Inversion of Teleseismic Polarization Data for Crustal Velocities in Norway

Anne Drottning

Thesis for the degree
Master of Science

Department of Earth Science
University of Bergen
May 31, 2017

Abstract

To characterize the crustal and upper mantle structure it is possible to perform a joint inversion of receiver functions and apparent S-velocities. The inversion is performed using a linearized weighted iterative least-squares inversion and estimates the S-velocity structure. By inverting for two data types the non-uniqueness and the non-linearity of the solution is reduced.

In this thesis the inversion method is automated to be applicable on large data sets through testing on synthetic data. The tests were done on different inversion parameters, to find the definitions giving the best results. The most important inversion parameters were the weights, the starting model and how the number of iterations was determined. Several situations for each of these were compared to find the definitions which automated and optimized the inversion. After the tests were done on the synthetic data, they were shortly repeated on real data from a single station. This confirmed that the inversion parameters found were optimal for real data as well. After that the inversion was performed on data from stations in Scandinavia.

The results for Scandinavia was compared to results from other studies, showing that this automated joint inversion provide well resolved results displaying complex structures without smoothing. Issues with the estimated results were only apparent when the data had a high signal-to-noise ratio, or if the maximum amplitude peak in the receiver function did not correspond to the Moho discontinuity.

Acknowledgements

First I would like to thank my supervisor Stphane Rondenay for all the help during the thesis. Further, I would like to thank my co-supervisor Christian Schiffer for helping me understand and work further with the code used in the thesis.

Thank you to all my fellow students and friends for the taco fridays, movie marathons and beers throughout my five years at UiB.

Last but not least I would like to thank my family. Thank you to my brothers and my sister-in-law for all the laughs during stressful times and thank you mom and dad for always supporting and inspiring me.

Contents

1 Introduction

Seismology is the study of elastic waves propagating through the Earth. Recordings in time of the ground motion produced by these elastic waves are called seismograms. Seismograms are used to examine different aspects within Earth science (e.g. magnitude of earthquakes, focal mechanism and imaging Earth structure). There exist many approaches to characterize the seismic structure beneath stations. Lei and Zhao (2005) used seismic tomography and vanderLee and Nolet (1997) inverted surface waves. In this study I will focus on two methods: receiver functions and apparent S-velocities.

By using several recordings at one station, one can compute the receiver functions for a station. Receiver functions are processed seismograms, and are mainly used in an inversion to estimate the crustal and upper mantle structure beneath stations. For receiver function inversion, the relative S-wave velocities are well constrained and usually layer depths are well resolved. The main issues with receiver function inversion is the non-linearity and non-uniqueness of the solution.

Svenningsen and Jacobsen (2007) computed apparent S-velocities for stations. The apparent S-velocities are computed using the different wavelengths of the propagating elastic waves and describe the development of S-velocities for the crust and upper mantle. Apparent S-velocity inversion resolves the absolute S-velocities for the subsurface. The non-linearity and non-uniqueness of these inversion results are smaller than for receiver function inversion.

These two types of inversion can be combined in a weighted linearized least-squares iteration algorithm. By combining the two in a joint inversion, the non-linearity and non-uniqueness are reduced and the estimated model parameters are better constrained.

The main objective of this thesis is to perform extensive tests on synthetic data using joint inversion of receiver functions and apparent S-velocity, because this has not been done to date. The goal of the tests is to adjust the inversion parameters so the inversion results are of high quality and the inversion method can be automated and applied to large data sets. The joint inversion is then applied to a set of data from stations in and around Norway, so my results can be compared to previous results for this area.

This thesis contains ten chapters and an appendix. Chapter 1 is the introduction. Here the motivation and explanation of the work done in the thesis is provided. In Chapter 2 I will explain essential background information about receiver functions, including how they are defined and computed. In Chapter 3 I define the apparent velocity and explain how it is computed. In

Chapter 4 I derive the weighted linearized least-squares iteration algorithm from the simple definition of an inverse problem. Here, I also provide insight into receiver function inversion and apparent S-velocity inversion. Chapter 5 provides a summary of the inversion method used, and a description of the synthetic and real data. In Chapter 6 I go through the extensive testing of the inversion method on the synthetic data. In Chapter 7 I show some of the same tests done with the synthetic data, but this time performed using data from a single station. In Chapter 8 I describe the results obtained when the automated inversion was performed on data from stations in and around Norway. In Chapter 9 I compare the results described in Chapter 8 to other inversion results for the same area. Chapter 10 provide a conclusion on the quality of the joint inversion method. In the last chapter, I also give some ideas on how the method can be further developed. In the Appendix, I include material provided through the thesis like mathematical derivations and non-essential tests.

2 Receiver Functions: Theory

2.1 Introduction

Analysis of teleseismic receiver functions is a method which has been used for approximately 50 years, to obtain information about the Earth's structure and velocities.The first to introduce the concept of using receiver functions to determine the structure beneath stations was Phinney (1964). He analyzed the spectrums for long-period body waves by investigating their frequency behaviour and comparing them to the spectrums for earth models. Work on receiver functions was furthered by Langston (1977) and Vinnik (1977). Langston (1977) estimated the effective source function and deconvolved it from the observations. Then the observed amplitude ratios were compared to theoretical ratios determined for different crustal models. He furthered his work in Langston (1979), by deconvolving the instrument response and the effective source time function from the theoretical displacement response to generate what he referred to, for the first time, as receiver functions. This technique has similarities to the technique used with spectral ratios. However the deconvolution was an improvement, because the phase information was conserved. Since these starting breakthroughs, the method of receiver functions have been further developed to become a major tool for investigating lateral variation of discontinuities and velocities in the crust and upper mantle.

In this thesis, the method to estimate velocity structures beneath stations is based on receiver functions and apparent S-wave velocities. Because of this, there will here be subchapters to describe and lay the basis for understanding receiver functions, and Chapter 3 will describe the apparent S-velocity. I will also present how to compute receiver functions and describe some of the ways they have been used.

2.2 Defining Receiver Functions

Receiver functions are computed using body waves, which are teleseismic waves traveling through the interior of the Earth. Teleseismic waves are defined in several different ways, but can be generally defined as waves which travel long distances. In a scientific definition, teleseismic waves are defined using the epicentral distance [°] of the waves. Langston (1977) defined teleseismic waves as waves which have an epicentral distance of more than 30°. For the method I have used in my thesis, teleseismic waves are defined as waves with an epicentral distance between 20°and 105°.

To understand the concept of teleseismic receiver functions, it is first necessary to understand converted seismic waves. When teleseismic waves en-

counter a discontinuity in seismic properties, the energy of the waves is partitioned to satisfy the boundary conditions at the discontinuity. The energy is partitioned into transmitted and reflected waves, which have various polarizations. This energy partitioning is controlled by the reflection (R) and transmission (T) coefficients, controlled by the type of discontinuity. This can vary between a free surface, liquid-liquid, liquid-solid or solid-solid interface. For this study, the case of interest is discontinuities which represent a solid-solid interface, also called a welded interface. For a welded interface, the boundary conditions demand that the components of the displacement and the stress vectors are continuous. The first boundary condition is to prevent disturbances of mass or voids at the boundary, and the second boundary condition restricts the stress tensor values. When the polarization of the waves change at the discontinuity, we say that the wave is converted.

As teleseismic waves travel through the Earth, they take different paths from the source to the receiver. The arrival of a teleseismic wave is called a phase and describe the path of the wave. One distinguish between primary and converted phases, where primary phases have the same polarization throughout the path. When working with receiver functions, we focus on the converted phases. Ideally these converted phases should be undisturbed by other phases. To satisfy this we mainly look at the P-coda and limit the depth to the crust and upper mantle. At teleseismic distances, the P-coda remains mostly undisturbed from the primary phases, yet it still contains the converted phases.

For welded boundaries, an incident P-wave is reflected and refracted partly as P-waves and party as converted reflected and refracted SV-waves. In Figure 1 the interface separates two layers of different seismic properties. It is seen here that there are no SH-waves present in Figure 1. This is because of the particle motion of the incident P-wave is confined in the plane shown, and no refractions at the interface produces motion in the x_2 plane, which is where the particle motion of SH-waves reside. The particle motion of SV-waves are in the same plane as for the P-wave. Conversions between P and SV are called mode conversions.

The reflection and transmission coefficients are the ratios of the amplitudes after the wave has been converted, over the amplitude of the incident wave. For simplicity I will describe the reflection and transmission coefficients for a welded boundary. These coefficients are quite complicated, so the equations are given in Table 1. The energy is partitioned to fit: $T - R = 1$. The reflection coefficient varies from -1 to $+1$ and the transmission coefficient varies from 0 to 2 depending on how the acoustic impedance change over the interface. The acoustic impedance is defined by $I = \rho\alpha$ and is related to the reflection and transmission coefficients by the following equations:

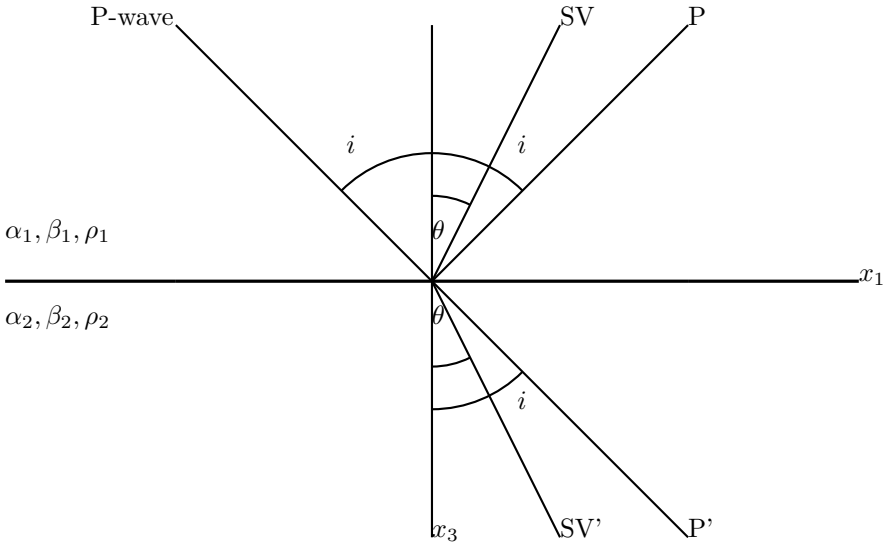


Figure 1: Incident P-wave on a welded boundary and the rays for waves generated at the interface. α is the P-wave velocity for the layer, β is the S-wave velocity for the layer and ρ is the density for the layer (Lay and Wallace (1995)).

$$R = \frac{\rho_1\alpha_1 - \rho_2\alpha_2}{\rho_1\alpha_1 + \rho_2\alpha_2}$$
$$T = \frac{\frac{1}{\alpha_1}}{\frac{1}{\alpha_2}} \frac{2\rho_1\alpha_1}{\rho_1\alpha_1 + \rho_2\alpha_2}$$

From these equations it is clear that the impedance contrast between the layers is significant in determining the partitioning of the wave energy. If the impedance contrast is large, then most of the waves will be reflected. And if the impedance contrast is small, most of the waves will be refracted. In other words, if the layers are similar in seismic properties, the waves will move through the boundaries more easily.

2.3 Calculation of Receiver Functions

When teleseismic waves arrive at seismic stations all around the Earth, they are registered in a three component system representing North-South, East-West and the vertical direction. This is denoted a N-E-Z system. To compute receiver functions, we want to isolate the different wavefields of the teleseismic waves. In a N-E-Z system, the phases and their associated wavefields are not well isolated. We therefore rotate the N-E-Z system to a radial,

Table 1: Reflection and Transmission coefficients for a solid-solid boundary modified from Lay and Wallace (1995). The subscripts PP and PS mean incident P-wave reflected as either a P wave or a converted SV wave. p denotes the ray parameter, and η denotes the slowness.

Coefficient	Formula
R_{PP}	$\left[(b\eta_{\alpha_1} - c\eta_{\alpha_2})F - (a + d\eta_{\alpha_1}\eta_{\beta_2})Hp^2\right]/D$
R_{PS}	$-\left[2\eta_{\alpha_1}(ab + cd\eta_{\alpha_2}\eta_{\beta_2})p(\alpha_1/\beta_1)\right]/D$
T_{PP}	$\left[2\rho_1\eta_{\alpha_1}F(\alpha_1/\alpha_2)\right]/D$
T_{PS}	$\left[2\rho_1\eta_{\alpha_1}Hp(\alpha_1/\beta_2)\right]/D$
$a = \rho_2(1 - 2\beta_2^2p^2) - \rho_1(1 - 2\beta_1^2p^2)$	$E = b\eta_{\alpha_1} + c\eta_{\alpha_2}$
$b = \rho_2(1 - 2\beta_2^2p^2) - 2\rho_1\beta_1^2p^2$	$F = b\eta_{\beta_1} + c\eta_{\beta_2}$
$c = \rho_1(1 - 2\beta_1^2p^2) + 2\rho_2\beta_2^2p^2$	$G = a - d\eta_{\alpha_1}\eta_{\beta_2}$
$d = 2(\rho_2\beta_2^2 - \rho_1\beta_1^2)$	$H = a - d\eta_{\alpha_2}\eta_{\beta_1}$
$D = EF + GHp^2$	

transverse and vertical (R-T-Z) system. In the R-T-Z system the different wavefields are well isolated and it is possible to examine them separately. This rotation is done according to Equation 1 from Rondenay (2009).

$$\begin{pmatrix} R \\ T \\ Z \end{pmatrix} = \begin{pmatrix} -cos\gamma & -sin\gamma & 0 \\ sin\gamma & -cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} N \\ E \\ Z \end{pmatrix} \tag{1}$$

In Equation 1, γ is the back azimuth of the incident ray. The back azimuth is the angle measured from the local North to the direction from which the energy arrives at the station. This can be seen in Figure 2. For the R-T-Z system the wavefields are separated so the incident wavefield is isolated on the vertical (Z) component. The wavefield containing the P to SV conversions is isolated on the radial (R) component and the SH wavefield is contained on the transverse (T) component. For this study, isotropy is assumed so only the radial and vertical components are of interest.

For a non-normal incident P-wave there will be some signal leakage between the radial and vertical component in the R-T-Z system. To avoid this leakage, the system can be rotated further, into the L-Q-T system or the P-SV-SH system (Rondenay (2009)). To perform the rotation to the L-Q-T system, an estimate of the near-surface P-wave velocity is required. The incident P-wave is isolated on the L-component, which is determined based on the polarization of this incident wave. The Q-component is perpendicular to the L-component, and the T-component is the transverse component also found in the R-T-Z system. For the P-SV-SH system, the polarization

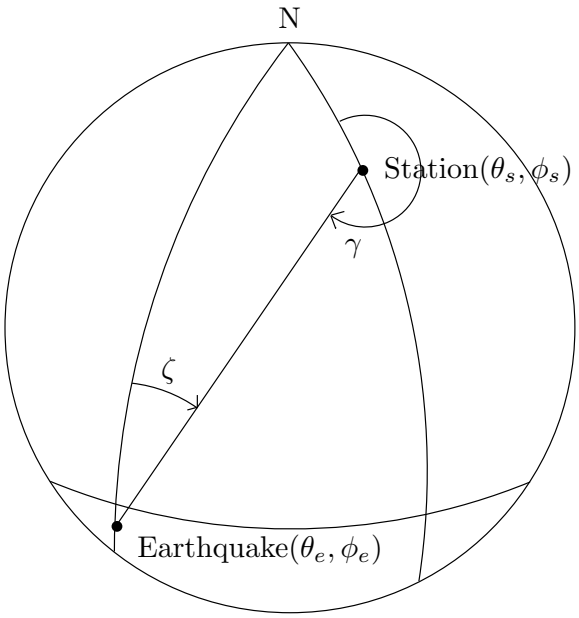


Figure 2: Figure describing the azimuth ζ and the back azimuth γ. θ is the latitude and ϕ is the longitude. Modified from Stein and Wysession (2003).

direction of the waves determine the components. For this study the R-T-Z system is used for two reasons.

First, the rotation to the R-T-Z system requires only the back azimuth and not any assumption or estimation of velocities. Secondly, the angle between the incident wave and the radial component is used in this study.

To obtain the receiver functions, the converted data needs to be source normalized. The signals are normalized to ensure high signal-to-noise ratio through deconvolution of the incident P-wave from the converted wavefield. In other words, the signals are source normalized by deconvolving the vertical component from the radial component. There are several different ways to perform deconvolution. For this thesis I will use the deconvolution technique from Rondenay (2009) which finds the receiver function ($\hat{r}$) through spectral division with a damping factor (δ). The equation is:

$$\hat{r} = \frac{d(\omega)w^*(\omega)}{w(\omega)w^*(\omega) + \delta} \tag{2}$$

Where $d(\omega)$ is the recorded signal, $w(\omega)$ is the combined source time function and the instrument response and ω denotes the frequency. The damping factor (δ) correspond to the pre-event noise which is found as the root mean square (rms) of the signal before the P-arrival.

To further increase the signal-to-noise ratio, the receiver functions can be stacked. Before the receiver functions can be stacked it is necessary to apply a moveout (timing) correction to each trace. For P-to-SV conversions, one looks at the time delay between the incident P-wave and the converted SV-wave. This delay time can be found through Equation 3.

$$T_{P_s}(p,h) = \int_{z=0}^{z=h} \left(\sqrt{\frac{1}{\beta^2(z)} - p^2} - \sqrt{\frac{1}{\alpha^2(z)} - p^2}\right) dz \tag{3}$$

Where p is the ray parameter of the incident wave, h is the conversion depth, β is the S-wave velocity and α is the P-wave velocity. The moveout correction is defined as the difference between the time delay of the observed signal and the time delay calculated for a reference ray parameter (p_0):

$$\Delta T_{P_s}(p,h) = T_{P_s}(p,h) - T_{P_s}(p_0,h) \tag{4}$$

The moveout correction is then applied to the data in the stacking of n traces as seen in Equation 5:

$$R(t) = IFT\Big[\sum_{k=1}^{n} \hat{r}_k(\omega) e^{i\omega \Delta T_{P_s}(p_k, h(t))}\Big] \tag{5}$$

where IFT is the Inverse Fourier Transform. These stacked traces form the basis for 1-D receiver function imaging.

2.4 How Receiver Functions can be Used

Receiver functions are used to image the structure or the velocities beneath a single seismic station or an array of seismic stations. There exist several studies where receiver function analysis have been applied to estimate velocity structures beneath seismic stations, some of which have already been mentioned. With relevance to the study performed in this thesis, I will also mention the studies of Ammon et al. (1990) and Xu et al. (2007).

Ammon et al. (1990) developed an inversion technique to estimate velocity structure directly from receiver functions. In this inversion they assumed a fixed layer thickness between 2 and 3 km resulting in a large number of layers. This resulted in velocity models where, in many cases, the Moho was difficult to determine because the velocity profiles were smoothed. Xu et al. (2007) used receiver functions to find 1-D S-wave velocity models beneath the southeastern Tibetan Plateau. They used the inversion method of Ammon et al. (1990), but modified the crustal model to be coarser. For

this coarser model, also a grid search was applied over the parameters. This caused the model to not have a fixed number of layers or fixed layer thicknesses.

For this study, I will use receiver functions to compute the apparent S-wave velocities corresponding to the station, and then invert the two in a joint inversion to estimate the S-velocity structure. Computation of the apparent S-wave velocities is described in Chapter 3. The math behind the inversion is described in Chapter 4 and then the method is summarized in Chapter 5.

3 Apparent Velocity: Theory

3.1 Introduction

The apparent velocity is the velocity a plane wave appears to travel at along a horizontal surface, while the true velocity is the actual velocity of the wave through a medium. For multiple layers of various velocities, the apparent velocity is frequency dependent. The apparent velocity can be used in different ways to characterize seismic properties of the substructure. Ghayamghamian and Nouri (2007) used the apparent velocity to estimate torsional ground motion, while Svenningsen and Jacobsen (2007) used the apparent S-wave velocity to invert for S-velocity structure.

Svenningsen and Jacobsen (2007) noted the three following good reasons for computing the apparent S-velocity:

1. The apparent velocity curves are direct indicators of expected absolute S-velocities at different depths.
2. The apparent velocity curves can be inverted for estimating S-velocity structures.
3. The apparent velocity can work as an additional constraint in receiver function inversion, to reduce the non-uniqueness.

In this study, the apparent S-velocity is used in a joint inversion with the radial receiver function. Here I will describe how to compute the apparent S-velocity and how it can be used to characterize the substructure.

3.2 Defining Apparent Velocity

To define the apparent velocity one needs to assume an incident P-wave traveling through multiple layers of different seismic properties, before being reflected at a free surface. The reflection of the incident P-wave is split into a reflected P-wave and a converted SV-wave. The apparent velocity is based on the thickness of the multiple layers versus the wavelength of the waves. The wavelength (λ) of a sinusoidal wave is related to the frequency (f) through Equation 6.

$$\lambda = \frac{V}{f} = V \cdot T \tag{6}$$

where V denotes the medium velocity and T denotes the period of the wave. From Equation 6 one can see that the apparent velocity depends on the frequency.

For increasing frequencies we get decreasing wavelengths. The apparent velocity is computed by averaging the velocities of the layers within the wavelength used. This means that if the wavelength is shorter than the thickness of the first layer, then the computed apparent velocity equals the true velocity of the first layer. When the wavelength is greater than the thickness of the first layer, another velocity is introduced and used in the computation. So for wavelengths surpassing multiple layers, the apparent velocity is estimated using the seismic properties of all the included layers.

3.3 Computing Apparent Velocity

There are two main ways to compute the apparent velocity: through travel time curves or by the incidence angle. When computing the apparent velocity using travel time curves, it is found as $\frac{1}{s}$ where s denotes the slope of the travel time curve in question. This method of computation is more challenging to automate as it depends on picking of arrival times at multiple stations. It would also require a dense station network.

Computing the apparent velocity using the incidence angle is simpler to automate. First we define the horizontal slowness of the incident P-wave using Snell's law:

$$p = \frac{\sin i}{\alpha} \tag{7}$$

according to notations in Figure 3. To relate Equation 7 and the apparent velocity we look at two points of the wavefront, seen in Figure 4. The velocity of the wavefront is defined as $\alpha = \frac{\Delta d}{\Delta t}$. Where Δd is the distance between the two points, and Δt is the time difference between the two points on the wavefront. As mentioned the apparent velocity is the velocity at which the plane wave appears to travel along a horizontal surface, which is defined mathematically as:

$$V_{app} = \frac{\Delta x_1}{\Delta t} = \frac{\Delta d / \sin i}{\Delta t} = \frac{\alpha}{\sin i} = \frac{1}{p} \tag{8}$$

To examine the relationship between the apparent and true velocities we need to look further into Equation 8, particularly the expression $V_{app} = \frac{\alpha}{\sin i}$. The incidence angle can vary between 0°and 90°, meaning the expression $\sin i$ can vary between 0 and 1. For a normal incidence wave ($i = 90°$), the apparent velocity is equal to the true velocity. For incidence angles below $90°(i < 90°)$ the apparent velocity is greater than the true velocity. As the incidence angle goes toward zero ($i \to 0°$) the apparent velocity goes

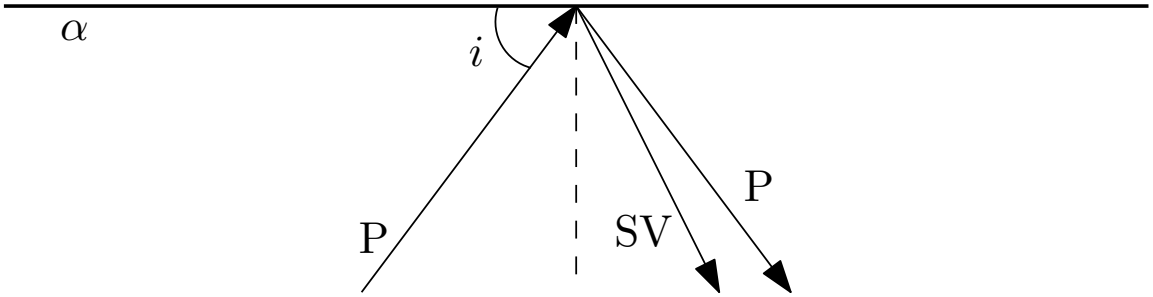


Figure 3: Incident P-wave reflected at a free surface as a P-wave and a converted SV-wave. The true incidence angle (i) is the angle between the incident P-wave and the surface, and α denotes the P-wave velocity of the medium.

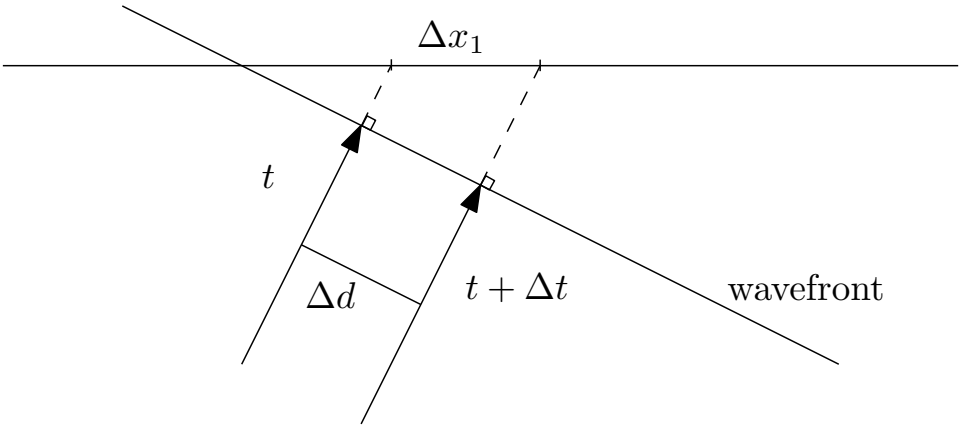


Figure 4: Geometry of two points on an incoming wavefront. Δd is the distance between the two points, and Δx_1 is the distance along the boundary. The time positions are t and $t + \Delta t$ (Pujol (2003))

towards infinity. In other words, the apparent velocity is greater or equal to the true velocity.

Svenningsen and Jacobsen (2007) suggested using another parameter called the apparent incidence angle to compute the apparent S-velocity. The apparent incidence angle (i_P) is defined by the surface particle motion occurring when an incident P-wave is reflected at a free surface, as a P-wave and as a converted SV-wave (Figure 3). The true incidence angle and the apparent incidence angle are related through Equation 9 (Nuttli and Whitmore (1961)).

$$i = \sin^{-1} \sqrt{\frac{(1 - \cos i_P) V_P^2}{2 V_S^2}} \tag{9}$$

From Equation 9 we get:

$$\frac{\sin i}{V_P} = \frac{\sin\left(\frac{1}{2}i_P\right)}{V_S} = p \tag{10}$$

where p is the horizontal slowness. See Appendix A for the derivation of Equation 10. By rearranging Equation 10 we get:

$$V_S = \frac{\sin\left(\frac{1}{2}i_P\right)}{p} \tag{11}$$

Equation 11 is used to compute the true S-velocity. However, by applying the principle here to a multilayered model, Equation 11 becomes:

$$V_{S,app} = \frac{\sin\left(\frac{1}{2}i_P\right)}{p} \tag{12}$$

To this point, we have computed the apparent S-velocity from the apparent incidence angle. So next I will describe how the apparent incidence angle is computed using data recorded over time (t). The apparent incidence angle is computed using the radial and vertical receiver functions, denoted R_{RF} and Z_{RF} respectively, by Equation 13.

$$i_P = \tan^{-1}\left[\frac{R_{RF}(t=0)}{Z_{RF}(t=0)}\right] \tag{13}$$

To gradually increase the periods at which the apparent incidence angle is computed, a low-pass filter (W) is introduced. The period of the cosine function in the filter is determined by the parameter T:

$$W(t) = \begin{cases} \cos^2\left(\frac{\pi t}{2T}\right), & for |t| < T \\ 0, & for |t| \geq T \end{cases} \tag{14}$$

By applying the cosine filter to the radial and vertical receiver functions, the apparent S-velocity can be computed for increasing depths. The apparent incidence angle is computed as a function of the low-pass filter-parameter T by:

$$i_P(T) = \tan^{-1}\left[\frac{\int_{-T}^{T} R_{RF}(\tau)\cos^2\left(\frac{\pi t}{2\tau}\right)d\tau}{\int_{-T}^{T} Z_{RF}(\tau)\cos^2\left(\frac{\pi t}{2\tau}\right)d\tau}\right] \tag{15}$$

The apparent velocity is then expressed as a function of the low-pass filter-parameter as well:

$$V_{S,app}(T) = \frac{\sin\left(\frac{1}{2} i_P(T)\right)}{p} \tag{16}$$

By increasing the period (T), the number of layers used to compute the apparent S-velocity is increased. Following this, the apparent S-velocities are estimated with more certainty for short periods. As the number of layers increases, so does the uncertainty in the computations. For all the periods, the wavelength of the P-wave is longer than the wavelength of the S-wave. When the period is longer, the wavelength of the P-wave may exceed the layer thickness, while the wavelength of the S-wave does not. For large velocity contrasts between the layers, this causes a decrease in the apparent S-velocity. A large velocity gradient may also cause interference between the incident P-wave and the converted P-to-SV wave from the Moho.

4 Least-Squares Inversion

4.1 Introduction

In geophysics, we want to understand how the Earth looks and works. To examine the Earth and it's structure, scientists monitor signals which occur either by a natural source (e.g. earthquakes, gravity, wind) or by man-made sources (e.g. exploration seismics, nuclear explosions, mining). But how can we use these observations to determine the Earth's properties? This problem is usually stated as:

$$\mathbf{d} = \mathbf{G}\mathbf{p} \tag{17}$$

where $\mathbf{d}$ denotes the observed measurements (e.g. receiver functions), $\mathbf{p}$ denotes the model parameters (e.g. the Earth model) causing the measurements and $\mathbf{G}$ denotes the model which relates the two. Equation 17 shows the forward problem, and can be rearranged to the equation for an inverse problem. Equation 17 is used to create synthetic observations based on chosen model parameters and compare them to the observed data. When solving an inverse problem one estimates the model parameters from the observations.

In this chapter, I will describe the inversion method used in this study, and how it is derived from Equation 17. The full derivation of the inversion method can be seen in Appendix B. The inversion preformed in this study is the weighted linearized least squares iterative method. I will also discuss receiver function inversion, apparent S-velocity inversion and the joint inversion combining the two.

4.2 Least Squares Inversion

When solving inverse problems, the goal is, as mentioned, to estimate the model parameters from observed data. The inverse problem can be solved as a discrete or a continuous problem. Backus and Gilbert (1970) described a method of solving the inverse problem when the model parameters are described as a continuous function, while the data is discrete. The simplest version of a solution is by rearranging Equation 17 to get

$$\mathbf{p} = \mathbf{G}^{-1}\mathbf{d} \tag{18}$$

However, Equation 18 assumes that $\mathbf{G}$ is a square nonsingular matrix, which is not the case for realistic problems. So to solve this, we need to find an

approximation to the inverse model ($\mathbf{G}^{-1}$) named the generalized inverse. The generalized inverse does not need to be square or nonsingular, but will still possess some of the properties of the exact inverse (Gupta (2011)).

There are several ways to determine the generalized inverse. I will use least-squares minimization, but another popular method is through Singular Value Decomposition (SVD) (Lines and Treitel (1984)). I will also develop further on the least-squares method to estimate the model parameters from the iterative weighted linearized least-squares solution.

To estimate the solution for the iterative weighted linearized least-squares problem, we start with the linear explicit form of the inverse problem, seen in Equation 17. In Equation 17 the vectors are defined as

$$\mathbf{d} = [d_1, d_2, \cdots, d_N]^T$$
$$\mathbf{p} = [p_1, p_2, \cdots, p_M]^T$$

where the model matrix $\mathbf{G}$ has size $M \times N$.

If $M = N$ and $\mathbf{G}$ is nonsingular, then the problem is even-determined, meaning that all the observations are independent. Usually we deal with over- or underdetermined problems. An inverse problem is overdetermined if $M < N$ meaning that the amount of data is larger than the amount of unknowns. An inverse problem is underdetermined if $M > N$, meaning that the amount of unknowns is larger than the amount of data. In many cases a problem is mixed-determined. In this case the inverse problem is not entirely over- or under-determined.

When solving the inverse problem using least-squares the goal is to minimize the length of the error. The linearized error ($\mathbf{e}$) is defined in Equation 19 and is found using Taylor series (Lines and Treitel (1984)):

$$e_i = d_i^{obs} - d_i^{est} \Leftrightarrow \mathbf{e} = \mathbf{d} - \mathbf{Gp} \tag{19}$$

where d_i^{obs} is the observed data and d_i^{est} is the estimated data. Here we use the Euclidean distance as the measure of length, though it is possible to use other measures as well (Menke (1989)). The Euclidean distance of the error in Equation 19 is defined as the L_2-norm:

$$||\mathbf{e}||_2 = \sqrt{\sum_{i=1}^{N} e_i^2}$$

The total error (E) is the squared Euclidean length, and is defined in Equation 20.

$$E = \mathbf{e}^T\mathbf{e} \tag{20}$$

To find the best fitting solution, we want to find the solution which minimizes the total error. This is found by differentiating the total error with respect to a model parameter, p_q and set the differentiated equation equal to zero to find the minimum point:

$$\frac{\partial E}{\partial p_q} = 2\sum_k^M p_k \sum_i^N G_{iq}G_{ik} - 2\sum_i^N G_{iq}d_i = 0 \tag{21}$$

In matrix notation Equation 21 becomes:

$$\mathbf{G}^T\mathbf{G}\mathbf{p} - \mathbf{G}^T\mathbf{d} = 0 \tag{22}$$

From Equation 22 it is possible to estimate the values for the model parameters ($\mathbf{p}$) by the linear least squares solution:

$$\mathbf{p}^{est} = [\mathbf{G}^T\mathbf{G}]^{-1}\mathbf{G}^T\mathbf{d} \tag{23}$$

Because the problem is rarely well-determined (i.e. $\mathbf{G}^T\mathbf{G}$ is square), but often singular, we also need to regularize the solution. This regularization means that we need to add a damped prediction error ($\epsilon^2 L$) to the total error (E). The damping is implemented through the damping factor (ϵ), which can be determined in different ways. A common method is by trial-and-error testing to find the damping factor which gives the smallest Euclidean length for the model parameters and the estimated data (Aster and Thurber (2013)). The total error is denoted Q_{tot} and defined as:

$$Q_{tot} = E + \epsilon^2 L = \mathbf{e}^T\mathbf{e} + \epsilon^2\mathbf{p}^T\mathbf{p} \tag{24}$$

To find the best damped least squares solution we minimize Equation 24 in the same way as for the error in Equation 20. The result can be seen as:

$$\frac{\partial Q_{tot}}{\partial p_q} = 2\sum_k^M p_k \sum_i^N G_{iq}G_{ik} - 2\sum_i^N G_{iq}d_i + 2\epsilon^2\sum_k^M p_k = 0 \tag{25}$$

In matrix notation Equation 25 becomes:

$$\mathbf{p}\mathbf{G}^T\mathbf{G} - \mathbf{G}^T\mathbf{d} + \epsilon^2\mathbf{I}\mathbf{p} = 0 \tag{26}$$

The estimated damped least squares solution can be found from Equation 26 and is defined as:

$$\mathbf{p}^{est} = [\mathbf{G}^T\mathbf{G} + \epsilon^2\mathbf{I}]^{-1}\mathbf{G}^T\mathbf{d} \tag{27}$$

There is still an uncertainty in the estimated model parameters when solving Equation 27. To reduce this uncertainty it is reasonable to introduce some *a priori* information. *A priori* information is knowledge about the model parameters from models or previous studies and is frequently used to constrain the solutions.

When introducing *a priori* knowledge it is also reasonable to weight the errors. This weighting makes it possible to determine whether the solution should be optimized for the estimated data, or the estimated model parameters compared to the *a priori* knowledge. To implement *a priori* information and weights, the total error is redefined from Equation 24 into:

$$Q_{tot} = \mathbf{e}^T\mathbf{W}_e\mathbf{e} + \epsilon^2[\mathbf{p} - \langle\mathbf{p}\rangle]^T\mathbf{W}_p[\mathbf{p} - \langle\mathbf{p}\rangle] \tag{28}$$

In Equation 28 the *a priori* information about the model parameters is denoted $\langle\mathbf{p}\rangle$ and the weights are implemented through the matrices $\mathbf{W}_e$ and $\mathbf{W}_p$. To find the best weighted damped least squares solution with *a priori* knowledge, we minimize Equation 28 the same way as previously done. The differentiated error in matrix notation is:

$$\frac{\partial Q_{tot}}{\partial p_q} = \mathbf{p}\mathbf{G}^T\mathbf{G}\mathbf{W}_e - \mathbf{G}^T\mathbf{d}\mathbf{W}_e + \epsilon^2\mathbf{W}_p[\mathbf{p} - \langle\mathbf{p}\rangle] = 0 \tag{29}$$

From Equation 29 we can then estimate the model parameters:

$$\mathbf{p}^{est} = \langle\mathbf{p}\rangle + [\mathbf{G}^T\mathbf{W}_e\mathbf{G} + \epsilon^2\mathbf{W}_p]^{-1}\mathbf{G}^T\mathbf{W}_e[\mathbf{d} - \mathbf{G}\langle\mathbf{p}\rangle] \tag{30}$$

All the solutions above have been for linear inverse problems. However, in realistic cases the inverse problems are nonlinear. For nonlinear inverse problems Equation 17 becomes:

$$\mathbf{d} = \mathbf{G}(\mathbf{p}) \tag{31}$$

When solving nonlinear (and linear) inverse problems it is common to linearize the problem. The linearization comes from how we want to optimize a model from the starting model. The model is perturbed iteratively to

minimize the error. To derive the solution for the iterative linearized inverse problem, we define some relationships:

$$\mathbf{x} = \begin{bmatrix}\mathbf{d} & \mathbf{p}\end{bmatrix}^T$$
$$C_x = \begin{bmatrix}C_d & C_{dp} \\ C_{pd} & C_p\end{bmatrix}$$

where C denotes covariance matrices, for the data and the model parameters the covariance matrices are C_d and C_p respectively. C_{dp} and C_{pd} are the cross-covariance matrices between the data and the model parameters. The vector $\mathbf{x}$ has the length $N + M = S$. The weighting of the problem will be implemented through the covariance matrices, which are defined from the variances. These notations are valid if the data ($\mathbf{d}$) and the *a priori* model parameters ($\langle\mathbf{p}\rangle$) have Gaussian distribution and we have the implicit constraint that $\mathbf{f}(\mathbf{d}, \mathbf{p}) = \mathbf{f}(\mathbf{x}) = 0$, where $\mathbf{f}$ has a length of $n \leq S$. Using these notations the total error is defined as:

$$Q_{tot} = [\mathbf{x} - \langle\mathbf{x}\rangle]^T C_x^{-1} [\mathbf{x} - \langle\mathbf{x}\rangle] \tag{32}$$

To find the best solution for an iterative linearized inverse problem, we minimize the error in Equation 32. Since Equation 32 has a constraint ($\mathbf{f}(\mathbf{x}) = 0$), the minimization needs to be done with consideration to the constraint. To differentiate an equation which has one or more constraints, we use Lagrange multipliers:

$$\nabla(Q_{tot} + \lambda \cdot \mathbf{f}(\mathbf{x})) = 0 \Leftrightarrow \frac{\partial Q_{tot}}{\partial x_i} + \sum_j^n \left(\frac{\partial \lambda f_j}{\partial x_i}\right) = 0 \tag{33}$$

where λ is the Lagrange multiplier. The differentiation in Equation 33 is done with respect to a parameter x_q, giving the equation:

$$2[\mathbf{x} - \langle\mathbf{x}\rangle]^T C_x^{-1} = -\lambda^T \mathbf{F} \tag{34}$$

where $\mathbf{F}$ is a matrix of partial derivatives ($F_{ji} = \sum_j^n \frac{\partial f_j}{\partial x_i}$). We need to remove the Lagrange multipliers in Equation 34, and solve the resulting equation simultaneously as $\mathbf{f}(\mathbf{x}) = 0$ is minimized. This is done by Equation 35.

$$[\mathbf{x} - \langle\mathbf{x}\rangle] = C_x \mathbf{F}^T [\mathbf{F} C_x \mathbf{F}^T]^{-1} \big(\mathbf{F}[\mathbf{x} - \langle\mathbf{x}\rangle] - \mathbf{f}(\mathbf{x})\big) \tag{35}$$

Making Equation 35 an iterative equation gives:

$$\mathbf{x}_{i+1}^{est} = \langle \mathbf{x} \rangle + C_x \mathbf{F}_i^T [\mathbf{F}_i C_x \mathbf{F}_i^T]^{-1} \big(\mathbf{F}_i [\mathbf{x}_i^{est} - \langle \mathbf{x} \rangle] - \mathbf{f}(\mathbf{x}_i^{est}) \big) \tag{36}$$

where i is the iteration. Equation 36 shows an iterative method to estimate $\mathbf{x}$. However, we want to estimate just the model parameters $\mathbf{p}$ which is possible if the constraint equation is written out explicitly: $\mathbf{f}(\mathbf{x}) = \mathbf{d} - \mathbf{G}(\mathbf{p}) = 0$, and the *a priori* model parameters are uncorrelated (Menke (1989)). To rewrite Equation 36 to fit the explicit constraint, we define the following relationships from Tarantola and Valette (1982a):

$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} \mathbf{d} & \mathbf{p} \end{bmatrix}^T \\ \mathbf{F} &= \begin{bmatrix} \mathbf{I} & -\mathbf{G} \end{bmatrix} \\ C_x &= \begin{bmatrix} C_d & C_{dp} \\ C_{pd} & C_p \end{bmatrix} \end{aligned} \tag{37}$$

Here $\mathbf{G} = G_{ij} = \frac{\partial g_j}{\partial p_i}$, meaning $\mathbf{G}$ is the Jacobian matrix for the data with respect to the model parameters. By assuming that the uncertainties in the data are independent of the uncertainties in the model parameters, the covariance matrix C_x can be simplified (Tarantola and Valette (1982a)):

$$C_{dp} = (C_{pd})^T = 0 \Rightarrow C_x = \begin{bmatrix} C_d & 0 \\ 0 & C_p \end{bmatrix}$$

Ideally the inversion should work without *a priori* information. Instead of defining *a priori* model parameters, we define a starting model of model parameters ($\mathbf{p}_0$). In Equation 35 the difference between the estimated $\mathbf{x}$ and the *a priori* $\mathbf{x}$ is computed, and then in Equation 36 it is defined in the second term. Now, we want the second term to define the difference between the iterations. Equation 36 then becomes the weighted linearized least-squares iterative algorithm:

$$\mathbf{p}_{i+1}^{est} = \mathbf{p}_i + [\mathbf{G}_i^T C_d^{-1} \mathbf{G}_i + C_p^{-1}]^{-1} \big(\mathbf{G}_i^T C_d^{-1} (\mathbf{d} - \mathbf{g}(\mathbf{p}_i^{est})) + C_p^{-1} (\mathbf{p}_i^{est} - \mathbf{p}_0) \big) \tag{38}$$

To simplify Equation 38 we define:

$$\begin{aligned} \mathbf{M} &= [\mathbf{G}^T C_d^{-1} \mathbf{G} + C_p^{-1}]^{-1} \\ \Delta\mathbf{d} &= \mathbf{d} - \mathbf{g}(\mathbf{p}^{est}) \\ \Delta\mathbf{p} &= \mathbf{p}^{est} - \mathbf{p}_0 \end{aligned}$$

which transform Equation 38 to:

$$\mathbf{p}_{i+1} = \mathbf{p}_i + \mathbf{M}[\mathbf{G}^T C_d^{-1} \Delta\mathbf{d} + C_p^{-1} \Delta\mathbf{p}] \tag{39}$$

Following the relationships shown in Equation 37 the total error in Equation 32 can be written as:

$$Q_{tot} = \Delta\mathbf{d}^T C_d^{-1} \Delta\mathbf{d} + \Delta\mathbf{p}^T C_p^{-1} \Delta\mathbf{p} \tag{40}$$

In the next sections we look at how the general inversion technique described is applied to receiver functions and apparent S-velocities.

4.3 Receiver Function Inversion

Receiver functions are dominated by Ps conversion phases and reverberation phases caused by discontinuities in the crust and the upper mantle. Because they are dominated by shear-conversions, receiver functions are sensitive to discontinuities causing S-wave velocity changes. In many cases one uses Poisson's ratio to determine the P-wave velocity from the S-wave velocities (e.g. Ottemoller and Midzi (2003)). In receiver function inversion the main issues are with the non-linearity of the inversion and the non-uniqueness of the solution (Ammon et al. (1990)).

The non-uniqueness and the non-linearity of the solution for receiver function inversion are due to how small changes of velocity in the starting model causes large differences in the results (Jacobsen and Svenningsen (2008)). These differences appear as a consequence of minimizing Equation 40 and converging to a local minima, and not the global minimum. Ammon et al. (1990) reduced the non-uniqueness by adding *a priori* information. Jacobsen and Svenningsen (2008) showed that increasing the number of iterations for the inversion, reduces the non-uniqueness.

Changes in the velocity cause changes in the travel times, which alters the arrival times of deeper conversions. These altered arrival times are the main reasons for non-linearity in receiver function inversion (Jacobsen and Svenningsen (2008)). Jacobsen and Svenningsen (2008) proved through non-linearity analysis, that to get a smaller non-linearity error, one can parameterize the model parameters in delay times instead of depth.

Using the method described, the weighted linearized iterative least-squares receiver function inversion is performed using Equation 39, where $\mathbf{d}$ denotes the receiver function.

To summarize, receiver function inversion has issues with non-uniqueness and non-linearity, but they can both be addressed. Since the receiver functions are sensitive to velocity contrasts for shear-converted waves, the inver-

sion results estimate relative S-wave velocities, especially in the near-surface layers.

4.4 Apparent S-Wave Velocity Inversion

Apparent S-wave velocities are computed from the receiver functions and contain information on the variations in the S-wave velocities beneath stations. Due to these velocity variations, the apparent S-velocity inversion results constrain the absolute S-velocities in the crust and uppermost mantle (Svenningsen and Jacobsen (2007)). Svenningsen and Jacobsen (2007) parameterized the model parameters in delay time to reduce the non-linearity. The non-uniqueness in apparent S-velocity inversion is small, so inversion results without any *a priori* information are still very good. The non-linearity of the inversion is much more reduced than for receiver function inversion, and the main source of non-linearity is in the determination of V_P/V_S ratio in the model layers.

The weighted linearized iterative least-squares inversion described in Equation 39 is performed on just the apparent S-wave velocity when $\mathbf{d}$ denotes the apparent S-velocities.

4.5 Joint Inversion of Receiver Functions and Apparent S-Wave Velocities

A combination of receiver function inversion and apparent S-wave velocity inversion is good, because their constraints complement each other. As mentioned the receiver function inversion constrains the relative S-velocities, while the apparent S-velocity inversion constrains the absolute S-velocities. The combination of these two will reduce the non-uniqueness without necessarily including *a priori* information. The non-linearity is reduced by parameterization in delay times for the model parameters.

The inversion is performed using Equation 39 where $\mathbf{d}$ denotes both the receiver function and the apparent S-velocities on the form $[\mathbf{RF}, \mathbf{V_{S,app}}]$.

For this study, I will use joint inversion of receiver functions and apparent S-wave velocities, to estimate S-velocity structures beneath stations.

5 Method and Data

5.1 Overview

In this chapter I will go through the method and the data used in this thesis. The method was explained in detail in Chapter 4, and will here be summarized. The method has been tested on both synthetic and real data. The main objective for testing on synthetic data was to improve and automate the method, so that the inversion can be applied to large data sets.

Here I describe the synthetic model, and how it was chosen, and the creation of the corresponding synthetic data. Further, I describe the real data used for testing and the data set where the automated inversion method was applied.

5.2 Method

The inversion method described in Chapter 4 is a weighted linearized iterative least squares inversion. I use this method in a joint inversion of receiver functions and apparent S-velocities. The joint inversion is aiming to reduce the non-uniqueness and non-linearity associated with the method. It is also interesting to see how the solution compares to other inversion methods.

5.3 The Synthetic Data

To create realistic synthetic data I used averages for seismic parameters in Norway. The synthetic S-velocity structure can be seen in Figure 5 and the corresponding model parameters can be seen in Table 2. To create the synthetic data I defined the following:

1. Layer depths
2. P- and S-velocities
3. Densities

To determine the layer depths I used the CRUST1.0 model [1] for Norway, and used the maximum Moho depth. I also included a mid-crustal layer depth from CRUST1.0 where the sediments and the crystalline crust is separated. For the seismic velocities I used global averages from Christensen

[1] Downloaded for free at `https://igppweb.ucsd.edu/~gabi/crust1.html`

and Mooney (1995), while the densities were estimated using the Poisson-Birch assumption ($\rho = 320 \cdot V_P + 770$) defined in Svenningsen and Jacobsen (2007).

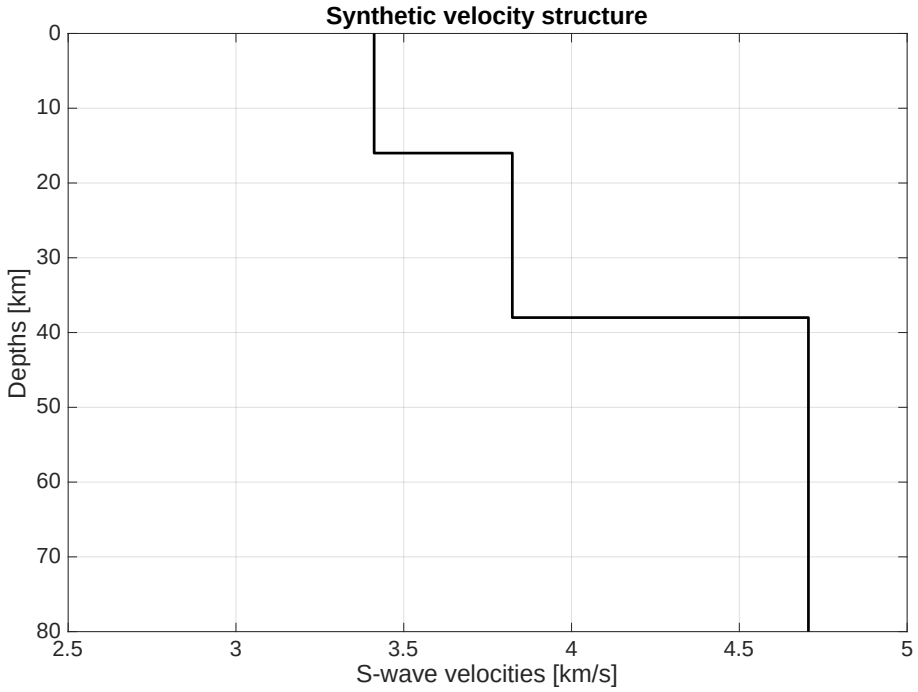


Figure 5: The S-wave velocity structure of the synthetic data. The model parameters are listed in Table 2.

Table 2: Values used for the synthetic structure. The layer depths correspond to the bottom of the layer.

Layer	Depth [km]	$V_P[km/s]$	$V_S[km/s]$	$\rho[kg/m^3]$
1	16	5.8	3.412	2626
2	38	6.5	3.824	2850
3	Mantle	8.0	4.706	3330

To create the synthetics I used the free software RAYSUM written by Andrew Frederiksen[2]. RAYSUM is used to create synthetic traces based on velocity structure, back-azimuths, ray parameters and the Gaussian pulse-width, and was applied to the structure described in Figure 5 and Table

[2]Available on `https://home.cc.umanitoba.ca/~frederik/Software`

2 assuming the layers to be horizontal, homogeneous and isotropic. The back-azimuths were ranging from 0°to 360°with a step of 10°, and the ray parameters were ranging from 0.04 km/s to 0.08 km/s randomly distributed among the synthetic events. The Gaussian pulse-width was set to 0.8 seconds so the discontinuities would not be averaged over depths.

The synthetic traces were source normalized according to Equation 2, with a damping factor of 100. A high damping factor reduces the possible noise while still showing the discontinuities in the receiver functions. The source normalization, or deconvolution, is performed in the frequency domain and the signals are converted back to the time domain afterwards. The synthetic receiver functions were then scaled, so the maximum amplitude of the vertical component were normalized to 1. The scaling was performed using Equation 41.

$$r(t) = \frac{r(t)}{\max(z(t))} \tag{41}$$

A synthetic trace is created for each back-azimuth, where the mean values of the radial and vertical receiver functions can be seen in Figure 6. From the radial receiver function in Figure 6, the Moho peak is clear at approximately 4 seconds, which fits well with a Moho depth of 38 km.

Since the method inverts for both the receiver function and the apparent S-velocity, I also computed the apparent S-velocities for the synthetic data. They were computed following the method described in Chapter 3. The synthetic apparent S-velocities can be seen in Figure 7.

The apparent S-velocities in Figure 7, range from 3.3 km/s to 5.3 km/s which fits well with the S-velocities in Table 2 considering the apparent velocities are higher or equal to the true velocities. The apparent S-velocity curves in Figure 7 also show that the computed apparent S-velocities are very stable for small periods and steadily grows more unstable for the longer periods, which is consistent with the solution stability mentioned in Chapter 3.

5.4 Real Data

5.4.1 Station IU_KONO

Further testing of the inversion method were done on real data from a single station; IU_KONO. It was chosen because it had recorded many (187) events, and so the observed receiver function was less disturbed by noise.

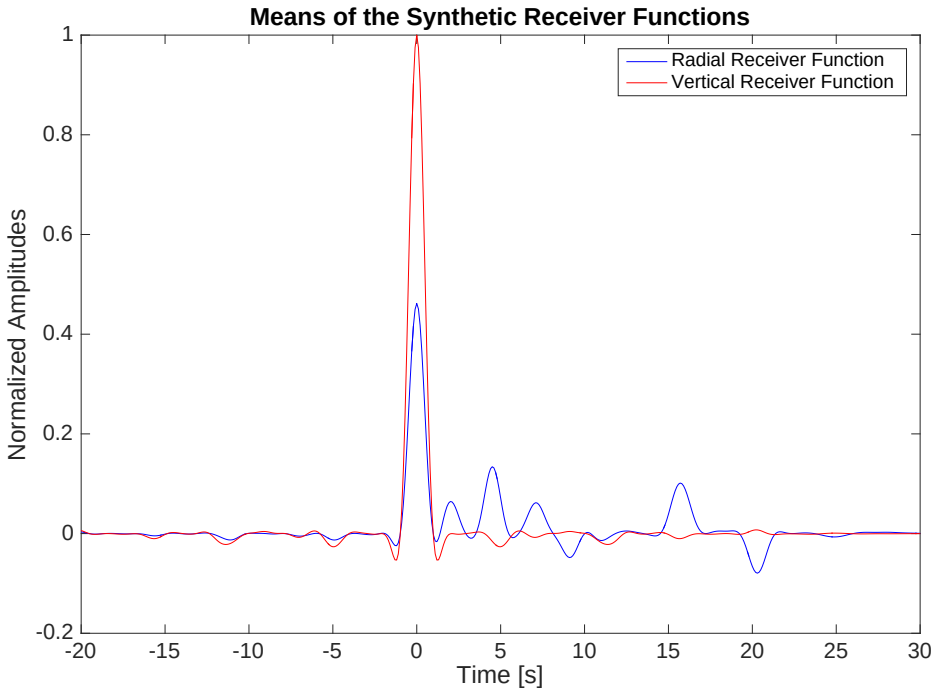


Figure 6: Mean radial (blue) and vertical (red) receiver functions corresponding to the synthetic structure described in Figure 5.

5.4.2 Scandinavian Stations

After testing the inversion method using data from a single station, I applied it to a set of stations in Scandinavia. The stations I used (Figure 8) are extracted from the GLImER database, and are located in and around Norway.

The station networks used are the Norwegian National Seismic Network (NNSN) (the stations available through orfeus), the Norwegian Seismic Array Network (NO), the Digital World-Wide Standardized Seismograph Network (DWWSSN), the Swedish National Seismic Network (SNSN), the Norway 05 network (YY05) and the Global Seismic Network (GSN). Details for the stations in Figure 8 are shown in Table 3.

The goal of applying the joint inversion to a subset of stations, is to compare my resulting velocity structures with the results of other studies. This shows whether the automated method has been successful, and could possibly be applied to a larger set of stations in the future.

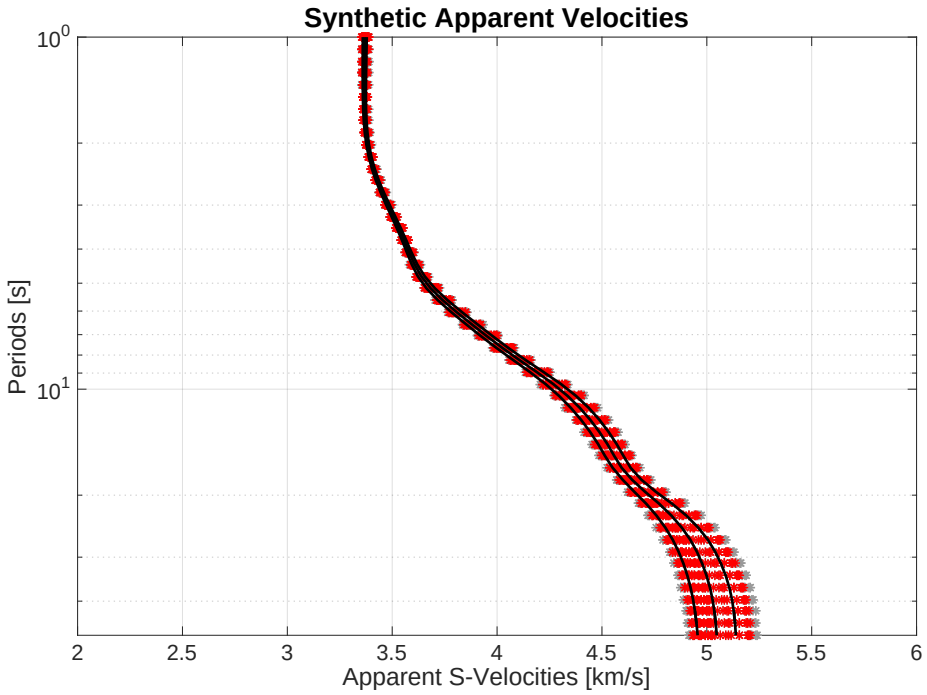


Figure 7: The apparent S-wave velocities for the synthetic events. The grey stars show all the computed apparent S-velocities, the red stars marks the 68% fraction of the estimates which are closest to the median. The mean apparent S-velocities and the mean $\pm$ the standard deviations for the 68% fraction are all shown as black lines.

The results of applying the joint inversion on the stations in Figure 8 is shown in Chapter 8.

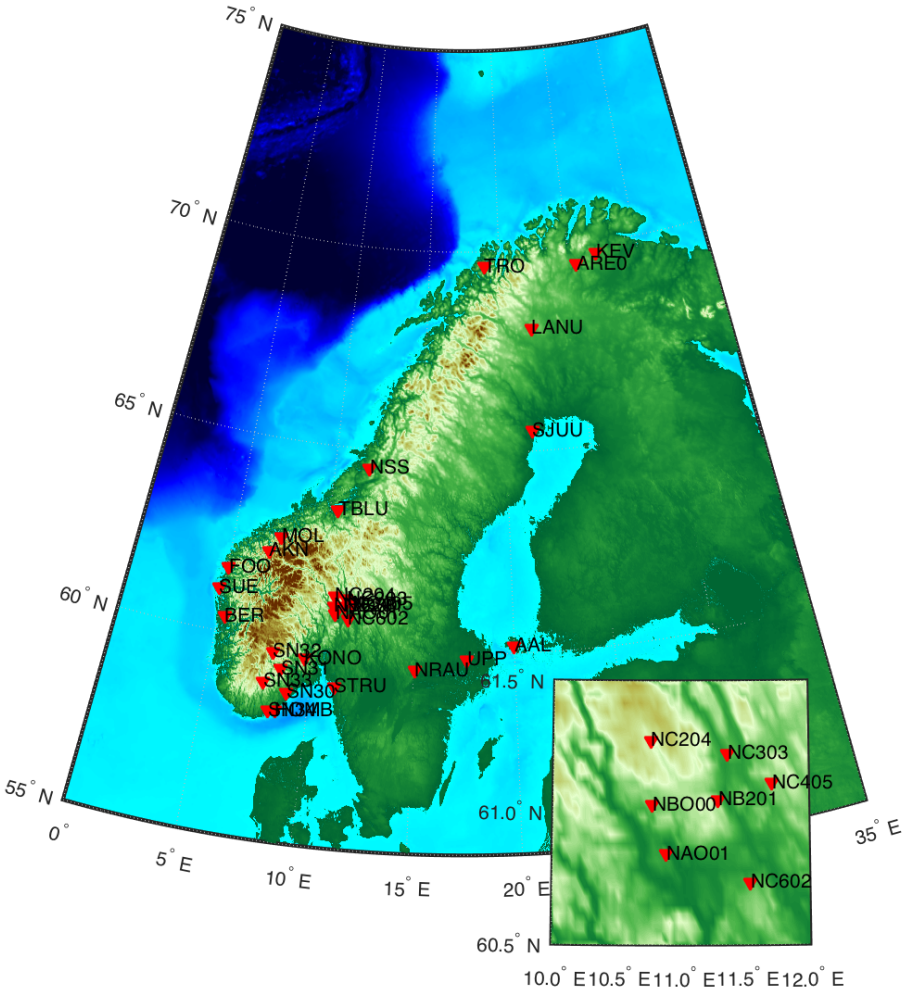


Figure 8: Map displaying the subset of stations, on which the joint inversion was performed.

Station	Latitude [°N]	Longitude [°E]	Number of recorded events	Geologic Domain
NS_BER	60.4	5.3	158	Caledonides
NS_FOO	61.6	5.0	11	Southwest Scandinavian Domain
NS_HOMB	58.3	8.5	132	Southwest Scandinavian Domain
NS_MOL	62.6	7.5	189	Southwest Scandinavian Domain
NS_NSS	64.5	12.0	7	Southwest Scandinavian Domain
NS_SUE	61.1	4.8	82	Southwest Scandinavian Domain
NS_TBLU	63.4	10.4	301	Caledonides
NS_TRO	69.6	18.9	30	Caledonides
NO_AKN	62.2	7.0	203	Southwest Scandinavian Domain
NO_ARE0	69.5	25.5	394	Archean Domain
NO_NAO01	60.8	10.9	342	Southwest Scandinavian Domain
NO_NB201	61.0	11.3	325	Caledonides
NO_NBO00	61.0	10.8	324	Caledonides
NO_NC204	61.3	10.8	351	Caledonides
NO_NC303	61.2	11.4	319	Caledonides
NO_NC405	61.1	11.7	307	Southwest Scandinavian Domain
NO_NC602	60.7	11.5	245	Southwest Scandinavian Domain
DW_KEV	69.8	27.0	25	Archean Domain
UP_AAL	60.2	20.0	151	Svecofennian Domain
UP_LANU	68.0	22.0	211	Archean Domain
UP_NRAU	59.6	15.0	111	Svecofennian Domain
UP_SJUU	65.5	21.6	94	Svecofennian Domain
UP_STRU	59.0	11.2	78	Southwest Scandinavian Domain
UP_UPP	59.9	17.6	151	Svecofennian Domain
YY05_SN30	58.8	8.9	26	Southwest Scandinavian Domain
YY05_SN31	59.3	8.5	32	Southwest Scandinavian Domain
YY05_SN32	59.7	8.0	26	Southwest Scandinavian Domain
YY05_SN33	58.9	7.8	22	Southwest Scandinavian Domain
YY05_SN34	58.2	8.2	13	Southwest Scandinavian Domain
IU_KONO	59.6	9.6	187	Southwest Scandinavian Domain

Table 3: Table of the stations used in the inversion, their coordinates, the number of events recorded on each station and the country where the station is. NS: Norwegian National Seismic Network, NO: Norwegian Seismic Array Network, DW: Digital World-Wide Standardized Seismograph Network, UP: Swedish National Seismic Network, YY05: Norway 05 and IU: Global Seismic Network.

6 Testing on Synthetic Data

6.1 Introduction

In the previous chapters I have described the joint inversion method for radial receiver functions and apparent S-velocities. Here I tested this method on synthetic data with the goal of automating the inversion, so that it can be applied to large data sets. To automate it I looked at the different inversion parameters to find out how they could improve the results, such as:

1. Weights on the receiver function and the apparent S-velocities
2. Starting model for the S-velocity structure
3. The number of iterations in the inversion
4. Constraints on the S-velocity structure

After determining the best inversion parameters, I added different levels of noise to the synthetic data, to determine if the inversion parameters were suitable for more realistic data.

6.2 The Original Inversion Parameters

Each inversion parameter was tested separately. The basic starting model was defined with five layers over a half-space. The layer depths are somewhat based on the peaks from the observed receiver functions and the seismic velocities used are global averages. The number of layers in the starting model also determines how many layers the resulting velocity structure will consist of. The inversion is set to iterate 20 times, and no constraints are set on the model parameters.

These values were used unless otherwise improved during the testing.

6.3 Choice of Weights

In a joint inversion, the weights determine how well each data set should influence the inversion. This means one weight for the receiver function and another weight for the apparent S-velocities. Higher weights means that the results are more constrained to the starting model than to the data. In other words, using a high weight on the receiver function means that the joint-inversion favoures the apparent S-velocity inversion, and vice versa. I will reference the synthetic radial receiver function and the synthetic apparent S-velocities as the observed data.

The goal of the testing, was to determine the optimal weights which provide best quality results. My testing strategy was as follows:

1. Determining if there is a maximum weight, where the results change very little if the weights are increased further.
2. Test if the results are best when the inversion focuses equally on both data sets, or if they are better when focusing on one of the data sets.
3. When using equal weights on both data sets, test if the results will benefit from increasing these weights.
4. Test to determine if there is a weighting ratio which improves the results further.

6.3.1 Maximum Weight

By determining a maximum weight, the number of weights used in the tests is limited. The initial weights I used was weights of 1 for both data types. The weights were increased on one data type at the time, from 5 to 100, with a step size of 5. For the other data type, the weight remained at 1. The inversion focuses on fitting the data type with the lowest weight to the observed data, while the data type with the highest weight is more constrained by the starting data.

First, I increased the weight on the receiver function. Figure 9 shows how the model parameters and the data error change as the weight on the receiver function is increased. The relative change (in %) is computed between two consecutive weight increases.

The total changes in the S-velocities as the weight on the receiver function was increased, plateaus at around 20% for weights above 15. For lower weights, the changes decrease. For the layer depth changes, the differences are more varying also for higher weights. The changes in the data error were increasing and varying for weights above 15. For lower weights, there was a clear decrease as the weights increased.

Next, I increased the weight on the apparent S-velocities. The results in Figure 10 show the difference in the estimated model parameters and the data error as the weight on the apparent S-velocity was increased.

For the S-velocity, the total changes remained almost constant at 2%. The changes in the depths decreased for weights towards 20, and for higher weights remained nearly constant.

There are changes in layer depths for weights below 20 are drastic. Whereas for higher weights, the changes are fairly small. The change in error varies

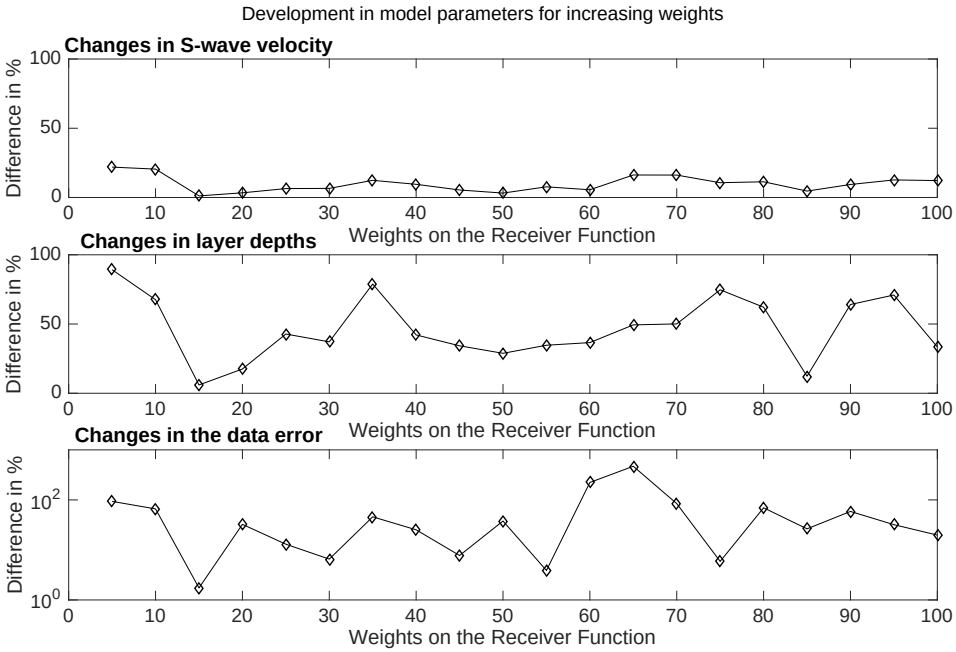


Figure 9: The difference (in %) between the estimated model parameters (S-velocity and layer depths) for increasing weight on the receiver function. The values show the total change in all the estimated S-velocities and depths. Bottom: The difference in the RMSE for increasing weights on the receiver function.

for all the weights.

While there were no very distinct maximum weight for either data type, there were indicators to help the selection. From Figure 9 the changes are more varying for weights above 15, while for weights below 15 the changes decrease dramatically. From this the maximum weight could be chosen to be 15.

From Figure 10 the changes are vary more weights above 20. From this the maximum weight could be chosen to be 20. Consequently the maximum weight was set to 20 for both data types.

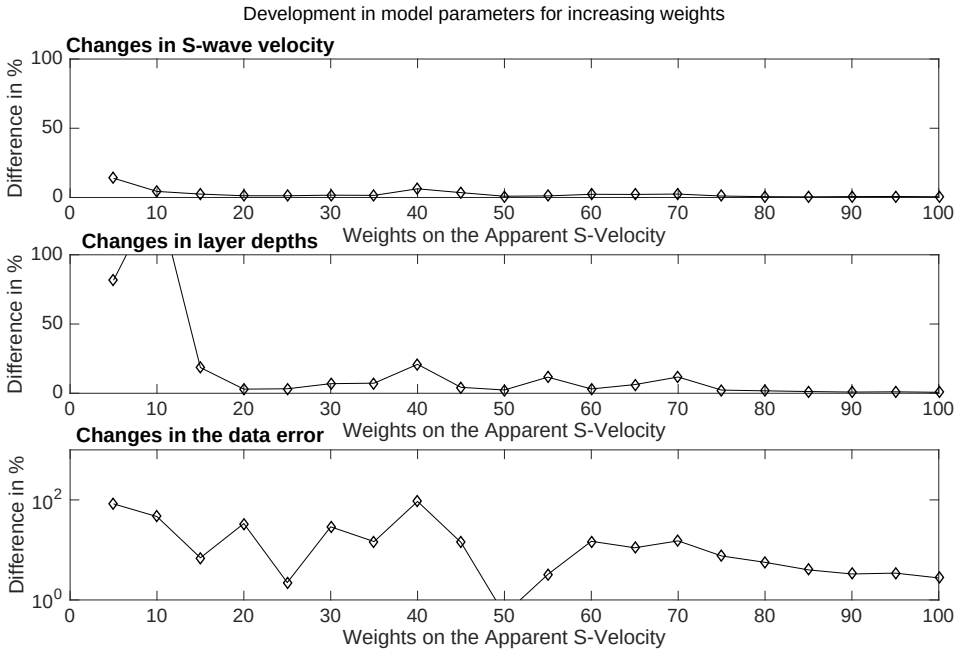


Figure 10: The difference (in %) between the estimated model parameters (S-velocity and layer depths) for increasing weight on the apparent S-velocity. The values show the total change in all the estimated S-velocities and depths. The difference in the RMSE for increasing weights on the receiver function.

6.3.2 Comparison with Synthetic Structure

In the previous section I looked at the error analysis when increasing the weight for one data type at the time. Here, I look at how well the predicted results compare to the true synthetic structure. The weight on one data type at the time was increased from 1 to 5 and then to 20 with a step size of 5. For the other data type, the weight remained at 1. To determine the quality of the results I looked at the estimated velocity structure, the fit to the data and the data error.

Similar to in the previous section, I started by increasing the weight on the radial receiver function. The fit of the model parameters when the weight on the radial receiver function is increased can be seen in Figure 11a. The estimated model parameters show a good fit to the true synthetic structure

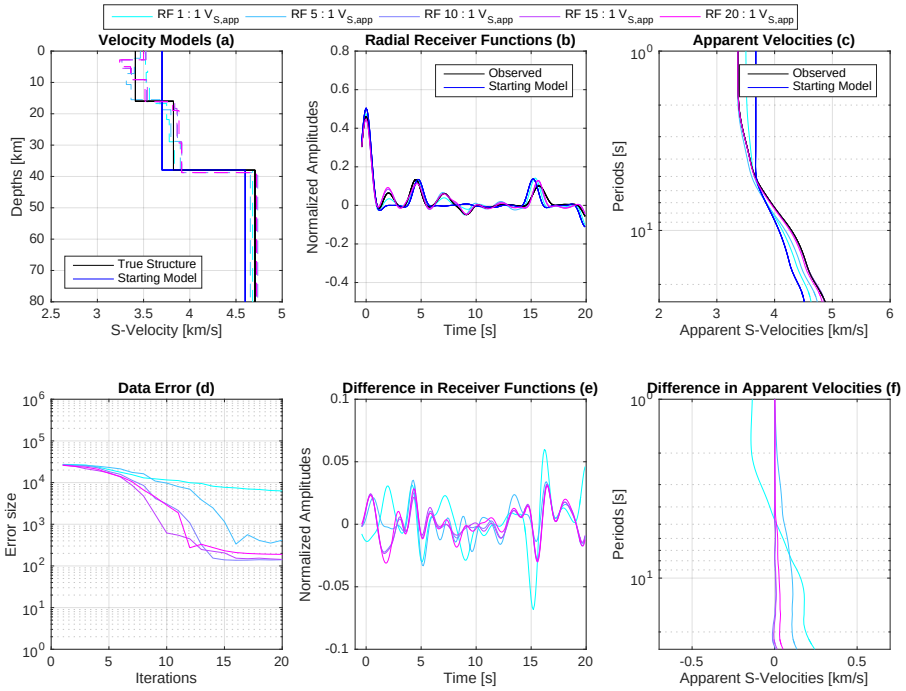


Figure 11: Results when increasing weights on the receiver function. **(a)** The true synthetic structure, the starting model and the estimated model parameters. **(b)** The fit of the estimated receiver functions to the observed and starting receiver function. **(c)** The fit of the estimated apparent S-velocities to the observed and starting ones. **(d)** The development of the data error during the iterations. **(e)** The difference between the observed radial receiver function and the estimated radial receiver functions. **(f)** The difference between the observed and estimated apparent S-velocities.

for all the weights.

To examine the data fit, I looked at Figure 11b - f. Figure 11b shows how the estimated receiver function fits the observed and the starting receiver function, while Figure 11e shows the difference between the observed and predicted data. It appears that the estimated radial receiver function is improved as the weight on the receiver function was increased.

Figure 11c displays the estimated apparent S-velocities compared to the observed and starting apparent S-velocities, while Figure 11f displays the difference between the observed and estimated data. Figure 11f show that as the weight on the receiver function is increased, the fit of the estimated

apparent S-velocity is improved.

Figure 11d displays the data error. The data error consists of the error for the radial receiver function and the apparent S-velocities. For all the weights used, the data error converges as the number of iterations increases. As the weight on the receiver function is increased, the data error decreases.

Next, I increased the weights on the apparent S-velocities (see Figure 12). For all the weights on the apparent S-velocities, the estimated model parameters fit the true synthetic structure very well.

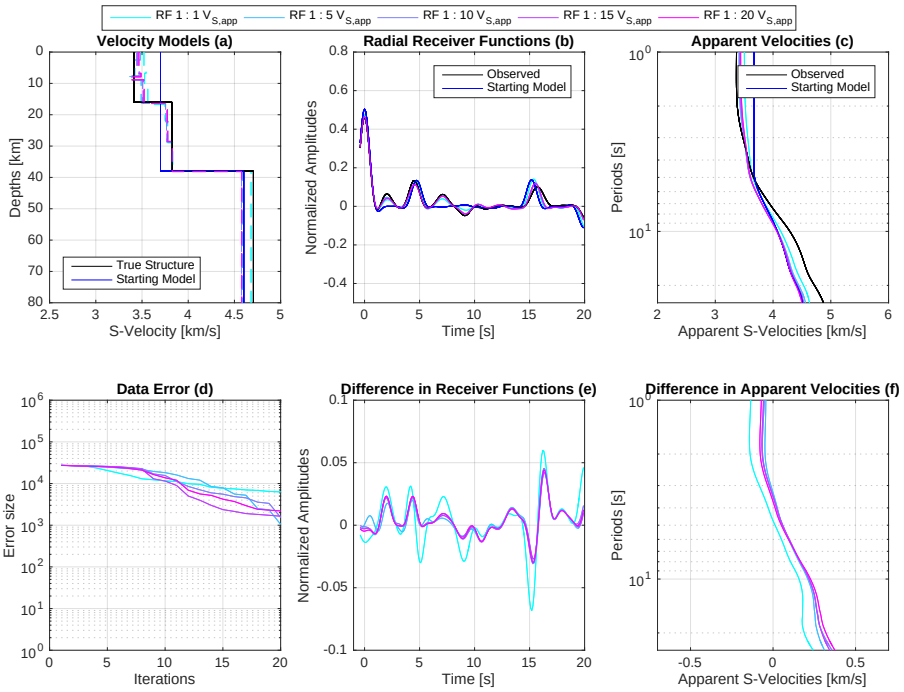


Figure 12: Results as the weight was increased on the apparent S-velocity. Explanation of plots found in description of Figure 11.

As the weight on the apparent S-velocity was increased, the difference between the observed and estimated receiver functions was decreased. The difference between the observed and estimated apparent S-velocities was small for short periods, and larger for longer periods. The data error generally decreased as the weights were increased.

As the weight on one data type was increased, the estimated results are quite similar. This is caused by the similarity between the observed and the

starting data. Because of this similarity I decided to revisit using a high weight on one data type when testing different starting models.

Otherwise it seems natural to use the weight 1 on both data types. By doing this, the non-uniqueness and non-linearity of the solution is reduced and the model parameters are estimated with the best trade-off between the data types.

6.3.3 Equal Weights on Both Data Types

When using equal weights on both data types in the inversion, the predictions should fit the observed data types equally as good. As in the previous tests, the weights were increased from 1 to 5 and then to 20 with a step size of 5. Figure 13 display the results when using equal weights on the data types, and increasing these weights.

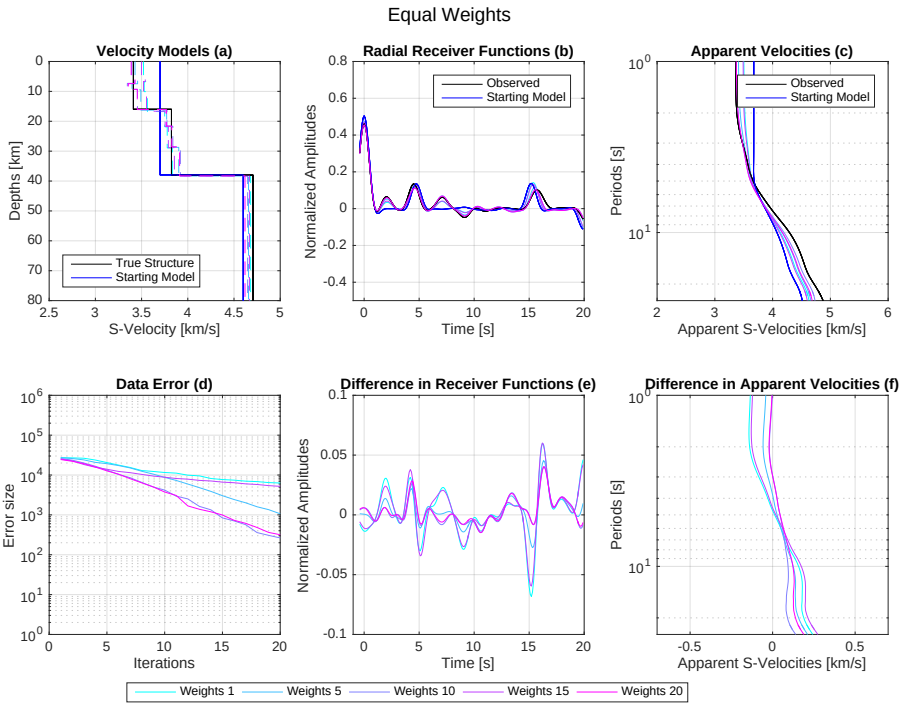


Figure 13: Results for increasing equal weights on the radial receiver function and apparent S-velocity. Explanation of each plot is found in description of Figure 11.

The estimated model parameters compared to the true and starting model

parameters are displayed in Figure 13a. Here the estimated model parameters are fit well to the synthetic structure, although the weights were increased.

The estimated radial receiver function compared to the observed and starting receiver functions are displayed in Figure 13b, and Figure 13e show the difference between the observed and estimated data. This difference decreased as the weights increased.

Figure 13c and f display the results for the apparent S-velocities as the equal weights were increased. In Figure 13f it can be seen that as the weights were increased, the difference decreased.

Figure 13d displays the data error. As the weights were increased, the data error generally decreased.

Similarly to when the weight on one data set was increased, the results when increasing both weights equally were very good. Again, the high quality of the results were caused by the similarity in the observed data and the starting data. This means that if the starting model is close to the real structure, the choice of weights is less critical.

6.3.4 Weighting Ratios Between the Data Types

When testing weighting ratios I examined using a high weight on one data type, while increasing the weight on the other data type. I used a high weight of 20, while increasing the other weight according to ratios 0.1 to 0.5.

First the highest weight was put on the radial receiver function, while adjusting the weight on the apparent S-velocity according to the ratios. The results can be seen in Figure 14.

Figure 14a compares the estimated model parameters to the true synthetic structure and the starting model parameters. Here one can see that for the increased weighting ratio, the model parameters remain well fitted to the true synthetic structure.

The radial receiver function fit is displayed in Figure 14b and e. Figure 14b displays the estimated radial receiver functions compared to the observed and starting receiver functions, while Figure 14e shows the difference between the observed and the estimated data. As the weighting ratio was increased, the difference in Figure 14e was smallest for the smallest weighting ratio (0.1).

Figure 14c displays the estimated apparent S-velocities compared to the observed and starting apparent S-velocities. The difference between the

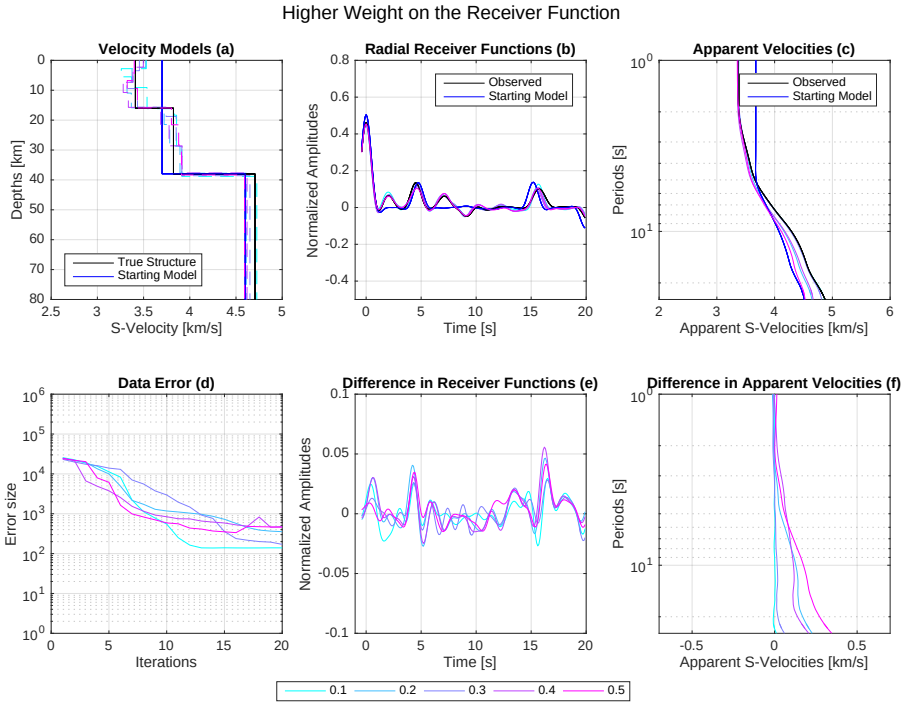


Figure 14: Results when using a high weight on the receiver function and increasing the weight on the apparent S-velocity. Explanation for each plot is found in the description of Figure 11.

observed and estimated data was smallest for the weighting ratio 0.1.

The smallest data error occurs for the smallest weighting ratio.

Next, I used the high weight on the apparent S-velocity and adjusted the weight on the receiver function according to the ratios (see Figure 15).

The estimated velocity structure is very similar to the true structure for all the weighting ratios. As the weighting ratio was increased, the difference displayed in Figure 15e decreased. As the weighting ratio was increased, the difference in Figure 15f decreased. The lowest data error occurs for the weighting ratio 0.5.

What these observations have in common is that for high weights on the receiver function, the error is small. As mentioned in the previous tests, the results are good although the weights are high because of the similar nature of the observed and starting data. This similarity is especially strong in the receiver function, where the Moho peak is almost a perfect match.

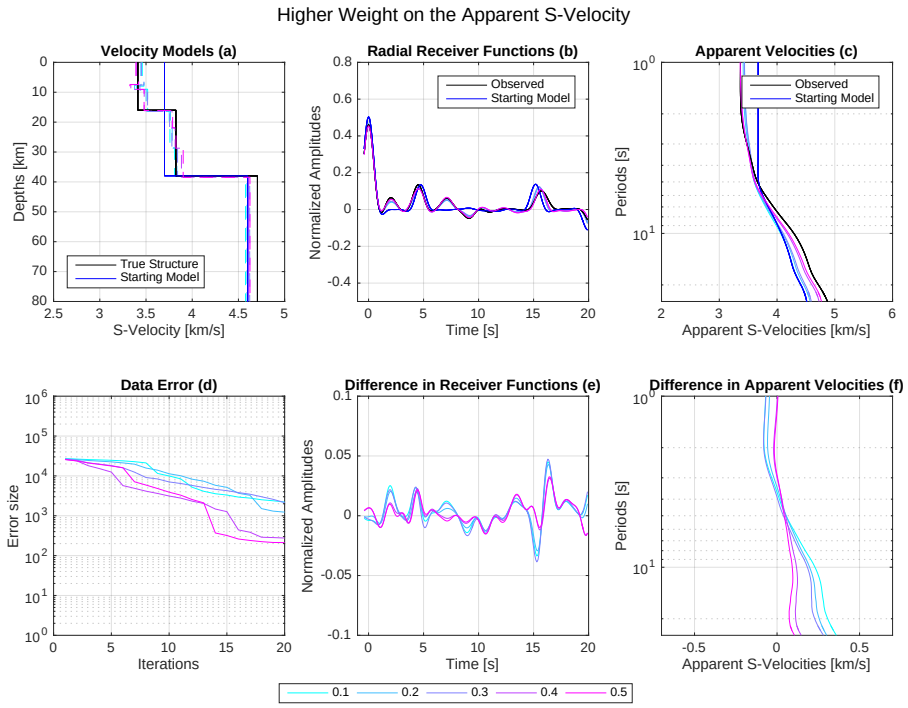


Figure 15: Results when using a high weight on the apparent S-velocities and increasing the weight on the receiver function. Description of the individual plots is the same as in the caption for Figure 11.

It is difficult to draw any conclusions regarding the choice of weights without investigating them for an alternative starting model. The tests were redone using only the extreme cases and not the steps between.

6.4 Testing Starting Models

The starting model is perturbed iteratively through the inversion, to approximate the observations. It can be used to include *a priori* information, and determine the number of estimated model parameters.

The inversion was tested using three different starting models:

1. A starting model where the delay times are picked from the observed receiver function and the crustal velocities are picked from the observed apparent S-velocities.
2. A starting model based on the global crust model CRUST1.0.

3. A starting model using global averages of the model parameters following Christensen and Mooney (1995).

Additionally I performed tests using unique starting models. The results of these can be seen in Appendix C.

While performing these tests, the weight on both of the data types were set to 1, while the other inversion parameters remained unchanged.

6.4.1 Starting Model 1: Peaks

This starting model is mainly based on observed data, where the delay times are picked from the observed receiver function and then converted to depths using the relationship from Zhu and Kanamori (2000):

$$h = \frac{\Delta t_{Ps}}{\sqrt{V_S^{-2} - p^2} - \sqrt{V_P^{-2} - p^2}} \tag{42}$$

where h is the layer thickness, Δt_{Ps} are the delay times of the Ps-phase, V_S is the S-wave velocity, V_P is the P-wave velocity and p is the ray parameter.

The median apparent S-velocity for 1 second was used for velocities above the Moho and below a global average was used. The global average was used because of the instability of computing apparent velocities for long periods. When using this starting model, the number of layers in the estimated results is equal to the number of peaks in the observed receiver function.

The results for this starting model is seen in Figure 16.

The estimated model parameters fit very well with the synthetic structure. The S-velocities fit the synthetic structure with a difference of $\pm 0.2 km/s$, and the depths with a difference of $\pm 1 km$.

The fit between observed and predicted receiver functions is very good, although the Moho peak is slightly shifted. The estimated apparent S-velocities are slightly lower than the observed apparent S-velocities. The data error converges after few iterations and remains stable.

6.4.2 Starting Model 2: CRUST

CRUST1.0 is a global crustal model for the entire Earth created by G. Laske, Z. Ma, G. Masters and M. Pasyanos. The starting model was created using average crustal values for Norway, with three layers over a half-space. Unlike in the Peaks starting model, here the depths are picked and the delay times

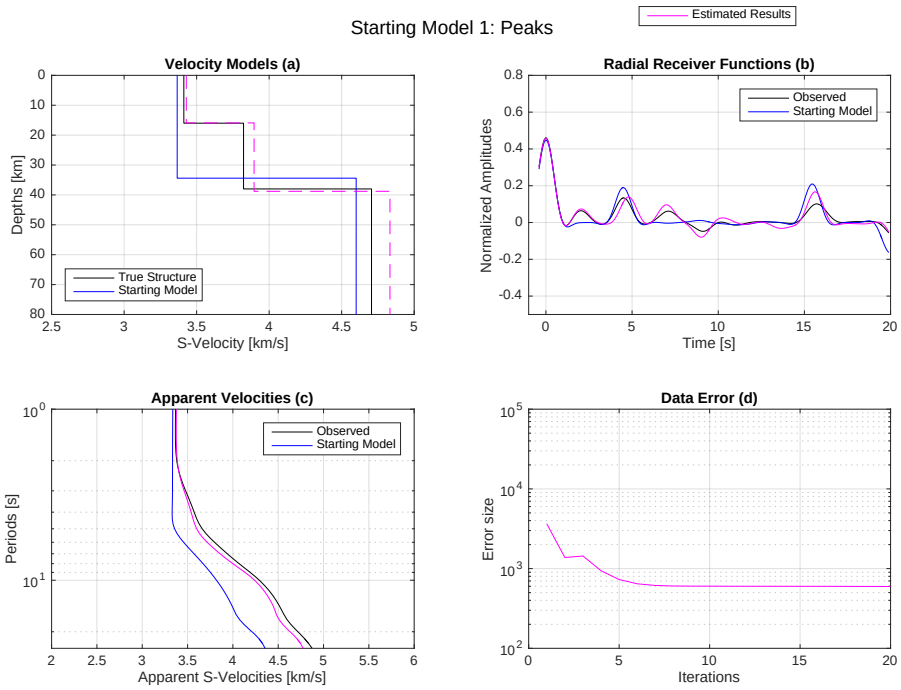


Figure 16: Results when using the starting model based on the observed data. **(a)** The estimated model parameters compared to the true synthetic structure and the starting model parameters. **(b)** The estimated radial receiver function compared to the observed and starting radial receiver functions. **(c)** The estimated apparent S-velocities compared to the observed and starting apparent S-velocities. **(d)** The data error.

are calculated by rewriting Equation 42. The results when using the CRUST starting model can be seen in Figure 17.

The estimated S-velocities are similar to the true S-velocities, with an error of $\pm 0.2 km/s$. The estimated layer depths are also well resolved, however a new layer appears at approximately $28 km$ depth. This extra layer appears because the number of layers is determined by the starting model.

The estimated receiver function fits the observations very well, even though the starting model places the Moho peak at a later time. The estimated apparent S-velocities are almost overlapping with the observed apparent S-velocities. The data error converges after 10 iterations.

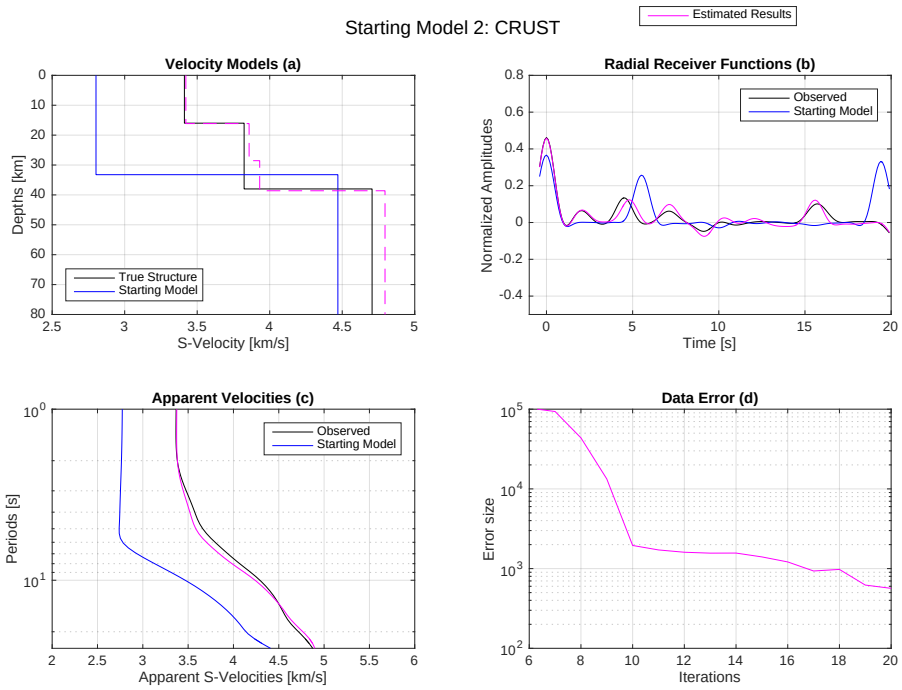


Figure 17: Results when using the starting model based on the CRUST1.0 database. Description of individual plots found in description of Figure 16.

6.4.3 Starting Model 3: Global

This starting model uses global averages for depths and velocities and assumed two layers over a half-space. I used the global averages found in Table 3 in Christensen and Mooney (1995). The results are seen in Figure 18.

The estimated S-velocities fit the true synthetic structure with an error of $\pm 0.3 km/s$. The Moho depth is well predicted, while the crustal layer is estimated too shallow.

The estimated receiver function only produces the Moho peak, while the estimated apparent S-velocities are well predicted. The data error converges after only a few iterations.

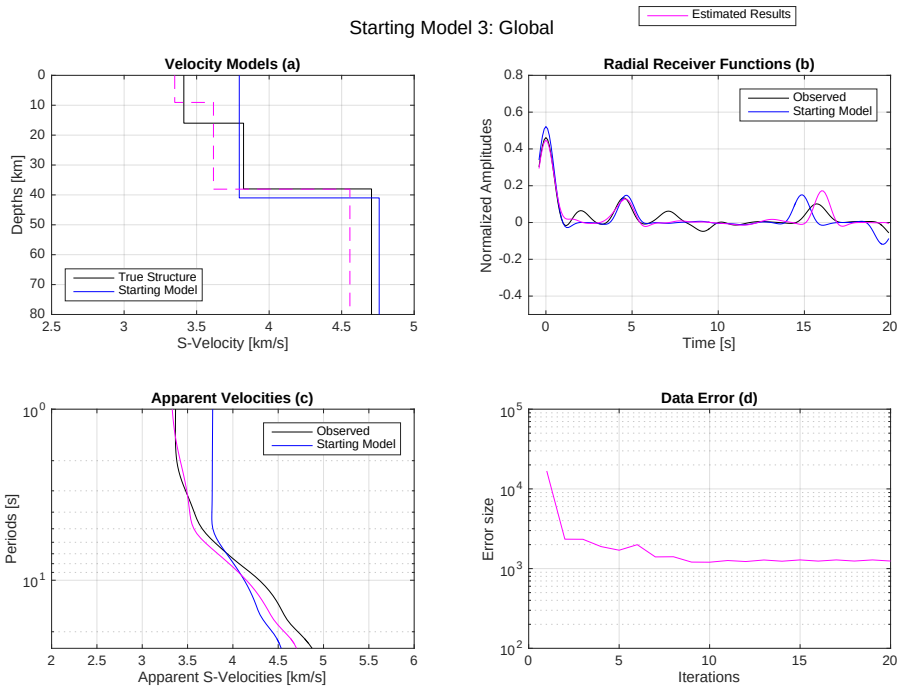


Figure 18: Results when using the starting model based on global averages. Description of individual plots found in caption of Figure 16.

6.4.4 Comparison and Results

Figure 19 shows the results for the three starting models.

The Peaks and CRUST starting models predict the best model parameters and very similar receiver functions, while the receiver function predicted by the Global starting model differ throughout. The estimated apparent S-velocities differ the least from the observations when using the CRUST starting model.

The data error now starts at different values, because the first error is computed for the starting model compared to the observed model. The overall lowest data error is for the Peaks starting model.

For all the starting models, the estimated results fit the observed data well. The best results were obtained using the Peaks and the CRUST starting models, but the data error for the Peaks starting model converges after less iterations.

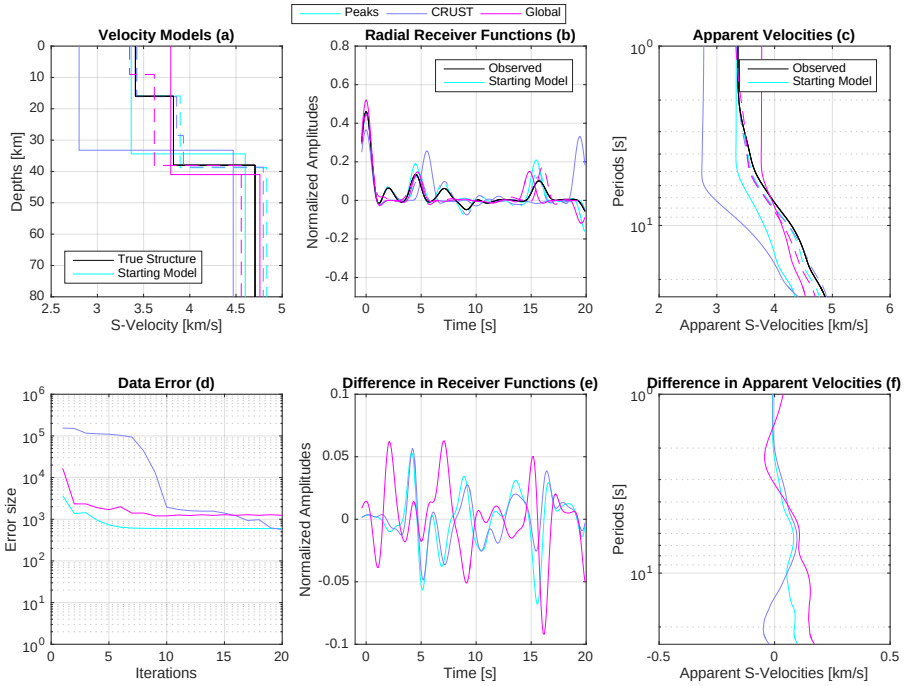


Figure 19: Results for the three starting models. Peaks is the starting model based on observations, CRUST is the starting model based on the CRUST1.0 model and Global is the starting model based on global averages. **(a)** The estimated model parameters compared to the true synthetic structure and the starting model parameters. **(b)** The estimated, observed and starting radial receiver functions for the three starting models. **(c)** The estimated, observed and starting apparent S-velocities for the three starting models. **(d)** The data error for the three starting models. **(e)** The difference between the observed and the estimated radial receiver functions. **(f)** The difference between the observed and the estimated apparent S-velocities.

The Peaks starting model was already automated to fit the observations and is not constrained to use a predetermined number of layers. Consequently the Peaks stating model will be used for the automated data analysis of this study.

6.4.5 Testing Weights for the CRUST Starting Model

The retesting of the weights was performed using the CRUST starting model, which is independent of the observations. The retesting is done shortly by just reproducing some of the results from Chapter 6.3. The tests were started by using a high weight (20) on one of the data types and a weight of 1 on the other data type.

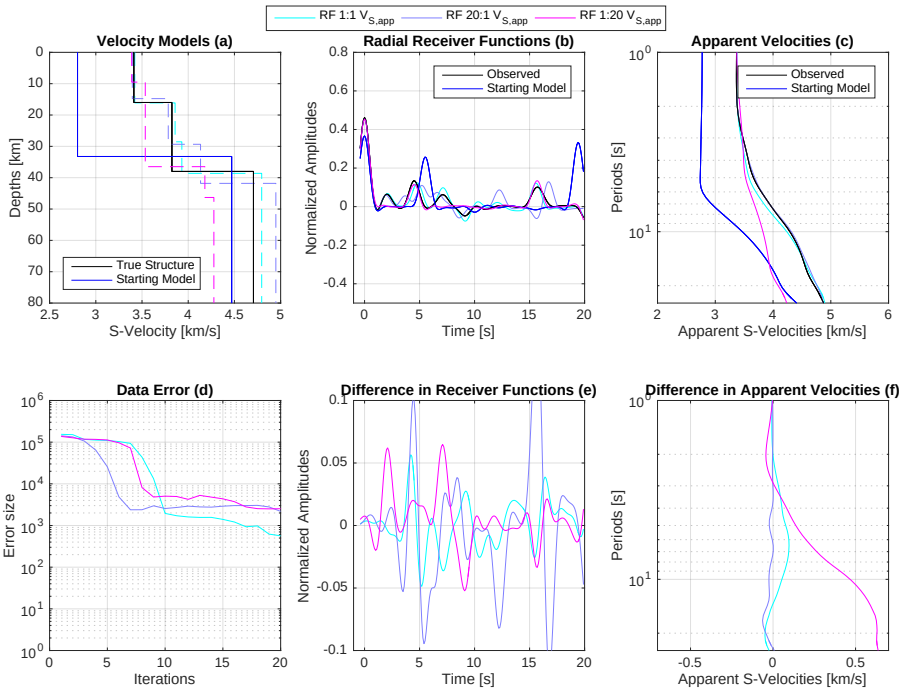


Figure 20: The results for three different weighting combinations. **(a)** The estimated model parameters compared to the true synthetic structure and the starting model parameters. **(b)** The estimated, observed and starting radial receiver functions. **(c)** The estimated, observed and starting apparent S-velocities. **(d)** The data error. The difference between the estimated and observed **(e)** radial receiver functions, and **(f)** apparent S-velocities.

Figure 20a show the resulting velocity structures. Using a high weight on the receiver function, predicts a deeper Moho and a high S-velocity in the mantle. These observations are consistent with the results in Figure 20b and c. The estimated radial receiver function fits the Moho peak of the starting model rather than the observed Moho peak, while the estimated apparent S-velocities are resolved extremely well. These observations are confirmed

in Figure 20e and f. The total data error in Figure 20d converges after less than ten iterations.

The velocity structure when using a high weight on the apparent S-velocity predict lower S-velocities and shallower layer depths.The fit of the predicted data confirm these observations. The difference for the radial receiver functions is quite small, especially for the Moho peak. For the apparent S-velocities, the difference increases for the longer periods. The data error converges after few iterations.

When using equal weights of 1, the results are superior. This stems from the difference between the starting model and the observed model.

6.5 Stopping the Iterations

Using a fixed number of iterations in the inversion does work well. However, it may cause the inversion to run too many iterations without improving the results. It is therefore possible to stop the iterations when the results are no longer improving. The following criteria can be used:

1. Changes in the model parameters
2. Change in the Root-mean square error ($RMSE$)

The changes in the model parameters (the layer depths and S-wave velocities) is defined as the relative difference between two consecutive iterations, computed using:

$$Change[\%] = \frac{|m_{i+1} - m_i|}{m_i} \cdot 100 \tag{43}$$

where m_i is the model parameters for iteration i. When the change is less than 1% for three consecutive iterations, the inversion is completed.

If the iteration causes the data error to increase, the change in the model parameters is reduced to avoid large faulty changes. Since this change might be less than 1%, the condition that the three consecutive iterations all had to be good was included.

The other method of stopping the iterations was based on the root-mean-square error ($RMSE$), which is computed using:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d - \hat{d}_i)^2} \tag{44}$$

where n is the number of samples in the data, d is the observed data and $\hat{d}_i$ is the estimated data at iteration i.

Figure 21 shows the results when stopping the number of iterations in the inversion using the methods described above, including a combination of the two.

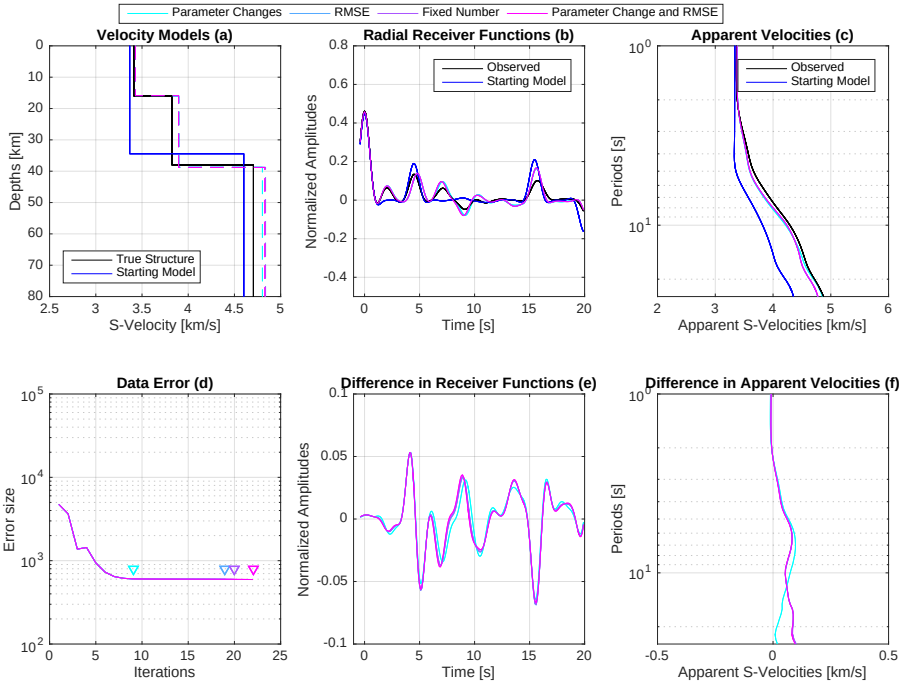


Figure 21: Results when using different methods to stop the iterations in the inversion. Descriptions of the plots are given in Figure 20. In **(d)** the triangles mark the number of iterations.

Figure 21 shows that the results are all very similar. Figure 21d shows how many iterations are performed using the different criteria. The number of iterations is low, when determined by the change in model parameters. It is higher for the other criteria.

The number of iterations in the inversion determines how much the data error converges and the time the inversion takes. By using the combination of $RMSE$ and model parameter change to determine the number of iterations in the inversion, the quality of the solution is based on the estimated model parameters and data. A maximum number of iterations (30) was also included in case the criteria could not be fulfilled.

6.6 Adding Constraints to the Results

Constraints prevent the estimated model parameters to vary a certain amount from the starting model, and are defined as the variance. To allow the crustal S-velocities and delay times to range, only the S-velocity in the mantle was constrained. Christensen and Mooney (1995) stated that the P-velocity in the mantle vary with approximately ± 0.2 km/s, so the mantle S-velocity was constrained with ± 0.1 km/s. Figure 22 shows the effect of the constraints for two weighting situations.

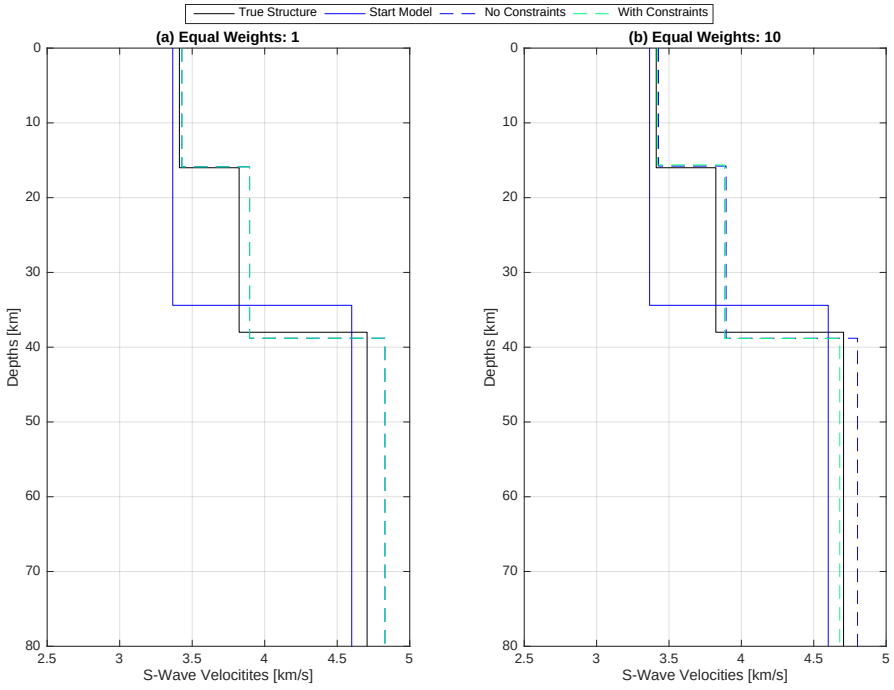


Figure 22: Figure showing the effect of adding constraints to the model parameters in the inversion. The plot to the left has equal weights of 1, while the plot to the right has equal weights of 10 in the inversion.

Figure 22a show that including constraints does not change the estimated model parameters. When the weights were increased, Figure 22b, the difference is clear. Meaning that when the constraints is included, the predictions are only influenced for high weights. Considering this, the constraint was included in the inversion.

6.7 Synthetics with Noise

Adding noise to the synthetic data will make it more realistic. By increasing the noise level, the effect of it could be examined.

The Gaussian coloured noise was added to the waveforms before calculating the receiver functions, the noise levels are shown in Table 4 and the data in Figure 23.

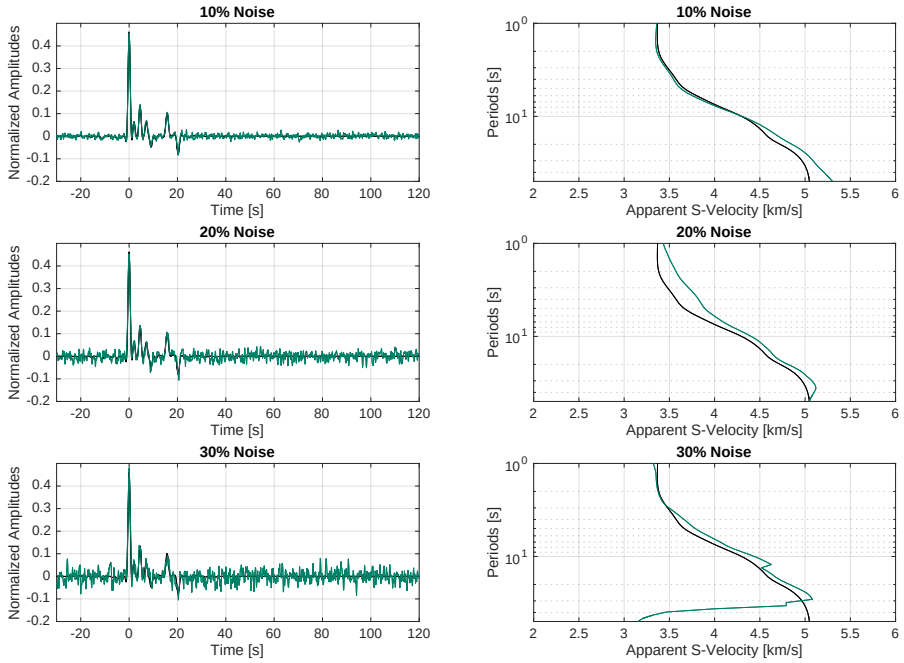


Figure 23: Figure showing the radial receiver functions (plots in the left column) and apparent S-velocities (plots in the right column) for the synthetics with increasing levels of noise added. The noised synthetics is shown in green, while the black graphs show the noise-free synthetic data.

As the noise level increased, the peaks in the receiver function were more disturbed. When the noise level was high, it was difficult to distinguish between the true peaks and the noise. Only the highest noise level made the apparent S-velocity computations drastically more unstable. Here the velocities for longer periods decrease because of interference between the incident P-wave and the Ps phase.

The inversion was performed on the synthetic data with noise using the

Table 4: Table displaying the noise level and the related signal-to-noise ratio (SNR) according to power. For all the noise levels, the standard deviation was 0.2.

Noise level	SNR
10%	$(\frac{10}{1})^2 = 100$
20%	$(\frac{10}{2})^2 = 25$
30%	$(\frac{10}{3})^2 \approx 11.11$

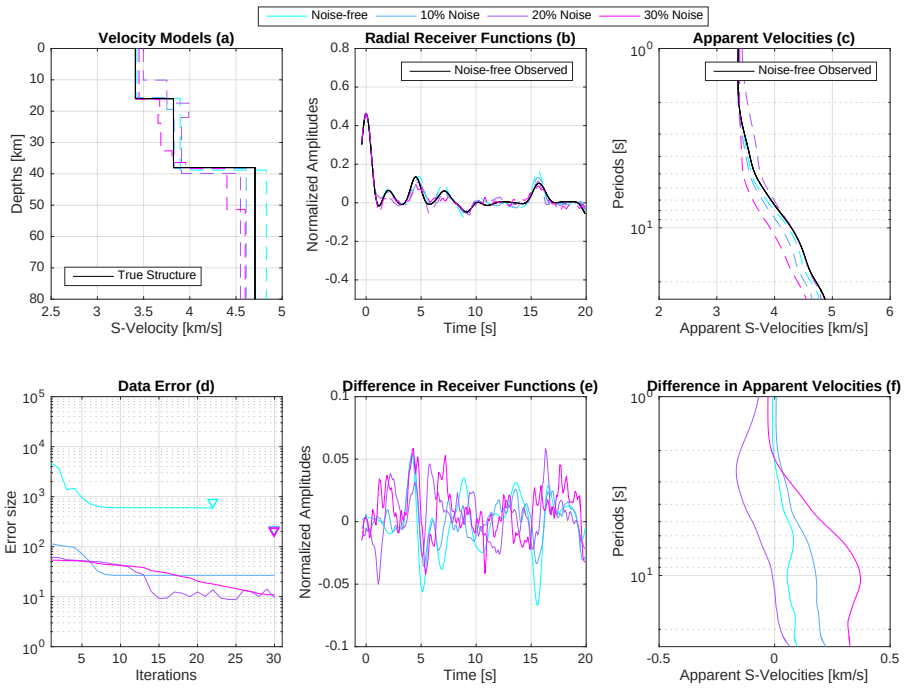


Figure 24: Comparison of results as the noise level is increased on the synthetics. Description of all the plots is found in Figure 20.

improved inversion parameters, the results of which are shown in Figure 24.

Figure 24a show that as the noise level was increased, so was the parameterization while the S-velocity in the mantle decreased. These observations are supported by the estimated data shown in Figure 24b and c, where the increased noise level cause disturbance in the radial receiver function, and reduce the apparent S-velocity for long periods. The data error (Figure 24d) was generally lower for the synthetic data with noise.

The results in Figure 24 confirm that the inversion parameters used work for more realistic data.

6.8 Summary

The goal of testing the inversion method on synthetic data was to automate it to be applicable on large data sets. The testing improved the inversion parameters which are compared to their original definitions in Table 5.

Different weights were tested to compare the results when the inversion favoured the receiver functions or apparent S-velocities, and both equally.

When the inversion favoured the apparent S-velocities, the estimated depths were more related to the starting receiver function than the observed while the estimated S-velocities where fitted very well. When the inversion favoured the receiver function, the estimated depths were related to the observed receiver function while the estimated S-velocities were influenced by the starting apparent S-velocities. These observations are clearly displayed in Figure 20.

The joint inversion favouring both data types equally provide predictions best fitting the data and the model parameters, and a lower data error.

The inversion is dependent on the starting model to determine the parameterization, meaning the number of layers in the estimated S-velocity structure. Using a fixed number of layers in the starting model (as for the CRUST and Global starting models) could be beneficial when comparing the results of stations within close proximity. However, there were three main reasons why using the starting model based on the observed data (the Peaks starting model) was better:

1. The estimated model parameters consist of delay times and S-velocities. In Chapter 4 it was explained that parameterization in delay times improve the non-linearity and non-uniqueness of the solution.

2. When the number of layers is determined by the peaks in the observed radial receiver function, it is possible to separate between areas of complicated or simple crustal structures.

3. When using the observed data to create the starting model, it is already automated.

The estimated model parameters and data are improved iteratively from the starting model. The inversion is stopped when the estimated results and the error change less than 1% for three consecutive iterations.

Constraining the S-velocity in the mantle only improved the results when the weights used were high.

After the inversion parameters had been improved, the inversion was performed on synthetic data with noise. The results indicated that the inversion parameters used should give satisfactory results for real data.

Table 5: Table describing the original and improved inversion parameters.

Inversion parameters	**Original versions**	**Improved versions**
Weights	Receiver Function: 20 Apparent S-Velocity: 1	Receiver Function: 1 Apparent S-Velocity: 1
Starting Model	Predetermined number of layers which meant peaks were fabricated. Velocities used were global averages.	Number of layers determined by number of peaks picked on the observed radial receiver function. Use apparent S-velocities for crust and global average for mantle.
Iterations	20	Stop when RMSE and model parameter changes are small.
Constraints	No constraints	Constrain on the mantle S-velocity.

7 Testing on Real Data

7.1 Introduction

In the previous chapter, the inversion method was improved through testing on synthetic data. Here it is applied to real data from a single station to determine if the inversion yield successful results for real data as well. As mentioned in Chapter 5 the data used was from the station IU_KONO. The mean of the computed radial receiver function and apparent S-velocities from the seismogram recorded at KONO are shown in Figure 25, and will be called the observed data.

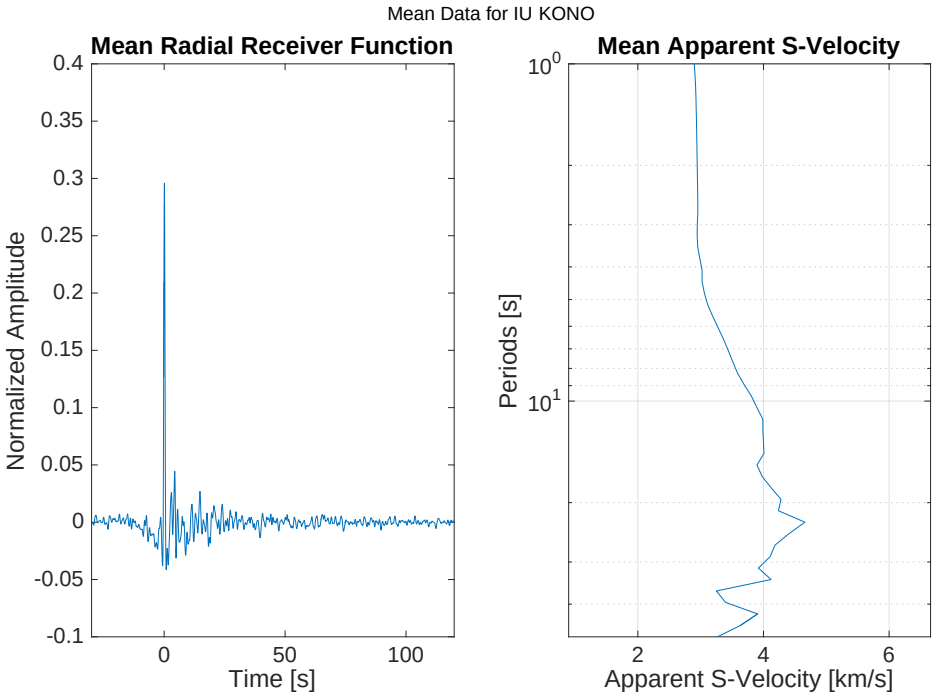


Figure 25: Figure displaying the mean of the observed data. The mean of the 187 observed radial receiver functions is seen in the left plot. The mean of the apparent S-velocity, computed according to Chapter 3, is seen in the right plot.

To determine if the Moho depth computed using the method of this study was realistic, it was compared to the Moho depth for KONO estimated by Ottemoller and Midzi (2003): 34 - 36 km.

The tests performed here are limited compared to the tests in Chapter 6 as only the weights and starting models were examined.

The weights tested for were equally low weights and then high on one data type at the time. Two different starting models were used: the CRUST starting model and the Peaks starting model. For the other inversion parameters, their improved definitions were used.

7.2 Weights

When testing the weights on the real data, three different combinations were used: equal weights of 1 and a high weight (20) on one data type and a low weight (1) on the other.

To examine the quality, the difference between the estimated and observed data, the estimated S-velocity structure and the data errors were compared. The results for the different weighting combinations are shown in Figure 26.

In the S-velocity structures in Figure 26a the estimated depths are similar, and the main difference is in the S-velocities. Figure 26b and c show the difference between the observed and estimated data. When the weights are equal the fit to both data types are good, while for the other weight combinations the fit is only good for one data type. The data error (Figure 26d) is smallest when the inversion is performed using equal low weights on both data types.

Based on these observations, the best results are achieved when using equal weight of 1 in the inversion.

7.3 Starting Model

The two starting models which were tested here, were described in detail in Chapter 6. The results for the inversion of data from KONO using these two starting models, are shown in Figure 27.

There are distinct differences in the estimated velocity structures considering the parameterization, the Moho depths and the average S-velocities. The CRUST starting model has three layers over a half-space, while the Peaks starting model has three crustal layers and one layer in the mantle. The Moho depth estimated from the CRUST starting model i shallower, and the S-velocities are much lower. Comparing the Moho depths to the one estimated by Ottemoller and Midzi (2003), the CRUST starting model estimate is too shallow.

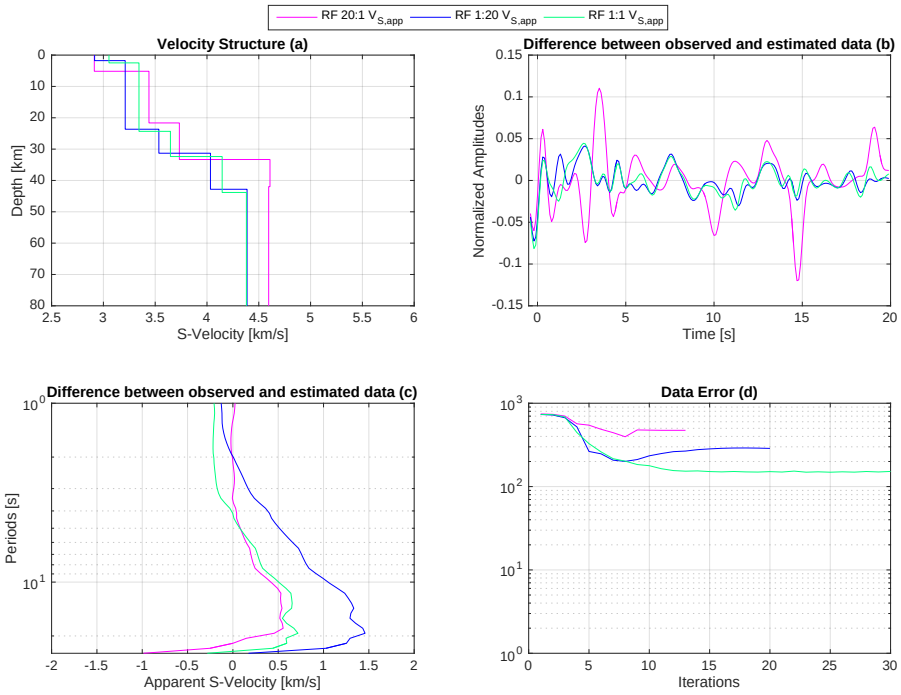


Figure 26: Comparison of the inversion results when using high weights on one of the data sets, and low equal weights on data from IU_KONO. **(a)** Displaying the estimated S-velocity structures for the different weights. **(b)** Shows the difference between the observed and estimated radial receiver function for the different weights. **(c)** Shows the difference between the observed and estimated apparent S-velocities for the different weights. **(d)** Shows the data error for the different weights.

The difference between the estimated and observed receiver functions was the smallest when the Peaks starting model was used, while for the apparent S-velocities the difference was the smallest when using the CRUST starting model. The data error was the lowest when using the Peaks starting model.

The choice of starting model largely influence the results. Based on the observations made from Figure 27, the Peaks starting model provides the most realistic Moho depth and the lowest data error.

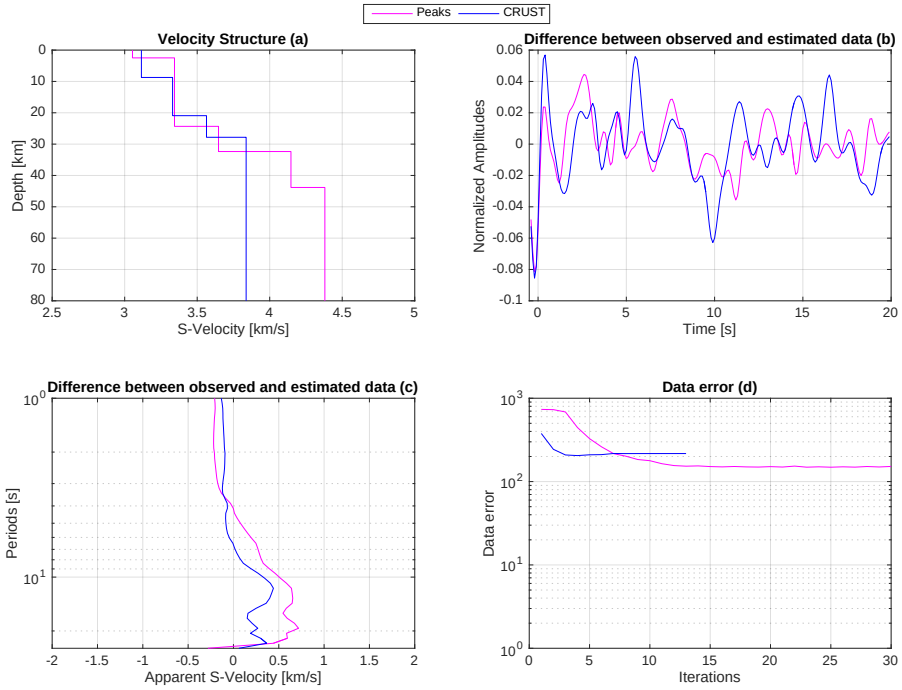


Figure 27: Comparison of the joint inversion for two different starting models. **(a)** The two estimated velocity structures. **(b)** The difference between the estimated and observed radial receiver functions. **(c)** The difference between the estimated and observed apparent S-velocities. **(d)** The data error for the two different starting models.

7.4 Summary

The weights and starting models were tested on the data from a single station to examine the quality of the inversion results for real data, before it was applied to a large data set. The tests confirmed that the inversion parameters found through testing on synthetic data were reasonable for real data as well.

8 Application to Scandinavian Data

8.1 Introduction

To this point, the inversion method has been tested on synthetic data and real data from a single station. Here, the joint inversion is performed on a subset of stations in and around Norway.

A short overview of the main geological areas in Scandinavia will be given before the inversion results are described.

The inversion results for the stations will be described within the geological domains and for stations within close proximity.

8.2 Geologic and Tectonic History of the Area

Scandinavia can be divided into five different geological zones (Gaal and Gorbatschev (1987)) shown in Figure 28: the Archean domain, the Caledonides, the Southwest Scandinavian domain, the Svecofennian domain and the Oslo Graben.

The Archean domain consists of Archean gneiss and early Proterozoic granite-greenstone (Gaal and Gorbatschev (1987)) and some of the oldest rocks in Scandinavia are located here.

The Caledonides was a Paleozoic mountain range which affected much of the Precambrian continental crust, causing metamorphism and complex folds (Gaal and Gorbatschev (1987)).

The Southwest Scandinavian domain consist of plutonic and metamorphic basement which was affected by the Sveconorwegian-Grenvillian and Caledonian orogenies (Frassetto and Thybo (2013)) and includes the Transscandinavian Igneous Belt (Kolstrup and Maupin (2013)).

The Svecofennian domain has a mainly metamorphosed plutonic crust (Frassetto and Thybo (2013)) with granite intrusions (Gaal and Gorbatschev (1987)).

The Oslo Graben is a rift zone believed to be associated with the TransEuropean rift system (Ottemoller and Midzi (2003)).

8.3 Results for the Stations

The automated joint inversion was applied to the stations listed in Chapter 5 using the inversion parameters determined through the synthetic and real

Figure 28: Map showing approximately the boundaries between different geologic domains. The stations used in the inversion are marked. The red marks with an A corresponds to the NNSN, the light blue marks corresponds to the SNSN, the dark blue marks corresponds to the NO network, the pink marks correspond to the Norway 05 network, the white mark corresponds to GSN and the green mark corresponds to DWWSSN.

testing described in Chapters 6 and 7. The estimated S-velocity structures for all the stations are shown in Figure 29, and the estimated Moho depth and average S-velocity for each station is given in Table 8. In Appendix D the results for each station is given in more detail.

The Moho depth varies between 21 and 52 km with a mean of 36 km, and the average S-velocity varies between 3.6 and 4.3 km/s with a mean of 3.9 km/s.

Table 6: Table showing the estimated Moho depths and estimated average S-velocities for the stations.

Station	Estimated Moho Depth [km]	Estimated Average S-Velocity [km/s]
NS_BER	29.3	4.0
NS_FOO	23.3	4.2
NS_HOMB	30.6	3.7
NS_MOL	38.8	4.0
NS_NSS	36.1	3.7
NS_SUE	31.2	4.2
NS_TBLU	45.4	4.1
NS_TRO	35.0	3.9
NO_AKN	40.1	4.1
NO_ARE0	43.3	3.8
NO_NAO01	34.6	4.0
NO_NB201	37.8	3.6
NO_NBO00	35.8	4.0
NO_NC204	35.2	3.9
NO_NC303	38.1	3.9
NO_NC405	21.8	3.9
NO_NC602	37.6	3.8
DW_KEV	44.2	3.9
UP_AAL	41.2	3.9
UP_LANU	51.6	3.7
UP_NRAU	44.5	4.0
UP_SJUU	42.8	3.7
UP_STRU	34.1	3.9
UP_UPP	44.4	3.9
YY05_SN30	33.9	4.0
YY05_SN31	32.7	3.8
YY05_SN32	35.1	3.9
YY05_SN33	25.3	4.3
YY05_SN34	31.7	3.7
IU_KONO	33.8	3.9

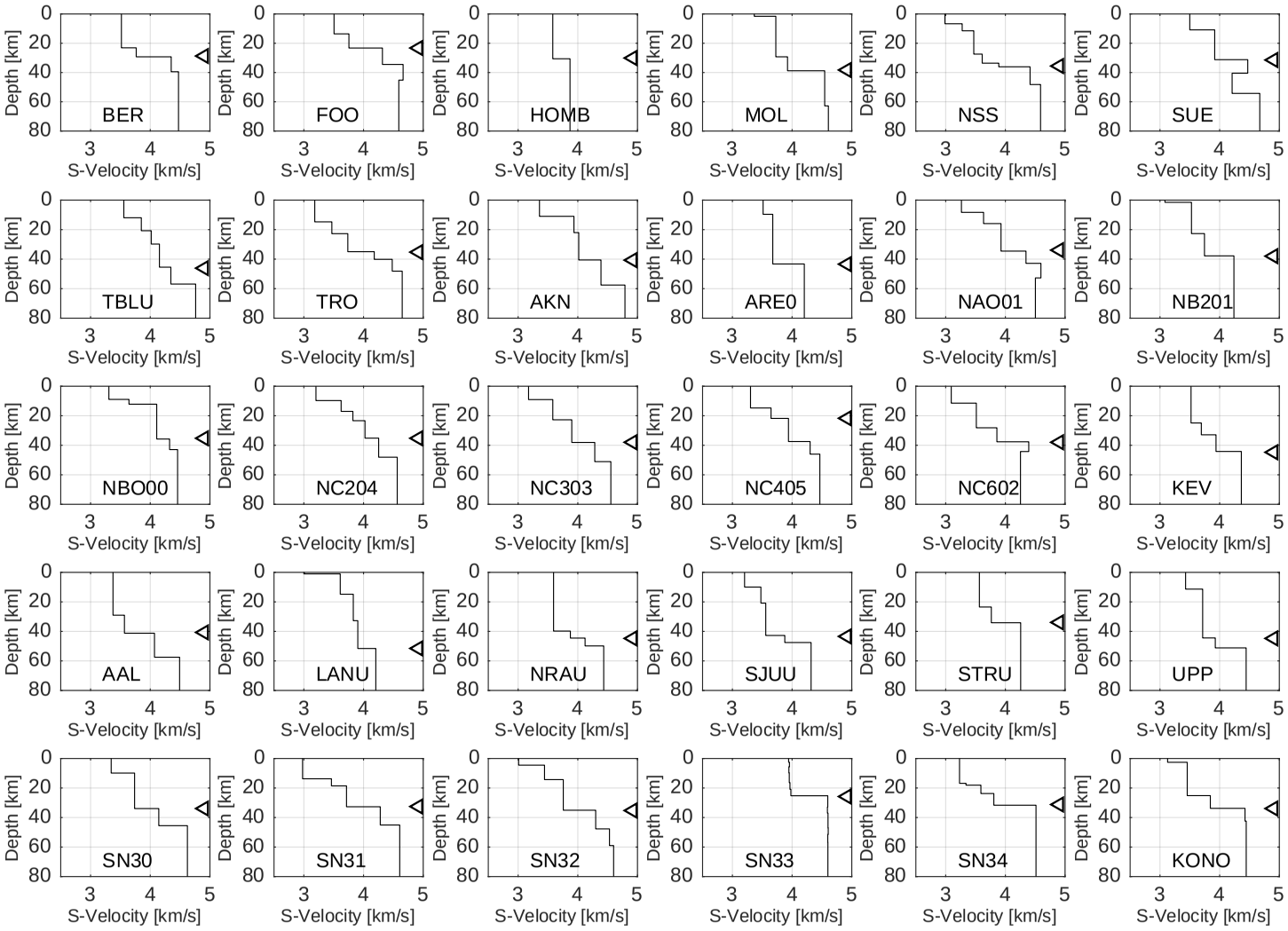


Figure 29: The estimated S-velocity structures for all the stations. The arrow points to the Moho.

8.3.1 Results within Geological Domains

Within the Archean domain the Moho was estimated at large depths and the average S-velocities were estimated close to 4.0 km/s. The S-velocity structures revealed some crustal layering for the area, with no apparent mantle structures.

Within the Caledonides the estimated Moho depths vary and the average S-velocities are generally high. The S-velocity structures reveal complex crustal layers and some mantle structure.

The majority of stations are within the Southwest Scandinavian domain, so the results showed a lot of variation. The Moho depths range from shallow to intermediate and the average S-velocities range from low (3.7 km/s) to high (4.3 km/s). For the stations located the farthest South, the complexity of the crustal layering is reduced.

The stations within the Svecofennian domain showed a thick layer above the deep Moho and a layer in the mantle, and high average S-velocities.

8.3.2 Results for Stations with Close Proximity

For two of the seismic networks, the stations within are located close together: The NO network, where all the stations, except AKN and ARE0, are located North of the Oslo Graben, and the Norway 05 network, where the stations are located West of the Oslo Graben. The stations KONO and HOMB are also located within the bounds of the Norway 05 network.

The stations North of the Oslo Graben have similar Moho depths at around $35km$, except for the station NC405 where the Moho is picked at $21km$. The automated method picks the Moho as the depth according to the peak with the highest amplitude in the observed radial receiver function. For NC405 the Moho should be deeper, which is apparent when looking at the velocity structure in Figure 29, however an earlier arriving peak has a higher amplitude. In the velocity structure for the stations NC204, NC303 and NC602 a layer interface at approximately 21 km is located.

For the stations located West of the Oslo Graben, the Moho depths becomes shallower to the South. All the stations in the proximity show crustal and mantle layering, except HOMB and SN33.

8.3.3 The Error of the Results

The data error of the inversion results differ from station to station, as seen in Table 7. The RMS error for the data varies between 60 and 461 for the

stations, with a mean error of 200. The error on the apparent S-velocity is much smaller than for the receiver function.

Table 7: Table showing the errors corresponding to the stations.

Station	Data Error	Receiver Function Error	Apparent S-Velocity Error
BER	316	205	119
FOO	217	211	6
HOMB	160	155	33
MOL	106	102	4
NSS	251	242	11
SUE	188	169	19
TBLU	61	51	10
TRO	260	253	6
AKN	225	213	14
ARE0	194	170	24
NAO01	71	61	10
NB201	177	145	35
NBO00	90	81	8
NC204	93	80	13
NC303	92	89	3
NC405	310	279	32
NC602	160	139	21
KEV	145	130	18
AAL	295	267	28
LANU	108	99	8
NRAU	162	145	18
SJUU	159	151	18
STRU	269	240	28
UPP	145	126	20
SN30	83	74	10
SN31	460	446	14
SN32	309	298	11
SN33	423	351	76
SN34	173	165	9
KONO	218	165	55

9 Discussion

9.1 Overview

In this chapter the results described in Chapter 8 are discussed. The estimated Moho depths are compared with Moho depths computed in the following studies: Kolstrup and Maupin (2013) who used a joint inversion of receiver functions and surface waves, Ottemoller and Midzi (2003) and Frassetto and Thybo (2013) who used receiver function inversion and Korsman et al. (1999) who used results for different geophysical methods. These comparisons are shown in Table 8.

The velocity structures for two stations are compared in detail with the velocity structures estimated by Ottemoller and Midzi (2003) and Kolstrup and Maupin (2013). The implications of the results and the method will also be discussed here before stating what further work can be performed on the method.

9.2 Comparison of Two Stations

The two stations which will be compared to other studies are: FOO and NAO01. These stations were chosen because Ottemoller and Midzi (2003) and Kolstrup and Maupin (2013) have both estimated velocity structures for them.

First, the results for NAO01 were compared, see Figure 30.

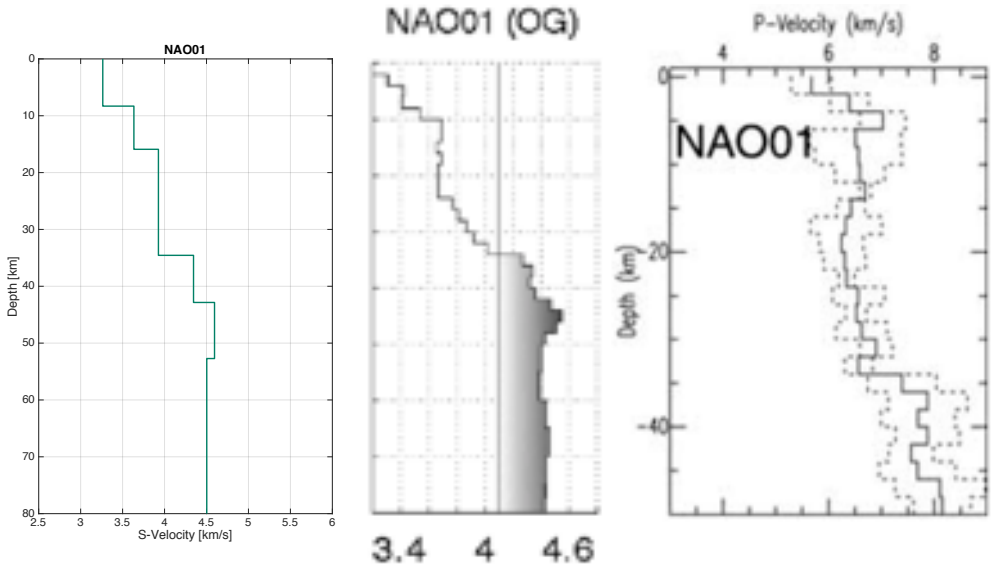
Figure 30a show the results from this thesis, where there are crustal layers at 9 and 16 km with a Moho depth of 34.6 km and an increase in the mantle velocity at 42 to 52 km. Through the crust, the S-velocity increases from 3.3 km/s to 4 km/s. In the upper mantle the velocity is 4.4 km/s before increasing to 4.6 km/s from 42 to 52 km depth, and then decreasing to 4.5 km/s.

The S-velocity structure estimated by Kolstrup and Maupin (2013) is shown in Figure 30b. Although there are many thin layers here, one can locate the same layer interfaces as in my results. There are crustal layers at 10 and 14 km and the Moho depth was estimated to 34 km. In the mantle there is a velocity increase between 42 and 47 km, similarly to in my results. In the crust the S-velocity increases from 3.4 to 4 km/s, and in the upper mantle the velocity is 4.3 km/s, 4.6 km/s in the mantle layer between 42 and 47 km, then 4.5 km/s below this.

Ottemoller and Midzi (2003) estimated the P-velocity structure shown in Figure 30c. The P-velocities are estimated with a constant Poisson's ratio of

Table 8: Table showing Moho depths found for this study compared to other studies.

Station	Moho depth [km], this study	Moho depth [km], Kolstrup and Maupin (2013)	Moho depth [km], Ottemoller and Midzi (2003)	Moho depth [km], Frassetto and Thybo (2013)	Moho depth [km], Korsman et al. (1999)
NS_BER	29.3	30	NA	27-34	NA
NS_FOO	23.3	33	32-36	30-40	NA
NS_HOMB	30.6	NA	NA	NA	NA
NS_MOL	38.8	39	38-40	35-40	NA
NS_NSS	36.1	NA	38-42	30-35	NA
NS_SUE	31.2	NA	30-34	30-35	NA
NS_TBLU	45.4	NA	NA	NA	NA
NS_TRO	35.0	NA	36-40	NA	NA
NO_AKN	40.6	NA	NA	NA	NA
NO_ARE0	43.3	NA	NA	NA	NA
NO_NAO01	34.6	34	34-36	33-37	NA
NO_NB201	37.8	NA	NA	NA	NA
NO_NBO00	35.8	NA	NA	NA	NA
NO_NC204	35.2	39	NA	NA	NA
NO_NC303	38.1	NA	NA	NA	NA
NO_NC405	21.8	NA	NA	NA	NA
NO_NC602	37.6	38	NA	NA	NA
DW_KEV	44.2	NA	44-50	NA	NA
UP_AAL	41.2	NA	NA	NA	44-46
UP_LANU	51.6	NA	NA	NA	47
UP_NRAU	44.5	NA	NA	NA	46
UP_SJUU	42.8	NA	NA	NA	44
UP_STRU	34.1	NA	NA	NA	36-38
UP_UPP	44.4	NA	NA	NA	44-46
IU_KONO	33.8	NA	34-36	35-40	NA
YY05_SN30	33.9	NA	NA	NA	NA
YY05_SN31	32.7	NA	NA	NA	NA
YY05_SN32	35.1	NA	NA	NA	NA
YY05_SN33	25.3	NA	NA	NA	NA
YY05_SN34	31.7	NA	NA	NA	NA



(a) My results (b) Kolstrup and Maupin (2013) (c) Ottemoller and Midzi (2003)

Figure 30: Estimated velocity structures for various studies for the station NAO01.

0.25 meaning that the P- and S-velocity structures are proportional. Here the Moho depth was estimated around 34 to 36 km. Through the crust and mantle the velocity is quite stable besides a velocity increase at 4 to 6 km and at the Moho.

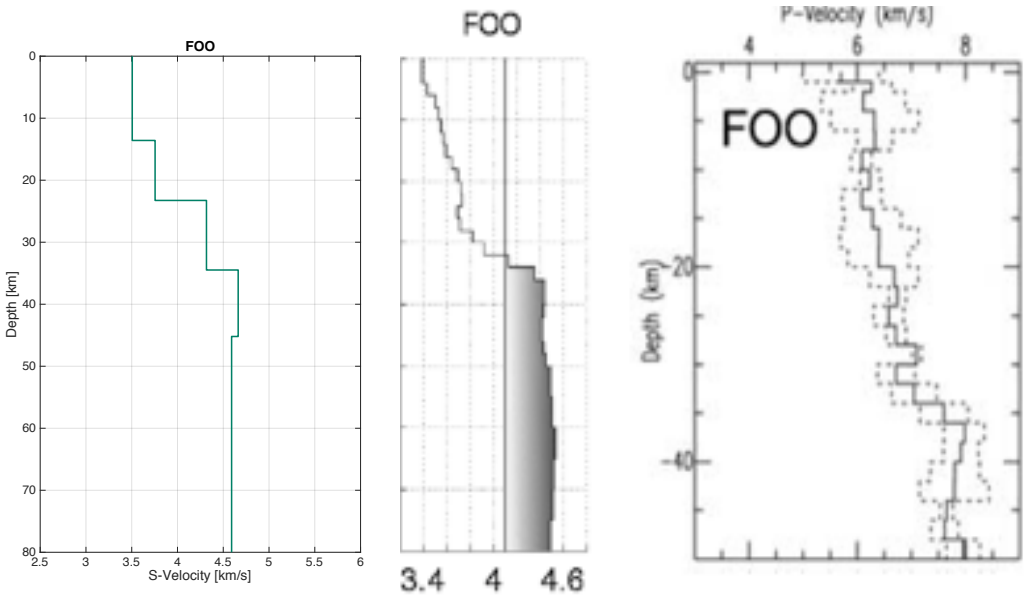
The velocity structures estimated for the station NAO01 in this thesis and by Kolstrup and Maupin (2013) are very similar, with the main difference being the parameterization, while the velocity structure estimated by Ottemoller and Midzi (2003) do not reveal the crustal and mantle layering.

Next, the results for the station FOO were compared, see Figure 31.

The estimated S-velocity structure for FOO from this study is shown in Figure 31a. Here the Moho is picked at 23.3 km depth, with a crustal layer at 14 km and a mantle layer at 34 km depth. Through the crust the S-velocity increases from 3.5 km/s to 4.4 km/s at the picked Moho. In the mantle, it increases to 4.7 km/s at 34 km and decrease to 4.6 km/s at 45 km depth.

The velocity structure estimated by Kolstrup and Maupin (2013) is shown in Figure 31b. Here the Moho depth is estimated to be at 33 km and a crustal layer can be interpreted at 24 km. The S-velocity increase from 3.4 km/s in the crust to 4.5 km/s in the mantle.

The P-velocity structure estimated by Ottemoller and Midzi (2003) is shown



(a) My results (b) Kolstrup and Maupin (2013) (c) Ottemoller and Midzi (2003)

Figure 31: Estimated velocity structures for various studies for the station FOO.

in Figure 31c. The Moho depth is estimated to be between 32 and 36 km, with no other clear layers.

The estimated Moho depth in this thesis is very different from the other studies. Here the automated inversion method has picked the Moho depth too shallow, as happened for NC405 as described in Chapter 8. Otherwise the structures are well matched.

9.3 Implications

Based on the results in Chapter 8 the following can be discussed:

1. Comparison to other results.
2. Implications for Scandinavia.
3. Application of the inversion method.

The results achieved in this thesis were well resolved compared to other studies and provided details about the crustal and mantle structure. The estimated velocity structures were not over parameterized making the structures more distinguished. The only apparent issue with this method is that the Moho depth might be picked wrong, however the real Moho discontinuity

still appears in the velocity structure.

For Scandinavia the results indicate that the crust is thicker and less complex beneath Sweden and become shallower towards the coast, especially in Southern Norway. The crust is very complex within the Caledonides and parts of the Southwest Scandinavian domain.

Based on the good comparison of the results with other studies, applying the automated joint inversion to a global data set would provide satisfactory estimations of S-velocity structures. The only issues were for the two stations where the Moho depth was picked too shallow, however the true Moho was still in the velocity structure.

9.4 Further Work

The method, as is, can be applied to global data sets with success. The method is being further developed by also inverting the apparent P-velocity making the V_P/V_S ratio a model parameter. The inversion would then estimate P- and S-velocity structures for all stations, which could then be examined for a much more thorough interpretation of the crustal and upper mantle structure.

10 Conclusion

The main objective of this thesis was to automate and improve the joint inversion of receiver functions and apparent S-velocities to be applied to real data from a set of stations. By performing the joint inversion, the benefits from inverting each data type are incorporated which reduces the non-uniqueness and non-linearity of the solution. In this joint inversion the delay times are estimated to fit the observed delay times and the conversion to depth uses the apparent S-velocity for short periods, meaning that the estimated depths and S-velocities are very well resolved. When the inversion method is automated it is possible to apply to a global data set without making specifications according to the different geological environments.

To automate the inversion method, tests of the inversion parameters were performed on synthetic and real data. The tests on the synthetic data revealed that the results where more accurate for the joint inversion than when inverting only one of the data types.

By letting the number of peaks in the observed receiver function determine the number of layers in the starting and estimated model, over parameterization and smoothing is avoided. When the starting model is similar to the observed data, the other inversion parameters become less important.

The main issues of the method arise when the data has a high SNR or when the highest amplitude peak does not mark the Moho, otherwise the results are

Through the work in this thesis an automated joint inversion method applicable to a large data set have been tested and improved. The choice of inversion parameters provide good estimations of Moho depths and describe complexities in the crustal and mantle structure. It would be interesting to see how the method holds up for a global data set.

References

Ammon, C. J., Randall, G. E., and Zandt, G. (1990). On the nonuniqueness of receiver function inversions. *Journal of Geophysical Research-Solid Earth and Planets*, 95(B10):15303–15318.

Aster, R. C. and Thurber, C. H. (2013). *Parameter estimation and inverse problems*. Parameter Estimation and Inverse Problems (Second Edition). Academic Press, Waltham, MA, 2nd ed. / richard c. aster, clifford h. thurber. edition.

Backus, G. and Gilbert, F. (1970). Uniqueness in inversion of inaccurate gross earth data. *Philosophical Transactions of the Royal Society of London Series a-Mathematical and Physical Sciences*, 266(1173):123–192.

Christensen, N. I. and Mooney, W. D. (1995). Seismic velocity structure and composition of the continental-crust - a global view. *Journal of Geophysical Research-Solid Earth*, 100(B6):9761–9788.

Frassetto, A. and Thybo, H. (2013). Receiver function analysis of the crust and upper mantle in fennoscandia - isostatic implications. *Earth and Planetary Science Letters*, 381:234–246.

Gaal, G. and Gorbatschev, R. (1987). An outline of the precambrian evolution of the baltic shield. *Precambrian Research*, 35:15–52.

Ghayamghamian, M. R. and Nouri, G. R. (2007). On the characteristics of ground motion rotational components using chiba dense array data. *Earthquake Engineering & Structural Dynamics*, 36(10):1407–1429.

Gupta, P. K. (2011). *Inverse Theory, Linear*, pages 632 – 639. Springer Netherlands.

Jacobsen, B. H. and Svenningsen, L. (2008). Enhanced uniqueness and linearity of receiver function inversion. *Bulletin of the Seismological Society of America*, 98(4):1756–1767.

Kolstrup, M. L. and Maupin, V. (2013). A proterozoic boundary in southern norway revealed by joint-inversion of p-receiver, functions and surface waves. *Precambrian Research*, 238:186–198.

Korsman, K., Korja, T., Pajunen, M., Virransalo, P., Ekdahl, E., Elo, S., Haapala, I., Heikkinen, P., Hjelt, S. E., Holtta, P., Huhma, H., Idman, H., Jokinen, J., Kahkonen, Y., Kilpelainen, T., Koistinen, T., Komminaho, K., Kontinen, A., Korhonen, J., Korja, A., Kousa, J., Kukkonen, I. T., Iurimo, M., Lahtinen, R., Leino, M. A. H., Lerssi, J., Lindkvist, K., Lukkarinen, H., Luosto, U., Luukas, J., Luukkonen, E., Malaska, J., Mertanen, S., Niemelaa, R., Nironen, M., Paavola, J., Papunen, H., Peltonen, P., Pesonen, L., Pietikainen, K., Pihlaja, P., Ramo, T., Rastas,

J., Ruotoistenmaki, T., Saavuori, H., Tervo, T., Tuisku, P., Vaasjoki, M., Vaisanen, M., Wasenius, P., Yliniemi, J., and Grp, G. S. W. (1999). The ggt/sveka transect: Structure and evolution of the continental crust in the paleoproterozoic svecofennian orogen in finland. *International Geology Review*, 41(4):287–333.

Langston, C. A. (1977). Corvallis, oregon, crustal and upper mantle receiver structure from teleseismic p-waves and s-waves. *Bulletin of the Seismological Society of America*, 67(3):713–724.

Langston, C. A. (1979). Structure under mount rainier, washington, inferred from teleseismic body waves. *Journal of Geophysical Research*, 84(NB9):4749–4762.

Lay, T. and Wallace, T. C. (1995). *Modern Global Seismology*. International Geophysics. Elsevier Science, Burlington.

Lei, J. S. and Zhao, D. P. (2005). P-wave tomography and origin of the changbai intraplate volcano in northeast asia. *Tectonophysics*, 397(3-4):281–295.

Lines, L. R. and Treitel, S. (1984). Tutorial - a review of least-squares inversion and its application to geophysical problems. *Geophysical Prospecting*, 32(2):159–186.

Menke, W. (1989). *Geophysical Data Analysis: Revised Edition*. United States: Academic Press, rev. ed. edition.

Nuttli, O. and Whitmore, J. D. (1961). An observational determination of the variation of the angle of incidence of p waves with epicentral distance. *Bulletin of the Seismological Society of America*, 51(2):269–276.

Ottemoller, L. and Midzi, V. (2003). The crustal structure of norway from inversion of teleseismic receiver functions. *Journal of Seismology*, 7(1):35–48.

Phinney, R. A. (1964). Structure of earths crust from spectral behavior of long-period body waves. *Journal of Geophysical Research*, 69(14):2997–3017.

Pujol, J. (2003). *Elastic wave propagation and generation in seismology*. Cambridge University Press, Cambridge.

Rondenay, S. (2009). Upper mantle imaging with array recordings of converted and scattered teleseismic waves. *Surveys in Geophysics*, 30(4-5):377–405.

Stein, S. and Wysession, M. E. (2003). *An introduction to seismology, earthquakes and earth structure*. Blackwell Science, Oxford.

Svenningsen, L. and Jacobsen, B. H. (2007). Absolute s-velocity estimation from receiver functions. *Geophysical Journal International*, 170(3):1089–1094.

Tarantola, A. and Valette, B. (1982a). Generalized non-linear inverse problems solved using the least-squares criterion. *Reviews of Geophysics*, 20(2):219–232.

Tarantola, A. and Valette, B. (1982b). Inverse problems = quest for information. *Journal of Geophysics-Zeitschrift Fur Geophysik*, 50(3):159–170.

vanderLee, S. and Nolet, G. (1997). Upper mantle s velocity structure of north america. *Journal of Geophysical Research-Solid Earth*, 102(B10):22815–22838.

Vinnik, L. P. (1977). Detection of waves converted from p to sv in the mantle. *Physics of the Earth and Planetary Interiors*, 15(1):39–45.

Xu, L. L., Rondenay, S., and van der Hilst, R. D. (2007). Structure of the crust beneath the southeastern tibetan plateau from teleseismic receiver functions. *Physics of the Earth and Planetary Interiors*, 165(3-4):176–193.

Zhu, L. P. and Kanamori, H. (2000). Moho depth variation in southern california from teleseismic receiver functions. *Journal of Geophysical Research-Solid Earth*, 105(B2):2969–2980.

Appendix

A The Relationship Between the True and Apparent Incidence Angle

Deriving the relationship between the true and apparent incidence angle from the following equation from Svenningsen and Jacobsen (2007):

$$i = \sin^{-1} \sqrt{\frac{(1 - \cos i_P)V_P^2}{2V_S^2}} \tag{A.1}$$

Rewriting Equation A.1 gives:

$$\sin i = \sqrt{\frac{(1 - \cos i_P)V_P^2}{2V_S^2}} \tag{A.2}$$

Raising Equation A.2 to the power of 2:

$$\sin^2 i = \frac{(1 - \cos i_P)V_P^2}{2V_S^2} \tag{A.3}$$

Rewriting Equation A.3 gives:

$$2V_S^2 \sin^2 i = (1 - \cos i_P)V_P^2 \tag{A.4}$$

By using the following trigonometric identity:

$$\sin^2 u = \frac{1 - \cos 2u}{2} \Rightarrow 1 - \cos u = 2\sin^2 (\frac{1}{2}u)$$

Equation A.4 can be written as

$$\begin{gathered} 2V_S^2 \sin^2 i = 2\sin^2 (\frac{1}{2}i_P)V_P^2 \\ \Rightarrow \\ V_S \sin i = \sin (\frac{1}{2}i_P)V_P \\ \Rightarrow \\ \frac{\sin i}{V_P} = \frac{\sin (\frac{1}{2}i_P)}{V_S} = p \end{gathered} \tag{A.5}$$

B Deriving the Linearized Weighted Least Squares Iteration

The derivation performed here follows the methods in Menke (1989), Tarantola and Valette (1982a) and Tarantola and Valette (1982b).

Start with the linear explicit form of an inverse problem:

$$\mathbf{Gp} = \mathbf{d} \tag{B.1}$$

where the arrays are defined as

$$\begin{aligned} \mathbf{d} &= [d_1, d_2, \cdots, d_N] \\ \mathbf{p} &= [p_1, p_2, \cdots, p_M] \end{aligned}$$

and the model matrix $\mathbf{G}$ is the Jacobian matrix, which relates the data and the model parameters.

The goal with Least-Squares inversion is to minimize the length of the error. The error is linearized using Taylor series of the data at the model parameters:

$$\begin{aligned} e_i &= d_i(p_j) - \frac{\partial d_i}{\partial p_j}(p_j) \\ \mathbf{e} &= \mathbf{d} - \mathbf{Gp} \\ e_i &= d_i^{obs} - d_i^{est} \end{aligned} \tag{B.2}$$

The length of this error is the Euclidean distance of the estimated data ($\mathbf{d}^{est}$) from the observed data ($\mathbf{d}^{obs}$). The Euclidean distance is also named the L_2-norm and is defined as:

$$||\mathbf{e}||_2 = \sqrt{\sum_{i=1}^{N} e_i^2} \Rightarrow E = \mathbf{e}^T\mathbf{e} \tag{B.3}$$

To find the solution which minimizes this length, one needs to differentiate the error equation with respect to a model parameter (p_q) and set differentiated error equal to zero. Unless otherwise specified, this is the method used for minimization further in this derivation. To do this, the notation of the error is changed:

$$
\begin{aligned}
E &= \mathbf{e}^T\mathbf{e} \\
&= (\mathbf{d}-\mathbf{Gp})^T(\mathbf{d}-\mathbf{Gp}) \\
&= \sum_i^N \Big[d_i - \sum_j^M G_{ij}p_j\Big]\Big[d_i - \sum_k^M G_{ik}p_k\Big] \\
&= \sum_i^N \Big[d_i d_i - \sum_k^M G_{ik}p_k d_i - \sum_j^M G_{ij}p_j d_i + \sum_j^M\sum_k^M G_{ij}G_{ik}p_j p_k\Big] \\
&= \sum_i^N d_i d_i - \sum_i^N\sum_k^M G_{ik}p_k d_i - \sum_i^N\sum_j^M G_{ij}p_j d_i + \sum_i^N\sum_j^M\sum_k^M G_{ij}G_{ik}p_j p_k \\
&= \sum_i^N d_i d_i - \sum_k^M p_k\sum_i^N G_{ik}d_i - \sum_j^M p_j\sum_i^N G_{ij}d_i + \sum_j^M p_j\sum_k^M p_k\sum_i^N G_{ij}G_{ik} \\
&= \sum_j^M\sum_k^M p_j p_k\sum_i^N G_{ij}G_{ik} - 2\sum_j^M p_j\sum_i^N G_{ij}d_i + \sum_i^N d_i d_i
\end{aligned}
\tag{B.4}
$$

To differentiate Equation B.4 I will differentiate each term separately before combining them.

Term 1:

$$
\begin{aligned}
&\frac{\partial}{\partial p_q}\Big[\sum_j^M\sum_k^M p_j p_k\sum_i^N G_{ij}G_{ik}\Big] \\
&= \sum_j^M\sum_k^M\Big(\frac{\partial p_j p_k}{\partial p_q}\Big)\sum_i^N G_{ij}G_{ik} \\
&= \sum_j^M\sum_k^M\Big(\frac{\partial p_j}{\partial p_q}\cdot p_k + p_j\cdot\frac{\partial p_k}{\partial p_q}\Big)\sum_i^N G_{ij}G_{ik} \\
&= \sum_j^M\sum_k^M\Big(\delta_{jq}p_k + p_j\delta_{kq}\Big)\sum_i^N G_{ij}G_{ik} \\
&= \sum_k^M\Big(p_k + p_k\Big)\sum_i^N G_{ij}G_{ik} \\
&= 2\sum_k^M p_k\sum_i^N G_{iq}G_{ik}
\end{aligned}
\tag{B.5}
$$

where δ_{jq} is the Kronecker delta defined as

$$\delta_{jq} = \begin{cases} 0, j \neq q \\ 1, j = q \end{cases}$$

Term 2:

$$\begin{aligned} & -2\frac{\partial}{\partial p_q}\Big[\sum_j^M p_j \sum_i^N G_{ij} d_i\Big] \\ & = -2\sum_j^M \frac{\partial p_j}{\partial p_q} \sum_i^N G_{ij} d_i \\ & = -2\sum_j^M \delta_{jq} \sum_i^N G_{ij} d_i \\ & = -2\sum_i^N G_{iq} d_i \end{aligned} \tag{B.6}$$

Term 3:

$$\frac{\partial}{\partial p_q}\Big[\sum_i^N d_i d_i\Big] = 0 \tag{B.7}$$

The results from these differentiations are collected and seen in Equation B.8.

$$\frac{\partial E}{\partial p_q} = 2\sum_k^M p_k \sum_i^N G_{iq} G_{ik} - 2\sum_i^N G_{iq} d_i = 0 \tag{B.8}$$

which in matrix notation becomes:

$$\mathbf{G}^T\mathbf{G}\mathbf{p} - \mathbf{G}^T\mathbf{d} = 0 \tag{B.9}$$

From Equation B.9 it is possible to estimate the values for the model parameters ($\mathbf{p}$) by

$$\mathbf{p}^{est} = [\mathbf{G}^T\mathbf{G}]^{-1}\mathbf{G}^T\mathbf{d} \tag{B.10}$$

Since the problem is rarely determined ($M = N$) the solution needs to be regularized. The solution is regularized by using a damping factor ϵ. To

implement this, the error is changed to also include the prediction error as seen in Equation B.11, then the total error is minimized.

$$Q_{tot} = E + \epsilon^2 L = \mathbf{e}^T\mathbf{e} + \epsilon^2 \mathbf{p}^T\mathbf{p} \tag{B.11}$$

For the minimization to be simpler, Equation B.11 is rewritten from matrix notation.

$$\begin{aligned} Q_{tot} &= \mathbf{e}^T\mathbf{e} + \epsilon^2\mathbf{p}^T\mathbf{p} \\ &= [\mathbf{d} - \mathbf{G}\mathbf{p}]^T[\mathbf{d} - \mathbf{G}\mathbf{p}] + \epsilon^2\mathbf{p}^T\mathbf{p} \\ &= \Big[\sum_i^N d_i - \sum_j^M G_{ij}p_j\Big]\Big[\sum_i^N d_i - \sum_k^M G_{ik}p_k\Big] + \epsilon^2 \sum_j^M p_j \sum_k^M p_k \end{aligned} \tag{B.12}$$

The first term in Equation B.12 is the same as in Equation B.4, so it is unnecessary to show this differentiation again. Instead the differentiation for the second term of Equation B.12 is shown.

$$\begin{aligned} &\epsilon^2 \frac{\partial}{\partial p_q}\Big[\sum_j^M \sum_k^M p_j p_k\Big] \\ &= \epsilon^2 \sum_j^M \sum_k^M \Big(\frac{\partial p_j p_k}{\partial p_q}\Big) \\ &= \epsilon^2 \sum_j^M \sum_k^M \Big(\frac{\partial p_j}{\partial p_q} \cdot p_k + p_j \frac{\partial p_k}{\partial p_q}\Big) \\ &= \epsilon^2 \sum_j^M \sum_k^M \Big(\delta_{jq} p_k + p_j \delta_{kq}\Big) \\ &= 2\epsilon^2 \sum_k^M p_k \end{aligned} \tag{B.13}$$

From these terms we get the following:

$$\frac{\partial Q_{tot}}{\partial p_q} = 2\sum_k^M p_k \sum_i^N G_{iq}G_{ik} - 2\sum_i^N G_{iq}d_i + 2\epsilon^2 \sum_k^M p_k = 0 \tag{B.14}$$

It is then possible to rewrite Equation B.14 to matrix notation:

$$\begin{aligned} \mathbf{p}\mathbf{G}^T\mathbf{G} - \mathbf{G}^T\mathbf{d} + \epsilon^2\mathbf{I}\mathbf{p} &= 0 \\ \mathbf{p}[\mathbf{G}^T\mathbf{G} + \epsilon^2\mathbf{I}] &= \mathbf{G}^T\mathbf{d} \end{aligned} \tag{B.15}$$

The estimated solution of Equation B.15 is defined in Equation B.16 and is called the damped least squares solution.

$$\mathbf{p}^{est} = [\mathbf{G}^T\mathbf{G} + \epsilon^2\mathbf{I}]^{-1}\mathbf{G}^T\mathbf{d} \tag{B.16}$$

Next *a priori* information is implemented to find a solution where the length is minimized by comparison with the *a priori* information ($\langle\mathbf{p}\rangle$). Weights will also be implemented.

Both are implemented in the total error equation (Equation B.11) by changing E and L.

$$Q_{tot}(\mathbf{p}) = \mathbf{e}^T\mathbf{W}_e\mathbf{e} + \epsilon^2[\mathbf{p} - \langle\mathbf{p}\rangle]^T\mathbf{W}_p[\mathbf{p} - \langle\mathbf{p}\rangle] \tag{B.17}$$

As previously, the solution is best for the estimation when Equation B.17 is minimized.

Term 1:

$$\begin{aligned} &\frac{\partial}{\partial p_q}\Big[\Big(\sum_i^N d_i - \sum_j^M G_{ij}p_j\Big)W_{e,ii}\Big(\sum_i^N d_i - \sum_k^M G_{ik}p_k\Big)\Big] \\ &= 2\mathbf{p}\mathbf{G}^T\mathbf{G}\mathbf{W}_e 2\mathbf{G}^T\mathbf{d}\mathbf{W}_e \end{aligned} \tag{B.18}$$

The calculations were shown more clearly in Equation B.4.

Term 2:

$$
\begin{aligned}
&\epsilon^2 \frac{\partial}{\partial p_q}\Big[\sum_j^M \Big(p_j\langle p\rangle_j\Big) W_{p,jk} \sum_k^M \Big(p_k - \langle p\rangle_k\Big)\Big] \\
&= \epsilon^2 \frac{\partial}{\partial p_q}\Big[\sum_j^M p_j \sum_k^M p_k W_{p,jk} - \sum_j^M \sum_k^M p_j\langle p\rangle_k W_{p,jk} - \sum_j \langle p\rangle_j \sum_k^M p_k W_{p,jk} \\
&+ \sum_j^M \langle p\rangle_j \sum_k^M \langle p\rangle_k W_{p,jk}\Big] \\
&= \epsilon^2 \sum_j^M \sum_k^M W_{p,jk}\Big[\frac{\partial p_j p_k}{\partial p_q} - \frac{\partial p_j\langle p\rangle_k}{\partial p_q} - \frac{\partial \langle p\rangle_j p_k}{\partial p_q} + \frac{\partial \langle p\rangle_j\langle p\rangle_k}{\partial p_q}\Big] \\
&= \epsilon^2 \sum_j^M \sum_k^M W_{p,jk}\Big[\Big(\frac{\partial p_j}{\partial p_q}\cdot p_k + p_j \cdot \frac{\partial p_k}{\partial p_q}\Big) - \Big(\frac{\partial p_j}{\partial p_q}\cdot \langle p\rangle_k + p_j \cdot \frac{\partial \langle p\rangle_k}{\partial p_q}\Big) \\
&- \Big(\frac{\partial \langle p\rangle_j}{\partial p_q}\cdot p_k + \langle p\rangle_j \cdot \frac{\partial p_k}{\partial p_q}\Big) + \Big(\frac{\partial \langle p\rangle_j}{\partial p_q}\cdot \langle p\rangle_k + \langle p\rangle_j \cdot \frac{\partial \langle p\rangle_k}{\partial p_q}\Big)\Big] \\
&= \epsilon^2 \sum_j^M \sum_k^M W_{p,jk}\Big[\delta_{jq} p_k + p_j\delta_{kq} - \delta_{jq}\langle p\rangle_k - \langle p\rangle_j \delta_{kq}\Big] \\
&= \epsilon^2 \sum_k^M W_{p,kk}[p_k + p_k - \langle p\rangle_k - \langle p\rangle_k] \\
&= 2\epsilon^2 \sum_k^M W_{p,kk}(p_k - \langle p\rangle_k) \\
&= 2\epsilon^2 \mathbf{W}_p[\mathbf{p} - \langle \mathbf{p}\rangle]
\end{aligned}
\tag{B.19}
$$

To find the minimized solution Equations B.18 and B.19 are combined to:

$$
\frac{\partial Q_{tot}}{\partial p_q} = \mathbf{p}\mathbf{G}^T\mathbf{G}\mathbf{W}_e - \mathbf{G}^T\mathbf{d}\mathbf{W}_e + \epsilon^2 \mathbf{W}_m[\mathbf{p} - \langle \mathbf{p}\rangle] = 0 \tag{B.20}
$$

From Equation B.20 it is possible to find the weighted damped least squares solution with *a priori* information:

$$
\begin{aligned}
\mathbf{p}\mathbf{G}^T\mathbf{G}\mathbf{W}_e - \mathbf{G}^T\mathbf{d}\mathbf{W}_e + \epsilon^2\mathbf{W}_m[\mathbf{p} - \langle\mathbf{p}\rangle] &= 0 \\
\mathbf{G}^T\mathbf{W}_e\mathbf{G}\mathbf{p} - \mathbf{G}^T\mathbf{W}_e\mathbf{d} + \epsilon^2\mathbf{W}_p[\mathbf{p} - \langle\mathbf{p}\rangle] - \mathbf{G}^T\mathbf{W}_e\mathbf{G}\langle\mathbf{p}\rangle + \mathbf{G}^T\mathbf{W}_e\mathbf{G}\langle\mathbf{p}\rangle &= 0 \\
\mathbf{G}^T\mathbf{W}_e\mathbf{G}[\mathbf{p} - \langle\mathbf{p}\rangle] + \epsilon^2\mathbf{W}_p[\mathbf{p} - \langle\mathbf{p}\rangle] - \mathbf{G}^T\mathbf{W}_e\mathbf{d} + \mathbf{G}^T\mathbf{W}_e\mathbf{G}\langle\mathbf{p}\rangle &= 0 \\
[\mathbf{p} - \langle\mathbf{p}\rangle]\left[\mathbf{G}^T\mathbf{W}_e\mathbf{G} + \epsilon^2\mathbf{W}_p\right] &= \mathbf{G}^T\mathbf{W}_e[\mathbf{d} - \mathbf{G}\langle\mathbf{p}\rangle] \\
[\mathbf{p} - \langle\mathbf{p}\rangle] &= \left[\mathbf{G}^T\mathbf{W}_e\mathbf{G} + \epsilon^2\mathbf{W}_p\right]^{-1}\mathbf{G}^T\mathbf{W}_e[\mathbf{d} - \mathbf{G}\langle\mathbf{p}\rangle] \\
&\Rightarrow \\
\mathbf{p}^{est} &= \langle\mathbf{p}\rangle + \left[\mathbf{G}^T\mathbf{W}_e\mathbf{G} + \epsilon^2\mathbf{W}_p\right]^{-1}\mathbf{G}^T\mathbf{W}_e[\mathbf{d} - \mathbf{G}\langle\mathbf{p}\rangle]
\end{aligned}
\tag{B.21}
$$

The weights are implemented through the covariance matrices by assuming a Gaussian distribution on the data and the *a priori* model parameters, and using the implicit constraint $\mathbf{f}(\mathbf{d}, \mathbf{p}) = 0$.

$\mathbf{x} = [\mathbf{d}, \mathbf{p}]^T$, and the covariance matrix, C_x, which contains both C_d and C_p are defined. The vector $\mathbf{x}$ then has the length $N + M = S$.

The total error is then redefined to:

$$Q_{tot} = [\mathbf{x} - \langle\mathbf{x}\rangle]^T C_x^{-1}[\mathbf{x} - \langle\mathbf{x}\rangle] \tag{B.22}$$

Minimizing Equation B.22 is different in this case because of the constraint: $\mathbf{f}(\mathbf{d}, \mathbf{p}) = \mathbf{f}(\mathbf{x}) = 0$ with length $n \leq S$. Because of this, the minimization is performed using Lagrange multipliers, a method for deciding extreme values for a function with more than one variable and with one or more constraints. The basic equation is:

$$\nabla(a + \lambda b) = 0 \tag{B.23}$$

where a is the equation we want to minimize, b is the constraint and λ is the Lagrange multiplier. Applying Equation B.23 to this problem we get:

$$\frac{\partial Q_{tot}}{\partial x_i} + \left(\frac{\partial \lambda f_j}{\partial x_i}\right) = 0 \tag{B.24}$$

The differentiations is performed on the terms of Equation B.24 individually and by a variable x_q.

Term 1:

$$
\begin{aligned}
\frac{\partial Q_{tot}}{\partial x_q} &= \frac{\partial}{\partial x_q}\Big[\sum_{j}^{S}(x_j - \langle x\rangle_j)C_x^{-1}\sum_{k}^{S}(x_k - \langle x\rangle_k)\Big] \\
&= C_x^{-1}\frac{\partial}{\partial x_q}\Big[\sum_{j}^{S}x_j\sum_{k}^{S}x_k - \sum_{j}^{S}x_j\sum_{k}^{S}\langle x\rangle_k - \sum_{j}^{S}\langle x\rangle_j\sum_{k}^{S}x_k \\
&+ \sum_{k}^{S}\sum_{j}^{S}\langle x\rangle_j\langle x\rangle_k\Big] \\
&= C_x^{-1}\sum_{j}^{S}\sum_{k}^{S}\Big[\frac{\partial(x_jx_k)}{\partial x_q} - \frac{\partial(x_j\langle x\rangle_k)}{\partial x_q} - \frac{\partial(\langle x\rangle_j x_k)}{\partial x_q} + \frac{\partial(\langle x\rangle_j\langle x\rangle_k)}{\partial x_q}\Big] \\
&= C_x^{-1}\sum_{j}^{S}\sum_{k}^{S}\Big[(\frac{\partial x_j}{\partial x_q}\cdot x_k + x_j\cdot\frac{\partial x_k}{\partial x_q}) - (\frac{\partial x_j}{\partial x_q}\cdot\langle x\rangle_k + x_j\cdot\frac{\partial\langle x\rangle_k}{\partial x_q}) \\
&- (\frac{\partial\langle x\rangle_j}{\partial x_q}\cdot x_k + \langle x\rangle_j\cdot\frac{\partial x_k}{\partial x_q}) + (\frac{\partial\langle x\rangle_j}{\partial x_q}\cdot\langle x\rangle_k + \langle x\rangle_j\cdot\frac{\partial\langle x\rangle_k}{\partial x_q})\Big] \\
&= C_x^{-1}\sum_{j}^{S}\sum_{k}^{S}\Big[\delta_{jq}x_k + x_j\delta_{kq} - \delta_{jq}\langle x\rangle_k - \langle x\rangle_j\delta_{kq}\Big] \\
&= C_x^{-1}\sum_{j}^{S}[x_j + x_j - \langle x\rangle_j - \langle x\rangle_j] \\
&= 2\cdot C_x^{-1}\sum_{j}^{S}[x_j - \langle x\rangle_j] \\
&\Rightarrow 2[\mathbf{x} - \langle\mathbf{x}\rangle]^T C_x^{-1}
\end{aligned}
\tag{B.25}
$$

Term 2:

$$
\begin{aligned}
\frac{\partial(\lambda f_j)}{\partial x_q} &= \frac{\partial}{\partial x_q}\Big(\sum_{j}^{n}\lambda f_j\Big) \\
&= \lambda\sum_{j}^{n}\frac{\partial f_j}{\partial x_q}
\end{aligned}
\tag{B.26}
$$

Combining the results from the two terms together in Equation B.24 gives:

$$
\begin{aligned}
\frac{\partial Q_{tot}}{\partial x_i} + \sum_{j}^{n}\lambda\frac{\partial f_j}{\partial x_i} &= 0 \\
2[\mathbf{x} - \langle\mathbf{x}\rangle]^T C_x^{-1} &= -\lambda^T\mathbf{F}
\end{aligned}
\tag{B.27}
$$

where $\mathbf{F}$ is a matrix of partial derivatives.

The next step is to eliminate the Lagrange multipliers (λ).

$$\begin{aligned} 2[\mathbf{x} - \langle\mathbf{x}\rangle]^T C_x^{-1} &= -\lambda^T \mathbf{F} \\ [\mathbf{x} - \langle\mathbf{x}\rangle]^T &= -\frac{1}{2}\lambda^T C_x \mathbf{F} \\ [\mathbf{x} - \langle\mathbf{x}\rangle] &= -\frac{1}{2}\lambda C_x \mathbf{F}^T (*) \\ \mathbf{F}[\mathbf{x} - \langle\mathbf{x}\rangle] &= -\frac{1}{2}\lambda \mathbf{F} C_x \mathbf{F}^T \\ \Rightarrow \lambda &= -2[\mathbf{F} C_x \mathbf{F}^T]^{-1} \mathbf{F}[\mathbf{x} - \langle\mathbf{x}\rangle] \end{aligned} \tag{B.28}$$

To remove the Lagrange multipliers, the final result in Equation B.28 is implemented into the equation marked with $(*)$ in Equation B.28:

$$\begin{aligned} [\mathbf{x} - \langle\mathbf{x}\rangle] &= -\frac{1}{2}\Big(- 2[\mathbf{F} C_x \mathbf{F}^T]^{-1} \mathbf{F}[\mathbf{x} - \langle\mathbf{x}\rangle]\Big) C_x \mathbf{F}^T \\ [\mathbf{x} - \langle\mathbf{x}\rangle] &= [\mathbf{F} C_x \mathbf{F}^T]^{-1} \mathbf{F}[\mathbf{x} - \langle\mathbf{x}\rangle] C_x \mathbf{F}^T \end{aligned} \tag{B.29}$$

Equation B.29 needs to be solved simultaneously as the constraint equation ($\mathbf{f}(\mathbf{x}) = 0$) to be minimized, this is done by combining them in the following manner:

$$[\mathbf{x} - \langle\mathbf{x}\rangle] = C_x \mathbf{F}^T [\mathbf{F} C_x \mathbf{F}^T]^{-1} \Big(\mathbf{F}[\mathbf{x} - \langle\mathbf{x}\rangle] - \mathbf{f}(\mathbf{x}) \Big) \tag{B.30}$$

To make Equation B.30 to an iterative linearized solution it is redefined, where i denotes the iteration.

$$\begin{aligned} \mathbf{x}_{i+1}^{est} - \langle\mathbf{x}\rangle &= C_x \mathbf{F}_i^T [\mathbf{F}_i C_x \mathbf{F}_i^T]^{-1} \Big(\mathbf{F}_i [\mathbf{x}_i^{est} - \langle\mathbf{x}\rangle] - \mathbf{f}(\mathbf{x}_i^{est}) \Big) \\ \mathbf{x}_{i+1}^{est} &= \langle\mathbf{x}\rangle + C_x \mathbf{F}_i^T [\mathbf{F}_i C_x \mathbf{F}_i^T]^{-1} \Big(\mathbf{F}_i [\mathbf{x}_i^{est} - \langle\mathbf{x}\rangle] - \mathbf{f}(\mathbf{x}_i^{est}) \Big) \end{aligned} \tag{B.31}$$

where $\mathbf{F}_i = \sum_j^S \frac{\partial f_j}{\partial x_i}$. In Equation B.31 it is clear that the estimated results depend on the results from the previous iteration and the *a priori* knowledge.

To estimate the model parameters, the solution is found with the explicit constraint $\mathbf{f}(\mathbf{x}) = \mathbf{d} - \mathbf{g}(\mathbf{p}) = 0$, and assuming that the *a priori* model parameters are uncorrelated (there are no linear relationships between the parameters). To rewrite Equation B.31 to an explicit form the following relationships from Tarantola and Valette (1982a) are defined:

$$\mathbf{x} = [\mathbf{d}, \mathbf{p}]^T$$
$$\mathbf{F} = \begin{bmatrix} \mathbf{I} & -\mathbf{G} \end{bmatrix}$$
$$C_x = \begin{bmatrix} C_d & C_{dp} \\ C_{pd} & C_p \end{bmatrix}$$

Assuming that uncertainties in the data ($\mathbf{d}$) are independent of the uncertainties in the model parameters ($\mathbf{p}$) leads to:

$$C_{dp} = (C_{pd}^T) = 0 \tag{B.32}$$

Note: To not use any *a priori* information on the data, we define that for iteration number $i+1$, $\langle \mathbf{d} \rangle = \mathbf{d}_i^{est}$.

Implementing these relationships into Equation B.31 leads to:

$$\begin{aligned}
\begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix}_{i+1}^{est} &= \left\langle \begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix} \right\rangle + \begin{bmatrix} C_d & 0 \\ 0 & C_p \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ -\mathbf{G}_i^T \end{bmatrix} \left(\begin{bmatrix} \mathbf{I} & -\mathbf{G}_i \end{bmatrix} \begin{bmatrix} C_d & 0 \\ 0 & C_p \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ -\mathbf{G}_i^T \end{bmatrix} \right)^{-1} \\
&\cdot \left(\begin{bmatrix} \mathbf{I} & -\mathbf{G}_i \end{bmatrix} \left(\begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix}_i^{est} - \left\langle \begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix} \right\rangle \right) - (\mathbf{d} - \mathbf{g}(\mathbf{p}_i^{est})) \right) \\
&= \begin{bmatrix} \langle \mathbf{d} \rangle \\ \langle \mathbf{p} \rangle \end{bmatrix} + \begin{bmatrix} C_d \\ -\mathbf{G}_i^T C_p \end{bmatrix} \left(\begin{bmatrix} C_d & -\mathbf{G}_i C_p \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ -\mathbf{G}_i^T \end{bmatrix} \right)^{-1} \\
&\cdot \left(\begin{bmatrix} \mathbf{I} & -\mathbf{G}_i \end{bmatrix} \begin{bmatrix} \mathbf{d}_i^{est} - \langle \mathbf{d} \rangle \\ \mathbf{p}_i^{est} - \langle \mathbf{p} \rangle \end{bmatrix} - (\mathbf{d} - \mathbf{g}(\mathbf{p}_i^{est})) \right) \\
&= \begin{bmatrix} \langle \mathbf{d} \rangle \\ \langle \mathbf{p} \rangle \end{bmatrix} + \begin{bmatrix} C_d \\ -\mathbf{G}_i^T C_p \end{bmatrix} \left(C_d + \mathbf{G} C_p \mathbf{G}^T \right)^{-1} \\
&\cdot \left(\left((\mathbf{d}_i^{est} - \langle \mathbf{d} \rangle) - \mathbf{G}_i^T C_p (\mathbf{p}_i^{est} - \langle \mathbf{p} \rangle) \right) - (\mathbf{d} - \mathbf{g}(\mathbf{p}_i^{est})) \right)
\end{aligned} \tag{B.33}$$

Working further from Equation B.33 the focus is on the estimation of the model parameters, by setting $\mathbf{d}_i^{est} - \langle \mathbf{d} \rangle = 0$.

$$\begin{aligned}
\mathbf{p}_{i+1}^{est} &= \langle \mathbf{p} \rangle - \mathbf{G}_i^T C_p \left[C_d + \mathbf{G}_i C_p \mathbf{G}_i^T \right]^{-1} \left(-\mathbf{G}_i [\mathbf{p}_i^{est} - \langle \mathbf{p} \rangle] - [\mathbf{d} - \mathbf{g}(\mathbf{p}_i^{est})] \right) \\
&= \langle \mathbf{p} \rangle + \mathbf{G}_i^T C_p \left[C_d + \mathbf{G}_i C_p \mathbf{G}_i^T \right]^{-1} \mathbf{G}_i [\mathbf{p}_i^{est} - \langle \mathbf{p} \rangle] \\
&+ \mathbf{G}_i^T C_p \left[C_d + \mathbf{G}_i C_p \mathbf{G}_i^T \right]^{-1} [\mathbf{d} - \mathbf{g}(\mathbf{p}_i^{est})] \\
&= \langle \mathbf{p} \rangle + \mathbf{G}_i^T C_p \left[C_d + \mathbf{G}_i C_p \mathbf{G}_i^T \right]^{-1} \left(\mathbf{G}_i [\mathbf{p}_i^{est} - \langle \mathbf{p} \rangle] + \mathbf{d} - \mathbf{g}(\mathbf{p}_i^{est}) \right)
\end{aligned} \tag{B.34}$$

To modify Equation B.34 further, the following identities from Tarantola and Valette (1982b) are defined:

$$\mathbf{G}^T C_d^{-1}[\mathbf{G}^T C_d^{-1}\mathbf{G} + C_p^{-1}]^{-1} = [C_d + \mathbf{G}C_p\mathbf{G}^T]^{-1}C_p\mathbf{G}^T \tag{B.35}$$

$$[\mathbf{G}^T C_d^{-1}\mathbf{G}^T + C_p^{-1}]^{-1} = C_p - C_p\mathbf{G}^T[C_d + \mathbf{G}C_p\mathbf{G}^T]^{-1}\mathbf{G}C_p \tag{B.36}$$

Using the identity in Equation B.35 in Equation B.34 gives:

$$\mathbf{p}_i^{est} = \langle\mathbf{p}\rangle + [\mathbf{G}_i^T C_d^{-1}\mathbf{G}_i + C_p^{-1}]^{-1}\mathbf{G}_i^T C_d^{-1}\big(\mathbf{d} - \mathbf{g}(\mathbf{p}_i^{est}) + \mathbf{G}_i[\mathbf{p}_i^{est} - \langle\mathbf{p}\rangle]\big) \tag{B.37}$$

Equation B.37 is the linearized, iterative, weighted least squares solution.

To make each iteration depend on the previous iteration and use a starting model ($\mathbf{p}_0$), which can be dependent on *a priori* knowledge or not, the *a priori* model is set to be the previous iteration. Except for the last term where the starting model is used. Then, the difference between the previous iteration and the start model will be used to find the new iteration.

The following relationship from Tarantola and Valette (1982b) will also be used:

$$C_p^{-1} = \mathbf{G}^T C_d^{-1}\mathbf{G} \tag{B.38}$$

The relationship defined in Equation B.38 is for linear problems, however if the non-linearity is weak Equation B.38 becomes an approximation which can be used.

Equation B.37 then becomes:

$$\mathbf{p}_i^{est} = \mathbf{p}_i + [\mathbf{G}_i^T C_d^{-1}\mathbf{G}_i + C_p^{-1}]^{-1}\big(\mathbf{G}_i^T C_d^{-1}(\mathbf{d} - \mathbf{g}(\mathbf{p}_i^{est})) + C_p^{-1}[\mathbf{p}_i^{est} - \mathbf{p}_0]\big) \tag{B.39}$$

To simplify Equation B.39 the following is defined:

$$\begin{aligned} \mathbf{M} &= [\mathbf{G}^T C_d^{-1}\mathbf{G} + C_p^{-1}]^{-1} \\ \Delta\mathbf{d} &= \mathbf{d} - \mathbf{g}(\mathbf{p}^{est}) \\ \Delta\mathbf{p} &= \mathbf{p}^{est} - \mathbf{p}_0 \end{aligned} \tag{B.40}$$

Equation B.39 can be simplified, using the formulas in Equation B.40:

$$\mathbf{p}_{i+1} = \mathbf{p}_i + \mathbf{M}[\mathbf{G}^T C_d^{-1}\Delta\mathbf{d} + C_p^{-1}\Delta\mathbf{p}] \tag{B.41}$$

C Testing Unique Starting Models

Here shortly describe the results when using particular starting models. The unique starting models I tested were:

1. A starting model with evenly distributed layers
2. A starting model with a layer in the mantle
3. A starting model with no Moho.

C.1 Evenly Distributed Layers

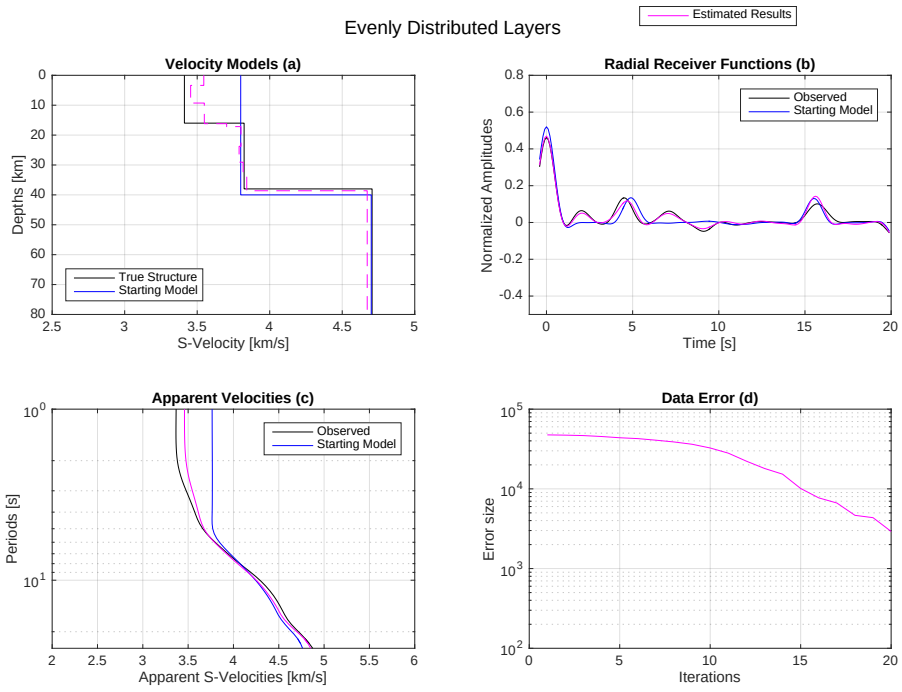


Figure 32: Results when using a starting model with evenly distributed layers. **(a)** The estimated velocity structure compared to the true synthetic and starting model. **(b)** The estimated, observed and starting radial receiver functions. **(c)** The estimated, observed and starting apparent S-velocities. **(d)** The data error.

For the starting model with evenly distributed layers I defined 8 layers over a half-space with a thickness of 5 km each. This placed the Moho at 40 km.

For the velocities I used global averages and used two S-velocities, one for the crust and one for the mantle. The results are seen in Figure 32.

Figure 32 show that the data and model parameters are well resolved. Although the resulting model parameters contain as many layers as the starting model, the Moho and crustal depths are well resolved.

C.2 Layer in the Mantle

To add a layer in the mantle, the Moho peak was located and peaks were fabricated at 1 second before and 2 seconds after it. These delay times correspond to layer depths of approximately 10 km above the Moho and 20 km below the Moho. The velocities I used were global averages, one for the crust and one for the mantle. The inversion results are shown in Figure 33.

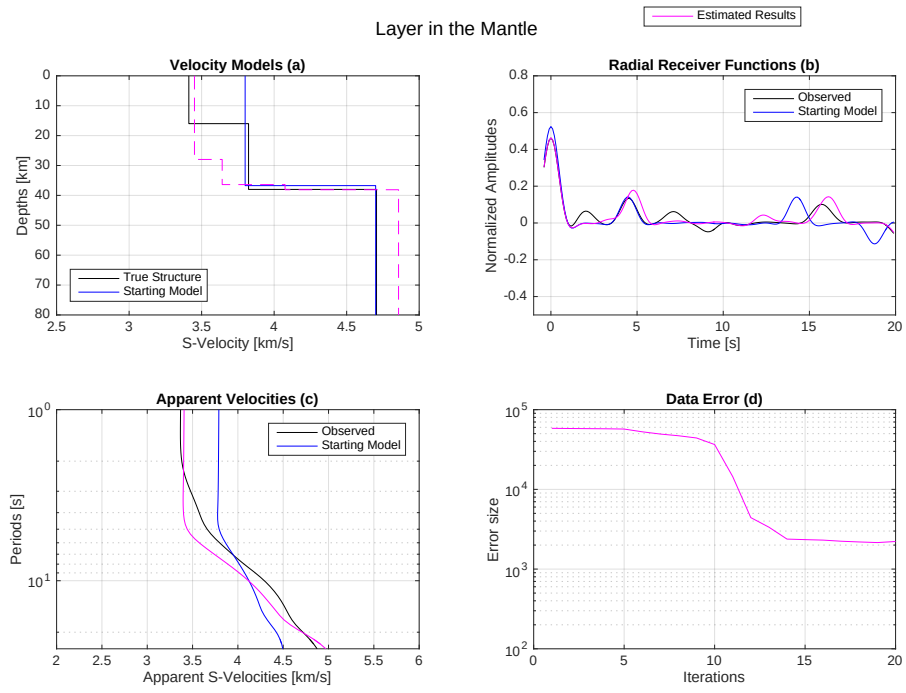


Figure 33: Results when using a starting model with a layer in the mantle. Description of individual plots seen in Figure 32.

Figure 33 show that the Moho peak is well resolved, although slightly shifted. The resulting velocity structure only tries to fit the Moho peak on the re-

ceiver function, while the apparent S-velocities for short periods are well resolved.

C.3 Invisible Moho

When creating a starting model with no visible Moho I used all the peaks picked from the observed radial receiver function. However, I only defined one crustal velocity for all the layers. This velocity was based on global averages. The results are shown in Figure 34.

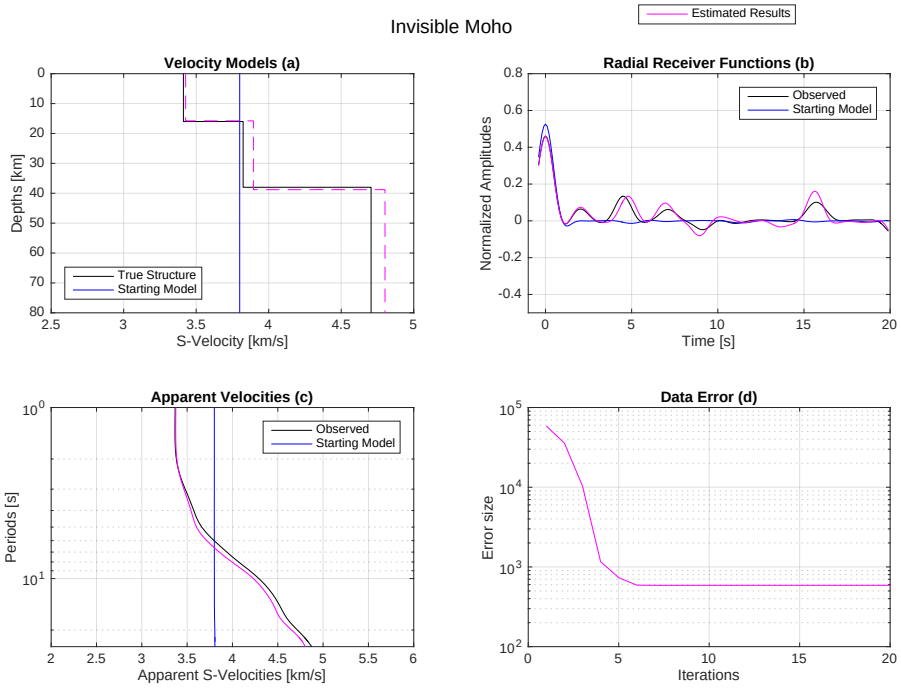


Figure 34: Results when using a starting model with no Moho discontinuity. Description of individual plots are in Figure 32.

In Figure 34 it is shown that the estimated data and model parameters when using a starting model with no Moho, fit the observed data extremely well. Although there was no Moho defined in the starting model, the inversion located and fitted the Moho and velocities from the observed data.

D Detailed Inversion Results for all the Stations

For all the figures, the colour of the stapled line shows the estimations as the iterations progress. The earliest iterations are in dark blue, developing towards yellow. The solid blue lines show the starting model and the starting data, while the solid black lines represent the observed data. For the error the solid black line represents the total data error, which is the receiver function error (in blue) and the apparent S-velocity error (in red) combined.

On the top line of plots, the first plot shows the S-velocity structures. One can see how the estimation develops from the starting model. The two next plots display the data, the receiver functions and the apparent S-velocities respectively. These can be compared to the observed data.

On the bottom line of plots, the first plot shows how the data error develops as the iterations progress. Here one can also see how many iterations were required for the station. The two next plots show the difference between the observed and estimated data, for the radial receiver functions and the apparent S-velocities, respectively.

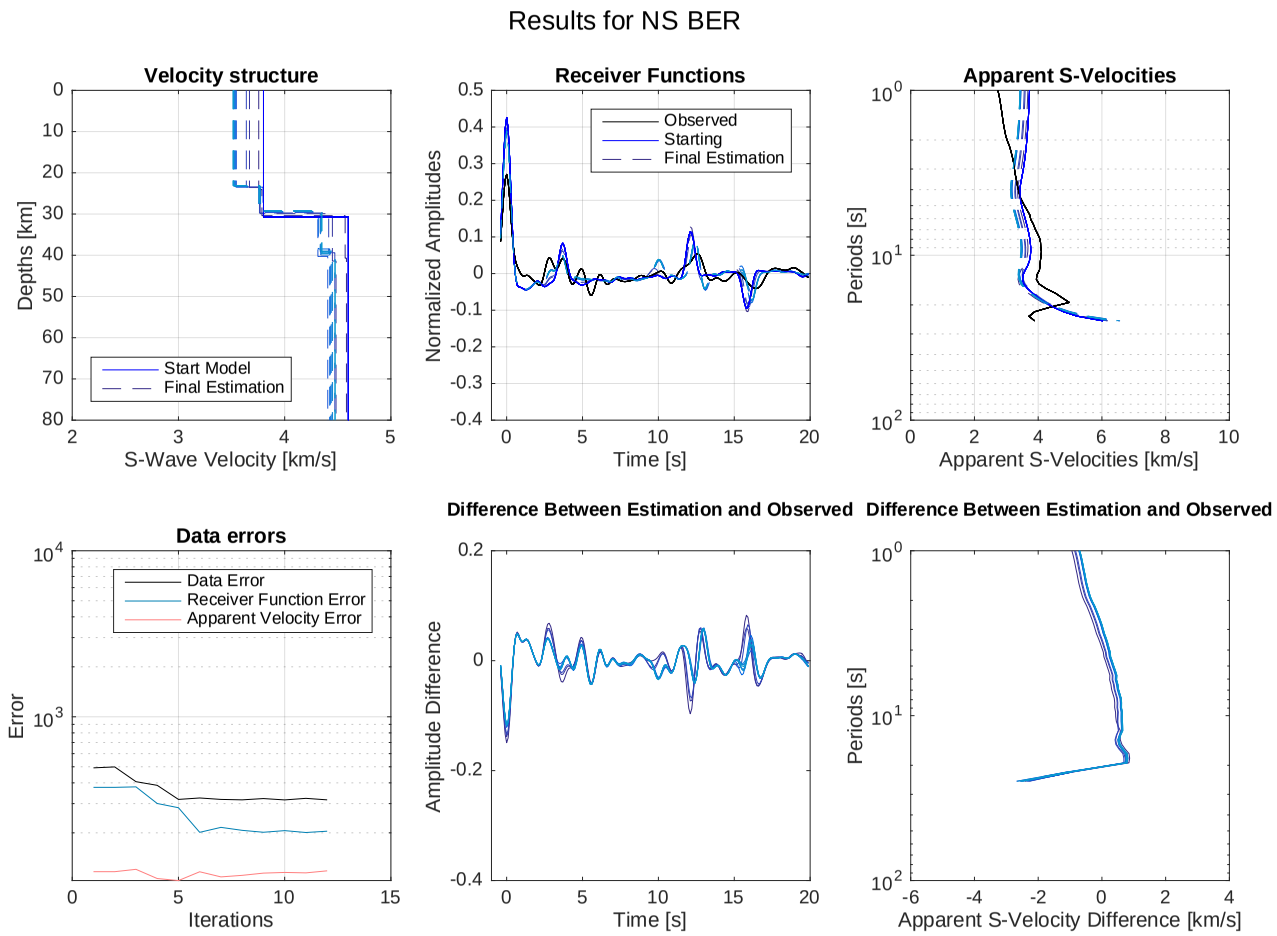


Figure 35: Station NS_BER

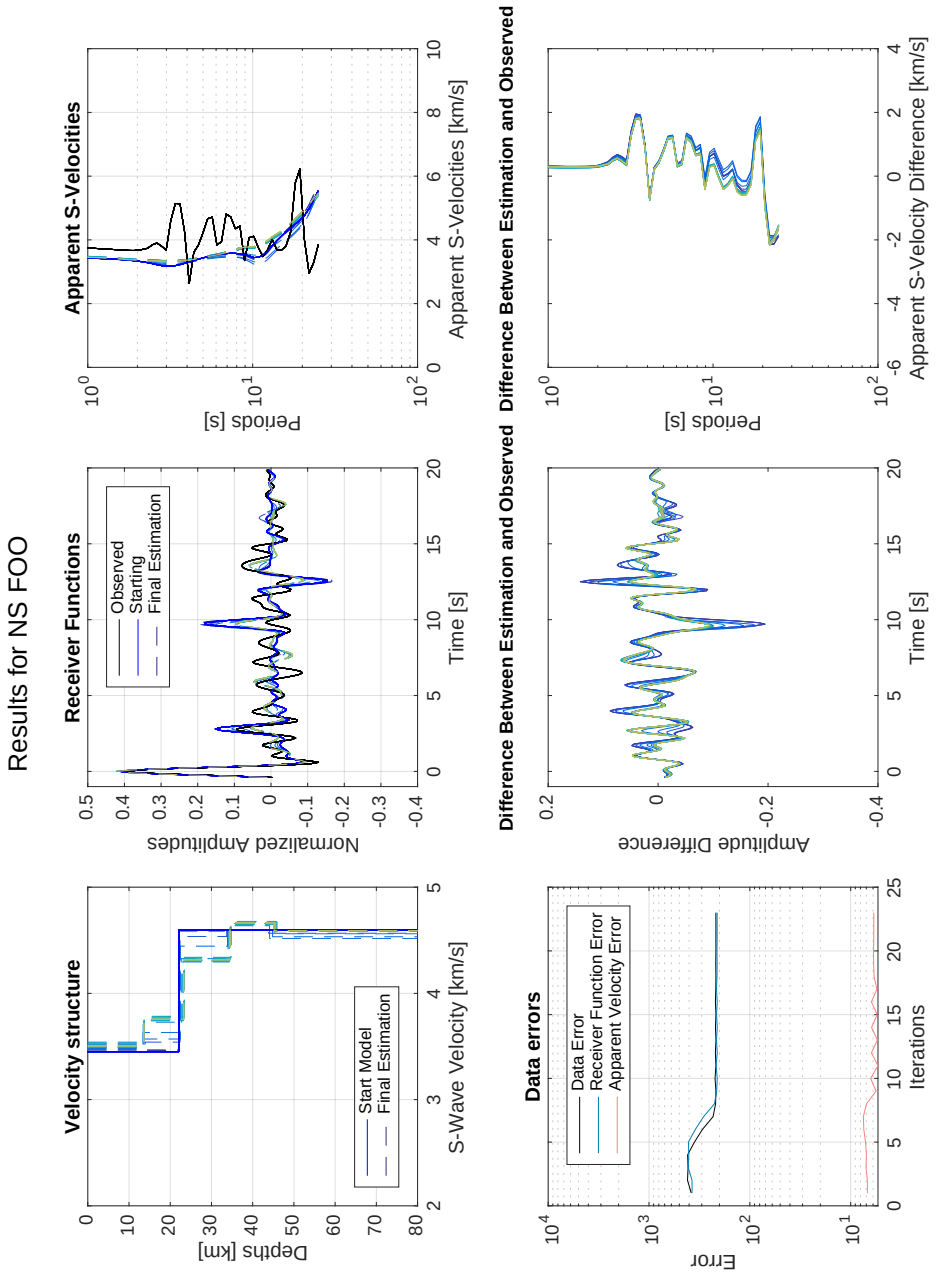


Figure 36: Station NS_FOO

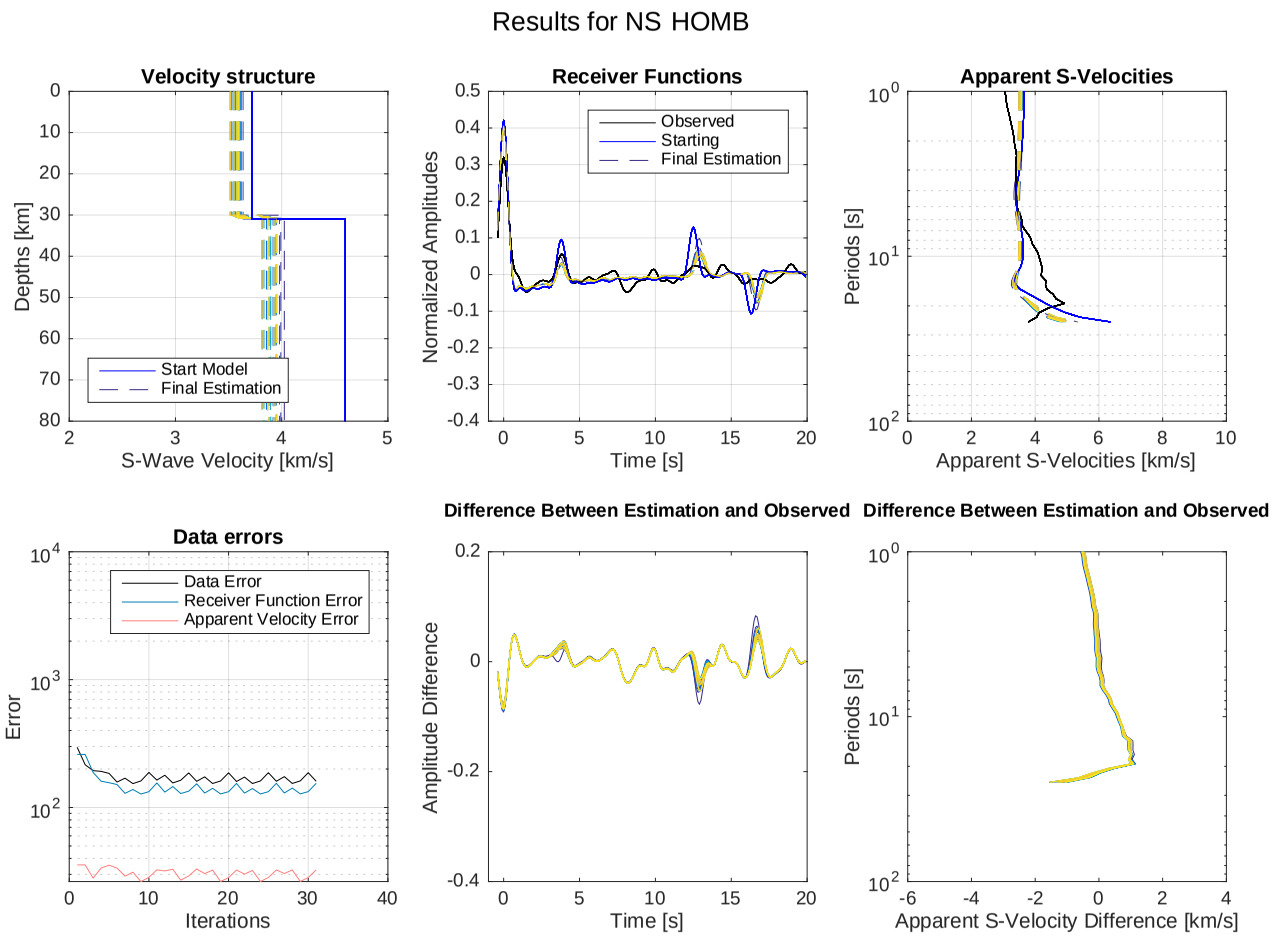


Figure 37: Station NS_HOMB

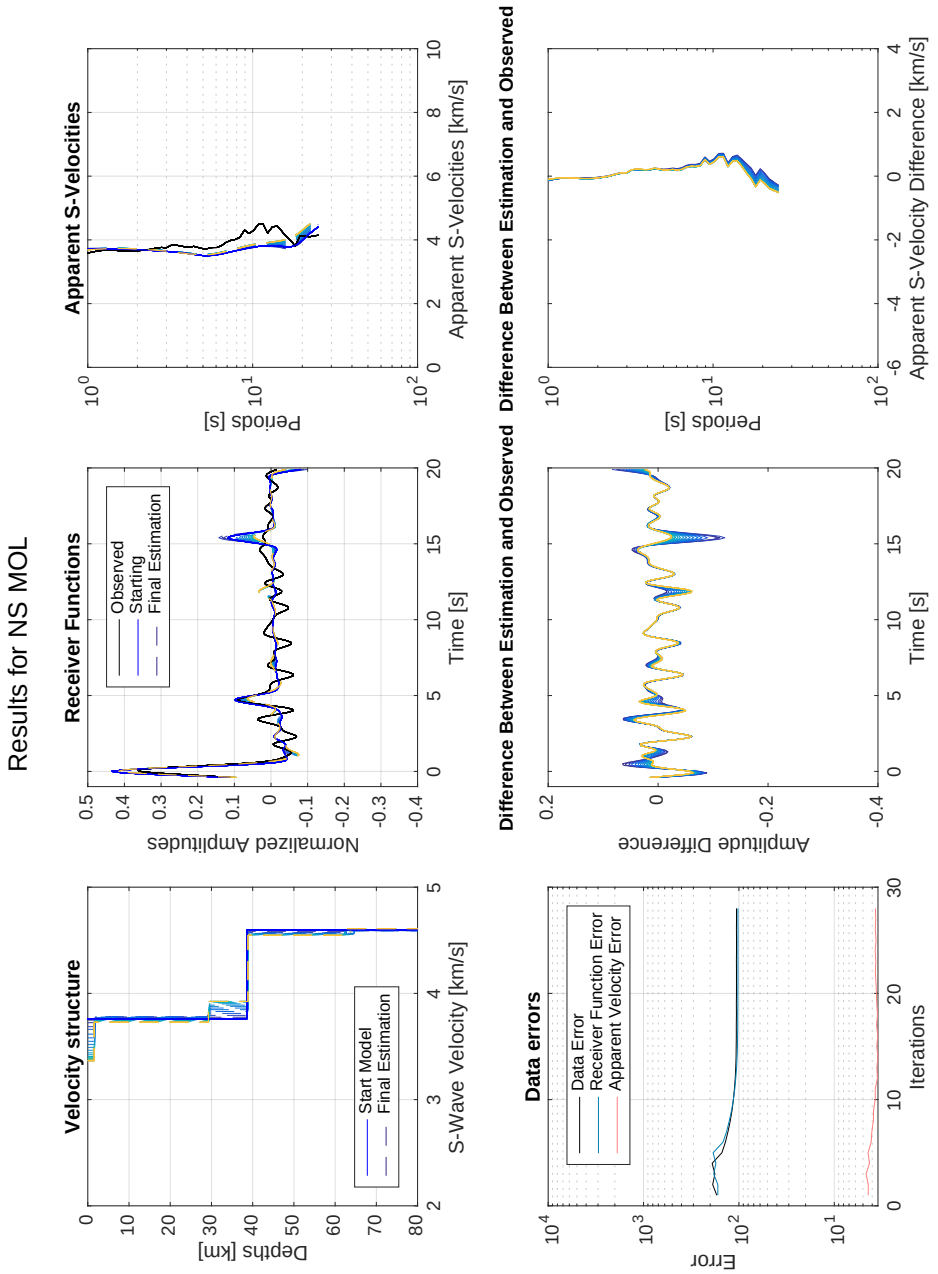


Figure 38: Station NS_MOL

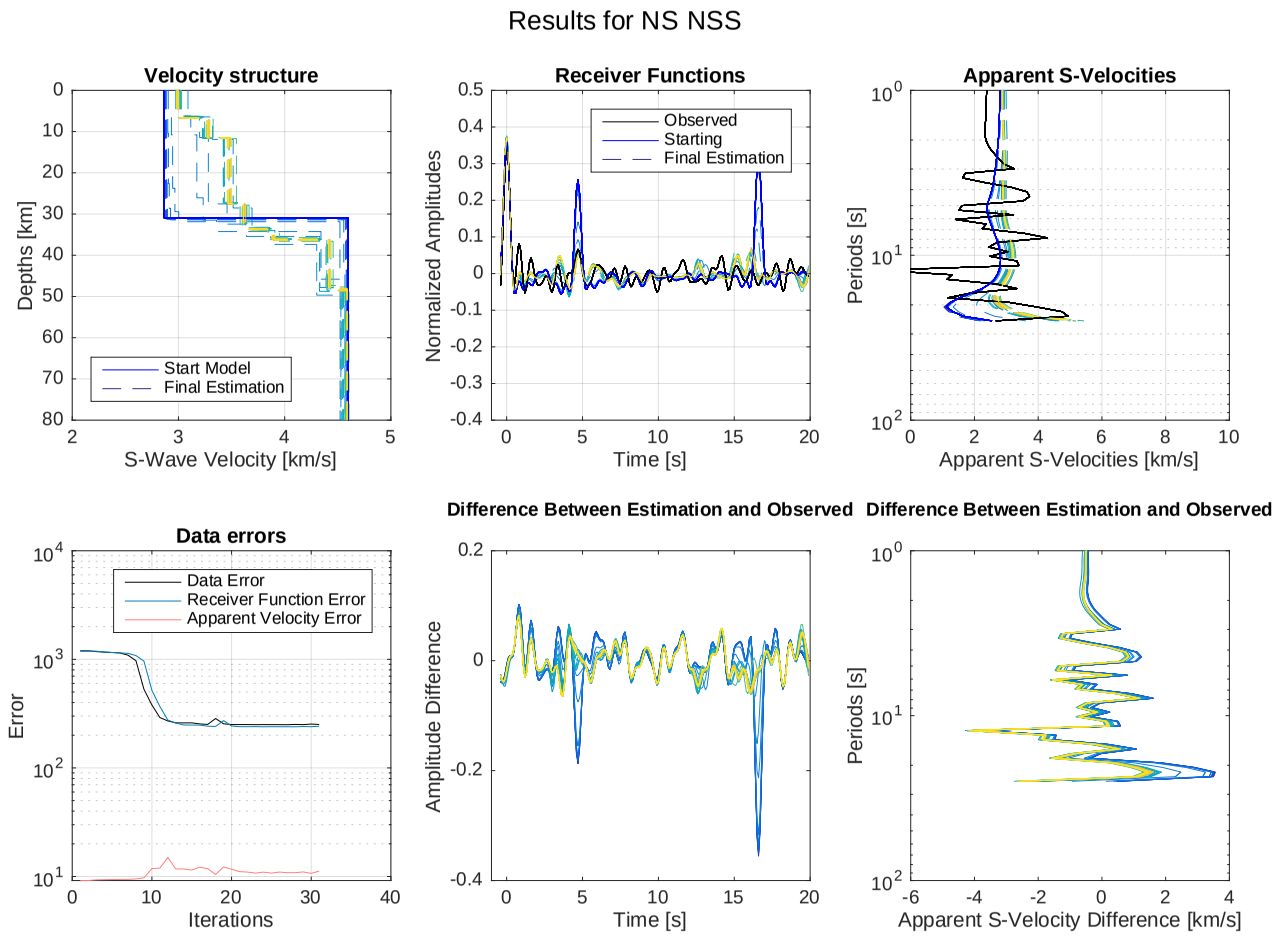


Figure 39: Station NS_NSS

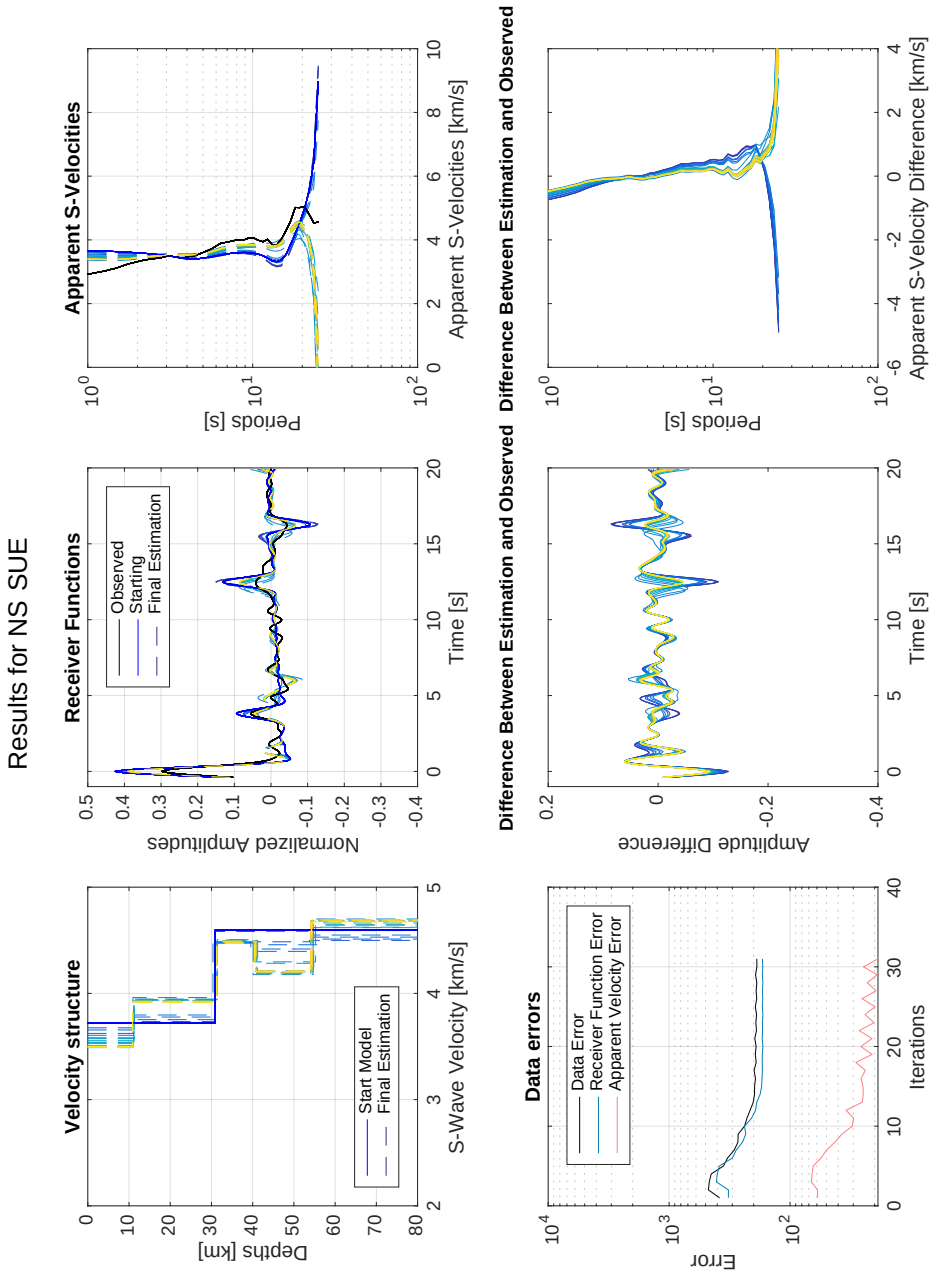


Figure 40: Station NS_SUE

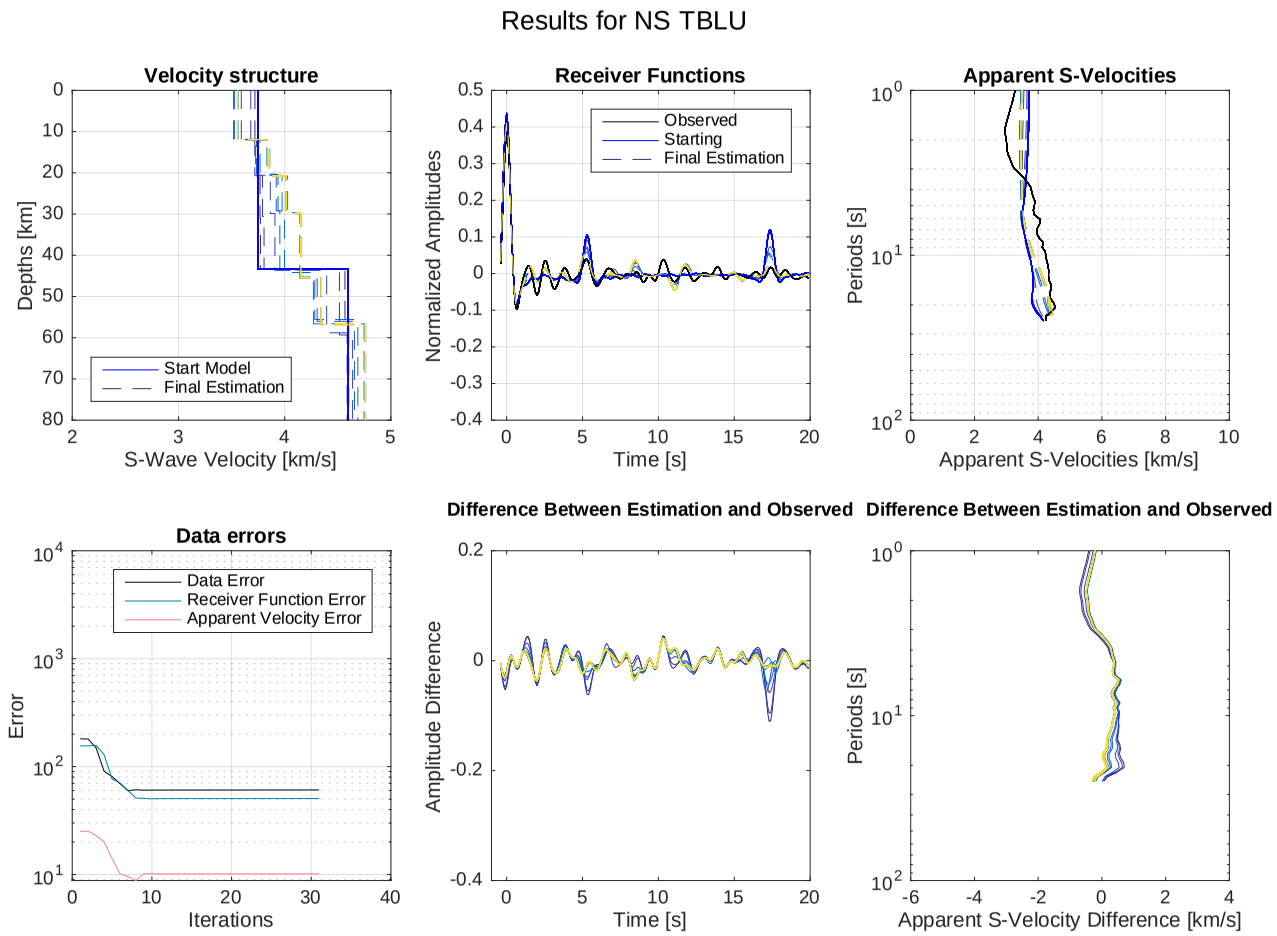


Figure 41: Station NS_TBLU

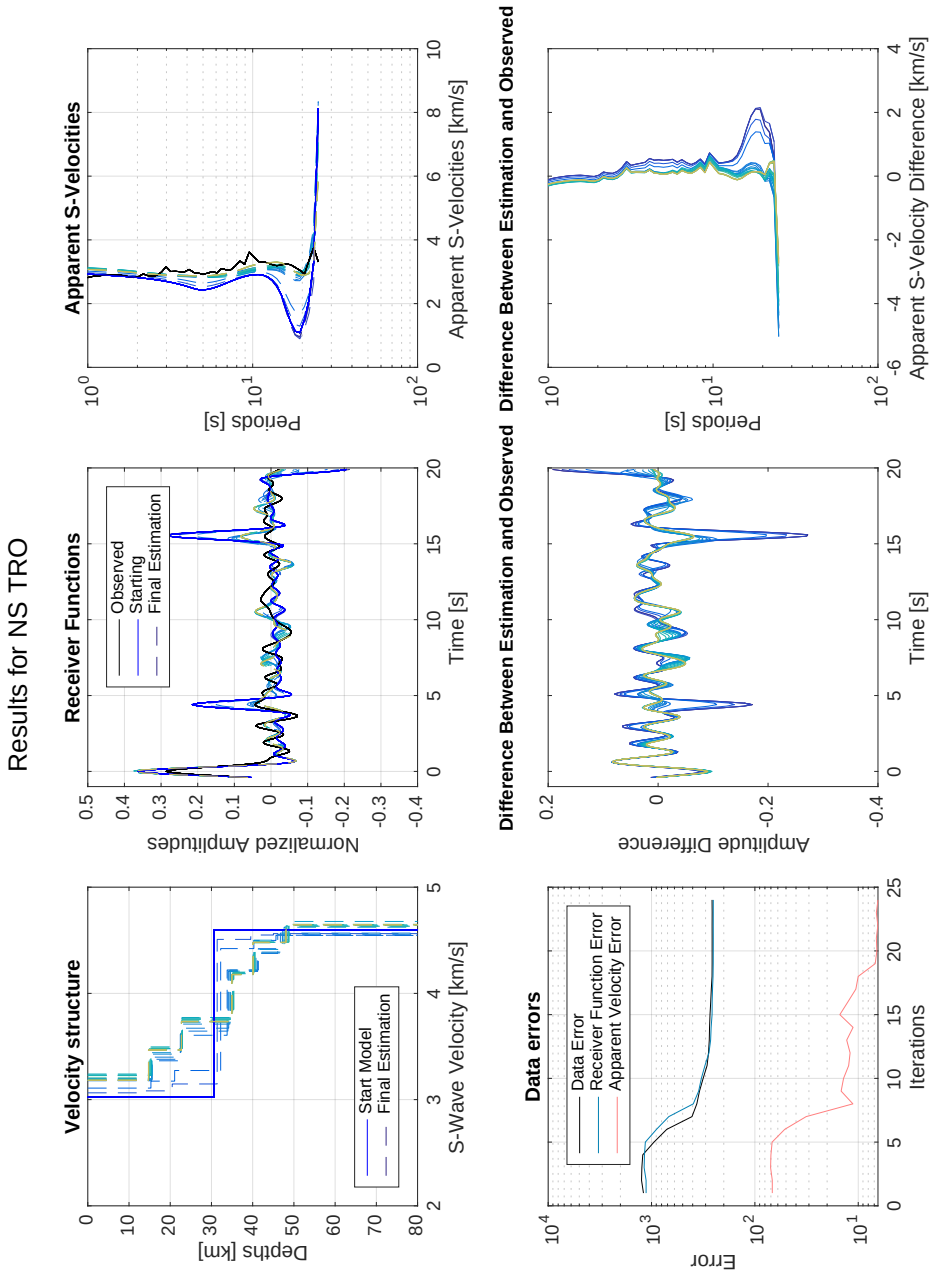


Figure 42: Station NS_TRO

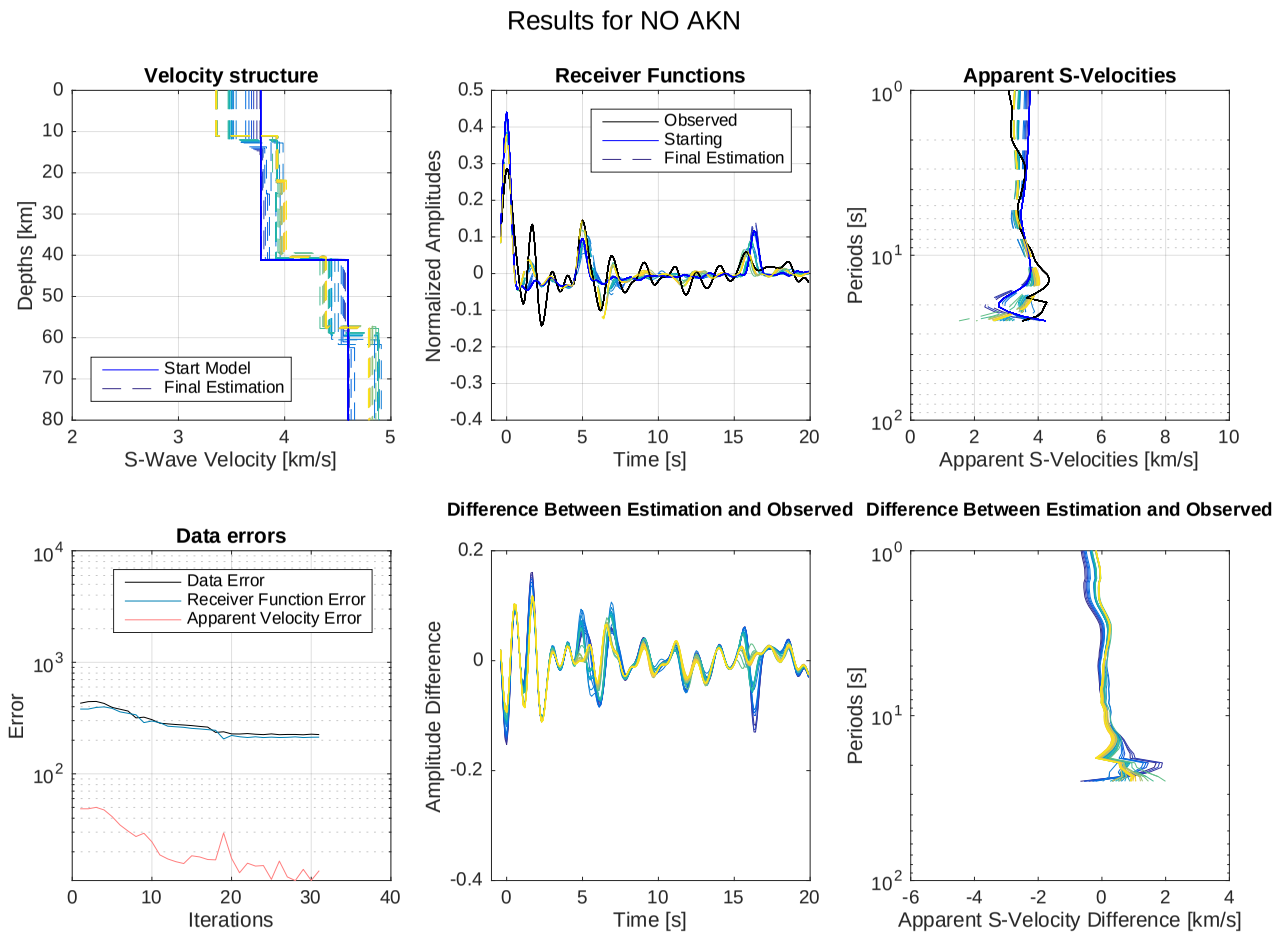


Figure 43: Station NO_AKN

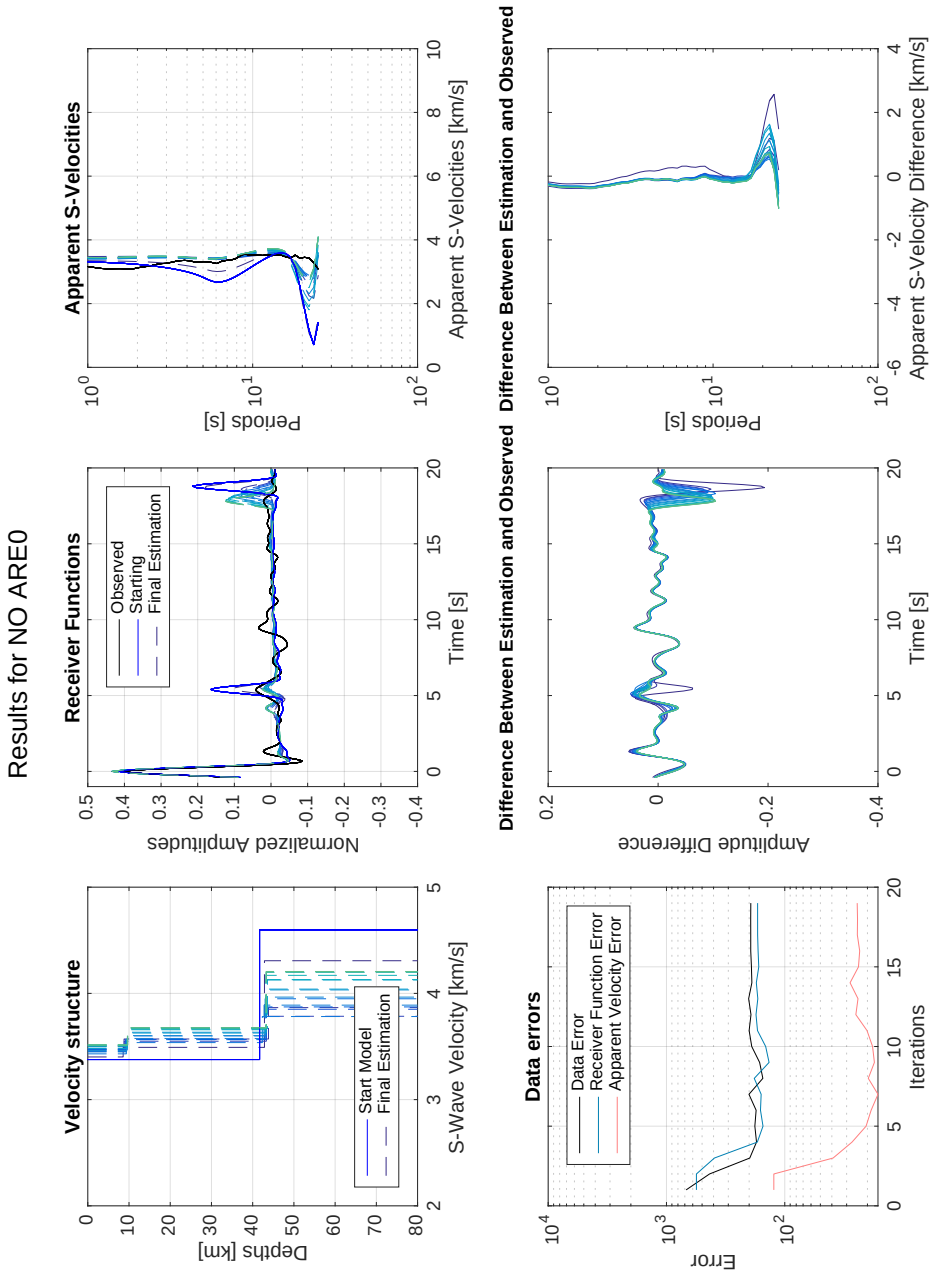


Figure 44: Station NO_ARE0

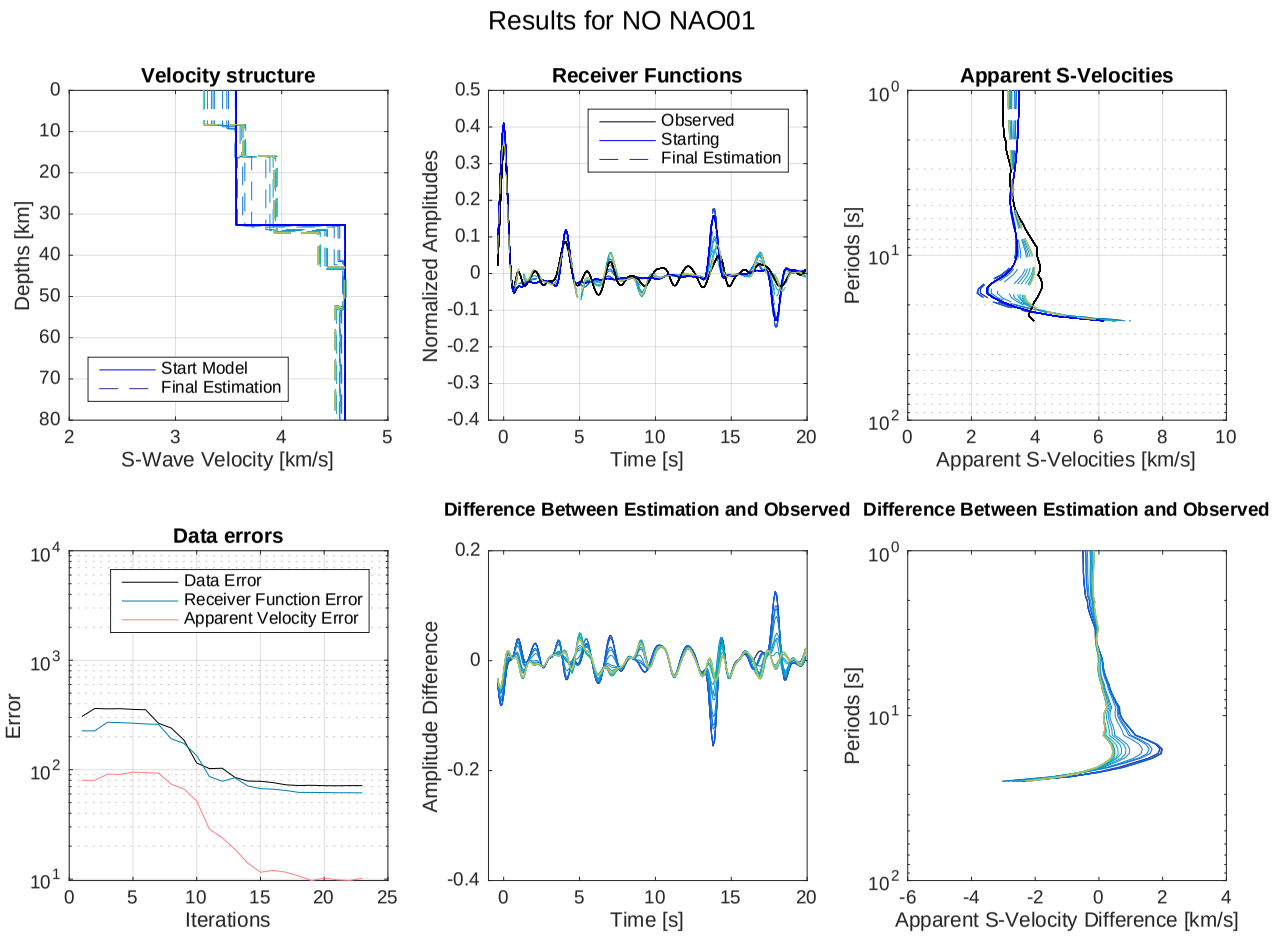


Figure 45: Station NO_NAO01

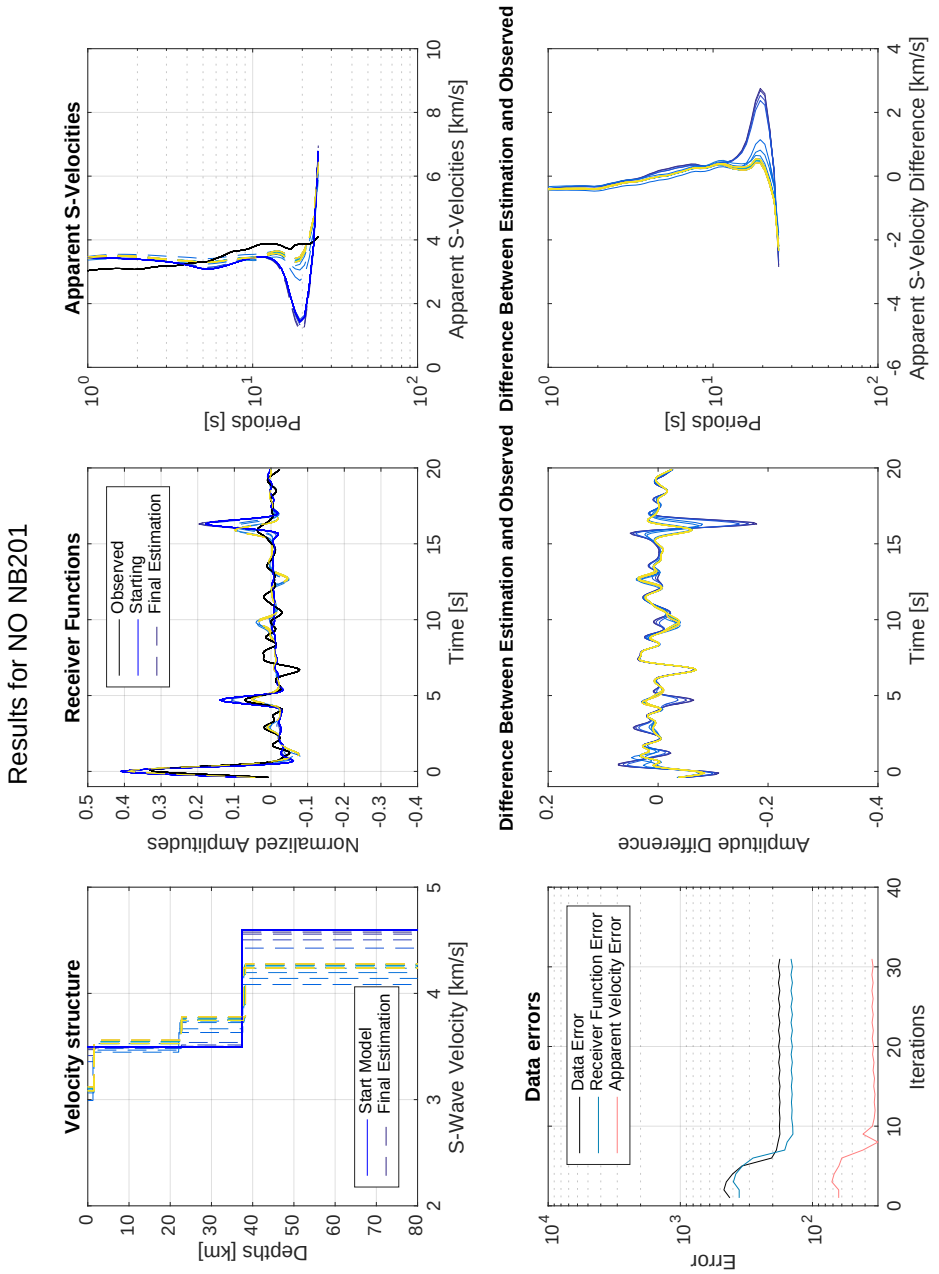


Figure 46: Station NO_NB201

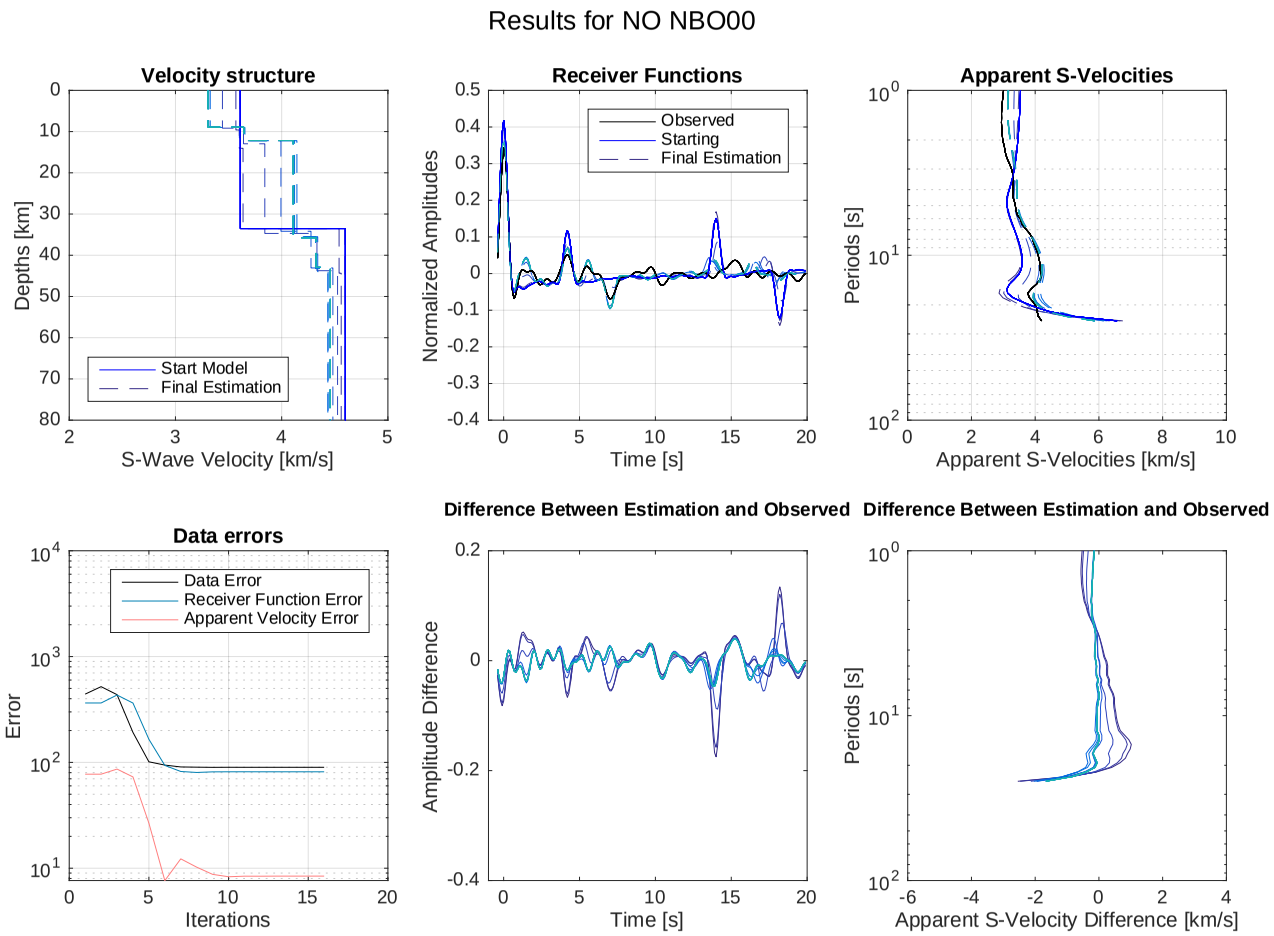


Figure 47: Station NO_NBO00

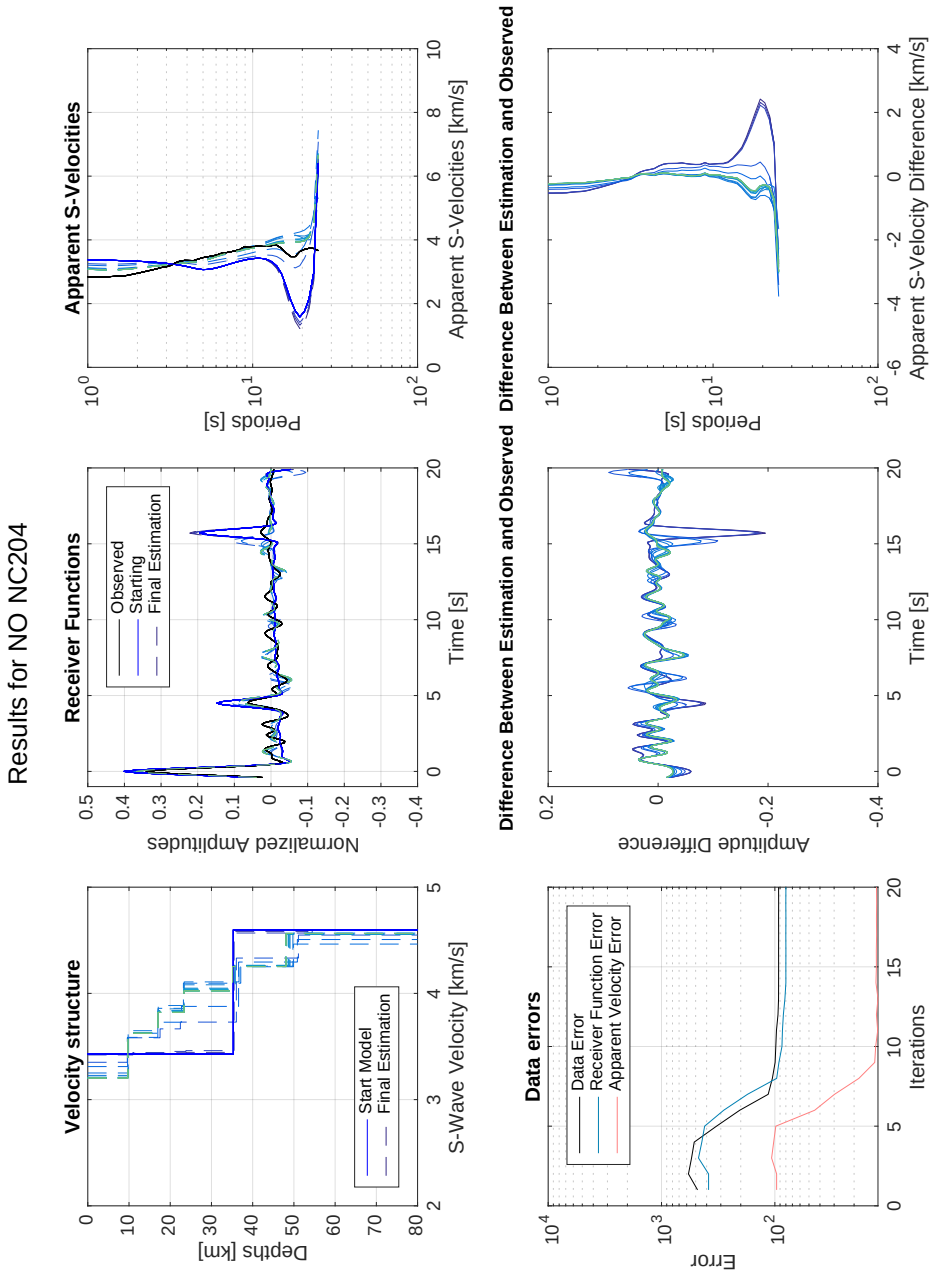


Figure 48: Station NO_NC204

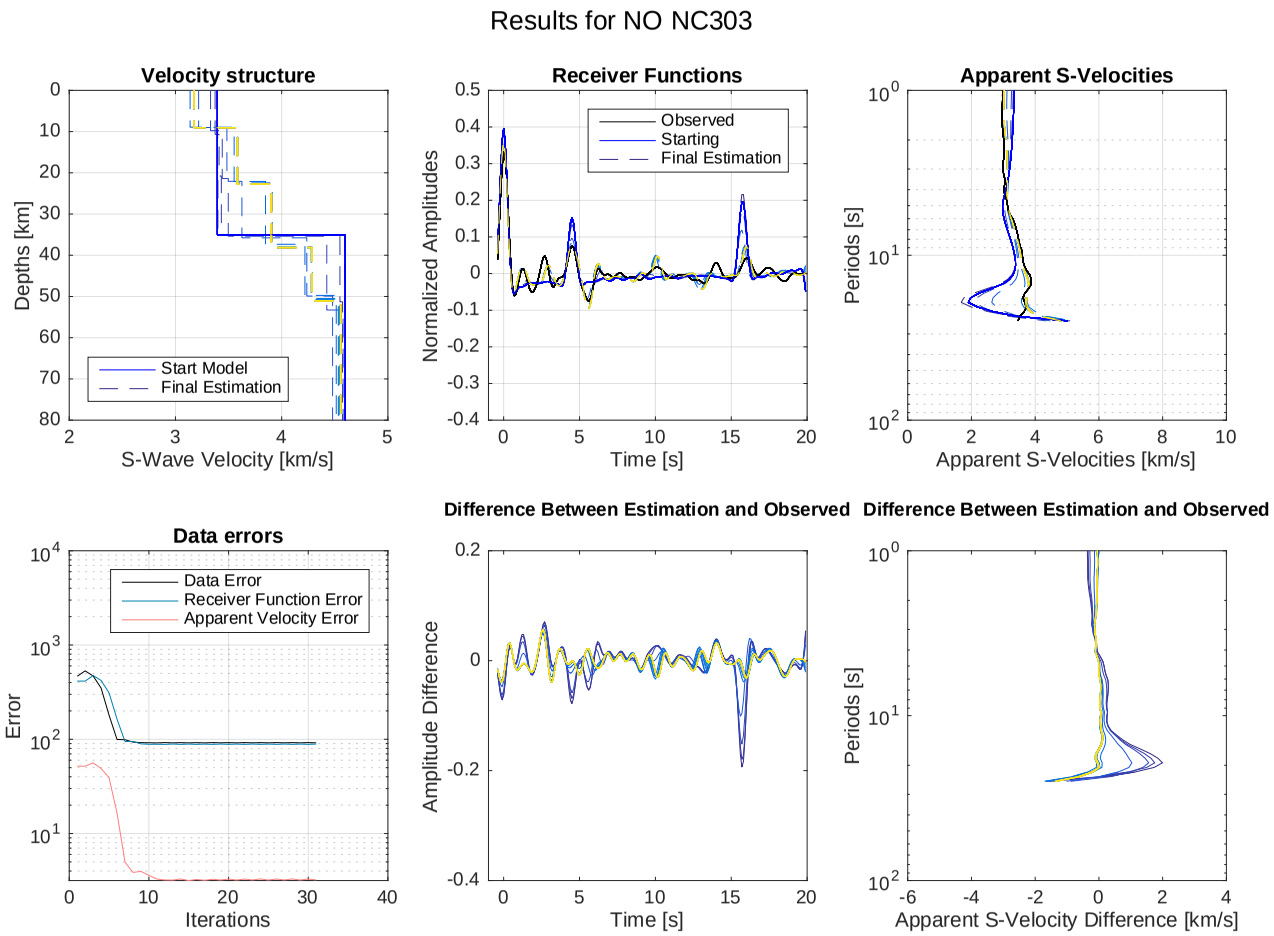


Figure 49: Station NO_NC303

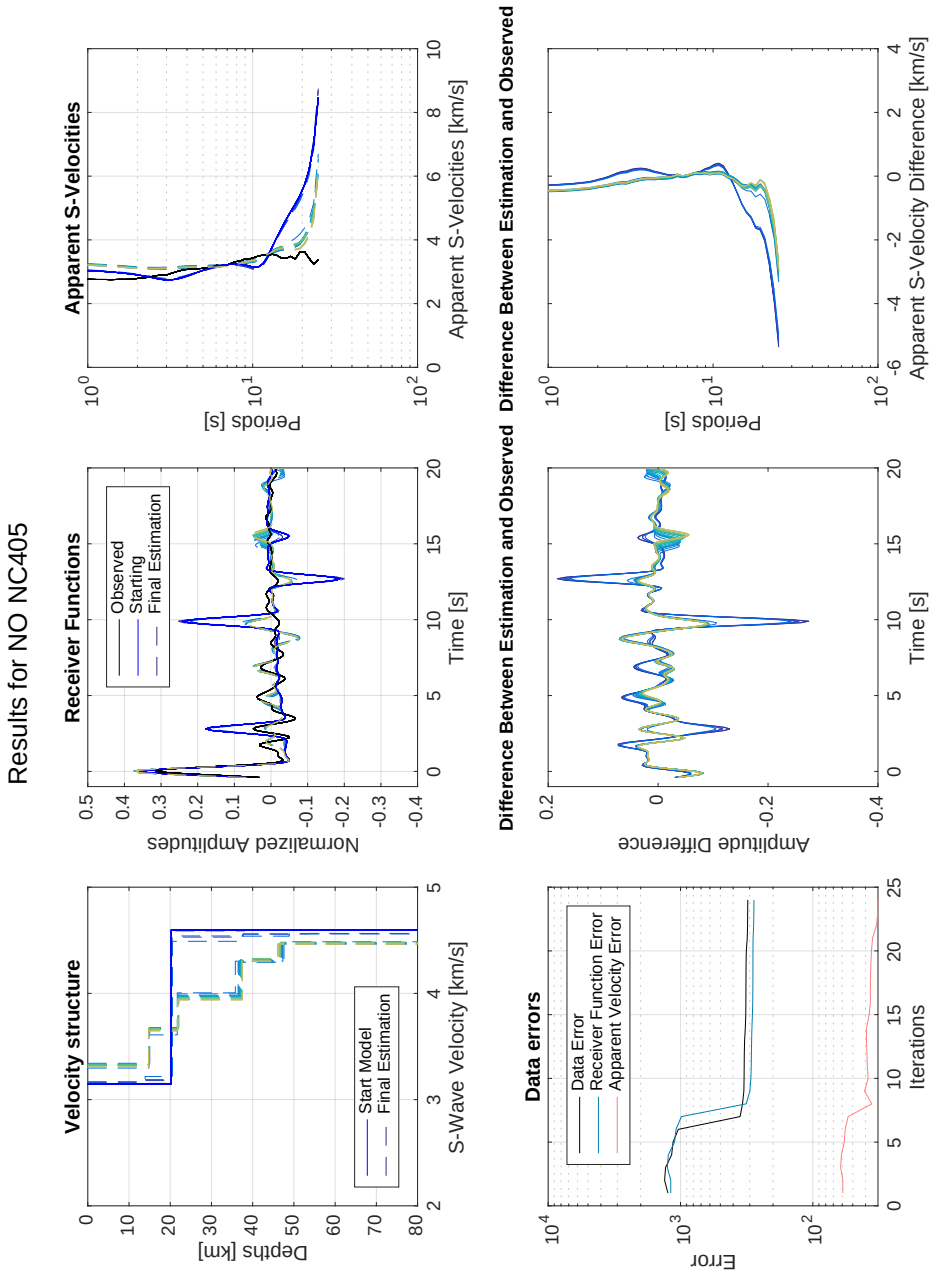


Figure 50: Station NO_NC405

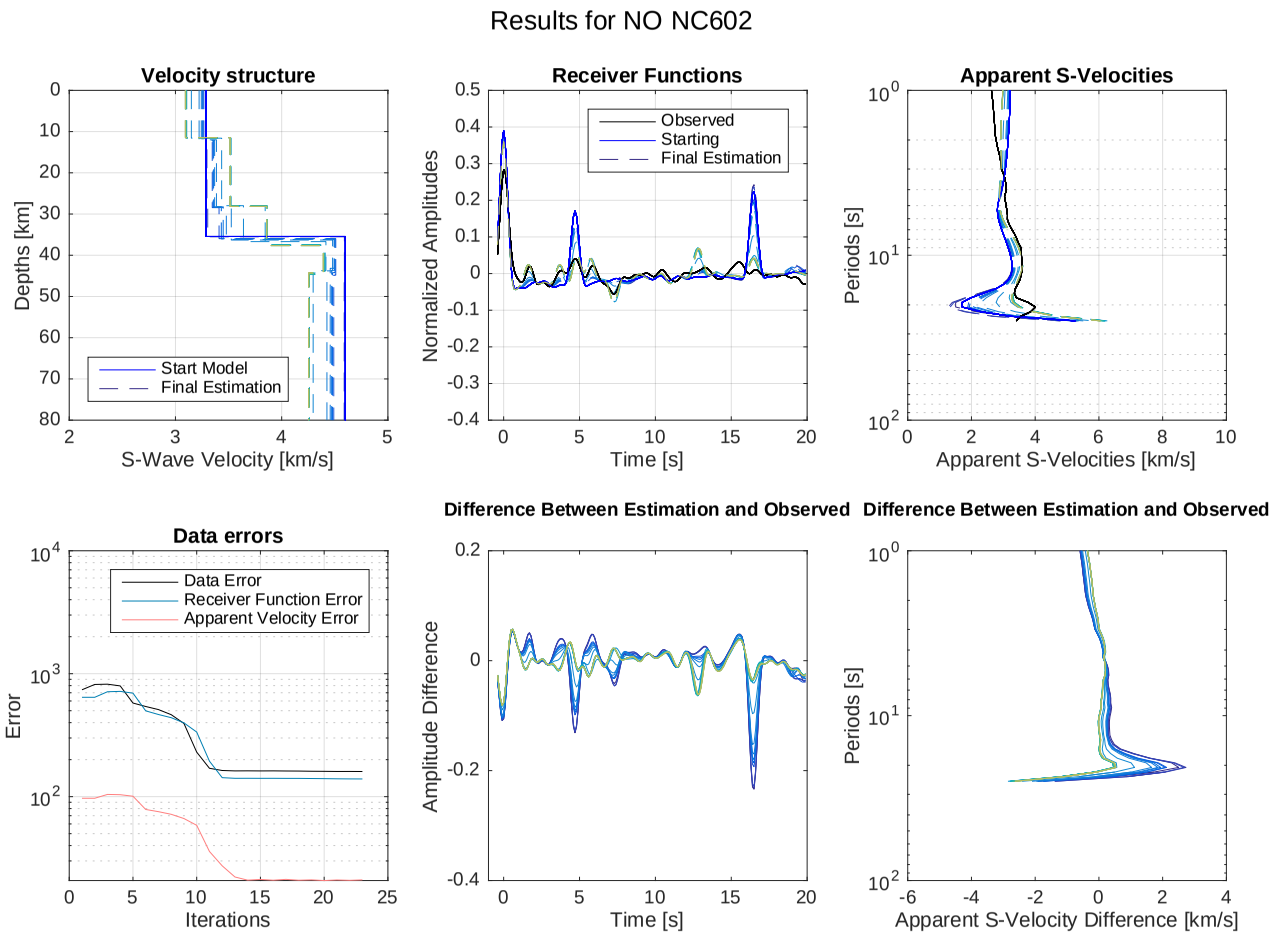


Figure 51: Station NO_NC602

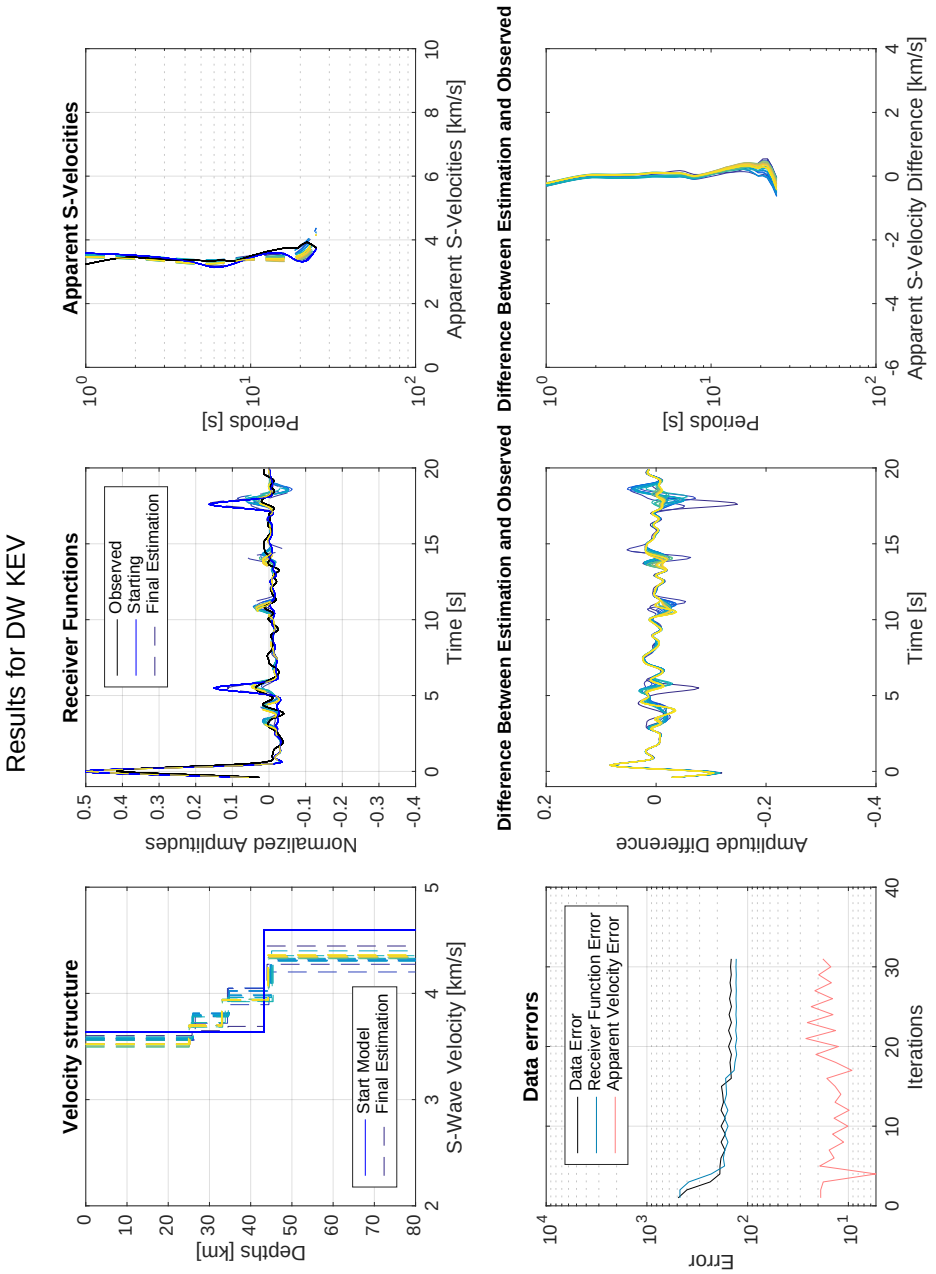


Figure 52: Station DW_KEV

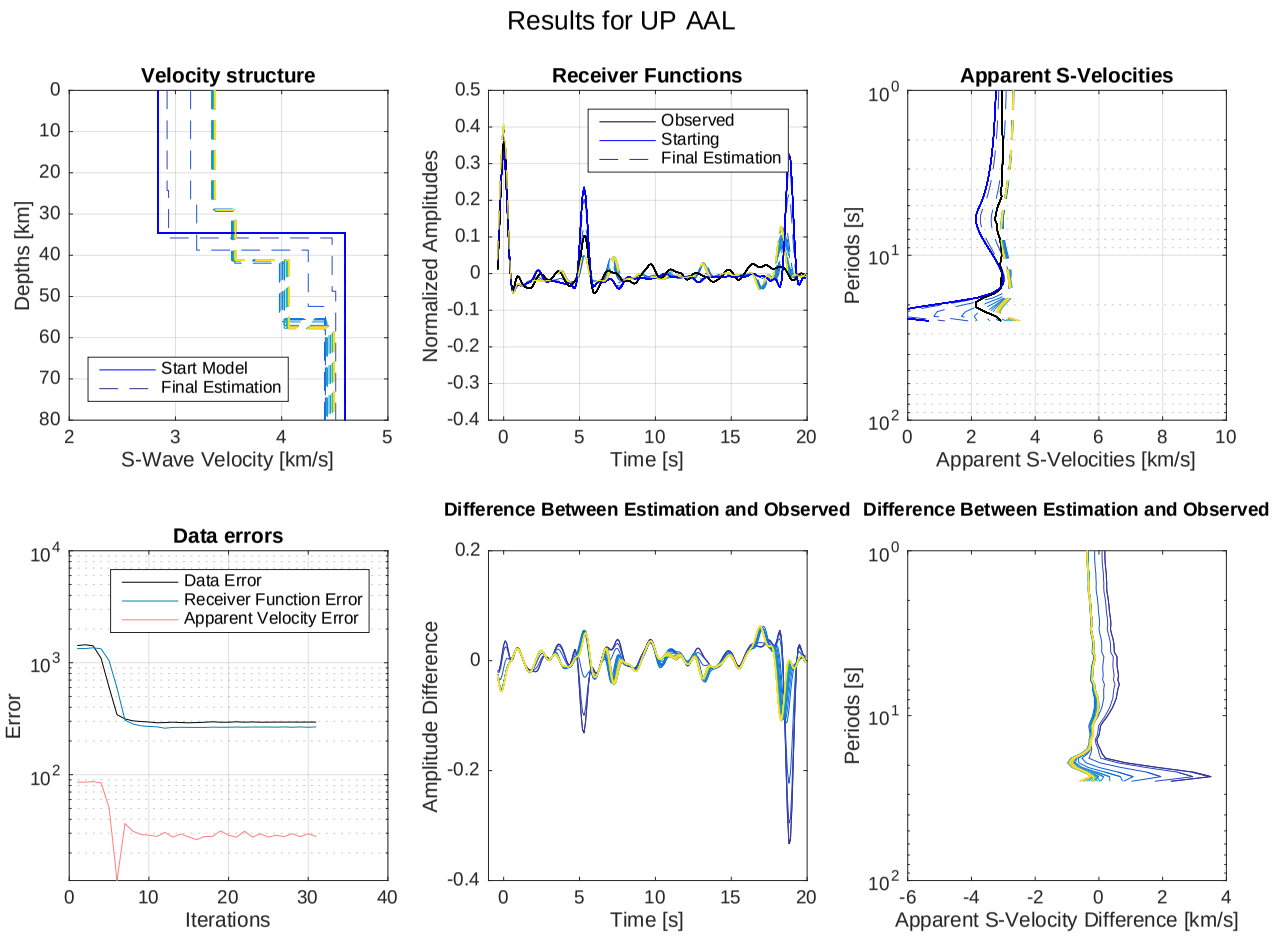


Figure 53: Station UP_AAL

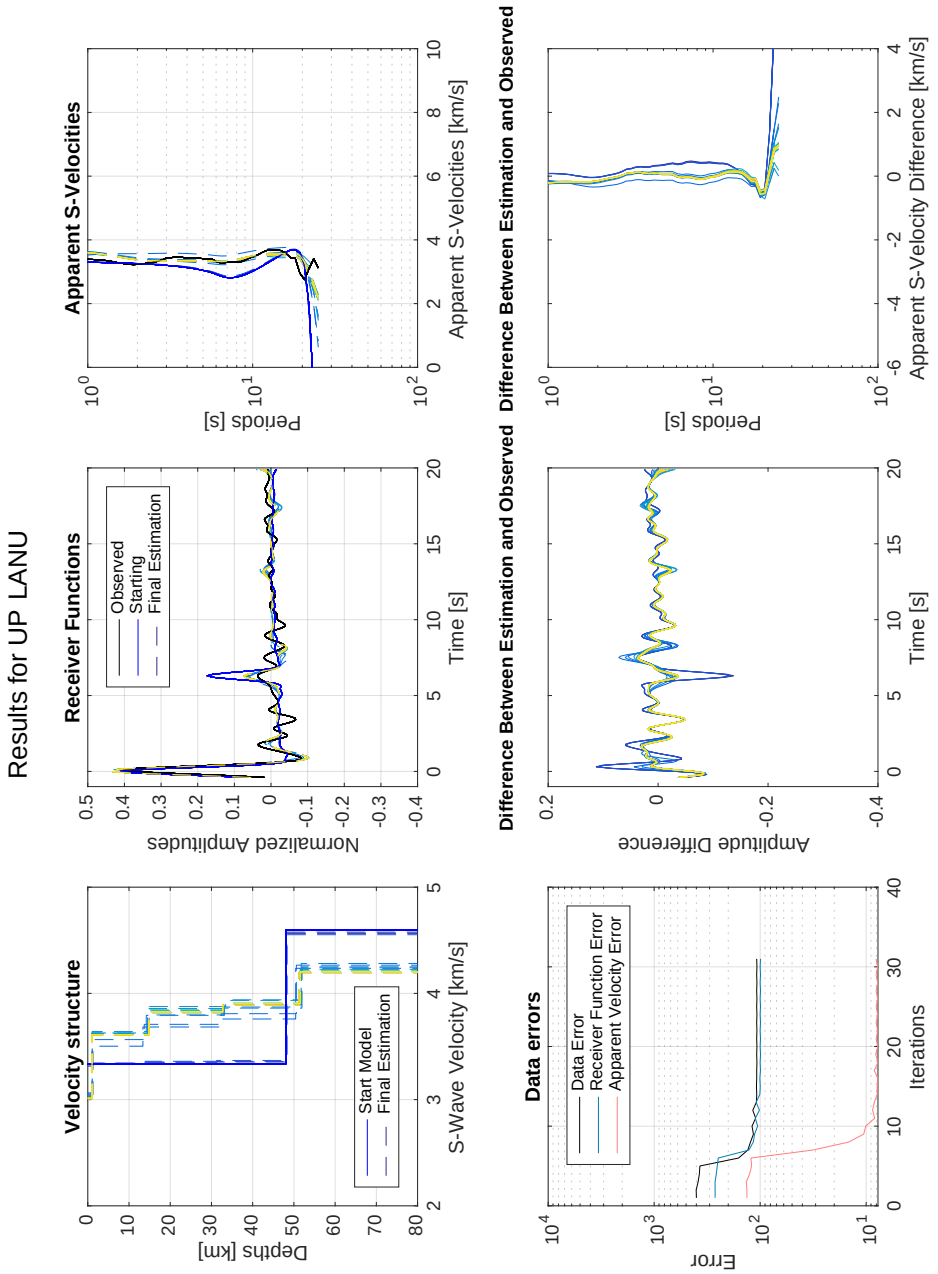


Figure 54: Station UP_LANU

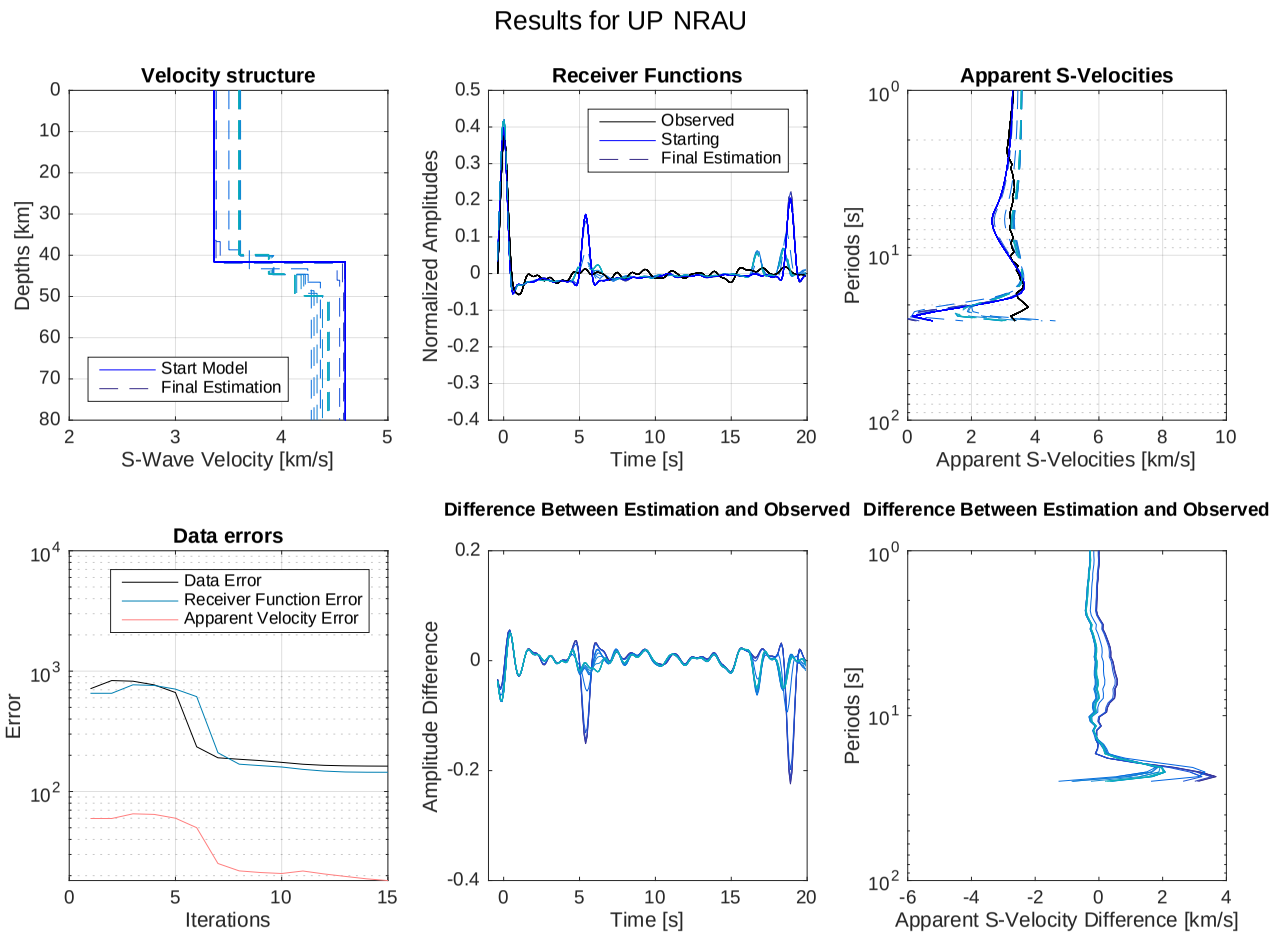


Figure 55: Station UP_NRAU

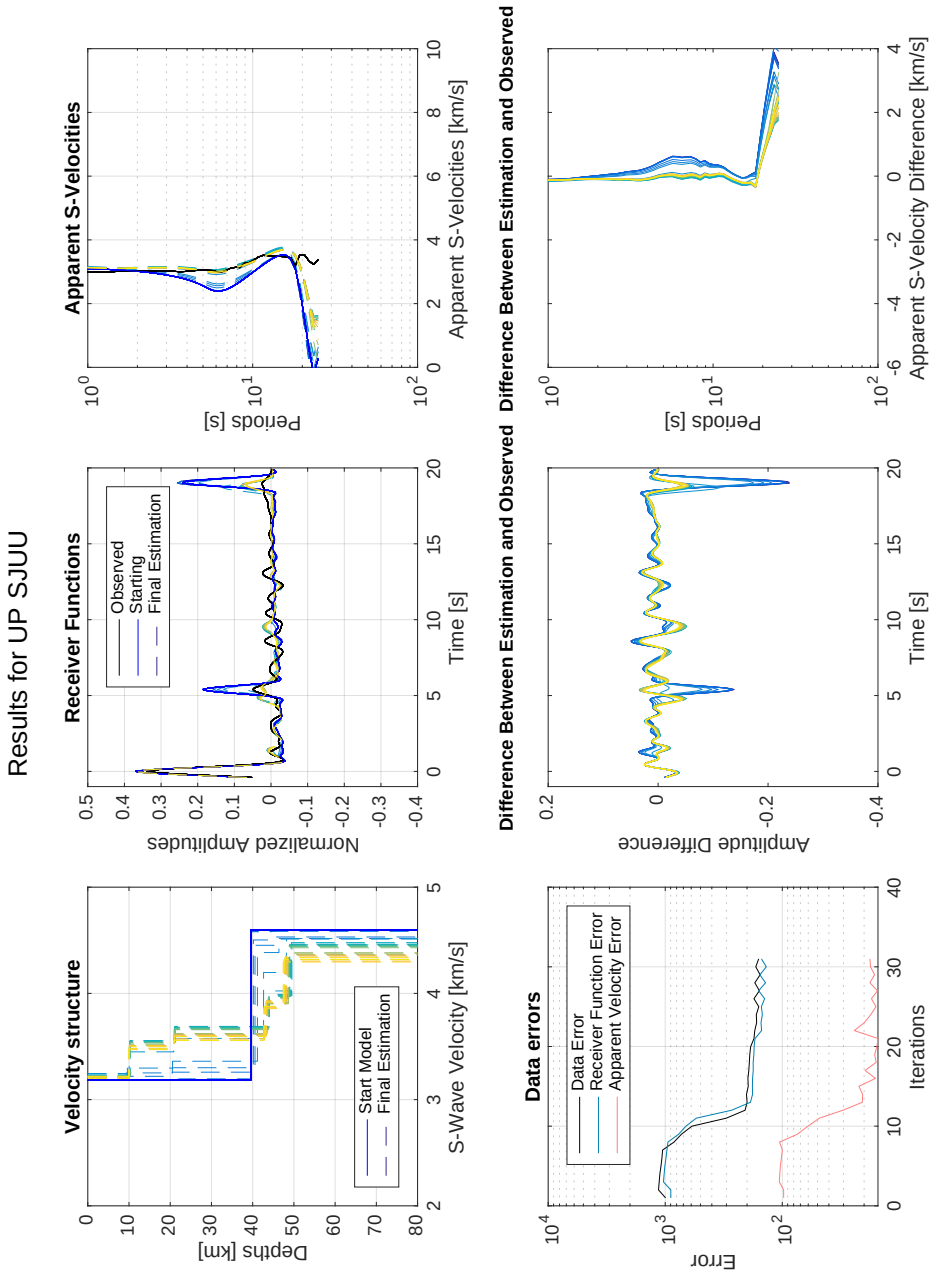


Figure 56: Station UP_SJUU

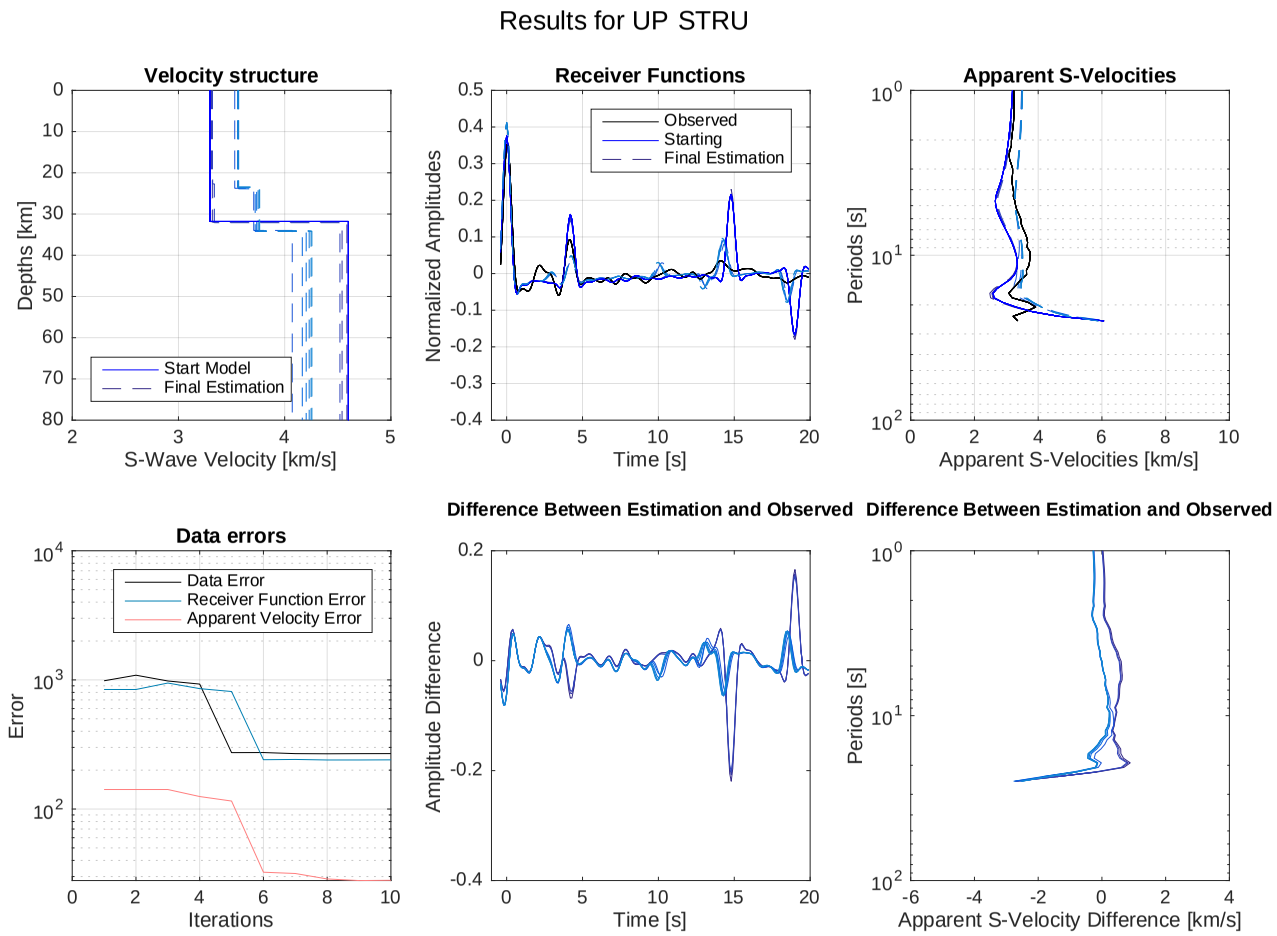


Figure 57: Station UP_STRU

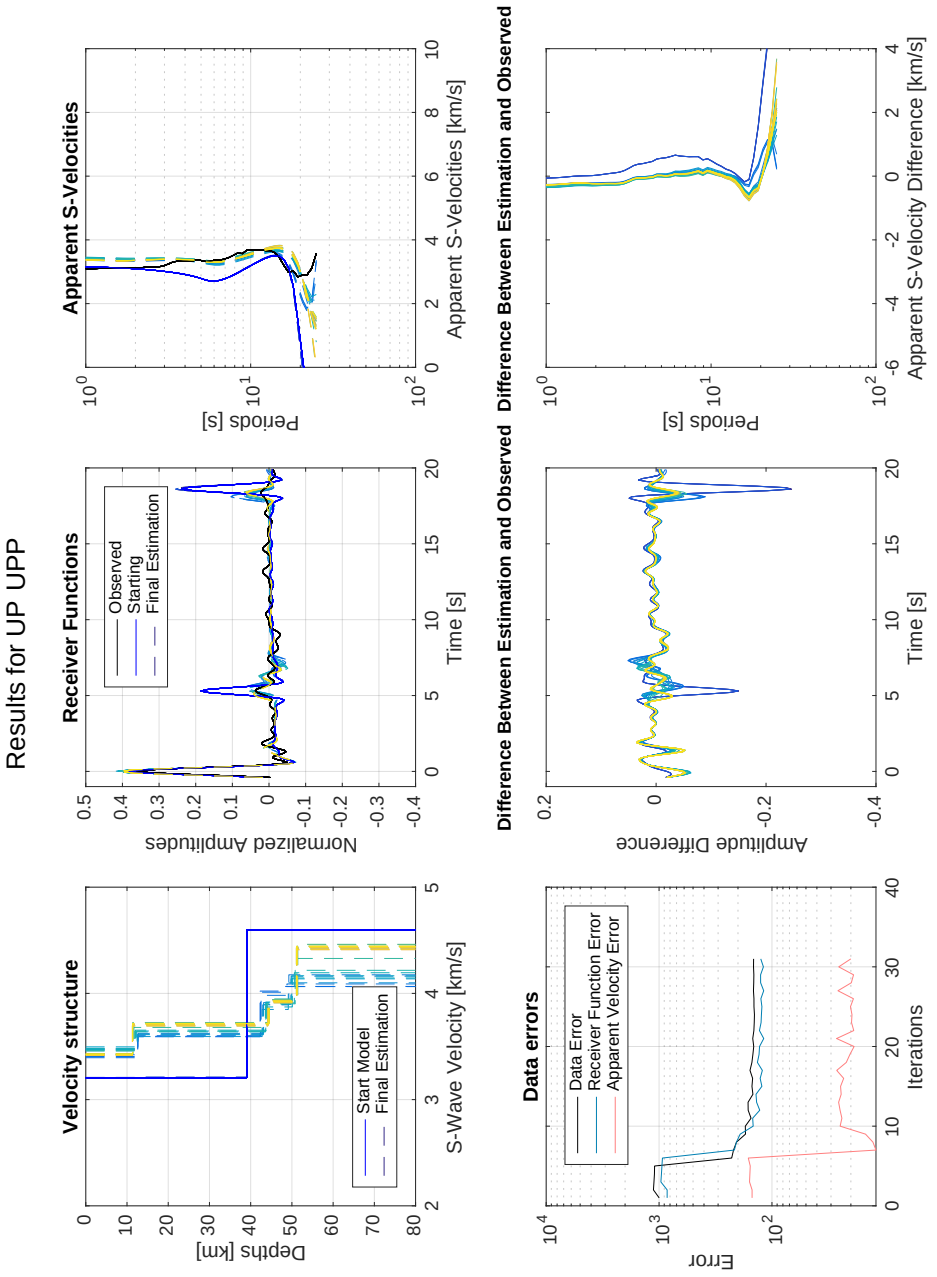


Figure 58: Station UP_UPP

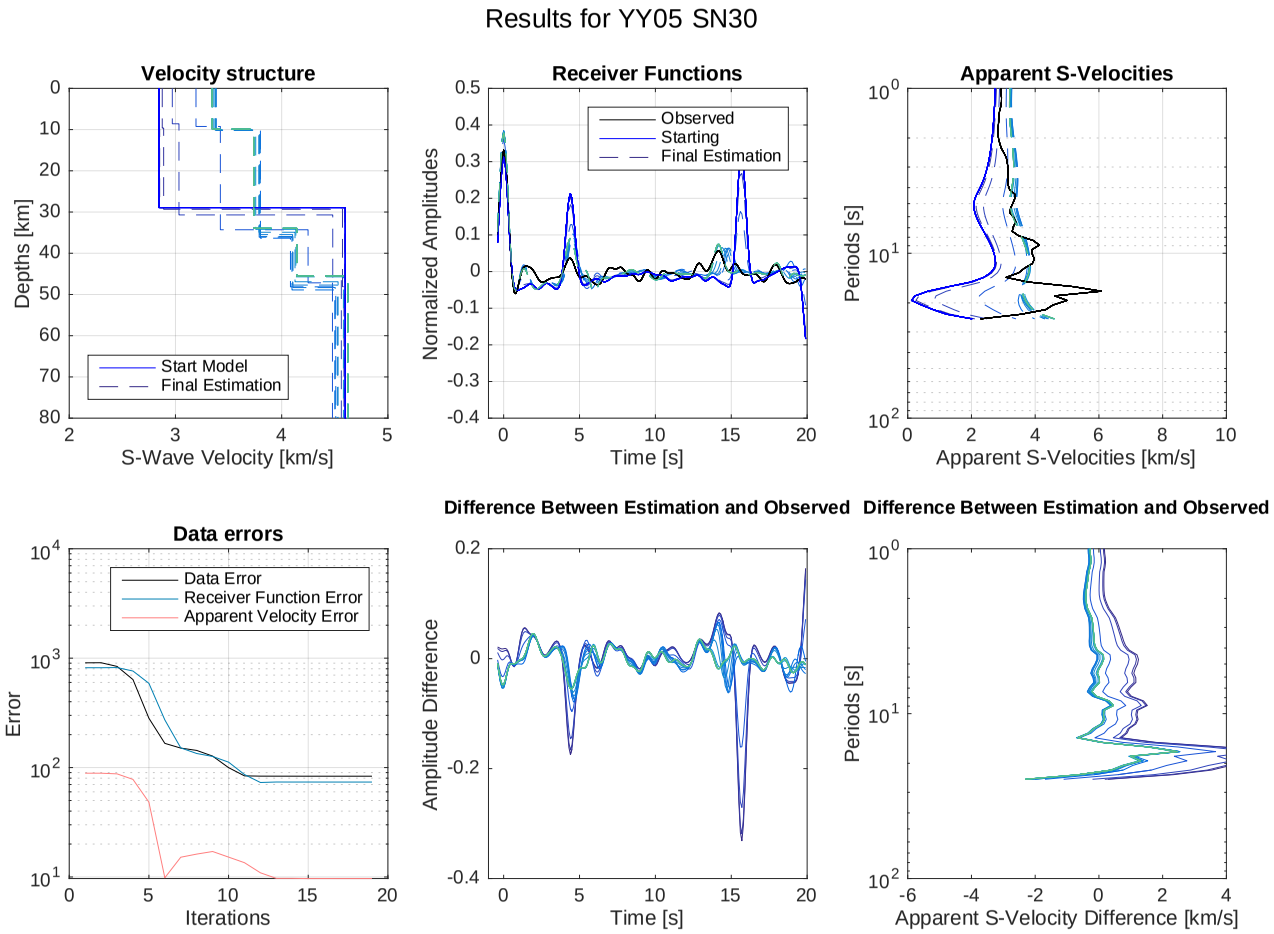


Figure 59: Station YY05_SN30

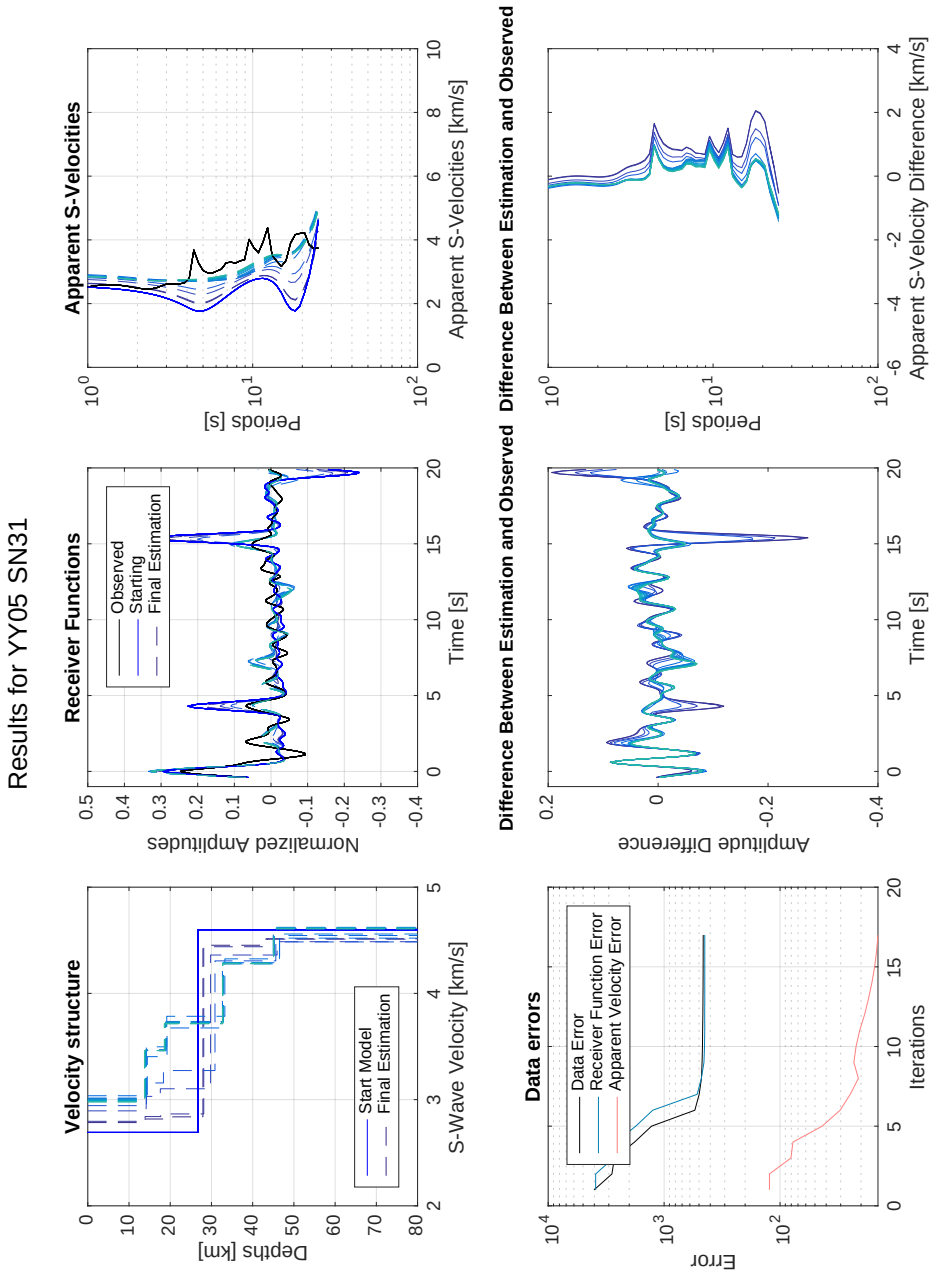


Figure 60: Station YY05_SN31

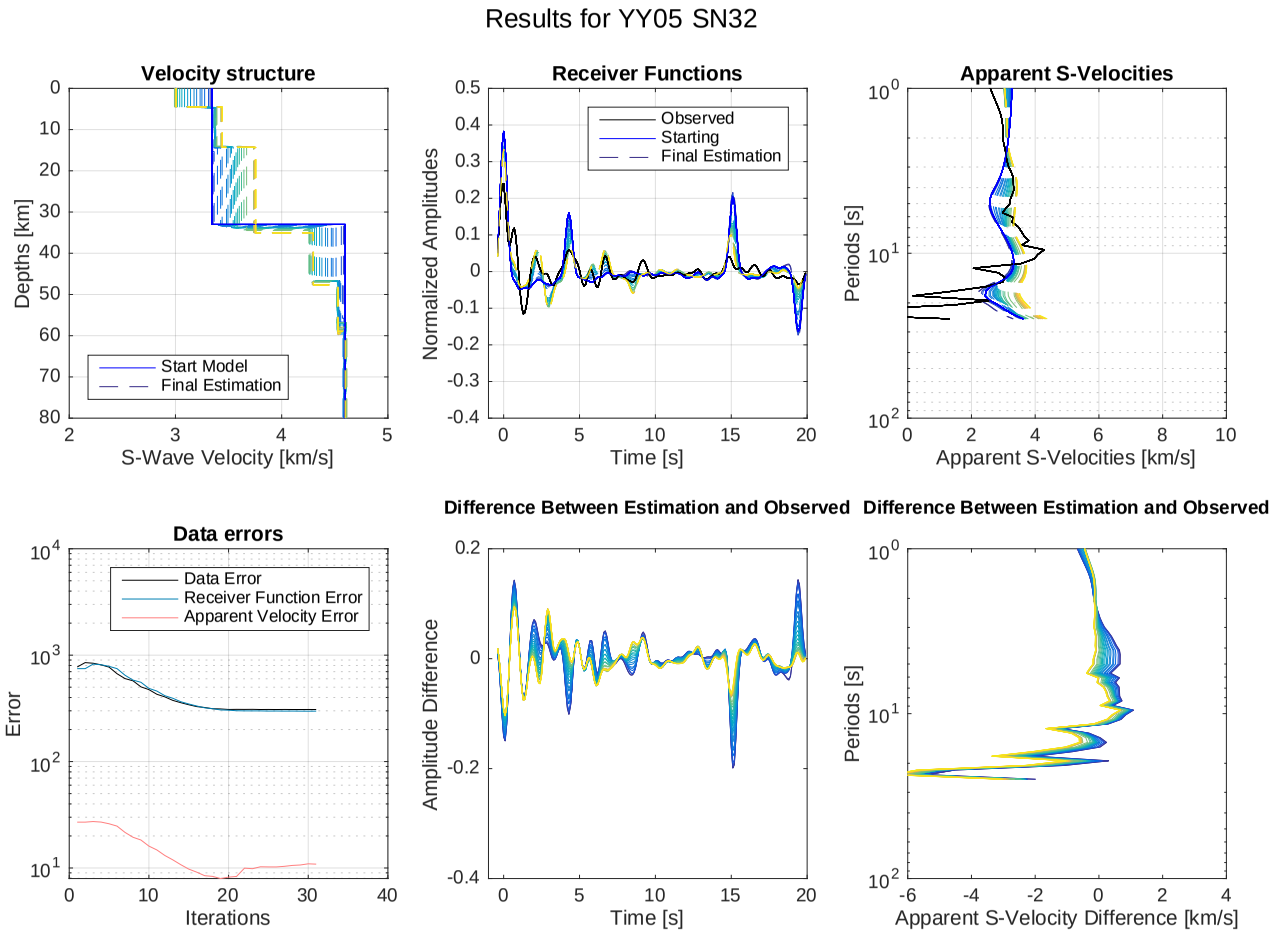


Figure 61: Station YY05_SN32

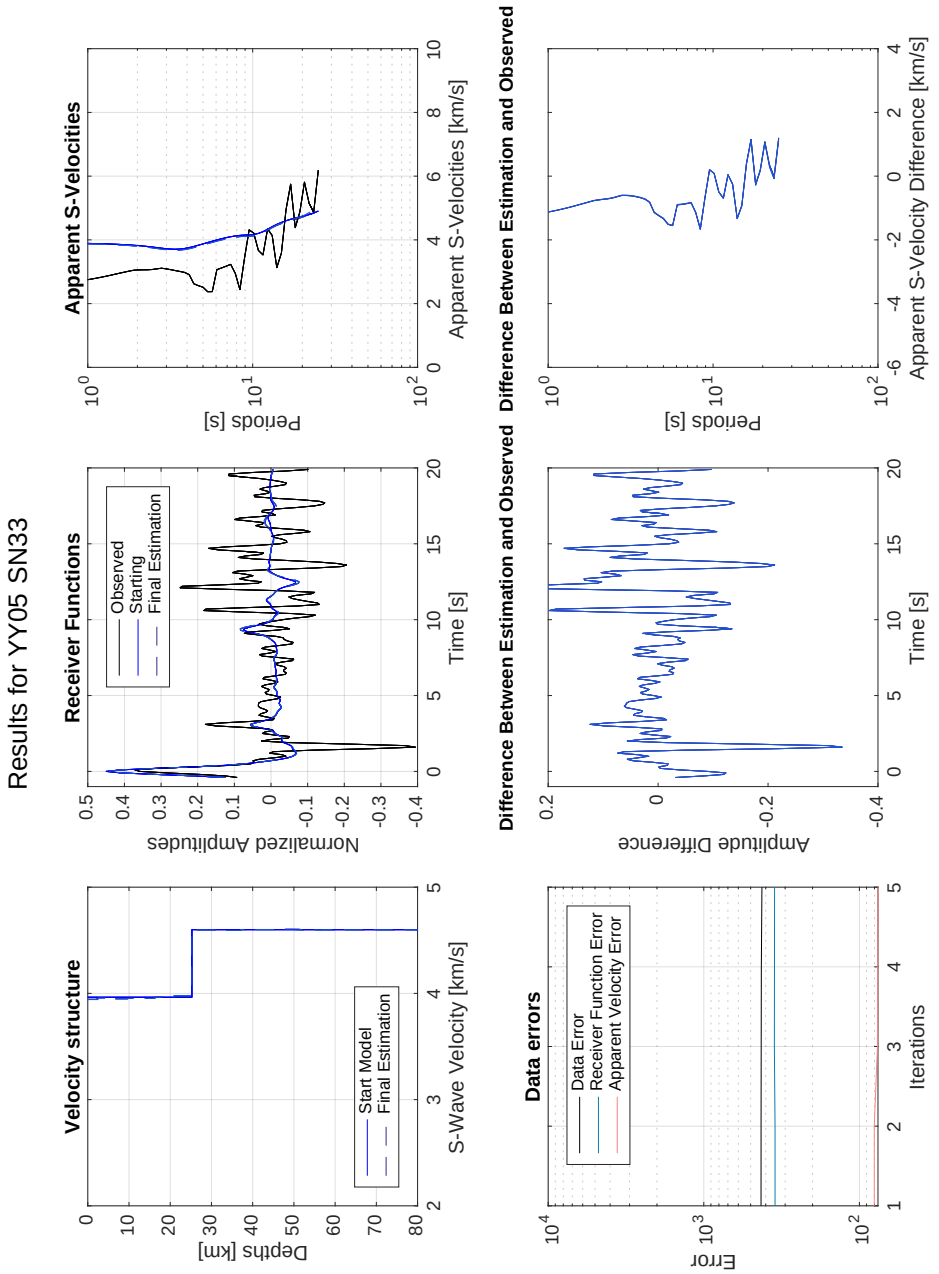


Figure 62: Station YY05_SN33

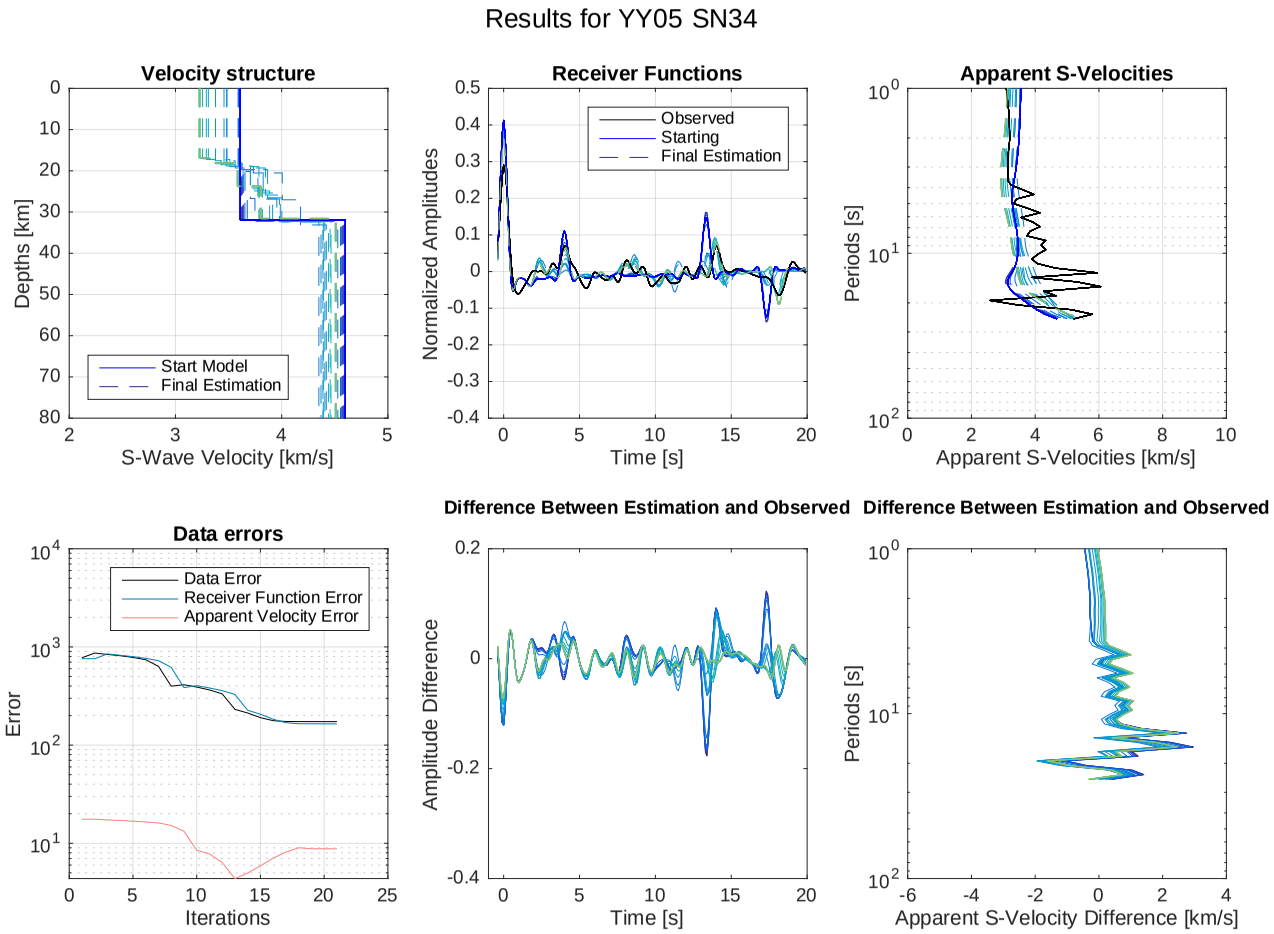


Figure 63: Station YY05_SN34

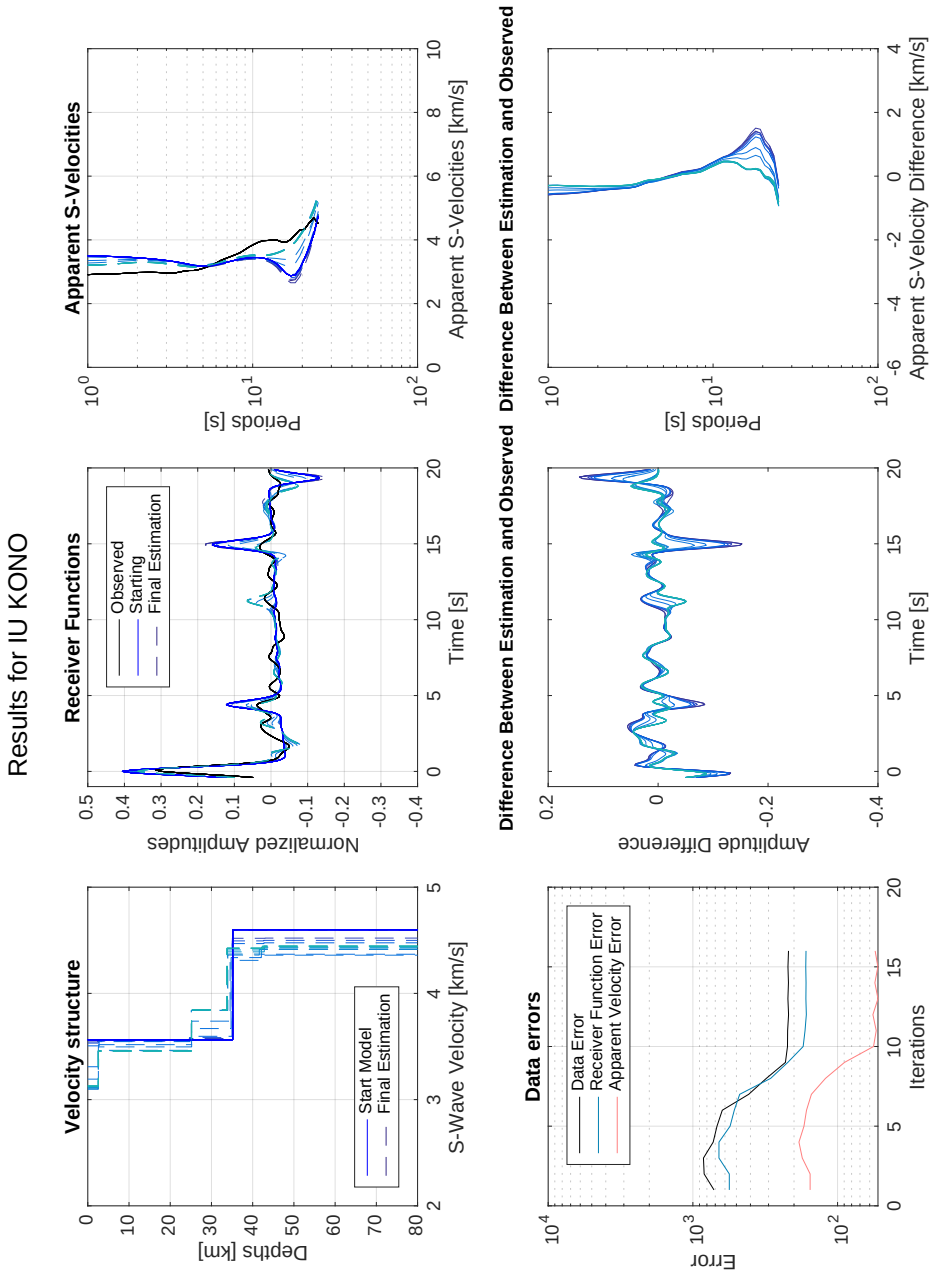


Figure 64: Station IU_KONO