next to a smaller road, in addition to a busy road at 400 m distance. Therefore, it is probable that the main source of noise are cars. The data display most of the energy above 3 Hz with a peak at 15 Hz.

2.2.4 Results

In the following section we are going to present the results for station TWI, where a circular array was installed, and station EGA, where the L-shaped array was set-up. At both locations three-component geophones were used, therefore, Love and Rayleigh dispersion curves could be derived. The 6C measurements were performed simultaneously.

2.2.4.1 Station TWI

As a first step dispersion curves were computed from the array data using HRFK and MSPAC analysis, as well as from the 6C data using the ROLODE method. The Love and Rayleigh dispersion curves are presented in the slowness domain, which is the reciprocal of velocity (Fig. 2.14). In addition, the array resolution limits k_{min} and k_{max} are computed following Wathelet et al. (2008). The Rayleigh dispersion curves show a similar trend in a frequency range of 8 Hz to 17 Hz. For frequencies above 11 Hz the 6C curve displays higher slowness values, which are however still in the range of the data´s standard deviation. The data point at 20 Hz has a larger variation. In the lower frequency range the computed curves strongly deviate. As normal dispersion is assumed, the slowness values should decrease towards lower frequencies. This is not the case for the 6C curve, which shows a strong increase below 8 Hz. This can be explained by rotational motion amplitudes that at this low frequencies are too small to be recorded by the blueSeis-3A and its self-noise level is reached. By looking at the power spectral densities (PSD) of the recorded data (Fig. 2.15) it can be seen that the self-noise exceeds the recorded rotational motions below 7 Hz to 8 Hz and thus makes the 6C method infeasible. Taking the computed array resolution limits into account, the HRFK and SPAC methods allow the computation of dispersion curves to lower frequencies. The limit k_{min} is crossing the curves at 5 Hz, while $k_{min}/2$ is crossing the curves at 3 Hz. For the HRFK method we expect reliable results down to $k_{min}/2$. The data below this limit are not reliable, which is also indicated by the variation of the HRFK and SPAC curves, and should not be included in the following inversion. Similar observations can be made for the Love dispersion curves. However, above 12 Hz the 6C curve shows larger deviations. One reason for the discrepancies specifically at high frequencies could be the different spatial resolution of the single-station 6C approach compared to the array approaches. The single-station 6C method provides a local estimation of the velocity profile, while the array methods yield an average velocity structure across the whole array. Another explanation could be the higher sensitivity of the rotational motions to small-scale structures compared to translational motions, as was shown by Singh et al. (2020).

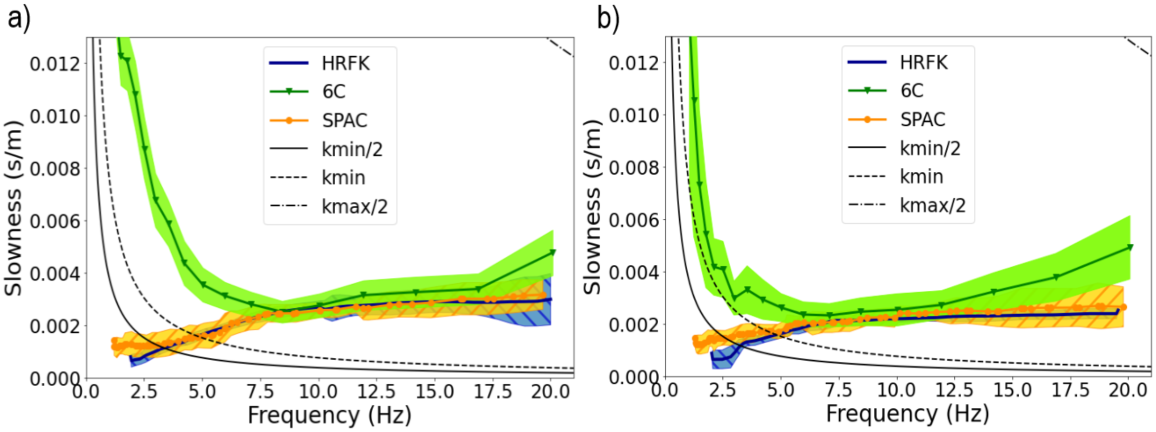

Figure 2.14: a) Rayleigh and b) Love dispersion curves for station TWI computed from the array data using HRFK analysis and MSPAC, and from the 6C data using the ROLODE method. The shadings represent the standard deviations of the data. In addition, the array resolution limits $k_{min}/2$, k_{min} and $k_{max}/2$ are plotted.

In order to obtain more information about the subsurface structure, H/V spectral ratios, as well as ellipticity curves, are computed from the three-component Trillium Compact data (Fig. 2.16). The shape of the two curves is very similar: both exhibit a first peak at about 2 Hz and a broader peak between 10 Hz and 20 Hz. The amplitudes of the ellipticity curve are consistently lower than the H/V ratio, especially in the troughs, due to the removed contribution of the Love waves on the horizontal components. In general, the

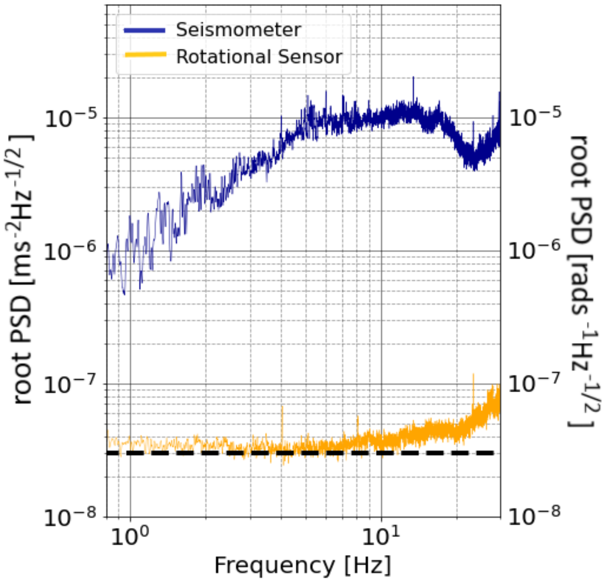

Figure 2.15: Root PSD plots of the translational and rotational data recorded at station TWI. The dotted line marks the self-noise level of the blueSeis-3A rotational sensor.

peak has a relatively low value of 1.8, which could be an indication for a small impedance contrast between two sedimentary layers. Nevertheless, Hobiger et al. (2013) showed that even ellipticity curves without strong peaks carry important information on the subsurface structure. In this case, the complete curve should be used for the inversion.

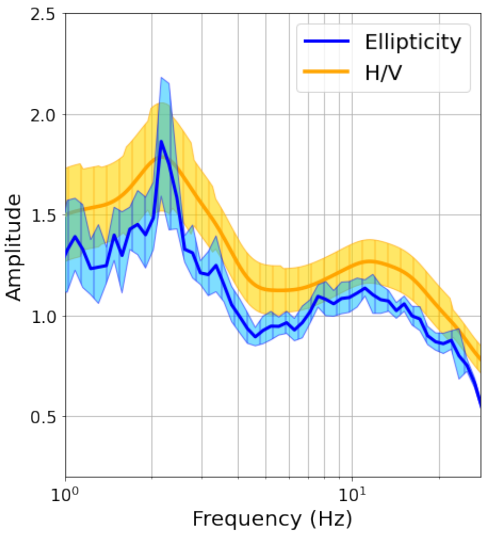

Figure 2.16: Comparison of classical H/V curves computed with GEOPSY (Wathelet et al. 2020) and RayDec (Hobiger et al. 2009) Rayleigh wave ellipticity curves at station TWI. The shading represents the standard deviation of the curves.

In the next step, the surface wave inversion is performed in order to obtain the local P- and S-wave velocity profiles. Different input data, as well as different parameter settings are tested to find the best configuration. Assuming that we lack any information on the subsurface structure, we start with a simple three layer model and link v_P to v_S, as the surface wave data provide more information about S-waves. In Fig. 2.17a), only the ellipticity curve was used as input target. There are several different velocity structures with a low misfit value visible. This indicates that the information from the ellipticity curve alone are not sufficient to resolve the subsurface structure. Nevertheless, the ellipticity curve in the inversion target plot (Fig. 2.A1a in appendix 2.2.A) is fitted very well for high frequencies. The first peak displays a lower fit, however, it still falls within the ensemble of computed models. In Fig. 2.17b), the Love and Rayleigh dispersion curves, derived from the 6C data, were used as input targets. The velocity structure is well resolved down to 15 m depth and displays a velocity step at 4 m. At greater depth the velocity estimates are widely scattered and no clear structure is visible, which is due to the missing information in the lower frequency range. Therefore, the second step at 70 m depth, indicated by the model with the lowest misfit, is not reliable. Both dispersion curves are fitted well in the inversion target plots (Fig. 2.A1b). In Fig. 2.17c) the dispersion curves were inverted together with the ellipticity curve. So far, this gives the best constrained velocity profiles, with two clear velocity steps at 3 m and at about 35 m depth. In order to fit the dispersion curves in the inversion target plots (Fig. 2.A1c) their weight in the inversion had to be increased. This results in a good fit for both of the dispersion curves, as well as the ellipticity curve. We additionally tested different parameter settings, such as varying number of layers and leaving v_P as a free parameter. Not linking v_P to v_S results in an ambiguous P-wave velocity profile. Increasing the number of layers additionally resolves shallower velocity steps, however, the main velocity contrasts stay the same. Furthermore, when adding more layers to the inversion one has to be careful not to over-fit the data. Based on these findings, for the following inversions we combine dispersion curves with ellipticity ratios, link v_P to v_S and constrain the model space to three layers.

In the next step, separate inversions are performed using the Love and Rayleigh dispersion curves from the different analysis methods together with the ellipticity ratios, in order to compare the resulting velocity profiles. In Fig. 2.18a), the best-fitting P- and S-wave velocity models are plotted together with their uncertainties. All three methods resolve an upper velocity step at 3 m to 4 m depth and estimate a P-wave velocity in the second layer of 670 m/s to 800 m/s and a S-wave velocity of 400 m/s to 490 m/s. The depth for the second velocity step ranges between 35 m to 40 m. Also, the velocity estimates have a larger range of 1470 m/s to 2150 m/s for v_P and 890 m/s to 1040 m/s for v_S, which could be due to the decreasing resolution with depth. However, considering the uncertainty of the inversion results, the velocity models from the different input data are in good agreement, despite the differences in the dispersion curves at high frequencies. Furthermore, the targets for the 6C inversion are fitted well (Fig. 2.A2a in appendix 2.2.A). The fit for the HRFK and SPAC targets is lower, however, they still fall within the ensemble of computed models (Fig. 2.A2b and c).

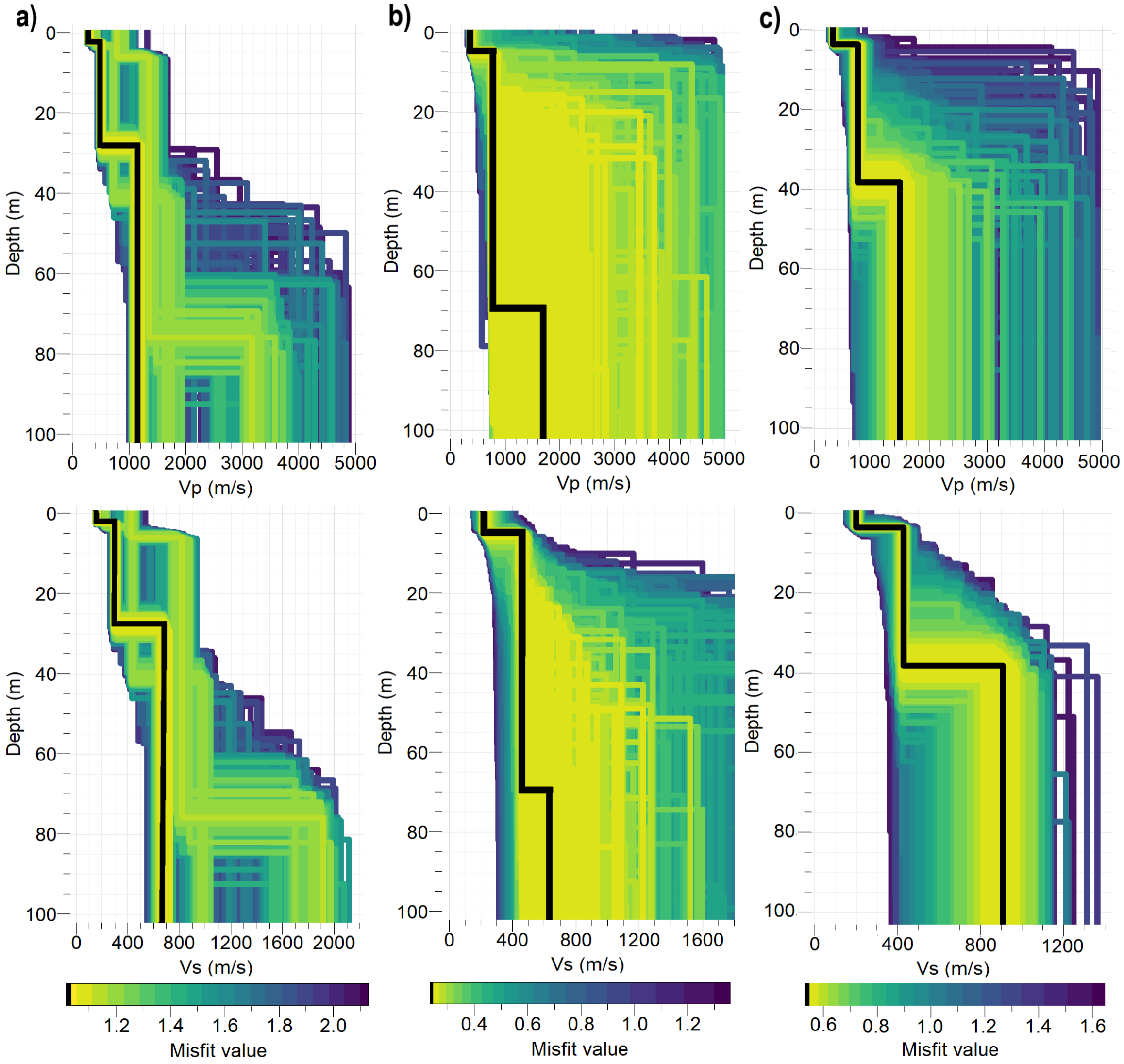

Figure 2.17: P- and S-wave velocity profiles at station TWI for different input data. The inversion runs were constrained to three layer models. a) shows the inversion results using only the ellipticity curve as input. b) shows the inversion results for the 6C-data derived Love and Rayleigh wave dispersion curves. c) shows the inversion results for the 6C-data derived Love and Rayleigh wave dispersion curves together with the ellipticity curve. The color scale gives the misfit of the computed models. The black line marks the model with the smallest misfit. The corresponding inversion targets and computed curves can be found in Fig. 2.A1 of the appendix 2.2.A.

In the next step, the 1D velocity profiles are compared to the lithologic profile in Fig. 2.18b), that was interpolated from borehole data during the GeoPot project of the TUM (Technical-University-Munich 2016-2019). It can be seen that the upper velocity step coincides with the groundwater table at 4 m depth, as well as with the lithologic change from gravel to sand. The second step, however, lies within a thick clay layer, where no layer boundary is present in the profile.

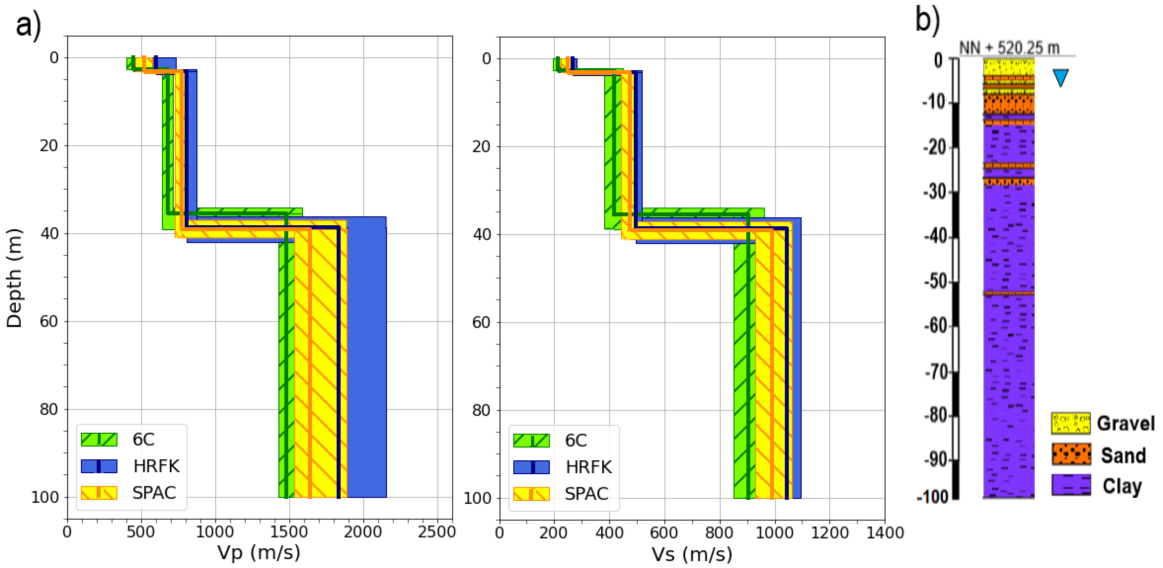

Figure 2.18: a) P- and S-wave velocity profiles for a three-layer inversion at station TWI using ellipticity ratios and the dispersion curves from the different analysis methods (6C, HRFK, SPAC). The solid lines mark the velocity structure of the model with the lowest misfit (6C = 0.54, HRFK = 0.76, SPAC = 0.42). The shaded areas include all profiles with a misfit lower than 1.05 times the smallest misfit to account for uncertainty. The corresponding inversion targets and computed curves can be found in Fig. 2.A2 of the appendix 2.2.A. b) lithologic profile of site TWI interpolated from borehole data during the GeoPot project of the TUM (Technical-University-Munich 2016-2019). The triangle marks the depth of the groundwater table.

We try to improve the inversion results by increasing the number of layers to four (Fig. 2.19a). The inversion targets (Fig. 2.A3a in appendix 2.2.A) are fitted well in this case. This inversion resolves once more a shallow velocity step at 4 m depth and an additional layer at 11 m depth, which coincides with the lithologic change from sand to clay. However, the dominant velocity step at 40 m depth still persists. This can have several reasons: 1) the seismic facies is not the same as the lithologic facies, 2) the lithological profile is estimated by interpolation, therefore thickness as well as exact depth of the layers may vary in reality or 3) another groundwater layer is present at this depth. In one groundwater monitoring well north of the site TWI the groundwater level in the tertiary sediments was measured at 30 m depth, while in one monitoring well south of the site it was measured at 55 m depth. However, since there are no data available close to the site this deeper groundwater layer cannot be properly constrained. Another discrepancy in the velocity profiles are the low P-wave velocities below the groundwater table. The sediments below the groundwater table are water saturated and therefore we expect v_P to be at least about 1400 m/s to 1500 m/s, which is equal to the compression wave velocity of water (e.g. Foti et al. 2018, Cox & Teague 2016). Because of that, an inversion is performed constraining v_P to be larger than 1400 m/s in the second layer in order to account for the presence of groundwater (Fig. 2.19b). The resulting velocity profiles display velocity steps in the same depth ranges as in Fig. 2.19a), as well as similar S-wave velocity values. The P-wave velocities are similar for the first and last layer, but increases in the second and third layer from 500 m/s and 750 m/s to 1400 m/s and 1450 m/s, respectively, due to the constraint. The fit of the Love dispersion curve in the inversion target plot is improved for high frequencies (Fig. 2.A3b). However, the ensemble of computed Rayleigh dispersion curves is shifted to smaller slowness values and therefore decreases the fit of the picked dispersion curve. In addition, the first peak of the ellipticity curve is fitted less well. These observations indicate that the P-wave velocity constraint does not improve the inversion results. Another observation from the velocity profiles is that the thin sandstone layers at about 30 m and 50 m depth cannot be resolved even if the number of layers in the inversion is increased. One explanation is that these layers are too thin for this depth ranges to be resolvable: According to Foti et al. (2018) the thickness of the first layer should be greater than half of the minimum available experimental wavelength. This provides an approximation of the thinnest resolvable layer, which is about 3 m for the site TWI. However, this is only valid for layers close to the surface. As resolution decreases with depth, deeper layers have to be thicker in order to be detected. If there are no additional layer boundaries below 15 m depth which influence the P- and S-wave velocities, it would be more appropriate to add a velocity gradient to the inversion, which models an increase of v_P and v_S due to soil compaction. The resulting velocity profiles (Fig. 2.19c) display an upper velocity step at 3 m depth, in accordance to the three and four layer inversions. The small and regular steps below this layer represent the continuous increase of velocity with depth. The inversion targets are fitted very well (Fig. 2.A3c in appendix 2.2.A), which indicates that a velocity gradient due to soil compaction could be possible.

As a last step, the Vs30 value is calculated from the S-wave velocity profile, which is defined as the average seismic shear-wave velocity from the surface to a depth of 30 meter. It is widely used as a parameter to characterize site response and is implemented in building codes (e.g. Borcherdt 2012). Therefore, it is important to calculate this parameter for our microzonation study in Munich, in order to estimate the amplification potential in case of induced seismicity. For the site TWI a Vs30 value of 355 m/s was calculated. According to the International Building Code (IBC), the site falls into group D, which represents stiff soils. Several studies (e.g. Hartzell et al. 2001, Boore 2004) have shown that such low velocities indeed pose a risk for site amplification.

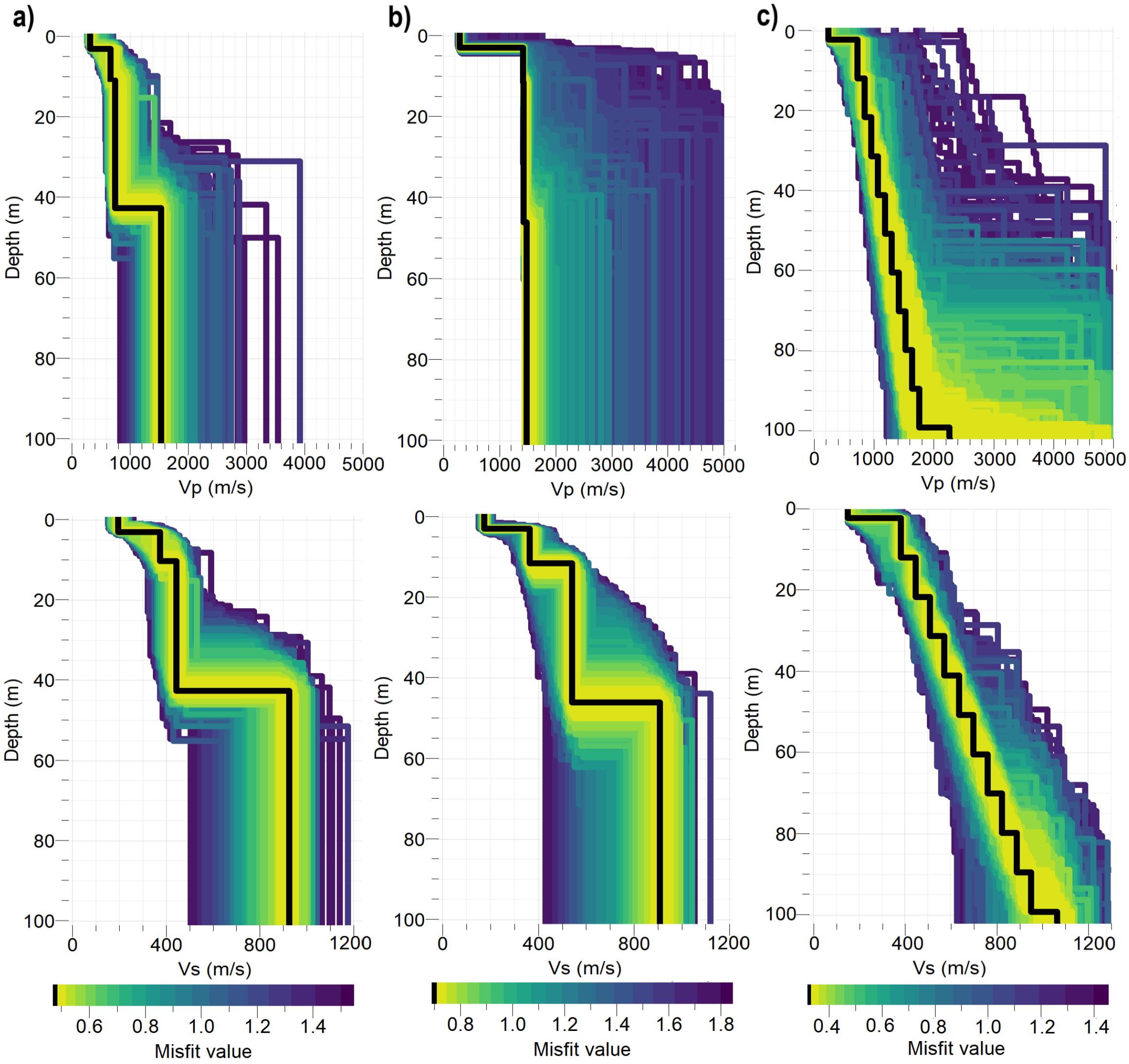

Figure 2.19: P- and S-wave velocity profiles at station TWI for different inversion constraints using ellipticity ratios and the 6C dispersion curves. In a) the inversion was constrained to a four layer model. Additionally, in b) the P-wave velocity was constrained to be larger than 1400 m/s below the groundwater table. In c) the inversion was constrained using a velocity gradient. The color scale gives the misfit of the computed models. The black lines mark the models with the smallest misfit. The corresponding inversion targets and computed curves can be found in Fig. 2.A3 of the appendix 2.2.A.

2.2.4.2 Station EGA

At site EGA, where the L-shaped array was set-up, the same analysis as for the site TWI, which was discussed in detail above, is done. However, the results will be discussed more concisely. First of all, the Love and Rayleigh dispersion curves were computed from the array data using HRFK and MSPAC analysis, as well as from the 6C data using the ROLODE method (Fig. 2.20). The Rayleigh dispersion curves show similar trends in a frequency range of 6 Hz to 18 Hz, where the higher slowness values of the 6C curve are still within the data´s standard deviation. The data point at 20 Hz shows larger differences. Furthermore, the 6C method once more shows a limitation in the lower frequency range, in this case starting at 6 Hz, which is the point where the self-noise of the rotational sensor exceeds the recorded signal, as shown in the PSD plot (Fig. 2.21). Due to the geometry, the array response is directional dependent. The resolution limits were computed for the most unfavorable direction. In this case, the array data can be interpreted to frequencies as low as 4 Hz, according to the computed resolution limit $k_{min}/2$. The limit k_{max} is crossing the dispersion curves at 14 Hz, which could lead to aliasing at higher frequencies. Similar observations can be made for the Love dispersion curves. Once more the 6C dispersion curve shows larger deviations at high frequencies. Reasons were discussed in section 2.2.4.1. Another observation is that with the SPAC method the Love dispersion curve could only be extracted up to 12.5 Hz, which is likely caused by the limited distribution of the co-arrays.

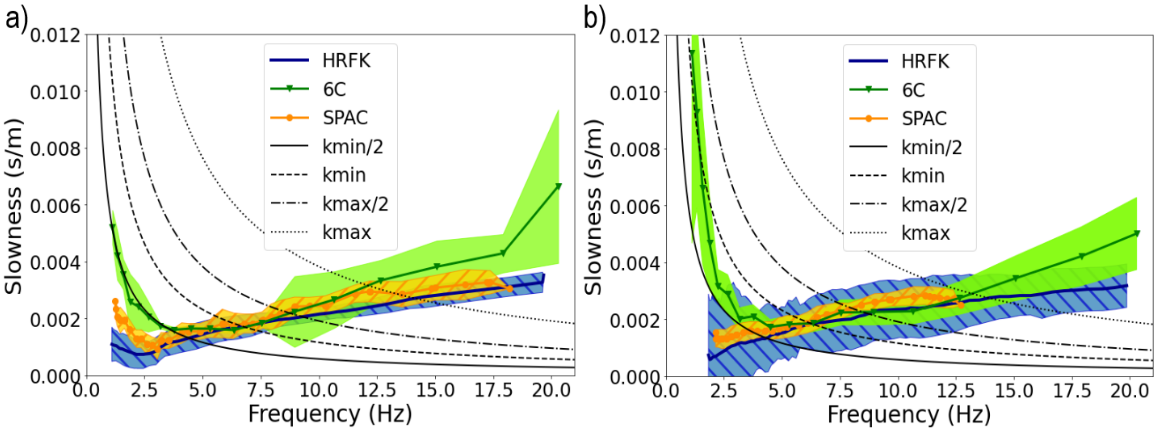

Figure 2.20: a) Rayleigh and b) Love dispersion curves for station EGA computed from the array data using HRFK analysis and MSPAC, and from the 6C data using the ROLODE method. The shadings represent the standard deviations of the data. In addition, the array resolution limits $k_{min}/2$, k_{min}, $k_{max}/2$ and k_{max} are plotted.

In the next step, the H/V and ellipticity curves are computed for this site (Fig. 2.B1 in appendix 2.2.B). The shape of the two curves is very similar, both exhibit a first peak at about 5 Hz and a second peak at 20 Hz. The amplitudes of the ellipticity curve are consistently lower than the H/V ratio, especially in the troughs, due to the removed contribution of the Love waves on the horizontal components. The complete ellipticity curve is used for

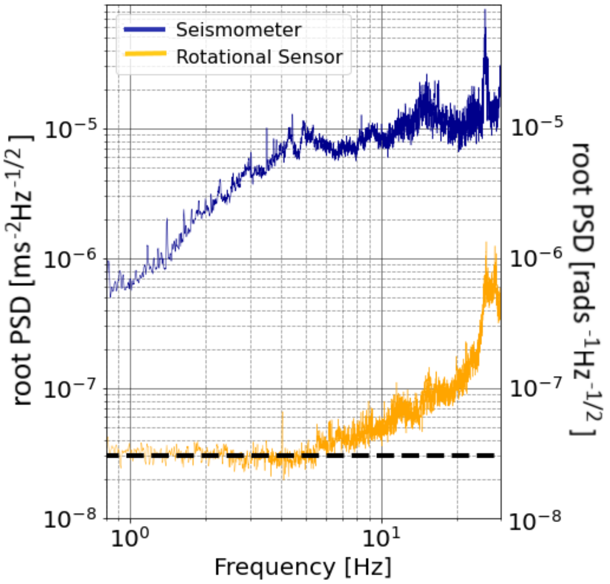

Figure 2.21: Root PSD plots of the translational and rotational data recorded at station EGA. The dotted line marks the self-noise level of the blueSeis-3A rotational sensor.

the following inversions. An inversion with the ellipticity curve alone results once more in an ambiguous velocity structure, indicating that it does not provide enough information to resolve the subsurface structure (Fig. 2.B2 in appendix 2.2.B). We again find that linking v_P to v_S, constraining the inversion to a three layer model, and using dispersion curves together with ellipticity ratios, results in the clearest velocity profiles at this site. Using this parameter configuration, we perform separate inversions for the dispersion curves of the different methods. The comparison of the resulting P- and S-wave velocity profiles is shown in Fig. 2.22a). The depth estimates for the velocity steps coincide very well. The upper step lies at about 3 m depth and the second one at 19 m to 22 m depth. The S-wave velocity profiles display similar trends with only small variations. The P-wave velocities in the first layer display a larger variation of 330 m/s to 550 m/s, where the 6C profile shows the lowest velocity. This is probably caused by the lower velocities in the 6C Love and Rayleigh dispersion curves at high frequencies. The P-wave estimates of the best fitting models in the second and third layer coincide well, however, the uncertainty in the third layer is very large. Nevertheless, considering the uncertainty of the inversion results, the velocity profiles from the three different data sets are in good agreement.

In the next step, the 1D P-and S-velocity profiles are compared to the lithologic profile in Fig. 2.22b). The upper step coincides with the groundwater table and the change from sand to clay. The second step lies in the range of the lower boundary of the thick sandstone layer, indicating the influence of the lithologic change from sand to clay on the velocity. The upper boundary of the sandstone layer can be resolved with a four layer inversion, however, the velocity increase is not as significant as for the lower boundary. Once more we notice that v_P in the second layer is too small for water saturated sediments. Therefore, we perform a new inversion constraining v_P in the second layer to be larger than 1400 m/s (Fig. 2.B4b in appendix 2.2.B). However, once more this does not improve the inversion results. On the contrary, the inversion targets are fitted less well, which increases the misfit values.

As a last step, the Vs30 value is calculated from the S-wave velocity profile in Fig. 2.22a), which yields a value of 400 m/s. Therefore, the site is categorized into site class C, which classifies the subsurface as very dense soil according to the IBC.

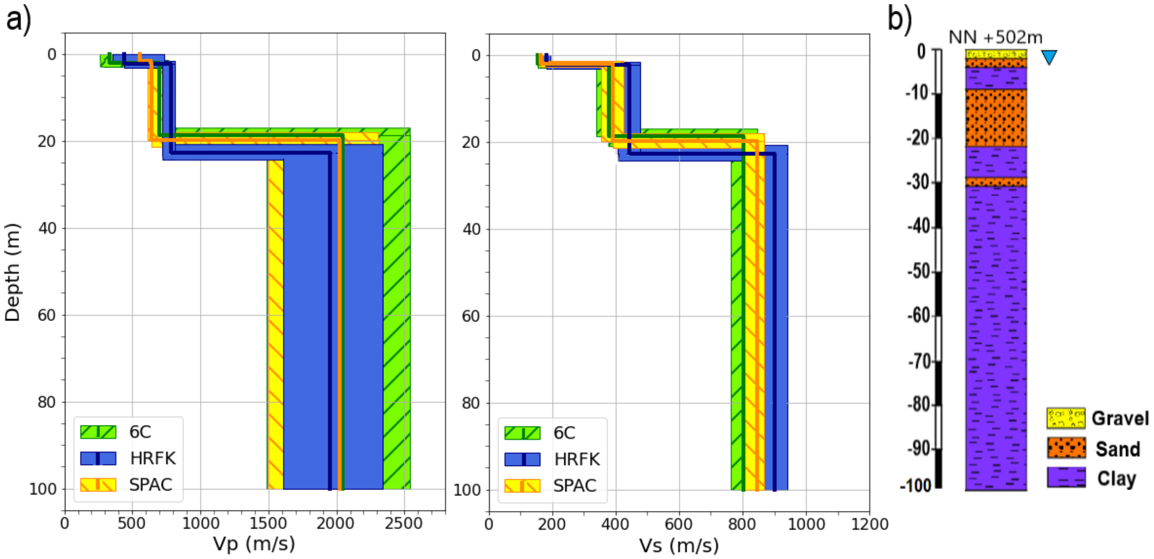

Figure 2.22: a) P- and S-wave velocity profiles for a three-layer inversion at station EGA using ellipticity ratios and the dispersion curves from the different analysis methods (6C, HRFK, SPAC). The solid lines mark the velocity structure of the model with the lowest misfit (6C = 0.68, HRFK = 0.44, SPAC = 0.59). The shaded areas include all profiles with a misift smaller than 1.05 times the lowest misfit to account for uncertainty. The corresponding inversion targets and computed curves can be found in Fig. 2.B3 of the appendix 2.2.B. b) lithologic profile of site EGA interpolated from borehole data during the GeoPot project of the TUM (Technical-University-Munich 2016-2019). The triangle marks the depth of the groundwater table.

2.2.5 Discussion

Love and Rayleigh wave dispersion curves were computed using the three different approaches of 1) HRFK analysis and 2) MSPAC for the analysis of array data and 3) the ROLODE method for the single-station 6C data. The dispersion curves from all three methods overlap in a certain frequency range. However, the 6C curves show consistently higher slowness values at high frequencies. This could be explained by the different spatial resolutions of the methods. The single-station 6C method provides a truly local estimate of the velocity profile directly below the station, while the array methods yield an average velocity over the whole array aperture. If there are small-scale lateral variations in the shallow subsurface structure the array methods would result in an altered velocity estimation compared to the 6C approach. In fact, such small-scale variations are possible within the Quaternary and Tertiary deposits below Munich. Another explanation for the differences of the 6C and array dispersion curves at high frequencies could be the higher sensitivity of the rotational motions to small-scale structures, compared to translational motions, as was shown in Singh et al. (2020). A third option could be the instrumental uncertainties in the recorded amplitudes. The 6C method relies on the comparison of amplitudes, while the array methods rely on phase estimations. For the Trillium Compact 120s an amplitude precision of 0.5% is specified, which only has a small effect on the computed phase velocities (± 1 m/s). The amplitude uncertainty for the rotational sensor is higher since the measured amplitudes are close to the self-noise level of the instrument. In this case we calculated a velocity variation of 10-80 m/s at frequencies between 17-20 Hz due to the uncertainties in the amplitudes. Furthermore, the array data in our case always allow the computation to lower frequencies. The resolvable frequency range hereby mainly depends on the array geometry and aperture. The limitation of the 6C method in the lower frequency range is due to the small rotational amplitudes at those frequencies and the self-noise level of the blueSeis-3A sensor, as was also observed in Keil et al. (2021). However, we overcome this limitation by adding the ellipticity ratio to the inversion, which provides information in the lower frequency range. This does not only resolves deeper velocity steps, but also increases the resolution of the velocity profiles, especially for v_P, as was also found in other studies (e.g. Fäh et al. 2008, Scherbaum et al. 2003, Arai & Tokimatsu 2005). In addition, the comparison of classical H/V ratios with the ellipticity curve shows a consistent difference in amplitude, which highlights the importance of correcting for the influence of Love waves. An inversion of the ellipticity curves alone resulted in non-unique velocity profiles, indicating the missing information on the subsurface structure. Furthermore, Hobiger et al. (2013) showed that several ground models can correspond to the same ellipticity curve, highlighting the importance of additional information on the subsurface structure. Testing different parameter settings and combinations of input data, we could show that a three layer inversion using both dispersion and ellipticity data, resulted in the most constrained velocity profiles. Because of that we also combined the dispersion curves computed from the array data with ellipticity ratios for the inversion. The resulting P-and S-wave velocity profiles are all in good agreement within their uncertainties, which highlights the potential of the novel 6C approach.

Comparing the three methods, a remarkable advantage of the 6C single-station approach is its simpler way of deployment, which especially eases the application in urban areas. This is in contrast to array deployments, which require a lot of space and efforts installing them. This is also the reason why we had to restrict the co-located measurements to parks and sports fields. Furthermore, an array measurement is time consuming, not only because of the duration of data acquisition but also due to the set-up of the array geometry. Especially the measurements with the circular array layout were very laborious and took up to eight hours. Compared to this, the 6C measurement in total takes only about three hours, including the set-up time of one hour. Therefore, this deployment just slightly takes longer than a simple 3C set-up for H/V measurements. However, the array measurements can be simplified by using alternative geometries.

Two different array set-ups were tested in order to evaluate the influence of the sensor geometry on the results and to find a simpler geometry for urban settings. First of all, we picked a circular sensor distribution, since it guarantees an equal response of the array for waves coming from all directions. Secondly, for the SPAC method, a circular array configuration with equal sensor spacing makes the computation of the azimuthal average of the autocorrelation function easier. For the second set-up we tested an L-shaped array geometry, which strongly reduces the space and time requirements and resulted in a similar data quality for the HRFK analysis. Only the MSPAC Love dispersion curve was limited to a smaller frequency range. This might be caused by less sensor pairs in the co-arrays and an irregular directional coverage. However, adding the ellipticity curve to the inversion process compensated for the missing information in the Love dispersion curve. The disadvantage of the L-shaped array is the directional dependency. Therefore, the orientation of the array has to be carefully selected by taking the main direction of the noise sources into account.

The main disadvantage of our 6C method is its limitation in the lower frequency range. At "noisy" sites we were able to compute dispersion curves to frequencies as low as 5 Hz, while at more remote sites the computation can be limited to > 10 Hz. However, in most cases the missing information in the lower frequency range can be compensated by the Rayleigh wave ellipticity.

Another shortcoming of the 6C method is that different modes cannot be distinguished, which could influence the dispersion curve estimation. However, from the array analysis it was determined that mainly fundamental modes are present in the wavefield, since only small parts of a higher mode could be identified. Furthermore, Wassermann et al. (2016) showed that with higher modes present in the wavefield the dispersion curve gets overestimated. This is, however, not the case in our results, as the phase velocities at high frequencies are lower than the ones from the array analysis. Therefore, in our case the higher modes in the wavefield do not seem to have a significant influence on the 6C results.

As a last step, we compared the estimated 1D P- and S-wave velocity profiles with interpolated lithologic profiles and found in general a positive correlation. We observed that the groundwater, as well as the change in lithology influences the wave velocity. At the site TWI one velocity step did not coincide with any layer boundary, which can have several reasons: 1) the seismic facies is not the same as the lithologic facies, 2) there is an error in the interpolated lithologic profile or 3) another deeper groundwater layer is present. In this case, we also tested an inversion using a velocity gradient to account for soil compaction, which resulted in a good fit of the inversion targets. Therefore, such a velocity gradient could be a good representation of the subsurface model. Another observation that is difficult to explain are the low P-wave velocities below the groundwater table. First of all, this is observed in the 6C data inversion, as well as in the array data inversion. In addition, constraining v_P to 1400 m/s below the groundwater table degrades the data fit. Because of that it might reflect an actual effect caused by the subsurface structure and therefore we advice against a v_P constraint in the inversion. Possible explanations are 1) a differing groundwater table during the measurements, as it is influenced by seasonal variations or 2) the soil is only partially saturated, as the P-wave velocity is very sensitive to changes in water content close to saturation (Allen et al. 1980, Valle-Molina & Stokoe 2012). This remains to be investigated in more detail in future studies. Another point to keep in mind is that in general the dispersion curves provide more information about the S-wave velocity structure than the P-wave structure. Therefore, larger uncertainties are contained in the P-wave velocity model. In addition, the S-wave velocity is the more important property for ground motion studies.

2.2.6 Conclusion

The objective of this study was to compare the novel method of single-station six-component measurements with established array measurements, including HRFK and MSPAC analysis, for the purpose of seismic microzonation. With the three methods we are able to estimate Love and Rayleigh dispersion curves. Overall we obtain similar results, however, there is a limitation of the rotational sensor in the lower frequency range, which is a combination of the sensors´ self-noise level and the small rotational amplitudes. This limitation can be overcome by computing ellipticity ratios, which provide additional information especially at lower frequencies. Furthermore, the higher slowness values in the 6C dispersion curves at high frequencies can be explained by the different spatial resolutions of the methods, as well as the higher sensitivity of rotational motions to small-scale structures. We find the most consistent inversion results, when inverting the dispersion curves together with ellipticity ratios, linking v_{P} to v_{S}, and constraining the parameter space to a three layer model. The resulting P- and S-wave velocity profiles are comparable for all methods and show in general a positive correlation to the lithology, which supports the inversion results. These are important findings, as a single-station measurement is very simple in terms of logistics compared to an array set-up, which makes measurements for example in an urban environment much easier. We conclude that the developed 6C method provides a fast and simple way for site characterization. As a further step, the Vs30 values are calculated from the velocity profiles and yield relatively low values, in accordance to the underlying sedimentary layers. Based on that the study area is mainly categorized into the site classes C and D, which characterize the subsurface as stiff to dense soils according to the IBC. These findings indicate an amplification potential for seismic waves, which should be investigated in future studies.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Acknowledgements

The research of this study was supported by the research project SEIGER (Project. no. 03EE4003G), funded by the German Federal Ministry of Economic Affairs and Climate Action and managed by the research institute Julich (PTJ). The authors are grateful to Manuel Hobiger, who helped to improve this paper with fruitful discussions. In addition, the authors thank the reviewer David Sollberger and an anonymous reviewer for the critical assessment of this work, which helped to improve this manuscript significantly.

2.2.A Appendix: Inversion targets station TWI

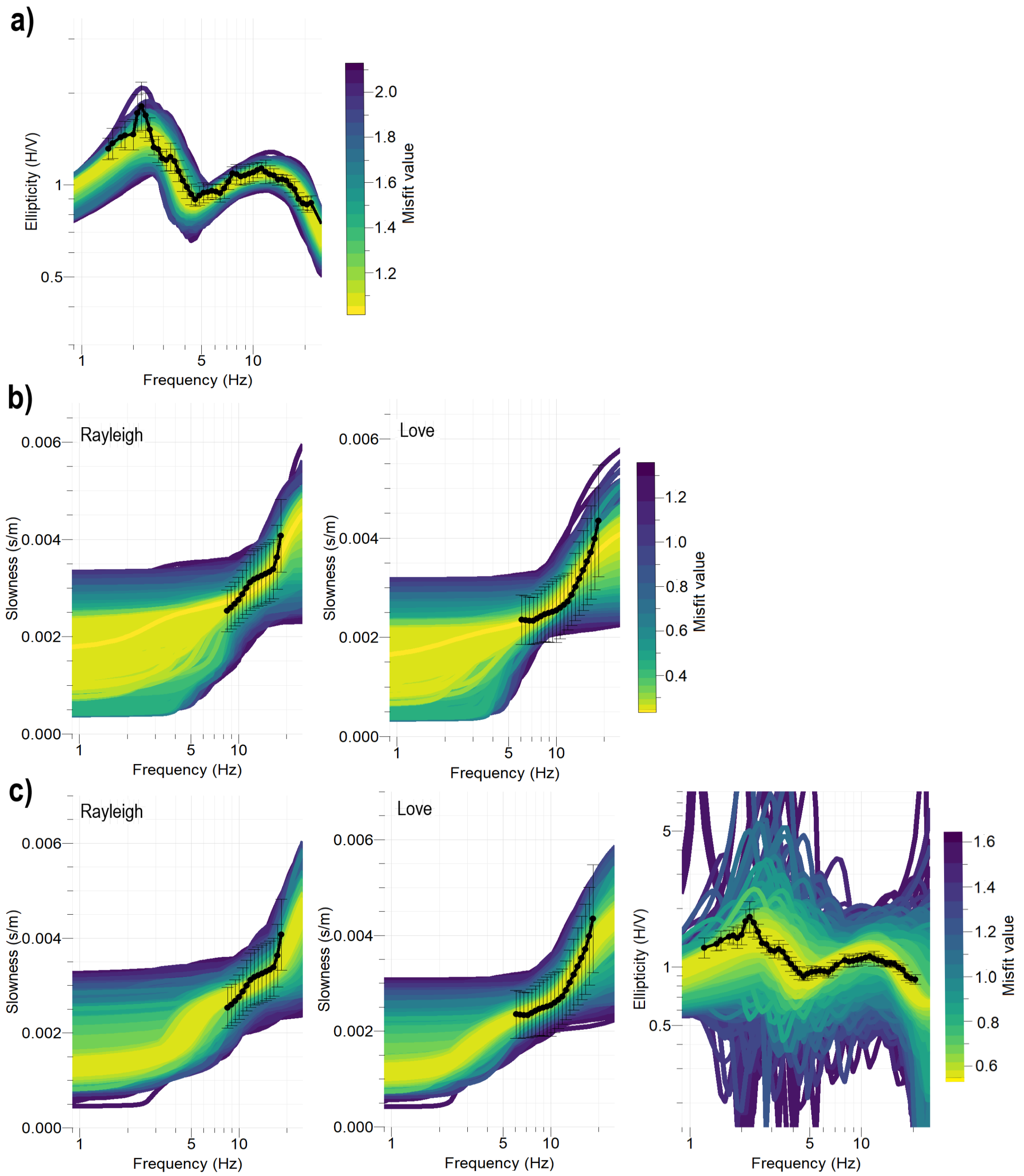

Figure 2.A1: The plots show the dispersion and ellipticity curves corresponding to the models generated during the respective inversions in Fig. 2.17 of the main text. The color indicates the misfit value of the computed model and the black line with error bars indicate the target curves used for the inversion. Subplot a) corresponds to Fig. 2.17a), where only the ellipticity curve was used as input data. b) corresponds to Fig. 2.17b), where the Rayleigh and Love dispersion curves were used as input data and c) corresponds to Fig. 2.17c), where both dispersion curves together with ellipticity ratios were used as input data.

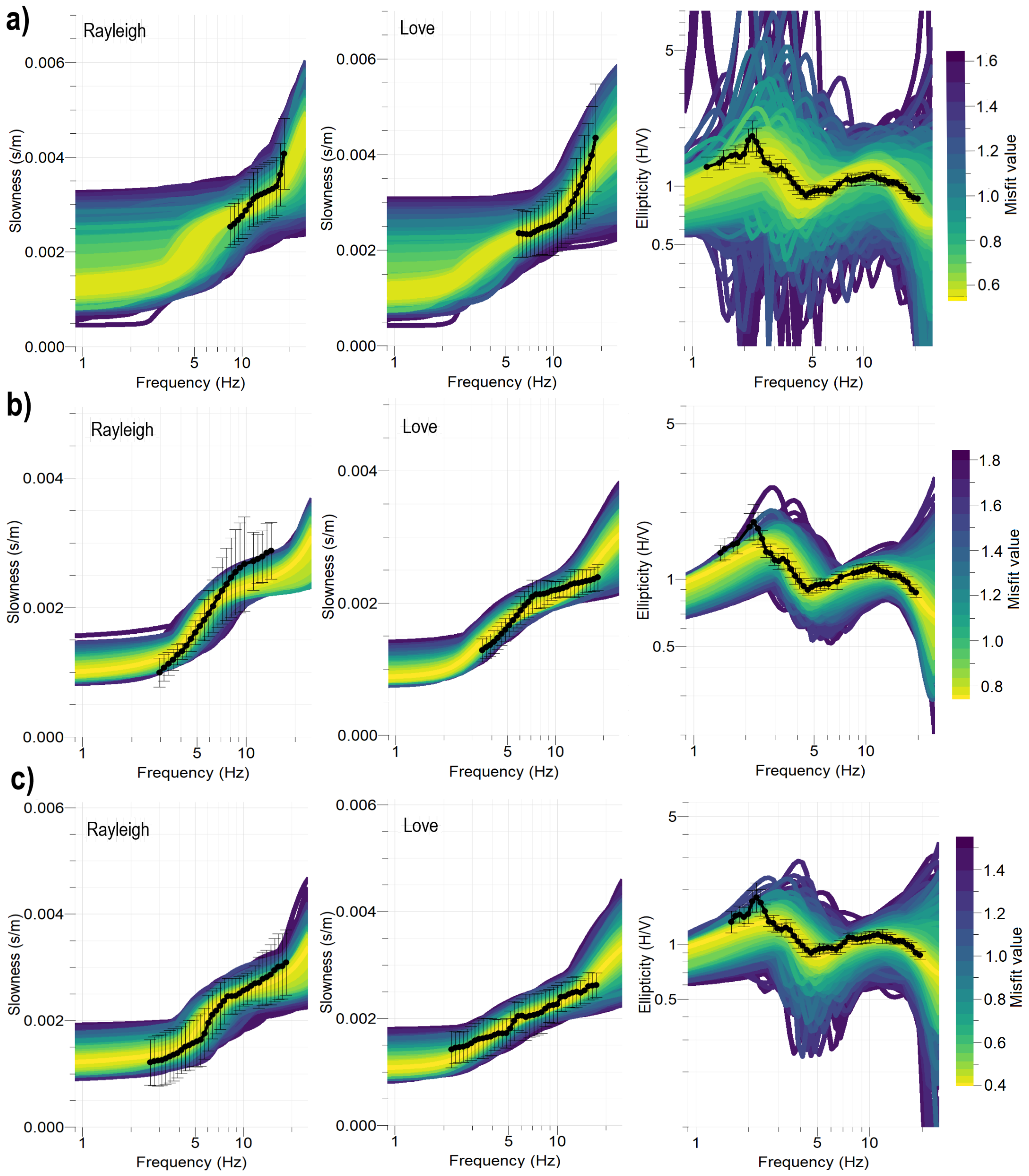

Figure 2.A2: The plots show the dispersion and ellipticity curves corresponding to the models generated during the respective inversions in Fig. 2.18 of the main text. The color indicates the misfit value of the computed model and the black line with error bars indicate the target curves used for the inversion. Subplot a) corresponds to the inversion run, where the 6C Rayleigh and Love dispersion curves were used as input data. b) corresponds to the inversion, where the HRFK dispersion curves were used and c) corresponds to the inversion, where the SPAC dispersion curves were used.

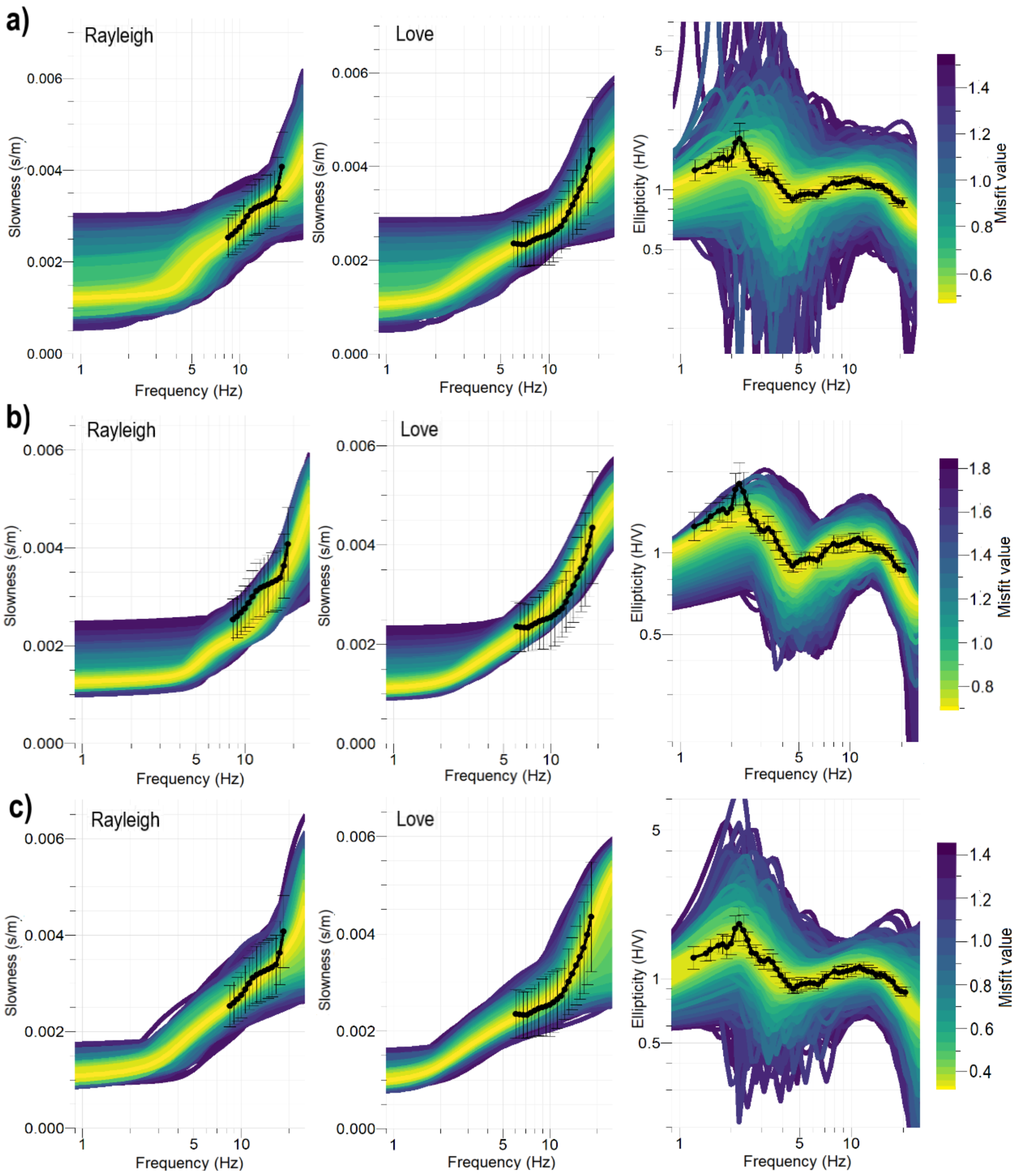

Figure 2.A3: The plots show the dispersion and ellipticity curves corresponding to the models generated during the respective inversions in Fig. 2.19 of the main text. The color indicates the misfit value of the computed model and the black line with error bars indicate the target curves used for the inversion. Subplot a) corresponds to the inversion run, where the inversion was constrained to a four layer model. b) corresponds to the inversion, where the P-wave velocity below the groundwater table was constrained to >1400 m/s and c) corresponds to the inversion, where a velocity gradient was used.

2.2.B Appendix: Additional inversion results and inversion targets station EGA

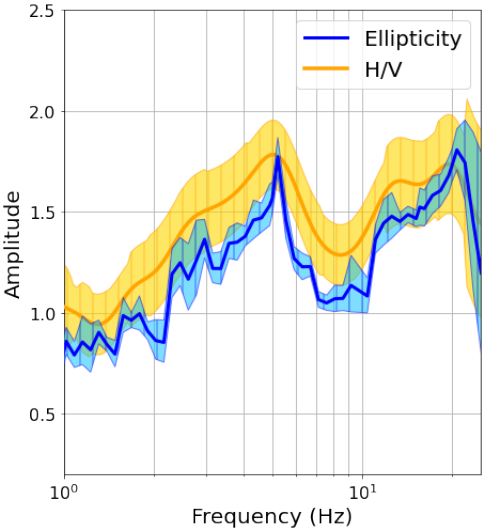

Figure 2.B1: Comparison of classical H/V curves computed with GEOPSY (Wathelet et al. 2020) and RayDec (Hobiger et al. 2009) Rayleigh wave ellipticity curves at station EGA. The shading represents the standard deviation of the curves.

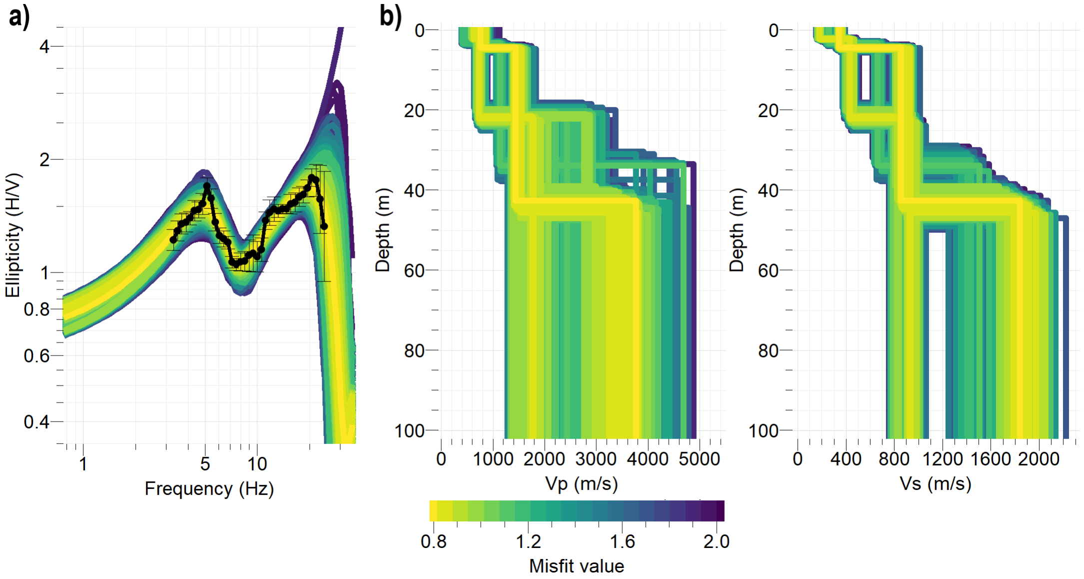

Figure 2.B2: Inversion results for a three layer inversion at station EGA using only the ellipticity curve as inversion target. The colors indicate the misfit value of the generated models. Subplot a) shows the ellipticity curves generated during the inversion. The black solid line shows the input data with their standard deviation. In b) the generated P- and S-wave velocity profiles are shown.

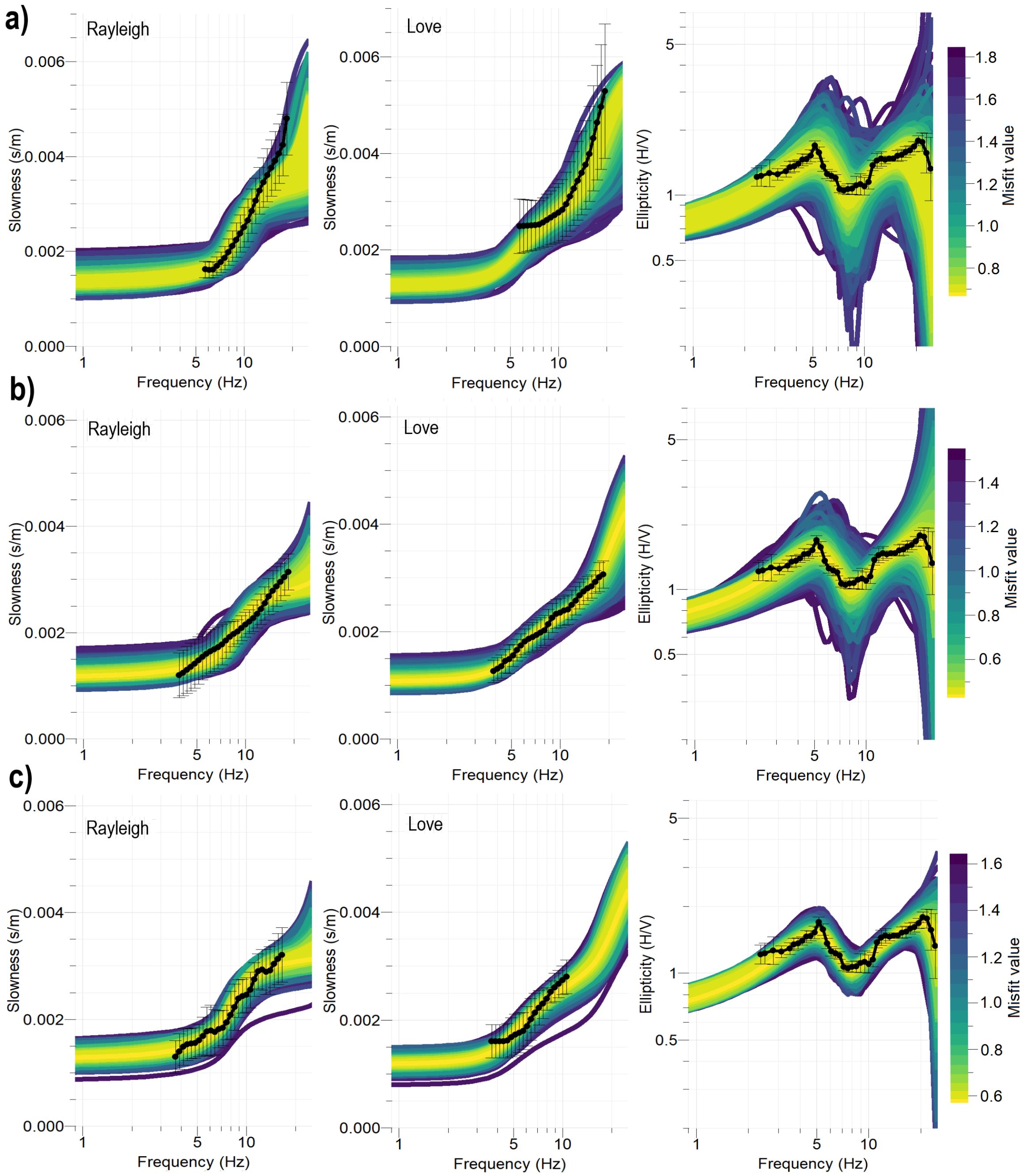

Figure 2.B3: The plots show the dispersion and ellipticity curves corresponding to the models generated during the respective inversions in Fig. 2.22 of the main text. The colors indicate the misfit value of the computed models and the black solid lines with error bars indicate the target curves used for the inversion. Subplot a) corresponds to the inversion, where the 6C Rayleigh and Love dispersion curves were used as input data. b) corresponds to the inversion, where the HRFK dispersion curves were used and c) corresponds to the inversion, where the SPAC dispersion curves were used.

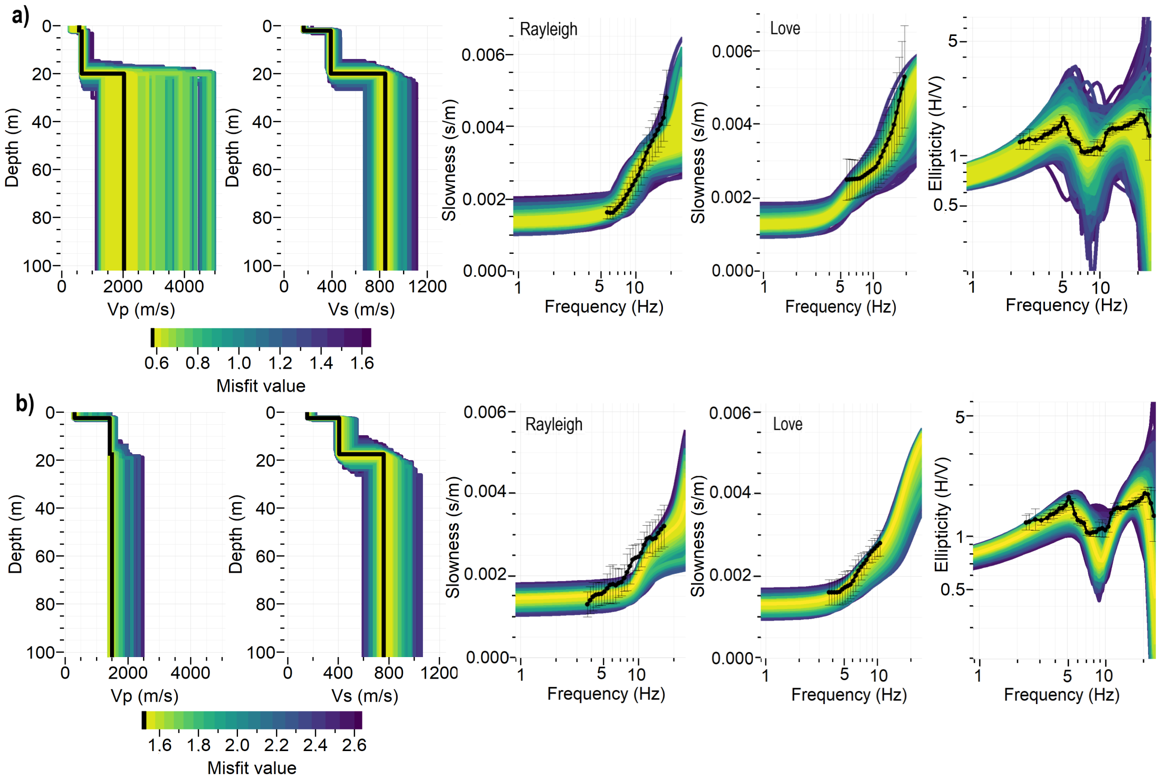

Figure 2.B4: P- and S-wave velocity profiles together with inversion targets for a three-layer inversion at station EGA using ellipticity ratios and the SPAC dispersion curves. The color scale gives the misfit of the computed models. The black line in the velocity profiles marks the model with the smallest misfit. The black lines in the inversion target plots show the input data together with the standard deviation. In a) the inversion was only constrained by linking v_P to v_S. In b) the inversion was additionally constrained by enforcing v_P to be larger than 1400 m/s below the groundwater table.

2.3 Surface wave dispersion curve inversion using mixture density networks

by Sabrina Keil and Joachim Wassermann

Published in Geophysical Journal International (2023)

Abstract

In many seismological, environmental and engineering applications a detailed S-wave velocity model of the shallow subsurface is required. This is generally achieved by the inversion of surface wave dispersion curves using various inversion methods. The classical inversion approaches suffer from several shortcomings, such as inaccurate solutions due to local minima or large computation times in case of a wide parameter space. A number of machine learning (ML) approaches have been suggested to tackle these problems, which however do not provide probabilistic solutions and/or constrain layer number and layer thickness to a fixed value. In this study, we develop a novel neural network (NN) approach in order to characterize the shallow velocity structure from Love and Rayleigh wave dispersion curves. The novelty of our method lies in the simultaneous estimation of layer numbers, layer depth and a complete probability distribution of the S-wave velocity structure. This is achieved by a two-step ML approach, where 1) a regular NN classifies the number of layers within the upper 100 m of the subsurface and 2) a mixture density network outputs the depth estimates together with a fully probabilistic solution of the S-wave velocity structure. We show the advantages of our ML approach compared to a conventional neighborhood inversion and a Markov chain Monte Carlo algorithm. Our ML approach is then applied to dispersion curves extracted from recorded noise data in Munich, Germany. The resulting velocity profile is in accordance with lithologic information at the site, which highlights the potential of our approach.

2.3.1 Introduction

Estimating the shallow velocity structure quickly and with high accuracy is important for many environmental and engineering applications, including the assessment of earthquake site response (e.g Karagoz et al. 2015, Hartzell et al. 2001), the characterization of landslide activity (e.g Hibert et al. 2012, Jongmans et al. 2009), the assessment of soil liquefaction potential (e.g Lin et al. 2004) and the monitoring of shallow water content (e.g Bergamo et al. 2016, Lecocq et al. 2017). For these near-surface studies, the velocity structure is typically retrieved by inversion of surface wave dispersion curves, which are extracted from ambient vibration recordings (e.g Fäh et al. 2008, Scherbaum et al. 2003). The derivation of 1-D velocity profiles from surface wave data is a classical non-linear inverse problem in geophysics, which is often poorly constrained and multi-modal (e.g Cox & Teague 2016).

Therefore, the solutions are often non-unique and the choice of the parameterization (number of layers, range of velocity, and thickness values for the layers) is of vital importance for obtaining reliable results (e.g. Renalier et al. 2010, Foti et al. 2009). To solve this inverse problem several approaches have been proposed, including linear methods, non-linear global search techniques and Bayesian methods (e.g. Gosselin et al. 2022, Sen & Stoffa 2013, Tarantola 2005).

Compared to non-linear methods, linearized inversion typically requires fewer forward operations since they exploit misfit gradient information rather than employ directed random searches, which makes them computationally efficient (e.g. Gosselin et al. 2022). Parolai et al. (2006) could show that, given a relatively accurate starting model, linearized inversion yielded similar S-wave profiles as non-linear methods. Nevertheless, linearized approaches are prone to become trapped in a local minimum if the problem is strongly non-linear and/or if the starting model is poor.

This problem is overcome by global search methods, which apply a direct random search over a wide parameter space. Methods which belong to this family are Simulated Annealing (SA) (Laarhoven & Aarts 1987), Genetic Algorithms (GA) (e.g Lomax & Snieder 1994) and the Neighborhood Algorithm (NA) (Sambridge 1999). Compared to basic Monte Carlo sampling, these approaches try to guide the random generation of samples based on the results of previous samples, which makes them computationally more efficient. The NA has been used in many 1D seismic site characterization studies (e.g. Wathelet et al. 2004, Keil et al. 2022a, Hobiger et al. 2021, Wood et al. 2019) due to the availability of the user-friendly software implementation GEOPSY (Wathelet 2008, Wathelet et al. 2020). Also SA has been used in many near-surface seismic studies (e.g. Beaty et al. 2002, Pei et al. 2007, Lu et al. 2016), as well as GA (e.g. Parolai et al. 2005, Yamanaka & Ishida 1996). One drawback of these methods is that the algorithm performance is based on the subjective choice of tuning parameters (e.g. annealing schedule in SA, evolutionary step in GA), which control the balance between exploration and exploitation of the parameter space. In addition, these methods only locate the optimal solution, but are not able to estimate uncertainties, which however is important due to the non-uniqueness in the inversion results.

An alternative approach is Bayesian inversion based on probabilistic sampling over the parameter space to estimate properties of the posterior probability density of the models (e.g. Gosselin et al. 2018, Molnar et al. 2010). Bayesian inversions provide parameter estimates together with a quantitative uncertainty analysis, typically employing Markov chain Monte Carlo (McMC) methods (Brooks et al. 2011). Studies that apply Bayesian inversion methods for probabilistic site characterization include e.g. Molnar et al. (2010), Gosselin et al. (2018) and Hallo et al. (2021). This approach provides the most informative solution, however, it is computationally demanding, especially when a large parameter space is sampled and/or 2D or 3D problems are solved, as is the case for tomography studies (e.g. Galetti et al. 2017).

To overcome these problems, several machine learning (ML) approaches have been applied to perform the non-linear inversion of surface wave dispersion curves to estimate the subsurface structure. Chen et al. (2022) desgined a convolutional neural network (CNN) to directly invert fundamental-mode Rayleigh wave dispersion curves for the shallow Vs structure. The advantage of this approach is that the neural network can be trained quickly and outputs directly the Vs model. Hu et al. (2020) trained a CNN using both Rayleigh wave phase and group velocities to invert for the deep velocity structure beneath continental China and southern California. They showed that such a ML approach is suitable to invert big datasets for tomography studies faster than conventional inversion methods. Yablokov et al. (2021) designed a fully connected artificial neural network (ANN) to predict the velocity and thickness values for near surface layers and applied it to data from West Siberia. They showed that the accuracy of the proposed approach outperforms that of a Monte Carlo inversion. The main disadvantage of such classical neural network structures (ANN, CNN) is that they only provide deterministic velocity estimates and are not able to model uncertainties. This problem can be solved using a specific class of NN, called mixture density networks (MDN) (Bishop 1994), which output fully probabilistic solutions. Earp et al. (2020) applied MDNs to estimate the deep velocity structure from surface wave dispersion data for a tomography study in the Grane field in the Norwegian North sea. They showed that this approach provides a more efficient way to obtain probabilistic solutions compared to Monte Carlo methods, which make it suitable for rapid, repeated 3D subsurface imaging and monitoring. Cao et al. (2020) applied MDNs to invert Rayleigh wave dispersion curves for the near-surface velocity structure. They demonstrated that the trained ML model is able to invert a large number of dispersion curves into a 3D velocity model in a matter of seconds. However, in both studies the MDNs are trained for a fixed number of subsurface layers and fixed thickness values, which is a problem at sites where these values are unknown. In addition, they train separate MDNs for each depth interval, which raises the training costs. Meier et al. (2007) applied MDNs to invert surface wave data for a global model of crustal thickness. They found that their results are in good agreement with other crustal models, while the MDN inversion significantly outperforms sampling based techniques. These MDN studies perform the inversion of layer thickness or velocities, but not both simultaneously, requiring always one of these parameters to be fixed.

To overcome this limitation we propose a ML approach for the estimation of the shallow subsurface structure, which outputs both the layer thickness and the S-wave velocity by inverting Love and Rayleigh wave dispersion curves in a frequency range of 1-20 Hz. This is a two-step approach, where 1) a common classification NN predicts the optimum number of layers within the upper 100 m of the subsurface and 2) a MDN outputs the layer depth estimates together with the probability distribution of the S-wave velocity structure. The NN is trained to distinguish between structures with 2 - 7 layers, including the half-space. For each number of layers individual mixture density networks are trained to output the subsurface structure. The MDNs invert for the complete velocity structure instead of single subsurface layers, as was done in e.g. Earp et al. (2020) and Cao et al. (2020), which makes the training process computationally more efficient. We test the trained networks on synthetic data, as well as on recorded noise data from Munich, Germany, and compare the results to a classical neighborhood inversion using the GEOPSY software (Wathelet 2008, Wathelet et al. 2020), as well as a Markov chain Monte Carlo (McMC) method implemented by Alder et al. (2021).

2.3.2 Methodology

2.3.2.1 Mixture density networks

Neural networks are mathematical models that can be trained to find a non-linear mapping between input data **d** and output parameters **m**. During the training the network is presented with a set of data-model pairs and minimizes a cost function that measures the difference between the NN output and the true output. While conventional neural networks can represent arbitrary non-linear functions, mixture density networks (MDN) can model complete probability distributions (Bishop 1994). They are trained on the same data-model pairs, but instead of providing one model estimate they output the Bayesian posterior probability distribution $p(\mathbf{m}|\mathbf{d})$. The probability density of the target data is represented by a mixture of Gaussian kernels (Bishop 1994):

$$p(\mathbf{m}|\mathbf{d}) = \sum_{i=1}^{M} \alpha_i(\mathbf{d})\phi_i(\mathbf{m}|\mathbf{d}) \tag{2.19}$$

where M is the number of Gaussians in the mixture, α_i are the mixing coefficients, which give relative importance to each Gaussian kernel and ϕ_i are the Gaussian density functions:

$$\phi_i(\mathbf{m}|\mathbf{d}) = \frac{1}{(2\pi)^{c/2}\sigma_i^c(\mathbf{d})} exp\Big\{-\frac{1}{2}\sum_{k=1}^{c}\frac{(m_k-\mu_{ik}(\mathbf{d}))^2}{\sigma_i^2(\mathbf{d})}\Big\} \tag{2.20}$$

where c is the dimensionality of **m**, μ_{ik} is the mean and σ_i the standard deviation. The number of mixtures M controls the complexity of the final probability distribution and therefore also the number of outputs, which is given by $(c+2)\times M$, compared to c outputs of a standard NN. The MDN estimates values for the mixture parameters α_i, means μ_{ik} and standard deviations σ_i, which fully define the set of Gaussian kernels. Hereby, it is assumed that the components of the output vector are statistically independent within each component of the distribution and can be described by a common standard deviation. The number of kernels that should be used in the mixture depends on the complexity of the problem. However, a greater number of kernels than necessary will have little negative effect, since the network can reduce the amplitude of any mixture parameter to near zero for redundant kernels. During the training of the network the weights are adjusted to maximize the likelihood of the desired probability density function. The error function to be minimized is the negative log likelihood function (Bishop 1994):

$$E = -\sum_{n=1}^{N} ln\Big\{\sum_{i=1}^{M}\alpha_i(d_n)\phi_i(m_n|d_n)\Big\} \tag{2.21}$$

For a more detailed description of MDNs we refer the reader to Bishop et al. (1995).
We train individual MDNs for each group of velocity structures, depending on the number of subsurface layers. Each MDN is parameterized by a mixture of 10 Gaussian kernels. This number was chosen in order to capture the complexity of the 7 layer structures. For structures with less subsurface layers the MDN generally assigns significant non-zero values only to a few kernels. Nevertheless, the additional kernels have no negative effect on the performance of the network. However, choosing a significantly higher number would increase the time needed for training. As an example, the MDN for velocity structures with 5 subsurface layers returns $(5 + 2) \times 10 = 70$ output parameters, which define the probability distribution of the S-wave velocity in each layer. As mean velocity estimate we take the maximum of the probability distribution. The layer depth estimates are output separately as deterministic values. The network structure is explained in more detail in section 2.3.2.3.

2.3.2.2 Creating the training sets

In order to train the neural networks, a set of velocity-dispersion curve pairs has to be created. We generate individual training sets for velocity structures with 2 - 7 layers. The velocity structures reach a depth of 100 m and are parameterized depending on the number of layers (Table 2.1). The S-wave velocity of the uppermost layer (v_{top}) was randomly selected from the uniform probability distribution $U(100m/s, 500m/s)$ to represent near-surface low velocity sediments. We force the uppermost layer to have the lowest S-wave velocity in order to create fundamental mode Love and Rayleigh waves (e.g Lowrie 2007). Furthermore, the uppermost layer has a minimum thickness of 5 m. The S-wave velocities of the following layers were randomly selected from the uniform probability distribution $U(v_{top} + 100m/s, 1600m/s)$ with minimum thicknesses of 10 m. We force the half-space to have the largest velocity by selecting a value from the uniform probability distribution $U(max(\mathbf{v}) + 200, 2000m/s)$, where $\mathbf{v}$ is the vector of velocities of the overlying layers. The layer boundary of the half-space has a maximum depth of 85 m, as we expect that the dispersion data in the selected frequency range are not able to resolve deeper layers. We generate different amounts of velocity models depending on the number of subsurface layers (Table 2.2). More models are created for complex structures (more layers) to cover the larger range of possible velocity profiles. From the generated S-velocity structures we compute the corresponding P-wave velocity and density profiles based on relations found for sedimentary rocks (Castagna et al. 1985, Brocher 2005):

$$v_P = 1.16 v_S + 1.36 \tag{2.22}$$

$$\rho = 1.74 v_P^{0.25} \tag{2.23}$$

The forward problem is solved for each of the synthetic 1D velocity structures using the package gplivemodel, which is part of the GEOPSY software (Wathelet et al. 2020). The resulting fundamental Love and Rayleigh wave dispersion curves are then logarithmically resampled with 100 samples in a frequency range of 1-20 Hz. The clean dispersion curves are then augmented with 0-5% varying levels of noise according to:

$$u_j = U(0, 0.05) \times d_j \tag{2.24}$$

$$\tilde{d}_j = U(-1, 1)u_j + d_j \tag{2.25}$$

where u_j is the uncertainty value of the noisy data $\tilde{d}_j$ and U is a random uniform distribution. Adding noise helps to generalize when the network is presented with new data and accounts for the data uncertainty in the Bayesian solution. An example for a generated synthetic velocity profile and its corresponding Love and Rayleigh wave dispersion curves with added noise are shown in Fig. 2.23. Earp et al. (2020) showed that including data uncertainties as separate input vectors in the mixture density network gives more reliable mean velocity estimates than only inputting the noisy dispersion curves. Therefore, we save the uncertainty values u_j in an uncertainty vector, which is then used as additional input into the NN. The final data sets are then standardized in order to remove the mean and scale to unit variance, which improves the network performance. As a final step the scaled data set is split into 90% training data, that is used to train the network and 10% validation set to check the performance during the training.

Table 2.1: Parameterization of the synthetic velocity structures used for the training of the neural networks.

Layer	**S-velocity (m/s)**	**Thickness (m)**
Top	$100 - 500$	$5 - 85$
2 - 6	$(v_{top} + 100) - 1600$	$10 - 80$
Half-space	$(max(\mathbf{v}) + 200) - 2000$	$infinite$

Table 2.2: Number of models in the training set of the different neural networks.

Network	**Number of models**
classification NN	600 000
2 layer MDN	100 000
3 layer MDN	150 000
4 layer MDN	250 000
5 layer MDN	340 000
6 layer MDN	400 000
7 layer MDN	450 000

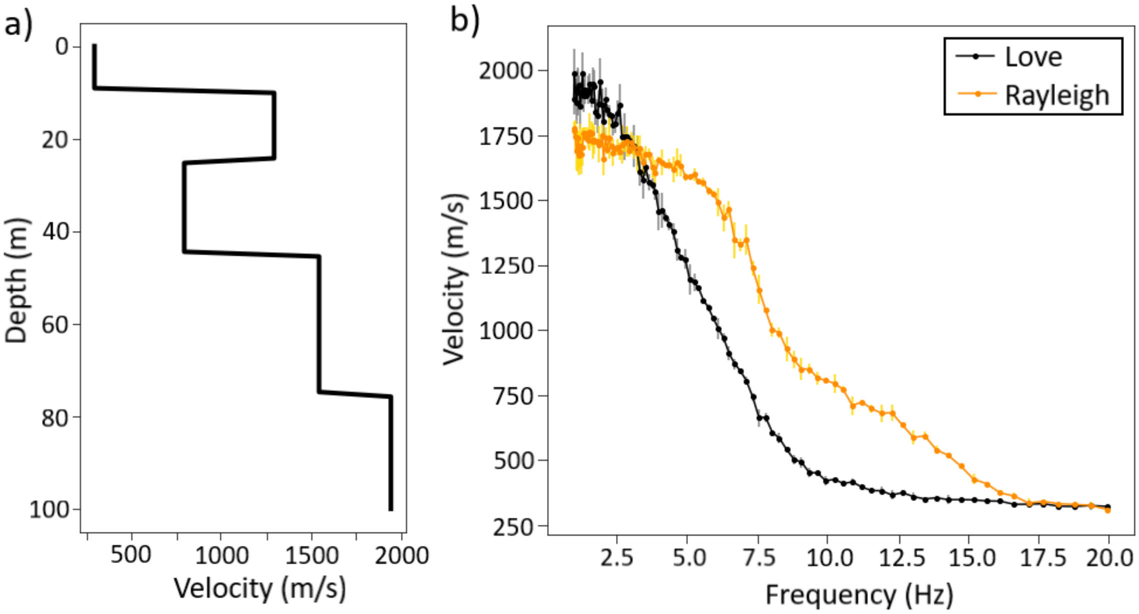

Figure 2.23: Example of training data pair, where a) shows a synthetic 5-layer S-wave velocity profile and b) are the corresponding forward modelled Love and Rayleigh dispersion curves with added noise. Vertical bars show the data uncertainties.

2.3.2.3 Network design and training

After creating the training data sets the individual mixture density networks and the classification NN are trained. A simplified network structure for the mixture density network is shown in Fig. 2.24. The network contains four input layers for the Love dispersion curve, Love uncertainty vector, Rayleigh dispersion curve and Rayleigh uncertainty vector. The input data are passed through one or two fully connected (dense) layers followed by a dropout layer for generalization. The outputs are step-wise concatenated and further passed trough several dense and dropout layers. Afterwards the network splits into one path leading to the layer depth output and a second path, which includes the MDN layer to output the probability distribution of the shear wave velocity structure. The MDN layer consists of three dense layers for the estimation of the mixture parameters α_i, means μ_i and standard deviations σ_i. The rectified linear unit (ReLU) activation function, which is a piece-wise linear function that will output the input directly if it is positive, otherwise, it will output zero, is implemented in each dense layer, except for the α_i layer in the MDN. Here a softmax activation function is used since the mixing coefficients must sum to unity. The loss function for the MDN output is the negative log likelihood (Eq. 2.21) and the loss function for the depth output is the mean squared error.

In order to find the optimum number of neurons in each dense layer and also the number of subsequent dense layers, the Python library Hyperas (Pumperla 2022) is used, which is a wrapper around the library Hyperopt (Bergstra et al. 2015) for hyperparameter optimization. The exact number of neurons in each layer and the number of dense layers in each trained network can be found in Table 2.C1 and 2.C2 in the appendix 2.3.C.

The classification NN is constructed in a similar way, but contains only one output layer for the prediction of the number of subsurface layers (Fig. 2.C1 in appendix 2.3.C). As loss function the mean squared error is implemented.

In order to test the performance of the trained neural networks, 4000 previously unseen dispersion curves and uncertainty vectors are input into each network. The classification NN predicts the number of subsurface layers. The difference between the predicted number compared to the true number of layers is shown in Fig. 2.25a). More than 60% of the test models are estimated correctly, while approximately 30% have an error of ±1. Only few percent have an error of ±2 or ±3. The correlation between the predicted layer number and the true number of subsurface layers can be used to evaluate the performance of the network (Fig. 2.25b). The corresponding Pearson correlation coefficient (r) has a value of 0.90. The fit for intermediate numbers of layers (4 - 6) seems to be lower than for 2, 3 and 7 layers.

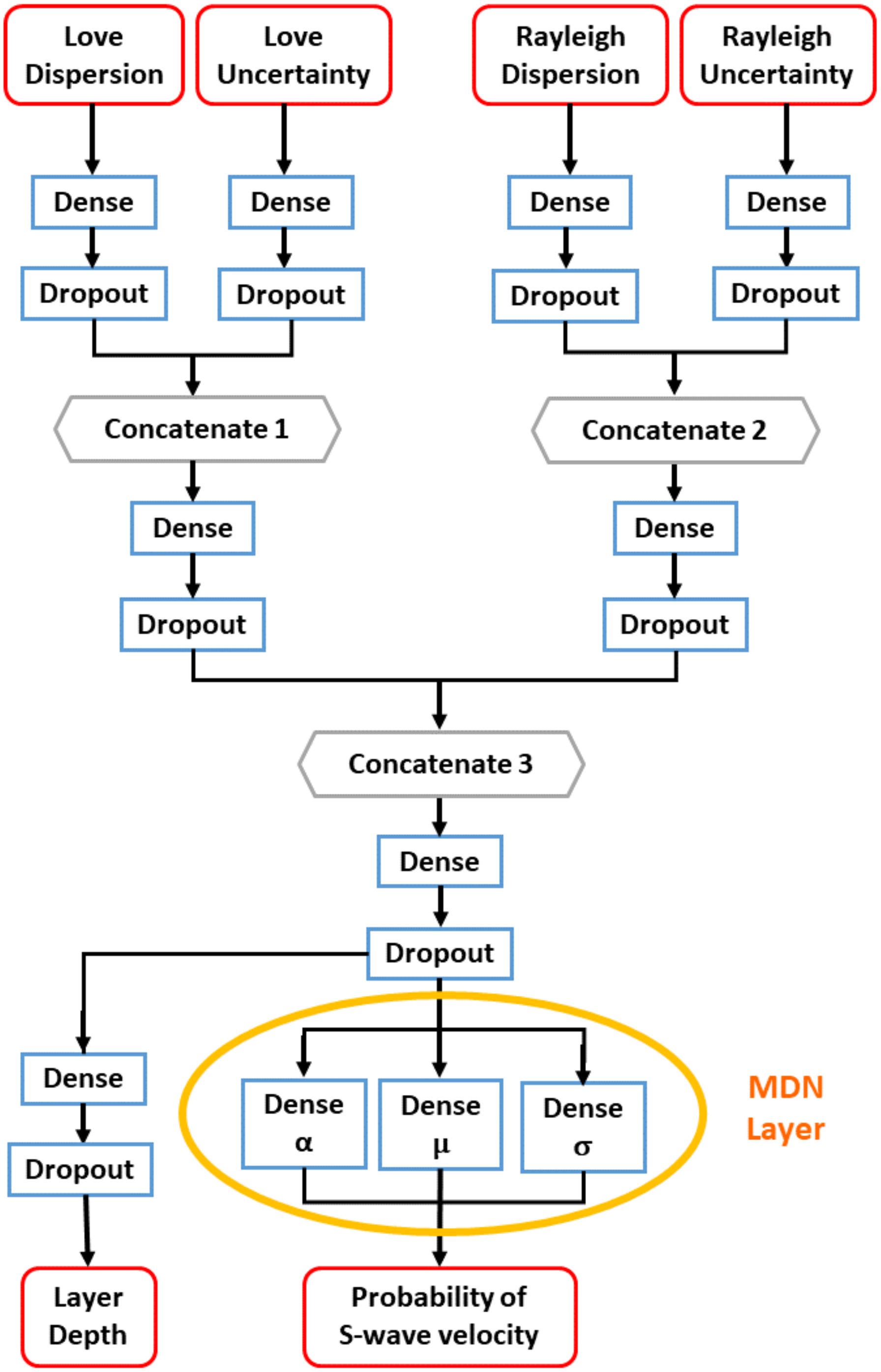

Figure 2.24: Schematic network structure used for the training of the mixture density network. Red rounded edged boxes represent the inputs and outputs of the network. Blue squared edged boxes represent one or more Dense and Dropout layers. The gray hexagonal boxes represent concatenation layers. The arrows represent the flow direction of data through the network. The MDN layer is highlighted and includes 3 Dense layers for the computation of α, μ and σ.

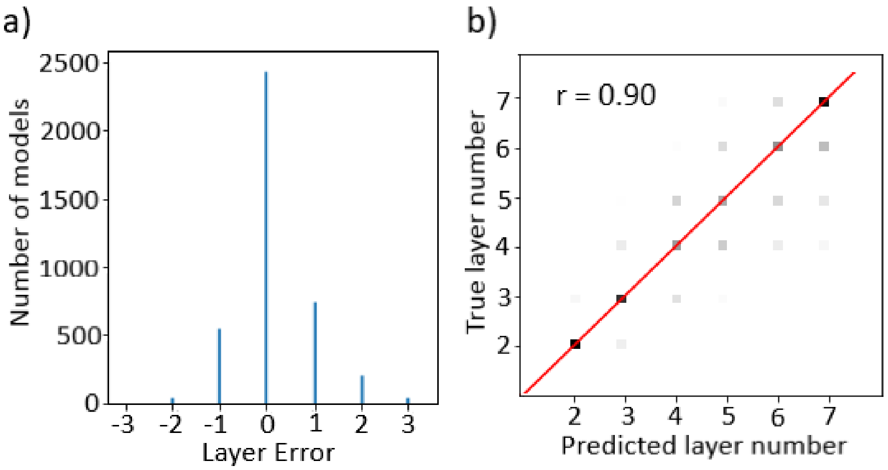

Figure 2.25: Performance of classification NN for a test data-set of 4000 models. a) Histogram of differences between predicted and true number of subsurface layers. b) Pearson correlation between predicted and true layer numbers. The shading represents the density of models that fall into each category. Darker colors correspond to more models. The Pearson correlation coefficient r is written in the upper left corner. The line shows the optimal correlation.

The MDN networks output the probability distribution of the S-wave velocity structure together with depth estimates for the layer boundaries. For each layer we pick the maximum of the probability density function as mean velocity estimate. To evaluate the performance of the individual MDNs the correlation between the true velocity and the predicted mean velocity are computed for each layer (Fig. 2.26). In general, the Pearson correlation coefficient for simple profiles (2 - 4 layers) is high in all of the layers ($r > 0.84$), while for more complex structures the correlation in the deeper layers decreases ($0.62 < r > 0.87$). The Pearson correlation coefficient for the half-space is high ($r > 0.84$) for all of the MDNs. The same holds for the depth estimates of the layer boundaries (Fig. 2.27). For simple profiles (2-4 layers) the Pearson correlation coefficient is high ($r > 0.80$), while for more complex structures the correlation is lower especially for deeper layer boundaries ($0.29 < r > 0.67$). Reasons for this will be discussed in section 2.3.4.

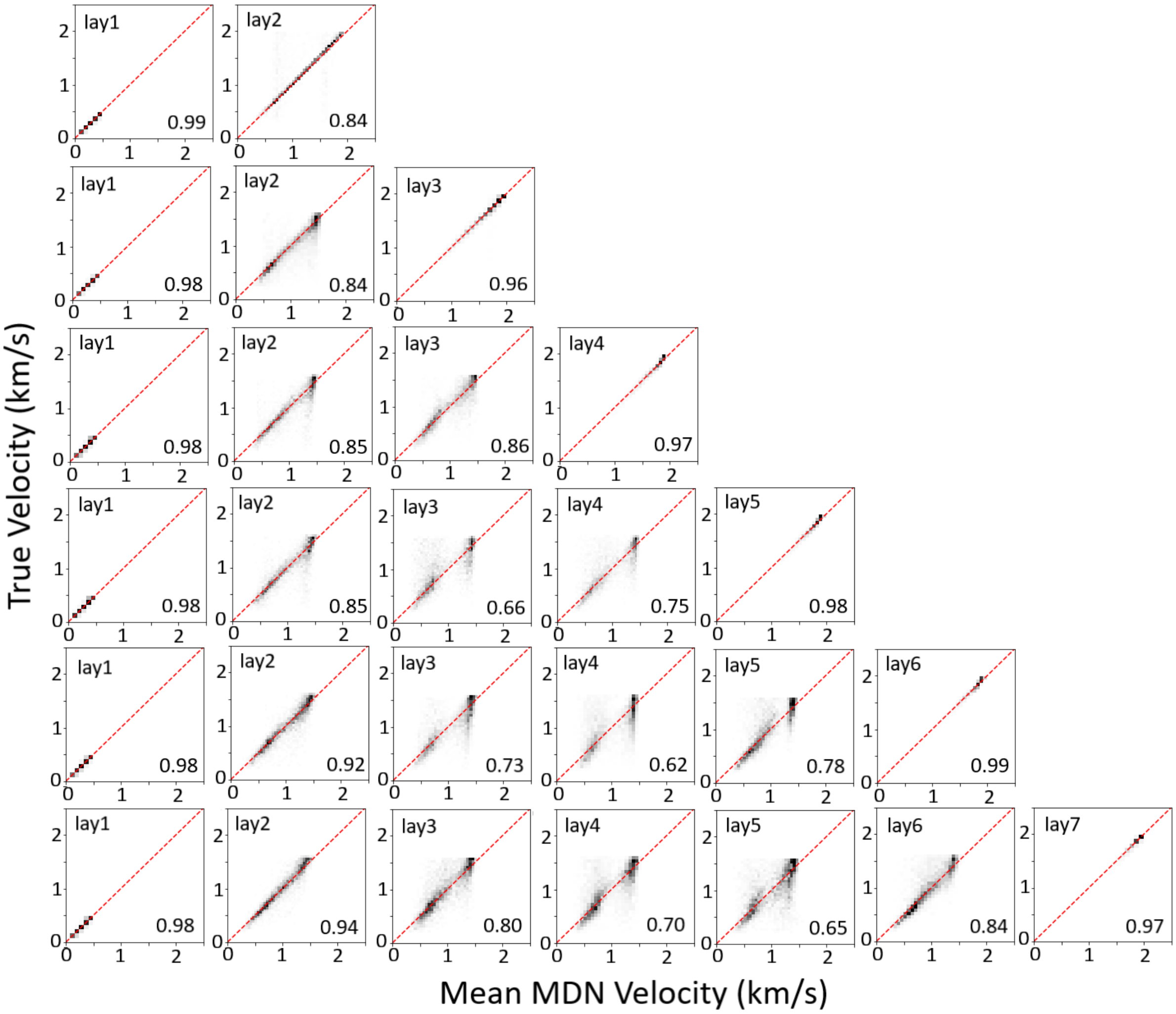

Figure 2.26: Correlation between the true velocity values and the mean velocity estimates from the MDNs for a test data-set of 4000 models. Each row shows the results of a separately trained MDN for the different numbers of subsurface layers, starting at the top row with 2 layer models until the bottom row for 7 layer models. The layer number within the profile is indicated in the upper left corner. The Pearson correlation coefficient (r) is given in the lower right corner. The line shows the optimal correlation.

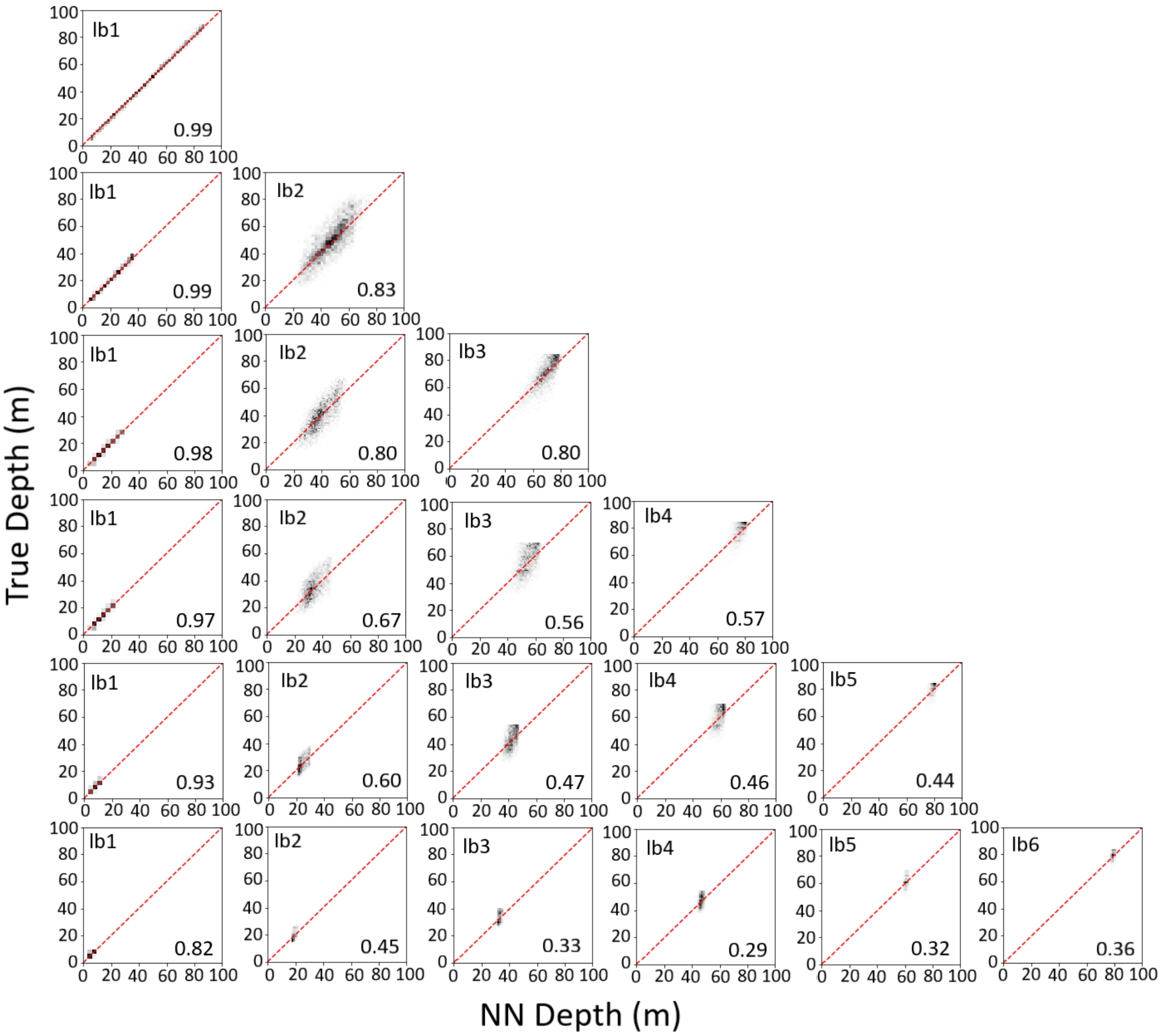

Figure 2.27: Correlation between the true layer boundary depth and the predicted layer depth from the MDNs for a test data-set of 4000 models. Each row shows the results of a separately trained MDN network for the different numbers of subsurface layers, starting at the top row with 2 layer models until the bottom row for 7 layer models. The layer boundary (lb) number within the profile is indicated in the upper left corner. The Pearson correlation coefficient (r) is given in the lower right corner. The line shows the optimal correlation.

2.3.3 Results

2.3.3.1 Synthetic data

First we test the trained models on synthetic dispersion data. In Fig. 2.28 two examples for predicted velocity profiles are shown. The MDN results are posterior distributions, subject to a prior distribution that is defined by the training set. The comparison of the posterior probability function and the prior probability function is plotted. The structure in Fig. 2.28a) was predicted by the MDN trained for 3 layer models. The posterior probability density distributions are narrow as only few kernels with different mean and standard deviations were assigned significantly non-zero amplitudes by the MDN. The true velocity values and depth values are matched closely by the MDN-predictions. The structure in Fig. 2.28b) was predicted by the MDN trained for 7 layer models. The posterior probability distributions are broader and the contribution of different kernels can be seen, which results in several peaks. This could indicate an ambiguity in the inversion results. Nevertheless, the true velocity and the layer depth are mostly recovered.

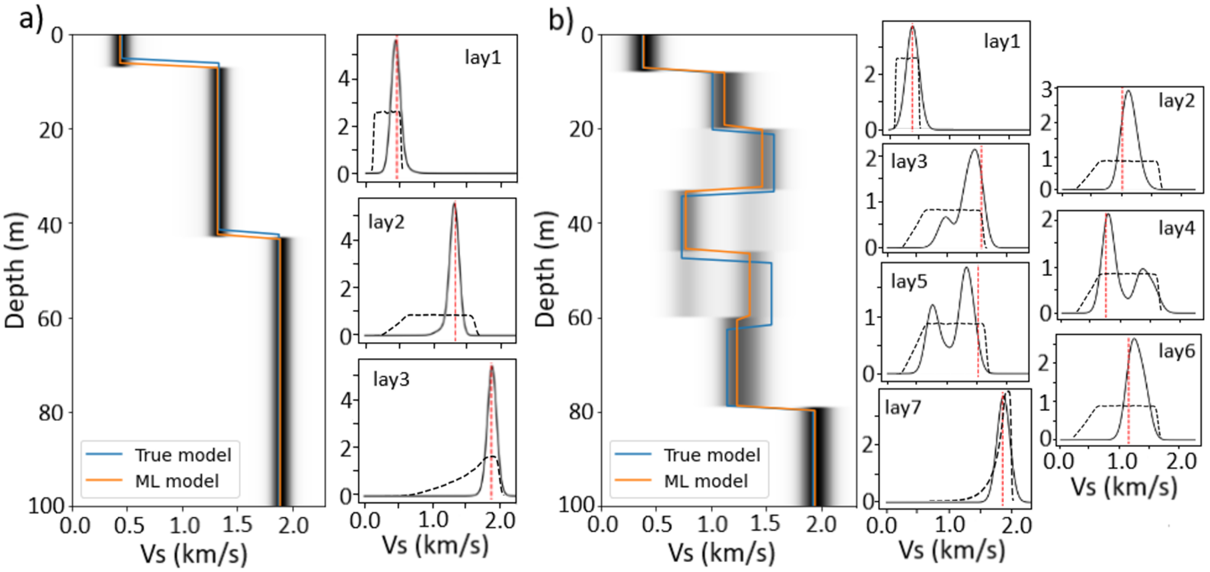

Figure 2.28: Examples for velocity structures predicted by the a) 3 layer MDN and the b) 7 layer MDN. The shading represents the probability density distribution, where darker colors represent areas of higher probability. The kernel density estimates for each layer are shown to the right. The solid line is the posterior probability density function from the MDN, the red vertical dashed line marks the true velocity value, and the black dashed line is the prior probability density function.

In the next step we compare our machine learning approach with two common inversion methods. First the DINVER module of the GEOPSY software (Wathelet 2008) is used, which implements a neighborhood algorithm. This is a user-friendly open-source tool set for ambient vibration processing (Wathelet et al. 2020), however, it does not allow the number of subsurface layers to be set as a free parameter. This leads to some extent to a subjective choice of the optimal number of subsurface layers. The algorithm returns the optimum solution together with a misfit value for the computed models. As second inversion method we use a Markov chain Monte Carlo (McMC) algorithm implemented by Alder et al. (2021). Hereby, the inverse problem is treated in a Bayesian framework, where the solution is represented by the probability of the computed models. The algorithm allows the velocity, layer thickness and number of layers to be set as free parameters. To have a fair comparison with our ML approach, we constrain the parameter space to a maximum of 7 subsurface layers and S-wave velocities in the range of 100 to 2000 m/s. We test the three methods on two synthetic structures, including a simple two layer model (Fig. 2.29a - c) and a more complex five layer model (Fig. 2.29d - e).

First of all, the classification NN predicts the correct number of layers for the two layer profile (Fig. 2.29a). In the next step, the MDN trained for two layer models estimates the layer boundary depth together with the S-wave velocity distribution in each layer. The resulting mean velocity values and the depth of the layer boundary are close to the true values (Fig. 2.29a). The inversion results are returned within a second. The GEOPSY inversion also returns a velocity profile close to the true structure (Fig. 2.29b) and the inversion targets (input dispersion curves) are fitted well (Fig. 2.D1b in the appendix 2.3.D). The inversion process took approximately 30 seconds for a total of 10 000 tested models. The probability distributions returned by the McMc algorithm resemble the true two layer structure (Fig. 2.29c). For a total of 200 000 visited models the algorithm required about 2 CPU hours.

In case of the five layer profile, the classification NN again predicts the correct number of subsurface layers. The probability distributions of the S-wave velocity are wider, however, the mean values are close to the true velocity structure and the depth estimates have errors of less than 5 m (Fig. 2.29d)). Once more the result is returned in a second. The results for the GEOPSY inversion are shown in Fig. 2.29e). Even though the inversion targets are fitted well (Fig. 2.D1e), the resulting velocity structure deviates from the true model. The uppermost velocity step is recovered well, while the true layer boundary at 35 m depth is shifted and the velocity step at 60 m is not resolved. For 30 000 tested models the algorithm required approximately 60 seconds. The probability distributions returned by the McMC algorithm mostly resemble the true structure (Fig. 2.29f). For a total of 500 000 visited models the algorithm required about 7 CPU hours.

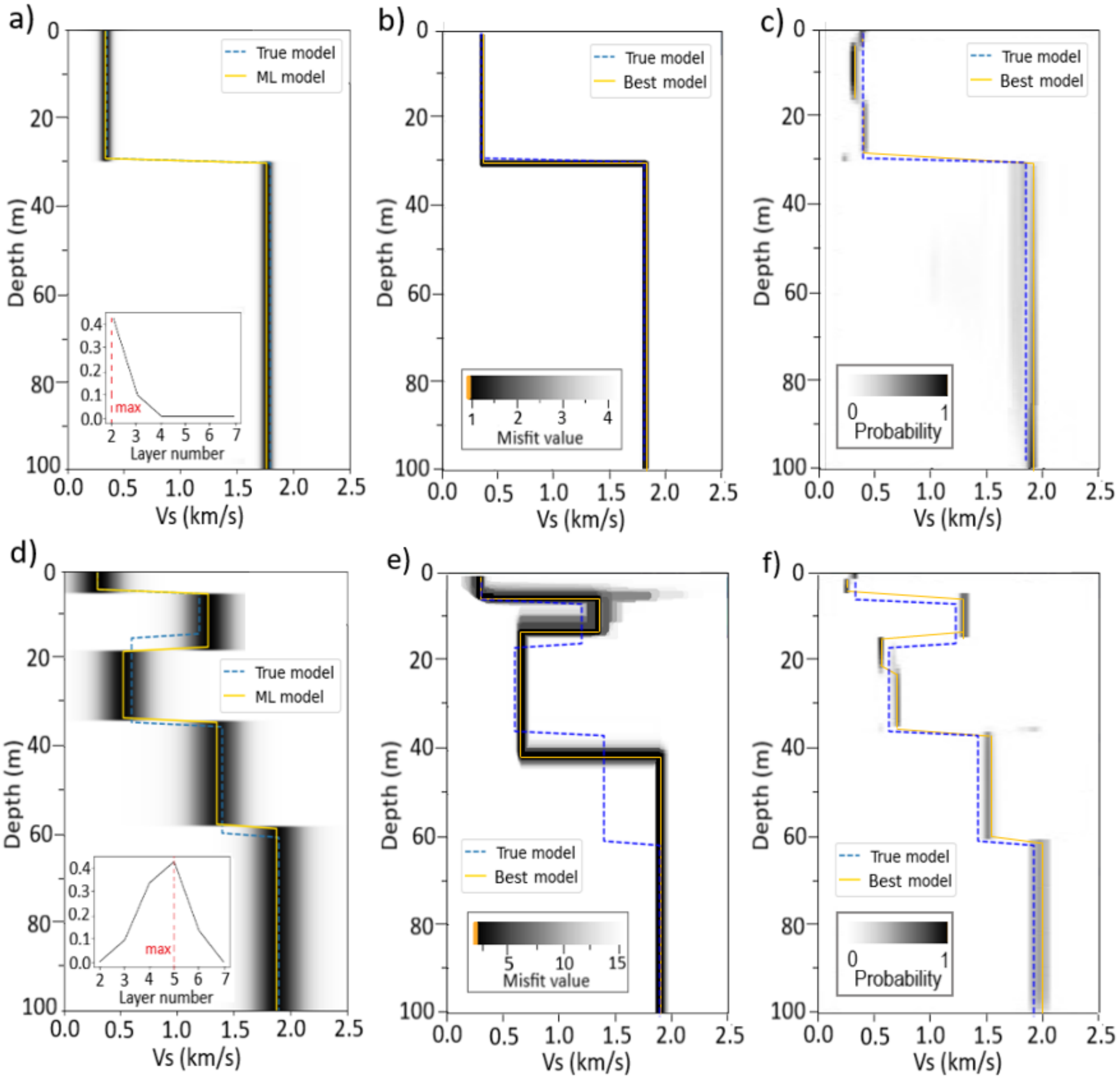

Figure 2.29: Inversion results for synthetic dispersion curves generated from 2 layer (a, b, c) and 5 layer (d, e, f) models. a) and d) show the predicted velocity structures from the MDN trained for 2 and 5 layer models respectively. The shading represents the probability distribution of the S-wave velocity, where darker colors represent areas of higher probability. In the lower left corner the output from the classification NN is shown. The red dashed line marks the predicted number of layers. b) and e) show the neighborhood inversion results using the DINVER module of the GEOPSY software (Wathelet 2008). The shading indicates the misfit values of the computed models. c) and f) show the inversion results from the Markov chain Monte Carlo algorithm (Alder et al. 2021). The shading indicates the probability of the computed models. The corresponding inversion targets and forward modelled dispersion curves can be found in Fig. 2.D1 of the appendix 2.3.D.

2.3.3.2 Field data

In the next step, we test our machine learning approach on dispersion curves extracted from ambient noise array measurements in Munich´s inner city (Fig. 2.30a). The measurements were conducted as part of a microzonation study in Munich, Germany, to investigate the shallow velocity structure for site effect characterization. The data used here were measured with a L-shaped array set-up with a side length of 100 m and a total of 11 three-component 4.5 Hz geophones. Ambient noise data were recorded for two hours and analyzed in a frequency range of 1 - 20 Hz. The Love and Rayleigh dispersion curves were computed from the data using high-resolution frequency wavenumber analysis (HRFK) (Capon 1969). A more detailed description of the setting, the measurements and the computation of the dispersion curves is given in Keil et al. (2022a). Using these input data, the classification NN predicts five subsurface layers (Fig. 2.30b). Therefore, the MDN trained for 5 layer models is used to estimate the velocity structure. The resulting profile has an upper velocity step at 5 m depth, followed by a more gradient-like velocity increase with depth. The inversion results are compared to a lithologic profile (Fig. 2.30c), which was extracted from the hydrogeological 3D model of Munich that was derived from borehole data during the GeoPot project (Technical-University-Munich 2016-2019). The

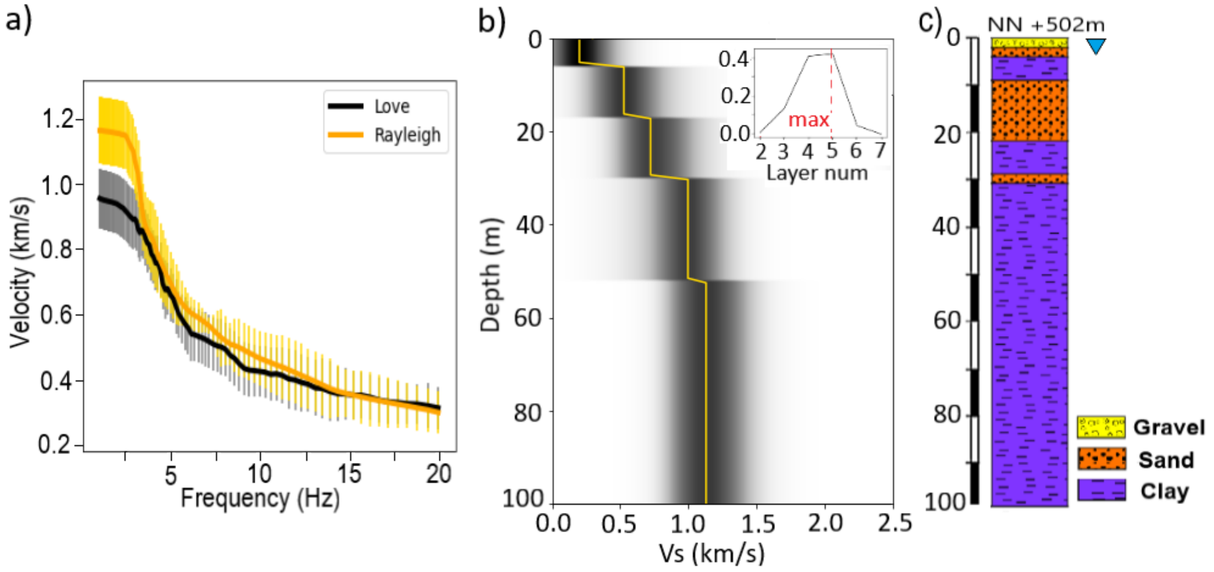

Figure 2.30: a) Love and Rayleigh dispersion curves extracted from ambient noise data in Munich, Germany. The vertical bars represent the data uncertainty. b) Predicted velocity structure from the MDN trained for 5 layer models. The shading represents the probability distribution of the S-wave velocity. In the upper right corner the output from the classification NN is shown. The dashed line marks the predicted number of layers. c) Lithologic profile at the site extracted from the hydrogeological 3D model of Munich (Technical-University-Munich 2016-2019). The triangle marks the depth of the groundwater table.

upper velocity step coincides with the groundwater table and the lithologic change from sand to clay. The second velocity step is in the range of the layer boundary between sand and clay at 20 m depth. Thereafter, no further lithological changes are visible in the profile (except for a very thin sandstone layer at 30 m depth). Therefore, a velocity gradient which is caused by compaction of the sedimentary material could be reasonable, as was also discussed in Keil et al. (2022a).

2.3.3.3 Robustness tests

In order to test the robustness of our trained ML models, we input synthetic dispersion curves generated from velocity structures that our network was not trained for. We test a smooth velocity gradient increasing from 200 to 1800 m/s (Fig. 2.31a). The classification NN predicts 7 subsurface layers, which is the maximum number of layers it was trained for. The resulting velocity structure, predicted by the MDN trained for 7 layer structures, resembles the velocity gradient quite close. The corresponding forward modelled Love and Rayleigh dispersion curves (Fig. 2.31b) are close to the input data, which highlights the ambiguity in the inversion results.

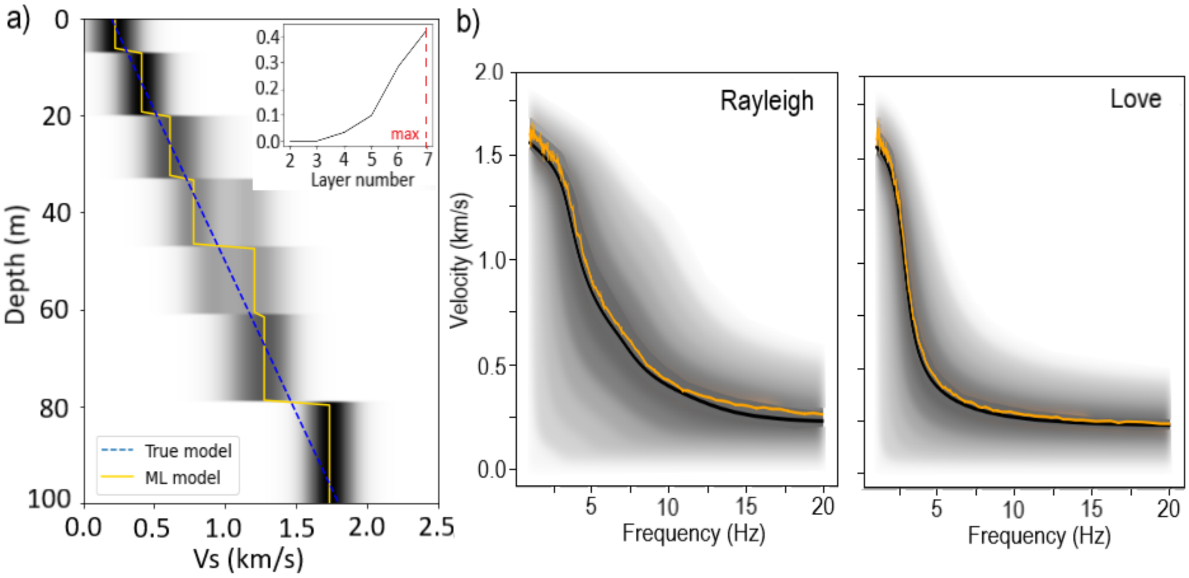

Figure 2.31: Inversion results for a synthetic velocity gradient. a) Predicted velocity structure from the MDN trained for 7 layer models. The shading represents the probability distribution of the S-wave velocity. In the upper right corner the output from the classification NN is shown. The dashed line marks the predicted number of layers. b) Forward computed Love and Rayleigh dispersion curves. The orange curves represent the input dispersion curves.

As a second robustness test we wanted to investigate how to handle data gaps, because often dispersion curve data cannot be computed down to 1 Hz, for example due to geometrical limitations of the array set-up. Therefore, we simulate a data gap between 1 - 3 Hz. Since the input vectors for the ML models need to contain values in a frequency range between 1 - 20 Hz, we substitute the missing information by extrapolating to a specific velocity value and increasing the standard deviation of the new data points. Since the true phase velocities at low frequencies are unknown we test four different extrapolation models (Fig. 2.32 a). For models 1 to 3 the classification NN predicts a three layer structure and for model 4 a four layer structure (Fig. 2.32 b). To get an average velocity structure, we add the probability distributions from the MDN predictions (Fig. 2.32 c). The shallow velocity structure is consistent for all four models. The velocity estimate in the first layer and the depth of the layer boundary correspond to the true model. Variations are visible between 30 and 60 m depth. However, due to the averaging of the MDN outputs, the maximum of the probability distribution is close to the true structure.

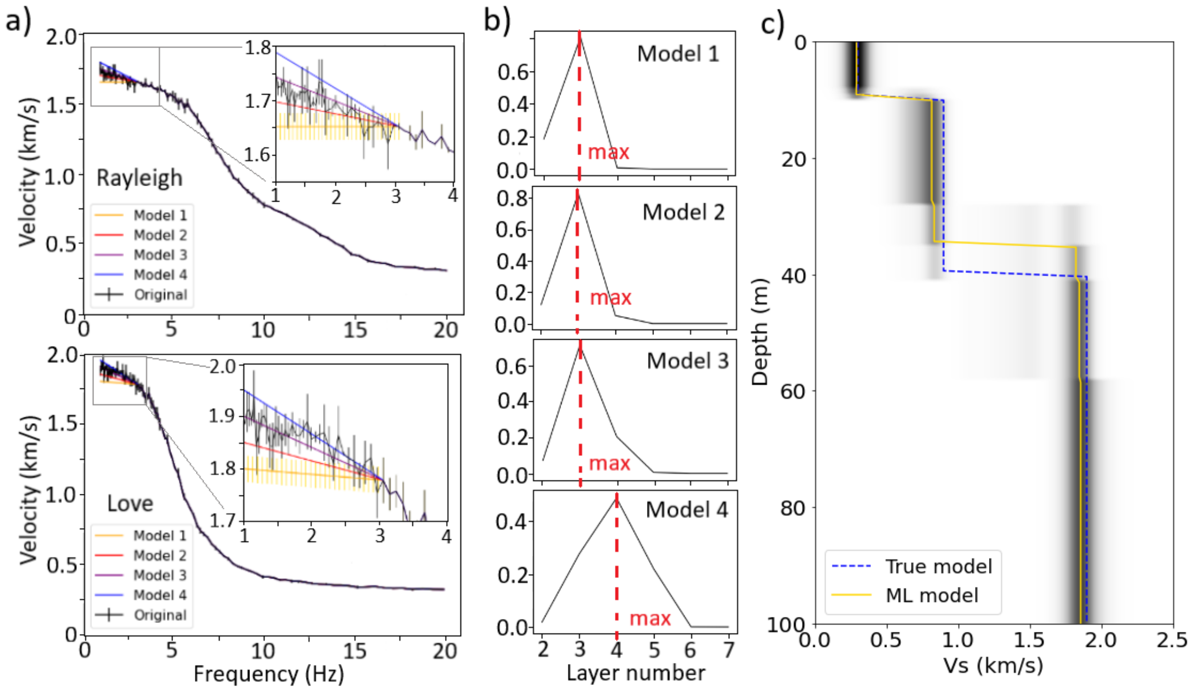

Figure 2.32: Inversion results for data gap between 1 - 3 Hz. a) Love and Rayleigh input dispersion curves. The black curves are the original dispersion curves. Models 1 to 4 extrapolate the data points below 3 Hz with different gradients. Vertical bars show the data uncertainty, where a standard deviation of 50 was assigned to the data points between 1 - 3 Hz. For clarity reasons only the standard deviation of model 1 is plotted. A magnification of the relevant frequency range is shown in the upper right corner of the plots. b) output from the classification NN for the different input models. The red dashed line marks the predicted number of layers. c) average velocity structure from the inversion results of the different models. The shading represents the probability distribution of the S-wave velocity, where darker colors represent areas of higher probability.

2.3.4 Discussion

We proposed a two-step machine learning approach in order to characterize the shallow subsurface structure from surface wave dispersion curves. The trained neural networks are able to predict the optimum number of layers within the upper 100 m of the subsurface, as well as the depth of the layer boundaries together with the probability distribution of the S-wave velocity. The trained classification NN performs well on previously unseen data, specifically for structures with few layers (2,3) and structures with the maximum number of layers (7). The NN seems to have more difficulties distinguishing between intermediate layer numbers (4-6), as their signature in the dispersion curves might be too similar. Nevertheless, in case of misclassification of the layer number, the error is mainly in the range of ±1.

The MDNs trained for simple structures (2-4 layers) perform well in terms of velocity prediction, as well as layer depth estimation. In general, the MDNs trained for more complex structures (5-7 layers) predict the S-wave velocity and the depth with less accuracy, especially for deeper layers. One reason could be the larger range of possible velocity structures, which we try to compensate by adding more models to the training set. Another reason could be the ambiguity in the inversion results, i.e. several sets of model parameters can give very similar dispersion curves (e.g. Foti et al. 2009). A third explanation is the decreasing resolution of the dispersion curve data with depth. Consequently the layer thickness has to increase with depth in order for the layer to be detected (Foti et al. 2014). Therefore, the minimum thickness of 10 m, which we defined in the construction of the velocity models, might be too thin for deep layers. In order to illustrate this problem, we forward modelled several velocity structures with a low velocity layer at depth with varying thickness between 10 to 30 m (Fig. 2.33). We compare the resulting Love dispersion curves with a structure that has no low velocity layer. Especially when noise is added, the shape of the Love dispersion curve resulting from the 10 m thick low velocity layer is barely distinguishable from the curve that was computed from the profile without low velocity layer. The difference is larger for the 20 m thick layer, but a clear variation is only visible for the 30 m thick layer. The shape of the dispersion curves is also influenced by other factors such as the velocity contrast between neighboring layers, but this test gives a first approximation on the sensitivity of the dispersion curves at depth. It also confirms that there is no need to train a network which can classify more than 7 layers, as a higher number would require thinner layers, which are not resolvable.

We compared our ML approach with a common neighborhood and McMC inversion. The great advantage of the ML approach is the extremely fast prediction of the inversion results. However, the time required for creating the data set and training of the models has to be taken into account. The training of the different NN took several hours, depending on the amount of training data. The creation of the data set, including the velocity structures and the forward modelling of the corresponding dispersion curves took several days. Nevertheless, once the networks are trained they can be easily applied to large data sets. For example, the 4000 test data were inverted within seconds. Compared to this the McMC inversion required several CPU hours for a single profile. The neighborhood

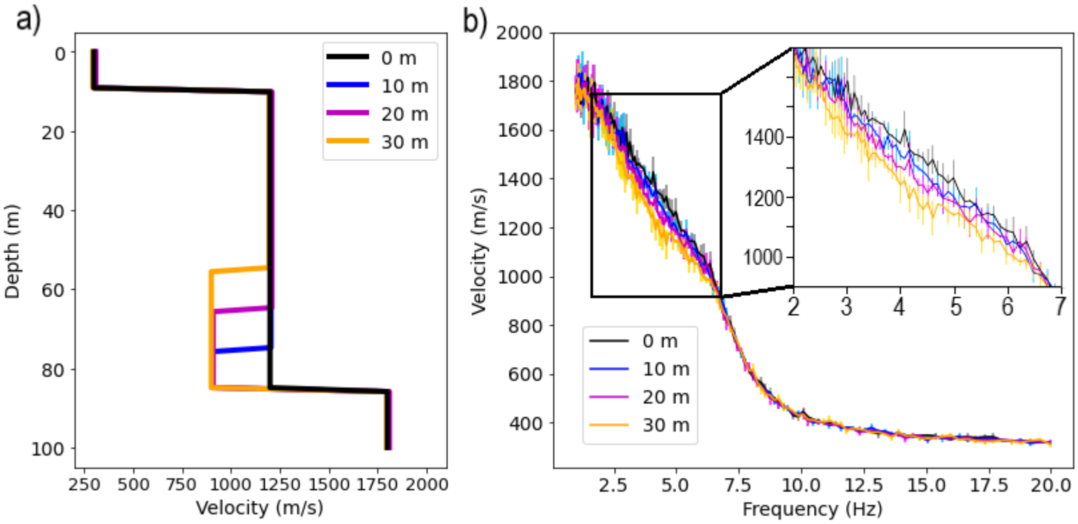

Figure 2.33: Sensitivity test for a low velocity layer at depth. a) Synthetic velocity profiles for a low velocity layer with varying thickness. b) Forward modelled Love dispersion curves with added noise. In the upper right corner of the plot a magnification of the relevant frequency range is shown.

inversion is also relatively fast (several tens of seconds for one profile), however, we could show that the results for complex structures are less accurate. Another point to discuss is the broader probability distribution of the MDN output in Fig. 2.29 d) compared to the McMC inversion result in Fig. 2.29 f). This can be explained by the density of samples that lie close to the observed data. As was noted by e.g. Käufl et al. (2016) posterior sampling methods, such as McMC, are implicitly conditioned on the observed data and are therefore designed to target sampling only in this area, whereas prior sampling methods, such as the MDN approach, are not. Therefore, the MDN result might be based on information provided by a small set of samples, which results in higher uncertainties. Nevertheless, we showed that the maximum value of the probability distribution is close to the true structure.

Furthermore, we showed that our machine learning approach is applicable to recorded noise data from Munich, Germany. The predicted velocity structure is consistent with the available lithologic profile at the site. In addition, our ML models are robust against untrained velocity structures, such as velocity gradients. The forward modelled Love and Rayleigh dispersion curves were close to the true data in this case, which represents an ambiguity in the inversion results. It was also shown by Ivanov et al. (2013) that velocity gradients and layered models can result in near identical dispersion curves within a specific frequency range. They suggest that differences can only be observed in the dispersion curve image patterns (e.g. energy blobs) at relatively low and high frequency ranges (3 to 70 Hz). These features cannot be resolved with our method. In addition, for our approach input data in the entire frequency range between 1 - 20 Hz are required in order to reliably recover the velocity structure down to a depth of 100 m. This could pose problems in situations where dispersion curves cannot be extracted across the whole frequencies domain. We have shown that data gaps can be filled by extrapolating the missing data points. Hereby, several inversions with different extrapolation values should be performed to obtain an average velocity structure.

In the future, the accuracy of the predicted velocity structure could be improved by adding horizontal-to-vertical (H/V) spectral ratios (Nogoshi & Igarashi 1971) to our models, since they provide additional information on the depth of layer boundaries. Combining dispersion curves and H/V ratios for surface wave inversion is already commonly applied in ambient vibration analysis (e.g. Scherbaum et al. 2003, Fäh et al. 2008). Also, our ML approach could be extended to higher mode surface waves to increase the resolution with depth, as they specifically provide further information on the deep velocity structure (e.g. Xia et al. 2003, Pan et al. 2019). In addition, working towards explainable neural networks (e.g. Samek et al. 2021) could improve our understanding of the NN inversion process and the sensitivity of the dispersion curve data on specific layer depth ranges. However, these prospects are outside of the scope of this paper.

2.3.5 Conclusion

To model the uncertainties in the inversion results and to put less constraints on the choice of the inversion parameters, in particular layer thickness and layer number, we proposed a two-step machine learning approach for the inversion of surface wave dispersion curves in a frequency range of 1-20 Hz. In the first step a classification NN is able to distinguish between 2 - 7 subsurface layers in the upper 100 m. In the second step, based on the number of layers, individually trained mixture density networks predict the layer depth, as well as a complete probability distribution of the S-wave velocity structure. The probability distribution gives an estimation of the ambiguities and uncertainties contained in the subsurface model. The lower prediction accuracy for deeper layers is most likely related to the decreasing sensitivity of the dispersion curve data with depth. Nevertheless, we showed that our ML approach has advantages over conventional inversion methods, as it is able to recover complex structures with higher accuracy compared to the applied neighborhood inversion. In addition, once the NNs are trained the inversion of the dispersion curves is performed much faster compared to a McMC inversion, which is specifically useful for multidimensional problems. Furthermore, we showed that our ML approach can be applied to recorded noise data, which has implications for any near-surface study, where a fast and accurate estimation of the shallow subsurface structure is required.

Data Availability

The trained ML models and python codes for the application of the inversion are available at https://doi.org/10.5281/zenodo.7670360.

Acknowledgements

This work was supported by the research project SEIGER (Project no. 03EE44003G), funded by the German Federal Ministry of Economic Affairs and Climate Action and managed by the research institute Jülich (PTJ). The authors thank Andrew Curtis and Stephanie Earp for providing parts of the MDN code and for the helpful discussions concerning the application of mixture density networks. Thanks to Max Moorkamp for the support and helpful discussions concerning machine learning in Python. In addition, the authors are grateful to Thomas Bodin for the support with the McMC algorithm. The authors thank the editor Dr. Andrew Valentine and two anonymous reviewers for their helpful comments, which improved the quality of the manuscript.

2.3.C Appendix: Network configuration

Table 2.C1: Network configuration of the classification neural network including the number of neurons in each dense layer (D).

Disp. Lov		**Unc. Lov**		**Disp. Ray**		**Unc. Ray**		**Concat1**	**Concat2**	**Concat3**		**Total parameters**
D1	D2	D1	D2	D1	D2	D1	D2	D1	D1	D1	D2	
418	716	369	769	117	396	216	828	879	832	332	845	4,104,297

Table 2.C2: Network configuration of the different mixture density networks including the number of neurons in each dense layer (D).

Network	**Disp. Lov**	**Unc. Lov**	**Disp. Ray**	**Unc. Ray**	**Concat 1**	**Concat 2**	**Concat 3**		**Depth**	**Total parameters**
	D1	D1	D1	D1	D1	D1	D1	D2	D1	
2 layer	544	271	696	565	622	254	479	814	281	2,110,758
3 layer	271	696	565	622	254	479	814	/	281	1,900,565
4 layer	936	296	532	168	707	880	655	299	817	3,186,399
5 layer	936	296	532	168	707	880	655	299	817	3,190,217
6 layer	936	296	532	168	707	880	655	299	817	3,194,035
7 layer	936	296	532	168	707	880	655	299	817	3,197,853

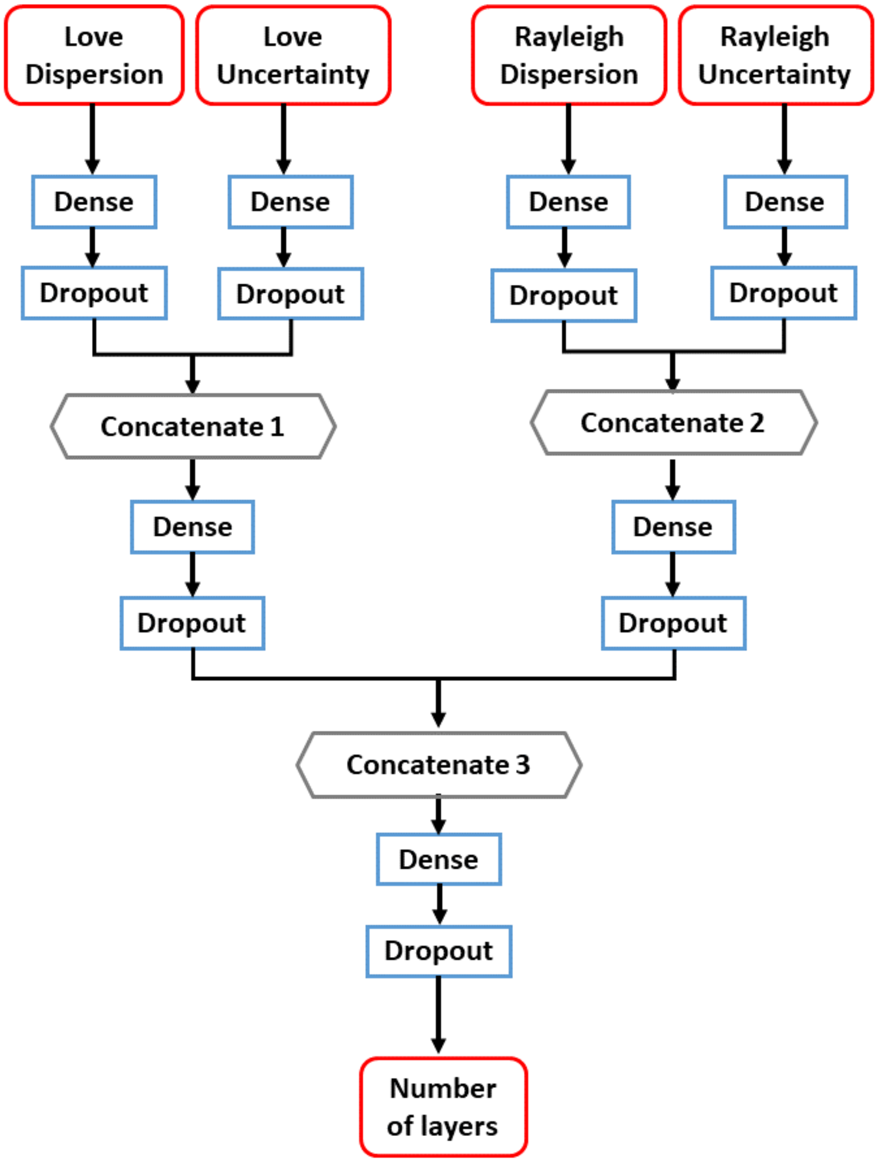

Figure 2.C1: Schematic network structure used for the training of the classification neural network. Red rounded edged boxes represent the inputs and outputs of the network. Blue squared edged boxes represent one or more Dense and Dropout layers. The gray hexagonal boxes represent concatenation layers. The arrows represent the flow direction of data through the network.

2.3.D Appendix: Inversion targets

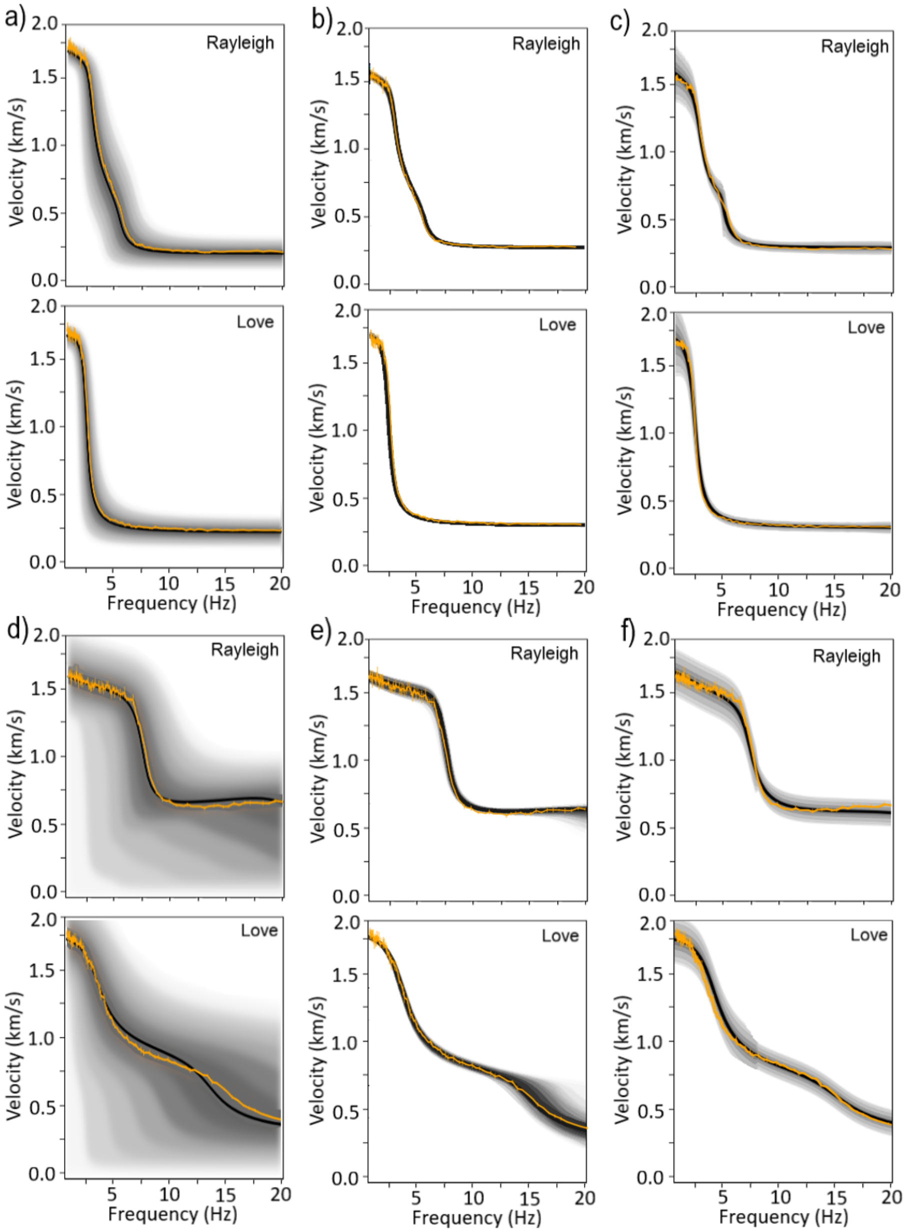

Figure 2.D1: Inversion targets and forward modelled dispersion curves corresponding to the inversion results shown in Fig. 2.29. a) and d) correspond to the ML inversion, where the orange lines represent the input data and the black lines are the forward modeled dispersion curves from the predicted mean velocity structure. The shading indicates the probability distribution of the computed models, where darker colors represent areas with higher probability. b) and e) correspond to the neighborhood inversion, where the orange lines represent the input data and the shading indicate the misfit value of the computed models. c) and f) correspond to the McMC inversion, where the orange lines indicate the input data and the black line is the best fitting dispersion curves from the computed models. The shading indicates the probability distribution of the computed models, where darker colors represent areas with higher probability.

Chapter 3

Estimation of Ground Motion due to Induced Seismicity at a Geothermal Power Plant near Munich, Germany, using Numerical Simulations

by Sabrina Keil, Joachim Wassermann and Tobias Megies

Published in Geothermics 106 (2022): 102577

Abstract

In the North Alpine Foreland Basin, especially in the greater Munich area, several geothermal plants exploit a deep hydrothermal reservoir. So far at three sites events with $M_L > 2.0$ were detected, in an otherwise seismically inactive region. In this study we investigate a site east of Munich, where moderate seismicity started to appear around five years after the beginning of geothermal circulation. Two larger events ($M_L 2.1/1.8$) occurred in December 2016, followed by an additional $M_L 2.1$ event about 10 months later in September 2017. All three events occurred at about 3 km depth and were felt by the population. As this is a densely inhabited area, the estimation of the maximum ground motion and its distribution are of great importance for the authorities and the public. We perform 3D seismic simulations of the main events using the spectral element code SALVUS. The results can supplement recorded and macroseismic data in order to estimate the possible seismic impact in the area. In addition, we evaluate the influence of the uncertainties contained in the event parameters and in the subsurface model on the maximum peak ground velocity (PGV) values to calibrate the simulations. The simulated waveforms are mostly in good agreement with the ground motion recordings. Furthermore, the ground motion

distribution coincides with the macroseismic data. According to the simulation results, even the largest event in the area did not exceed the critical PGV value of 5 mm/s defined by German norms and therefore had no damage potential. Such a numerical approach can help to improve the seismic monitoring network, identify affected zones and mitigate the seismic risk.

3.1 Introduction

The Bavarian Molasse Basin, especially the greater Munich area, is well suited for geothermal energy production, as an aquifer at depth provides a natural reservoir with both high water temperatures (90 - 120 °C) and high flow rates (up to 120 l/s). These favorable conditions led to the development of many geothermal projects in the region. Currently 17 geothermal power plants are operating in and around Munich (Agemar et al. 2014). Even though the region shows only scarce background seismicity (Leydecker 2011), induced events were recorded at some of the power plants. For example, at the Unterhaching site microseismicity was already detected during the first years of operation, including a magnitude 2.4 event in 2008 (Megies & Wassermann 2014). In comparison, the onset of seismicity in Poing occurred around 5 years after the beginning of geothermal circulation with two larger events (M_L 2.1 and 1.8) that occurred on the 7th and 20th of December 2016, followed by an additional M_L 2.1 event about 10 months later on the 9th of September 2017 (Megies & Wassermann 2016, 2017*a*). Due to the shallow hypocentral depth of a few kilometers, even such relatively small earthquakes can lead to ground shaking, which is felt by the population. For example, the Poing M_L2.1 event in 2017 led to around 140 felt intensity reports, as well as several damage claims, even tough the maximum recorded peak ground velocity (PGV) value was 1.6 mm/s at about 500 m distance from the epicenter, which does not indicate a damage potential (Megies & Wassermann 2017*a*). According to the German DIN 4150-3, which regulates how the effects of vibrations on structures are measured and assessed, damage to buildings is unlikely for ground motions below 5 mm/s in a frequency range of 1-10 Hz. However, because of the sparse station network in this area the actual maximum ground motion that the event induced might have not been recorded. This problem is regularly encountered at geothermal sites, as a dense network is neither feasible nor required by law. In this case, numerical simulations are useful in order to estimate the damage potential and the distribution of the ground motion. These results can supplement recorded and macroseismic data in order to identify affected areas according to the German mining laws ("Einwirkungsbereichs-Bergverordnung"), which only recently were amended to account for potential damage by induced seismicity at geothermal wells. According to this law affected areas are defined as zones, where the ground motion after an induced event exceeds 5 mm/s in a frequency range of 1 to 10 Hz and/or intensities of V (EMS-98) or larger were reported. Damage within these defined zones is then accounted to the geothermal power plant, meaning the operating company has to disprove damage claims brought forward by local population and companies or otherwise has to compensate for it.

In this study the ground motion generated by induced seismicity at the Poing geothermal site, which is situated 15 km east of Munich, is investigated in more detail using 3D numerical simulations. The simulations are performed using the spectral element code SALVUS (Afanasiev et al. 2019). This powerful and user-friendly software package, together with a vast amount of information on the deep, as well as on the shallow subsurface structure in the greater Munich area, enables the performance of realistic and reproducible full waveform modelling. First of all, a detailed 3D-subsurface model is constructed using the available geological and geophysical data. As a second step a parameter study is conducted to quantify the impact of uncertainties in the event parameters and in the subsurface model on the simulation results. After that, the numerical simulations are performed for the M_L 1.8 event in 2016, as it is the best recorded earthquake at this site and the M_L 2.1 event in 2017, as it is one of the largest ones. Thereafter, the simulated waveforms are compared to the recorded data in order to validate the simulation results. As a last step shakemaps are constructed from the simulation results to estimate the distribution of the ground motion and the damage potential of the events.

3.2 Geological setting and seismicity

The study area is situated 15 km east of Munich within the Molasse basin, which is a foreland basin north of the Alps (Fig. 3.1a). The basin has an E-W extension of approximately 700 km and a maximum N-S extension of 130 km. It is filled with up to 5000 m thick Tertiary sediments, comprising mainly conglomerates, sandstones, as well as silt- and claystones. They can be subdivided into five groups, in order of stratigraphic age, from oldest to youngest: Lower Marine Molasse (Rupelian), Lower Freshwater Molasse (Chattian–Aquitanian), Upper Marine Molasse (Burdigalian), Freshwater–Brackish Molasse (late Burdigalian) and Upper Freshwater Molasse (Langhian–Tortonian) (e.g. Kuhlemann & Kempf 2002). These "Molasse" sediments are underlain by 500 - 1000 m thick Mesozoic limestone layers and the Variscian crystalline basement (Fig. 3.1b) (e.g. Bachmann et al. 1987). Basin formation and sedimentation was primarily due to the flexure of the European plate as a result of the northward thrusting of the Alps, which also resulted in a dominant structural southward dipping of the basement and Mesozoic layers and the formation of mainly normal faults striking parallel to the Alpine Orogen (e.g. Bachmann et al. 1987, Kuhlemann & Kempf 2002). These E-W trending steeply dipping fault structures can be traced from the basement, through the Mesozoic layers, up to the Miocene strata (Bachmann & Müller 1992, Moeck et al. 2015). The main fault structure in the study area is a synthetic normal fault with an offset of up to 200 m. It traces with an ENE trend through Poing, where it is named "Markt-Schwabener Verwurf" (Dussel et al. 2012). These Tertiary sediments are overlain by the so called Munich gravel plain, which is comprised of loose alluvial, fine- to coarse-grained sediments and formed during the Pleistocene glacial periods (Bauer et al. 2006). Since Poing is situated within this flat gravel plain, the area has almost no topographic features.

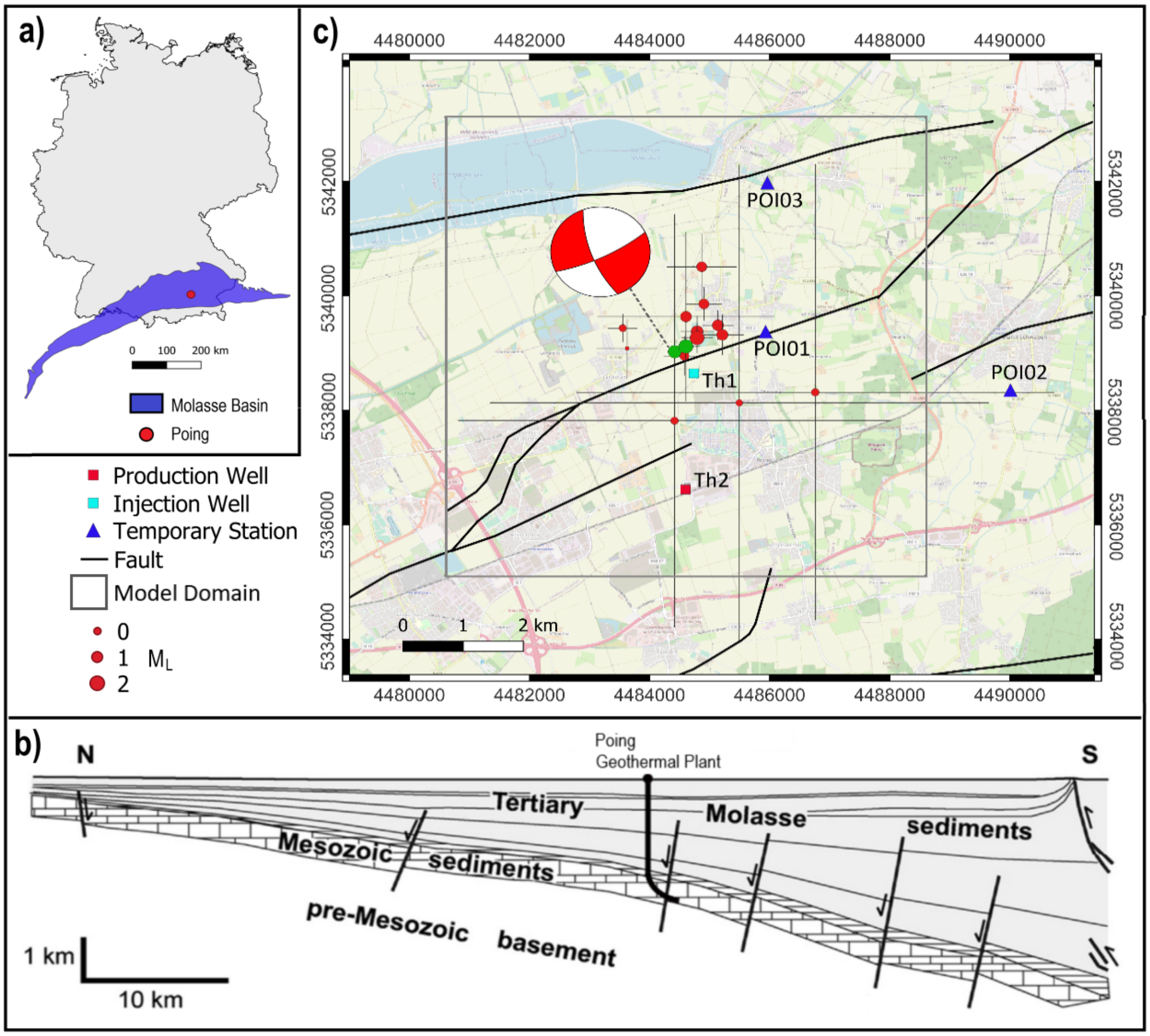

Figure 3.1: a) Overview map of the Molasse basin in the south of Germany. The location of the study area is marked by the dot. b) Schematic N-S cross-section through the Molasse basin (modified after Reinecker et al. (2010), also see Lemcke (1988)). The Tertiary Molasse sediments are underlain by Mesozoic sediments, which form the target horizon for geothermal wells in southern Germany, and the pre-Mesozoic basement. The roughly E-W striking normal faults and the geothermal well in Poing are sketched schematically. c) Study area with fault structure (Bayer. Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie (2010)), microseismic events and seismic stations. The horizontal and vertical lines represent the location uncertainty of the events. The simulated events are highlighted in green and the estimated focal mechanism is shown. The gray box marks the boundaries of the model domain. The POI-Stations form a temporary network. Th1 is the location of the injection well, where also a permanent station is installed. Th2 is the production well. The map is referenced by the Gauss-Krüger coordinate system (GK zone 4, EPSG code 31468).

The target horizon for deep geothermal projects in southern Germany is the Malm formation. It constitutes series of limestones and marls, which were deposited in the Upper Jurassic and were later exposed and subjected to erosion and karstification (Bachmann et al. 1987). The Malm formation constitutes a natural reservoir with locally varying permeability, depending on the porosity of the limestone facies, the degree of karstification and the occurrence of faults and fractures (Birner 2013). Due to the southward dipping of the limestone layers, the depth of the aquifer and subsequently the water temperature in the reservoir are increasing towards the Alps. In the study area the reservoir reaches a depth of 2.3 - 2.5 km, where water temperatures of 85 ℃ are encountered. The Poing geothermal project consists of two wells which were drilled in 2008, the injection well Th1 and the production well Th2. The wells were drilled to depths of 3050 and 3014 m TVD, respectively. Circulation started in December 2012 with maximum flow rates of 100 l/s (Agemar et al. 2014).

First seismicity was recorded at the end of 2016, including two larger events with local magnitudes of 2.1 and 1.8 and several smaller aftershocks (Megies & Wassermann 2016). Another M_L 2.1 event occurred on the 9th of September 2017 (Megies & Wassermann 2017*a*). All the events that were detected using data from the Bavarian earthquake service, the research project MAGS2 EP2 (Megies & Wassermann 2017*b*) of the Ludwig-Maximilians-University Munich (LMU) and stations from surrounding geothermal power plants, are plotted in Fig. 3.2. It is very likely that additional microseismic events occured, however, they could not be detected due to the small amount of monitoring stations in the area (Megies & Wassermann 2017*b*). In addition, this causes large uncertainties of several kilometers in the determination of the hypocenter location. After the M_L 2.1 mainshock on the 7th of December 2016, a temporary network of three stations (POI01, POI02, POI03) was installed in the study area together with a permanent station at the re-injection well Th1 (Megies & Wassermann 2016). This significantly decreased the location uncertainty for subsequent events and also allowed the detection of smaller earthquakes. The M_L 1.8 event on the 20th of December 2016 could be analyzed in more detail: Megies & Wassermann (2016) calculated a hypocentral depth of 3.4($\pm$1) km. In addition, the focal mechanism was determined, which indicates a left lateral strike-slip mechanism with a normal faulting component (64 ± 7 strike, 76 ± 5 dip, -22 ± 6 rake), that is also in accordance with the main fault geometry in the area.

The maximum recorded peak ground velocity (PGV) value of 1.6 mm/s (unfiltered) was induced by the M_L 2.1 event on the 9th of September 2017 at the station "Poing" that is located at the re-injection well Th1 at approximately 500 m epicentral distance. In addition, this event was felt by residents in Poing and Pliening, which resulted in about 140 felt intensity reports with the majority of intensities between III-IV (Megies & Wassermann 2017*a*). The seismic events occurred below the reservoir within the crystalline basement along the "Markt-Schwabener-Verwurf" in the vicinity of the re-injection well Th1 (Fig. 3.1c) (Megies & Wassermann 2017*b*). This observation, together with a lack of natural background seismicity in the study area, that is reflected in the German earthquake catalog maintained by Leydecker (2011), suggests a correlation between the seismic events and the re-injection of fluids at the well Th1. The time delay of the onset of seismicity could

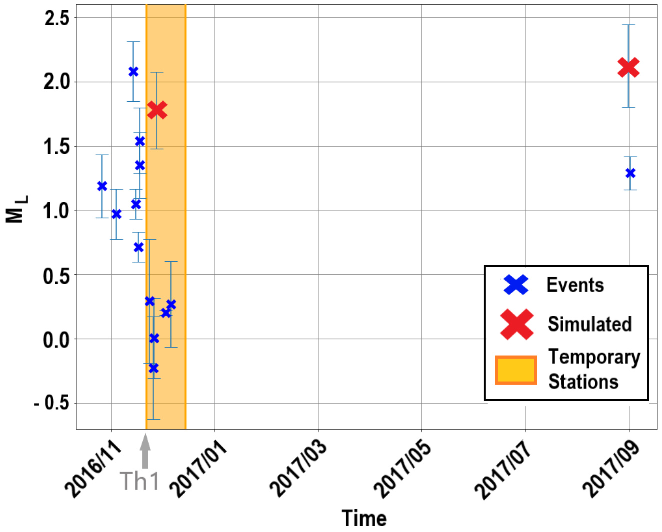

Figure 3.2: Detected microseismic events at the Poing geothermal power plant. The vertical lines represent the magnitude uncertainty. The events that are simulated in this study are marked by the larger red crosses. The shaded area shows the time span when all three temporary stations (POI01, POI02, POI03) were installed. The arrow shows the time when the permanent station at the re-injection well Th1 got installed.

be explained by long term effects such as delayed stress field rotation by thermo-elastic effects and fault weakening by carbonate dissolution (Seithel et al. 2019).
The estimated event parameters for the $M_L1.8$ and the 2017 $M_L2.1$ earthquakes, which are simulated in this study, are listed in Table 3.1.

Table 3.1: Estimated event parameters together with their uncertainties for the two induced earthquakes that are simulated in this study. The depth of the events is given as depth below ground level. The x-y-coordinates are referenced by the Gauss-Krüger coordinate system (GK zone 4, EPSG code 31468).

Date	20.12.2016	09.09.2017
$\mathbf{M_L}$	1.8 ± 0.3	2.1 ± 0.3
Depth	3.4 ± 1.0 km	3.1 ± 1.0 km
x-coordinate	4484418 ± 200 m	4484604 ± 470 m
y-coordinate	5339022 ± 200 m	5339114 ± 470 m
Focal mechanism	64° ± 7°, 76° ± 5°, -22° ± 6°	/

3.3 Set-up of the 3D numerical model

The numerical simulations of seismic wave propagation in Poing are carried out using the spectral element code SALVUS (Afanasiev et al. 2019), which is a software suite applicable for full waveform modeling in 2D and 3D. It is well suited for our modeling purposes, as it allows the simulation to high frequencies and includes a toolbox for meshing, which facilitates the generation of highly efficient spectral-element meshes including heterogeneous media with complex topography or internal structures. In addition, SALVUS offers a lot of flexibility and supports various parameterizations of the model domain. For this study the model domain is dicretized using the P- and S-wave velocity structure (Vp and Vs), density (rho), as well as seismic attenuation (Q_P and Q_S).

For the set-up of the 3D model domain different geological and geophysical data that are available from previous studies are combined. First of all, information on the depth of the main layer boundaries are extracted from a structural model developed by the Bavarian State Office for Environment (Bayerisches Landesamt für Umwelt 2012), which covers the study area and includes the main reflectors that are encountered in seismic surveys in the Molasse Basin. The three layer boundaries that form the basic framework of the 3D numerical model are the Upper Marine Molasse (OMM), the Lithothamnium limestones and the crystalline basement (Fig. 3.3a). The two deeper layers have a typical southward dipping geometry and include the fault line of the "Markt-Schwabener Verwurf" with vertical offsets of approximately 200 m. The Lithothamnium limestone is characterized by a distinct reflection below the Tertiary sandstones and mudstones of the Molasse sediments and is often used as a reference horizon for interpretation of seismic data in the Molasse basin (e.g Lüschen et al. 2014). Due to scattering within the karstified Malm carbonates, reflections at the Base of the Malm formation are only poorly resolved, however, a clear contrast to the crystalline basement is discernible (Budach et al. 2018).

The P-wave velocities within the layers are based on a 3D seismic survey conducted in 2015/16 as part of the GRAME project, which covered 170 km^2 in the southern and western parts of Munich. Since the study area is only about 15 km east of this seismic survey, we assume that the velocity structure within the layers does not change significantly. Therefore, we take average P-wave velocities from each layer and assign them to the numerical model. In Fig. 3.3b) a 1D profile extracted from the complete 3D numerical model is shown, including the P- and S-wave velocity structure, as well as the density gradient. The uppermost layer has a thickness of approximately 100 m and represents the low velocity Quaternary sediments of the Munich gravel plain together with weathered Tertiary sediments. This layer is followed by the Upper Freshwater Molasse (OSM). The first large velocity step near sea level marks the top of the OMM. Below this layer boundary the seismic velocities increase within the Molasse sediments until the Lithothamnium limestones are reached at about 1900 m below sea level (bsl). There is a change in the velocity gradient at approximately 1000 m bsl, which is identified in the 3D seismic survey as the top of the Chattian formation. However, this layer boundary is not included in the structural model provided by the LfU and therefore the depth of this layer boundary is extrapolated. For simplicity, the velocities within the Mesozoic limestones are modelled to be nearly constant until the basement is reached. The 3D seismic survey estimates an average Vp value for the basement of 5200 m/s, however, we would expect larger values for unweathered crystalline rocks. For example, Grad et al. (2009) estimated a P-wave velocity in the crystalline upper crust of the Molasse basin of about 6150 m/s. This discrepancy will be evaluated in section 3.4.

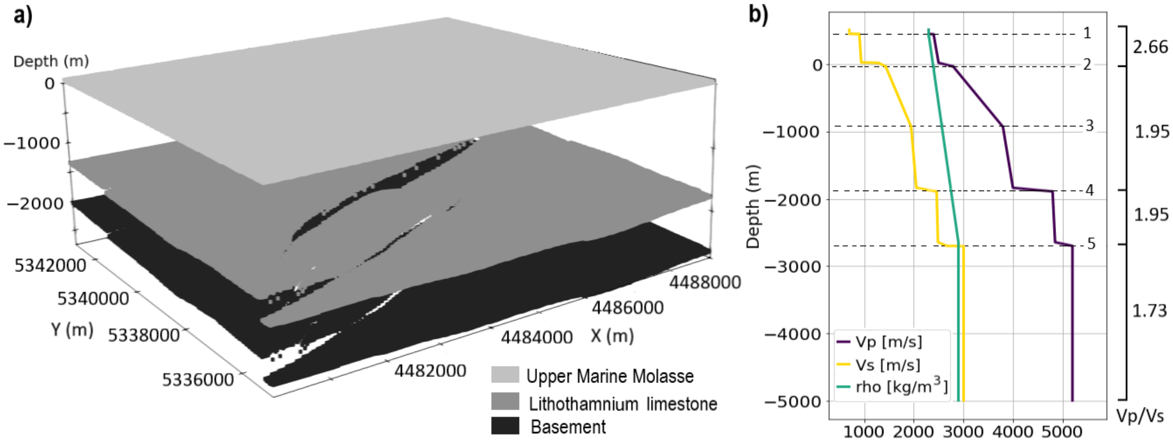

Figure 3.3: a) Main geologic interfaces that are used for the construction of the numerical 3D model, extracted from a structural model developed by the LfU (Bayerisches Landesamt für Umwelt 2012). The layer boundaries of the Lithothamnium limestones and the basement show a southward dipping, as well as the fault line of the "Markt-Schwabener Verwurf". The coordinates are referenced by the Gauss-Krüger coordinate system (GK zone 4, EPSG code 31468) and the depth is given in meters below sea level. b) 1D example profile extracted from the complete numerical 3D model, including the P- and S-wave velocity structure (Vp, Vs), as well as the density (rho). The main velocity steps can be accounted to the geologic interfaces of the 1: Upper Freshwater Molasse, 2: Upper Marine Molasse, 3: Chattian formation, 4: Lithothamnium limestones and 5: basement. The Vp/Vs ratios that are used to calculate the S-wave velocities in each layer interval are listed on the right.

The S-wave velocity model is constructed from the P-wave velocity model using Vp/Vs ratios estimated by Wawerzinek et al. (2021) for the Munich area. They estimated a median value of 2.66 for the uppermost 500 m of the Molasse sediments. For the deeper Molasse units above the carbonate platform the Vp/Vs ratio varies between 1.8 and 2.1, with a median of 1.92, which is in agreement with a Vp/Vs ratio of 1.9 - 2.0 found by Megies & Wassermann (2017*b*). Therefore, we take an average Vp/Vs value of 1.95 for the interval between the OMM and the Lithothamnium limestone formation. The Vp/Vs ratio of the Mesozoic carbonates shows a stronger variation between 1.5 and 2.2, depending on the carbonate lithology and porosity (Wawerzinek et al. 2021). For simplicity we pick a median value of 1.95, which was also estimated by Megies & Wassermann (2017*b*). The Vp/Vs ratio in the basement is set to 1.73, which is a standard value for the upper European crust (e.g. Grad et al. 2009).

For site effects the shallow velocity structure plays a major role, as low velocity sediments can amplify seismic waves. Keil et al. (2021) conducted a microzonation study in Munich and estimated locally differing S-wave velocity values in the upper 30 m (Vs30) of 200 - 450 m/s. Since Poing is also located on the Munich gravel plain, the shallow subsurface structure is assumed to be similar to the one in Munich. Therefore, an average S-wave velocity value of 350 m/s is implemented in the upper 30 m of the model domain to account for site effects.

Estimating the density structure is more difficult, as it depends on many different factors, such as mineral composition, porosity and density of pore fluids. Since no detailed information is available the density model is constructed using a linear gradient between 2300 kg/m^3 at the surface to 2900 kg/m^3 in the basement, which includes typical values for sedimentary as well as crystalline rocks. This range also includes density values of the Molasse sediments found in Leu et al. (2006) and density values of the Malm carbonates estimated by Homuth (2014).

Seismic attenuation is an intrinsic property of rocks causing dissipation of energy as seismic waves propagate through the subsurface. It results in the decay of amplitude of the seismic waves and therefore is an important property for the correct estimation of the ground motion. Eulenfeld & Wegler (2016) estimated the seismic attenuation of shear waves for a site located in the Molasse basin, about 20 km southwest of our study area. They estimated a mean S-wave quality factor (Qs) of 100, averaged over the whole ray path, which is constant for frequencies lower than 8 Hz. Due to the lack of further information on the attenuation of P-waves, we set the P-wave quality factor (Qp) to 200, as literature suggests that Qp is approximately two times higher than Qs (e.g. Fowler 1990).

Furthermore a flat surface is implemented in the 3D model, since there are no strong topographic features in the area. Once all the different information are gathered and combined, the 3D model domain can be constructed using the SALVUS mesher. We define a domain with 8 km x 8 km extension and 5 km depth, which is large enough to contain the earthquake hypocenters, as well as the main distribution of the ground motion at the surface. We apply absorbing boundary conditions on the sides of the domain, as well as on the bottom, and a free-surface condition on the top. The goal is to simulate frequencies of up to 10 Hz, as this is a common frequency range of induced events and is most important for the building response.

The numerical dispersion error from the spatial discretization of the spectral element scheme can be controlled by the size of the elements in the mesh and the polynomial order of the finite-element test functions. Increasing the number of elements and/or the polynomial degree leads to more accurate results, however, the computational costs are also increased. A mesh convergence study was conducted to evaluate the influence of the polynomial degree and number of elements on the simulation results and estimate the dispersion error (Fig. 3.4). The simulated PGV value at station Th1 is taken as a reference for comparison. In general, a convergence to a PGV value of about 0.121 mm/s, in a frequency range of 1 to 5 Hz, with increasing polynomial order and number of elements is observed. In total the PGV values have a variation of 13%. Increasing the number of

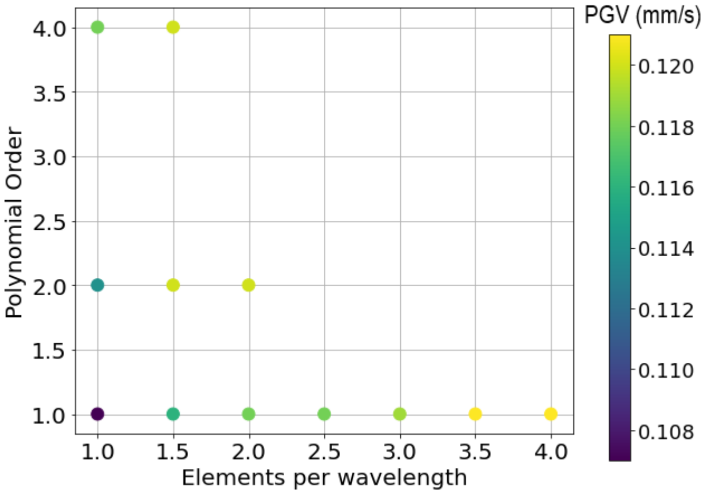

Figure 3.4: Results of mesh convergence study. Peak ground velocity (PGV) value at station Th1 with varying polynomial order of the finite-element test functions and number of elements per wavelength in the mesh.

elements per wavelength from 1 to 1.5 already significantly improves the results. In this case the difference is less than 5%.

Because of that and in order to save computational costs, a polynomial degree of 1 and a minimum of 1.5 elements per wavelength is implemented. Nevertheless, this only holds for the smallest wavelengths, while larger wavelengths are sampled with more elements and are therefore modelled with higher accuracy.

Another feature that is implemented in the SALVUS mesher is adaptive meshing. The element size coarsens with depth as the velocity and subsequently the wavelengths increase, which reduces the total number of elements and grid points in the mesh significantly and therefore reduces the computational costs. Our final mesh has a number of elements and number of grid points on the order of 4×10^6 (Fig. 3.5).

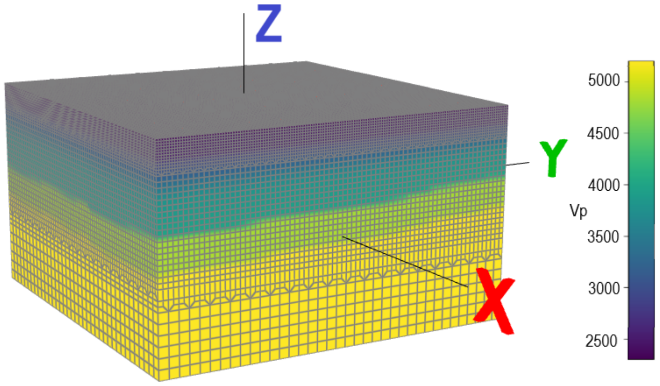

Figure 3.5: 3D mesh constructed with the SALVUS mesher. The colors show the P-wave velocity model. An increase of element size with depth is visible.

The domain is constructed around the source, which is implemented using the event parameters estimated for the recorded earthquakes. The focal mechanism and the magnitude are modeled using a moment tensor source. Therefore, the local magnitude M_L has to be converted to the moment magnitude M_w. We set M_w equal to M_L, since a study using European earhtquake data by Grünthal et al. (2009) shows that both scales are in agreement down to magnitudes of around 2, which includes the main events in this area. However, for smaller events the local magnitude scale yields lower values than the moment magnitude scale. For a small earthquake, the fault is considered to be a single point source, where rupture occurs instantaneous. In that case the source time function is simply a delta pulse (e.g. Stein & Wysession 2009). We model this delta pulse using a sinc-function with a flat power spectrum up to 10 Hz (Fig. 3.6). The simulated waveforms are afterwards bandpass filtered between 1 - 10 Hz.

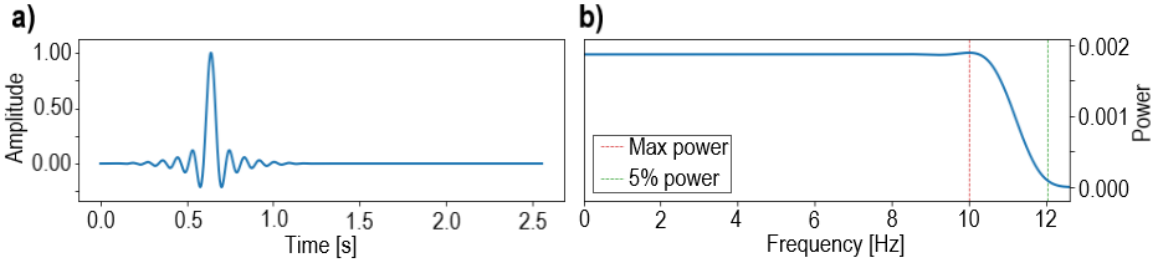

Figure 3.6: Source time function (stf) of the simulated events in the a) time domain and b) frequency domain. The stf is modelled by a sinc-function.

3.4 Parameter study

A parameter study was conducted in order to estimate the influence of the different event parameters and the properties of the subsurface model on the maximum peak ground velocity. To reduce the computation costs, the simulations are performed only up to 8 Hz using the subsurface model developed in section 3.3. In the first part the influence of the event parameters is investigated including depth, location, focal mechanism and magnitude. In every simulation run one event parameter of the M_L 1.8 event on the 20th of December 2016 is varied within the determined uncertainties. The summary of this parameter study is shown in Fig. 3.7a). It can be seen that the variation of the magnitude within the uncertainties (1.8 ± 0.3) has the largest influence on the maximum PGV. As expected, higher magnitudes result in larger ground motion values, in total there is an increase of 700%. The variation of the event depth (3.4 ± 1 km) has the second largest impact. Shallower events cause an increase in the maximum PGV value of 290%. Hereby, a linear trend can be observed between 3000 m below ground level (bgl) and 4400 m bgl, while above 3000 m there is a jump towards higher PGV values. The reason for this is the high impedance contrast between the basement and the overlying Malm formation, which lies at about 2800 m bgl. If the source is located in the basement more seismic energy is reflected at the boundary than transmitted. In contrast, if the source is located above the basement more energy is transmitted to the surface resulting in higher PGV values. However, a hypocenter location above the basement is rather unlikely. In comparison, changing the horizontal location of the epicenter and the focal mechanism within the uncertainties, has a rather low effect on the overall maximum PGV value. A variation of 16% and 7% is observed. However, the spatial variation can be higher in this case, since the lateral distribution of the ground motion changes. For example, the difference at station POI01 for a varying focal mechanism is 110% and for a varying epicenter location 50%.

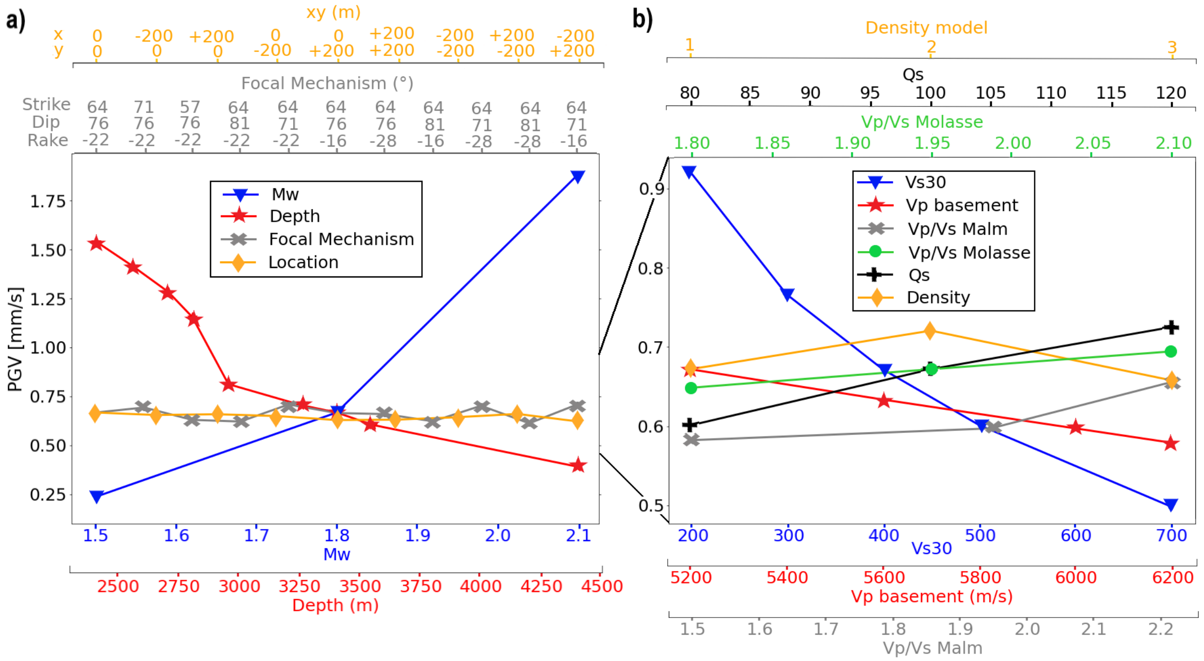

Figure 3.7: a) Maximum PGV for varying values of the event parameters of the $M_L 1.8$ event on 20.12.2016. The parameters are varied within their determined uncertainties (see Table 3.1). The epicenter location is given relative to the original xy-coordinates. b) Maximum PGV for varying values of the velocity, density and attenuation structure. The parameters are varied within the uncertainties discussed in section 3.3.

In the second part of the parameter study, different properties of the subsurface structure are varied including Vs30, Vp/Vs ratios, density and attenuation (Fig. 3.7b). In this case the variation of the shallow velocity structure has the largest effect on the maximum PGV. Varying Vs30 between 200 m/s and 700 m/s causes a clear increase of the PGV value for lower velocities. An amplification of 85% is calculated, which highlights the importance of accounting for site effects. Furthermore, the P-wave velocity (and consequently also the S-wave velocity) within the basement is varied between 5200 m/s, according to the GRAME model, up to 6200 m/s, which is close to the value found by Grad et al. (2009). This causes a decrease of the PGV value of 16% for higher velocities. This can be explained by the larger impedance contrast to the overlaying Malm formation and consequently a larger reflection coefficient. The Vp/Vs ratios within the lower Molasse sediments and the Mesozoic limestones are varied within the corresponding ranges found in Wawerzinek et al. (2021). In both cases, higher Vp/Vs ratios lead to higher PGV values, however, the variation is rather low with 7% for the Molasse sediments and 12% for the Mesozoic layers. The shear wave attenuation values Qs are varied between 80 - 120, while Qp in each case is set to be twice as large as Qs. Larger Q values result in higher PGV values, because attenuation decreases. In total there is a variation of 20%. As a last step different density models are implemented. The first density model has a linear gradient between 2300 kg/m^3 at the surface and 2900 kg/m^3 in the basement, as described in section 3.3. The second density model has a steeper gradient between 2000 kg/m^3 and 2900 kg/m^3 to account for loose sediments close to the surface. The third density model has a linear gradient of 2400 - 2600 kg/m^3 in the upper Molasse sediments and a gradient of 2800 - 2900 kg/m^3 in the Mesozoic layers. Between the different density models there is a small PGV variation of 9%.

These findings help to quantify the impact of the uncertainties contained in the event parameters and in the subsurface model on the simulation results and are used in the following to fine-tune the seismic simulations of the main events. In the following section the simulation results for both main events are presented and discussed.

3.5 Simulation results and discussion

The 3D subsurface model constructed in section 3.3 is implemented together with the seismic sources according to the estimated event parameters in Table 3.1. Different event and subsurface parameter combinations are tested to find the best match with the recorded data and to calibrate the numerical simulations. For a fair comparison both recorded and simulated data are bandpass filtered between 1-10 Hz.

In order to compare the recorded and the simulated data we adapt the time-frequency goodness-of-fit (GOF) criteria developed in Kristeková et al. (2009), which quantifies the level of agreement between two signals. This method is based on the time-frequency (TF) representation of the seismograms obtained as the continuous wavelet transform. This representation has advantages over a simple visual comparison, as it allows a proper quantification and characterization of the differences. Therefore, Kristeková et al. (2009) defined locally and globally normalized time-frequency envelope (TFEM) and phase (TFPM) misfit criteria, which characterize the difference between the envelope/phase of two signals as a function of time and frequency. We apply the global normalization as it allows to account for both the envelope/phase differences at a time-frequency point and the significance of the envelope at that point with respect to the maximum envelope of the signal. Therefore, less weight is given to those parts of the signal where the amplitudes are small compared to the maximum amplitude of the reference signal. This is useful for our case, as we are most interested in the peak ground motion, rather than in particular wave groups with relatively small amplitudes. The projection of the TF misfits onto the time domain gives time-dependent envelope (TEM) and phase (TPM) misfits. Likewise, the projection of the TF misfits onto the frequency domain gives frequency-dependent envelope (FEM) and phase (FPM) misfits. Based on these misfit calculations the GOF criteria can be introduced in order to focus on the level of agreement rather than disagreement between the two signals. Depending on the misfit value, a numerical GOF value between 0 and 10 is assigned, where the GOF criteria approaches zero with an increasing level of disagreement. Furthermore, a verbal evaluation of the fit can be added: we define a score below 4 as a poor fit, a score of 5-6 as a fair fit, a score of 7-8 as a good fit and a score above 9 as an excellent fit (Table 3.2).

Table 3.2: Discrete goodness-of-fit values with verbal evaluation after Kristeková et al. (2009)

Numerical value	**Verbal value**
10	excellent
9	
8	good
7	
6	fair
5	
4	poor
3	
2	
1	
0	

3.5.1 Simulation of the event on 20.12.2016

The M_L1.8 event on the 20th of December 2016 is simulated, because it is the best recorded earthquake in this area. For this event recorded data of the three temporary stations (POI01, POI02, POI03) and the permanent station at the well Th1 are available. However, station POI02 lies outside of the model domain and is therefore not considered in the following. Different simulation runs are performed, while varying the event and subsurface parameters within their uncertainty ranges discussed in section 3.4, to find the best match with the recorded data at station POI01, POI03 and Th1. As main criteria for the acceptance of one model over another we take the maximum PGV values at the stations. The parameter study in section 3.4 gives a good approximation how much each parameter changes the PGV value.

First of all, the P-wave velocity in the basement was corrected to 6100 m/s, as this is a more reasonable value for the crystalline upper crust of the Molasse basin (Grad et al. 2009), and the Vs30 value was set to 350 m/s. The density structure and Vp/Vs ratios were implemented as shown in Fig. 3.3b). The final attenuation values are 100 for Qs and 180 for Qp. In the next step the event parameters are varied in order to match the recorded PGV values on all three components as closely as possible. However, due to the high computationally costs of our method it is not feasible to test all possible parameter combinations. The final event parameters that were used are listed in Table 3.3.

Table 3.3: Adjusted event parameters for the simulation of the induced earthquake on 20.12.2016.

M_W	1.90
Depth	3.1 km
x-coordinate	4484418 + 200 m
y-coordinate	5339022 - 200 m
Focal mechanism	71°, 81°, -16°

The comparison and goodness-of-fit evaluation for all three components of the recorded data and the simulated waveforms at station POI01 are shown in Fig. 3.8. The middle panel in each column shows the time signals, which allows a visual inspection of the similarities and differences. The P- and S-wave onsets are overall in good agreement. In addition, the maximum recorded peak ground velocity value of 0.578 mm/s is recovered very well, as the simulation gives a PGV value of 0.583 mm/s. Differences can be observed in the shape of the waveforms, especially after the first arrivals, which can be better quantified in the GOF TF representation. The upper panels show the time-frequency envelope GOF (TFEG) evaluation and the lower panels the phase (TFPG) evaluation. The TFEG and TFPG indicate an excellent fit for the vertical components (HHZ). This is mainly due to the global normalization scheme that was applied, which gives less weight to low amplitudes. Nevertheless, the PGV value for this component is also matched very well (0.040 mm/s measured, 0.042 mm/s simulated). The level of overall agreement for the east component (HHE) is excellent for the phases and for the envelopes. Slight differences can be observed in the frequency content between 6-8 Hz and below 4 Hz. The phases for the north components (HHN) show an excellent fit and the envelopes a good fit. There are differences observed in the frequency content mainly between 1.5 - 7 Hz. The GOF for the TFEG is lower because not every peak in the S-wave phase is matched. In addition, most weight is given to this component because it recorded the highest amplitudes.

The same evaluation is done for the recorded and simulated waveforms at station POI03 (Fig. 3.A1). The recorded maximum PGV value of 0.101 mm/s is once more recovered very well, here the simulation gives a PGV value of 0.1008 mm/s. However, the waveforms show in general larger differences compared to station POI01. Nevertheless, the level of overall agreement for the N- and E-component is good for the phases and for the envelopes. The TFPG indicates a good fit for the Z-component, while the TFEG only indicate a fair fit. This is mainly due to the overestimation of the amplitudes in the simulated waveform. The evaluation for the simulated and recorded waveforms at station Th1 are shown in Fig. 3.A2 of the appendix 3.A. The level of overall agreement for the N- and E-component is excellent for the phases and good for the envelopes. For the Z-component the TFEG and TFPG both indicate an excellent fit, which is mainly due to the global normalization scheme. Once more the maximum recorded PGV value of 0.344 mm/s is recovered well by

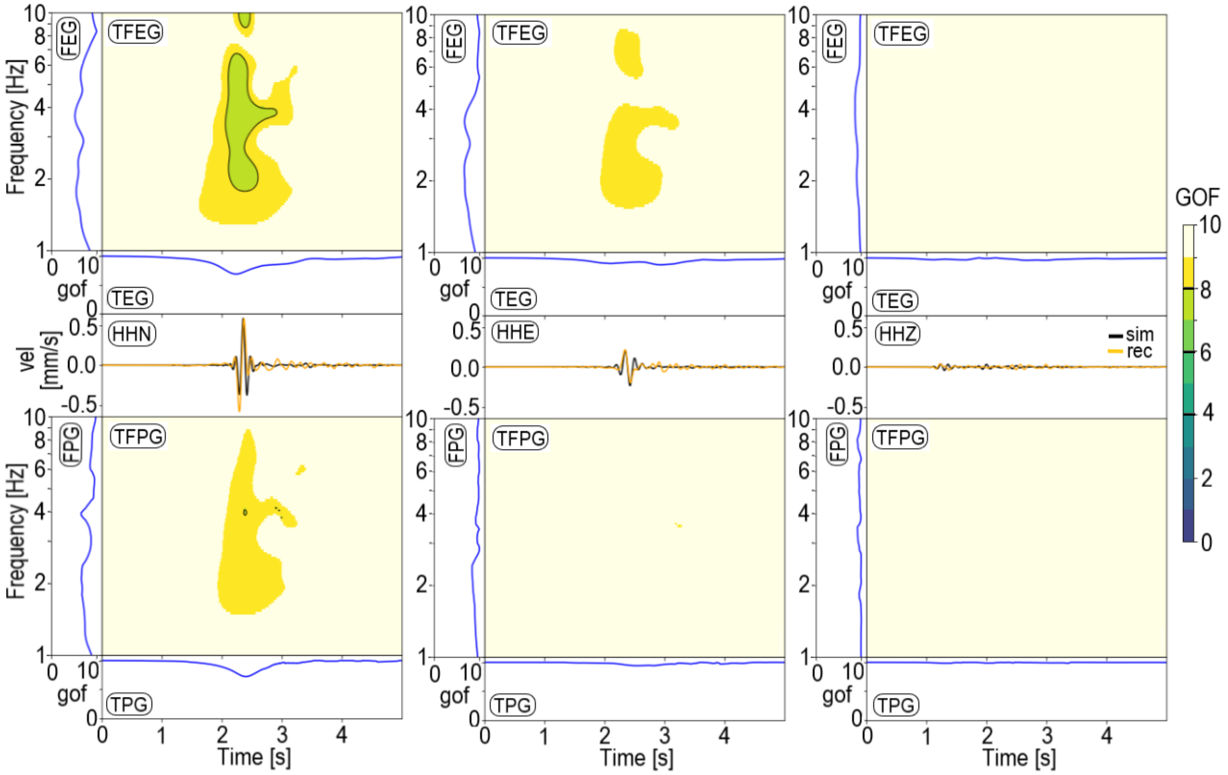

Figure 3.8: Goodness-of-fit evaluation for all three components between recorded data and simulated waveforms at station POI01 for the event on 20.12.2016. The top panel shows the globally normalized time-frequency envelope GOF (TFEG) together with the frequency-dependent envelope GOF (FEG) and time-dependent envelope GOF (TEG). The middle panel shows the time signals of the simulated (sim) and recorded (rec) data. The lower panel shows the globally normalized time-frequency phase GOF (TFPG) together with the frequency-dependent phase GOF (FPG) and time-dependent phase GOF (TPG). The colorscale represents the GOF score between 0 and 10. The black contour lines mark the transition between the verbal evaluations.

the simulation (0.340 mm/s).

In general, the differences in the simulated waveforms compared to the recorded data can be explained by inaccuracies in the subsurface model and the event parameters. Especially the lack of more detailed information on the attenuation structure can have a strong influence on the waveforms and the frequency content. In this case, depth and frequency dependent attenuation values could improve the simulation results. In addition, the stations are located in the junction of the SV and SH radiation pattern of the double couple (Fig. 3.9) and are therefore more sensitive to changes in the focal mechanism. This highlights the importance of correctly estimating the focal mechanism, which however is challenging with a limited number of stations. Nevertheless, overall we obtain a good to excellent fit for stations POI01 and Th1 and a good to fair fit for station POI03. In addition, the maximum PGV value is matched closely for all stations, which is most important for ground motion

studies.
As a last step, a shakemap is constructed from the simulation results, which shows the peak ground velocity values at the surface (Fig. 3.9). Based on this map the distribution of the ground motion can be evaluated. A typical SV/SH radiation pattern is observed, where the largest ground motions occur east and south of the epicenter in accordance with the focal mechanism and the fault geometry. The largest simulated ground motion occurs about 3 km east of the epicenter with a PGV value of 1.52 mm/s between 1-10 Hz, which lies below the critical DIN 4150-3 value of 5 mm/s. Therefore, according to the simulation results this event did not have a damage potential to buildings. Furthermore, the seismic stations POI01, POI03 and Th1 do not lie in the areas of largest ground motion. These findings indicate that the ground motions recorded at the stations were not representative of the largest ground motion induced by the event, which highlights the importance of numerical simulations.

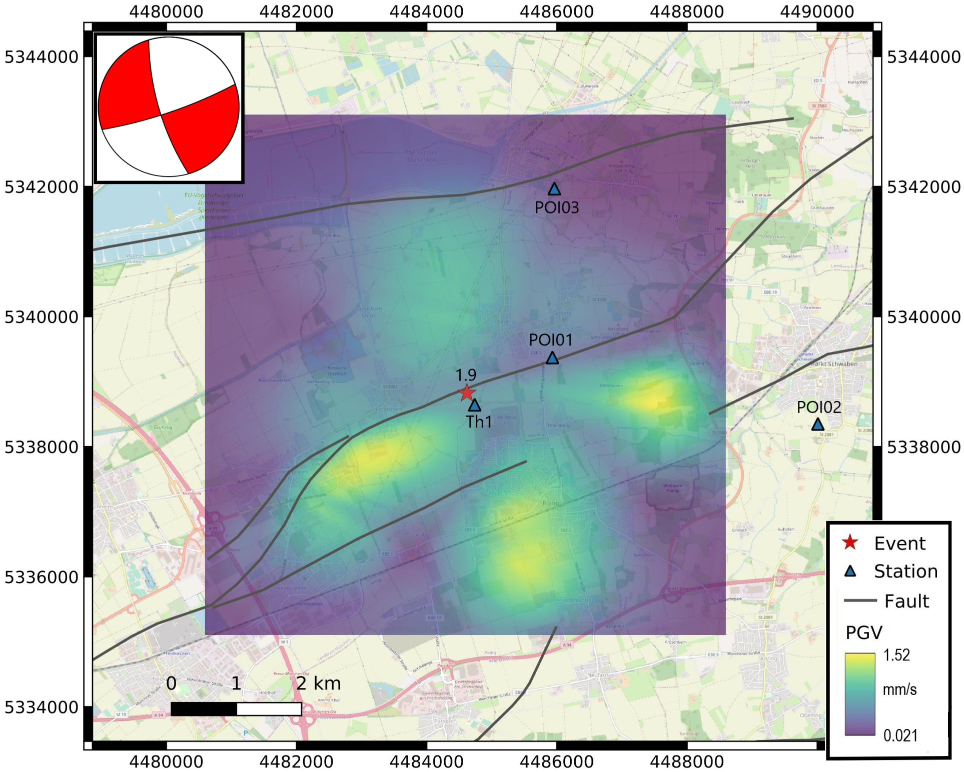

Figure 3.9: Shakemap of the simulated 20.12.2016 event. The colorscale represents the peak ground velocity values at the surface. The fault structure, stations and epicenter are drawn in the map. In the upper left corner the focal mechanism of the event can be seen. The map is referenced by the Gauss-Krüger coordinate system (GK zone 4, EPSG code 31468).

3.5.2 Simulation of the event on 09.09.2017

The M_L2.1 event on the 9th of September 2017 is simulated, because it was one of the largest induced earthquakes in the area. Therefore, we implement the seismic source according to the estimated event parameters in Table 3.1. For the focal mechanism we adapt the values estimated for the M_L1.8 event, as it occurred along the same fault line. For this event recorded data are only available for station Th1. The variation of the event parameters within their uncertainties results in the final event parameters listed in Table 3.4. The subsurface parameters were adapted from the previous simulation in section 3.5.1.

Table 3.4: Adjusted event parameters for the simulation of the induced earthquake on 09.09.2017.

M_W	2.14
Depth	3.2 km
x-coordinate	4484418 - 60 m
y-coordinate	5339022 + 300 m
Focal mechanism	71°, 71°, -22°

The comparison and goodness-of-fit evaluation for all three components of the recorded data and the simulated waveforms at station Th1 is shown in Fig. 3.10. The P- and S-wave onsets in the time signals are in good agreement and the maximum recorded PGV value of 0.821 mm/s is closely recovered by the simulation (0.835 mm/s). The level of overall agreement for the Z-component is excellent for the phases and the envelopes. There are only slight differences in the frequency content visible. The recorded PGV value for this component is 0.224 mm/s, while the simulation gives a value of 0.220 mm/s. The TFEG and TFPG for the N-component indicate an excellent fit with small differences at frequencies above 4 Hz. The phase GOF for the E-component also has an excellent fit, while the envelops have a good fit. There are discrepancies across the whole frequency domain (1-10 Hz), however, most importantly the main peak is recovered.

Once more the shakemap is computed from the simulation results (Fig. 3.11). The largest ground motion occurs approximately 3 km south of the epicenter near the city of Poing. The distribution of the ground motion is influenced by fault effects, which could explain the gap in the SV/SH radiation pattern near Poing. Furthermore, a number of felt intensity reports were transmitted by the population after this event, which resulted in intensities ranging between 1 to 5. Most of the reports were transmitted from Poing, which coincides with the areas of largest ground motion. However, the distribution of the reports can be biased due to the location of residential areas. In the study area there are several smaller communities, which are surrounded by farmland. Nevertheless, in general the distribution of the intensity reports coincides with the simulation results. An exception is the intensity 4-5 estimate, that was transmitted from Kirchheim, south-west of the epicenter. However, since this is an isolated report the intensity could have been over-estimated, as it is a

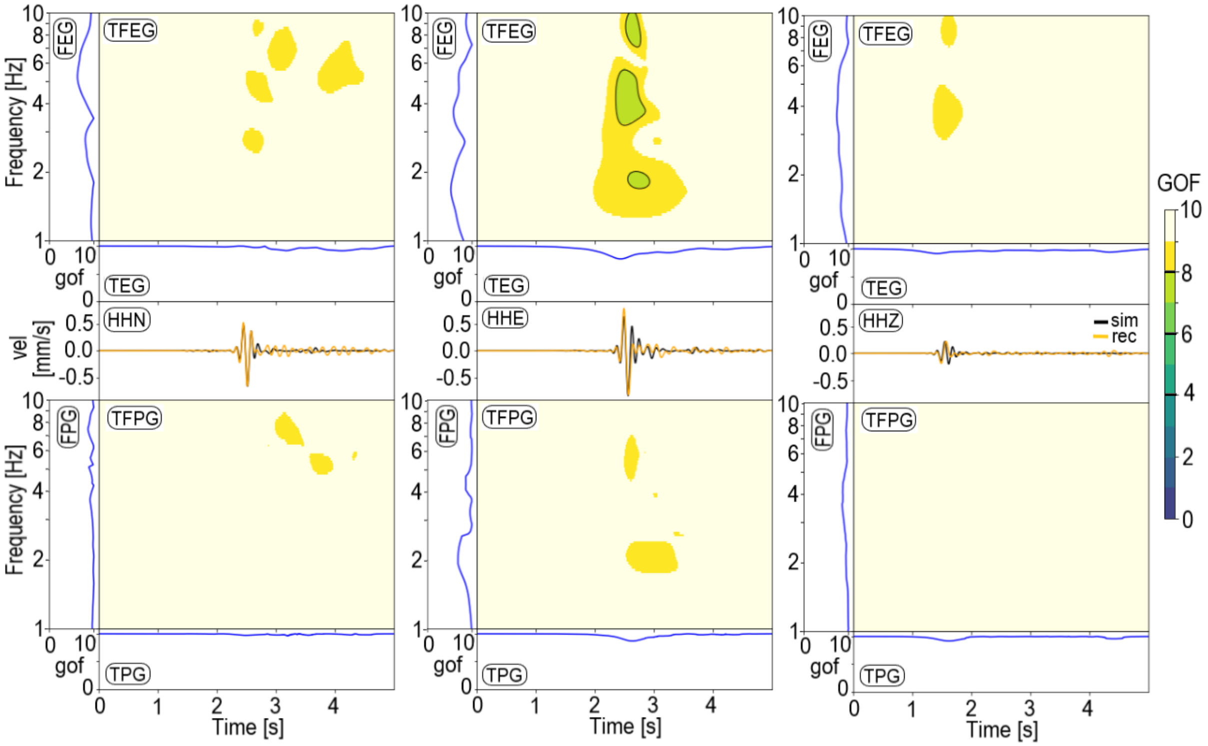

Figure 3.10: Goodness-of-fit evaluation for all three components between recorded data and simulated waveforms at station Th1 for the event on 09.09.2017. The top panel shows the globally normalized time-frequency envelope GOF (TFEG) together with the frequency-dependent envelope GOF (FEG) and time-dependent envelope GOF (TEG). The middle panel shows the time signals of the simulated (black) and recorded (orange) data. The lower panel shows the globally normalized time-frequency phase GOF (TFPG) together with the frequency-dependent phase GOF (FPG) and time-dependent phase GOF (TPG). The colorscale represents the GOF score between 0 and 10. The black contour lines mark the transition between the verbal evaluations.

subjective perception. Once more, the seismic station Th1 does not lie in the area of largest ground motion, which indicates that the ground motion recorded at a single station is not representative of the largest ground motion induced by an event. The maximum PGV value for this simulation is 2.20 mm/s, which lies below the critical DIN 4150-3 value of 5 mm/s. This indicates that even the largest induced event in the area did not have damage potential.

Comparing the shakemap for this event with the one in Fig. 3.9, significant differences in the ground motion distribution can be observed. Small variations can be explained by the varying focal mechanism. However, we find that the main differences are caused by the hypocenter location. In case of the $M_L 2.1$ event, the hypocenter lies north of the main fault line, while in the simulation of the $M_L 1.8$ the hypocenter lies south of it. In this case, more energy is reflected by the fault towards the south, which causes the shift in the ground motion pattern. This highlights the importance of a precise estimation of the hypocenter location, which could be achieved with a larger number of monitoring stations.

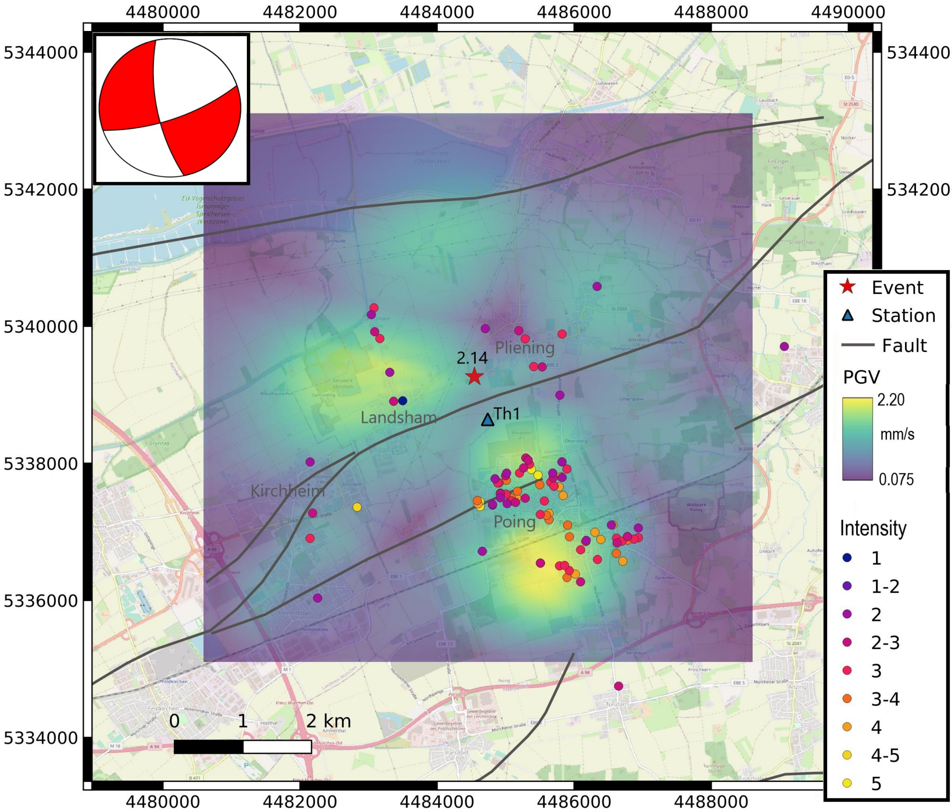

Figure 3.11: Shakemap of the simulated 09.09.2017 event. The colorscale represents the peak ground velocity values at the surface. The fault structure, station, epicenter and location of intensity reports are drawn in the map. In the upper left corner the focal mechanism of the event can be seen. The map is referenced by the Gauss-Krüger coordinate system (GK zone 4, EPSG code 31468).

3.6 Conclusion

We performed numerical simulations of the main induced events that occurred at the geothermal power plant near Poing, using the spectral element code SALVUS. This software package allows the implementation of a detailed 3D subsurface model, which was constructed from different geologic and geophysical data. This enables the performance of realistic and reproducible full waveform modelling. Besides the foreseeable dominant influence of the hypocenter location and the magnitude, the shallow velocity structure (Vs30) also has a significant effect on the peak ground velocity and therefore should not be neglected. The findings from the parameter study are further used to fine-tune the seismic simulations. The goodness-of-fit criteria mostly indicates an excellent to good fit between the simulated waveforms and the recorded data, which demonstrates the potential of seismic simulations for the estimation of seismic shaking scenarios. Larger differences are only observed at the more distant station POI03. In general, the differences can be explained by inaccuracies in the subsurface model and the focal mechanism. Especially the lack of more detailed information on the seismic attenuation structure can detoriate the results. Nevertheless, the peak ground velocity values at the different seismic stations are recovered well by the simulations, which is most important for such ground shaking studies. In addition, the simulated ground motion distribution of the $M_L2.1$ event coincides well with the locations of the felt intensity reports that were transmitted by the population. However, there is a significant difference compared to the shakemap of the $M_L1.8$ event. The different ground motion distributions can be explained by the varying hypocenter location relative to the main fault line and the acting fault effects. Therefore, a precise estimation of the hypocenter location and the focal mechanism are of vital importance. This can only be achieved with a higher number of monitoring stations. Furthermore, the simulation results indicate that even the largest event in the area did not exceed the critical PGV value of 5 mm/s in a frequency range of 1-10 Hz (DIN 4150-3) and therefore did not have damage potential. We showed that numerical simulations are useful to supplement recorded and macroseismic data in order to identify affected areas, especially when the monitoring network is scarce.

Data availability

An input file for running one of the simulations in SALVUS is provided in the supplementary material.

Acknowledgements

This work was supported by the research project SEIGER (Project no. 03EE44003G), funded by the German Federal Ministry of Economic Affairs and Climate Action and managed by the research institute Jülich (PTJ). The authors thank the SALVUS-Team, in particular Lion Krischer, Michael Afanasiev and Christian Boehm, for their technical support. In addition, the authors thank the two reviewers for their helpful comments which improved this manuscript.

3.A Appendix: Additional simulation results of the event on 20.12.2016

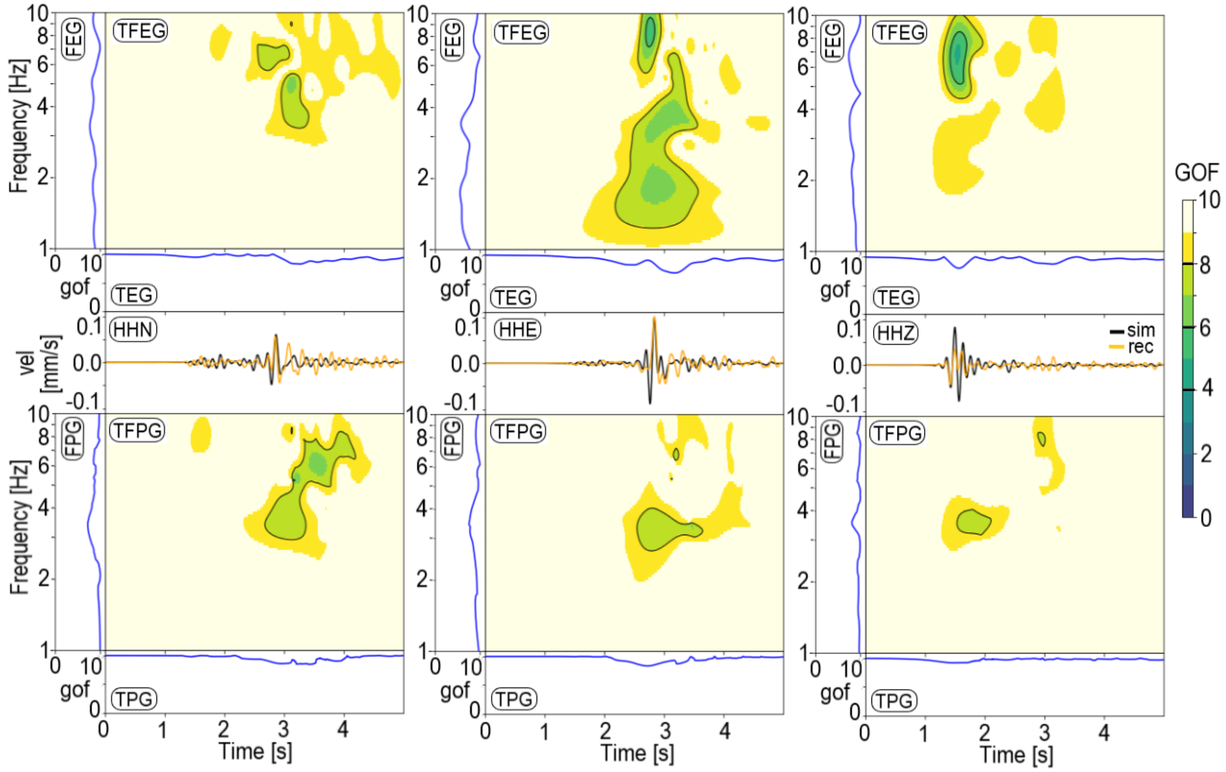

Figure 3.A1: Goodness-of-fit evaluation for all three components between recorded data and simulated waveforms at station POI03 for the event on 20.12.2016. The top panel shows the globally normalized time-frequency envelope GOF (TFEG) together with the frequency-dependent envelope GOF (FEG) and time-dependent envelope GOF (TEG). The middle panel shows the time signals of the simulated (black) and recorded (orange) data. The lower panel shows the globally normalized time-frequency phase GOF (TFPG) together with the frequency-dependent phase GOF (FPG) and time-dependent phase GOF (TPG). The colorscale represents the GOF score between 0 and 10. The black contour lines mark the transition between the verbal evaluations.

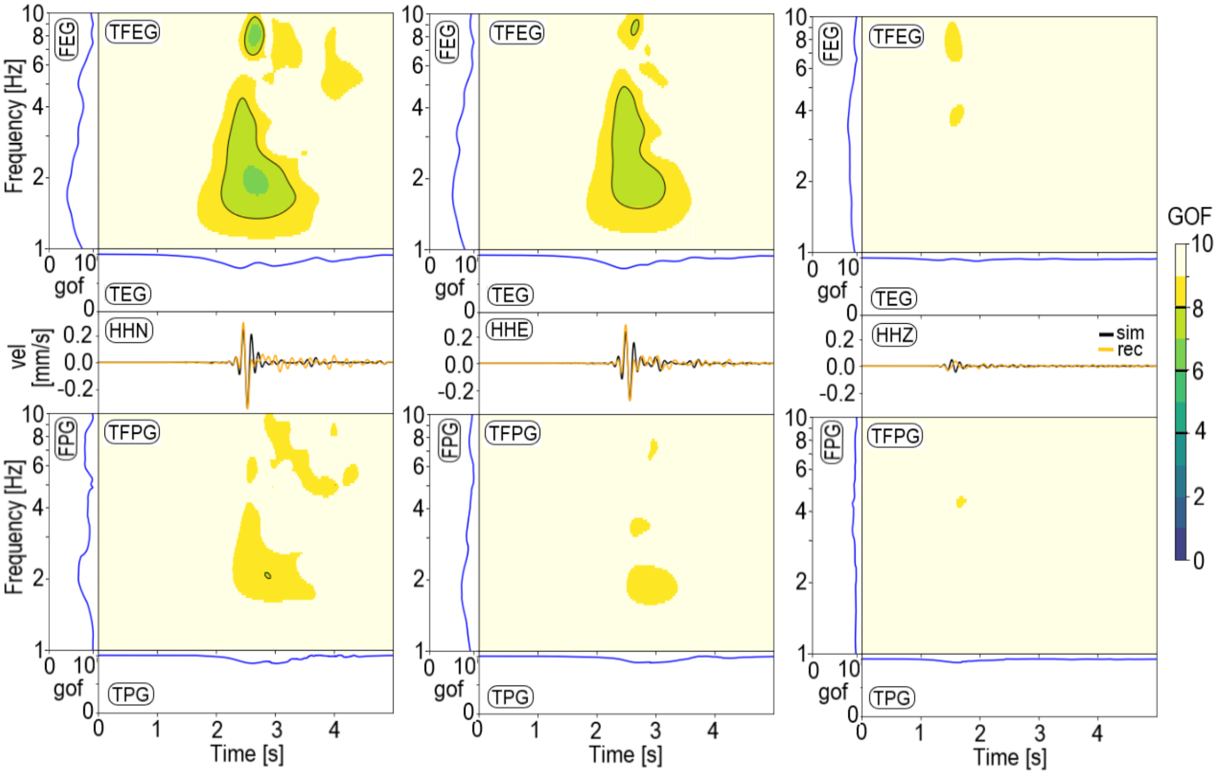

Figure 3.A2: Goodness-of-fit evaluation for all three components between recorded data and simulated waveforms at station Th1 for the event on 20.12.2016. The top panel shows the globally normalized time-frequency envelope GOF (TFEG) together with the frequency-dependent envelope GOF (FEG) and time-dependent envelope GOF (TEG). The middle panel shows the time signals of the simulated (black) and recorded (orange) data. The lower panel shows the globally normalized time-frequency phase GOF (TFPG) together with the frequency-dependent phase GOF (FPG) and time-dependent phase GOF (TPG). The colorscale represents the GOF score between 0 and 10. The black contour lines mark the transition between the verbal evaluations.

Chapter 4

Optimal Network Design for Microseismic Monitoring in Urban Areas - A Case Study in Munich, Germany

by Sabrina Keil, Joachim Wassermann, Tobias Megies and Toni Kraft

Submitted to Seismica (2023)

Abstract

Well-designed monitoring networks are crucial for obtaining precise locations, magnitudes and source parameters, both for natural and induced microearthqakes. The performance of a seismic network depends on many factors, including network geometry, signal-to-noise ratio (SNR) at the seismic station, instrumentation and sampling rate. Therefore, designing a high-quality monitoring network in an urban environment is challenging due to the high level of anthropogenic noise and dense building infrastructure, which can impose geometrical limitations and elevated construction costs for sensor siting. To address these challenges, we apply a numerical optimization approach to design a microseismic surveillance network for induced earthquakes in the metropolitan area of Munich (Germany), where several geothermal plants exploit a deep hydrothermal reservoir. First of all, we develop a detailed noise model for the city of Munich, to capture the heterogeneous noise conditions. Then, we calculate the expected location precision for a randomly chosen network geometry from the body-wave amplitudes and travel times of a synthetic earthquake catalog considering the modeled local noise level at each network station. In the next step, to find the optimum network configuration, we use a simulated annealing approach in order to minimize the error ellipsoid volume of the linearized earthquake location problem. The results indicate that a surface station network cannot reach the required location precision (0.5 km in epicentre and 2 km in source depth) and detection capability (magnitude of completeness $M_c = 1.0$) due to the city´s high seismic noise level. In order to reach this goal, borehole stations need to be added to increase the SNR of the microearthquake recordings, the accuracy of their body-wave arrival times and source parameters. The findings help to better quantify the seismic monitoring requirements for a save operation of deep geothermal projects in urban areas.

4.1 Introduction

The main purpose of seismic networks is to determine earthquake locations and magnitudes, which is important for earthquake characterization, hazard assessment and emergency response both for natural and induced seismicity (e.g. Trnkoczy et al. 2009, Lomax et al. 2009). Specifically, induced seismicity caused by geothermal energy production is a growing concern, since the number of geothermal projects is raising in search of carbon-free heat and electricity generation (Hirschberg et al. 2014, Lund & Toth 2021). In most cases the induced events have small magnitudes ($M_L < 2$) and are not felt by the local population (Evans et al. 2012). However, examples like the Deep Heat Mining Project in Basel, Switzerland (Häring et al. 2008), and geothermal projects near Strasbourg, France (Schmittbuhl et al. 2021), where induced events with magnitudes $M_L > 3$ were recorded, highlight the importance of managing the induced seismicity risk. Hereby, a good monitoring network is a necessary component of the risk governance strategy to detect and locate small magnitude earthquakes, which enable the functioning of magnitude-based traffic light systems (Kraft et al. 2020). The precision of earthquake location depends on several factors, such as the distribution of seismic stations, detection of seismic waves and the accuracy of their observed and calculated arrival times (e.g. Bondár et al. 2004, Trnkoczy et al. 2009). However, low-SNR recordings hamper the detection of small magnitude events and lead to high location uncertainties, which result in a poor performance of the monitoring network (e.g. Bormann & Wielandt 2013). This is especially an issue in urban areas where often high seismic noise levels are encountered. Even though, well-designed monitoring networks are fundamental to allow the detection of weak seismic signals, seismic network planning is still mainly performed as a manual task based on simple design rules, which may fail in complex settings. Several approaches have been proposed to solve this problem in the framework of D-optimal network design (e.g Steinberg & Rabinowitz 2003, Kijko 1977). In this study we are applying the method of Kraft et al. (2013), which builds on the simulated annealing approach proposed by Hardt & Scherbaum (1994). This approach allows the optimization of seismic networks in complex settings, taking into account user-specified velocity models and heterogeneous noise conditions, as well as already existing monitoring stations. The program returns expected location uncertainties and detection thresholds of the resulting network.

We apply this method to the metropolitan area of Munich, where currently 17 deep geothermal power plants operate (Agemar et al. 2014). This includes the geothermal project in Schäftlarnstraße (SLS), which is located in Munich´s inner-city with a total of six deep wells (3 production, 3 re-injection) and a footprint of several square kilometers (Lentsch & Schweingruber 2022). Since induced earthquakes were observed at surrounding geothermal power plants with magnitudes up to 2.4 (Megies & Wassermann 2014, Seithel et al. 2019), the induced seismicity risk needs to be considered also at this recently realized project, which rises the requirement for a high quality monitoring network. The monitoring network for the geothermal power plants south of Munich was already optimized during the MAGS2 project (Megies & Wassermann 2017*b*), however, the inner-city project SLS had not been constructed at that time.

Since the number of geothermal projects in Germany is raising and consequently the risk of induced seismicity increases, Baisch et al. (2012) proposed a number of seismic monitoring recommendations for induced seismicity for the German Research College Physics of the Earth (FKPE). They recommend a monitoring network that is able to reliably detect and locate all earthquakes with magnitudes $M_L \geq 1$ with epicentral uncertainties of less than 500 m and vertical uncertainties of less than 2 km. These thresholds should be reached in an area of 5 km surrounding the target areas of the geothermal project. For the following quality assessment of the monitoring network in the Munich area, we are taking these recommendations into account.

First of all, we construct a detailed model of anthropogenic noise in the metropolitan region of Munich to capture its heterogeneous noise conditions. In the next step the quality of the existing monitoring network is evaluated according to the FKPE recommendations. Afterwards, a number of numerical network optimization runs are performed that test how the FKPE recommendations can be met by adding new surface and boreholes stations to the existing network.

4.2 Methodology

For the network optimization, we use a simulated annealing code initially developed by Hardt & Scherbaum (1994) that was substantially extended by Kraft et al. (2013) (hereafter referred to as NetOpt3D). Due to license issues, NetOpt3D was recently rewritten by Antuens et al. (2023) using open software libraries. For the current analysis, the python wrapper pyNetOpt3D (Megies et al. 2023) was built around the binaries of Antuens et al. (2023) to handle the input and output of the optimization code more easily. In the following, we briefly describe the concept of NetOpt3D and pyNetOpt3D.

NetOpt3D finds the D-optimal design by minimizing the volume of the error ellipsoid of the linearized earthquake location problem (D-criterion) (e.g. Kijko 1977) using a simulated annealing approach (Kirkpatrick et al. 1983). In order to solve the optimization problem the program computes traveltimes of seismic body waves using the finite difference ray tracer of Podvin & Lecomte (1991) and a user-defined velocity model. Furthermore, to evaluate the detectability of an event at the seismic stations body wave amplitudes are calculated based on earthquake source processes and wave propagation effects. Path effects are only treated in an approximate way by geometrical spreading, constant attenuation and free-surface amplification. The Brune model (Brune 1970) is implemented as seismic source. The SNR is defined as the ratio of the synthetic body wave amplitude and the observed or estimated long-term root-mean-squared ground velocity at the station. We choose a SNR of 5 as the threshold for an earthquake to be observed at a certain station. In general a SNR ≥ 3 is considered being sufficient to reliably detect a seismic phase onset in a seismogram (e.g. Hardt & Scherbaum 1994, Baisch et al. 2012). However, we chose a more conservative threshold as the estimated signal amplitude in our optimization approach corresponds to the maximum expected amplitude of the considered body wave at the recording station, which may be significantly larger than the amplitude of the phase onset (Kraft et al. 2013). The estimated SNR of a seismic phase at a station is then used to calculate the expected uncertainty of the phase's onset time following the approach of Aki (1976). The signal bandwidth needed for this estimate is approximated by the max(fc, fmax), where fc is the Brune corner frequency of the event, and fmax is the high-frequency band-limitation of the radiated field estimated from the attenuation model of Edwards et al. (2011) for Switzerland. More details about the NetOpt3D program, including the annealing schedule and the calculation of body wave traveltimes and amplitudes are given in Kraft et al. (2013).

In its current form, NetOpt3D is lacking usability and it is time consuming to set up new optimization problems. Input files (e.g. velocity models, synthetic earthquake catalogs) have to be set up manually in fixed legacy ASCII formats defined by the underlying C codes and a large number of helper programs (e.g. Linux shell scripts) are used for preparational steps and for analysis and visualization of results. Therefore, the consistent and easy-to-use Python Application Programming Interface (API) pyNetOpt3D was developed that internally uses NetOpt3D C codes but hides all unwieldy steps from the user. It enables the start of a complete optimization run with a single, short Python script using the newly developed API. All coordinate conversions from global geographic coordinates (WGS84) to local geodetic coordinates (e.g. UTM, Gauß-Krüger, Swiss Grid, ...) and vice versa are handled automatically. Functionalities to calculate convex hulls, buffers and equistant station grids are included. It also enhances reproducibility by providing (de)serialization of a full optimization run including all input data and results into a single file. Furthermore, pyNetOpt3D provides command line tools to quickly plot optimization results from a serialized file on disk.

In order for the NetOpt3D program to perform the optimization a number of user-specified input data is required, which will be discussed in detail in the next sections.

4.3 Ambient noise analysis

The detectability of an event at a specific station depends on the amplitude of the earthquake signal and the noise level at the site. Therefore, an estimate of the background noise at the existing stations and the potential new network sites is required. First of all, we investigate the frequency content of the seismic noise by computing power spectral densities (PSD) at several stations located in the Munich city (Fig. 4.1). The surface station SYBOB[1] clearly shows higher PSD values for frequencies above 3 Hz compared to the underlying 180m-deep borehole station SYBAD. The highest power at SYBOB is observed between 10-20 Hz, while the PSD values at SYBAD decrease for frequencies above 6 Hz. The PSD values of the surface station MNH are high for frequencies larger than 2 Hz. Above 5 Hz the temporary installed station EGA displays PSD values lower than SYBOB and MNH, which can be explained by the installation within a park area. From these observations it can be inferred that the anthropogenic noise sources (e.g. trains, vehicles, construction work, industrial operation) influence the noise amplitudes at high frequencies

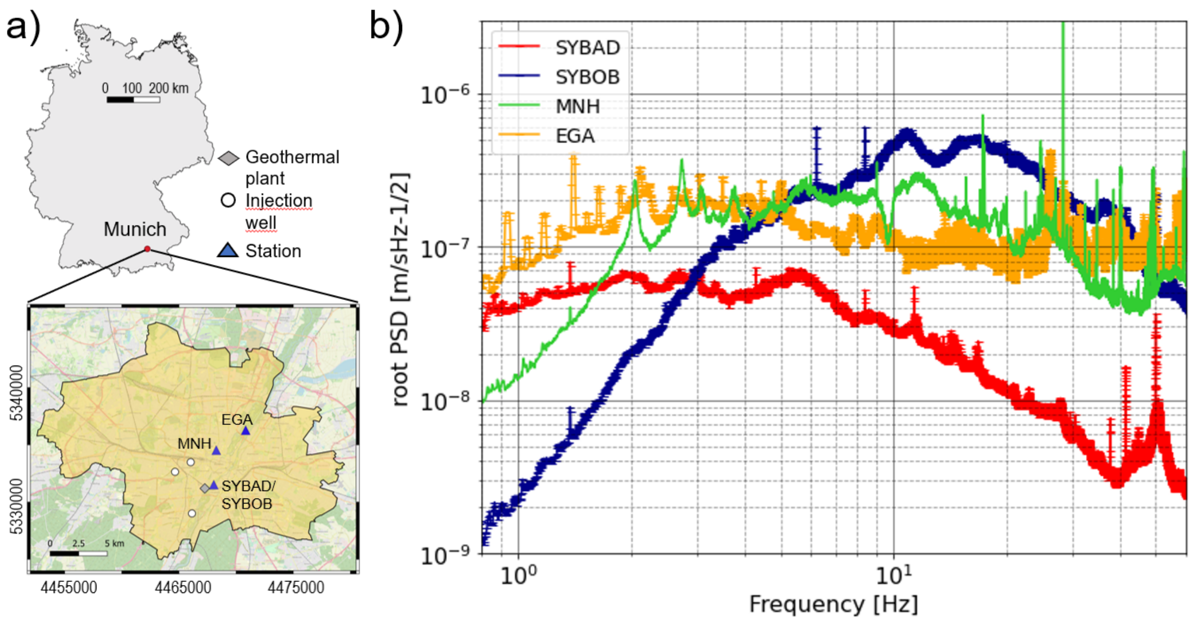

Figure 4.1: a) The upper map shows an overview of Germany with the location of Munich marked. The lower map shows an overview of the Munich city with the geothermal power plant at Schäftlarnstraße and its three injection wells. The locations of several seismic stations in Munich are marked. SYBAD corresponds to a 180m-deep borehole station and SYBOB is its overlying surface station. MNH is a permanent surface station and EGA was temporarily installed within a park area. At SYBOB a 4.5 Hz geophone is installed, at MNH a Mark L4-3D 1 Hz seismometer, at EGA a Trillium Compact 120s seismometer, and at SYBAD a Trillium Compact PH 20s seismometer. The coordinate system is Gauss-Krüger (GK4). b) Root Power spectral density (PSD) plots of data recorded at the seismic stations marked in a). The PSDs were computed from the vertical component for one day of data.

---

[1] Note that at SYBOB a 4.5 Hz geophone is installed, therefore the data should not be interpreted for frequencies much lower than 4.5 Hz.

(>1 Hz), which is consistent with findings of other authors (e.g. Asten & Henstridge 1984, Groos & Ritter 2009). In addition, the noise amplitudes at the seismic station can be reduced through installation in boreholes and more isolated areas, like parks and green spaces.

Another measure to evaluate the noise level at a site is the I95 value, which represent the 95^{th} percentiles of the ground velocity amplitude recordings (Fig. 4.2 a). We calculate the I95 values in a frequency range of 1-20 Hz, which contains the dominant amplitudes of the cultural noise and corresponds to the main frequency range of the induced events observed in the Munich area (Megies & Wassermann 2017*b*). To investigate the variation of anthropogenic noise, the I95 values are computed at the surface station SYBOB and the borehole station SYBAD for 10-minute time windows during the daytime and nighttime, respectively. The computed I95 values are summed in violin plots and the median is taken as a representative value for the noise amplitude at the site (Fig. 4.2 b). A clear variation between daytime and nighttime is visible. For the surface station SYBOB the median noise amplitudes are reduced by a factor of 2 during the night. In addition, the noise amplitudes at the surface station SYBOB are by a factor of 10 larger for the vertical component compared to the borehole station SYBAD. This value is close to a factor of 13 that is estimated using the simple assumption that the noise level in the borehole decreases by a

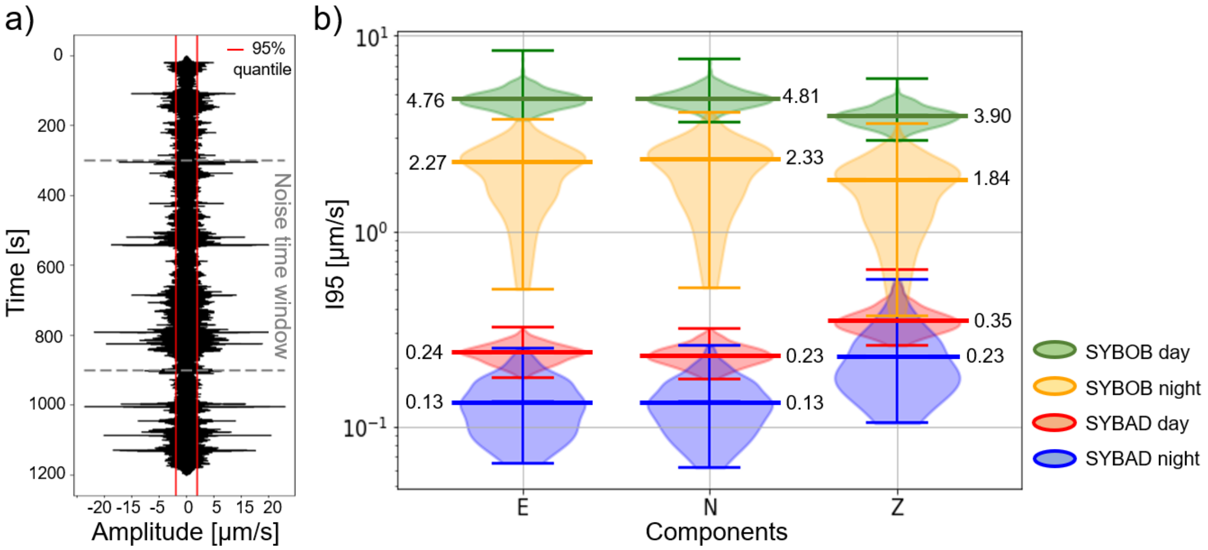

Figure 4.2: a) Seismic noise recorded at the surface station SYBOB in Munich in a frequency range of 1 - 20Hz. The 95% quantile of the data is shown by the red lines. The I95 value is computed from the 95% quantiles for 10 minute time windows. b) Violin plots of I95 values calculated at the surface station SYBOB and the underlying 180m-deep borehole station SYBAD for the east (E), north (N) and vertical (Z) component. The I95 values were calculated over 5 days in 10 minute time windows and were separated into daytime and nighttime. The median values are marked in the plot.

factor of $\sqrt{depth[m]}$. Assuming the most inconvenient noise conditions for the detection of microseismic events, we take the median I95 value during the day as a measure for the noise amplitudes at the site. In order to implement the calculated noise values into the pyNetOpt3D program the I95 values have to be converted to root-mean-square (RMS) ground velocity values. Assuming that the noise distribution is Gaussian, the I95 values can be converted by $RMS = I95/2$ (Neuffer & Kremers 2017).

To estimate the background noise at the potential new network sites, a noise map for the Munich area has to be developed. Kraft (2014, 2016) developed an ambient seismic noise model for Europe based on land-use data derived from satellite imagery by the European Commission project CORINE (Büttner et al. 2004) and open GIS data on infrastructure from the OpenStreetMap project. The model is available for Europe in a $250m \times 250m$ resolution and divides the surface into three classes that represent good, intermediate and bad ambient noise conditions. Kraft (2014, 2016) defined following RMS bounds for each noise class: Low: $RMS \leq 30nm/s$, Middle: $30nm/s < RMS \leq 120nm/s$, High: $RMS > 120nm/s$. Almost the entire Munich city is characterized by high ambient noise values (Fig. 4.3). By comparing the measured noise values at the stations with the values assigned in the noise map, we see that they are mainly underestimated in the model. Therefore, for optimizing the seismic monitoring network in the urban area of Munich a more detailed noise model is required in order to capture the small-scale heterogeneous noise conditions.

We develop such a noise model for the Munich city extending the approach of Kraft (2014, 2016). First of all, land-use data from the Bavarian surveying administration (see data availability) is used to categorize the area into different classes including industrial buildings, residential buildings, sports and recreation areas, vegetation or water bodies and based on that assign a minimum noise level (Table 4.1). In the second step, different types of roads are identified as noise sources and subdivided into different classes based on OpenStreetMap data (see data availability). Highways are assumed to have a higher noise contribution, compared to intercity roads, railways or residential streets. In order to account for noise propagation away from these sources, we implement noise-distance relations, that were derived from seismic measurements at distinct noise features (Riedl 2017). Hereby, several seismometers were installed with increasing distance from the source to map the decreasing amplitude of the ambient vibrations. As last input traffic volume data from the city of Munich (see data availability) are implemented to adjust the noise level for busy roads. The overall noise value at one point is then calculated by adding the minimum noise level assigned from the land-use data and the noise contribution of the main sources scaled by the noise-distance relation. The resulting noise model of Munich´s inner-city (Fig. 4.4) has a resolution of 5 × 5 m. To verify the calculated noise levels we compare them to the measured noise values at permanent and temporary installed stations. For sites with low noise level the calculated values are mostly close to the measured values. For sites with high noise level our model underestimates the RMS value, which is most likely due to noise sources and site effects that are not mapped into our model. As can be seen in Fig. 4.4, our noise model for Munich is dominated by street traffic noise. In addition,

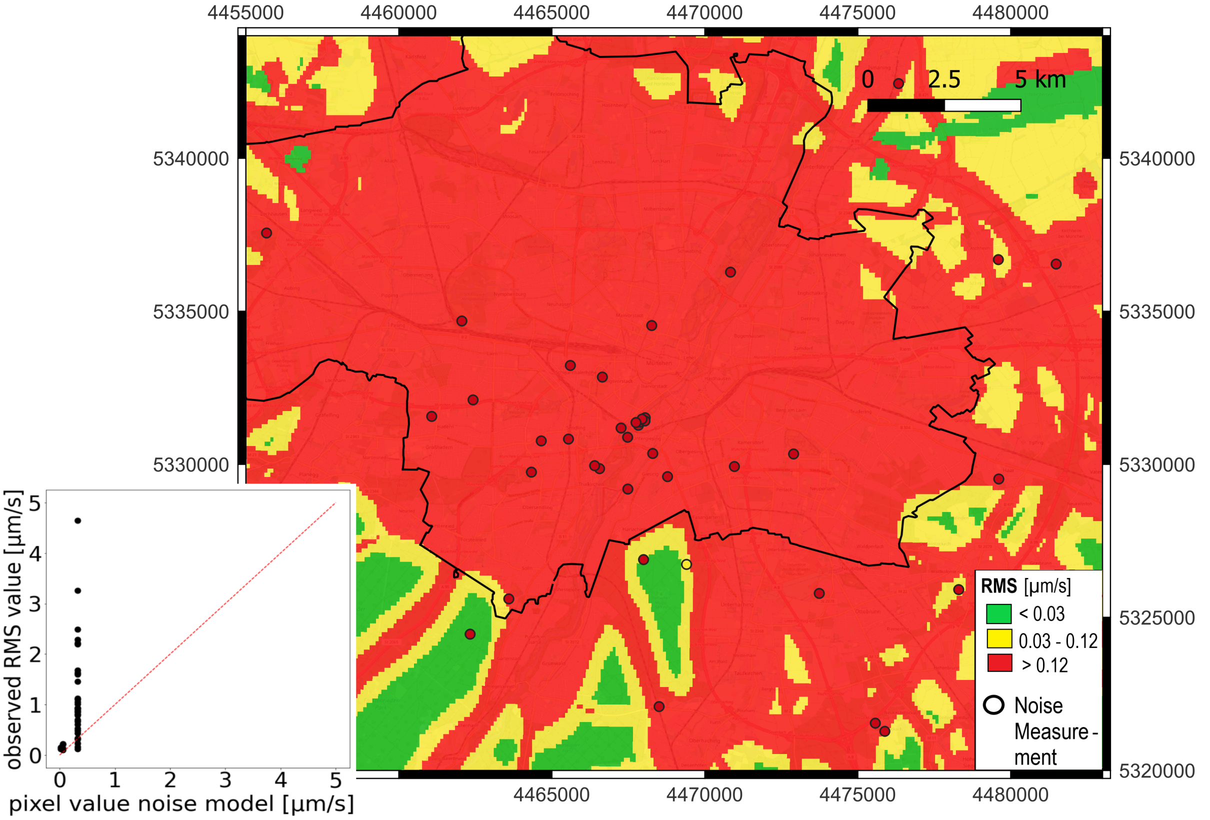

Figure 4.3: Noise map of Munich after Kraft (2014, 2016). The city boundary is outlined by the black line. The area is divided into three noise classes with low, intermediate and high noise values. The green areas are assigned a value of 0.015 µm/s, the yellow area 0.06 µm/s and the red area a value of 0.325 µm/s. The small circles show noise measurements at permanent and temporary installed seismic stations. The coordinate system is Gauss-Krüger (GK4). In the lower left corner the observed RMS value at the seismic stations are plotted against the calculated pixel value in the noise model.

the overall noise level in the city center is higher compared to the surroundings. Nevertheless, even within the city low noise areas are identified, which might be suitable for the installation of monitoring stations. We implement the high-resolution noise map of the Munich city into the larger-scale background noise map of Kraft (2014, 2016) for the surrounding areas.

Table 4.1: Land use classes with assigned minimum I95 noise level after Riedl (2017).

Land use class	**Noise value [μm]**
Industrial usage	1.2
Housing	0.6
Sports/recreation	0.3
Vegetation, water	0.15

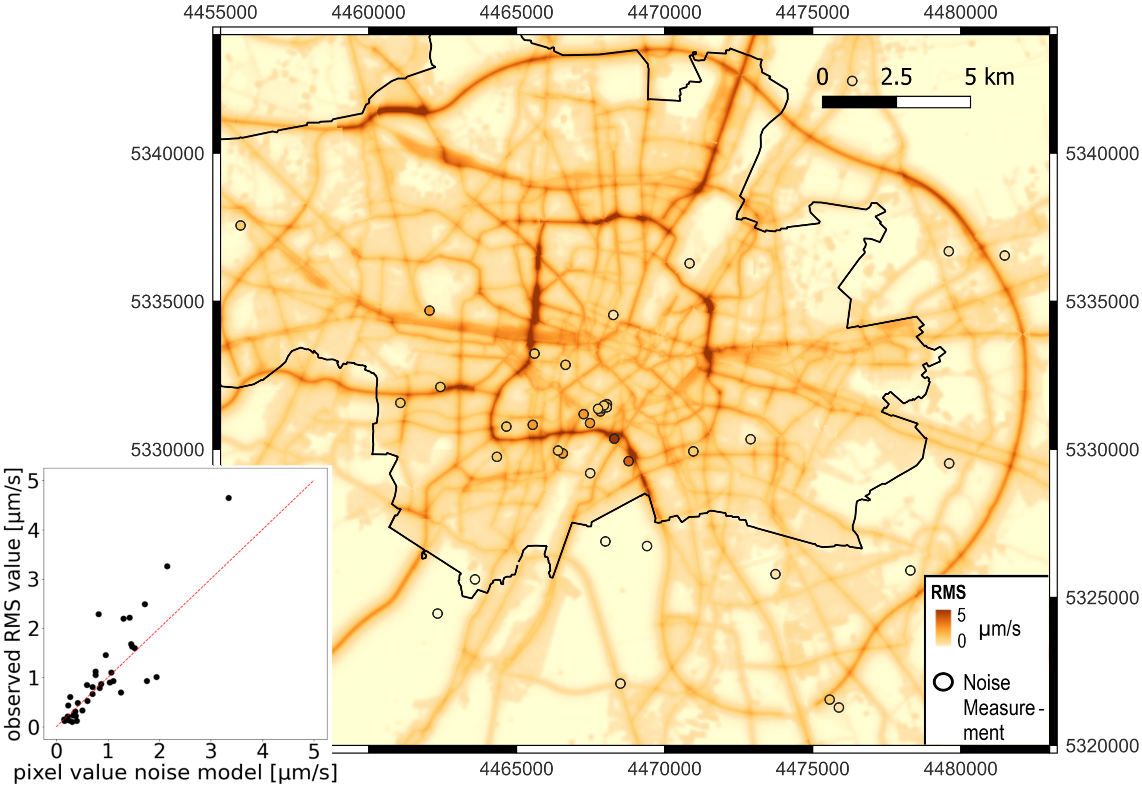

Figure 4.4: High-resolution noise map of Munich´s inner-city. The city boundary is outlined by the black line. Colors represent the noise level, which is calculated as I95 values in a frequency range of 1-20 Hz and converted to RMS. Circles show locations of noise measurements from permanent and temporary installed seismic stations. The coordinate system is Gauss-Krüger (GK4). In the lower left corner the observed RMS value at the seismic stations are plotted against the calculated pixel value in the noise model.

4.4 Model set-up

To calculate the signal-to-noise ratio at the potential station, we implement the high resolution noise model developed in section 4.3. As next step, in order to calculate body wave traveltimes, a velocity model has to be implemented. In the Munich area, information on the boundaries of the main geological units are available from a structural model developed by the Bavarian State Office for Environment (Bayerisches Landesamt für Umwelt 2012). The P-wave velocities within the layers are based on a 3D seismic survey conducted in 2015/16 as part of the GRAME project (Hecht & Pletl 2015), which covered 170 km^2 in the southern and western parts of Munich. The S-wave velocities are calculated from Vp/Vs ratios found by Wawerzinek et al. (2021) for the Munich area. The NetOpt3D program is able to implement 3D velocity models, however, in this study we only consider a 1D velocity profile (Fig. 4.5) since we assume that 3D effects only have a minor influence on the results.

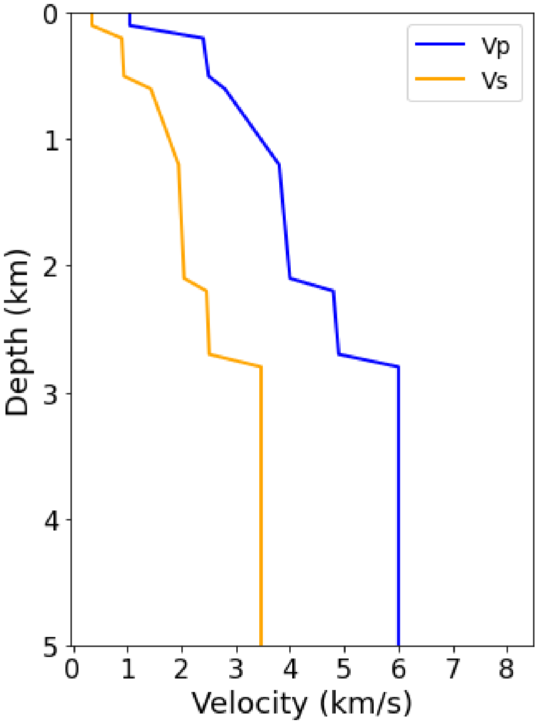

Figure 4.5: 1D P- and S-wave velocity profiles (Vp, Vs) implemented into pyNetOpt3D for the calculation of body wave amplitudes and traveltimes.

Seismic waves attenuate while propagating and their amplitudes usually decrease with propagation distance. To account for seismic attenuation, we implement the attenuation model of Eulenfeld & Wegler (2016) for the geothermal project in Unterhaching south of Munich, since the ray geometry and geologic setting at this site is very similar to the one expected for other locations in the study area. They estimated a mean S-wave quality factor (Qs) of 100 averaged over the whole ray path, which is constant for frequencies lower than 8 Hz. Due to the lack of further information on the attenuation of P-waves, we set the P-wave quality factor (Qp) to 200, as literature suggests that Qp is approximately two times higher than Qs (e.g. Fowler 1990).

For the network optimization a synthetic earthquake catalogue has to be generated. We place the events in the crystalline basement at 3 - 4 km depth underneath the re-injection wells of the geothermal power plants, as most of the recorded induced seismicity occurred close to these locations (Megies & Wassermann 2014, Seithel et al. 2019). The focal mechanisms for the events were chosen to resemble those of the known induced earthquakes and the fault geometry in the study area, which generally corresponds to left-lateral strike-slip mechanisms with normal faulting component. We implement the events with M_W 1.3, which was converted from M_L 1.0 according to the relation found by Grünthal & Wahlström (2003) for earthquakes in central Europe.

As the optimization algorithm is able to take already existing stations into account, we implement the existing surface and borehole stations in the area with their observed noise levels.

As a last step, we have to define the geographical region for possible new station locations. We set the station perimeter with a maximum distance of 8 km to the earthquake epicenters, which corresponds to approximately twice the maximum hypocentral depth. Placing the stations at greater distance would not improve the network performance, as will be shown in section 4.5. The station perimeter was then filled by a grid of possible station locations with a spacing of 100 m, which is enough to cover the low-noise areas within the city. With decreasing station spacing the computational costs increase since a larger number of network configurations has to be tested. Locations where it would be impossible to install a station, e.g. in water bodies, were already excluded from this grid.

The final set-up of the input data, generated by pyNetOpt3D and used by the binaries of Antuens et al. (2023) for the optimization, is shown in Fig. 4.6.

4.5 Optimization results and discussion

First of all, the performance of the existing network with a focus on the area surrounding the recently installed SLS power plant is tested using the NetOpt3D program without optimization option. The performance in case of a M_W 1.3 event at 3 km depth is tested, which corresponds to the minimum detectable M_W-converted magnitude recommended by the FKPE (Baisch et al. 2012), and is hereafter referred to as target event. The program returns expected location uncertainties and number of detections (i.e., recordings with SNR ≥ 5). The largest number of P-wave detections per event is reached for events occurring south of the city-center, while this number decreases significantly in the northeast and in the surrounding of SLS (Fig. 4.7 a). In the northernmost part of the 5 km radius surrounding SLS, the target events would be detected by even less than 3 stations. The location uncertainties are calibrated using recorded events at the geothermal plants in the southern part of the study area (Megies & Wassermann 2014). The threshold for the FKPE-recommended epicentral uncertainty of < 500 m is only reached south of the

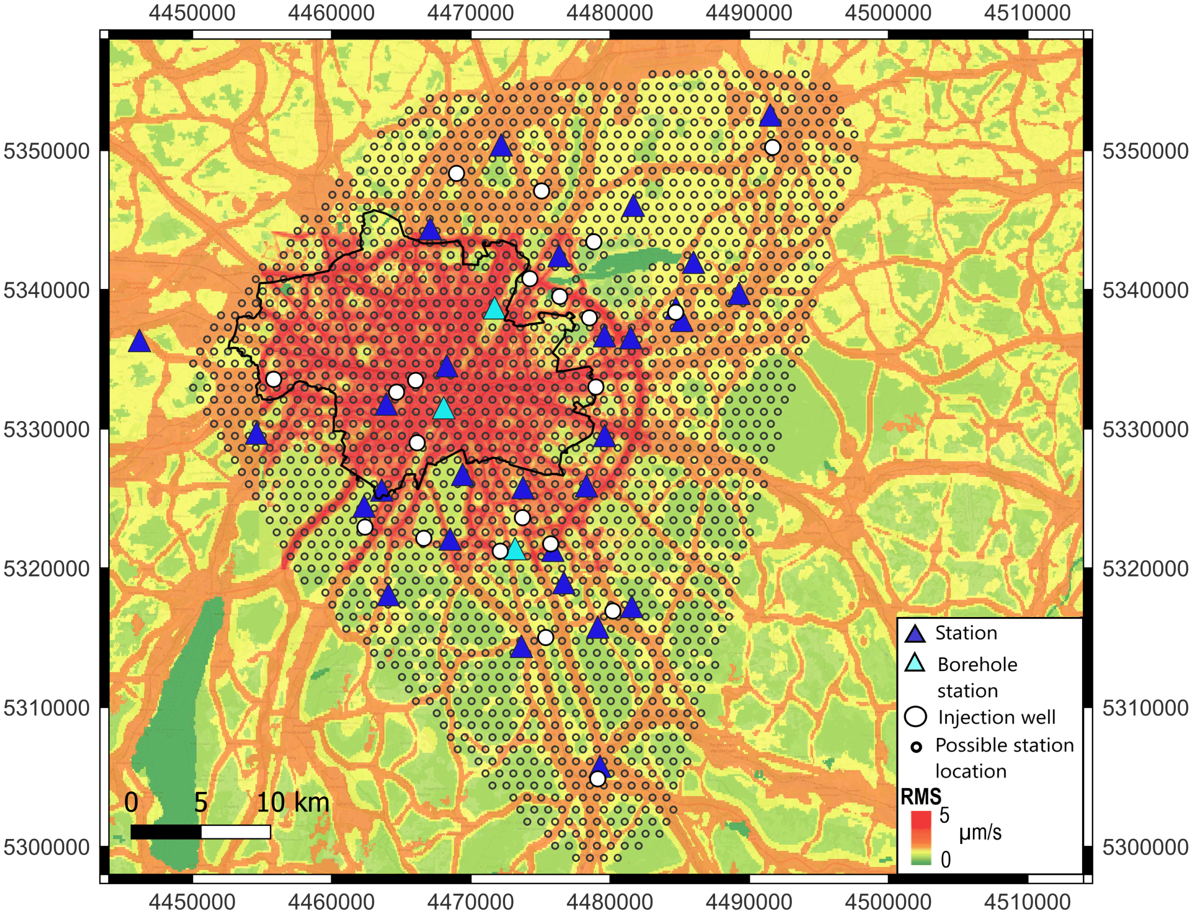

Figure 4.6: Set-up of the input data for the network optimization program. The Munich city boundary is outlined by the black line. Existing surface and borehole stations, as well as location of injection wells are plotted. The event locations are placed at the injection wells. The colors show the computed background noise level as RMS ground velocity. Small circles represent schematically the grid of possible station locations that can be selected during the optimization process. The coordinate system is Gauss-Krüger (GK4).

city-center, while in the vicinity of SLS epicentral uncertainties of more than 2 km are computed. The 2 km threshold for the vertical uncertainty is once more mostly reached south of the city-center. In general, the poor performance of the existing network in the SLS area can be explained by 1) a lack of monitoring stations in the northwest and a consequent azimuthal gap in this region and 2) the high noise levels in the inner-city, which cause low-SNR recordings resulting in poor onset-time precision and consequently higher location uncertainties.

Considering these observations, we next evaluate how to improve the seismic network by adding new stations. We perform an optimization run for the randomly chosen number of 5 new surface stations, implementing the input data as shown in Fig. 4.6. The NetOpt3D

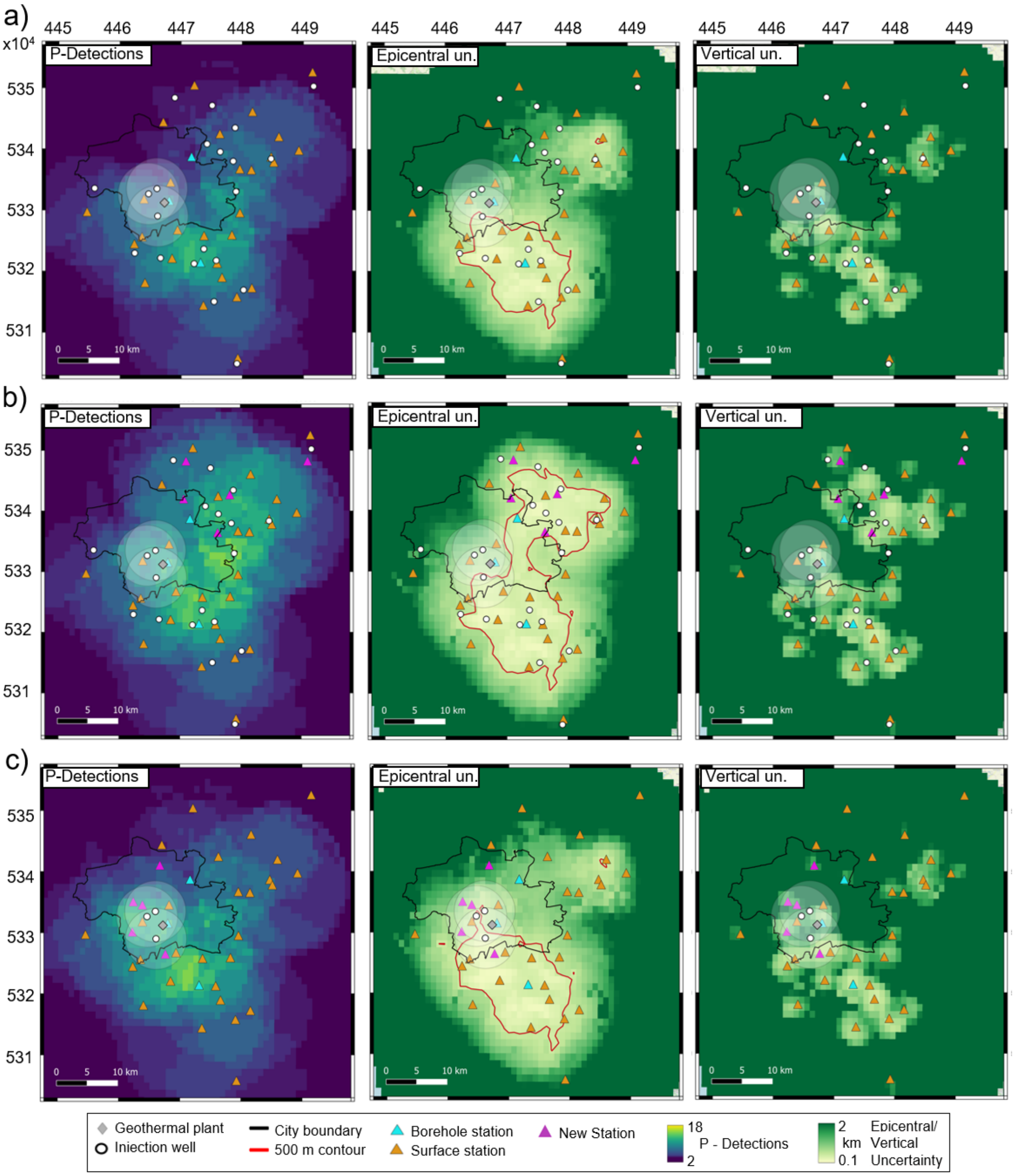

Figure 4.7: Evaluation of monitoring performance for a M_W 1.3 event at 3 km depth. The performance of the a) existing network, b) optimized network with 5 new stations under consideration of all re-injection wells in the region, c) optimized network with 5 new stations and focus on the three inner-city re-injection wells, is shown. The panels from left to right show the number of P-arrival detections (i.e., recordings with SNR ≥ 5), the epicentral uncertainty and the vertical uncertainty. The location of the inner-city geothermal power plant SLS is plotted. The shaded circles around the three SLS re-injection wells mark a radius of 5 km. The red outline in the epicentral uncertainty plots mark the 500 m contour line. The coordinate system is Gauss-Krüger (GK4).

program performs the simulated annealing and returns the optimal locations for these 5 new stations (Fig. 4.7 b). All the new stations are placed in the north-northeast, which increases the number of P-wave detections and decreases the epicentral and vertical uncertainties in this area significantly. Nevertheless, in the vicinity of SLS the performance only slightly improved, since none of the stations was placed in the city center. The algorithm placed most of the stations in the north-northeast as the noise levels are lower compared to the city-center and the code tends to locate stations in the quietest sites only (Kraft et al. 2013). Furthermore, it resulted in the largest improvement of the network performance since the improved SNR at a quiet site overrules the lower SNR at a geometrically more optimal site (Kraft et al. 2013).

In order to improve the network specifically in the city center, we perform a new optimization run with 5 new surface stations, but only considering the three SLS re-injection wells as event locations. Therefore, the grid of possible station locations only samples the city center. This time the algorithm places the 5 stations closer to SLS (Fig. 4.7 c). Accordingly, the number of P-wave detections increases in this region. In addition, the epicentral and vertical uncertainties decrease, however, it is not enough to reach the FKPE-recommended location accuracy. The reason are the relatively low SNR values, which results in a poor onset-time precision. Again, the algorithm places the new stations in low noise areas (Fig. 4.8), which mainly correspond to park areas within the city. This highlights the importance of a high-resolution noise map.

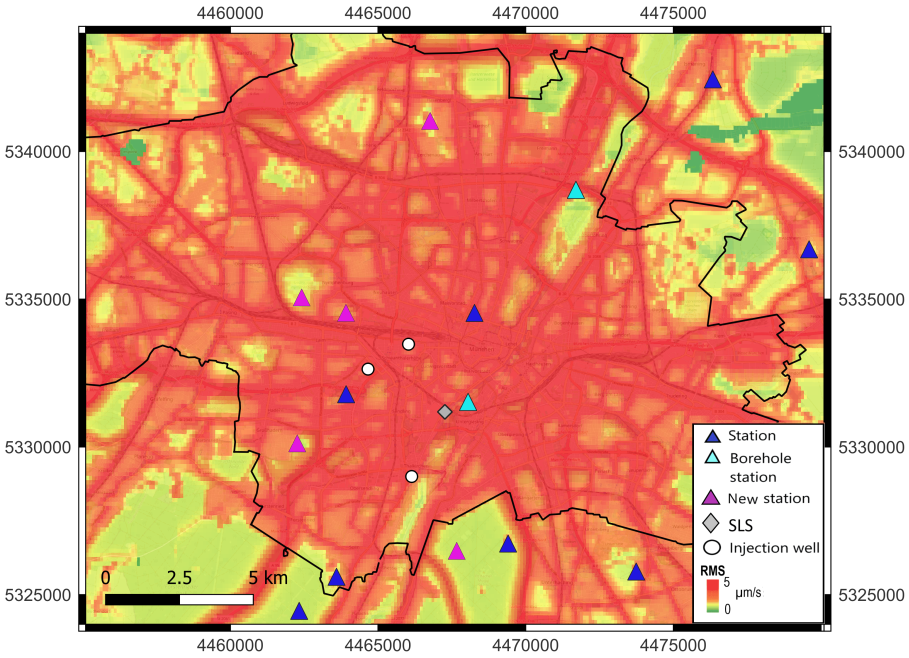

Figure 4.8: Zoom into the network optimization result for the scenario shown in Fig. 4.7 c). The pink triangles mark the 5 new surface stations placed by the algorithm. The colors show the computed background noise level as RMS ground velocity. The coordinate system is Gauss-Krüger (GK4).

To allow the algorithm to choose low-noise areas outside of the city, we increase the station perimeter from 8 km to 12 km. Nevertheless, the algorithm still places four of the new surface stations close to the SLS power plant and only one station closer to the edge of the city (Fig. 4.9 a). The resulting epicentral and vertical uncertainties are similar to the values in Fig. 4.7 c). Therefore, we have shown that considering a station perimeter of 8 km is enough, as placing station at larger distance does not improve the monitoring per-

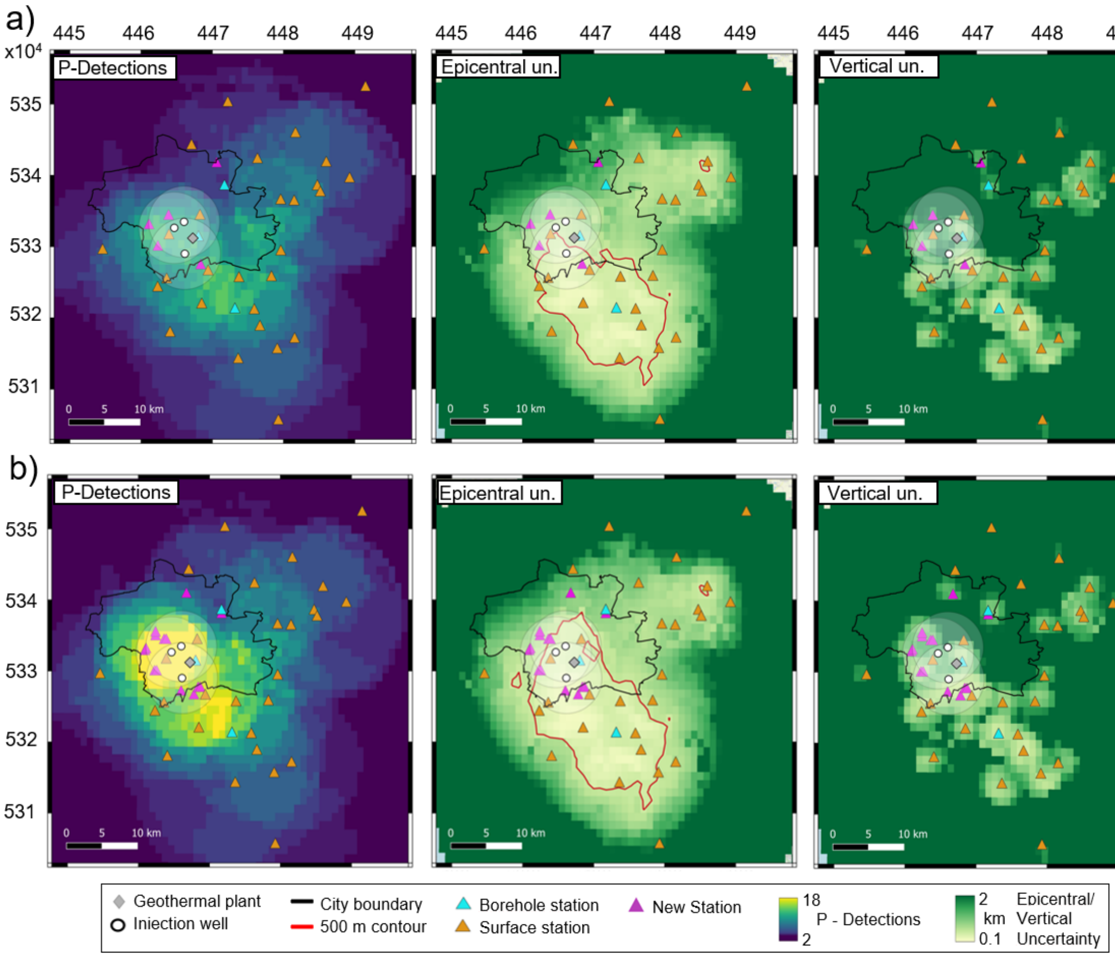

Figure 4.9: Evaluation of monitoring performance for a M_W 1.3 event at 3 km depth. The performance of the a) optimized network with 5 new stations considering a station perimeter of 12 km and focus on the three inner-city re-injection wells, b) optimized network with 15 new stations and focus on the three inner-city re-injection wells, is shown. The panels from left to right show the number of P-arrival detections (i.e., recordings with SNR $\geq$ 5), the epicentral uncertainty and the vertical uncertainty. The location of the inner-city geothermal power plant SLS is plotted. The shaded circles around the three SLS re-injection wells mark a radius of 5 km. The red outline in the epicentral uncertainty plots mark the 500 m contour line. The coordinate system is Gauss-Krüger (GK4).

formance significantly. This is most likely related to the decreasing amplitude of the ground motion away from the epicenter.

To see if a larger number of surface stations could reach the recommended location precision, the same optimization run is performed using 15 new stations (Fig. 4.9 b). The number of P-wave detections significantly increases. Nevertheless, even though the epicentral and vertical uncertainties improve it is not sufficient to reach the FKPE-recommended location precision in the vicinity of SLS. In fact, adding even more stations does not significantly improve the location precision any further.

In order to increase the SNR and allow a more accurate determination of the event location, borehole stations are considered in the next step of the optimization. In section 4.3 the 180m-deep borehole station SYBAD was compared to the overlying surface station SYBOB. We observed that for the vertical component the noise level in the borehole is a factor of 10 lower than at the surface. Therefore, to simulate the noise level for borehole stations in Munich we divide the noise model by a factor of 10 and input it into the NetOpt3D program. Then a network optimization for borehole stations is performed. We find that at least 5 new borehole stations are sufficient to reach the recommended epicentral uncertainty of less than 500 m in the surroundings of the SLS re-injection wells (Fig. 4.10 a). Additionally, the vertical uncertainty threshold of < 2 km is reached almost within the entire 5 km radius, except for some outermost parts.

To estimate the minimum required borehole depth we stepwise decrease the scaling factor for the noise map. We find that a scaling factor of 6 is sufficient to reach the recommended location accuracy (Fig. 4.10 b). Assuming the simple relation of noise decreasing with depth by a factor of $\sqrt{depth[m]}$ this would correspond to a borehole depth of 36 m.

Even though, borehole stations significantly improve the quality of the monitoring network, their installation is not always feasible due to high costs and infrastructural limitations. Therefore, we test if less borehole stations in combination with additional surface stations could also reach the FKPE-recommended location precision. At first, the optimization is performed for 3 new borehole stations by scaling the noise map with a factor 10. This is followed by an optimization run with 5 new surface stations, while fixing the previously determined borehole stations. The recommended epicentral and vertical uncertainty thresholds are reached in this case (Fig. 4.10 c).

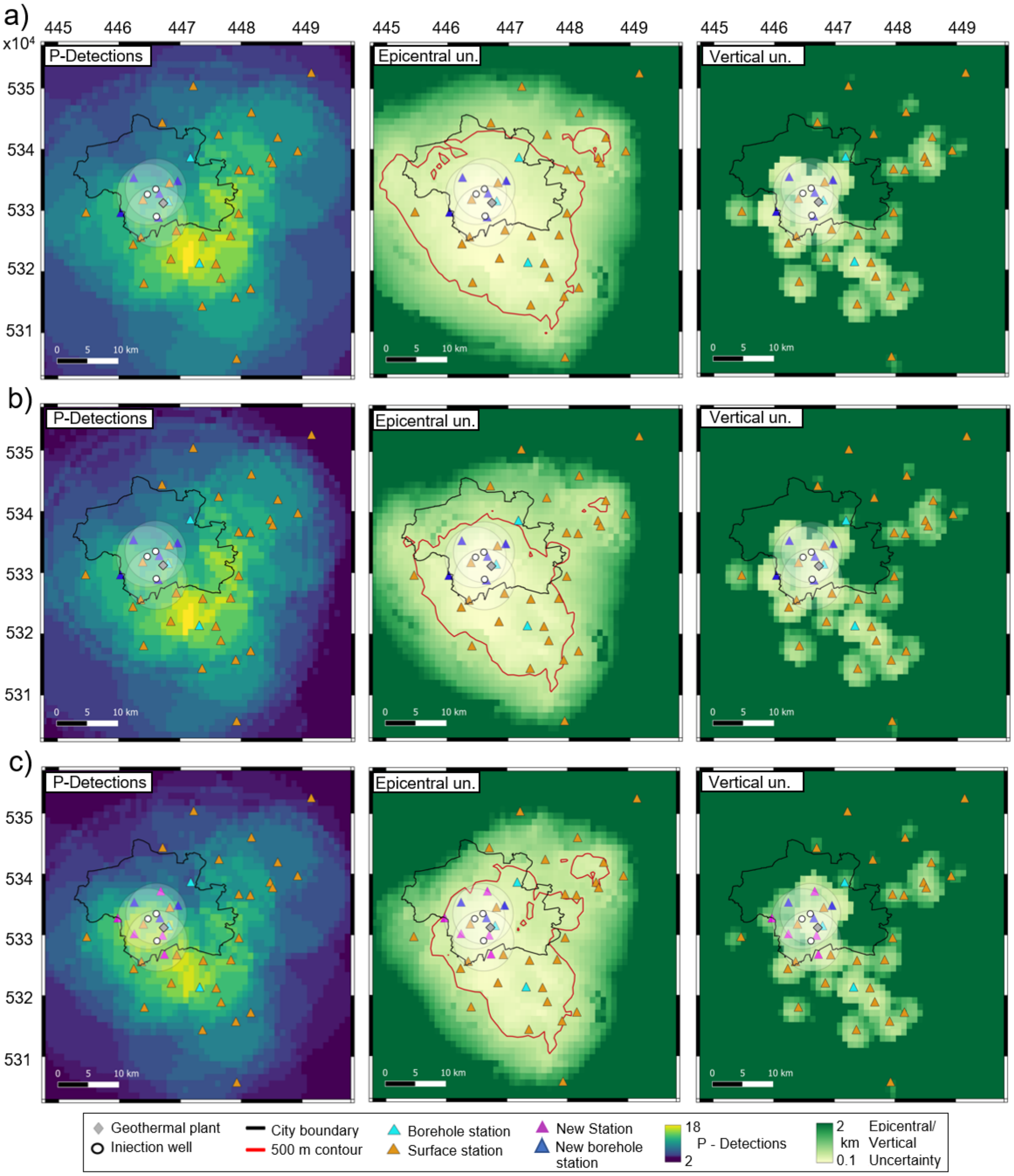

Figure 4.10: Evaluation of monitoring performance for a M_W 1.3 event at 3 km depth. The performance of the a) optimized network with 5 new borehole stations and focus on the three inner-city re-injection wells, b) optimized network with 5 new shallower borehole stations and focus on the three inner-city re-injection wells, c) optimized network with 3 new borehole stations and 5 new surface stations and focus on the three inner-city re-injection wells. The panels from left to right show the number of P-arrival detections, the epicentral uncertainty and the vertical uncertainty. The location of the inner-city geothermal power plant SLS is plotted. The shaded circles around the three SLS re-injection wells mark a radius of 5 km. The red outline in the epicentral uncertainty plots mark the 500 m contour line. The coordinate system is Gauss-Krüger (GK4).

4.6 Conclusion

We performed a network optimization using the python wrapper pyNetOpt3D around the NetOpt3D program in order to improve the microseismic monitoring for a save operation of deep geothermal plants in Munich´s inner-city. In the first step we constructed a noise model for the Munich area in order to capture the heterogeneous noise conditions. This high resolution noise model enabled the algorithm to find suitable station locations even within the city center. The results suggest that the current monitoring network is not suitable to locate M_L 1 earthquakes with a FKPE-recommended epicentral uncertainty of < 500 m and vertical uncertainty of < 2 km. We showed, that adding solely surface stations to the inner-city network is not sufficient to reach the recommended thresholds. The addition of borehole stations significantly improved the quality of the monitoring network, which indicates that borehole installations are indispensable in urban environments. However borehole installations are not always feasible and come with high costs. We were able to show that a combination of new borehole and new surface stations can be used to record and locate M_L 1 events in Munich with the recommended location precision. Our results indicate solutions for improving the microseismic monitoring within urban areas both for induced and natural seismicity. Nevertheless, we emphasise that proper seismic monitoring is only one component of a comprehensive risk governance strategy for induced seismicity.

Data and resources

The Geographical base data from the Bavarian administration for geographical surveying (Geobasisdaten Bayerische Vermessungsverwaltung 2017) was requested at
https://www.ldbv.bayern.de/ (last request July 30, 2017).
The OpenStreetMap data were downloaded from:
https://www.openstreetmap.org/export#map = 11/48.0290/11.6331 (last request July 29, 2017).
The traffic volume data from the city of Munich for 2019 were searched at
https://stadt.muenchen.de/infos/verkehrsdaten.html (last accessed July 10, 2020).
The pyNetOpt3D code and the binaries of NetOpt3D from Antuens et al. (2023) can be downloaded from *https://doi.org/*10.5281/*zenodo*.763885.

Acknowledgements

This work was supported by the research project SEIGER (Project no. 03EE44003G), funded by the German Federal Ministry of Economic Affairs and Climate Action and managed by the research institute Jülich (PTJ). T. Megies was supported by the GEOBEST-CH project of the Swiss Seismological Service funded by the Swiss Energy Program of the Swiss Federal Office of Energy for the development of pyNetOpt3D.

Chapter 5

Conclusion and Outlook

The main goal of this thesis was to develop new tools to manage the risk of induced seismicity at inner-city deep geothermal power plants. In doing so, the existing challenges for the microseismic monitoring in urban environments and the estimation of the ground motion already in the planning stages of a geothermal project, as well as after an event, were addressed on the example of Munich, Germany.

First of all, a new single-station method to estimate the shallow velocity structure was developed to make seismic microzonation feasible in urban environments and to remove the ambiguity in the inversion results. I could show that it is possible to compute Love and Rayleigh wave dispersion curves from single-station six-component ambient noise recordings, combining three translational and three rotational motions. However, a limitation exists in the lower frequency range, which is a combination of the self-noise level of the blueSeis-3A rotational sensor and the small rotational amplitudes at low frequencies. However, this limitation is overcome by adding H/V ratios to the inversion, which then allows the computation of a 1D velocity profile for the upper 100 m of the subsurface. The comparison to established array measurements shows consistent results, which indicates the potential of our new method. Differences in the dispersion curves are only observed at high frequencies, which can be explained by the different spatial resolutions of the methods, as well as the higher sensitivity of rotational motions to small-scale structures. The great advantage of the single-station 6C method over array measurements is the fast and easy installation, which allows the performance of noise measurements even in environments with limited open space.
Furthermore, the comparison of the computed velocity structure with borehole derived lithologic profiles gives a positive correlation. Therefore, in the future these 1D velocity profiles might be linked to the available 3D hydro-geologic Geopot model (Technical-University-Munich 2016-2019) of Munich´s shallow subsurface in order to inter- and extrapolate the velocity structure to retrieve a complete microzonation map for Munich. This significantly reduces the amount of required noise measurements for the spatial interpre-

tation of local site effects.
The neighborhood inversion that was applied in the microzonation study for the inversion of surface wave dispersion curves presented some challenges, such as the subjective choice of the number of subsurface layers, tuning parameters and velocity constraints. To overcome the limitations of conventional inversion methods, a machine learning algorithm (ML) was trained based on a mixture density network to invert for the shallow velocity structure. The ML model is able to distinguish between subsurface structures with 2 - 7 layers and returns a complete probability distribution of the S-wave velocity without fixing the layer depth. I was able to show that complicated structures are predicted with higher accuracy compared to a regular neighborhood inversion. In addition, once the ML models are trained the inversion is performed much faster compared to a Markov chain Monte Carlo inversion, which makes it applicable to multidimensional problems. Furthermore, the ML models are robust against previously untrained structures, such as velocity gradients, and work well on the recorded microzonation data from Munich. Nevertheless, the depth and velocity of deeper layers are resolved with less accuracy, which is most likely related to the decreasing sensitivity of the dispersion curve data with depth. Therefore, in the future the models could be extended to higher mode surface waves, since they are more sensitive to deeper structures (Pan et al. 2019) and/or H/V spectral ratios, which provide information in the lower frequency range.

In the next step, the shallow velocity structure estimated from the microzonation study was implemented together with additional geological and geophysical information into a 3D model to perform numerical simulations with the spectral element code SALVUS. The parameter study showed that a low velocity Vs30 layer can have a significant effect on wave amplification, which highlights the importance of seismic microzonation for the estimation of the ground motion. The Vs30 values computed from the 1D velocity profiles in Munich classify the subsurface into site class C and D, which indicates a moderate amplification effect. Therefore, this low velocity layer must be considered in the seismic simulations.
Since during the period of this thesis no induced events were recorded at the Schäftlarnstraße (SLS) geothermal power plant, I simulated two induced events that occurred at the geothermal power plant in Poing, located 15 km east of Munich, in order to verify the simulation results. The comparison of the simulated waveforms with the recorded data gives an overall good to excellent fit. Specifically, the amplitude values were recovered with a high accuracy, which is most important for the ground motion analysis. From the simulation results shakemaps were computed to estimate the spatial distribution of the ground motion, which coincides well with macroseismic data. The computed PGV values are below the critical 5 mm/s (DIN 4150-3), which indicates that even the largest event in the area (M_L 2.1) did not have damage potential.
After I showed that the numerical simulations give reliable results, they can be applied in future studies to the geothermal power plant at SLS in Munich´s inner-city. Different scenarios can be simulated in order to estimate the maximum ground motion and its spatial distribution, as well as the maximum tolerable magnitude of induced events. This is an important step in estimating the seismic risk already in the planning stages of a geothermal

project and to develop risk governance strategies. In addition, the simulations can be used to evaluate the shaking effects after an event in order to determine affected areas according to the German mining laws. A spatial interpretation of the ground motion is otherwise difficult due to the sparse station network.
The drawback of the simulation approach are the relatively high computational costs, which make the computation of many scenarios, as well as the application to problems that need real-time solutions, infeasible. Therefore, in future studies alternative methods are to be tested, such as the reduced-order modeling (ROM) approach that can be used to create simulated PGV maps for different earthquake sources in real-time (Rekoske et al. 2022). Hereby, a number of simulations are preformed to create PGV maps for different source parameters, which are then decomposed using the proper orthogonal decomposition (POD) technique, a type of surrogate modeling approach. The ROM is then created by fitting function approximators that predict the POD coefficients for different values of the input source parameters. The ROM can then be used to instantaneously predict PGV maps for new earthquakes that were not modeled.

In addition, the simulation results have implications for the seismic monitoring. Once the distribution of the ground motion is estimated, stations can be placed in the areas where the largest ground motion is expected in order to reliably record it. However, the small number of stations together with high noise levels within the city make microseismic monitoring in urban environments challenging. Therefore, a network optimization approach based on simulated annealing was applied to find the optimum number and distribution of seismic stations in the Munich area. The optimization program requires a high resolution noise map, in order to find suitable station locations even in environments with heterogeneous noise conditions. The results indicate that a surface station network cannot reach the required location precision and detection capability due to the city´s high seismic noise level. Instead, borehole stations must be added in order to increase the signal-to-noise ratio of the microearthquake recordings. The comparison of a 180m-deep borehole station in Munich with an overlying surface station showed that the noise level in the borehole is a factor of 10 lower than at the surface, which significantly improves the detectability of small earthquakes. However, the high construction costs and infrastructural limitations make the installation of borehole stations not always feasible. Therefore, alternative approaches such as small surface arrays could be tested to improve the SNR through stacking of the data.

The dissertation results have implications for future geothermal projects, as they provide solutions for existing challenges: 1) The new single-station 6C method allows a fast and easy site characterization in urban environments. 2) The information on the shallow velocity structure should then be used for the simulation of earthquake scenarios, which is an important part for the estimation of the seismic risk already in the planning stages of a geothermal project. 3) The seismic simulations allow a spatial interpretation of the ground motion, which can help public authorities to make fast and precise decisions after damaging events. 4) Through the application of the network optimization approach the seismic monitoring within urban environments can be improved in order to detect small magnitude events with sufficient accuracy, which is a necessary component of the risk governance strategy.

Bibliography

Afanasiev, M., Boehm, C., van Driel, M., Krischer, L., Rietmann, M., May, D. A., Knepley, M. G. & Fichtner, A. (2019), 'Modular and flexible spectral-element waveform modelling in two and three dimensions', *Geophysical Journal International* **216**(3), 1675–1692.

Agemar, T., Weber, J. & Schulz, R. (2014), 'Deep geothermal energy production in germany', *Energies* **7**(7), 4397–4416.

Aki, K. (1957), 'Space and time spectra of stationary stochastic waves, with special reference to microtremors', *Bull. Earthquake Res. Inst.* **35**.

Aki, K. (1976), 'Signal to noise ratio in seismic measurements', *Volcanoes and Tectonos· phere, Tokai Univ. Press, Tokyo* pp. 187–192.

Alder, C., Debayle, E., Bodin, T., Paul, A., Stehly, L., Pedersen, H. & Group, A. W. (2021), 'Evidence for radial anisotropy in the lower crust of the apennines from bayesian ambient noise tomography in europe', *Geophysical Journal International* **226**(2), 941–967.

Allen, N. F., Woods, R. D. & Richart Jr, F. E. (1980), 'Fluid wave propagation in saturated and nearly saturated sands', *Journal of the Geotechnical Engineering Division* **106**(3), 235–254.

Antuens, V., Toledo, T., Kraft, T., Reyes, C. & Megies, T. (2023), 'Optimal design and ground truth performance test for deep geothermal seismic monitoring networks. (in preparation)'.

Arai, H. & Tokimatsu, K. (2005), 'S-wave velocity profiling by joint inversion of microtremor dispersion curve and horizontal-to-vertical (h/v) spectrum', *Bulletin of the Seismological Society of America* **95**(5), 1766–1778.

Asten, M. (1984), 'Array estimators and use of microseisms for reconnaissance of sedimentary basins', *Exploration Geophysics* **15**.

Asten, M. W. & Henstridge, J. (1984), 'Array estimators and the use of microseisms for reconnaissance of sedimentary basins', *Geophysics* **49**(11), 1828–1837.

Bachmann, G. & Müller, M. (1992), 'Sedimentary and structural evolution of the german molasse basin', *Eclogae Geologicae Helvetiae* **85**(3), 519–530.

Bachmann, G., Müller, M. & Weggen, K. (1987), 'Evolution of the molasse basin (germany, switzerland)', *Tectonophysics* **137**(1-4), 77–92.

Baisch, S., Carbon, D., Dannwolf, U., Delacou, B., Devaux, M., Dunand, F., Jung, R., Koller, M., Martin, C., Sartori, M. et al. (2009), 'Deep heat mining basel: seismic risk analysis', *SERIANEX Group, Departement für Wirtschaft, Soziales und Umwelt des Kantons Basel-Stadt, Basel* .

Baisch, S., Fritschen, R., Groos, J., Kraft, T., Plenefisch, T., Plenkers, K., Ritter, J. R. & Wassermann, J. (2012), 'Empfehlungen zur überwachung induzierter seismizitätpositionspapier des fkpe', *DGG Mitteilungen* (3), 17–31.

Bard, P.-Y., Acerra, C., Aguacil, G., Anastasiadis, A., Atakan, K., Azzara, R., Basili, R., Bertrand, E., Bettig, B., Blarel, F., Bonnefoy-Claudet, S., Paola, B., Borges, A., Sørensen, M., Bourjot, L., Cadet, H., Cara, F., Caserta, A., Chatelain, J.-L. & Zacharopoulos, S. (2008), 'Guidelines for the implementation of the h/v spectral ratio technique on ambient vibrations measurements, processing and interpretation', *Bulletin of Earthquake Engineering* **6**, 1–2.

Bauer, M., Thuro, K., Scholz, M. & Neumann, P. (2006), 'The geology of munich (germany) and its significance for ground modelling in urban areas.', *Proceedings of the 10th IAEG International Congress* **Paper 454**.

Bayer. Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie (2010), *Bayerischer Geothermieatlas*, Bayer. Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie), München.

Bayerisches Landesamt für Umwelt (2012), 'Geothermische charakterisierung von karstkluft-aquiferen im großraum münchen.', *Endbericht* .

Beaty, K. S., Schmitt, D. R. & Sacchi, M. (2002), 'Simulated annealing inversion of multimode rayleigh wave dispersion curves for geological structure', *Geophysical Journal International* **151**(2), 622–631.

Bergamo, P., Dashwood, B., Uhlemann, S., Swift, R., Chambers, J. E., Gunn, D. A. & Donohue, S. (2016), 'Time-lapse monitoring of climate effects on earthworks using surface wavestime-lapse seismic monitoring with sw', *Geophysics* **81**(2), EN1–EN15.

Bergstra, J., Komer, B., Eliasmith, C., Yamins, D. & Cox, D. D. (2015), 'Hyperopt: a python library for model selection and hyperparameter optimization', *Computational Science & Discovery* **8**(1), 014008.

Bernauer, F., Wassermann, J., Guattari, F., Frenois, A., Bigueur, A., Gaillot, A., de Toldi, E., Ponceau, D., Schreiber, K. & Igel, H. (2018), 'Blueseis3a: Full characterization of a 3c broadband rotational seismometer', *Seismological Research Letters* **89**.

Bettig, B., Bard, P., Scherbaum, F., Riepl, J., Cotton, F., Cornou, C. & Hatzfeld, D. (2001), 'Analysis of dense array noise measurements using the modified spatial autocorrelation method (spac): application to the grenoble area', *Bollettino di Geofisica Teorica ed Applicata* **42**(3-4), 281–304.

Birner, J. (2013), Hydrogeologisches Modell des Malmaquifers im Süddeutschen Molassebecken, PhD thesis.

Bishop, C. M. (1994), 'Mixture density networks.', *Technical Report. Aston University, Birmingham* .

Bishop, C. M. et al. (1995), *Neural networks for pattern recognition*, Oxford university press.

Bondár, I., Myers, S. C., Engdahl, E. R. & Bergman, E. A. (2004), 'Epicentre accuracy based on seismic network criteria', *Geophysical Journal International* **156**(3), 483–496.

Boore, D. M. (2004), 'Ground motion in anchorage, alaska, from the 2002 denali fault earthquake: Site response and displacement pulses', *Bulletin of the Seismological Society of America* **94**(6B), S72–S84.

Borcherdt, R. D. (2012), Vs30–a site-characterization parameter for use in building codes, simplified earthquake resistant design, gmpes, and shakemaps, *in* 'The 15th World Conference on Earthquake Engineering'.

Bormann, P. & Wielandt, E. (2013), Seismic signals and noise, *in* 'New manual of seismological observatory practice 2 (NMSOP2)', Deutsches GeoForschungsZentrum GFZ, pp. 1–62.

Brocher, T. M. (2005), 'Empirical relations between elastic wavespeeds and density in the earth's crust', *Bulletin of the seismological Society of America* **95**(6), 2081–2092.

Brokešová, J., Málek, J. & Kolínský, P. (2012), 'Rotaphone, a mechanical seismic sensor system for field rotation rate measurements and its in situ calibration', *Journal of seismology* **16**(4), 603–621.

Brooks, S., Gelman, A., Jones, G. & Meng, X.-L. (2011), *Handbook of markov chain monte carlo*, CRC press.

Brune, J. N. (1970), 'Tectonic stress and the spectra of seismic shear waves from earthquakes', *Journal of geophysical research* **75**(26), 4997–5009.

Budach, I., Moeck, I., Lüschen, E. & Wolfgramm, M. (2018), 'Temporal evolution of fault systems in the upper jurassic of the central german molasse basin: case study unterhaching', *International Journal of Earth Sciences* **107**(2), 635–653.

Büttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G. & Soukup, T. (2004), 'The corine land cover 2000 project', *EARSeL eProceedings* **3**(3), 331–346.

Cao, R., Earp, S., de Ridder, S. A., Curtis, A. & Galetti, E. (2020), 'Near-real-time near-surface 3d seismic velocity and uncertainty models by wavefield gradiometry and neural network inversion of ambient seismic noise', *Geophysics* **85**(1), KS13–KS27.

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

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

Chen, X., Xia, J., Pang, J., Zhou, C. & Mi, B. (2022), 'Deep learning inversion of rayleigh-wave dispersion curves with geological constraints for near-surface investigations', *Geophysical Journal International* **231**(1), 1–14.

Cochard, A., Igel, H., Schuberth, B., Suryanto, W., Velikoseltsev, A., Schreiber, U., Wassermann, J., Scherbaum, F. & Vollmer, D. (2006), Rotational motions in seismology: theory, observation, simulation, *in* 'Earthquake source asymmetry, structural media and rotation effects', Springer, pp. 391–411.

Cox, B. R. & Teague, D. P. (2016), 'Layering ratios: a systematic approach to the inversion of surface wave data in the absence of a priori information', *Geophysical Journal International* **207**(1), 422–438.

Dussel, M., Lüschen, E., Schulz, R., Thomas, R., Wenderoth, F., Fritzer, T., Birner, J., Schneider, M., Wolfgramm, M., Bartels, J. et al. (2012), 'Geothermische charakterisierung von karstig-klüftigen aquiferen im großraum münchen: Endbericht; laufzeit des vorhabens: 01.05. 2008-31.12. 2011'.

Earp, S., Curtis, A., Zhang, X. & Hansteen, F. (2020), 'Probabilistic neural network tomography across grane field (north sea) from surface wave dispersion data', *Geophysical Journal International* **223**(3), 1741–1757.

Edwards, B., Fäh, D. & Giardini, D. (2011), 'Attenuation of seismic shear wave energy in switzerland', *Geophysical Journal International* **185**(2), 967–984.

Ellsworth, W. L., Giardini, D., Townend, J., Ge, S. & Shimamoto, T. (2019), 'Triggering of the pohang, korea, earthquake (m w 5.5) by enhanced geothermal system stimulation', *Seismological Research Letters* **90**(5), 1844–1858.

Eulenfeld, T. & Wegler, U. (2016), 'Measurement of intrinsic and scattering attenuation of shear waves in two sedimentary basins and comparison to crystalline sites in germany', *Geophysical Journal International* **205**(2), 744–757.

Evans, K. F., Zappone, A., Kraft, T., Deichmann, N. & Moia, F. (2012), ‘A survey of the induced seismic responses to fluid injection in geothermal and co2 reservoirs in europe’, *Geothermics* **41**, 30–54.

Fäh, D., Kind, F. & Giardini, D. (2001), ‘A theoretical investigation of average h/v ratios’, *Geophysical Journal International* **145**(2), 535–549.

Fäh, D., Stamm, G. & Havenith, H.-B. (2008), ‘Analysis of three-component ambient vibration array measurements’, *Geophysical Journal International* **172**(1), 199–213.

Ferreira, A. M. & Igel, H. (2009), ‘Rotational motions of seismic surface waves in a laterally heterogeneous earth’, *Bulletin of the Seismological Society of America* **99**(2B), 1429–1436.

Field, E. H., Johnson, P. A., Beresnev, I. A. & Zeng, Y. (1997), ‘Nonlinear ground-motion amplification by sediments during the 1994 northridge earthquake’, *Nature* **390**(6660), 599–602.

Foti, S., Comina, C., Boiero, D. & Socco, L. (2009), ‘Non-uniqueness in surface-wave inversion and consequences on seismic site response analyses’, *Soil dynamics and earthquake engineering* **29**(6), 982–993.

Foti, S., Hollender, F., Garofalo, F., Albarello, D., Asten, M., Bard, P.-Y., Comina, C., Cornou, C., Cox, B., Di Giulio, G. et al. (2018), ‘Guidelines for the good practice of surface wave analysis: a product of the interpacific project’, *Bulletin of Earthquake Engineering* **16**(6), 2367–2420.

Foti, S., Lai, C. G., Rix, G. J. & Strobbia, C. (2014), *Surface wave methods for near-surface site characterization*, CRC press.

Fowler, C. M. R. (1990), *The solid earth: an introduction to global geophysics*, Cambridge University Press.

Galetti, E., Curtis, A., Baptie, B., Jenkins, D. & Nicolson, H. (2017), ‘Transdimensional love-wave tomography of the british isles and shear-velocity structure of the east irish sea basin from ambient-noise interferometry’, *Geophysical Journal International* **208**(1), 36–58.

Gercek, H. (2007), ‘Poisson’s ratio values for rocks’, *International Journal of Rock Mechanics and Mining Sciences* **44**(1), 1–13.

Gosselin, J. M., Cassidy, J. F., Dosso, S. E. & Brillon, C. (2018), ‘Probabilistic seismic-hazard site assessment in kitimat, british columbia, from bayesian inversion of surface-wave dispersion’, *Canadian Geotechnical Journal* **55**(7), 928–940.

Gosselin, J. M., Dosso, S. E., Askan, A., Wathelet, M., Savvaidis, A. & Cassidy, J. F. (2022), 'A review of inverse methods in seismic site characterization', *Journal of Seismology* pp. 1–41.

Grad, M., Brückl, E., Majdański, M., Behm, M., Guterch, A., 2000, C. & Groups, A. . W. (2009), 'Crustal structure of the eastern alps and their foreland: seismic model beneath the cel10/alp04 profile and tectonic implications', *Geophysical Journal International* **177**(1), 279–295.

Groos, J. & Ritter, J. (2009), Time domain classification and quantification of seismic noise, *in* 'Noise and diffuse wavefields: extended abstracts of the Neustadt workshop; Neustadt an der Weinstraße, Germany, 5-8 July 2009. Ed.: Ch. Sens-Schönfelder', p. 25.

Grünthal, G. & Wahlström, R. (2003), 'An m w based earthquake catalogue for central, northern and northwestern europe using a hierarchy of magnitude conversions', *Journal of seismology* **7**, 507–531.

Grünthal, G., Wahlström, R. & Stromeyer, D. (2009), 'The unified catalogue of earthquakes in central, northern, and northwestern europe (cenec)—updated and expanded to the last millennium', *Journal of Seismology* **13**(4), 517–541.

Gutenberg, B. (1958), Microseisms, *in* 'Advances in Geophysics', Vol. 5, Elsevier, pp. 53–92.

Hallo, M., Imperatori, W., Panzera, F. & Fäh, D. (2021), 'Joint multizonal transdimensional bayesian inversion of surface wave dispersion and ellipticity curves for local near-surface imaging', *Geophysical Journal International* **226**(1), 627–659.

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

Häring, M. O., Schanz, U., Ladner, F. & Dyer, B. C. (2008), 'Characterisation of the basel 1 enhanced geothermal system', *Geothermics* **37**(5), 469–495.

Hartzell, S., Carver, D. & Williams, R. A. (2001), 'Site response, shallow shear-wave velocity, and damage in los gatos, california, from the 1989 loma prieta earthquake', *Bulletin of the Seismological Society of America* **91**(3), 468–478.

Hecht, C. & Pletl, C. (2015), 'Das verbundprojekt grame–wegweiser für eine geothermische wärmeversorgung urbaner ballungsräume', *Geothermische Energie* **82**(2), 02.

Hibert, C., Grandjean, G., Bitri, A., Travelletti, J. & Malet, J.-P. (2012), 'Characterizing landslides through geophysical data fusion: Example of the la valette landslide (france)', *Engineering Geology* **128**, 23–29.

Hirschberg, S., Wiemer, S. & Burgherr, P. (2014), *Energy from the Earth: Deep Geothermal as a Resource for the Future?*, Vol. 62, vdf Hochschulverlag AG.

Hobiger, M., Bard, P.-Y., Cornou, C. & Le Bihan, N. (2009), 'Single station determination of rayleigh wave ellipticity by using the random decrement technique (raydec)', *Geophysical Research Letters* **36**(14).

Hobiger, M., Bergamo, P., Imperatori, W., Panzera, F., Marrios Lontsi, A., Perron, V., Michel, C., Burjánek, J. & Fäh, D. (2021), 'Site characterization of swiss strong-motion stations: The benefit of advanced processing algorithms', *Bulletin of the Seismological Society of America* **111**(4), 1713–1739.

Hobiger, M., Cornou, C., Wathelet, M., Giulio, G. D., Knapmeyer-Endrun, B., Renalier, F., Bard, P.-Y., Savvaidis, A., Hailemikael, S., Le, B. N. et al. (2013), 'Ground structure imaging by inversions of rayleigh wave ellipticity: sensitivity analysis and application to european strong-motion sites', *Geophysical Journal International* **192**(1), 207–229.

Homuth, S. (2014), 'Aufschlussanalogstudie zur charakterisierung oberjurassischer geothermischer karbonatreservoire im molassebecken', *PhD Thesis, Technische Universität Darmstadt* .

Hu, J., Qiu, H., Zhang, H. & Ben-Zion, Y. (2020), 'Using deep learning to derive shear-wave velocity models from surface-wave dispersion data', *Seismological Research Letters* **91**(3), 1738–1751.

Ivanov, J., Schwenk, J. T., Miller, R. D. & Peterie, S. (2013), Dispersion-curve imaging nonuniqueness studies from multi-channel analysis of surface waves (masw) using synthetic seismic data, *in* '2013 SEG Annual Meeting', OnePetro.

iXblue (2017), 'blueseis-3a'.
URL: *www.blueseis.com, accessed: 01.10.2019*

Jongmans, D., Bièvre, G., Renalier, F., Schwartz, S., Beaurez, N. & Orengo, Y. (2009), 'Geophysical investigation of a large landslide in glaciolacustrine clays in the trièves area (french alps)', *Engineering geology* **109**(1-2), 45–56.

Karagoz, O., Chimoto, K., Citak, S., Ozel, O., Yamanaka, H. & Hatayama, K. (2015), 'Estimation of shallow s-wave velocity structure and site response characteristics by microtremor array measurements in tekirdag region, nw turkey', *Earth, Planets and Space* **67**(1), 1–17.

Käufl, P., P. Valentine, A., W. de Wit, R. & Trampert, J. (2016), 'Solving probabilistic inverse problems rapidly with prior samples', *Geophysical Journal International* **205**(3), 1710–1728.

Keil, S. & Wassermann, J. (2023), 'Surface wave dispersion curve inversion using mixture density networks', *Geophysical Journal International* p. ggad227.

Keil, S., Wassermann, J. & Igel, H. (2021), 'Single-station seismic microzonation using 6c measurements', *Journal of Seismology* **25**(1), 103–114.

Keil, S., Wassermann, J. & Megies, T. (2022b), ‘Estimation of ground motion due to induced seismicity at a geothermal power plant near munich, germany, using numerical simulations’, *Geothermics* **106**, 102577.

Keil, S., Wilczek, A., Wassermann, J. & Kremers, S. (2022a), ‘Comparing single-station 6c measurements and array measurements for seismic microzonation in munich, germany’, *Geophysical Journal International* **231**(3), 1634–1652.

Kijko, A. (1977), ‘An algorithm for the optimum distribution of a regional seismic network—i’, *pure and applied geophysics* **115**(4), 999–1009.

Kim, K.-H., Ree, J.-H., Kim, Y., Kim, S., Kang, S. Y. & Seo, W. (2018), ‘Assessing whether the 2017 m w 5.4 pohang earthquake in south korea was an induced event’, *Science* **360**(6392), 1007–1009.

Kirkpatrick, S., Gelatt Jr, C. D. & Vecchi, M. P. (1983), ‘Optimization by simulated annealing’, *science* **220**(4598), 671–680.

Köhler, A., Ohrnberger, M., Scherbaum, F., Wathelet, M. & Cornou, C. (2007), ‘Assessing the reliability of the modified three-component spatial autocorrelation technique’, *Geophysical Journal International* **168**(2), 779–796.

Kraft, T. (2014), ‘A high-resolution ambient seismic noise model for europe’, *EGU General Assembly, EGU2014-2282* **27**.

Kraft, T. (2016), ‘A high-resolution and calibrated model of man-made seismic noise for europe’, *Jahrestagung der Deutschen Geophysikalischen Gesellschaft* pp. S2–A.

Kraft, T., Mignan, A. & Giardini, D. (2013), ‘Optimization of a large-scale microseismic monitoring network in northern switzerland’, *Geophysical Journal International* **195**(1), 474–490.

Kraft, T., Roth, P. & Wiemer, S. (2020), Good practice guide for managing induced seismicity in deep geothermal energy projects in switzerland: Version 2, Technical report, ETH Zurich.

Kristeková, M., Kristek, J. & Moczo, P. (2009), ‘Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals’, *Geophysical Journal International* **178**(2), 813–825.

Kuhlemann, J. & Kempf, O. (2002), ‘Post-eocene evolution of the north alpine foreland basin and its response to alpine tectonics’, *Sedimentary Geology* **152**(1-2), 45–78.

Kurrle, D., Igel, H., Ferreira, A., Wassermann, J. & Schreiber, K. (2010), ‘Can we estimate local love wave dispersion properties from collocated amplitude measurements of translations and rotations?’, *Geophysical Research Letters* **37**.

Laarhoven, P. J. v. & Aarts, E. H. (1987), *Simulated annealing: Theory and applications*, Springer.

Lachet, C. & Bard, P.-Y. (1994), 'Numerical and theoretical investigations on the possibilities and limitations of nakamura's technique', *Journal of Physics of the Earth* **42**(5), 377–397.

Lecocq, T., Longuevergne, L., Pedersen, H. A., Brenguier, F. & Stammler, K. (2017), 'Monitoring ground water storage at mesoscale using seismic noise: 30 years of continuous observation and thermo-elastic and hydrological modeling', *Scientific reports* **7**(1), 1–16.

Lefevre, H. C. (2014), *The fiber-optic gyroscope*, Artech house.

Lemcke, K. (1988), *Das bayerische Alpenvorland vor der Eiszeit*, Schweizerbart'sche Verlagsbuchhandlung, Stuttgart.

Lentsch, D. & Schweingruber, M. (2022), 'First multilateral deep geothermal well in the south german molasse basin', *Eropean Geothermal Congress* .

Leu, W., Mégel, T. & Schärli, U. (2006), 'Geothermische eigenschaften der schweizer molasse tiefenbereich 0–500 m)—datenbank für wärmeleitfähigkeit, spezifsche wärmekapazität, gesteinsdichte und porosität.', *. Bericht Schweizer Bundesamt für Energie* .

Leydecker, G. (2011), 'Erdbebenkatalog für deutschland mit randgebieten für die jahre 800 bis 2008', *Geologisches Jahrbuch E 59.* .

Lin, C.-J., Huang, H.-P., Pham, N. D., Liu, C.-C., Chi, W.-C. & Lee, W. H. K. (2011), 'Rotational motions for teleseismic surface waves', *Geophysical Research Letters* **38**(15).

Lin, C.-P., Chang, C.-C. & Chang, T.-S. (2004), 'The use of masw method in the assessment of soil liquefaction potential', *Soil Dynamics and Earthquake Engineering* **24**(9-10), 689–698.

Lomax, A., Michelini, A., Curtis, A. & Meyers, R. (2009), 'Earthquake location, direct, global-search methods', *Encyclopedia of complexity and systems science* **5**, 2449–2473.

Lomax, A. & Snieder, R. (1994), 'Finding sets of acceptable solutions with a genetic algorithm with application to surface wave group dispersion in europe', *Geophysical Research Letters* **21**(24), 2617–2620.

Lowrie, W. (2007), 'Fundamentals of geophysics', *2nd, Cambridge UK* .

Lu, Y., Peng, S., Du, W., Zhang, X., Ma, Z. & Lin, P. (2016), 'Rayleigh wave inversion using heat-bath simulated annealing algorithm', *Journal of Applied Geophysics* **134**, 267–280.

Lund, J. W. & Toth, A. N. (2021), 'Direct utilization of geothermal energy 2020 worldwide review', *Geothermics* **90**, 101915.

Luo, Y., Huang, Y., Yang, Y., Zhao, K., Yang, X. & Xu, H. (2022), 'Constructing shear velocity models from surface wave dispersion curves using deep learning', *Journal of Applied Geophysics* **196**, 104524.

Lüschen, E., Wolfgramm, M., Fritzer, T., Dussel, M., Thomas, R. & Schulz, R. (2014), '3d seismic survey explores geothermal targets for reservoir characterization at unterhaching, munich, germany', *Geothermics* **50**, 167–179.

Malischewsky, P. & Scherbaum, F. (2004), 'Love's formula and h/v-ratio (ellipticity) of rayleigh waves', *Wave Motion* **40**, 57–67.

Marano, S. & Fäh, D. (2014), 'Processing of translational and rotational motions of surface waves: Performance analysis and applications to single sensor and to array measurements', *Geophysical Journal International* **196**, 317–339.

Megies, T., Kraft, T. & Reyes, C. (2023), 'pynetopt3d', *Zenodo* .
URL: *https://doi.org/10.5281/zenodo.7638856*

Megies, T. & Wassermann, J. (2014), 'Microseismicity observed at a non-pressure-stimulated geothermal power plant', *Geothermics* **52**, 36–49.

Megies, T. & Wassermann, J. (2016), 'Erste auswertung des erdbebens im raum pliening/poing vom 07.12.2016 mit auswertung des nachbebens am 20.12.2016', *Bayerischer Erdbebendienst* .

Megies, T. & Wassermann, J. (2017*a*), 'Erste auswertung des erdbebens im gemeindegebiet pliening vom 09.09.2017', *Bayerischer Erdbebendienst* .

Megies, T. & Wassermann, J. (2017*b*), 'Verbundprojekt mags2 - vom einzelsystem zur großräumigen nutzung- ep2: Untersuchungen zur optimierten seismischen Überwachung hydrogeothermaler systeme bei dichter räumlicher lage der bohrerlaubnisfelder am beispiel der situation im süden münchens', *Abschlussbericht* .

Meier, U., Curtis, A. & Trampert, J. (2007), 'Global crustal thickness from neural network inversion of surface wave data', *Geophysical Journal International* **169**(2), 706–722.

Moeck, I., Uhlig, S., Loske, B., Jentsch, A., Ferreiro Maehlmann, R. & Hild, S. (2015), Fossil multiphase normal faults-prime targets for geothermal drilling in the bavarian molasse basin, *in* 'Proceedings, World Geeothermal Congress'.

Molnar, S., Dosso, S. E. & Cassidy, J. F. (2010), 'Bayesian inversion of microtremor array dispersion data in southwestern british columbia', *Geophysical Journal International* **183**(2), 923–940.

Naik, S. P., Gwon, O., Porfido, S., Park, K., Jin, K., Kim, Y.-S. & Kyung, J.-B. (2020), 'Intensity reassessment of the 2017 pohang earthquake mw= 5.4 (south korea) using esi-07 scale', *Geosciences* **10**(11), 471.

Nakamura, Y. (1989), 'A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface', *Railway Technical Research Institute, Quarterly Reports* **30**(1).

Neuffer, T. & Kremers, S. (2017), 'How wind turbines affect the performance of seismic monitoring stations and networks', *Geophysical Journal International* **211**(3), 1319–1327.

Nogoshi, M. & Igarashi, T. (1971), 'On the amplitude characteristics of microtremor (part 2)', *Zisin (Journal of the Seismological Society of Japan. 2nd ser.)* **24**, 26–40.

Ohrnberger, M., Schissele, E., Cornou, C., Bonnefoy-Claudet, S., Wathelet, M., Savvaidis, A., Scherbaum, F. & Jongmans, D. (2004), Frequency wavenumber and spatial autocorrelation methods for dispersion curve determination from ambient vibration recordings, *in* 'Proceedings of the 13th World Conference on Earthquake Engineering', Vol. 946, Vancouver Canada.

Pan, L., Chen, X., Wang, J., Yang, Z. & Zhang, D. (2019), 'Sensitivity analysis of dispersion curves of rayleigh waves with fundamental and higher modes', *Geophysical Journal International* **216**(2), 1276–1303.

Parolai, S., Picozzi, M., Richwalski, S. & Milkereit, C. (2005), 'Joint inversion of phase velocity dispersion and h/v ratio curves from seismic noise recordings using a genetic algorithm, considering higher modes', *Geophysical research letters* **32**(1).

Parolai, S., Richwalski, S. M., Milkereit, C. & Fäh, D. (2006), 'S-wave velocity profiles for earthquake engineering purposes for the cologne area (germany)', *Bulletin of Earthquake Engineering* **4**, 65–94.

Pei, D., Louie, J. N. & Pullammanappallil, S. K. (2007), 'Application of simulated annealing inversion on high-frequency fundamental-mode rayleigh wave dispersion curves', *Geophysics* **72**(5), R77–R85.

Podvin, P. & Lecomte, I. (1991), 'Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools', *Geophysical Journal International* **105**(1), 271–284.

Poggi, V., Fäh, D., Burjanek, J. & Giardini, D. (2012), 'The use of rayleigh-wave ellipticity for site-specific hazard assessment and microzonation: application to the city of lucerne, switzerland', *Geophysical Journal International* **188**(3), 1154–1172.

Pumperla, M. (2022), 'hyperas: Keras+ hyperopt', *Astrophysics Source Code Library* pp. ascl–2205.

Reinecker, J., Tingay, M., Müller, B. & Heidbach, O. (2010), 'Present-day stress orientation in the molasse basin', *Tectonophysics* **482**(1-4), 129–138.

Rekoske, J. M., Gabriel, A.-A. & May, D. A. (2022), 'Instantaneous physics-based ground motion maps using reduced-order models', *arXiv preprint arXiv:2212.11335* .

Renalier, F., Jongmans, D., Savvaidis, A., Wathelet, M., Endrun, B. & Cornou, C. (2010), 'Influence of parameterization on inversion of surface wave dispersion curves and definition of an inversion strategy for sites with a strong vs contrast', *Geophysics* **75**(6), B197–B209.

Reynolds, J. M. (2011), *An introduction to applied and environmental geophysics*, John Wiley & Sons.

Riedl, C. (2017), A seismic noise map for the greater munich area, Master's thesis, LMU Munich.

Sambridge, M. (1999), 'Geophysical inversion with a neighbourhood algorithm—i. searching a parameter space', *Geophysical journal international* **138**(2), 479–494.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J. & Müller, K.-R. (2021), 'Explaining deep neural networks and beyond: A review of methods and applications', *Proceedings of the IEEE* **109**(3), 247–278.

Scherbaum, F., Hinzen, K.-G. & Ohrnberger, M. (2003), 'Determination of shallow shear wave velocity profiles in the cologne, germany area using ambient vibrations', *Geophysical Journal International* **152**(3), 597–612.

Schmittbuhl, J., Lambotte, S., Lengliné, O., Grunberg, M., Jund, H., Vergne, J., Cornet, F., Doubre, C. & Masson, F. (2021), 'Induced and triggered seismicity below the city of strasbourg, france from november 2019 to january 2021', *Comptes Rendus. Géoscience* **353**(S1), 561–584.

Seithel, R., Gaucher, E., Mueller, B., Steiner, U. & Kohl, T. (2019), 'Probability of fault reactivation in the bavarian molasse basin', *Geothermics* **82**, 81–90.

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

Singh, S., Capdeville, Y. & Igel, H. (2020), 'Correcting wavefield gradients for the effects of local small-scale heterogeneities', *Geophysical Journal International* **220**(2), 996–1011.

Stein, S. & Wysession, M. (2009), *An introduction to seismology, earthquakes, and earth structure*, John Wiley & Sons.

Steinberg, D. M. & Rabinowitz, N. (2003), 'Optimal seismic monitoring for event location with application to on site inspection of the comprehensive nuclear test ban treaty', *Metrika* **58**(1), 31–57.

Sylvette, B.-C., Cécile, C., Pierre-Yves, B., Fabrice, C., Peter, M., Jozef, K. & Fäh, D. (2006), 'H/V ratio: a tool for site effects evaluation. Results from 1-D noise simulations', *Geophysical Journal International* **167**(2), 827–837.

Tarantola, A. (2005), *Inverse problem theory and methods for model parameter estimation*, SIAM.

Technical-University-Munich (2016-2019), 'Geopot project', *funded by the Bavarian State Ministry of Environment and Public Health (StMUV)* .

Tokimatsu, K. (1997), 'Geotechnical site characterization using surface waves', *in 1st International Conference on Earthquake geotechnical engineering* **3**, 1333–1368.

Trnkoczy, A., Havskov, J. & Ottemöller, L. (2009), Seismic networks, *in* 'New Manual of Seismological Observatory Practice (NMSOP)', Deutsches GeoForschungsZentrum GFZ, pp. 1–60.

Valle-Molina, C. & Stokoe, K. H. (2012), 'Seismic measurements in sand specimens with varying degrees of saturation using piezoelectric transducers', *Canadian geotechnical journal* **49**(6), 671–685.

Wassermann, J., Wietek, A., Hadziioannou, C. & Igel, H. (2016), 'Toward a single-station approach for microzonation: Using vertical rotation rate to estimate love-wave dispersion curves and direction finding', *Bulletin of the Seismological Society of America* **106**(3), 1316–1330.

Wathelet, M. (2008), 'An improved neighborhood algorithm: parameter conditions and dynamic scaling', *Geophysical Research Letters* **35**(9).

Wathelet, M., Chatelain, J.-L., Cornou, C., Giulio, G. D., Guillier, B., Ohrnberger, M. & Savvaidis, A. (2020), 'Geopsy: A user-friendly open-source tool set for ambient vibration processing', *Seismological Research Letters* **91**(3), 1878–1889.

Wathelet, M., Guillier, B., Roux, P., Cornou, C. & Ohrnberger, M. (2018), 'Rayleigh wave three-component beamforming: signed ellipticity assessment from high-resolution frequency-wavenumber processing of ambient vibration arrays', *Geophysical Journal International* **215**(1), 507–523.

Wathelet, M., Jongmans, D. & Ohrnberger, M. (2004), 'Surface-wave inversion using a direct search algorithm and its application to ambient vibration measurements', *Near surface geophysics* **2**(4), 211–221.

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

Wawerzinek, B., Buness, H., von Hartmann, H. & Tanner, D. C. (2021), 'S-wave experiments for the exploration of a deep geothermal carbonate reservoir in the german molasse basin', *Geothermal Energy* **9**(1), 1–21.

Wood, C. M., Deschenes, M., Ledezma, C., Meneses, J., Montalva, G. & Morales-Velez, A. C. (2019), 'Dynamic site characterization of areas affected by the 2017 puebla-mexico city earthquake', *Soil Dynamics and Earthquake Engineering* **125**, 105704.

Woods, J. W. & Lintz, P. R. (1973), 'Plane waves at small arrays', *Geophysics* **38**(6), 1023–1041.

Xia, J., Miller, R. D., Park, C. B. & Tian, G. (2003), 'Inversion of high frequency surface waves with fundamental and higher modes', *Journal of Applied Geophysics* **52**(1), 45–57.

Yablokov, A. V., Serdyukov, A. S., Loginov, G. N. & Baranov, V. D. (2021), 'An artificial neural network approach for the inversion of surface wave dispersion curves', *Geophysical Prospecting* **69**(7), 1405–1432.

Yamanaka, H. & Ishida, H. (1996), 'Application of genetic algorithms to an inversion of surface-wave dispersion data', *Bulletin of the Seismological Society of America* **86**(2), 436–444.

Supplementary Material

Input file for the seismic simulation of the event on 09.09.2017 in Chapter 3

(applicable for SALVUS version 0.11.45)

[1]:

```
from scipy import signal
import numpy as np
import xarray as xr
import salvus.namespace as sn
import salvus.flow.simple_config as sc
from salvus.project.configuration.model.utils import derive_bm_file
```

[2]:

```
#limits for dataset
xmin=4477641.764412108
xmax=4491591.764412108
ymin=5332138.290354518
ymax=5346088.290354518

x=np.linspace(xmin,xmax,280)
y=np.linspace(ymin,ymax,280)
z=np.linspace(515,-5000,1104)
```

[3]:

```
#load velocity models
vp = np.load("vp_model.npy")
vs = np.load("vs_model.npy")
rho = np.load("rho_model.npy")

#load attenuation models
qkappa = np.load("qk_model.npy")
qmu = np.load("qm_model.npy")
```

[4]:
```
# Assemble into a dataset
ds_model = xr.Dataset(
    data_vars={
        "vp": (["x", "y", "z"], vp),
        "vs": (["x", "y", "z"], vs),
        "rho": (["x", "y", "z"], rho),
        "qkappa": (["x", "y", "z"], qkappa),
        "qmu": (["x", "y", "z"], qmu),
    },
    coords={"x": x, "y": y, "z": z},
)
```

[5]:
```
#extend of model domain
xmin_domain=4480604
xmax_domain=4488604
ymin_domain=5335113
ymax_domain=5343113
```

[6]:
```
# create a new project.
p = sn.Project.from_domain(
    path="Simulation",
    domain=sn.domain.dim3.BoxDomain(x0=xmin_domain, x1=xmax_domain,␣
 ↪y0=ymin_domain, y1=ymax_domain,z0=min(z), z1=max(z)),
)
```

[7]:
```
#add volume model to project
p += sn.model.volume.cartesian.GenericModel(
    name="3d_model", data=ds_model,mode="linear",
)
```

[8]:
```
#source (71,71,-22)  Mw 2.14
xx=0.74 * 1.3863e12
yy=-0.44 * 1.3863e12
zz=-0.30 * 1.3863e12
yz=0.50 * 1.3863e12
xz=0.25 * 1.3863e12
xy=-1.00 * 1.3863e12

#use moment tensor as source
src = sn.simple_config.source.cartesian.MomentTensorPoint3D(
    x=4484358, y=5339322, z=-2700.0,
    mxx=xx, myy=yy, mzz=zz,
```

```
    myz=yz, mxz=xz, mxy=xy)


# Receivers
recs=[]


recs.append(sn.simple_config.receiver.cartesian.Point3D(
        x=4485932.046, y=5339370.658, z=max(z),
        station_code="POI01", fields=["velocity"]
    ))


recs.append(sn.simple_config.receiver.cartesian.Point3D(
        x=4485962.270, y=5341968.086, z=max(z),
        station_code="POI03", fields=["velocity"]
    ))

recs.append(sn.simple_config.receiver.cartesian.Point3D(
        x=4484737.401, y=5338643.533, z=max(z),
        station_code="Th1", fields=["velocity"]
    ))

p += sn.Event(event_name="event_09092017", sources=src, receivers=recs)
```

[9]:

```
#source-time-function
dt = 0.002
tdelay = 0.5
tsteps = 1000
sousig = np.zeros(tsteps)
width = 29.6

for n in range(tsteps):
    t = n * dt
    if (n < int(tdelay/dt)):
        window_taper = (1+np.cos((2*np.pi/tdelay*dt/2)*((n-1)-tdelay/
 ↪dt)))/2
        sousig[n] = np.sin(np.pi*(t-tdelay)*width)/(np.
 ↪pi*(t-tdelay)*width)*window_taper
    elif (n == int(tdelay/dt)):
        sousig[n] = 1.
    elif (n >= int(tdelay/dt)+1 and n <= int(tdelay/dt*2)+1):
```

```
        window_taper = (1+np.cos((2*np.pi/tdelay*dt/2)*((n-1)-tdelay/
 ↪dt)))/2
        sousig[n] = np.sin(np.pi*(t-tdelay)*width)/(np.
 ↪pi*(t-tdelay)*width)*window_taper
    elif (n > int(2*tdelay/dt)+1):
        sousig[n] = 0

f, spec = signal.welch(sousig, 1/dt, nperseg=512)
```

[10]:

```
ec = sn.EventConfiguration(
    wavelet=sc.stf.Custom.from_array(sousig,sampling_rate_in_hertz=391,
    start_time_in_seconds=0.0),
    waveform_simulation_configuration=sn.WaveformSimulationConfiguration(
        end_time_in_seconds=5.5,
        attenuation=True
    ),
)
```

[11]:

```
#absorbing boundaries
abp = sn.AbsorbingBoundaryParameters(
    reference_velocity=1500.0,
    number_of_wavelengths=0.5,
    reference_frequency=6.0,
)
```

[12]:

```
#model configuration
mc = sn.ModelConfiguration(
        background_model=sn.model.background.one_dimensional.
 ↪FromBm(filename="background_model.bm"),
        volume_models="3d_model",
        linear_solids=sn.LinearSolids(reference_frequency=1.0),
    )
```

[13]:

```
#simulation configuration
p.add_to_project(
    sn.SimulationConfiguration(
        name="sim_event_09092017",
        tensor_order=1,
        max_frequency_in_hertz=10.0,
        elements_per_wavelength=1.5,
        model_configuration=mc,
        event_configuration=ec,
```

```
        absorbing_boundaries=abp,
    )
)
```

[14]:

```
#start simulation
p.simulations.launch(
    ranks_per_job=150,
    site_name=os.environ.get("SITE_NAME", "site_name"),
    events="event_09092017",
    simulation_configuration="sim_event_09092017",
    extra_output_configuration={
        "surface_data": {
            "sampling_interval_in_time_steps": 15,
            "fields": ["velocity"],
            "side_sets": ["z1"],
        },
    }
)
```

Acknowledgements

An dieser Stelle möchte ich mich bei allen bedanken, die mich bei der Anfertigung dieser Doktorarbeit unterstützt haben.

Mein besonderer Dank gilt Joachim Wassermann für die ausgezeichnete Betreuung und Unterstützung während der gesamten Arbeit. Durch Ihn wurde erst mein Interesse an der Geophysik und letztendlich dem Thema der Dissertation geweckt. Er stand mir stets bei Fragen zur Seite und gab mir zugleich die notwendige Freiheit, um an neuen Aufgaben zu wachsen. Vor allem möchte ich mich für die kleine Auszeit auf Stromboli bedanken.

Des Weiteren bedanke ich mich bei Heiner Igel für die vielen wertvollen Diskussionen und motivierenden Worte. Ich durfte während meiner Zeit an der LMU viel von Ihm lernen. Außerdem bedanke ich mich für seine Unterstützung bei der Erstellung der vielen Video-Präsentationen. Ich bin dankbar für die Teilnahme an diversen Workshops und Konferenzen, die mir Heiner und Jo ermöglicht haben.

Ebenfalls bedanke ich mich bei Eva Eibl für das Interesse an meiner Arbeit und für die Bereitstellung als Zweitgutachterin.

Ich bedanke mich beim gesamten SEIGER-Team für die freundliche Zusammenarbeit innerhalb des Projekts und die zahlreichen spannenden Projekttreffen. Ohne die Förderung des SEIGER Projekts durch das Bundesministerium für Wirtschaft und Klima wäre die Durchführung dieser Doktorarbeit nicht möglich gewesen.

Außerdem möchte ich mich beim gesamten Observatoriums-Team in Fürstenfeldbruck für die positive Arbeitsatmosphäre und die stetige Unterstützung bei jeglichen Feldeinsätzen bedanken. Hierfür bedanke ich mich insbesondere bei Jürgen, Sven, Marcos und Renate.

Danke an Andreas Brotzer für die vielen tollen Gespräche und Felix Bernauer für die Hilfe bei jeglichen blueSeis-Problemen. Vielen Dank an Tobias Megies, der mir immer bei Fragen zur Seite stand.

Insbesondere gehört mein Dank Aniko Wirp, die über die Jahre das Büro mit mir teilte und immer ein offenes Ohr hatte.

Nicht zuletzt muss ich mich bei Berta Vilacis und Nicolas Hayek bedanken, die mich bereits während des Master Studiums begleitet haben und mir immer mit lieben Worten zur Seite standen.

Natürlich möchte ich auch meinen Eltern und Geschwistern danken, die mich Zuhause immer mit offenen Armen empfangen und meinen Lebensweg nie angezweifelt haben.