Inversion of surface wave dispersion and reflectivity imaging using ambient noise seismic interferometry

Master of Science Thesis

Dirk Meine van Haeringen, Department of Earth Sciences, Universiteit Utrecht
Supervised by Arie Verdel (TNO), Stefan Carpentier (TNO), Hanneke Paulssen (UU)

Abstract

Seismic exploration typically measures signals originating from a carefully constructed seismic source. If the same results can be accomplished without requiring a man-made source, but rather relying on randomly distributed ambient noise signals, significant cost reductions are possible.

The main goal of this research is to efficiently retrieve near-surface shear wave velocity profiles from ambient seismic noise. Shear wave velocities allow for computation of rock properties that are key factors in e.g. building-construction as well as in mining, drilling and reservoir production activities.

A line of seismic receivers was set up near 'De Grote Peel' national park in the Netherlands, which recorded ambient seismic noise over a period of 35 days. Surface wave signals are extracted from passive seismic data through the process of 'ambient noise seismic interferometry' (ANSI). This involves cross-correlating the signal responses recorded with a line of receivers in order to make virtual source gathers.
Surface wave components with different wavelengths will travel at different speeds and disperse. Dispersion curves are generated from the surface waves and inverted using a neighborhood algorithm. The resulting shear wave velocity profiles are successfully linked to formation transition depths from borehole data.

As a side-project, zero-offset autocorrelations are used to image deep reflecting layers. The resulting reflectors are successfully linked to vintage active seismic data.

This research has been successful in finding efficient methods for imaging both deep reflecting layers and near-surface shear wave velocity profiles, solely from ambient seismic noise recordings.

Contents

Page:

1 Introduction

Seismic exploration has been a much exploited method since the early 20th century. The usual process includes an active seismic source which creates a signal that can be observed with seismic receiver stations. The time-delay of the signal for different receivers can be used to image the subsurface. The need for an active source, be it a large hammer, a seismic vibrator or an explosive charge, complicates exploration surveys and makes up a large part of their financial budget. A rather new method relies on seismic background noise rather than an active source to image the subsurface, by the process of ambient noise seismic interferometry (ANSI). The ANSI theory was first described by Claerbout (1968) and has been commercially exploited since 2006. The implementation of this method (without requiring active sources) allows for a great reduction of exploration costs.

This research aims to extract surface waves from background noise recorded in the Dutch province of Noord-Brabant, and measure their dispersion. Using an inversion, shear-wave velocity profiles for the shallow subsurface are produced (up to a depth of 200m). Shear-wave velocities are a key factor in calculating subsurface properties such as Poisson's ratio and rigidity. Furthermore, near-surface velocity information greatly improves imaging of the deep subsurface using exploration surveys. The focus of this research lies on finding an efficient and effective method to ultimately extract near surface shear-wave velocity profiles from ambient seismic noise. As little to no similar studies have been performed in the survey area and no near-surface velocity information is available, the results are linked to nearby shallow borehole studies.

Deep body wave reflections are also briefly investigated, and correlated with existing active seismic survey results.

2 Background

The area to be studied was chosen to be a 750m line just west of the national park ‘De Grote Peel.’ Figure 1 shows a satellite image of the area in question, which is subject to possible future geothermal exploration. The location fits a number of low-risk criteria; including a simple horizontally layered geology and minimal large regional faults. Figure 2 shows vintage seismic data that has been shot nearby the study area.

Figure 1: Satellite image of the survey area, with the 750 receiver line colored yellow. The equipment was controlled from the farm slightly west of the line. Points A and B represent the locations of two borehole surveys. From Google Maps (2017).

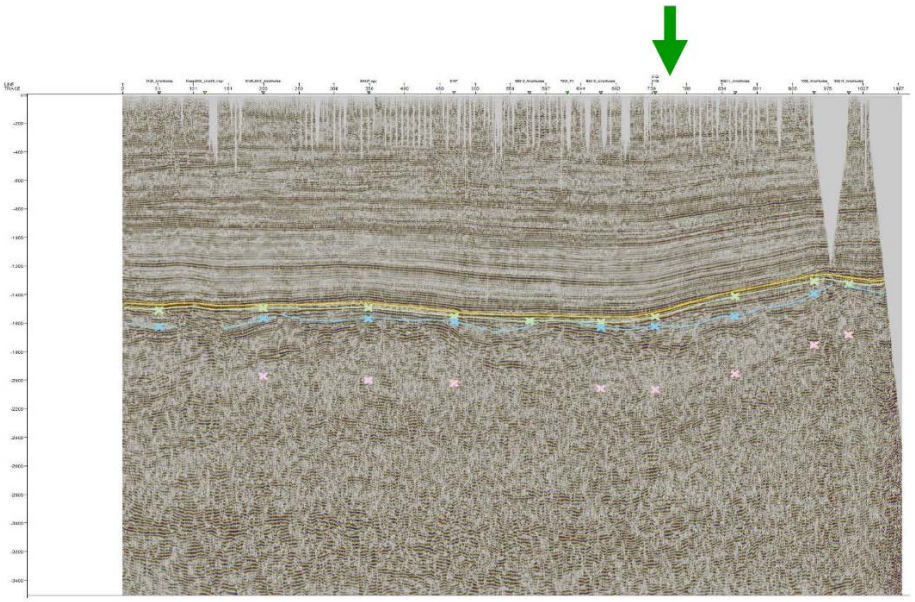


Figure 2: Vintage active seismic section, shot parallel to the passive survey at a distance of ±500m. The green arrow roughly indicates the location of the passive survey. From BP Nederland b.v. (1982).

A small number of geological borehole surveys have been performed in the vicinity of the line. Their locations relative to the studied line are shown in Figure 1, with their results shown in Figure 3.

There is a distinct transition of the formation of Sterksel (ST; coarse river sands) into the formation of Stramproy (SY; fine sands and clays) at a depth of around 55m.

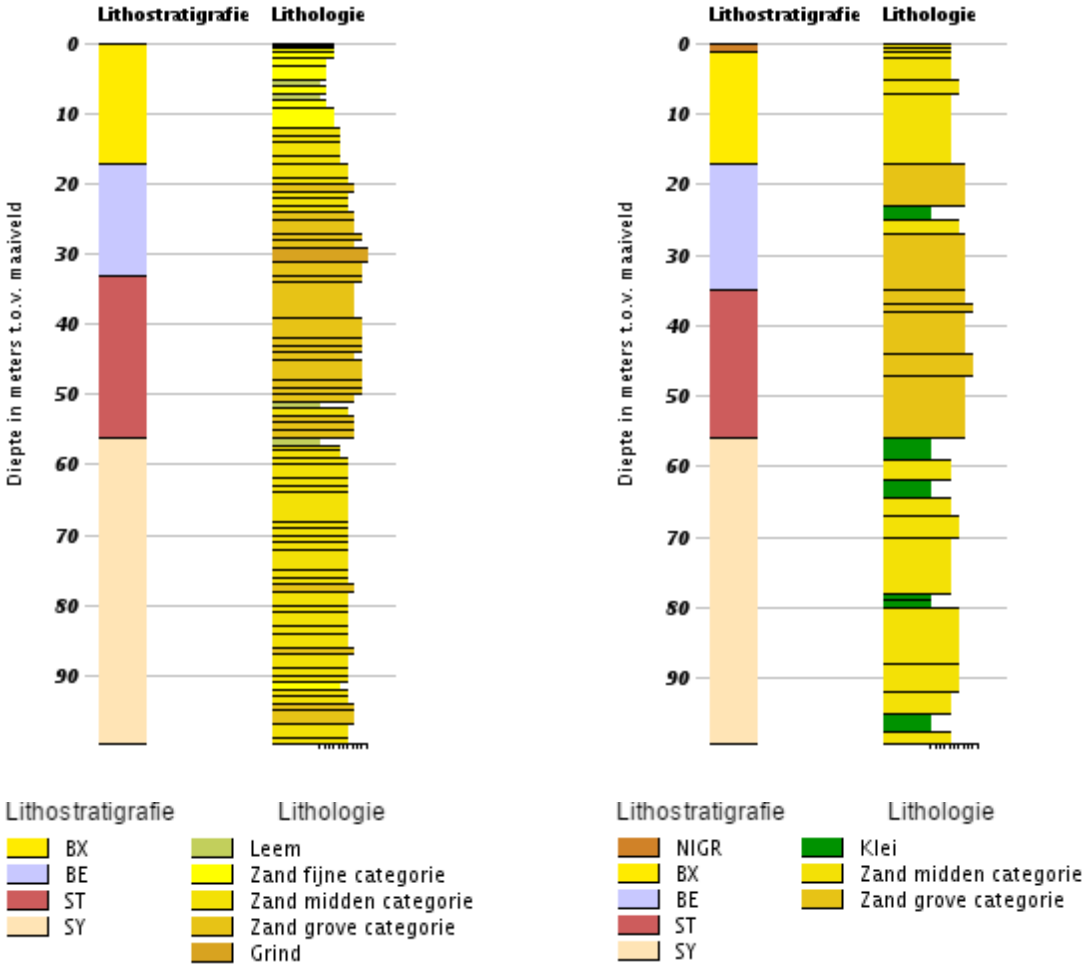


Figure 3: Borehole survey results for the locations A (left) and B (right) in Figure 1. ST and SY represent the formations of Sterksel and Stramproy, respectively, with a transition depth of ±55m. From www.dinoloket.nl/ondergrondgegevens

3 Methods

Many processing steps are required when attempting to extract useful information from ambient seismic noise. The data processing flow that is used in this research is shown in Figure 4, where the numbered processing steps refer to their respective subchapters in the methods section.

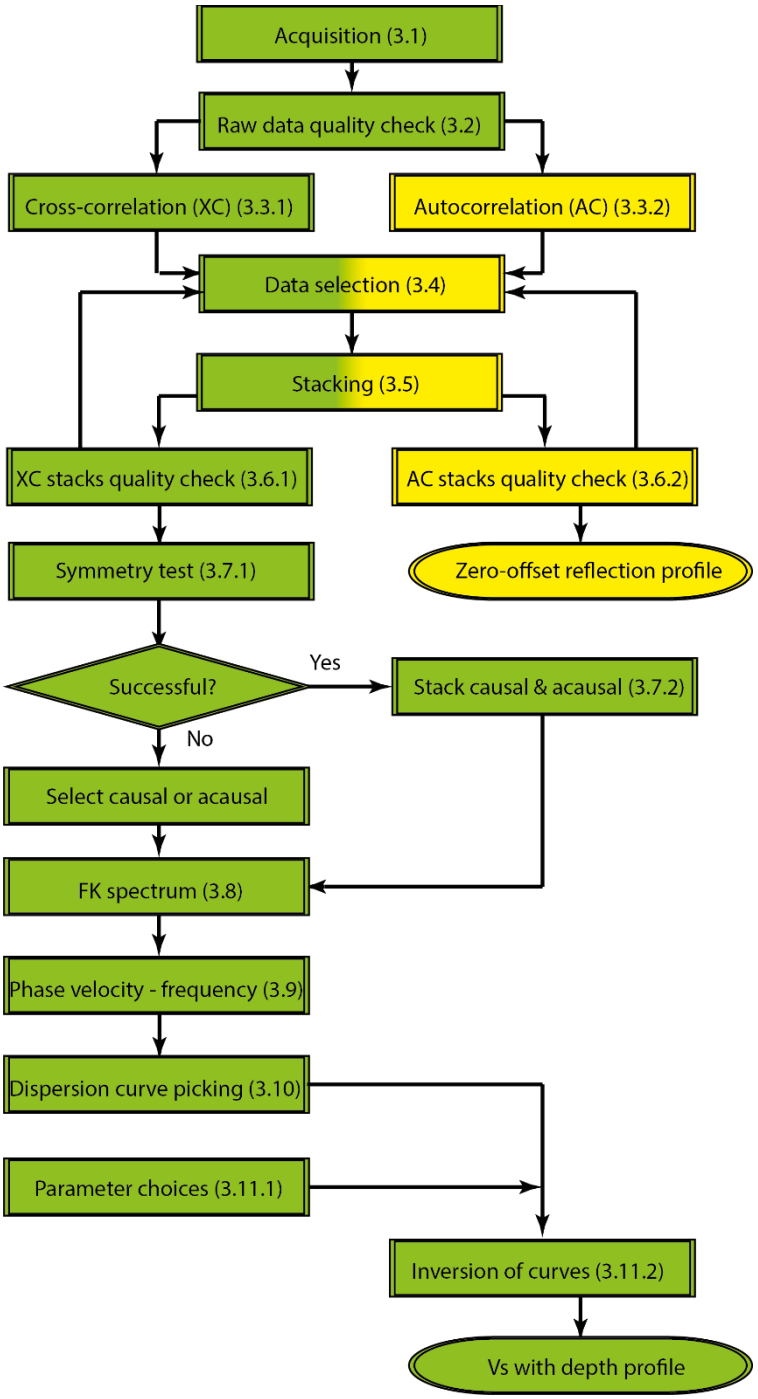


Figure 4: Data processing flow that is followed in this research. The numbers used refer to the subchapters in the methods section. Green represents the cross-correlation processing flow for surface wave dispersion analysis. Yellow represents the autocorrelation processing flow which focusses on reflecting body waves.

3.1 Data acquisition

Seismic background noise was measured over a total period of 40 days using 31 seismic receiver stations (numbered 3-33). The stations were placed along a strip of farmland directly next to national park “De Grote Peel” with a spacing of 25m, totaling to a line with a length of 750m. Each station consisted of a DSU-3 three-component accelerometer that recorded with a sampling rate of 2 ms, over a bandwidth of 0 to 100 Hz. Each DSU-3 unit was attached to a battery powered RAU-D unit, with GPS and WIFI capabilities, that stored and transmitted the recorded data. Only the recorded vertical component of each receiver station is used in this research. The final output of this setup was a dataset of around 55.000 ‘ambient noise panels’ of 31 traces wide and one minute long. An example of one of these panels is shown in Figure 5.

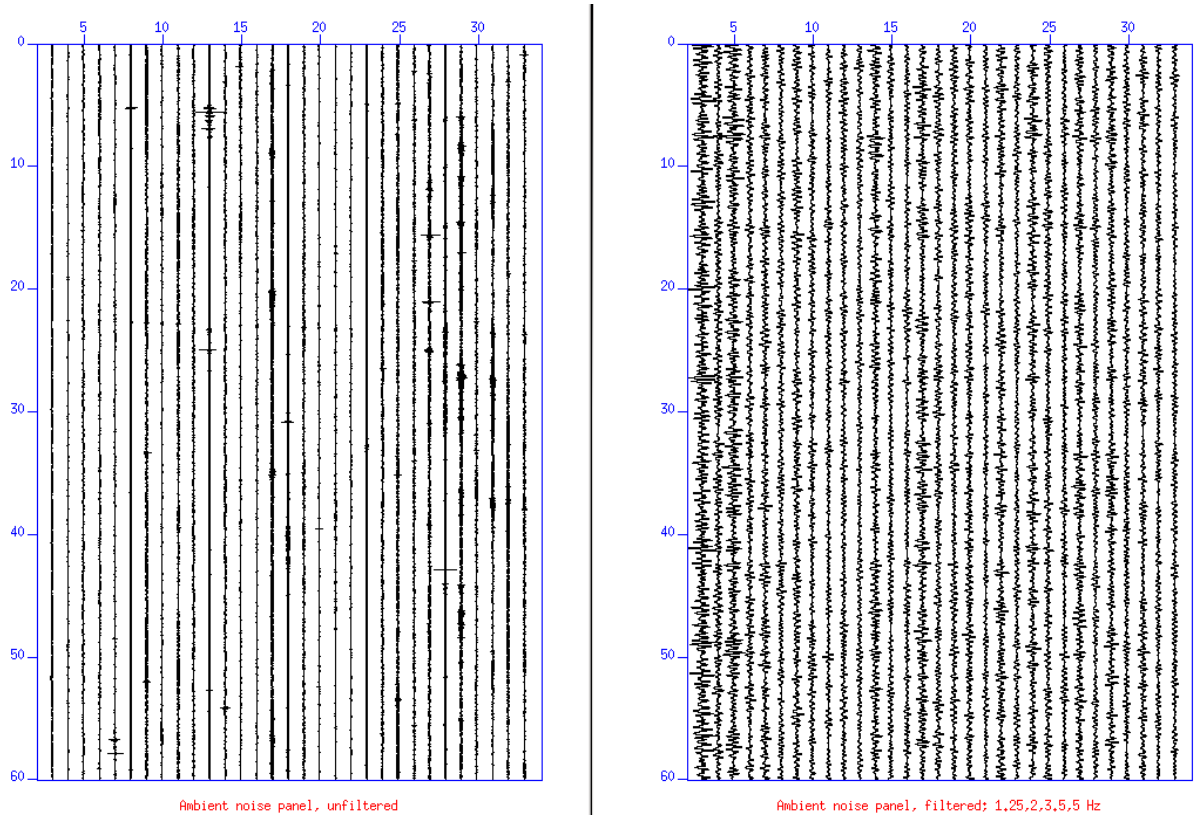


Figure 5: Typical recorded raw ambient noise panel. It consists of 31 traces with a spacing of 25 meters, with 60 seconds of recorded noise. The panel on the right shows the same data, but has been bandpass filtered between 2 Hz and 3.5 Hz with filter slopes ending at 1.25 Hz and 5 Hz, respectively. This frequency band is chosen in order to best show surface waves.

3.2 Raw data quality check

In the case of the ‘De Peel’ acquisition, occasionally the desired noise data is completely overshadowed by strong repetitive signals from a nearby drainage pump. In order to not ‘ruin’ the rest of the data, noise panels that have recorded these vibrations must be excluded from further processing. Therefore, before cross-correlation, each noise panel with a maximum amplitude above a certain threshold is discarded.

3.3.1 Cross-correlation

The recorded ambient noise panels show the response of 31 seismic stations to noise sources that are ideally randomly distributed in the subsurface. Through the process of seismic interferometry the differences in response between two separate stations can be compared and normalized. One of the two stations will then act as a virtual source for the other station. When assuming stationary, spatially uncorrelated noise sources that illuminate two receiver stations (x_A and x_B) from all directions, this process can applied to the ambient noise panels with the relation:

$$\{G(x_B, x_A, t) + G(x_B, x_A, -t)\} * S_N(t) \approx \langle u(x_B, t) * u(x_A, -t)\rangle \qquad \text{(Wapenaar et al. 2010)}$$

Where $\{G(x_B, x_A, t) + G(X_B, x_A, -t)\}$ represents the causal and time-reversed Green's tensors and $S_N(t)$ represents the autocorrelation of the source time function of noise sources. $\langle u(x_B, t) * u(x_A, -t)\rangle$ represents the averaged ensemble of the vertical particle velocity components. A visual representation is shown in Figure 6.

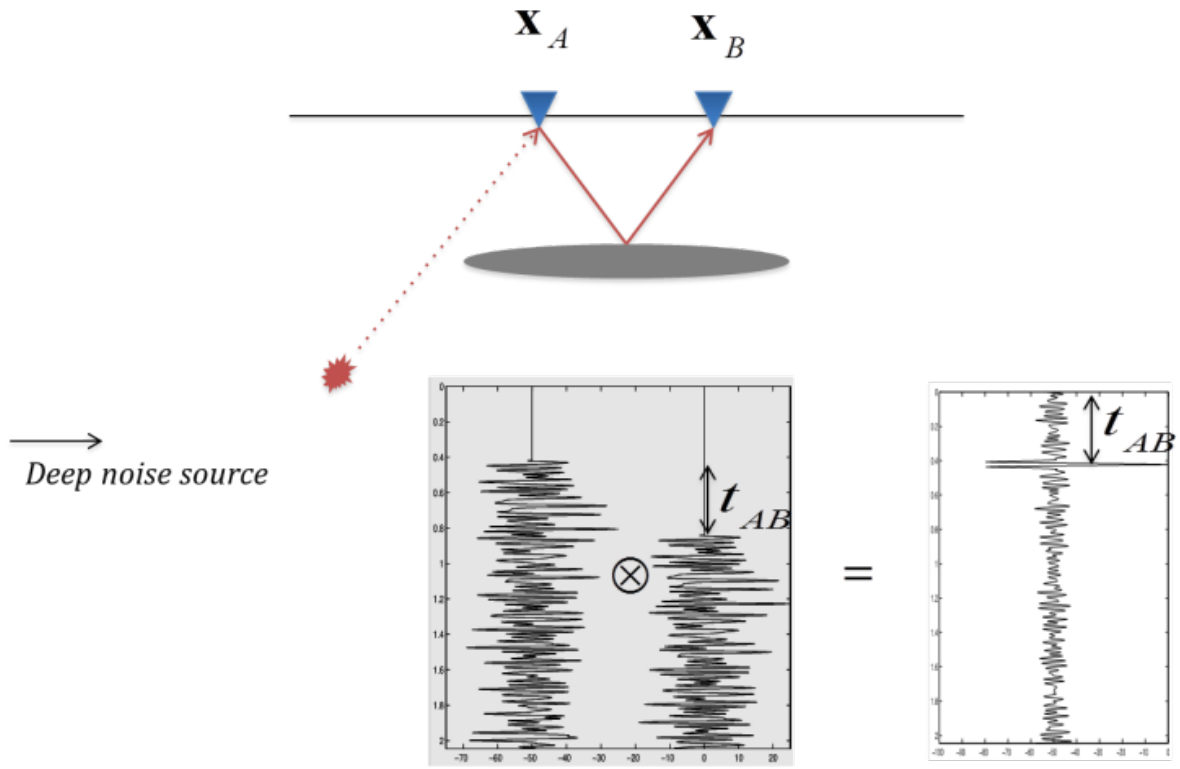
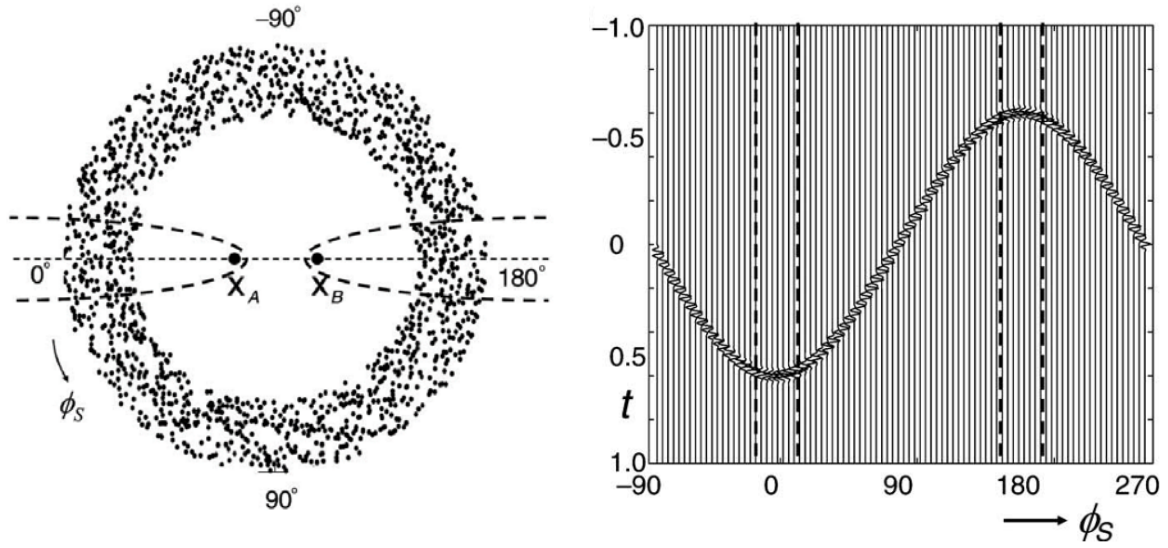


Figure 6: Visual representation of ambient noise seismic interferometry (ANSI). The cross-correlation of two recorded signals from separate receivers results in the response of one receiver to a virtual source at the other. From Wapenaar (2004).

With the receiver stations in a straight line, and noise sources from all directions, the main noise source contributions to a signal after cross-correlation are located in the zones in the extension of the line; 'Fresnel zones' (Wapenaar et al. 2010). Figures 7.1 shows two receivers with randomly distributed noise sources. The contributions of each noise source to the responses of the receivers as a function of its polar coordinates is shown in Figures 7.2. As a result, the contribution of the noise sources to the cross-correlation is greatest in the so-called Fresnel zones.



Figures 7.1 - 7.2: The image on the left shows two receivers (X_A and X_B) with randomly distributed noise sources. The image on the right shows the contributions of noise sources to the cross-correlation of the two receivers, as a function of the sources' polar coordinates. The largest contributions come from the areas enclosed by the striped lines in both images; the so-called Fresnel zones. From Wapenaar et al., 2010

The raw noise panels that have been acquired consist of 31 traces, each containing 60 seconds of noise data. One of these 31 traces is chosen as a 'master trace' which is then cross-correlated with all the traces in the panel. The result is a 120 seconds long panel of 31 traces with a virtual source on the 'master' trace at t=0. Because the source is virtual, signals can be observed propagating in both the positive and negative time; the 'causal' and 'acausal' parts of a virtual source panel, respectively.

For each panel of raw data, each of the 31 traces is sequentially chosen as the 'master' trace and cross-correlated. The result is 31 cross-correlated panels, all with a virtual source location on a different trace. When correlating 60 seconds long raw noise panels, one day of raw data will require 24 * 60 * 31 * 31 = 1.383.840 cross-correlations. In order to save computational time, the cross-correlation is done as a multiplication in the frequency domain; The 'master' trace is flipped in time (-t) and Fourier transformed, and multiplied with each trace of the Fourier transformed raw noise panel. The inverse Fourier transform of this product gives the same result as a cross-correlation in the time domain, but significantly faster (Cooley & Tukey, 1965)

3.3.2 Autocorrelation

An autocorrelation is a special case of cross-correlation, where a signal is cross-correlated with itself (zero-offset). During the cross-correlation process (see 3.3.1), each trace is sequentially chosen as the 'master' trace and cross-correlated with every trace in a panel, including that same 'master' trace. In a cross-correlated virtual source panel the virtual source trace is therefore the autocorrelation of that trace. From each of the 31 cross-correlated virtual source panels the virtual source traces (autocorrelations) are taken and added together into an autocorrelated panel consisting of 31 traces.

As no signals recorded by different receivers are cross-correlated with each other, each trace of an auto-correlation panel will show only the signals that pass a single receiver station more than once. Therefore, autocorrelation panels are very useful for inspecting reflections of underground layers.

3.4.1 Data selection – Pre-stack

Data obtained from cross-correlating 60 seconds of noise will contain a limited amount of useful information. However, when summing several cross-correlated panels (with the same virtual source) together their signal will be amplified ('stacking', see 3.5).

Choices must be made when selecting the data that is to be used for further processing. 'Good' data will amplify the desired signal, whereas 'bad' data will suppress it. Each day of recorded data leads to over 40.000 cross-correlated panels, making a manual selection impractical. Therefore a small range of simple selections is made, for both the cross-correlations and autocorrelations; All data together (Dall), separate days, and separate hours.

3.4.2 Data selection – Post-stack

It is observed (see 3.6.1) that adding more cross-correlated data to a stack does not necessarily improve its quality. Therefore, after comparing the data quality of separate hours, the 'best' hours are combined into new data selections;

> Method 1: The first of these new selection criteria are automated and require little effort, but improve the quality substantially; all daytime hours (08:00 to 20:59) during weekdays (Monday to Friday) (Working hours, WH).
>
> Method 2: A second selection of higher quality (but requiring considerably more effort) is a manual selection of the highest quality hours for each day of data (Manual selection, MS).
>
> Method 3: An even more selective dataset contains only the highest quality hours for the five 'best' days of data (Manual selection 2, MS2).

In the case of autocorrelations it is observed that a refined selection of separate hours is not necessary in order to improve their quality.

3.5 Stacking

In order to amplify the desired signal originating from a virtual source, multiple panels with the same virtual source point are summed together, or 'stacked'. Any randomly distributed 'noise' will instead be cancelled out. Eventually, after enough stacking, the desired signal will stand out from the unwanted 'noise' by which it is being suppressed. Each ambient noise panel of 60 seconds that is recorded in a one hour window (and cross-correlated) is used to create a stack of 60 x 60 seconds (hourly stack) and all hourly stacks of one day are added together into a 24 x 60 x 60 seconds (daily) stack. The full range of all available panels is used to create a stack that contains around one month of recorded data.

In theory, stacking more panels should improve the quality of the dataset. However, it is observed (see 3.6.1) that this is not necessarily the case, due to variations in the quality of the data being stacked. Thus, new data selections are made (see 3.4.2) and stacked.

3.6.1 Cross-correlated data selections quality check

The stacks of the first data selections (see 3.4.1) are visually inspected. A bandpass filter between the values of 1.25, 2, 3.5, 5 Hz is applied in order to focus on surface waves and an automatic gain control (agc) is used to display them more clearly. Figure 8 shows a stack of the cross-correlations of one hour, one day, and one month of recorded data.

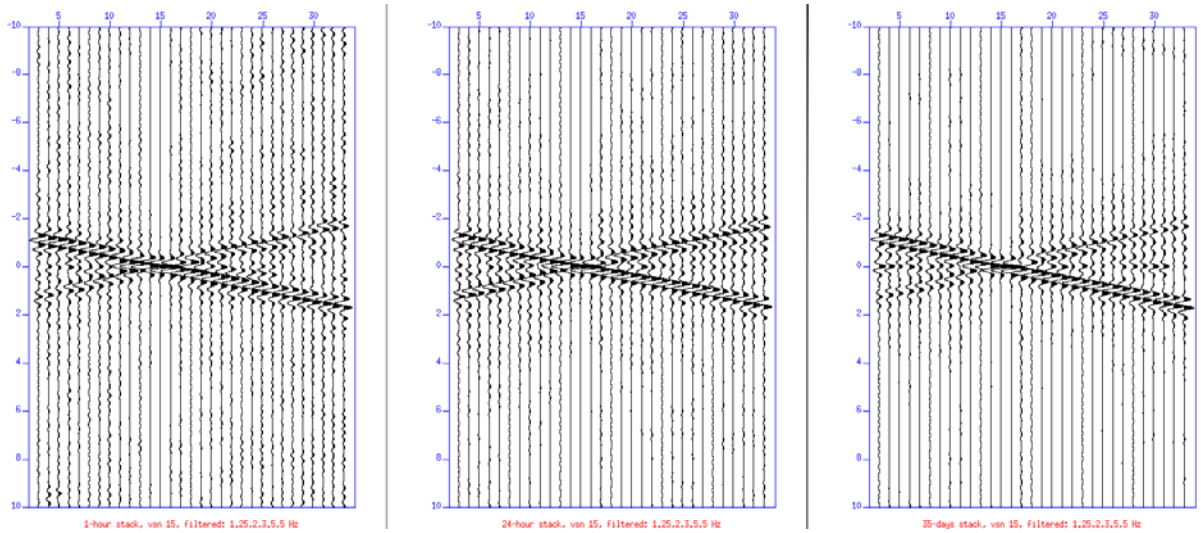


Figure 8: Stacks of cross-correlated noise panels with a virtual source on the 15th trace, windowed between -10 and 10 seconds. Positive and negative times represent the causal and acausal part of the virtual source panel, respectively. It is observed that stacking more cross-correlated data amplifies the surface wave signal (x-shape), when comparing the 1-hour, 24-hour and 35-days stacks.

Qualitative inspection per hour allows for new data selections to be made out of the hourly stacks with the clearest desired surface wave signal. Figures 13.1-13.6 show hourly stacks for one virtual source, between the hours of 03:00 and 10:00 AM on Wednesday the 27th of May 2016. Hourly stacks created from data recorded at night, or on weekend days, show a significantly weaker signal than those created from data recorded during weekdays, at daytime. This difference is most likely caused by a higher amount of noise sources, induced by the higher amount of traffic that occurs during weekdays. (see 4.3)

3.6.2 Autocorrelated data selections quality check

Autocorrelation stacks of one hour, one day and one month of recorded data are shown in Figure 15. Significant differences in the separate hourly stacks of autocorrelations are not observed; rather, their quality seems to be positively linked to the amount of data that is stacked. Therefore, no new autocorrelation data selections are made.

3.7.1 Symmetry test

A virtual source panel will consist of signals travelling in both the positive and the negative time; the causal and acausal parts of a seismogram, respectively. If there are no directional differences in the travel times of a wave propagating through a medium (reciprocity), the acausal part of a virtual source panel can be 'flipped' in the time domain and stacked with the causal part to improve the signal even more. In order to test for reciprocity the causal and 'flipped' acausal parts of a virtual source panel are overlain and displayed together (see Figure 16).

3.7.2 Causal – Acausal Stacking

If the symmetry test (see 3.7.1) is positive the causal and acausal parts of a virtual source panel can be stacked, which will improve the signal further. However, any difference in directional magnitude of signals caused by a non-uniform distribution of noise sources will be lost. As this research focuses on exploring subsurface properties rather than the distribution of noise sources, causal and acausal stacking is performed.

The acausal part is cut off from the causal part, and reversed in the time domain. The causal and 'flipped' acausal part are then stacked, creating a virtual source panel with signals travelling only in the positive time.

Stacking the causal and acausal parts of an autocorrelation panel will neither improve or reduce the quality of the signal, as autocorrelations are (by definition) identical in both the positive and negative time direction.

3.8 Frequency - wavenumber spectrum

The used receiver spacing of 25m limits the maximum apparent wavenumber to be measured to:

$$\frac{1}{2 * 25m} = \frac{0{,}02}{m}$$

According to the dispersion relation $\omega = \frac{v}{\lambda}$, this results in maximum frequencies to be able to be measured for each wave propagation velocity by the relation:

$$Frequency = velocity * \frac{0{,}02}{m}$$

Any signals with a frequency/velocity ratio above this 0.02/m threshold will be aliased and therefore unreliable. This energy must be disregarded and frequency vs wavenumber plots are made in order to visually verify this threshold.

3.9 Plotting Phase velocity against frequency

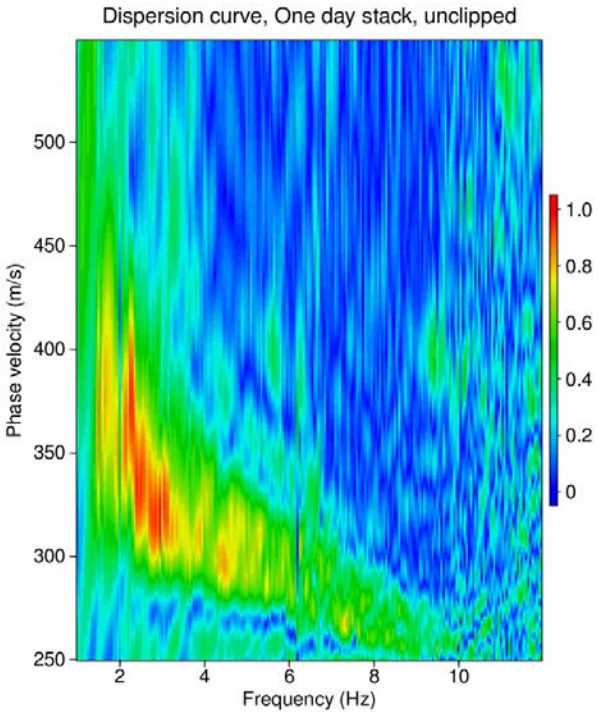
Wave propagation speeds vary, and usually increase, with depth (Haskell, 1953). Surface waves components with a large wavelength will travel deeper in the subsurface than those with smaller wavelengths, therefore travelling faster. As a result, wave components originating at a point source will propagate at different speeds based on their wavelengths. This phenomena is known as dispersion, and can be visualized by plotting a signal's phase velocity against frequency.

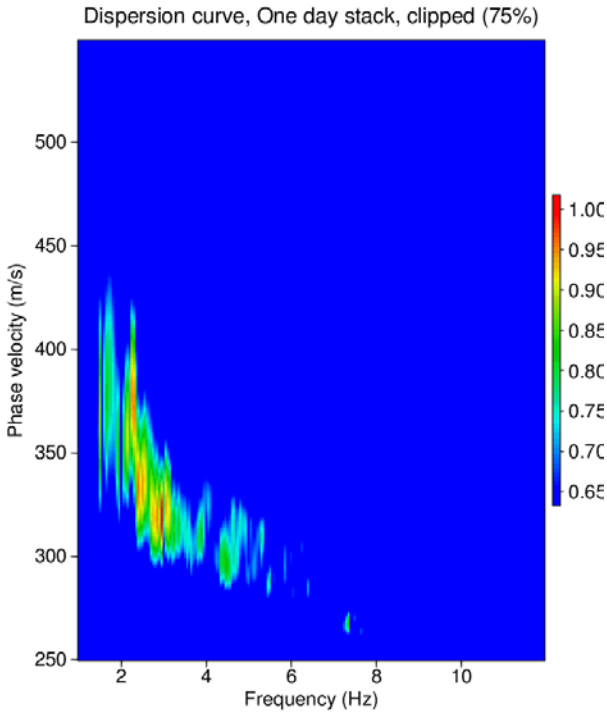
3.10 Dispersion Curve Picking.

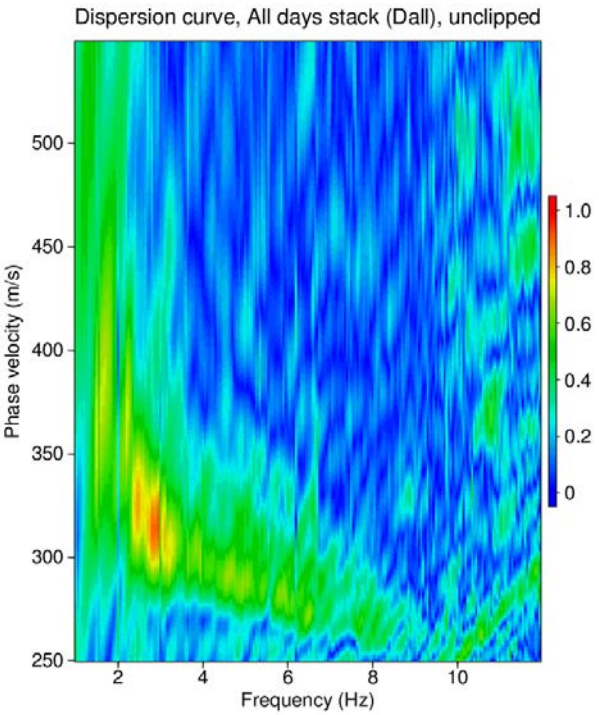
A phase velocity against frequency amplitude spectrum plot is created from the virtual source panels, showing their dispersion curves (Figures 9.1-9.3). The lower 60-90% of its amplitude spectrum is removed ('clipped'), so that only the most reliable points are left over (see Figures 10.1-10.3) This research focusses only on the fundamental mode dispersion curves.

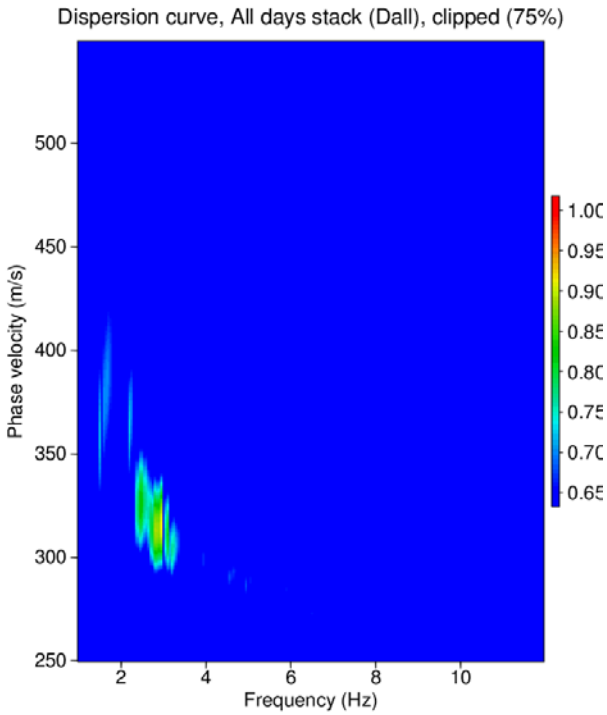
Figures 9.1 - 9.3: Phase velocity against frequency plots. The high amplitude area visible in the lower left corner represents the so-called 'dispersion curve' of the fundamental mode.

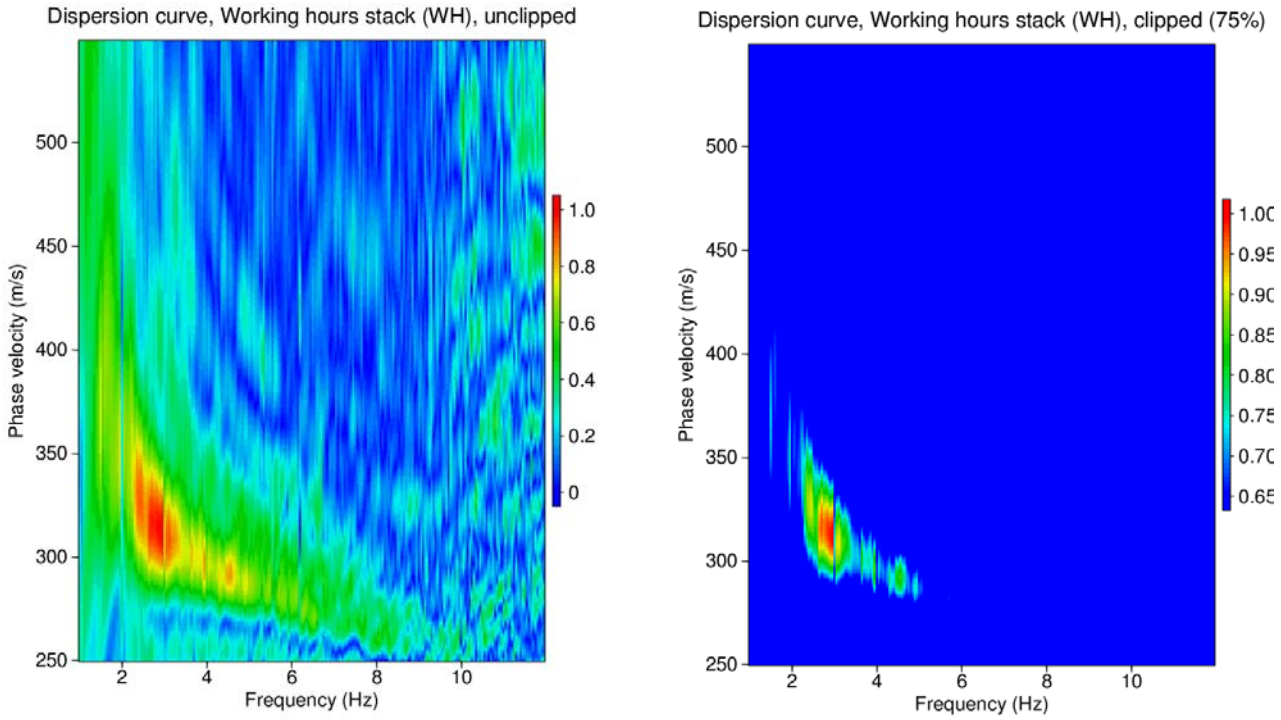
Figures 10.1 - 10.3: The same phase velocity against frequency plots as in Figures 9.1-9.3, but with the lower 75% of the amplitude spectrum removed ('clipped'). Note that the Working Hours data selection (WH) shows the smoothest curve.











Then, for each frequency step of 0.025 Hz, the velocity for which the amplitude in the clipped data is at its maximum is picked and saved. Any picks with a frequency/velocity ratio above 0.02/m are discarded (aliasing, see 3.8). The result is a dispersion curve as shown in Figures 19.1-19.3.
This process is repeated for the stacks of every data selection described in 3.4, for several clipping percentages ranging between 60-90%.

3.11.1 Parameter choices

The inversion of dispersion curves can be finetuned through a number of subsurface parameters. The parameters that are taken into account when performing a dispersion curve inversion are compression-wave velocity (Vp), shear-wave velocity (Vs), density and Poisson's ratio (the ratio of transverse strain to axial strain). Two parameter selections are made; one for a two-layered and one for a three-layered subsurface. In both cases Vs for each layer is chosen to lie between 50 and 800 m/s, and Vp is set to 200 to 5000 m/s for each layer. Poisson's ratio is constrained to a range of 0.2 to 0.5, and density to a range of 1500 to 2500 kg/m3. The two-layered model will have a layer-transition at a depth of 1 to 150m. The three-layered model is set to have its first transition at a depth of 1 to 150m, and its second transition at a depth of 1 to 200m. Realistic values for these parameters lie well within the chosen minima and maxima (Prasad et al., 2004)

3.11.2 Inversion of picked curves

Using the parameters chosen in 3.11.1, the picked dispersion curves (see 3.10) are inverted with the goal of obtaining profiles of shear-wave velocity with depth. An inversion is performed using Wathelet's modification (2008) of Sambridge's Neighborhood Algorithm (1999) considering 50 cells (nr) with 50 new random samples (ns) being generated and added with each iteration for a maximum of 50 iterations (itmax), until an acceptable solution is reached. This is done for all virtual source positions, for the stacks of each data selection (Dall, WH, MS, MS2).

4 Results

4.1 Acquisition

The panels that were recorded (see 3.1) contain ambient noise signals originating from sources all around the 750m long line of seismic receiver stations. The bulk of the recordings are similar to the example shown in Figure 5 (3.1). Occasionally, however, the raw recordings are corrupted by signals originating from some repetitive source (Figure 11) . This result is obviously not desired, therefore panels such as in Figure 11 are discarded and not used for further processing.

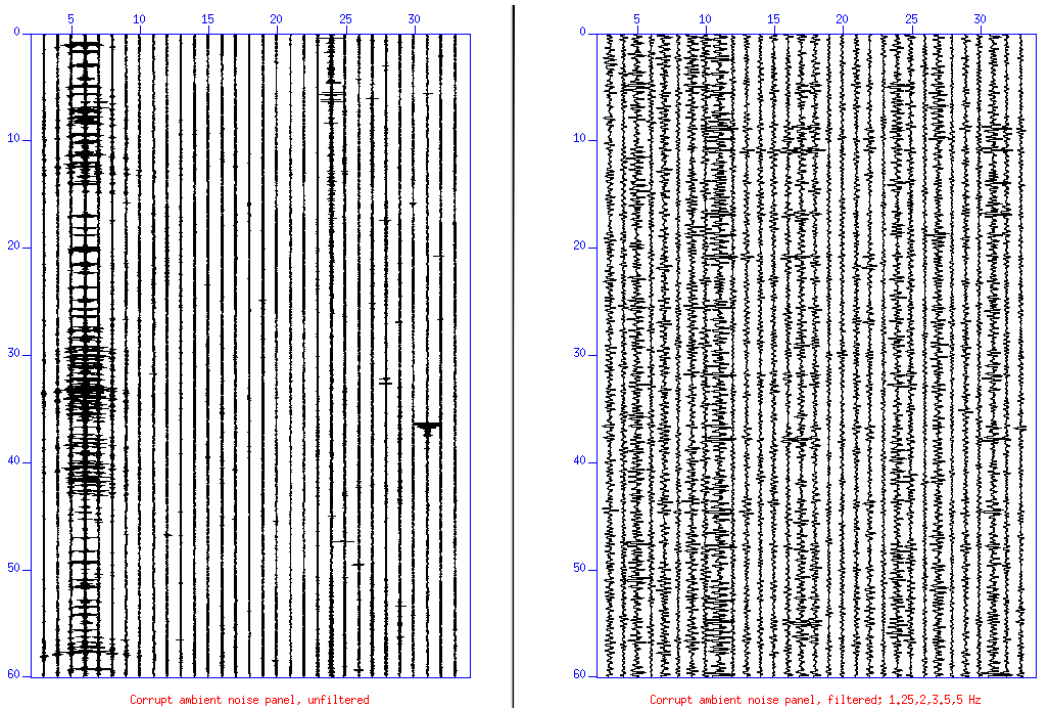


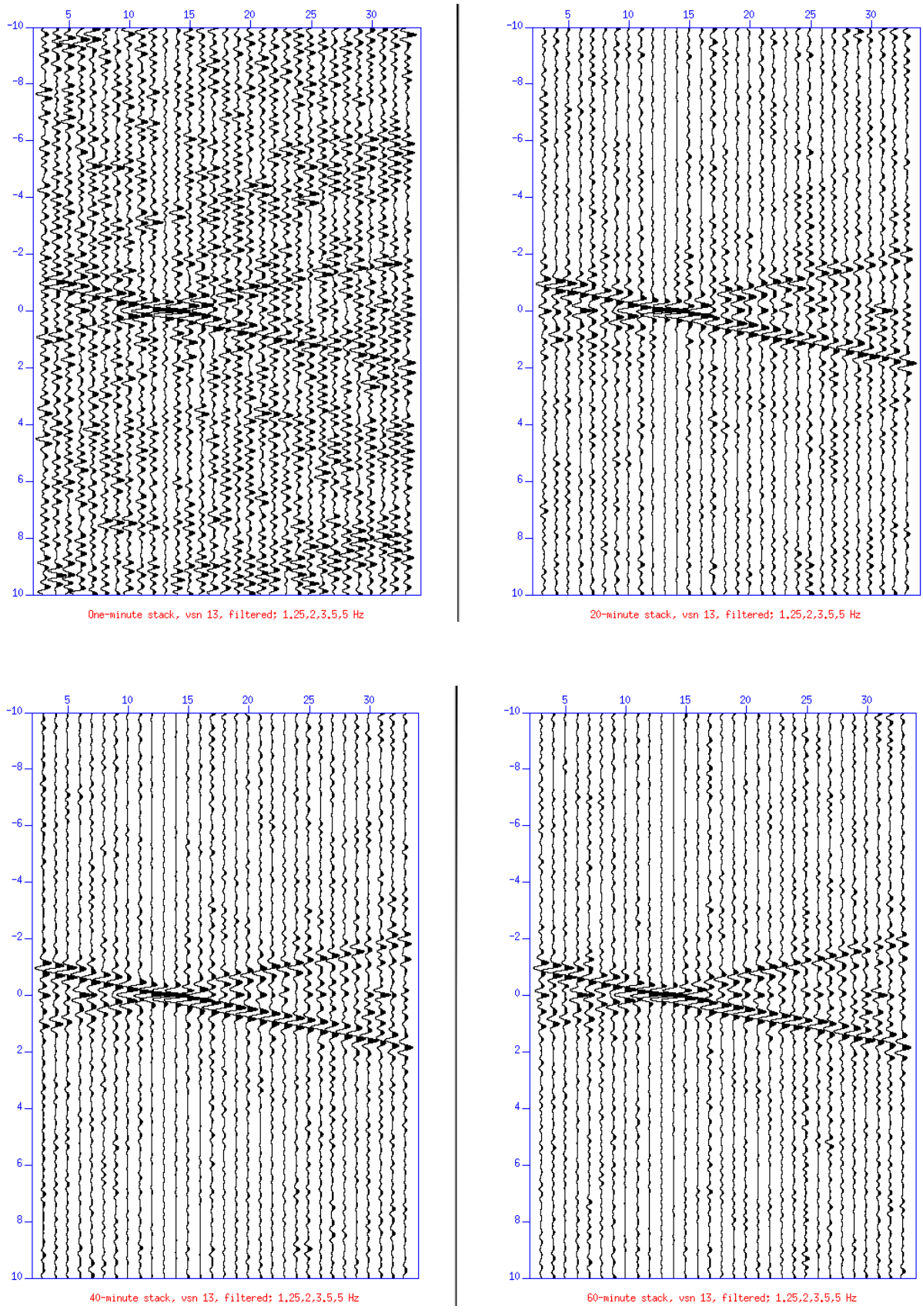
Figure 11: Example of a corrupt recorded raw ambient noise panel (left). A repetitive source is observed on the 6th trace. The bandpass filtered panel is shown in the right.

4.2 Cross-correlation

Cross-correlation (see 3.3.1) of raw recordings results in panels that resemble active seismic data. However, signals can be observed propagating in both the positive and negative time. Figures 12.1 shows the result of a cross-correlation of 1 minute of ambient noise data, with the 'master' trace at receiver number 13. The acausal (negative time) part of the panel is represented by t=-10 to t=0 and t=0 to t=10 shows the causal (positive time) part. The V-shaped surface waves in an active seismic shot become an X-shape in a virtual shot panel, and can already be seen after cross-correlating just one minute of data.

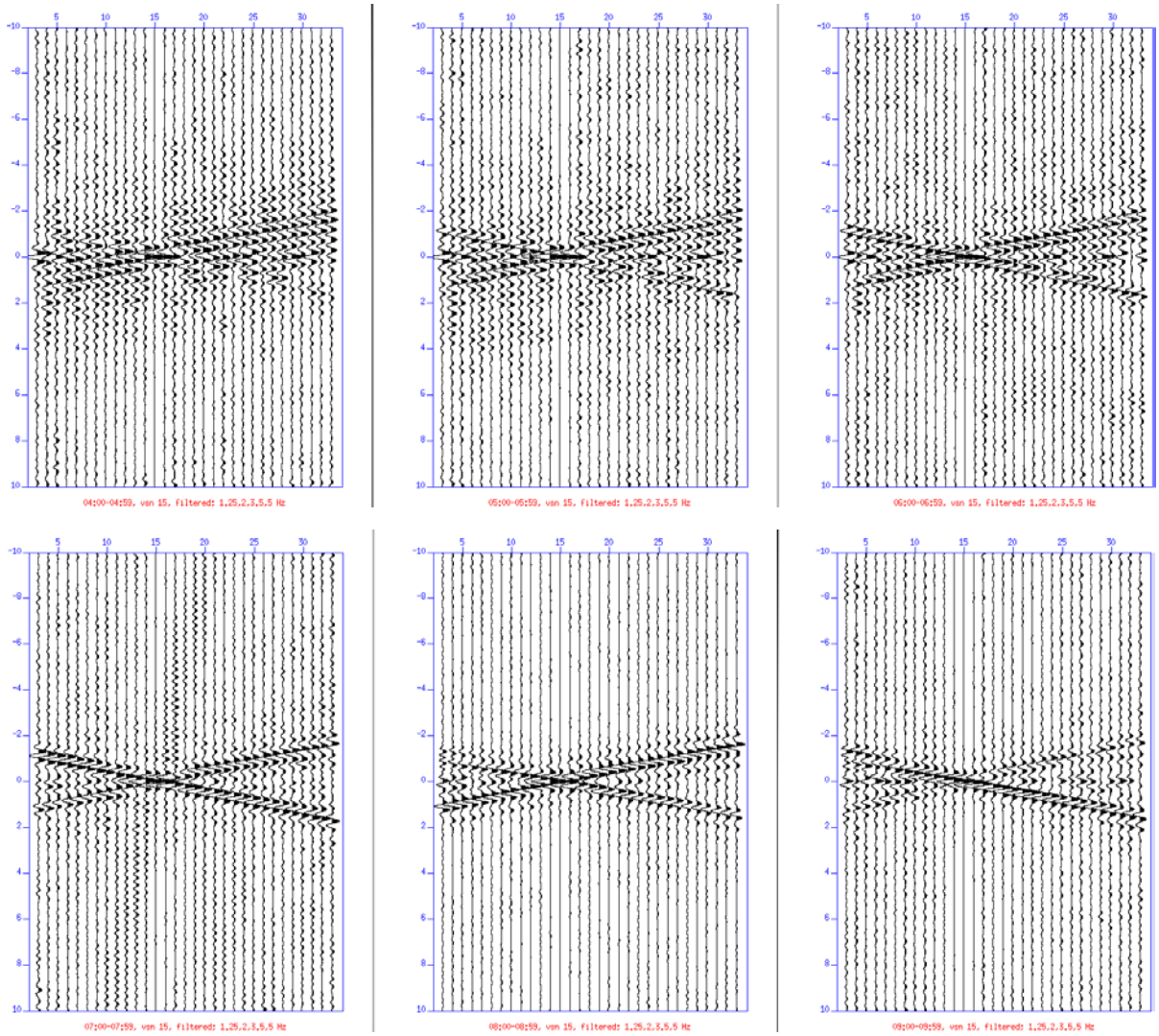
4.3 Stacking

Stacking (see 3.4) allows for an amplification of the desired impulse response, and just a small amount of panels stacked together significantly increases the quality of the data, as is shown in Figures 12.1-12.4.



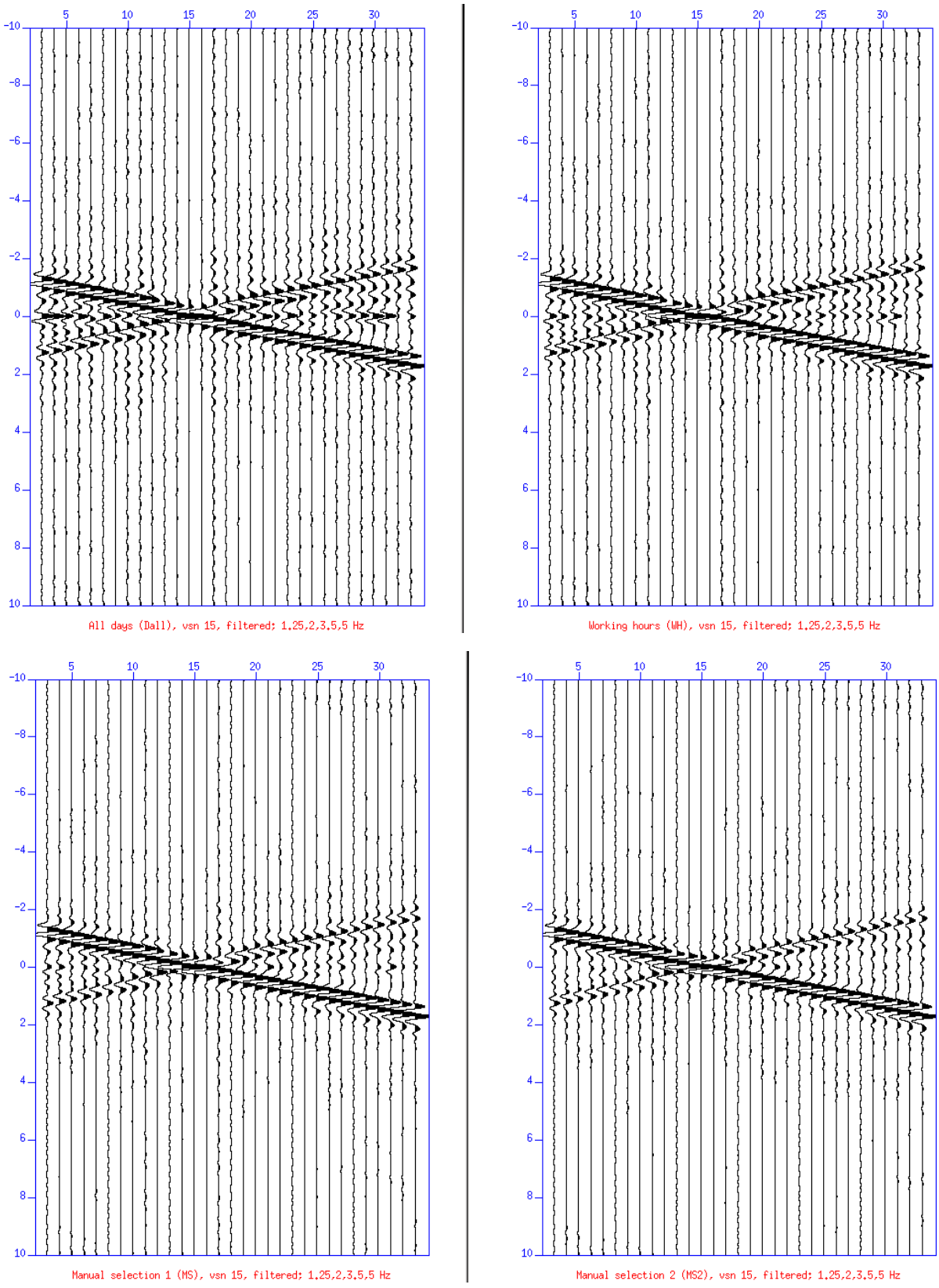
Figures 12.1 - 12.4: Stacks of 1, 20, 40 and 60 minutes of cross-correlated data, with a virtual source on the 13th trace and windowed between -10 and 10 seconds. Bandpass filtered between 2 Hz and 3.5 Hz to focus on surface waves (x-shape). The amplification of the desired impulse signal by stacking is clearly visible.

While this is the case for most panels that are stacked, not all cross-correlated virtual source panels improve the signal quality. Especially data recorded at night-time and during weekends seems less useful than data recorded during weekdays, between the hours of 08:00-20:59. 1-hour stacks of the same cross-correlated virtual source, but recorded at different times are shown in Figures 13.1-13.6.



Figures 13.1 – 13.6: 1-Hour stacks of cross-correlated noise panels recorded between 04:00 and 10:00 AM on a Wednesday. Windowed between -10 and 10 seconds, with a virtual source on the 15th trace and bandpass filtered between 2 Hz and 3.5 Hz. The quality of the cross-correlated data clearly improves with the time of day, indicating that noise sources are more abundant during daytime. This is most likely linked to the higher amount of traffic that occurs.

After stacking the data selections mentioned in 3.4 , an improvement in quality of the desired surface wave signal is observed (Figures 14.1-14.4). An especially visible improvement with each more selective stack is the removal of the high amplitude signal around t=0 in the 24-hour/day stack (Figures 14.1).



Figures 14.1 – 14.4: Cross-correlated panel stacks of increasingly selective data selections, with a virtual source at the 15th trace and windowed between -10 and 10 seconds. Bandpass filtered between 2 Hz and 3.5 Hz. The 'All days' selection (Dall) consists of all the cross-correlated noise data that has been recorded over 35 days. 'Working hours' (WH) contains data recorded over the same 35 days, but only between the hours of 08:00-20:59 on Mondays through Fridays. 'Manual selection 1' (MS) is a more selective version of the 'Working hours' stack, containing only the very best hours of each recorded day. 'Manual Selection 2' (MS2) consists of only the very best hours of the 5 best days of data. Note that with each selection the desired impulse signal is amplified and the surface wave signal (x-shape) becomes clearer.

The stacks of the manual selection MS and MS2 are of only slightly higher quality than WH (hours 08:00-20:59, weekdays), but require intensive effort. As the goal of this research is to explore to which extent subsurface properties can be acquired with reasonable effort, the results of the selections MS and MS2 are not primarily focused on.

4.4 Autocorrelation

Autocorrelations are affected more by noise signals coming from the subsurface directly below each receiver station, rather than nearby (probably traffic-induced) surface wave signals. Therefore there is no difference in their quality during nighttime or daytime. The stacks of autocorrelations of 1 hour, 1 day and 1 month of recorded data are shown in Figure 15.

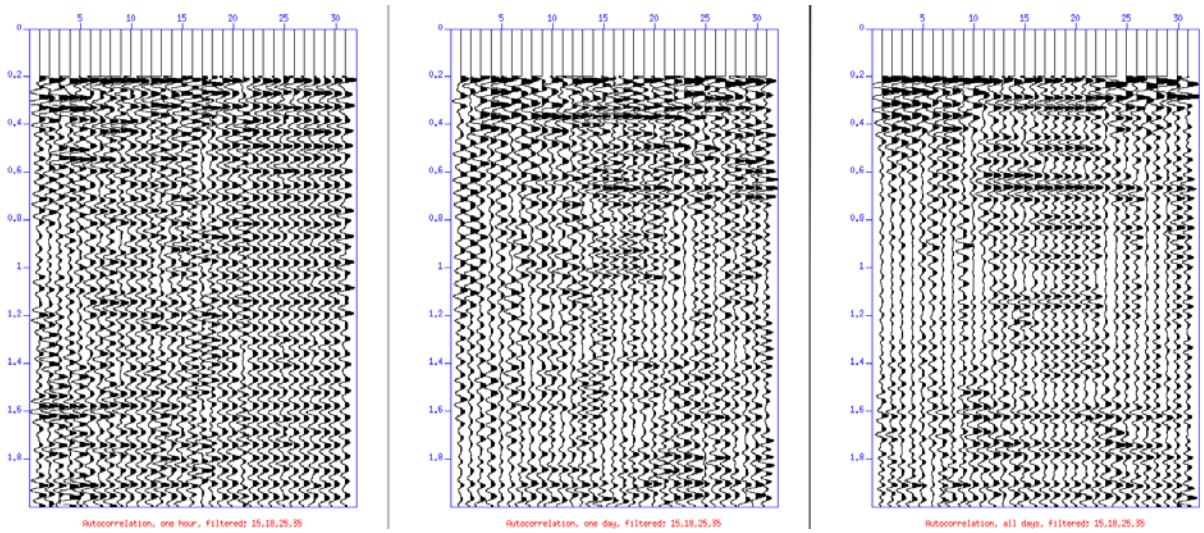


Figure 15: Stacks of autocorrelations of 1 hour, 1 day and 35 days of recorded noise, windowed to 2.0 seconds. Bandpass filtered between 18 Hz and 25 Hz, with filter slopes ending at 15 Hz and 35 Hz, respectively. The first 0.2 seconds of the autocorrelation stacks have been muted, as they contain very high amplitude signals which overshadow the rest of the panel. With more data stacked reflecting layers become increasingly visible.

4.5 Symmetry test

A symmetry test is performed in order to test the reciprocity of the subsurface. Figure 16 shows the acausal part plotted over the causal part. Their signals overlap sufficiently (less than 5% misfit) to assume reciprocity, therefore causal and acausal stacking is performed (see 3.7.2). The causal and acausal stacks of the panels in Figures 14.1-14.4 are shown in Figures 17.1-17.4.

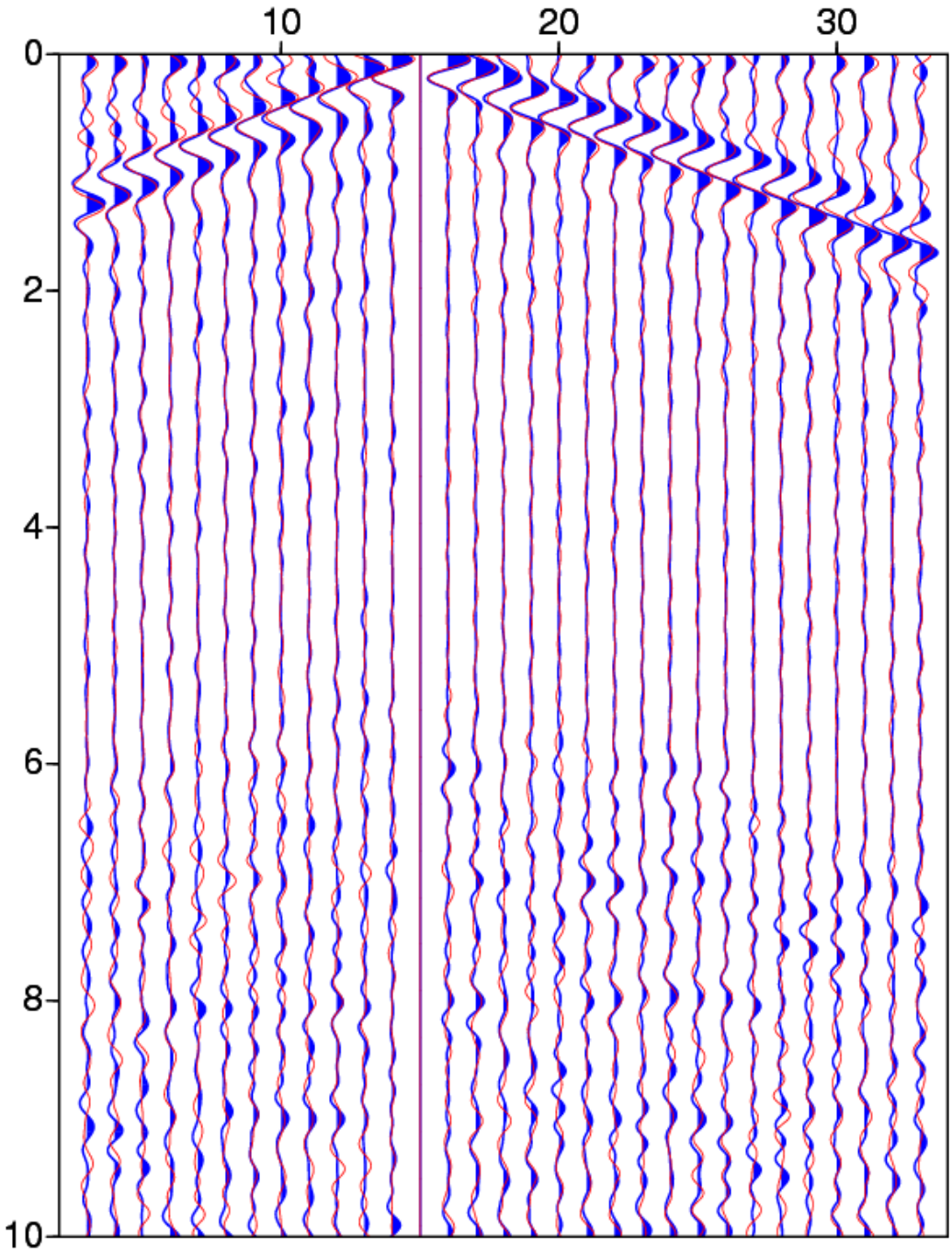
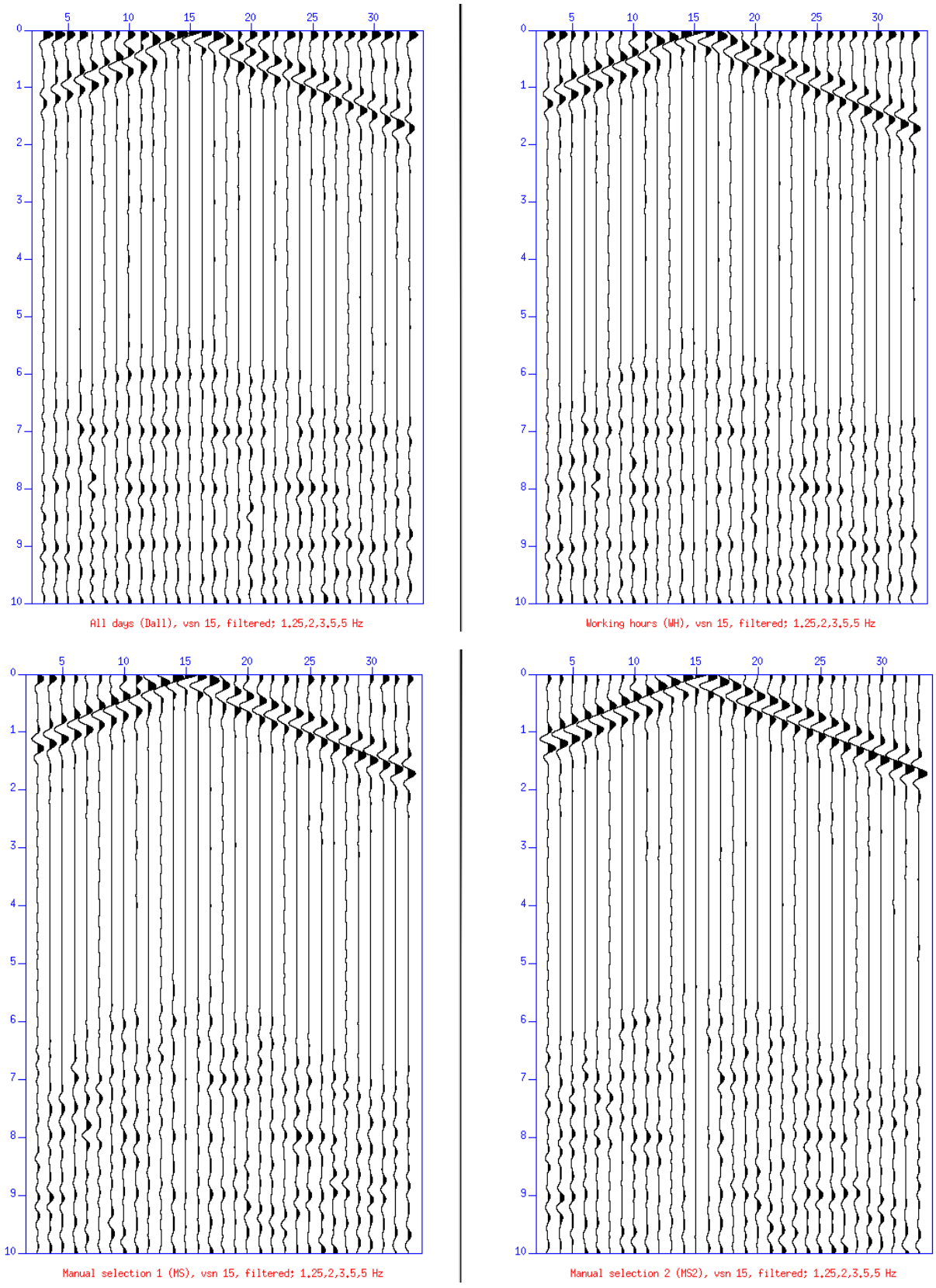


Figure 16: Symmetry test; The causal (red) and flipped acausal (blue) parts of a cross-correlated panel with a virtual source on the 15th trace are overlain. The peaks of the surface wave signals overlap, thus reciprocity is assumed. Stacking of the causal and acausal parts of the cross-correlated panels will further amplify the desired impulse signal to 'noise' ratio.



Figures 17.1 – 17.4: Stacks of the causal (positive time) and flipped acausal (negative time) parts of the cross-correlated virtual source panels shown in Figures 14.1-14.4.

4.6 Frequency – wavenumber spectrum

The energy distribution in the frequency – wavenumber spectrum plot (see 3.9) of a virtual source panel is shown in Figure 18 (data selection WH). Note that most of the energy is concentrated between lines a and b and c and d, corresponding to wave velocities of 235 m/s, 420 m/s, -235 m/s and -420 m/s, respectively. The energy 'wrapped' around the edges (concentrated between lines a' and b' and c' and d') is an effect of aliasing.

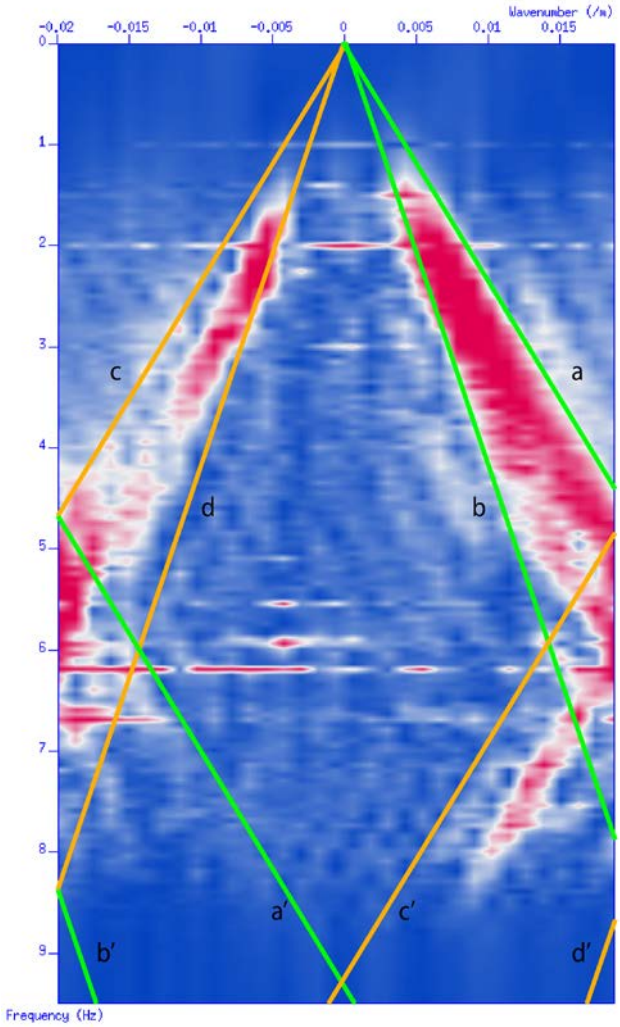
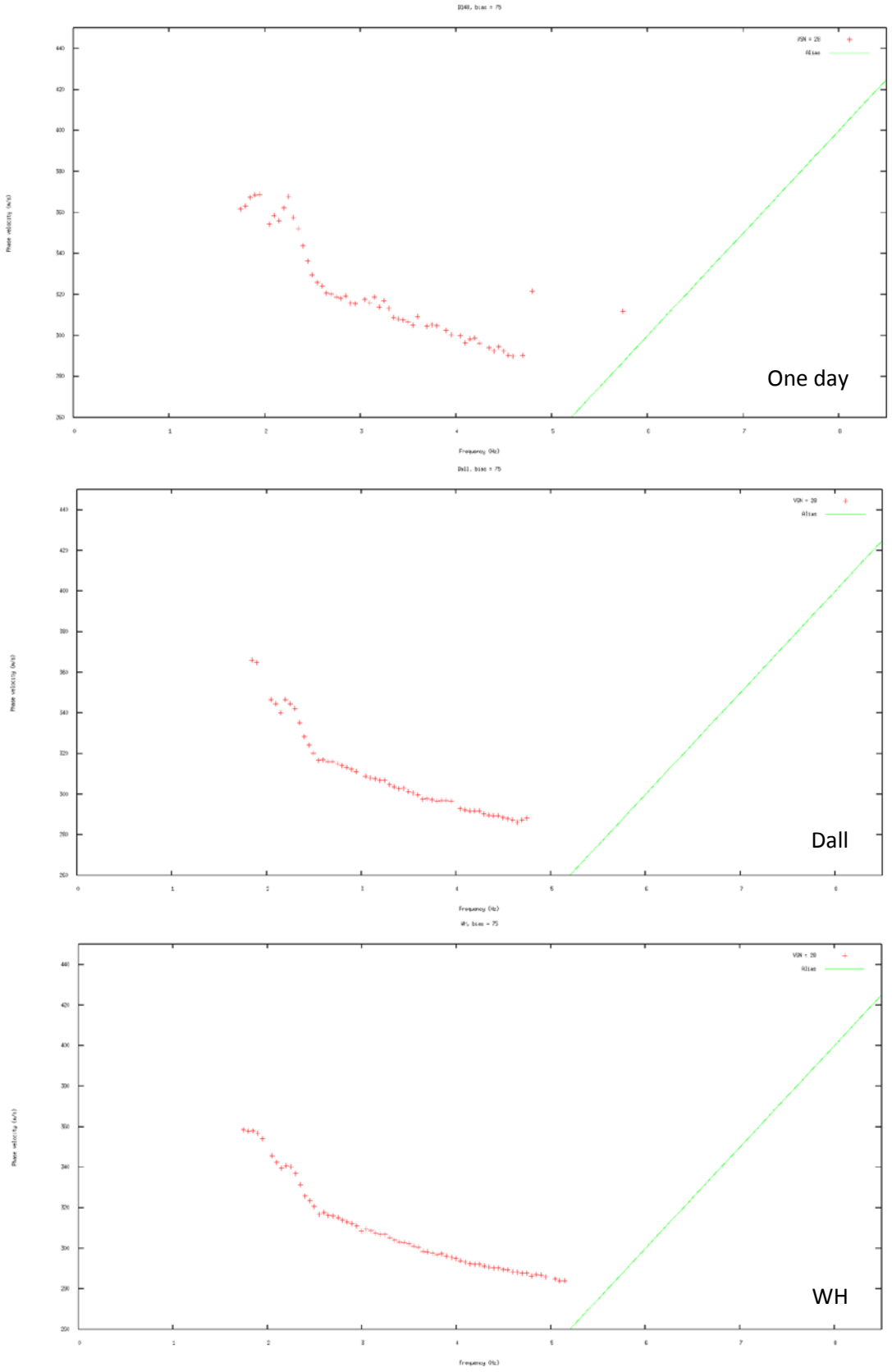


Figure 18: Frequency against wavenumber spectrum plot of a cross-correlated panel stack (WH) with a virtual source on the 15th trace. Most of the energy is concentrated between lines a and b and between lines c and d, which represent wave velocities of 235 m/s, 420 m/s, -235 m/s and -420 m/s, respectively. The frequency at which these lines hit the wavenumber=+-0.02/m threshold is their aliasing frequency, above which the data becomes unreliable. Energy above the aliasing threshold will 'wrap' around the plot. For lines a, b, c and d this phenomena is represented by lines a', b', c' and d'.

4.7 Phase velocity against frequency

Phase velocity against frequency plots are created for each virtual source number of each data selection stack (Dall, WH, MS, MS2). Figures 9.1-9.3 show this plot for virtual source number 28, for several datasets. For each frequency the phase velocity with the highest amplitude is picked. The resulting picked dispersion curves for the phase velocity against frequency plots in Figures 9.1-9.3 are shown in Figures 19.1-19.3, with a 'clipping' percentage of 75% (see 3.10 & 5). The green line represents the limit above which the picks are aliased.



Figures 19.1 – 19.3: Picked dispersion curves from the phase velocity against frequency plots in Figures 9.1-9.3 and 10.1-10.3 (stacks of one day of data, Dall selection and WH selection). The green line in each figure represents the aliasing threshold, above which any picks are unreliable and should not be included in the inversion process. Note that a single day of recorded data results in a 'messy' dispersion curve. The working hours selection (WH) results in the smoothest curve.

4.8 Inversion

Inversion is applied to each acquired dispersion curve (see 3.11.2). The resulting curve fits for the dispersion curves in Figures 19.1-19.3, and a dispersion curve of the MS2 data selection, are shown in Figure 20.1-20.4, for the 2-layered model (see 3.11.1). Note that in these figures slowness (s/m) is used rather than velocity (m/s). Each line represents a dispersion curve, modelled from possible subsurface properties that are limited to the parameters defined in 3.11.1. The color of each line shows its misfit (normalized by slowness) with the originally picked dispersion curve, the red lines having the lowest misfit (see 4). The picks of the original dispersion curve are represented by the dotted black line. The fitted curves are linked to the profiles of shear-wave velocity with depth shown in Figure 21.1-21.4, using the same colors.

A comparison with the 3-layered model is shown in Figure 22. Because the 2-layered model shows a more uniform dispersion curve fit, this model is preferred for describing the subsurface rather than the 3-layered model.

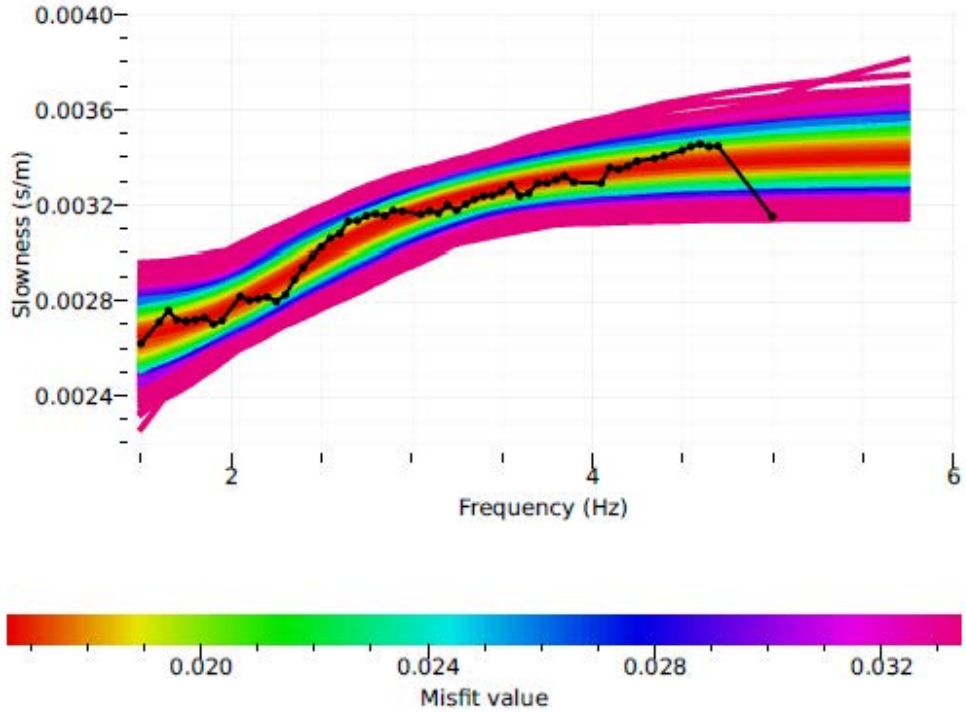


*Figure 20.1: Inverted dispersion curve fits for **one day of recorded data** (colored curves). The dotted black line represents the picked dispersion curve that was subjected to inversion. The red curves represent the inverted curves with the best fit.*

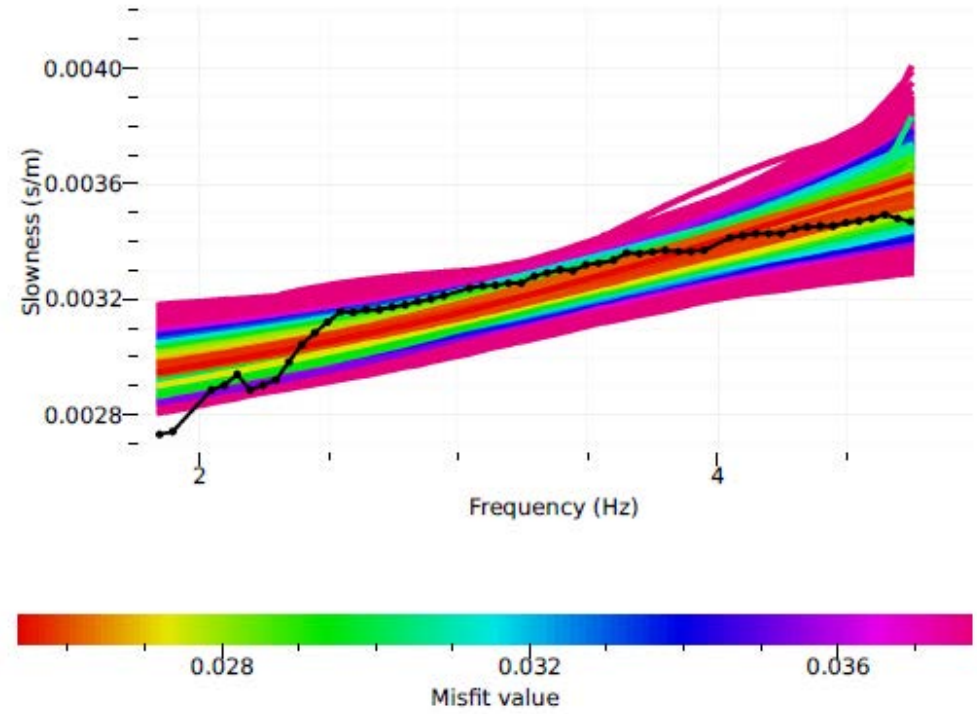


*Figure 20.2: Inverted dispersion curve fits for the **Dall data selection**, i.e. 35 full days of recorded data (colored curves). The dotted black line represents the picked dispersion curve that was subjected to inversion.*

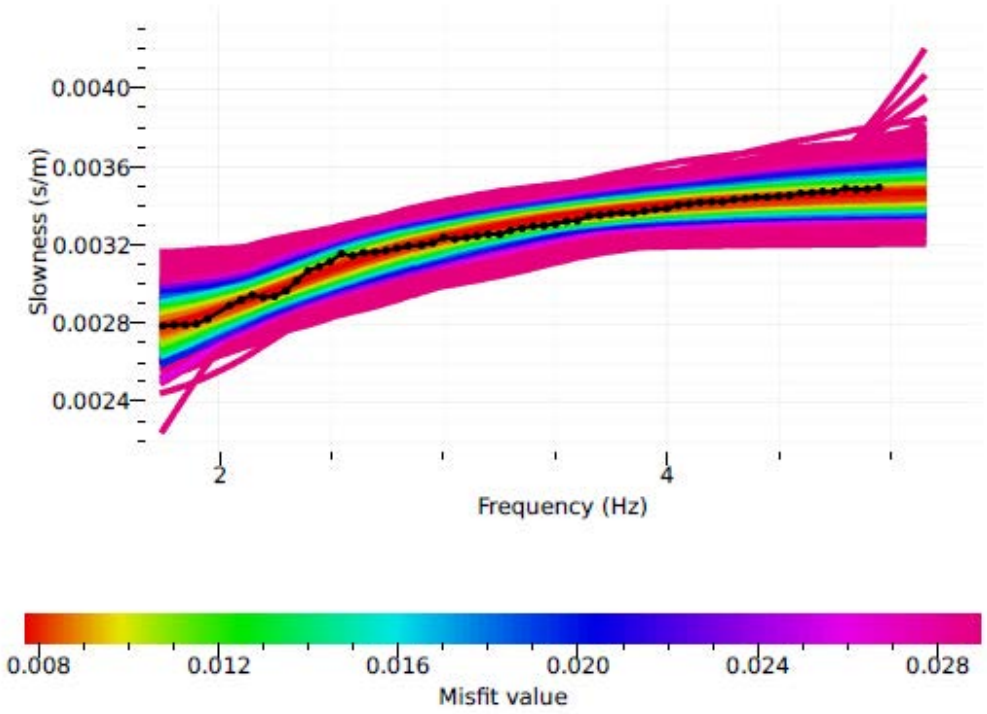


*Figure 20.3: Inverted dispersion curve fits for the **WH data selection**, i.e. all data recorded between 08:00-20:59 on weekdays (colored curves). The dotted black line represents the picked dispersion curve that was subjected to inversion.*

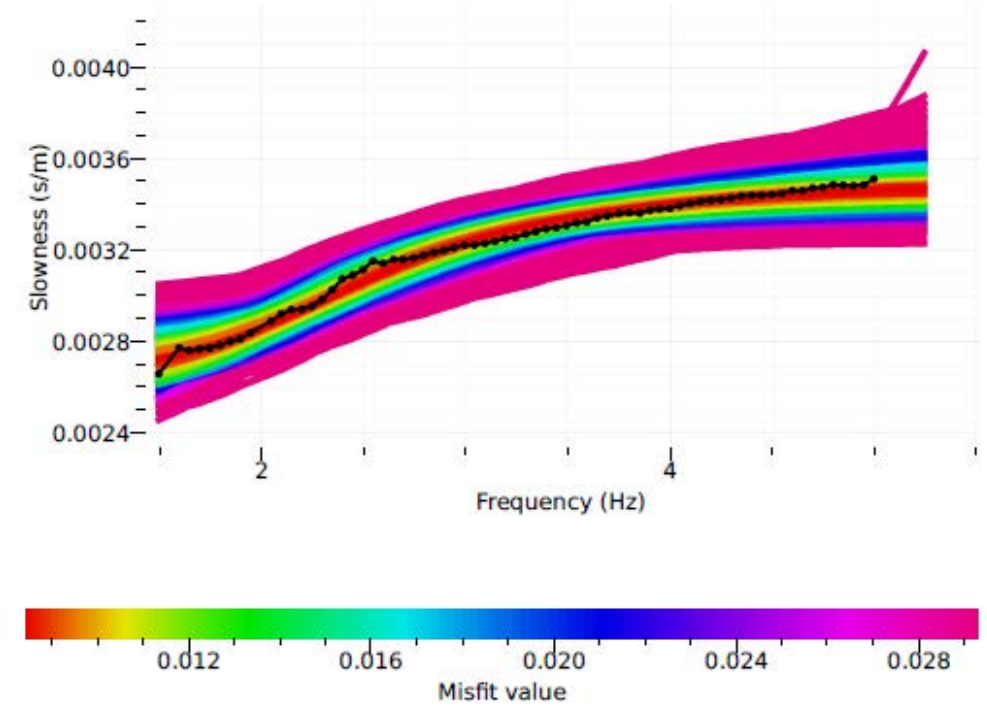


*Figure 20.4: Inverted dispersion curve fits for the **MS2 data selection**, i.e. the most selective manual data selection (colored curves). The dotted black line represents the picked dispersion curve that was subjected to inversion.*

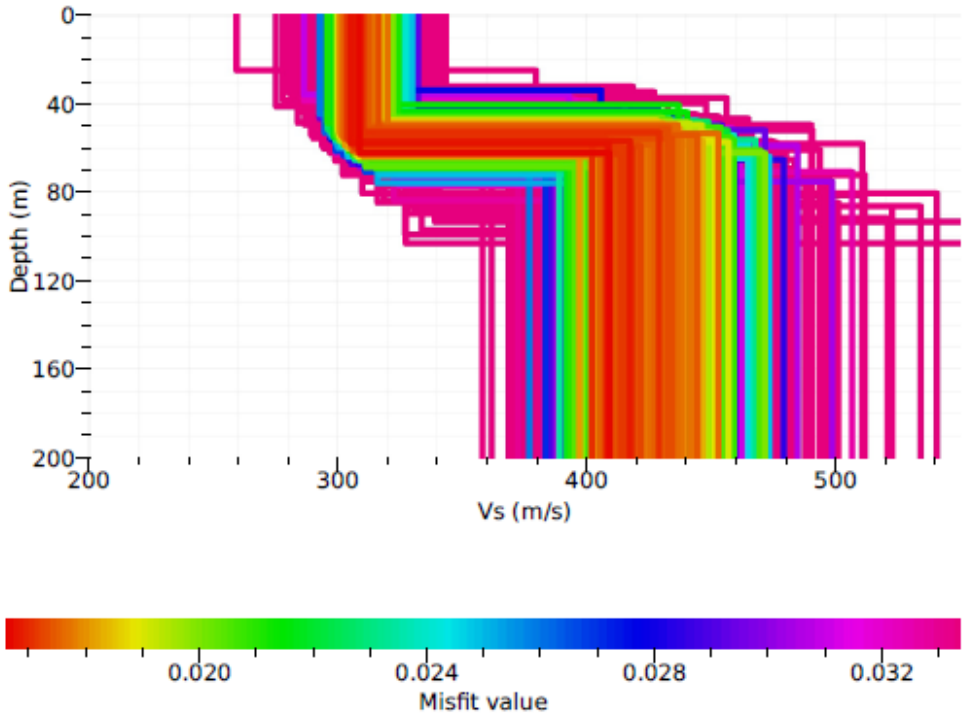


*Figure 21.1: Shear wave velocity with depth profiles linked to the dispersion curve fits in Figure 20.1 (**one day of data**).*

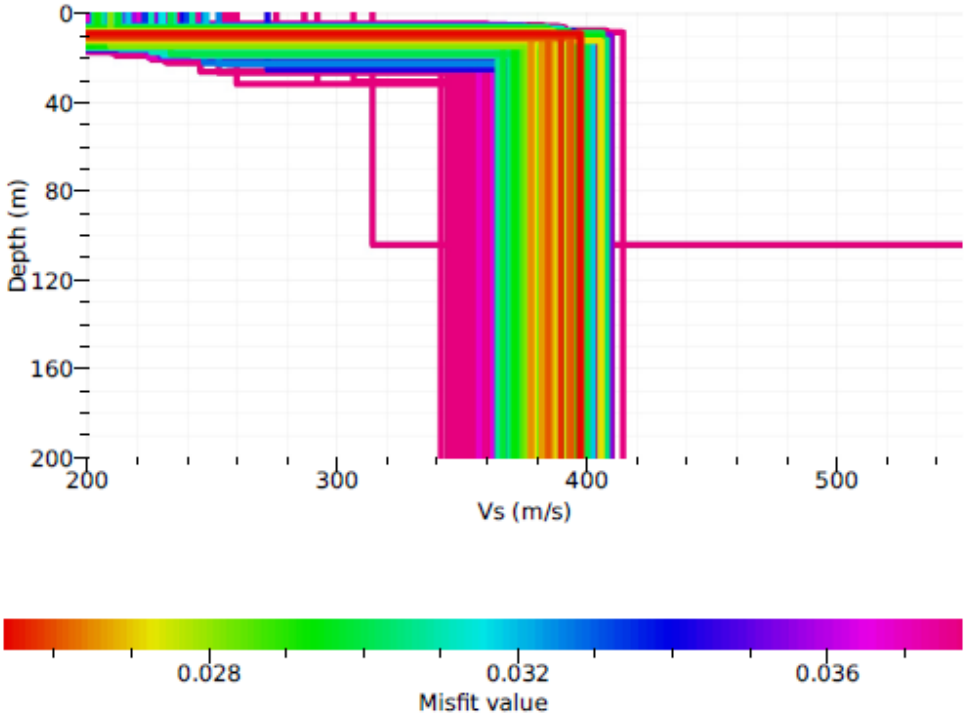


*Figure 21.2: Shear wave velocity with depth profiles linked to the dispersion curve fits in Figure 20.2 (**Dall selection**).*

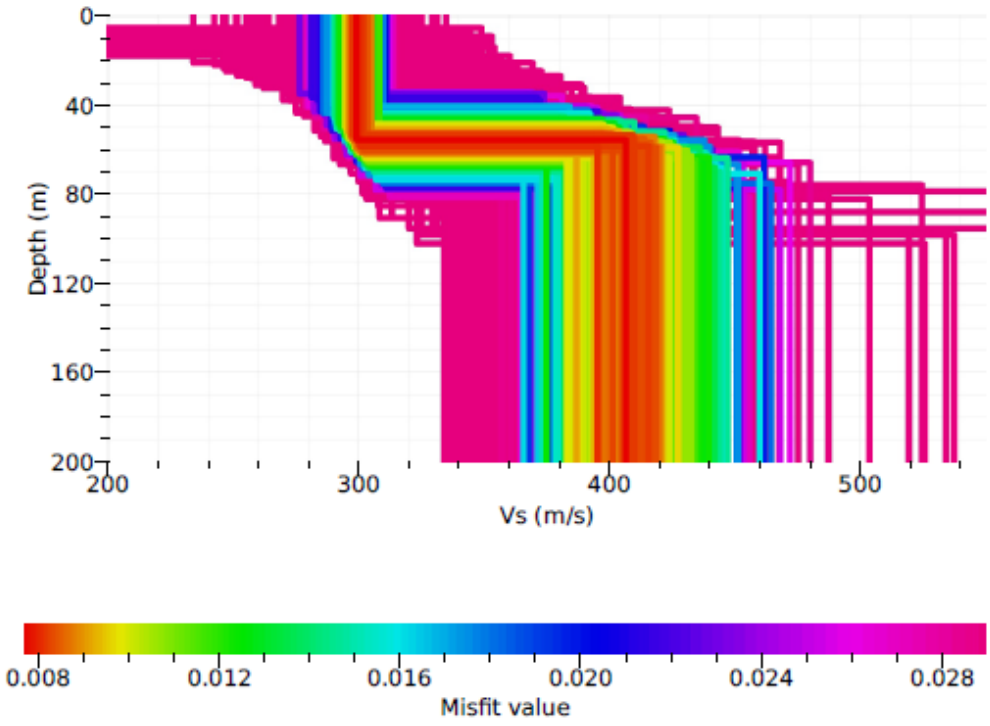


*Figure 21.3: Shear wave velocity with depth profiles linked to the dispersion curve fits in Figure 20.3 (**WH selection**).*

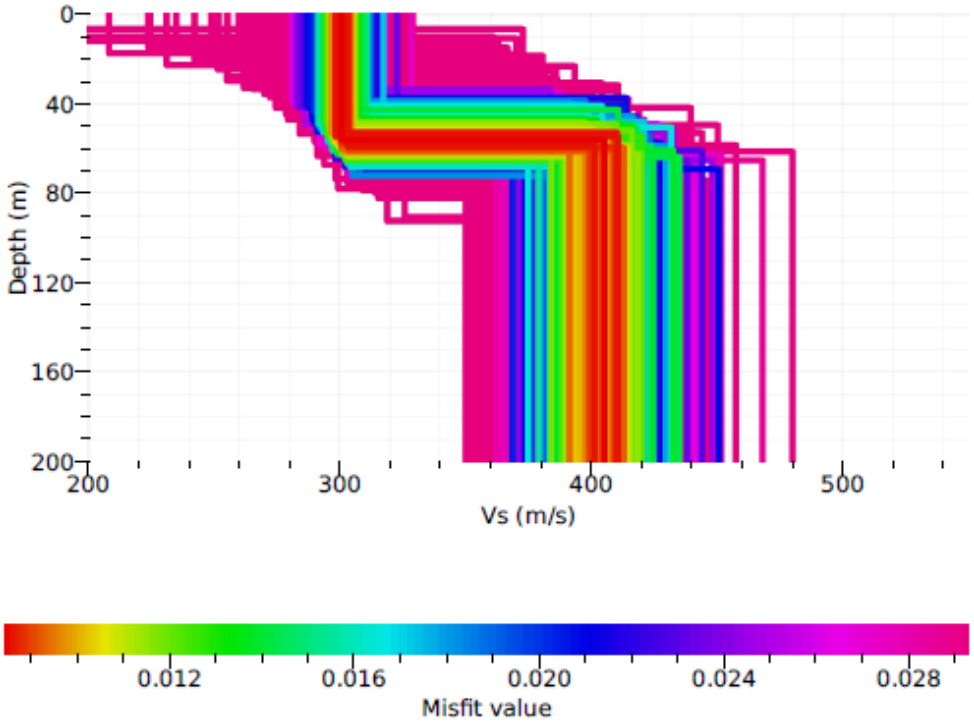


*Figure 21.4: Shear wave velocity with depth profiles linked to the dispersion curve fits in Figure 20.4 (**MS2 selection**).*

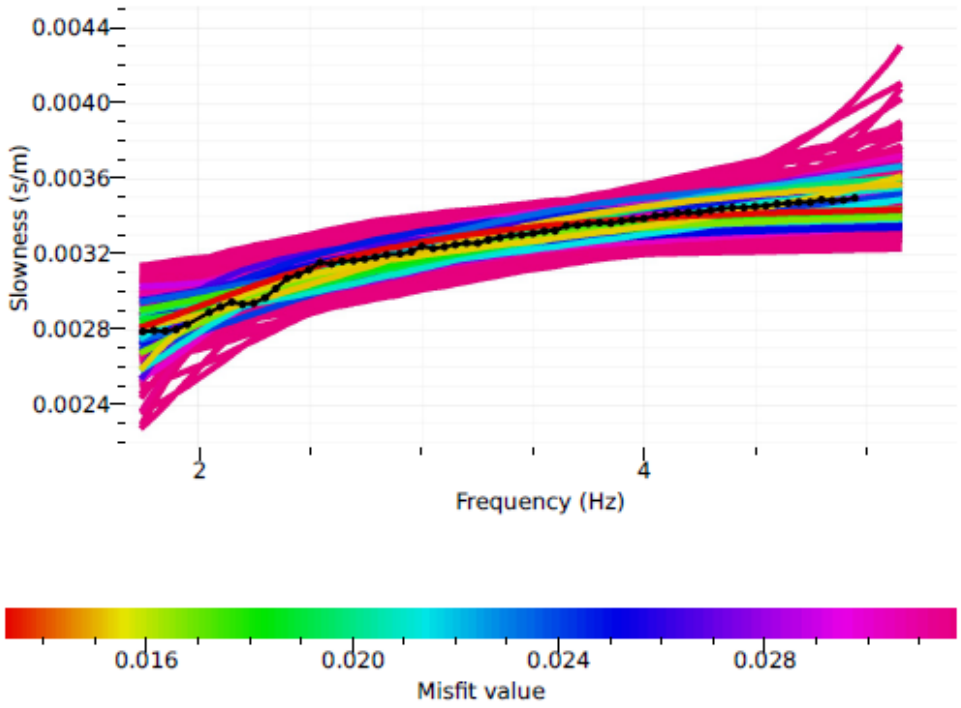


*Figure 22.1: Inverted dispersion curve fits for the **WH data selection**, for a 3-layered underground model (colored curves). The dotted black line represents the picked dispersion curve that was subjected to inversion.*

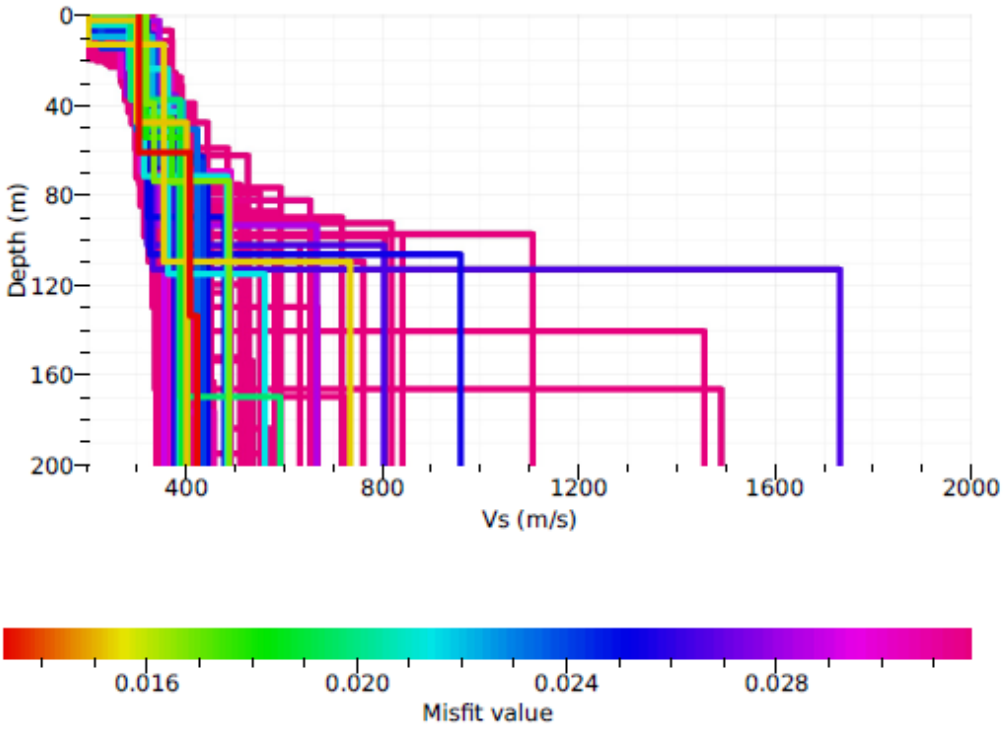


*Figure 22.2: Shear-wave velocity with depth profile corresponding to the dispersion curve fits in Figure 22.1 (**WH**). Note that the misfits are larger than those of the 2-layered model (Figure 21.3), and velocities with depth are not as well constrained.*

5 Discussion

The main difficulty in this research has proven to be selecting data to be used for processing, starting with the 1 minute raw ambient noise panels that sometimes contain a repetitive signal. This signal, very likely originating from a nearby drainage pump, corrupts the recorded data in such a way that the entire panel is rendered unusable; The cross-correlation of such panels would result in not only the difference in arrival time for a single signal across multiple traces, but also the difference between arrival times of multiple separate signals. If no check is performed to exclude cross-correlated 'corrupt' data panels from the stacking process (see 3.5), no useful information might be retrieved at all. An example of a stack containing cross-correlated corrupt data panels can be seen in Figure 11.

After cross-correlation another data selection (rather than stacking all panels together), is not necessarily required, but might improve the data quality of stacks even further. Checking the quality of each cross-correlated panel of data manually would be an inefficient and time-consuming task. Therefore, 1-hour stacks are inspected individually. As can be seen in Figures 13.1-13.4, ambient noise recorded during different hours of the day results in different levels of data quality. The 'best' hours of data for surface waves seem to lie between the hours of 08:00-20:59, from Monday to Friday. This difference in quality is most likely due to the larger contribution of noise sources from nearby farms and traffic during weekdays. However, as this research does not attempt to map the distribution or directions of ambient noise source locations, no certainty can be given to this matter. 'Beamforming' of cross-correlated noise data could give more insight into the origin of the recorded noise (Ruigrok et al. 2016). However, due to the receiver stations used for this research being set up in a straight line, one could only investigate the contribution of noise sources in the extension of the line (i.e. a left-right difference).

From inspecting hourly stacks the decision is made to create new stacks, containing only the hours that show the clearest signal. The results of these new data selections are shown in Figures 14.1-14.4. The most straightforward secondary data selection, consisting of data recorded between the hours of 08:00 – 20:59 from Monday to Friday, shows a clear contrast with the stack of 'all' data (Figures 14.1-14.2). The signal of the stacks of the data selections MS and MS2 (see 3.4.2) sharpen the signal even further, but at a cost. For these selections each individual hourly stack of recorded data used for WH was inspected manually, and discarded if its signal was not deemed sufficiently clear. A list of the hourly stacks selected for MS and MS2 is shown in appendix 1.

The results of stacking the MS and MS2 data selections are shown in Figures 14.3-14.4 and their signal is surprisingly clear. However, when inverting the dispersion curves created from these stacks no significant difference from the WH selection is observed. Taking into account the amount of effort necessary for a manual selection, the obtained results do not weigh up against the time spent. As this research aims to find an efficient method for extracting information from surface waves obtained from ambient seismic noise, its focus lies on the results of the Dall and WH data selections rather than on MS and MS2.

The autocorrelations show the response to signals that pass a single receiver multiple times, by being reflected by sedimentary layers below the receiver. After sufficient stacking these 'reflectors' show up in the autocorrelations, and any surface waves (propagating horizontally) will not be visible. As opposed to the cross-correlated panels, autocorrelations rely on noise source contributions from directly below the receivers, within the earth. Therefore, rather than improving from a secondary data selection, they show an increase with quality that is directly linked to the total number of panels stacked (Figure 15). Figure 23 shows a comparison between the stack of the autocorrelations

of all recorded data and the processed results of an active seismic survey, both filtered with the same frequency values. This active survey was performed in 1981, parallel to the passive array line, at a distance of approximately 500 m (BP Nederland b.v. 1982). The strongest reflectors are highlighted by green lines. The similarity in reflectors found using both methods suggests that autocorrelated seismic noise data can provide equally qualitative information on reflectors as data acquired by an active seismic survey. The remarkably low cost of a passive survey, relative to that of an active survey, suggests that this could prove to be a viable exploration method. In addition, passive seismic data can easily be used to increase the bandwidth of an active survey at the lower end of the frequency spectrum.

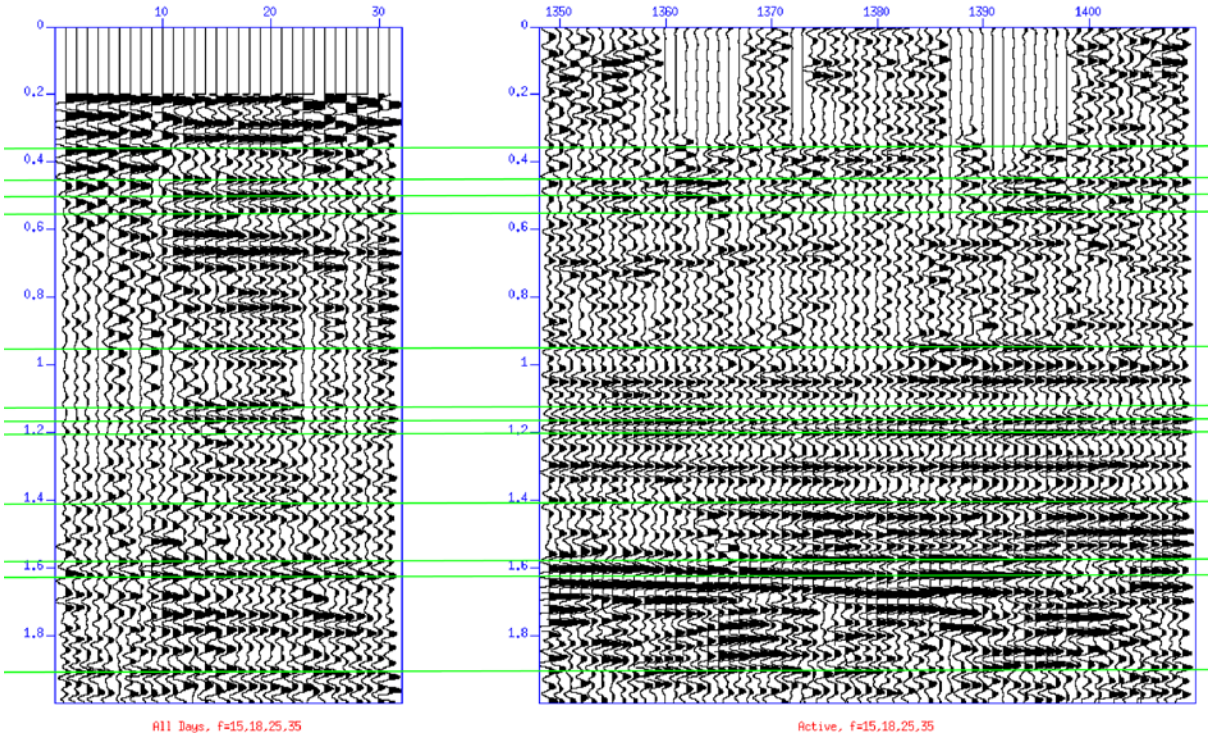


Figure 23: Stack of autocorrelations of 35 days of recorded noise data compared to the results of a nearby active survey. Both panels were bandpass filtered using the same frequency values. Observed reflecting layers are surprisingly similar in the results of both methods and the clearest ones are highlighted by green lines. Adapted from BP Nederland b.v. (1982).

The symmetry test performed (see 3.7.1) has shown that the subsurface in the surveyed area is reciprocal, i.e. there are no directional travel time differences for waves propagating through the medium. This means that the dispersion of surface waves is equal in both directions along the line. Because of this, stacking of the causal and acausal parts of each cross-correlated panel will further amplify the dispersed surface wave signal. However, the distribution of noise sources might not be similar in the extensions of the receiver line and any information on the contribution of noise sources from different directions will be lost.

This passive seismic survey was set up with a receiver spacing of 25 meters, in order to cover a 750m line. A longer line allows for a larger investigation area for reflecting body waves extracted from ambient seismic noise. However, the chosen receiver spacing might limit the bandwidth of surface wave frequencies that can be measured. The maximum wavenumber that can be measured is directly linked to the receiver spacing, i.e. 0.02/m. The velocity of a surface wave multiplied with its

wavenumber gives the frequency at which it is recorded, thus limiting the frequency of surface waves that can be measured. Higher frequency signals will become 'aliased' and will give distorted results (see 3.8). This can be confirmed by plotting the frequency energy spectrum of a virtual source panel against its wavenumber energy spectrum. The resulting spectrum shows how the energy of propagating waves is distributed. Figure 18 shows such a frequency-wavenumber spectrum. Surface wave energy shows up between straight lines determined by constant velocities and is concentrated between lines a and b and c and d. Their energy is significantly higher than that of body waves, rendering the body wave energy invisible relative to the energy of surface waves. Line a represents a propagation velocity of 235 m/s, line b a velocity of 420 m/s. These lines hit the wavenumber of 0.02/m threshold at frequencies of 4.6 Hz and 8.4 Hz, respectively. At higher frequencies the energy relates to wavenumbers that cannot be measured and is instead observed at lower wavenumbers. This energy 'wrapping' around the 0.02/m threshold is an effect of aliasing. As expected, aliasing occurs when the frequency/velocity ratio of a surface wave becomes greater than 0.02/m.

The dispersion of surface waves is due to a variation in propagation velocity for different frequencies. As density increases with depth, wave propagation velocities typically increase also. Surface wave components with a greater wavelength but the same total energy as a small wavelength wave component will have a greater amplitude and therefore travel deeper through the subsurface (Haskell, 1953). As a result wave components with a greater wavelength will propagate at greater velocities than small wavelength wave components.

Dispersion of the surface waves is already noticed in Figure 18, as different propagation velocities can be observed. A more straightforward method of identifying dispersion is by plotting the propagation velocity of waves against their frequency. Figures 9 shows a phase velocity against frequency amplitude spectrum for a virtual source panel. The curved red area shows the highest amplitudes and its shape is defined by the dispersion of surface waves. For each frequency the phase velocity with the highest amplitude is picked automatically, so that each frequency corresponds to a single phase velocity. The result is a curved line commonly referred to as a surface wave dispersion curve. In order to pick only the most reliable points (i.e. the highest amplitudes) the lower percentage (60-90%) of the amplitude spectrum is set to zero, or 'clipped' (Figures 10). The clipping percentages are varied until one is found that results in the maximum amount of picks, but with minimal extreme outliers. A clipping percentage of 75% meets these requirements best and is the preferable choice for this research. Examples of the resulting picked dispersion curve are shown in Figures 19.1-19.3. The green line represents a frequency/velocity ratio of 0.02/m above which aliasing occurs. Any picks above this threshold are unreliable and are not included in the inversion process.

Several parameter constraints are required before an inversion can be performed (3.11.1). Any known properties of the subsurface can be taken into account and allow for the inversion to be narrowed down to a solution with the smallest misfit. One of the main goals of this research is to investigate to which extent dispersion curves acquired through ambient noise seismic interferometry can successfully be inverted. If it is possible to apply this method without the need of alternative measurements and obtain reliable results, unnecessary costs and effort can be kept at a minimum. Thus, from a research point of view, the parameters are chosen with minimal assumptions on subsurface properties. Layer transition depths are limited to a maximum of 200m, in order to be able to verify the inversion results with borehole sampling.

Wathelet's modification (2008) of Sambridge's Neighborhood Algorithm (1999) is an often-used, computationally fast method of inversion. It is applied to all virtual source panels of each data selection, in order to compare the differences in their outcome. The inversion results in dispersion curve fits, accompanied by velocity profiles with depth. Compressional wave velocities are not easily defined when inverting surface wave dispersion curves. Shear-wave velocities, however, are constrained quite well. Figure 20.1-20.4 show examples of picked dispersion curves, together with the curves obtained from the inversion. Each of the modelled curves has a 'misfit' value representing how far off solution lies from the original dispersion curve (normalized by slowness), ranging between 0.0 (no misfit) and 1.0 (100% misfit). Modelled curves with a misfit of more than 0.05 (5%) from the original curve are not shown. Note that the colors of the lines do not represent the same misfits for each of the data selections. The WH data selection shows a significantly better fit than the Dall data selection. The MS and MS2 data selection, however, show little improvement from the WH selection. The shear wave velocity with depth profiles in Figure 21.1-21.4 are directly linked to the modelled curves in Figure 20.1-20.4. Upon inspection of the modelled profiles of the WH data selection, a transition depth is found of approximately 55 meters, where the shear wave velocity jumps from approximately 300 m/s to 410 m/s. The same depth and velocities are found when inspecting the results from virtual source positions along the line. A transition depth of 55 meters correlates with the transition of the Sterksel formation into the Stramproy formation shown in Figure 3 (see Figure 24).

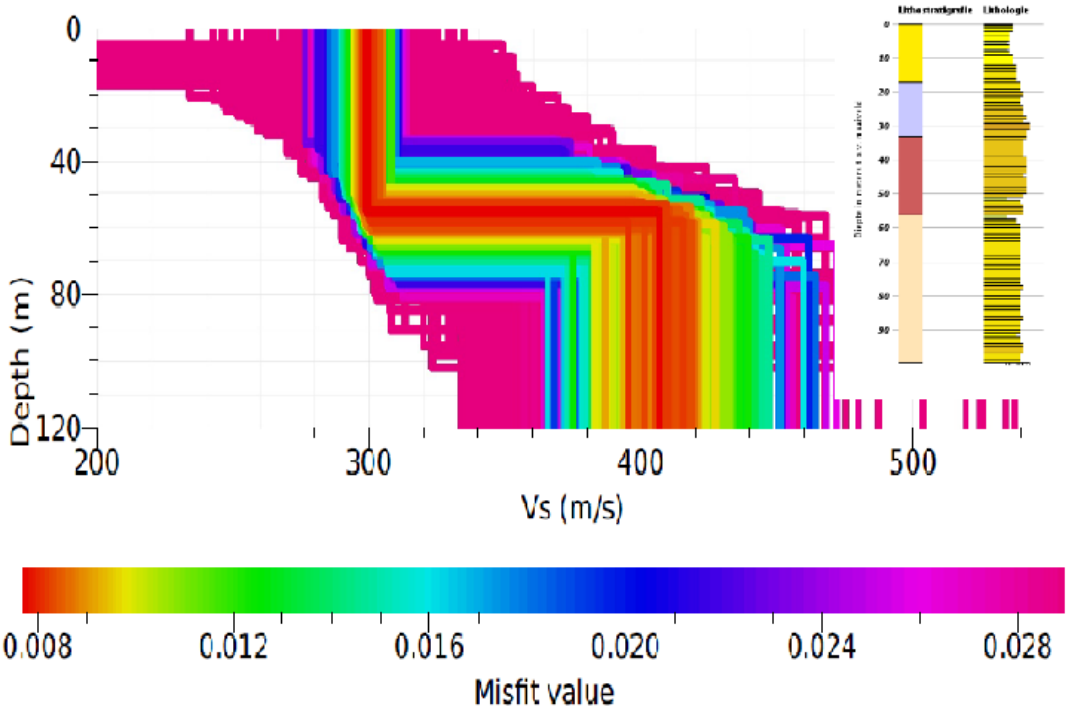


Figure 24: Shear wave velocity with depth profiles linked to the curve fits in Figure 20.3 (WH data selection) compared to borehole survey results (Figure 3). The transition found using inversion at a depth of ±55m is linked to the transition of the Sterksel Formation (dark brown) into the Stramproy formation (light brown).
Adapted from www.dinoloket.nl/ondergrondgegevens

Shear-wave velocities allow for the computation of subsurface properties such as rigidity and shear modulus, which are key factors in construction and related subsurface activities.

6 Conclusions

Ambient noise seismic interferometry (ANSI) has proven to be an effective and efficient method for obtaining surface wave dispersion curves in the studied area. Inversion of these dispersion curves results in shear-wave velocity models, which can be used to give much needed insight into shallow subsurface properties.

Autocorrelations produced during this research have resulted in two-way travel times of deep reflecting layers that correlate quite well with those found with active seismic exploration. As the ANSI method requires a relatively small fraction of the cost of an active seismic survey, it could be a valuable addition to future exploration. A combination of passive and active data allows for broadband subsurface analysis in the frequency spectrum.

When stacking autocorrelation panels, making a selection of data to be used does not appear to improve the quality of the results. Cross-correlation panels however, show stronger surface wave signals during different hours of the week. Several data selections have been made, the most efficient one including the data recorded between 08:00-20:59 on Mondays through Fridays. A manual data selection improves the signal further, but without a significant improvement in the final results.

Inversion of the dispersion curves generated from cross-correlated ambient noise panels has resulted in shear-wave velocity profiles for the shallow subsurface (up to ±200m). A velocity transition was found at a depth of ± 55 m for the Working Hours (WH) data selection, which correlates with the transition of the Sterksel formation into the Stramproy formation as found in borehole studies. When inverting the dispersion curves extracted from stacks of all recorded data (Dall data selection) no fixed transition depth is found, accentuating the importance of selecting the right data.

The shear-wave velocities found in this study can be used to compute rigidity and shear modulus of the subsurface which are vital properties for the construction industry.

All in all ANSI is an extremely effective and low-cost method for obtaining shear-wave velocity profiles (and therefore valuable subsurface properties), as well as for locating deep reflecting layers.

7 Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement No. 608553 (Project IMAGE).

8 References

BP Nederland b.v., (1982). Active seismic survey; identification L2BPE1981A, from http://www.nlog.nl/nlog/requestData/nlogp/seismiek2D/metaData.jsp?table=SmcSurvey&id=1618710213

Claerbout, J. F. (1968). Synthesis of a layered medium from its acoustic transmission response. *Geophysics*, *33*(2), 264-269.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, *19*(90), 297-301.

Dinoloket, (2017). Geological borehole research; identification B58A0122 (1985), from https://dinoloket.nl/ondergrondgegevens

Dinoloket, (2017). Geological borehole research; identification B58A0160 (1993), from https://dinoloket.nl/ondergrondgegevens

Google, (2017). Satellite imagery of De Grote Peel national park. Retrieved March 15, 2017, from https://www.google.nl/maps/@51.3441905,5.7831941,2522m/data=!3m1!1e3?hl=en

Haskell, N. A. (1953). The dispersion of surface waves on multilayered media. *Bulletin of the seismological Society of America*, *43*(1), 17-34.

Prasad, M., Zimmer, M. A., Berge, P. A., & Bonner, B. P. (2004). Laboratory measurements of velocity and attenuation in sediments. *Near surface geophysics: SEG Investigations in Geophysics Series*, *13*, 71-87.

Ruigrok, E., Gibbons, S., & Wapenaar, K. (2016). Cross-correlation beamforming. *Journal of Seismology*, 1-14.

Sambridge, M. (1999). Geophysical inversion with a neighbourhood algorithm—II. Appraising the ensemble. *Geophysical Journal International*, *138*(3), 727-746.

Wapenaar, K. (2004). Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation. Physical review letters, 93(25), 254301.

Wapenaar, K., Draganov, D., Snieder, R., Campman, X., & Verdel, A. (2010). Tutorial on seismic interferometry: Part 1—Basic principles and applications.*Geophysics*, *75*(5), 75A195-75A209.

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

Appendix 1

Data included in 'Manual selection 1' (MS)

Day	Hours included
March 17th, 2016	8 9 10 11 12 13 14 15 16 17 18
March 21st, 2016	8 9
March 23rd, 2016	13 14 15 16 17
March 24th, 2016	8 9 10 11 14
March 30th, 2016	8 9 10 11 12 13 14 15 16 17
March 31st, 2016	8 9 10 15 18
April 3rd, 2016	8 10 11 15 16 17
April 4th, 2016	8 9 10 11 12 13 14 15
April 5th, 2016	8 9 10 11 12 13 14 15 16 17 18
April 6th, 2016	8 9 10 12 13 14 15 16
April 7th, 2016	8 9 10 11 12 13 14 15 16 17 18
April 8th, 2016	8 9 10 11 12 13 14 15
April 9th, 2016	13 14 15 16 17
April 10th, 2016	8 9 10 11 13 14 15 16 17 18
April 11th, 2016	8 14
April 16th, 2016	8 9 10 12 13 14 15 16 17 18 19 20
April 17th, 2016	8 9 10 11 12 13 14 15 16
April 18th, 2016	8 12 13 14 15 16 17 18
April 21st, 2016	8 9 10 11 12 17 19
April 22nd, 2016	8 9 10 11 12 13 14 15 16 17 18

Data included in 'Manual selection 2' (MS2)

Day	Hours included
March 17th, 2016	8 9 10 11 12 13 14 15 16 17 18
March 30th, 2016	8 9 10 11 12 13 14 15 16 17
April 16th, 2016	8 9 10 12 13 14 15 16 17 18 19 20
April 17th, 2016	8 9 10 11 12 13 14 15 16
April 22nd, 2016	8 9 10 11 12 13 14 15 16 17 18