NTNU
Norwegian University of
Science and Technology

# Using the Composite Likelihood Method on 4D AVA Seismic Data

**Yngve Borgan**

Master of Science in Physics and Mathematics
Submission date: June 2011
Supervisor: Jo Eidsvik, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

# Problem Description

The aim of this thesis is to investigate the performance of the composite likelihood method on 4D AVA seismic data. The method is tested for parameter estimation and prediction of the elastic parameters.

Assignment given: 20. January 2011
Supervisor: Jo Eidsvik

**Acknowledgements**

This thesis was carried out at the Department of Mathematical Sciences and Technology at the Norwegian University of Science and Technology (NTNU) during the period January 2011 to June 2011.

I would first like to thank my supervisor Jo Eidsvik for his excellent guidance and support. I would also like to thank my fellow students in reading hall 393C for motivational digressions and morale support.

Trondheim, June 2011

Yngve Borgan

**Abstract**

This thesis is concerned with 4D AVA seismic inversion problems. By comparing two seismic surveys done over the same area, but at different times, one hopes to discover untapped pockets of oil or gas. Using the full likelihood to analyse 4D AVA seismic data is impossible in practice due to memory and computational restrictions. The goal of the thesis is to find a useful framework for parameter estimation and predictions for 4D AVA seismic data, and the composite likelihood is introduced as a possible solution. The composite likelihood method takes in pairs of data points and sums over them instead of taking in all the data as is the case for the full likelihood. This makes calculations fast while avoiding matrix operations on large matrices.

The composite likelihood method is tested on a data set from the Norne field for parameter estimations and predictions. Eight variations of the model are tested, the variations being the exponential or Matern correlation function, one or two data columns used as a data point in the composite likelihood, and a simple or wavelet convoluted noise term. The composite likelihood method is shown to perform well; it is fast and the estimates found agree well with previous experience. Comparison of the different models indicate that the choice of correlation function has little effect on the results, that the noise term should be kept simple, and that it is sufficient to use one data column.

# Contents

# 1 Introduction

The field of statistics has changed drastically with the availability of computational power, which allows us to generate and store vast amounts of data. Examples arise in navigation with GPS, weather measurements, seismic data and in many other fields. Large data sets can easily be stored and used for calculations in ways thought to be of only theoretical interest not many years ago. But the new possibilities introduced by great computational power comes with new problems and restrictions as well. Data sets of great size can now be used for calculations, but there are still limits to what computers can accomplish.

This thesis will concern itself with seismic reflection, a field in which large data sets are unavoidable. Seismic data are for instance used to determine whether there are reservoirs of oil or gas by analysing the reflection of sound waves. In later years a new technique called 4D seismic AVA inversion has been developed that is not used for exploration, but rather for reservoir monitoring. AVA, or AVO which is basically the same, refers to amplitude versus angle, or offset in the case of AVO. What it means is that the strength of seismic waves measured in their amplitude is related to the angle of the wave [4]. When the oil or gas is extracted from the reservoir it is often replaced by water. Comparing two seismic surveys done in the same area, but at different times makes it possible to ascertain whether this change has actually taken place. 4D seismic AVA inversion is therefore a valuable tool when finding new places to drill in an existing reservoir. The main references on 4D seismic AVA inversion include [4] and [10]

Using statistics in 4D seismic AVA inversion consists of creating a statistical model based on prior knowledge of the reservoir, then finding the inversion results by some statistical method using a physically based model for the likelihood of the data. Seismic data will typically have dimensions $100 \times 100 \times 100$ or larger leading to a covariance matrix of size $10^6 \times 10^6$. Doing calculations involving such a large matrix is memory demanding, computationally slow or even numerically impossible.

It is therefore of interest to find an alternative likelihood method that avoids computations involving the full covariance matrix. For this purpose the composite likelihood method is introduced. The method uses the covariance for pairs of data entries and sums over all pairs instead of using the full covariance matrix. When assuming that the covariance depends on the distance between data points, not all pairs of data need to be included in the calculation to obtain adequate results and the method becomes fast. The main references used on the composite likelihood and similar likelihood methods for large spatial data sets include [6], [14], [9] and [13].

The composite likelihood method will be tested on a 4D seismic data set from the Norne field. This will be the main focus of the thesis, investigating the performance of the composite likelihood method on real seismic data. Different variations of the statistical model will be tested for parameter estimation and prediction.

This thesis is organised as follows: Section 2 starts with an introduction to 4D seismic, why it is used and what challenges arise in connection with the method. Next the physics behind seismic reflection is summarised and the statistical model for both 3D and 4D seismic data is explained. In Section 3 the reader is introduced to the composite likelihood method. Parameter estimation and asymptotic properties of the composite

likelihood estimators are also discussed. Section 4 explains how to use the composite likelihood method to do predictions and ways of assessing the results. In section 5 the data set from Norne is introduced and the model for the data described. The different model variations will also be described. In Section 6 results of the parameter estimation and the results of the predictions are presented and discussed. Section 7 summarises the results and what has been accomplished in this thesis. A discussion of possible improvements and potential further work is also included.

# 2 Seismic Model

## 2.1 Background

Seismic exploration is most commonly associated with discovering new reservoirs of oil and gas beneath the sea floor. A seismic vessel emits sound waves that are reflected and then read by sensors that trail the vessel. Rock layers containing oil or gas have distinct elastic properties so analysis of the reflected waves can be used to determine if there is oil or gas present. The waves are reflected in the subsurface rock layers and will vary in amplitude and angle depending on the properties of the rock layers that reflected them.

In recent years a new way of using seismic readings has been introduced, the 4D seismic AVA inversion method. The 4D seismic AVA inversion method makes further use of the original survey, called the base survey, by comparing it with a new seismic survey, called the monitor survey, in the same area. The data in a seismic survey is usually gathered over a three dimensional grid and the two different time values are the fourth dimension thus explaining the name.

The approach is primarily used to investigate where there is remaining oil or gas in a reservoir. When a reservoir is emptied water is usually injected to push the oil or gas out. By repeated seismic surveying it is possible to compare the results for two different time periods. The reflected seismic waves are different for water than for oil or gas so the technique is able to tell if all the oil or gas in a reservoir has been replaced by water or not.

4D seismic surveys have also helped improve existing 3D analysis methods. In several instances unexplained anomalies that were categorised as noise have shown themselves to appear again in 4D surveys and hence the anomalies could not be explained as random noise. The same is true for the opposite case where events that were thought to be significant could be dismissed as noise after performing a 4D survey. In this respect the 4D seismic approach has contributed to improve analysis of 3D seismic data [10].

The method has become a great success and is an important asset in prolonging a field's economic life. In 2001 there were 75 active projects worldwide with an annual expenditure of $50-100 million US [10] and the amount of projects have risen steadily since. Since the world's oil reserves are steadily diminishing and new fields are harder to come by techniques for extracting as much as possible from existing fields, like 4D AVA, will become more important in the years to come.

## 2.2 Challenges

There are many challenges regarding the 4D seismic approach which makes understanding and analysing the changes that arise difficult. It is impossible for the seismic vessel to collect data from the exact same locations both times because of the heavy boat traffic and new installations that have been built. When performing the monitor survey there will naturally have been built an installation in connection with the pocket of oil that was discovered in the original survey. The difference in measurement locations

and angles will naturally increase measurement noise and cause discrepancies between the measurements that are not due to actual changes in the rock-layers.

The challenges with regard to collecting the data present new problems when it comes to the processing of the data. By processing the data one aims to create individual 3D seismic images for the different surveys. It is important that the images are created so they are fairly similar in areas where it is known to be no subsurface changes. Keeping this in mind the goal of the data processing is to make the data sets comparable by standardising the data with respect to amplitude, phase, spectral bandwidth etc. [10]. This is a time consuming and intricate process and is of great importance to the analysis.

In a 4D setting it is usually assumed that the geology of the area investigated is time-invariant. This is usually the case and makes changes in saturation, pressure and temperature easier to identify. But time-invariant geology is not always the case and the assumption can then cause analysis of the data to give erroneous results. For more information on 4D seismic and the challenges it presents the reader is referred to [10].

In spite of the many difficulties associated with 4D seismic it does a good job in comparing data from measurements done at different times. By subtracting the data for the baseline survey from the data for the monitor survey a data set with differences is obtained. This new combined data set is analysed using a variety of different geophysical and statistical methods. Some of these methods include Bayesian time-lapse inversion [5], AVA inversion in the Fourier domain [3], block composite likelihood [6] and several others found in [10]. This thesis will focus on the composite likelihood method and explore it's performance in connection with 4D seismic.

## 2.3 Gaussian model for linear 3D seismic AVA inversion

In a seismic AVA setting it is natural to consider a Gaussian response variable $\boldsymbol{Y}$ containing all reflected data in a two- or three dimensional grid. The response variable depends on a set of elastic parameters. By using knowledge of the rock physics involved it is possible to write out a model for the response variable. The grid for the data is illustrated in Figure 1 and has dimensions $n_1 \times n_2 \times n_3 = n$ where $n_3$ is the depth.

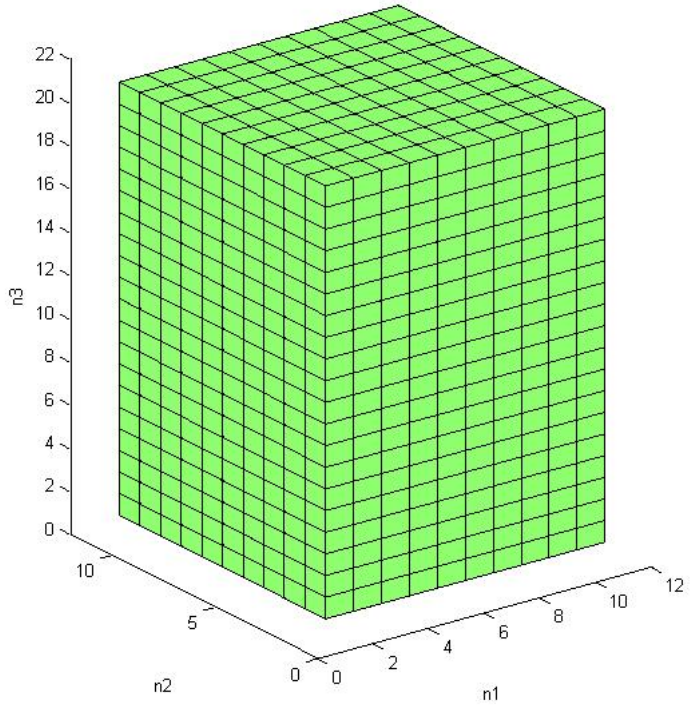



Figure 1: An example of the form of the data. In the figure $n_1$, $n_2$ and $n_3$ are 10, 10 and 20 respectively, but they take on different values throughout this thesis

To create the Gaussian model for the data an understanding of the elastic parameters is needed. For each point in the grid there are three elastic parameters of interest in a seismic setting, p-wave velocity ($\alpha$), s-wave velocity ($\beta$) and density ($\rho$). The p-wave, or compressional wave, velocity is the speed of waves oscillating in the direction of the waves energy. S-wave, or shear wave, velocity is the speed of waves oscillating perpendicular to the waves direction. The density is the density of the rock layer from which the wave is reflected.

The expressions for the p-wave velocity and the s-wave velocity and their relation

to the density are

$$\alpha = \sqrt{\frac{K_{sat} + \frac{4}{3}\mu_{sat}}{\rho}} \tag{1}$$

$$\beta = \sqrt{\frac{\mu_{sat}}{\rho}} \tag{2}$$

as given in [11]. $K_{sat}$ is the bulk modulus and $\mu_{sat}$ the shear modulus. $K_{sat}$ and $\mu_{sat}$ are determined using Gassmann's equation and depend on the porosity and saturation of the rock. Assuming homogeneous mineral modulus and saturation, the quasi-static variant of Gassmann's equation is given by

$$\begin{aligned} \frac{K_{sat}}{K_0 - K_{sat}} &= \frac{K_{dry}}{K_0 - K_{dry}} + \frac{K_{fluid}}{\psi(K_0 - K_{fluid})} \\ \mu_{sat} &= \mu_{dry} \end{aligned} \tag{3}$$

and can with the adequate knowledge be used to find the bulk modulus of fully saturated rock [1]. In equation 3 $K_0$ is the bulk moduli of the mineral grain, $K_{dry}$ the bulk modulus of dry rock, $K_{fluid}$ the bulk modulus of the pore-fluid and $\psi$ the porosity. Often the dry bulk modulus $K_{dry}$ is not available, but the bulk modulus of a fully brine-saturated rock is, usually obtained from a well log. Brine is basically water with a high concentration of salt. Using the new bulk modulus equation 3 can be modified

$$\begin{aligned} \frac{K_{sat1}}{K_0 - K_{sat1}} - \frac{K_{fluid1}}{\psi(K_0 - K_{fluid1})} &= \frac{K_{sat}}{K_0 - K_{sat}} - \frac{K_{fluid}}{\psi(K_0 - K_{fluid})} \\ \mu_{sat} &= \mu_{dry} \end{aligned} \tag{4}$$

where $K_{sat1}$ and $K_{fluid1}$ are the rock bulk modulus and pore fluid modulus respectively for the fully brine-saturated rock. An expression for $K_{fluid}$, the pore fluid modulus for the partially saturated rock, is given by

$$K_{fluid}(s) = \left(\frac{s}{K_{brine}} + \frac{1-s}{K_{hc}}\right)^{-1} \tag{5}$$

where $K_{brine}$ and $K_{hc}$ are the bulk moduli of brine and the given hydrocarbon respectively and $s$ is the saturation. It is now possible to calculate $K_{sat}$ using the expressions from equations 4 and 5 together with prior knowledge of the rock layers. For more on the rock physics behind seismic exploration the reader is referred to [1] and [11].

By finding estimates of the elastic parameters, their mean, variance, etc. it is possible to determine whether or not all oil or gas has been replaced by water in an assumed empty oil pocket. In order to obtain information on these parameters a suitable model is needed. The three elastic parameters of interest are placed in a vector

$$\boldsymbol{x} = \begin{bmatrix} \ln(\alpha) \\ \ln(\beta) \\ \ln(\rho) \end{bmatrix}$$

The elastic parameters $\alpha$, $\beta$, and $\rho$ are positive, but by taking logarithms the values of $\boldsymbol{x}$ have no such restriction. Using vector $\boldsymbol{x}$ it is possible to determine the reflection coefficient for a data point

$$z_{ik}(\omega_p) = \boldsymbol{a}_{\omega_p}^T \cdot (\boldsymbol{x}_{ik} - \boldsymbol{x}_{i,k-1}) \,. \tag{6}$$

Here $i$ and $k$ denote the position of a measurement where $i$ is the index of the data column and $k$ the index of the depth. $\boldsymbol{a}_{\omega_p}$ is a vector of constants dependant only on the angle $\omega_p$ giving the linear relationship between the three parameters for a given angle. This is based on the Aki Richards equations which are linearisations of the more general Zoeppritz equations [12].

$$\begin{aligned} a_\alpha(\omega_p) &= \frac{1}{2}(1 + \tan^2 \omega_p) \\ a_\beta(\omega_p) &= -4\frac{\bar{\beta}_k^2}{\bar{\alpha}_k^2} \sin^2 \omega_p \\ a_\rho(\omega_p) &= \frac{1}{2}(1 - 4\frac{\bar{\beta}_k^2}{\bar{\alpha}_k^2} \sin^2 \omega_p) \end{aligned}$$

The values for $\bar{\alpha}_k$ and $\bar{\beta}_k$ are held constant for a given depth and derived from a known well log. Alternatively they can be held constant for all depths. The angle $\omega_p$ is the reflection angle of the seismic wave. $p$ takes on the values from 1 to $P$, where $P$ is the number of angles from the seismic data gathering.

From equation 2.3 it is now possible to write out an expression for finding the reflection coefficients for different angles

$$\begin{bmatrix} \boldsymbol{z}_i(\omega_1) \\ \vdots \\ \boldsymbol{z}_i(\omega_P) \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{\alpha\omega_1} & \boldsymbol{A}_{\beta\omega_1} & \boldsymbol{A}_{\rho\omega_1} \\ \vdots & \vdots & \vdots \\ \boldsymbol{A}_{\alpha\omega_P} & \boldsymbol{A}_{\beta\omega_P} & \boldsymbol{A}_{\rho\omega_P} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_\alpha \\ \boldsymbol{x}_\beta \\ \boldsymbol{x}_\rho \end{bmatrix} \tag{7}$$

where the reflection coefficients $\boldsymbol{z}_i(\omega)$ are built by using the expression from equation 6

$$\boldsymbol{z}_i(\omega_p) = \begin{bmatrix} z_{i1}(\omega_p) \\ z_{i2}(\omega_p) \\ \vdots \\ \vdots \\ z_{in_3}(\omega_p) \end{bmatrix}$$

and

$$\boldsymbol{x}_\alpha = \begin{bmatrix} x_{i1\alpha} \\ x_{i2\alpha} \\ \vdots \\ \vdots \\ x_{in_3\alpha} \end{bmatrix} \quad \boldsymbol{x}_\beta = \begin{bmatrix} x_{i1\beta} \\ x_{i2\beta} \\ \vdots \\ \vdots \\ x_{in_3\beta} \end{bmatrix} \quad \boldsymbol{x}_\rho = \begin{bmatrix} x_{i1\rho} \\ x_{i2\rho} \\ \vdots \\ \vdots \\ x_{in_3\rho} \end{bmatrix} .$$

$x_{ik\alpha}$ is $\ln \alpha_{ik}$ where $i$ is the column index and $k$ is the depth in that column as before. The $\boldsymbol{A}_{\omega_p\alpha}$, $\boldsymbol{A}_{\omega_p\beta}$ and $\boldsymbol{A}_{\omega_p\rho}$ from equation 7 are given by

$$\boldsymbol{A}_{\omega_p\alpha} = \begin{bmatrix} -a_{\alpha\phi_p} & a_{\alpha\omega_p} & 0 & & 0 & 0 \\ 0 & -a_{\alpha\omega_p} & a_{\alpha\omega_p} & \dots & 0 & 0 \\ 0 & 0 & -a_{\alpha\omega_p} & & 0 & 0 \\ & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -a_{\alpha\omega_p} & a_{\alpha\omega_p} \\ a_{\alpha\omega_p} & 0 & 0 & \dots & 0 & -a_{\alpha\omega_p} \end{bmatrix}$$

$$\boldsymbol{A}_{\omega_p\beta} = \begin{bmatrix} -a_{\beta\omega_p} & a_{\beta\omega_p} & 0 & & 0 & 0 \\ 0 & -a_{\beta\omega_p} & a_{\beta\omega_p} & \dots & 0 & 0 \\ 0 & 0 & -a_{\beta\omega_p} & & 0 & 0 \\ & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -a_{\beta\omega_p} & a_{\beta\omega_p} \\ a_{\beta\omega_p} & 0 & 0 & \dots & 0 & -a_{\beta\omega_p} \end{bmatrix} \tag{8}$$

$$\boldsymbol{A}_{\omega_p\rho} = \begin{bmatrix} -a_{\rho\omega_p} & a_{\rho\omega_p} & 0 & & 0 & 0 \\ 0 & -a_{\rho\omega_p} & a_{\rho\omega_p} & \dots & 0 & 0 \\ 0 & 0 & -a_{\rho\omega_p} & & 0 & 0 \\ & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -a_{\rho\omega_p} & a_{\rho\omega_p} \\ a_{\rho\omega_p} & 0 & 0 & \dots & 0 & -a_{\rho\omega_p} \end{bmatrix}$$

The amplitude of the seismic wave is modelled by a convolutional model given by

$$y_{ik} = \Sigma_{j=-b}^{b} w_{j\omega_p} \cdot z_{i,k-j} + \epsilon_i, \tag{9}$$

where $w_{j\omega_p}$ is a wavelet function for the given angle $\omega_p$ and $\epsilon_i$ is a noise term. $b$ is a number giving the length of the wavelet in each direction. The expression for an entire data column now becomes

$$\boldsymbol{Y}_i = \begin{bmatrix} \boldsymbol{y}_{i\omega_1} \\ \vdots \\ \boldsymbol{y}_{i\omega_P} \end{bmatrix} = \begin{bmatrix} \boldsymbol{W}_{\omega_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boldsymbol{W}_{\omega_P} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{\alpha\omega_1} & \boldsymbol{A}_{\beta\omega_1} & \boldsymbol{A}_{\rho\omega_1} \\ \vdots & \vdots & \vdots \\ \boldsymbol{A}_{\alpha\omega_P} & \boldsymbol{A}_{\beta\omega_P} & \boldsymbol{A}_{\rho\omega_P} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{i\alpha} \\ \boldsymbol{x}_{i\beta} \\ \boldsymbol{x}_{i\rho} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{i\omega_1} \\ \vdots \\ \boldsymbol{\epsilon}_{i\omega_P} \end{bmatrix}$$

where $\boldsymbol{W}_{\omega_p}$ is an $n_3 \times n_3$ matrix where the weights from the wavelet function is placed on and around the diagonal and is zero elsewhere. The block diagonal matrix containing $\boldsymbol{W}_{\boldsymbol{\omega_p}}$, denoted $\boldsymbol{W}$, is an $n_3 \cdot P \times n_3 \cdot P$ matrix. The $\boldsymbol{A}_{\cdot,\omega_p}$ matrices are the blocks from equation 8 where each block is a $n_3 \times n_3$ matrix. The whole matrix of $\boldsymbol{A}_{\cdot,\omega_p}$s, denoted $\boldsymbol{A}$, is an $n_3 \cdot P \times n_3 \cdot 3$ matrix. $\boldsymbol{\epsilon}_{i\omega_p}$ is an $n_3 \times 1$ vector with noise terms for the given angle $\omega_p$.

The data vector $\boldsymbol{Y}_i$ for a single data column is given by

$$\boldsymbol{Y}_i = \begin{bmatrix} y_{i\omega_1 1} \\ y_{i\omega_1 2} \\ \vdots \\ y_{i\omega_1 n_3} \\ \vdots \\ y_{i\omega_P 1} \\ y_{i\omega_P 2} \\ \vdots \\ y_{i\omega_P n_3} \end{bmatrix} \tag{10}$$

It is built up as a single vector with length $n_3 \cdot P$. The first $n_3$ elements of the vector will be the $y_{ik}$'s for $i = 1$ and $k$ going from $1 \rightarrow n_3$. Then the $\boldsymbol{Y}_i$ for each angle is appended to the vector and then the process is repeated for all $i$ so that the complete data vector $\boldsymbol{Y}$ is given by

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_{n_1 n_2} \end{bmatrix}$$

The $\boldsymbol{X}_i$ vector is similarly built

$$\boldsymbol{X}_i = \begin{bmatrix} x_{i\alpha 1} \\ x_{i\alpha 2} \\ \vdots \\ x_{i\alpha n_3} \\ x_{i\beta 1} \\ x_{i\beta 2} \\ \vdots \\ x_{i\beta n_3} \\ x_{i\rho 1} \\ x_{i\rho 2} \\ \vdots \\ x_{i\rho n_3} \end{bmatrix} \tag{11}$$

It is a vector of length $n_3 \cdot 3$ where the first $n_3$ elements are the $\ln(\alpha)$ values for each point in the grid followed by the $n_3$ elements for $\ln(\beta)$ and lastly the $n_3$ for $\ln(\rho)$. The entire $\boldsymbol{X}$ vector is built the same way as $\boldsymbol{Y}$ by appending $\boldsymbol{X}_i$ together for all $i$ so that the complete vector of elastic parameters $\boldsymbol{X}$ is given by

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_{n_1 n_2} \end{bmatrix}$$

Denoting $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{A}$ it is now possible to write out the Gaussian model

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{G} & & \\ & \ddots & \\ & & \boldsymbol{G} \end{bmatrix} \boldsymbol{X} + \boldsymbol{\epsilon}. \tag{12}$$

The noise vector $\boldsymbol{\epsilon}$ is built the same way as $\boldsymbol{Y}$ and is assumed independent of $\boldsymbol{X}$. The noise parameter is assumed to be normally distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}_\epsilon$.

In a 3D seismic setting the prior of $\boldsymbol{X}$ is usually selected as

$$\boldsymbol{X} \sim N(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X(\boldsymbol{\theta})),$$

where $\boldsymbol{\mu}_X$ is the expectation vector of $\boldsymbol{X}$ and $\boldsymbol{\Sigma}_X(\boldsymbol{\theta})$ is the covariance matrix. $\boldsymbol{\theta}$ is a vector of parameters that are involved in the parametrisation of $\boldsymbol{\Sigma}_X$. Using this prior distribution the marginal distribution of $\boldsymbol{Y}$ becomes

$$\boldsymbol{Y} \sim N(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y(\boldsymbol{\theta}))$$

where

$$\boldsymbol{\mu}_Y = \begin{bmatrix} \boldsymbol{G} & & \\ & \ddots & \\ & & \boldsymbol{G} \end{bmatrix} \boldsymbol{\mu}_X$$

$$\boldsymbol{\Sigma}_Y(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{G} & & \\ & \ddots & \\ & & \boldsymbol{G} \end{bmatrix} \boldsymbol{\Sigma}_X(\boldsymbol{\theta}) \begin{bmatrix} \boldsymbol{G} & & \\ & \ddots & \\ & & \boldsymbol{G} \end{bmatrix}^T + \boldsymbol{\Sigma}_\epsilon$$

## 2.4 Gaussian Model in a 4D seismic setting

For the 4D seismic case the basics of the model stays the same. The model discussed in the previous section holds for each of the surveys independently. The difference is that now the response $\boldsymbol{Y}$ is the difference between to surveys

$$\boldsymbol{Y} = \boldsymbol{Y}_{\text{Monitor}} - \boldsymbol{Y}_{\text{Base}}$$

The model for a 4D response is

$$\boldsymbol{Y}_{\text{Monitor}} - \boldsymbol{Y}_{\text{Base}} = \boldsymbol{G}_{\text{Monitor}}\boldsymbol{X}_{\text{Monitor}} + \boldsymbol{\epsilon}_{\text{Monitor}} - \boldsymbol{G}_{\text{Base}}\boldsymbol{X}_{\text{Base}} - \boldsymbol{\epsilon}_{\text{Base}}$$

The prior of $\boldsymbol{X}_{monitor}$ and $\boldsymbol{X}_{base}$ is assumed equal, so the prior of $\boldsymbol{X} = \boldsymbol{X}_{monitor} - \boldsymbol{X}_{base}$ becomes

$$\boldsymbol{X} \sim (\boldsymbol{0}, \boldsymbol{\Sigma}_X)$$

where $\boldsymbol{\Sigma}_X$ is the covariance matrix of the differences between the two surveys. In this thesis $\boldsymbol{G}_{\text{Monitor}}$ and $\boldsymbol{G}_{\text{Base}}$ are assumed equal for both surveys. The mean of the 4D data then becomes

$$\begin{aligned} E(\boldsymbol{Y}) &= E(\boldsymbol{G}(\boldsymbol{X}_{\text{Monitor}} - \boldsymbol{X}_{\text{Base}}) + \boldsymbol{\epsilon}_{\text{Monitor}} - \boldsymbol{\epsilon}_{\text{Base}}) \\ &= \boldsymbol{G}E(\boldsymbol{X}_{\text{Monitor}} - \boldsymbol{X}_{\text{Base}}) \\ &= \boldsymbol{G}(\boldsymbol{\mu}_{\text{Monitor}} - \boldsymbol{\mu}_{\text{Base}}), \end{aligned}$$

which assuming equal expected value for both surveys leads to $\boldsymbol{\mu}_Y = \boldsymbol{0}$. The variance of $\boldsymbol{Y}$ is now given by

$$\text{var}(\boldsymbol{Y}) = \text{var}(\boldsymbol{G}(\boldsymbol{X}_{\text{Monitor}} - \boldsymbol{X}_{\text{Base}})) + \text{var}(\boldsymbol{\epsilon}_{\text{Monitor}} - \boldsymbol{\epsilon}_{\text{Base}})$$

Setting $\boldsymbol{X} = \boldsymbol{X}_{\text{Monitor}} - \boldsymbol{X}_{\text{Base}}$ and $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}_{\text{Monitor}} - \boldsymbol{\epsilon}_{\text{Base}}$ gives

$$\begin{aligned} \text{var}(\boldsymbol{Y}) &= \boldsymbol{G}\text{var}(\boldsymbol{X})\boldsymbol{G}^T + \text{var}(\boldsymbol{\epsilon}) \\ &= \boldsymbol{G}\boldsymbol{\Sigma}_X\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon \end{aligned}$$

which means that the variance of $\boldsymbol{Y}$ is the variance of changes in noise and the elastic parameters. For more on the Gaussian 4D model is referred to [5]. For the remainder of the thesis $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{\epsilon}$ will refer to differences in the data, elastic parameters and noise respectively.

# 3 Composite likelihood

## 3.1 Likelihood

One of the goals of this thesis is to obtain estimates for the parameter vector $\boldsymbol{\theta}$ through likelihood methods. The parameter vector is crucial for obtaining reliable inversion results of elastic properties for seismic data. In the general case a normally distributed response vector $\boldsymbol{y}$ has the likelihood function

$$L(\boldsymbol{y};\boldsymbol{\theta}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}_y(\boldsymbol{\theta})|^{-1/2}\exp(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Sigma}_y(\boldsymbol{\theta})^{-1}\boldsymbol{y})$$

where $\boldsymbol{\theta}$ is the vector of parameters involved in the expression of the covariance of $\boldsymbol{y}$. Ignoring a scalar depending neither on $\boldsymbol{\theta}$ or $\boldsymbol{y}$ the negative log-likelihood function is given by

$$l(\boldsymbol{y};\boldsymbol{\theta}) = \frac{1}{2}\ln|\boldsymbol{\Sigma}_y(\boldsymbol{\theta})| + \frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Sigma}_y(\boldsymbol{\theta})^{-1}\boldsymbol{y}. \tag{13}$$

When $n$ becomes large using the covariance matrix for calculations becomes time consuming and sometimes impossible due to the matrix operations on the covariance matrix. Finding the inverse of a matrix for example when $n$ becomes large enough is numerically impossible and therefore alternatives to the full covariance matrix must be found.

## 3.2 Composite likelihood

Composite likelihood is an alternative to ordinary likelihood described in [14] and [9]. The idea of the composite likelihood method is to break up the calculation of the log-likelihood function into smaller pieces and then summing them. It especially becomes efficient when one assumes the data to be spatially correlated, meaning that the corraltion between data points is dependent on the distance between them. Spatial correlation is usually assumed for seismic data. It is then possible to consider only the closest neighbourhood without much loss of accuracy.

For the general case the composite likelihood method sums up the likelihood function for each pair of data points, instead of finding the likelihood function for the whole data set at once. Assuming a response vector $\boldsymbol{y}$ with elements $y_i$ for $i = 1 \to N$ the composite likelihood function becomes

$$CL(\boldsymbol{y};\boldsymbol{\theta}) = \prod_i \prod_j f(y_i, y_j;\boldsymbol{\theta}) \tag{14}$$

where $f(y_i, y_j|\boldsymbol{\theta})$ is the joint probability distribution for a data pair. The negative log-composite-likelihood function becomes

$$cl(\boldsymbol{y};\boldsymbol{\theta}) = \sum_i \sum_j \left(\frac{1}{2}\ln|\boldsymbol{\Sigma}_{yij}(\boldsymbol{\theta})| + \frac{1}{2}y_{ij}^T\boldsymbol{\Sigma}_{yij}(\boldsymbol{\theta})^{-1}y_{ij}\right),$$

where $y_{ij} = \begin{bmatrix} y_i \\ y_j \end{bmatrix}$ and $\boldsymbol{\Sigma}_p$ is the $2 \times 2$ covariance matrix for $y_i$ and $y_j$.

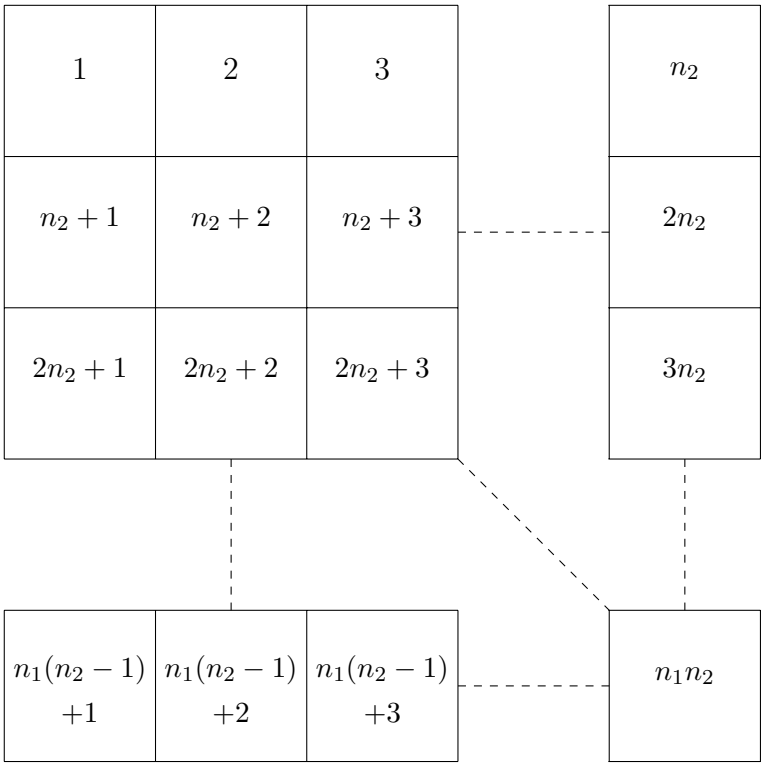



Figure 2: An illustration of the lateral view of the data. For each square in the figure there is an $n_3 \times 1$ column.

In this thesis each data point will be a column in the data, i.e. $\boldsymbol{Y_i} \in \mathbb{R}^{(n_3 \times 1)}$. Figure 2 shows how the data are indexed. Considering the large dimensions of seismic data composite likelihood greatly simplifies calculations because the covariance matrix which is an $(n_1 \cdot n_2 \cdot n_3) \times (n_1 \cdot n_2 \cdot n_3)$, or $n \times n$, matrix is replaced by many $(2 \cdot n_3) \times (2 \cdot n_3)$ matrices. This also requires far less memory since only parts of the covariance matrix needs to be included for each sum. The negative log-composite-likelihood function becomes

$$cl(\boldsymbol{Y};\boldsymbol{\theta}) = \sum_i \sum_j \left( \frac{1}{2} \ln |\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})| + \frac{1}{2} \boldsymbol{Y}_{ij}^T \boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1} \boldsymbol{Y}_{ij} \right) \tag{15}$$

where $\boldsymbol{Y}_{ij} = \begin{bmatrix} \boldsymbol{Y}_i \\ \boldsymbol{Y}_j \end{bmatrix}$ and $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$ are the data vectors corresponding to the data points $i$ and $j$ respectively. $\boldsymbol{\Sigma}_{Yij}$ is the $2 \cdot n_3 \times 2 \cdot n_3$ covariance matrix for $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$.

As mentioned earlier the composite likelihood method is particularly useful when the data are spatially correlated. When the data are spatially correlated the closest data points will affect the covariance the most and one can use only the closest neighbours when summing over all pairs for a given data column $i$. Defining a neighbourhood can be done in different ways and depend on the grid selected for the data and the amount of time available for calculations. The larger the selected neighbourhood is the more time is needed for calculations since the the computation time is dependant on the number of data pairs involved. In this thesis a neighbourhood will mean the square of data columns surrounding data column $i$ from a lateral view point. Looking at Figure 2 this means that the neighbourhood for column $i = n_2 + 2$ is $\mathbb{N} = [1, 2, 3, n_2 + 1, n_2 + 3, 2n_2 + 1, 2n_2 + 2, 2n_2]$.

## 3.3 Optimisation using the Gauss Newton method

It is of interest to find the maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$, for the model selected. In order to find $\hat{\boldsymbol{\theta}}$ an optimisation method is needed. The goal of the optimisation is to find

$$\arg\min_{\boldsymbol{\theta}} \{l(\boldsymbol{Y};\boldsymbol{\theta})\}$$

or in the case of composite likelihood

$$\arg\min_{\boldsymbol{\theta}} \{cl(\boldsymbol{Y};\boldsymbol{\theta})\}$$

where $l(\boldsymbol{Y};\boldsymbol{\theta})$ and $cl(\boldsymbol{Y};\boldsymbol{\theta})$ are given by equation 13 and 15 respectively. The following describes the method used in this thesis, the Gauss Newton method. First it will be explained for the normal likelihood function and then for the composite likelihood function.

The Gauss-Newton method is a Fisher-scoring method that solves non-linear least squares problems [8], finding the minimum. The method uses the score vector defined by

$$\boldsymbol{g} = \frac{\partial l(\boldsymbol{Y})}{\partial \boldsymbol{\theta}}$$

and the Hessian matrix defined by

$$\overline{\boldsymbol{H}} = \frac{\partial^2 l(\boldsymbol{Y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}.$$

Many optimisation algorithms need to calculate the full Hessian matrix. For large data sets this is impractical and time consuming. The Gauss Newton method eliminates the need for calculating second derivatives and instead considers the expected value of the Hessian matrix denoted $\boldsymbol{H}$. Not only does this simplify calculations but it also makes the optimisation method more stable. For more on the Gauss Newton method the reader is referred to [8].

The method takes in an initial guess, $\boldsymbol{\theta}_0$, and then updates the value of $\boldsymbol{\theta}$ for each step in the algorithm using the formula

$$\boldsymbol{\theta}_{m+1} = \boldsymbol{\theta}_m - \psi \boldsymbol{H}_m^{-1} \boldsymbol{g}_m, \tag{16}$$

where $\boldsymbol{g}_m$ is the score vector evaluated for $\boldsymbol{\theta}_m$ and $\boldsymbol{H}_m$ is the expected Hessian evaluated for $\boldsymbol{\theta}_m$

$$\boldsymbol{H} = E\left(\frac{\partial^2 l(\boldsymbol{Y};\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m^2}\right). \tag{17}$$

$\psi$ is called the pessimistic factor. It adjusts the step length of the iterative method in an effort to prevent divergence. A shorter step length will mean slower convergence, but is needed if the system diverges for $\psi = 1$.

To find $\arg\min_\theta\{l(\boldsymbol{Y};\boldsymbol{\theta})\}$, which is equivalent to finding the maximum of the positive log-likelihood function, the derivatives of $l(\boldsymbol{Y};\boldsymbol{\theta})$ with respect to all $\theta_k$'s must be found. In the following Tr refers to the trace of a matrix, the sum of a matrix's diagonal elements. The score function becomes

$$g_k = \frac{\partial l(\boldsymbol{Y};\boldsymbol{\theta})}{\partial \theta_k} = \frac{1}{2}\frac{\partial}{\partial \theta_k}\left(\ln|\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})| + \frac{1}{2}\boldsymbol{Y}^T \frac{\partial}{\partial \theta_k}\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})^{-1}\boldsymbol{Y}\right). \tag{18}$$

Using the relations

$$\begin{aligned}
\frac{\partial}{\partial \theta_k}\ln|\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})| &= Tr\left(\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}_Y(\boldsymbol{\theta})}{\partial \theta_k}\right) \\
\frac{\partial}{\partial \theta_k}\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})^{-1} &= -\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}_Y(\boldsymbol{\theta})}{\partial \theta_k}\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})^{-1}
\end{aligned} \tag{19}$$

found in [8], equation 18 can be written as

$$g_k = \frac{1}{2}Tr\left(\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}_Y(\boldsymbol{\theta})}{\partial \theta_k}\right) - \frac{1}{2}\boldsymbol{Y}^T\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}_Y(\boldsymbol{\theta})}{\partial \theta_k}\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})^{-1}\boldsymbol{Y} \tag{20}$$

yielding the score vector $\boldsymbol{g}$.

The Hessian is often impractical or computationally expensive to calculate and is therefore estimated by the expected Hessian given in equation 17. Each element of the matrix is given by

$$H_{kl} = E\left(\frac{\partial^2 l(\boldsymbol{Y};\boldsymbol{\theta})}{\partial\theta_k \partial\theta_l}\right). \tag{21}$$

The expectation in equation 21 is calculated

$$H_{kl} = \frac{1}{2} Tr\left(\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})}{\partial\theta_k}\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})}{\partial\theta_l}\right).$$

The final iteration step of the Gauss Newton algorithm then becomes as in equation 16.

The algorithm terminates when convergence is reached, when $\boldsymbol{\theta}_{m+1} - \boldsymbol{\theta}_m < t$ where $t$ is some chosen tolerance. The algorithm for the Gauss Newton Method for the full likelihood can be seen in Algorithm 1.

**Algorithm 1** The Gauss Newton algorithm using the full covariance matrix.

**Input:**
data dimensions, $n_1$, $n_2$, $n_3$
initial guess, $\boldsymbol{\theta}_0$
**Do:**
decide on tolerance $t$
set $m = 0$
set $\boldsymbol{H} = 0$
set $\boldsymbol{g} = 0$
**while** $\hat{\boldsymbol{\theta}}_{m+1} - \hat{\boldsymbol{\theta}}_m > t$ **do**
  calculate $\boldsymbol{\Sigma}_Y(\hat{\boldsymbol{\theta}}_m)$, $\frac{\partial\boldsymbol{\Sigma}_Y(\hat{\boldsymbol{\theta}}_m)}{\partial\theta_k}$, $k = 1, ..., length(\boldsymbol{\theta})$
  calculate $g_k = \frac{1}{2}Tr\left(\boldsymbol{\Sigma}_Y^{-1}\frac{\partial\boldsymbol{\Sigma}_Y}{\partial\theta_k}\right) - \frac{1}{2}\boldsymbol{Y}^T\boldsymbol{\Sigma}_Y^{-1}\frac{\partial\boldsymbol{\Sigma}_Y}{\partial\theta_k}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{Y}$ $k = 1, ..., length(\boldsymbol{\theta})$
  calculate $H_{kl} = \frac{1}{2}Tr\left(\boldsymbol{\Sigma}_Y^{-1}\frac{\partial\boldsymbol{\Sigma}_Y}{\partial\theta_k}\boldsymbol{\Sigma}_Y^{-1}\frac{\partial\boldsymbol{\Sigma}_Y}{\partial\theta_l}\right)$, $k, l = 1, ..., length(\boldsymbol{\theta})$
  set $\hat{\boldsymbol{\theta}}_{m+1} = \hat{\boldsymbol{\theta}}_m - \boldsymbol{\Sigma}_{Ym}^{-1}\boldsymbol{g}_m$
**end while**
**Return:**
$\hat{\boldsymbol{\theta}}$

For the composite likelihood described in Section 3.1 the Gauss Newton method is quite similar. The task is now to find $\arg\min_\theta\{cl(\boldsymbol{Y};\boldsymbol{\theta})\}$ and so the score function becomes

$$g_k = \frac{\partial pl}{\partial\theta_k} = \sum_i\sum_j\left(\frac{1}{2}Tr\left(\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})}{\partial\theta_k}\right) - \frac{1}{2}\boldsymbol{Y}_{ij}^T\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})}{\partial\theta_k}\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\boldsymbol{Y}_{ij}\right), \tag{22}$$

where the relations from equation 19 are used. Similarly to the the regular likelihood case equation 22 gives us $\boldsymbol{g}$.

The matrix $\boldsymbol{H}$ is found in a similar way as before and is given by

$$H_{kl} = \sum_i \sum_j \frac{1}{2} Tr \left( \boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})}{\partial \theta_k} \boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})}{\partial \theta_l} \right). \tag{23}$$

The iteration step is the same as in 16 and terminates when $\boldsymbol{\theta}_{m+1} - \boldsymbol{\theta}_m < t$ as before. To save computation time, it is advisable to calculate the $\boldsymbol{\Sigma}_{Yij}$s and $\boldsymbol{H}$s outside the loop that goes through the data pairs. This can be done since only the closest neighbours are used and the expressions for $\boldsymbol{\Sigma}_{Yij}$ and $\boldsymbol{H}$ are only dependant on the distance between the data columns and the vector of parameters $\boldsymbol{\theta}$. $\boldsymbol{\Sigma}_{Yij}$ and $\boldsymbol{H}$ are therefore calculated for all different distances that can occur. The number of different distances will depend on the way the neighbourhood is chosen. The algorithm for the Gauss Newton method using composite likelihood can be seen in Algorithm 2.

**Algorithm 2** The Gauss Newton algorithm for the composite likelihood approach. $\boldsymbol{\Sigma}_{Yij}$ is the $2 \cdot n_3 \times 2 \cdot n_3$ composite covariance matrix for $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$ where only neighbouring pairs are included.

**Input:**
data dimensions, $n_1$, $n_2$, $n_3$
initial guess, $\boldsymbol{\theta}_0$
**Do:**
calculate $\boldsymbol{\Sigma}_{Yij}$ for all data pair distances
decide on tolerance $t$
set $m = 0$
set $\boldsymbol{H} = 0$
set $\boldsymbol{g} = 0$
**while** $\hat{\boldsymbol{\theta}}_{m+1} - \hat{\boldsymbol{\theta}}_m > t$ **do**
  calculate $\boldsymbol{\Sigma}_{Yij} = cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$ and $\frac{\partial \boldsymbol{\Sigma}_{Yij}}{\partial \theta_k}$ for all data pair distances and for all $k$
  calculate $H_{kl} = \frac{1}{2} Tr \left( \boldsymbol{\Sigma}_{Yij}^{-1} \frac{\partial \boldsymbol{\Sigma}_{Yij}}{\partial \theta_k} \boldsymbol{\Sigma}_{Yij}^{-1} \frac{\partial \boldsymbol{\Sigma}_{Yij}}{\partial \theta_l} \right), \quad k, l = 1, ..., length(\boldsymbol{\theta})$ for all data pair distances
  **for** $i$ in $1 : n_1 \cdot n_2$ **do**
    **for** $j$ in $\mathbb{N}$ **do**
      calculate $g_k = g_k + \frac{1}{2} Tr \left( \boldsymbol{\Sigma}_{Yij}^{-1} \frac{\partial \boldsymbol{\Sigma}_{Yij}}{\partial \theta_k} \right) - \frac{1}{2} \boldsymbol{Y}_{ij}^T \boldsymbol{\Sigma}_{Yij}^{-1} \frac{\partial \boldsymbol{\Sigma}_{Yij}}{\partial \theta_k} \boldsymbol{\Sigma}_{Yij}^{-1} \boldsymbol{Y}_{ij}, \quad k = 1, ..., length(\boldsymbol{\theta})$
      add $H_{kl}$ with the appropriate data pair distance to the existing $H_{kl}$
    **end for**
  **end for**
  set $\hat{\boldsymbol{\theta}}_{m+1} = \hat{\boldsymbol{\theta}}_m - \boldsymbol{H}_m^{-1} \boldsymbol{g}_m$
**end while**
**Return:**
$\hat{\boldsymbol{\theta}}$

## 3.4 Properties

According to [9] the composite likelihood estimator is unbiased if the score vector has expected value zero. Using equation 22 and taking the expectation yields

$$\begin{aligned}
E(g_k) &= E\left(\sum_i\sum_j\left(\frac{1}{2}Tr\left(\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})}{\partial\theta_k}\right)-\frac{1}{2}\boldsymbol{y}_{ij}^T\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})}{\partial\theta_k}\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\boldsymbol{y}_{ij}\right)\right)\\
&= \frac{1}{2}\sum_i\sum_j Tr\left(\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})}{\partial\theta_k}\right)-\frac{1}{2}\sum_i\sum_j E\left(\boldsymbol{y}_{ij}^T\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})}{\partial\theta_k}\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\boldsymbol{y}_{ij}\right)\\
&= \frac{1}{2}\sum_i\sum_j Tr\left(\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})}{\partial\theta_k}\right)-\frac{1}{2}\sum_i\sum_j Tr\left(\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})}{\partial\theta_k}\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})^{-1}\boldsymbol{\Sigma}_{Yij}(\boldsymbol{\theta})\right)\\
&= 0.
\end{aligned}$$

Thus the composite likelihood estimator is an unbiased estimator.

The information matrix for the composite likelihood, called the Godambe information, is given by

$$\boldsymbol{I}_G(\boldsymbol{\theta}) = \boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{J}(\boldsymbol{\theta})^{-1}\boldsymbol{H}(\boldsymbol{\theta}) \tag{24}$$

where $\boldsymbol{H}(\boldsymbol{\theta})$ is the expected Hessian and $\boldsymbol{J}(\boldsymbol{\theta})$ is given by $\boldsymbol{J}(\boldsymbol{\theta}) = \mathrm{Var}(\frac{\partial cl(\boldsymbol{\theta},\boldsymbol{Y})}{\partial\boldsymbol{\theta}})$. The vector of parameters $\hat{\boldsymbol{\theta}}$ is asymptotically distributed as $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \boldsymbol{I}_G^{-1})$. For more on the Godambe information the reader is referred to [14] and [7]. In the previous section an estimate for the Hessian was obtained through $\boldsymbol{H}$ from equation 23, so all that is needed to calculate the Godambe information is $\boldsymbol{J}(\boldsymbol{\theta})$, which is given by

$$\begin{aligned}
\boldsymbol{J}(\boldsymbol{\theta}) &= \mathrm{Var}\left(\frac{\partial cl(\boldsymbol{Y};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\\
&= \mathrm{Var}\left(\sum_i\sum_j -\frac{1}{2}\boldsymbol{Y}_{ij}^{\mathrm{T}}\boldsymbol{\Sigma}_{Yij}^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_{Yij}^{-1}\boldsymbol{Y}_{ij}\right)
\end{aligned} \tag{25}$$

where $\boldsymbol{Y}_{ij} = \begin{bmatrix}\boldsymbol{Y}_i\\ \boldsymbol{Y}_j\end{bmatrix}$ is the vector of length $2\cdot n_3$ of $\boldsymbol{Y}$'s that correspond to columns $i$ and $j$. Continuing from the expression in equation (25)

$$\begin{aligned}
\boldsymbol{J}(\boldsymbol{\theta}) = &\sum_{\mathbb{K}} Tr\left(\boldsymbol{\Sigma}_{Yij}^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_{Yij}^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}}{\partial\boldsymbol{\theta}}\right)\\
&+\sum_{\mathbb{K}}\sum_{\mathbb{L}} Tr\left(\begin{bmatrix}\boldsymbol{\Sigma}_{Yij}^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_{Yij}^{-1} & 0\\ 0 & 0\end{bmatrix}\boldsymbol{\Sigma}_{\mathbb{KL}}\begin{bmatrix}0 & 0\\ 0 & \boldsymbol{\Sigma}_{Yij}^{-1}\frac{\partial\boldsymbol{\Sigma}_{Yij}}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_{Yij}^{-1}\end{bmatrix}\boldsymbol{\Sigma}_{\mathbb{KL}}\right).
\end{aligned}$$

Here $\mathbb{K}$ is the set of all pairs in the desired neighbourhood, i.e. the set of all pairs $(i,j)$, $i = 1,...,n_1\cdot n_2$ and $j \in$ [neighbours of $i$]. $\mathbb{L}$ is the set of all pairs that contain one of the values $i$ or $j$ for a given pair in $\mathbb{K}$, or in other words, all neighbouring pairs of the given pair in $\mathbb{K}$. $\boldsymbol{\Sigma}_{\mathbb{KL}}$ is the $4\cdot n_3 \times 4\cdot n_3$ covariance matrix between the two pairs in $\mathbb{K}$ and $\mathbb{L}$. It is now possible to calculate the Godambe information by equation (24).

For the full likelihood $\boldsymbol{J}(\boldsymbol{\theta}) = \boldsymbol{H}(\boldsymbol{\theta})$ and so $\boldsymbol{I}_G(\boldsymbol{\theta}) = \boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{J}(\boldsymbol{\theta})^{-1}\boldsymbol{H}(\boldsymbol{\theta}) = \boldsymbol{H}(\boldsymbol{\theta})$ which demonstrates the difference in the asymptotic variance for the full likelihood parameter estimates and the composite likelihood parameter estimates.

# 4 Prediction

## 4.1 Full likelihood

In a seismic setting it is of interest to predict the variables $\boldsymbol{X}$ using the values of the data $\boldsymbol{Y}$. Being able to predict the values for $\alpha$, $\beta$ and $\rho$ based on the values of the seismic measurements makes finding changes in reservoirs much easier.

In statistical prediction the objective is to obtain values for an unknown entity using the information that is available. Following this trail of thought the optimal predictor for $\boldsymbol{X}_i$ is then the conditional expectation $E(\boldsymbol{X}_i|\boldsymbol{Y})$.

In order to predict $\boldsymbol{X}_i$ for a certain data column it is assumed that the distribution of $\boldsymbol{X}_i$ is known. This assumption makes it possible to write out the covariance matrix for $\boldsymbol{X}_i$ and all the data $\boldsymbol{Y}$

$$\begin{bmatrix} \boldsymbol{X}_i \\ \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_{n_1 n_2} \end{bmatrix} \sim N\left( \begin{bmatrix} \boldsymbol{\mu}_{X_i} = 0 \\ \boldsymbol{\mu}_Y = \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{Xii} & \boldsymbol{\Sigma}_{Xi,\cdot}\boldsymbol{G}^T \\ \boldsymbol{G}\boldsymbol{\Sigma}_{Xi,\cdot} & \boldsymbol{\Sigma}_Y \end{bmatrix} \right) \tag{26}$$

$\boldsymbol{\Sigma}_Y$ is the $n_1 \cdot n_2 \cdot n_3 \times n_1 \cdot n_2 \cdot n_3$ covariance matrix of all the data columns $\boldsymbol{Y}$. $\boldsymbol{\Sigma}_{Xii}$ is the covariance matrix for the covariate column $\boldsymbol{X}_i$. By using the distribution from equation 26 it is now possible to write out the full likelihood for $\boldsymbol{Y}$ and $\boldsymbol{X}_i$

$$l(\boldsymbol{Y}, \boldsymbol{X}_i; \boldsymbol{\theta}) = -\frac{1}{2} \begin{bmatrix} \boldsymbol{X}_i \\ \boldsymbol{Y} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{\Sigma}_{Xii} & \boldsymbol{\Sigma}_{Xi,\cdot}\boldsymbol{G}^T \\ \boldsymbol{G}\boldsymbol{\Sigma}_{Xi,\cdot} & \boldsymbol{\Sigma}_Y \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}_i \\ \boldsymbol{Y} \end{bmatrix}$$

The matrix $\boldsymbol{G}\boldsymbol{\Sigma}_{Xi,\cdot}$ is a $n_3 \times n_1 \cdot n_2 \cdot n_3$ block matrix where each block is the covariance matrix for $\boldsymbol{X}_i$ and $\boldsymbol{Y}_j$ for all values of $j$. Each block is given by

$$\text{cov}(\boldsymbol{X}_i, \boldsymbol{Y}_j) = \text{cov}(\boldsymbol{X}_i, \boldsymbol{G}\boldsymbol{X}_j + \boldsymbol{\epsilon}) = \boldsymbol{G}\,\text{cov}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \boldsymbol{G}\boldsymbol{\Sigma}_{Xij}$$

where $\boldsymbol{\Sigma}_{Xij}$ is the covariance matrix between $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$. Finding the predicted value of $\boldsymbol{X}_i$ is achieved by finding the derivative of the likelihood function with respect to $\boldsymbol{X}_i$ and setting it equal to zero yields the well known result

$$\hat{\boldsymbol{X}}_i = \boldsymbol{\Sigma}_{Xi,\cdot}\boldsymbol{G}^T \cdot \boldsymbol{\Sigma}_Y^{-1}\boldsymbol{Y}$$

Solving this equation for the full likelihood is not possible when $n_1 \cdot n_2 \cdot n_3$ becomes large, and it again becomes necessary to use composite likelihood.

## 4.2 Composite likelihood

The composite likelihood function for the distribution of $\begin{bmatrix} \boldsymbol{X}_i \\ \boldsymbol{Y} \end{bmatrix}$ given in equation 26 is given by

$$\text{cl}(\boldsymbol{Y}, \boldsymbol{X}_i; \boldsymbol{\theta}) = \Sigma_{j \in \mathbb{N}} - \frac{1}{2} \begin{bmatrix} \boldsymbol{X}_i \\ \boldsymbol{Y}_i \\ \boldsymbol{Y}_j \end{bmatrix}^T \begin{bmatrix} \boldsymbol{\Sigma}_{Xii} & \boldsymbol{\Sigma}_{Xii}\boldsymbol{G}^T & \boldsymbol{\Sigma}_{Xij}\boldsymbol{G}^T \\ \boldsymbol{G}\boldsymbol{\Sigma}_{Xii} & \boldsymbol{\Sigma}_{Yii} & \boldsymbol{\Sigma}_{Yij} \\ \boldsymbol{G}\boldsymbol{\Sigma}_{Xij} & \boldsymbol{\Sigma}_{Yij} & \boldsymbol{\Sigma}_{Yjj} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}_i \\ \boldsymbol{Y}_i \\ \boldsymbol{Y}_j \end{bmatrix}$$

where $\boldsymbol{G}\boldsymbol{\Sigma}_{Xij}$ is the $n_3 \times n_3$ covariance matrix for $\boldsymbol{X}_i$ and $\boldsymbol{Y}_j$

$$\text{cov}(\boldsymbol{X}_i, \boldsymbol{Y}_j) = \text{cov}(\boldsymbol{X}_i, \boldsymbol{G}\boldsymbol{X}_j + \boldsymbol{\epsilon}_j) = \boldsymbol{G}\,\text{cov}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \boldsymbol{G}\boldsymbol{\Sigma}_{ij}$$

By first denoting

$$\begin{bmatrix} \boldsymbol{\Sigma}_{Xii} & \boldsymbol{\Sigma}_{Xii}\boldsymbol{G}^T & \boldsymbol{\Sigma}_{Xij}\boldsymbol{G}^T \\ \boldsymbol{G}\boldsymbol{\Sigma}_{Xii} & \boldsymbol{\Sigma}_{Yii} & \boldsymbol{\Sigma}_{Yij} \\ \boldsymbol{G}\boldsymbol{\Sigma}_{Xij} & \boldsymbol{\Sigma}_{Yij} & \boldsymbol{\Sigma}_{Yjj} \end{bmatrix}^{-1} = \boldsymbol{Q}_{ji}$$

it is now possible to find the derivative with respect to $\boldsymbol{X}_i$

$$\frac{\partial cl(\boldsymbol{Y}, \boldsymbol{X}_i; \boldsymbol{\theta})}{\partial \boldsymbol{X}_i} = \Sigma_{j \in \mathbb{N}} \left( \boldsymbol{Q}_{ji}(0)\boldsymbol{X}_i + \boldsymbol{Q}_{ji}(0,1)\boldsymbol{Y}_i + \boldsymbol{Q}_{ji}(0,2)\boldsymbol{Y}_j \right) \tag{27}$$

By setting the expression in equation 27 equal to zero and solving for $\boldsymbol{X}_i$ an expression for $\hat{\boldsymbol{X}}_i$ is obtained

$$\hat{\boldsymbol{X}}_i = -\left[\Sigma_{j \in \mathbb{N}} \boldsymbol{Q}_{ji}(0)\right]^{-1} \Sigma_{j \in \mathbb{N}} (\boldsymbol{Q}_{ji}(0,1)\boldsymbol{Y}_i + \boldsymbol{Q}_{ji}(0,2)\boldsymbol{Y}_j) \tag{28}$$

where $\boldsymbol{Q}_{ji}(0)$, $\boldsymbol{Q}_{ji}(0,1)$ and $\boldsymbol{Q}_{ji}(0,2)$ are the values in the given area of $\boldsymbol{Q}_{ji}$ given by the following pattern

$$\boldsymbol{Q}_{ji} = \begin{bmatrix} \boldsymbol{Q}_{ji}(0) & \boldsymbol{Q}_{ji}(0,1) & \boldsymbol{Q}_{ji}(0,2) \\ \boldsymbol{Q}_{ji}(1,0) & \boldsymbol{Q}_{ji}(1) & \boldsymbol{Q}_{ji}(1,2) \\ \boldsymbol{Q}_{ji}(2,0) & \boldsymbol{Q}_{ji}(2,1) & \boldsymbol{Q}_{ji}(2) \end{bmatrix}.$$

$\hat{\boldsymbol{X}}_i$ can now be predicted for any data column $i$. The algorithm for performing the prediction is given in Algorithm 3

## 4.3 Leave one out prediction

It is also possible to predict $\boldsymbol{X}_i$ when $\boldsymbol{Y}_i$ is taken out of the dataset. This approach is tested to see how well the predictions perform for unknown data. Removing $\boldsymbol{Y}_i$ from the dataset alters the composite likelihood function

$$cl(\boldsymbol{Y}_{-i}, \boldsymbol{X}_i; \boldsymbol{\theta}) = \Sigma_{j \in \mathbb{N}} - \frac{1}{2} \begin{bmatrix} \boldsymbol{X}_i \\ \boldsymbol{Y}_j \end{bmatrix}^T \begin{bmatrix} \boldsymbol{\Sigma}_{Xii} & \boldsymbol{\Sigma}_{Xij}\boldsymbol{G}^T \\ \boldsymbol{G}\boldsymbol{\Sigma}_{Xij} & \Sigma_{Yjj} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{X}_i \\ \boldsymbol{Y}_j \end{bmatrix}$$

**Algorithm 3** The algorithm used for predicting $\hat{\boldsymbol{X}}_i$ and for finding the Godambe information

**Input:**
data dimensions, $n_1$, $n_2$, $n_3$
estimated parameters $\hat{\boldsymbol{\theta}}$
**Do:**
calculate $\boldsymbol{\Sigma}_{Xii}$, $\boldsymbol{\Sigma}_{XYij}$ and $\boldsymbol{\Sigma}_{Yij}$ for all data pair distances
calculate $\boldsymbol{Q}_{ij}(0)$, $\boldsymbol{Q}_{ij}(0,1)$ and $\boldsymbol{Q}_{ij}(0,2)$ for all data pair distances
set $m = 0$
set $\boldsymbol{H} = 0$
set $\boldsymbol{g} = 0$
set $\overline{\boldsymbol{Q}}_{ij}(0) = 0$
set $\overline{\boldsymbol{Q}}_{ij}(0,1) = 0$
set $\overline{\boldsymbol{Q}}_{Yij} = 0$
**for** $i$ in $1 : n_1 \cdot n_2$ **do**
  **for** $j$ in $\mathbb{N}$ **do**
    calculate $\overline{\boldsymbol{Q}}_{ij}(0) = \overline{\boldsymbol{Q}}_{ij}(0) + \boldsymbol{Q}_{ij}(0)$
    calculate $\overline{\boldsymbol{Q}}_{ij}(0,1) = \overline{\boldsymbol{Q}}_{ij}(0,1) + \boldsymbol{Q}_{ij}(0,1)$
    calculate $\overline{\boldsymbol{Q}}_{Yij} = \overline{\boldsymbol{Q}}_{Yij} + \boldsymbol{Q}_{ij}(0,1) \cdot \boldsymbol{Y}_i + \boldsymbol{Q}_{ij}(0,2) \cdot \boldsymbol{Y}_j$
  **end for**
  set $\boldsymbol{H} = \overline{\boldsymbol{Q}}_{ij}(0)$
  set $\boldsymbol{B}_{ij} = [\overline{\boldsymbol{Q}}_{ij}(0), \overline{\boldsymbol{Q}}_{ij}(0,1)]$
  calculate $\boldsymbol{J} = \boldsymbol{B}_{ij}\boldsymbol{\Sigma}_{XYii}\boldsymbol{B}_{ij}^T$
  **for** $j$ in $\mathbb{N}$ **do**
    calculate $\boldsymbol{J} = \boldsymbol{J} + \boldsymbol{B}_{ij}[\boldsymbol{\Sigma}_{XYij}, \boldsymbol{\Sigma}_{Yij}]\boldsymbol{Q}_{ij}^T(0,2)$
    **for** $j'$ in $\mathbb{N}$ **do**
      calculate $\boldsymbol{J} = \boldsymbol{J} + \boldsymbol{Q}_{ij}(0,2)\boldsymbol{\Sigma}_{Yjj'}\boldsymbol{Q}_{ij}^T(0,2)$
    **end for**
  **end for**
  set $\hat{\boldsymbol{X}}_i = -\overline{\boldsymbol{Q}}_{ij}^{-1}(0) \cdot \overline{\boldsymbol{Q}}_{Yij}$
  set $\boldsymbol{I}_{Gi} = \boldsymbol{H}\boldsymbol{J}^{-1}\boldsymbol{H}$
**end for**
**Return:**
$\hat{\boldsymbol{X}}$
$\boldsymbol{I}_G$

where $\mathbb{N}$ is the set of desired neighbours for column $i$. Similar to the previous section the inverse covariance matrix is denoted

$$\begin{bmatrix} \boldsymbol{\Sigma}_{Xii} & \boldsymbol{\Sigma}_{Xij}\boldsymbol{G}^T \\ \boldsymbol{G}\boldsymbol{\Sigma}_{Xij} & \Sigma_{Yjj} \end{bmatrix}^{-1} = \boldsymbol{Q}_{ji}$$

As before the derivative of $cl(\boldsymbol{Y}, \boldsymbol{X}_i; \boldsymbol{\theta})$ is found

$$\frac{\partial cl(\boldsymbol{Y}, \boldsymbol{X}_i; \boldsymbol{\theta})}{\partial \boldsymbol{X}_i} = \Sigma_{j\in\mathbb{N}} \left( \boldsymbol{Q}_{ji}(0)\boldsymbol{X}_i - \boldsymbol{Q}_{ji}(0,1)\boldsymbol{Y}_j \right) \tag{29}$$

A predicted value for $\boldsymbol{X}_i$ can now be found by setting the derivative equal to zero and solving for $\boldsymbol{X}_i$

$$\hat{\boldsymbol{X}}_i = -\left[\Sigma_{j\in\mathbb{N}}\boldsymbol{Q}_{ji}(0)\right]^{-1} \Sigma_{j\in\mathbb{N}}\boldsymbol{Q}_{ji}(0,1)\boldsymbol{Y}_j \tag{30}$$

$\boldsymbol{Q}_{ji}(0)$ and $\boldsymbol{Q}_{ji}(0,1)$ are the values in the given area of $\boldsymbol{Q}_{ji}$ given by the following pattern

$$\boldsymbol{Q}_{ji} = \begin{bmatrix} \boldsymbol{Q}_{ji}(0) & \boldsymbol{Q}_{ji}(0,1) \\ \boldsymbol{Q}_{ji}(1,0) & \boldsymbol{Q}_{ji}(1) \end{bmatrix}.$$

$\hat{\boldsymbol{X}}_i$ can now be predicted for any data column $i$.

## 4.4 Sandwich estimation of prediction variance

The sandwich estimate, or the Godambe information, as described in Section 3.4 can be used to estimate the variance of the prediction result. As mentioned the Godambe information is given by

$$\boldsymbol{I}_G = \boldsymbol{H}\boldsymbol{J}^{-1}\boldsymbol{H}. \tag{31}$$

To find $\boldsymbol{G}$ expressions for $\boldsymbol{H}$ and $\boldsymbol{J}$ are needed. These are given by

$$\boldsymbol{H} = E\left(\frac{\partial^2 cl(\boldsymbol{Y}, \boldsymbol{X}_i; \boldsymbol{\theta})}{\partial \boldsymbol{X}_i^2}\right)$$
$$\boldsymbol{J} = \mathrm{Var}\left(\frac{\partial cl(\boldsymbol{Y}, \boldsymbol{X}_i; \boldsymbol{\theta})}{\partial \boldsymbol{X}_i}\right)$$

The expressions for $\boldsymbol{H}$ and $\boldsymbol{J}$ will be slightly different depending on which prediction technique is chosen, but the procedure for finding them is the same. To find $\boldsymbol{H}$ it is necessary to find the the second derivative of the composite likelihood function and then taking the expectation. This is achieved by finding the derivative of the composite likelihood function, given in equation 27 or 29 depending on the prediction technique,

with respect to $\boldsymbol{X}_i$. For equation 29 $\boldsymbol{H}$ is given by

$$\begin{aligned}\boldsymbol{H} &= E\left(\frac{\partial^2 cl(\boldsymbol{Y}, \boldsymbol{X}_i; \boldsymbol{\theta})}{\partial \boldsymbol{X}_i^2}\right) \\ &= E\left(\frac{\partial\left(\frac{\partial cl(\boldsymbol{Y}, \boldsymbol{X}_i; \boldsymbol{\theta})}{\partial \boldsymbol{X}_i}\right)}{\partial \boldsymbol{X}_i}\right) \\ &= E\left(\frac{\partial}{\partial \boldsymbol{X}_i}\Sigma_{j\in\mathbb{N}}\left(\boldsymbol{Q}_{ji}(0)\boldsymbol{X}_i - \boldsymbol{Q}_{ji}(0,1)\boldsymbol{Y}_j\right)\right) \\ &= \Sigma_{j\in\mathbb{N}}\boldsymbol{Q}_{ji}(0)\end{aligned}$$

The result is exactly the same for equation 27 since the only difference between the two expressions are terms not related to $\boldsymbol{X}_i$.

To find $\boldsymbol{J}$ the variance of the derivative of the composite likelihood function with respect to $\boldsymbol{X}_i$, given in equation 27 or 29 depending on the prediction technique, needs to be found. For equation 29 $\boldsymbol{J}$ is given by

$$\begin{aligned}\boldsymbol{J} =& \text{var}\left(\frac{\partial \text{cl}}{\partial \boldsymbol{X}_i}\right) \\ =& \text{var}\left(\Sigma_{j\in\mathbb{N}}\boldsymbol{Q}_{ji}(2)\boldsymbol{X}_i - \boldsymbol{Q}_{ji}(2,1)\boldsymbol{Y}_j\right) \\ =& \overline{\boldsymbol{Q}}_{ij}(0)\text{var}(\boldsymbol{X}_i)\overline{\boldsymbol{Q}}_{ij}(0)^T \\ &+\Sigma_{j\in\mathbb{N}}\overline{\boldsymbol{Q}}_{ij}(0)\text{cov}(\boldsymbol{X}_i, \boldsymbol{Y}_j)\boldsymbol{Q}_{ij}(0,1)^T \\ &+\Sigma_{j\in\mathbb{N}}\Sigma_{j'\in\mathbb{N}}\boldsymbol{Q}_{ij}(0,1)\text{cov}(\boldsymbol{Y}_j, \boldsymbol{Y}_{j'})\boldsymbol{Q}_{ij'}(0,1)^T\end{aligned}$$

where $\overline{\boldsymbol{Q}}_{ij}(0) = \Sigma_{j\in\mathbb{N}}\boldsymbol{Q}_{ij}(0)$.

For equation 27 the expression is quite similar, but some elements are altered. $\overline{\boldsymbol{Q}}_{ij}(0)$ is replaced with the matrix $\boldsymbol{B}_{ij} = \left[\overline{\boldsymbol{Q}}_{ij}(0), \overline{\boldsymbol{Q}}_{ij}(0,1)\right]$ where $\overline{\boldsymbol{Q}}_{ij}(0,1) = \Sigma_{j\in\mathbb{N}}\boldsymbol{Q}_{ij}(0,1)$. $\boldsymbol{X}_i$ is replaced by $\begin{bmatrix}\boldsymbol{X}_i \\ \boldsymbol{Y}_i\end{bmatrix}$ resulting in the following expression for $\boldsymbol{J}$

$$\begin{aligned}\boldsymbol{J} =& \boldsymbol{B}_{ij}\text{var}\left(\begin{bmatrix}\boldsymbol{X}_i \\ \boldsymbol{Y}_i\end{bmatrix}\right)\boldsymbol{B}_{ij}^T \\ &+\Sigma_{j\in\mathbb{N}}\boldsymbol{B}_{ij}\text{cov}\left(\begin{bmatrix}\boldsymbol{X}_i \\ \boldsymbol{Y}_i\end{bmatrix}, \boldsymbol{Y}_j\right)\boldsymbol{Q}_{ij}(0,2)^T \\ &+\Sigma_{j\in\mathbb{N}}\Sigma_{j'\in\mathbb{N}}\boldsymbol{Q}_{ij}(0,2)\text{cov}(\boldsymbol{Y}_j, \boldsymbol{Y}_{j'})\boldsymbol{Q}_{ij'}(0,2)^T\end{aligned}$$

The resulting Godambe information $\boldsymbol{I}_G = \boldsymbol{H}\boldsymbol{J}^{-1}\boldsymbol{H}$ is, when the inverse is taken, the asymptotic covariance matrix of the predicted $\hat{\boldsymbol{X}}_i$ as discussed in section 3.4.

## 4.5 Prediction residuals

In order to assess the results of the predictions it is interesting to look at the residuals

$$\hat{\boldsymbol{r}}_i = \boldsymbol{Y}_i - \boldsymbol{G}_i\hat{\boldsymbol{X}}_i. \tag{32}$$

The residuals should not exceed the estimated noise term $\hat{\boldsymbol{\epsilon}}$ and should represent random noise.

As described in subsection 4.4 it is possible to calculate the Godambe information for the predictions. Using the Godambe information it is possible to create prediction intervals for the residuals. The interval for a given data column is found using

$$-\text{abs}\left(1.96\sqrt{\text{diag}(\boldsymbol{G}\boldsymbol{I}_{Gi}^{-1}\boldsymbol{G}^T+\boldsymbol{\epsilon})}\right) \leq \hat{\boldsymbol{r}}_i \leq \text{abs}\left(1.96\sqrt{\text{diag}(\boldsymbol{G}\boldsymbol{I}_{Gi}^{-1}\boldsymbol{G}^T+\boldsymbol{\epsilon})}\right) \tag{33}$$

where abs means taking the absolute value and diag means the diagonal of the given matrix. Approximately 95% of the residuals $\hat{\boldsymbol{r}}_i$ should lie within this interval.

# 5 Data Set from Norne

The data set is gathered from the Norne field, located in the North Sea off the coast of Norway. See Figure 3 for a map showing the location. The base survey was performed

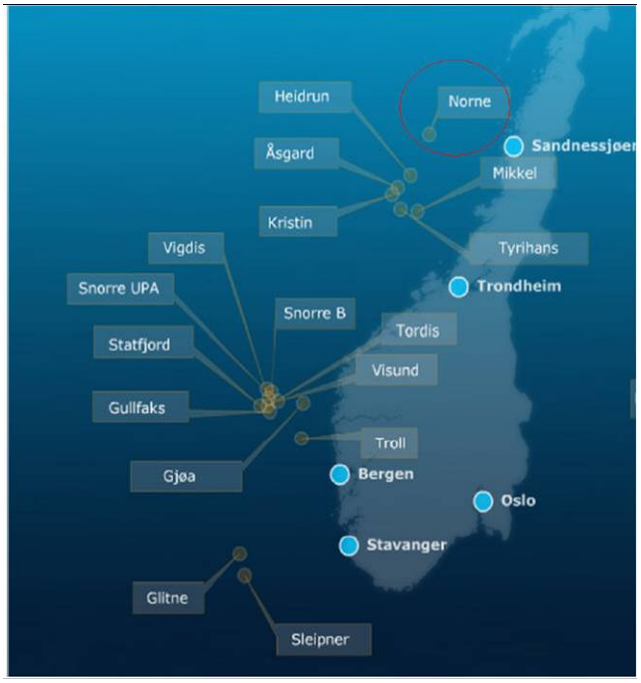



Figure 3: A map showing the location of the Norne Field

in 2001 and the monitor survey in 2003. For each survey data is gathered for different angles, which are placed into three categories, near, mid and far. Approximate values for each angle category is needed to calculate the expressions in equation 2.3, and are set to 5°, 15° and 30° for near, mid and far respectively. A plot of the data for the three angles from the 2001 survey can be seen in Figure 4. A similar plot for the 2003 survey can be seen in Figure 5. In this data set only the depth and the inline directions are included, thus making this a 2D data set with two measurement times. The response variable $\boldsymbol{Y}$ is the difference between the 2003 and 2001 survey as discussed in Section 2.4. A plot of the difference between the two surveys can be seen in Figure 6. The dimensions of the data are $150 \times 100$ where 150 is the dimension of the depth and 100 is the dimension in the inline direction, meaning that $n_1 = 100$, $n_2 = 1$ and $n_3 = 150$. The reservoir is located at approximately depth 90-110 and can be seen best for the far angle in Figure 6. The noise at depth 60 is not related to the reservoir, but is rather noise the processing was not able to remove and is referred to as processing artefacts.

Histograms for the three different angles are made to see if the assumption of normally distributed data seems reasonable and are shown in Figure 7. It is clear from the histograms that a normal distribution for the data is an acceptable choice. The model selected for the data is based on the more general model described in Section 2.4. The

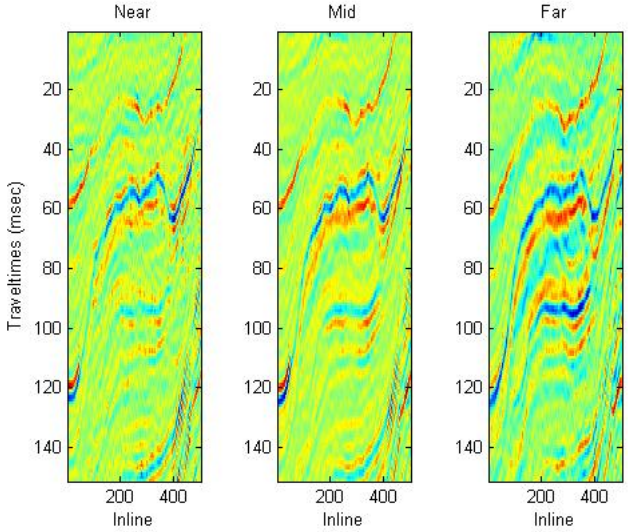



Figure 4: Plot of the seismic data from the survey performed in 2001.

model for each data column is given by

$$\boldsymbol{Y}_i = \boldsymbol{G}\boldsymbol{X}_i + \boldsymbol{\epsilon}_i. \tag{34}$$

The prior of $\boldsymbol{X}_i$ is

$$\boldsymbol{X}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_X). \tag{35}$$

and the marginal distribution of $\boldsymbol{Y}_i$ is then given by

$$\boldsymbol{Y}_i \sim N(\boldsymbol{0}, \boldsymbol{G}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{G}^T + \boldsymbol{\Sigma}_{\epsilon i})$$

$\boldsymbol{G} = \boldsymbol{W}\boldsymbol{A}$ is the design matrix based on the behaviour of seismic waves described in section 2.3. $\boldsymbol{W}$ is based on well logs from the Norne field, where the estimated wavelet function is equal for all three angles and can be seen in Figure 8. The covariance matrix $\boldsymbol{\Sigma}_X$ is assumed to be spatially correlated and also needs to reflect the covariance between the elastic parameters. In order to match the build of the $\boldsymbol{X}_i$ vector described in equation 11 the covariance matrix of $\boldsymbol{X}_i$ needs to be of the form

$$\boldsymbol{\Sigma}_X = \boldsymbol{\Sigma}_0 \otimes \boldsymbol{C}_{ii} \tag{36}$$

where $\boldsymbol{\Sigma}_0$ is the covariance matrix of the elastic parameters and $\boldsymbol{C}_{ii}$ is the correlation matrix for $\boldsymbol{X}_i$. The $\otimes$ operator in equation 36 refers to the Kronecker product which works by multiplying the last matrix in the Kronecker product in every value of the

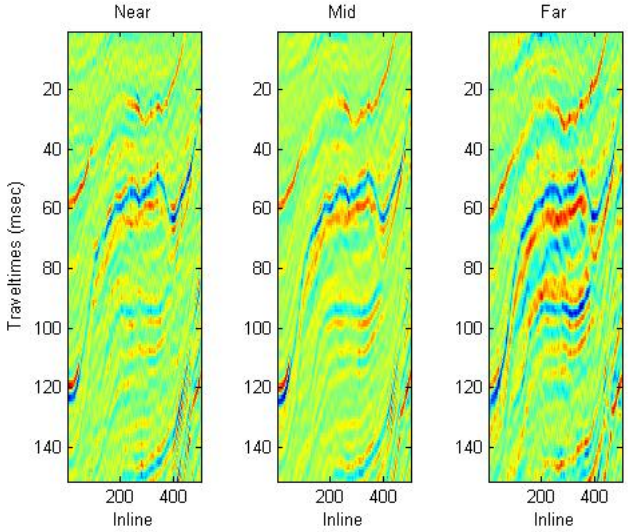



Figure 5: Plot of the seismic data from the survey performed in 2003

first matrix. The Kronecker product is illustrated in the small example below

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{11} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{bmatrix}$$

The derivative of a Kronecker product is given by

$$(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{B}' \tag{37}$$

$\boldsymbol{\Sigma}_0$ is the covariance matrix of the elastic parameters given by

$$\boldsymbol{\Sigma_0} = \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} & \sigma_{\alpha\rho} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 & \sigma_{\beta\rho} \\ \sigma_{\alpha\rho} & \sigma_{\beta\rho} & \sigma_\rho^2 \end{bmatrix}$$

In order to estimate the entries in $\Sigma_0$ a parametrisation of the covariance matrix of $\alpha$, $\beta$ and $\rho$ is needed and is chosen to be

$$\hat{\boldsymbol{\Sigma}}_0 = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix} \tag{38}$$

$\boldsymbol{C}$ is a distance dependant correlation matrix, meaning that the correlation between data points is determined by the distance between the data points through a correlation function. This will be discussed in the following.

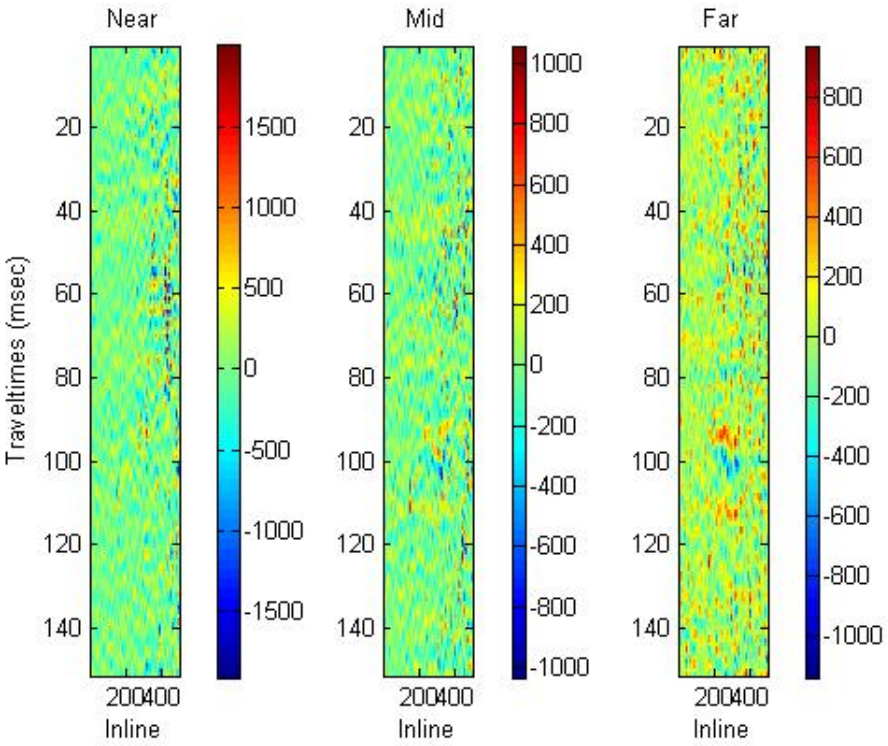



Figure 6: Plot of the difference between the 2003 survey and the 2001 survey

By trying different variations of the model described above it can be seen which one gives the best estimates of parameters and the best predictions. There are three aspects of the model that will vary.

## 5.1 Correlation functions

The correlation matrix of the elastic parameters $\boldsymbol{C}$ is assumed to be dependent on the distance between the data, i.e. for each pair of data points $p$ and $q$ the correlation between them is dependant on the distance between them. There are many correlation functions dependent on distance. One of the most common is the exponential correlation function given by

$$c(p,q;\phi) = \exp\left(\frac{-3|p-q|}{\phi}\right) \tag{39}$$

Another well known correlation function is the Matern correlation function

$$c(p,q;\phi) = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(2\sqrt{\nu}\frac{|p-q|}{\phi}\right)^{\nu} K_\nu\left(2\sqrt{\nu}\frac{|p-q|}{\phi}\right)$$

where $\nu$ is a smoothing parameter, $\Gamma(\nu)$ the gamma function and $K_\nu$ is a modified Bessel function as described in [2]. A special case is when $\nu = \frac{3}{2}$ and the Matern covariance

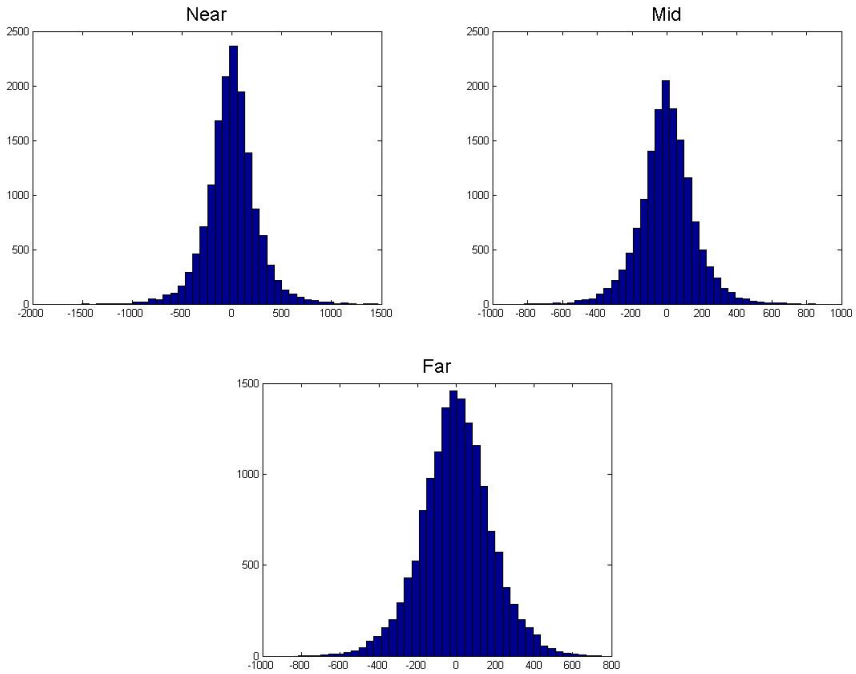



Figure 7: Histograms of the data for each angle. Apart from heavy tails for the near angle, the assumption of normally distributed data seems reasonable.

function becomes

$$c(p,q;\phi,\nu=\frac{3}{2}) = (1+|p-q|\frac{3}{\phi})\exp\left(\frac{-3|p-q|}{\phi}\right) \tag{40}$$

For both correlation functions $\phi$ is parameter giving the range of the correlation. For more on distance dependent correlation functions and additional examples the reader is referred to [2]. In the model variations the correlation function will vary between the exponential correlation function from equation 39 and the Matern correlation function with $\nu = \frac{3}{2}$ from equation 40. Figure 9 shows how the two correlation functions behave as a function of distance when $\phi = 3$ for the exponential correlation function and $\phi = 2$ for the Matern correlation function.

## 5.2 The noise term $\epsilon$

The second aspect of the model that varies is related to the noise term. There are many ways to model noise, but this thesis will focus on two models for the noise term

$$\hat{\boldsymbol{\epsilon}}_1 \sim N(0, \boldsymbol{\tau}_1^2 \otimes \boldsymbol{I}) \tag{41}$$

$$\hat{\boldsymbol{\epsilon}}_2 \sim N(0, \boldsymbol{\tau}_2^2 \otimes \boldsymbol{I} + \boldsymbol{W}(\boldsymbol{\tau}_3^2 \otimes \boldsymbol{I})\boldsymbol{W}^T) \tag{42}$$

where $\boldsymbol{\tau}_1^2$, $\boldsymbol{\tau}_2^2$ and $\boldsymbol{\tau}_3^2$ are $3 \times 3$ diagonal matrices and $\boldsymbol{W}$ is the matrix described in Section 2.3. The reason they are $3 \times 3$ matrices is because the noise is estimated

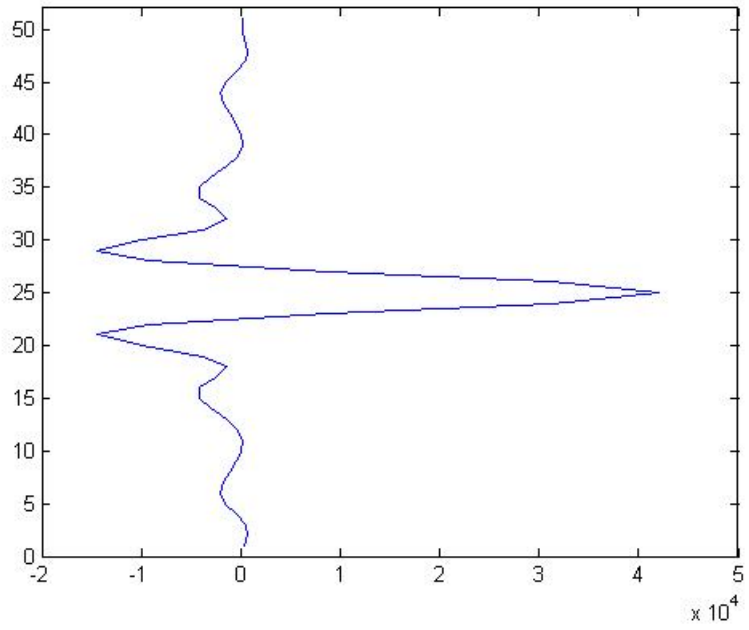



Figure 8: The wavelet function used in $\boldsymbol{W}$

separately for the three different angles. The first model is simple and uses a single noise term equal for all the data points for a given angle. The second model expands on the first by adding correlation in the noise depending on the wavelet matrix $\boldsymbol{W}$.

The values of the diagonal elements in $\boldsymbol{\tau}_1^2$ are not found using the Gauss Newton iteration technique, but rather estimated using the data themselves. This is done by assuming that there has been little or no change in the top layers of the data, usually referred to as the overburden, and that the changes found here can be attributed to noise. For this data set the number of layers used for this estimation is 40. Using this assumption the $\tau_1$ parameter for each of the angles can be estimated by finding the empirical variance of the top levels from the data. The resulting $\hat{\boldsymbol{\tau}}$ is then a $3 \times 3$ matrix with the variance for each angle on the diagonal and zero elsewhere.

When finding $\boldsymbol{\tau}_2^2$ and $\boldsymbol{\tau}_3^2$ a simple linear regression problem is solved. The elements of the two matrices are found for each of the three angles separately. First the empirical covariance matrix for the overburden is found for the first angle and then placed in a vector, $\boldsymbol{c}$. This is done by taking out the first row of the matrix and then appending each of the following rows of the covariance matrix. The same is done for the $\boldsymbol{W} \cdot \boldsymbol{W}^T$ product for the first 40 rows. This is now set equal to the corresponding term containing

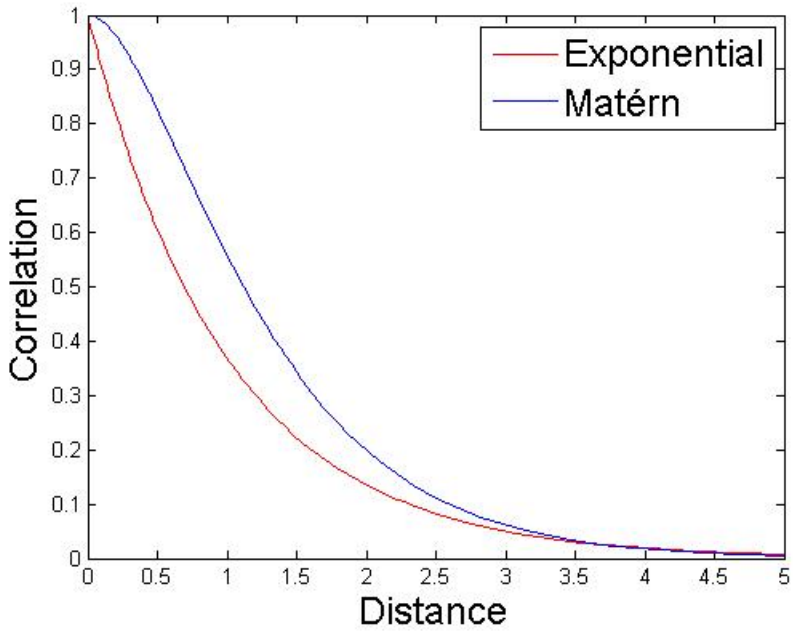



Figure 9: The exponential and Matern correlation functions with $\phi = 3$ and $\phi = 2$ respectively

$\tau^2_{2\theta_p}$ and $\tau^2_{3\theta_p}$ given by the following equation

$$\boldsymbol{c} = \begin{bmatrix} c_{1,1} \\ c_{1,2} \\ \vdots \\ c_{1,40} \\ c_{2,1} \\ c_{2,2} \\ \vdots \\ c_{2,40} \\ c_{40,1} \\ c_{40,2} \\ \vdots \\ c_{40,40} \end{bmatrix} = \begin{bmatrix} 1 & w_{1,1} \\ 0 & w_{1,2} \\ \vdots & \vdots \\ 0 & w_{1,40} \\ 0 & w_{2,1} \\ 1 & w_{2,2} \\ \vdots & \vdots \\ 0 & w_{2,40} \\ 0 & w_{40,1} \\ 0 & w_{40,2} \\ \vdots & \vdots \\ 1 & w_{40,40} \end{bmatrix} \begin{bmatrix} \tau^2_{2\omega_j} \\ \tau^2_{3\omega_j} \end{bmatrix}$$

The first column of the matrix with ones and zeros ensures that $\tau^2_{2\omega_p}$ is included for all the variance terms. The system can now be solved using the least squares method to obtain estimates for $\tau_{1\omega_p}$ and $\tau_{2\omega_p}$ and this is repeated for all three angles.

For the data set from Norne the values for $\boldsymbol{\tau}^2_1$, $\boldsymbol{\tau}^2_2$ and $\boldsymbol{\tau}^2_3$ are

$$\boldsymbol{\tau}^2_1 = 10^4 \begin{bmatrix} 6.079 & 0 & 0 \\ 0 & 1.884 & 0 \\ 0 & 0 & 2.955 \end{bmatrix} \tag{43}$$

$$\boldsymbol{\tau}_2^2 = 10^4 \begin{bmatrix} 2.299 & 0 & 0 \\ 0 & 1.206 & 0 \\ 0 & 0 & 1.002 \end{bmatrix} \tag{44}$$

$$\boldsymbol{\tau}_3^2 = 10^{-5} \begin{bmatrix} 0.805 & 0 & 0 \\ 0 & 0.253 & 0 \\ 0 & 0 & 0.405 \end{bmatrix} \tag{45}$$

The values agree well with prior experience in seismic reflection that the noise is highest for the near angle. This is mostly due to disturbances from the sensors and also from waves being reflected twice. When the seismic waves reach the sensor, some of that wave will be reflected back to the subsurface and back again, much weaker, but still strong enough to create noise in the measurements.

## 5.3 Number of data columns used for each $i$

The last aspect of the model to vary concerns the data column $\boldsymbol{Y}_i$. In Section 2.3 $\boldsymbol{Y}_i$ is given as a single data column as shown in equation 10. An alternative is to select $\boldsymbol{Y}_i$ as multiple data columns, i.e. each $\boldsymbol{Y}_i$ is a small neighbourhood of columns, primarily $2 \times 2$ columns in the 3D case. For the 2D case studied here this means $\boldsymbol{Y}_i$ consists of two data columns. Both alternatives are shown for the 2D case in Figure 10. The resulting $\boldsymbol{Y}_i$, from now on denoted $\boldsymbol{Y}_i^2$, is

$$\boldsymbol{Y}_i^2 = \begin{bmatrix} y_{i\omega_1 1} \\ y_{i\omega_1 2} \\ \vdots \\ y_{i\omega_1 n_3} \\ \vdots \\ y_{i\omega_P 1} \\ y_{i\omega_P 2} \\ \vdots \\ y_{i\omega_P n_3} \\ y_{i+1\omega_1 1} \\ y_{i+1\omega_1 2} \\ \vdots \\ y_{i+1\omega_1 n_3} \\ \vdots \\ y_{i+1\omega_P 1} \\ y_{i+1\omega_P 2} \\ \vdots \\ y_{i+1\omega_P n_3} \end{bmatrix} \tag{46}$$

This results in the number of $\boldsymbol{Y}_i^2$'s being cut in half compared to the $\boldsymbol{Y}_i$s, but the dimensions of the covariance matrix for each $\boldsymbol{Y}_i^2$ will double in both directions in comparison

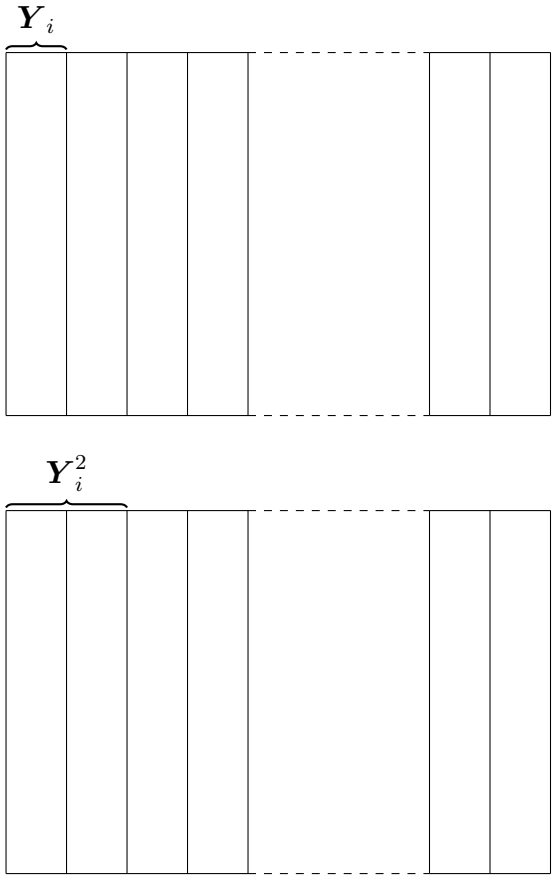



Figure 10: An illustration of the data columns and the effect of having a model with $\boldsymbol{Y}$ or $\boldsymbol{Y}^2$

thus increasing computation time. The covariance matrix of $\boldsymbol{Y}_i^2$ is

$$\begin{bmatrix} \boldsymbol{G}(\boldsymbol{\Sigma_0} \otimes \boldsymbol{C}_{ii})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma_0} \otimes \boldsymbol{C}_{i,i+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon \\ \boldsymbol{G}(\boldsymbol{\Sigma_0} \otimes \boldsymbol{C}_{i,i+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma_0} \otimes \boldsymbol{C}_{i+1,i+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon \end{bmatrix}$$

It is worth noting that the size of covariance matrix of $\boldsymbol{Y}_i^2$ is four times larger than the covariance matrix of $\boldsymbol{Y}_i$ and will therefore cause computation time to increase.

# 6 Parameter estimation and prediction results

## 6.1 Parameter estimation

The three varying aspects of the model all have two variations, meaning all combinations of variations yield $2^3 = 8$ different models. For easier representation of the results in this section each model is given a name that reflects the different variations it contains. The eight models are

$$
\begin{aligned}
&M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{exp}) \\
&M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{exp}) \\
&M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{Matern}) \\
&M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern}) \\
&M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{exp}) \\
&M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{exp}) \\
&M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{Matern}) \\
&M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})
\end{aligned}
$$

For each of the models $M$ the chosen alternative for the three variations is described in the parenthesis. $\boldsymbol{Y}$ or $\boldsymbol{Y}^2$ refers to whether $\boldsymbol{Y}_i$ consists of one or two data columns respectively. $\boldsymbol{\tau}_1^2$ means that the noise term is only the diagonal matrix from equation 41, while $\boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2$ means the noise term also includes a wavelet convoluted term as in equation 42. Exp or Matern gives the correlation function used in the model.

Having defined the eight different models it is desirable to estimate the vector of parameters $\boldsymbol{\theta}$ using the Gauss Newton method. The parameter vector $\boldsymbol{\theta}$ contains seven parameters regardless of which model is being used and is given by

$$
\boldsymbol{\theta} = \begin{bmatrix} \phi \\ \sigma_1^2 \\ \sigma_2^2 \\ \sigma_3^2 \\ \rho_{12} \\ \rho_{13} \\ \rho_{23} \end{bmatrix}
$$

$\phi$ is the range parameter from the correlation function used and is described in section 5.1. The $\sigma_i^2$ and the $\rho_{ij}$ make up the covariance matrix of the underlying parameters as shown in equation 38. The covariance matrices of $\boldsymbol{Y}$, $\boldsymbol{\Sigma}_{Yij}$ for $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$ and $\boldsymbol{\Sigma}_{Yij}^2$ for $\boldsymbol{Y}_i^2$ and $\boldsymbol{Y}_j^2$ need to be found. $\boldsymbol{\Sigma}_{Yij}^2$ refers to the covariance matrix of two $\boldsymbol{Y}_i^2$s. These covariance matrices are given by

$$
\boldsymbol{\Sigma}_{Yij} = \begin{bmatrix} \boldsymbol{G}(\boldsymbol{\Sigma_0} \otimes \boldsymbol{C}_{ii})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma_0} \otimes \boldsymbol{C}_{ij})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon \\ \boldsymbol{G}(\boldsymbol{\Sigma_0} \otimes \boldsymbol{C}_{ij})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma_0} \otimes \boldsymbol{C}_{jj})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon \end{bmatrix}
$$

and

$$\boldsymbol{\Sigma}^2_{Yij} = \begin{bmatrix} \boldsymbol{G}\boldsymbol{\Sigma}_{Xii}\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xi,i+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xi,j})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xi+1,j+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon \\ \boldsymbol{G}(\boldsymbol{\Sigma}_{Xi,i+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xi+1,i+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xi+1,j})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xi+1,j+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon \\ \boldsymbol{G}(\boldsymbol{\Sigma}_{Xi,j})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xi+1,j})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xjj})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xj,j+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon \\ \boldsymbol{G}(\boldsymbol{\Sigma}_{Xi,j+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xi+1,j+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xj,j+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon & \boldsymbol{G}(\boldsymbol{\Sigma}_{Xj+1,j+1})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon \end{bmatrix}$$

When implementing the Gauss Newton method derivatives of these matrices with respect to all parameters in $\boldsymbol{\theta}$ are needed. In order to find these derivatives the expression $\boldsymbol{G}(\boldsymbol{\Sigma_0} \otimes \boldsymbol{C}_{ij})\boldsymbol{G}^T + \boldsymbol{\Sigma}_\epsilon$ needs to be differentiated with respect to each of the seven parameters. The noise terms will always become zero since they are independent of all the parameters. The $\boldsymbol{G}$ matrix is also independent of all the parameters and will therefore remain constant during differentiation. The task is therefore restricted to finding the derivative of $\boldsymbol{\Sigma_0} \otimes \boldsymbol{C}_{ij}$ with respect to all the parameters. From equation 37 the derivative of $\boldsymbol{\Sigma_0} \otimes \boldsymbol{C}_{ij}$ is $\frac{\partial \boldsymbol{\Sigma_0}}{\partial \theta_k} \otimes \boldsymbol{C}_{ij} + \boldsymbol{\Sigma_0} \otimes \frac{\boldsymbol{C}_{ij}}{\partial \theta_k}$. That means that in order to use the Gauss Newton method expressions for $\frac{\partial \boldsymbol{C}}{\partial \theta_k}$ and $\frac{\partial \boldsymbol{\Sigma_0}}{\partial \theta_k}$. $\boldsymbol{C}_{ij}$ are only dependent on $\phi$ so all that is needed is finding $\frac{\partial \boldsymbol{C}_{ij}}{\partial \phi}$. Instead of using $\phi$ when running the algorithm $\phi' = 3/\phi$ is used for simplicity and stability. Using this substitution the derivative of the correlation functions become

$$\begin{aligned} \frac{\partial}{\partial \phi'} \exp\left(-|p-q|\phi'\right) &= -|p-q| \exp\left(|p-q|\phi'\right) \\ \frac{\partial}{\partial \phi'}(1 + |p-q|\phi') \exp\left(-|p-q|\phi'\right) &= |p-q|^2 \phi' \exp\left(-|p-q|\phi'\right) \end{aligned} \tag{47}$$

and from these $\frac{\partial \boldsymbol{C}}{\partial \phi'}$ can easily be found. From equation 38 it is easy to find the expressions for $\frac{\partial \boldsymbol{\Sigma_0}}{\partial \theta_k}$

$$\begin{aligned} \frac{\partial \hat{\boldsymbol{\Sigma}}_0}{\partial \sigma_1^2} &= \begin{bmatrix} 1 & \frac{\rho_{12}\sigma_2}{2\sigma_1} & \frac{\rho_{13}\sigma_3}{2\sigma_1} \\ \frac{\rho_{12}\sigma_2}{2\sigma_1} & 0 & 0 \\ \frac{\rho_{13}\sigma_3}{2\sigma_1} & 0 & 0 \end{bmatrix} \quad \frac{\partial \hat{\boldsymbol{\Sigma}}_0}{\partial \rho_{12}} = \begin{bmatrix} 0 & \sigma_1\sigma_2 & 0 \\ \sigma_1\sigma_2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ \frac{\partial \hat{\boldsymbol{\Sigma}}_0}{\partial \sigma_2^2} &= \begin{bmatrix} 0 & \frac{\rho_{12}\sigma_1}{2\sigma_2} & 0 \\ \frac{\rho_{12}\sigma_1}{2\sigma_2} & 1 & \frac{\rho_{23}\sigma_3}{2\sigma_2} \\ 0 & \frac{\rho_{23}\sigma_3}{2\sigma_2} & 0 \end{bmatrix} \quad \frac{\partial \hat{\boldsymbol{\Sigma}}_0}{\partial \rho_{13}} = \begin{bmatrix} 0 & 0 & \sigma_1\sigma_3 \\ 0 & 0 & 0 \\ \sigma_1\sigma_3 & 0 & 0 \end{bmatrix} \\ \frac{\partial \hat{\boldsymbol{\Sigma}}_0}{\partial \sigma_3^2} &= \begin{bmatrix} 0 & 0 & \frac{\rho_{13}\sigma_1}{2\sigma_3} \\ 0 & 0 & \frac{\rho_{23}\sigma_2}{2\sigma_3} \\ \frac{\rho_{13}\sigma_1}{2\sigma_3} & \frac{\rho_{23}\sigma_2}{2\sigma_3} & 1 \end{bmatrix} \quad \frac{\partial \hat{\boldsymbol{\Sigma}}_0}{\partial \rho_{23}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \sigma_2\sigma_3 \\ 0 & \sigma_2\sigma_3 & 0 \end{bmatrix} \end{aligned} \tag{48}$$

Using the definition of the derivative of a Kronecker product given in equation 37 the

Table 1: Table of the estimated $\hat{\boldsymbol{\theta}}$ for the four models with one data column

| | $M(\boldsymbol{Y};\boldsymbol{\tau}_1^2;\text{exp})$ | $M(\boldsymbol{Y};\boldsymbol{\tau}_1^2;\text{Matern})$ | $M(\boldsymbol{Y};\boldsymbol{\tau}_2^2,\boldsymbol{\tau}_3^2;\text{exp})$ | $M(\boldsymbol{Y};\boldsymbol{\tau}_2^2,\boldsymbol{\tau}_3^2;\text{Matern})$ |
|---|---|---|---|---|
| $\phi$ | 3.4562 | 1.5981 | 2.2876 | 1.2898 |
| $\sigma_1^2$ | $9.2866 \cdot 10^{-4}$ | $9.0358 \cdot 10^{-4}$ | $1.0860 \cdot 10^{-3}$ | $1.0269 \cdot 10^{-3}$ |
| $\sigma_2^2$ | $7.6670 \cdot 10^{-4}$ | $7.8589 \cdot 10^{-4}$ | $9.9114 \cdot 10^{-4}$ | $9.4554 \cdot 10^{-4}$ |
| $\sigma_3^2$ | $1.1760 \cdot 10^{-3}$ | $1.1300 \cdot 10^{-3}$ | $1.2779 \cdot 10^{-3}$ | $1.2063 \cdot 10^{-3}$ |
| $\rho_{12}$ | 0.9338 | 0.9365 | 0.9584 | 0.9587 |
| $\rho_{13}$ | -0.9909 | -0.9910 | -0.9910 | -0.9910 |
| $\rho_{23}$ | -0.9097 | -0.9114 | -0.9286 | -0.9291 |

Table 2: Table of the estimated $\hat{\boldsymbol{\theta}}$ for the four models with two data columns

| | $M(\boldsymbol{Y}^2;\boldsymbol{\tau}_1^2;\text{exp})$ | $M(\boldsymbol{Y}^2;\boldsymbol{\tau}_1^2;\text{Matern})$ | $M(\boldsymbol{Y}^2;\boldsymbol{\tau}_2^2,\boldsymbol{\tau}_3^2;\text{exp})$ | $M(\boldsymbol{Y}^2;\boldsymbol{\tau}_2^2,\boldsymbol{\tau}_3^2;\text{Matern})$ |
|---|---|---|---|---|
| $\phi$ | 3.8342 | 1.6579 | 3.0513 | 1.4438 |
| $\sigma_1^2$ | $1.0657 \cdot 10^{-3}$ | $9.9404 \cdot 10^{-4}$ | $9.7163 \cdot 10^{-4}$ | $9.6120 \cdot 10^{-4}$ |
| $\sigma_2^2$ | $9.5925 \cdot 10^{-4}$ | $9.2527 \cdot 10^{-4}$ | $8.2247 \cdot 10^{-4}$ | $8.5664 \cdot 10^{-4}$ |
| $\sigma_3^2$ | $1.3155 \cdot 10^{-3}$ | $1.2189 \cdot 10^{-3}$ | $1.1678 \cdot 10^{-3}$ | $1.1405 \cdot 10^{-3}$ |
| $\rho_{12}$ | 0.9379 | 0.9396 | 0.9550 | 0.9573 |
| $\rho_{13}$ | -0.9908 | -0.9909 | -0.9904 | -0.9908 |
| $\rho_{23}$ | -0.9100 | -0.9120 | -0.9257 | -0.9282 |

final expressions for the derivatives are given by

$$\begin{aligned}
\frac{\partial \hat{\boldsymbol{\Sigma}}_{Yij}}{\partial \phi'} &= \boldsymbol{G}(\boldsymbol{\Sigma_0} \otimes \frac{\partial \boldsymbol{C_{ij}}}{\partial \phi'})\boldsymbol{G}^T \\
\frac{\partial \hat{\boldsymbol{\Sigma}}_{Yij}}{\partial \sigma_1^2} &= \boldsymbol{G}(\frac{\partial \boldsymbol{\Sigma_0}}{\partial \sigma_1^2} \otimes \boldsymbol{C_{ij}})\boldsymbol{G}^T \\
\frac{\partial \hat{\boldsymbol{\Sigma}}_{Yij}}{\partial \sigma_2^2} &= \boldsymbol{G}(\frac{\partial \boldsymbol{\Sigma_0}}{\partial \sigma_2^2} \otimes \boldsymbol{C_{ij}})\boldsymbol{G}^T \\
\frac{\partial \hat{\boldsymbol{\Sigma}}_{Yij}}{\partial \sigma_3^2} &= \boldsymbol{G}(\frac{\partial \boldsymbol{\Sigma_0}}{\partial \sigma_3^2} \otimes \boldsymbol{C_{ij}})\boldsymbol{G}^T \\
\frac{\partial \hat{\boldsymbol{\Sigma}}_{Yij}}{\partial \rho_{12}} &= \boldsymbol{G}(\frac{\partial \boldsymbol{\Sigma_0}}{\partial \rho_{12}^2} \otimes \boldsymbol{C_{ij}})\boldsymbol{G}^T \\
\frac{\partial \hat{\boldsymbol{\Sigma}}_{Yij}}{\partial \rho_{13}} &= \boldsymbol{G}(\frac{\partial \boldsymbol{\Sigma_0}}{\partial \rho_{13}^2} \otimes \boldsymbol{C_{ij}})\boldsymbol{G}^T \\
\frac{\partial \hat{\boldsymbol{\Sigma}}_{Yij}}{\partial \rho_{23}} &= \boldsymbol{G}(\frac{\partial \boldsymbol{\Sigma_0}}{\partial \rho_{23}^2} \otimes \boldsymbol{C_{ij}})\boldsymbol{G}^T
\end{aligned}$$

where the expressions from 47 and 48 are used.

The optimisation is run using $\psi = 0.4$ and over 40 steps to ensure convergence. After 40 steps the changes in the parameters are of the order $10^{-5}$, and even lower for the $\sigma_{ij}$, for all eight models. The results of the Gauss Newton runs done for the eight models

can be seen in Table 1 and Table 2. The tables show that the different models return fairly similar parameters and especially the correlation coefficients are very similar. This means the p-wave speed and the s-wave speed are positively correlated and the density is negatively correlated with both the p-wave speed and the s-wave speed. The correlation is very high in absolute value for all three parameters, especially the correlation between s-wave velocity and the density. Explaining the correlation between the three elastic parameters is complicated with many factors involved. There are still certain things that can be looked at to help explain the results. In section 2.3 the relationships between $\alpha$, $\beta$ and $\rho$ are given as

$$\alpha = \sqrt{\frac{K + \frac{4}{3}\mu}{\rho}}$$
$$\beta = \sqrt{\frac{\mu}{\rho}}$$

and looking at these expressions it seems only reasonable that the correlation coefficients gain the values they do. Both $\alpha$ and $\beta$ are both function of the inverse of the square root of $\rho$.

From the results it appears that the choice of correlation function has little impact on the estimates. Apart from the value of $\phi$ which naturally changes due to the differences in the expressions for the correlation, the other parameters have more or less the same value.

Comparing the models with $\boldsymbol{Y}_i$ to the models with $\boldsymbol{Y}_i^2$ it is clear that having $\boldsymbol{Y}_i$ results in greater correlation range for all the models, especially for the models with exponential correlation function. With regard to the variance estimates they appear to increase when using the simple noise term, but decrease when when the convoluted noise term is used.

## 6.2 Asymptotic variance of parameter estimations

The Godambe information described in section 3.4 is together with the parameter estimations themselves calculated for each of the models. Using the Godambe information matrix as described in section 3.4 and its asymptotic properties gives an estimate of the variance in the parameter estimation. By finding the inverse of the Godambe information matrix an estimate for the covariance matrix of the parameters are found and it is then possible to construct confidence intervals for the parameters. To find the 95% confidence intervals for each parameter 1.96 times the estimated standard deviation estimated by the Godambe information matrix is added and subtracted to give the interval. The results can be seen in Tables 3 to 6 with 95% confidence interval for each of the parameters in each of the eight models.

Table 3: Table of the 95% confidence intervals of $\hat{\boldsymbol{\theta}}$ for $M(\boldsymbol{Y};\boldsymbol{\tau}_1^2;\exp)$ and $M(\boldsymbol{Y};\boldsymbol{\tau}_1^2;\text{Matern})$

| $M(\boldsymbol{Y};\boldsymbol{\tau}_1^2;\exp)$ | $M(\boldsymbol{Y};\boldsymbol{\tau}_1^2;\text{Matern})$ |
|---|---|
| $3.2492 \leq \phi \leq 3.6914$ | $1.5330 \leq \phi \leq 1.6690$ |
| $6.512 \cdot 10^{-4} \leq \sigma_1^2 \leq 1.206 \cdot 10^{-3}$ | $6.471 \cdot 10^{-4} \leq \sigma_1^2 \leq 1.160 \cdot 10^{-3}$ |
| $3.639 \cdot 10^{-4} \leq \sigma_2^2 \leq 1.169 \cdot 10^{-3}$ | $4.132 \cdot 10^{-4} \leq \sigma_2^2 \leq 1.159 \cdot 10^{-3}$ |
| $8.970 \cdot 10^{-4} \leq \sigma_3^2 \leq 1.455 \cdot 10^{-3}$ | $8.720 \cdot 10^{-4} \leq \sigma_3^2 \leq 1.388 \cdot 10^{-3}$ |
| $0.9158 \leq \rho_{12} \leq 0.9518$ | $0.9190 \leq \rho_{12} \leq 0.9540$ |
| $-0.9926 \leq \rho_{13} \leq -0.9892$ | $-0.9926 \leq \rho_{13} \leq -0.9894$ |
| $-0.9296 \leq \rho_{23} \leq -0.8898$ | $-0.9306 \leq \rho_{23} \leq -0.8922$ |

Table 4: Table of the 95% confidence intervals of $\hat{\boldsymbol{\theta}}$ for $M(\boldsymbol{Y};\boldsymbol{\tau}_2^2,\boldsymbol{\tau}_3^2;\exp)$ and $M(\boldsymbol{Y};\boldsymbol{\tau}_2^2,\boldsymbol{\tau}_3^2;\text{Matern})$

| $M(\boldsymbol{Y};\boldsymbol{\tau}_2^2,\boldsymbol{\tau}_3^2;\exp)$ | $M(\boldsymbol{Y};\boldsymbol{\tau}_2^2,\boldsymbol{\tau}_3^2;\text{Matern})$ |
|---|---|
| $2.1035 \leq \phi \leq 2.5070$ | $1.2149 \leq \phi \leq 1.3745$ |
| $7.800 \cdot 10^{-4} \leq \sigma_1^2 \leq 1.392 \cdot 10^{-3}$ | $7.380 \cdot 10^{-4} \leq \sigma_1^2 \leq 1.316 \cdot 10^{-3}$ |
| $5.535 \cdot 10^{-4} \leq \sigma_2^2 \leq 1.429 \cdot 10^{-3}$ | $5.324 \cdot 10^{-4} \leq \sigma_2^2 \leq 1.359 \cdot 10^{-3}$ |
| $9.711 \cdot 10^{-4} \leq \sigma_3^2 \leq 1.585 \cdot 10^{-3}$ | $9.166 \cdot 10^{-4} \leq \sigma_3^2 \leq 1.496 \cdot 10^{-3}$ |
| $0.9468 \leq \rho_{12} \leq 0.9700$ | $0.9471 \leq \rho_{12} \leq 0.9703$ |
| $-0.9928 \leq \rho_{13} \leq -0.9891$ | $-0.9929 \leq \rho_{13} \leq -0.9892$ |
| $-0.9458 \leq \rho_{23} \leq -0.9115$ | $-0.9462 \leq \rho_{23} \leq -0.9120$ |

Table 5: Table of the 95% confidence intervals of $\hat{\boldsymbol{\theta}}$ for $M(\boldsymbol{Y}^2;\boldsymbol{\tau}_1^2;\exp)$ and $M(\boldsymbol{Y}^2;\boldsymbol{\tau}_1^2;\text{Matern})$

| $M(\boldsymbol{Y}^2;\boldsymbol{\tau}_1^2;\exp)$ | $M(\boldsymbol{Y}^2;\boldsymbol{\tau}_1^2;\text{Matern})$ |
|---|---|
| $3.5287 \leq \phi \leq 4.1976$ | $1.5751 \leq \phi \leq 1.7499$ |
| $6.510 \cdot 10^{-4} \leq \sigma_1^2 \leq 1.480 \cdot 10^{-3}$ | $6.172 \cdot 10^{-4} \leq \sigma_1^2 \leq 1.371 \cdot 10^{-3}$ |
| $3.576 \cdot 10^{-4} \leq \sigma_2^2 \leq 1.561 \cdot 10^{-3}$ | $3.774 \cdot 10^{-4} \leq \sigma_2^2 \leq 1.473 \cdot 10^{-3}$ |
| $8.983 \cdot 10^{-4} \leq \sigma_3^2 \leq 1.733 \cdot 10^{-3}$ | $8.399 \cdot 10^{-4} \leq \sigma_3^2 \leq 1.598 \cdot 10^{-3}$ |
| $0.9134 \leq \rho_{12} \leq 0.9624$ | $0.9156 \leq \rho_{12} \leq 0.9636$ |
| $-0.9931 \leq \rho_{13} \leq -0.9885$ | $-0.9932 \leq \rho_{13} \leq -0.9886$ |
| $-0.9381 \leq \rho_{23} \leq -0.8820$ | $-0.9395 \leq \rho_{23} \leq -0.8844$ |

Table 6: Table of the 95% confidence intervals of $\hat{\boldsymbol{\theta}}$ for $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \exp)$ and $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$

| $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \exp)$ | $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$ |
|---|---|
| $2.7500 \leq \phi \leq 3.4267$ | $1.3443 \leq \phi \leq 1.5593$ |
| $5.297 \cdot 10^{-4} \leq \sigma_1^2 \leq 1.414 \cdot 10^{-3}$ | $5.505 \cdot 10^{-4} \leq \sigma_1^2 \leq 1.372 \cdot 10^{-3}$ |
| $1.896 \cdot 10^{-4} \leq \sigma_2^2 \leq 1.455 \cdot 10^{-3}$ | $2.686 \cdot 10^{-4} \leq \sigma_2^2 \leq 1.445 \cdot 10^{-3}$ |
| $7.247 \cdot 10^{-4} \leq \sigma_3^2 \leq 1.611 \cdot 10^{-3}$ | $7.287 \cdot 10^{-4} \leq \sigma_3^2 \leq 1.552 \cdot 10^{-3}$ |
| $0.9368 \leq \rho_{12} \leq 0.9731$ | $0.9400 \leq \rho_{12} \leq 0.9747$ |
| $-0.9934 \leq \rho_{13} \leq -0.9875$ | $-0.9936 \leq \rho_{13} \leq -0.9880$ |
| $-0.9527 \leq \rho_{23} \leq -0.8988$ | $-0.9536 \leq \rho_{23} \leq -0.9028$ |

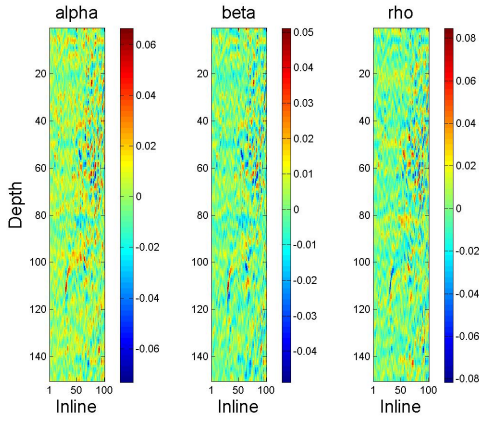
## 6.3 Composite prediction

Using the approach described in Section 4 predictions for the elastic parameters $\boldsymbol{X}_i$ are found and are denoted $\hat{\boldsymbol{X}}_i$. The parameters found in the previous section are used in their respective prediction models. To visualise the results four plots are made for each model. First a plot of the predicted $\hat{\boldsymbol{X}}_i$ for all $i$ is shown. The second plot is a plot of the residuals, $\boldsymbol{Y}_i - \boldsymbol{G}\hat{\boldsymbol{X}}_i$ for all $i$. The last two plots show the lower and upper bounds of the predicted $\hat{\boldsymbol{X}}_i$. The bounds are given by

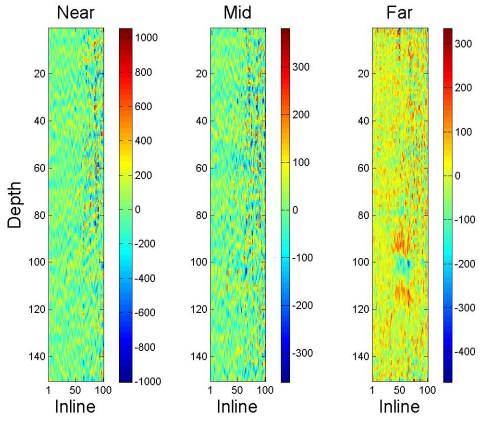
$$\hat{\boldsymbol{X}}_i \pm 1.96\sqrt{\text{diag}(\boldsymbol{I}_G^{-1})}.$$

Apart from the columns at each end of the data set, the prediction variance is the same for all the data columns. It varies slightly with the depth, so for each model the mean of the prediction variance is found for $\alpha$, $\beta$ and $\rho$. The prediction variance varies little for each elastic parameter and the mean gives a good and easy presentable indication of the magnitude of the uncertainty for the prediction results together with the plots.

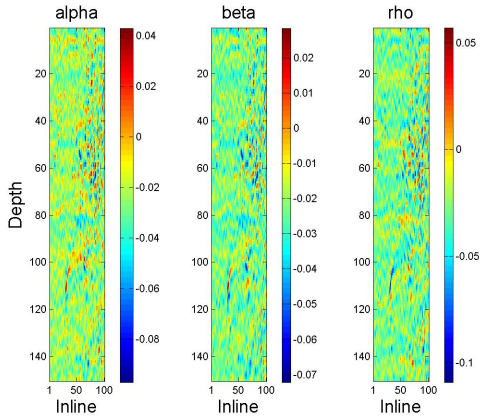
First the plots and prediction variances for each of the eight models are presented. This is followed by analysis and discussion, focusing on the performance of the composite likelihood method and comparison of the different models.



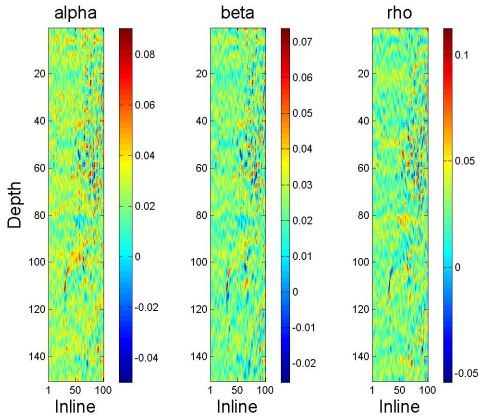

(a) The predicted $\hat{\boldsymbol{X}}_i$ for all $i$




(b) The residuals for all $i$




(c) The 95% prediction interval lower bound as given by the Godambe information

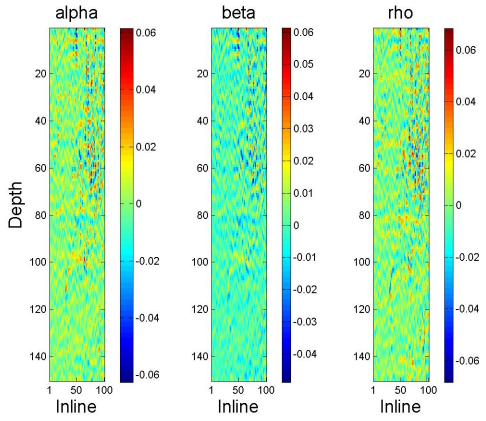



(d) The 95% prediction interval upper bound as given by the Godambe information

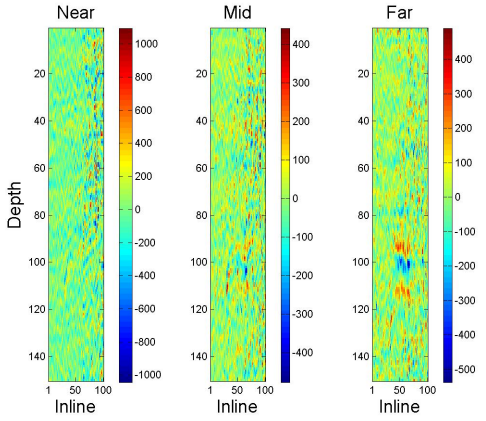
Figure 11: Prediction plots for the model $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{exp})$

The plots in Figure 11 show the predicted $\hat{\boldsymbol{X}}_i$, the residuals, and the upper and lower bound of the predictions for all $i$. The model used is $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{exp})$ which means the model uses single data columns for each $i$, the noise is not wavelet convoluted and the correlation function used is the exponential correlation function.

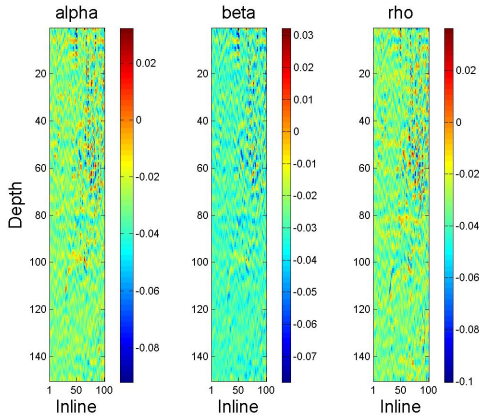
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.569 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.542 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 0.774 \cdot 10^{-3}$.



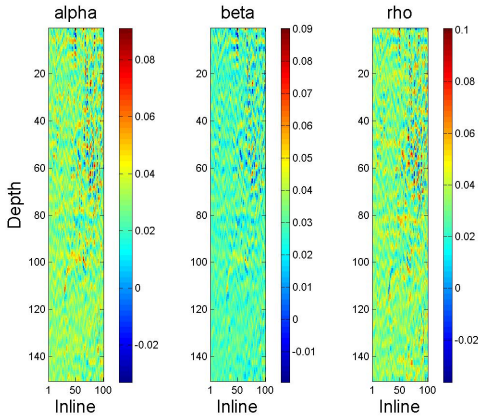

(a) The predicted $\hat{\boldsymbol{X}}_i$ for all $i$




(b) The residuals for all $i$




(c) The 95% prediction interval lower bound as given by the Godambe information

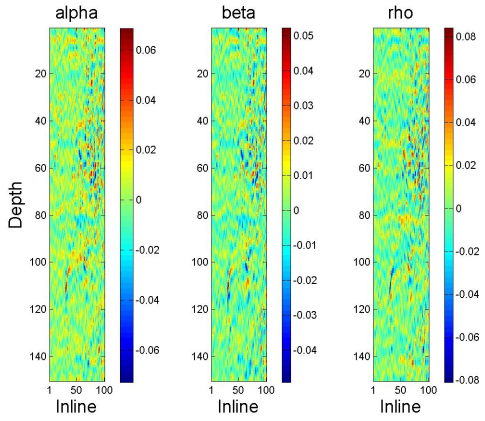



(d) The 95% prediction interval upper bound as given by the Godambe information

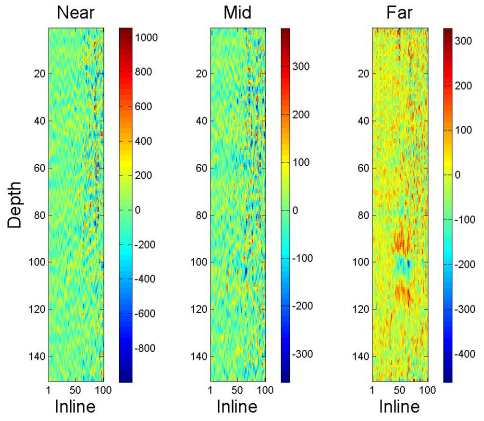
Figure 12: Prediction plots for the model $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \exp)$

The plots in Figure 12 show the predicted $\hat{\boldsymbol{X}}_i$, the residuals, and the upper and lower bound of the predictions for all $i$. The model used is $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \exp)$ which means the model uses single data columns for each $i$, the noise is wavelet convoluted and the correlation function used is the exponential correlation function.

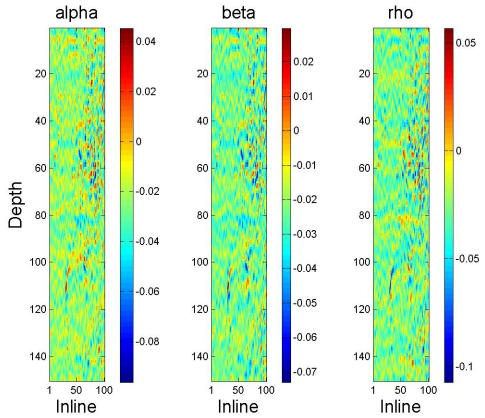
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.892 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.870 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.057 \cdot 10^{-3}$.



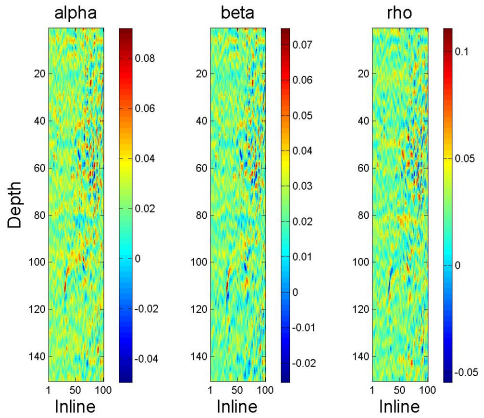

(a) The predicted $\hat{\boldsymbol{X}}_i$ for all $i$




(b) The residuals for all $i$




(c) The 95% prediction interval lower bound as given by the Godambe information

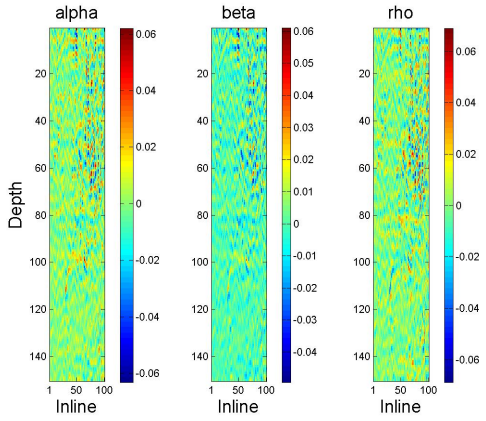



(d) The 95% prediction interval upper bound as given by the Godambe information

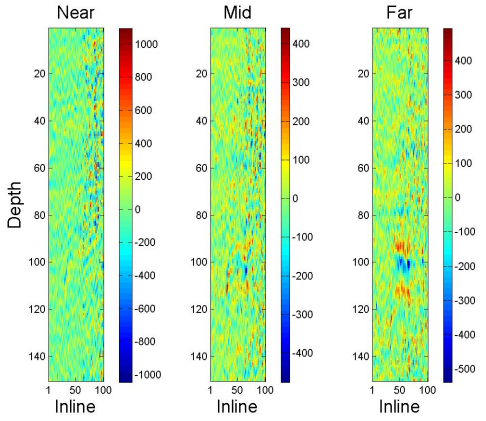
Figure 13: Prediction plots for the model $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{Matern})$

The plots in Figure 13 show the predicted $\hat{\boldsymbol{X}}_i$, the residuals, and the upper and lower bound of the predictions for all $i$. The model used is $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{Matern})$ which means the model uses single data columns for each $i$, the noise is not wavelet convoluted and the correlation function used is the Matern correlation function.

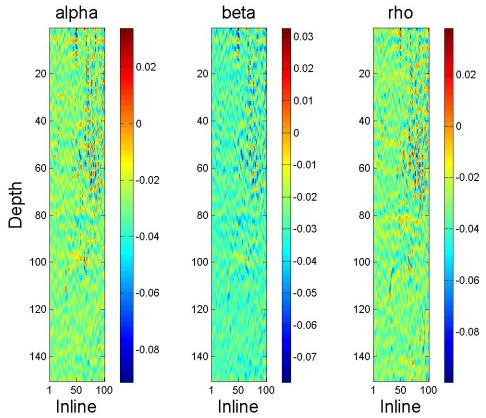
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.547 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.545 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 0.723 \cdot 10^{-3}$.



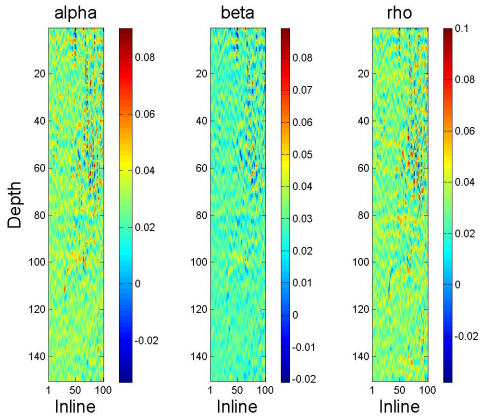

(a) The predicted $\hat{\boldsymbol{X}}_i$ for all $i$




(b) The residuals for all $i$




(c) The 95% prediction interval lower bound as given by the Godambe information

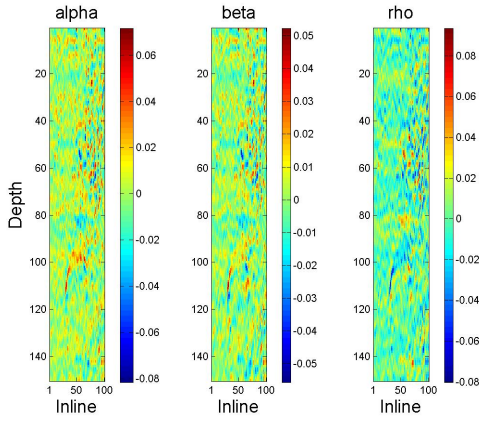



(d) The 95% prediction interval upper bound as given by the Godambe information

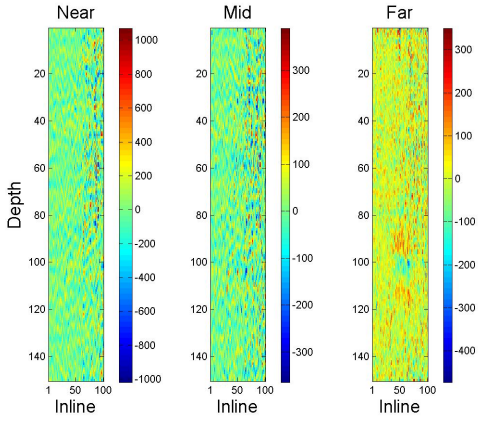
Figure 14: Prediction plots for the model $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$

The plots in Figure 14 show the predicted $\hat{\boldsymbol{X}}_i$, the residuals, and the upper and lower bound of the predictions for all $i$. The model used is $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$ which means the model uses single data columns for each $i$, the noise is wavelet convoluted and the correlation function used is the Matern correlation function.

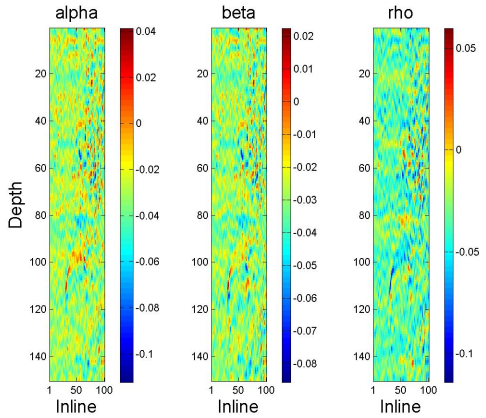
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.822 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.815 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 0.977 \cdot 10^{-3}$.



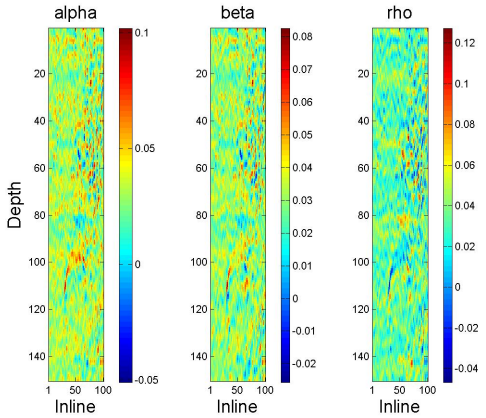

(a) The predicted $\hat{\boldsymbol{X}}_i$ for all $i$




(b) The residuals for all $i$




(c) The 95% prediction interval lower bound as given by the Godambe information

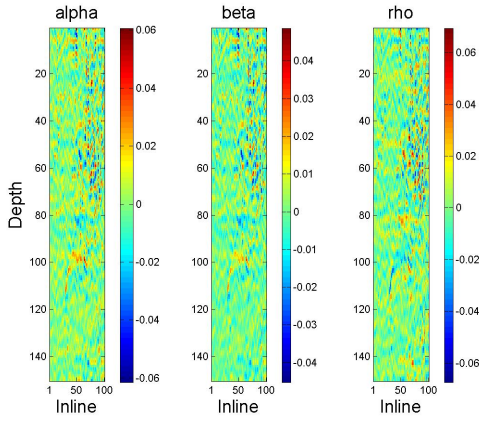



(d) The 95% prediction interval upper bound as given by the Godambe information

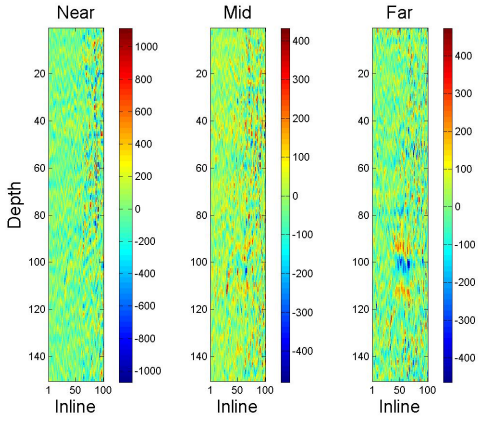
Figure 15: Prediction plots for the model $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \exp)$

The plots in Figure 15 show the predicted $\hat{\boldsymbol{X}}_i$, the residuals, and the upper and lower bound of the predictions for all $i$. The model used is $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \exp)$ which means the model uses two data columns for each $i$ as discussed in Section 5.3 the noise is not wavelet convoluted and the correlation function used is the exponential correlation function.

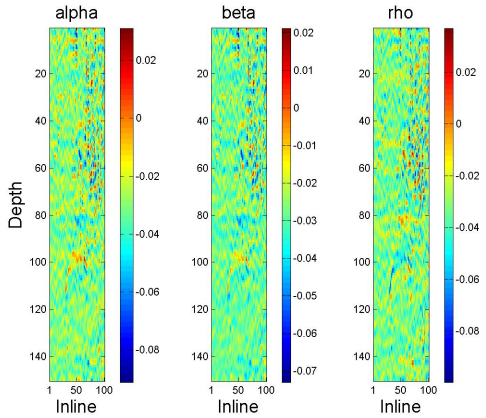
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.954 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.902 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.153 \cdot 10^{-3}$.



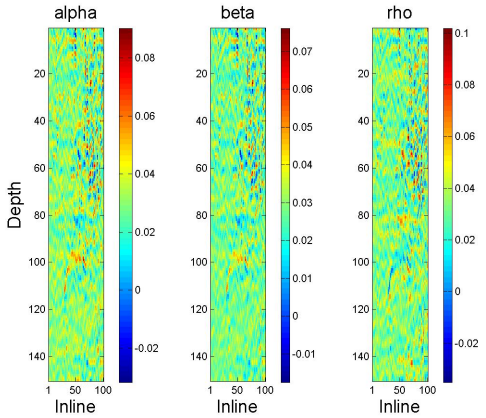

(a) The predicted $\hat{\boldsymbol{X}}_i$ for all $i$




(b) The residuals for all $i$




(c) The 95% prediction interval lower bound as given by the Godambe information

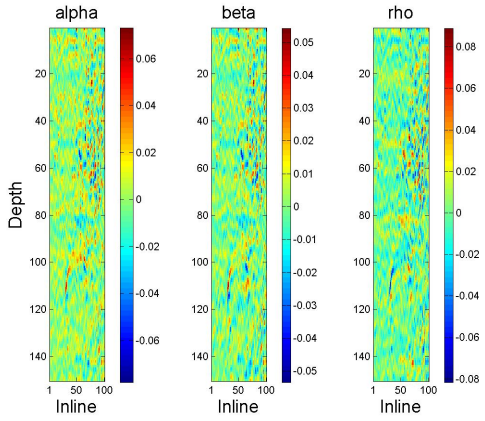



(d) The 95% prediction interval upper bound as given by the Godambe information

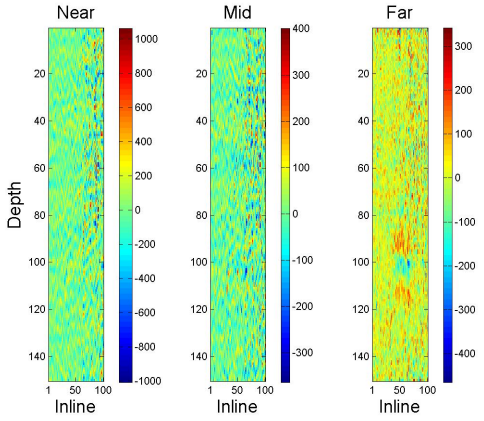
Figure 16: Prediction plots for the model $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \exp)$

The plots in Figure 16 show the predicted $\hat{\boldsymbol{X}}_i$, the residuals, and the upper and lower bound of the predictions for all $i$. The model used is $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \exp)$ which means the model uses two data columns for each $i$ as discussed in Section 5.3 the noise is wavelet convoluted and the correlation function used is the exponential correlation function.

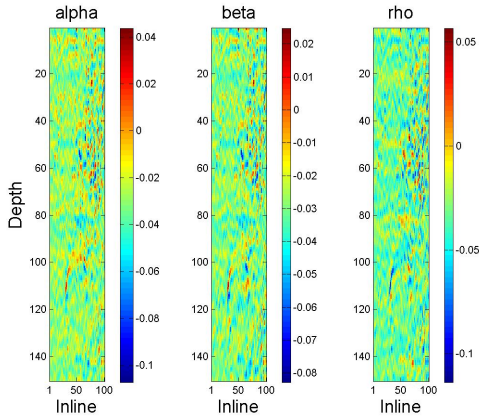
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.897 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.783 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.067 \cdot 10^{-3}$.



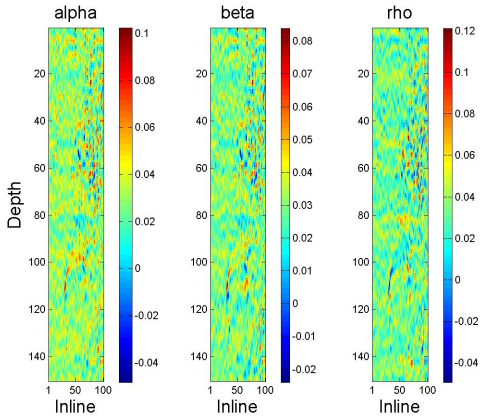

(a) The predicted $\hat{\boldsymbol{X}}_i$ for all $i$




(b) The residuals for all $i$




(c) The 95% prediction interval lower bound as given by the Godambe information

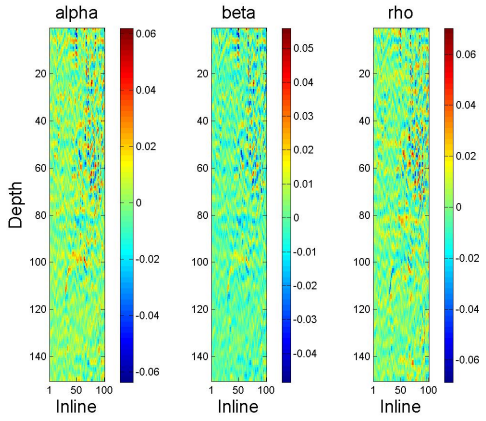



(d) The 95% prediction interval upper bound as given by the Godambe information

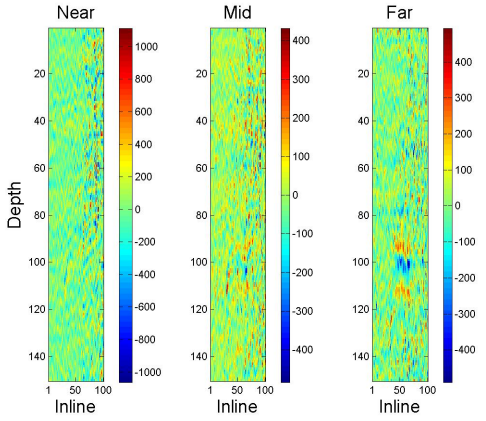
Figure 17: Prediction plots for the model $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{Matern})$

The plots in Figure 17 show the predicted $\hat{\boldsymbol{X}}_i$, the residuals, and the upper and lower bound of the predictions for all $i$. The model used is $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{Matern})$ which means the model uses two data columns for each $i$ as discussed in Section 5.3 the noise is not wavelet convoluted and the correlation function used is the exponential correlation function.

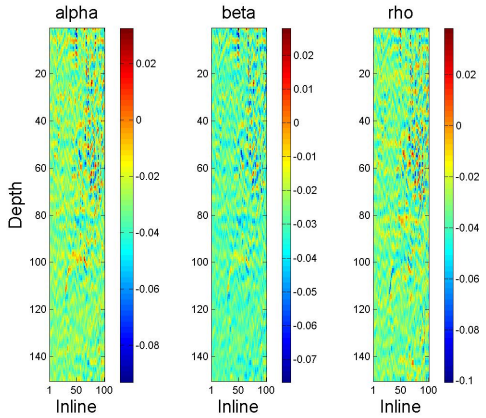
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.882 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.865 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.058 \cdot 10^{-3}$.



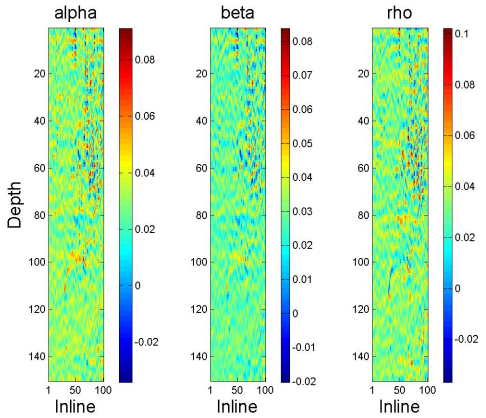

(a) The predicted $\hat{\boldsymbol{X}}_i$ for all $i$




(b) The residuals for all $i$




(c) The 95% prediction interval lower bound as given by the Godambe information




(d) The 95% prediction interval upper bound as given by the Godambe information

Figure 18: Prediction plots for the model $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$

The plots in Figure 18 show the predicted $\hat{\boldsymbol{X}}_i$, the residuals, and the upper and lower bound of the predictions for all $i$. The model used is $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{exp})$ which means the model uses two data columns for each $i$ as discussed in Section 5.3 the noise is wavelet convoluted and the correlation function used is the exponential correlation function.

The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.886 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.814 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.039 \cdot 10^{-3}$.

Comparing the different models it seems evident that the convoluted sound term is a poor choice. The prediction variance is much larger when comparing to the same models with the simple noise term without any clear improvements in the predictions themselves. The processing artefacts in the data are a possible explanation for this. The wavelet convoluted noise term is not designed to model the processing artefacts and this causes the convoluted noise term to be ineffective at modelling the data.

Using $\boldsymbol{Y}_i^2$ instead of $\boldsymbol{Y}_i$ appears to contribute little to the improvement of the predictions which is unexpected. Remembering that it also entails added computation time, using $\boldsymbol{Y}^2$ is not recommended. It suggests that that the added column that needs predicting evens out the advantages of using more data columns in the prediction process. A better approach for improvement might therefore be to consider single data columns $\boldsymbol{Y}_i$, but include a larger neighbourhood. The range of the correlation functions is however quite small, so the effect of a larger neighbourhood might be limited. Another way of improving the predictions would be to include a non-stationary model. By letting the mean or the variance of the elastic parameters be dependant on the depth in accordance with prior knowledge of the data the irregularities in the data could possibly be captured better. With regard to choice of correlation function, it appears that both perform quite equally. There are no noticeable differences in the predictions themselves and the prediction variances are more or less the same.

As mentioned in the beginning of this section, the prediction variance for each model is not an exact value, but rather an indicator of the magnitude. One may therefore argue that it is not advisable to put too much emphasis on their values and that the values for the models presented here are more or less the same. This is important to keep mind, but this does not alter any of the suggestions with regard to choice of model. In both the case of noise term and number of data columns the easiest and computationally fastest model is chosen due to lack of any evident improvement of the results, not because the results are thought to be better for the model suggested.

When doing composite prediction it is possible to compare the results using the Fourier transform inversion method [3]. The method has become somewhat of a benchmark in seismic inversion problems, but the method is only viable under certain assumptions. The mean and variance must be stationary and with seismic data that is often not the case. Since stationarity is assumed in this thesis however, the Fourier transform inversion method can help evaluate the results of the composite predictions.

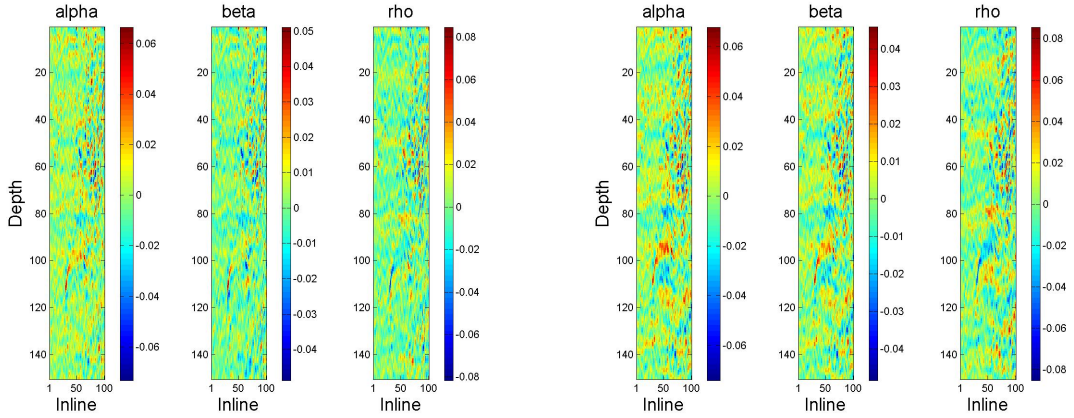
Stationary AVA inversion problems can be solved exactly using the Fourier transform inversion method. The idea is find the posterior distribution of $\boldsymbol{X}$ from a prior distribution and a likelihood model. In order to obtain the posterior distribution the calculations are done in the Fourier domain, allowing for fast calculations. A short summary of the method is given, but the reader is referred to [3] for a detailed and thorough explanation of the method.

First a prior distribution for $\boldsymbol{X}$ must be found together with a suitable likelihood model. The prior of $\boldsymbol{X}$ and the marginal distribution of $\boldsymbol{Y}$ from section 2.4 are used for this purpose. The next step is to transform all the elements of the model to the Fourier domain. $\boldsymbol{\mu}_X = \boldsymbol{0}$ for 4D data so the Fourier transform of $\boldsymbol{\mu}_X$, denoted $\tilde{\boldsymbol{\mu}}_X$, is also $\boldsymbol{0}$. For the method to work the two covariance matrices $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_\epsilon$ first need to be made into circular matrices [3]. The circular matrices can now be transformed to

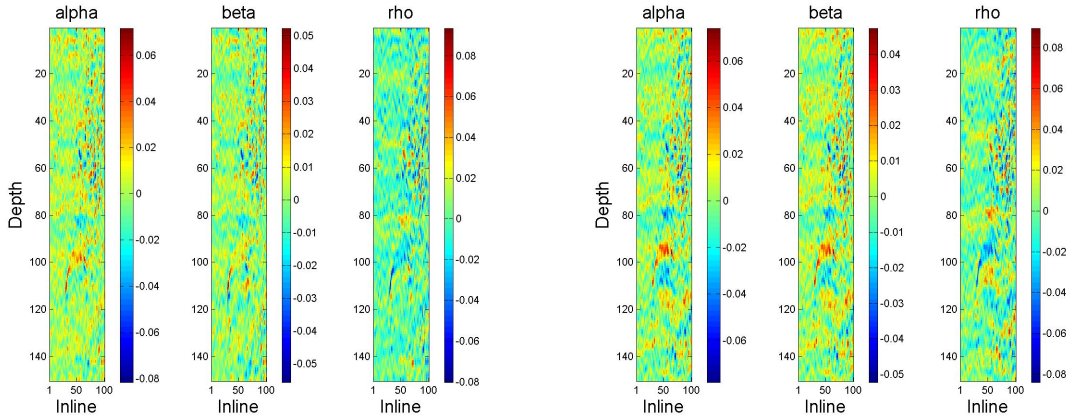
the Fourier domain, yielding $\tilde{\boldsymbol{\Sigma}}_X$ and $\tilde{\boldsymbol{\Sigma}}_\epsilon$. The design matrix $\boldsymbol{G}$ is also transformed to the Fourier domain. This is achieved by only transforming the wavelet function in $\boldsymbol{W}$ and multiplying this by the corresponding elements in $\boldsymbol{A}$. Now in the Fourier domain it is possible to calculate the posterior mean and covariance matrix

$$\tilde{\boldsymbol{\mu}}_{X|Y} = \tilde{\boldsymbol{\mu}}_X + (\tilde{\boldsymbol{G}}\tilde{\boldsymbol{\Sigma}}_X)^* \tilde{\boldsymbol{\Sigma}}_Y^{-1}(\tilde{\boldsymbol{Y}}) - \tilde{\boldsymbol{\mu}}_Y$$
$$\tilde{\boldsymbol{\Sigma}}_{X|Y} = \tilde{\boldsymbol{\Sigma}}_X - (\tilde{\boldsymbol{G}}\tilde{\boldsymbol{\Sigma}}_X)^* \tilde{\boldsymbol{\Sigma}}_Y^{-1} \tilde{\boldsymbol{G}}\tilde{\boldsymbol{\Sigma}}_X$$

where the $*$ operator denotes the Hermitian or the conjugate transpose of the matrix. In a 4D model both $\tilde{\boldsymbol{\mu}}_X$ and $\tilde{\boldsymbol{\mu}}_Y$ will be $\mathbf{0}$ vectors. By inverse Fourier transformation it is now possible to find $\boldsymbol{\mu}_{X|Y}$ and $\boldsymbol{\Sigma}_{X|Y}$ which give us the predictions and the variance of the predictions.




(a) The predicted $\hat{\boldsymbol{X}}_i$ for all $i$ for $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{exp})$ (b) The predicted $\boldsymbol{X}_{Fi}$ for all $i$ for $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{exp})$




(c) The predicted $\hat{\boldsymbol{X}}_i$ for all $i$ for $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{exp})$ (d) The predicted $\boldsymbol{X}_{Fi}$ for all $i$ for $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{exp})$

Figure 19: Plots comparing $\hat{\boldsymbol{X}}_i$ and $\boldsymbol{X}_{Fi}$ for all $i$ for the two models $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{exp})$ and $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{exp})$

Figure 19 shows the predictions using composite prediction and using the Fourier inversion method. The two different prediction techniques yield similar plots with the same distinct patterns and it appears that the composite likelihood method does a good job of predicting.

## 6.4 Leave one out prediction results

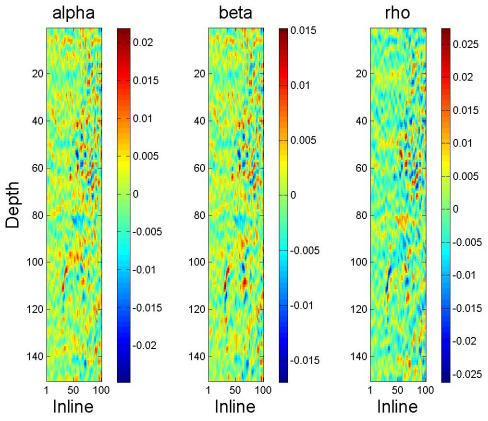
For the leave on out predictions the parameters found in the Section 6.1 are used in their respective prediction models. As discussed in section 4.5 it is interesting to look at the residuals when evaluating leave one out predictions. In order to visualise the results three different plots for each model are shown. First the estimation of the elastic parameters $\boldsymbol{X}_i$ themselves for all $i$ are plotted. Next the residuals $\hat{\boldsymbol{r}}_i$ as given by equation 32 are plotted. The last plot aims to visualise whether or not the residuals lie within the prediction intervals derived from the Godambe information as given in equation 33. The plot shows where the true values of $\boldsymbol{Y}_i$ are covered by the prediction intervals by having different colours for the two alternatives. As in the previous section the prediction variance derived from the Godambe information described in Section 4.4 is calculated for each of the elastic parameters for each of the models.

The mean square error (MSE) is calculated for each angle so it can be compared to the estimated noise estimated in Section 5.2. For the models with the simple noise term given in equation 41 it is sufficient to compare the MSE for each angle with each diagonal element in $\boldsymbol{\tau}_1^2$. When the model uses the noise term given in equation 42 the entire matrix is first calculated. Then the mean of the diagonal for each angle is used for comparison. The values can be seen in Table 7.

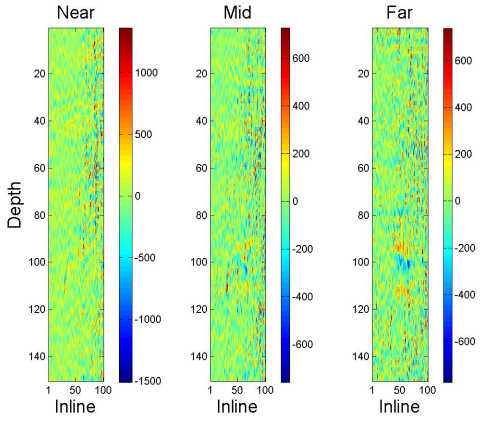
Table 7: Table of the estimated noise for the two noise terms for each of the three angles.

| | Near | Mid | Far |
|---|---|---|---|
| Simple noise | $6.079 \cdot 10^4$ | $1.884 \cdot 10^4$ | $2.955 \cdot 10^4$ |
| Convoluted noise | $6.130 \cdot 10^4$ | $1.874 \cdot 10^4$ | $2.932 \cdot 10^4$ |

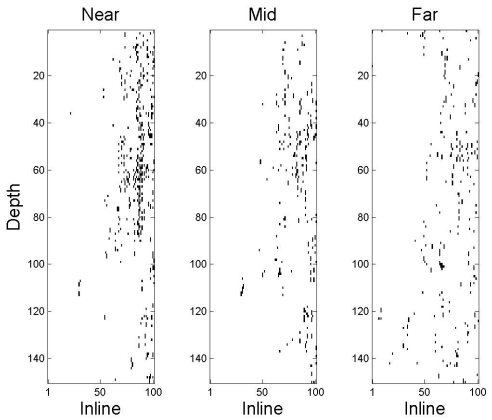
Similar to the previous section plots and results will be presented for each model first, then they will be discussed and analysed.




(a) The values of $\hat{\boldsymbol{X}}_i$ for all $i$




(b) Residuals for $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \exp)$

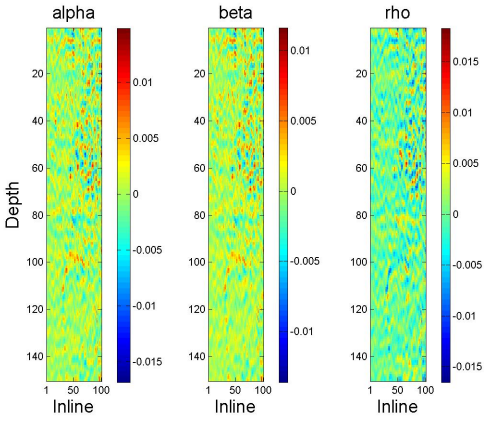



(c) Plot of whether the residuals lie within the confidence interval defined by the Godambe information. White means yes, black no.

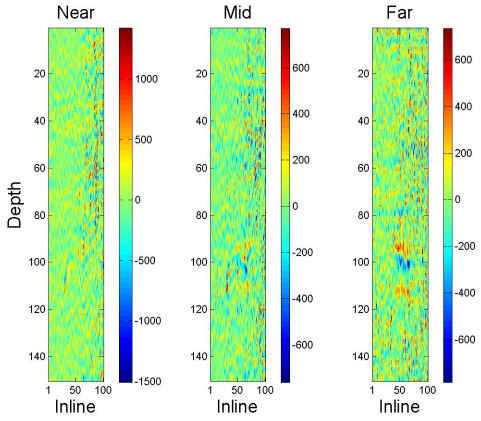
Figure 20: Prediction plots for the model $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \exp)$

The plots in Figure 20 show the predicted $\hat{\boldsymbol{X}}_i$ for all $i$, the residuals $\hat{\boldsymbol{r}}_i$ for all $i$ and a plot of whether the residuals lie within a 95% prediction interval. The model used is $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \exp)$ which means the model uses single data columns for each $i$, the noise is not wavelet convoluted and the correlation function used is the exponential correlation function.

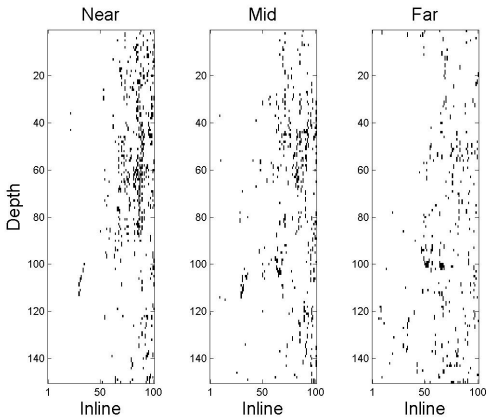
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.950 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.778 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.221 \cdot 10^{-3}$.




(a) The values of $\hat{\boldsymbol{X}}_i$ for all $i$




(b) Residuals for $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{exp})$

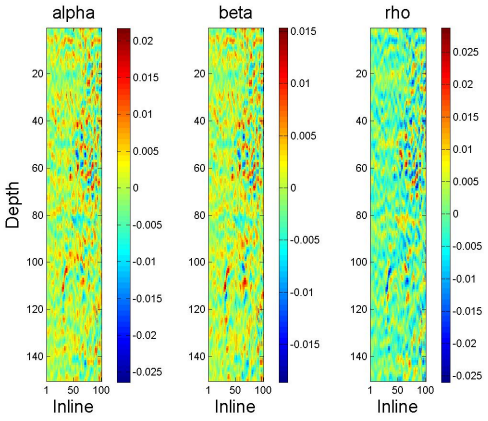



(c) Plot of whether the residuals lie within the confidence interval defined by the Godambe information. White means yes, black no.

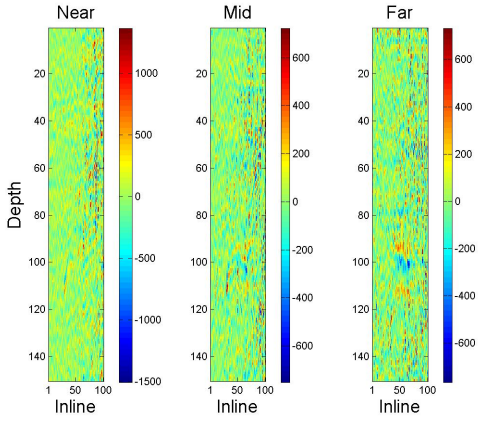
Figure 21: Prediction plots for the model $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{exp})$

The plots in Figure 21 show the predicted $\hat{\boldsymbol{X}}_i$ for all $i$, the residuals $\hat{\boldsymbol{r}}_i$ for all $i$ and a plot of whether the residuals lie within a 95% prediction interval. The model used is $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{exp})$ which means the model uses single data columns for each $i$, the noise is wavelet convoluted and the correlation function used is the exponential correlation function.

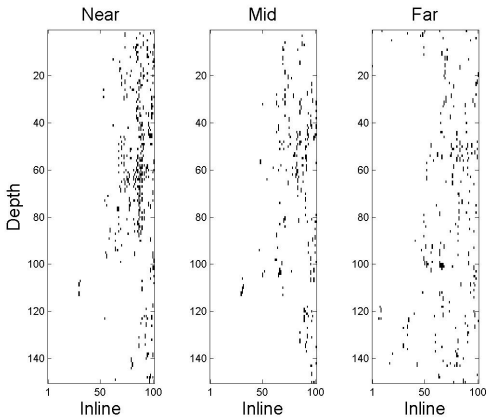
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 1.088 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.994 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.281 \cdot 10^{-3}$.




(a) The values of $\hat{\boldsymbol{X}}_i$ for all $i$




(b) Residuals for $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{Matern})$

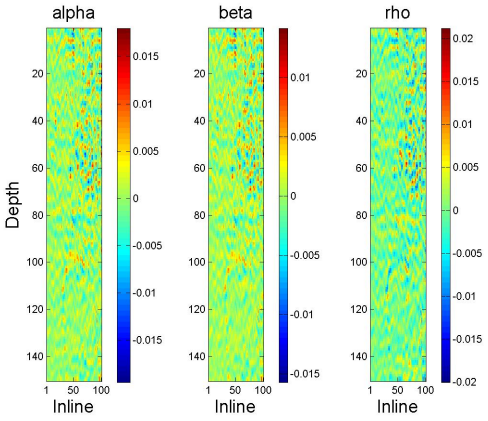



(c) Plot of whether the residuals lie within the confidence interval defined by the Godambe information. White means yes, black no.

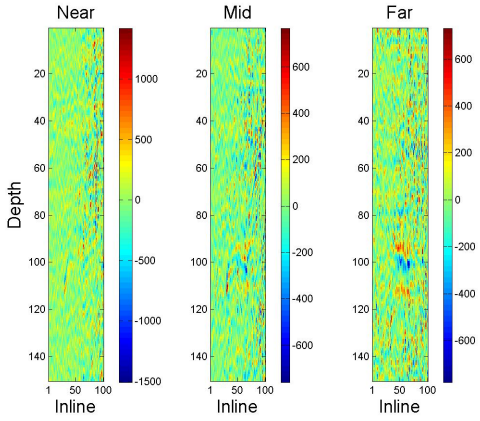
Figure 22: Prediction plots for the model $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{Matern})$

The plots in Figure 22 show the predicted $\hat{\boldsymbol{X}}_i$ for all $i$, the residuals $\hat{\boldsymbol{r}}_i$ for all $i$ and a plot of whether the residuals lie within a 95% prediction interval. The model used is $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{Matern})$ which means the model uses single data columns for each $i$, the noise is not wavelet convoluted and the correlation function used is the Matern correlation function.

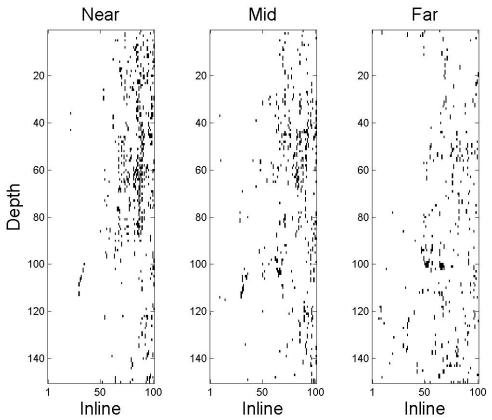
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.903 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.787 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.155 \cdot 10^{-3}$.




(a) The values of $\hat{\boldsymbol{X}}_i$ for all $i$




(b) Residuals for $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$

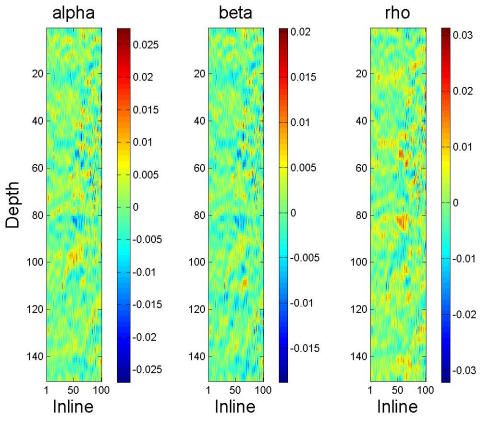



(c) Plot of whether the residuals lie within the confidence interval defined by the Godambe information. White means yes, black no.

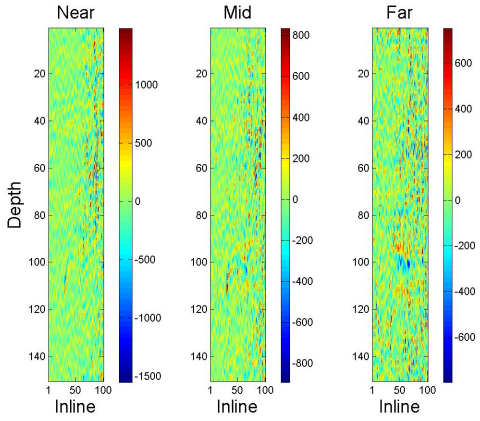
Figure 23: Prediction plots for the model $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$

The plots in Figure 23 show the predicted $\hat{\boldsymbol{X}}_i$ for all $i$, the residuals $\hat{\boldsymbol{r}}_i$ for all $i$ and a plot of whether the residuals lie within a 95% prediction interval. The model used is $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$ which means the model uses single data columns for each $i$, the noise is wavelet convoluted and the correlation function used is the Matern correlation function.

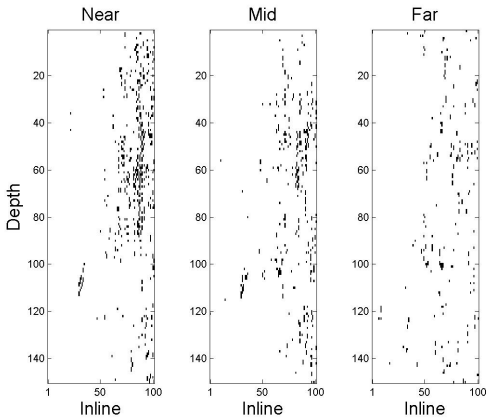
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 1.026 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.947 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.205 \cdot 10^{-3}$.




(a) The values of $\hat{\boldsymbol{X}}_i$ for all $i$




(b) Residuals for $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \exp)$

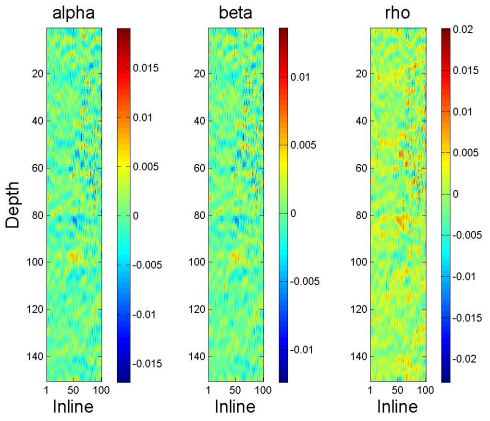



(c) Plot of whether the residuals lie within the confidence interval defined by the Godambe information. White means yes, black no.

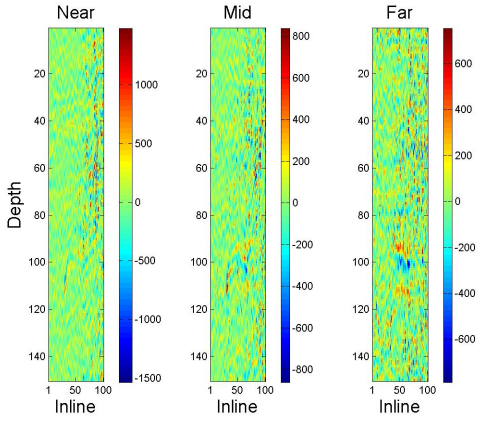
Figure 24: Prediction plots for the model $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \exp)$

The plots in Figure 24 show the predicted $\hat{\boldsymbol{X}}_i^2$ for all $i$, the residuals $\hat{\boldsymbol{r}}_i^2$ for all $i$ and a plot of whether the residuals lie within a 95% prediction interval. The model used is $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{Matern})$ which means the model uses two data columns for each $i$ as discussed in Section 5.3, the noise is not wavelet convoluted and the correlation function used is the exponential correlation function.

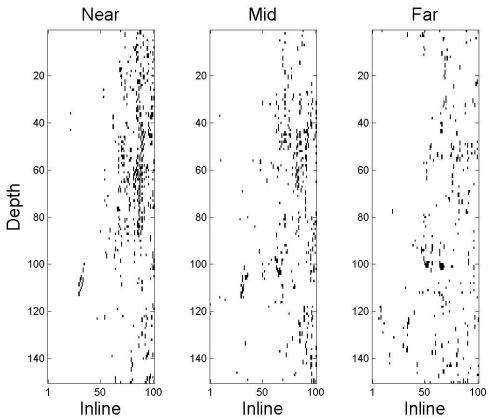
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 1.100 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.977 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.361 \cdot 10^{-3}$.




(a) The values of $\hat{\boldsymbol{X}}_i$ for all $i$




(b) Residuals for $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \exp)$

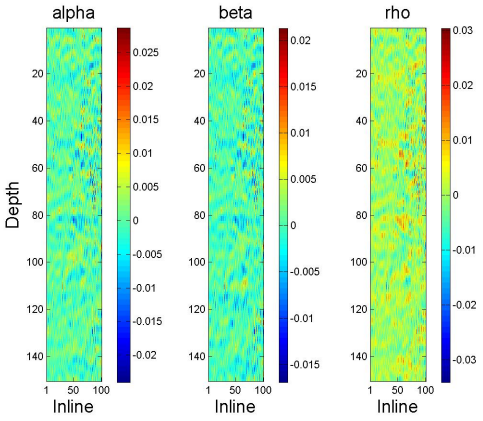



(c) Plot of whether the residuals lie within the confidence interval defined by the Godambe information. White means yes, black no.

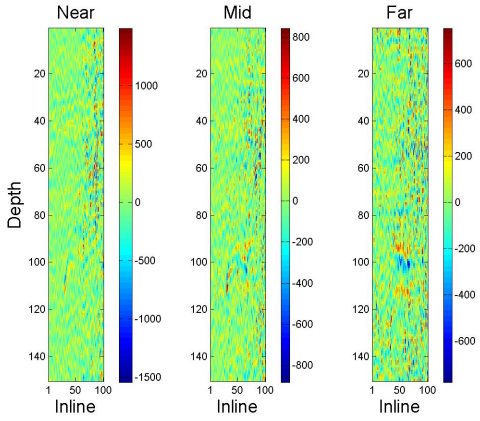
Figure 25: Prediction plots for the model $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \exp)$

The plots in Figure 25 show the predicted $\hat{\boldsymbol{X}}_i^2$ for all $i$, the residuals $\hat{\boldsymbol{r}}_i^2$ for all $i$ and a plot of whether the residuals lie within a 95% prediction interval. The model used is $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$ which means the model uses two data columns for each $i$ as discussed in Section 5.3, the noise is wavelet convoluted and the correlation function used is the exponential correlation function.

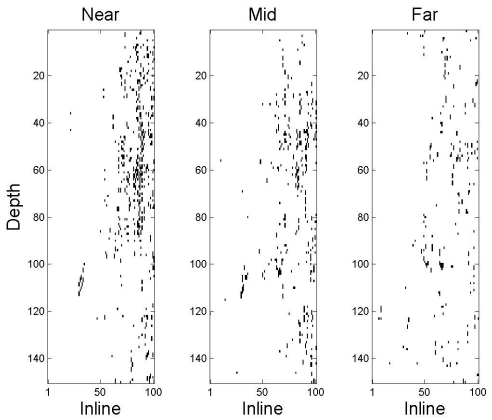
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.985 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.829 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.185 \cdot 10^{-3}$.




(a) The values of $\hat{\boldsymbol{X}}_i$ for all $i$




(b) Residuals for $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{Matern})$

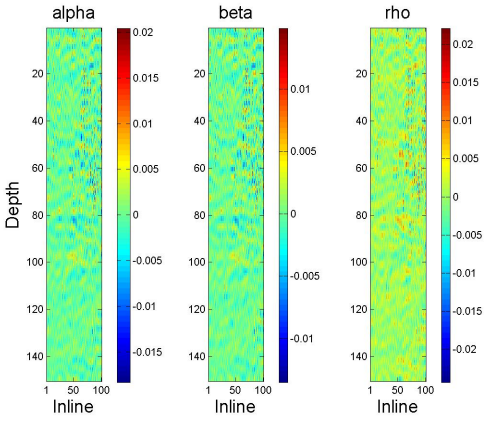



(c) Plot of whether the residuals lie within the confidence interval defined by the Godambe information. White means yes, black no.

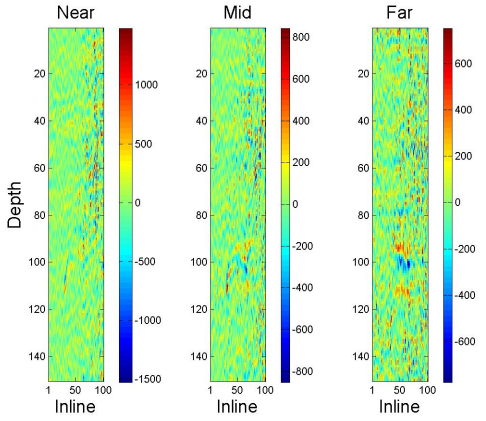
Figure 26: Prediction plots for the model $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{Matern})$

The plots in Figure 26 show the predicted $\hat{\boldsymbol{X}}_i^2$ for all $i$, the residuals $\hat{\boldsymbol{r}}_i^2$ for all $i$ and a plot of whether the residuals lie within a 95% prediction interval. The model used is $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{Matern})$ which means the model uses two data columns for each $i$ as discussed in Section 5.3, the noise is not wavelet convoluted and the correlation function used is the Matern correlation function.

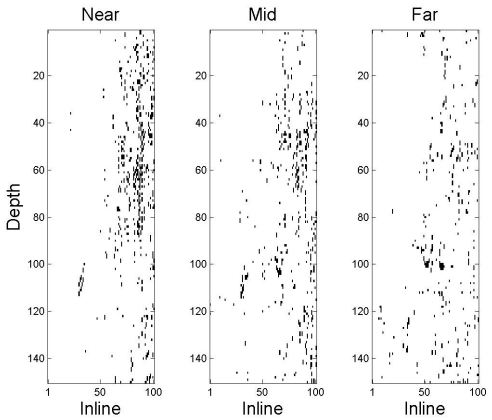
The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 1.015 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.937 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.249 \cdot 10^{-3}$.




(a) The values of $\hat{\boldsymbol{X}}_i$ for all $i$




(b) Residuals for $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$




(c) Plot of whether the residuals lie within the confidence interval defined by the Godambe information. White means yes, black no.

Figure 27: Prediction plots for the model $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$

The plots in Figure 27 show the predicted $\hat{\boldsymbol{X}}_i^2$ for all $i$, the residuals $\hat{\boldsymbol{r}}_i^2$ for all $i$ and a plot of whether the residuals lie within a 95% prediction interval. The model used is $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$ which means the model uses two data columns for each $i$ as discussed in Section 5.3, the noise is wavelet convoluted and the correlation function used is the Matern correlation function.

The mean of the prediction variance using the Godambe information is $\text{var}(\hat{\boldsymbol{X}}_\alpha)_{\text{mean}} = 0.972 \cdot 10^{-3}$, $\text{var}(\hat{\boldsymbol{X}}_\beta)_{\text{mean}} = 0.863 \cdot 10^{-3}$ and $\text{var}(\hat{\boldsymbol{X}}_\rho)_{\text{mean}} = 1.155 \cdot 10^{-3}$.

Table 8: Table of the MSE values for each of the three angles for all eight models.00

| | Near | Mid | Far |
|---|---|---|---|
| $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{exp})$ | $5.813 \cdot 10^4$ | $2.104 \cdot 10^4$ | $2.765 \cdot 10^4$ |
| $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{exp})$ | $5.994 \cdot 10^4$ | $2.255 \cdot 10^4$ | $3.022 \cdot 10^4$ |
| $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{Matern})$ | $5.825 \cdot 10^4$ | $2.118 \cdot 10^4$ | $2.791 \cdot 10^4$ |
| $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$ | $5.977 \cdot 10^4$ | $2.244 \cdot 10^4$ | $3.007 \cdot 10^4$ |
| $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{exp})$ | $6.166 \cdot 10^4$ | $2.378 \cdot 10^4$ | $3.134 \cdot 10^4$ |
| $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{exp})$ | $6.178 \cdot 10^4$ | $2.392 \cdot 10^4$ | $3.179 \cdot 10^4$ |
| $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{Matern})$ | $6.195 \cdot 10^4$ | $2.395 \cdot 10^4$ | $3.183 \cdot 10^4$ |
| $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$ | $6.193 \cdot 10^4$ | $2.402 \cdot 10^4$ | $3.205 \cdot 10^4$ |

The MSE for each model can be seen in Table 8 and are very close to the estimated noise from the overburden seen in Table 7. This is encouraging and implies that all the models do a fairly good job of predicting. The values of the MSE appear to vary less for each angle than the estimated noise.

As in the previous section the choice of using two data columns for each $i$ does little to improve the model, but increases computation time. The prediction variances are more or less the same, the MSEs are actually a little higher and the coverage of the prediction intervals about the same. Much of the same arguments apply for the leave-one-out prediction technique as in the previous section as to why this is so. It appears that the difficulty of predicting two data columns at once negates the advantages of using more data columns in the prediction process. Again it is advisable to instead attempt to predict just one data column at a time, while using a larger neighbourhood in order to improve on the model.

For the leave one out predictions it would also be interesting to look at a non-stationary model. Since the model used in this thesis is stationary, some areas of the data are not modelled well. By letting the mean or the variance of the elastic parameters be dependant on the depth in accordance with prior knowledge of the data these effects could possibly be captured better. This is especially true around the processing artefacts and the reservoir, and this could have been incorporated in a non-stationary model.

The results of the choice of noise term appear to be dependant on the number of data columns included in the model when considering the prediction variance. The differences are not great and considering that the simple noise term performs slightly better than the convoluted in the model with $\boldsymbol{Y}$, the simple noise term is the preferred term. The choice of correlation function appears to have little effect on the results and there is no evidence that one is better than the other.

Table 9: Coverage of the 95% prediction intervals

| | Coverage |
|---|---|
| $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{exp})$ | 0.9788 |
| $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{exp})$ | 0.9673 |
| $M(\boldsymbol{Y}; \boldsymbol{\tau}_1^2; \text{Matern})$ | 0.9782 |
| $M(\boldsymbol{Y}; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$ | 0.9674 |
| $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{exp})$ | 0.9670 |
| $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{exp})$ | 0.9647 |
| $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_1^2; \text{Matern})$ | 0.9670 |
| $M(\boldsymbol{Y}^2; \boldsymbol{\tau}_2^2, \boldsymbol{\tau}_3^2; \text{Matern})$ | 0.9640 |

The coverage of the prediction intervals, seen in Table 9, are surprisingly high for all eight models, all above 95%. This is most likely due to the processing artefacts discussed in Section 5. The variance in this area is very high and causes the estimates of the prediction variance to become higher. It is clear when looking at the plots that the coverage is poor in the area where the processing artefacts and the reservoir are located, but otherwise quite good. The prediction variance is estimated equally for all data columns, so the processing artefacts will influence the prediction intervals for all the data columns and thereby cause the intervals to be too large for most of the data.

# 7 Conclusion

This thesis has concerned itself with the performance of the composite likelihood method used in connection with 4D seismic data sets with regards to prediction and parameter estimation. The method enables the use of likelihood methods by estimating the covariance matrix through a close neighbourhood, assuming spatial correlation. This allows for quite fast calculations on data sets that would normally be impossible to use because of their high dimension. There are other useful methods for seismic data inversion problems, but the composite likelihood method could provide a very general framework that imposes few restrictions and yet yield satisfying results.

First in this thesis 4D seismic reflection is explained. It allows to locate untapped pockets of oil or gas and also serves a diagnostics tool for regular 3D seismic reflection. A Gaussian model for 4D seismic data is created, while attempting to explain some of the physics that dictates the form of the model. The composite likelihood method and its asymptotic properties are explained together with techniques for doing parameter estimation and prediction.

The method is then tested on a real data set from the oil field Norne where seismic data was collected in 2001 and again in 2003. The method is tested for eight different models, all variations of the general Gaussian model introduced earlier. Variations are made in the form of the noise estimate, the correlation function used and the number of data columns used as one data vector. Parameter estimations and predictions for both prediction techniques are carried out for all eight models.

The results from the parameter estimation and the predictions are encouraging. Most importantly the performance of the composite likelihood method appears to be satisfactory. The parameter estimations coincide well with previous experience and the asymptotic variance of the estimates is quite small. Also the predictions yield good results, for both techniques. The composite predictions are compared to predictions done using the Fourier Domain AVA inversion technique, regarded as a benchmark method in stationary seismic inversion problems and the predictions from each of the two methods appear to coincide well. There are still differences between the methods, but they appear to give similar results. The results from comparing the eight models are not as expected. By looking at the results of the predictions and the parameter estimations for the eight different models it seems that none of the alternatives that were meant to improve on the model do any improvement. Both the convoluted noise and the inclusion of more data columns do little to improve on the results and the only reasonable conclusion is to use the simplest model. The convoluted noise term is most likely affected by the artefacts in the data and should be tested more before a definite conclusion about it is reached. The reason the inclusion of more data columns has little effect is partly because of the short range of the correlation. In other models with high correlation range the approach might yield better results.

The results are nonetheless encouraging in the sense that the composite likelihood method performs quite well. It is a method that should arouse interest in spatial statistics due to its simple implementation and fast computation time. In the field of seismic reflection it is certainly an interesting method and further investigations should be made.

In this thesis the model investigated has been a stationary model. For instance it would be interesting to use the composite likelihood method with a non-stationary model. Using knowledge of the data this could help improve the model fit, especially in areas with anomalies and in and around the reservoir itself. It would also be interesting to look at a full 4D data set, meaning the data in each survey is 3D. The method has no problem with 3D data sets, and it would be interesting to try a full scale data set.

It is possible to estimate the noise parameter using the Gauss Newton method instead of using the overburden. It might also be interesting to use a noise term that is dependent on position as the noise in the data seems to vary for different areas in the Norne data set. These artefacts in the Norne data set affect many of the results found in this thesis. It would therefore be interesting to find the results for a different data set without any artefacts.

Another possible improvement on the method might be to include a larger neighbourhood. This would increase computation time, but it might affect the results for real data, especially if considering a data set with greater correlation range than the Norne data set.

# References

[1] Ran Bachrach. Joint estimation of porosity and saturation using stochastic rock-physics modelling. *Geophysics*, 2006.

[2] Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand. *Hierarchical modelling for spatial data*. Chapman Hall, 1 edition, 2004.

[3] Arild Buland, Odd Kolbjørnsen, and Henning Omre. Rapid spatially coupled AVO inversion in the Fourier domain. *Geophysics*, 68(3):824–836, 2003.

[4] Arild Buland and Henning Omre. Bayesian linearized AVO inversion. *Geophysics*, 68:185–198, 2003.

[5] Arild Buland and Youness El Ouair. Bayesian time-lapse inversion. *Geophysics*, 71:43–48, 2006.

[6] Jo Eidsvik, Benjamin A. Shaby, Brian J. Reich, Matthew Wheeler, and Jared Niemi. Estimation and prediction in spatial models with block composite likelihoods using parallel computing (preprint). *Statistics*, 2011.

[7] V. P. Godambe and C. C. Heyde. Quasi-likelihood and optimal estimation. *International Statistical Review*, 55(3):231–244, 1987.

[8] Peter K. Kitandis and Robert W. Lane. Maximum likelihood parameter estimation of hydrologic spatial processes be the Gauss-Newton method. *Journal of Hydrology*, (79):53–71, 1985.

[9] Bruce G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80:221–240, 1988.

[10] David E. Lumey. Time-lapse seismic reservoir monitoring. *Geophysics*, 2001.

[11] Gary Mavko. Time-lapse seismic reservoir monitoring. *Rock Physics Laboratory Stanford University*, 2001.

[12] Robert E. Sheriff and L. P. Geldart. *Exploration seismology*. Cambridge University Press, 2 edition, 1995.

[13] Micheal L. Stein, Zhiyi Chi, and Leah J. Welty. Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society*, 66(2):275–296, 2004.

[14] Cristiano Varin. On composite marginal likelihoods. *Advances in Statistical Analysis*, (92):1–28, 2008.