NTNU

Norwegian University of
Science and Technology

A blockyness Constraint for seismic AVA Inversion

Ingrid Østgård Jensås

Master of Science in Physics and Mathematics
Submission date: June 2008
Supervisor: Jo Eidsvik, MATH
Co-supervisor: Ulrich Theune, StatoilHydro

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

Develop a method for blocky Cauchy inversion of seismic amplitude-versus-angle data.

Compare the blocky inversion method with non-blocky Gaussian inversion.

Test the methods on synthetic and real seismic data.

Assignment given: 10. January 2008
Supervisor: Jo Eidsvik, MATH

PREFACE

This report is the result of the course *TMA4905 - Statistics, Master Thesis*. It counts for 30 points, which means the study load for one regular semester. Industrial mathematics is a continuation of the study program *Applied Physics and Mathematics* at NTNU.

This work is a combination of the two disciplines statistics and geophysics. It has been performed in cooperation with StatoilHydro's R&D centre at Rotvoll. Most of the work is done by using the numerical computing environment and programming language MATLAB.

In order to provide a more complete picture of the theoretical background, some parts from my project thesis, *Methods for blocky seismic amplitude inversion*, 2007, are included in this report.

I want to thank StatoilHydro ASA for providing this project with seismic and well log data. I also want to thank my supervisors, Ulrich Theune at StatoilHydro and Jo Eidsvik at NTNU for good advice, whenever needed. They have both contributed with good ideas to help bring the project forward, and our discussions have been truly enlightening.

Trondheim, June 2008 Ingrid Østgård Jensås

ABSTRACT

The aim of seismic inversion is to determine the distribution of elastic parameters from recorded seismic reflection data. If a combination of elastic parameters is known, they indicate a certain fluid or lithology. Elastic parameters can therefore be very good hydrocarbon indicators. Although it is possible to interpret the reflection data from seismic acquisitions after processing, an improved analysis can be achieved by inverting for elastic properties. This can contribute to improved vertical resolution of the image.

This work applies different applications of the blocky seismic inversion technique, which is based on Bayesian inversion. Generally, a Gaussian prior for the three elastic parameters P-wave velocity, S-wave velocity and density is assumed in inverse problems. This assumption does not always provide sharp edges between layers, and the idea of the work reported here is to improve this by assuming a prior distribution for the contrasts in the elastic parameters with more probability of high contrasts. Since the Cauchy distribution has heavier tails than the normal distribution, the idea for the blocky inversion is to assume a Cauchy prior distribution for the contrasts in the elastic parameters.

Inversion is a non-unique process, hence, the more reasonable prior information we use, the better the result. When using statistical inversion based on Bayes' rule, the prior distribution is used to shape the solution, and the modified Cauchy norm can help provide a solution with better focused layer boundaries. The scale parameter in the Cauchy distribution is not very easy to estimate, and different methods are tested.

Spatial coupling of the model parameters $\boldsymbol{m}$ is introduced along a line to provide lateral consistency and robust results from inverse problems. The 2D inversion was done by assuming a Markov model where the inversion result at one location depends only on the neighbouring traces. This implies a sparse structure of the matrix to be inverted, and Cholesky factorization was used as a computational tool. This method allows tracewise nesting in contrast to setting up the whole operation matrix for all traces at a line, and therefore reduces the computational time significantly. The aim of this approach was to consider the use of lateral correlation while inverting data as a sophisticated way of stacking data to improve the signal to noise ratio. To get a picture of the uncertainties in

the inversion result, different methods, such as importance sampling was performed, even though the answers were unreasonable large. This remains a topic for further work. The data used in this work are a synthetic created case and real seismic data from the Kvitebjørn field in the North Sea.

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

Seismic data analysis is a geophysical tool to examine the underground. The results from collecting seismic signals are maps of the geological structures. These maps are interesting for oil companies because they are very useful in the search for hydrocarbons. Inversion of seismic data in order to extract elastic parameters as P-wave velocity, S-wave velocity and density, is provided in order to get an improved analysis of the seismic data. Amplitude versus offset (AVO) or Amplitude versus angle (AVA) analysis was first defined in 1984, and was a break-through method on this topic. Ostrander (1984), showed that the presence of gas in a sand capped by a shale would cause an amplitude variation with offset. After that, AVO has become a common technique for pre-stack qualitative seismic interpretation.

The aim of this work has been to explore the possibility of improving the well known Bayesian inversion technique. This technique uses prestack seismic reflection data to provide an image of the distribution of the P-wave velocity (v_p), the S-wave velocity (v_s) and the density (ρ) in the subsurface. The traditional assumption is a Gaussian prior distribution of the elastic parameters. This work will compare regular Gaussian inversion to blocky inversion. Blocky inversion assumes Cauchy distributed contrasts for the elastic parameters in order to get a more "blocky" inversion result. Lateral correlation along a 2D seismic line, parameter estimation and estimation of uncertainties will also be discussed.

Standard Bayesian inversion is, among others, performed by Buland and Omre (2003a), and the blocky inversion technique is tested by Theune (2007). The idea of applying the Cauchy norm for sharper contrasts is also used in other subject areas, such as edge preserving imaging, Charbonnier et al. (1997), Youzwishen and Sacchi (2006) and Portniaguine and Zhdanov (1999). Lateral correlation was discussed by Buland et al. (2003), but in a different setting compared to this work. The work presented here experiments with different methods related to the blocky inversion technique for inversion of elastic parameters.

Next chapter is an introduction to seismic inversion and the linearized model for seismic data. Chapter 3 describes Gaussian Bayesian inversion and estimation methods related to this inversion, and Chapter 4 is about blocky Bayesian inversion. Chapters 5 and 6 describe the results from a synthetic created case

and data from the Kvitebjørn field respectively.

2 SEISMIC AVO INVERSION

There are two types of seismic waves, body waves and surface waves. Geophysicists are mainly interested in the body waves, which propagate through the earth and are reflected at structural interfaces of interest. Longitudinal P-waves (Primary waves or Pressure waves) are the fastest of the body waves and arrive first at the recording system. These waves have particle movement along the direction of propagation. The other type of body waves are the transversal S-waves (Secondary waves or Shear waves). These waves are slower, and the particle movement is perpendicular to the direction of propagation. For land seismics and ocean bottom seismics, the receivers are geophones that record particle movement, whilst hydrophones are used in marine seismic acquisition to get information regarding pressure changes. The explosions, which cause the seismic input signals, are mostly generated by air guns or vibrators (land seismics). Recorded amplitudes represent the seismic data, which depend on the distance between the source and the receiver (offset) and the properties in the rocks. Before seismic inversion, the offset dependency is transformed to angle of incidence. The seismic inversion process aims to determine the reason for the recorded seismic signal. This includes characteristics of rocks and fluids in the underground.

To extract information about the subsurface parameters from seismic reflection data is called quantitative seismic interpretation. There are methods that analyse the strength of the reflected amplitudes and detect the presence of different fluids in the rock's pore space. Prior to AVO analysis and linear AVO inversion, the recorded seismic data need to be processed.

AVO explains the direct impact of changes in the rock properties on the amplitudes, and is a useful part of the total seismic interpretation. However, it should not be considered as a stand-alone technique. This is because the AVO analysis is very sensitive to the quality of the pre-processing, which includes the estimation of the signal-to-noise ratio and the wavelet phase, correction for absorption, anisotropy and geometrical spreading, top and bottom muting, normal moveout (NMO) correction, dip moveout (DMO) correction, time migration and multiple removal. Anisotropy means that the properties of the rocks vary with the measurement direction (Ayzenberg and Ursin (2007)).

The aim of seismic inversion is to obtain an image of the velocities and densities in the subsurface based on seismic reflection data. An isotropic, elastic medium is completely described by the three elastic material parameters, P- and S-wave velocity and density. These parameters influence the two-way vertical seismic traveltime (twt) and depend on the lateral location (Buland and El Quair (2006)).

2.1 Linearized model of the seismic data

A basic assumption in seismic processing is that the recorded seismic signal can be expressed as a convolution between the wavelet representing the input signal and the impulse responses of the earth (reflectivities). From an exploration point of view, the interesting information about the subsurface is contained in the impulse responses. Therefore, the aim is to extract the impulse responses from the recorded seismic signal. This is called deconvolution (Landrø (2007)).

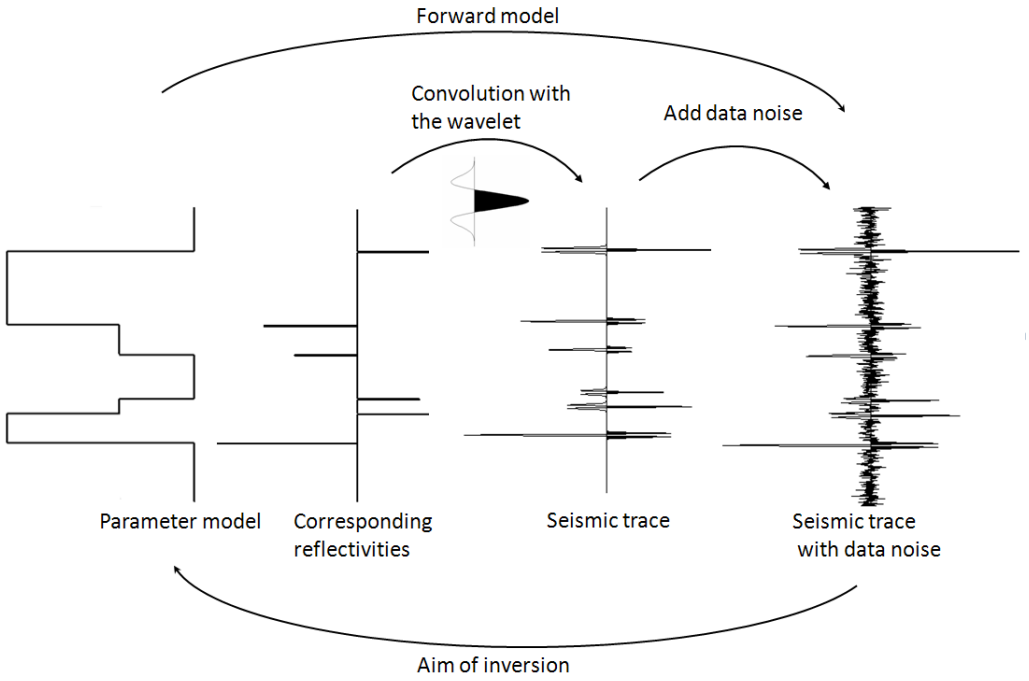

Figure 2.1: Synthetic example to show the steps in modelling and inversion. From left: Changes in an elastic parameter, corresponding reflectivities, seismic trace after convolution with the wavelet, seismic trace with data noise.

Figure 2.1 shows (from left) how elastic parameters with corresponding reflectivities convolved with a wavelet constitute the seismic trace. Seismic inversion

aims to go backwards, to extract the reflectivities and the distribution of the elastic parameters from the recorded seismic trace. The reflection coefficient for zero offset reflected P-wave data is given as

$$R_{PP}(0) = \frac{z_{i+1} - z_i}{z_{i+1} + z_i}, \tag{2.1}$$

where z_i, is the acoustic impedance $v_p\rho$ for layer i at the reflector. Figure 2.2 shows a sketch of the first two layers ($i = 1, 2$). The non-linear Zoeppritz equations describe the reflection coefficient from such two-layer systems with reflection angle θ. They are valid under assumptions of elastic homogenous materials and incoming plane waves.

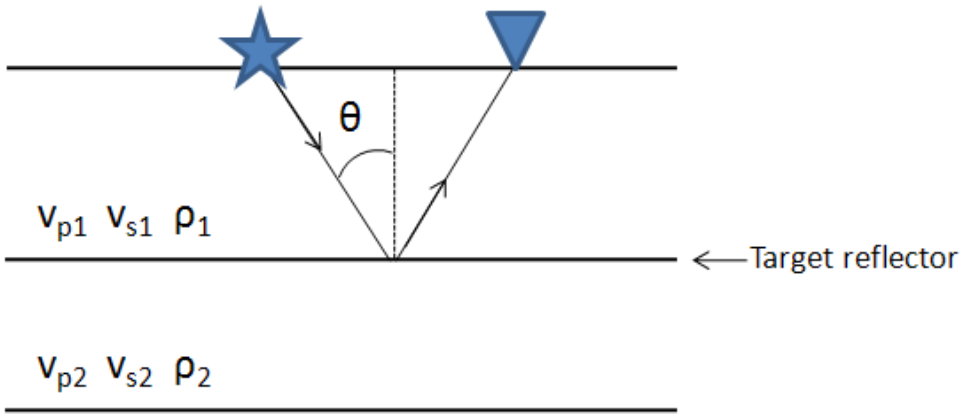

Figure 2.2: Sketch of the two-layer system described by the Zoeppritz equations.

Different linear approximations to the Zoeppritz equations are commonly used, such as the Aki & Richards approximations (Aki and Richards (2002)). By using this approximation, the reflected P-wave impulse ($P_{down}P_{up}$) for a target reflector as the one in Figure 2.2 is given by

$$R_{PP}(\theta) = a(\theta)\frac{\Delta v_p}{\overline{v_p}} + b(\theta)\frac{\Delta v_s}{\overline{v_s}} + c(\theta)\frac{\Delta \rho}{\overline{\rho}}, \tag{2.2}$$

where

$$
\begin{aligned}
a(\theta) &= \frac{1}{2}(1+\tan^2\theta), \\
b(\theta) &= -4\frac{\overline{\boldsymbol{v_s}^2}}{\overline{\boldsymbol{v_p}^2}}\sin^2\theta, \\
c(\theta) &= \frac{1}{2}(1-4\frac{\overline{\boldsymbol{v_s}^2}}{\overline{\boldsymbol{v_p}^2}}\sin^2\theta), \\
\Delta\xi &= \xi_2 - \xi_1 \text{ and} \\
\overline{\xi} &= \frac{\xi_1+\xi_2}{2}
\end{aligned}
\tag{2.3}
$$

for $\xi \in \{\boldsymbol{v_p}, \boldsymbol{v_s}, \boldsymbol{\rho}\}$ (Buland and Omre (2003a)). This approximation is valid for small reflection angles (up to approximately $40°$) and weak contrasts between layers. The coefficients $b(\theta)$ and $c(\theta)$ are dependent on the ratio between $\overline{\boldsymbol{v_s}}$ and $\overline{\boldsymbol{v_p}}$. This ratio is commonly assumed constant for inversion. In some parts of this work, however, $\overline{\boldsymbol{v_s}}$ and $\overline{\boldsymbol{v_p}}$ are approximated by a background trend extracted from borehole data. This is a slowly varying curve made by applying a lowpass filter to the borehole data. The coefficients in Equation 2.3 then become time dependent as $a(t,\theta)$, $b(t,\theta)$ and $c(t,\theta)$ where t is the traveltime along the trace. Since the aim is to invert for velocities and densities directly, and not reflectivities, the contrasts $\Delta\xi, \xi \in \{v_p, v_s, \rho\}$ can be approximated by the derivative. This gives

$$
\frac{\Delta\xi}{\xi} \approx \frac{\partial}{\partial t}\ln\xi,
$$

and Equation 2.2 becomes

$$
R_{PP}(t,\theta) = a(t,\theta)\frac{\partial}{\partial t}\ln v_p(t) + b(t,\theta)\frac{\partial}{\partial t}\ln v_s(t) + c(t,\theta)\frac{\partial}{\partial t}\ln\rho(t). \tag{2.4}
$$

In order to get an expression for the recorded seismic data, the reflection coefficients in Equation 2.4 is convolved with the wavelet. The wavelets have to be estimated from the seismic data, and this can cause some uncertainties. Buland and Omre (2003b) and Gunning and Glinsky (2006) are works regarding wavelet estimation. The reflectivities in Equation 2.4 and the convolution operation have to be discretized in order to express this as a linear problem. The linearized version of the convolution operation is a convolution matrix $\boldsymbol{W}$, with one wavelet for each chosen angle, and the reflection coefficients $a(t,\theta)$, $b(t,\theta)$ and $c(t,\theta)$ in Equation 2.4, are stored in the reflection matrix $\boldsymbol{A}$. Then the linearized model can be written as

$$\boldsymbol{d} = \boldsymbol{WADm} + \boldsymbol{n}, \tag{2.5}$$

where $\boldsymbol{d} = (d_1, ..., d_{S \cdot T})^T$ is the pre-processed seismic data, $\boldsymbol{D}$ is a differential matrix and length $N = 3 \cdot T$ vector $\boldsymbol{m}$ is arranged as

$$\boldsymbol{m} = \begin{bmatrix} \ln(v_{p_1}) \\ \ln(v_{p_2}) \\ \vdots \\ \ln(v_{p_T}) \\ \ln(v_{s_1}) \\ \ln(v_{s_2}) \\ \vdots \\ \ln(v_{s_T}) \\ \ln(\rho_1) \\ \ln(\rho_2) \\ \vdots \\ \ln(\rho_T) \end{bmatrix}, \tag{2.6}$$

and contains the logarithm of the three elastic parameters $\boldsymbol{v_p}$, $\boldsymbol{v_s}$ and $\boldsymbol{\rho}$ for each time sample in the seismic trace, $t = 1, 2, ..., T$. S is the number of angles. The additive noise term $\boldsymbol{n}$ is added to describe model and measurement noise. The noise terms, $n_1, ..., n_{S \cdot T}$ are considered independent and normally distributed with zero mean and variances $\sigma_1^2, ..., \sigma_S^2$ for the S different angles. The noise terms are assumed equal for all traveltimes, $t = 1, ..., T$. This can be written as

$$n_i \sim N(0, \sigma_i^2), \quad i = 1, ..., S.$$

The derivatives of the properties with respect to time are approximated by a first order differential operator $\boldsymbol{D}$ as

$$\begin{bmatrix} \frac{\partial m_1}{\partial t} \\ \frac{\partial m_2}{\partial t} \\ . \\ . \\ . \\ . \\ . \\ . \\ \frac{\partial m_N}{\partial t} \end{bmatrix} \approx \boldsymbol{Dm} = \begin{bmatrix} -1 & 1 & 0 & . & . & . & . & . & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 & . & . & . & . & 0 \\ 0 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & . & . & . & 0 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ . & . & . & . & . & . & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ . \\ . \\ . \\ . \\ . \\ . \\ m_N \end{bmatrix}. \tag{2.7}$$

This represents forward difference as an approximation for the first component, backward for the last, and central differences for the rest of the components.

3 GAUSSIAN BAYESIAN INVERSION

Inverse problems can be defined by finding the cause of an observed effect. In seismic data analysis, the observed effect is the recorded seismic amplitudes and the cause is the elastic properties in the subsurface. The linear problem in Equation 2.5 shows the relation between the observed effect ($\boldsymbol{d}$) and the cause ($\boldsymbol{m}$) that are sought. Knowledge and uncertainties regarding the parameter $\boldsymbol{m}$ are summarized in the Bayesian approach as probability distributions. The prior distribution $P(\boldsymbol{m})$ indicates the knowledge prior to observations, the likelihood distribution $P(\boldsymbol{d}|\boldsymbol{m})$ expresses how the data depend on the parameters $\boldsymbol{m}$, and the posterior distribution $P(\boldsymbol{m}|\boldsymbol{d})$ is the answer to the inverse problem from the Bayesian point of view. The posterior describes the distribution of the parameters given the data, and the aim is to maximize this posterior to obtain the most probable solution (Kolbjørnsen (2002b)). This solution is called the maximum posterior (MAP) solution. A Bayesian setting is a natural choice for many geophysical inverse problems since it is possible to combine available prior knowledge with the information in measured data. The possibility of calculating uncertainties in the results is also an important advantage of Gaussian Bayesian inversion. Bayes' rule can be written as

$$P(\boldsymbol{m}|\boldsymbol{d}) = \frac{P(\boldsymbol{m},\boldsymbol{d})}{P(\boldsymbol{d})} = \frac{P(\boldsymbol{d}|\boldsymbol{m})P(\boldsymbol{m})}{P(\boldsymbol{d})}, \tag{3.1}$$

for every random variables $\boldsymbol{d}$ and $\boldsymbol{m}$.

The following theory is first presented for one trace, and then extended to 2D in Section 3.1. As mentioned before, the common assumption in Bayesian linear AVA inversion is that the prior distribution of the logarithm of the elastic parameters $\boldsymbol{m}$ is Gaussian, given by

$$P(\boldsymbol{m}) = \frac{1}{(2\pi)^{N/2}(det(\boldsymbol{C_m}))^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{m}-\boldsymbol{m_p})^T \boldsymbol{C_m}^{-1}(\boldsymbol{m}-\boldsymbol{m_p})\right\}, \tag{3.2}$$

where $\boldsymbol{m_p}$ is the prior expectation of $\boldsymbol{m}$. In this work, the low frequency background model derived from lowpass filtering of well log data, is used as $\boldsymbol{m_p}$

to guide the inversion. The symmetric 3×3 covariance matrix Σ_0 for the three elastic parameters $\boldsymbol{v_p}$, $\boldsymbol{v_s}$ and $\boldsymbol{\rho}$,which is estimated from well logs, is defined as

$$\boldsymbol{\Sigma_0} = \begin{bmatrix} \tau^2_{v_p} & \eta_{v_p,v_s} & \eta_{v_p,\rho} \\ \eta_{v_s,v_p} & \tau^2_{v_s} & \eta_{v_s,\rho} \\ \eta_{\rho,v_p} & \eta_{\rho,v_s} & \tau^2_{\rho} \end{bmatrix}, \tag{3.3}$$

where τ^2_ξ, $\xi \in \{v_p, v_s, \rho\}$ are the variances, and $\eta_{\alpha,\beta}$, $\alpha, \beta \in \{v_p, v_s, \rho\}$ are the covariances between α and β. The properties in the covariance matrix $\boldsymbol{\Sigma_0}$ are assumed equal for all time samples $t = 1, 2, ..., T$. Each time sample is further assumed independent, and this gives a sparse symmetric $N \times N$ covariance matrix $\boldsymbol{C_m}$ given as

$$\boldsymbol{C_m} = \begin{bmatrix} \tau^2_{v_p} & & & \eta_{v_p,v_s} & & & \eta_{v_p,\rho} & & \\ & \ddots & & & \ddots & & & \ddots & \\ & & \tau^2_{v_p} & & & \eta_{v_p,v_s} & & & \eta_{v_p,\rho} \\ \eta_{v_s,v_p} & & & \tau^2_{v_s} & & & \eta_{v_s,\rho} & & \\ & \ddots & & & \ddots & & & \ddots & \\ & & \eta_{v_s,v_p} & & & \tau^2_{v_s} & & & \eta_{v_s,\rho} \\ \eta_{\rho,v_p} & & & \eta_{\rho,v_s} & & & \tau^2_{\rho} & & \\ & \ddots & & & \ddots & & & \ddots & \\ & & \eta_{\rho,v_p} & & & \eta_{\rho,v_s} & & & \tau^2_{\rho} \end{bmatrix}. \tag{3.4}$$

The likelihood model describes how likely the observed data are for specified parameters. For the linearized model of the seismic data, the likelihood model is already defined in Equation 2.5. Due to the assumption of independent and normally distributed noise terms, the seismic data given the properties $\boldsymbol{m}$ are Gaussian distributed with mean $\boldsymbol{WADm} = \boldsymbol{Gm}$ and variances σ_i^2, $i = 1, ..., S$ for the different angles. Independency for the noise terms in time is assumed, and the likelihood can be written as

$$P(\boldsymbol{d}|\boldsymbol{m}) = \frac{1}{(2\pi)^{ST/2}(det(\boldsymbol{C_d}))^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{d} - \boldsymbol{Gm})^T \boldsymbol{C_d}^{-1}(\boldsymbol{d} - \boldsymbol{Gm})\right\}, \tag{3.5}$$

where $\boldsymbol{C_d}$ is a diagonal $ST \times ST$ covariance matrix with the noise for different angles, $\sigma_1, ..., \sigma_S$ on the diagonal. These variances are estimated from the seismic data, and they are assumed equal for all time samples $t = 1, 2, ..., T$. $\boldsymbol{C_d}$ could possibly be a full matrix with correlation between the noise terms.

Independency is chosen here for simplicity, which gives fewer parameters to estimate. In this model, the variances of the noise terms are the only unknown parameters.

By using Bayes' rule in Equation 3.1, the aposteriori distribution is given by

$$P(\boldsymbol{m}|\boldsymbol{d}) \propto P(\boldsymbol{d}|\boldsymbol{m})P(\boldsymbol{m}).$$

By using the expressions in equations 3.2 and 3.5, the posterior distribution for the Gaussian inversion becomes

$$P(\boldsymbol{m}|\boldsymbol{d}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m})^T{\boldsymbol{C_d}}^{-1}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m}) - \frac{1}{2}(\boldsymbol{m}-\boldsymbol{m_p})^T{\boldsymbol{C_m}}^{-1}(\boldsymbol{m}-\boldsymbol{m_p})\right\}. \tag{3.6}$$

The maximum posterior (MAP) solution for the Gaussian inversion is calculated by maximizing Equation 3.6 with respect to $\boldsymbol{m}$. This gives the most probable solution, and can be found by minimizing

$$f(\boldsymbol{m}) = \frac{1}{2}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m})^T{\boldsymbol{C_d}}^{-1}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m}) + \frac{1}{2}(\boldsymbol{m}-\boldsymbol{m_p})^T{\boldsymbol{C_m}}^{-1}(\boldsymbol{m}-\boldsymbol{m_p}).$$

Derivation with respect to $\boldsymbol{m}$ gives

$$\frac{\partial f(\boldsymbol{m})}{\partial \boldsymbol{m}} = ({\boldsymbol{C_m}}^{-1} + \boldsymbol{G}^T{\boldsymbol{C_d}}^{-1}\boldsymbol{G})\boldsymbol{m} - (\boldsymbol{G}^T{\boldsymbol{C_d}}^{-1}\boldsymbol{d} + {\boldsymbol{C_m}}^{-1}\boldsymbol{m_p}).$$

The MAP solution $\widehat{\boldsymbol{m}}_{MAP}$ is calculated by setting this equation equal to zero. This gives

$$\widehat{\boldsymbol{m}}_{MAP} = \left[{\boldsymbol{C_m}}^{-1} + \boldsymbol{G}^T{\boldsymbol{C_d}}^{-1}\boldsymbol{G}\right]^{-1}\left[\boldsymbol{G}^T{\boldsymbol{C_d}}^{-1}\boldsymbol{d} + {\boldsymbol{C_m}}^{-1}\boldsymbol{m_p}\right] \tag{3.7}$$

since $\boldsymbol{C_m}$ and $\boldsymbol{C_d}$ are symmetric matrices. The Gaussian inversion finds the conditional expectation $E[\boldsymbol{m}|\boldsymbol{d}]$. With the notation used in this report, this posterior expectation from Equation 30 in Buland and Omre (2003a) is given as

$$E\left[\boldsymbol{m}|\boldsymbol{d}\right] = \boldsymbol{m}_p + \boldsymbol{C}_m\boldsymbol{G}^T\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]^{-1}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m}_p). \tag{3.8}$$

In order to show that the MAP solution is equal to the posterior expectation when the prior distribution is Gaussian, Equation 3.8 is multiplied with

$\left[\boldsymbol{C_m}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right]$. If this is equal to $\left[\boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{d} + \boldsymbol{C_m}^{-1}\boldsymbol{m_p}\right]$, the proof is completed. This is legal because $\left[\boldsymbol{C_m}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right]$ is already defined invertible in Equation 3.7. This means that the statement that has to be proved is

$$\left[\boldsymbol{C_m}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right]\boldsymbol{m}_p + \left[\boldsymbol{C_m}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right]\boldsymbol{C}_m\boldsymbol{G}^T\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]^{-1}(\boldsymbol{d} - \boldsymbol{G}\boldsymbol{m}_p) = \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{d} + \boldsymbol{C_m}^{-1}\boldsymbol{m_p}.$$

Multiplying out some parentheses yields

$$\boldsymbol{C_m}^{-1}\boldsymbol{m_p} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\boldsymbol{m_p} + \left[\boldsymbol{C_m}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right]\boldsymbol{C}_m\boldsymbol{G}^T\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]^{-1}(\boldsymbol{d} - \boldsymbol{G}\boldsymbol{m}_p) = \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{d} + \boldsymbol{C_m}^{-1}\boldsymbol{m_p}.$$

Subtracting $\boldsymbol{C}_m^{-1}\boldsymbol{m_p}$ on both sides and moving $\boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\boldsymbol{m_p}$ to the right side gives

$$\left[\boldsymbol{C_m}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right]\boldsymbol{C}_m\boldsymbol{G}^T\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]^{-1}(\boldsymbol{d} - \boldsymbol{G}\boldsymbol{m}_p) = \boldsymbol{G}^T\boldsymbol{C}_d^{-1}(\boldsymbol{d} - \boldsymbol{G}\boldsymbol{m}_p).$$

Now it is enough to show that $\left[\boldsymbol{C_m}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right]\boldsymbol{C}_m\boldsymbol{G}^T\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]^{-1}$ is equal to $\boldsymbol{G}^T\boldsymbol{C}_d^{-1}$ since these two expressions are both multiplied to the right with the vector $\boldsymbol{d} - \boldsymbol{G}\boldsymbol{m}_p$. $\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]$ is assumed invertible in Equation 3.8, and multiplying to the right with $\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]$ on both sides gives

$$\left[\boldsymbol{C_m}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right]\boldsymbol{C}_m\boldsymbol{G}^T = \boldsymbol{G}^T\boldsymbol{C}_d^{-1}\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right].$$

Multiplying out the parenthesis now gives

$$\boldsymbol{G}^T + \boldsymbol{G}^T\boldsymbol{C}_d^{-1}\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T = \boldsymbol{G}^T\boldsymbol{C}_d^{-1}\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{G}^T,$$

and the statement is proved.

For the Gaussian case, the covariance matrix is

$$Var(\widehat{\boldsymbol{m}}_{MAP}) = \left[\frac{\partial^2 f(\boldsymbol{m})}{\partial^2 \boldsymbol{m}}\right]^{-1} = \left[\boldsymbol{C_m}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right]^{-1}. \tag{3.9}$$

The conditional variance $Var[\boldsymbol{m}|\boldsymbol{d}]$ is given in Equation 31 in Buland and Omre (2003a). With notation used here, this is

$$Var\left[\boldsymbol{m}|\boldsymbol{d}\right] = \boldsymbol{C}_m - \boldsymbol{C}_m\boldsymbol{G}^T\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]^{-1}\boldsymbol{G}\boldsymbol{C}_m. \tag{3.10}$$

In order to show that the posterior variance in Equation 3.10 is equal to the MAP variance in Equation 3.9, Equation 3.10 is multiplied with the inverse of Equation 3.9. This must be shown to be equal to the identity matrix. The conditional variance in Equation 3.10 multiplied with $\left[\boldsymbol{C_m}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right]$ means that the statement which has to be proved is

$$\left[\boldsymbol{C_m}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right]\left[\boldsymbol{C}_m - \boldsymbol{C}_m\boldsymbol{G}^T\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]^{-1}\boldsymbol{G}\boldsymbol{C}_m\right] = I.$$

Multiplying out some parenthesis, subtracting an identity matrix on both sides and moving some parts to the right side gives

$$\boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]^{-1}\boldsymbol{G}\boldsymbol{C}_m = -\boldsymbol{G}^T\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]^{-1}\boldsymbol{G}\boldsymbol{C}_m + \boldsymbol{G}^T\boldsymbol{C}_d^{-1}\boldsymbol{G}\boldsymbol{C}_m.$$

Setting $\boldsymbol{G}\boldsymbol{C}_m$ outside on the right and then multiplying the rest to the right with $\left[\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C_d}\right]$ gives

$$\boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T = -\boldsymbol{G}^T + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}(\boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C}_d),$$

and the statement is proved.

3.1 Lateral correlation along a 2D line

In order to provide lateral consistency and robust results from inverse problems, spatially coupling of the model parameters $\boldsymbol{m}$ may be a possibility to ensure this. Considering the circumstances in seismic data, such laterally coupling between traces is an obvious option, since the lithology at a certain depth is strongly correlated to the lithology at the same depth for neighbouring traces. One exception from this can be at faults, where the layers are shifted upwards or downwards. At the edge of faults, direct correlation with neighbouring traces is not preferred.

The following theory is inspired by Rue and Held (2005). Here it is applied to each trace and in a new setting. Markov properties are assumed for simplicity, implying that the inversion result for one trace is only dependent on the neighbouring traces. This can also be denoted as conditional independence. Elastic parameters at trace i and all other traces except traces $i-1$ and $i+1$, are conditional independent given the traces $i-1$ and $i+1$. Determining the properties for the traces except traces $i-1$ and $i+1$ does not give additional information about the distribution of the model parameters for trace i. This corresponds to the model of autoregressive time series. The vectors $\boldsymbol{m_j}$, $j = 1, 2, ..., M$, where

M is the number of traces, represent the elastic parameters for trace j at all time steps. The full conditional probability distribution for the laterally correlated $\boldsymbol{m} = (\boldsymbol{m}_1, \boldsymbol{m}_2, ..., \boldsymbol{m}_M)^T$ is then

$$P(\boldsymbol{m}) = P(\boldsymbol{m}_1)P(\boldsymbol{m}_2|\boldsymbol{m}_1) \cdot .... \cdot P(\boldsymbol{m}_M|\boldsymbol{m}_{M-1}). \tag{3.11}$$

The vector $\boldsymbol{m}$ therefore consists of $N \cdot M$ values since each $\boldsymbol{m}_j$ is an $N \times 1$ vector. An autoregressive process of order 1 with standard normal errors $\boldsymbol{\epsilon_j}$ is expressed as

$$\boldsymbol{m}_j = \phi \boldsymbol{m}_{j-1} + \boldsymbol{\epsilon_j}, \quad \boldsymbol{\epsilon_j} \overset{iid}{\sim} N(0, I) \tag{3.12}$$

where iid means *independent identical distributed*, and $\phi \in [0, 1]$ is the lateral correlation coefficient. $\phi = 0$ corresponds to zero lateral correlation. In conditional form, for $j = 2, ..., M$, this can be written as

$$\boldsymbol{m}_j|\boldsymbol{m}_1, \boldsymbol{m}_2, ..., \boldsymbol{m}_{j-1} \sim N(\boldsymbol{m}_{\boldsymbol{p}_j} + \phi(\boldsymbol{m}_{j-1} - \boldsymbol{m}_{\boldsymbol{p}_{j-1}}), I), \tag{3.13}$$

which is common in time series (Rue and Held (2005)). By assuming $\boldsymbol{m}_1 \sim N(\boldsymbol{m}_{\boldsymbol{p}_1}, \frac{1}{1-\phi^2} I)$, the full conditional probability distribution can be written as

$$P(\boldsymbol{m}) = \frac{1}{(2\pi)^{M/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{m} - \boldsymbol{m_p})^T \boldsymbol{\Sigma^{-1}} (\boldsymbol{m} - \boldsymbol{m_p})\right\}, \tag{3.14}$$

where $\boldsymbol{\Sigma^{-1}}$ is the precision matrix or the inverse covariance matrix. This consists of two elements, $\boldsymbol{Q}$ and $\boldsymbol{C}_m^{-1}$ that is explained next. The covariance matrix is a full matrix, whereas the precision matrix for each trace is sparse due to the Markov assumptions, with the structure

$$\boldsymbol{Q} = \begin{bmatrix} 1 & -\phi & & & & & & \\ -\phi & 1+\phi^2 & -\phi & & & & & \\ & \cdot & \cdot & \cdot & & & & \\ & & \cdot & \cdot & \cdot & & & \\ & & & \cdot & \cdot & \cdot & & \\ & & & & \cdot & \cdot & \cdot & \\ & & & & & -\phi & 1+\phi^2 & -\phi \\ & & & & & & -\phi & 1 \end{bmatrix}. \tag{3.15}$$

Equation 3.15 gives the correlation structure between the traces laterally, and corresponds to a covariance matrix with $\frac{1}{1-\phi^2}$ on the diagonal. This can be extended to 2D (and 3D) by assuming $\boldsymbol{m}_j$ as a vector of length $N = 3T$, where T is the number of time samples. As before, independence between the elastic parameters in time is assumed. Then, the result from the inversion is a vector of length $N \cdot M$. Equation 3.15 is extended to 2D by multiplying each entry with the inverse of the $N \times N$ covariance matrix $\boldsymbol{C_m}$ for each trace (Equation 3.4). This is the Kronecker product of $\boldsymbol{Q}$ and $\boldsymbol{C}_m^{-1}$. The resulting covariance matrix is $NM \times NM$, and can be written as

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{C_m}^{-1} & -\phi\boldsymbol{C_m}^{-1} & & & & & \\ -\phi\boldsymbol{C_m}^{-1} & (1+\phi^2)\boldsymbol{C_m}^{-1} & -\phi\boldsymbol{C_m}^{-1} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -\phi\boldsymbol{C_m}^{-1} & (1+\phi^2)\boldsymbol{C_m}^{-1} & -\phi\boldsymbol{C_m}^{-1} & \\ & & & & -\phi\boldsymbol{C_m}^{-1} & \boldsymbol{C_m}^{-1} \end{bmatrix}. \tag{3.16}$$

$\boldsymbol{\Sigma}^{-1}$ is an extension of the precision matrix in Equation 3.15 in 2D. $\boldsymbol{\Sigma}$ includes both the variances and covariances between the three elastic parameters $\boldsymbol{v_p}$, $\boldsymbol{v_s}$ and $\boldsymbol{\rho}$, and the covariances between neighbouring traces laterally. The covariance matrix for the noise terms, $\boldsymbol{C}_d$, is also extended to 2D. This becomes an $NM \times NM$ diagonal matrix since independency between traces is assumed.

The MAP solution in 2D is given as an extension of Equation 3.7. This gives

$$\boldsymbol{R}\widehat{\boldsymbol{m}}_{MAP} = \boldsymbol{b},$$

where

$$\boldsymbol{R} = \left[\boldsymbol{\Sigma}^{-1} + \boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{G}\right] \tag{3.17}$$

and

$$\boldsymbol{b} = \left[\boldsymbol{G}^T\boldsymbol{C_d}^{-1}\boldsymbol{d} + \boldsymbol{\Sigma}^{-1}\boldsymbol{m_p}\right].$$

$\boldsymbol{G}$ is a sparse and block diagonal $NM \times NM$ matrix with the matrix $\boldsymbol{G}_j = \boldsymbol{W}\boldsymbol{A}_j\boldsymbol{D}$ for each trace on the main diagonal:

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}_1 & & & & \\ & \boldsymbol{G}_2 & & & \\ & & \boldsymbol{G}_3 & & \\ & & & \ddots & \\ & & & & \boldsymbol{G}_M \end{bmatrix}. \tag{3.18}$$

In this work, the convolution matrix $\boldsymbol{W}$ and the differential matrix $\boldsymbol{D}$ are equal for all traces, and the reflectivity matrix $\boldsymbol{A}$ depends on the low frequency background model for the current trace. Due to instability in the inversion result when varying the $\boldsymbol{A}$ matrix for each trace, $\boldsymbol{G}_i = \boldsymbol{G}_j, \forall i, j$ is used for the Gaussian inversion. In cases with enough information available to estimate different wavelets for each trace, $\boldsymbol{W}$ could vary along the line as well.

3.1.1 Numerical solution

If the number of traces on the seismic line is large, $\boldsymbol{G}$ is a very large matrix. Therefore, a numerical solution that avoids setting up the entire matrix $\boldsymbol{R}$ is required in order to determine the solution on a computer. Several methods have been used such as Sequential Gaussian Simulation or Cholesky factorization of the $\boldsymbol{R}$ matrix which is used here. This factorization determines a lower triangular matrix $\boldsymbol{L}$ which satisfies

$$\boldsymbol{R} = \boldsymbol{L}\boldsymbol{L}^T.$$

This is a special case of the $\boldsymbol{LU}$ factorization where the upper triangular matrix $\boldsymbol{U}$ is equal to the transpose of the lower triangular matrix $\boldsymbol{L}$ due to the sparseness in the Markov problem. A lower triangular matrix is a matrix with nonzero elements only on or below the diagonal. Because of the assumed Markov property in this case, the matrix $\boldsymbol{R}$ has a sparse block tri-diagonal structure. The Cholesky factorization is therefore a good choice due to computational efficiency. The lower triangular matrix $\boldsymbol{L}$ inherits the structure from $\boldsymbol{R}$, and $\boldsymbol{L}$ can be written as

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{L}_{1,1} & & & & & & & \\ \boldsymbol{L}_{2,1} & \boldsymbol{L}_{2,2} & & & & & & \\ & \boldsymbol{L}_{3,2} & \boldsymbol{L}_{3,3} & & & & & \\ & & \cdot & \cdot & & & & \\ & & & \cdot & \cdot & & & \\ & & & & \cdot & \cdot & & \\ & & & & & \boldsymbol{L}_{M-1,M-2} & \boldsymbol{L}_{M-1,M-1} & \\ & & & & & & \boldsymbol{L}_{M,M-1} & \boldsymbol{L}_{M,M} \end{bmatrix}. \tag{3.19}$$

$\boldsymbol{L}_{i,j}$, $i, j \in \{1, 2, ..., M\}$ are $N \times N$ matrices. By multiplying $\boldsymbol{L}$ and $\boldsymbol{L}^T$, the matrix $\boldsymbol{R}$ becomes

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{L}_{1,1}\boldsymbol{L}_{1,1}^T & \boldsymbol{L}_{1,1}\boldsymbol{L}_{2,1}^T & & \\ \boldsymbol{L}_{2,1}\boldsymbol{L}_{1,1}^T & \boldsymbol{L}_{2,1}\boldsymbol{L}_{2,1}^T + \boldsymbol{L}_{2,2}\boldsymbol{L}_{2,2}^T & \boldsymbol{L}_{2,2}\boldsymbol{L}_{3,2}^T & \\ & \ddots & \ddots & \ddots \\ & & \boldsymbol{L}_{M,M-1}\boldsymbol{L}_{M-1,M-1}^T & \boldsymbol{L}_{M,M-1}\boldsymbol{L}_{M,M-1}^T + \boldsymbol{L}_{M,M}\boldsymbol{L}_{M,M}^T \end{bmatrix}$$

Since most of the lower triangular matrix $\boldsymbol{L}$ contains only zeros, the non-zero elements of $\boldsymbol{L}$ is of size $NM \times 2N$. These elements, $\boldsymbol{L}_{i,j}$, $i, j \in \{1, 2, ..., M\}$, can be constructed sequential in this order:

- $\boldsymbol{R}_{1,1} = \boldsymbol{L}_{1,1}\boldsymbol{L}_{1,1}^T \Rightarrow \boldsymbol{L}_{1,1} = chol(\boldsymbol{R}_{1,1})$
- $\boldsymbol{R}_{1,2} = \boldsymbol{L}_{1,1}\boldsymbol{L}_{2,1}^T \Rightarrow \boldsymbol{L}_{2,1}^T = \boldsymbol{L}_{1,1}^{-1}\boldsymbol{R}_{1,2}$
- $\boldsymbol{R}_{2,2} = \boldsymbol{L}_{2,1}\boldsymbol{L}_{2,1}^T + \boldsymbol{L}_{2,2}\boldsymbol{L}_{2,2}^T \Rightarrow \boldsymbol{L}_{2,2} = chol(\boldsymbol{R}_{2,2} - \boldsymbol{L}_{2,1}\boldsymbol{L}_{2,1}^T)$

$$\vdots$$

$$\vdots$$

- $\boldsymbol{R}_{M-1,M} = \boldsymbol{L}_{M-1,M-1}\boldsymbol{L}_{M,M-1}^T \Rightarrow \boldsymbol{L}_{M,M-1}^T = \boldsymbol{L}_{M-1,M-1}^{-1}\boldsymbol{R}_{M-1,M}$
- $\boldsymbol{R}_{M,M} = \boldsymbol{L}_{M,M-1}\boldsymbol{L}_{M,M-1}^T + \boldsymbol{L}_{M,M}\boldsymbol{L}_{M,M}^T \Rightarrow \boldsymbol{L}_{M,M} = chol(\boldsymbol{R}_{M,M} - \boldsymbol{L}_{M,M-1}\boldsymbol{L}_{M,M-1}^T)$

$chol(\boldsymbol{A})$ means the Cholesky factorization of the matrix $\boldsymbol{A}$. The matrices $\boldsymbol{R}_{i,j}$ are taken from the definition of $\boldsymbol{R}$ in Equation 3.17, and are given by

$$\boldsymbol{R}_{1,1} = \boldsymbol{C}_m^{-1} + \boldsymbol{G}_1^T\boldsymbol{C}_{d_1}^{-1}\boldsymbol{G}_1,$$

$$\boldsymbol{R}_{M,M} = \boldsymbol{C}_m^{-1} + \boldsymbol{G}_M^T\boldsymbol{C}_{d_M}^{-1}\boldsymbol{G}_M,$$

$$\boldsymbol{R}_{i,i} = (1+\phi^2)\boldsymbol{C}_m^{-1} + \boldsymbol{G}_i^T \boldsymbol{C}_{d_i}^{-1} \boldsymbol{G}_i,$$

and

$$\boldsymbol{R}_{i,i+1} = -\phi \boldsymbol{C}_m^{-1},$$

where $\boldsymbol{G}_i$ and $\boldsymbol{C}_{d_i}$ correspond to trace i.

The intention with the Cholesky factorization is, as mentioned, to avoid setting up the full $NM \times NM$ matrix $\boldsymbol{R}$, but rather do this sequentially trace by trace. The problem to be solved is

$$\boldsymbol{R}\widehat{\boldsymbol{m}}_{\boldsymbol{MAP}} = \boldsymbol{b}. \tag{3.20}$$

After Cholesky factorization of $\boldsymbol{R}$, the problem can be solved by first calculating

$$\boldsymbol{L}\boldsymbol{v} = \boldsymbol{b},$$

and then, using the resulting vector $\boldsymbol{v}$, solve

$$\boldsymbol{L}^T \widehat{\boldsymbol{m}}_{\boldsymbol{MAP}} = \boldsymbol{v}.$$

These two steps can be done trace by trace by solving $\boldsymbol{v} = \boldsymbol{L}^{-1}\boldsymbol{b}$ forward and then $\boldsymbol{m} = \boldsymbol{L}^{-T}\boldsymbol{v}$ backward in for-loops for all traces $j = 1, 2, ..., M$. Within these two for-loops, there is just need for inverting $N \times N$ matrices $\boldsymbol{L}_{j,j}$, $j = 1, 2, ..., M$, which are small enough for the memory on a regular computer.

Solving $\boldsymbol{L}\boldsymbol{v} = \boldsymbol{b}$ forward

The forward problem can be written as

$$\begin{bmatrix} \boldsymbol{L}_{1,1} & & & & & & & \\ \boldsymbol{L}_{2,1} & \boldsymbol{L}_{2,2} & & & & & & \\ & \boldsymbol{L}_{3,2} & \boldsymbol{L}_{3,3} & & & & & \\ & & \cdot & \cdot & & & & \\ & & & \cdot & \cdot & & & \\ & & & & \cdot & \cdot & & \\ & & & & & \boldsymbol{L}_{M-1,M-2} & \boldsymbol{L}_{M-1,M-1} & \\ & & & & & & \boldsymbol{L}_{M,M-1} & \boldsymbol{L}_{M,M} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{v}_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{b}_M \end{bmatrix}, \tag{3.21}$$

where $\boldsymbol{v}_j$ and $\boldsymbol{b}_j$, $j = 1, 2, ..., M$ are $N \times 1$ vectors for each trace. This can be solved tracewise by

- Initialize

$$\boldsymbol{v}_1 = \boldsymbol{L}_{1,1}^{-1}\boldsymbol{b}_1$$

- Iterate forward, $j = 2, 3, ..., M$

$$\boldsymbol{v}_j = \boldsymbol{L}_{j,j}^{-1}(\boldsymbol{b}_j - \boldsymbol{L}_{j,j-1}\boldsymbol{v}_{j-1})$$

Solving $\boldsymbol{L}^T\boldsymbol{m} = \boldsymbol{v}$ backward

When the vector $\boldsymbol{v}$ is built in the forward loop, $\boldsymbol{L}^T\boldsymbol{m} = \boldsymbol{v}$ can be solved backward. This problem can be written as

$$\begin{bmatrix} \boldsymbol{L}_{1,1}^T & \boldsymbol{L}_{2,1}^T & & & & & & \\ & \boldsymbol{L}_{2,2}^T & \boldsymbol{L}_{3,2}^T & & & & & \\ & & \boldsymbol{L}_{3,3}^T & \boldsymbol{L}_{4,3}^T & & & & \\ & & & \cdot & \cdot & & & \\ & & & & \cdot & \cdot & & \\ & & & & & \cdot & & \\ & & & & & & \boldsymbol{L}_{M-1,M-1}^T & \boldsymbol{L}_{M,M-1}^T \\ & & & & & & & \boldsymbol{L}_{M,M}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{m}_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{v}_M \end{bmatrix}, \tag{3.22}$$

and is solved stepwise by

- Initialize

$$\boldsymbol{m}_M = \boldsymbol{L}_{M,M}^{-T}\boldsymbol{v}_M$$

- Iterate backward, $j = M-1, M-2, ..., 1$

$$\boldsymbol{m}_j = \boldsymbol{L}_{j,j}^{-T}(\boldsymbol{v}_j - \boldsymbol{L}_{j+1,j}^T\boldsymbol{m}_{j+1})$$

3.2 Estimation of Σ_0 from well logs

The estimation of the covariance matrix in Equation 3.3 from well logs can be performed straight forward by using standard estimates for variances and covariances as

$$\widehat{\boldsymbol{\tau}}_\xi^2 = \frac{1}{T-1}\sum_{i=1}^{T}(m_{\xi_i} - \overline{m_\xi})^2, \quad \xi \in \{\boldsymbol{v_p}, \boldsymbol{v_s}, \boldsymbol{\rho}\}, \tag{3.23}$$

and

$$\widehat{\boldsymbol{\eta}}_{\alpha,\beta} = \frac{1}{T-1}\sum_{i=1}^{T}(m_{\alpha_i} - \overline{m_\alpha})(m_{\beta_i} - \overline{m_\beta}), \quad \alpha,\beta \in \{\boldsymbol{v_p}, \boldsymbol{v_s}, \boldsymbol{\rho}\}, \tag{3.24}$$

where $\overline{m}$ is the mean of the m_i's from $i = 1, 2, ..., T$.

3.3 Estimation of Σ_0 and ϕ from seismic data

Because of the completely different scales for well log data and seismic data, the covariance matrix in Equation 3.3 should be estimated from seismic data instead of well log data in order to achieve consistency in the results. One alternative method to do this is described by the Figure 3.1 workflow. In this case, the covariance matrix from well logs is multiplied with a scaling parameter γ, $\boldsymbol{\Sigma_0} = \gamma\boldsymbol{\Sigma}_{0_{well}}$. $P(\boldsymbol{d})$ is Gaussian distributed, $P(\boldsymbol{d}) \sim N(\boldsymbol{G}\boldsymbol{m}_p, \boldsymbol{G}\boldsymbol{C}_m\boldsymbol{G}^T + \boldsymbol{C}_d)$. From Bayes rule in Equation 3.1, this can also be written as

$$P(\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{m})P(\boldsymbol{m})}{P(\boldsymbol{m}|\boldsymbol{d})}, \tag{3.25}$$

where the mean and the variance of the Gaussian density $P(\boldsymbol{m}|\boldsymbol{d})$ is given in Equations 3.8 and 3.10. $P(\boldsymbol{d}) = P_{\gamma,\phi}(\boldsymbol{d})$ can now be calculated for a range of γ and ϕ values used in the inversion. The best choice of γ and ϕ was assumed to be the pair of parameter values that corresponds to a maximization of $P_{\gamma,\phi}(\boldsymbol{d})$

3.4 Simulation from a Gaussian Markov Random Field

One way of introducing inference in complex hierarchical models is to use simulation-based methods. Variances and means can be calculated by creating several realizations from the target distribution, and use Monte Carlo methods on them. In general this means that the target integral

$$I = \int_D g(\boldsymbol{m})P(\boldsymbol{m}|\boldsymbol{d})d\boldsymbol{m},$$

where D is a high dimensional region, is approximated by

$$\widehat{I}_n = \frac{1}{n}\{g(\boldsymbol{m}^{(1)}) + ... + g(\boldsymbol{m}^{(n)})\}.$$

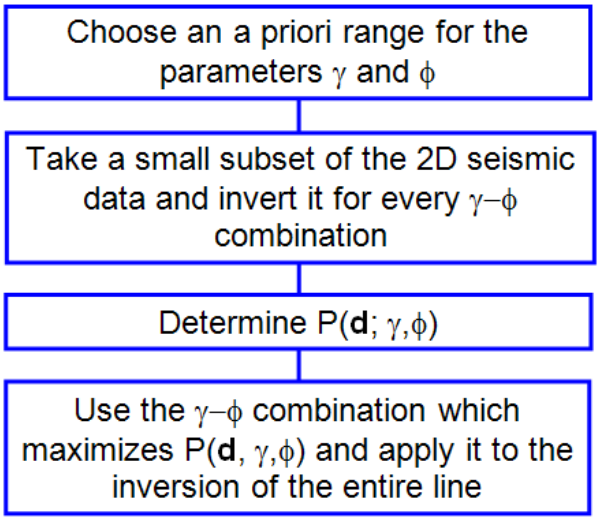

Figure 3.1: Generalized workflow of the estimation method for the parameters in the Gaussian inversion.

This approximation is exact when the number of samples, n, approaches infinity. For the Gaussian inversion, there is need for simulation from a Gaussian Markov Random Field (GMRF). One way of doing this is to simulate from a zero mean GMRF with precision matrix $\boldsymbol{R}$, and add the MAP solution as a mean afterwards. This can be done by computing the Cholesky triangle of $\boldsymbol{R}$ ($\boldsymbol{L}$) and solve

$$\boldsymbol{L}^T\boldsymbol{x} = \boldsymbol{z},$$

where $\boldsymbol{z}$ is a vector of independent standard normal variables, and $\boldsymbol{x}$ is the model parameters $\boldsymbol{m}$ before the low frequency model is added, see Rue and Held (2005). The sparse structure of $\boldsymbol{L}$ will make this step efficient. This method can also be used as an approximation for the blocky inversion, with the Cholesky factorization of the $\boldsymbol{R}$ matrix in Equation 4.11. The algorithm for making realizations from the posterior distribution $P(\boldsymbol{m}|\boldsymbol{d})$ is

- Set up the lower triangular matrix $\boldsymbol{L}$ created from Cholesky factorization of the precision matrix
- Make realizations from a standard normal distribution, $\boldsymbol{z} \sim N(0, \boldsymbol{I})$
- Iterate backward to get realizations from a zero mean GMRF $\boldsymbol{x}$

  Initalize $\boldsymbol{x}_M = \boldsymbol{L}_{M,M}^{-T}\boldsymbol{z}_M$

  Iterate backward, $j = M-1, M-2, ..., 1$

$$\boldsymbol{x}_j = \boldsymbol{L}_{j,j}^{-T}(\boldsymbol{z}_j - \boldsymbol{L}_{j+1,j}^{T}\boldsymbol{x}_{j+1})$$

- Add the inversion result $\widehat{\boldsymbol{m}}_{MAP}$ as a mean to the zero mean vector $\boldsymbol{x}$

After doing this, a realization of the posterior distribution for all traces $j = 1, 2, ..., M$ is available. This can be repeated, and inference on the realizations will provide estimates of the variance of the MAP-solution. This is accomplished by calculating the inverse of the lognormal cumulative distribution function, since the interesting property is the variance of $\exp\{\boldsymbol{m}|\boldsymbol{d}\}$.

4 BLOCKY BAYESIAN INVERSION

In order to achieve a more realistic image of the changes in the properties $\boldsymbol{m}$ in the subsurface, the idea is to assume a Cauchy distribution for the contrasts $\boldsymbol{Dm}$. The principle idea of this approach is to enforce a sparse distribution of the model gradients, thereby sharpening the contrast of boundaries between adjacent layers. This is based on the assumption that mainly intermediate gradients contribute to the apparent smoothing in seismic inversion. In theory, the inverted properties of a layer have then sharp, flat boundaries at the top and bottom, while they are constant within the layer. The Cauchy distribution can be characterized mathematically as a probability distribution with heavier tails than the Gaussian distribution, which gives more probability to larger contrasts, whilst suppressing intermediate values. This can contribute to sharper edges between the layers. Figure 4.1 shows a comparison of the Gaussian and the Cauchy distributions. The blue curve is the Gaussian distribution and the red curve is the Cauchy distribution for corresponding maximum values.

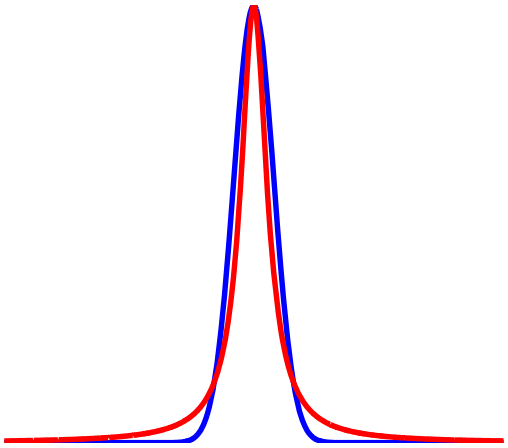

Figure 4.1: Comparison of the Gaussian and the Cauchy distributions. The blue curve is the Gaussian distribution and the red curve is the Cauchy distribution.

The Cauchy distribution is in this work applied to the contrasts $\boldsymbol{Dm}$ in the blocky inversion. Kolbjørnsen (2002a), also used the Cauchy prior, but for the parameters directly, and not for the gradients. For each trace, the prior distribu-

tion is given as

$$
\begin{aligned}
P(\boldsymbol{m}) &= \prod_{i=1}^{N} \frac{1}{\pi\kappa\left(1+\left(\frac{(Dm)_i-(Dm)_{p_i}}{\kappa}\right)^2\right)} \\
&= \frac{1}{(\pi\kappa)^N}\exp\left\{-\sum_{i=1}^{N}\ln\left(1+\left(\frac{(Dm)_i-(Dm)_{p_i}}{\kappa}\right)^2\right)\right\},
\end{aligned}
\tag{4.1}
$$

where $(Dm)_{p_i}$ is the expected value for the contrast $(Dm)_i$ given by the low frequency model $(Dm)_{p_i} = Dm_{p_i}$ and κ is the scale parameter in the Cauchy distribution. No dependency between the $(Dm)_i$'s in time is assumed. The expression for the contrasts $\boldsymbol{Dm}$ is given in Equation 2.7. The likelihood $P(\boldsymbol{d}|\boldsymbol{m})$ is not changed from the Gaussian inversion, and is given in Equation 3.5. The resulting posterior distribution in 1D is then given as

$$
\begin{aligned}
P(\boldsymbol{m}|\boldsymbol{d}) &\propto P(\boldsymbol{d}|\boldsymbol{m})P(\boldsymbol{m}) \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{WADm})^T\boldsymbol{C}_d^{-1}(\boldsymbol{d}-\boldsymbol{WADm})\right\} \\
&\cdot \exp\left\{-\sum_{i=1}^{N}\ln\left(1+\left(\frac{(Dm)_i-(Dm)_{p_i}}{\kappa}\right)^2\right)\right\}.
\end{aligned}
\tag{4.2}
$$

4.1 Lateral correlation along a 2D line

To introduce lateral correlation along a 2D line to the blocky inversion scheme, the prior distribution is assumed proportional to a product of a Gaussian and a Cauchy distribution. The model for the lateral correlation is described by Equation 3.12, while the prior model for each trace vertically is Cauchy distributed as in Equation 4.1. This gives Gaussian distribution laterally with Cauchy constraints vertically, and can be expressed by

$$
P(\boldsymbol{m}) \propto P_G(\boldsymbol{m})P_C(\boldsymbol{m}), \tag{4.3}
$$

where $P_G(\boldsymbol{m})$ is the Gaussian prior given in Equation 3.14 and $P_C(\boldsymbol{m})$ is the product of the 1D prior distribution in Equation 4.1 for all traces $j = 1, 2, ..., M$.

This means that the Gaussian distribution takes care of the lateral correlation, while the Cauchy distribution is independent for the traces. The full prior distribution for the elastic parameters for all traces can then be expressed as

$$
\begin{aligned}
P(\boldsymbol{m}) &\propto \frac{1}{(2\pi)^{M/2}}|\boldsymbol{Q}|^{1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{m}-\boldsymbol{m_p})^T\boldsymbol{Q}(\boldsymbol{m}-\boldsymbol{m_p})\right\} \\
&\cdot\prod_{j=1}^{M}\prod_{i=1}^{N}\left(\frac{1}{\pi\kappa\left(1+\left(\frac{(Dm)_i-(Dm)_{p_i}}{\kappa}\right)^2\right)}\right)_j \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{m}-\boldsymbol{m_p})^T\boldsymbol{Q}(\boldsymbol{m}-\boldsymbol{m_p})-\sum_{j=1}^{M}\sum_{i=1}^{N}\ln\left(1+\left(\frac{(Dm)_i-(Dm)_{p_i}}{\kappa}\right)^2\right)_j\right\},
\end{aligned}
\tag{4.4}
$$

where the precision matrix $\boldsymbol{Q}$ from Equation 3.15 is extended to 2D by multiplying each entry with an $N\times N$ identity matrix. Since the likelihood distribution is still assumed Gaussian, the posterior distribution is given by

$$
\begin{aligned}
P(\boldsymbol{m}|\boldsymbol{d}) &\propto P(\boldsymbol{d}|\boldsymbol{m})P(\boldsymbol{m}) \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m})^T{\boldsymbol{C_d}}^{-1}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m})-\frac{1}{2}(\boldsymbol{m}-\boldsymbol{m_p})^T\boldsymbol{Q}(\boldsymbol{m}-\boldsymbol{m_p})\right\} \\
&\cdot\exp\left\{-\sum_{j=1}^{M}\sum_{i=1}^{N}\ln\left(1+\left(\frac{(Dm)_i-(Dm)_{p_i}}{\kappa}\right)^2\right)_j\right\}.
\end{aligned}
\tag{4.5}
$$

As before, this expression needs to be maximized in order to find the maximum posterior solution for the properties $\boldsymbol{m}$, or equally, minimizing the cost function $O(\boldsymbol{m})$ given by

$$
\begin{aligned}
O(\boldsymbol{m}) &= \frac{1}{2}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m})^T{\boldsymbol{C_d}}^{-1}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m})+\frac{1}{2}(\boldsymbol{m}-\boldsymbol{m_p})^T\boldsymbol{Q}(\boldsymbol{m}-\boldsymbol{m_p}) \\
&+\sum_{j=1}^{M}\sum_{i=1}^{N}\ln\left(1+\left(\frac{(Dm)_i-(Dm)_{p_i}}{\kappa}\right)^2\right)_j.
\end{aligned}
\tag{4.6}
$$

To differentiate $O(\boldsymbol{m})$ with respect to $\boldsymbol{m}$, a linearization of the last part of this function is necessary. The following is done according to Charbonnier et al. (1997). Because of the non-linearity in the resulting equation, this needs to be solved iteratively. By Taylor expansion around the previous step in the iteration, $O(\boldsymbol{m})$ can be written approximately by

$$\begin{aligned} O(\boldsymbol{m}) \approx \frac{1}{2}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m})^T {\boldsymbol{C_d}}^{-1}(\boldsymbol{d}-\boldsymbol{G}\boldsymbol{m}) + \frac{1}{2}(\boldsymbol{m}-\boldsymbol{m_p})^T\boldsymbol{Q}(\boldsymbol{m}-\boldsymbol{m_p}) \\ + 2(\boldsymbol{m}-\boldsymbol{m_p})^T\boldsymbol{D}^T\boldsymbol{B}\boldsymbol{D}(\boldsymbol{m}-\boldsymbol{m_p}), \end{aligned} \tag{4.7}$$

where $\boldsymbol{m}$ now is the result from the previous iteration step, $\boldsymbol{m_p}$ is the prior low frequency expected model, and $\boldsymbol{B}$ is a sparse diagonal matrix with non-zero elements given by

$$B_{ii} = \frac{1}{\kappa^2 + ((Dm)_i - (Dm)_{p_i})^2}. \tag{4.8}$$

This approximation is legitimate because of the weak non-linearity in the problem. Therefore, the inverse problem can be treated as a quasi-linear problem. Then, differentiation of $O(\boldsymbol{m})$ with respect to $\boldsymbol{m}$ is

$$\frac{\partial O(\boldsymbol{m})}{\partial \boldsymbol{m}} = -\boldsymbol{G}^T{\boldsymbol{C_d}}^{-1}\boldsymbol{d} + \boldsymbol{G}^T{\boldsymbol{C_d}}^{-1}\boldsymbol{G}\boldsymbol{m} + \boldsymbol{Q}\boldsymbol{m} - \boldsymbol{Q}\boldsymbol{m_p} + 4(\boldsymbol{D}^T\boldsymbol{B}\boldsymbol{D}\boldsymbol{m} - \boldsymbol{D}^T\boldsymbol{B}\boldsymbol{D}\boldsymbol{m_p}). \tag{4.9}$$

The MAP solution is achieved by setting this equal to zero. This yields

$$\widehat{\boldsymbol{m}}_{MAP} = \left[\boldsymbol{G}^T{\boldsymbol{C_d}}^{-1}\boldsymbol{G} + \boldsymbol{Q} + 4\boldsymbol{D}^T\boldsymbol{B}\boldsymbol{D}\right]^{-1}\left[\boldsymbol{G}^T{\boldsymbol{C_d}}^{-1}\boldsymbol{d} + \boldsymbol{Q}\boldsymbol{m_p} + 4\boldsymbol{D}^T\boldsymbol{B}\boldsymbol{D}\boldsymbol{m_p}\right]. \tag{4.10}$$

The size of the matrices in Equation 4.10 is $NM \times NM$. The $\boldsymbol{G}$ matrix is given in Equation 3.18 and $\boldsymbol{C}_d$ is the data noise covariance matrix for all traces. For simplicity, this is assumed diagonal, which corresponds to non-correlated random noise. The data vector $\boldsymbol{d}$ is a $TSM \times 1$ vector with seismic data for all traces and angles. $\boldsymbol{D}$ is the differential matrix defined in Equation 2.7, extended to 2D by a block diagonal matrix with $N \times N$ matrices on the main diagonal. The $N \times N$ matrix $\boldsymbol{B}$ with non-zero elements from Equation 4.8 is also extended to 2D by making a block diagonal $NM \times NM$ matrix for all traces. As for the Gaussian

problem, Cholesky factorization is used to find the MAP solution in Equation 4.10. Figure 4.2 shows a generalized workflow of the blocky inversion scheme. The difference from the Gaussian inversion is in the $\boldsymbol{B}$ matrix, which needs to be updated iteratively due to the non-linearity. It is shown in Theune (2007) that there is just need for two updates. More updates do not improve the inversion result further. The Cholesky solution of

$$\boldsymbol{R}\widehat{\boldsymbol{m}}_{MAP} = \boldsymbol{b}$$

is then performed with

$$\boldsymbol{R} = [\boldsymbol{G}^T \boldsymbol{C_d}^{-1} \boldsymbol{G} + \boldsymbol{Q} + 4\boldsymbol{D}^T \boldsymbol{B} \boldsymbol{D}] \tag{4.11}$$

and

$$\boldsymbol{b} = [\boldsymbol{G}^T \boldsymbol{C_d}^{-1} \boldsymbol{d} + \boldsymbol{Q}\boldsymbol{m_p} + 4\boldsymbol{D}^T \boldsymbol{B} \boldsymbol{D}\boldsymbol{m_p}].$$

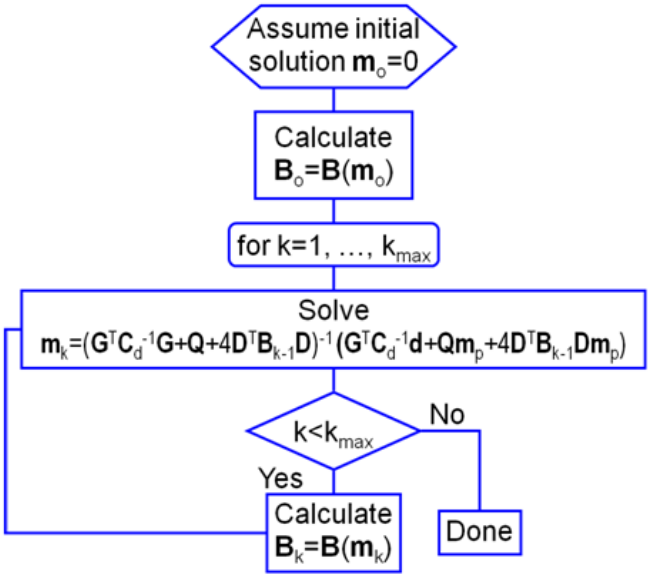

Figure 4.2: Generalized workflow of the blocky inversion scheme.

Asymptotic theory says that the variance of the MAP solution can be expressed by the fitted curvature at the mode. Theorem 5.1 in Carlin and Louis (2000) states that the posterior distribution can be expressed by a normal distribution with mean equal to the posterior mode (MAP solution) and covariance matrix equal to the negative inverse of the second derivative of the log posterior evaluated at the mode. This statement is exact if the likelihood and prior models both are Gaussian. It can also be used as an approximation if the number of

data points is much greater than the number of variables. This theory can be used to calculate an approximation of the covariance matrix for the MAP solution in the blocky inversion, and gives

$$Var(\widehat{\boldsymbol{m}}_{MAP}) \approx \left[\frac{\partial^2 O(\boldsymbol{m})}{\partial^2 \boldsymbol{m}}\right]^{-1}_{|\boldsymbol{m}=\widehat{\boldsymbol{m}}_{MAP}} = \left[\boldsymbol{G}^T {\boldsymbol{C_d}}^{-1}\boldsymbol{G} + \boldsymbol{Q} + 4\boldsymbol{D}^T\boldsymbol{B}\boldsymbol{D}\right]^{-1}. \quad (4.12)$$

The Gauss approximation for the blocky MAP solution is checked by importance sampling described in Section 4.4.

4.2 Maximum likelihood estimation of the scale parameter $\boldsymbol{\kappa}$ from well logs

The aim of the maximum likelihood estimation (MLE) method is to find the most reasonable parameters, i.e. to find the parameters that cause a maximization of the product of the probability density functions for all the stochastic variables represented. Since one different $\boldsymbol{\kappa}$ value for each of the elastic parameters is assumed, the maximum likelihood estimation can be divided into three parts. The maximum likelihood function for the contrasts $\boldsymbol{Dm}$ for each of the elastic parameters is defined as

$$L(\boldsymbol{Dm}) = \prod_{t=1}^{T} f((Dm)_t) \quad (4.13)$$

where $f((Dm)_t)$ is the probability density function for the stochastic variables $(Dm)_t$, $t = 1, 2, ..., T$. A maximization of this function with respect to the sought parameters will give the most probable solution. In this case, the contrasts $\boldsymbol{Dm}$ are Cauchy distributed,

$$(Dm)_t \sim Cauchy((Dm)_{p_t}, \kappa),$$

with the low frequency prior model used as a mode for $\boldsymbol{Dm}$. The Cauchy distribution is an example of a distribution that has no mean, variance or higher order moments, but the mode and median are both well defined. The parameter

$\boldsymbol{\kappa}$ is called the scale parameter, and specifies the width of the distribution function. Since all the $(Dm)_t$'s in time are assumed independent for all $t = 1, 2, ..., T$, the likelihood function becomes

$$L((Dm)_1, (Dm)_2, ..., (Dm)_T; (Dm)_p, \boldsymbol{\kappa}) = \prod_{t=1}^{T} f((Dm)_t; (Dm)_{p_t}, \kappa),$$

where

$$f((Dm)_t; (Dm)_{p_t}, \kappa) = \frac{1}{\pi\kappa(1 + ((Dm)_t - (Dm)_{p_t})^2/\kappa^2)}.$$

This included in the likelihood function yields

$$L((Dm)_1, (Dm)_2, ..., (Dm)_T; (Dm)_p, \boldsymbol{\kappa}) = \prod_{t=1}^{T} \frac{1}{\pi\kappa(1 + ((Dm)_t - (Dm)_{p_t})^2/\kappa^2)}.$$

It is often easier to maximize the logarithm of the likelihood function instead of the likelihood itself. This will cause no problems since the logarithm function is strongly increasing. For the well log estimation, different κ's for the three elastic parameters were assumed. The log likelihood function for each of the three parameters is given by

$$l = \ln(L) = -T\ln(\pi\kappa_\xi) - \sum_{t=1}^{T} \ln(1 + (((Dm)_t - (Dm)_{p_t})^2/\kappa_\xi^2)), \tag{4.14}$$

where $\xi \in \{\boldsymbol{v_p}, \boldsymbol{v_s}, \boldsymbol{\rho}\}$. The maximum likelihood estimator for κ_ξ results from maximizing the log likelihood, i.e. setting the derivative of Equation 4.14 equal to zero,

$$\frac{\partial l}{\partial \kappa_\xi} = -\frac{T}{\kappa_\xi} + \frac{2}{\kappa_\xi^3} \sum_{t=1}^{T} \frac{((Dm)_t - (Dm)_{p_t})^2}{1 + ((Dm)_t - (Dm)_{p_t})^2/\kappa_\xi^2} = 0.$$

This can be solved with respect to κ_ξ numerically by a grid of κ values or by other numerical methods.

4.3 Estimation of κ and ϕ from seismic data

A similar technique as the one in Section 3.3 may be used to estimate the scale parameter $\boldsymbol{\kappa}$ and the lateral correlation coefficient ϕ for the blocky inversion. The workflow in Figure 3.1 can be followed with the parameter κ instead of γ. The posterior distribution $P(\boldsymbol{m}|\boldsymbol{d})$ can be approximated with a normal distribution with mean at the MAP-solution and covariance matrix $\boldsymbol{R}^{-1}$. $\boldsymbol{R}$ for the blocky inversion is given in Equation 4.11. This is approximation is noted as $\widehat{P}_{\kappa,\phi}(\boldsymbol{m}|\boldsymbol{d})$. The prior distribution $P_{\kappa,\phi}(\boldsymbol{m})$ is given in Equation 4.4. The inversion is then run for a range of $\boldsymbol{\kappa}$ and ϕ values to choose the most probable pair of parameters by maximizing

$$P_{\kappa,\phi}(\boldsymbol{d}) \approx \frac{P(\boldsymbol{d}|\boldsymbol{m})P_{\kappa,\phi}(\boldsymbol{m})}{\widehat{P}_{\kappa,\phi}(\boldsymbol{m}|\boldsymbol{d})}_{|\boldsymbol{m}=\widehat{\boldsymbol{m}}_{MAP}}. \tag{4.15}$$

$P(\boldsymbol{d}|\boldsymbol{m})$ is needed since $\widehat{\boldsymbol{m}}_{MAP}$ depends on the κ and ϕ values used in the prior distribution, and $\boldsymbol{G}\widehat{\boldsymbol{m}}_{MAP}$ takes a role in $P(\boldsymbol{d}|\boldsymbol{m})$. As mentioned above, the approximation of $P(\boldsymbol{m}|\boldsymbol{d})$ is an asymptotic approximation, but in our case it is more unclear if a Gaussian approximation in the denominator gives an accurate approximation of the likelihood $P(\boldsymbol{d})$.

4.4 Importance sampling

Importance sampling is performed to focus on the correct regions when using Monte Carlo methods, i.e. to weight the most important simulations heavier before doing inference with the simulated realizations (Liu (2001)). The target property is

$$I = E(h(\boldsymbol{m})) = \int h(\boldsymbol{m})P(\boldsymbol{m}|\boldsymbol{d})d\boldsymbol{m},$$

where $P(\boldsymbol{m}|\boldsymbol{d})$ is the distribution function of $h(\boldsymbol{m})$.

The algorithm to determine the importance sampling weights and calculate an estimate of I, is

- Draw n realizations from a trial distribution $g(\cdot)$

- Calculate the importance weights

$$w^{(j)} = \frac{P(\boldsymbol{m}|\boldsymbol{d}^{(j)})}{g(\boldsymbol{m}^{(j)})}, \text{ for } j = 1, 2, ..., n$$

- Approximate I by

$$\widehat{I} = \frac{w^{(1)}h(\boldsymbol{m}^{(1)}) + ... + w^{(n)}h(\boldsymbol{m}^{(n)})}{w^{(1)} + ... + w^{(n)}} \tag{4.16}$$

To make the estimation error small, $g(\boldsymbol{m})$ should have as similar shape as possible to $h(\boldsymbol{m})P(\boldsymbol{m}|\boldsymbol{d})$. This method can be used to make estimates of the mean of the posterior by making realizations from the posterior, and then calculating the weights as shown in the algorithm above. For the blocky inversion case, the trial distribution is the assumed normal distribution for $\boldsymbol{m}$ given $\boldsymbol{d}$, $g(\boldsymbol{m}) = \widehat{P}(\boldsymbol{m}|\boldsymbol{d}) \sim N(\widehat{\boldsymbol{m}}_{MAP}, \boldsymbol{R}^{-1})$. For the Gaussian inversion, all the weights will be equal because the target distribution is Gaussian.

If $h(\boldsymbol{m}) = \boldsymbol{m}^2$, $\hat{I}$ will be an estimate of the variance of $\boldsymbol{m}$ from the definition of the variance. An estimate of the variance of the MAP solution from importance sampling is therefore

$$Var(\boldsymbol{m}|\boldsymbol{d}) \approx \frac{w^{(1)}[\boldsymbol{m}^{(1)} - E(\boldsymbol{m}|\boldsymbol{d})]^2 + ... + w^{(n)}[\boldsymbol{m}^{(n)} - E(\boldsymbol{m}|\boldsymbol{d})]^2}{w^{(1)} + ... + w^{(n)}}. \tag{4.17}$$

5 SYNTHETIC EXAMPLE

5.1 Data description

In order to check the inversion and estimation methods, a synthetic example was created. Figure 5.1 shows the created model for $\boldsymbol{v_p}$, $\boldsymbol{v_s}$ and $\boldsymbol{\rho}$. The model consists of 7 horizontal layers, where each layer is a Gaussian Random Field with a certain lateral and vertical variation. These layer boundaries are marked as black lines in Figure 5.1. The lateral range is related to the lateral correlation coefficient ϕ by the exponential autocorrelation function

$$\phi = \exp\left\{-\frac{h}{\epsilon}\right\},$$

where ϵ is the correlation range, and h is the distance. This means, as expected, that the correlation decreases with the distance. The correlation range, ϵ, controls how long the correlation has effect before it fades out. To generate the synthetic data in this case, the correlation range was $\epsilon = 30$ and the standard deviation was 0.04. A corresponding criterion was used in the vertical direction for each of the 7 layers, and the vertical range was set to 13.

A 30 Hz Ricker wavelet was used in the convolution matrix $\boldsymbol{W}$, and the Aki and Richards approximation in Equation 2.4 with four reflection angles from 10 to 40 degrees were used to set up the reflectivity matrix $\boldsymbol{A}$. The time vector consists of 250 samples ranging from 0 to 1 seconds with a 4 ms sampling interval. In order to create the seismic data, the models shown in Figure 5.1 were multiplied with the operator $\boldsymbol{G} = \boldsymbol{WAD}$. Subsequential, noise with standard deviation σ was added to make synthetic seismic data. Since the same operator $\boldsymbol{G}$ was used in the modelling of the data and in the inversion, the results are assumed to be better than for more realistic cases with real seismic data. In addition, the noise level σ is known from the modelling, which is very advantageous for the inversion. For real seismic data, processing of the data can cause big uncertainties. This synthetic case avoids this problem.

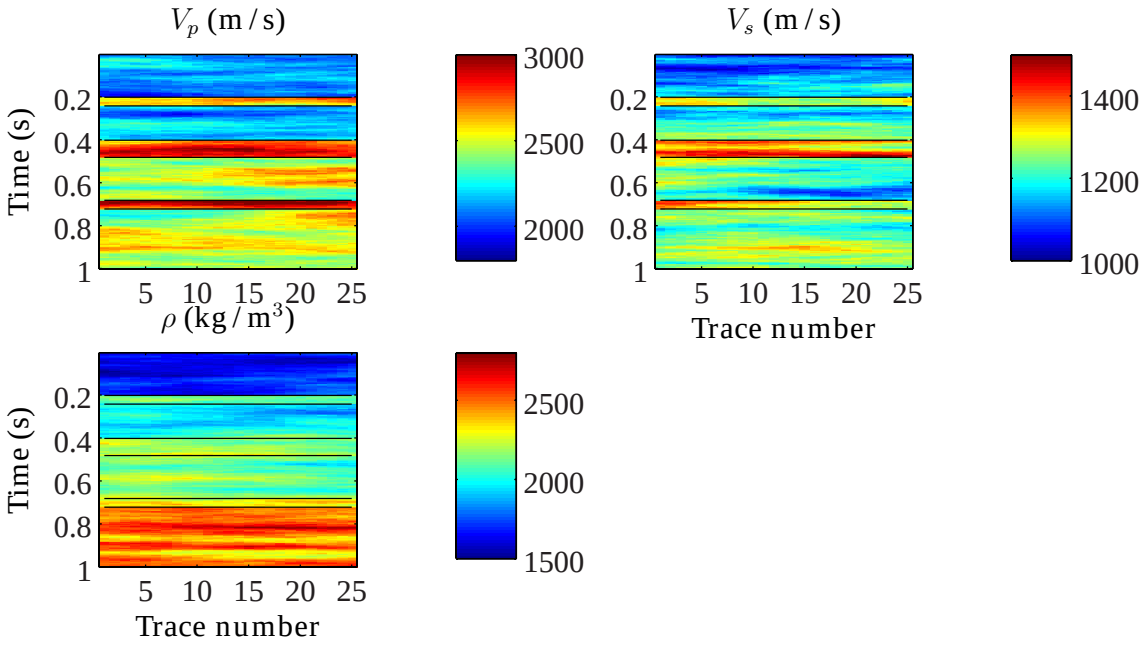

Figure 5.1: Synthetic model for $\boldsymbol{v_p}$, $\boldsymbol{v_s}$ and $\boldsymbol{\rho}$. Black horizontal lines indicate the boundaries between the 7 layers.

5.2 Results

5.2.1 Gaussian inversion

The Gaussian inversion was performed with the same $\boldsymbol{G}$ matrix and low frequency model $\boldsymbol{m}_p$ for all traces due to instability in the solution. The low frequency model $\boldsymbol{m}_p$ was created by calculating the mean of the lowpass-filtered input models for all traces. Several values of the lateral correlation coefficient ϕ and the scaling parameter γ for the covariance matrix Σ_0 were tested on a grid, and the estimation workflow described in Figure 3.1 was used. In the following results, the standard deviation of the noise is equal to 0.01. Figure 5.2 shows the synthetic seismic data for the last trace without noise in blue and random noise with standard deviation 0.01 in red for the four angles. The noise level is significantly smaller than the main amplitudes, and we assume the noise to be in a realistic area. Some tests showed that the estimation algorithm resulted in a ϕ value very close to the edge at $\phi = 1$. Since ϕ is limited to a lower bound 0 and upper bound 1, this was reparametrized to the correlation range

$$\epsilon = -\frac{h}{\ln(\phi)} \in (0, \infty),$$

in order to easier find the maximum close to 1. Figure 5.3 shows two visualizations of $const + \ln P(\boldsymbol{d})$ for a range of parameters, where $const$ is an unknown positive or negative constant. The maximum value for $P(\boldsymbol{d})$ corresponds to $\phi = 0.99$ and $\gamma = 0.005$ ($\ln(\gamma) = -5.2983$). It is not so easy to see from Figure 5.3, but the surface in the right panel dips downwards for the last two ϕ values close to 1 tested. The inversion result for the estimated parameters is shown in Figure 5.4, where the horizontal black lines indicate the original layer boundaries in the input model.

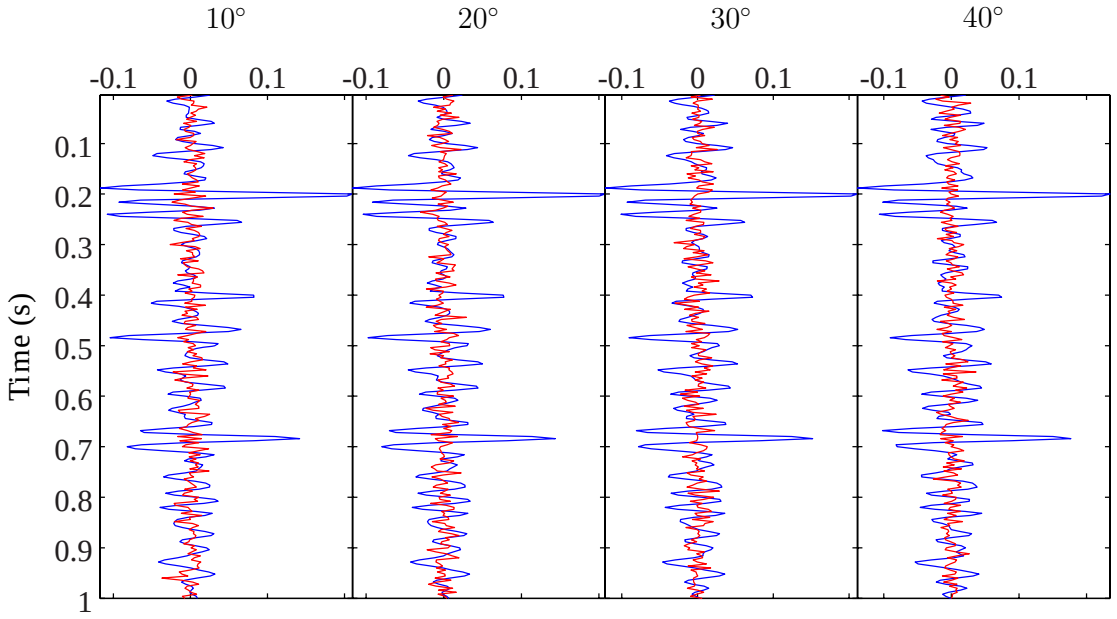

Figure 5.2: The created seismic data for the last trace and the four angles (blue). Random noise with standard deviation 0.01 in red.

A comparison between the input model in Figure 5.1 and the inversion results in Figure 5.4 shows that the Gaussian inversion with the estimated parameters reconstruct the input model reasonable well. The colour scale is the same in both figures, which means that the inversion returns values with the correct magnitude by using the estimated parameters. It seems as the inversion result is more continuous horizontally compared to the input model, but this could be due to the high correlation coefficient. Some of the details laterally from the input model are not recreated in the inversion result, and some values are smaller in magnitude, but the boundaries between the layers seem to show very clear after the inversion. Figure 5.5 shows the inversion result for the estimated $\gamma = 0.005$ and zero correlation. The inversion result with no lateral correlation seems to be smeared out at the layer boundaries. The input model for $\boldsymbol{v_p}$ in

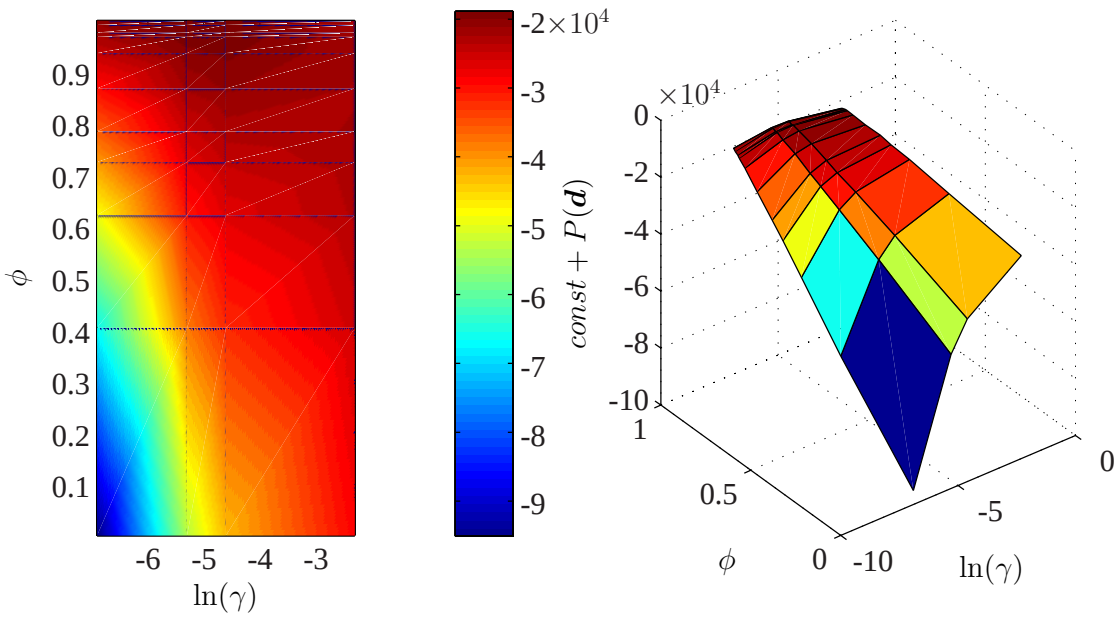

Figure 5.3: Two realizations of $const + \ln P(\boldsymbol{d})$ for the Gaussian inversion as a function of ϕ and $\ln(\gamma)$ on a grid.

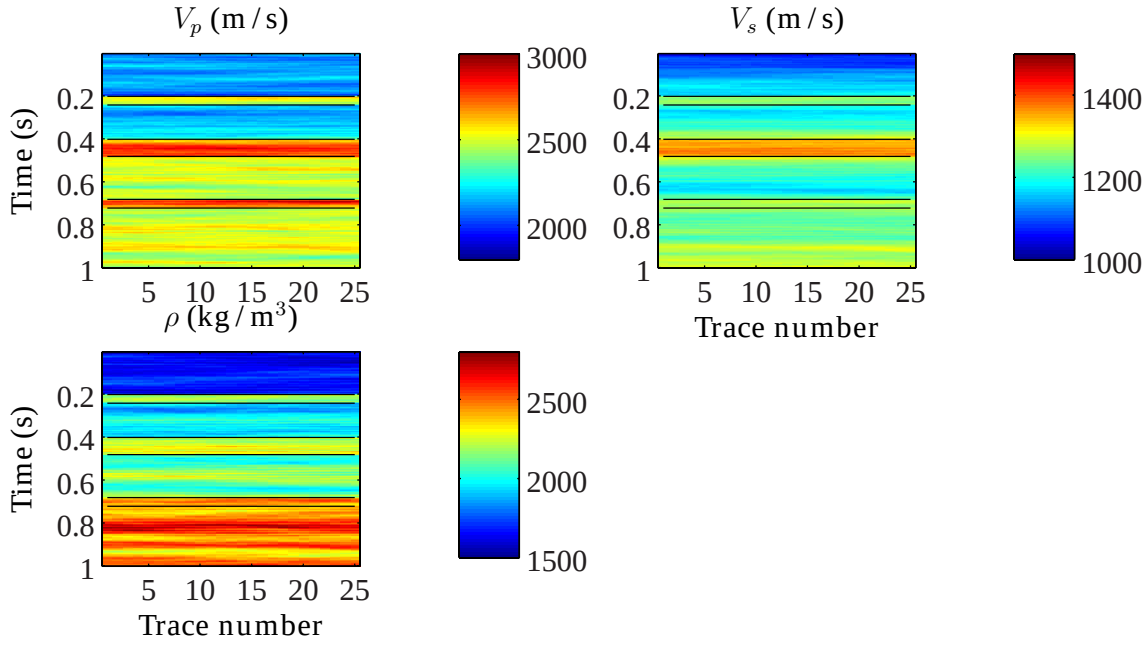

Figure 5.4: Inversion result from the Gaussian inversion with the estimated parameters, $\gamma = 0.005$ and $\phi = 0.99$.

Figure 5.1 shows two areas with higher velocities at about 0.8 s and 0.9 s traveltime. These events both appear in the inversion result with lateral correlation in Figure 5.4, but for zero lateral correlation in Figure 5.5, the event are hardly shown. In the input model, the event at 0.8 s is dipping, and at trace number 15, the two events have merged. The inversion with lateral correlation does not detect this merge. This can be one of the disadvantages by introducing lateral correlation; in some areas correlation is needed, whereas in others clearly not because of e.g. dipping events or faults. To check if the layer boundaries appear at the correct place, the result for the last trace is plotted in 1D. Figure 5.6, 5.7 and 5.8 show the inversion result for the last trace for $\boldsymbol{v_p}$, $\boldsymbol{v_s}$ and $\boldsymbol{\rho}$ respectively together with the input model and the low frequency background model for the estimated parameters. These show that the input data is well repeated after the inversion and that the contrasts between the layers are well defined in the inversion results, especially for $\boldsymbol{v_p}$ and density. To visualize the difference between zero correlation and $\phi = 0.99$, the last trace for the inversion result with $\gamma = 0.005$ and $\phi = 0$ is shown in Figure 5.9. This shows clearly that the boundaries between the layers are smeared out compared to Figure 5.6.

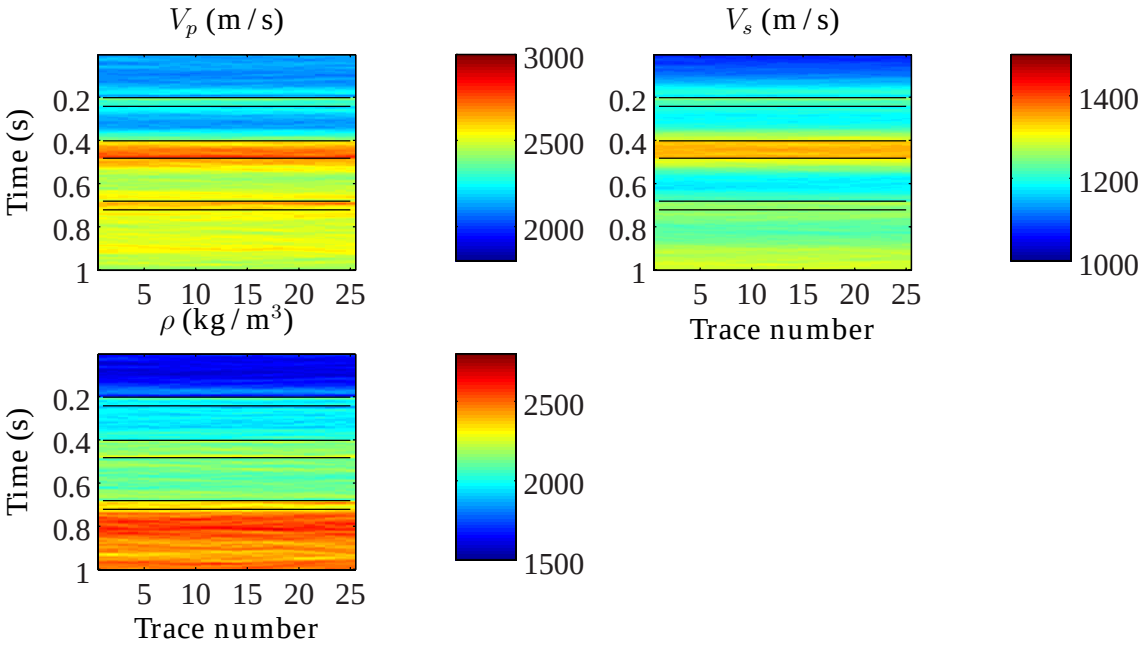

Figure 5.5: Inversion result from the Gaussian inversion with $\gamma = 0.005$ and $\phi = 0$.

The variance of the MAP solution for the Gaussian inversion is given in Equation 3.9. Figure 5.10 shows the inversion result for the last trace with the estimated parameters for the three elastic parameters in blue and the calculated

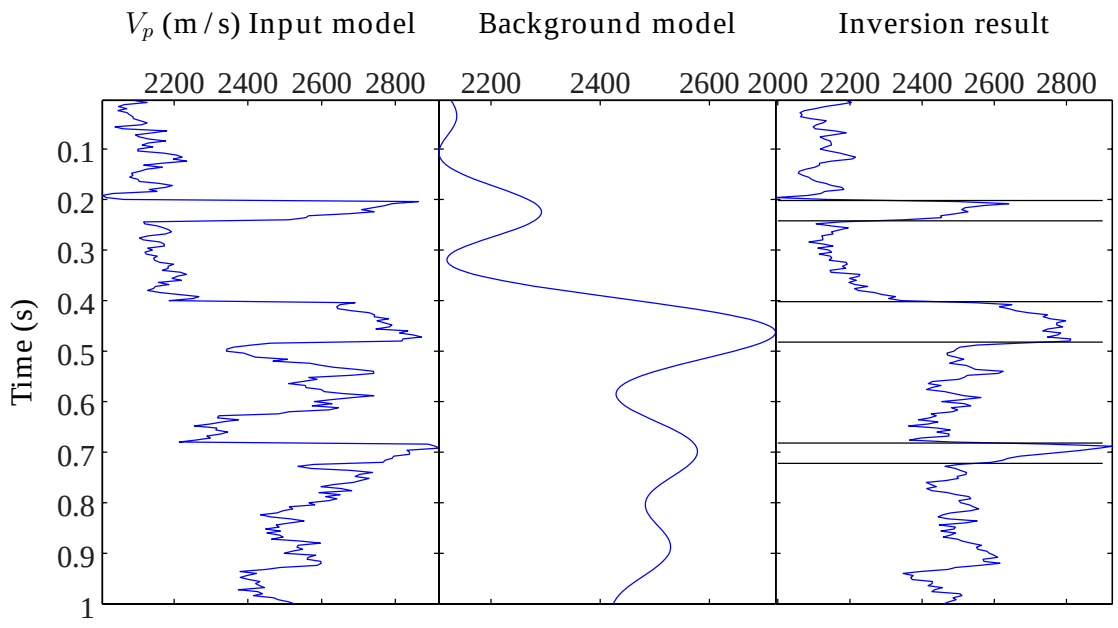

Figure 5.6: From left: Input model, low frequency background model, inversion result for $\boldsymbol{v_p}$ with $\gamma = 0.005$ and $\phi = 0.99$ for the last trace from the Gaussian inversion with the boundaries from the input model indicated by black lines.

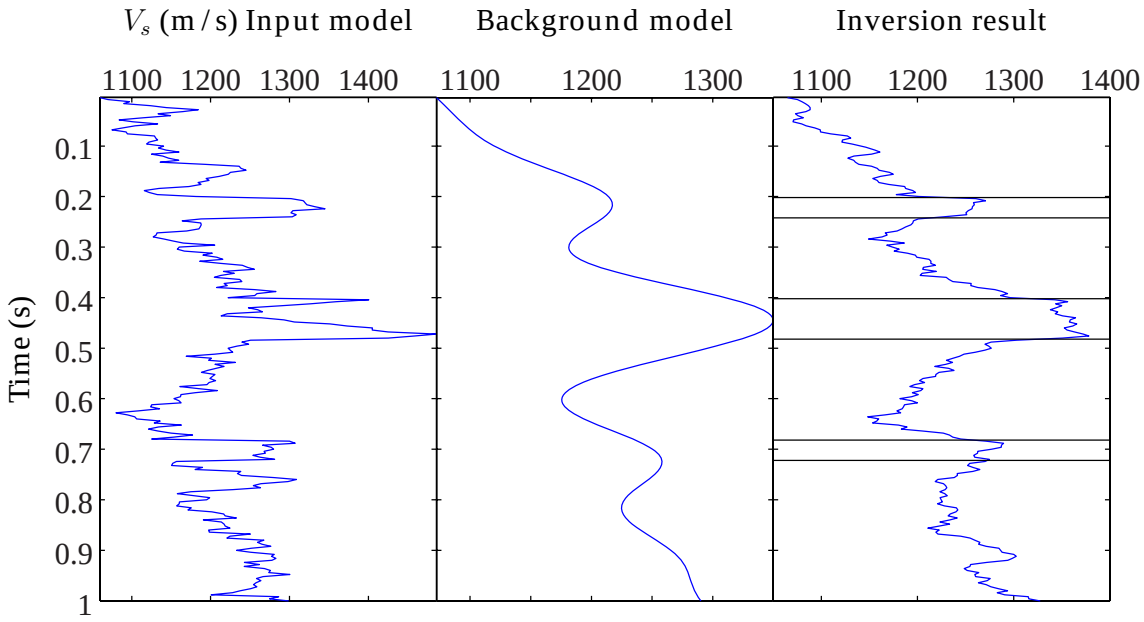

Figure 5.7: From left: Input model, low frequency background model, inversion result for $\boldsymbol{v_s}$ with $\gamma = 0.005$ and $\phi = 0.99$ for the last trace from the Gaussian inversion.

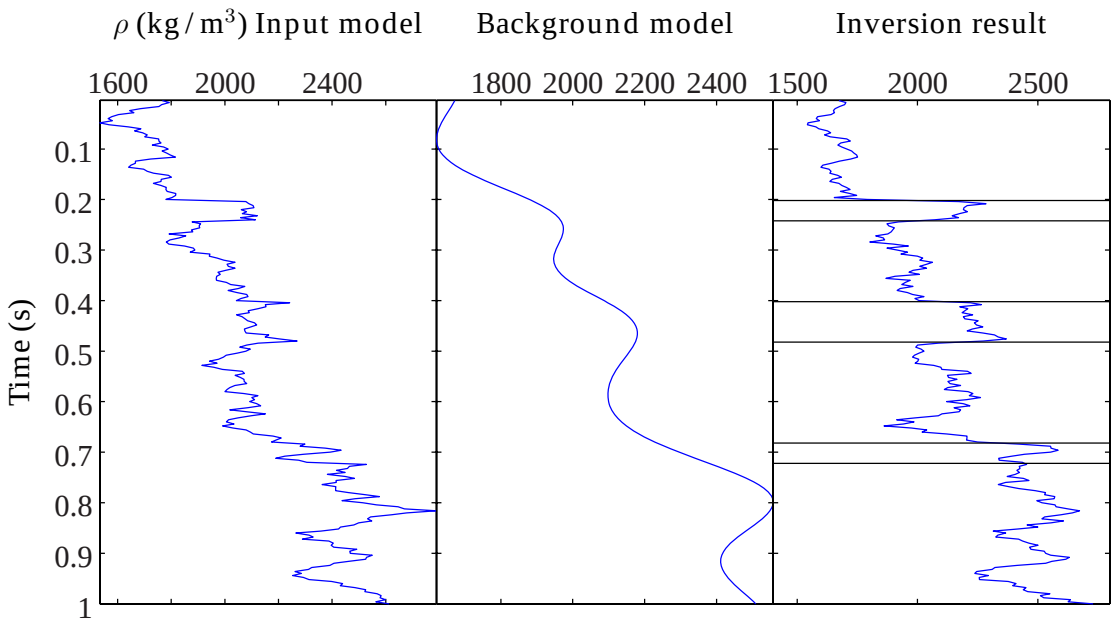

Figure 5.8: From left: Input model, low frequency background model, inversion result for $\boldsymbol{\rho}$ with $\gamma = 0.005$ and $\phi = 0.99$ for the last trace from the Gaussian inversion.

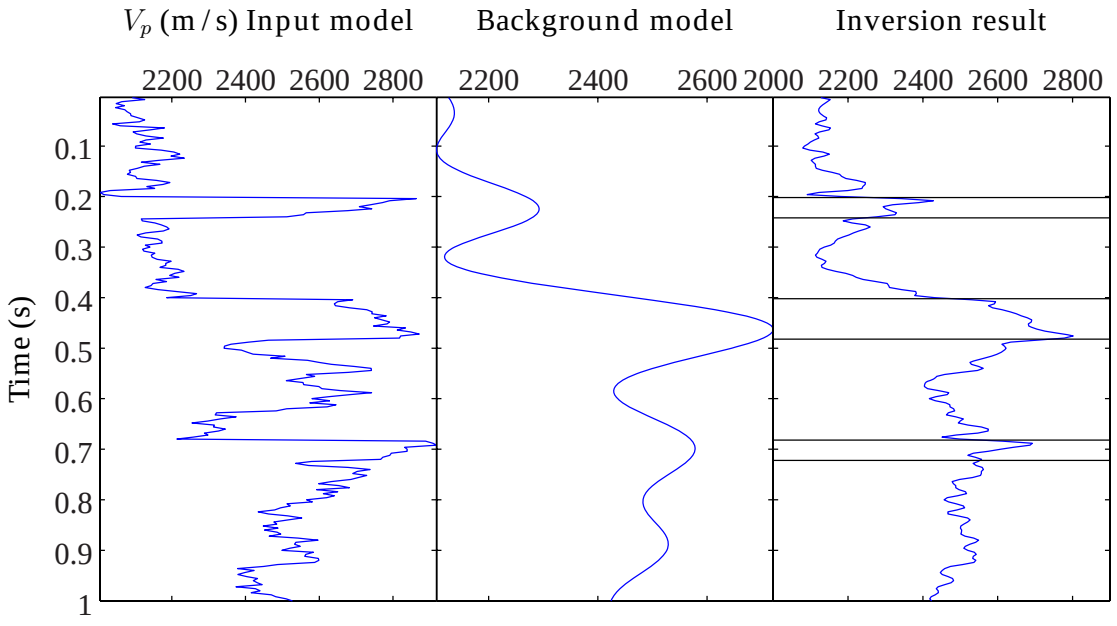

Figure 5.9: From left: Input model, low frequency background model, inversion result for $\boldsymbol{v_p}$ with $\gamma = 0.005$ and $\phi = 0$ for the last trace from the Gaussian inversion with the boundaries from the input model indicated by black lines.

95% prediction intervals for $\exp\{\boldsymbol{m}\}$ in red. The 95% prediction interval for $\widehat{\boldsymbol{m}}_{MAP}$ is given as

$$\widehat{\boldsymbol{m}}_{MAP} \pm z_{\frac{\alpha}{2}}\widehat{\sigma},$$

where $z_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ quantile of the Gaussian distribution, and $\widehat{\sigma}$ is the calculated standard deviation of the MAP solution by the square root of the expression in Equation 3.9. To get a 95% prediction interval, α is equal to 0.05 and $z_{\frac{\alpha}{2}} = 1.96$. The intervals seem reasonable, but maybe too small ranges between the lower and the upper confidence bounds of about 40 m / s for $\boldsymbol{v_p}$, 15 m / s for $\boldsymbol{v_s}$ and 60 kg / m^3 for $\boldsymbol{\rho}$. The different ranges for the three parameters are mainly because of the different ranges for the three parameters in the input model. The range of the S-wave velocity in the input model is about 500 m / s, while the range for the density is almost 1500 kg / m^3.

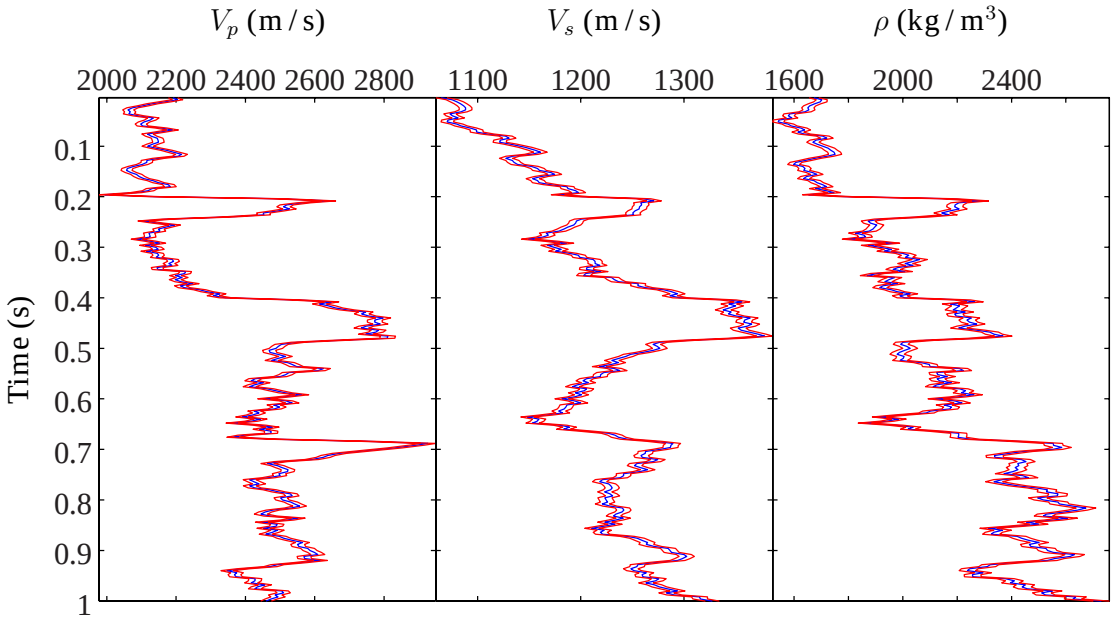

Figure 5.10: Inversion results from the last trace for the three parameters after Gaussian inversion with 95% prediction intervals included in red.

5.2.2 Blocky inversion

For the blocky inversion, different $\boldsymbol{G}$ matrices and low frequency models for each trace were used. The low frequency model was made by applying a low

pass filter of 3.4 Hz to the input model. A range of values for the lateral correlation coefficient ϕ and the scaling parameter κ were tested on a grid, using the estimation method described in Section 4.3. The noise level was unchanged from the Gaussian inversion, and is shown in Figure 5.2. Figure 5.11 shows two visualizations of $const + \ln P(\boldsymbol{d})$ as a function of ϕ and κ values, where $const$ is an unknown positive or negative constant.

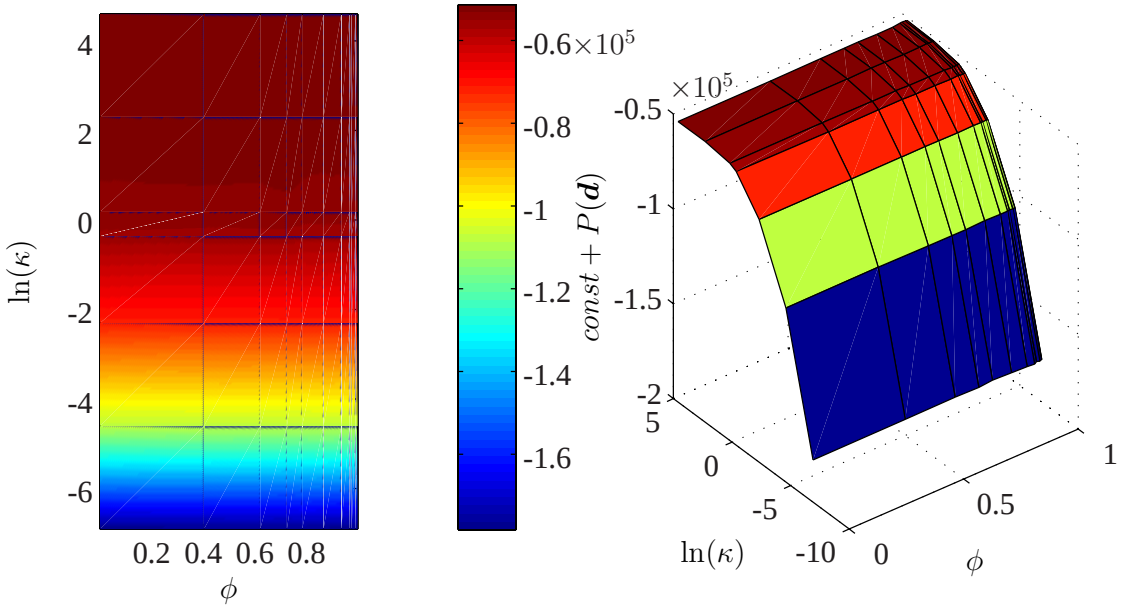

Figure 5.11: Two representations of $const + \ln P(\boldsymbol{d})$ for the blocky inversion as a function of ϕ and $\ln(\kappa)$ values on a grid.

It is easy to see from the left panel that $P(\boldsymbol{d})$ hardly change with ϕ compared to the change with κ. The maximum values from estimation were $\phi = 0.98$ and $\kappa = 100$ ($\ln(\kappa) = 4.6052$). Note that the chosen κ value is the largest value tested in this grid. This means that a local maximum is not found among the values tested. Several larger κ values were tested, although $P(\boldsymbol{d})$ continues to increase with κ. Since the surface in the right panel of Figure 5.11 seems to flatten out when κ increases, the last value on the grid, $\kappa = 100$ is considered as the estimated κ. The chosen ϕ value is not the largest value tested, but it is still very close to the limit 1. The inversion result for the estimated parameters with similar colour ranges as in Figure 5.1 is shown in Figure 5.12, and Figure 5.13 shows the inversion result for the last trace for $\boldsymbol{v_p}$ with the estimated parameters, $\phi = 0.98$ and $\kappa = 100$. In Figure 5.13, big and rapid oscillations are shown in the inversion result in the right panel. The boundaries between layers can almost not be seen. The values are completely out of acceptable range, and do

not correspond to the input data in the left panel.

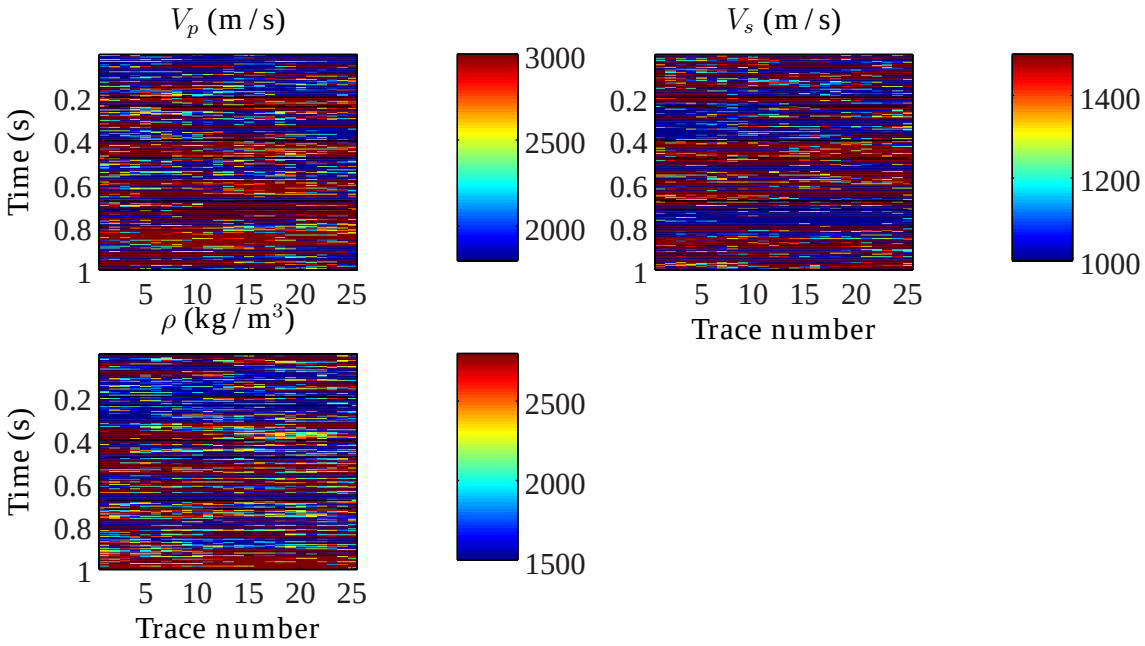

Figure 5.12: Inversion result for the blocky inversion with the estimated parameters, $\phi = 0.98$ and $\kappa = 100$.

Since the estimated parameters did not give satisfactory inversion results, a smaller value of the scale parameter κ was tested. As a first approximation, the fourth κ value was tested because this corresponds almost to area where the surface in Figure 5.11 flattens out. Figure 5.14 shows the inversion result for $\phi = 0.98$ and $\kappa = 0.7$ for $\boldsymbol{v_p}$ for the last trace. This result is closer to the magnitude of the input model, but still too large values and many oscillations lead to further research for better parameter choices. The method of trial and error with different parameter combinations resulted in the choice $\phi = 0.4$ and $\kappa = 0.04$. Figure 5.15 shows the inversion result for these chosen parameter values. The inversion result is obviously more likely due to comparison with the input model in Figure 5.1. The range of the values after inversion is similar to the input data, and the boundaries between the layers are clearly shown. Figure 5.16-5.18 show the inversion result for the last trace with $\phi = 0.4$ and $\kappa = 0.04$. The boundaries between the original layers in the input model are marked as black horizontal lines in the right panels. The inversion results for $\boldsymbol{v_p}$ and $\boldsymbol{\rho}$ are very similar to the input model. One exception are the rapid oscillations in $\boldsymbol{\rho}$ and some in $\boldsymbol{v_p}$ after inversion. These may arise because of the multimodal behaviour of the posterior by using the Cauchy distribution as the prior. Kolbjørnsen (2002a) also commented on these oscillations, and suggested

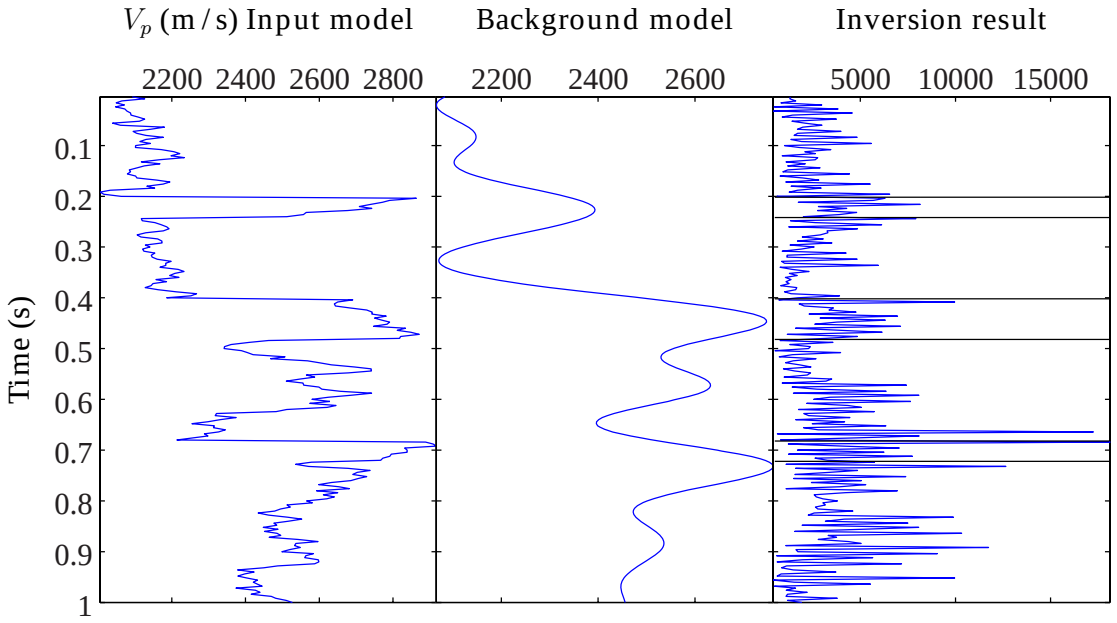

Figure 5.13: Inversion result for $\boldsymbol{v_p}$ for the last trace from the blocky inversion with the estimated parameters $\phi = 0.98$ and $\kappa = 100$.

that a completely flat model would probably be just as likely in the areas with the rapid oscillations. The main trend from the input model is though recreated very well after the inversion, but the boundaries between the adjacent layers do not show very clear for $\boldsymbol{v_s}$. Here, the inversion result is very similar to the low frequency background model, so the sharp contrasts between the layers are smeared out. The same observations were done for the Gaussian inversion, so $\boldsymbol{v_s}$ is probably the most difficult parameter to invert for.

The estimation of the parameters by the method described in Section 4.3 did not give reasonable results for the blocky inversion. One reason for this could be that the assumption about the posterior as a Gaussian distribution with mean at the MAP solution and variance $\boldsymbol{R}^{-1}$ from Equation 4.11, is not valid. Therefore, this approximation of the posterior distribution was checked by importance sampling described in Section 4.4. Importance sampling could also be used to provide an estimate of the uncertainties in the inversion results. $\boldsymbol{R}^{-1}$ for the blocky inversion is not so easy to determine as $\boldsymbol{R}^{-1}$ for the Gaussian inversion given in Equation 3.17, and after all, $\boldsymbol{R}^{-1}$ from Equation 4.11 is just an estimate of the variance of the MAP solution. The assumed posterior distribution, $N(\widehat{\boldsymbol{m}}_{MAP}, \boldsymbol{R}^{-1})$, was used as a trial distribution, and 100 realizations from this was simulated by the method described in Section 3.4. Figure 5.19 shows the estimated variance of the MAP solution with $\kappa = 0.04$ and $\phi = 0.4$ for the

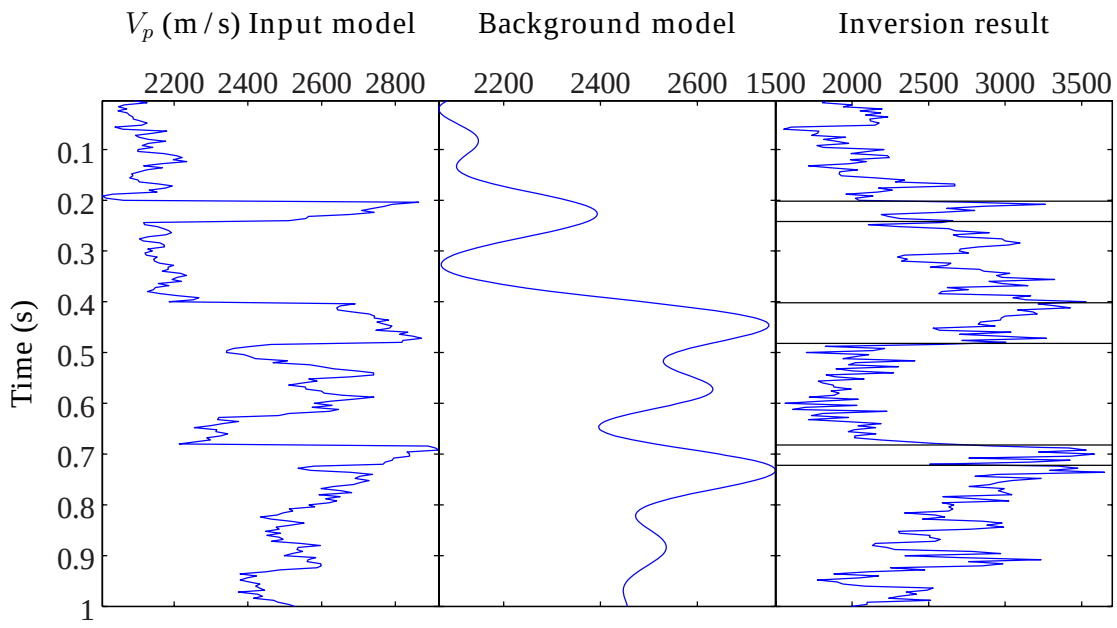

Figure 5.14: Inversion result for $\boldsymbol{v_p}$ for the last trace from the blocky inversion with $\kappa = 0.7$ and $\phi = 0.98$.

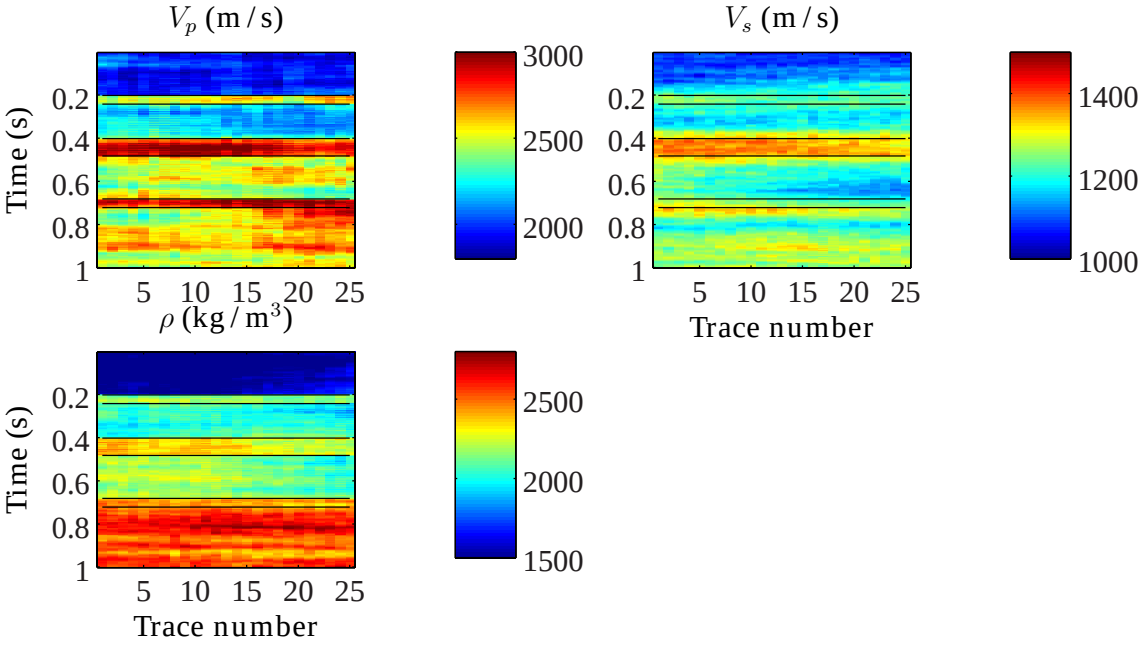

Figure 5.15: Inversion result for the blocky inversion with $\phi = 0.4$ and $\kappa = 0.04$.

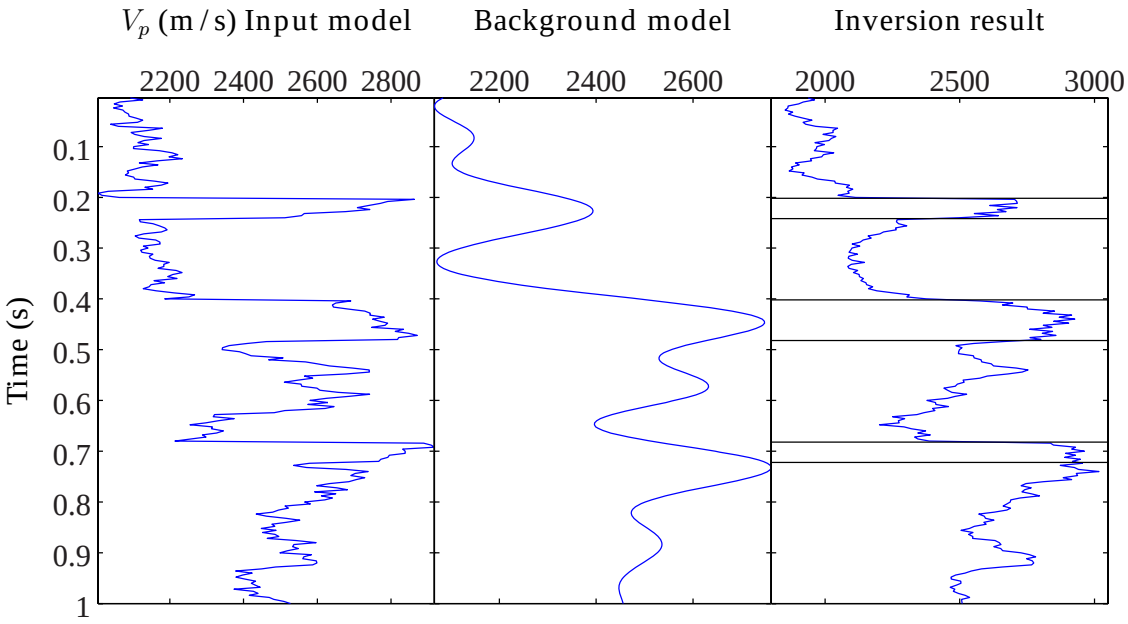

Figure 5.16: Inversion result for $\boldsymbol{v_p}$ the last trace from the blocky inversion with $\phi = 0.4$ and $\kappa = 0.04$.

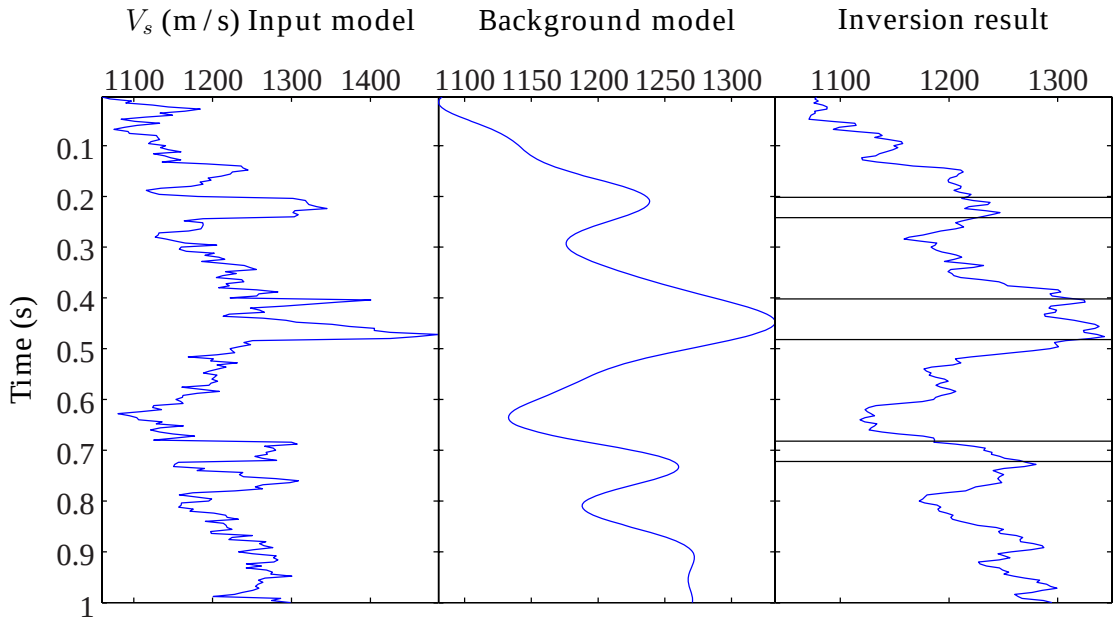

Figure 5.17: Inversion result for $\boldsymbol{v_s}$ the last trace from the blocky inversion with $\phi = 0.4$ and $\kappa = 0.04$.

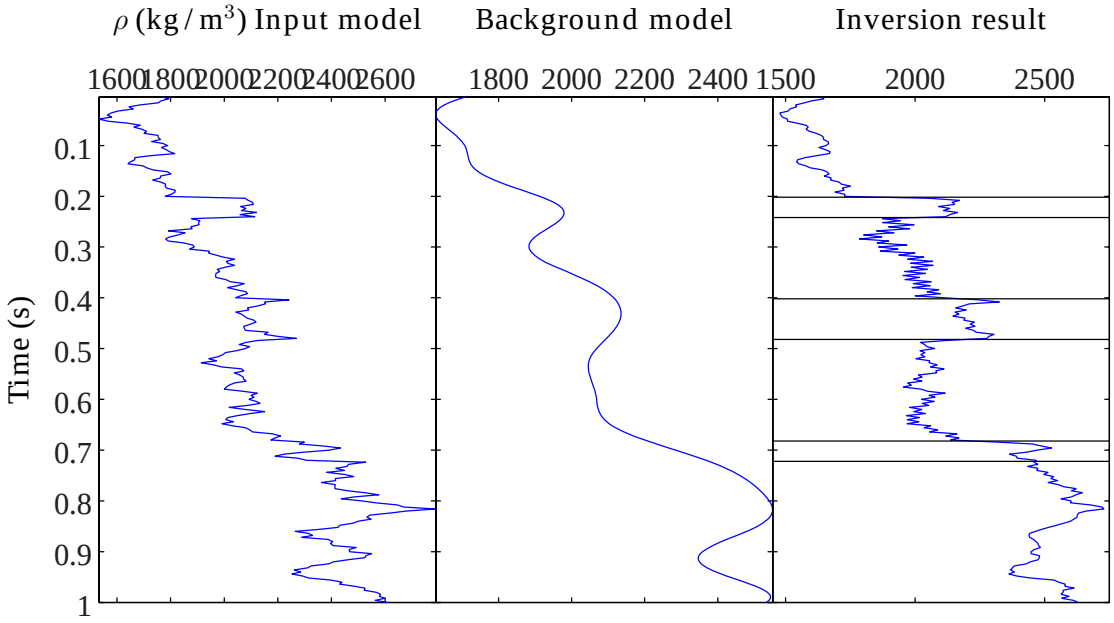

Figure 5.18: Inversion result for $\boldsymbol{\rho}$ the last trace from the blocky inversion with $\phi = 0.4$ and $\kappa = 0.04$.

last trace. This is valid under assumptions of a normal approximation for the posterior. The ranges from the lower to the upper bound of the prediction interval were around 1000 m / s for $\boldsymbol{v_p}$, 600 m / s for $\boldsymbol{v_s}$ and 1000 kg / m^3 for the density. These ranges were significantly larger then for the Gaussian case in Figure 5.10. Figure 5.20 shows the scaled importance weights multiplied with 100 calculated for blocky inversion with $\kappa = 0.04$ and lateral correlation $\phi = 0.4$. The weights are almost similar, which means that the normal distribution assumed for the posterior distribution, should be reliable. Figure 5.21 shows the expected model calculated by the importance weights from Figure 5.20 and Equation 4.16. Since the weights for the 100 simulated realizations in Figure 5.20 are quite similar, the expected value from importance sampling is not very different from the inversion result in Figure 5.15. Figure 5.22 shows the MAP solution for the three parameters for the last trace in red and 95% prediction intervals calculated by importance sampling included in cyan. The intervals are very wide, even wider than the intervals calculated directly from the simulated realizations in Figure 5.19. Importance sampling is possibly not capturing the heavy tails of the true posterior in the right way.

To compare the Gaussian and the blocky inversion schemes, the inversion results from the two methods were plotted together. This is shown in Figure 5.23. The scaling parameter γ in the Gaussian inversion was 0.005 and the lateral

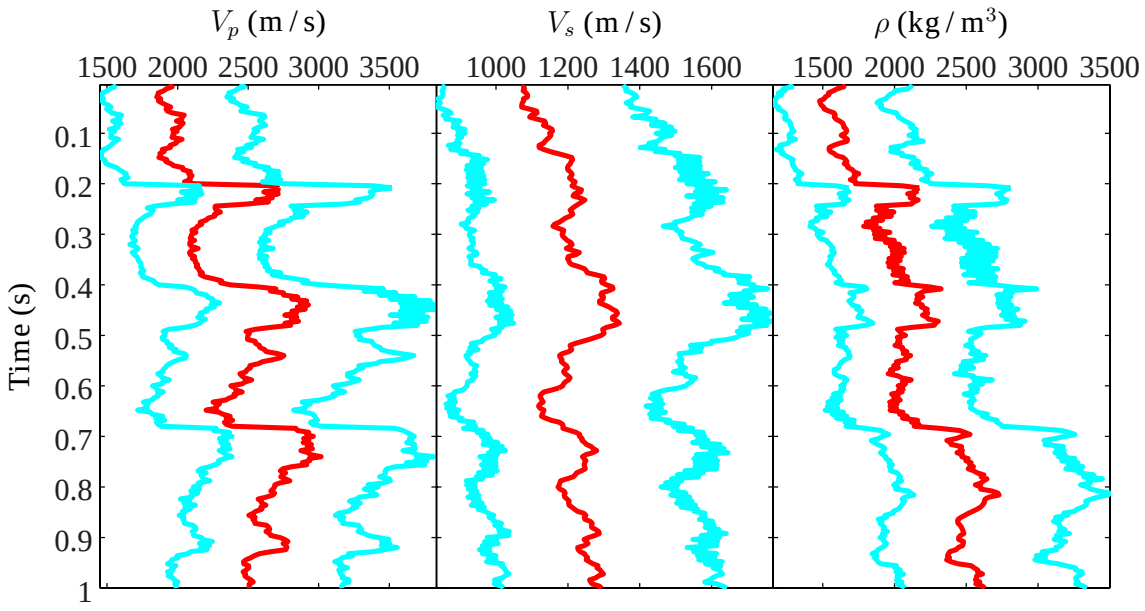

Figure 5.19: 95% prediction interval for the last trace in the synthetic model with $\kappa = 0.04$ and $\phi = 0.4$, calculated from simulated realizations.

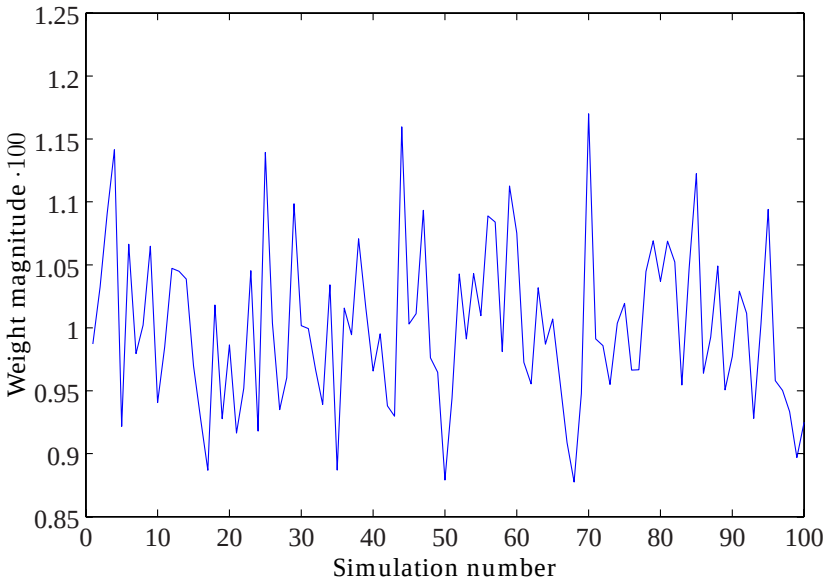

Figure 5.20: Importance weights calculated by importance sampling for the posterior distribution by blocky inversion with $\phi = 0.4$ and $\kappa = 0.04$.

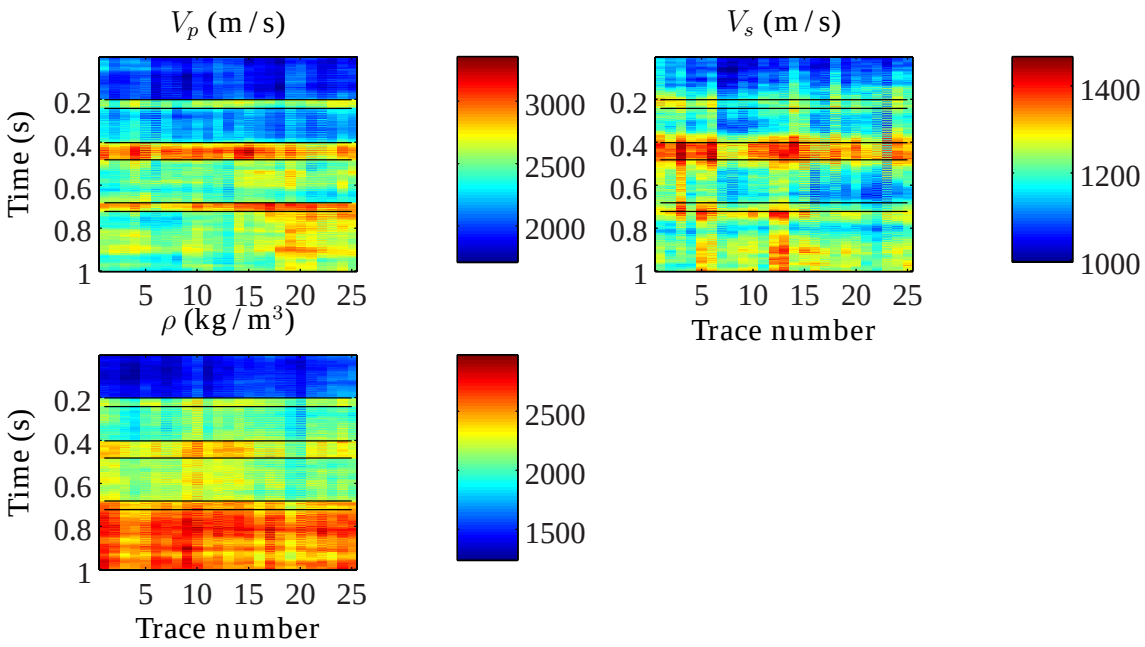

Figure 5.21: Expected inversion solution calculated by importance sampling for the blocky inversion with $\phi = 0.4$ and $\kappa = 0.04$.

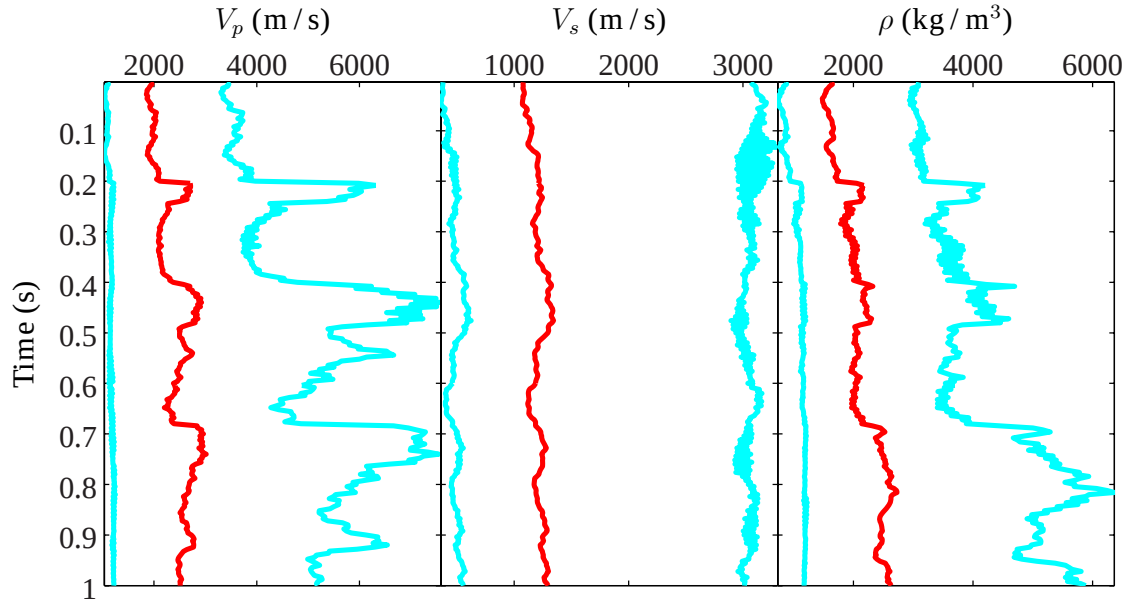

Figure 5.22: MAP solution with 95% prediction interval calculated by importance sampling for the last trace for the blocky inversion with $\phi = 0.4$ and $\kappa = 0.04$.

correlation was $\phi = 0.98$. For the blocky inversion, κ was 0.04 and the lateral correlation coefficient was 0.4. The green curve in Figure 5.23 is the created synthetic input model, the blue curve is the Gaussian inversion result and the red curve is the blocky inversion result. It is not easy to decide if the Gaussian or the blocky inversion is the best method from these results. In some areas of the inverted data, the blocky inversion result follows the input model better than the blue Gaussian curve, but especially for the density, the blocky inversion result contains rapid oscillations that are probably just artefacts due to a difficult posterior surface with Cauchy prior constraint. Such artefacts do not appear in the Gaussian inversion result. For $\boldsymbol{v_p}$ and $\boldsymbol{\rho}$, the contrasts between the layers are clearly reconstructed both for the Gaussian and the blocky inversion. Especially for the density, the blocky inversion result follows the input model very well, maybe even better than the Gaussian inversion if the rapid oscillations are ignored. Another criterion could be to detect the removal of side lobes. That is though almost impossible, because side lobes could be mistaken for actual changes in the properties.

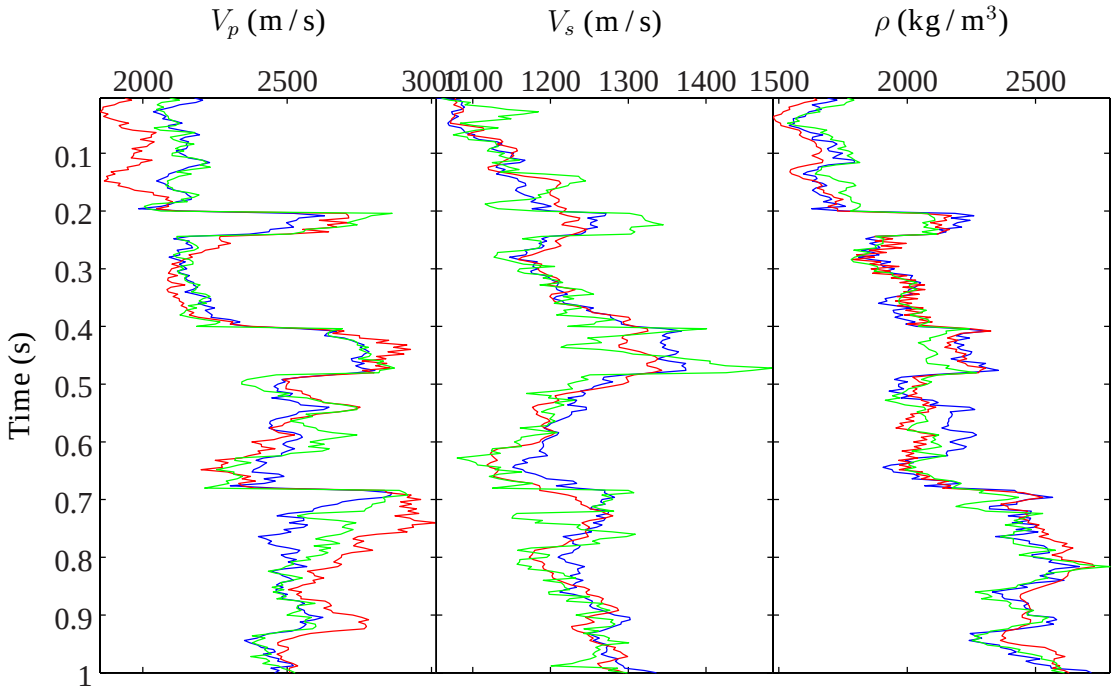

Figure 5.23: Comparison of Gaussian and blocky inversion results for the synthetic model. Input model (green), Gaussian inversion with $\gamma = 0.005$ and $\phi = 0.98$ (blue) and blocky inversion with $\kappa = 0.04$ and $\phi = 0.4$ (red).

6 KVITEBJØRN EXAMPLE

6.1 Data description

The Kvitebjørn field was discovered in 1994. It is located just east of the Gullfaks field, and contains hydrocarbons characterised as gas-condensate. Figure 6.1 shows where the field is located in the North Sea.

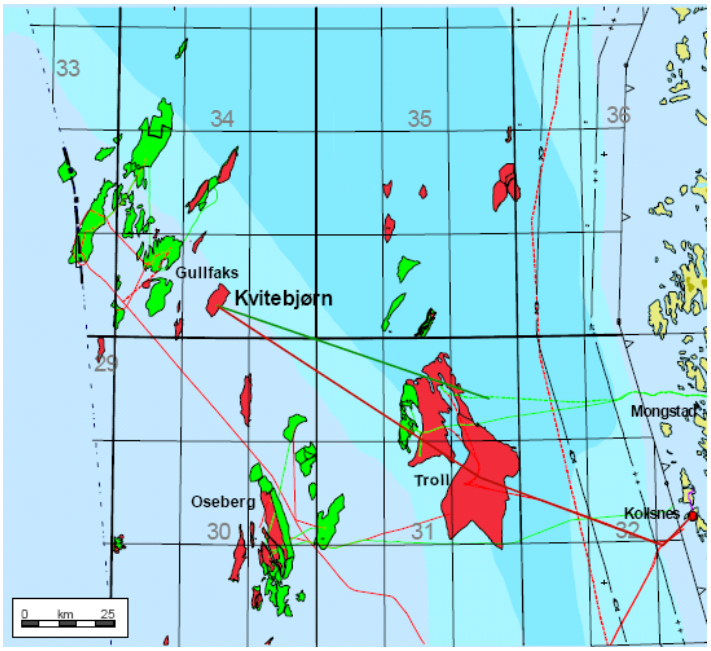

Figure 6.1: Location map of the Kvitebjørn field in the North Sea.

The field is a so-called HPHT (High Pressure High Temperature) field which is defined as: "Exploration and appraisal wells where the undisturbed bottom hole temperature at prospective reservoir depth or total depth is greater than $300°F$ ($149°C$) and, either the maximum anticipated pore pressure of any porous formation to be drilled through exceeds a hydrostatic gradient of 0.8 psi/ft (1.85 SG) or pressure control equipment with a rated working pressure in excess of 10000 psig (690 bar) is required (HPHT Well Control Manual, Ab-

erdeen Petroleum Training International)." Clay coating has resulted in good reservoir quality with relatively high average porosity. The drainage strategy is based on pressure depleting (reduction) which causes the drilling to be complex.

After a 3D survey from 2002, seven horizons were mapped in two-way time. These were Seabed, Balder Formation, Svarte Formation, Base Cretaceous Unconformity (BCU), Heather Formation, Brent Group and Etive Formation. Dunlin Group and Statfjord Formation are poor reflections, and therefore the interpretation of them is very uncertain.

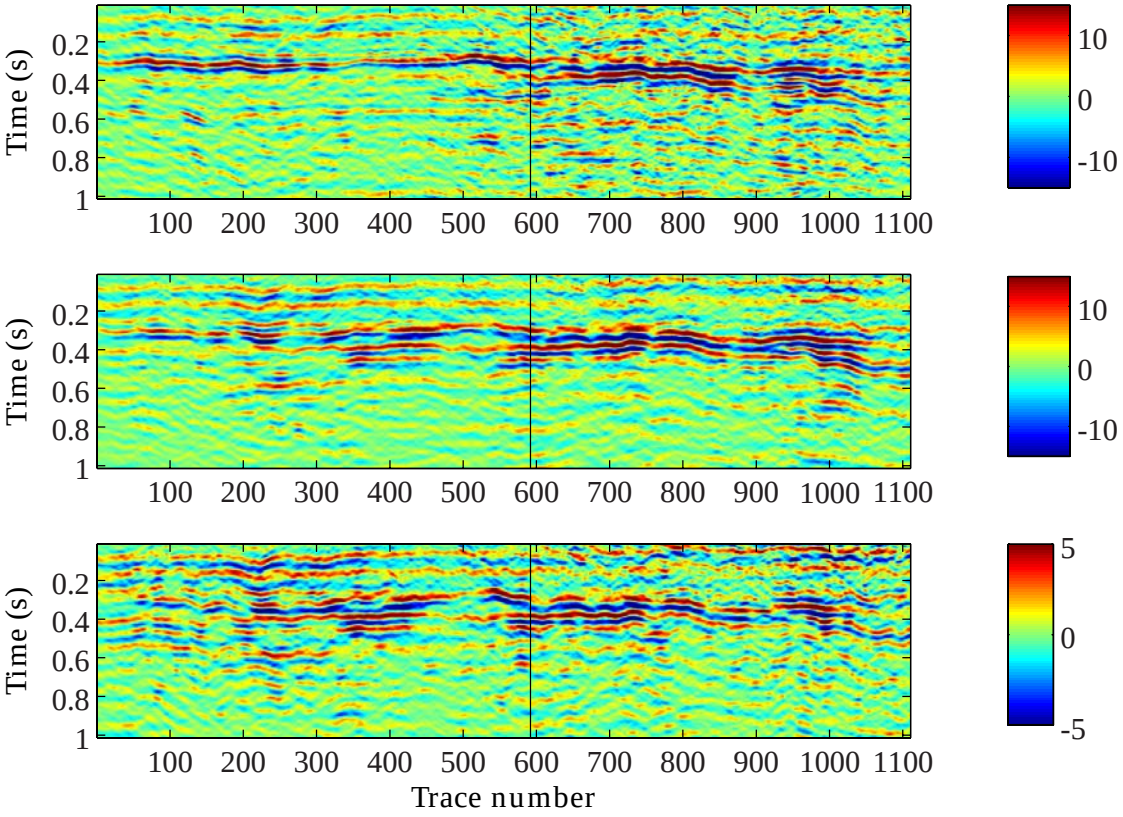

Figure 6.2: Real seismic stacked data from a seismic line on the Kvitebjørn field. From top: near angle, mid angle and far angle. Five interpreted horizons are included. These are (from top) Svarte, BCU, Brent, Drake and Statfjord. The location of well A is indicated by the black vertical line.

In this study, real seismic data were used. Well log data from well A were used to create the low frequency background model, and to compare the inversion results. This well is located almost at the intersection between a seismic inline

and crossline. Figure 6.2 shows the real seismic data from this specific crossline with the interpreted horizons Svarte, BCU, Brent, Drake and Statfjord included. The traveltimes are not real, but at a correct range. Well A's location is marked by the black vertical line. The incidence angles are stacked together to near, mid and far angles.

The wavelets and the standard deviation of the data $\boldsymbol{C}_d$ used in the inversion were estimated from the seismic data. Estimating random noise from seismic data is a difficult, yet important step. In order to obtain a rough estimate of the noise level, an average of three neighbouring traces was calculated. The difference from this mean value was used as an estimate of the standard deviation of the data for each trace. However, this noise estimate resulted in noise values that were of the same magnitude as the seismic data. This fact could possibly be explained by dipping layers in the data set used in the noise estimation process. Such dipping layers would violate the assumption of flat reflections in the seismic data, which is a fundamental assumption in this approach. It was therefore decided to divide the noise level by two, although this value is open to debate. On the other hand, an accurate determination of the noise level is probably not very important within the framework of this project and therefore, this value can be considered as sufficiently appropriate.

The low frequency expected background model, $\boldsymbol{m_p}$, was extracted from the well log data by a low pass filter of 4 Hz. This model was stretched according to the interpreted horizons in order to fit the geology along the line.

6.2 Results

6.2.1 Gaussian inversion

For the case with the real data from Kvitebjørn, the covariance matrix Σ_0 from Equation 3.3 was estimated from well log data as described in Section 3.2, and this gave

$$\boldsymbol{\Sigma_0} = \begin{bmatrix} 0.0139 & 0.0181 & 0.0016 \\ 0.0181 & 0.0299 & 0.0008 \\ 0.0016 & 0.0008 & 0.0021 \end{bmatrix}.$$

Several values for the correlation coefficient ϕ were tested, but there were no big differences. Figure 6.3 shows the inversion result for the three elastic parame-

ters using Σ_0 from well logs and $\phi = 0.4$ for 26 traces around the borehole location. 95% prediction intervals were calculated by Equation 3.9. The inversion result for the trace nearest the borehole with the prediction intervals included, is shown in Figure 6.4. The borehole data is also included as black dots. For the last part of the time interval, there is no well log data available. Therefore, the low frequency background model is not very reliable in this region. The interval ranges are about $1000m/s$ for $\boldsymbol{v_p}$ and $\boldsymbol{v_s}$, and $400kg/m^3$ for $\boldsymbol{\rho}$. These ranges are, as expected, significantly larger than for the synthetic example due to several additional noise sources in real data. For comparison, the wavelets and the noise levels are not known from the modelling, and seismic processing could have contributed to further uncertainties. The inversion result with $\phi = 0.4$ and Σ_0 from well logs in Figure 6.3 shows values for the three parameters with reasonable magnitudes, and Figure 6.4 shows that the inversion result is in the correct area, due to the well observations from the current area. Most of the well observations are inside the limits of the 95% prediction intervals. Therefore, the whole seismic line with 1110 traces was inverted with this parameter combination. Figure 6.5 shows the inversion results for the Gaussian inversion with $\phi = 0.4$ and Σ_0 from well logs for $\boldsymbol{v_p}$, $\boldsymbol{v_s}$ and density for all the 1110 traces. It is not easy to decide whether this is a good inversion result, because the truth is not known as it is for the synthetic example. Although, compared to the seismic data in Figure 6.2, the result looks reasonable. The layered structure in the inversion result corresponds to the interpreted horizons from seismic data.

Figure 6.6 shows the inversion result for $\boldsymbol{v_p}$ from the Gaussian inversion with different values for the lateral correlation coefficient. There are no main differences between the results, but at some points the inversions with lateral correlation seem to show some events clearer. The second downmost of the five interpreted horizons from Figure 6.5 around time 0.6 s do not appear very clearly for $\phi = 0$ and $\phi = 0.4$. However, for $\phi = 0.8$ in the bottom panel, this layer show at least in the middle part of the line between trace number 450 and 650.

6.2.2 Blocky inversion

The estimation of the parameters in the blocky inversion, even for the synthetic example, appeared to be very difficult. The model itself might cause some problems because of the assumption of correlation only in the lateral direction. Anisotropic properties then are not taken into account. The wavelet and the estimation of the wavelet also often cause problems and uncertainties. A combination of parameters which gave good results was found simply by trial and error. Figure 6.7 shows the inversion result for the same 26 traces near the bore-

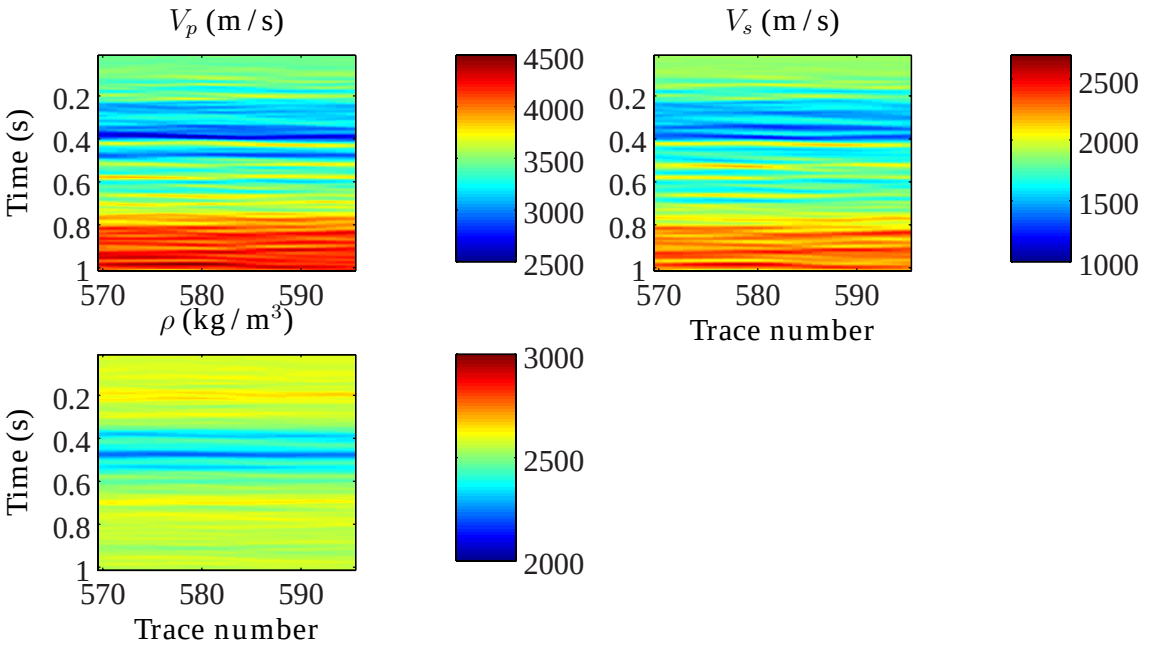

Figure 6.3: Inversion result for the Gaussian inversion of Kvitebjørn data with $\phi = 0.4$ and Σ_0 estimated from well logs.

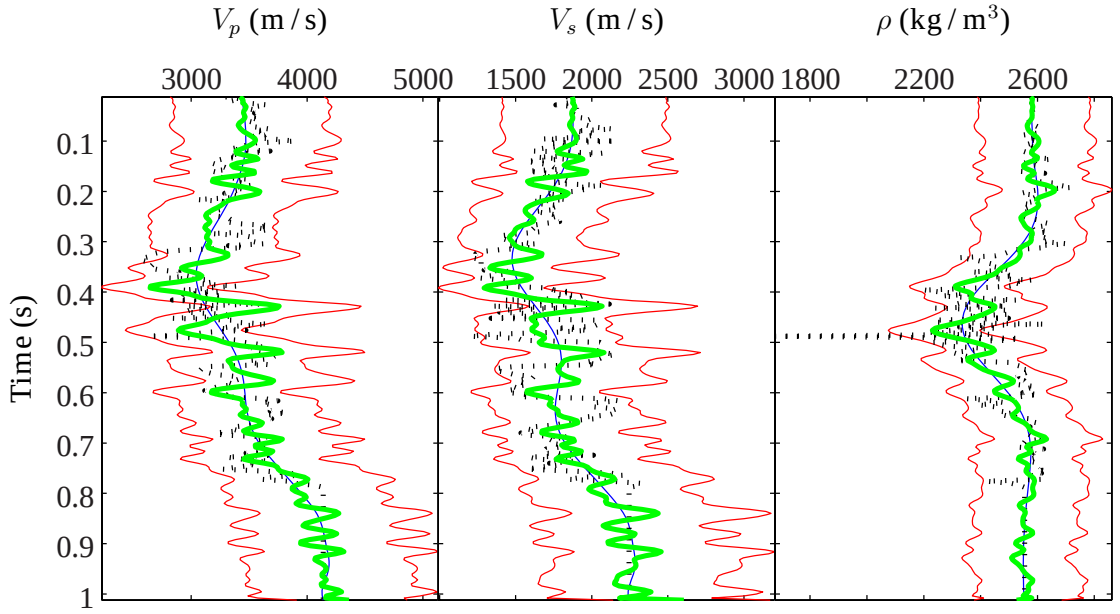

Figure 6.4: Gaussian inversion of Kvitebjørn data with $\phi = 0.4$ and Σ_0 estimated from well logs for the trace nearest the borehole location with 95% prediction intervals (red) and borehole data (black dots).

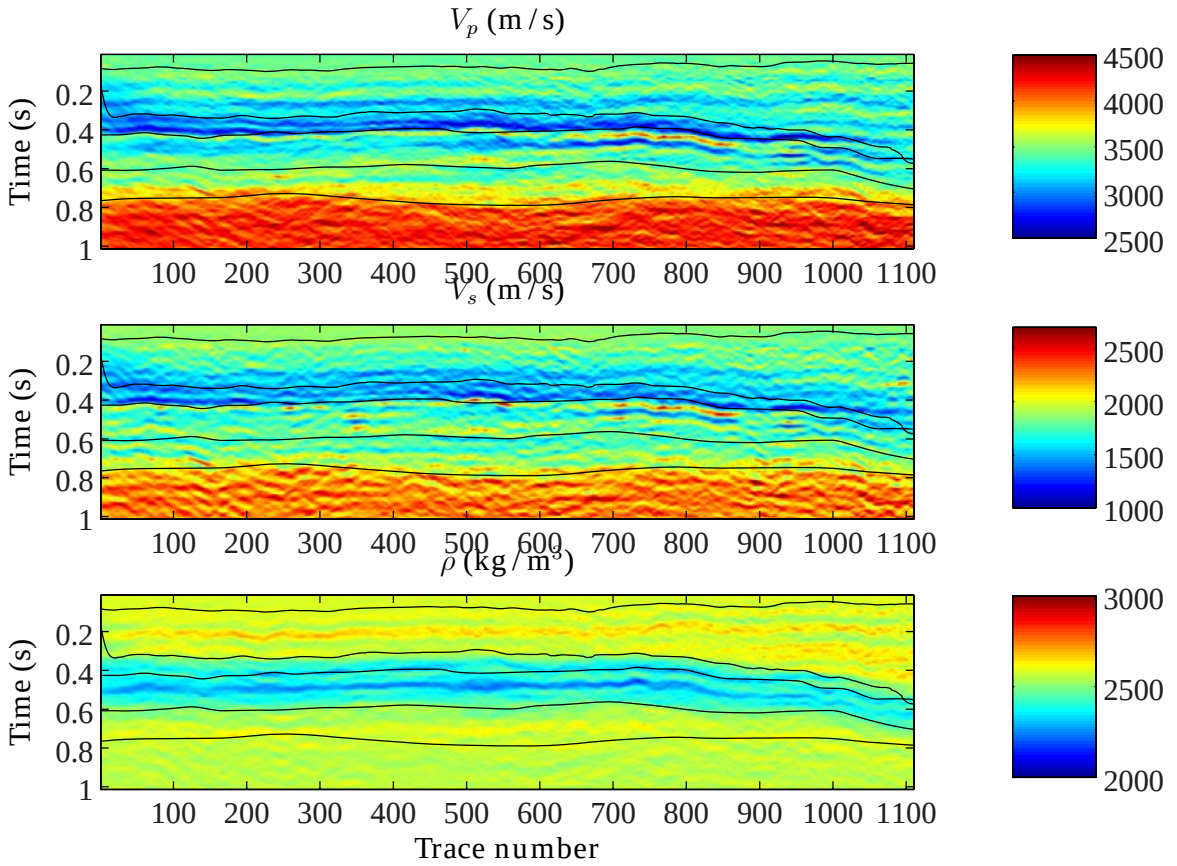

Figure 6.5: Gaussian inversion of Kvitebjørn data for the whole line with $\phi = 0.4$ and Σ_0 estimated from well logs.

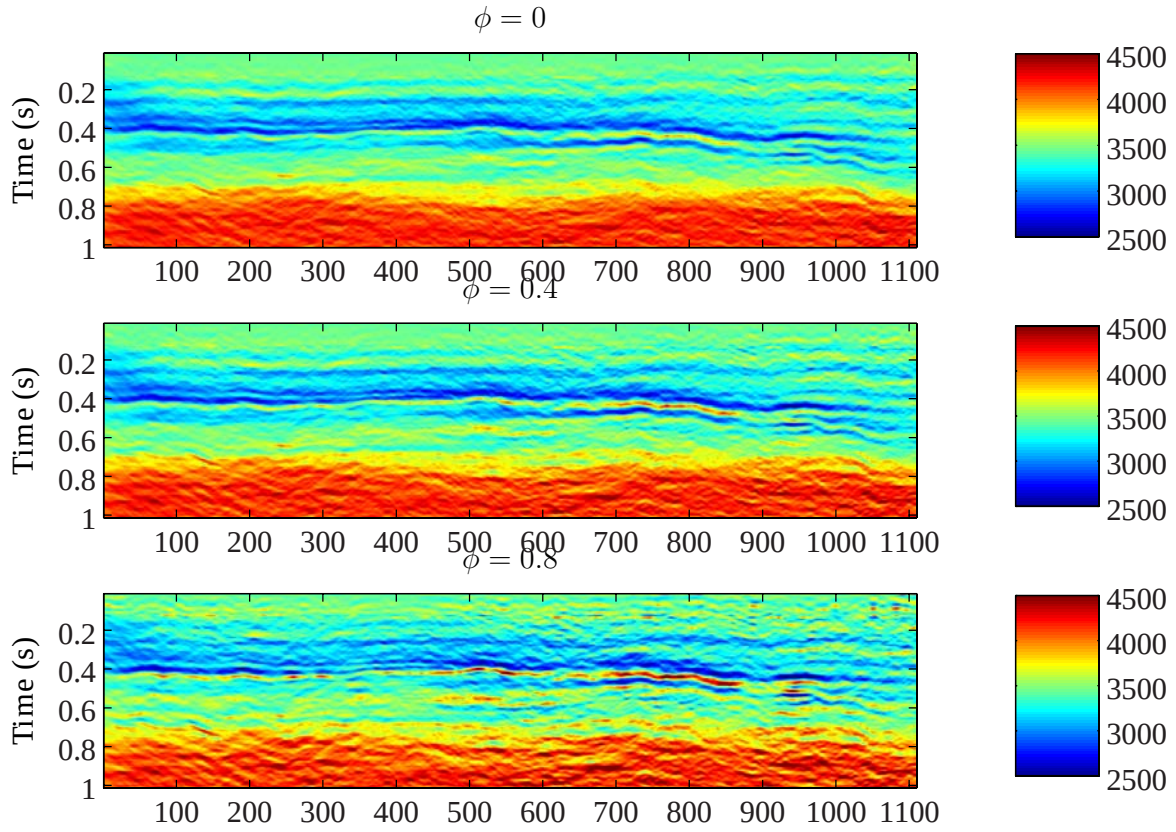

Figure 6.6: Comparison of the Gaussian inversion result for $\boldsymbol{v_p}$ for different ϕ values, from top: $\phi = 0$, $\phi = 0.4$ and $\phi = 0.8$.

hole location as in Figure 6.3 for $\kappa = 0.7$ and $\phi = 0.4$. This combination of parameters provides inversion results in the right order of magnitude, and the layered structure is comparable to the Gaussian result in Figure 6.3.

In order to provide an estimate of the uncertainties in the blocky inversion, importance sampling was performed for the MAP solution with $\kappa = 0.7$ and $\phi = 0.4$. Figure 6.8 shows an estimate of the variance of the MAP solution calculated from 100 simulated realizations from the approximate Gaussian posterior using the algorithm described in Section 3.4. These intervals are unrealistically large. Figure 6.9 shows the importance weights for 100 simulated realizations calculated by the algorithm described in Section 4.4. Figure 6.10 shows the expected values of $\boldsymbol{m}$ calculated by importance sampling and the weights from Figure 6.9. The weights are very similar for all the simulated realizations. This should mean that the posterior is not very different from a Gaussian distribution. Figure 6.11 shows the variance estimate for one of the inverted traces, shown as 95% prediction intervals around the MAP solution. The intervals are very wide, with ranges around 6000 for $\boldsymbol{v_p}$, 3500 for $\boldsymbol{v_s}$ and 4500 for $\boldsymbol{\rho}$. Both the intervals calculated from the simulated realizations and by importance sampling show unrealistically large values, and can not be used as a direct estimate for the uncertainties in the model.

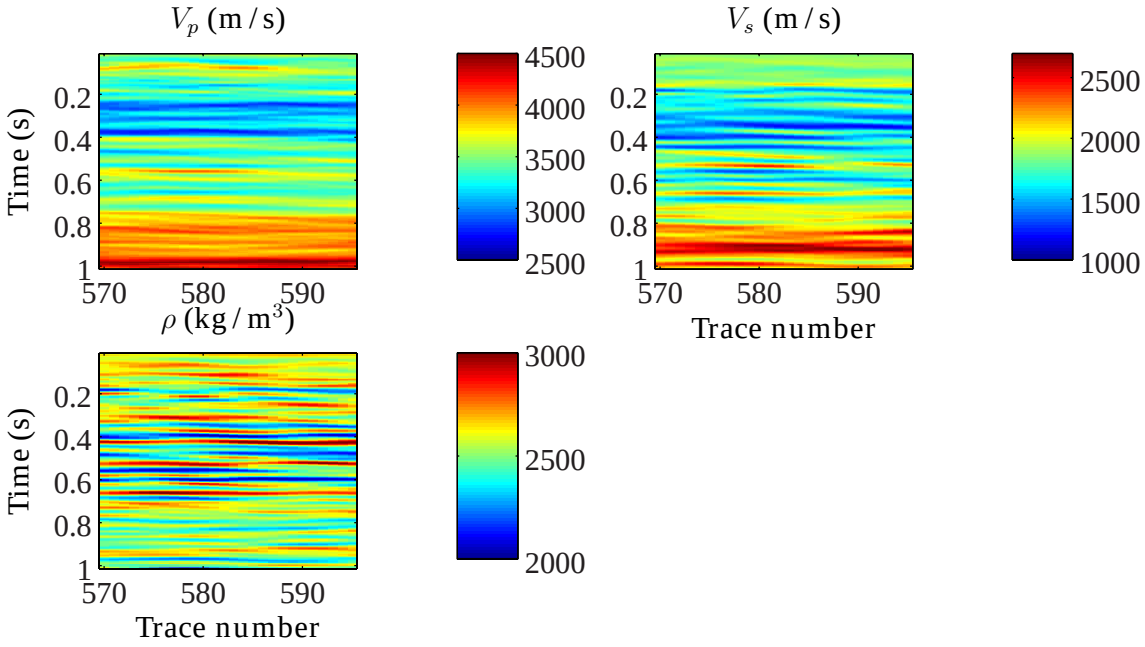

Figure 6.7: Inversion result from the blocky inversion for 26 traces around the borehole with $\kappa = 0.7$ and $\phi = 0.4$.

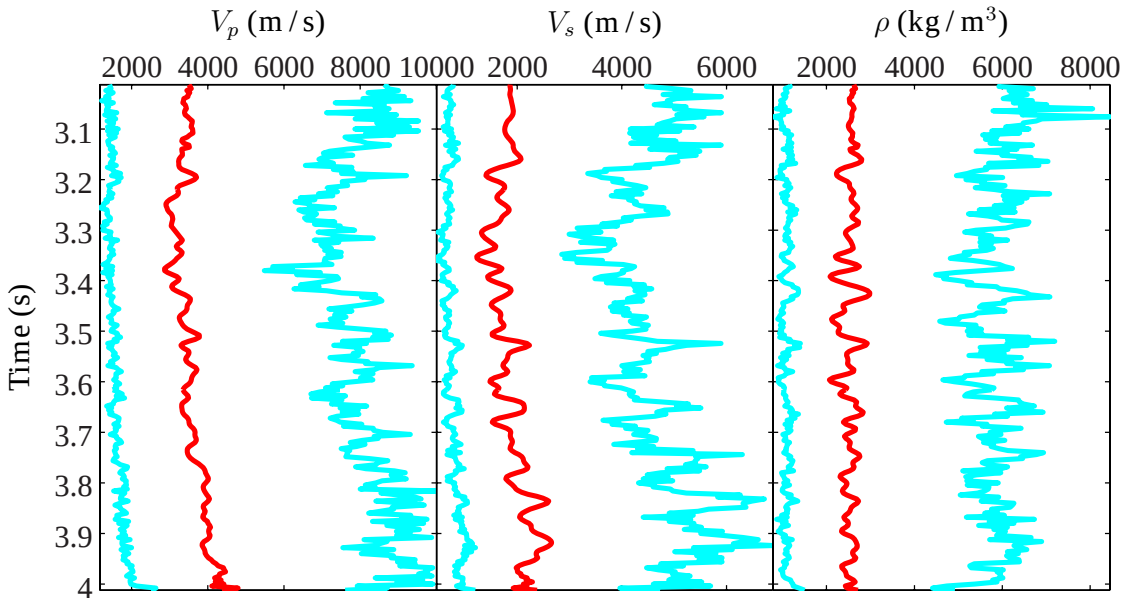

Figure 6.8: 95% prediction interval for the last trace used in the importance sampling with $\kappa = 0.7$ and $\phi = 0.4$, calculated from simulated realizations.

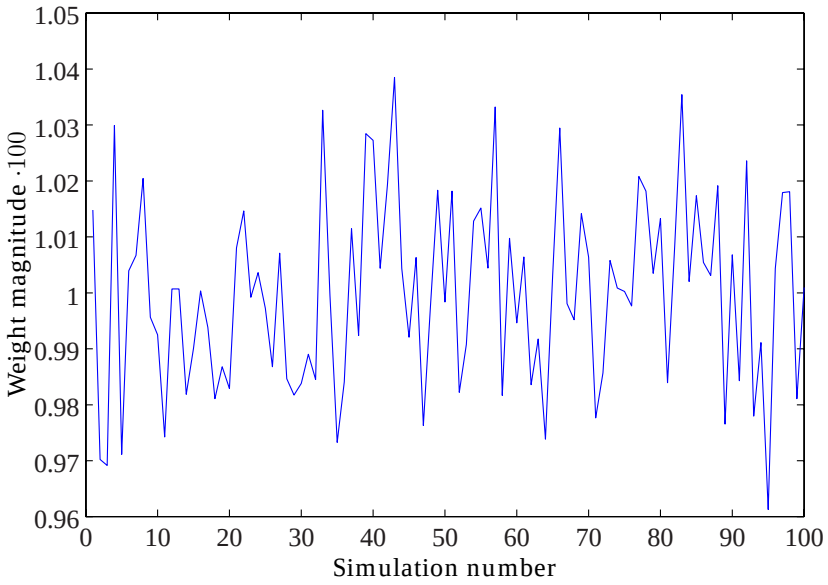

Figure 6.9: Importance weights for the MAP solution of the Kvitebjørn data calculated with $\kappa = 0.7$ and $\phi = 0.4$.

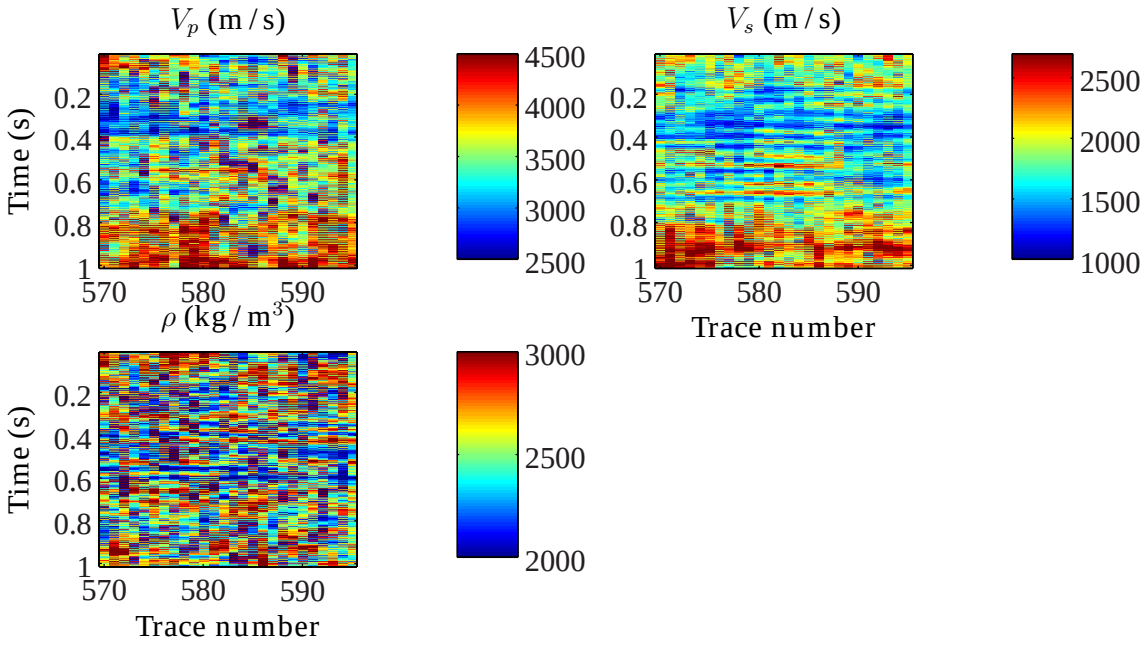

Figure 6.10: Expected $\boldsymbol{m}$ created by importance sampling and the weights from Figure 6.9.

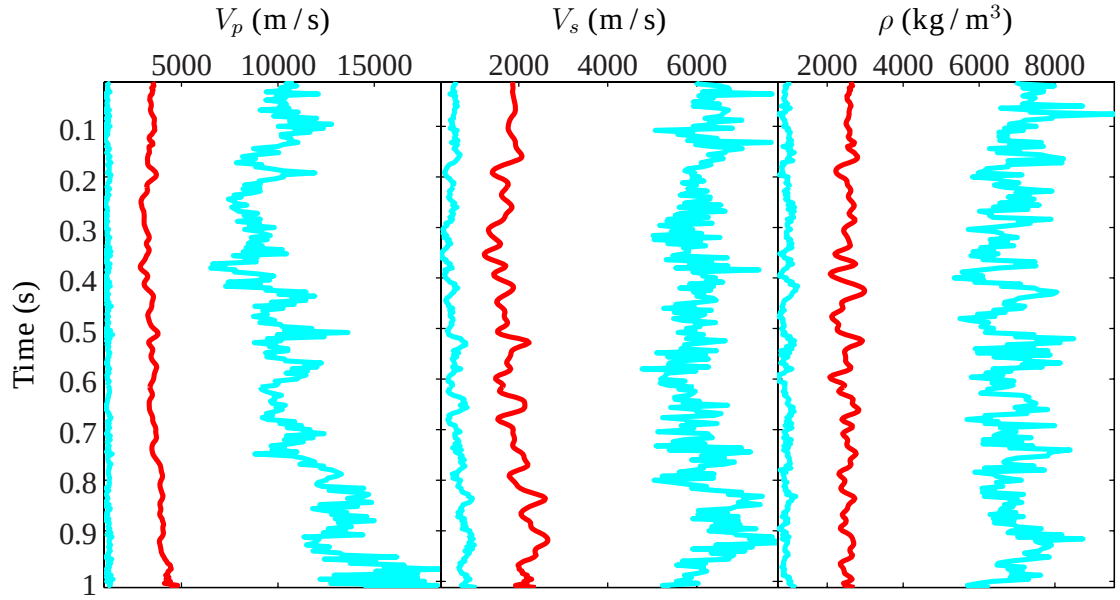

Figure 6.11: MAP solution from the blocky inversion for (from left) $\boldsymbol{v_p}$, $\boldsymbol{v_s}$ and $\boldsymbol{\rho}$ with $\kappa = 0.7$ and $\phi = 0.4$. 95% prediction intervals calculated from importance sampling are included in cyan.

Figures 6.12, 6.13, 6.14 and 6.15 show inversion results for $\boldsymbol{v_p}$, $\boldsymbol{v_s}$ and $\boldsymbol{\rho}$ from blocky inversion of the whole line (1110 traces) with different κ and ϕ values for comparison. Figure 6.12 shows the result by using the estimated κ values from maximum likelihood estimation of well log data and $\phi = 0.4$. The estimation method described in Section 4.2 was used with well log data rescaled to a similar range as the seismic data. The results from this estimation was

$$\begin{aligned} \kappa_{v_p} &= 0.0101 \\ \kappa_{v_p} &= 0.0146 \\ \kappa_{\rho} &= 0.0045. \end{aligned} \tag{6.1}$$

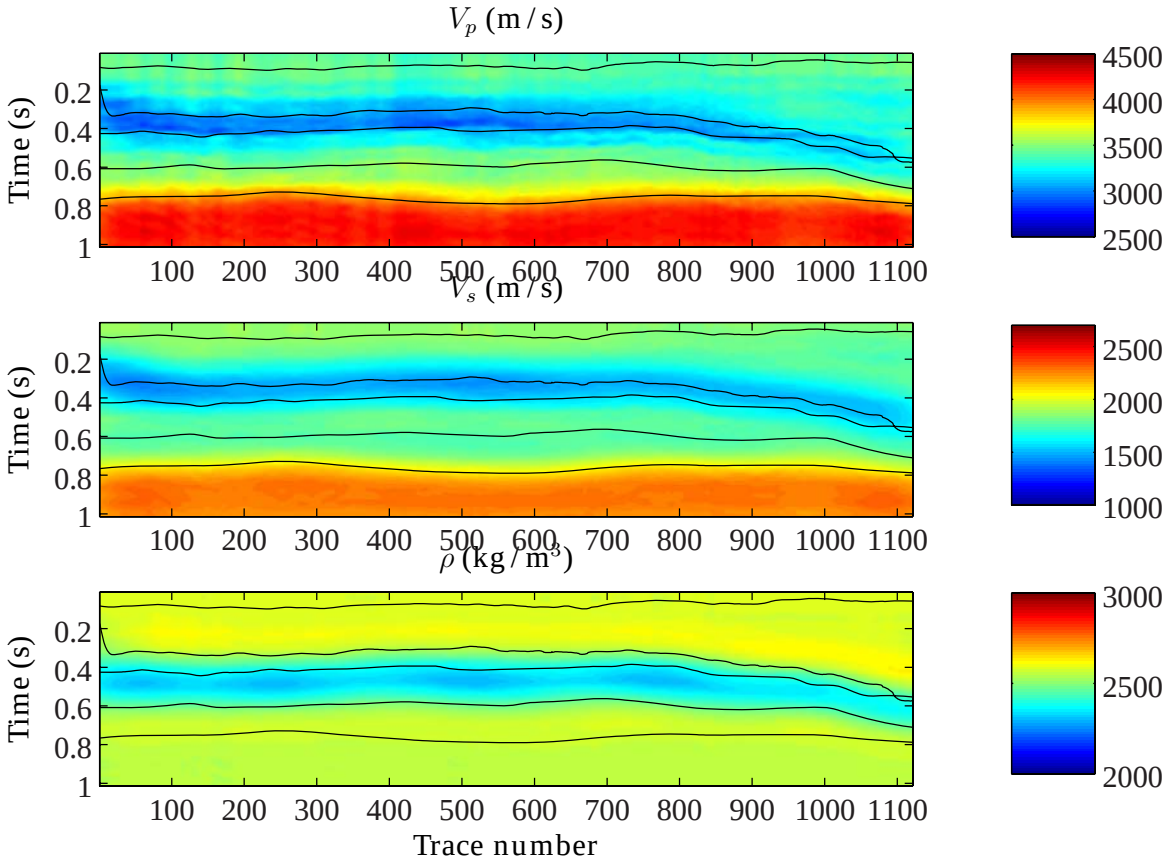

Figure 6.12: Inversion result with κ from MLE estimation and lateral correlation $\phi = 0.4$.

As mentioned before, too small κ values give almost just the low frequency background model after inversion. The κ's calculated from maximum likelihood estimation of well log data showed this behaviour. Therefore, larger κ

values were tested in order to get more structure out of the inversion. Figure 6.13, 6.14 and 6.15 show the blocky inversion result for $\kappa = 0.7$ and lateral correlation 0, 0.4 and 0.8 respectively. The κ values seems to be in the right order of magnitude since the structure from the inversion is not very different from the Gaussian inversion result. However, some more details seem to appear after blocky inversion. For example the interpreted horizon at about 0.6 s travel-time appear very clear after blocky inversion, both for $\boldsymbol{v_p}$ and $\boldsymbol{v_s}$. Compared to the Gaussian inversion in Figure 6.5, this horizon is much easier to detect after blocky inversion. Figure 6.16 shows a comparison between the inversion result for $\boldsymbol{v_p}$ for different κ values (from top: $\kappa = 0.3$, $\kappa = 0.7$ and $\kappa = 1.2$). All these values give reasonable inversion results and there are very small differences. It seems as if κ is in the right magnitude, it will not influence the inversion very much, but too small or too large κ values result in almost no structure and too large values respectively.

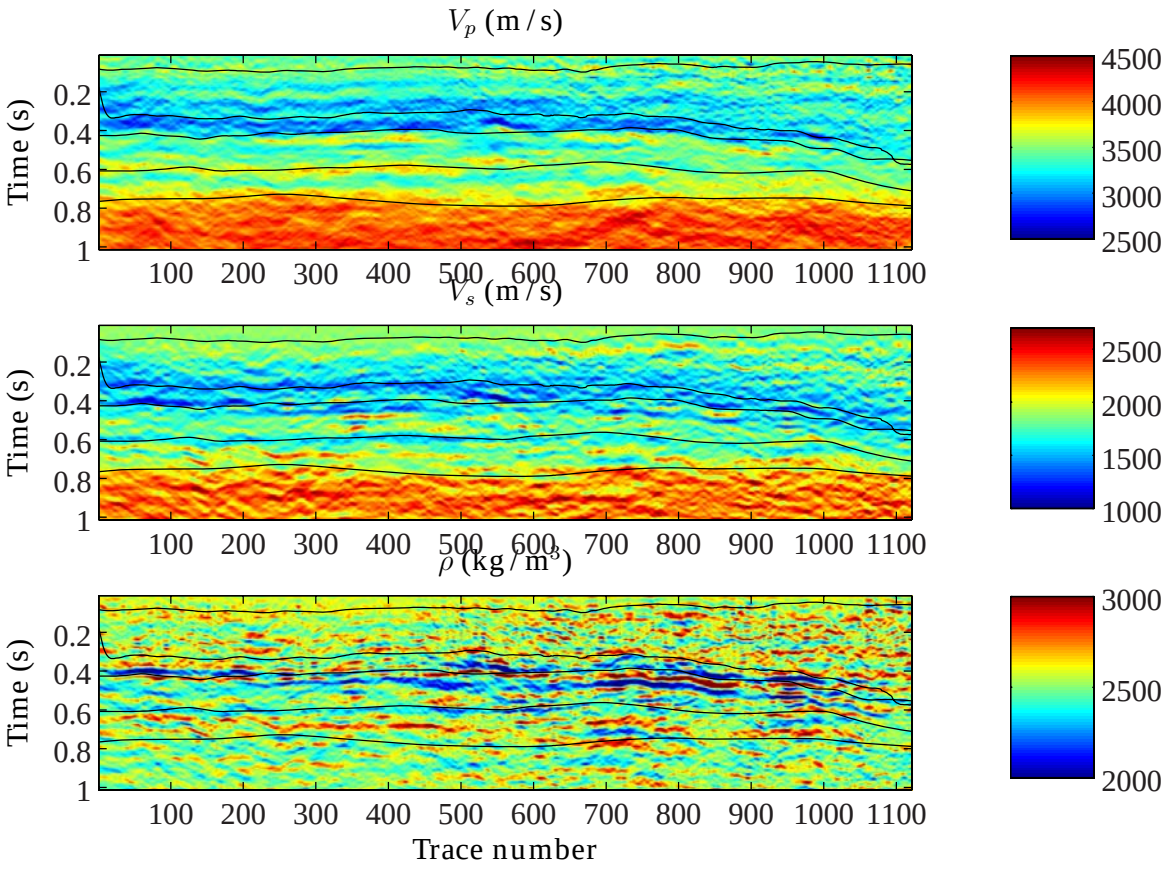

Figure 6.13: Blocky inversion result with $\kappa = 0.7$ and lateral correlation $\phi = 0$.

It is not very easy to conclude on which of the two Bayesian inversion methods

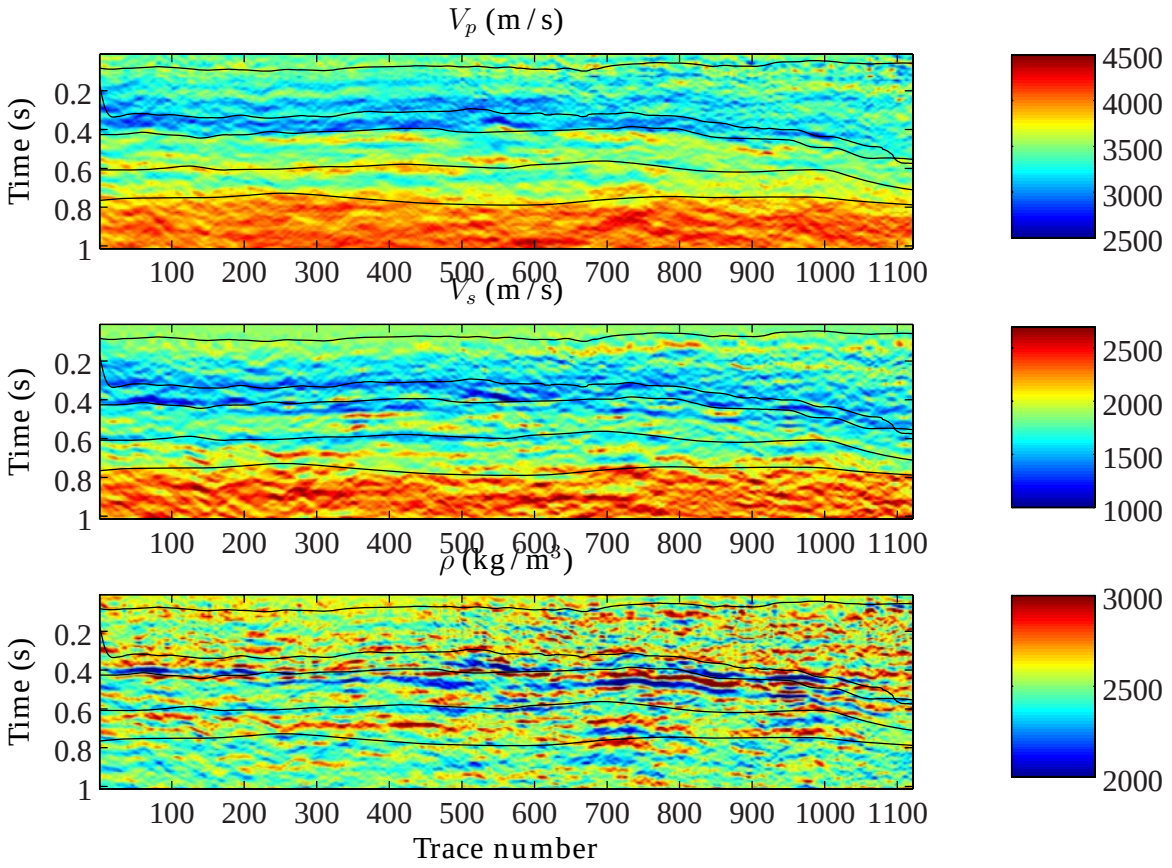

Figure 6.14: Blocky inversion result with $\kappa = 0.7$ and lateral correlation $\phi = 0.4$.

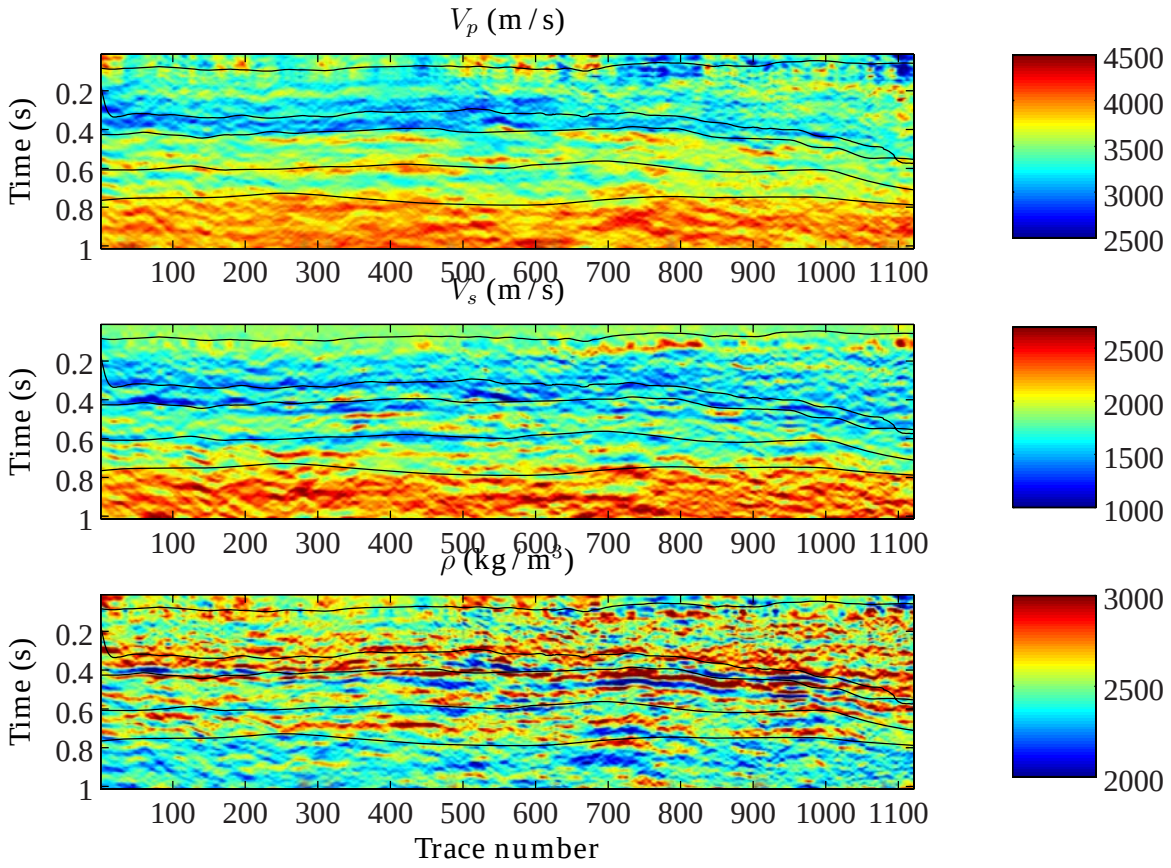

Figure 6.15: Blocky inversion result with $\kappa = 0.7$ and lateral correlation $\phi = 0.8$.

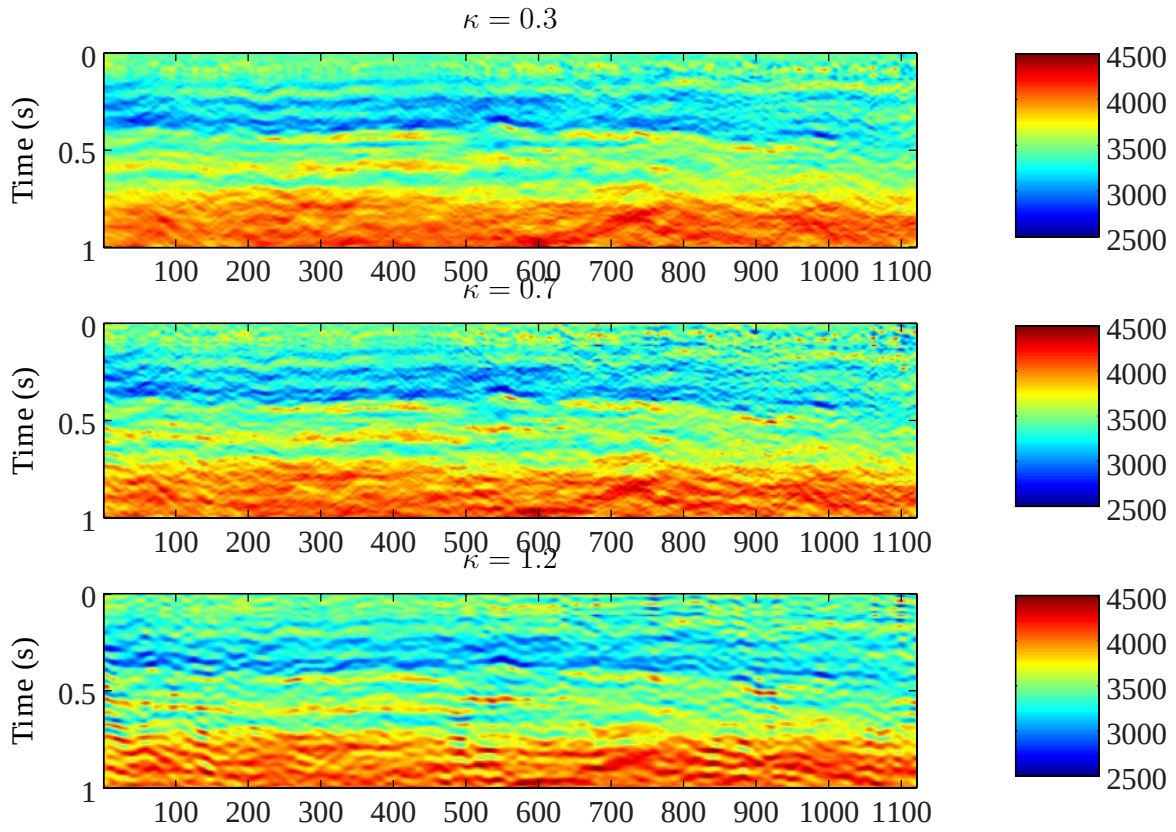

Figure 6.16: Blocky inversion result for $\boldsymbol{v_p}$ with $\kappa = 0.3$ (top), $\kappa = 0.7$ (middle) and $\kappa = 1.2$ (bottom). The lateral correlation was set to 0.4.

tested here is the best. In order to perform a comparison of the Gaussian and the blocky inversion for the Kvitebjørn data, the inversion result from the two methods for the trace nearest the borehole was plotted together. This is shown in Figure 6.17. Borehole data are included as black dots. The Gaussian inversion was performed with Σ_0 from borehole data, and the scale parameter in the blocky inversion was set to $\kappa = 0.7$. The lateral correlation coefficient was 0.4 for both. For the P-wave velocity and the S-wave velocity, the two inversion results are quite concurrent, whereas the blocky inversion gives much more big oscillations for the density. These oscillations are maybe not as reasonable as the Gaussian curve, but for the P- and S-velocity, both inversion methods seem to work fairly well.

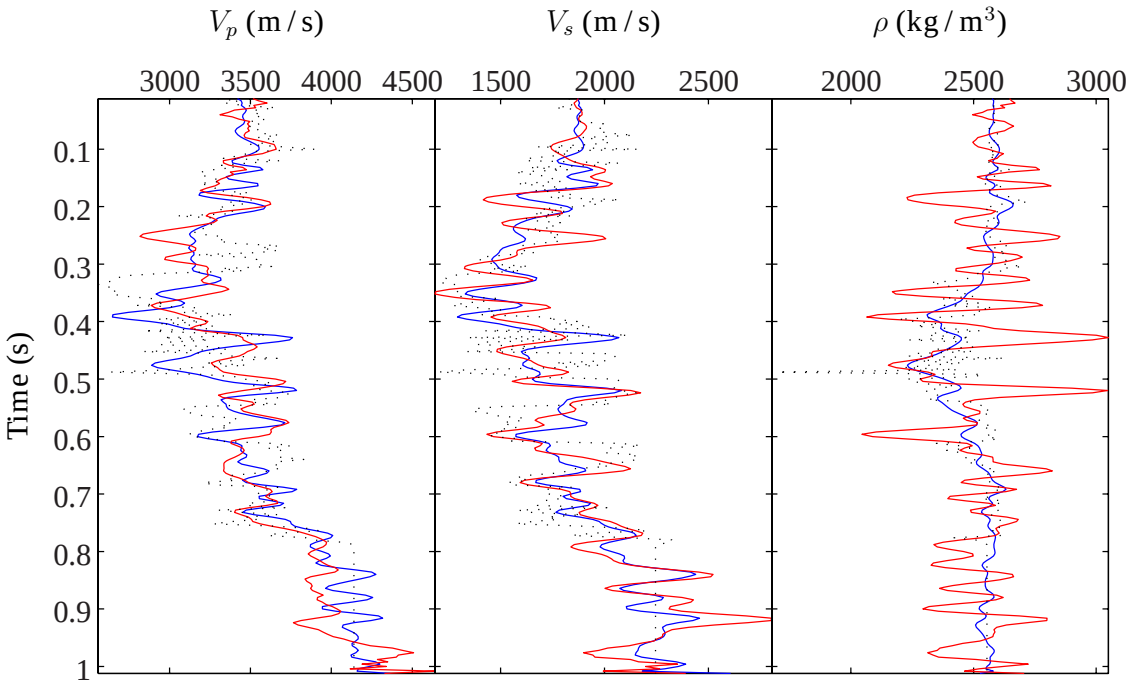

Figure 6.17: Comparison of the Gaussian and the blocky inversion result for the trace near well A. Σ_0 for the Gaussian inversion was estimated from borehole data and the scale parameter κ in the blocky inversion was set to 0.7. Black dots indicate well observations.

7 CLOSING REMARKS

This work includes inversion of a synthetic created case and pre-stack real seismic data from the Kvitebjørn field. Both standard Gaussian Bayesian inversion and blocky inversion are applied to compare the results in terms of achieved vertical resolution. The two methods differ due to different prior models for the elastic parameters $\boldsymbol{v_p}$, $\boldsymbol{v_s}$ and $\boldsymbol{\rho}$. The Gaussian inversion assumes a normal distribution as prior information of the parameters $\boldsymbol{m} = \ln(\boldsymbol{\xi})$, $\boldsymbol{\xi} \in \{\boldsymbol{v_p}, \boldsymbol{v_s}, \boldsymbol{\rho}\}$, whereas the blocky inversion assumes a Gaussian lateral correlation model with Cauchy constraints on the vertical contrasts. This approach is chosen because the Cauchy distribution has heavier tails than the normal distribution, which causes greater probability of large contrasts. This again can provide sharper edges between adjacent layers after inversion. Lateral correlation was included in the algorithm to provide lateral consistency. The following are attempted estimated; the covariance matrix for the Gaussian inversion, the scale parameter κ in the Cauchy distribution, the lateral correlation coefficient and the variance of the MAP solution.

The use of Cholesky factorization as a computational tool for inverting a large sparse matrix was very appropriate for this purpose. It was quite fast, and the size of the matrices to invert was $N \times N$ instead of $NM \times NM$ as it would have been for the direct solution. Due to memory problems in MATLAB, storage of the non-zero elements of the $\boldsymbol{L}$ matrix for all traces was a problem. For the Kvitebjørn data with 251 time samples, this problem arose when the number of traces was 30 or more. To invert the complete seismic line, the $\boldsymbol{L}$ matrix had to be saved piecewise to keep the lateral dependency for the forward and backward iterations. This problem could be solved by using a computer with more memory. Compared to the direct solution by setting up the whole $\boldsymbol{G}$ matrix for all traces and inverting a matrix of size $NM \times NM$, the Cholesky factorization still is a much better choice.

The results from the synthetic created example show satisfactory inversion results for both the Gaussian and the blocky inversion by using appropriate values for the parameters. For the Gaussian inversion, the estimation method for γ and ϕ described in Section 3.3 was working, whereas for the blocky model, parameter estimation did not work so well. The results for specified parameters

seem to recreate the truth satisfactory for both methods. It is not easy to decide which one is the best from these results because there are no major differences. The Gaussian result look more stable without the small rapid oscillations that appeared in the blocky inversion. The synthetic example also shows that the blocky inversion can provide results in the right magnitude even though the constraints were applied to the contrasts and not on the parameters themselves. The results recreate the blocky structure in the input model well, but with small-scale fluctuations in between. These may be because the posterior distribution when using a Cauchy constraint for the prior, is multimodal. A completely flat model between the layers might be just as likely. Estimation of uncertainties gave proper results for the Gaussian case, but the variances for the blocky inversion were unreasonable large.

For the Kvitebjørn data, problems regarding parameter estimation were experienced, for both the Gaussian and the blocky model. This might occur because of the scales of the real data caused by wavelet or prior levels, and the estimation possibly attempts to take this out with extreme sets of parameters. Appropriate parameters were found by trial and error. Then the inversion results for both the Gaussian and the blocky inversion were qualitatively compared to the recorded seismic amplitude data. The interpreted horizons were easier to detect from the inversion results than from the seismic data themselves, which could be an indicator of satisfactory results. Introducing the lateral correlation did not provide the expected differences, but some parts of the section appeared clearer with higher correlation coefficient.

After working with this inversion method, several challenging topics for further research stood out. Some of the issues are listed below.

- One possibility to improve the lateral correlation, could be to extend the neighbourhood in the Markov model from the nearest neighbours
- Estimation of parameters in the blocky inversion model was very demanding
- Estimation of uncertainties could be tested on other sampling based approaches besides importance sampling, which seemed to give unrealistically large uncertainties
- The Cauchy constraint seems to cause instabilities in the inversion results
- Other probability distributions could possibly be used in order to stabilize the inversion results and facilitate for estimation of uncertainties and other parameters

- The Cauchy assumption led to a weak non-linear problem, and well known numerical methods were easy to apply. This may not be the issue for e.g. the Laplace distribution

REFERENCES

Aki, K. and P. G. Richards (2002). *Quantitative Seismology* (2 ed.). University Science Books.

Ayzenberg, M. and B. Ursin (2007). *Seismic amplitude analysis*. Course material, TPG 4170, Reservoir Seismics, NTNU.

Buland, A. and Y. El Quair (2006). Bayesian time-lapse inversion. *Geophysics 71*, R43–R48.

Buland, A., O. Kolbjørnsen, and H. Omre (2003). Rapidly spatially coupled AVO inversion in the Fourier domain. *Geophysics 68*(3), 824–836.

Buland, A. and H. Omre (2003a). Bayesian linearized AVO inversion. *Geophysics 68*, 185–198.

Buland, A. and H. Omre (2003b). Joint AVO Inversion, Wavelet Estimation, and Noise Level Estimation using a Spatially Coupled Hierarchical Bayesian Model. *Geophysical Prospecting 51*, 531–550.

Carlin, B. P. and T. A. Louis (2000). *Bayes and empirical Bayes methods for data analysis* (2 ed.). Chapman & Hall / CRC.

Charbonnier, P., L. Blanc-Feraud, G. Aubert, and B. M. (1997). Deterministic Edge-Preserving Regularization in Computed Imaging. *IEEE Transactions on image processing 6*(2), 298–311.

Gunning, J. and M. E. Glinsky (2006). Wavelet extractor: a Bayesian well-tie wavelet derivation program. *Computers & Geosciences 32*, 681–695.

Kolbjørnsen, O. (2002a). *Nonlinear topics in the Bayesian approach to inverse problems with applications to seismic inversion, Cauchy prior for Bayesian linearized seismic AVO inversion*. Ph. D. thesis.

Kolbjørnsen, O. (2002b). *Nonlinear topics in the Bayesian approach to inverse problems with applications to seismic inversion, Fundamentals of inverse problems*. Ph. D. thesis.

Landrø, M. (2007). *Seismic Data Acquisition and Imaging* (4 ed.). Course material, TPG 4190, Seismic Data Acquisition and Processing, NTNU.

Liu, S. J. (2001). *Monte Carlo Strategies in Scientific Computing* (1 ed.). Springer.

Ostrander, W. J. (1984). Plane-wave reflection coefficients for gas sands at non-normal angles of incidence. *Geophysics 49*, 1637–1648.

Portniaguine, O. and M. S. Zhdanov (1999). Focusing geophysical inversion images. *Geophysics 64*(3), 874–887.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields, Theory and Applications* (1 ed.). Chapman & Hall / CRC.

Theune, U. (2007). *Blockyness constrains to enhance vertical resolution of AVA inversion*. Statoil internal report.

Youzwishen, C. F. and M. D. Sacchi (2006). Edge Preserving Imaging. *Journal of Seismic Exploration 15*(4), 45–58.

A EAGE POSTER

Some of the results from my project and master thesis have been presented at the EAGE conference in Rome, June 2008. Attached is a pdf version of the poster presentation.

Methods for Blocky

Seismic AVA Inversion

Introduction

One of the important challenges in seismic inversion is to resolve finer structures from band-limited seismic data. The resolution of seismic inversion is dominantly limited due to a) the lack of necessary high frequencies in the seismic data and b) artefacts introduced by mathematical inversion algorithms that are based on optimization techniques. Such artefacts can be smoothing or introduction of side lobes (Figure 1). Our work concentrates on an inversion scheme that aims at reducing the appearance of such artefacts and thereby improves the vertical resolution.

Several approaches have been taken to achieve this goal. On this poster, we discuss a stochastic approach to an improved vertical resolution, whose idea has been inspired by image processing algorithms (e.g., Charbonnier et al., 1997). Figure 1 illustrates this approach schematically. The principle idea of this approach is to enforce a sparse distribution of the model gradients, thereby sharpening the contrast of boundaries between adjacent layers. This is based on the assumption that mainly intermediate gradients contribute to the apparent smoothing in seismic inversion. In theory, the inverted properties of a layer have then sharper, flatter boundaries at the top and bottom, while they are constant within the layer. We may therefore name this inversion scheme a "blocky seismic amplitude inversion method".

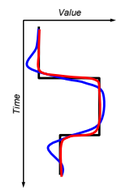

Fig.1: Illustrative example of the blocky inversion method. The black line indicates a property profile along with a sketch of inversion with substantial artefacts (blue line) and an inversion with reduced side-lobes (red line).

The Cauchy norm

The Cauchy norm can be characterized mathematically as a probability distribution with "heavier tails", which gives higher probability to larger gradients, whilst suppressing intermediate values. Figure 2 shows a comparison between the well-known normal distribution and the Cauchy distribution.

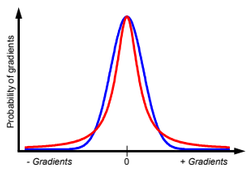

Fig. 2: Schematic comparison between normal (blue) and Cauchy distribution (red).

Using the Cauchy norm in an inversion scheme leads to a non-linear problem that can only be solved numerically. The "blocky" inversion algorithm can be derived as an optimum solution of Bayes' theorem. We will show different ways of determining the controlling statistical parameters of the blocky inversion. In order to stabilize the inversion, we also include a local lateral correlation into the inversion scheme.

Ingrid Østgård Jensås (NTNU and StatoilHydro R&D, Trondheim, Norway), Jo Eidsvik (NTNU Trondheim, Norway), Ulrich Theune (StatoilHydro R&D, Trondheim, Norway)

Theoretical considerations

Model

As usual, we assume a linear relationship between the observed data **d** and the model of the subsurface properties **m**:

$$\mathbf{d} = \mathbf{G}\cdot\mathbf{m} + \mathbf{n}$$

with
- **d**: seismic angle stacks
- **G**: the modelling operator
- **m**: model of the subsurface (P and S velocity, and density), and
- **n**: the omnipresent random noise

Data and model are statistically linked by Bayes' theorem

$$P(\mathbf{m}\,|\,\mathbf{d}) \propto P(\mathbf{d}\,|\,\mathbf{m})\cdot P(\mathbf{m})$$

In here, we assume a normal distribution for the data likelihood $P(\mathbf{d}|\mathbf{m})$ and a Cauchy distribution for the gradients of **m** as the a priori model assumption

$$P(\mathbf{d}\,|\,\mathbf{m}) \propto \exp\left\{-\tfrac{1}{2}(\mathbf{d}-\mathbf{G}\mathbf{m})^T\Sigma^{-1}(\mathbf{d}-\mathbf{G}\mathbf{m})\right\}$$

$$P(\mathbf{m}) \propto \exp\left\{-\sum_{i=1}^{N}\ln\left(1+\frac{\left((Dm)_i-(Dm_p)_i\right)^2}{\kappa_i^2}\right)\right\}$$

Where $\mathbf{m}_p$ is an aptly chosen low frequency or background model and κ is the so called scale parameter of the Cauchy distribution.

Solution of the inverse problem

In order to determine the posterior solution $P(\mathbf{m}|\mathbf{d})$, we calculate the maximum posterior estimate $\mathbf{m}_{MAP}$ by the means of optimization. This leads to

$$\mathbf{m}_{MAP}=(\mathbf{G}^T\Sigma^{-1}\mathbf{G}+4\mathbf{D}^T\mathbf{B}(\mathbf{m})\mathbf{D})^{-1}(\mathbf{G}^T\Sigma^{-1}\mathbf{d}+4\mathbf{D}^T\mathbf{B}(\mathbf{m})\mathbf{D}\mathbf{m}_p)$$

where the matrix $\mathbf{B}(\mathbf{m})$ contains the contribution to the solution that enforces blockyness:

$$B_{ii}(\mathbf{m}) = \frac{1}{\kappa_i^2+((\mathbf{Dm})_i-(\mathbf{Dm}_p)_i)^2}.$$

Note, that this matrix depends on the solution itself, and hence we are dealing with a non-linear inversion problem. However, the non-linearity is weak and we can treat the blocky inversion as a quasi-linear problem, for which solution algorithms exist.

Lateral correlation along a 2D line

In order to provide lateral consistency and robust results in the inversion, spatially coupling of the model parameters **m** is included in the algorithm.

Markov properties are assumed for simplicity, implying that the inversion result for one trace is only dependent on the neighbouring traces. In this case, we can impose lateral correlation by including a *precision matrix* **Q**, which has a sparse structure and is defined as

$$Q=\begin{bmatrix} 1 & -\phi & & & \\ -\phi & (1+\phi^2) & -\phi & & \\ & \ddots & \ddots & \ddots & \\ & & -\phi & (1+\phi^2) & -\phi \\ & & & -\phi & 1 \end{bmatrix}, \quad \phi\in[0\ \ 1]$$

We can now use φ as a parameter that controls the lateral correlation between neighbouring traces.

With this formulation, the prior model $P(\mathbf{m})$ is a joint Gaussian in the lateral direction, with Cauchy constraints in the vertical direction to ensure sharp layer contrasts. **Q** equal to zero gives a blocky trace by trace solution, while κ equal to infinity implies no Cauchy constraint.

The MAP solution for the blocky inversion in 2D can then be found by solving

$$\mathbf{m}_{MAP} = \mathbf{R}^{-1}\mathbf{b}$$
$$\mathbf{R} = [\mathbf{G}^T\Sigma^{-1}\mathbf{G}+\mathbf{Q}+4\mathbf{D}^T\mathbf{B}\mathbf{D}]$$
$$\mathbf{b} = [\mathbf{G}^T\Sigma^{-1}\mathbf{d}+\mathbf{Q}\mathbf{m}_p+4\mathbf{D}^T\mathbf{B}\mathbf{D}\mathbf{m}_p]$$

The precision matrix **Q** is implemented in 2D by multiplying each entry with an identity matrix.

Numerical aspects

The solution of the inverse problem cannot be obtained analytically. We used an iteratively reweighted method to determine the solution **m**:

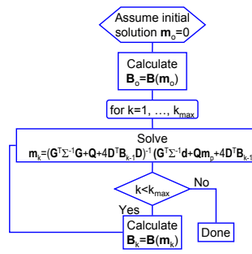

Fig. 3: Generalized workflow of the blocky inversion scheme.

Parameter estimation

Maximum likelihood estimation of κ from well logs

In order to estimate the controlling parameters of the inversion (κ, and φ in case of laterally correlated data) we suggest two statistical methods. If we are only interested in estimating κ, we can make use of borehole data. For a range of κ-values, we determine a maximum likelihood function $L(\mathbf{Dm})$ and choose the optimum κ_{opt} as that one, which maximizes L:

$$L(\mathbf{Dm},\kappa_j)=\prod_{i=1}^{N}\frac{1}{\pi\kappa_j(1+(Dm)_i^2/\kappa_j^2)}, \qquad j=1,\ldots,n_\kappa$$

$$\frac{\partial}{\partial\kappa}\ln(L(\mathbf{Dm})) = -\frac{N}{\kappa}+\frac{2}{\kappa^3}\sum_{i=1}^{N}\frac{(D_{ij}m_j)^2}{1+(D_{ij}m_j)^2/\kappa}=0$$

$$for\ \kappa=\kappa_{opt}$$

Estimation of the scale parameter κ from seismic data

Because of the completely different scales for well log data and seismic data, the scale parameter κ may be better estimated from seismic data than well log data to achieve consistency in the results. One alternative method is to determine an estimate of $P(\mathbf{d})$ and maximize this with respect to the scaling parameter κ and the lateral correlation coefficient φ. From Bayes' rule, an approximation of $P(\mathbf{d})$ can be written as

$$\hat{P}(\mathbf{d},\kappa,\phi)\approx\left.\frac{P(\mathbf{d}\,|\,\mathbf{m})P_\kappa(\mathbf{m})}{\hat{P}_{\phi,\kappa}(\mathbf{m}\,|\,\mathbf{d})}\right|_{\mathbf{m}=\mathbf{m}_{MAP}}$$

where the notation $P(\mathbf{d};\kappa,\phi)$ is used to highlight that this probability distribution has been determined for a certain κ,φ pair.

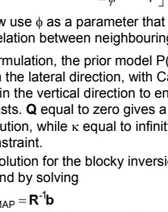

Fig 4.: Inversion workflow with initial parameter estimation

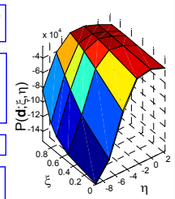

Fig 5.: Data probability distribution as a function of two parameters.

Examples

Synthetic example

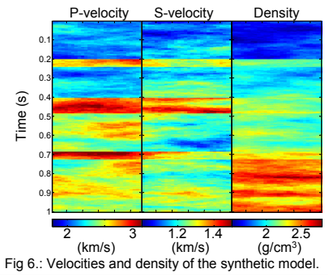

Fig 6.: Velocities and density of the synthetic model.

Inversion results

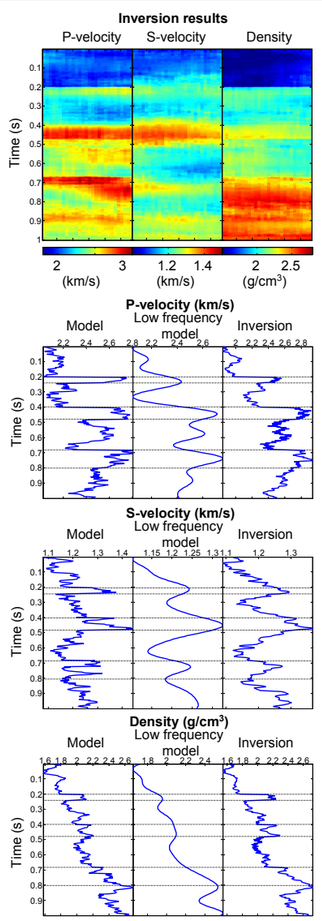

Fig 7.: Inversion results. Top: Line display, bottom: detailed comparison between model and inversion result.

Field data example from the Norwegian North Sea

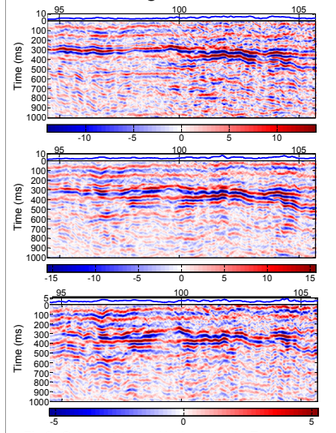

Fig 8.: Seismic data used in the example. Top: near angle stack, middle: mid angle stack, bottom: far angle stack. The estimated noise is shown above the panels.

From inversion tests we found that the following values for the parameters κ and φ gave satisfactory results:

$$\kappa = 1.2 \qquad \phi = 0.4.$$

In order to demonstrate the significance of the correlation parameter, we compare the inversion results for these parameters also with inversions for φ=0 and φ=0.8.

Inverted seismic properties

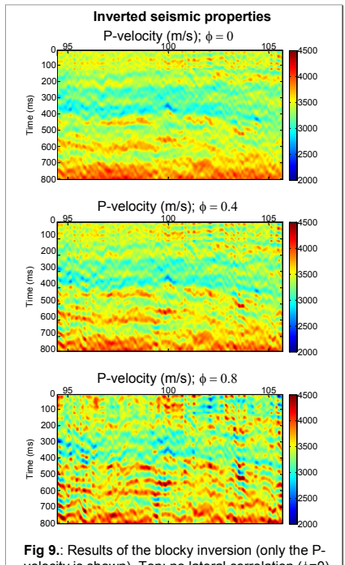

Fig 9.: Results of the blocky inversion (only the P-velocity is shown). Top: no lateral correlation (φ=0), middle: intermediate correlation (φ=0.4), bottom: strong lateral correlation (φ=0.8).

Summary

- The Cauchy norm applied to the gradients of the model parameters as stochastic a priori assumption can lead to inversion results with improved contrasts at layer boundaries.
- The formulation of the solution of the inverse problem leads to a non-linear solution that can only be calculated numerically.
- However, a solution based on the Cauchy norm can be transformed such that it allows us to use standard solution algorithms for the optimization of l^2-problems using iterative solvers.
- Our experience shows that two iterations in the IRLS loop are sufficient to obtain a stable solution.
- We included a local lateral correlation in the algorithm to enhance the stability of the inversion. If required, we can vary the lateral correlation weight within the cube to be inverted.
- Determination of the parameters κ and φ may be a challenge. We presented two statistical methods to estimate these parameters from log and seismic data, respectively.
- These parameter estimations are notably noise sensitive.
- The method is computationally demanding, mainly because of the iterative solution approach.
- The synthetic example demonstrates that the blocky inversion method indeed is able to create inversion results with sharp layer contrasts.
- Inversion tests for different correlation parameters φ for the field data example showed that this value influences the inversion considerably.
- Posterior uncertainties are difficult to assess directly. The posterior curvature at the inversion result carries some information about the uncertainty, but the posterior is not Gaussian, and the usual 5 and 95 percentiles are not directly available.

References

Aki, K. and Richards, P.G. [2002], *Quantitative Seismology (2.ed)*, **University Science Books**.

Charbonnier, P., Blanc-Féraud, L., Aubert, G. and Barlaud, M. [1997] *Deterministic Edge-Preserving Regularization in Computed Imaging*. **IEEE Transaction on image processing**, 6, 298-311

Rue, H. and Held, L., [2005], *Gaussian Markov Random Fields – Theory and Applications*,**Chapman & Hall/CRC**

Scales, J., Gersztenkorn, A., and Treitel, S. [1988], *Fast l_p solution of large, sparse, linear systems: Application to seismic travel time tomography*, **J. Comput. Phys.**, 75

Contact information

Ingrid Østgård Jensås (ioj@statoilhydro.com)
Jo Eidsvik (joeid@math.ntnu.no)
Ulrich Theune (uth@statoilhydro.com)

Acknowledgements

Ingrid Ø. Jensås likes to thank StatoilHydro for supporting her master project and her attendance at the EAGE conference.

NTNU
Norwegian University of Science and Technology

StatoilHydro

NTNU
Norwegian University of Science and Technology

StatoilHydro