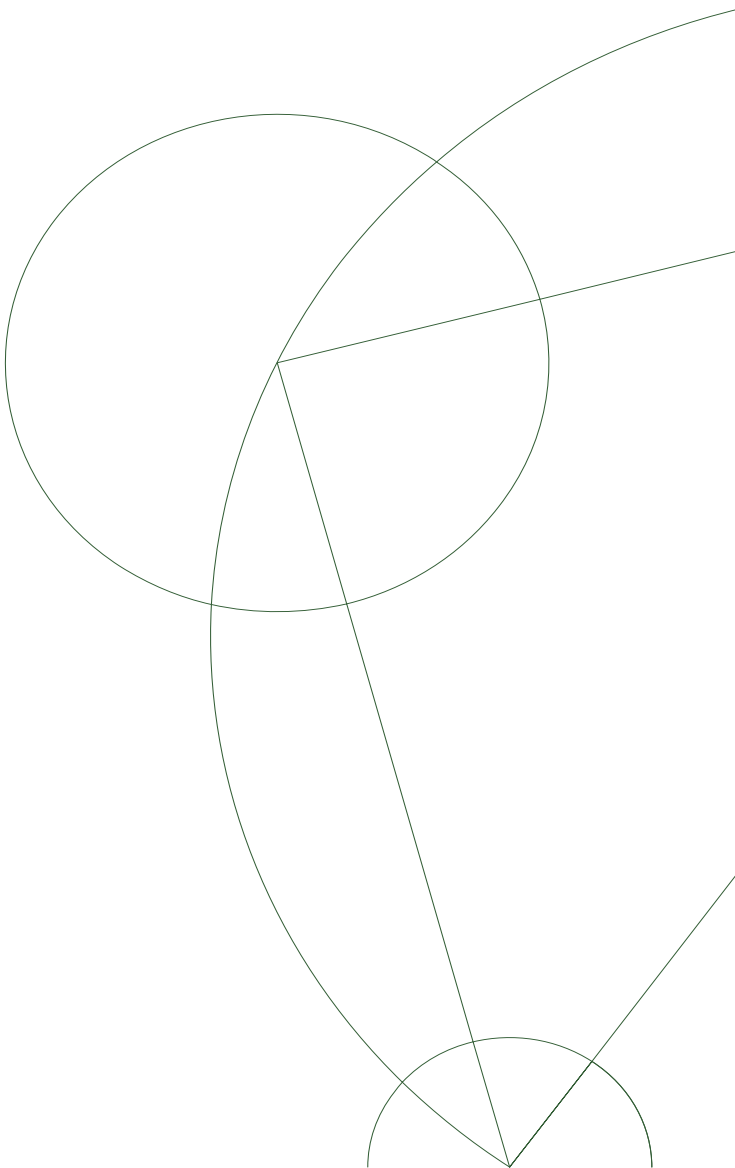
Probabilistic Linear Inversion of Reflection Seismic Data

Rasmus Bødker Madsen

Ph.D. Thesis

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen

Niels Bohr Institute
Date: June 1, 2018
Supervisor: Associate professor Thomas Mejer Hansen



Doctor of Philosophy Thesis in Computational Geophysics
Title: Probabilistic Linear Inversion of Reflection Seismic Data
Danish title: Probabilistisk lineær inversion af reflektionsseismiske data

Institute: Niels Bohr Institute
Section: Climate and Computational Geophysics
Group: Inverse Modeling and Geostatistics Project
Task delivery date: June 1, 2018

Author: Ph.D. candidate, M.Sc. Rasmus Bødker Madsen, University of Copenhagen
E-mail: rasmus.madsen@nbi.ku.dk

Supervisor: Associate professor, Ph.D. Thomas Mejer Hansen, University of Copenhagen

Thesis advisor: Thomas Mejer Hansen Rasmus Bødker Madsen

Probabilistic Linear Inversion of Reflection Seismic Data

Abstract

Seismic acquisition is the major provider of geophysical data for large-scale hydrocarbon exploration purposes. Large quantities of data-rich seismic surveys hampers a proper integration of geophysical data and geological prior knowledge, due to computational demands. To remedy this, simplified and approximate models, and mathematically convenient priors are applied to allow computationally faster solutions. It is in particular difficult to obtain reliable uncertainty estimates of the physical parameters in the subsurface using linear-least squares solutions. These models present a trade-off between obtaining a realistic depiction of the subsurface and having an analytic (computationally efficient) solution to the inverse problem. This thesis addresses this issue and presents new methodologies and work-flows, which could potentially enable a better description of the subsurface. The work is manifested in three main directions.

Firstly, errors from using a simplified linear forward are considered. The work is split between linear assumptions of the non-linear Zoeppritz equations and errors arising during processing of seismic data. Errors from both sources are deemed significant in seismic inversion, and in both cases, the error is estimated and described with a Gaussian distribution. This enables more trustworthy posterior solutions.

Secondly, non-stationarity in the Gaussian prior distribution is considered. Using localized marginal maximum likelihood estimators, a non-stationary global measure of the variance is obtained. This estimate is used as a plug-in variance in the prior distribution. Improved posterior results are obtained using the non-stationary estimate compared to a traditional stationary measure.

Finally, the strategy of inferring the noise model as part of the inversion result is investigated. It is demonstrated that the shape of the noise is decisive for the reliability of the obtained posterior solutions. In all three presented work areas,

the results indicate that a proper description of the noise on the data is vital in bridging the gap between computationally feasible solutions and realistic posterior distributions. A better description and accounting for modeling errors can potentially significantly improve currently available linear probabilistic inversion methods.

Probabilistisk lineær inversion af reflektionsseismiske data

Dansk Resumé

Seismisk dataindsamling er den primære geofysiske datakilde ved stor-skala efterforskning af hydrokarboner. Beregningsmæssige vanskeligheder ved store mængder af seismisk data hæmmer en passende integration af seismisk data med forudgående geologisk viden. Dette problem kan afhjælpes ved at benytte approksimative modeller og matematisk praktiske priorfordelinger til at opnå beregningslette løsninger. Det er i særdeleshed vanskeligt at opnå pålidelige usikkerhedsestimater for de fysiske parametre i undergrunden ved brug af mindste-kvadraters metoder. Brugen af mindste-kvadraters metoder er dermed en afvejning mellem realisme i inversionsresultaterne og muligheden for at have en analytisk løsning. Denne afhandling beskæftiger sig med denne problemstilling og præsenterer nye metoder og arbejdsgange til en bedre afbildning af undergrunden. Arbejdet manifesterer sig i tre hovedretninger.

For det første undersøges fejl, der opstår som følge af brugen af simplificerede lineære modeller til at løse det direkte problem. Dette arbejde er delt mellem linæariserede versioner af de ikke-lineære Zoeppritz-ligninger og fejl der opstår som følge af processering af seismisk data. I begge tilfælde er fejlen estimeret og beskrevet med en Gaussisk fordeling, hvilket muliggør mere troværdige løsninger af det inverse problem.

For det andet, behandles ikke-stationaritet i Gaussiske priorfordelinger. Ved brug af lokale maksimum sandsynlighedsestimatorer kan et globalt ikke-stationært mål for variansen opnås. Dette mål bruges som et estimat for variansen i priorfordelingen. Ved at benytte dette estimat opnås mere troværdige løsninger til det inverse problem end med en traditionel stationær varians.

Slutteligt afprøves strategien om at udlede støj-modellen som en del af inversionsalgoritmen. Det er her demonstreret, hvorledes formen på støjen har betydning for pålideligheden af de opnåede inversionsresultater. Resultaterne, for alle tre arbejdsområder indikerer, at en passende beskrivelse af støjen på data er

altafgørende for at bygge bro mellem de beregningslette løsninger og realistiske inversionsresultater. En bedre beskrivelse af støjen, hvor den samtidig bliver håndteret i inversionen, vil potentielt kunne forbedre nuværende lineære probabilistiske inversionsmetoder signifikant.

Contents

Listing of figures

This Ph.D. dissertation is dedicated to my grandparents: Birthe, Knud, Ruth, and Alf

Preface

The present Ph.D. thesis entitled "Probabilistic Linear Inversion of Reflection Seismic Data" finalize a 3-year Ph.D. program, and is part of the requirements to complete a Ph.D. degree at the University of Copenhagen. The work has been carried out at the Inverse Modeling and Geostatistics Project (IMGP) group at the Niels Bohr Institute (NBI), University of Copenhagen. The initial title for the Ph.D. project was *Probabilistic Seismic Prospect Assessment* (PSPA) and has been carried out in the period between 2015 and 2018. The Ph.D. project is funded by the Innovation Foundation Denmark, grant number: 53-2014-3. The original project formulation was to develop software that could provide an effective tool when large amounts of geophysical data from the subsurface are pieced together. The PSPA project is split between two parties, the IMGP at NBI and Qeye Labs, which is a small seismic inversion contractor based in Copenhagen. Qeye Labs' role in the project is to provide software and practical experience with inversion of seismic data on large scale problems. The objective for the IMGP group is to develop numerical methods and algorithms for probabilistic inversion. The work

carried out through this Ph.D. study falls in the latter category.

The main focus of this Ph.D. study has been to investigate modeling errors in seismic AVO data and their effect on probabilistic inversion. During my four months abroad at the Norwegian Science and Technology University (NTNU), the work was extended to cover work with non-stationarity in the variance of physical parameters as well. This work was done in collaboration with professor Henning Omre. The seismic data and scenarios for this Ph.D. are based on exploration geophysics on kilometer-scale, which is mainly of interest for the oil and gas industry.

Most computations have been carried out on a conventional 3,1 GHz Intel Core i7 laptop with 16 GB 1867 MHz DDR3 RAM. Heavier computational calculations, for example running multiple Finite Differences modeling and Monte-Carlo based inversion, have been performed on local in-house workstations.

The thesis consist of a summary of the work over the course of the Ph.D. project and six appendices with scientific work either submitted or currently undergoing review in geophysical journals. Appendix H.1 is an unaccepted expanded abstract for the Society of Exploration Geophyiscs (SEG) annual meeting in 2016. This work, along with a conference abstract for the GEOSTATS conference in 2016 (Appendix H.2), would later create the basis for Appendix H.4. Appendices H.3-H.5 are reviewed and published in SEG annual meeting 2017 proceedings, GEOPHYSICS and European Association for Geoscientists and Engineers (EAGE) annual meeting 2018 proceedings respectively. Appendix H.6 is written in collaboration with Henning Omre from NTNU and is submitted to GEOPHYSICS, currently undergoing review.

- **Appendix H.1: Seismic forward modeling errors from linear approximations to Zoeppritz equations** Rasmus Bødker Madsen & Thomas Mejer Hansen, submitted to SEG annual meeting 2016
- **Appendix H.2: The interplay between geostatistical prior information and modeling error in seismic data** Rasmus Bødker Madsen & Thomas Mejer Hansen, Abstracts from Valencia GEOSTATS 2016, page 179
- **Appendix H.3: On inferring the noise in probabilistic seismic AVO inversion using hierarchical Bayes** Rasmus Bødker Madsen, Andrea Zunino & Thomas Mejer Hansen, SEG annual meeting 2017 proceedings, page 601-605
- **Appendix H.4: Estimation and accounting for the modeling error in probabilistic linearized AVO inversion** Rasmus Bødker Madsen & Thomas Mejer Hansen, GEOPHYSICS, 2018, Volume 83, page N15-N30
- **Appendix H.5: Accounting for processing errors in AVO/AVA data** Rasmus Bødker Madsen, Egon Nørmark & Thomas Mejer Hansen, EAGE 80th annual meeting 2018, accepted 4 page expanded abstract
- **Appendix H.6: Prediction of acoustic impedance with non-stationary variance from seismic data** Rasmus Bødker Madsen, Thomas Mejer Hansen, & Henning Omre, GEOPHYSICS (In review)

Acknowledgments

I WOULD LIKE TO ACKNOWLEDGE all involved parties at the Niels Bohr Institute who helped me during my Ph.D. studies; Thomas Mejer Hansen for supervision and the occasional hourly conversation about distorted guitars and music in general; Andrea Zunino for being a great companion (i.e. listening to all my complaints) in office 108 and brewing some good espresso; Søren Granat for crisis management when dealing with the administration and for always being an all-round good guy; Knud Cordua for being a friendly face when I first came to Copenhagen, and for all the subsequent tennis/geophysics related activities; Klaus Mosegaard for being a source of inspiration and challenging me with thought-provoking statistical dilemmas/paradoxes; and the list continues...

I would also like to thank Qeye Labs for a good and constructive partnership during the Ph.D. A special gratitude towards Henning Omre and the Applied Statistics group at NTNU in Norway for taking good care of me while staying abroad in 2017. Especially, for encouraging me to apply an analytical approach to inverse problems, and for interesting inputs and discussions along the way. Your

hospitality was a great inspiration for my personal life as well as professional. Lastly, a warm thanks to all of my family, in particular Emma, and friends outside the university. Thank you for always having a positive effect on my mood, even during bad spells.

"I couldn't claim that I was smarter than sixty-five other guys - but the average of sixty-five other guys, certainly!"

Richard Feynman, Surely You're Joking, Mr. Feynman!: Adventures of a Curious Character

1 Introduction

The most important job for a geophysicist is to obtain knowledge and build reliable models of the Earth. Many methods exist to acquire information about the subsurface, but very few have been as successful and long-lasting as seismic exploration and seismology in general. The popularity of seismology is mainly rooted in the fact that it offers a unique source of information of deep-lying features of the subsurface. Secondly, but almost equally important, it provides a way of acquiring huge amounts of relatively high resolution data, that cover large spatial distances. The second property has made seismic data the main tool for hydrocarbon exploration purposes in the industry. The acquisition

of seismic data is a topic in itself and will not be covered in this thesis, although a brief overview of seismic acquisition will be given in the next section. The focus of this thesis is instead how to build reliable models from the acquired seismic data. Data does not usually contain direct measurements of the subsurface parameters of interest. Instead, data will contain indirect information of the subsurface parameters. One solution for building models of the subsurface through indirect information is formulating the retrieval of a model as an inverse problem. In this thesis, a probabilistic setting for inverse problems is used. A probabilistic setting allows combining information from the indirect measurements (i.e. data) with direct information of the subsurface (i.e. a priori information) to build models. The aim of the thesis is to investigate some of the main challenges involved with probabilistic linear inverse problems, namely quantifying the effect (the error) of linearizing a non-linear problem and dealing with non-stationarity of statistical properties of the subsurface and noise in general. In the following chapter, a short introduction to seismic data is given along with the processed entity called Amplitude Variation with Offset (AVO) seismic data. Thereafter, a section follows describing the general inverse problem and the typical approach to invert seismic data.

1.1 Seismic Data

The basic requirement for any seismic investigation of a medium (here the subsurface) is merely a source that can generate a pulse of energy (the signal) which travels through the medium. The acoustic properties of the medium affects the arrival and appearance of the signal. The altered signal is recorded with displacement sensitive microphones (the seismometers) as a time series of

displacement and/or pressure variation amplitudes. The seismic data thereby provide information which allows one to infer knowledge on properties of the medium. In the discipline of seismology the source is natural and is provided by earthquakes that propagate through the entire Earth. The arrivals of different wave phases have given much insight to the internal structure of the planet. Seismologic studies of the arrival of primary waves enabled the Danish seismologist Inge Lehmann to discover the Earth's inner core (Kölbl-Ebert, 2001). In general, three major investigation depths can be identified for different exploration purposes. Besides the seismologic studies regarding the whole Earth, two shallower depths can be identified. On meter-scale, very high-frequent seismic data is gathered to investigate near-surface properties. These experiments are usually conducted with the purpose of groundwater mapping and foundation of buildings. On kilometer-scale, high-frequent data are gathered usually with the aim of identifying hydrocarbons in the subsurface (i.e. exploration seismology). In this thesis the main target will be seismic data on the kilometer-scale. The inverse methods developed here can nevertheless be applied on different scales.

To obtain the high-frequent and precise data in exploration seismology, the energy source must be more predictable than earthquakes and is instead fabricated. This has over the years been done with everything from sledge hammers to explosives (Miller et al., 1986). In Russia, some exploration studies have even been carried out using smaller nuclear bombs as a source (Scheimer and Borg, 1984). Seismic data can be gathered both on land or at sea. The data collected at sea are important for exploring hydrocarbons in basins formed by tectonic activity (Allen and Allen, 2004), as is the case in e.g. the North Sea. Due to the water column at sea, an air gun is usually applied as the energy source. The

air gun is towed by a survey ship along with a streamer containing pressure sensitive microphones for marine usage (i.e. hydrophones). A sketch of this basic setup is shown in Figure 1.1. The energy leaves the air gun as pressure waves and travels through the subsurface. Depending on the seismic velocities (which depend on the rock properties) of each layer the seismic energy will either be reflected or refracted back according to Snell's law. A more detailed description of Snell's law can be found in Appendix A. A sudden shift in rock properties, and hence elastic properties, will result in energy being reflected back to the surface. These shifts in rock properties can therefore be detected on the resulting seismic data as large amplitudes arriving from the same depth, but only if the shifts are thick and rapid enough to reflect detectable amounts of energy back. The detectable "packages of energy" are often referred to as seismic reflectors. Because data acquired for exploration purposes are coupled to reflections in the subsurface they can perhaps more appropriately be named reflection seismic data.

An example of a raw reflection seismic data set can be seen in Figure 1.2. A raw reflection seismic data set can also more specifically be named a raw seismic shot gather or simply "raw seismic data". These three terms will be used interchangeably throughout the thesis. The term "raw" here refers to acquired seismic data which have not yet been subject to any processing. Usually seismic data are processed using signal processing techniques to lower noise levels to emphasize seismic reflectors. The raw seismic data display the recorded signal from each hydrophone as a function of time. Hydrophones are usually equally spaced along the seismic streamer, as illustrated in Figure 1.1, but the geometry can vary a lot between different surveys. The hydrophone number on the

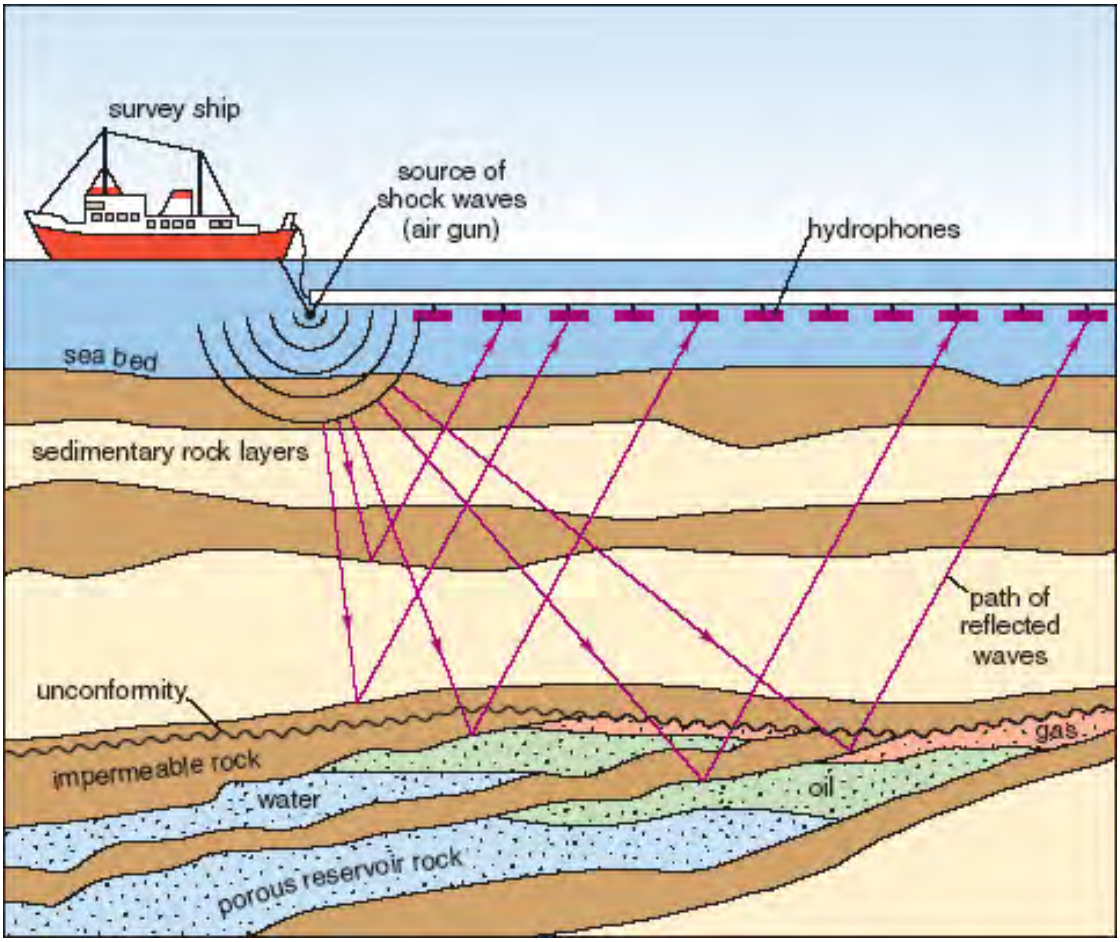


Figure 1.1: Sketch rendition of a seismic acquisition survey at sea. The survey ship is towing an air gun to inject seismic energy and a streamer with hydrophones for recording. Picture courtesy of Openlearn.

abscissa-axis is thereby a measure of horizontal location (i.e. source offset). The time on the ordinate-axis is a measure on the time it takes the pulse of energy to reach a seismic reflector and then get back to the surface, where it is recorded by the hydrophones. This time is commonly recognized as the *Two Way Travel time* (TWT) and acts as a pseudo-measure of depth. A proper conversion to depth in meters requires a known velocity field of the subsurface. Seismic velocities are typically in the thousand of meters per second scale. The sampling interval is often recorded in milliseconds while having a recording interval of a few seconds. In Figure 1.2 the sampling interval is for instance 2 milliseconds, and the recording interval is approximately 4 seconds. Inspecting the raw data in Figure 1.2, seismic reflectors at a certain depth can be identified by their hyperbolic movement/delay as a function of offset, because it takes increasingly more time to reach hydrophones at greater offset. In fact, the arrival of the signal is approximately proportional with time squared. A few seismic reflectors are marked on the raw data with yellow on the right. A wave of energy also travels in the water directly from the air gun to the hydrophones. This wave is moving proportional with time and is easily detectable in the seismic raw data, as it appears as a straight line. The direct wave is marked with blue in Figure 1.2. The direct wave is usually the first arrival at low offset hydrophones. However, as hinted in Figure 1.2, the direct wave may not necessarily be the first arriving wave at larger offsets because the velocity at which waves can travel in rocks is greater than in water.

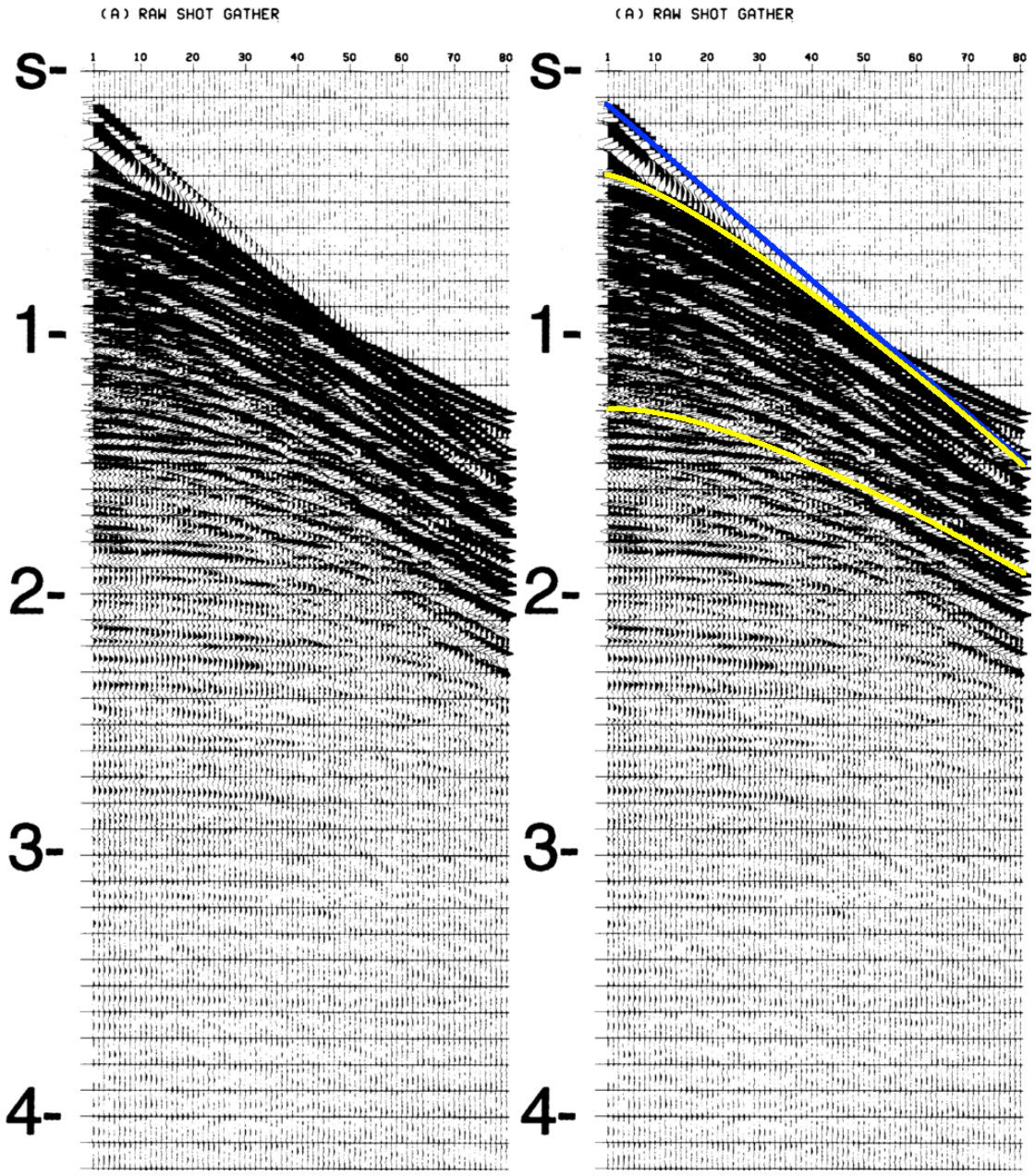


Figure 1.2: Example of raw seismic data. Left: raw seismic data as taken from Yilmaz (2001). Right: interpretations of arrivals of different phases colored on the raw data, where blue is the direct surface wave and yellow shows two seismic reflectors.

1.2 AVO Seismic Data

As stated in the previous section, seismic shot gathers are usually not used in their raw format but are instead processed to obtain lower noise levels. The data are then further processed using different processing techniques depending on the requirements for the specific seismic discipline. A wide variety of seismic disciplines require different processed data sets (see e.g. Yilmaz (2001) for examples). Some seismic data sets are for instance processed to amalgamate information from several raw data shot gathers into one section to showcase the subsurface geometry. This is done in order to enable easier structural geological interpretation.

In the following we shall explore a particular processed type of seismic data named Amplitude Variation with Offset or simply *Amplitude Versus Offset* data. In particular the *Amplitude Versus Angle* (AVA) data will be presented, because these serve as a basis for much of the work carried out in the thesis. AVO/AVA data are processed as a gather of traces. Each trace represents a different offset, that enables amplitude effects due to offset (or angles) to be investigated. An example of an AVA data set is shown in Figure 1.3. The gather shows amplitude at a specific location in the subsurface (here the *Common Depth Point* - CDP). Time is, as for the raw data set, used as measure of depth on the ordinate-axis. The raw data is processed such that the abscissa-axis shows the reflections as a function of incidence angle of the incoming wave front.

The main motivation for processing data into this form is primarily because of the rise of the seismic discipline AVO analysis. Since 1982, when it was first demonstrated that reflection coefficients of seismic boundaries change in an

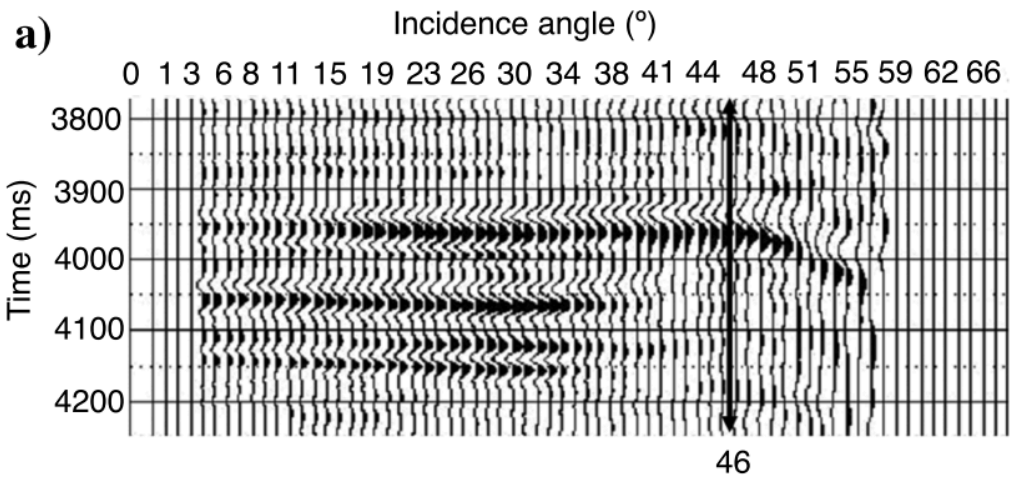


Figure 1.3: Example of seismic AVA data. Taken from (Contreras et al., 2007). The reflection amplitudes on the ordinate-axis are shown as a function of incidence angle on the abscissa-axis. A single trace is representing each incidence angle. The black arrow at incidence angle $= 46^{\circ}$ is an arbitrarily chosen cut-off value where the assumptions of AVA data breaks down (Contreras et al., 2007).

anomalous fashion with the presence of hydrocarbons, AVO analysis has become increasingly popular (Ostrander, 1982; Castagna and Backus, 1993). The anomalous behavior of hydrocarbons makes AVO/AVA data perfect for hydrocarbon detection. Hydrocarbons were discovered to locally show either high or low amplitudes in the seismic data depending on the elastic properties of the surroundings. These anomalous areas were given the popular names *Brigt-* and *Dim-spots* respectively. The fact that seismic amplitudes are affected by variations in the physical properties around layer boundaries also makes AVO data useful for lithology identification, porosity identification, and fluid parameter analysis (Castagna and Backus, 1993; Zhang and Brown, 2001). Even though there should be a clear distinction between AVO and AVA data, AVA data have historically been referred to as AVO data in AVO analysis (Zhang and Brown, 2001). To avoid any confusion within the following, AVA data are here referring to the variation of reflection and transmission coefficients with angle of incidence as seen in Figure 1.3.

AVO analysis rely on one fundamental assumption: Reflections between two elastic media are assumed to arise from plane waves at a planar interface. This assumption lays the foundation of Zoeppritz (1919) equations, which forms the relationship between incident wave and the scattered wave phases at the boundary. Following Zoeppritz equations, reflection amplitudes are then dependent on the incidence angle of the wave. This explains why AVA data, such as in Figure 1.3, often show amplitude variation as a function of incidence angle for a specific seismic reflector. In practice, an approximation to Zoeppritz equations is typically used in AVO analysis, as will be elaborated in Section 2.3.1. AVA gathers can also be used to predict elastic properties (velocitiy, impedance,

density, etc.) (Buland and Omre, 2003a) and/or petro-physical properties (porosity, water saturation, clay volume, etc.) (Grana, 2016) using inversion techniques. Obtaining reasonable inversion results for elastic properties of seismic AVA data will be a recurring theme in this thesis.

1.3 Probabilistic Inverse Problems

In geophysical exploration, data are most likely not directly measuring the physical quantity of interest. Instead, information about *model parameters* (or physical parameters) has to be inferred from the available data. Given a specific model parametrization $\mathbf{m}$, data $\mathbf{d}$ can be modeled using the governing physical equations g describing the relationship between model parameters and data:

$$\mathbf{d} = g(\mathbf{m}) \tag{1.1}$$

This unique data response is typically known as the *forward problem* or the *direct problem*. A visual representation of this relationship is shown in Figure 1.4. In most circumstances, if an experiment is repeated with the same instrumental setup, the data response is never exactly the same. In reality, observed data $\mathbf{d}_{\text{obs}}$ are always associated with information not explained by the forward model. Assuming that this unexplained information is additive noise $\boldsymbol{\varepsilon}$, the observed data can be computed as:

$$\mathbf{d}_{\text{obs}} = g(\mathbf{m}) + \boldsymbol{\varepsilon} \tag{1.2}$$

The *inverse problem* of predicting the values of $\mathbf{m}$ given some observed data $\mathbf{d}_{\text{obs}}$ is therefore inherently non-unique. I.e. since noisy data are not fully reproducible, neither will the inversion results be. The non-unique nature of the inverse

problem makes exact information about the values of model parameters unobtainable (Menke, 2012). There are two main ways of solving these inverse problems. The first option is to obtain the best possible model which minimizes the data misfit. The second option is to create an ensemble of models which all could potentially be a realistic depiction of the subsurface. Consider first trying to obtain the best possible model, i.e. a *deterministic* inverse problem. The typical property of deterministic inverse problems is their ill-posed nature (Bertero and Boccacci, 1998). The ill-posed nature (due to noisy data) leads to instabilities in the obtained solutions, and a certain tolerance towards misfit with the physical model must be accepted (Parker, 1994). These solutions can be obtained through optimization, regularization, and other data fitting techniques (see e.g. Nocedal and Wright (1999); Aster et al. (2004)). This way of formulating an inverse problem is deterministic in the sense that there will be only one final model which gives the lowest accepted misfit with the observed data. However, the choice of smoothing in e.g. regularization methods is a heuristic procedure based on arbitrary and subjective choices without reference to empirical observations (Mosegaard and Hansen, 2016).

In order to eliminate some of these subjective choices, the inverse problem can instead be formulated *probabilistically*. Acknowledging that information about model parameters and data is always incomplete, the model parameters are represented by stochastic variables and their corresponding probability distribution (Mosegaard and Hansen, 2016). The issue of ill-posedness of inverse problems within the probabilistic framework is summarized in the following sentence by Mosegaard and Tarantola (1995):

> "...there are no well-posed questions or ill-posed questions, but just

questions that have a probabilistic answer."

A probabilistic formulation of inverse problem theory allows combining several sources of information that can be quantified probabilistically. This is typically in the form of a *prior* probability $\rho_m(\mathbf{m})$ and a *likelihood* function $\mathrm{L}(\mathbf{m})$. The prior distribution describes prior independent information about $\mathbf{m}$. The likelihood describes how well the data $\mathbf{d}$, associated to a specific model $\mathbf{m}$ through the forward problem $\mathbf{d} = \mathrm{g}(\mathbf{m})$, match some observed data. The solution to such an inverse problem is the *posterior* probability density, $\sigma_m(\mathbf{m})$, which allows a full uncertainty characterization:

$$\sigma_m(\mathbf{m}) = k\,\rho_m(\mathbf{m})\mathrm{L}(\mathbf{m}) \tag{1.3}$$

where k is a normalization constant. In general, the likelihood function is given by (Tarantola and Valette, 1982)

$$\mathrm{L}(\mathbf{m}) = \int_{\mathbb{D}} \mathrm{d}\mathbf{d} \frac{\rho_d(\mathbf{d})\theta(\mathbf{d}|\mathbf{m})}{\mu_{\mathrm{D}}(\mathbf{d})} \tag{1.4}$$

where $\rho_d(\mathbf{d})$ reflect measurement uncertainties, $\mu_{\mathrm{D}}(\mathbf{d})$ is the homogeneous probability density, and $\theta(\mathbf{d}|\mathbf{m})$ reflect modeling uncertainties describing any imperfections in solving the forward problem, i.e. theoretical uncertainties. The right plot in Figure 1.4 shows a visual representation of the modeling uncertainties of the forward model. In most cases the modeling error is ignored (left plot in Figure 1.4), in which case Equation 1.4 reduces to $\mathrm{L}(\mathbf{m}) = \rho_d(\mathbf{d})$, and hence Equation 1.3 reduces to $\sigma_m(\mathbf{m}) = k\,\rho_m(\mathbf{m})\rho_d(\mathbf{d})$. This can be due to the difficulty of quantifying the modeling error through $\theta(\mathbf{d}|\mathbf{m})$, or it can relate

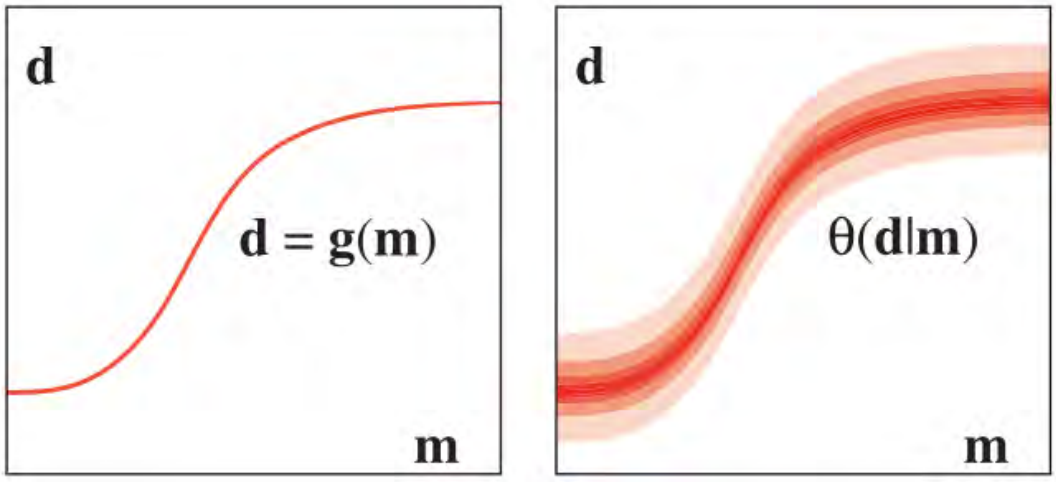


Figure 1.4: Left: Visual representation of the forward model (Equation 1.1) which describes the relationship between data and model parameters. Right: Relationship between data and model parameters with modeling uncertainties $\theta(\mathbf{d}|\mathbf{m})$ in the forward model. Figure taken from (Tarantola, 2005).

to the non-trivial problem of evaluating the integral in Equation 1.4, even in cases where $\theta(\mathbf{d}|\mathbf{m})$ is available.

Nevertheless, when solving a probabilistic inverse problem, a correct evaluation of the likelihood $L(\mathbf{m})$, and hence a good model of the uncertainties related to both measurements and theory, is essential. Without a correct assessment of uncertainties, and hence Equation 1.4, the variability of the posterior probability will also be incorrect (Gerstoft and Mecklenbräuker, 1998; Ulrych et al., 2001; Riedel et al., 2003; Tarantola, 2005; Bosch et al., 2010). Similarly, a good representation of the prior distribution $\rho_{\mathrm{M}}(\mathbf{m})$ is needed, because the posterior results will be equally affected by the likelihood and prior model through Equation 1.3. The problem of constructing realistic priors and data statistics is discussed in detail in e.g. Scales and Snieder (1997).

As mentioned earlier, the information of $\mathbf{d}$ and $\mathbf{m}$ should per definition be

independent (Tarantola, 2005). Two marginal probability densities $p(\mathbf{u})$ and $p(\mathbf{v})$ are deemed independent if the joint probability density $p(\mathbf{u}, \mathbf{v})$ equals the product of the two marginal probability densities

$$p(\mathbf{u}, \mathbf{v}) = p(\mathbf{u})p(\mathbf{v}) \tag{1.5}$$

The independence of information from different sources shall be explored in Section 5.3.1. It will be shown that this one demand is difficult to abide in practice. In statistics there is a long tradition for using data to obtain some idea of the prior distribution. From a practitioners viewpoint it is in fact almost impossible to keep a priori and data information completely separate. If the independence of information is violated, correlations between $\mathbf{d}$ and $\mathbf{m}$ will occur in the combined information and hence some degrees of freedom are consequently lost.

1.3.1 The Bayesian approach

Before moving on, a small remark should be made considering what is known as *Bayesian* inversion. The probabilistic framework introduced above follows from the principles of conjunction of information formulated and developed in Tarantola and Valette (1982). The theory is set up with a basic structure to the space of all probability distributions as characterized by Kolmogoroff (1933), with an additional requirement of *invariance of form* under coordinate transformation. The invariance ensures that two observers of the same experiment do not get different results just because they use two different coordinate systems. Bayesian inversion methods are usually developed using the notion of Bayes' theorem which is a simple consequence of conditional

probabilities (see e.g. Box and Tiao (1992)). A conditional probability is defined in the Kolmogoroff space. The idea of Bayes' theorem is that probabilities can be updated when new information arises. Bayes' theorem in its beautiful simplicity states the probability $p(A|B)$ of an event A occurring given that event B is true is:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \tag{1.6}$$

where $p(B|A)$ is the probability of B occurring given that A is true, and $p(A)$ and $p(B)$ are the probabilities of observing each event, A and B, independently of each other. $p(A)$ and $p(B)$ are also known as the marginal probabilities. Combining Bayes' theorem in Equation 1.6 with probability densities allow a reformulation in terms of vectors. Using the data and model parameters as vectors yields:

$$p(\mathbf{m}|\mathbf{d}) = c \times p(\mathbf{d}|\mathbf{m})p(\mathbf{m}) \tag{1.7}$$

where the marginal probability of the data $p(\mathbf{d})$ is written as a normalization constant $p(\mathbf{d}) = c^{-1}$ (Box and Tiao, 1992). In the Bayesian approach to inference the distribution $p(\mathbf{m})$ represents the density assigned to $\mathbf{m}$, i.e. the prior information. By introducing a likelihood probability distribution $L(\mathbf{m}|\mathbf{d})$ for the observed data information about $\mathbf{m}$, the prior knowledge of the model parameters $\mathbf{m}$ can be updated using the information contained in the likelihood. The updating mechanism is then given by Bayes' theorem (Givens and Hoeting, 2012):

$$p(\mathbf{m}|\mathbf{d}) = c \times p(\mathbf{m})L(\mathbf{m}|\mathbf{d}) \tag{1.8}$$

The careful reader will by now have noticed the resemblance to the solution of the inverse problem in Equation 1.3. Although one may arrive at the same

outcome, there is a difference in the two paths leading to these results. The difference lies in the distinction between conjunction of information (in this case the prior and likelihood) and updating information (prior) with a likelihood, as in Bayesian approach. Generally, the conjunction of two probability distributions $p_1(\mathbf{m})$ and $p_2(\mathbf{m})$ states (Tarantola and Valette, 1982):

$$p_1(\mathbf{m}) \wedge p_2(\mathbf{m}) = k\frac{p_1(\mathbf{m})p_2(\mathbf{m})}{\mu(\mathbf{m})} \tag{1.9}$$

where $\wedge$ means conjunction, k is a normalization constant, and $\mu(\mathbf{m})$ is a homogeneous probability density. $\mu(\mathbf{m})$ ensures that the result is invariant to changes in coordinate system. This in principle allows a solution in which one can change coordinate system and still obtain a solution which is consistent. In the Bayesian formulation the coordinate system in which $p(\mathbf{m})$ and $L(\mathbf{m}|\mathbf{d})$ in Equation 1.8 is found, is implicitly selected through the formulation itself. The implicit coordinate system within the Bayesian framework leads to the conditional probabilities used in Bayes' theorem. In fact, Bayes theorem can be computed as a special case of the conjunction of information as shown in (Tarantola and Valette, 1982). But, one cannot deduce the conjunction of information from Bayes' theorem without introducing extra information. In that sense the probabilistic formulation of Tarantola and Valette (1982) is more general, as the probability density distributions introduced here are invariant. Since most inverse problems are formulated in a fixed coordinate system, this has no apparent effect in practice, and both methods are applicable and would yield the same result. To explore this further the reader is referred to Mosegaard and Hansen (2016). In the following there will be made no further distinction

between these two ways of formulating the problem. The general solution in the probabilistic formulation of Tarantola and Valette (1982) is preferred in this thesis, because it contains a probabilistic description of the imperfect forward models, as shown in Figure 1.4.

1.4 Linear Inverse Problems

If the forward relationship between data and model parameters is described by a linear mapping function $\mathbf{G}$, the forward problem in Equation 1.1 can be written as:

$$\mathbf{d} = \mathbf{G}\mathbf{m} \tag{1.10}$$

The simplest inverse problems to solve generally involve linear forward models. Rewriting Equation 1.10 and using observed data ($\mathbf{d}_{\text{obs}}$) one can obtain an estimate of the model parameters $\mathbf{m}_{\text{est}}$:

$$\mathbf{m}_{\text{est}} = \mathbf{G}^{-1}\mathbf{d}_{\text{obs}} \tag{1.11}$$

Although this representation is slightly simplistic, this is essentially the idea of linear deterministic inversion. It also gives a good insight to the problem with instabilities in the solution of $\mathbf{m}_{\text{est}}$ as presented in the previous section. The solution can be achieved by introducing a regularization parameter as proposed by Tikhonov (1963), or solved by some iterative updating method (Nocedal and Wright, 1999; Aster et al., 2004). A popular solution could be *Generalized Linear Inversion* (GLI) which is an iterative technique refining an initial user-supplied guess until the response matches sufficiently (Cooke and Schneider, 1983). The refining part typically involves finding the least squares error solution, i.e.

minimizing the second norm of the residual $\|\mathbf{G}\mathbf{m}_{\text{est}} - \mathbf{d}_{\text{obs}}\|$ (Aster et al., 2004).

The solution to the probabilistic inverse problem in Equation 1.3 can also be obtained with a linear forward model using a general least-squares solution as described in Tarantola and Valette (1982). To obtain the least-squares solution in this case, a multivariate Gaussian assumption of the distribution of both observations and model parameters is needed, i.e. they can be described by a mean and a covariance function (Appendix B). The prior distribution of n_m model parameters is then described by $\rho_m(\mathbf{m}) \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}_m, \mathbf{C}_m)$. The n_d data points are distributed by $\mathbf{d}_{\text{obs}} \sim \mathcal{N}_{n_d}(\mathbf{d}_{\text{obs}}, \mathbf{C}_D)$. Implying that the noise is a zero-mean *centred* Gaussian distribution: $\boldsymbol{\varepsilon} \sim \mathcal{N}_{n_d}(0, \mathbf{C}_D)$. The solution to the inverse problem can then be analytically obtained as a multivariate Gaussian probability distribution $\mathbf{m}_{\text{est}} \sim \mathcal{N}_{n_m}(\widetilde{\boldsymbol{\mu}}_m, \widetilde{\mathbf{C}}_m)$, with mean:

$$\widetilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_m + \widetilde{\mathbf{C}}_m \mathbf{G}^{\mathsf{T}} \mathbf{C}_D^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{G}\boldsymbol{\mu}_m) \tag{1.12}$$

and covariance:

$$\widetilde{\mathbf{C}}_m = \left(\mathbf{G}^{\mathsf{T}} \mathbf{C}_D^{-1} \mathbf{G} + \mathbf{C}_m^{-1}\right)^{-1} \tag{1.13}$$

The probabilistic least-squares solution in Equation 1.12 and 1.13 can also be obtained using conditional Gaussian probabilities (e.g. Buland and Omre (2003a)).

1.5 Inversion of Seismic Data

A wide variety of inversion methods for seismic data exists today. Inversions typically try to obtain reasonable estimates of subsurface properties. No prevalent method is yet to emerge which is supported by the broad collection of

approaches that now exist within industry as well as academia (see e.g. Yilmaz (2001)). The deterministic GLI method is probably the most commonly used seismic inversion technique (Cooke and Cant, 2010). GLI has been the industry standard for many years (see e.g. Cooke and Schneider (1983)) and its popularity is possible due the computational efficiency compared to e.g. probabilistic inversion methods. *Full Waveform Inversion* (FWI), which is also a deterministic inversion technique, has seen a rise in popularity throughout the last decade (Datta and Sen, 2016). In principle, a posterior covariance can also be estimated, i.e. solving the FWI probabilistically. However, in practice it is not so easy due to heavy computational demands in solving the forward problem. The basic deterministic formulation follows from GLI as a data-fitting least-squares minimization problem. FWI is in contrast to GLI based on a non-linear forward model, as in Tarantola (1986). FWI incorporates all wave phase information, which sets it apart from traditional seismic data inversion methods. By the inclusion of the full wave field in the forward modeling, FWI is trying to limit wave-number sensitivity issues in the regular deterministic inversion approaches (Virieux et al., 2017). A detailed review of the method can be found in Virieux and Operto (2009).

The popularity of deterministic inversion methods as GLI and FWI can also potentially be ascribed to the typical "geological" mindset within the exploration industry. In geology one model of the subsurface is usually sufficient. For instance, consider a geological outcrop. The geological interpretation of this outcrop is deterministic and only depends on the observers interpretation. Layer boundaries, layer types, facies etc. are all determined without uncertainty in a single final geological model, which represents all available information of the

outcrop. Similarly, this mindset leads to only one true model of the Earth. Accepting the fact that there exists not only one best model, but rather an ensemble of equally probable models is not an easy transition. This also goes for decision making in industry. How are one supposed to make a decision whether to drill or not based on several models? A shift in paradigm to (computational) time consuming probabilistic methods is also a practical challenge. However, the need for reliable risk assessments within the exploration seismic community have slowly increased the interest in probabilistic methods (Cooke and Cant, 2010). While deterministic methods have been successful, they have their limitations in seismic inversion. Mainly, there is no way to ensure correct uncertainty quantification as well as ensuring that the final model is consistent with all information available.

Due to these limitations, probabilistic inversion methods have gained increased attention in later years. Large data sets have so far been an issue due to the computational cost of computing the full uncertainty. The linearization of a seismic forward model provided by Buland and Omre (2003a) allowed a computationally efficient method of obtaining probabilistic results, even for large data sets. The linearization of the seismic forward has led to a wide variety of probabilistic inversion techniques. In recent years, it has also been used in combination with a linearized rock physics model to introduce Bayesian linearized rock-physics inversion (Grana, 2016). One of the main issues with probabilistic linearized seismic inversion is the assumption of Gaussian distributions of both data residuals and model parameters. The Gaussian distributions can be transformed (e.g. normal score transformation) in order to produce non-Gaussian results. The Gaussian random fields can also be

generalized in order to induce skewness or multi-modality into the models (Rimstad and Omre, 2014a). The linear seismic forward has for instance been used with Gaussian mixture random fields to obtain lithological inversion results (Grana et al., 2017). Alternatively, a sampling method can be applied in order to allow for both a non-linear forward relation and non-Gaussian distributions (Mosegaard and Tarantola, 1995).

Even though the linear seismic forward model has been extremely popular in recent years, it is still a linear approximation to a non-linear problem. The reliability of the linearization still remains unresolved. For probabilistic methods in general there is also an issue regarding how to obtain reliable probability distributions describing data residuals (including modeling uncertainties) and model parameters. When working with data, one inevitably introduce distortions in the results, which can be attributed to presumptions and/or methodical errors. These distortions will in this thesis be referred to as biases. This Ph.D. thesis explores some of these biases and issues of obtaining trustworthy posterior uncertainties, when doing probabilistic linear inversion of reflection seismic data. The center of attention is biases arising from modeling errors in the forward problem, but work is also carried out in regards to non-stationarity in the prior model and the noise model in general.

"Go back?" he thought. "No good at all! Go sideways? Impossible! Go forward? Only thing to do! On we go!" So up he got, and trotted along with his little sword held in front of him and one hand feeling the wall, and his heart all of a patter and a pitter."

J.R.R. Tolkien (1937), The Hobbit

2

Forward models

Just as the hobbit did, one must first be able to go forward in order to solve the inverse problem. For a mathematician there is a certain amount of ambiguity when describing one problem as the inverse problem and one as the forward problem. In mathematics, there is an exchanging role of data and model parameters where it seems quite arbitrary to decide a forward (or direct) relation and an inverse relation respectively (Bertero and Boccacci, 1998). For a physicist however, one relation is deemed more fundamental. Data are measured as a result of some physical process. If the data were not measured, they would not exist. It is this causality which traditionally has deemed one relationship more fundamental

than the other. Forward modeling is essentially the physics of any inverse problem, wherein a link between model parameters and the measured data is presented. As put by Tarantola (2005):

> "Forward modeling: discovery of the physical laws allowing us, for given values of the model parameters, to make predictions on the results of measurements on some observable parameters."

Recall the forward model defined in Equation 1.1.

$$\mathbf{d} = g(\mathbf{m})$$

Raw seismic data $\mathbf{d}$ are, as stated in the introduction, recordings of pressure variation over time at different locations. These data arise from recordings of energy traveling from a source through the subsurface. We parametrize the problem as a subsurface with some elastic properties $\mathbf{m}$. In this study, the elastic properties used is the pressure wave velocity (v_p), shear wave velocity (v_s), and density (ρ). The perhaps most correct forward model g to obtain seismic data is to calculate the propagation of the full wave front from a known source in the subsurface through space and time. This is also known as *Full Waveform Modeling* (FWM). The wave propagation can for instance be simulated by approximating the elastic wave equation using a finite difference scheme. This technique has previously been successfully applied in order to simulate seismic shot gathers in e.g. Charara et al. (1996). For a general review of the use of finite-difference in modeling of seismic wave propagation please refer to Moczo et al. (2007). For AVO cases, the raw seismic data from the FWM need to be processed such that they resemble AVO data. The processing then implicitly

becomes part of the forward model and involves compensation for spherical divergence and *Normal MoveOut* (NMO) correction. If an AVA gather is needed, an offset to angle conversion is also applied. Alternatively, seismic gathers can also be modeled using a convolution between a proper wavelet and a reflectively series. For AVA data the reflectively series is angle-dependant. In the following section a short review is given of the FWM used in this thesis. This is followed by the introduction of the convolutional forward model along with different options of calculating the necessary reflectivity series. Finally, the processing techniques used in order to convert raw seismic data to AVA seismic data are presented.

2.1 Full Waveform Modeling

The elastic wave equation is used to describe the propagation of waves through an elastic medium. In order to obtain the elastic wave equation (used as the seismic wave equation), the momentum equation must be considered as well as a definition of the relationship between stress (i.e. the force per area) and strain (i.e. the deformation of the medium).

2.1.1 Elastic wave equation

From Appendix C the momentum equation for a continuous medium is established as:

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \partial_j \tau_{ij} + f_i \tag{2.1}$$

where u_i denotes displacements, ρ is density, τ_{ij} is the stress-tensor, and f_i is body forces (e.g. gravity) acting on the medium. Assuming a linear elastic relationship between stress and strain, the stress tensor in Equation 2.1 can be replaced by the

strain tensor:

$$\tau_{ij} = \begin{bmatrix} \lambda tr(\mathbf{e}) + 2\mu e_{11} & \lambda 2\mu e_{12} & \lambda 2\mu e_{13} \\ \lambda 2\mu e_{21} & \lambda tr(\mathbf{e}) + 2\mu e_{22} & \lambda 2\mu e_{23} \\ \lambda 2\mu e_{31} & \lambda 2\mu e_{32} & \lambda tr(\mathbf{e}) 2\mu e_{33} \end{bmatrix} \tag{2.2}$$

where λ and μ are Lamé parameters and ε_{ij} is the strain tensor, and $\mathbf{e} = [\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}]$ is the diagonal elements of the strain tensor. This assumption holds true if the medium is homogeneous and isotropic. Using index notation the expression of the strain tensor can be written as:

$$\tau_{ij} = \lambda \varepsilon_{kk} \delta_{ij} + 2\mu \varepsilon_{ij} \tag{2.3}$$

where $\varepsilon_{kk} = \mathbf{e}$ and δ_{ij} is the Kronecker delta (a function providing ones only where indices are equal, otherwise zero). The strain tensor is described using the spatial derivative of the deformation u (Aki and Richards, 2002; Shearer, 2009):

$$\varepsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \tag{2.4}$$

In total, Equation 2.1, 2.3, and 2.4 form a set of coupled differential equations which describes the (seismic) elastic wave equation.

2.1.2 Finite difference formulation

The full waveform seismic modeling is performed by solving the partial derivatives in the elastic wave equation using a finite-difference scheme with a staggered grid. The finite difference method is presented in Appendix E. The finite-difference solver is designed based on the work presented in Virieux

(1986); Levander (1988). The elastic parameters are defined on a 2D staggered-grid. The finite difference implementation follows a first order Taylor series truncation as in Equation E.6. With the proposed finite-difference scheme, the main source of error is numerical dispersion (Bohlen et al., 2015). This error is an effect of the discretization of the partial derivatives. The problem can be lessened by changing the discretization in time Δt according to Courant's constraint $C_{\max}$:

$$\frac{v_p \, \Delta t}{\Delta x} \leq C_{\max} \tag{2.5}$$

where Δx is the discretization in space. Courant's constraint $C_{\max}$ basically sets an upper level of accepted numerical error in each simulation based on the velocity in the model and the discretization hereof. Alternatively, the amount of points to be included in the finite difference operators (i.e. the order of accuracy) both in space and time can be changed. More sophisticated finite difference implementations of the first-order coupled elastic equations with higher accuracy are available (Bohlen et al., 2015; Igel, 2017). However, higher accuracy implementations are also computationally demanding. Choosing a suitable level of accuracy presents a trade-off with wall-clock computational efficiency. For the general purposes in the present work, a sampling accuracy in fourth order space and second order time is deemed sufficient.

In order to suppress waves at the boundaries of the model a strategy of *Perfectly Matched Layers* (PML) is adapted from Collino and Tsogka (2001). The general idea is to split all displacements and stresses into two components. For a continuous formulation, this has the property of not generating artificial reflections at the interface between the dampening layers and the unaffected area. Due to the dampening factor in the discretization of the coupled elastic wave

equation in the finite difference scheme, small reflections are however bound to occur. A trade-off therefore exists between magnitude of the dampening factor and the size of the discretization step (the finite difference). However, the dampening factor cannot be chosen too low, otherwise a large PML boundary is needed in order to dampen unwanted waves.

In this thesis, the PML have a dampening factor of 0.001 as suggested by Collino and Tsogka (2001) for isotropic elastic media in order to avoid any unwanted reflections in the PML boundary. Heuristically, 40 grid nodes in the PML are chosen. Smaller boundaries work pretty well from approximately 20 grid nodes and upwards in size. However, to enable an almost perfect absorption at the boundaries, 40 grid nodes are chosen as a precaution. Albeit, this is slightly more wall-clock time consuming as a bigger grid is needed. An example of wave propagation when applying the finite difference forward model is demonstrated in Figure 2.1. The 1D vertical elastic log in Figure 2.2 is used to generate the 2D elastic field by applying it at all horizontal locations. The subsurface thereby becomes a stack of horizontal, homogeneous, and isotropic layers. This construction of the elastic properties has the added benefit that all recordings of reflected waves at the surface will share a CRP. More specifically, all reflected waves will have a *common depth point* (CDP). In a real seismic survey, the raw data must be stacked during the processing phase in order to obtain a CDP shot gather.

The divergence and curl of the finite difference solution, and hence the propagation of pressure and shear waves is shown as a series of time snapshots in Figure 2.1. Shear waves are not present in the first water layer, but arises when the

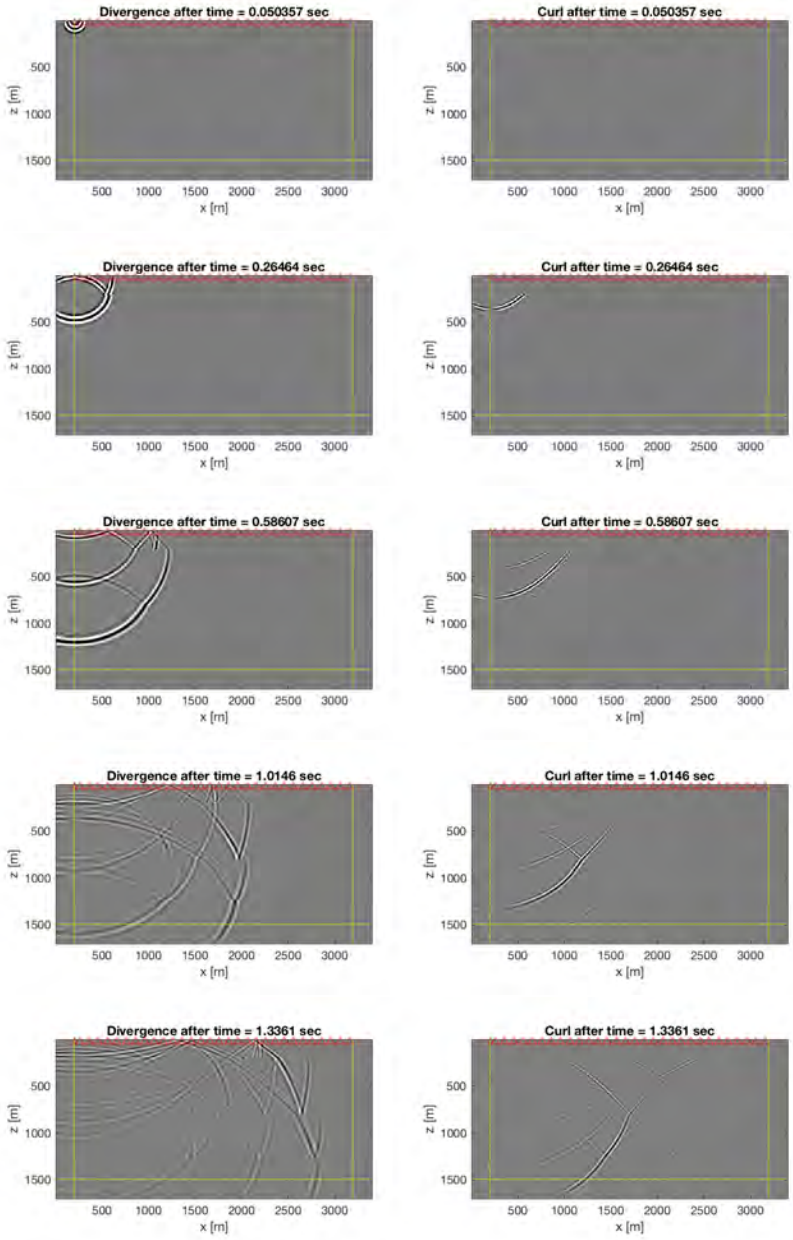


Figure 2.1: Snapshots of wave propagation using full waveform modeling. The figures in the left column show the divergence (pressure waves) of the wave field, and the right column shows the curl (shear waves) of the wave field. The red stars in the top of each plot indicate the location of receivers, and the yellow lines illustrate the position of absorbing PML boundaries. The elastic log in Figure 2.2 is used to construct the elastic model.

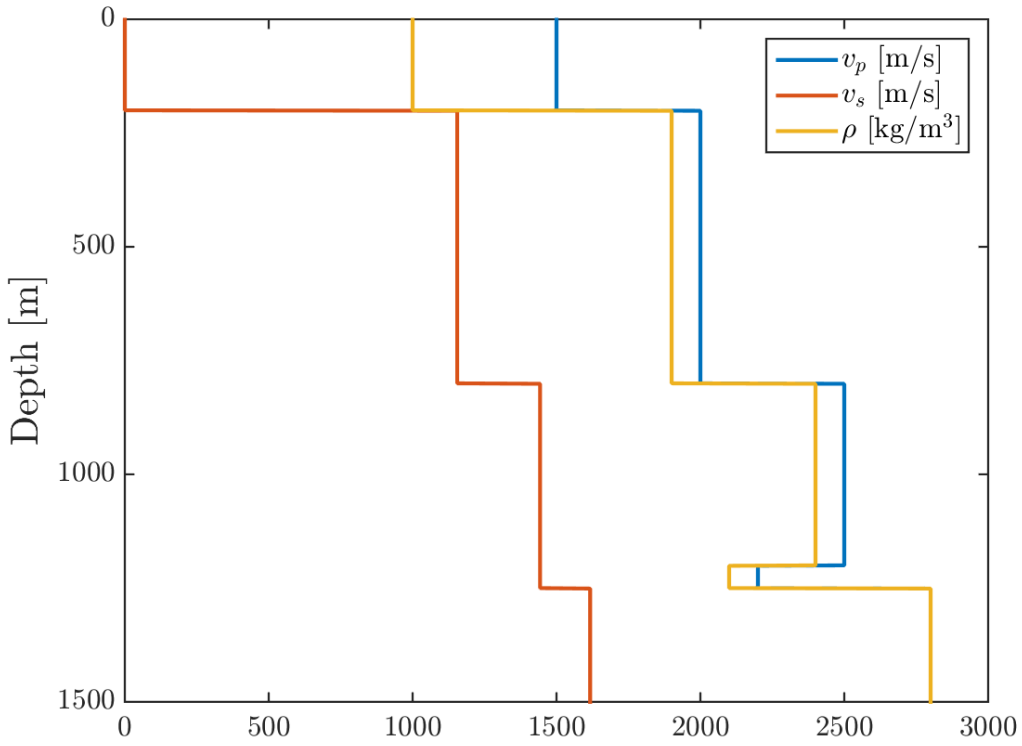


Figure 2.2: Simple elastic log (v_p, v_s, and ρ) used for full waveform modeling in Figure 2.1. The top layer is water with zero shear wave velocity because water is a liquid.

wave has propagated to 200 m depth in accordance with the elastic log in Figure 2.2. The PML boundaries for this setup are indicated with yellow lines on the sides and on the bottom. The top boundary is a free surface. The PML boundaries are reasonable at absorbing the waves, as indicated by the snapshots. The wave propagation also reveals that the pressure waves are propagating faster than the shear waves both in accordance with the elastic log and the theory of seismic velocities (see e.g. Aki and Richards (2002); Shearer (2009)). The red stars at the top of the setup indicate locations of receivers. If the divergence (i.e. the pressure) is recorded at these locations for each time step in the full waveform model, a synthetic seismic shot gather can be emulated. The raw seismic shot gather recorded from the finite difference forward run in Figure 2.1, is presented in Figure 2.3. The resulting seismic shot gather is denoted $\mathbf{d}_{\text{raw}}$ and the FWM which is then considered the forward model g_{fwm}, produces this data from an elastic 1D log $\mathbf{m}$:

$$\mathbf{d}_{\text{raw}} = g_{fwm}(\mathbf{m}) \tag{2.6}$$

2.2 Convolutional Model

Simulating shot gathers using FWM is computationally costly (Virieux and Operto, 2009), especially if the full wave field is to be computed at each iteration in e.g. Monte-Carlo sampling techniques. A seismic trace can instead be modeled as a convolution between a reflectivity series (i.e. the earth response) and a seismic wavelet (Yilmaz, 2001). This is known as the *convolutional model*, which assumes that the earth is composed of a series of flat, homogeneous, and isotropic

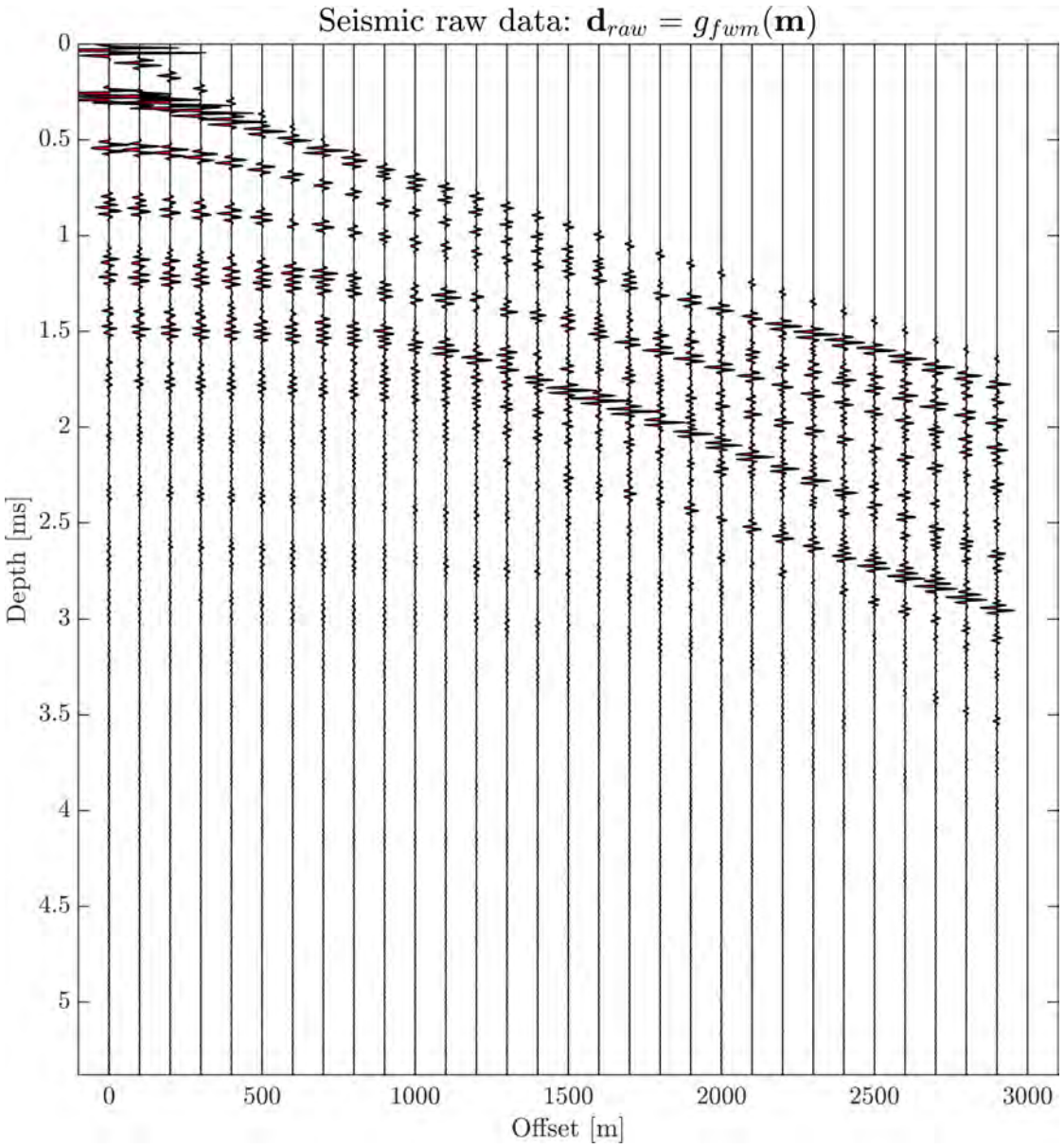


Figure 2.3: Synthetic seismic shot gather obtained using FWM on the elastic log in Figure 2.2. Snapshots of the executing code and the receiver positions can be seen in Figure 2.1. The offset in meters is displayed on the abscissa-axis. The depth in TWT is displayed on the ordinate-axis.

layers (Downton, 2005). The origin of the convolutional model for generating seismic traces is not well known. The litterature is also very sparse on the physical rightness of the convolutional model, although it can be shown that the convolutional model is a high-frequency approximation to the linearized wave equation (Winslow, 2000). Regardless, the convolutional model remains a very popular, if not the most popular, method for computing seismic data, primarily because it allows a computationally efficient alternative to full-waveform modeling. The convolutional model for one seismic trace $S(t)$ can be stated as:

$$S(t) = W(t) \circledast R(t) \equiv \int_0^{t_s} W(\tau) R(t-\tau)\, \mathrm{d}\tau \tag{2.7}$$

where $R(t)$ is the reflectivity series, $W(t)$ is the wavelet (source-time function), t_s is the duration of the source input, and $\circledast$ denotes a convolution. The source-time function is usually either estimated from observed seismic data or approximated by a Ricker-wavelet. The Ricker-wavelet is a common pulse shape used in reflection seismic modeling (Shearer, 2009). The Ricker-wavelet can be calculated as (Aki and Richards, 2002):

$$W(t) = (1 - 2\pi^2 f_c^2 t^2) \exp\{-\pi^2 f_c^2 t^2\} \tag{2.8}$$

where f_c is the center (i.e. peak) frequency in the frequency spectrum of the Ricker wavelet.

2.3 Reflection Coefficients

In so-called "reflecticity inversion", the reflectivity series $R(t)$ used in the convolutional model (Equation 2.7) is the desired quantity to retrieve in the

inversion (see e.g. Downton and Lines (2004); Ulrych and Sacchi (2005)). Reflectivity inversion has been shown to provide useful inversion results for e.g. interpretation purposes (Yilmaz, 2001). However, the reflection coefficients themselves do not carry much physical meaning. In this thesis, retrieval of elastic parameters (v_p, v_s, and ρ) is instead the desired outcome of the inversion. The reflectivity series must therefore be coupled to the elastic properties of the subsurface through some function, in order to make use of the convolutional model.

The reflectivity used in the convolutional model is that arising from reflections of pressure waves R_{pp}. In order to couple elastic properties to reflection coefficients, several assumptions are needed. As for the momentum equation (Equation C.12), used in the full wave-form modeling, the subsurface must behave isotropic and elastic (Mavko et al., 2009). Furthermore, the wave front is assumed plane. This assumption is increasingly more accurate with increasing distance to the origin (i.e. the source) of the energy wave because the curvature is diminishing.

Consider a reflection R_{pp} at a single interface between two elastic media with different elastic properties. If an incoming pressure wave is "hitting" the interface between the two media at an angle perpendicular to the interface, the reflection coefficient of the interface is described as:

$$R_{pp} = \frac{Z_2 - Z_1}{Z_2 + Z_1} \tag{2.9}$$

where $Z = v_p \rho$ is the acoustic impedance of each medium. This relationship is sometimes also referred to as the *zero-offset* reflection. The full reflectivity series $R(t)$ can simply be computed for an entire elastic log (with depth in TWT) by

applying Equation 2.9 for all layer interfaces. The reflection coefficient in Equation 2.9 is large if the elastic contrast between the two media is large. This result can be used as a general rule of thumb for more complex systems.

2.3.1 Zoeppritz equations

If the incoming wave is hitting the interface at a different angle than perpendicular to the interface, the simple relationship in Equation 2.9 does not hold. To describe the full partitioning (in amplitudes of reflected and transmitted waves) of a plane-wave hitting the interface, Zoeppritz (1919) derived a set of equations from Snell's law. Considering only the amplitudes of the waves arising from an incident pressure wave, the reflection coefficients for reflected (R_{pp} and R_{ps}) and transmitted waves (T_{pp} and T_{ps}) as a function of incident angle of the incoming wave (θ) is given by (Aki and Richards, 2002; Mavko et al., 2009):

$$\begin{bmatrix} R_{\mathrm{pp}} \\ R_{\mathrm{ps}} \\ T_{\mathrm{pp}} \\ T_{\mathrm{ps}} \end{bmatrix} = \begin{bmatrix} -\sin\theta_1 & -\cos\varphi_1 & \sin\theta_2 & \cos\varphi_2 \\ \cos\theta_1 & -\sin\varphi_1 & \cos\theta_2 & -\sin\varphi_2 \\ \sin 2\theta_1 & \frac{v_{p1}}{v_{s1}}\cos 2\varphi_1 & \frac{\rho_2 v_{s2}^2 v_{p1}}{\rho_1 v_{s1}^2 v_{p2}}\cos 2\varphi_1 & \frac{\rho_2 v_{s2} v_{p1}}{\rho_1 v_{s1}^2}\cos 2\varphi_2 \\ -\cos 2\varphi_1 & \frac{v_{s1}}{v_{p1}}\sin 2\varphi_1 & \frac{\rho_2 v_{p2}}{\rho_1 v_{p1}}\cos 2\varphi_2 & \frac{\rho_2 v_{s2}}{\rho_1 v_{p1}}\sin 2\varphi_2 \end{bmatrix} \begin{bmatrix} \sin\theta_1 \\ \cos\theta_1 \\ \sin 2\theta_1 \\ \cos 2\varphi_1 \end{bmatrix} \tag{2.10}$$

The angles θ_1, θ_2, φ_1, and φ_2 represent angles of outgoing waves (see e.g. Figure A.1). These angles are related to the incident angle θ according to Snell's law

(Appendix A):

$$\theta_1 = \theta, \tag{2.11}$$

$$\theta_2 = \sin^{-1}\left(\frac{v_{p2}}{v_{p1}}\sin(\theta)\right), \tag{2.12}$$

$$\varphi_1 = \sin^{-1}\left(\frac{v_{s1}}{v_{p1}}\sin(\theta)\right), \tag{2.13}$$

$$\varphi_2 = \sin^{-1}\left(\frac{v_{s2}}{v_{p1}}\sin(\theta)\right) \tag{2.14}$$

Because Zoeppritz equations define a relationship between elastic parameters and reflection coefficients which is angle dependent, these equations can also be used to model AVA data. Evaluating the AVA forward problem using Zoeppritz equations amounts to first computing the $R_{pp}(\theta, t)$ reflection coefficients using Zoeppritz equations for each desired angle stack. These series of reflection coefficients are convolved with a wavelet for each angle using Equation 2.7. Since the convolution is one dimensional, different wavelets can be applied for each angle. This process of obtaining an AVA shot gather is inherently a non-linear process as the Zoeppritz equations are non-linear.

2.3.2 Approximate forms of Zoeppritz equations

Due to the non-linear nature of Zoeppritz equations, several authors have presented and suggested different linearizations, and hence approximations, to Zoeppritz equations (Aki and Richards, 1980; Stolt and Weglein, 1985; Shuey, 1985; Wang, 1999). Another motivation for approximating the Zoeppritz equations is the complexity of the resulting equations. As pointed out by Mavko et al. (2009), the approximations typically prove more physically insightful than

the Zoeppritz equations. By assuming a small contrast between the layer properties, Aki and Richards (1980) provided a successful approximation to Zoeppritz equations. For retrieval of the $R_{pp}(\theta)$ coefficients the following relationship holds:

$$R_{pp}(\theta) \approx a_{v_p}(\theta)\frac{\Delta v_p}{\overline{v_p}} + a_{v_s}(\theta)\frac{\Delta v_s}{\overline{v_s}} + a_\rho(\theta)\frac{\Delta \rho}{\overline{\rho}} \tag{2.15}$$

where the coefficients are given by:

$$\begin{aligned} a_{v_p}(\theta) &= \frac{1}{2\cos^2\theta} \\ a_{v_s}(\theta) &= -\frac{4\overline{v_s}^2}{\overline{v_p}^2}\sin^2\theta \\ a_\rho(\theta) &= \frac{1}{2}(1 - 4\frac{\overline{v_s}^2}{\overline{v_p}^2}\sin^2\theta) \end{aligned} \tag{2.16}$$

where $\overline{v_p}$, $\overline{v_s}$, and $\overline{\rho}$ represent average pressure wave, shear wave, and density over the interface, while Δv_p, Δv_s, and $\Delta\rho$ represent elastic contrasts over the interface. A detail which is sometimes overlooked is that θ does not represent incident angle in the Aki and Richards (1980) formulation, but instead is the mean of transmitted and incidence angle as noted by Downton and Ursenbach (2006). The transmitted wave requires prior knowledge of the velocity field which is usually not readily available in inversion cases. The reflection angle θ is therefore often approximated as the incidence angle (Mavko et al., 2009). This point is also discussed a bit further in Appendix H.4.

In this thesis, the main approximate form of Zoeppritz equations is provided by Stolt and Weglein (1985). This formulation expands the Aki and Richards

approximation, so that the reflection coefficients are now also time-dependent:

$$R_{pp}(t,\theta) = a_{v_p}(t,\theta)\frac{\partial}{\partial t}\ln v_p(t) + a_{v_s}(t,\theta)\frac{\partial}{\partial t}\ln v_s(t) + a_{\rho}(t,\theta)\frac{\partial}{\partial t}\ln \rho(t) \tag{2.17}$$

The difference terms in Equation 2.15 (Δv_p, Δv_s, and $\Delta\rho$) are in Equation 2.17 substituted with the partial derivative of the logarithmic value of each material parameter, e.g. $\frac{\partial}{\partial t}\ln v_p(t)$ replaces $\frac{\Delta v_p}{\overline{v_p}}$. This substitution is valid only for small contrasts in the elastic parameters. Using the Stolt and Weglein (1985) formulation, the small contrast approximation is thereby reinforced. For a constant or slowly varying background velocity model, $\overline{v_p}(t)$ and $\overline{v_s}(t)$, which are the coefficients in Equation 2.16, can be precalculated before inversion. Precalculating the coefficients enable them to be put on matrix form ($\mathbf{A}$), and thus a linear mapping between the time derivative of the logarithm of the elastic parameters exists in the following form:

$$\mathbf{R} = \mathbf{A}\mathbf{m}' = \mathbf{A}\mathbf{D}\mathbf{m} \tag{2.18}$$

where $\mathbf{D}$ is a differential matrix. By constructing a matrix containing the wavelets for each model parameter $\mathbf{W}$, the convolution is coupled to the reflection coefficients:

$$\mathbf{d}_{\text{ava}} = \mathbf{W}\mathbf{A}\mathbf{D}\mathbf{m} = \mathbf{G}\mathbf{m} \tag{2.19}$$

where the forward model is given by $\mathbf{G} = \mathbf{W}\mathbf{A}\mathbf{D}$. A full linear forward relationship then exists between AVA data $\mathbf{d}_{\text{ava}}$ and model parameters $\mathbf{m}$. This result and approach is used frequently throughout this thesis because the linear forward in Equation 2.19 enables linear probabilistic inversion of AVA data. The

reader is referred to Buland and Omre (2003a) for specific construction of the matrices **W** and **A**. Equation 2.19 is used jointly with Equation 2.26 for the same elastic log in Appendix H.5 to assess differences between processing raw data and using the convolutional model as a forward model.

2.4 Processing

Processing raw seismic reflection data into any form of data set is not a trivial task. The goal of the processing also depends on the seismic discipline for which it should be used. The raw seismic data can be processed for visual purposes (i.e. interpretation) or in compliance with the convolutional model for AVO analysis, or for inversion purposes. A detailed walk-through of the many different processing steps is beyond the scope of this thesis but can be found in e.g. Yilmaz (2001). The focus here is the processing of raw seismic data to AVA seismic data with the purpose of inversion for elastic parameters. There are three important aspects of a processing sequence for AVO analysis in general (Feng and Bancroft, 2006):

- Relative amplitudes of the seismic data must be preserved throughout the processing sequence as all AVO methodologies rely on the variation as a function of offset.
- The processing should make as minimal an impact as possible on the original frequency spectrum during any form of band-pass filtering, i.e. filters should be broad in frequencies.
- CRP gathers should be used to derive any AVO attributes in pre-stack inversion

multiple interference	array effects
temporal tuning	instrumentation
mode conversions	source strength & consistency
transmission losses	receiver coupling
effect of overburden	RNMO
reflector curvature	processing algorithms
spherical divergence	NMO stretch
phase changes with offset	geology
noise and interference	ground roll
attenuation, dispersion, absorption	random noise

Figure 2.4: Factors influencing the amplitudes of reflectivity as a function of offset. Table taken from Downton et al. (2000)

Castagna and Backus (1993); Downton et al. (2000) recognize the trade-off between preserving the relative amplitudes and reducing noise as the most important aspect in processing AVA data. Downton et al. (2000) offers a brief summary of some of the factors influencing the amplitudes of reflectivity as a function of offset. These factors are displayed in Figure 2.4. Processing of real world seismic data therefore requires a lot of expertise when balancing out these different factors. Ideally, the processor is able to preserve all amplitude changes related to geology while removing all possible noise sources. However, processing itself usually leads to introduction of some biases in the form of signal corruption. Due to the amount of raw seismic data available in a seismic survey, the processing has to be automated for the majority of the survey. In practice, only a few different raw data sections are processed individually in order to setup a quasi-robust processing sequence for the entire data set. Quasi-robust implies, that the processing sequence most likely work more efficiently for some sections

of the seismic data than others, depending on the factors displayed in Figure 2.4.

An additional challenge in the processing of seismic data is the heuristic nature of the method in general. The processing of a specific seismic section depends on subjective decisions made by the processor. These subjective choices are usually based on previous experience and on what looks ”realistic” in terms of expected geology and seismic reflectors. The problems regarding processing of seismic data is covered in Section 5.2.

The essential processing steps when processing raw seismic data $\mathbf{d}_{\text{raw}}$ from the proposed FWM (Equation 2.6) into AVA gathers are covered in the following. The raw data from the FWM are an idealized version of what would be expected in a real world seismic data set. Therefore, some additional processing steps would be needed for real data in order to suppress noise that are not present in the FWM results. The noise on $\mathbf{d}_{\text{raw}}$ will exclusively be attributed to modeling errors from the FWM. Those can roughly be divided into two categories: The modeling errors from inaccurate physics (mainly describing the wave propagation with the elastic wave equation), and the simulation errors due to the finite difference method. The fourth order space and second order time accuracy combined with the PML strategy for unwanted wave suppression makes the noise from the finite difference modeling g_{fwm} supposedly negligible compared to errors from processing. Additionally, using a PML boundary at the top surface of the modeling grid allows the suppression of multiples directly in the forward modeling and not in the subsequent processing. Multiples are usually a big issue in seismic processing (Yilmaz, 2001). Because $\mathbf{d}_{\text{raw}}$ is considered basically ”noise-free”, the following processing steps can be thought of as the minimal (or essential) steps in transforming the raw data into AVA shot gathers.

2.4.1 Correction for spherical divergence

As a wave propagates away from the source it spreads out spherically as seen in Figure 2.1. The wave front is diverging as a function of distance r to the source. A proper compensation for this spherical effect is an absolute necessity if reflection amplitudes are to be used in further analysis (Newman, 1973). This is also known as correction for *geometrical spreading*.

In a simple case with a homogeneous medium, the amplitude should attenuate proportionally with distance (the radius from the source point) squared r^2. However, since the subsurface is made of complex layering and is not a homogenous medium, this is considered more a rule of thumb for the diverging wave field. Velocities usually increase with depth causing a more rapid decay in amplitudes than $1/r^2$ (Yilmaz, 2001). To compensate for the wave attenuation, a divergence factor D_{div} can be introduced. Newman (1973) introduced a full description of the wave attenuation as a function of horizontal distance to the source x (i.e. the offset) and incidence angle θ

$$D_{div}(\theta, x) = \frac{2x}{\tan^2(\theta_1)} \sum_{i=1}^{n} \frac{d_i \sin(\theta_i)}{\cos^3(\theta_i)} \tag{2.20}$$

where θ_i and d_i are the incidence angle and thickness of the ith layer respectively. To apply Equation 2.20, a priori knowledge of interval velocity and the thickness of each layer must exist. In practice, some approximation to this divergence measure is instead used. A common approach is to approximate the divergence $D_{div,0}$ from TWT and *root-mean-squared* (rms) velocities at zero-offset and apply

these for all offsets (Wang and McCowan, 1988; Yilmaz, 2001):

$$D_{div,0}(t_0) = t_0 \frac{\overline{v_p}^2}{v_{p1}} \tag{2.21}$$

where t_0 is the TWT, v_{p1} is the velocity in the first layer, and $\overline{v_p}$ is the time-weighted rms velocities. These velocities are usually estimated as part of the pre-processing phase. However, since the raw seismic data here are based on a known elastic model, these velocities are readily available, and thus reduces the need for pre-processing. A caveat of applying Equation 2.21 to accommodate the wave attenuation is that it is known to overcompensate the correction (Yilmaz, 2001). A common approach is furthermore to apply a simple correction of the following form (Wang and McCowan, 1988; Yilmaz, 2001).

$$D_{div,0}(t_0) = \beta t_0^{\alpha} \tag{2.22}$$

where α and β are arbitrarily chosen scalars. Figure 2.5 shows a raw data set before and after being corrected for spherical divergence using Equation 2.22 with $\alpha = 1.1$ and $\beta = 1$. These values are chosen empirically with the goal of having as consistent amplitude of the direct wave as possible within the correction. It should now be apparent for the reader that the subject of compensation for spherical divergence is not straight-forward and requires a series of subjective choices ranging from decisions about which compensation method to apply, to choosing compensation factors within each method. Alternatively, correction for divergence can also be performed in the slowness domain (Wang and McCowan, 1988). Simply using an *Automatic Gain Control* (AGC) filter does not provide a

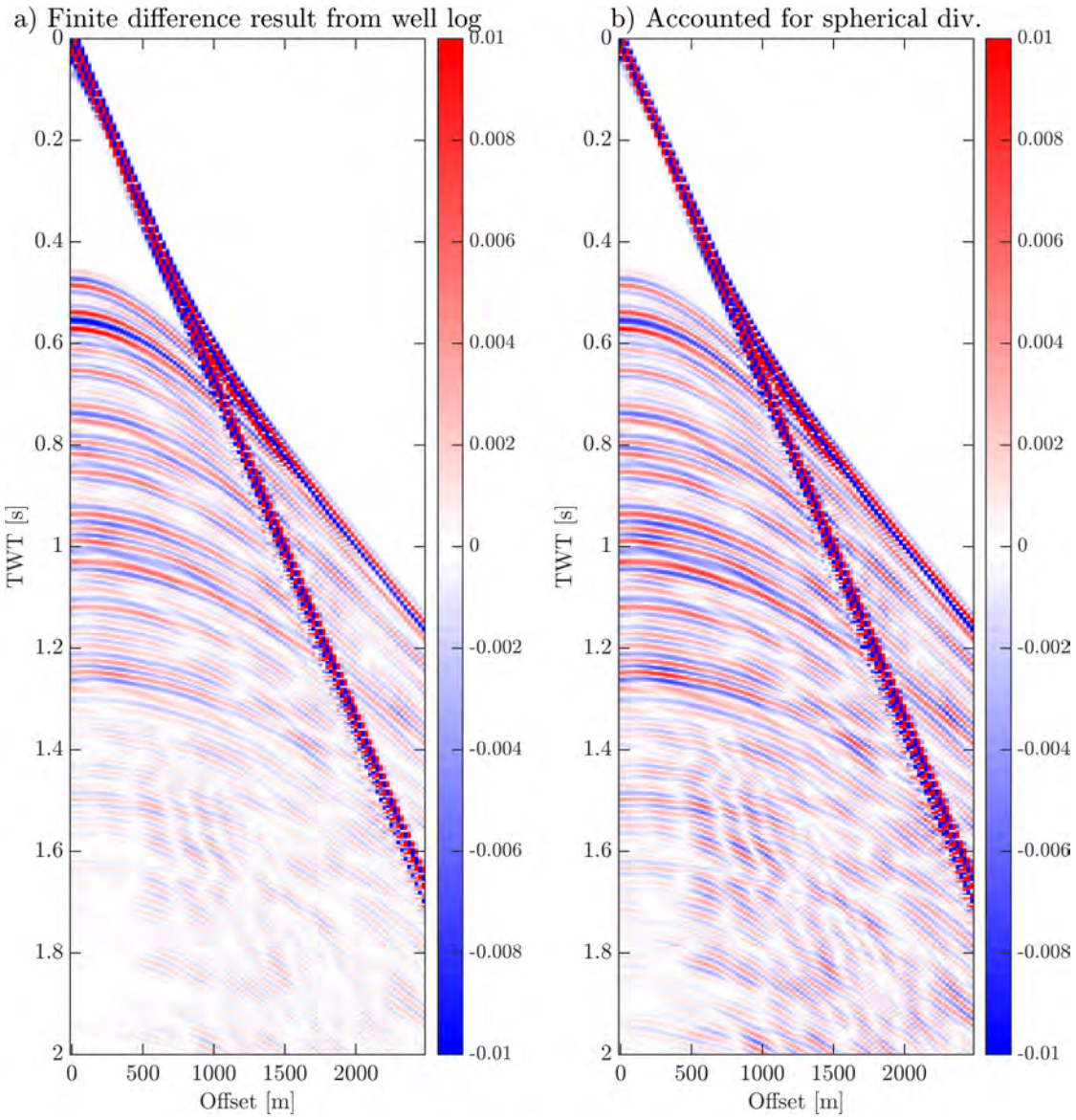


Figure 2.5: Example of correction for spherical divergence. a) Raw data from finite difference simulation. b) Raw data corrected for spherical divergence using Equation 2.22. Abscissa-axis displays offset in meters while the ordinate-axis displays TWT depth in seconds.

correction for spherical divergence because it is not based on any physical understanding of the problem, but rather a way of leveling out amplitudes. Applying an AGC filter would therefore tamper with the original amplitude variations, and hence violate the assumptions of AVO data.

2.4.2 Normal Moveout Correction

The raw data in e.g. Figure 2.3 and Figure 2.5a display a hyperbolic trajectory of seismic reflectors as a function of offset. The time delay experienced at a receiver at a given offset compared with the arrival time at zero-offset is known as the *Normal MoveOut* (NMO) and is denoted Δt_{NMO}. The idea behind *NMO-correction* is to account for this time delay and hence flatten the seismic reflectors, in other words shifting the seismic records to their zero-offset equivalent (Shearer, 2009). From Appendix F the following relationship (Equation F.3) between TWT and arrival time at a certain offset $t(x)$ is established as a function offset x:

$$t(x) = \sqrt{t_0^2 + \frac{x^2}{v^2}} \tag{2.23}$$

where v is the velocity in the layer above the reflector. This relationship can be used to correct the arrival times by calculating the NMO time delay Δt_{NMO} as:

$$\Delta t_{\text{NMO}} = t(x) - t_0 \tag{2.24}$$

Equation 2.23 is derived under the assumption that the velocity is constant in the layers above the reflector. However, the velocity profile of the subsurface is never constant. The velocity model v therefore needs to satisfy the assumption of

constant velocity while still approximating the true velocity of the subsurface. If the velocity is too low, the arrival time $t(x)$ in Equation 2.23 becomes larger than intended, and hence Δt_{NMO} in Equation 2.24 becomes too large. The result is an overcompensation. Similarly, for a velocity that is higher than intended, the result is an undercompensation of the reflections. The default choice for velocities is to use the time-weighted rms velocities obtained during pre-processing of the data. This satisfies both assumptions fairly, because the velocity is not constant, while at all depths the velocity is an average of all the above-laying layers. As for the spherical divergence correction, the velocity log already exists for the synthetic data which enables a much better NMO-correction.

A well-known problem with applying NMO-corrections is non-stationarity of the compensation level, i.e. depths and offsets are not compensated equally. A frequency distortion is therefore to be expected. This distortion is particularly present for shallow reflectors and at large offsets (Yilmaz, 2001). At larger offsets the NMO-correction is heavier due to the hyperbolic trajectory of the reflectors. The reflectors are stretched and thus the frequency is lowered. This is also known as NMO-stretching and is an unfortunate side-effect of NMO-corrections. In Figure 2.6 the raw data corrected for geometric spreading is NMO-corrected using the available elastic log. The frequency distortions are easily detectable in the NMO-corrected data set (i.e. an AVO gather). The effects at shallow depths and large offsets are sometimes muted out of the resulting AVO gather in order to avoid any issues from these distortions in the AVO analysis or in the further processing. This can be done simply by introducing a top mute which sets all affected values to zero. However, the lowering of the frequency due to

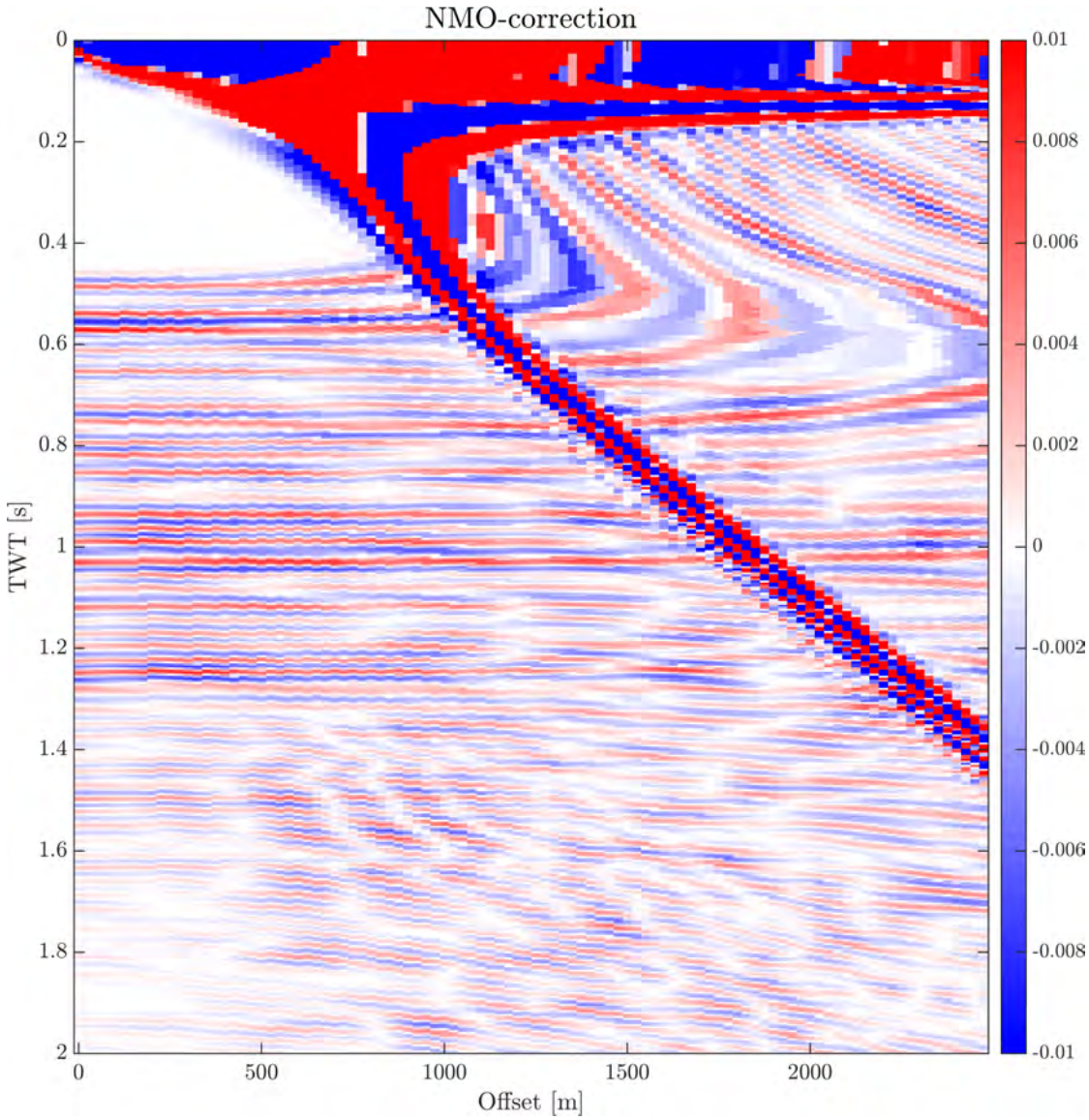


Figure 2.6: Example of NMO-correction. The data in Figure 2.5b is corrected for normal moveout. Frequency disturbance effects are easily detectable in the NMO-corrected shot gather. The abscissa-axis displays offset in meters while the ordinate-axis displays TWT depth in seconds.

NMO-stretching is not possible to remove completely. Spectral balancing is commonly used to compensate the high frequency loss (Xu and Chopra, 2007). Nevertheless, a residual amount of NMO-stretching must be expected in the resulting AVO seismic data. A stretching of the seismic wavelet as a function of offset can accommodate some of this bias in the further analysis (Xu and Chopra, 2007). Yet, this wavelet stretching is still stationary in depth (i.e. time) and not straight-forward.

2.4.3 AVO gather to AVA gather conversion

Finally, the AVO gather is converted to an AVA gather. This offset-to-angle conversion follows from combining offset traces to angle traces. It is therefore important to establish a relationship between incoming angles and offsets. Castagna and Backus (1993) provide the following equation between the incidence angle θ and offset x:

$$\sin(\theta) = \left(\frac{v_{int}}{v_{smooth}}\right)\left(\frac{x}{\sqrt{\left[x^2 + (v_{smooth}t_0)^2\right]}}\right) \tag{2.25}$$

where v_{smooth} is a smoothed stacking velocity profile (e.g. the rms velocities) and v_{int} is a velocity profile for each layer following from Hewitt Dix (1955) formula. The second paranthesis follows from simple trigonometry and the first paranthesis is a term correcting the use of smoothed velocities. Using Equation 2.25, a single angle trace can then be constructed by ray-tracing the angle through the offset domain. In practice, a central angle of interest is chosen including nearby angles in a panel for which ray-tracing is performed (Castagna and Backus, 1993). By stacking all the traces within the panel, an angle trace is

constructed for the central angle. An example of panels for ray-tracing can be seen in Figure 2.7. Applying the panels shown in Figure 2.7 to construct an AVA gather from the AVO gather in Figure 2.6 is seen in Figure 2.8a. The zero-offset trace is taken directly from the AVO gather. The NMO-correction plus angle conversion leads to increasing biases as a function of offset. Especially when anisotropy in the subsurface is considered. Mukhopadhyay and Mallick (2011) propose a ray-based offset-to-angle transform, allowing the generation of AVA gathers using raw seismic data uncorrected for normal moveout. However, the standard procedure still remains to apply an NMO-correction with a subsequent offset-to-angle conversion. The final essential processing sequence for raw seismic data to AVA gathers can be summarized as in Figure 2.9.

A final remark on processing: The processing done in Appendix H.5 is performed using the industry software package Promax (Landmark, 1999). The specific implemented processing algorithms are therefore unknown but should follow the physical theory outlined in the presented corrections.

Using the FWM (Equation 2.6) to create a raw seismic data set, a processing sequence g_{proc} can be set up to create a forward model for the full forward modeling of AVA shot gathers from an elastic model:

$$\mathbf{d}_{\text{ava}} = g_{proc}(\mathbf{d}_{\text{raw}}) = g_{proc}(g_{fwm}(\mathbf{m})) \tag{2.26}$$

g_{proc} should, as mentioned earlier, as a minimum include the processing steps in Figure 2.9.

A comparison of obtaining the $\mathbf{d}_{\text{ava}}$ from both processing raw data ($\mathbf{d}_{\text{ava,proc}}$)

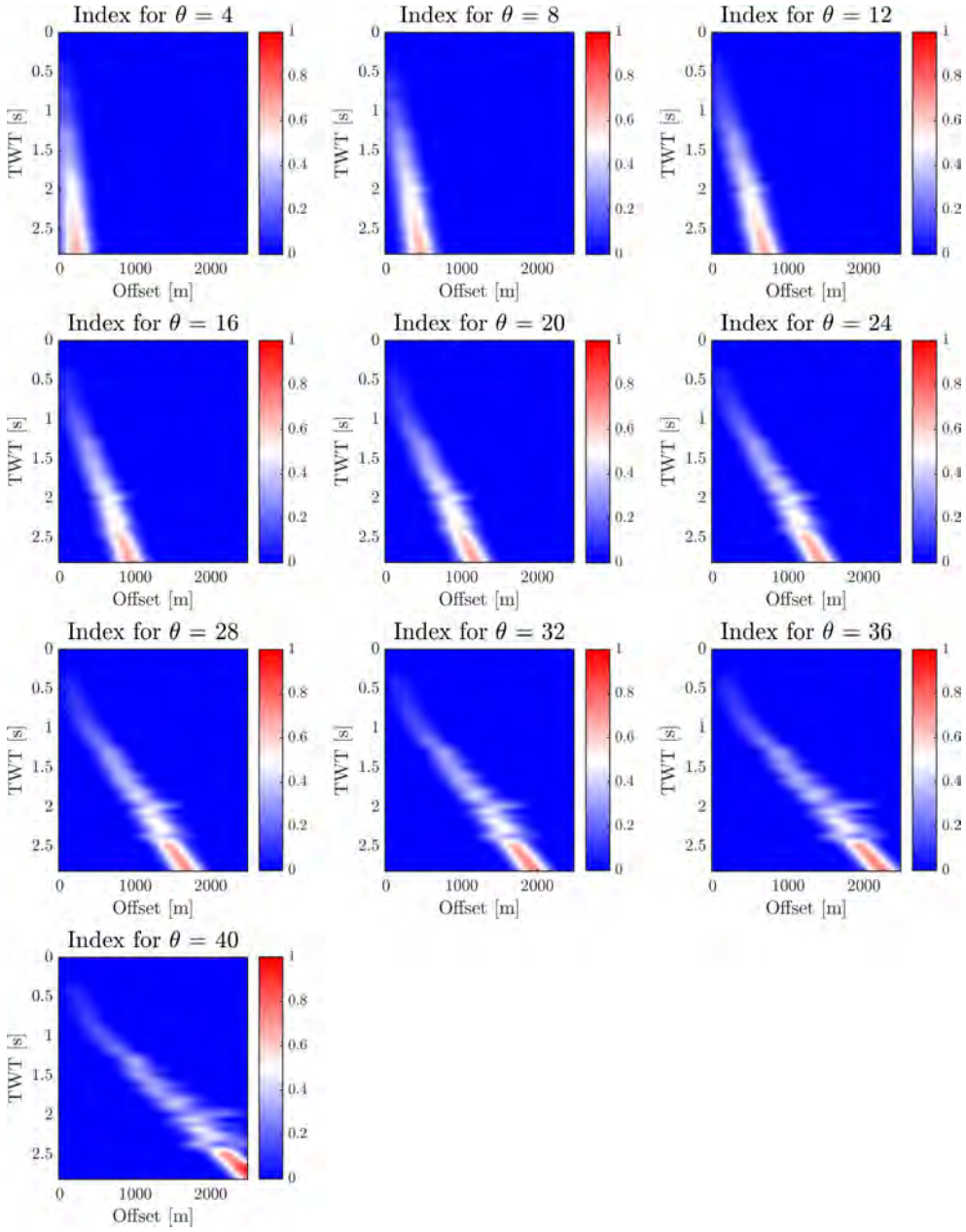


Figure 2.7: Example of angle panels to be used for ray-tracing in the AVO gather displayed in Figure 2.6. The used panels for each of the angles $\theta = 4, 8, \ldots, 40$ are shown in each subfigure. The abscissa-axis displays offset in meters while the ordinate-axis displays TWT depth in seconds.

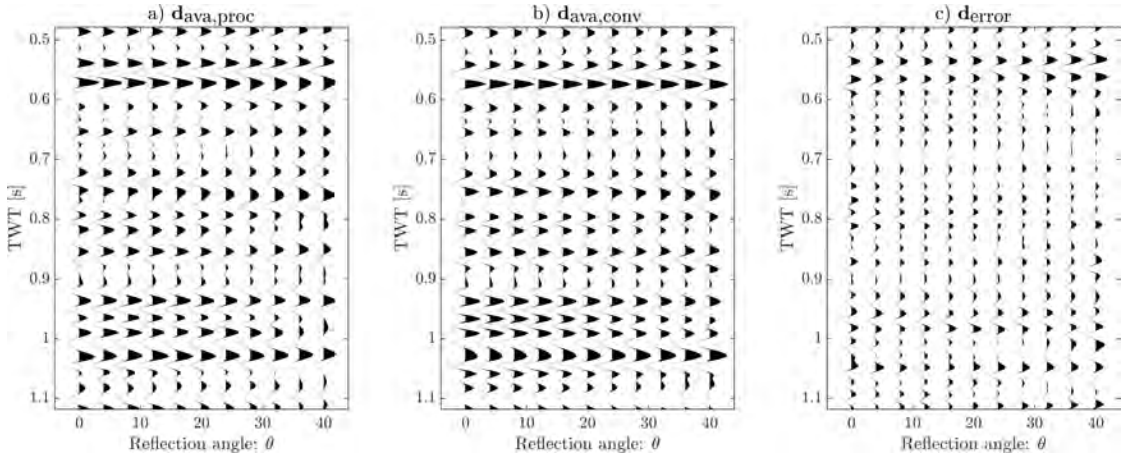


Figure 2.8: AVA gather from same elastic log using different forward models. a) FWM plus processing $\mathbf{d}_{\text{ava,proc}}$. AVA gather constructed from the AVO gather in Figure 2.6 applying the angle panels shown in Figure 2.7. The zero-offset trace is taken directly from the AVO gather. b) Convolutional model $\mathbf{d}_{\text{ava,conv}}$. c) Difference between the two AVA data sets $\mathbf{d}_{\text{error}}$. The abscissa-axis displays reflection angles θ while the ordinate-axis displays TWT in seconds.

Figure 2.9: Flowchart of necessary processing steps from raw seismic data to AVA gathers.

and convolutional model ($\mathbf{d}_{\text{ava,conv}}$) can be seen in Figure 2.8a and Figure 2.8b respectively. The difference between the two AVA data sets ($\mathbf{d}_{\text{error}} = \mathbf{d}_{\text{ava,proc}} - \mathbf{d}_{\text{ava,conv}}$) is shown in Figure 2.8c. Even for a scenario with low noise, low variability in the subsurface and optimal conditions for processing, the error $\mathbf{d}_{\text{error}}$ is still relevant. How relevant will be discussed in 5.2.

"[t]oday's posterior distribution is tomorrow's prior"

Lindley (1970), Bayesian Analysis in Regression Problems

3 Prior models

In probabilistic inversion methods, our expectation to the behaviour of the subsurface must be quantified. The general idea is to incorporate other information on the model parameters **m** in the final model than what comes from the observed data. These direct sources of information must be independent of the data and should be based on prior experience, knowledge from experts, physical constraints or alternative measurements of **m**. In seismic exploration one might for instance know something about the distribution of the model parameters from geological experts and/or from nearby well logs. The direct a priori information on the model parameters **m** are gathered in a *prior* probability

distribution often simply denoted the *prior model*. For probabilistic linear inversion of AVA seismic data as suggested by Buland and Omre (2003a), the prior model is Gaussian. This prior model is adapted and used frequently throughout the work in this thesis and is simply referred to as the *Buland and Omre* prior model. The Gaussian distribution has a series of beneficial mathematical properties, one of which allows an analytical solution to the probabilistic inverse problem. In much of the following an analytical solution is preferred because it provides computationally fast solutions which can also serve as a reference.

One major downside of this prior model is the general smoothness associated with realizations from Gaussian distributions. Sharp transitions between layers are to be expected in the subsurface, and the Gaussian prior distribution cannot be expected to fully encompass these changes. More complex prior models are applied in the literature in order to better describe realistic subsurface variability.

Skewness has been introduced into the Gaussian prior model (Rimstad and Omre, 2014b). Gaussian mixture models have also been proposed to allow a prior model which is locally smooth and globally discrete (Grana et al., 2017). Zunino et al. (2015) use geostatistical algorithms to generate realizations of prior models (from training images) based on multiple-point statistics.

In this thesis, a *pluri-Gaussian* prior model is used in order to create sharp transitions in the elastic properties in Appendix H.1 and Appendix H.4. The above-mentioned alternative prior models are probably more adequate at describing the true variability of the subsurface. They also have one thing in common, that is they do not allow an analytical solution to the inverse problem. Instead, some sampling technique must be applied in order to solve the

probabilistic inverse problem. In the following the Buland and Omre prior model is introduced along with the pluri-Gaussian prior model. Finally, the prior model used to create the elastic subsurface models used for the FWM is presented.

3.1 Buland and Omre Prior Model

The main prior model used in this thesis is a discrete Gaussian distribution, i.e. it can be described by a mean vector $\boldsymbol{\mu}_m$ and a covariance matrix $\mathbf{C}_m$ (see Appendix B). The prior distribution $\rho_M(\mathbf{m})$ of the model parameters $\mathbf{m}$ is then described by the multivariate Gaussian probability density function:

$$\begin{aligned}\rho_M(\mathbf{m}) &\sim \mathcal{N}_{n_m}\{\boldsymbol{\mu}_m, \mathbf{C}_m\} \\ &= (2\pi)^{-\frac{n_m}{2}} |\mathbf{C}_m|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{m}-\boldsymbol{\mu}_m)^\mathsf{T}\mathbf{C}_m^{-1}(\mathbf{m}-\boldsymbol{\mu}_m)\right]. \end{aligned} \tag{3.1}$$

where n_m is the number of model parameters. The linear forward model based on Stolt and Weglein (1985) presented in Section 2.3.1 allows a linear relationship between the logarithm of the elastic properties v_p, v_s, and ρ and seismic data. The model parameters are therefore defined as a continuous vector field:

$$\mathbf{m} = [\ln(v_p), \ln(v_s), \ln(\rho)]^\mathsf{T} \tag{3.2}$$

The following is largely based on Buland and Omre (2003a). Describing this set of model parameters with a multivariate Gaussian distribution requires a discrete mean vector and a discrete covariance matrix as established previously. The mean value in this proposal is simply the expectation value of each elastic variable.

$$\boldsymbol{\mu}_m = \mathrm{E}\{\mathbf{m}\} = [\mathrm{E}\{\ln(v_p)\}, \mathrm{E}\{\ln(v_s)\}, \mathrm{E}\{\ln(\rho)\}]^\mathsf{T} \tag{3.3}$$

This formulation also allows for slightly varying (i.e. smooth) expectation functions, which would still satisfy the small-contrast approximation of the forward model. However, in the following the mean value is constant as suggested in Buland and Omre (2003a):

$$\boldsymbol{\mu}_m = [8.006, 7.313, 7.719]^{\mathsf{T}} \tag{3.4}$$

These values correspond to actual velocities: $v_p = 3000\,\mathrm{m\,s^{-1}}$, $v_s = 1500\,\mathrm{m\,s^{-1}}$, and density: $\rho = 2500\,\mathrm{kg\,m^{-3}}$. The covariance matrix of the model parameters is constructed using a stationary covariance function:

$$\mathbf{C}_m = \mathbf{C}_0 v_t(t_0) \tag{3.5}$$

where $v_t(t_0)$ is a stationary correlation function of TWT and $\mathbf{C}_0$ is an invariant covariance matrix containing the correlation and variance between the elastic model parameters. The correlation function $v_t(t_0)$ needs to be positive definite and should abide $v_t(0) = 1$. Several correlation functions have these properties, e.g. the Exponential type, Gaussian type and Spherical type spatial correlation functions (see e.g. Journel and Huijbregts (1989)). A spherical type spatial correlation function with anistropic features is for instance used in Appendix H.6. For the Buland and Omre prior model, a second-order exponential function (i.e. the Gaussian correlation function) is used:

$$v_t(t_0) = \exp\left[-\left(\frac{t_0}{r}\right)^2\right] \tag{3.6}$$

where r is a range parameter that can be set arbitrarily, depending on the desired correlation in TWT (i.e. depth). A relatively small range compared with the

TWT sampling interval produces small correlations between the elastic parameters and vice versa. In the Buland and Omre prior model, a value of $r = 5$ ms is implemented. The sampling interval is set to $dt = 2$ ms, making the correlation in depth last over several sampling points.

In the Buland and Omre prior model, the stationary covariance matrix $\mathbf{C}_0$ is set up using an assumption of stationary variance for each elastic parameter ($\sigma^2_{v_p}$, $\sigma^2_{v_s}$, and σ^2_ρ). There is furthermore an assumption of a stationary correlation between each elastic parameter (ν_{v_p,v_s}, $\nu_{v_p,\rho}$, and $\nu_{v_s,\rho}$). The elastic covariance matrix for the elastic parameters is then:

$$\mathbf{C}_0 = \begin{Bmatrix} \sigma^2_{v_p} & \sigma_{v_p}\sigma_{v_s}\nu_{v_p,v_s} & \sigma_{v_p}\sigma_\rho\nu_{v_p,\rho} \\ \sigma_{v_p}\sigma_{v_s}\nu_{v_p,v_s} & \sigma^2_{v_s} & \sigma_{v_s}\sigma_\rho\nu_{v_s,\rho} \\ \sigma_{v_p}\sigma_\rho\nu_{v_p,\rho} & \sigma_{v_s}\sigma_\rho\nu_{v_s,\rho} & \sigma^2_\rho \end{Bmatrix} \tag{3.7}$$

If zero correlation is assumed between the elastic parameters (i.e. $\nu_{v_p,v_s} = \nu_{v_p,\rho} = \nu_{v_s,\rho} = 0$), Equation 3.7 reduces to:

$$\mathbf{C}_0 = \begin{Bmatrix} \sigma^2_{v_p} & 0 & 0 \\ 0 & \sigma^2_{v_s} & 0 \\ 0 & 0 & \sigma^2_\rho \end{Bmatrix} \tag{3.8}$$

The used correlation between the elastic parameters in the Buland and Omre prior equals that of "Well B" in Buland and Omre (2003a) ($\nu_{v_p,v_s} = \nu_{v_p,\rho} = \nu_{v_s,\rho} = 0.7$), which is relatively strong. The variance of each individual elastic parameter is set to $\sigma^2_{v_p} = 0.0074$, $\sigma^2_{v_s} = 0.0074$, and $\sigma^2_\rho = 0.0024$.

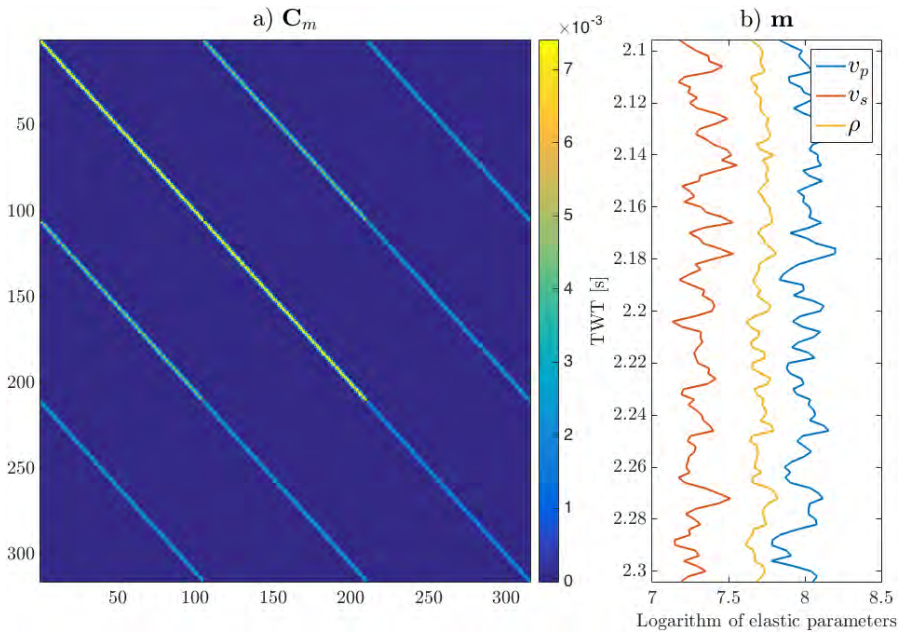


Figure 3.1: Buland and Omre prior model. a) Covariance matrix $\mathbf{C}_m$. b) Realization m from Buland and Omre prior using Cholesky decomposition. The logarithm of the elastic parameters is displayed on the abscissa-axis. The TWT is displayed on the ordinate-axis.

Using either Equation 3.7 or Equation 3.8 to compute $\mathbf{C}_o$ the final stationary covariance matrix of the model parameters $\mathbf{C}_m$ is then symmetric, positive definite, and of size $n_m \times n_m$. The final covariance matrix $\mathbf{C}_m$ for the Buland and Omre prior model is displayed in Figure 3.1a. The assumed stationarity in the covariance matrix is not a requirement for the whole prior model to be normally distributed. It is probably more realistic to expect non-stationarity in the subsurface and hence model parameters. However, a priori information about such non-stationarity is rarely available leading to stationary assumptions. In Appendix H.6, a new methodology for estimating the non-stationarity in the prior variance is proposed using maximum likelihood estimators in a sliding window approach.

A realization from a prior distribution is often needed for various reasons. Since the covariance matrix in a Gaussian multivariate distribution is positive definite and symmetric, a realization can be generated using Cholesky decomposition (Higham, 2009). The Cholesky decomposition of the prior covariance $\mathbf{C}_m$ is given by:

$$\mathbf{C}_m = \mathbf{U}^\mathsf{T}\mathbf{U} \tag{3.9}$$

where $\mathbf{U}$ is a non-singular upper-triangular matrix. This factorization is only possible if the prior covariance is positive definite. A realization $\mathbf{m}$ can then be generated from the Gaussian prior distribution $\mathcal{N}_{n_m}(\boldsymbol{\mu}_m, \mathbf{C}_m)$ as (Aster et al., 2004):

$$\mathbf{m} = \mathbf{U}^\mathsf{T}\mathbf{s} + \boldsymbol{\mu}_m \tag{3.10}$$

where $\mathbf{s}$ is a vector of size n_m consisting of random values that are independent and normally distributed with a mean of zero and standard deviation of one. The approach of generating or "drawing" a realization of a discrete Gaussian distribution using Cholesky decomposition can be applied to any symmetric, positive definite matrix $\mathbf{A}$ of size $n \times n$.

Figure 3.1b displays a realization drawn from the Buland and Omre prior model using Cholesky decomposition (Equation 3.10). The analytical posterior distribution of the Buland and Omre prior and the linear forward model is e.g. also multivariate Gaussian and thereby possible to generate realizations from.

In order to create less smooth a priori models than the Gaussian distribution a pluri-Gaussian prior model is adapted in Appendix H.1 and Appendix H.4 to describe a discrete prior model. Pluri-Gaussian prior models are developed and used frequently in the french geostatistical community in order to simulate facies and lithotypes (Armstrong et al., 2011). Pluri-Gaussian priors follow as a natural extension to truncated Gaussian random fields.

Consider a realization $\mathbf{r}$ from a central Gaussian distribution (i.e. with zero mean) and standard deviation of one ($\mathcal{N}(0,1)$). Such a realization $\mathbf{r}$ is shown in Figure 3.2a. The histogram of this realization is shown in Figure 3.2b. By choosing a set of truncation points, the continuous realization $\mathbf{r}$ can be discretized into subcategories, which could then for instance represent facies or lithotypes. If for instance two truncation points are applied of 1 and -0.5, the Gaussian distribution can be divided into three categories (1,2, and 3), where for all points in $\mathbf{r}$ a vector $\mathbf{m}$ can be constructed as:

$$\mathbf{m} = \begin{cases} 1, & \text{if } \mathbf{r} \leq -0.5 \\ 2, & \text{if } -0.5 \leq \mathbf{r} \leq 1 \\ 3, & \text{if } \mathbf{r} \geq 1 \end{cases} \tag{3.11}$$

An example of how such a truncation looks in 1D is shown in Figure 3.2c. The discrete nature of the truncated Gaussian is clearly visible to the naked eye. The distribution of the three categories is shown in 3.2d as a histogram. The pluri-Gaussian method is basically just an extension of this truncation

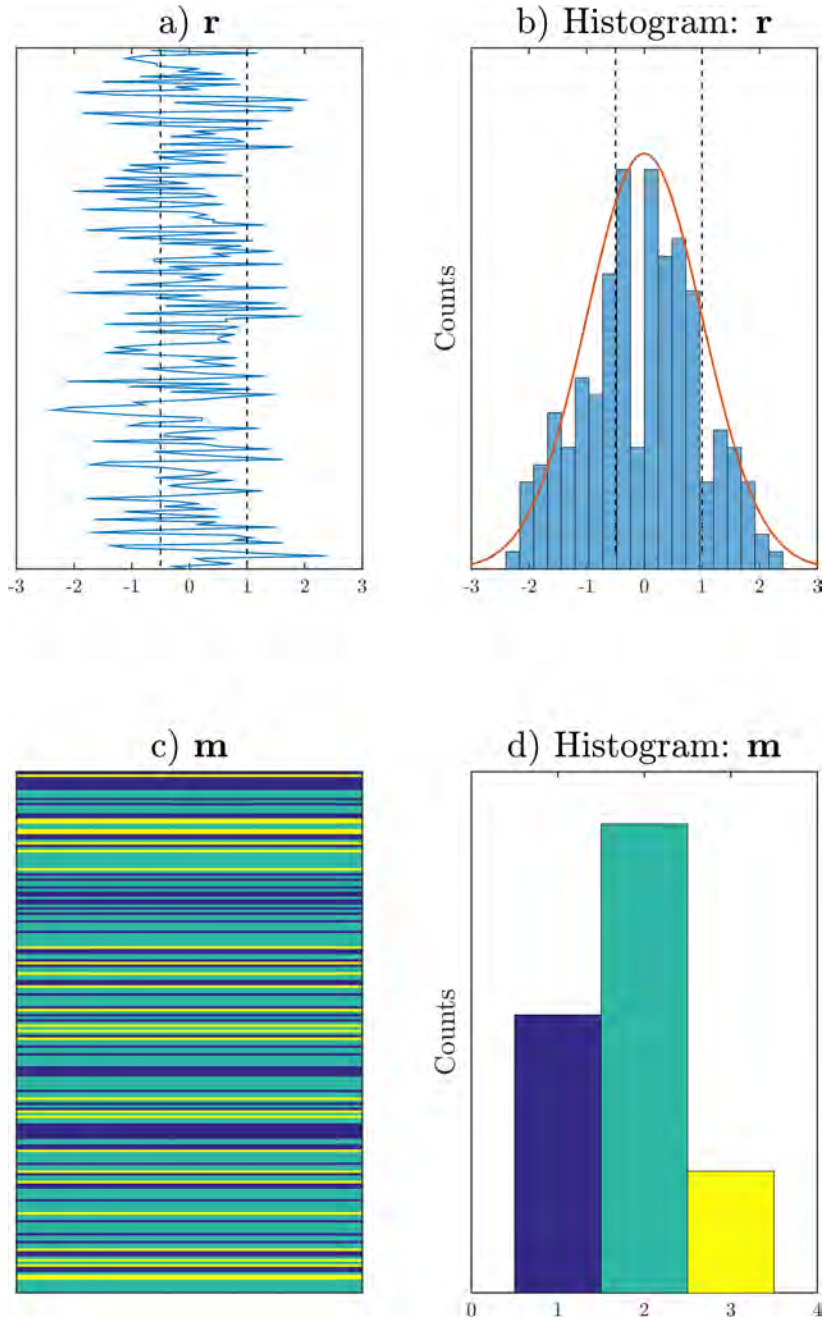


Figure 3.2: Example of a truncated realization $\mathbf{m}$ from a Gaussian random field. a) The un-truncated realization $\mathbf{r}$. b) Histogram of $\mathbf{r}$. c) Truncated vector $\mathbf{m}$ consisting of three categories. d) Distribution of three categories represented by a histogram.

methodology with several Gaussian distributions to create more complex patterns. Using instead two Gaussian distributions, the truncation can be split into more intricate patterns.

Consider two realizations $\mathbf{r}_1$ and $\mathbf{r}_2$ from two Gaussian distributions $\mathcal{N}_1$ and $\mathcal{N}_2$. A truncation rule is then set up for both Gaussian distributions. This is done by constructing a 2D truncation map determining the truncation for $\mathbf{r}_1$ and $\mathbf{r}_2$ simultaneously. An example of a truncation map with four different layers is seen in Figure 3.3. If for instance the first element in the two realizations are $\mathbf{r}_1(1) = 1.5$ and $\mathbf{r}_2(1) = 0.5$, the first element in the pluri-Gaussian realization $\mathbf{m}(1)$ is layer 4 (yellow) according to the truncation map (see the black cross in Figure 3.3). The discrete realization $\mathbf{m}$ can then be constructed element-wise following this approach. In paper Appendix H.4 two central Gaussian distributions are used to generate the realizations $\mathbf{r}_1$ and $\mathbf{r}_2$. The covariance matrices are setup as:

$$\mathbf{C}_r = \sigma_r^2 \nu_t(t_0) \tag{3.12}$$

where $\nu_t(t_0)$ is the spatial correlation function and σ_r^2 is the variance of $\mathbf{r}$. The variance is set to $\sigma_r^2 = 1$ in both distributions. For the correlation function $\nu_t(t_0)$ a Gaussian type spatial correlation function as in Equation 3.6 is chosen for $\mathcal{N}_1$ with range $r = 1$. An exponential type spatial correlation function is chosen for $\mathcal{N}_2$ with range $r = 10$.

The truncation map in Figure 3.3 and the outlined distributions are equal to those used in Appendix H.4 to generate discrete prior realizations. A discrete truncated pluri-Gaussian realization $\mathbf{m}$ from this setup, can be seen in Figure 3.4a. The histogram of $\mathbf{m}$ in Figure 3.4b reveals that the distribution of layers in a

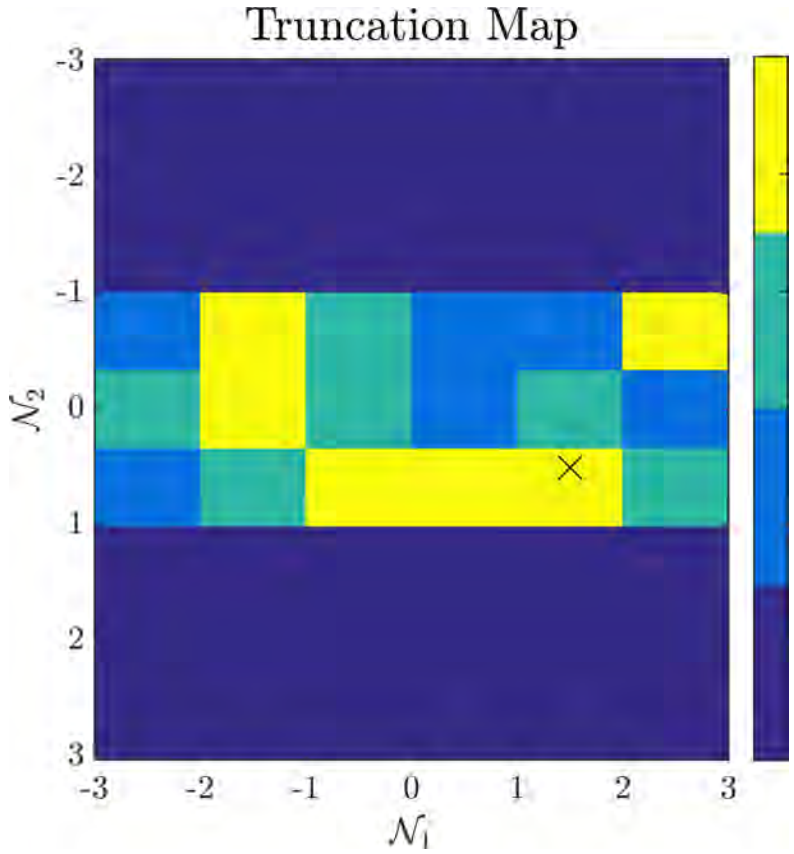


Figure 3.3: Truncation map used in Appendix H.4 to generate discrete realizations of the subsurface with four different layers. Black dot indicates $\mathbf{r}_1(1) = 1.5$ and $\mathbf{r}_2(1) = 0.5$.

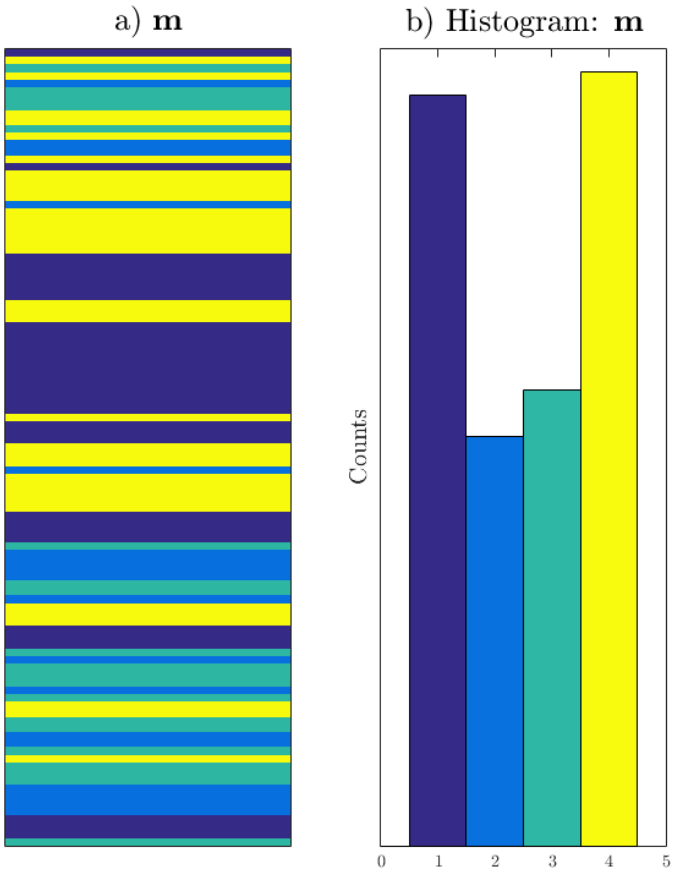


Figure 3.4: a) Pluri-Gaussian realization $\mathbf{m}$ created using the two Gaussian distributions $\mathcal{N}_1$ and $\mathcal{N}_2$, and the truncation map in Figure 3.3. b) Histogram of the four different layers in $\mathbf{m}$.

pluri-Gaussian realization does not need to be Gaussian even if the method is based on Gaussian distributions.

The pluri-Gaussian method can also be used in order to generate discrete realization in more dimensions by defining the Gaussian distributions accordingly. Examples can be found in Le Loc'h and Galli (1997); Mariethoz et al. (2009); Armstrong et al. (2011).

The truncation of the Gaussian distributions makes the pluri-Gaussian

methods inherently non-linear and non-Gaussian as established previously. It was therefore not used for any inversion purposes during this PhD work as the focus has been on linearized AVO inversion. For the purpose of generating rapidly changing layers with correlations, the pluri-Gaussian method is nonetheless very capable, as witnessed in the realization **m** in Figure 3.4a. This proved very useful in Appendix H.4, where a subsurface prior model with rapid changes in elastic properties was needed to test the small-contrast approximation of Zoeppritz equations.

3.3 Prior Model for Full Waveform Modeling

An elastic model (v_p, v_s, and ρ) is needed for the full-waveform modeling. The full-waveform modeling is used to generate raw seismic data in Appendix H.5. Here the effect of processing seismic raw data to AVA data is investigated. The subsurface variability is assumed to follow the Buland and Omre prior model with a sampling rate of $dt = 4$ ms. To mimic an offshore seismic acquisition scenario, a realization **m** from the Buland and Omre prior model was added a water layer ($v_p = 1500\ \mathrm{m\,s^{-1}}$, $v_s = 0\ \mathrm{m\,s^{-1}}$, and $\rho = 1000\ \mathrm{kg\,m^{-3}}$) on top. At the boundary between water and rocks (i.e. the sea-bottom), a hard "kick" in the seismic response is to be expected due to the large difference in elastic properties. The processing engine within the Promax software (Landmark, 1999) does not handle these hard kicks particularly well, leading to a lot of heuristic processing steps to compensate for the effect (e.g. top muting and removing larger offsets). Instead of introducing biases in the AVA data from adding these processing steps, a smooth transition between water and rocks in the elastic log is introduced. An example of a resulting elastic log is shown in Figure 3.5a.

The smooth transition at the sea-bottom transitions does not carry any physical meaning. It is in fact rather un-physical. The transition is purely added as a necessary step in designing a semi-realistic elastic log with water is on top of the realization $\mathbf{m}$, which does not produce a hard kick. A small consolation is that a layer of unconsolidated sediment is usually found on the sea-bottom in real geological settings. This layer also acts as a smooth transition between water and rocks at the sea-bottom. The problem with the hard kick is therefore especially an issue for synthetic data. Mukhopadhyay and Mallick (2011) also utilize a smooth transition at the sea-bottom based on an-isotropic velocity analysis of actual seismic data. The final subsurface model passed on the to the full-waveform algorithm needs to be converted from depth in TWT to depth in meters. Since the v_p velocity is available, this time-depth conversion is trivial. The resulting conversion of the log in Figure 3.5a is displayed in Figure 3.5b.

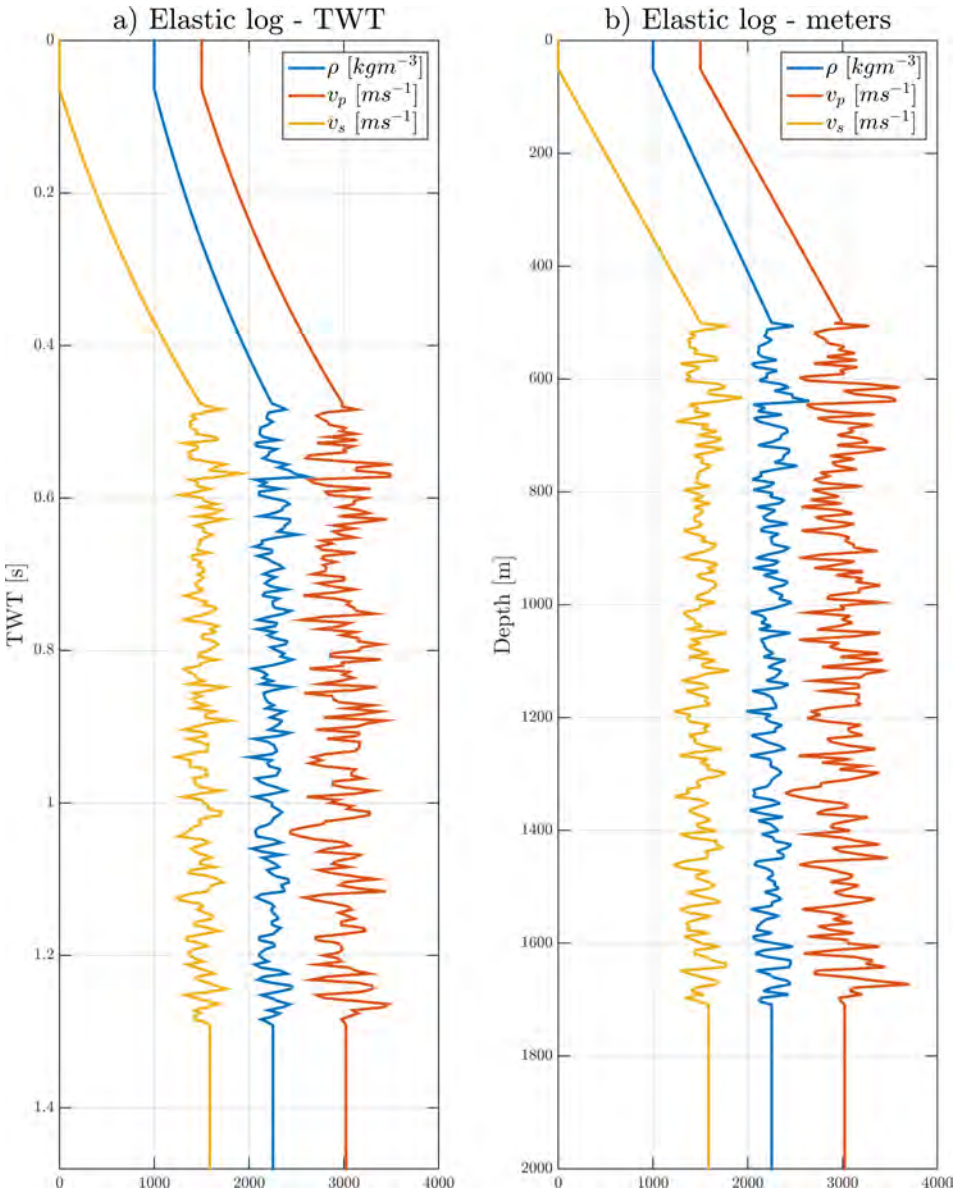


Figure 3.5: a) Elastic log with smooth transition between subsurface and water column. The ordinate-axis shows depth in TWT. b) Elastic log from subplot a), that is converted to depth in meters. This model can be used as input in the full-waveform modeling. The ordinate-axis shows depth in meters.

"That's what I do. Like to make noise. Like to go out and be a boogie boy"

Iggy Pop (1993), Boogie Boy

4

Noise models

Observations of geophysical data, including seismic data, are associated with uncertainty. If data were not associated with uncertainty, and thus would be absolutely reproducible, then all models whose response did not match the data perfectly would be incorrect (Scales and Snieder, 1998). The uncertainty may arise from a variety of different sources ranging from imprecise measurement equipment to approximate or even faulty physical relationships between data $\mathbf{d}$ and model parameters $\mathbf{m}$. As established in the introduction, an observed data set $\mathbf{d}_{\text{obs}}$ can be thought of as a physical forward response $g(\mathbf{m})$ and an additive noise

component $\boldsymbol{\varepsilon}$ (Equation 1.2):

$$\mathbf{d}_{\text{obs}} = g(\mathbf{m}) + \boldsymbol{\varepsilon}$$

The noise is assumed additive but could in principle also be multiplicative or have an even more complex relationship to the forward response. Because the noise is deemed additive, the practical understanding of 'noise' is then the residual between the observed and the predicted measurements $(\mathbf{d}_{\text{obs}} - g(\mathbf{m}))$, i.e. all that cannot be accounted for by the forward relation g. This is probably not an unreasonable assumption and the additive noise is also the standard formulation in inverse problem theory (Sen and Stoffa, 1996; Mosegaard and Tarantola, 2002; Ulrych and Sacchi, 2005; Tarantola, 2005; Menke, 2012; Bodin et al., 2012). Our inability to fully predict the outcome of an experiment boils down to two main reasons (Scales and Snieder, 1998):

- Inaccurate physics limitations: Current physical models are not capable at perfectly describing the complexity of the true physical system. The current physical models can therefore be considered to be approximations of the true systems in nature.
- Computational limitations: Our current physical models are sometimes also further approximated on purpose, in order to make them computationally feasible (i.e. computationally inexpensive to use).

If one could e.g. describe all particles and their internal relationship for all time and space, a perfect physical forward model would be available. Even, if such a function was available, it would probably be highly impractical because it would

require possibly infinitely many model parameters. In terms of human understanding of the physical relations at play, the physical model must be somehow simple enough for us to grasp the basic concepts. In this sense, our understanding of nature will probably always be inept. A physical model needs to be simple enough for geophysicists to reasonably understand and predict the fundamental outcomes of nature, not describing all particles' positions at all times. In our current paradigm, the noise becomes all that can not be directly mapped with a physical model and the chosen parametrization. By acknowledging that data is infected with noise of some form or another, a realistic noise model must be set up in order to account for data uncertainties during inversion of seismic data.

4.1 The Human Aspect

What is deemed signal and noise on a specific data set is ultimately decided by the geophysicist. Humans have excelled in nature by our ability to recognize and discern "patterns". What is meant by patterns is neatly summarized by Heile et al. (2017)

> "A pattern is what enables prediction at above the chance level. Brains, human or animal, have evolved as pattern recognition machines. That's because survival is enhanced if successful predictions can be made. The organism will be more likely to survive if it can predict where and when food will be found and if it can avoid predators. The prediction can be in the time domain or in the space domain or both."

All technological and scientific progress are build upon reproduction and

modification of patterns encountered in nature and the fabrication of new patterns (Mattson, 2014). The structure of the human mind also allows us to pattern-complete possibly noisy and incomplete signals (Battaglia et al., 2012). This is a natural part of human behavior and is also what has made us successful so far in evolutionary terms. The appreciation of the pattern-seeking brain has lead to the development of the mathematical discipline of pattern theory (Mumford, 1997).

As much as this property of the human brain is an extremely helpful property in interpretation of data, it also has a build-in fallacy in the discrimination of signal from noise. Because the human brain is wired to detect patterns, these patterns are usually weighted higher in terms of signal value. Conversely, data with no apparent patterns are more likely to be selected as noise or deemed unimportant. This fallacy is illustrated in the following example.

Consider the data recorded (Data 1) in Figure 4.1a. The data displays a strong sinusoidal pattern. The normalized power spectrum for Data 1 is shown in Figure 4.1b in blue. The sinusoidal pattern clearly stands out on the power spectrum as a low frequency component. The data also show a strong high-frequent component. A logical and natural separation of the signal and noise would be to consider the high-frequent component as noise and the sinusoidal component as the signal. A band-pass butterworth filter (red line in Figure 4.1b) is applied to the data in order to suppress the assumed noise and highlight the signal. The resulting processed data is shown in Figure 4.1c. The filtered data clearly match the original signal (as illustrated by the red line), when the high-frequent noise is filtered out.

Now, consider the data recorded (Data 2) in Figure 4.1d. Although slightly

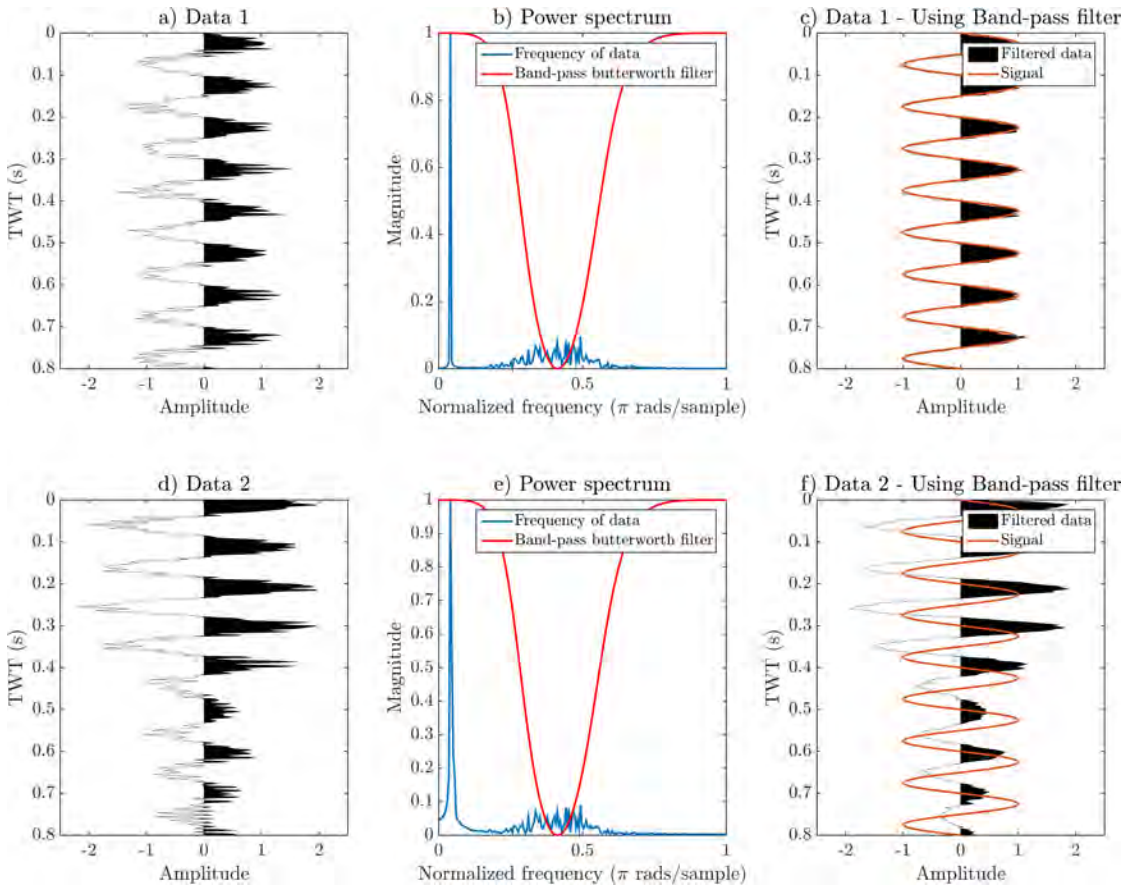


Figure 4.1: Example of filtering of two noisy data sets (Data 1 and Data 2). a) Unfiltered data (Data 1). b) Normalized power spectrum for Data 1. c) Data 1 after band-pass filter. d)-f) Same figures as in a)-c) for Data 2. For a),b),d) and f) the abscissa-axis displays amplitude, and the ordinate-axis displays depth in TWT. The two power spectra in b) and e) display normalized frequency on the abscissa-axis and magnitude on the ordinate-axis.

more complex than Data 1, a sinusoidal pattern is definitely discernible for the naked eye. The normalized power spectrum for this signal in Figure 4.1e shows a similar pattern to the power spectrum for the first data. Applying a similar band-pass filter for Data 2, to suppress the assumed noise reveals a smooth data set, as visualized in Figure 4.1f. Even though the filtered data looks smooth and apparently noise-free as in the first case, the signal (red line) does not match the filtered data at all. In fact, the signal is exactly the same as used in Data 1. The only difference between the two data sets is the noise added. The actual noise in Data 2 contains a low-frequent component which is inseparable from the signal both in the unfiltered data and the power spectrum hereof.

This small example, although a bit simplistic, illustrates two major points regarding noise in data. Firstly, that our pattern recognizing brains can be deceived by apparent structures in the data. This may ultimately result in a bias or tendency in the way we perceive noise. Are outliers part of the signal or not? Is noise mainly something irregular and high-frequent which clearly stands out from the overall pattern in the data? Secondly, that some noise sources may resemble the signal to a degree where they are basically "hiding" within the data, by having the same frequency. A known example from seismic data is for instance recurring multiples. How can one take such correlated effects into account when building a noise model of the data uncertainty?

The answers to the questions above are not straight-forward and should be considered when building a trustworthy noise model. The example does demonstrate that it is possible and relatively easy to remove noise which looks much different from the signal. For seismic data, white noise is usually filtered out during processing (Landrø, 2008). If data is repeatedly recorded at a certain

location in space, the results can also be stacked to further suppress uncorrelated noise. This presupposes that the noise in the data does not correlate over time (Scales and Snieder, 1998).

Both signal filtering and trace-stacking are frequently applied to raw seismic data in order to suppress noise. One should therefore be aware of noise with a strong correlation, because uncorrelated noise is likely removed from processed data. One should especially be alert for noise which is indiscernible from the signal. This is discussed further in Section 4.3.

4.2 Gaussian Data Uncertainty

In order to describe and account for noise in probabilistic inversion, a probability density function $\rho_d(\mathbf{d})$ over the data parameter space is introduced. The simplest probabilistic model to describe residual noise between predicted values $\mathbf{d}$ and observed data $\mathbf{d}_{\text{obs}}$ is a Gaussian model (see Appendix B):

$$\rho_d(\mathbf{d}) = k \exp\left(-\frac{1}{2}(\mathbf{d} - \mathbf{d}_{\text{obs}} - \boldsymbol{\mu}_D)^{\mathsf{T}} \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{d}_{\text{obs}} - \boldsymbol{\mu}_D)\right) \tag{4.1}$$

where $k = (2\pi)^{-n/2} |\mathbf{C}_d|^{-1/2}$ is a normalization constant, n is the dimension of the Gaussian, $\mathbf{C}_d$ is the covariance of the data uncertainties, and $\boldsymbol{\mu}_d$ is the mean in the data uncertainty. For most problems, the noise is centred Gaussian, meaning that the noise is thought to have zero mean and hence Equation 4.1 reduces to (Mosegaard and Tarantola, 2002):

$$\rho_d(\mathbf{d}) = k \exp\left(-\frac{1}{2}(\mathbf{d} - \mathbf{d}_{\text{obs}})^{\mathsf{T}} \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{d}_{\text{obs}})\right) \tag{4.2}$$

The assumption of centred Gaussian data uncertainties can be partially justified by the *central limit theorem* (Gelman et al., 2014). The central limit theorem basically states that the sum of numerous independent random variables produce an approximately Gaussian random variable (Aster et al., 2004). The central limit theorem applies regardless of the distribution of the independent random variables. The random variables could therefore come from all possible discrete or continuous distributions and the sum would still be Gaussian.

As a small intuitive example, consider $\overline{X}$ to be the sum of two independent random variables X_1 and X_2, i.e. $\overline{X} = X_1 + X_2$. Then the distribution of $\overline{X}$ is given by $p(\overline{X}) = p(X_1) \circledast p(X_2)$, where $\circledast$ is a convolution between the two distributions (Doyle, 2006). Since a convolution is a smoothing operation, $p(\overline{X})$ must be smoother than both $p(X_1)$ and $p(X_2)$ individually. This smoothing from convolution gives an intuitive idea of how the central limit theorem works, however it does not follow strictly from this result, that the final distribution approaches a Gaussian for large numbers of variables (Ulrych and Sacchi, 2005).

Recall from the introduction that if modeling errors are assumed negligible, Equation 1.4 reduces to $\mathrm{L}(\mathbf{m}) = \rho_d(\mathbf{d})$. The likelihood over the model parameter space $\mathrm{L}(\mathbf{m})$ is then also Gaussian for a Gaussian data uncertainty. This result is used frequently throughout different probabilistic inversion schemes (Buland and Omre (2003a); Malinverno and Briggs (2004); Downton (2005); Jullum and Kolbjørnsen (2016); to name a few). Mosegaard and Tarantola (1995) exhibit a case with non-Gaussian data uncertainty, but this can be considered an exception, because most geophysical data uncertainty is assumed to be Gaussian. Especially for seismic data.

Cordua et al. (2009) recognize the effect of correlated noise effects in cross-borehole ground penetrating radar data. They argue that higher resolution images is obtained when correlation of the data errors are accounted for. This result can probably be generalized to other ground penetrating methods as well.

In seismic inversion, data uncertainties are however usually assumed to be uncorrelated and stationary (Rabben et al., 2008; Singleton, 2009; Alemie and Sacchi, 2011; Aune et al., 2013; Grana, 2016). Explicitly dealing with correlated noise is suggested to be beneficial in seismic inversion (Ray et al., 2013). Although spatial correlations of noise in seismic data are crucial, they are difficult to quantify (Gouveia and Scales, 1998), and this leads to the convenient uncorrelated noise model. Uncorrelated noise basically means that noise associated with a specific data point does not have any relation to noise on all other data points, i.e. the noise is independent.

When little is known about the noise a priori, this can be used as an argument for applying an uncorrelated noise model for the data uncertainties. Because noise is uncorrelated and independent, the number of degrees of freedom for the noise is kept at the maximum (same as the dimension of the distribution). In other words, uncorrelated noise maximizes the information entropy (Ulrych et al., 2001; Aster et al., 2004). Uncorrelated (or white) noise is typically obtained by assuming a centred (zero-mean) Gaussian distribution of data uncertainties with a fixed (i.e. stationary) variance σ_e^2. The data covariance matrix then takes the following form:

$$\mathbf{C}_d = \sigma_e^2 \mathbf{I}_n \tag{4.3}$$

where $\mathbf{I}_n$ is an identity matrix with dimension $n \times n$.

From the previous sections, we have established that it is important to consider noise which is correlated and, if possible, not only assume uncorrelated noise in the data. Correlated noise can also be labeled colored noise as opposed to white noise. In seismic exploration some specific correlated noise components are also known as "coherent" noise (Castagna, 1996; Downton, 2005), and white noise is sometimes known as "generic" or random noise (Cambois, 2001; Downton, 2005). These quantities sometimes refer to noise that is and is not reoccurring in time, and hence deemed removable by stacking. In this study, these terms are refrained from being used, because they typically introduce some unwanted ambiguity in what is exactly meant. Instead, the distinction here is made between uncorrelated (i.e. independent) and correlated (i.e. dependent) noise.

One important source of correlated errors, which is considered in this study, is the use of inaccurate physics in the forward model g or "theory" errors. If these modeling errors $\theta(\mathbf{d}|\mathbf{m})$ are assumed Gaussian, then

$$\theta(\mathbf{d}|\mathbf{m}) = k \exp\left(-\frac{1}{2}(g(\mathbf{m}) - \mathbf{d}_{\text{obs}} - \boldsymbol{\mu}_T)^\top \mathbf{C}_T^{-1}(g(\mathbf{m}) - \mathbf{d}_{\text{obs}} - \boldsymbol{\mu}_T)\right) \quad (4.4)$$

where $\mathbf{C}_T$ is the covariance of the modeling errors, and $\boldsymbol{\mu}_T$ is the mean of the modeling errors. For infinite variances of the Gaussian distribution the homogeneous probability distribution $\mu_{\text{D}}(\mathbf{d}) = k$ (Hansen et al., 2014). The diagonal elements of the resulting covariance matrix are the variance of the total data error, and off-diagonal elements indicate correlation in the error. Combining the Gaussian modeling errors $\theta(\mathbf{d}|\mathbf{m})$ with the data uncertainties in Equation 4.1

yields the following likelihood $L(\mathbf{m})$:

$$\begin{aligned} \mathrm{L}(\mathbf{m}) &= \int_{\mathbb{D}} \mathrm{d}\mathbf{d} \frac{\rho_d(\mathbf{d})\theta(\mathbf{d}|\mathbf{m})}{\mu_{\mathrm{D}}(\mathbf{d})} \\ &= k \exp\left(-\frac{1}{2}(\mathbf{d} - g(\mathbf{m}) - \boldsymbol{\mu}_D)^{\mathsf{T}} \mathbf{C}_D^{-1} (\mathbf{d} - g(\mathbf{m}) - \boldsymbol{\mu}_D)\right) \end{aligned} \tag{4.5}$$

where the covariance matrix $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_T$ and mean vector $\boldsymbol{\mu}_D = \boldsymbol{\mu}_d + \boldsymbol{\mu}_T$ is an addition of the data uncertainties and modeling errors covariance, and mean respectively. The result that Gaussian data uncertainty and modeling errors are additive is simple yet non-trivial (Mosegaard and Tarantola, 2002). If both the data uncertainties and modeling errors are assumed to be centred Gaussian, i.e. $\boldsymbol{\mu}_D = 0$, Equation 4.5 reduces to:

$$\mathrm{L}(\mathbf{m}) = k \exp\left(-\frac{1}{2}(\mathbf{d} - g(\mathbf{m}))^{\mathsf{T}} \mathbf{C}_D^{-1} (\mathbf{d} - g(\mathbf{m}))\right) \tag{4.6}$$

The results in Equation 4.5 and Equation 4.6 is used frequently throughout the work in this thesis as a way of incorporating modeling errors into the likelihood.

Because most probabilistic inversion methods of seismic data involve uncorrelated noise, the shape (or color) of the data uncertainty and modeling errors is not well established. Riedel et al. (2003); Chen et al. (2007) proposes to describe the shape of the correlated noise from theory as proportional to the apparent covariance of the data. This is done using a fixed maximum likelihood estimate (Gerstoft and Mecklenbräuker, 1998). Gouveia and Scales (1998) suggest to treat different sources of noise on seismic data with an individual covariance matrix. Under the assumption of independence between the errors, the covariance matrices are added and used in a Gaussian likelihood. The

individual covariance matrices contain information on ambient noise, residual statics, scaling factors, and discretization errors. However, the caveat here is that all of these covariance matrices need to be determined before the inversion.

To avoid too many subjective choices in the noise model, data uncertainties have also been added as an unknown parameter in a Bayesian inversion framework (Dosso and Holland, 2006). Dettmer and Dosso (2012) suggested an extension to this paradigm with an autoregressive model which enables capturing of the general covariance structure. Noise estimation has also been applied in the frequency domain (Ray et al., 2016)

In this thesis, the covariance of the data uncertainty is assumed to be uncorrelated white noise as in Equation 4.3. The validity of the Gaussian assumption is also discussed. The correlated part of the noise is associated with theoretical errors (modeling errors). The modeling error is estimated prior to the inversion in e.g. Appendix H.4 and Appendix H.5. In Appendix H.3 and Appendix H.6 the correlated noise is instead assumed to imitate data as:

$$\mathbf{C}_T = \sigma_T^2 \mathbf{C}_{T,shape} = \sigma_T^2 \frac{\mathbf{H}\mathbf{C}_m\mathbf{H}^\mathsf{T}}{\max(\mathbf{H}\mathbf{C}_m\mathbf{H}^\mathsf{T})} \tag{4.7}$$

where $\mathbf{H}$ is a linear seismic forward operator, $\mathbf{C}_m$ is the prior covariance model of the model parameters and σ_T^2 is the variance of the theoretical errors. Assuming that the noise is imitating the shape of the data is completely subjective and in some sense arbitrary. However, noise that tends to have the same frequency and shape as the signal is very difficult to distinguish during data processing, as demonstrated in the example in Figure 4.1. Noise that resembles the data structure follows logically within this category, and is probably the worst possible

scenario in terms of filtering noise from a signal. Within this premise, the data covariance in Equation 4.7 is actually taking the worst possible noise scenario into account and can conversely be considered a conservative choice in terms of correlated noise.

"Don't raise your voice, improve your argument."

Desmond Tutu (2004), Address at the Nelson Mandela Foundation in Houghton, Johannesburg, South Africa, 23th of November 2004

5 Discussion

Various challenges exist when inverting seismic data for elastic parameters. First an inversion method must be chosen. The method might require more or less approximative physics in order to solve the forward problem. Deterministic solutions are bias-prone and often ill-posed simply because of the noise on the data. If instead probabilistic inversion is desired, a probabilistic formulation of the model parameters is needed. Although this solves the problem of ill-posedness, it adds several challenges in terms of providing reasonable a priori estimates of both prior and noise model. The large volumes of seismic data usually require the solutions to be analytically obtainable.

For a linear-least squares solution to the inverse problem, the forward is constrained to be linear, and the likelihood and prior must be Gaussian. Wave-propagation however is not linear, and subsurface variability is very unlikely smooth and Gaussian. By introducing a linear forward along with a Gaussian prior and noise model for a system that is likely both non-linear and non-Gaussian, one inevitably introduce biases in the posterior solutions.

With all the above-mentioned caveats and known pit-falls one might consider the feasibility of seismic inversion in the first place. Nonetheless, inversion of seismic data is still widely applied in geophysical exploration throughout the industry. Inversion of seismic data has also sparked, and still sparks, a lot of interest in the academic field, as witnessed by the extensive literature on the subject. Because seismic inversion is still around it can safely be concluded that the methodology has its merits, and have so far provided geophysicists with adequate models of the subsurface.

However, as easily accessible hydrocarbons become more scarce, an increased demand for improved solutions is consequently needed. For probabilistic linear inversion of seismic data, the question is how much the results are biased and what can be done in order to prevent some of these biases? Furthermore, since the solution is probabilistic (and not deterministic), it is pivotal to ask whether the resulting posterior uncertainties can be trusted? In the following, the challenges in probabilistic linear inverse problems, namely quantifying the effect (i.e. the error) of linearizing a non-linear problem and dealing with non-stationarity of statistical properties of the subsurface, is discussed. Finally, the method of estimating the noise model as part of the inversion is covered.

The forward model provides the link between physics and statistics. The physical system can probably never be fully described. This is partially because it is not possible to obtain the true physics and partially because simplistic models sometimes are preferred, as pointed out in the previous chapter.

Seismic raw data are recorded as the source wave propagates the subsurface. Numerous state-of-the-art software and codes are available to model wave propagation (Igel, 2017). This allows a near perfect forward model between elastic parameters and seismic raw data.

Following the probabilistic framework outlined by Tarantola and Valette (1982), it is theoretically possible to set up an inverse problem directly for seismic raw data. Any imperfections in the forward model is then the modeling error, and one needs to define a suitable noise and prior model. Due to the computational inefficiency of using full waveform modeling as the forward model, this is still in practice not an option. Some data reduction in the form of processing is needed in order to make inversion of seismic data computationally feasible when dealing with these quantities of data. Processing raw seismic data allows the use of computationally efficient forward models (e.g. the convolutional model).

Two main views on processing should be recognized here. Processing can either, in the view of Tarantola and Valette (1982), be viewed as ”throwing out” important information from the raw data while introducing biases. A forward describing the raw data would hence be preferable. Alternatively, processing is also argued to be advantageous to inversion because it does not require a noise model. The processing is, in contrast to inversion, thought to be free of

assumptions and should always be used prior to any inversion (Claerbout et al., 2004).

In this thesis, processing is seen as a necessary evil that allows the use of a simpler forward. The quality of processed data relies heavily on the processor and the processing sequence, and one should per default assume the data to be both inaccurate and imprecise (Castagna, 1996). In Figure 2.8 it is shown that processing and the use of the linear forward induce a difference between what is expected and what is the actual data. In general one is bound to expect biases from processing raw seismic data (see Figure 2.8 or Appendix H.5).

For an inversion purist, this might be unacceptable, but seeing processing as a necessary step sometimes allow solutions to be obtainable at all, and is therefore preferable for a practitioner. The amount of efficient forward solvers in existence combined with increased computer power would potentially lead to less processing and instead the use of raw seismic data directly in probabilistic inversion (Igel, 2017). A probabilistic inversion approach with full waveform modeling of seismic data is still in the distant future. In the mean time it might instead be beneficial to concentrate on quantifying the effect of using simpler forward models.

5.2 Modeling Errors

Modeling errors generally arise from imperfect theory as established in the previous chapters. Hansen et al. (2014) propose to estimate and describe these errors with a Gaussian distribution. The basic idea is to generate a subsurface realization from the prior distribution, where both the inexact and "exact" forward response is calculated. The difference between the two is then a

realization of the error for the specific choice of prior model. An example of the difference between two AVA forward responses $\mathbf{d}_{error}$ is depicted in Figure 2.8. Generating a sample of the modeling error by repeating this process makes the estimation of a Gaussian distribution possible by approximately describing this sample. A similar approach is used to quantify "non-linearity errors" (Møller et al., 2001). The non-linearity error refers to the difference between the simple linear forward used in the inversion and a more complex non-linear forward model which is more realistic.

The proposed methodology has several advantages. Firstly, it allows for a Gaussian estimation of potentially non-linear errors. Secondly, if the modeling errors are Gaussian, they can be accounted for in a linear least-squares solution. Finally, the methodology is independent of data because the sample is based solely on the chosen prior and forward. Hence, the modeling error can be estimated prior to inversion.

In e.g. Riedel et al. (2003) they estimate the modeling errors directly from the seismic data. The argument here is that the estimate is independent of personal choices for the prior information. On the contrary, it can be argued that a prior model must be chosen regardless when doing probabilistic inversion. Furthermore, in the methodology presented in Riedel et al. (2003), the data is inherently used twice, both for the estimation of modeling errors and again in the inversion. This conflicts with the idea of completely independent sources of information. Based on the reasoning above, the methodology of Hansen et al. (2014) is adapted in this thesis in order to quantify and estimate modeling errors in seismic AVA data.

Modeling errors for processed seismic AVA data are commonly acknowledged

in the literature (Gerstoft and Mecklenbräuker, 1998; Buland and Omre, 2003a; Riedel et al., 2003; Dosso and Holland, 2006; Chen et al., 2007; Rabben et al., 2008; Bosch et al., 2010; Aune et al., 2013). Many sources of modeling errors are recognized, including; using the 1D convolutional model to reflect a 3D physical system, the use of the acoustic-wave equation as opposed to the anisotropic visco-elastic wave equation, imperfections in data processing, general anisotropy considerations, the effects of processing the raw data, uncertain wavelet estimates, and uncertainty on the background velocity model (Ball et al., 2015; Li et al., 2015; Thore, 2015).

The background velocity model is necessary for all the critical processing steps shown in Figure 2.9. The background velocity is thus essential and is routinely estimated as part of pre-processing the raw seismic data. The importance of the velocity model in seismic processing is illustrated in the works of Weibull (2013).

Modeling errors due to processing are often neglected when the data is used in an inversion setup (or in other seismic disciplines). This is perhaps done out of convenience and/or because processing errors are difficult to quantify. The work in Appendix H.5 shows that modeling errors are to be expected on seismic AVA data even for processing of almost "noise-free" raw data with a limited amount of bias-infusing processing steps. Using the methodology in Hansen et al. (2014) a Gaussian model is constructed, which enables a better description of the modeling errors and hence improves the inversion results.

The processing is conducted knowing the true background velocity model. Since processing is dependent on the velocity model, the significance of the modeling errors are expected to increase significantly if the velocity is not known but rather estimated during pre-processing. Quantifying the effect of imperfect

velocity models is a natural extension to the presented work in Appendix H.5, and could potentially yield much improved inversion results.

A remark should be made concerning the size of the sample (N) for which the covariance of the modeling errors is based on. Figure 5.1 displays the estimated covariance for four different sample sizes of the modeling errors from the processing as in Figure 2.8. As N increases, the estimated covariance matrices converge towards a stable solution. The difference between Figure 5.1c and Figure 5.1d is minuscule compared with N = 50 in Figure 5.1a. In this study, N = 1000 is the preferred option, as a conservative choice for the sample size. However, the results in Figure 5.1 indicate that N = 500 is a reasonable threshold for the minimum size of the sample. This number corresponds well with results obtained in Hansen et al. (2014) for cross-hole tomography. Møller et al. (2001) used a smaller sample size adding some uncertainty to the reliability of the estimate.

Linear approximations to Zoeppritz equations are usually considered valid for a subsurface with low variability in the subsurface properties. In Appendix H.4 it is shown that even using a linear approximation to Zoeppritz equations, when the subsurface is smoothly varying (in this case Gaussian), results in significant modeling errors, especially, for angles above 30°. This is in accordance with previous results (Buland and Omre, 2003a; Mavko et al., 2009). If the subsurface is non-Gaussian, in this case pluri-Guassian, the modeling errors are significant for angles well below 30°.

The work in Appendix H.4 assumes perfect processing of the seismic data, i.e. that modeling errors can only be attributed to linearization of Zoeppritz

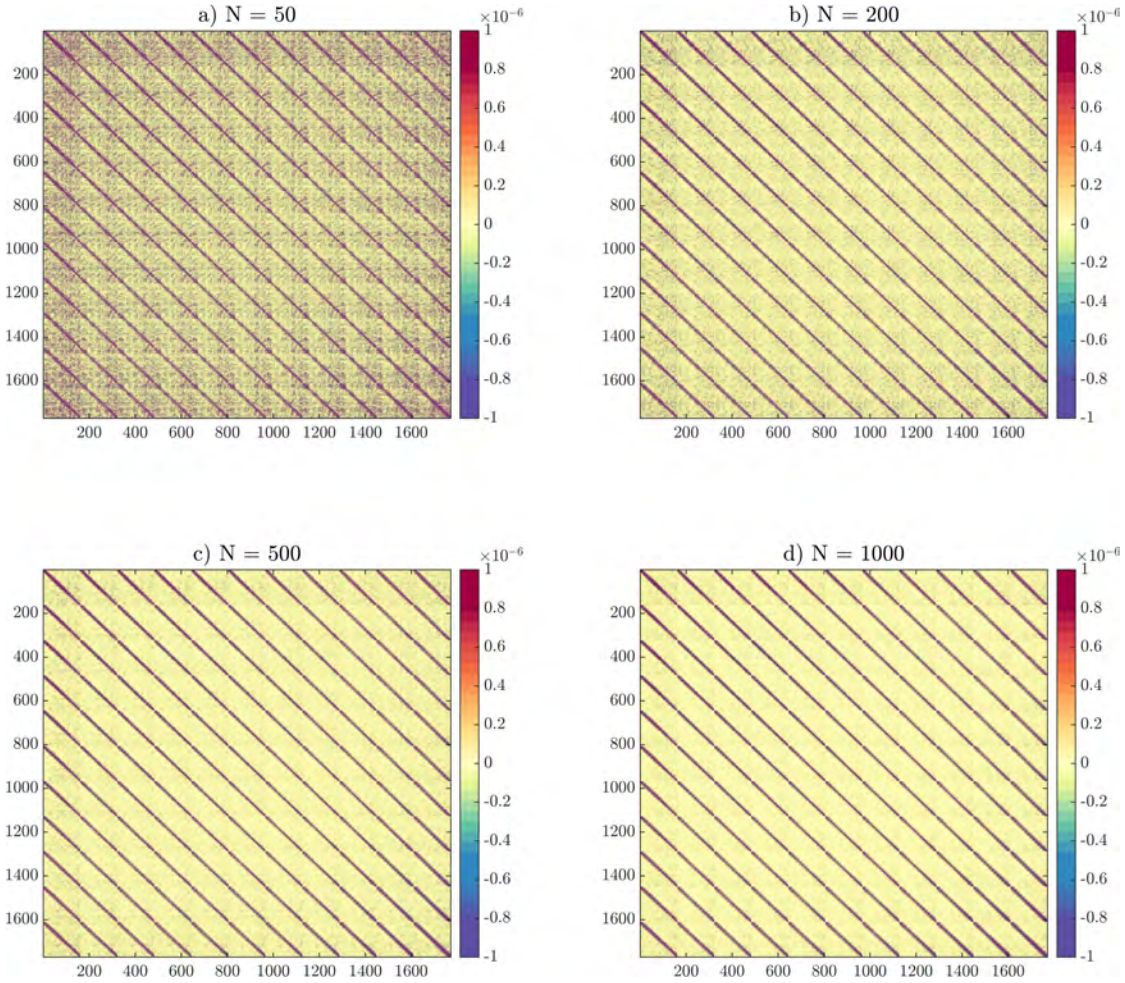


Figure 5.1: Estimated covariance models for processing error with different sample size. a) N = 50, b) N = 200, c) N = 500, d) N = 1000

equations. This indicates that even modeling errors, which may be interpreted as a lower limit of what can be expected for seismic data, are important to recognize and account for. This is the case even if the modeling errors at first may seem insignificant. The modeling errors for both processing and for the linearization of Zoeppritz equation increase with increasing angle. Larger angles of incidence include information about shear wave velocity (v_s) and density (ρ), whereas information about acoustic impedance is coupled to smaller angles. Modeling errors are therefore important to account for in order to obtain reasonable prediction of e.g. the v_p/v_s ratio, whereas acoustic impedance is not as prone to modeling errors.

One could argue that the combination of different sources of modeling errors may cancel out. If this is not the case, which there is no immediate evidence supporting, then modeling errors are perhaps the most important, yet unresolved, problems of probabilistic inversion of seismic data. The proposed method of approaching quantification/estimation of modeling errors has shown promising results so far for seismic data. Estimating the modeling errors makes them possible to be accounted for in probabilistic inversion and seems crucial in avoiding biases using approximative physics. This would ultimately lead to trustworthy predictions, and perhaps more importantly, trustworthy uncertainties. Until it becomes computationally viable to perform probabilistic full waveform inversion on raw seismic data, the methodology offers a step in the right direction. Basically, estimating modeling errors allows the geophysicist to use a "cheap" forward, while acknowledging the more exact forward solution.

Using a standard Monte-Carlo sampling technique (Mosegaard and Tarantola, 1995), each set of data would have to be sampled with approximately one million

iterations. With an expensive forward model, this process is unfeasible. However, the Gaussian model for modeling errors can be obtained with a sample size of 500, as shown earlier. Enormous computation savings can therefore be made by allowing the use of a much faster forward model. The slight caveat for using the approximate forward model with the inclusion of modeling errors comes in the form of posterior resolution. The resolution can never become as good as applying the more exact forward. The alternative of ignoring modeling errors is most likely prone to be both less accurate and produce false posterior uncertainties.

In the least-squares solution to the probabilistic inverse problem, modeling errors are trivially accounted for in Equation 4.6 by addition of the covariance matrices. For an inversion case with a non-Gaussian prior model sampled with a Monte-Carlo technique, the inclusion of the modeling errors would potentially easy the sampling, because more realistic uncertainties on the model are used. It would actually allow for a larger misfit between proposed forward models and the observed data. Although slightly un-intuitive, this result is confirmed in Appendix H.5 to be beneficial in obtaining more realistic posterior solutions. The results would not only be more accurate but also less time-consuming to obtain. Preliminary results have so far confirmed this hypothesis, but would be an interesting topic for further research.

Using a Monte-Carlo sampling technique, a non-Guassian likelihood can also be used. This would make the modeling error sample possible to describe with a more precise distribution. The 1D marginal distribution of the modeling errors in both Appendix H.4 and Appendix H.5 shows that some outliers can be detected. These outliers generally drag the estimated Gaussian model in the

direction of higher variance. This trend could perhaps be modeled with e.g. a Voigt profile, Poisson distribution, or a Cauchy distribution. Even if the 1D marginal distribution can be mapped better using either of these distributions, there is no guaranty that the spatial distribution is better matched. One might also apply a normal score transformation of a Gaussian distribution to keep some of the spatial correlations in the modeling errors, while providing a better match to the 1D marginal distribution.

5.3 Non-stationarity in Physical Parameters

Statistical properties of physical parameters of the subsurface (acoustic impedance, density, porosity, resistivity etc.) can generally be considered to be spatially non-stationary (Bongajum et al., 2013; Bouchedda et al., 2017; Sabeti et al., 2017). Although methods exist for incorporating non-stationary in both prior and likelihood models (Aune et al., 2013), non-stationarity in statistical properties of the subsurface are typically ignored in seismic AVA inversion. Inverse solutions are instead conditioned on global stationary probability density functions (Sabeti et al., 2017). These global estimates could for instance be obtained from nearby wells.

As for the modeling errors, non-stationarity in the subsurface properties is probably ignored because it is too difficult in quantify. A stationary Gaussian prior model is furthermore convenient in terms of solving the probabilistic linear least-squares problem, and is positive definite if the spatial correlation function is positive definite. Positive definiteness is a useful mathematically property that makes stationarity desirable. Stationary Gaussian distributions are also easy to sample using e.g. Fast Fourier Transform Moving Average or Cholesky

decomposition. Cholesky decomposition makes generating realizations from a non-stationary Gaussian model possible, but is prone to be very inefficient when sampling large models (Aune et al., 2013; Hansen et al., 2016). Although there are some mathematical advantages and efficiency bonuses by using a stationary prior model, it is ultimately unrealistic in terms of the complex geology of the subsurface.

Aune et al. (2013) incorporate non-stationarity in the prior model by conditioning it on nearby well logs with non-stationary in the physical parameters. Sabeti et al. (2017) apply a spatial direct sequential simulation technique with local variograms, which reflects the non-stationary in the physical parameters. Nearby well logs plays a big role using both methodologies. The results are therefore dependent on the quality of the well logs. Spatial anisotropy is also considered in linearized seismic inversion (Bongajum et al., 2013). Non-stationarity is achieved by combining different covariance matrices, which are determined using the direction of maximum continuity for the seismic data. However, this method requires the spatial anisotropy to be defined prior to inversion both locally and globally. Using local rock-physical and geophysical likelihoods, Jullum and Kolbjørnsen (2016) assesses local quantities related to the posterior distribution of rock properties.

In contrast to these earlier approaches, the work in Appendix H.6 couples non-stationarity to the variance of the physical parameter. The correlation function is assumed to be known and stationary. The variance is analytically obtainable through the statistical model and is coupled to the variance in the seismic data. A localized variance estimator is defined within a neighbourhood. The full non-stationary variance pattern is obtained using a sliding-window

approach.

The variance estimates can be locally constrained in a hierarchical model (see e.g. Appendix G). A variation of this methodology is applied by Asfaw and Omre (2016) in order to predict non-stationarity in a precipitation data set. In Asfaw and Omre (2016), the physical parameters are also what is being measured in the data, which differs from the proposed methodology in Appendix H.6, where the prediction is conditioned on indirect seismic data. The variance is always positive, and hence the resulting non-stationary covariance matrix is positive definite. This allows the estimated variance to be used as a plug-in estimate in a Gaussian prior model, and hence in the analytical solution for the probabilistic linear least-squares inverse problem.

5.3.1 The philosophical crossroad

The proposed method has one major weakness. Because the data is used to obtain information on the variance of the model parameters and subsequently predict the values of the model parameters, this methodology clearly violates that sources of information should be independent. In other words, data is used for two purposes. This violation may reasonable be considered as a stopping point for further discussion and implementation. As stated in the introduction the posterior solution requires independence of information to provide correct results. One should of course not violate this prerequisite, if the theory breaks down. Yet, this puristic viewpoint is very difficult to fulfill in reality. It would, in principle, require the person doing the inversion to never look at the data. A slight glimpse of the data would probably directly or indirectly effect the choice of forward model, prior model, and noise model, which should be established

before the data enters the stage.

The traditional view of "full independence" generally leads to non-robust and inconsistent methods because the independence is most likely always violated (Journel, 2002). In practice, a lot of people in the inversion community tend to violate the independence of information, simply because it is highly impractical. Furthermore, how does one determine in the first place whether or not information is dependent or not?

Consider a person which explicitly assumes a Gaussian prior for a certain experiment. Before the inversion the person looks at the data and then establishes a mean for the model. Is the established mean then dependent on the data? The choice of the mean, let alone the use of a Gaussian model, is still completely arbitrary. On the other hand, the person has probably also acquired some knowledge by looking at the data, so in some sense the choice of the mean is not completely independent. In other words, it might be difficult to actually establish what independence is in the first place.

Inversion results may have such apparent biases, that they can simply be disregarded from having any physical substance. Is it then reasonable to redo the inversion with a different prior model or noise model? If so, how much is the independence of information then violated? Probably a lot, because the previous inversion results would have guided the new choice of prior and noise model. Does it make the new inversion results completely useless simply because of this violation?

The dilemma above presents a philosophical crossroad for the person doing the inversion. How much is one allowed to look at the data in order to determine things that should in principle be independent of the data. None? a little? some?

or is everything allowed? The answer to this question is non-trivial and is still unresolved. Consider the current problem of predicting acoustic impedance from a line of seismic data. If one instead used the adjacent seismic line to infer the variance in the prior model, would the sources of information then be independent? The data would certainly not be used twice. Nonetheless, the noise between the two data sets would definitely be correlated both in space and time and would hence not be independent.

In statistics there is a long tradition for using data multiple times for different purposes with the bi-product of loosing a certain amount of degrees of freedom. This is a very practical approach to the issue, but does not answer the above questions fully. As stated in Scales and Snieder (1998) one should at least make an effort in not making the distinction between prior and data undesirably fuzzy. This does not provide a definitive answer though. The so-called *tau-model*, originally proposed by Journel (2002) is partly trying to resolve the issue of dependency in the sources of information. In the tau-model the different sources of information are aggregated given a certain scaling (i.e. "tau-factor").

The probability aggregation method is formalized by Allard et al. (2012). If the information is completely independent the tau-factor is set to 1. The tau-model then collapses down to the conjunction of information. For all dependent cases, the tau-factor can be set to lower values. In this sense, the tau-model can be considered a more general description of the inverse problem. Although this formulation is quite elegant, it is difficult to use in practice. Firstly, how does one tune these tau-factors? We have no direct measure to quantify these in the first place. Secondly, the interpretation of tau-factors is also tricky. What does a tau-factor of 0.7 signify exactly? How dependent are the data and in which sense?

5.3.2 Prediction using the non-stationary variance estimates

In the following the practical notion that the data are allowed to be used for several purposes is adopted while acknowledging that there is an issue of dependency in the sources of information in the proposed methodology. The results from Appendix H.6 indicate that a heterogeneous variance pattern of the physical parameters (in this case acoustic impedance) can be estimated directly from seismic data using such an approach. The estimates are slightly underestimating the true variance, which is in accordance with the loss of degrees of freedom. The inverse solution shows a greatly improved posterior resolution compared with the use of a prior model stationary variance.

The method is applied on a real data set, situated around a salt-diapir in the danish North Sea. It is expected that physical parameters around the salt-diapir should behave non-stationary. In accordance with this, the results show that higher variances are found around the diapir. If the non-stationary variance estimate is used for inversion, the resulting posterior distribution is most likely a better description of the subsurface than using a prior model with stationary variance. What is meant here by a ”better” description, is not only increased precision in the prediction (i.e. the most probable model), but more importantly, more trustworthy posterior uncertainty estimates. Since the diapir is of interest in terms of hydrocarbon exploration, a better description provides a better risk assessment for drilling, etc.

The results depend on the quality of the data and level of heterogeneity in the subsurface. The method should obviously not be used if one expects the behavior of the subsurface to be stationary. Using the constrained estimates from the hierarchical estimator showed however that only small contrasts in subsurface

variance is actually needed for the method to provide a better description of the subsurface than a corresponding model with stationary variance. The final results are unsurprisingly sensitive to the size of the local neighbourhood. A trade-off should be made between having a certain minimum size of the neighbourhood and not smoothing the results too much by having a large neighbourhood. One has to assume a relation between the variance of the noise and variance of the physical parameter. In Appendix H.6 a proportional relationship is assumed. This poses the problem of determining a correct proportionality constant. The constant is approximated using the assumed noise levels.

From the above it is clear that the method relies on an accurate description of the noise on the data. Because the variance estimate of the physical parameter is assumed to have a proportional relationship to the data variance, any unmodeled feature in the seismic data will be interpreted in terms of variance of the physical parameter. Regions of data with different noise, e.g. a ”bad” trace, will therefore affect the variance estimate. If the method is used in conjunction with e.g. a better description of the modeling errors, the results may be substantially improved. A better insight into the noise on the data would also provide a more reasonable approximation of the proportionality constant. In the real data case, a relatively conservative choice of noise levels is chosen. A better description of the noise (including modeling errors) in space and time would probably create less dependence between different data sets. Thus, the issue of independence would be reduced.

Proper estimation of data uncertainties is still basically unsolved in many geophysical disciplines, including seismic inversion (Igel, 2017). From the previous sections, the importance of correctly describing the uncertainties in seismic data and the background theory should be clear. Traditionally, the noise model is established prior to inversion. The uncertainty of specific measurement equipment might be easily quantifiable. Many sources of errors in seismic data are however more complex and correlated and are non-trivial to quantify.

In order to circumvent this obstacle, a hierarchical model can be combined with a sampling strategy to infer properties of the noise model as part of the inversion (Gelman et al., 2014). In this paradigm, the data is used to ”decide” what is signal and what is noise. The major upside of estimating the noise model directly during the inversion is the removal of potential biases which may arise due to the subjective choices when constructing the noise model in the classic paradigm.

The methodology can also be extended to cover the inference of information on the prior model (Malinverno and Briggs, 2004) and even the parameterization of the problem (Sambridge et al., 2006). The latter is known as trans-dimensional inversion and has recently been applied for large-scale seismic problems (Cho et al., 2017). Regardless of which entity (e.g. prior model, noise model, parameterization, etc.) is deemed inferable using a hierarchical approach, it adds complexity to the stochastic model which may or may not be needed. The elicitation of hyper distributions is another issue related to the methodology.

Buland and Omre (2003b) estimated seismic noise variance on AVA seismic data using such an approach. Malinverno and Briggs (2004); Malinverno and

Parker (2006) formalized the method for probabilistic linear least-squares problems. Recently, properties of the noise model has been estimated as part of the inversion in combination with trans-dimensional inversion (Bodin et al., 2012; Dettmer and Dosso, 2012; Ray et al., 2013, 2016).

In Appendix H.3 the methodology proposed by Malinverno and Briggs (2004); Malinverno and Parker (2006) is used in order to infer the variance of the noise model as part of the inversion. The variance is treated as a hyper parameter for the hyper distribution over the variance of the noise. A schematic representation of the stochastic model can be seen in Figure G.1. Synthetic AVA data is constructed using realizations from the Buland and Omre prior model (Section 3.1) and the forward relationship in Equation 2.19. Using the arguments from Section 4.3, correlated noise is primarily added to the forward response in order to obtain the final AVA data sets. Three inversion case studies are presented corresponding to different colors of Gaussian noise $\mathbf{C}_D$:

$$\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_T = \sigma_d^2 \mathbf{I}_n + \sigma_T^2 \mathbf{C}_{T,shape} \tag{5.1}$$

In the first case, the color is assumed white (i.e. uncorrelated), thus $\sigma_T^2 = 0$ and σ_d^2 is treated as a hyper parameter. In the second case the noise shape $\mathbf{C}_{T,shape}$ is identical to the noise shape used to create noise in the data. This noise resembles the shape of the data as in Equation 4.7. In the third case, an approximate knowledge of the shape of the noise is used. Both σ_T^2 and σ_d^2 are treated as hyper parameters in the two last cases.

The inversion results show that it is possible in all cases to accurately estimate the variance of the uncorrelated noise σ_d^2 on the data. However, because the uncorrelated part is not the main component of the noise added to the data, it is

not as crucial to estimate, as the correlated part. A simple uncorrelated noise model does not allow the total variance of the noise to be recovered, and as a consequence significant overfitting of the data is demonstrated. One cannot expect the classic implementation of the inversion algorithm to fully capture the complexity of the noise on the data.

If the correct shape of the noise $\mathbf{C}_{T,shape}$ is known prior to inversion, it is possible to estimate the correct level of both correlated and uncorrelated noise on the data. Inferring the noise level with the approximate noise shape, σ_T^2 becomes underestimated. The overfitting of data is not as significant as for the case with the assumed uncorrelated noise shape. In summary, the results from Appendix H.3 indicate that correlated noise amplitudes are difficult to estimate as part of the inversion. The result will generally depend on the shape of the noise on the data and the noise which is assumed. If such an approach is considered, good a priori knowledge of the shape is required.

A counterargument to these findings is that unrealistic amounts of correlated noise is perhaps added to the synthetic data compared to real seismic data, i.e. the quality of the data is much lower than for real seismic data. It is difficult to determine whether this is a valid point. On one hand, noisy data are processed by experienced processors, and one would like to believe that the signal to noise ratio is reasonable in the final data intended for inversion. On the other hand, as described earlier it is difficult for humans to separate correlated noise from signal. The gut-feeling that one might have regarding the level of noise on the data might be misconstrued. Even data that look noise-less might be noisy.

The importance of the correlated noise should not be neglected in any case. However, knowing the shape of the noise prior to inversion is a problem in itself,

and as shown in Appendix H.3, only if the shape is described correctly will the variance estimate be accurate. The methodology can be extended to estimate properties of the color of noise as well. Consider the covariance matrix $\mathbf{C}_{T,shape}$ describing the shape of the noise model:

$$\mathbf{C}_{T,shape} = \frac{\mathbf{H}\mathbf{C}_m\mathbf{H}^\top}{\max(\mathbf{H}\mathbf{C}_m\mathbf{H}^\top)} \tag{5.2}$$

If the range parameter (i.e. the correlation in time) in the $\mathbf{C}_m$ is treated as a hyper parameter, the color can then be inferred in the proposed setup. The marginal likelihood follows from Equation 4.6:

$$\mathrm{L}(\mathbf{m}) = (2\pi)^{-n/2}|\mathbf{C}_D|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{G}\mathbf{m})^\top\mathbf{C}_D^{-1}(\mathbf{d}-\mathbf{G}\mathbf{m})\right) \tag{5.3}$$

By sweeping over the hyper parameters a likelihood "map" can be constructed, showing regions of high and low likelihood depending on σ_T^2 and σ_d^2. Figure 5.2 shows such a sweep of the hyper parameters σ_T^2 and σ_d^2 for different choices of range parameter. The target amplitude of the noise variance is shown with a black dot. The reason that the amplitude of uncorrelated noise σ_d^2 is easy to estimate using this approach is clearly visible in Figure 5.2. The peak likelihood value is centered around a narrow band trending up and down, which indicates a high sensitivity towards σ_d^2. The peak likelihood is placed around the target variance for all possible combinations of σ_T^2 and range. In contrast, less sensitivity in the likelihood is shown for the variance of the correlated noise σ_T^2.

For the correct range of $r = 0.005$ ms, the highest possible likelihood is found around the target value. This explains why it is possible to obtain the correct

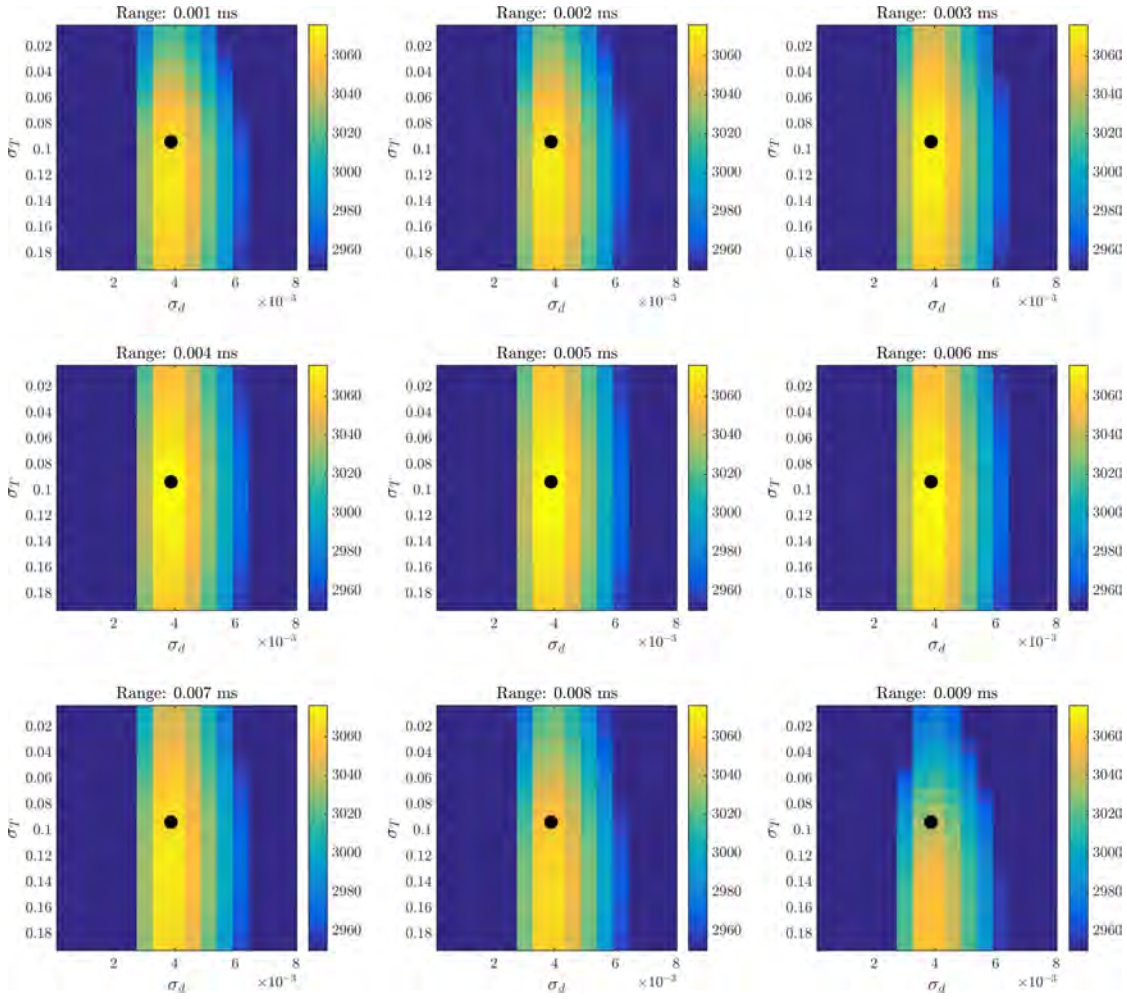


Figure 5.2: Marginal likelihood for a sweep of hyper parameters (σ_T^2 and σ_d^2) using different Gaussian noise models with different ranges. The noise model is constructed using the forward model. Each subplot displays the likelihood for a different range parameter. The true variance level is marked with a black dot. The correct range, and hence true color of the noise is $r = 0.005$ ms.

estimate of σ_T^2 if the correct shape is known in advance. Unfortunately, the likelihood maps are almost identical for all choices of range parameter. This indicates that there is practically zero sensitivity towards the range parameter with this setup of the noise model. A sampling technique with a random walk (e.g. the Metropolis-Hastings algorithm) would have difficulties in effectively estimating the correlation range, and hence the color of the noise.

One might also be tempted to estimate the noise in a broader context to remove further subjective prior assumptions. Let instead the shape covariance $\mathbf{C}_{T,shape}$ be described by a Gaussian correlation function as in Equation 3.6:

$$\mathbf{C}_{T,shape} = \exp\left[-\left(\frac{TWT}{r}\right)^2\right] \tag{5.4}$$

By doing this the forward is left out of the noise model and the only remaining assumption is the chosen correlation function, which in this case is Gaussian. Because the angle stacks are correlated in time, the covariance matrix is correlated for each trace. The marginal likelihood for different correlation lengths is shown in Figure 5.3. As in the previous case, the amplitude of the uncorrelated noise σ_d^2 is easy to determine and coincides with the target value for all choices of σ_T^2 and range. The likelihood maps indicate relatively strong sensitivity towards the range parameter where the highest values are found around $r = 0.008$ ms. Unfortunately, the likelihood is highest for low values of σ_T^2. This means that σ_T^2 is bound to be underestimated using a sampling technique with a random walk using this noise model. This is also in accordance with the findings in Appendix H.3.

The examples with the likelihood for different covariance matrices indicate

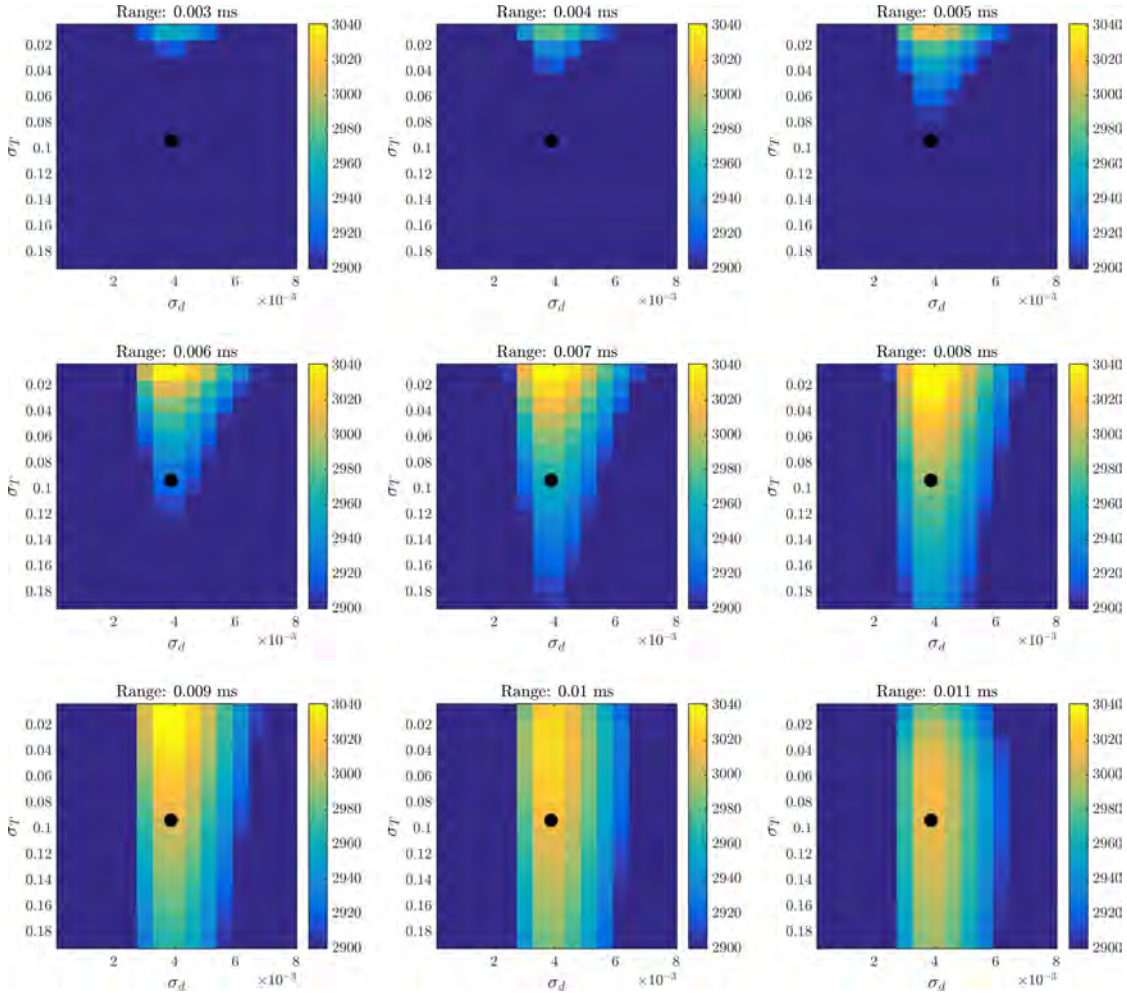


Figure 5.3: Marginal likelihood for a sweep of hyper parameters (σ_T^2 and σ_d^2) using different Gaussian noise models with different ranges. Each subplot displays the likelihood for a different range parameter. The true variance level is marked with a black dot.

that care should be taken when trying to infer the noise model as part of the inversion. The less information provided prior to the inversion about the shape of the noise, the worse the final estimation of the variance and the color most likely is. A wrong shape of the noise could potentially lead to biases in the inversion results even if the variance amplitudes are correctly estimated as demonstrated in Appendix H.3. If the methodology is extended to e.g. estimate non-stationarity in noise, this might further reduce the trustworthiness of the posterior models. The results indicate that one is probably better off by providing a reasonable estimate of the shape of the noise before applying the methodology. This estimate could e.g. come from estimating modeling errors as suggested previously.

The estimate of correlated noise binds together all the proposed efforts to improve linear probabilistic seismic inversion.

"Once again you've put your keen and penetrating mind to the task and as usual come to the wrong conclusion!"

J.K. Rowling (1999), Harry Potter and the Prisoner of Azkaban

6 Conclusion

This thesis investigates and identifies some of the major challenges in probabilistic inversion of seismic data. The main focus is on the popular linear least-squares solution and the inversion of AVA seimic data, which requires a Gaussian prior model and a linear-Gaussian likelihood.

Non-stationarity in the variance of the prior Gaussian model can be estimated directly from the data prior to inversion using maximum marginal likelihood estimators and a sliding window technique. A non-stationary variance model is more in accordance with the expected heterogeneous appearance of the subsurface. Improved posterior resolution of inversion results are acquired

compared with using a stationary prior model.

Seismic data is likely to contain noise which is correlated. If not accounted for, correlated errors are seemingly a hugely important issue in probabilistic inversion of seismic data. Correlated noise might be coupled to the use of inaccurate linear physics and the processing sequence of the raw data. Results show that it is possible to quantify and estimate the effect of both sources of errors. This gives rise to a better estimation of the uncertainty on seismic data. Taking modeling and processing errors into account improves posterior resolution and ultimately leads to more trustworthy inversion results.

Properties of the noise model can be estimated as part of the inversion of seismic data. In such a setting the complexity of the likelihood plays an important role in correctly estimating the properties of the noise model. The appearance of the likelihood is highly affected by correlated noise, as witnessed in Figure 5.2-5.3. If such an approach is used in estimating the properties of the noise model, care should be taken because overfitting data is a latent risk of the method.

Until a computationally feasible non-linear probabilistic inversion strategy based on FWM becomes available, least-squares solutions will still be the primary tool for large scale seismic inversion. This requires that methods are developed in order to quantify some of the errors that are associated with using linear forward approximations and Gaussian distributions. If these effects can be accounted for, a temporary middle ground can potentially be reached.

Previous efforts have mainly focused on providing more realistic Gaussian prior models (e.g. bimodal, skewed, etc.) which may provide more realistic depictions of the subsurface. The noise model and hence likelihood also plays a huge part in seismic inversion. A proper description of the correlated noise will

provide more trustworthy posterior uncertainty estimates. Yet, this seems to be an area which has been slightly overlooked in previous research. A fundamental change in the view of correlated noise might be both tedious and unpleasant but is necessary because demands for accurate descriptions of the subsurface increases. Quantifying modeling errors and taking them into account is an important first step in this direction.

References

Keiiti Aki and Paul G. Richards. *Quantitative seismology, theory and methods: Vol I and II*. W.H. Freedman Publishing, 1 edition, 1980.

Keiiti Aki and Paul G. Richards. *Quantitative Seismology*. University Science Books, 2 edition, 2002. ISBN 0-935702-96-2.

Wubshet Alemie and Mauricio D. Sacchi. High-resolution three-term AVO inversion by means of a Trivariate Cauchy probability distribution. *Geophysics*, 76(3):R43–R55, 2011. ISSN 00168033. doi: 10.1190/1.3554627.

D. Allard, A. Comunian, and P. Renard. Probability Aggregation Methods in Geoscience. *Mathematical Geosciences*, 44(5):545–581, 2012. ISSN 18748961. doi: 10.1007/s11004-012-9396-3.

Philip A. Allen and John R. Allen. *Basin Analysis - Principles and Applications*. Blackwell Publishing, 2 edition, 2004. ISBN 978-0-632-05207-3.

Margaret Armstrong, Alain Galli, Gaëlle Le Loc'h, Didier Renard, Brigitte Deligez, Francois Geffroy, and Rémi Eschard. *Plurigaussian Simulations in Geosciences*, volume 2. Springer, 2011. ISBN 9783642196065. doi: 10.1007/978-3-642-19607-2.

Zeytu Gashaw Asfaw and Henning Omre. Localized/Shrinkage Kriging Predictors. *Mathematical Geosciences*, 48(5):595–618, 2016. ISSN 18748953. doi: 10.1007/s11004-015-9626-6.

R C Aster, B Borchers, and C Thurber. *Parameter Estimation and Inverse Problems*. Academic Press, 1 edition, 2004.

Erlend Aune, Jo Eidsvik, and Bjørn Ursin. Three-dimensional non-stationary and non-linear isotropic AVA inversion. *Geophysical Journal International*, 194(2): 787–803, 2013. ISSN 0956540X. doi: 10.1093/gji/ggt127.

Vaughn Ball, Luis Tenorio, Christian Schiott, J P Blangy, and Michelle Thomas. Uncertainty in inverted elastic properties resulting from uncertainty in the low-frequency model. *The Leading Edge*, 34(9):1028–1035, 2015. doi: https://doi.org/10.1190/tle34091028.1.

Francesco P. Battaglia, Gideon Borensztajn, and Rens Bod. Structured cognition and neural systems: From rats to language. *Neuroscience and Biobehavioral Reviews*, 36(7):1626–1639, 2012. ISSN 01497634. doi: 10.1016/j.neubiorev.2012.04.004. URL http://dx.doi.org/10.1016/j.neubiorev.2012.04.004.

Mario Bertero and Patrizia Boccacci. *Introduction to Inverse Problems in Imaging*. Institute of Physics Publishing, 1 edition, 1998. ISBN 0750304391.

Thomas Bodin, Malcolm Sambridge, Nick Rawlinson, and Pierre Arroucau. Transdimensional tomography with unknown data noise. *Geophysical Journal International*, 189(3):1536–1556, 2012. ISSN 0956540X. doi: 10.1111/j.1365-246X.2012.05414.x.

Thomas Bohlen, K Daniel, and Stefan Jetschny. SOFI2d , seismic modeling with finite differences, 2D - acoustic and viscoelastic version, Users Guide, 2015. URL http://www.gpi.kit.edu/Software.php.

E. L. Bongajum, J. Boisvert, and M. D. Sacchi. Bayesian linearized seismic inversion with locally varying spatial anisotropy. *Journal of Applied Geophysics*, 88:31–41, 2013. ISSN 09269851. doi: 10.1016/j.jappgeo.2012.10.001. URL http://dx.doi.org/10.1016/j.jappgeo.2012.10.001.

Miguel Bosch, Tapan Mukerji, and Ezequiel F. Gonzalez. Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review. *Geophysics*, 75(5):A165–A176, 2010. ISSN 00168033. doi: 10.1190/1.3478209.

Abderrezak Bouchedda, Giroux Bernard, and Erwan Gloaguen. Constrained electrical resistivity tomography Bayesian inversion using inverse Matérn covariance matrix. *Geophysics*, 82(3):E129–E141, 2017.

George E.P. Box and George C. Tiao. Bayesian Inference in Statistical Analysis. *A Wiley-Interscience Publication*, page 608, 1992. doi: 10.1002/9781118033197. URL http://doi.wiley.com/10.1002/9781118033197.

Arild Buland and Henning Omre. Bayesian linearized AVO inversion. *Geophysics*, 68(1):185–198, 2003a. ISSN 00168033. doi: 10.1190/1.1543206.

Arild Buland and Henning Omre. Bayesian wavelet estimation from seismic and well data. *Geophysics*, 68(6):2000, 2003b. ISSN 00168033. doi: 10.1190/1.1635053.

G. Cambois. AVO processing: Myths and reality. *CSEG Recorder*, 26(3):30–33, 2001. ISSN 10523812. doi: 10.1190/1.1816505. URL http://link.aip.org/link/?SGA/20/189/1&Agg=doi.

John P. Castagna. AVO ANALYSIS: TUTORIAL AND REVIEW: 1995-1996 SEG Distinguished Lecture condensed from J.P. Castagna and MM. Backus, eds., Offset-Dependent Reflectivity—Theory and Practice of AVO Analysis Society of Exploration Geophysicists, 1993. In Jory A Pacht, Robert E Sheriff, and Bob F Perkins, editors, *Stratigraphic Analysis Utilizing Advanced Geophysical, Wireline and Borehole Technology for Petroleum Exploration and Production*. SEPM Society for Sedimentary Geology, 1996.

John P. Castagna and Milo M. Backus. *Offset-Dependent Reflectivity: Theory and Practice of AVO Analysis*. Society of Exploration Geophysicists, 1 1993. ISBN

978-1-56080-059-0. doi: 10.1190/1.9781560802624. URL http://library.seg.org/doi/book/10.1190/1.9781560802624.

Marwan Charara, Christophe Barnes, and Albert Tarantola. The state of affairs in inversion of seismic data: An OVSP example. In *66th Annual International Meeting: Society of Exploration Geophysics*, pages 1999–2002, 1996.

M A Chen, Michael Riedel, R D Hyndman, and Stan E. Dosso. AVO inversions of BSRs in marine gas hydrate studies. *Geophysics*, 72(2):C31–C43, 2007. ISSN 0016-8033. doi: 10.1190/1.2435604.

Youngchae Cho, Dejan Zhu, and Richard L. Gibson. 3-D transdimensional Markov chain Monte Carlo seismic inversion with uncertainty analysis. In *SEG Technical Program Expanded Abstracts 2017*, pages 607–611, 2017.

Jon F. Claerbout, Cecil Green, and Ida Green. *EARTH SOUNDINGS ANALYSIS: Processing versus Inversion*. Stanford University, 2004.

F. Collino and C. Tsogka. Application of the PML absorbing layer model to the linear elastodynamic problem in anisotropic heteregeneous media. *Geophysics*, 66(1):294–307, 2001. URL papers2://publication/uuid/8B8FC95D-3893-4A0C-8745-4584DAED0687.

Arturo Contreras, Carlos Torres-Verdín, and Tim Fasnacht. Sensitivity analysis of data-related factors controlling AVA simultaneous inversion of partially stacked seismic amplitude data: Application to deepwater hydrocarbon reservoirs in the central Gulf of Mexico. *Geophysics*, 72(1):C19, 2007. ISSN 00168033. doi: 10.1190/1.2399353.

Dennis Cooke and John Cant. Model-based Seismic Inversion : Comparing deterministic and probabilistic approaches. *CSEG Recorder*, 35(4):28–39, 2010.

Dennis Cooke and William A. Schneider. Generalized linear inversion of reflection seismic data. *Geophysics*, 48(6):665–676, 1983.

Knud Skou Cordua, Lars Nielsen, Majken C. Looms, Thomas Mejer Hansen, and Andrew Binley. Quantifying the influence of static-like errors in least-squares-based inversion and sequential simulation of cross-borehole ground penetrating radar data. *Journal of Applied Geophysics*, 68(1):71–84, 2009. ISSN 09269851. doi: 10.1016/j.jappgeo.2008.12.002.

Debanjan Datta and Mrinal K. Sen. Estimating a starting model for full-waveform inversion using a global optimization method. *Geophysics*, 81(4):R211–R223, 2016. ISSN 0016-8033. doi: 10.1190/geo2015-0339.1. URL http://library.seg.org/doi/10.1190/geo2015-0339.1.

Jan Dettmer and Stan E. Dosso. Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains. *Journal of the Acoustical Society of America*, 132(4):2239–50, 2012. ISSN 1520-8524. doi: 10.1121/1.4746016. URL http://www.ncbi.nlm.nih.gov/pubmed/23039421.

Stan E. Dosso and Charles W. Holland. Geoacoustic uncertainties from viscoelastic inversion of seabed reflection data. *IEEE Journal of Oceanic Engineering*, 31(3):657–671, 2006. ISSN 03649059. doi: 10.1109/JOE.2005.858358.

Jonathan E Downton. *Seismic parameter estimation from AVO inversion*. PhD thesis, University of Calgary, Calgary, Alberta., 2005. URL http://adsabs.harvard.edu/abs/2005PhDT.....87D.

Jonathan E Downton and Laurence R Lines. Three term AVO waveform inversion. In *CSEG National Convention*, pages 1–5, 2004. doi: 10.1190/1.1851205.

Jonathan E Downton and Charles P Ursenbach. Linearized amplitude variation with offset (AVO) inversion with supercritical angles. *Geophysics*, 71(5): E49–E55, 2006. ISSN 00168033. doi: 10.1190/1.2227617.

Jonathan E Downton, Brian H Russell, and Laurence R Lines. AVO for managers : pitfalls and solutions. *CREWES Research Report*, 12, 2000. URL http://nut.geo.ucalgary.ca/ForOurSponsors/ResearchReports/2000/2000-46.pdf.

Peter G. Doyle. *Grinstead and Snell's Introduction Introduction to Probability*. GNU Free Documentation License, 2006. ISBN 0821807498. doi: 10.1016/S1363-4127(97)81322-2. URL http://www.ncbi.nlm.nih.gov/pubmed/16758893.

Hong Feng and John C Bancroft. AVO principles , processing and inversion. *CREWES Research Report — Volume 18 (2006)*, 18:1–19, 2006. URL file:///c:/Users/Juan/Documents/INTEVEP/3_Proyectos/Tesistas/LeydiRamos/literature_repository/ArticulosAVO/AVO_Processing.pdf.

Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian Data Analysis*. CRC Press, 3 edition, 2014. ISBN 978-1-4398-9820-8. doi: 10.1007/s13398-014-0173-7.2.

Peter Gerstoft and Christoph F Mecklenbräuker. Ocean acoustic inversion with estimation of a posteriori probability distributions. *Journal of Acoustical Society America*, 104(2):808–819, 1998. ISSN 00014966. doi: 10.1121/1.423355.

Taras Gerya. *Introduction to Numerical Geodynamic Modelling*. Cambridge University Press, 2010. ISBN 9780521887540. doi: 10.1017/CBO9780511809101.

Geof H. Givens and Jennifer A. Hoeting. *Computational Statistics*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 10 2012. ISBN 9781118555552. doi: 10.1002/9781118555552. URL http://doi.wiley.com/10.1002/9781118555552.

Wences P. Gouveia and John A. Scales. Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis. *Journal of Geophysical Research*, 103(B2):2759–2779, 1998.

Dario Grana. Bayesian linearized rock-physics inversion. *Geophysics*, 81(6): D625–D641, 2016. ISSN 0016-8033. doi: 10.1190/geo2016-0161.1. URL http://library.seg.org/doi/10.1190/geo2016-0161.1.

Dario Grana, Torstein Fjeldstad, and Henning Omre. Bayesian Gaussian Mixture Linear Inversion for Geophysical Inverse Problems. *Mathematical Geosciences*, 49(4):1–23, 2017. ISSN 18748953. doi: 10.1007/s11004-016-9671-9.

Thomas Mejer Hansen, Knud Skou Cordua, Bo Holm Jacobsen, and Klaus Mosegaard. Accounting for imperfect forward modeling in geophysical inverse problems - Exemplified for crosshole tomography. *Geophysics*, 79(3):H1–H21, 2014. ISSN 00168033. doi: 10.1190/GEO2013-0215.1. URL http://library.seg.org/doi/abs/10.1190/geo2013-0215.1.

Thomas Mejer Hansen, Knud Skou Cordua, Andrea Zunino, and Klaus Mosegaard. Probabilistic Integration of Geo-Information. In Max Moorkamp, Peter G. Lelièvre, Niklas Linde, and Amir Khan, editors, *Integrated Imaging of the Earth: Theory and Applications*, chapter 6, pages 93–116. John Wiley & Sons, 1 edition, 2016. ISBN 9781118929063. doi: 10.1002/9781118929063.ch6.

Frank Heile, Robert Rapplean, David Jacobs, Craig Weinberg, and Alon Amit. What is a pattern?, 2017. URL https://www.quora.com/What-is-a-pattern.

C Hewitt Dix. Seismic velocities from surface mesurements. *Geophysics*, XX(1): 68–86, 1955.

Nicholas J Higham. Cholesky factorization. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(2):251–254, 2009. doi: 10.1002/wics.018.

Heiner Igel. *Computational Seismology*. Oxford University Press, 2017.

Richard A. Johnson and Dean W. Wichern. *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, 6th edition, 2007. ISBN 978-0-13-187715-3.

A. G. Journel. Combining knowledge from diverse sources: An alternative to traditional data independence hypotheses. *Mathematical Geology*, 34(5): 573–596, 2002. ISSN 08828121. doi: 10.1023/A:1016047012594.

A. G. Journel and Ch. J. Huijbregts. *Mining Geostatistcs*. Academic Press, Inc., London, 1 edition, 1989. ISBN 9780123910561.

Martin Jullum and Odd Kolbjørnsen. A Gaussian-based framework for local Bayesian inversion of geophysical data to rock properties. *Geophysics*, 81(3): R75–R87, 2016. ISSN 0016-8033. doi: 10.1190/geo2015-0314.1. URL http://library.seg.org/doi/10.1190/geo2015-0314.1.

Martina Kölbl-Ebert. Inge Lehmann ' s paper : " P '" (1936). *Episodes*, 24(4): 262–267, 2001.

Andrey Kolmogoroff. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Julius Springer, 1 edition, 1933.

Pijush Kundu, Ira M. Cohen, and David R. Dowling. *Fluid Mechanics*. Elsevier, 5 edition, 2012. ISBN 9780123821003.

Landmark. *ProMAX VSP User Training Manual*. Landmark Graphics Corporation, b edition, 1999.

Martin Landrø. The effect of noise generated by previous shots on seismic reflection data. *Geophysics*, 73(3):Q9–Q17, 2008. ISSN 0016-8033. doi: 10.1190/1.2894453. URL http://library.seg.org/doi/10.1190/1.2894453.

G. Le Loc'h and G.A. Galli. Truncated plurigaussian method: theoretical and practical points of view. *Geostatistics wollongong*, 96:211–222, 1997.

Alan R Levander. Fourth-order finite-difference P-W seismograms. *Geophysics*, 53(11), 1988.

Lewis Li, Jef Caers, and Paul Sava. Assessing seismic uncertainty via geostatistical velocity-model perturbation and image registration: An application to subsalt imaging. *The Leading Edge*, 34(9):1064–1070, 9 2015. ISSN 1070-485X. doi: 10.1190/tle34091064.1. URL http://library.seg.org/doi/10.1190/tle34091064.1.

Alberto Malinverno and V. A. Briggs. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics*, 69(4): 1005–1016, 2004. ISSN 00168033. doi: 10.1190/1.1778243.

Alberto Malinverno and Robert L. Parker. Two ways to quantify uncertainty in geophysical inverse problems. *Geophysics*, 71(3):W15–W27, 2006. ISSN 0016-8033. doi: 10.1190/1.2194516. URL http://library.seg.org/doi/10.1190/1.2194516.

Grégoire Mariethoz, Philippe Renard, Fabien Cornaton, and Olivier Jaquet. Truncated plurigaussian simulations to characterize aquifer heterogeneity. *Ground water*, 47(1):13–24, 2009. ISSN 1745-6584. doi: 10.1111/j.1745-6584.2008.00489.x. URL http://www.ncbi.nlm.nih.gov/pubmed/18793207.

Mark P. Mattson. Superior pattern processing is the essence of the evolved human brain. *Frontiers in Neuroscience*, 8(8 AUG):1–17, 2014. ISSN 1662453X. doi: 10.3389/fnins.2014.00265.

Gary Mavko, Tapan Mukerji, and Jack Dvorkin. *The Rock Physics Handbook*. Cambridge University Press, 2. edition, 2009. ISBN 9780521861366.

William Menke. *Geophysical Data Analysis: Discrete Inverse Theory MATLAB edition*. Academic Press, 3 edition, 2012. ISBN 978-0-12-397160-9.

R. D. Miller, S. E. Pullan, J. S. Waldner, and F. P. Haeni. Field comparison of shallow seismic sources. *Geophysics*, 51(11):2067–2092, 1986. ISSN <null>. doi: 10.1190/1.1442061.

Peter Moczo, Johan O.A. Robertsson, and Leo Eisner. The Finite-Difference Time-Domain Method for Modeling of Seismic Wave Propagation. *Advances in Geophysics*, 48(06):421–516, 2007. ISSN 00652687. doi: 10.1016/S0065-2687(06)48008-0.

Ingelise Møller, Bo Holm Jacobsen, and Niels Bøje Christensen. Rapid inversion of 2-D geoelectrical data by multichannel deconvolution. *Geophysics*, 66(3): 800–808, 2001. ISSN 00168033. doi: 10.1190/1.1444969.

Klaus Mosegaard and Thomas Mejer Hansen. Inverse Methods: Problem Formulation and Probabilistic Solutions. In *Integrated Imaging of the Earth: Theory and Applications*, pages 9–27. Wiley, 1 edition, 2016.

Klaus Mosegaard and Albert Tarantola. Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, 100(12):431–447, 1995. doi: 10.1029/94JB03097.

Klaus Mosegaard and Albert Tarantola. 16 Probabilistic approach to inverse problems. In William H. K. Lee, Hiroo Kanamori, Paul C. Jennings, and Carl Kisslinger, editors, *International Handbook of Earthquake and Engineering Seismology, Part A*, pages 237–265. Academic Press, 2002. ISBN 0124406521. doi: 10.1016/S0074-6142(02)80219-4. URL http://linkinghub.elsevier.com/retrieve/pii/S0074614202802194.

Pradip Kumar Mukhopadhyay and Subhashis Mallick. Exact angle-mute pattern for a transversely isotropic medium with vertical symmetry axis and its implication in offset-to-angle transform. *Geophysics*, 76(3):C41–C51, 2011. ISSN 10523812. doi: 10.1190/1.3628045. URL http://library.seg.org/doi/abs/10.1190/1.3628045.

David Mumford. Pattern Theory: A Unifying Perspective. In *Fields Medallists' Lectures*, pages 226–261. World Scientific, 1997.

P. Newman. Divergence effects in a layered earth. *Geophysics*, 38(3):481–488, 1973.

Jorge Nocedal and Stpehen J. Wright. *Numerical Optimization*. Springer, 1 edition, 1999. ISBN 0387987932.

W. J. Ostrander. Method for interpretation of seismic records to yeild indication of gaseous hydrocarbons, 1982.

Robert L. Parker. *Geophysical Inverse Theory*. Princeton University Press, 1 edition, 1994. ISBN 0-691-03634-9.

Tor Erik Rabben, Håkon Tjelmeland, and Bjørn Ursin. Non-linear Bayesian joint inversion of seismic reflection coefficients. *Geophysical Journal International*, 173(1):265–280, 2008. ISSN 0956540X. doi: 10.1111/j.1365-246X.2007.03710.x.

Anandaroop Ray, Kerry Key, and Thomas Bodin. Hierarchical Bayesian inversion of marine CSEM data over the Scarborough gas field — A lesson in correlated noise. In *SEG Technical Program Expanded Abstracts*, pages 723–727, 2013. doi: 10.1190/segam2013-1439.1. URL http://library.seg.org/doi/abs/10.1190/segam2013-1439.1.

Anandaroop Ray, Anusha Sekar, G. Michael Hoversten, and Uwe Albertin. Frequency domain full waveform elastic inversion of marine seismic data from the Alba field using a Bayesian trans-dimensional algorithm. *Geophysical Journal International*, 205(2):915–937, 2016. ISSN 1365246X. doi: 10.1093/gji/ggw061.

Michael Riedel, Stan E. Dosso, and Laurens Beran. Uncertainty estimation for amplitude variation with offset (AVO) inversion. *Geophysics*, 68(5):

1485–1496, 9 2003. ISSN 0016-8033. doi: 10.1190/1.1620621. URL http://library.seg.org/doi/10.1190/1.1620621.

Kjartan Rimstad and Henning Omre. Generalized Gaussian Random Fields using hidden selections. *arXiv preprint arXiv:1402.1144*, 2014a.

Kjartan Rimstad and Henning Omre. Skew-Gaussian random fields. *Spatial Statistics*, 10(7491):43–62, 2014b. ISSN 22116753. doi: 10.1016/j.spasta.2014.08.001.

Yousef Saad. *Iterative Methods for Sparse Iterative Methods*. Society for Industrial and Applied Mathematics, 2 edition, 2003.

Hamid Sabeti, Ali Moradzadeh, Faramarz Doulati Ardejani, Leonardo Azevedo, Amilcar Soares, Pedro Pereira, and Ruben Nunes. Geostatistical seismic inversion for non-stationary patterns using direct sequential simulation and co-simulation. *Geophysical Prospecting*, pages 1–24, 2017. ISSN 13652478. doi: 10.1111/1365-2478.12502.

Malcolm Sambridge, Kerry Gallagher, A. Jackson, and P. Rickwood. Trans-dimensional inverse problems, model comparison and the evidence. *Geophysical Journal International*, 167(2):528–542, 2006. ISSN 0956540X. doi: 10.1111/j.1365-246X.2006.03155.x.

John A. Scales and Roel Snieder. To Bayes or not to Bayes …. *Geophysics*, 62(4): 1045–1046, 1997.

John A. Scales and Roel Snieder. What is noise? *Geophysics*, 63(4):1122–1124, 1998.

J. F. Scheimer and I. Y. Borg. Deep Seismic Sounding with Nuclear Explosives in the Soviet Union. *Science*, 226(4676):787–792, 1984.

Mrinal K. Sen and Paul L. Stoffa. Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. *Geophysical Prospecting*, 44 (2):313–350, 1996. ISSN 00168025. doi: 10.1111/j.1365-2478.1996.tb00152.x.

Peter M. Shearer. *Introduction to Seismology*. Cambridge University Press, 2 edition, 2009.

R.T. Shuey. A simplification of the Zoeppritz equations. *Geophysics*, 50(4): 609–614, 1985.

Scott Singleton. The Effects of Seismic Data Conditioning on Prestack Simultaneous Impedance Inversion. *The Leading Edge*, 28(7):260–267, 2009. ISSN 1070485X. doi: 10.1190/1.3167776. URL http://library.seg.org/doi/10.1190/tle34091064.1.

R. H. Stolt and A. B. Weglein. Migration and inversion of seismic data. *Geophysics*, 50(12):2458–2472, 1985.

Albert Tarantola. A strategy for nonlinear elastic inversion of seismic reflection data, 1986. ISSN 1070485X. URL http://library.seg.org/doi/pdf/10.1190/1.1442046.

Albert Tarantola. *Inverse problem theory and Methods for Model Parameter Estimation*. SIAM, 1 edition, 2005. ISBN 0898715725. doi: 10.1137/1.9780898717921. URL http://www.ipgp.fr/~tarantola/Files/Professional/Teaching/Diverse/Complements_2/Lessons/Lessons_1-3.pdf.

Albert Tarantola and Bernard Valette. Inverse Problems = Quest for Information. *Journal of Geophysics*, 50(3):159–170, 1982.

Pierre Thore. Uncertainty in seismic inversion: What really matters? *The Leading Edge*, 34(9):1000–1004, 2015. ISSN 1070-485X, 1938-3789. doi: 10.1190/tle34091000.1. URL http://library.seg.org/doi/10.1190/tle34091000.1.

A.N. Tikhonov. On the solution of ill-posed problems and the method of regularization. *Matematicheskii Sbornik (in Russian)*, 151(3):501–504, 1963. doi: 10.18287/0134-2452-2015-39-4-459-461.

Tadeusz J. Ulrych and Mauricio D. Sacchi. *Information-based inversion and processing with applications*. Elsevier, 2005. ISBN 9780080447216. doi: 10.1016/S0950-1401(13)70016-8. URL http://linkinghub.elsevier.com/retrieve/pii/S0950140113700168.

Tadeusz J. Ulrych, Mauricio D. Sacchi, and Alan Woodbury. A Bayes tour of inversion : A tutorial. *Geophysics*, 66(1):55–69, 2001.

Jean Virieux. P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, 51(4):889, 1986. ISSN 1070485X. doi: 10.1190/1.1442147.

Jean Virieux and Stéphane Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009. ISSN 00168033. doi: 10.1190/1.3238367.

Jean Virieux, A. Asnaashari, Romain Brossier, L. Métivier, L. Isterre, and W. Zhou. An introduction to full waveform inversion. In *Encyclopedia of Exploration Geophysic;*, chapter 6, pages 1–1. Society of Exploration Geophysicists, 2017. ISBN 9781560803027. doi: 10.1190/1.9781560803027.entry6.

David. Y Wang and Douglas W. McCowan. Spherical divergence correction for seismic reflection data using slant stacks. *Geophysics*, 54(5):563–569, 1988. ISSN 0016-8033. doi: 10.1190/1.1442683.

Yanghua Wang. Approximations to the Zoeppritz equations and their use in AVO analysis. *Geophysics*, 64(6):1920–1927, 1999.

Wiktor Waldemar Weibull. *Seismic Processing and Velocity Analysis Using Reverse-Time Migration*. PhD thesis, NTNU: Norwegian University of Science and Technology, 2013.

Nathan W. Winslow. *Joint Inversion using the convolutional model*. PhD thesis, Rice University, 2000.

Yong Xu and Satinder Chopra. Improving AVO fidelity by NMO stretching and offset-dependent tuning corrections. *CSEG Recorder*, 32(9):30–34, 2007. ISSN 1070-485X. doi: 10.1190/1.2821941. URL http://library.seg.org/doi/10.1190/1.2821941.

Ozdogan Yilmaz. *Seismic Data Analysis*, volume 2. Society of Exploration Geophysicists, 1 2001. ISBN 978-1-56080-094-1. doi: 10.1190/1.9781560801580. URL http://library.seg.org/doi/book/10.1190/1.9781560801580.

Hongbo Zhang and R James Brown. A review of AVO analysis. *CREWES Research Report*, 13(1):357–380, 2001.

K. Zoeppritz. Über Reflexion und Durchgang seismischer Wellen durch Unstetigkeitsflächen. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pages 66–84, 1919.

Andrea Zunino, Klaus Mosegaard, Katrine Lange, Yulia Melnikova, and Thomas Mejer Hansen. Monte Carlo reservoir analysis combining seismic reflection data and informed priors. *Geophysics*, 80(1):R31–R41, 2015. ISSN 0016-8033. doi: 10.1190/geo2014-0052.1. URL http://library.seg.org/doi/10.1190/geo2014-0052.1.

Appendices

A

Snell's Law

Snells' law describes the refraction and reflection of an incoming wave (the incident wave) at an interface between two continuous media with different acoustic, and hence elastic properties. Consider two media with different elastic properties (v_p, v_s, and ρ) as illustrated in Figure A.1. The incoming pressure wave is split into four components when reaching the interface between the two media. Two waves are reflected back into medium 1. The first one remains a pressure wave and is denoted R_{pp}. The second reflected wave is transformed into a shear wave denoted R_{ps}. Similarly as for the two reflected waves, the incident wave is refracted into medium 2 in two different wave phases. We describe these as T_{pp}

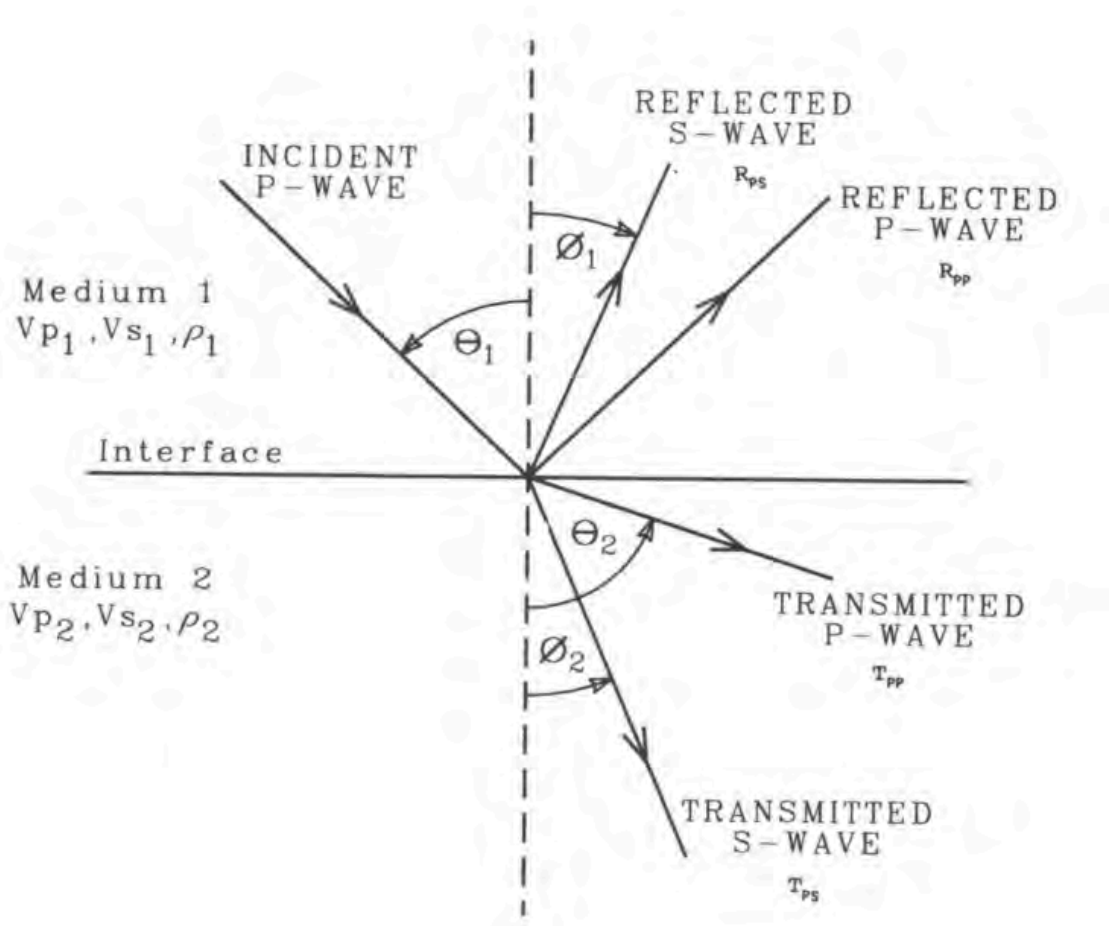


Figure A.1: Reflection and refraction of incoming pressure wave at an interface. Snells' law describes the relationship between incoming (indicent pressure wave) and outgoing wave-phases. Illustration is taken from Yilmaz (2001).

and T_{ps} for the pressure and shear wave components respectively. Snell's law simply states, that the ratio between sin() of the outgoing wave and the velocity in the transporting medium is constant:

$$p = \frac{\sin(\Theta_1)}{v_{p1}} = \frac{\sin(\Theta_2)}{v_{p2}} = \frac{\sin(\varphi_1)}{v_{s1}} = \frac{\sin(\varphi_2)}{v_{s2}} \tag{A.1}$$

The constant p is often referred to as the ray-parameter in the literature (see e.g. Zhang and Brown (2001); Shearer (2009)). In common words, the first term $p = \frac{\sin(\Theta_1)}{v_{p1}}$ states that the incoming angle is equal to the outgoing angle for the pressure wave reflection. This should analogously be familiar to many pool players. The general assumption in Snell's law is that no energy is lost at the interface. In reality, there is a small dissipation of energy at the interface due to friction of the molecules in the media. However, on a larger scale, this energy loss is deemed negligible and Snell's law is assumed valid.

B

The Gaussian Distribution

The multivariate Gaussian probability density function $p(\mathbf{r})$ for a variable $\mathbf{r}$ with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$ is given by (Johnson and Wichern, 2007):

$$p(\mathbf{r}) = (2\pi)^{-n/2}|\mathbf{C}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{C}^{-1}(\mathbf{r}-\boldsymbol{\mu})\right] \tag{B.1}$$

where n is the dimension of the vector $\mathbf{r}$ and $|\mathbf{C}|$ denotes the determinant of the covariance matrix $\mathbf{C}$. A compact notation form is $\mathbf{r} \sim \mathcal{N}_n(\boldsymbol{\mu}, \mathbf{C})$.

C

Momentum Equation

The following derivation of the momentum equation is mainly based on Shearer (2009). The momentum equation (or equation of motion) follows from applying Newton's second law of motion on a continuous medium. The law simply states that any force F on a body is proportional to the mass m and acceleration a of the body:

$$F = ma \tag{C.1}$$

The law thereby offers a relationship between forces and motion. The total forces F from a stress field can be subdivided into forces acting on the surfaces F_i and body forces F_i^{body} acting on the whole body (e.g. gravity or electromagnetic

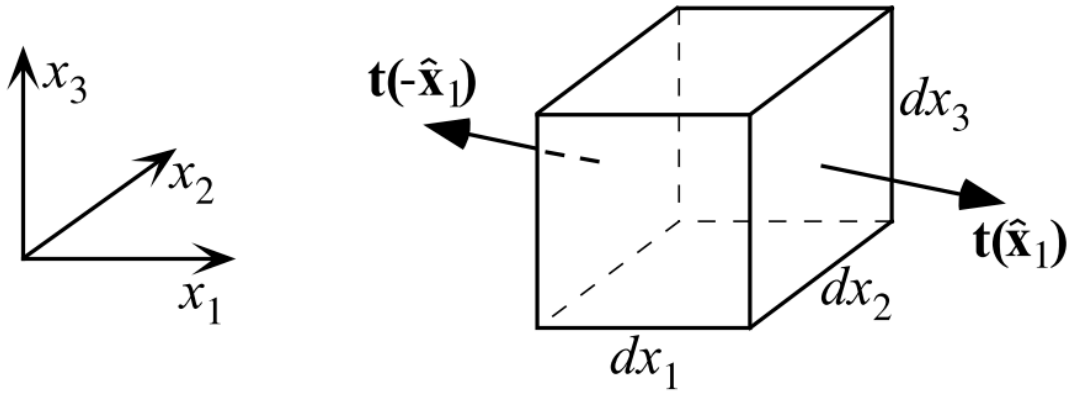


Figure C.1: Surface forces on an infinitesimal cube $dx_1 dx_2 dx_3$. Surface forces acting on the $dx_2 dx_3$ planes in regards to the $\hat{\mathbf{x}}_1$ direction. The force is given as the traction $\mathbf{t}(\hat{\mathbf{x}}_1)$ over the surface area $dx_2 dx_3$. Illustration is taken from Shearer (2009).

forces):

$$F = F_i + F_i^{body} \tag{C.2}$$

Let us consider an infinitesimal body with volume $dx_1 dx_2 dx_3$ as illustrated in Figure C.1. The force acting on any surface can be described as the product between the perpendicular traction vector $\mathbf{t}$ and the surface area (Kundu et al., 2012), because the traction is force per unit area. The force $\mathbf{F}(\hat{\mathbf{x}}_1)$ acting on the surface $dx_2 dx_3$ can e.g. be calculated as:

$$\mathbf{F}(\hat{\mathbf{x}}_1) = \mathbf{t}(\hat{\mathbf{x}}_1) dx_2 dx_3 = \boldsymbol{\tau} \hat{\mathbf{x}}_1 dx_2 dx_3 = \begin{bmatrix} \tau_{11} \\ \tau_{21} \\ \tau_{31} \end{bmatrix} dx_2 dx_3 \tag{C.3}$$

The last two equations follow from the linear stress tensor operator $\boldsymbol{\tau}$ defined in Equation D.2. As noted in Shearer (2009), a net force is only exerted if a spatial

gradient exists in the stress field. The traction at each end side of the cube is then different and hence $\mathbf{F}(\hat{\mathbf{x}}_1) \neq \mathbf{F}(-\hat{\mathbf{x}}_1)$. The traction vector $\mathbf{t}(\hat{\mathbf{x}}_1)$ can be written as the change in the stress field:

$$\mathbf{t}(\hat{\mathbf{x}}_1) = \frac{\partial}{\partial x_1} \begin{bmatrix} \tau_{11} \\ \tau_{21} \\ \tau_{31} \end{bmatrix} dx_1 = \frac{\partial}{\partial x_1} \hat{\boldsymbol{\tau}}_1 dx_1 \tag{C.4}$$

The surface force $\mathbf{F}(\hat{\mathbf{x}}_1)$ can then be written as:

$$\mathbf{F}(\hat{\mathbf{x}}_1) = \frac{\partial}{\partial x_1} \hat{\boldsymbol{\tau}}_1 dx_1 dx_2 dx_3 = \frac{\partial \tau_{i1}}{\partial x_1} dx_1 dx_2 dx_3 \tag{C.5}$$

where τ_{i1} is index notation. The forces acting on the other sides of the infinitesimal cube follow analogously from the above. The total sum of forces acting on the surface of the cube F_i is then given by:

$$F_i = \mathbf{F}(\hat{\mathbf{x}}_1) + \mathbf{F}(\hat{\mathbf{x}}_2) + \mathbf{F}(\hat{\mathbf{x}}_3) \tag{C.6}$$

$$= \left(\frac{\partial \tau_{i1}}{\partial x_1} + \frac{\partial \tau_{i2}}{\partial x_2} + \frac{\partial \tau_{i3}}{\partial x_3} \right) dx_1 dx_2 dx_3 \tag{C.7}$$

$$= \sum_{j=1}^{3} \frac{\partial \tau_{ij}}{\partial x_j} dx_1 dx_2 dx_3 \tag{C.8}$$

$$= \partial_j \tau_{ij} dx_1 dx_2 dx_3 \tag{C.9}$$

where $\partial_j \tau_{ij}$ implies summation over all surfaces. The sum of body forces F_i^{body} acting on the infinitesimal cube is proportional to the volume of the cube:

$$F_i^{body} = f_i dx_1 dx_2 dx_3 \tag{C.10}$$

where f_i is the body force. Finally, the mass m of the cube can more generally be written as the density ρ times the volume. The acceleration a of the cube can also be written as the second time derivative of the displacement u_i. Substituting all the above into Equation C.1 yields the momentum equation:

$$\rho dx_1 dx_2 dx_3 \frac{\partial^2 v_i}{\partial t^2} = \partial_j \tau_{ij} dx_1 dx_2 dx_3 + f_i dx_1 dx_2 dx_3 \tag{C.11}$$

Letting the common term $dx_1 dx_2 dx_3$ cancel out, the momentum equation for a continuous medium of any size is given by:

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \partial_j \tau_{ij} + f_i \tag{C.12}$$

A comprehensive derivation of the equation of motion can be found in e.g. Aki and Richards (2002). The momentum equation can be split into solutions for the P- and S-waves respectively (Shearer, 2009) by taking the divergence $(\nabla \cdot)$ and curl $(\nabla \times)$ of the equation, where ∇ is the vector differential operator:

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \tag{C.13}$$

D

The Stress Tensor

The following is based on Shearer (2009). A traction vector $\mathbf{t}(\hat{\mathbf{n}}) = [t_1, t_2, t_3]$ describes the force per unit area exerted on a surface of a continuous medium in direction 1,2, and 3. The traction vector is illustrated in Figure D.1 for a small infinitesimal plate. The direction of the traction is referring to a unit vector $\hat{\mathbf{n}} = [\hat{n}_1, \hat{n}_2, \hat{n}_3]$ pointing orthogonally to the plane. The traction acting ortogonally to the plane (along $\hat{\mathbf{n}}$) is known as normal stress, whereas traction acting along the plane is known as shear stress. In order to describe the traction $\mathbf{t}$ as a function of $\hat{\mathbf{n}}$ we need the introduction of a stress matrix $\boldsymbol{\tau}$ describing all the stresses acting on the plane. The stress tensor defines a linear mapping of the

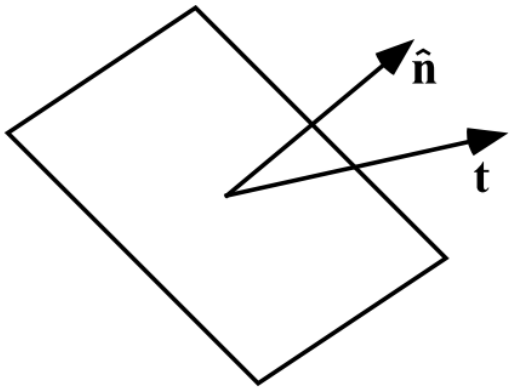


Figure D.1: Traction vector $\mathbf{t}$ on a infinitesimal plane. The orientation of the plate is described by the normal vector $\hat{\mathbf{n}}$. Illustration is taken from Shearer (2009).

traction to normal vectors.

$$\mathbf{t}(\hat{\mathbf{n}}) = \boldsymbol{\tau}\hat{\mathbf{n}} \tag{D.1}$$

The stress tensor consists of nine stress components τ_{ij} that define the full stress state in a continuous medium (Shearer, 2009; Kundu et al., 2012):

$$\boldsymbol{\tau} = \tau_{ij} = \begin{Bmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{12} & \tau_{13} \\ \tau_{31} & \tau_{12} & \tau_{13} \end{Bmatrix} \tag{D.2}$$

where the indices represent the three dimensions. A traction across any plane with a certain orientation $(\hat{\mathbf{n}})$ is then given by combining Equation D.1 and Equation D.2:

$$\mathbf{t}(\hat{\mathbf{n}}) = \begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{12} & \tau_{13} \\ \tau_{31} & \tau_{12} & \tau_{13} \end{bmatrix} \hat{\mathbf{n}} = \begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{12} & \tau_{13} \\ \tau_{31} & \tau_{12} & \tau_{13} \end{bmatrix} \begin{bmatrix} \hat{n}_1 \\ \hat{n}_2 \\ \hat{n}_3 \end{bmatrix} \tag{D.3}$$

E

Finite Difference Method

The finite difference method (FDM) is a method for approximating partial derivatives and is considered a fundamental method in numerical calculus. Many physical phenomenas are connected with *gradients*, e.g. the heat equation, momentum equation, continuity equation, etc. The method therefore has many applications in physics. It is postulated as the simplest and typical way of approximating partial differential operators (Saad, 2003).

Beacuse infinite limits (e.g. $\delta x \to 0$) are not allowed in numerical calculus, the FDM uses $\Delta x > \varepsilon$, where the ε expresses a round-off that is associated with the machine precision and Δx is the finite difference. In other words, the numerical solution may never reach the infinite limit but will only reach an infinitesimal

small number depending on the operating machine. The FDM introduces two types of errors which should be accounted for during usage:

- The first error is the above-mentioned round-off that introduces loss off precision due to computers rounding of decimal quantities. The error is relatively small but can easily increase through accumulation due to reuse of round-off numbers through extended modeling.
- The second error is known as the truncation error or discretization error. The FDM is based upon discretizing a function into small parts to be evaluated. The parts are separated by node-points, which can be seen as the $(x_0, x_1, x_2, \ldots)$ in figure E.1. The error is introduced through the Taylor series that gives the value for the next step between two node-points on the grid:

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + \frac{f''(x)}{2!}\Delta x^2 + \frac{f'''(x)}{3!}\Delta x^3 + \ldots + R_n \quad \text{(E.1)}$$

  where R_n is the truncation error. The size of this error is therefore dependent on the level at which the Taylor series is truncated.

The level of accuracy should be counter weighted by the need for the model to be simple, i.e. it is of course possible to create a model that only produces an insignificant error by including a lot of terms, but that would also mean that the model would become more complex to evaluate. The level of truncation should also be considered with regards to the type of differential equation one intend to solve. For instance, a third-order differential equation would need the inclusion of the first four terms (i.e. the third order) of Equation E.1 to account for the

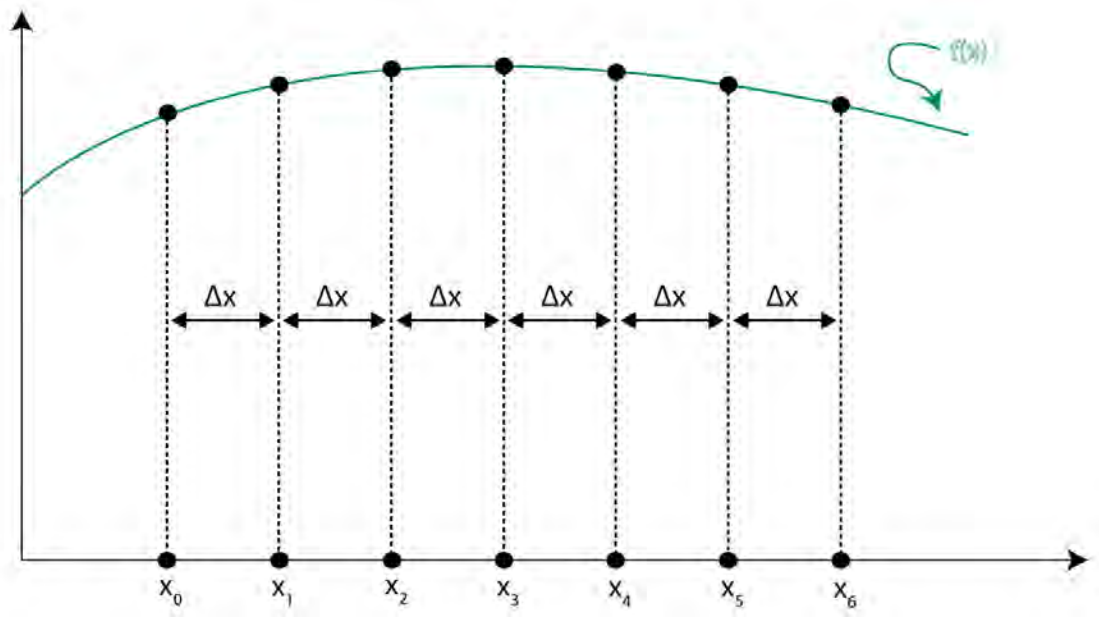


Figure E.1: The figure illustrates the way the FDM makes use of discretizing a function $f(x)$ into a grid of equal spacing.

third-order differential. To solve a simple parabolic differential equation the use of the first three terms (i.e. the second order) of the Taylor series (Equation E.1) proves to be a stable solution:

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + \frac{f''(x)}{2!}\Delta x^2 + R_n \tag{E.2}$$

By the use of the central difference, the value of $f'(x)$ can be evaluated to relatively high accuracy:

$$f(x + \Delta x) - f(x - \Delta x) \simeq 2f'(x)\Delta x \tag{E.3}$$

$$\Rightarrow f'(x) \simeq \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x} \tag{E.4}$$

If the finite difference Δx is sufficiently small, the first derivative can be directly

approximated by a truncation after the first two terms in Equation E.1 (the first order Taylor series):

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + R_n \tag{E.5}$$

$$\Rightarrow f'(x) \simeq \frac{f(x + \Delta x) - f(x)}{\Delta x} \tag{E.6}$$

In summary, there are four logical steps when applying finite differences (from Gerya (2010)):

1. The infinite number of geometrical points in a continuum needs to be replaced by a finite amount of grid nodes.
2. At these nodes, physical properties must be defined.
3. Partial differential equations at these nodes are substituted with discretized finite difference linear equations. These linear equations should make a relation between the physical properties
4. By solving the resulting set of linear equations, unknowns for the node-points can be obtained.

F

Normal Moveout

A schematic representation of the hyperbolic trajectory of arrival time t as a function of offset for a wave traveling between source and receiver can be seen in Figure F.1. Consider the ray passing through the subsurface from the source and being reflected back to a receiver at the surface. From Pythogaras' theorem follows:

$$\begin{aligned} d^2 &= h^2 + (x/2)^2 \\ 4d^2 &= 4h^2 + x^2 \end{aligned} \tag{F.1}$$

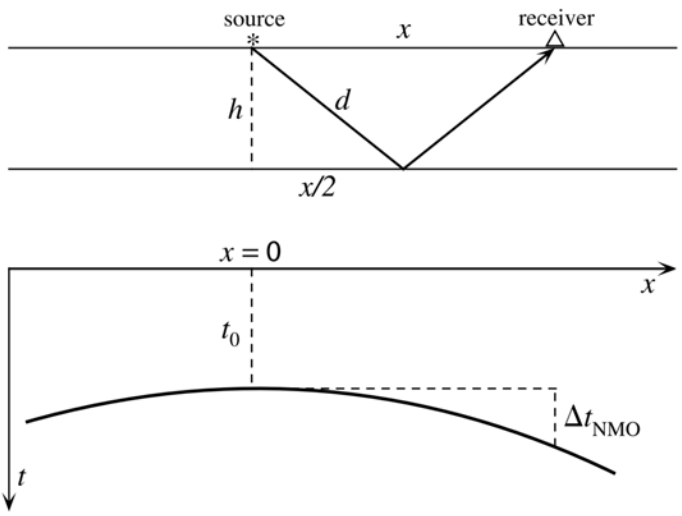


Figure F.1: Top: A reflected ray path, Bottom: Corresponding travel time curve as a function of offset. Figure taken from Shearer (2009).

If a constant velocity is assumed within the layer, then $2d = vt$, where v is the velocity in the layer. Similarly for a receiver at zero-offset, $2h = vt_0$ where t_0 is the two-way travel time (TWT). Substituting this into Equation F.1 yields:

$$v^2t^2 = t_0^2 + x^2 \tag{F.2}$$

$$\Rightarrow t(x) = \sqrt{t_0^2 + \frac{x^2}{v^2}} \tag{F.3}$$

Using the relation in Equation F.3 offers a way of computing corrected travel times from the two-way-travel time and offset under the assumption of a known velocity field.

G

Hierarchical Models

In statistics, a classical stochastic model contains observable variables which are conditioned on some conditional parameters. Usually these parameters are considered non-probabilistic. Alternatively, the parameters can be treated with a probabilistic specification. They will therefore depend on another set of parameters. In principle, these parameters could also be specified probabilistically and thus depend on yet another set of parameters. The additional parameters are typically known as hyper parameters to discern them from the original parameters. The hyper parameters follow a distribution, which is similarly known as a hyper distribution. The additional hyper parameters and distributions form a hierarchy of dependency between parameters in the

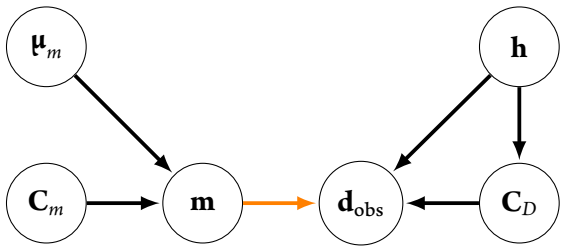


Figure G.1: The stochastic model as a directed acyclic graph. The nodes represent stochastic variables and the black arrows show probability dependencies. The orange arrow between model parameters $\mathbf{m}$ and observed data $\mathbf{d}_{\text{obs}}$ indicates the deterministic relationship of the forward problem.

stochastic model. Stochastic models of this kind are therefore known as hierarchical models. Hierarchical models have increased in popularity in recent years for understanding multi-parameter problems (Gelman et al., 2014).

As a simple example, consider the model parameters $\mathbf{m}$. Let $\mathbf{m}$ be Gaussian with mean $\boldsymbol{\mu}_m$ and covariance $\mathbf{C}_m$. The model parameters have a deterministic relationship with observed data $\mathbf{d}_{\text{obs}}$ through a forward model. Let the data $\mathbf{d}_{\text{obs}}$ be Gaussian as well, with covariance $\mathbf{C}_D$. Say, one is unsure of the appearance of the noise covariance $\mathbf{C}_D$. For instance, we might not know the exact variance of the noise σ_d^2 or exact range in the correlation function, but might instead be certain that it is not zero and it will not exceed some extreme threshold value. We can then replace the variance or range parameter in the correlation function with hyper parameters $\mathbf{h}$. The noise model and thereby the data, are then probabilistically dependent on the hyper parameters $\mathbf{h}$. The hyper distribution of the hyper parameters could then be modeled with an uniform distribution with our threshold values. The proposed stochastic dependencies in the model are visualized in the directed acyclic graph in Figure G.1.

H

Scientific work

H.1 Seismic forward modeling errors from linear approximations to Zoeppritz equations

Rasmus Bødker Madsen & Thomas Mejer Hansen, submitted to SEG annual meeting 2016

Seismic forward modeling errorsfrom linear approximations to Zoeppritz equations

Rasmus Bødker Madsen and Thomas Mejer Hansen, University of Copenhagen*

SUMMARY

A linearised form of Zoeppritz equations combined with the convolution model is widely used in inversion of amplitude versus offset (AVO) seismic data. This introduces a 'modeling error' related to the forward model compared to the full Zoeppritz equations. Here a methodology for quantifying the modeling error caused by using such an approximation in AVO modeling is demonstrated. First, a sample of the error related to using a linearised approximation to the Zoeppritz equations is generated. It is demonstrated that the modeling error depends on the degree of subsurface variability as described by the prior assumptions. Then, it is illustrated how this sample can be described by a correlated Gaussian probability density. Finally, we demonstrate how the modeling error can be accounted for in linearised AVO inversion, and in this way avoid biases in the inversion results.

INTRODUCTION

Reflection seismic AVO data is sensitive to relative changes in the acoustic impedance and the ratio between p-wave and s-wave velocities v_p/v_s, which can be heavily impacted by the e.g. water saturation. Therefore, inversion of reflection seismic AVO data is widely used in petroleum exploration, as variations in v_p/v_s ratio is potentially resolvable. Any inverse approach is based on some choice of how to solve the forward problem, i.e. for the present problem, how to compute the AVO data set related to a specific choice of subsurface model of elastic properties. Typically the parametrisation is some variation of a triplet of elastic parameters. For instance p-wave velocity v_p, s-wave velocity v_s and density ρ or acoustic impedance (AI), v_p/v_s-ratio and density.

In this study case the forward problem amounts to computing the AVO response from a specific elastic model. The best solution to the forward problem would be to use some form of full waveform solution, which in principle should provide the correct forward response to the subsurface model. Even though methods exist to perform full waveform inversion, the computational costs restrict the application for small problems at lower resolutions (Virieux and Operto, 2009). In practice, a widely used forward model is a combination of a linearised approximation to Zoeppritz equations combined with the convolution model.

Zoeppritz equations are used to describe the partitioning of seismic wave energy at different angles of incident at a single interface between two layers with different elastic properties (Aki and Richards, 2002). Zoeppritz equations lay the foundation for AVO analysis.

For practical purposes however, some form of linear approximation is usually applied to the full Zoeppritz equations. The success of the simpler approximations can be attributed to their usefulness in practical applications as well as their ease of evaluation and interpretion (Castagna and Backus, 1994; Mavko et al., 2009). Among the more popular ones are Aki and Richards equation (Aki and Richards, 1980) and a further simplification done by Shuey (1985). A detailed review of these can for instance be found in (Castagna and Backus, 1994). Buland and Omre (2003) introduced a Bayesian linearised AVO inversion technique by adapting an extension to the Aki and Richards approximation originally proposed by Stolt and Weglein (1985). This technique renders the AVO inverse problem into a linear inverse Gaussian problem, for which an analytical solution, in form of a Gaussian model, can be found.

Modeling error

In reality any specific choice of forward modeling, leads to a 'modeling error' (Tarantola and Valette, 1982; Hansen et al., 2014). Modeling errors related to AVO data can stem from a variety of sources such as using a 1D forward code to reflect a 3D physical system; the use of the acoustic wave equation as opposed to the anisotropic visco-elastic wave-equation; imperfections in data processing; uncertain wavelet estimates. See also Thore (2015).

Modeling errors are also connected to the use of approximations to Zoeppritz equations. Approximations to Zoeppritz equations are, in general, associated with the least amount of modeling error for smaller incidence angles. As a general rule of thumb, most approximations of the Zoeppritz equations are assumed valid (implying insignificant modeling errors) for incidence angles less than around 30° (Shuey, 1985; Castagna and Backus, 1993; Mavko et al., 2009). However, if the variations in the elastic properties are very smooth, implying small contrasts, the approximation may be valid for incidence angles above 30°. This relationship is confirmed when comparing the reflection coefficient for different angles at a single interface. Using the 4 AVO classes proposed by Castagna and Backus (1994), Haase (2004) showed that not only does the error depend on the contrast between the two layers at the interface, but also the type of lithology at the boundary.

However, to our knowledge, little has been published quantifying the modeling error combined with the convolution model for a time-series of elastic properties. Buland and Omre (2003) did a qualitative measure on the modeling error related to a smooth Gaussian distribution of elastic parameters, by visually comparing the results of using the Zoeppritz equations and their own approximation. They concluded that no significant modeling error arises from the forward approximation using such a Gaussian type a priori model. In general though, little work has been done to quantify modeling errors related to the use of approximations to Zoeppritz equations combined with the convolution model.

Recently a number of studies have emphasized the need to quantify and account for modeling errors in general in geophysical inverse problems (Hansen et al., 2014; Thore, 2015;

Valentine and Trampert, 2015). Here the approach developed by Hansen et al. (2014) will be adapted to simulate, model and account for modeling errors caused by the use of approximations to the full Zoeppritz equations combined with convolution model.

THEORY

The Zoeppritz equations (Zoeppritz, 1919) provide the full angular reflectivity at an interface between two media with different elastic properties for a plane interface wave.

Aki and Richards

The 'classic' approximation proposed by Aki and Richards (1980), stems from a parametrisation of p-wave and s-wave velocities and the density (v_p, v_s and ρ). The Aki and Richards approximation make use of the averages of these elastic parameters over the interface ($\overline{v_p}$, $\overline{v_s}$ and $\overline{\rho}$) as well as the contrasts Δv_p, Δv_p and $\Delta\rho$. For a small percentile change in elastic properties the reflectivity can be approximated by

$$R(\theta) \approx a_{v_p}(\theta)\frac{\Delta v_p}{\overline{v_p}} + a_{v_s}(\theta)\frac{\Delta v_s}{\overline{v_s}} + a_\rho(\theta)\frac{\Delta\rho}{\overline{\rho}} \tag{1}$$

where the coefficients are given by

$$a_{v_p}(\theta) = \frac{1}{2\cos^2\theta} \tag{2}$$

$$a_{v_s}(\theta) = -\frac{4\overline{v_s}^2}{\overline{v_p}^2}\sin^2\theta \tag{3}$$

$$a_\rho(\theta) = \frac{1}{2}(1 - 4\frac{\overline{v_s}^2}{\overline{v_p}^2}\sin^2\theta). \tag{4}$$

The full derivation can be seen in e.g. Aki and Richards (2002); Mavko et al. (2009). A popular way of rearranging the terms in the Aki and Richards approximation, into three major terms, was performed by Shuey (1985). These terms are the AVO intercept, gradient and a third order correction term. Here we will focus on the further linear approximation done by Buland and Omre.

Buland and Omre

Buland and Omre (2003) adapted the following forward approximation by Stolt and Weglein, that expands the Aki and Richards approximation, where the reflection now also is time-dependant

$$\begin{aligned} R(t,\theta) \approx\ & a_{v_p}(t,\theta)\frac{\partial}{\partial t}\ln v_p(t) \\ & + a_{v_s}(t,\theta)\frac{\partial}{\partial t}\ln v_s(t) + a_\rho(t,\theta)\frac{\partial}{\partial t}\ln v_\rho(t) \end{aligned} \tag{5}$$

The time-dependency in the coefficients $a_{v_p}(t,\theta)$, $a_{v_s}(t,\theta)$ and $a_\rho(t,\theta)$ is simply a generalisation of the Aki & Richards coefficients (eq. 2-4) with the averages $\overline{v_p}$ and $\overline{\rho}$ being time-dependent. In order to solve this equation, it is here assumed that these averages can be described by a known background-model. In practice this can be achieved using the mean value of all v_p and ρ values as the averages $\overline{v_p}$ and $\overline{\rho}$. By making the material parameters follow a log-gaussian distribution, they can be restricted from attaining negative values (Buland and Omre, 2003). Despite this further requirement of a known background-model, the Buland and Omre approximation offers one great advantage. It provides an analytical solution to the forward problem and provides a way of doing Bayesian linear inversion of seismic AVO data.

Both approximations are based on the common assumption of small-contrasts in the elastic properties at all interfaces.

Convolutional Model

A seismic trace can be obtained using the convolution model:

$$S(t) = W(t) * R(t) \equiv \int_0^{t_s} W(\tau)R(t-\tau)\,\mathrm{d}\tau \tag{6}$$

where $S(t)$ is the seismic trace, $R(t)$ is the reflectivity series (earth response), $W(t)$ is the wavelet (source-time function), and t_s is the duration of the source input. The convolutional model requires the wavelet to be known prior to convolution. This is often not the case and the wavelet is estimated from nearby wells, i.e. an *effective wavelet* is used. The convolutional model is discussed in more detail by Yilmaz and Doherty (2000). In this study a Ricker wavelet with decreasing center frequency from 50Hz at zero-offset to 25Hz at the highest incident angle is used.

Calculating modeling error

For one specific 1D model of elastic parameters the modeling error, $S_{error}(t)$, can simply be calculated by taking the difference in the seismic response from a forward approximation and the full Zoeppritz equations for all traces in a seismic AVO gather:

$$S_{error}(t) = S_{zoep}(t) - S_{approx}(t) \tag{7}$$

$$= W(t) * R_{zoep}(t) - W(t) * R_{approx}(t) \tag{8}$$

In case an a priori probability, that quantifies prior expectations for subsurface variability in the elastic parameters exists, then a sample of the corresponding (and unknown) probability density describing the modeling error can be simulated. As suggested by Hansen et al. (2014), a number of realizations from the prior can be generated for each of which a realization of the corresponding probability density describing the modeling error can be computed using Eqn. 7. This will provide a sample of the probability density describing the modeling error. Thus, the existence of a prior probability over the elastic model parameters must be available, in order to be able to quantify the modeling error in this way. Using e.g. Buland and Omre (2003) such a prior assumption is chosen explicit prior to inversion. In other cases a prior may be implicit in the choice of inversion model.

The approach described above will provide a sample of the probability density describing the modeling error. In some

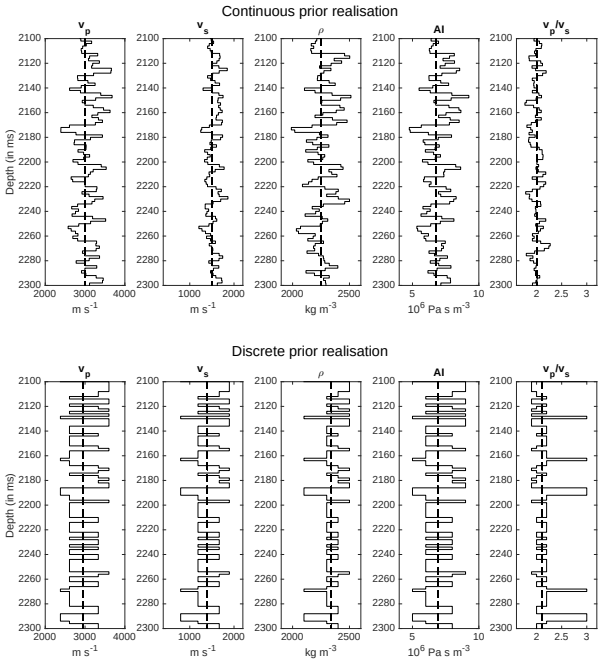


Figure 1: This figure displays one realisation from the continuous (top row) and the discrete (bottom row) prior considered.

cases such a sample can be described by a Gaussian probability, in which case a full description of a Gaussian modeling error can be estimated from the sample of the modeling error as a mean and a covariance (Hansen et al., 2014). If the Gaussian model can be adopted, then it can account for the Gaussian modeling in a linear inverse Gaussian problem, such as the one considered by Buland and Omre (2003), simply by addition of the mean and the covariance of the measurement uncertainty and the modeling error (see Tarantola (2005) for details).

RESULTS

Examples

In order to quantify the prior expectations for the subsurface variability, we have chosen one continuous and one discrete distribution of the elastic properties. The priors are chosen to represent extremes of how the subsurface could be represented. Figure 1 shows a realisation from both priors.

The continuous prior is identical to the one presented by Buland and Omre (2003), and is a Guassian type model with a Gaussian covariance type. The covariance type has a correlation coefficient of 0.7 between the elastic parameters. The continuous prior with strong correlation between the model parameters is one extreme case of the subsurface, in the sense that it produces very smooth transitions. The contrasts in elastic parameters at an interface is therefore relatively small and can be regarded as a 'best case' scenario for the forward approximations based on the small-contrast approximation.

The other extreme is the discrete case. Here we have implemented a truncated plurigaussian prior as described by Oliver et al. (2008) that allows simulation of complex arrangements of lithofacies. As can be seen in the lower plot in Figure 1, a set of four unique discrete layers are simulated with rapid shifts in lithofacies. The elastic parameters of the discrete prior emulates a simplified model of what can be found in North Sea green-sand environments.

In a real geological setting we would suspect the elastic parameters of the subsurface to be somewhere in between these two extremes, i.e. a mixture of smoothly varying elastic parameters within each layer and sharp discontinuous boundaries between some layers.

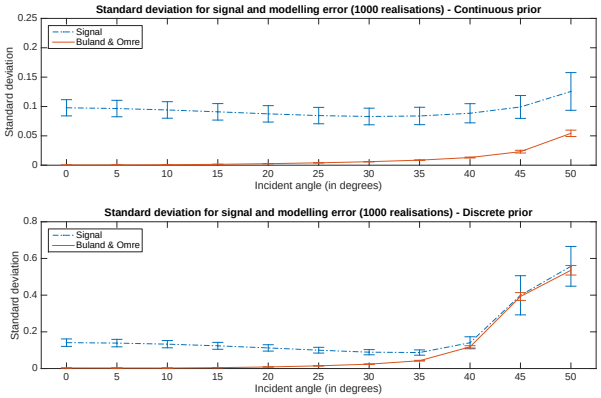


Figure 2: Standard deviation of forward responses from 1000 prior realisations plotted against incident angle using the Zoeppritz equations (dashed blue line). Also plotted is the standard deviation of a 1000 realisations sample of the forward-modeling error applying the Buland and Omre approximation (Red). On top is shown the continuous case while the bottom plot shows the discrete case

Forward modeling error

1000 realisations of the forward modeling error due to the use of the linearised small contrast approximation to the Zoeppritz equations as opposed to the full Zoeppritz equations is obtained using Eqn. 7. These realisations represent a sample of the modeling error. The standard deviation of these realisations at different incident angles are calculated for the Buland and Omre forward approximation and shown in Figure 2. For a relative measure on the severity of the modeling error we have also plotted the standard deviation from 1000 forward realisations using Zoeppritz equations as the signal. The signal to 'modeling error noise' ratio is generally very high for angles close to zero-offset and decreases with increasing incident angles. This is true for both cases, but the decrease is unsurprisingly more rapid for extreme case of a discrete prior. Using the mean of the signal and the modeling error as a signal-to-noise ratio (SN) $= 15$ is reached in the $25-30^\circ$ interval and SN $= 5$ within the $40-45^\circ$ for the continuous case. Meanwhile, the SN $= 15$ is already reached between $15-20^\circ$ and decreases to SN $= 5$ at $25-30^\circ$ for the discrete case.

Forward-modeling error covariance model

Assuming the modeling error is Gaussian, a covariance matrix describing the forward-modeling error is constructed from the sample of 1000 modeling error realisations, as proposed by

Hansen et al. (2014). Three realisations from this covariance model are simulated and plotted in Figure 3 against the actual modeling error. The realisations from the forward modeling error shows the same characteristics as the actual modeling error for both cases. Visually it seems that the low and high-frequencies in the forward-modeling error are well resolved for both the discrete and continuous case. In other words, it is very hard, if not impossible, to distinguish between the actual modeling error and the realisations of the modeling error in case its prior position is unknown, which suggest that the choice of a Gaussian model to represent the modeling error may be valid.

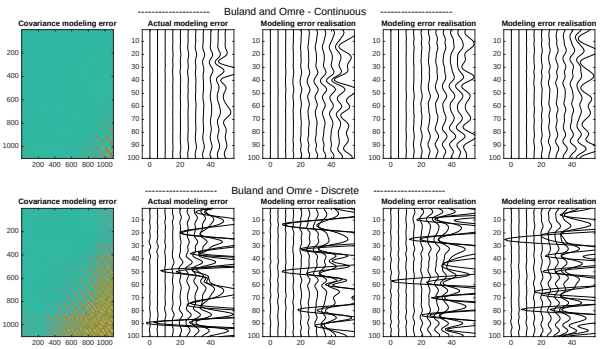


Figure 3: Actual modeling error plotted against three realisations from a Gaussian random field described by a covariance matrix. The modeling error is shown for the Buland and Omre forward. Each row represents a different forward approximation. The left column shows the estimated covariance matrix. The second column is the actual modeling error from a forward. The last three columns are realisations from the Gaussian probability representing the modeling error. The scaling is the same for both cases of realisations and covariance matrices.

Inversion

The upper plot in Figure 4 displays the results from 1000 posterior realisations of a linear AVO bayesian inversion as proposed by Buland and Omre (2003). The Zoeppritz forward response of the continuous prior realisation seen in Figure 1 with added uncorrelated noise is used as reference data. The noise level is set according to a SN = 15 for each incident angle, where the signal is the mean of 1000 forward realisations as described earlier. Initially (top figure) the modeling error is disregarded during inversion. This results in some features that seem well resolved but lie well beyond the 95% confidence interval and represents modeling errors being fitted as data. This bias happens especially where the contrasts in model parameters are fairly high, for instance around 2170-2190 ms and also around 2250 ms. On the lower plot in Figure 4 we have accounted for this modeling error under the assumption that it is Gaussian and that it can be described by the covariance matrix shown in Figure 3. Here it seems that there are no artifacts in the inversion results and that modeling error is handled. The reference model generally lie within the 95% confidence interval as would be expected. The inverted results indicate that the modeling error is properly accounted for.

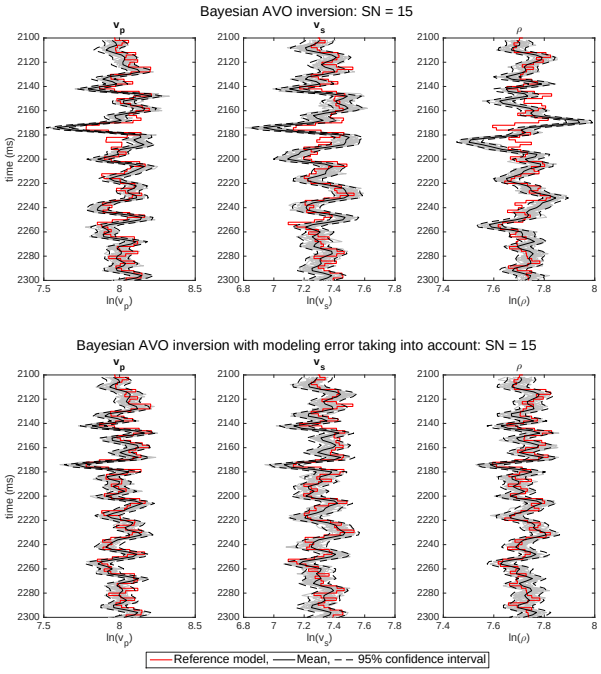


Figure 4: Linear bayesian inversion performed on reference data set (red line) with and without accounting for forward-modeling errors. Posterior realisations (grey), mean (black) and 95 % confidence interval (dashed line) is shown.

CONCLUSION

We have successfully illustrated and quantified the modeling error from applying a linear approximate solution to Zoeppritz equations. The error depends on the degree of subsurface variability as described by the prior assumptions. Assuming the error to be Gaussian distributed, we propose a correction for this modeling error. The error is significantly larger when using discrete a priori models as opposed to a continuous model as shown in Figure 2. Approximating the modeling error by a Gaussian covariance and a mean makes it possible to account for the error in inversion algorithms based on least-squares. Finally it is shown that even small modeling errors related to a smooth Gaussian prior can contribute to significant biases in inversion results.

ACKNOWLEDGMENTS

We would like to thank Henrik Juhl Hansen at Qeye Labs for providing us with realistic properties and real world problems of discrete transitions in the North Sea. Calculations were performed using the Matlab toolbox SIPPI (Hansen et al., 2013). This work is funded by Innovation Fund Denmark.

REFERENCES

Aki, K., and P. Richards, 1980, Quantitative seismology: Theory and methods.

———, 2002, Quantitative seismology: University Science Books. Geology (University Science Books).: Seismology.

Buland, A., and H. Omre, 2003, Bayesian linearized AVO inversion: Geophysics, **68**, 185.

Castagna, J., and M. Backus, 1994, Offset dependent reflectivity: Theory and practice of avo analysis: Investigations in geophysics.

Castagna, J. P., and M. Backus, 1993, Avo analysis-tutorial and review: Offset-dependent reflectivity: theory and practice of AVO analysis, 3–36.

Haase, A. B., 2004, Modelling of linearized zoeppritz approximations: Technical report, AVO model. CREWES Research Report, 2004, 16, 1-4.

Hansen, T. M., K. S. Cordua, B. H. Jacobsen, and K. Mosegaard, 2014, Accounting for imperfect forward modeling in geophysical inverse problemsExemplified for crosshole tomography: Geophysics, **79**.

Hansen, T. M., K. S. Cordua, M. C. Looms, and K. Mosegaard, 2013, SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 1-Methodology: Computers and Geosciences, **52**, 470–480.

Mavko, G., T. Mukerji, and J. Dvorkin, 2009, The Rock Physics Handbook, 2. ed.: Cambridge University Press.

Oliver, D. S., A. C. Reynolds, and N. Liu, 2008, Inverse theory for petroleum reservoir characterization and history matching: Cambridge University Press.

Shuey, R., 1985, A simplification of the Zoeppritzt equations: Geophysics, **50**, 609–614.

Stolt, R., and A. Weglein, 1985, Migration and inversion of seismic data: Geophysics, **50**, 2458–2472.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: Society For Industrial and Applied Mathematics.

Tarantola, A., and B. Valette, 1982, Inverse problems= quest for information: J. geophys, **50**, 150–170.

Thore, P., 2015, Uncertainty in seismic inversion: What really matters?: The Leading Edge, **34**, 1000–1004.

Valentine, A. P., and J. Trampert, 2015, The impact of approximations and arbitrary choices on geophysical images: Geophysical Journal International, **204**, 59–73.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1.

Yilmaz, O., and S. Doherty, 2000, Seismic data analysis: Processing, inversion and interpretation of seismic data: Society of Exploration Geophysicists. Investigations in Geophysics Series.

Zoeppritz, K., 1919, Vii b. über reflexion und durchgang seismischer wellen durch unstetigkeitsflächen: Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, **1919**, 66–84.

H.2 The interplay between geostatistical prior information and modeling error in seismic data

Rasmus Bødker Madsen & Thomas Mejer Hansen, Abstracts from Valencia GEOSTATS 2016, page 179

The interplay between geostatistical prior information and modeling error in seismic data

Rasmus Bødker Madsen[1], Thomas Mejer Hansen[1]

[1] *Niels Bohr Institute, University of Copenhagen*

Keywords: AVO, Zoeppritz, modeling, error

Abstract Inversion of AVO (Amplitude vs. Offset) seismic data is widely used to infer information about reservoir properties from reflection seismic data. The most common method for solving the AVO forward problem (computing the AVO response from an Earth model) is applying a small-contrast approximation to the Zoeppritz equations. This allows a linear formulation of the forward problem, and hence an efficient linear AVO inversion.

However, the small-contrast approximation will introduce a modeling error when solving the forward problem. In this study we demonstrate that the modeling error is closely linked to the choice of prior information, in form of a geostatistical model. The modeling error for a number of different geostatistical models, continuous or discrete, reflecting realistic subsurface, is analyzed, and the modeling error is quantified in form of a 'signal to modeling error' estimate. It is found that when the prior model is continuous and smoothly varying (as when considering a normal distribution with a Gaussian covariance model) then the small-contrast approximation is quite accurate. However, when the prior model contains discrete layer transitions, then modeling error due to using the small-contrast approximation increases to a level that, if not accounted for, could significantly affect linear AVO inversion results. The modeling error is increasingly significant for larger offsets.

Finally, we demonstrate how the modeling error due to the use of the small-contrast approximation can be characterized by a correlated Gaussian model. This allows accounting for the modeling error in linearized AVO inversion, without introducing artefacts or bias in the inversion results. This would further allow for data from larger offsets to be included in the inversion.

H.3 On inferring the noise in probabilistic seismic AVO inversion using hierarchical Bayes

Rasmus Bødker Madsen, Andrea Zunino & Thomas Mejer Hansen, SEG annual meeting 2017 proceedings, page 601-605

On inferring the noise in probabilistic seismic AVO inversion using hierarchical Bayes

Rasmus Bødker Madsen, Andrea Zunino and Thomas Mejer Hansen*
Niels Bohr Institute, University of Copenhagen

SUMMARY

A realistic noise model is essential for trustworthy inversion of geophysical data. Sometimes, as in case of seismic data, quantification of the noise model is non-trivial. To remedy this, a hierarchical Bayes approach can be adopted in which properties of the noise model, such as the amplitude of an assumed uncorrelated Gaussian noise model, can be inferred as part of the inversion. Here we demonstrate how such an approach can lead to substantial overfitting of noise when inverting a 1D reflection seismic NMO data set. We then argue that usually the noise model is correlated, and suggest to infer the amplitude of a correlated Gaussian noise model. This provides better results than assuming an uncorrelated model. In general though, the results suggest that care should be taken using the hierarchical Bayes approach to infer the noise model.

INTRODUCTION

Data ($\mathbf{d} \in \mathbb{R}^N$) are the unique forward response from a physical model (g) given some model parameters ($\mathbf{m} \in \mathbb{R}^M$):

$$\mathbf{d} = g(\mathbf{m}) \tag{1}$$

In nature, observed data ($\mathbf{d}_{\text{obs}}$) are not noise-free, i.e. $\mathbf{d}_{\text{obs}} = g(\mathbf{m}) + \varepsilon$. The inverse problem of inferring values of the model parameters from the observed data are therefore non-unique (Sen and Stoffa, 1996; Tarantola, 2005).

The chosen likelihood function, which measures how well the model parameters match the observed data, is dependent on the distribution of noise (Box and Tiao, 1992). The likelihood, and hence the final posterior distribution of model parameters, can be very sensitive to noise-level and noise models (Thore, 2015). The noise model should ideally include information about measurement and experimental errors as well as account for imperfections in the forward model and/or simplifications due to the parametrization (Dosso and Holland, 2006). However, it is often argued that since no theory is exact, all features that are not captured by the theory are just observational errors, and no distinction should be made between different types of noise (Sen and Stoffa, 1996).

In the classical Bayesian stochastic inference paradigm information on the uncertainty distribution of the data should be established a priori as part of the likelihood function. If for instance a Gaussian noise model is assumed, the corresponding Gaussian likelihood takes the following form (Box and Tiao, 1992)

$$p(\mathbf{d}_{\text{obs}}|\mathbf{m}) = \exp\left(-\frac{1}{2}(\mathbf{d}_{\text{obs}} - g(\mathbf{m}))^T \mathbf{C}_D (\mathbf{d}_{\text{obs}} - g(\mathbf{m}))\right) \tag{2}$$

where $\mathbf{C}_D$ is the data covariance model. The data covariance can be split into contributions from measurement error $\mathbf{C}_d$ and theory error $\mathbf{C}_T$, assuming independence of the two. (Mosegaard and Tarantola, 2002)

$$\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_T \,. \tag{3}$$

An extension to the classical paradigm is offered through the use of the hierarchical Bayesian approach, where parameters of the noise (and of the prior) can be inferred from the observed data (Gelman et al., 2014). A hierarchical scheme has been used in various geophysical inverse problems in order to infer information about noise level from data (Buland and Omre, 2003b; Malinverno and Briggs, 2004; Malinverno and Parker, 2006; Bodin et al., 2012; Dettmer and Dosso, 2012; Ray et al., 2013).

Buland and Omre (2003b) explicitly used a hierarchical Bayes formulation of the inverse problem, enabling noise-level estimation as well as wavelet estimation in a joint AVO (Amplitude Versus Offset) inversion scheme. They concluded that estimating the seismic noise model as part of the probabilistic inversion is viable. However, in their real data case only negligible improvements were gained on posterior variance of model parameters, compared to estimating the noise covariance and wavelet prior to AVO inversion. As opposed to Buland and Omre (2003b), we will analyze the posterior resolution (not the posterior variance) obtained using hierarchical Bayes with different assumptions about the noise model. Specifically we will investigate whether the noise is underestimated, which will result in an apparent smaller posterior variability. However, such reduced posterior variability might reflect non-existent features appearing as a result of fitting noise as data. We propose a set of synthetic tests similar to that of Buland and Omre (2003a). The synthetic tests are based on a reference model from which possible posterior biases in the model parameters can be assessed alongside the posterior variance.

THEORY AND METHOD

Forward model and prior information

The synthetic data (AVO gather) are created following the methodology of Buland and Omre (2003a). The problem here is to infer information about the three elastic parameters: P-wave (v_p), S-wave (v_s), and density (ρ) from a seismic AVO gather. The observed data can be expressed as a linear convolution forward problem with some added noise:

$$\mathbf{d}_{\text{obs}} = \mathbf{G}\mathbf{m} + \mathbf{e} = \mathbf{W}\mathbf{A}\mathbf{D}\mathbf{m} + \mathbf{e} \tag{4}$$

where $\mathbf{W}$ is the wavelet matrix which is convolved with the reflectivity series $\mathbf{AD}\mathbf{m}$. The prior distribution of model parameters $p(\mathbf{m})$ is Gaussian and the three elastic parameters have an internal correlation of 0.7. This prior corresponds to the "Well B" scenario of (Buland and Omre, 2003a). In order to allow

a least-squares solution, the prior distribution of the model parameters is assumed to be log-Gaussian, so:

$$\mathbf{m} = [\ln(v_p/v_p^0)^{\mathrm{T}}, \ln(v_s/v_s^0)^{\mathrm{T}}, \ln(\rho/\rho^0)^{\mathrm{T}}]^{\mathrm{T}} \sim \mathrm{Gauss}(\boldsymbol{\mu}, \mathbf{C}_{\mathrm{M}}) \tag{5}$$

where $\boldsymbol{\mu}$ is the prior expectation of the model parameters, v_p^0, v_s^0, ρ^0 are reference values, and $\mathbf{C}_{\mathrm{M}}$ describes the prior covariance. The noise on the observed data is $\mathbf{e} = \mathbf{e}_{\mathrm{uncor}} + \mathbf{e}_{\mathrm{cor}}$, where the noise model is Gaussian and split into two components as in equation 3. Here we let $\mathbf{C}_{\mathrm{d}}$ be the uncorrelated white noise component and $\mathbf{C}_{\mathrm{T}}$ be the correlated colored noise component, so that:

$$\mathbf{e}_{\mathrm{uncor}} \sim \mathrm{Gauss}(0, \mathbf{C}_{\mathrm{d}}),\ \mathbf{e}_{\mathrm{cor}} \sim \mathrm{Gauss}(0, \mathbf{C}_{\mathrm{T}})\,. \tag{6}$$

A tricky aspect of noise in seismic AVO data is the fact that the error is often systematic (Riedel et al., 2003). In order to test the hierarchical Bayesian approach of inferring noise, the colored noise is set to:

$$\mathbf{C}_{\mathrm{T}} = \sigma_{\mathrm{T}}^2 \mathbf{C}_{\mathrm{T,shape}} = \sigma_{\mathrm{T}}^2 \frac{\mathbf{WADC}_{\mathrm{M}}(\mathbf{WAD})^{\mathrm{T}}}{\max[\mathbf{WADC}_{\mathrm{M}}(\mathbf{WAD})^{\mathrm{T}}]}\,. \tag{7}$$

This covariance matrix gives the covariance of the "prior data distribution". Noise realizations from this distribution would tend to imitate data. By normalizing with the maximum value, the variance of the noise can be set according to σ_{T}^2. Colored noise with signal-to-noise ratio (SNR) = 1.25 is added to the synthetic data, which is in the poor end of what can be expected from seismic data. In practice this is achieved by having the standard deviation σ_{T} as the standard deviation of the reference model's forward response (signal) divided by 1.25. The covariance of the uncorrelated noise is simply the identity matrix $\mathbf{I}$ times the variance:

$$\mathbf{C}_{\mathrm{d}} = \sigma_{\mathrm{d}}^2 \mathbf{C}_{\mathrm{d,shape}} = \sigma_{\mathrm{d}}^2 \mathbf{I}. \tag{8}$$

A SNR = 30 is chosen for the uncorrelated noise. This number is perhaps slightly generous towards the filtering processes. On the other hand, white uncorrelated noise is different in wavelength from the seismic signal and can often easily be filtered out during processing of the raw data (Vecken and Da Silva, 2004). Most of the remaining noise would therefore tend to resemble the observed data in the frequency domain.

Bayesian linearized AVO inversion

Since the inverse problem is linear Gaussian, the Gaussian likelihood in equation 2 can be used with the linear operator $\mathbf{G}$. The posterior distribution $p(\mathbf{d}_{\mathrm{obs}}|\mathbf{m}) \sim \mathrm{Gauss}(\hat{\boldsymbol{\mu}}, \hat{\mathbf{C}}_{\mathrm{M}})$ is then also a multivariate Gaussian distribution (Tarantola and Valette, 1982), where the expectation and covariance are given by:

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + (\mathbf{WADC}_{\mathrm{M}})^{\mathrm{T}} \mathbf{C}_{\mathrm{D}}^{-1} (\mathbf{d}_{\mathrm{obs}} - \mathbf{WAD}\boldsymbol{\mu}) \tag{9}$$

$$\hat{\mathbf{C}}_{\mathrm{M}} = \mathbf{C}_{\mathrm{M}} - (\mathbf{WADC}_{\mathrm{M}})^{\mathrm{T}} \mathbf{C}_{\mathrm{D}}^{-1} \mathbf{WADC}_{\mathrm{M}} \tag{10}$$

This analytical solution of the Bayesian linear inverse problem, which is also referred to as the least-squares solution, describes the full posterior distribution of model parameters with uncertainty under a Gaussian assumption.

Hierarchical Bayes

In a hierarchical model, the conditional parameters (e.g. model parameters or noise model) for the observed data are themselves given a probabilistic specification. Consequently, the conditional parameters are then also dependent on another set of parameters (Gelman et al., 2014). These additional parameters are typically known as hyperparameters $\mathbf{h} = [h_1, h_2, \ldots]$. Uncertainty now includes both model parameters and hyperparameters. Therefore a prior distribution $p(\mathbf{h})$ should be set for the hyperparameters (hyperprior) that reflects the initial uncertainty on these. The posterior distribution of hyperparameters (hyperposterior) is then determined by inversion of the observed data. For linear inverse Gaussian problems, as the one outlined above, Malinverno and Briggs (2004) propose a computationally efficient approach of hieracrchical Bayes, that we adopt here. The marginal likelihood of the observed data conditional on the hyperparameters, in case a linear Gaussian solution to the problem exists, is given by

$$p(\mathbf{d}_{\mathrm{obs}}|\mathbf{h}) = \left[\frac{\det \hat{\mathbf{C}}_{\mathrm{M}}}{\det \mathbf{C}_{\mathrm{M}}}\right]^{\frac{1}{2}} \exp\left[-\frac{1}{2}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{C}_{\mathrm{M}}^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\right] \frac{1}{[(2\pi)^N \det \mathbf{C}_{\mathrm{D}}]^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{d}_{\mathrm{obs}} - \mathbf{G}\hat{\boldsymbol{\mu}})^{\mathrm{T}} \mathbf{C}_{\mathrm{D}}^{-1} (\mathbf{d}_{\mathrm{obs}} - \mathbf{G}\hat{\boldsymbol{\mu}})\right]. \tag{11}$$

Using the sampling strategy suggested by Malinverno and Briggs (2004), the posterior probability distribution of the hyperparameters are essentially sampled using a Metropolis-Hastings algorithm (Mosegaard and Tarantola, 1995). At each step a random walk goes through an exploration phase where a candidate value of the hyperparameters is proposed in vicinity of the current. Thereafter, an exploitation phase either accepts or rejects the candidate based on the marginal likelihood in equation 11. The posterior for the model parameters is sampled with a Gibbs sampler (Sen and Stoffa, 1996). The model parameters are at each step conditioned on the current accepted set of hyperparameters. For our proposed sampling strategy we let the step-length of the random walk in the hyperparameters be dynamic for the first 1000 iterations, with a target acceptance rate of 30%. This is a commonly used practice in Metropolis-Hastings algorithms in order to secure a more efficient burn-in period and subsequent sampling algorithm (Gelman et al., 1996).

SYNTHETIC TESTS

In the following, we present three case studies: the first assuming uncorrelated data noise, the second assuming a known shape of the noise, and the third assuming only an approximate knowledge of the shape of the noise. Regarding the simulated data, the standard deviations of the realization of the error are $\sigma_{\mathrm{d}} = 0.0039$ and $\sigma_{\mathrm{T}} = 0.0949$, respectively, so the noise is mainly correlated. The standard deviation on the final combined noise realization is $\sigma_{\mathrm{d+T}} = 0.0947$.

Hierarchical Bayes inversion - Case 1

In the first inversion example we assume that all noise on the data is uncorrelated. The only hyperparameter that is needed in the hierarchical model is then $h_1 = \sigma_{\mathrm{d}}$ in equation 8. We want

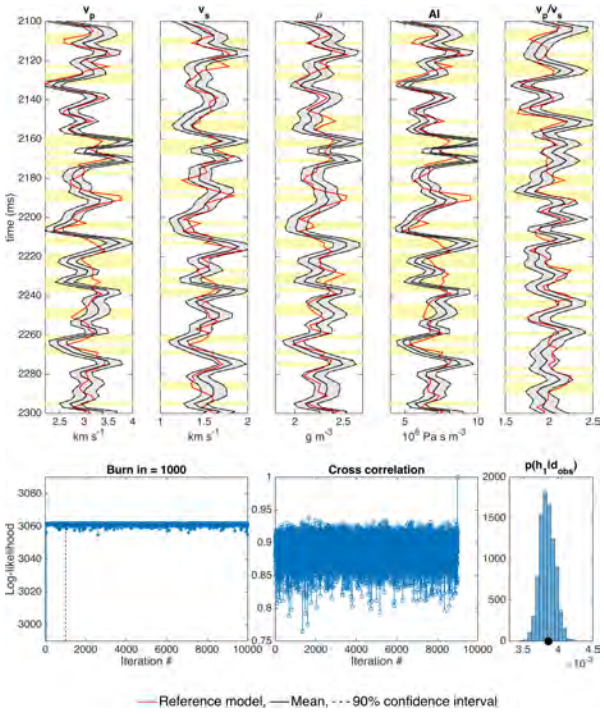


Figure 1: Posterior distribution of $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$ and $p(\mathbf{h}|\mathbf{d}_{\text{obs}})$, log-likelihood and cross-correlation for case 1 of hierarchical Bayes inversion using $\mathbf{C}_{\text{D}} = h_1^2\mathbf{C}_{\text{d}}$

to replicate a setting where little prior knowledge of the noise level exists, but the noise is assumed to be uncorrelated (e.g. Buland and Omre (2003b); Malinverno and Briggs (2004); Bodin et al. (2012)). In these circumstances it seems appropriate to have a relative non-informative hyperprior distribution. We therefore set a rather wide uniform hyperprior distribution for the standard deviation of the uncorrelated noise: $p(h_1) \sim \text{Unif}(0.0001, 1)$. Realizations and statistics of the resulting posterior distribution of model parameters $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$ and hyperparameters $p(\mathbf{h}|\mathbf{d}_{\text{obs}})$ are shown in Figure 1, as well as analytically calculated values for acoustic impedance (AI) and the ratio between P-wave and S-wave velocity (v_p/v_s). Areas where the reference model (red line) are not found within the 90% confidence interval is marked with yellow. Ideally, 90% of the profile should be non-yellow for the posterior resolution to be trustworthy for this interval. However, this is not the case as the reference model is found inside the confidence interval considerably less than 90%. This is especially true for v_p/v_s where the reference model is only found within the 90% confidence interval 46% of the time. This indicates a severe case of overfitting the data. The hyperparameter h_1, instead, is actually well determined when comparing the true standard deviation (black dot) to the histogram in Figure 1. However, the red dot only indicates the level of uncorrelated noise on the data σ_{d}. The total standard deviation of the error realization $\sigma_{\text{d+T}}$ (blue dot) is heavily underestimated as seen in Figure 2. Since the noise on the data is apparently too similar in its structure to the actual signal, the colored noise is then considered as part of the signal by the algorithm. This explains both the

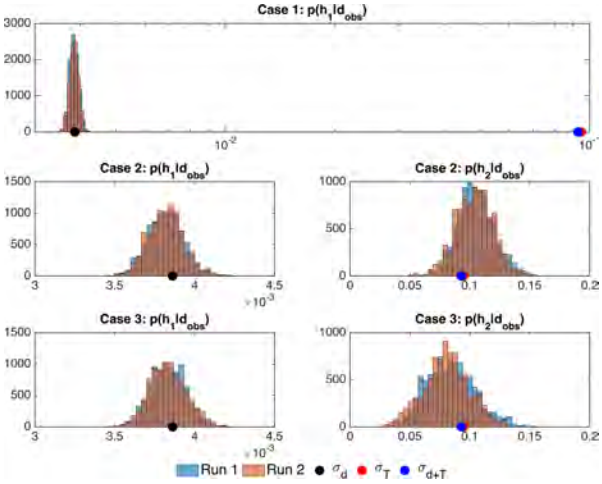


Figure 2: Histograms of the hyperposterior distributions for all MCMC runs $p(\mathbf{h}|\mathbf{d}_{\text{obs}})$. Notice the logarithmic scale used for Case 1 (top plot).

overfitting of the posterior for the model parameters and the algorithms ability to correctly determine the variance of uncorrelated noise on the data while underestimating the actual total noise level.

The correlation between adjacent samples is not high, as the correlation decreases to the average level quickly form the last sample. The pattern of log-likelihood and cross-correlation is similar for all the following MCMC runs in general and are therefore not shown.

The overfitting of the hierarchical Bayes inversion with the uncorrelated noise model becomes even more apparent in Figure 3. Here the percentage of the reference model being inside the confidence interval is plotted as a function of the size of the confidence interval. It is clearly visible that the posterior distribution of the elastic variables is not capturing the reference model for all confidence intervals. The apparently small posterior variability is actually reflecting non-existent features, i.e. noise being fitted as data. The result of the hierarchical inversion is similar to just applying a linear AVO inversion with $\sigma_{\text{d}} = 0.0039$. To evaluate the consistency of the results from case 1, the algorithm is run an additional time. The histogram of the hyperposterior of h_1 in Figure 2 for the secondary run shows the same pattern as for the first run. This indicates a certain level of consistency in the MCMC results.

Hierarchical Bayes inversion - Case 2

In the second test case we assume a known shape of colored noise $\mathbf{C}_{\text{T,shape}}$, i.e. we test whether it is possible at all to infer the variance of the noise knowing the reference color. The standard deviation of the colored noise is set as an additional hyperparameter $\sigma_{\text{T}} = h_2$. The noise model then takes the following form: $\mathbf{C}_{\text{D}} = h_1^2\mathbf{C}_{\text{d,shape}} + h_2^2\mathbf{C}_{\text{T,shape}}$. Again, we set a rather broad uniform hyperprior distribution for the standard deviation of the correlated noise: $p(h_2) \sim \text{Unif}(0.0001, 1)$. The result for two runs are summarized in Figure 2 and 3. Both MCMC runs show approximately the same hyperposterior dis-

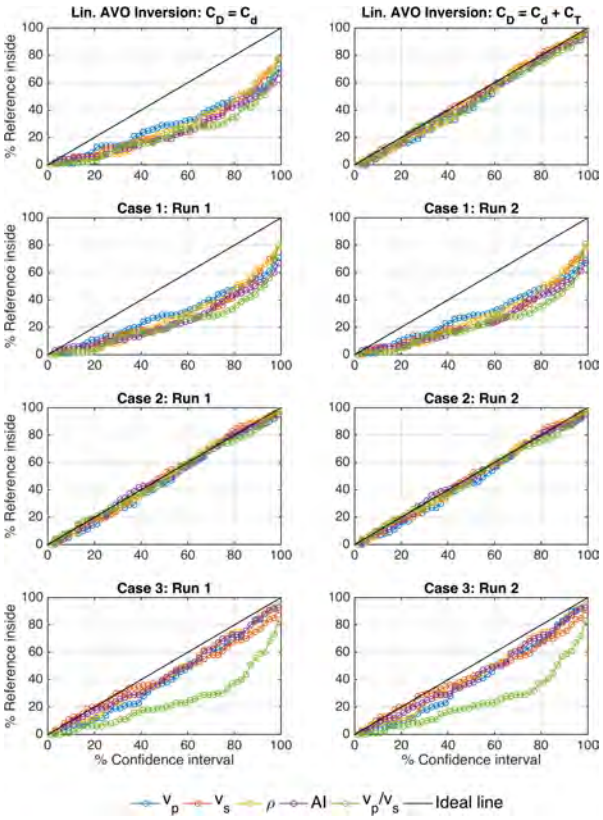


Figure 3: Percentage of reference model being inside the confidence interval, i.e. non-yellow areas in Figure 1), as a function of confidence interval. The uppermost figures are calculated using Bayesian linearized AVO inversion for reference.

tributions $p(\mathbf{h}|\mathbf{d}_{\text{obs}})$. The variance of the correlated noise h_2 (red dot) is slightly overestimated, whereas the variance of the uncorrelated noise h_1 is well-determined. The nearly-correct estimation of variance level, results in the reference model being within the confidence interval following the ideal line in Figure 3. The model therefore appears to neither over- nor underfit the data. The results are similar to doing a linear AVO inversion with the reference noise model.

Hierarchical Bayes inversion - Case 3

Since the correct shape of the noise is never readily available in a real-world scenario, we propose to estimate or assume some correlation of the noise prior to inversion. We assume that the noise is showing smoothness comparable with the wavelet. Furthermore, the noise is correlated between the individual angle-stacks. The correlation between angle stacks is believed to vary slightly as a function of angle. This is set up in the following manner

$$\mathbf{W}_{\text{corr}} = \begin{bmatrix} \beta_1\mathbf{I} & \frac{\beta_1+\beta_2}{2}\mathbf{I} & \dots & \frac{\beta_1+\beta_n}{2}\mathbf{I} \\ \frac{\beta_2+\beta_1}{2}\mathbf{I} & \beta_2\mathbf{I} & \dots & \frac{\beta_2+\beta_n}{2}\mathbf{I} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\beta_n+\beta_1}{2}\mathbf{I} & \frac{\beta_n+\beta_2}{2}\mathbf{I} & \dots & \beta_n\mathbf{I} \end{bmatrix}. \quad (12)$$

where $\mathbf{I}$ is the identity matrix with the size of one individual angle stack. A simple model can easily be derived by setting $\boldsymbol{\beta} = \beta_1, \beta_2, \dots, \beta_n = 1$. Using the variance of each angle stack on the observed data offers an estimate for $\boldsymbol{\beta}$. These beta values could possibly also be obtained using optimization of the marginal likelihood in equation 11. In our case we use $\boldsymbol{\beta} = [1, 0.99, 0.98, 0.95, 0.92, 0.90, 0.88]$. The final approximate shape of the colored noise is then:

$$\mathbf{C}_{\text{W,shape}} = \frac{\mathbf{W}\mathbf{W}_{\text{corr}}\mathbf{W}^{\text{T}}}{\max[\mathbf{W}\mathbf{W}_{\text{corr}}\mathbf{W}^{\text{T}}]} \quad (13)$$

Using the same approach as for case 2, the approximate noise model takes the following form: $\mathbf{C}_{\text{D}} = h_1^2\mathbf{C}_{\text{d,shape}} + h_2^2\mathbf{C}_{\text{W,shape}}$. The variance of the uncorrelated noise is as for case 1 and 2 correctly estimated as both histograms are centered around the correct value (black dot) in Figure 2. However, the variance of the uncorrelated noise is underestimated when comparing the hyperposterior distribution of $p(h_2|\mathbf{d}_{\text{obs}})$ with the correct value (blue dot). As for both case 1 and case 2, there is consistency between the results from the two MCMC runs. Figure 3 shows improvements of case 3 compared to case 1 for all elastic parameters but the v_p/v_s ratio, which is still not captured by the posterior distribution. This indicates that overfitting of the data is still present using the approximate shape, but is nevertheless reduced compared to simply assuming an uncorrelated shape of the noise.

CONCLUSION

Our results in general indicate that caution should be taken when inferring the noise as an additional parameter in inversion. It seems that the assumption of uncorrelated noise in case 1 is not good for inferring the correct noise level on data with correlated noise. The hierarchical Bayes approach was in all cases able to accurately estimate the variance of the uncorrelated noise on the data. However, using the approach of case 1 the total variance of the noise is not recovered and significant overfitting of the data was demonstrated. Choosing the correct shape of the noise, as in case 2, eliminates the overfitting, and a correct variance is estimated even for a wide hyperprior. The results from case 2 indicate that it is possible to estimate a correct variance of the noise model using the data. Finally, for case 3 with an approximate shape of the noise, the variance estimate of the noise is improved compared to case 1. The model is however still overfitting the data. Especially v_p/v_s shows an apparent smaller posterior variability that is not capturing the true model. In a real world case it is probably reasonable to assume that substantial knowledge about the general noise-level is available to constrain the wide hyperprior distribution (Gelman et al., 2014). This could potentially improve the result for an approximate noise model. For our synthetic tests, using an approximate shape of the correlated noise offers an improvement on the noise-level estimate compared with using an uncorrelated noise model. Further work could nonetheless be done to obtain more reliable estimates for an approximate shape of the noise, which could further improve the posterior resolution of the hierarchical Bayes approach in general.

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2017 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Bodin, T., M. Sambridge, N. Rawlinson, and P. Arroucau, 2012, Transdimensional tomography with unknown data noise: Geophysical Journal International, **189**, 1536–1556, http://doi.org/10.1111/j.1365-246X.2012.05414.x.

Box, G. E., and G. C. Tiao, 1992, Bayesian inference in statistical analysis: A Wiley-Interscience Publication, 608.

Buland, A., and H. Omre, 2003a, Bayesian linearized AVO inversion: Geophysics, **68**, 185–198, http://doi.org/10.1190/1.1543206.

Buland, A., and H. Omre, 2003b, Joint AVO inversion, wavelet estimation and noise-level estimation using a spatially coupled hierarchical Bayesian model: Geophysical Prospecting, **51**, 531–550, http://doi.org/10.1046/j.1365-2478.2003.00390.x.

Dettmer, J., and S. E. Dosso, 2012, Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains: Journal of the Acoustical Society of America, **132**, 2239–2250, http://doi.org/10.1121/1.4746016.

Dosso, S. E., and C. W. Holland, 2006, Geoacoustic uncertainties from viscoelastic inversion of seabed reflection data: IEEE Journal of Oceanic Engineering, **31**, 657–671, http://doi.org/10.1109/JOE.2005.858358.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2014, Bayesian data analysis, 3rd ed: Chapman and Hall/CRC.

Gelman, A., G. O. Roberts, and W. R. Gilks, 1996, Efficient metropolis jumping rules: Bayesian Statistics, **5**, 42.

Malinverno, A., and V. A. Briggs, 2004, Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes: Geophysics, **69**, 1005–1016, http://doi.org/10.1190/1.1778243.

Malinverno, A., and R. L. Parker, 2006, Two ways to quantify uncertainty in geophysical inverse problems: Geophysics, **71**, no. 3, W15–W27, http://doi.org/10.1190/1.2194516.

Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, **100**, 12431–12447, http://doi.org/10.1029/94JB03097.

Mosegaard, K., and A. Tarantola, 2002, Probabilistic approach to inverse problems, International handbook of earthquake and engineering seismology, part A: Academic Press, 237–265.

Ray, A., K. Key, and T. Bodin, 2013, Hierarchical Bayesian inversion of marine CSEM data over the Scarborough gas field A lesson in correlated noise: 83rd Annual International Meeting, SEG, Expanded Abstracts, 723–727, http://doi.org/10.1190/segam2013-1439.1.

Riedel, M., S. E. Dosso, and L. Beran, 2003, Uncertainty estimation for amplitude variation with offset (AVO) inversion: Geophysics, **68**, 1485–1496, http://doi.org/10.1190/1.1620621.

Sen, M. K., and P. L. Stoffa, 1996, Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion: Geophysical Prospecting, **44**, 313–350, http://doi.org/10.1111/j.1365-2478.1996.tb00152.x.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation, 1st ed.: SIAM.

Tarantola, A., and B. Valette, 1982, Inverse problems = quest for information: Journal of Geophysics, **50**, 159–170.

Thore, P., 2015, Uncertainty in seismic inversion: What really matters?: The Leading Edge, **34**, 1000–1004, http://doi.org/10.1190/tle34091000.1.
Vecken, P. C. H., and M. Da Silva, 2004, Seismic inversion methods and some of their constraints: First Break, **22**, 47–70, http://doi.org/10.3997/1365-2397.2004011.

H.4 Estimation and accounting for the modeling error in probabilistic linearized AVO inversion

Rasmus Bødker Madsen & Thomas Mejer Hansen, GEOPHYSICS, 2018, Volume 83, page N15-N30

GEOPHYSICS, VOL. 83, NO. 2 (MARCH-APRIL 2018); P. N15–N30, 10 FIGS., 2 TABLES.
10.1190/GEO2017-0404.1

Estimation and accounting for the modeling error in probabilistic linearized amplitude variation with offset inversion

Rasmus Bødker Madsen[1] and Thomas Mejer Hansen[1]

ABSTRACT

A linearized form of Zoeppritz equations combined with the convolution model is widely used in inversion of amplitude variation with offset (AVO) seismic data. This is shown to introduce a "modeling error," compared with using the full Zoeppritz equations, whose magnitude depends on the degree of subsurface heterogeneity. Then, we evaluate a methodology for quantifying this modeling error through a probability distribution. First, a sample of the unknown probability density describing the modeling error is generated. Then, we determine how this sample can be described by a correlated Gaussian probability distribution. Finally, we develop how such modeling errors affect the linearized AVO inversion results. If not accounted for (which is most often the case), the modeling errors can introduce significant artifacts in the inversion results, if the signal-to-noise ratio is less than 2, as is the case for most AVO data obtained today. However, if accounted for, such artifacts can be avoided. The methodology can easily be adapted and applied to most linear AVO inversion methods, by allowing the use of the inferred modeling error as a correlated Gaussian noise model.

INTRODUCTION

Amplitude variation with offset (AVO) reflection seismic data, and the corresponding amplitude variation with angle (AVA), can be used to analyze how the reflected energy from a layer boundary depends on the offset between a source and a receiver (or the angle of incidence ϕ at the boundary) and the elastic properties around a specific layer boundary (Castagna and Backus, 1993). AVO analysis has been successfully used to identify possible hydrocarbon reservoirs directly from seismic data (Castagna and Backus, 1993; Castagna et al., 1998). Inversion of NMO corrected prestack seismic data (AVO data) has also been widely used to infer information about the elastic parameters in a deterministic (e.g., Cooke and Schneider, 1983) and a probabilistic framework (e.g., Buland and Omre, 2003).

AVO data are always associated with uncertainty, which can be classified into "measurement uncertainty" and modeling error. The term modeling error (sometimes also referred to as "modelization error" or "theoretical error") is, in this paper, referring to the use of an inexact theory in the modeling (prediction) of the result of measurements as opposed to errors arising from inaccurate measurements due to, e.g., instrument errors (Tarantola and Valette, 1982b; Sen and Stoffa, 1996; Downton, 2005; Tarantola, 2005).

An abundance of work exists that acknowledges modeling errors related to AVO modeling (Gerstoft and Mecklenbräuker, 1998; Buland and Omre, 2003; Riedel et al., 2003; Dosso and Holland, 2006; Chen et al., 2007; Rabben et al., 2008; Bosch et al., 2010; Aune et al., 2013).

A widely considered modeling error with respect to AVO data is related to the use of approximations to the Zoeppritz (1919) equations, which allow the AVO/AVA response at a plane-layer interface to be computed analytically. These approximations are typically based on the small-contrast approximation given by Aki and Richards (1980) and are further developed by, e.g., Shuey (1985) and Castagna and Backus (1993). These approximations are used for AVO analysis and inversion.

Most approximations of the Zoeppritz equations are assumed to be valid (implying insignificant modeling errors) for incidence angles $\theta < 30°$ (Shuey, 1985; Castagna and Backus, 1993; Buland and Omre, 2003; Mavko et al., 2009). If the variations in the elastic properties are assumed to be very smooth, implying small contrasts, the approximation may be valid for incidence angles $\theta > 30°$. Furthermore, by using the average angle between the incidence and transmitted angle, the modeling error can be decreased in general (Downton and Ursenbach, 2006).

Even for small incidence angles (less than 30°), systematic errors have been shown to arise in linearized AVO inversion due to the

Manuscript received by the Editor 26 June 2017; revised manuscript received 19 September 2017; published ahead of production 21 November 2017; published online 11 January 2018.
[1]University of Copenhagen, Niels Bohr Institute, Copenhagen, Denmark. E-mail: rasmus.madsen@nbi.ku.dk; tmeha@nbi.ku.dk.

modeling error of applying a linear operator as a forward model (Downton, 2005; Rabben et al., 2008). The problem is most severe using two-term approximations (Downton, 2005). In this case, significant systematic errors for the gradient estimate can be detected. Further, linear approximations tend to produce false predictions, or artifacts in noncontinuous and nonsmooth subsurface models (Stolt and Weglein, 1985).

Thus, modeling errors related to using approximations to the Zoeppritz equations have been widely considered. Yet, little work has been done to quantify this modeling error and to account for this modeling error as part of inverting AVO data.

Here, we develop the approach proposed by Hansen et al. (2014) to simulate, model, and account for modeling errors in relation to AVO modeling. Specifically, and as a first example, the widely used small-contrast approximation of the Zoeppritz equations will be investigated.

We expect such modeling errors to represent a lower limit of the full set of modeling errors inherent in AVO data. A variety of sources for modeling errors exist in seismic data, such as using a 1D convolutional model to reflect a 3D physical system, the use of the acoustic-wave equation as opposed to the anisotropic visco-elastic wave equation, imperfections in data processing, general anisotropy considerations, the effects of processing the raw data, the coupling of data to physics within the forward model, uncertain wavelet estimates, and uncertainty on the low-frequency model. (see also Ball et al., 2015; Li et al., 2015; Thore, 2015). For example, we expect higher magnitude modeling errors related to using the Zoeppritz equations as opposed to using the full wave equation. However, as shown in this paper, even disregarding this lower limit of modeling errors can potentially lead to significant biases when such AVO data are inverted without accounting for modeling errors.

First, we introduce the approximate forward model and describe the modeling error associated with it. This is followed by a discussion on how to estimate and quantify this error. This is done in a probabilistic framework. First, a set of realizations representing a sample of a (unknown) probability distribution describing the modeling error is generated. Then, it is demonstrated that this sample can be reasonably described by a multivariate Gaussian probability density. Finally, we investigate the possibility of accounting for (and ignoring) the forward modeling error during probabilistic linear inversion of seismic AVO data (as in Buland and Omre, 2003), considering different levels of measurement uncertainty.

LINEARIZED AVO FORWARD MODELING

In the following, we will be using the term AVO as synonymous with AVA. Perhaps, the most correct approach to solve the AVO forward problem (i.e., simulating an AVA gather) is to use some form of full-waveform modeling to simulate shot gathers, followed by a NMO correction and sorting according to angle of incidence ϕ, to obtain a set of AVA gathers. It is however computationally demanding, and especially if used as part of an inverse problem, it becomes impractical (Virieux and Operto, 2009). It is also argued that in real-world cases, processing the raw seismic data such that it enables the use of a simpler forward problem is often advantageous (Claerbout et al., 2004). AVO data are, for example, processed such that each data point can be associated with the reflection at a specific point in the subsurface. The Zoeppritz equations allow an analytical relation between the amplitude of reflections and elastic parameters around an interface in the subsurface. In practice, a widely used forward model to simulate AVO data is a combination of a linearized approximation to Zoeppritz equations combined with the convolution model.

The success of the simpler approximations can be attributed to their ease of evaluation and interpretation (Castagna and Backus, 1993; Mavko et al., 2009), and their ability to linearize the AVO inverse problem (Buland and Omre, 2003).

1D linearized AVO forward modeling

The Zoeppritz (1919) equations describe the full set of angle-dependent reflectivities at a plane interface between two media with different elastic properties for a plane wave.

The Aki and Richards forward model

Aki and Richards (1980) propose a small-contrast approximation to Zoeppritz equations, in which the reflection coefficient, as a function of incidence angle ϕ can be computed from the elastic parameters above and below the interface using

$$R(\phi) \asymp a_{v_P}(\phi)\frac{\Delta v_P}{\overline{v_P}} + a_{v_S}(\phi)\frac{\Delta v_S}{\overline{v_S}} + a_{\rho}(\phi)\frac{\Delta \rho}{\overline{\rho}}, \quad (1)$$

where the coefficients are given by

$$a_{v_P}(\phi) = \frac{1}{2\cos^2\phi}, \quad (2)$$

$$a_{v_S}(\phi) = -\frac{4\overline{v_S}^2}{\overline{v_P}^2}\sin^2\phi, \quad (3)$$

$$a_{\rho}(\phi) = \frac{1}{2}\left(1 - 4\frac{\overline{v_S}^2}{\overline{v_P}^2}\sin^2\phi\right), \quad (4)$$

where $\overline{v_P}$, $\overline{v_S}$, and $\overline{\rho}$ represent the average P-wave, S-wave, and density over the interface, whereas Δv_P, Δv_P, and $\Delta\rho$ represent elastic contrasts over the interface. Equation 1 is valid for a small percentile change in elastic properties, i.e., for a small contrast between layers.

The assumption of small contrasts is linked to the relative change in contrasts $\Delta v_P/\overline{v_P}$, $\Delta v_S/\overline{v_S}$, and $\Delta\rho/\overline{\rho}$. We will refer to this choice of AVO forward model of reflectivities as the "Aki and Richards forward model."

Note that equation 1 is a better approximation if the average angle at the interface is used, as opposed to the incidence angle (Downton and Ursenbach, 2006).

The Buland and Omre forward model

Buland and Omre (2003) adapt the following forward approximation proposed by Stolt and Weglein (1985), which expands the Aki and Richards approximation, in which reflection coefficients are now also time dependent:

$$R(t,\phi) \asymp a_{v_P}(t,\phi)\frac{\partial}{\partial t}\ln\, v_P(t) + a_{v_S}(t,\phi)\frac{\partial}{\partial t}\ln\, v_S(t) + a_\rho(t,\phi)\frac{\partial}{\partial t}\ln\, v_P. \quad (5)$$

The difference terms in equation 1 are in equation 5 substituted with the partial derivative of the logarithmic value of each material parameter, e.g., $\partial/\partial t \ln v_P(t)$ replaces $\Delta v_P/\overline{v_P}$. This substitution is valid only for small contrasts in the elastic parameters.

The time dependency in the coefficients $a_{v_P}(t,\phi)$, $a_{v_S}(t,\phi)$, and $a_\rho(t,\phi)$ is a generalization of the Aki and Richards coefficients (equations 2–4) with the time-dependent averages $\overline{v_P}$, $\overline{v_S}$, and $\overline{\rho}$. To solve this equation, it is here assumed that these averages are described by a known background model.

This allows a linear relation between the derivative of the model parameters (the logarithm of the elastic parameters) and the reflection coefficients

$$\mathbf{R} = \mathbf{ADm}, \quad (6)$$

where $\mathbf{A}$ is the linear operator composed of the coefficients $a_{v_P}(t,\phi)$, $a_{v_S}(t,\phi)$, and $a_\rho(t,\phi)$ from equation 5 and $\mathbf{D}$ is the derivative matrix. We will refer to this choice of AVO forward model for reflectivities as the "Buland and Omre forward model."

The linear convolution model

A seismic trace is obtained using the convolution model:

$$S(t) = W(t) * R(t) \equiv \int_0^{t_s} W(\tau)R(t-\tau)\mathrm{d}\tau, \quad (7)$$

where $S(t)$ is the seismic trace, $R(t)$ is the reflectivity series (earth response), such as given in for example equations 1 and 5, and $W(t)$ is the wavelet (source-time function), where t_s is the duration of the source input. Evaluating the AVO forward problem using the Zoeppritz equations (to obtain a seismic trace $S_{\mathrm{zoep}}(t)$) thus amounts to first computing the reflection coefficients using the Zoeppritz equations, followed by a convolution with a wavelet. This is a nonlinear process.

By applying the convolution model for several angles of incidence, each with a possibly unique wavelet, a linear relationship between AVO seismic data and the elastic model parameters using equation 6 is written as

$$\mathbf{d}_{\mathrm{AVO}} = \mathbf{WR} = \mathbf{WADm}, \quad (8)$$

where $\mathbf{W}$ is a convolution matrix containing the wavelet. Buland and Omre (2003) introduce a Bayesian linearized AVO inversion technique by adapting the linear relation in equation 8 when the wavelet is known. This relation is considered the "full Buland and Omre AVO forward model" and allows a full description between model parameters and AVA data using a linear theory.

Calculating the forward modeling error

For one specific 1D elastic model, the modeling error S_{error} (related to a specific elastic model) is calculated as the difference (residual) between the seismic signal from the full Zoeppritz equations $S_{\mathrm{zoep}}(t)$ and the seismic signal from the approximate forward model $S_{\mathrm{app}}(t)$:

$$S_{\mathrm{error}}(t) = S_{\mathrm{zoep}}(t) - S_{\mathrm{app}}(t), \quad (9)$$

$$= W(t) * R_{\mathrm{zoep}}(t) - W(t) * R_{\mathrm{app}}(t), \quad (10)$$

where $R_{\mathrm{zoep}}(t)$ is the reflectivity series calculated with the Zoeppritz equations and $R_{\mathrm{app}}(t)$ is the one calculated with an approximate forward model (equation 1 or 6). For a single realization from a Gaussian probability density identical to "well B" in Buland and Omre (2003), the AVO forward response calculated using Zoeppritz equation and the Buland and Omre forward model are shown alongside the modeling error (equation 9) in Figure 1. The modeling error is increasing with angle of incidence.

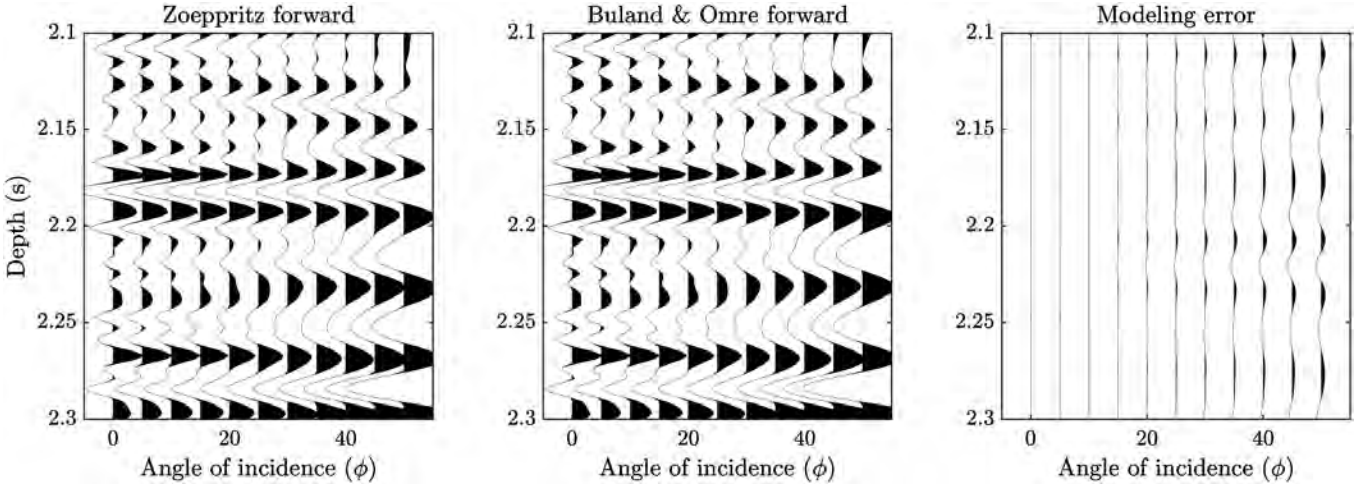


Figure 1. AVO (AVA) data obtained using the full Zoeppritz and the Buland and Omre forward model for a realization from a Gaussian probability density (the small reflectivity model) identical to that of Buland and Omre (2003). On the right, the corresponding forward modeling error is shown. The elastic model used to generate these forward responses is shown in Figure 2.

QUANTIFYING THE FORWARD-MODELING ERROR

Hansen et al. (2014) demonstrate how to quantify the modeling error probabilistically through a probability density, $\theta(\mathbf{d}|\mathbf{m})$. The main idea is to generate a large sample of an assumed (unknown) probability density reflecting the modeling error. To do this, a source of the modeling error has to be identified, and quantified probabilistically, such that realizations of the probability density describing the source of the modeling error are generated. An algorithm that can sample the probability density will suffice, and the actual probability density itself need not be known. The source of the modeling error can for example be subsurface variability and uncertainty related to the wavelet.

In some cases, the obtained sample of the modeling error can be described by a Gaussian probability, in which case, a full description of a Gaussian modeling error can be estimated from the sample of the modeling error as a mean and a covariance (Hansen et al., 2014). If the Gaussian model is adopted, then it can account for the Gaussian modeling error in a linear inverse Gaussian problem, such as the one considered by Buland and Omre (2003), by addition of the mean and the covariance of the measurement uncertainty and the modeling error (for details, see Mosegaard and Tarantola, 1995; Tarantola, 2005).

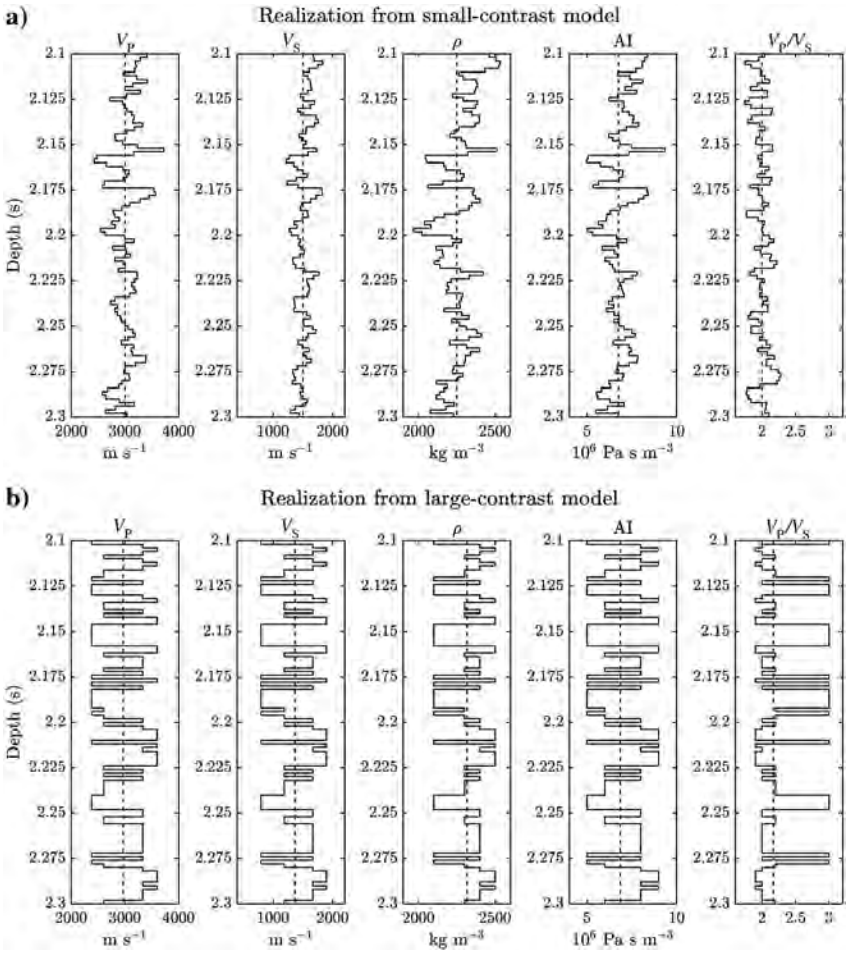


Figure 2. (a) One realization from the small-contrast statistical model and (b) the large-contrast statistical model considered.

Forward-modeling error

As an example, we consider two different types of geostatistical subsurface models to illustrate how to sample and quantify the associated modeling error. These two models, a small-contrast and a large-contrast model, respectively, represent extreme cases of subsurface variability. Figure 2 shows a realization from both models. Arbitrarily large amounts of realizations of these geostatistical models are generated that all respect the assumed statistical properties.

The "small-contrast" model is identical to the one presented by Buland and Omre (2003), and it is defined as a Guassian probability density. The correlation between the elastic parameters is 0.7. The choice of a Gaussian-type covariance model results in smooth models, in which the contrasts in elastic parameters at an interface are relatively small. The small-contrast prior model can therefore be regarded as a "best-case" scenario for the forward approximations based on the small-contrast approximation.

The other extreme is a "large-contrast" prior model, in which contrasts in the elastic parameters are high. The model is based on a truncated plurigaussian prior distribution that allows simulation of complex arrangements of lithofacies (Armstrong et al., 2011). As seen in the lower plot in Figure 2, a set of four unique discrete layers are simulated with rapid shifts in lithofacies. The elastic parameters of the large-contrast prior emulate a simplified model of what is found in North Sea green-sand environments (Svendsen et al., 2012). The large-contrast prior model may represent a "worst-case" scenario for the use of the small-contrast approximation.

In a real geologic setting, the elastic parameters of the subsurface are expected to be somewhere in between these two extremes.

The convolution model requires the wavelet to be known. Wavelet estimation is an inverse problem in itself. Here, it is assumed that the wavelet is known and is a Ricker wavelet with linearly decreasing center frequency from 50 Hz at zero offset to 25 Hz at the largest incidence angle ($\phi = 50°$). The interval between each trace is $\Delta\phi = 5°$. This yields 11 traces. For 100 time samples, the data ($\mathbf{d}_{\mathrm{AVO}}$) are of size $N_d = 1100$. An example of synthetic data calculated using the small-contrast prior is shown in Figure 1. In total, 100 time samples times the three elastic parameters gives $N_m = 300$ model parameters ($\mathbf{m} = [\ln(v_P)^\top, \ln(v_S)^\top, \ln(\rho)^\top]^\top$).

Using one realization from either model, one realization of the modeling error $\mathbf{d}_e = [S_{\mathrm{error},\phi=0}, S_{\mathrm{error},\phi=5}, \ldots, S_{\mathrm{error},\phi=50}]$ is calculated using equation 9. One thousand of these modeling error realizations are computed, which represent a sample of the (unknown) probability distribution describing the modeling error. The average standard deviation of this sample at different incidence angles is calculated for the Buland and Omre forward (equation 8) and the Aki and Richards forward (equation 1 plus 7), and it is shown in Figure 3. For the Aki and Richards forward, the average angle ϕ_{avg} between the incidence angle and the transmitted angle is used instead of the incidence angle, as described by

Downton and Ursenbach (2006). For the Buland and Omre forward, results are shown using the incidence angle and the average angle.

The average standard deviation introduces a way to measure the amplitude of the modeling error at different incidence angles. For a relative measure on the severity of the modeling error, we compare with the magnitude of the data signal. The data signal is obtained as the average standard deviation of a sample of 1000 forward realizations using the Zoeppritz equations (i.e., to obtain the average standard deviation of the signal, not the error). Comparing this signal (black line) with the modeling error of each approximation (dashed, dotted, and light-gray line) in Figure 3, a visual signal-to-noise ratio (S/N) is obtained.

Both forward approximations show increasing amplitude of the modeling error with the increasing angle of incidence, in agreement with Figure 1. For the small-contrast model (the uppermost plot), the modeling error of the Buland and Omre forward model (the dotted and dashed lines) is higher than the Aki and Richards forward model (the light gray line) for all incidence angles. This is due to the necessary additional small-contrast assumption (logarithmic approximation) in the Buland and Omre forward model. Especially, for angles of incidence $\phi > 35°$, the S/N becomes low for the Buland and Omre forward model because the average modeling error rises. Meanwhile, the average S/N is relatively high for the Aki and Richards forward model due to a generally low modeling error.

Similar trends are visible for the large-contrast model (the lowermost plot). However, the S/N is considerably lower and more significant for both forward models. This is because both forward models rely on the assumption of small contrasts in the elastic parameters, which is not provided by the large-contrast prior model. The S/N is still high for both models for lower angles of incidence ($\phi < 20°$). At larger angles of incidence, the S/N becomes considerably lower for both forward models. This culminates in an $S/N = 1$ for the Buland and Omre forward model with incidence angle (dotted line) for angles of incidence $\phi > 40°$. This implies that the average modeling error has the same amplitude as the seismic signal for these wide incidence angles. Interestingly, the average modeling error is actually higher for the forward using an average angle (the dashed and light gray lines) for angles of incidence between $10° < \phi < 35°$. A possible explanation for this curiosity is that a phase component arises in the solution when using the average angle in the forward models for large contrasts in the elastic parameters. To remedy this, we use the magnitude of the reflection coefficient's real and complex part as suggested by Lay and Wallace (1995).

In general, the most erroneous of the two approximations is the Buland and Omre forward model when using the incidence angle (the dotted line). However, because the Buland and Omre forward allows a linear relationship between AVO seismic data and the elastic model parameters (equation 8), consequently allowing a Bayesian linearized AVO inversion; we will be using this forward model in the following calculations.

The results also indicate that using the average angle, instead of the incidence angle, reduces the modeling error substantially, as suggested by Downton and Ursenbach (2006). But, as will be discussed later, the average angle is typically not known when performing linearized AVO inversion because the elastic parameters that are needed to compute the transmission angle are not known prior to inversion. Further, if an elastic a priori model is assumed, it is most often assumed to be smoothly varying, such that there will be little difference between the angle of incidence and the average angle.

A Gaussian model of the forward-modeling error

Assuming that the modeling error is Gaussian, a Gaussian model in the form of a mean and a covariance $\mathcal{N}(\mathbf{d}_{T\text{app}}, \mathbf{C}_{T\text{app}})$ describing the forward-modeling error $\theta(\mathbf{d}|\mathbf{m}) \sim \mathcal{N}(\mathbf{d}_{T\text{app}}, \mathbf{C}_{T\text{app}})$ can be constructed from a large sample of modeling error realizations, as proposed by Hansen et al. (2014) (see Appendix A). The estimated covariance matrix $\mathbf{C}_{T\text{app}}$ of the modeling error using the Buland and Omre forward approximation is depicted in Figure 4 for the two prior models. Both covariance matrices show a heavily banded

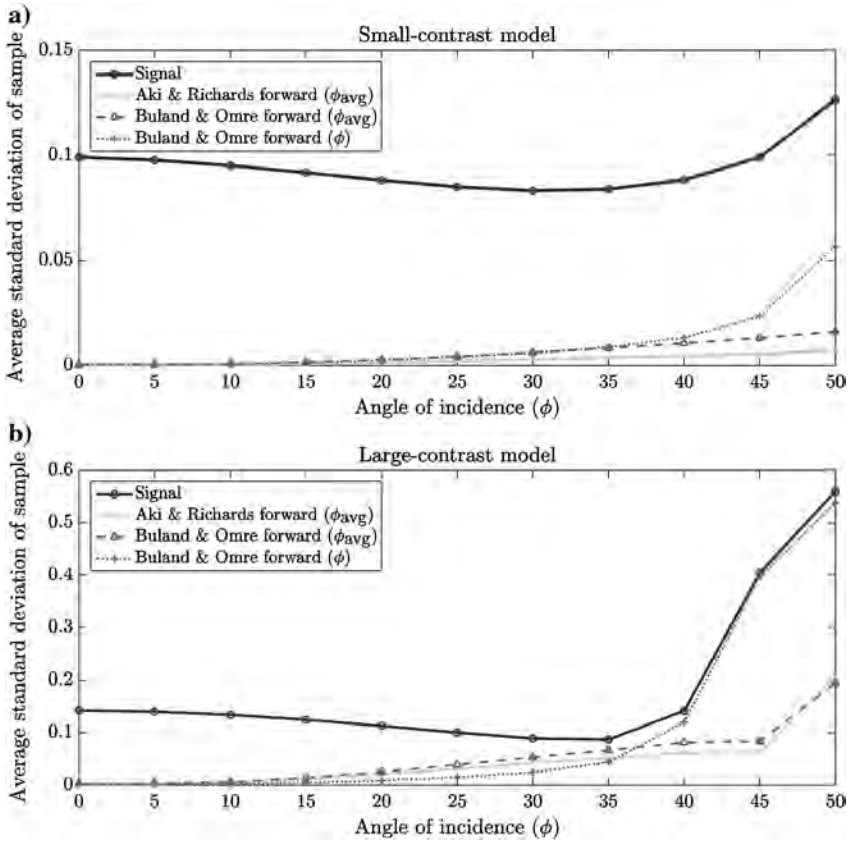


Figure 3. Average standard deviation of the modeling error applying the Aki and Richards approximation using the average angle (the light gray line), and the Buland and Omre approximation using the average and the incidence angle (dashed and dotted line) plotted against the incidence angle. The lines represent the average of 1000 realizations of the modeling error. Also plotted is the forward response from 1000 subsurface realizations using the Zoeppritz equations (the black line) as the signal. Results are shown for (a) the small-contrast model and (b) the large-contrast model.

structure, which indicates a high degree of correlated modeling error. The bands are especially distinguishable at intervals of 100 data points, indicating the correlation between the same time samples at different incidence angles. The covariance matrix also shows that adjacent time samples are correlated within these bands. The amplitude of the modeling error increases along the diagonal, i.e., with an increasing angle of incidence, as is also shown in Figure 3.

In addition to the estimated Gaussian model $\mathcal{N}(\mathbf{d}_{T\text{app}}, \mathbf{C}_{T\text{app}})$, we also examine a Gaussian model that contains only the uncorrelated part of the estimated covariance matrix. This is done to assess the importance of the apparent correlation of the modeling error and to evaluate the common practice, as mentioned earlier, of describing the overall data uncertainty with an uncorrelated Gaussian model with a known variance. The full estimated Guassian model (using equations A-3 and A-4) containing the correlated and uncorrelated parts will henceforth be known as $\mathbf{C}_{T\text{app1}}$, and the model containing only the uncorrelated part will be known as $\mathbf{C}_{T\text{app2}}$. However, the question now remains: Is the Gaussian model assumption for the modeling error reasonable? This is investigated further in the following by assessing the Gaussian assumption qualitatively and quantitatively.

Qualitative assessment — Visual comparison

A qualitative assessment of the validity of the Gaussian model assumption $\theta(\mathbf{d}|\mathbf{m}) \sim \mathcal{N}(\mathbf{d}_{T\text{app}}, \mathbf{C}_{T\text{app}})$ is obtained by visually comparing realizations of the observed (actual) modeling error (obtained using equation 10), with realizations of the Gaussian model describing the modeling error, which should show similar characteristics. Three realizations from $\mathbf{C}_{T\text{app1}}$ and $\mathbf{C}_{T\text{app2}}$ are simulated and plotted in Figure 5 alongside the observed modeling error. Realizations are shown for the small-contrast and large-contrast prior model.

Realizations from $\mathbf{C}_{T\text{app1}}$ generally show the same characteristics as the observed modeling error for both prior cases. Visually, it seems that the amplitudes and patterns are alike for corresponding incidence angles of the observed and simulated realizations. The main variability of the modeling error seems to be imitated and it is very hard, if not impossible, to distinguish between the observed modeling error and the realizations of $\mathbf{C}_{T\text{app1}}$. This suggests qualitatively that the choice of a Gaussian model to represent the modeling error is valid. On the contrary, the realizations from $\mathbf{C}_{T\text{app2}}$ are not able to reproduce the pattern of the observed modeling error for either prior model. The amplitudes are similar to the observed modeling error, but the white noise is a poor imitation of the overall characteristics.

In summary, a qualitative assessment of the modeling errors through visual comparison shows that the distribution and frequency content of the observed modeling error is fairly well-represented by the estimated $\mathcal{N}(\mathbf{d}_{T\text{app1}}, \mathbf{C}_{T\text{app1}})$. A quantitative assessment revealed that the correlated Gaussian model cannot completely describe the observed modeling error, but that it is a much better representation than a simple uncorrelated Gaussian model (see Appendix B). The tails of the observed modeling error $\mathbf{d}_{e,\text{obs}}$ in the 1D marginal distribution could potentially be fitted better with a distribution capable of producing outliers more regularly (Cauchy distribution, Voigt profile, etc.). Whether the Gaussian description of the modeling error is "good enough" will ultimately be determined by what it will be used for, as, for example, solving the inverse AVO problem.

FORWARD-MODELING ERROR IN BAYESIAN LINEARIZED AVO INVERSION

The linear relation in equation 8 allows formulating a linear inverse problem, in which the elastic parameters are estimated directly from AVO data. Buland and Omre (2003) solve this inverse problem in a probabilistic framework following Tarantola and Valette (1982a). Here, the noise on the AVO data is assumed to follow a Gaussian distribution with a mean of zero and a covariance matrix of $\mathbf{C}_D$ ($\varepsilon \sim \mathcal{N}(0, \mathbf{C}_D)$). The prior information on the logarithm of

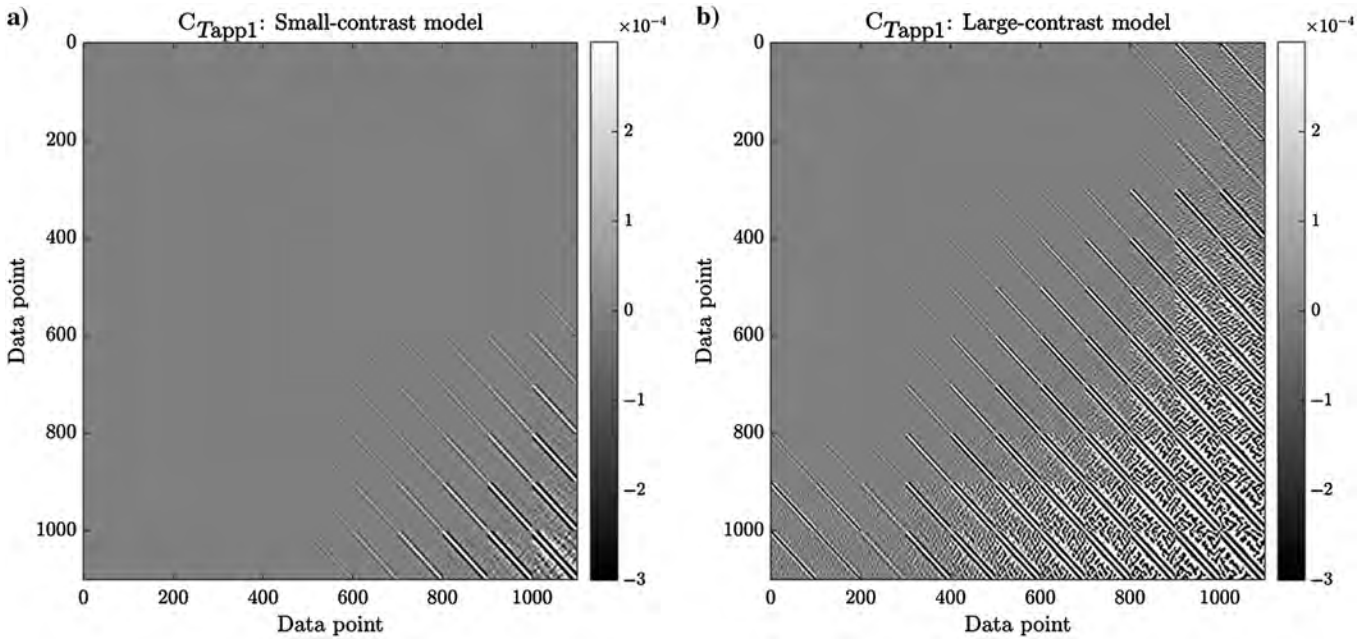


Figure 4. Estimated covariance matrices $\mathbf{C}_{T\text{app}}$ of the forward-modeling error for both prior models. The covariances are estimated from a large sample (1000 realizations) of the modeling error. The same color scaling is used for both plots. The covariances show a banded structure, indicating that the estimated modeling error is correlated.

the elastic parameters is assumed to follow a Gaussian distribution with mean vector $\boldsymbol{\mu}_M$ and covariance matrix $\mathbf{C}_M$ ($\mathbf{m} \sim \mathcal{N}(\boldsymbol{\mu}_M, \mathbf{C}_M)$), identical to the small-contrast statistical model presented earlier. The posterior probability density of the model parameters $\tilde{\mathbf{m}}$ is described as a Gaussian probability distribution $\mathcal{N}(\tilde{\mathbf{m}}, \tilde{\mathbf{C}}_M)$ with mean

$$\tilde{\mathbf{m}} = \boldsymbol{\mu}_M + (\mathbf{WADC}_M)^\top \mathbf{C}_D^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{WAD}\boldsymbol{\mu}_M) \qquad (11)$$

and covariance

$$\tilde{\mathbf{C}}_M = \mathbf{C}_M - (\mathbf{WADC}_M)^\top \mathbf{C}_D^{-1}\mathbf{WADC}_M, \qquad (12)$$

where $\mathbf{d}_{\text{obs}}$ is the observed data.

The least-squares solution described in equations 11 and 12 allows taking the modeling errors into account quite easily, as long as the modeling error can be described by a Gaussian probability density (Mosegaard and Tarantola, 2002; Tarantola, 2005). In that case, the mean (here zero) and covariance describing the measurement and modeling error combine through addition of the mean and covariances as

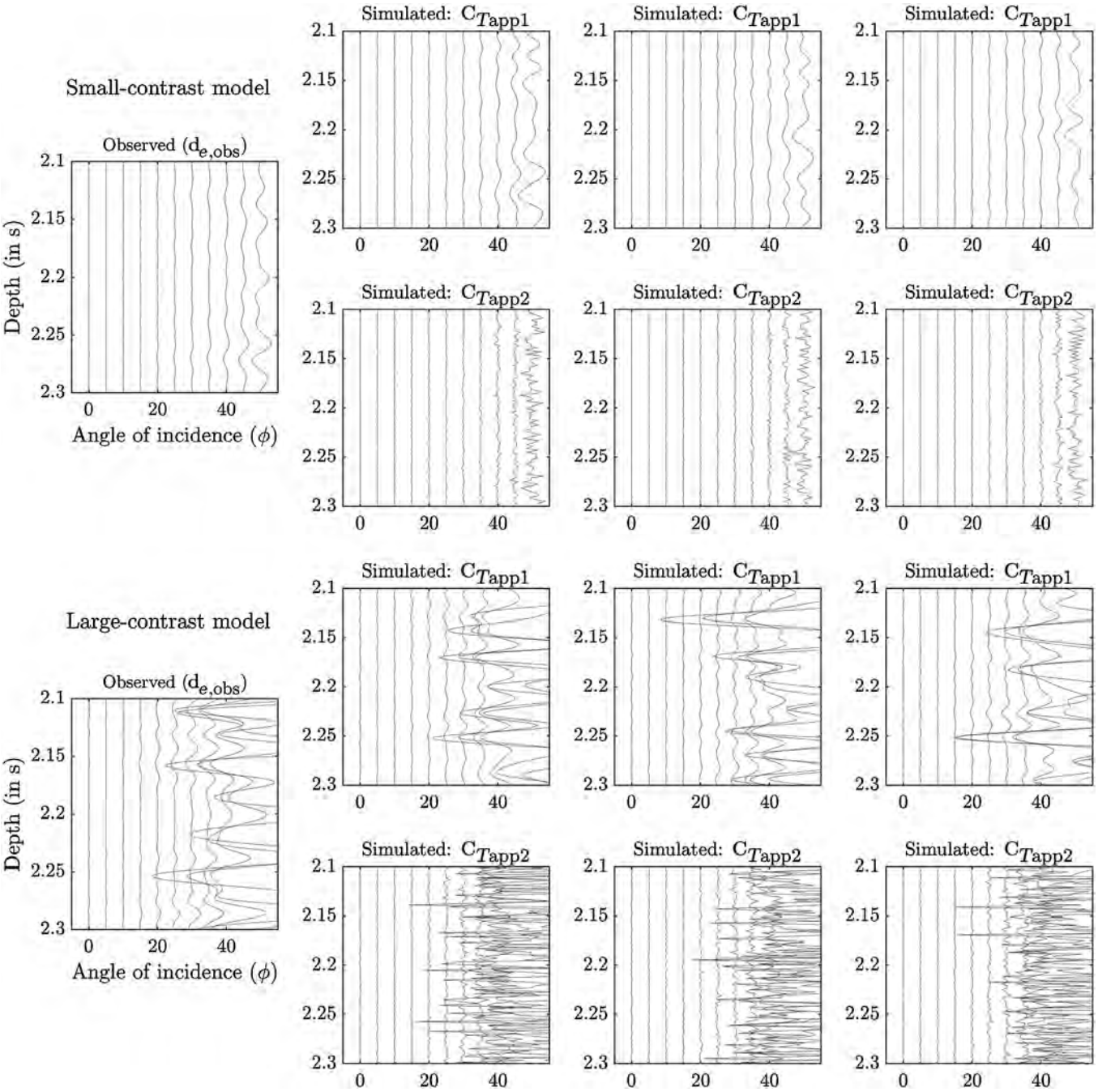


Figure 5. Observed modeling error (left column) plotted against three realizations from a Gaussian random field described by the estimated covariance matrices $\mathbf{C}_{T\text{app1}}$ and $\mathbf{C}_{T\text{app2}}$. The scaling is the same for all realizations.

$$\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_{T\text{app}}, \tag{13}$$

where $\mathbf{C}_d$ is the covariance matrix describing the measurement errors. In our case, we let the measurement error be represented by uncorrelated noise; i.e., $\mathbf{C}_d$ is the variance along the diagonal. As demonstrated previously, the linear Buland and Omre forward model is improved using the average angle as opposed to the incidence angle because the forward model is implicitly based on the average angle. AVO angle stacks can, in principle, be processed to reflect the average and incidence angles. But, as the offset-to-angle conversion is based on a typically smooth velocity field (in the present case, a constant field), a typical AVO angle gather will represent an approximation to either the actual incidence or the average angle. Therefore, we will in the remainder of the paper in reference to the linearized inverse problem, refer to an reflection angle. Through a numerical example, the effect of accounting for and discarding the modeling error will now be considered.

Bayesian linearized AVO inversion — Numerical example

The prior realizations in Figure 2 are used as two reference models $\mathbf{m}_{\text{ref}}$. A set of synthetic data is calculated using the Zoeppritz equations for 11 reflection angles (identical to Figure 1). Uncorrelated noise $\varepsilon \sim \mathcal{N}(0, \mathbf{C}_D)$ with $\text{S/N} = 5$ between the standard deviation of the forward response and the standard deviation of the noise is added to the data to obtain the "observed" data $\mathbf{d}_{\text{obs}}$.

The small-contrast model presents an ideal case for linearized AVO inversion because the variations in the elastic parameters are smooth (Gaussian), and it is known to be a realization of a Gaussian probability density with known mean and covariance, $\mathcal{N}(\boldsymbol{\mu}_M, \mathbf{C}_M)$. The a priori model is thus known. Because the exact noise model is also known, $\mathcal{N}(0, \mathbf{C}_D)$, the only unknown factor in the linearized AVO inversion method for the smooth prior is the effect of the modeling error.

The large-contrast model cannot be described fully by a Guassian model, which prohibits assessing the effect of the forward-modeling error in the linearized inversion directly. However, for comparison, a Gaussian prior model has been inferred that best matches the reference model using traditional semivariogram analysis. An exponential type of Gaussian distribution with 0.5 correlation coefficient between the elastic parameters and a range of 12 ms was found. The estimated variances for the elastic parameters are $\sigma^2_{v_{\text{P}}} = 0.01$, $\sigma^2_{v_{\text{S}}} = 0.05$, and $\sigma^2_{\rho} = 0.002$, respectively. In a real-world setting, a Gaussian model may also not be the obvious choice to describe the prior model, but the Gaussian prior model assumption is needed to make use of equations 11 and 12. Recently, Grana et al. (2017) propose a method that allows using Gaussian prior models with a non-Gaussian 1D marginal distribution. Sabeti et al. (2017) perform direct sequential simulation to allow using non-Gaussian 1D distributions. These methods may allow a better prior model describing models, such as the large-contrast model.

SMALL-CONTRAST MODEL

Figure 6 displays the results from the Bayesian linearized AVO inversion on $\mathbf{d}_{\text{obs}}$ from the small-contrast realization. Initially (the top figure), the modeling error is disregarded during inversion, i.e., $\mathbf{C}_D = \mathbf{C}_d$ in equations 11 and 12. This result in some features that seem to be well-resolved but that lie well beyond the 95% confidence interval and represent modeling errors being fitted as data. This bias happens especially where the contrasts in model parameters are fairly high, for instance, at approximately 2150–2220 ms.

The poorest performance is the posterior distribution of the density ρ, where serious bias effects are recognized for large parts of the posterior mean prediction compared with the reference model. This is a worrisome example of an apparently well-resolved feature, which is actually noise (modeling error) being fitted as data. Because the modeling error is increasing for far reflection angles, which are important for the density estimate, this could explain this bias. The P-wave velocity v_{P} and the velocity ratio $v_{\text{P}}/v_{\text{S}}$ are a bit better resolved, however still showing the noise being fitted as well-resolved features. The acoustic impedance and S-wave velocity are relatively well-resolved, but the reference model is still not fully captured by the uncertainty bands (confidence intervals). In fact, if the 95% confidence intervals were to be an accurate depiction of the uncertainty, one would expect the reference model to be contained within these uncertainty bands at approximately 95% of the samples, which is not the case. This makes the reference model a highly improbable realization of the posterior distribution obtained. Thus, if not accounted for, the modeling error related to the use of the linear Buland and Omre forward is able to create significant biases in inversion results for $\text{S/N} = 5$. The lower plots in Figure 6 show the corresponding inversion results accounting for the modeling error using equation 13 and the inferred covariance matrix $\mathbf{C}_{T\text{app1}}$ shown in Figure 4. The reference model generally lies within the 95% confidence interval. No artifacts are visibly present in the inversion results, and the modeling error seems to be properly accounted for. This is true even at approximately 2160–2180 ms, where the previous nonaccounting posterior prediction failed to resolve the reference model.

By calculating the root-mean-square deviation (rmsd) between the predicted values and the reference model, we can quantify how well each inversion scheme is predicting the reference model:

$$\text{rmsd} = \sqrt{\frac{\sum_{i=1}^{n}(\mathbf{m}_i - \tilde{\mathbf{m}}_i)^2}{n}}, \tag{14}$$

where $\mathbf{m}$ is the known reference values, $\tilde{\mathbf{m}}$ is the posterior prediction (mean), and n is the number of elements. Because the realization $\mathbf{m}_{\text{ref}}$ comes from a Gaussian distribution and the noise distribution is Gaussian, we can also quantify how likely $\mathbf{m}_{\text{ref}}$ is as a realization from the posterior distribution $\mathcal{N}(\tilde{\mathbf{m}}, \tilde{\mathbf{C}}_M)$ by

$$f(\mathbf{m}_{\text{ref}}|\mathcal{N}(\tilde{\mathbf{m}}, \tilde{\mathbf{C}}_M)) \sim \exp(-0.5(\mathbf{m}_{\text{ref}} - \tilde{\mathbf{m}})^{\top}\tilde{\mathbf{C}}_M^{-1}(\mathbf{m}_{\text{ref}} - \tilde{\mathbf{m}})). \tag{15}$$

Given a large sample of $\mathbf{m}_{\text{ref}}$, $\log(f)$ should follow a Gaussian distribution $\mathcal{N}(-N_m/2, \sqrt{N_m/2})$ if the degrees of freedom of the model parameters are sufficiently high (Hansen et al., 2016). In our case, the expected distribution is $\mathcal{N}(-150, \sqrt{150})$ because $N_m = 300$. In other words, the $\log(f)$ value should be in the interval of -150 ± 25 to be a likely realization from the posterior distribution $\mathcal{N}(\tilde{\mathbf{m}}, \tilde{\mathbf{C}}_M)$.

The quantitative measures are summarized in Table 1. The rmsd values and the $\log(f)$ value underline the visual results from Figure 6. By including the modeling error in the inversion, rmsd is reduced as much as 40%–50%. The largest improvement is seen

in the density prediction, which comply with the visual results that the poorest prediction of $\mathbf{m}_{\text{ref}}$ is offered for the density. Furthermore, the $\log(f)$ value for the case of $\mathbf{C}_D = \mathbf{C}_d$ reveals that $\mathbf{m}_{\text{ref}}$ is an very unlikely realization of the posterior distribution $\mathcal{N}(\tilde{\mathbf{m}}, \tilde{\mathbf{C}}_M)$. In other words, the posterior uncertainty does not capture the reference model. By accounting for the modeling error, $\mathbf{m}_{\text{ref}}$ can become a highly likely realization from that posterior distribution, as shown by the $\log(f)$ close to -150. By calculating the $\log(f)$ value for multiple realizations of $\mathbf{m}_{\text{ref}}$, it is determined that the result is not dependent on the specific realization used in this inversion (see Appendix C). In summary, the inclusion of the estimated modeling error in the linearized AVO inversion for the smooth prior does offer better predictions and a realistic uncertainty band.

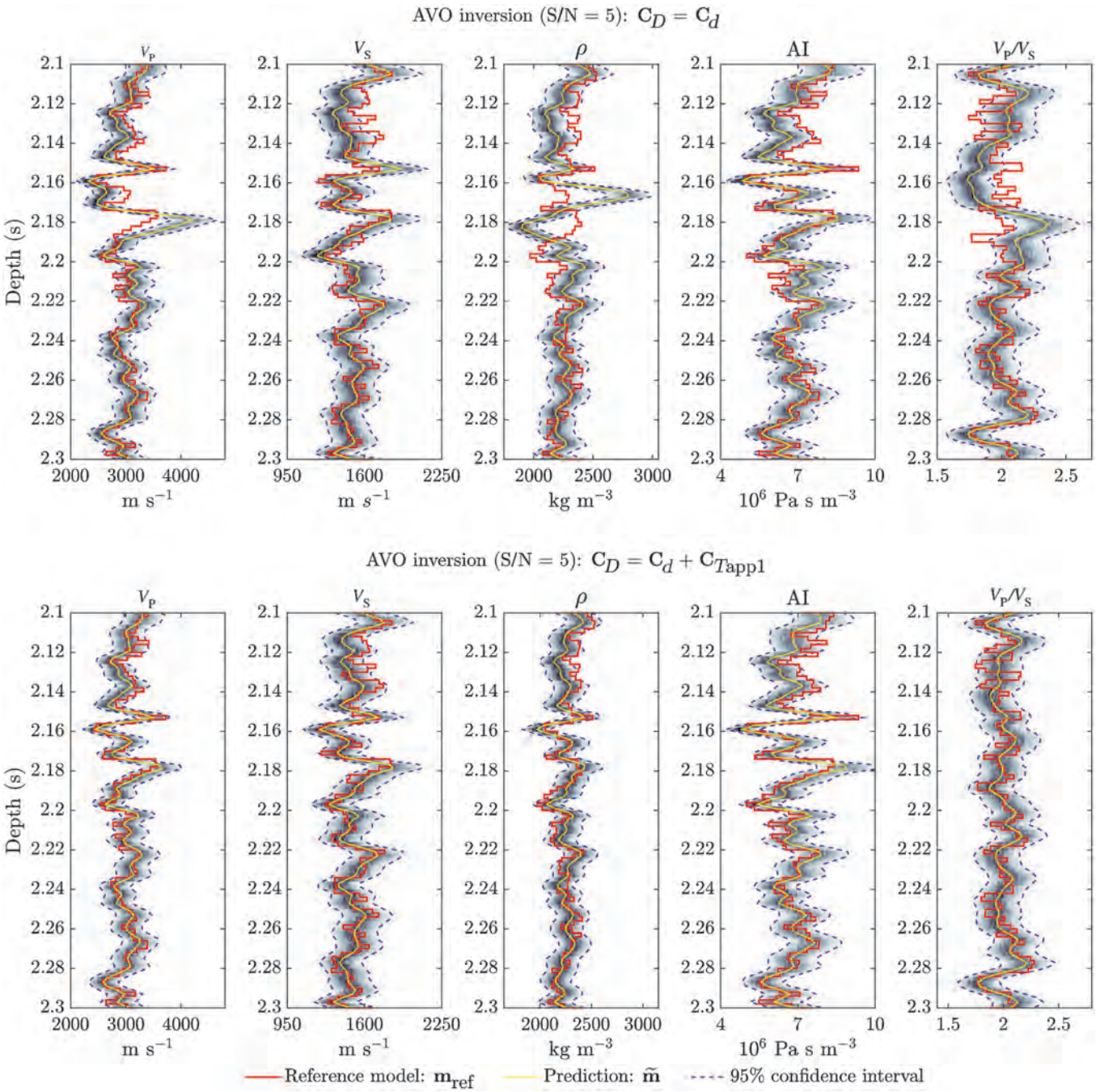


Figure 6. Linear Bayesian inversion performed on reference data calculated from the small-contrast reference model $\mathbf{m}_{\text{ref}}$ presented in Figure 2 (red line) with and without accounting for forward modeling errors. Posterior density (black = high density, white = low density), mean (yellow), and 95% confidence interval (dashed blue line) is shown. Before the inversion, the reference data were added random uncorrelated white noise with a standard deviation five times less than the standard deviation of the Zoeppritz forward response of the reference model; i.e., the $\text{S}/\text{N} = 5$.

LARGE-CONTRAST MODEL

Figure 7 displays the results from the Bayesian linearized AVO inversion on $\mathbf{d}_{\text{obs}}$ from the large-contrast realization and the estimated prior distribution described earlier. For the case of neglecting the forward-modeling error (top figure), i.e., $\mathbf{C}_D = \mathbf{C}_d$, the reference model $\mathbf{m}_{\text{ref}}$ is extremely poorly resolved. The reference model is rarely, if ever, within the 95% confidence interval. This happens despite the posterior distribution having a much lower resolution (higher uncertainty) than for the small-contrast prior inversion in Figure 6.

At certain depths, the predicted value is in fact opposite of the reference model. The predictions are for example consistently higher than the reference value for all elastic parameters at approximately 2150 ms. The high-prediction anomaly, which indicates a potential zone of interest, turns out to be an artifact from the inversion. The rmsd values in Table 2 show that predicted values are very far from the reference model for all elastic parameters, especially the S-wave velocity. The mismatch between the predicted values and the reference value is partly expected due to the use of the smooth Gaussian prior model to describe the abrupt changes in elastic properties of the subsurface realization. However, when comparing these results with the case of accounting for the modeling error $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_{T\text{app1}}$ (bottom figure), the wrong prior distribution can only explain a certain amount of the misfit. Accounting for the modeling error does, as for the small-contrast case, significantly improve the visual results and the rmsd values. Many of the major trends do seem to be captured by the posterior model, and the uncertainty bands often contain the reference model. Surprisingly, accounting for the modeling error even catches some of the major peaks in the $v_{\text{P}}/v_{\text{S}}$ ratio. The $\log(f)$ values are unsurprisingly low for both inversion cases as seen in Table 2 because we are fitting a discrete reference model with a smooth model. However, the large drop in $\log(f)$ when accounting for the modeling error further emphasizes the above-mentioned points.

DISCUSSION

Many types of inversion of AVO data has been proposed (Simmons and Backus, 1996; Buland and Omre, 2003; Vecken and Da Silva, 2004; Downton, 2005; Rabben et al., 2008; Wilson, 2010; Alemie and Sacchi, 2011; Aune et al., 2013; Grana, 2016). Usually, the data uncertainty is assumed to be Gaussian with zero mean and a known variance (Rabben et al., 2008; Singleton, 2009; Alemie and Sacchi, 2011; Aune et al., 2013; Grana, 2016). Sometimes, the modeling errors due to the choice of forward model are also considered as part of the general data uncertainty (e.g., Downton, 2005).

It has been proposed to account for systematic errors in Bayesian inversion related to "theory errors" (Riedel et al., 2003; Chen et al., 2007). This is achieved by adding a correlated covariance model describing the modeling uncertainty to the general uncorrelated data uncertainty. This covariance model is obtained using a fixed maximum likelihood estimate (Gerstoft and Mecklenbräuker, 1998). The shape of the correlated covariance modeling errors is in this case assumed to be proportional to the apparent covariance of the data. Alternatively, the data uncertainty including modeling errors has been added as an unknown parameter in a Bayesian inversion framework (Dosso and Holland, 2006).

The presented method of quantifying the forward-modeling error presents a straightforward workflow. In a Bayesian framework, a choice of prior must be performed in all circumstances. A sufficiently large sample of prior realizations and the subsequent modeling error can then be simulated to infer a Gaussian model describing the error. This Gaussian model can be added to the observational uncertainties, and the modeling errors are accounted for as long as the likelihood is Gaussian. Although the investigated errors are not strictly Gaussian, the inversion results indicate that the method offers a significant improvement compared with neglecting the error.

The key benefit of the proposed method is therefore that it allows a correct quantification of uncertainty while avoiding over-fitting the data, i.e., avoiding fitting noise. In addition, the proposed method avoids making any assumptions about the shape of the modeling error, other than it needs to be described by a Gaussian model. In comparison with the previously considered approaches by Riedel et al. (2003) and Chen et al. (2007), this method of quantifying the modeling error is independent of data. The quantification of the modeling error and thus the total data uncertainty can also be described prior to inversion as opposed to inverting for the magnitude of a correlated Gaussian model from data, as proposed by Dosso and Holland (2006).

Table 1. Small-contrast prior inversion. The rmsd calculated using equation 14 for the three elastic parameters v_{P}, v_{S}, and ρ divided by the average value to obtain the relative parameters $v_{\text{P}}/\overline{v_{\text{P}}}$, $v_{\text{S}}/\overline{v_{\text{S}}}$, $\rho/\bar{\rho}$, and the log(f) value calculated using equation 15. The values are calculated for both inversion cases ($\mathbf{C}_D = \mathbf{C}_d$ and $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_{T\text{app1}}$) shown in Figure 6.

	rmsd	rmsd	rmsd	
$\mathbf{C}_D$	$v_{\text{P}}/\overline{v_{\text{P}}}$	$v_{\text{S}}/\overline{v_{\text{S}}}$	$\rho/\bar{\rho}$	$\log(f)$
$\mathbf{C}_d$	0.0756	0.0760	0.0775	−1428.7
$\mathbf{C}_d + \mathbf{C}_{T\text{app1}}$	0.0444	0.0573	0.0377	−151.3

Handling significant forward-modeling errors for large reflection angle in Bayesian linearized inversion

The results indicate that the forward-modeling error of applying the Buland and Omre forward is significant for larger reflection angle, especially if the prior model has a large contrast in the elastic parameters. This is not surprising because the Buland and Omre forward model depends on a small contrast approximation. The modeling error posses a real concern for reflection angle $\phi > 20°$ if a non-Gaussian prior model is used. For the better case of a Gaussian prior model, our results are in agreement with earlier results suggesting the modeling error is negligible for data sets with $\phi < 30°$ (e.g., Shuey, 1985; Buland and Omre, 2003). Modern collection of seismic data allow processed AVO data to have a reflection angle well exceeding 30°. It is not uncommon that inversion is performed for data sets containing reflection angle of up to 50° (e.g., Barclay et al., 2008).

In practice, when dealing with AVO data with a large reflection angle, (at least) three possible approaches exist.

Solution 1: Neglect modeling error

One can neglect the possible modeling error of applying an imperfect forward and perform the inversion nonetheless. However, as shown, this strategy leads to significant biases in the posterior distribution compared with the true model. Most concerning is that apparently well-resolved posterior features can in fact, as demonstrated, be due to fitting modeling noise as though it was data. This can have a fatal effect on subsequent decision making. When random noise on the data is sufficiently large (small S/N), the theoretical errors drown in the random noise and the posterior distribution will not be biased significantly from the modeling error. Depending on the subsurface variability, the results in Appendix C suggest that a maximum of $S/N = 0.5$ is used, even for a smooth prior model when the Buland and Omre forward is used in Bayesian linearized AVO inversion for the considered model and configuration.

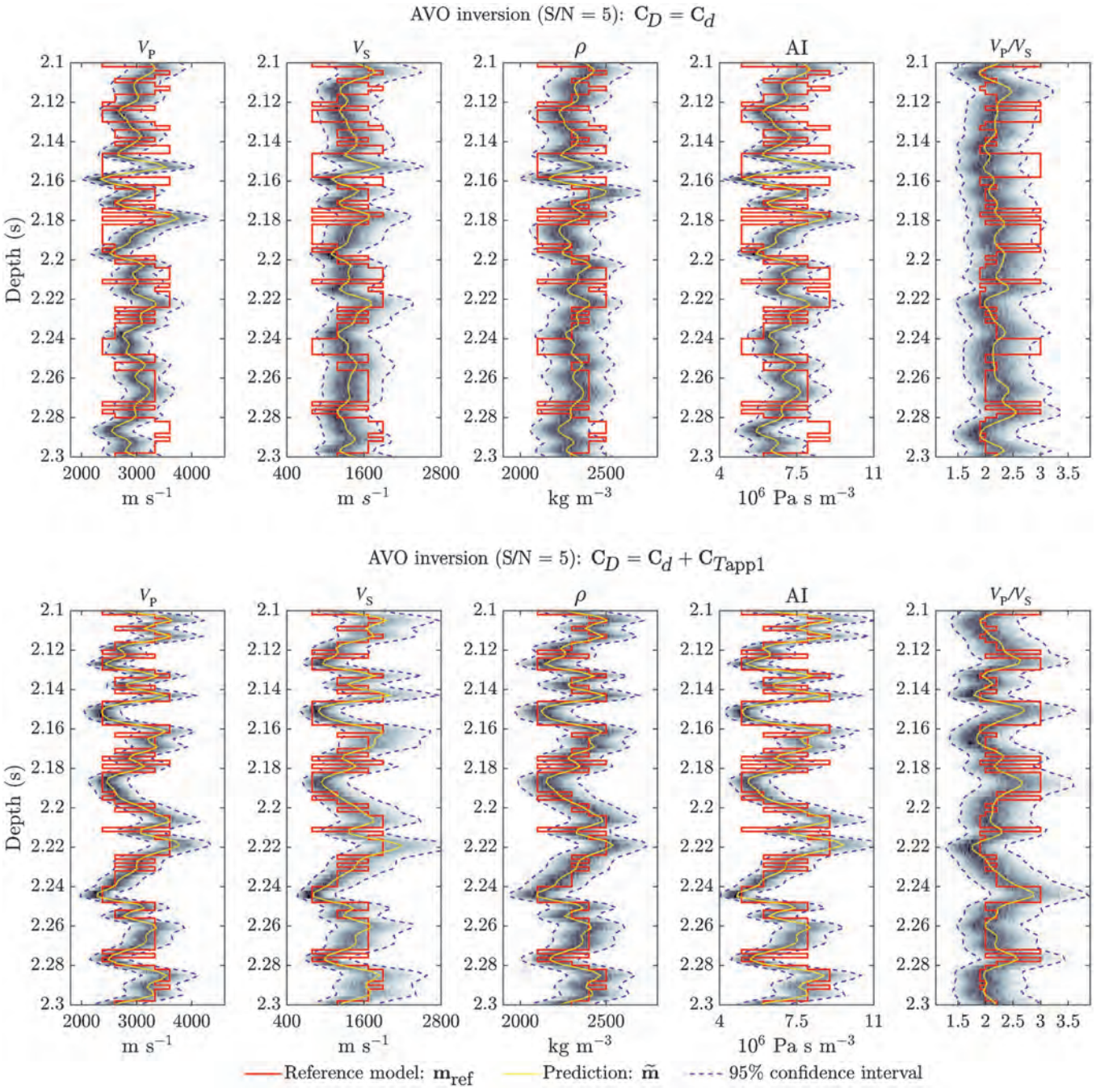


Figure 7. Linear Bayesian inversion performed on reference data calculated from the large-contrast reference model $\mathbf{m}_{ref}$ presented in Figure 2 (red line) with and without accounting for forward modeling errors. Posterior density (black = high density, white = low density), mean (yellow), and 95% confidence interval (dashed blue line) is shown. Before the inversion, the reference data were added random uncorrelated white noise with a standard deviation five times less than the standard deviation of the Zoeppritz forward response of the reference model; i.e., the $S/N = 5$.

Solution 2: Remove larger reflection angle

To avoid introducing significant forward-modeling errors, data from larger reflection angle are neglected (Stolt and Weglein, 1985). Two major issues arise using this strategy. First, seismic data, which are extremely difficult and expensive to acquire, are thrown away. Second, small reflection angles are mostly sensitive to changes in acoustic impedances, whereas larger reflection angles are potentially sensitive to changes in the S- and P-wave velocities (Shuey, 1985). By throwing away the wide reflection angle, significant information about the v_S and v_P/v_S ratio is lost.

Solution 3: Quantify and account for modeling error

Because the effect of the modeling error is significant for $S/N > 0.5$, good data with low noise are therefore explicitly an issue; i.e., the better the data, the bigger the biases in the posterior results. It should once again be stressed that these biases occur even in a best-case scenario with a small-contrast smooth prior. Because the magnitude of the modeling error is higher for the large-contrast prior, one can expect results with even more significant bias. By inferring a Gaussian model directly from a sample of the modeling error as demonstrated here, it has been shown that S/N less than 20 produce reasonable results. If the correlated part is ignored, a $S/N < 2$ has been shown to produce reasonable results (Appendix C). However, because the modeling error is correlated to a large degree, it is rather "naive" to only consider it to be uncorrelated. Our results suggest that the correlated part of the modeling contains significant information, which improves the inversion result.

Future work

In this paper, the focus has been on the modeling error related to the use of a linear approximation of Zoeppritz equations, especially the Buland and Omre forward. This is, as mentioned earlier, just one out of many potential uncertainties regarding AVO data. In the presented Bayesian linear inversion scheme, this modeling error is implicit in the scheme and hence unavoidable. It therefore represents a minimum level of modeling error, which is often neglected. The modeling error associated to for example ignoring anisotropy, imperfect NMO correction, or the use of uncertain wavelets (but to name a few significant sources of modeling errors) may very well be larger than the modeling error associated with the use of a linear small-contrast approximation to Zoeppritz equations, as argued by, e.g., Downton (2005). However, the examples demonstrated that modeling errors of these types can only be ignored when the S/N is less than 0.5, which is rarely the case in practice of marine seismic data.

The presented methodology can be extended to investigate, quantify, and possibly account for the effect of some of the other AVO modeling errors. Therefore, the next logical step would be to investigate and quantify the effect of the aforementioned sources of AVO modeling error. In particular, the modeling error associated with using the convolutional model as supposed to a NMO-corrected shot gather from waveform simulated seismics. The main requirement is that one must be able to provide a statistical description of the source of the modeling error, from which realizations of the modeling error can be computed. In the present case, the modeling error was linked to the subsurface variability. Further, the methodology can be used to quantify and account for modeling errors also in a full nonlinear, non-Gaussian inversion performed using, e.g., the extended metropolis sampler (see, e.g., Zunino et al., 2015).

A probabilistic Monte Carlo-based sampling strategy would for instance be able to sample the joint posterior distribution for the case of making use of the large-contrast prior (Mosegaard and Tarantola, 1995).

Table 2. Large-contrast prior inversion. The rmsd calculated using equation 14 for the three elastic parameters v_P, v_S, and ρ divided by the average value to obtain the relative parameters $v_P/\overline{v_P}$, $v_S/\overline{v_S}$, $\rho/\overline{\rho}$, and the log(f) value calculated using equation 15. The values are calculated for both inversion cases ($\mathbf{C}_D = \mathbf{C}_d$ and $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_{T\text{app1}}$) shown in Figure 7.

	rmsd	rmsd	rmsd	
$\mathbf{C}_D$	$v_P/\overline{v_P}$	$v_S/\overline{v_S}$	$\rho/\overline{\rho}$	$\log(f)$
$\mathbf{C}_d$	0.1867	0.3229	0.0636	−12893.0
$\mathbf{C}_d + \mathbf{C}_{T\text{app1}}$	0.1044	0.2008	0.0382	−282.9

CONCLUSION

We have simulated and quantified the modeling error related to using a linear approximate solution to Zoeppritz equations. The modeling error depends on the degree of subsurface variability and increases with angle of incidence. The Aki and Richards forward is shown to be less inaccurate than the Buland and Omre small-contrast forward, especially using the average angle as opposed to the incidence angle.

A Gaussian model describing this modeling error is inferred from an obtained sample of the modeling error. Realizations from this distribution resemble and mimic the observed values of the modeling error to a degree, where visual discrimination is difficult. A quantitative analysis reveals that the Gaussian model, while not perfect, provides a good description of the generated sample of the modeling error.

A Gaussian model of the modeling error is trivially accounted for as part of linearized AVO inversion. It has been shown that even small modeling errors from a linear forward model, related to a smooth Gaussian prior, can contribute to significant biases in inversion results. Apparently, well-resolved features in the posterior distributions may be caused by fitting modeling errors. In the case there is little subsurface variability expected (assuming a very smooth Gaussian-type a priori covariance model), our results indicate that the modeling error can be disregarded for the considered setup when the S/N of the AVO data is equal to or less than 0.5. Accounting for the modeling error, through the use of the inferred Gaussian model of modeling errors, improves the inversion results dramatically.

If more subsurface variability is expected, the effect of the modeling errors in inversion is even more severe. Our results indicate that accounting for the modeling error in this case provides a more reasonable prediction of the subsurface properties. Furthermore, it limits the amount of major artifacts in the inversion.

ACKNOWLEDGMENTS

We would like to thank H. J. Hansen at Qeye Labs for providing us with realistic properties and real world problems of discrete tran-

sitions in the North Sea. J. Downton for insightful comments during the review process. Calculations were performed using the MATLAB toolbox SIPPI (Hansen et al., 2013). This Ph.D. research is funded by the Innovation Fund Denmark (grant no. 53-2014-3).

APPENDIX A

ESTIMATING MODELING ERROR WITH A GAUSSIAN MODEL

Consider a sample consisting of N realizations of the modeling error $(\mathbf{d}_e^1, \mathbf{d}_e^2, \ldots, \mathbf{d}_e^N)$. This sample is set up in matrix form as

$$\mathbf{D}_e = [\mathbf{d}_e^1, \mathbf{d}_e^2, \ldots, \mathbf{d}_e^N]. \tag{A-1}$$

The mean-modeling error is estimated for each individual point j as

$$d_{T_{\text{app}}}^j = \frac{1}{N}\sum_{i=1}^{N} \mathbf{D}_e^{i,j}, \tag{A-2}$$

where the mean for the jth data point is the arithmetic mean of all $i = 1, \ldots, N$ realizations from the sample. The mean vector for all data points is then achieved by combining the mean of the individual data points calculated in equation A-2:

$$\mathbf{d}_{T\text{app}} = [d_{T_{\text{app}}}^1, d_{T_{\text{app}}}^2, \ldots, d_{T_{\text{app}}}^{N_d}]. \tag{A-3}$$

The covariance of the modeling error is estimated as

$$\mathbf{C}_{T\text{app}} = \frac{1}{N}[\mathbf{D}_e - \mathbf{D}_{T\text{app}}][\mathbf{D}_e - \mathbf{D}_{T\text{app}}]^\top, \tag{A-4}$$

where $\mathbf{D}_{T_{\text{app}}} = [\mathbf{d}_{T_{\text{app}}}^\top, \mathbf{d}_{T_{\text{app}}}^\top, \ldots, \mathbf{d}_{T_{\text{app}}}^\top]$ is a matrix containing N repetitions of the mean vector of equation A-3. The mean $\mathbf{d}_{T\text{app}}$ does in practice tend toward zero for large N as the modeling error is either negative or positive depending on the elastic parameters, which have a fixed mean. The $\mathbf{D}_{T\text{app}}$ can therefore be excluded from equation A-4 for large N.

APPENDIX B

QUANTITATIVE ASSESSMENT OF GAUSSIAN MODEL

A quantitative measure of the validity of the Gaussian assumption on the modeling error can be obtained by analyzing how probable the simulated modeling errors are as a realization of the inferred Gaussian model. The definition of the Gaussian model on a modeling error realization $\mathbf{d}_e$ is

$$f(\mathbf{d}_e|\mathcal{N}(\mathbf{d}_{T\text{app}}, \mathbf{C}_{T\text{app}})) \sim \exp(-0.5(\mathbf{d}_e - \mathbf{d}_{T\text{app}})^\top \mathbf{C}_{T\text{app}}^{-1}(\mathbf{d}_e - \mathbf{d}_{T\text{app}})), \tag{B-1}$$

$$= \exp(-0.5\ \Omega), \tag{B-2}$$

where $\Omega = (\mathbf{d}_e - \mathbf{d}_{T\text{app}})^\top \mathbf{C}_{T\text{app}}^{-1}(\mathbf{d}_e - \mathbf{d}_{T\text{app}})$ will then be distributed according to a χ^2-distribution with N_d number of degrees of freedom, which for large N_d tend to follow a Gaussian distribution $\Omega \sim \mathcal{N}(N_d, 2N_d)$ (Tarantola, 2005). This also means that $\log(f(\mathbf{d}_e|\mathcal{N}(\mathbf{d}_{T\text{app}}, \mathbf{C}_{T\text{app}}))) = -0.5\Omega$, for large N_d, will tend to follow a Gaussian distribution $\mathcal{N}(-N_d/2, \sqrt{N_d/2})$ (Hansen et al., 2016). Specifically, in our example, $\log(f)$ should follow the Gaussian distribution $\mathcal{N}(-550, \sqrt{550})$ because $N_d = 1100$.

Figure B-1 shows the histograms of $\log(f(\mathbf{d}_e|\mathcal{N}(\mathbf{d}_{T\text{app1}}, \mathbf{C}_{T\text{app1}})))$ for actual simulated modeling errors $(\mathbf{d}_{e,\text{obs}})$, realizations of $\mathcal{N}(\mathbf{d}_{T\text{app1}}, \mathbf{C}_{T\text{app1}})$ $(\mathbf{d}_{e,\text{app1}})$ and samples of $\mathcal{N}(\mathbf{d}_{T\text{app2}}, \mathbf{C}_{T\text{app2}})$ $(\mathbf{d}_{e,\text{app2}})$. The histogram of $\log(f(\mathbf{d}_{e,\text{app1}}))$ values (dark gray) follows the Gaussian distribution as described above, with a mean approximately -550. This is expected as we are comparing realizations from $\mathbf{C}_{T\text{app1}}$ with itself.

If the observed modeling error is described perfectly by the estimated Gaussian covariance model, the sample of the observed modeling error (black) would plot on top of the $\mathbf{C}_{T\text{app1}}$ sample (dark gray)

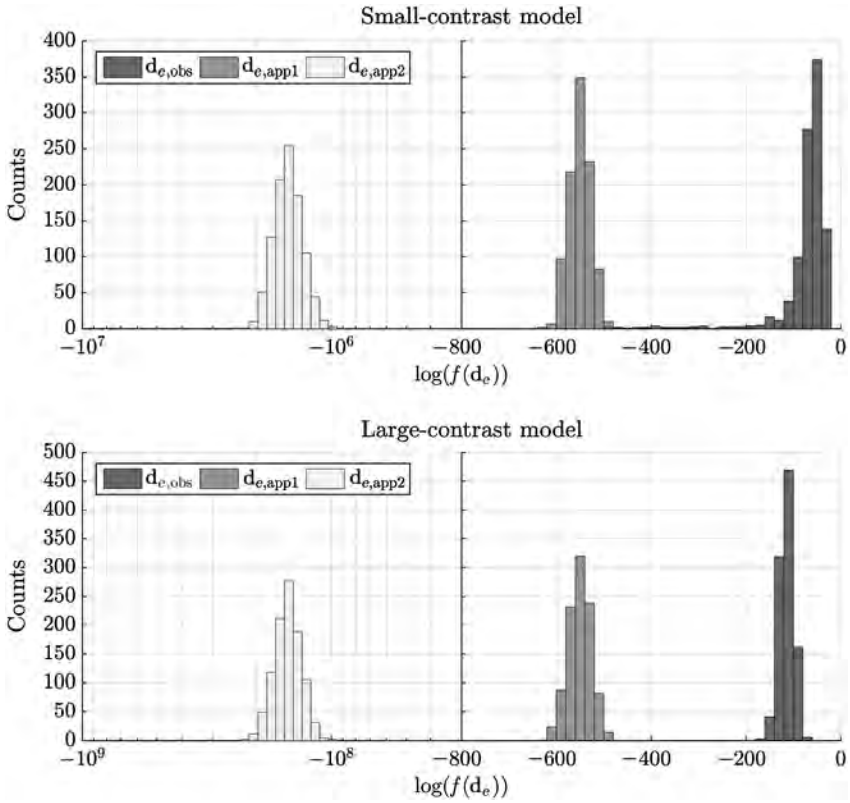


Figure B-1. Log-likelihood values calculated for 1000 realizations of the observed modeling error (black) and the two samples of $\mathbf{C}_{T\text{app1}}$ (dark gray), $\mathbf{C}_{T\text{app2}}$ (light gray), respectively. The results are shown for the two types of prior models, (top row) the small-contrast prior distribution and (base row) the large-contrast prior distribution. Notice the logarithmic scale used on the left side plots.

in Figure B-1. The histogram of the observed modeling error for the small-contrast prior is left-skewed with values mostly lying at approximately −60 with a tail in the range between −200 and −500. In the large-contrast case, a tail is not present, but the distribution is otherwise similar, with values of approximately −120. Such consistently higher log-likelihood values suggest that the sample values are consistently closer to the mean (zero) value than is expected from a normal distribution. For comparison, the log-likelihood histograms of realizations from the uncorrelated Gaussian model $\mathbf{C}_{T\text{app2}}$ (light gray) show values exceeding the previous two with several orders of magnitude. Therefore, the uncorrelated Gaussian model $\mathbf{C}_{T\text{app2}}$ is not a very good description of the actual noise sample, whereas the correlated Gaussian $\mathbf{C}_{T\text{app1}}$ while not perfect, does a much better job.

Figure B-2 shows the 1D marginal distribution of the observed modeling error $\mathbf{d}_{e,\text{obs}}$ (dashed black) compared with the estimated models below (Figure B-2a and B-2d) and greater than $\phi = 30°$ (Figure B-2b and B-2e) for the two subsurface models. The distribution of $\mathbf{d}_{e,\text{app1}}$ (dark gray) and $\mathbf{d}_{e,\text{app2}}$ (dotted light gray) shows identical distributions on both histograms because they are both from a Gaussian model with a similar variance. Most noticeable for the larger angles of incidence and for the large-contrast prior distribution in general, Figure B-2b, B-2d, and B-2e demonstrates that the observed modeling error produce a slimmer 1D distribution, with longer tails that cannot be completely described by the Gaussian model. This slight discrepancy to the Gaussian model explains the discrepancy observed in Figure B-1.

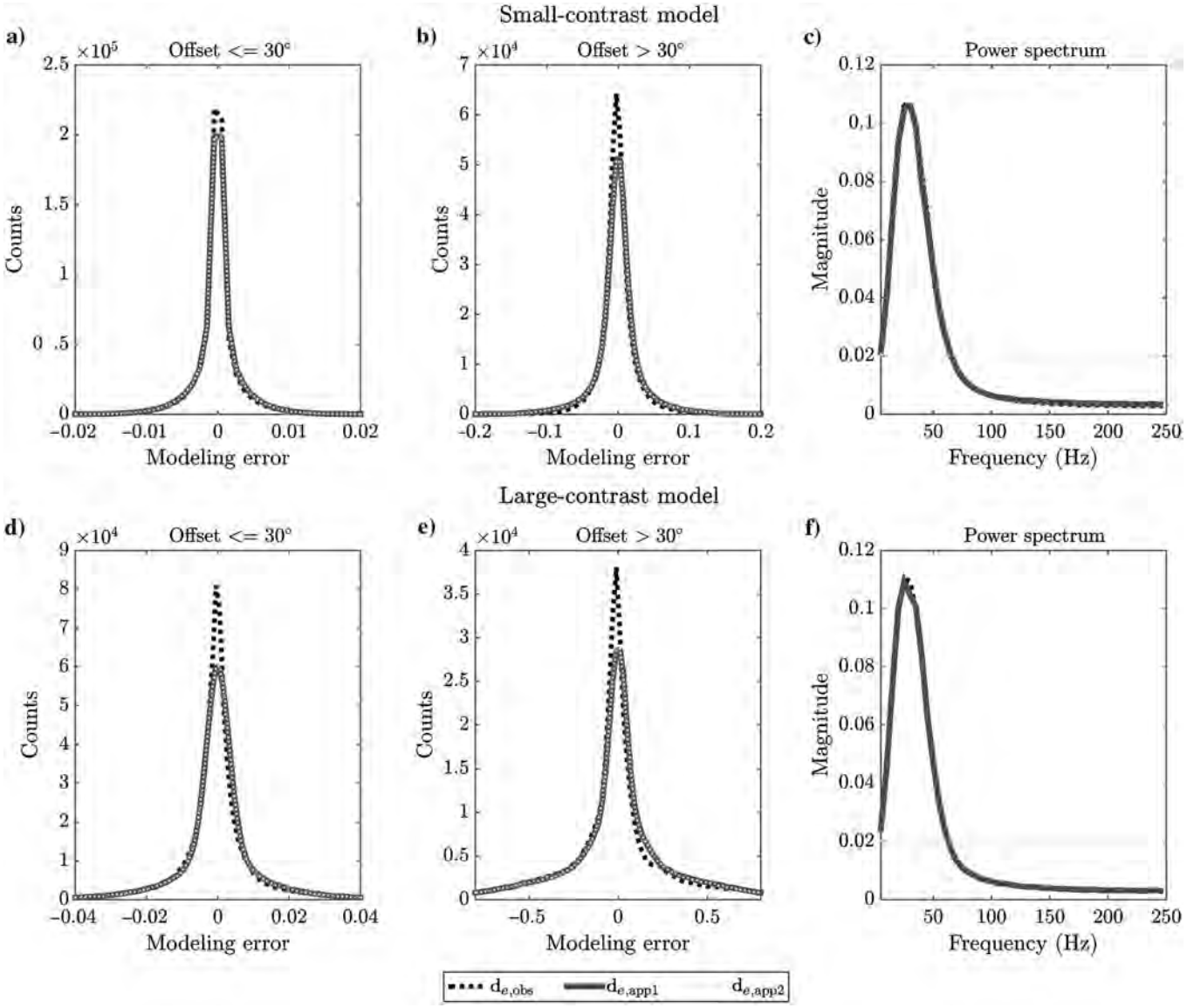


Figure B-2. (a, b, d, and e) Histogram of modeling error and (c and f) power spectrum shown for three samples consisting of 1000 realizations of the observed modeling errors $\mathbf{d}_{e,\text{obs}}$ (dashed black), $\mathbf{d}_{e,\text{app1}}$ (dark gray), and $\mathbf{d}_{e,\text{app2}}$ (light dots), respectively. The histograms and power spectrum are shown for (a-c) the small-contrast prior distribution and (d-f) the large-contrast prior distribution. The histograms of the modeling error are split between angles of incidence above and below $\phi = 30°$. The final column shows a normalized frequency spectrum of all angle of incidence traces.

In addition, a power spectral density (power spectrum) is calculated using a fast Fourier transformation on the three samples. The normalized power spectrum for the average of all incidence angles is shown in Figure B-2c and B-2f. Realizations from the estimated Gaussian model $\mathbf{d}_{e,\text{app1}}$ show a similar frequency pattern to the observed modeling error $\mathbf{d}_{e,\text{obs}}$, in which frequencies at approximately 30 Hz contain the most power. On the contrary, all frequencies, as expected, are equally represented for the sample of the uncorrelated Gaussian model $\mathbf{d}_{e,\text{app2}}$. This result underpins the qualitative result from Figure 5, where visual discrimination between the observed modeling error and realizations from $\mathbf{C}_{T\text{app1}}$ is difficult, if not impossible.

APPENDIX C

MULTIPLE INVERSION RESULTS

To determine whether the bias found in Figure 6 was an isolated or a more general problem, we repeat the inversion 1000 times, using 1000 different realizations of the small-contrast prior as reference models. The Zoeppritz forward response for each realization is calculated and noise is added, as described previously, to obtain 1000 observed reference AVO data sets.

If the estimated posterior Gaussian probability density adequately describes the solution to the inverse problem, then each of the 1000 reference models should be realistic realizations of the corresponding posterior Gaussian probability. This is quantified using equation 15 for each realization and the corresponding posterior distribution.

Figure C-1 shows histograms of the distribution of $\log(f)$ for pairs of 1000 reference models, and their corresponding Gaussian posterior distributions. Three different noise models ($\mathbf{C}_D$) are used for inverting the data. In the first case (left column), the noise model is equal to the observational uncertainty $\mathbf{C}_D = \mathbf{C}_d$ (i.e., modeling errors are ignored). For $S/N = 5$, the histogram is not following the expected Gaussian distribution (black line) and the $\log(f)$ tend to have much lower values. These outliers correspond to reference models that are inconsistent with the posterior distribution. Only a few prior realizations are found within the expected distribution. This confirms the result of heavy biases from the single prior realization in Figure 6. To avoid biases completely, the results from the likelihood histograms indicate that only $S/N = 1$ or less should be used if only the uncorrelated observational uncertainty is used as data uncertainty.

In the second case (the middle column), the data uncertainty consists of the observational uncertainty and the inferred Gaussian model for the modeling error $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_{T\text{app1}}$. For $S/N = 5$, the histogram is shifted to more or less fit inside the expected distribution. By accounting for the modeling error, the bias in results has more or less vanished. This confirms the visual results from Figure 6. The distribution of $\log(f)$ in Figure C-1 suggests that by accounting for the modeling error, it is possible to drastically improve inversion results in terms of the $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_{T\text{app1}}$ being an accurate quantification of the uncertainty. Even for $S/N = 20$, the inversion results indicate that the proposed data uncertainty is a relatively accurate description of the uncertainty.

Finally, the inversion is also performed with data uncertainty $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_{T\text{app2}}$, i.e., the assumption of uncorrelated modeling errors. The resulting histograms (the right column) show a similar distribution as for using $\mathbf{C}_D = \mathbf{C}_d$. In contrast to the first case, the distribution of $\log(f)$ generally attains higher values for $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_{T\text{app2}}$. In practice, this means that adding the uncorrelated Gaussian model, the inversion will provide a posterior distribution inconsistent with the reference model, though not to the same extent as for the observational uncertainties only. To avoid biases in the inversion results, it is here indicated that an $S/N = 2$ should not be exceeded. Because $\mathbf{C}_{T\text{app2}}$ is basically just white noise with differing variance, adding this to data uncertainty is somewhat comparable with just increasing the magnitude of the observational uncertainties. Therefore, the resolution (variance of the posterior) of $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_{T\text{app2}}$ is also lower than for the case of $\mathbf{C}_D = \mathbf{C}_d$. This could explain why the results are generally better because the lowered resolution allows for reference models to be a realistic realization from the inversion result.

At $S/N = 0.5$, the observational data noise is large enough such that the distribution of $\log(f)$ is similar for all three uncertainty

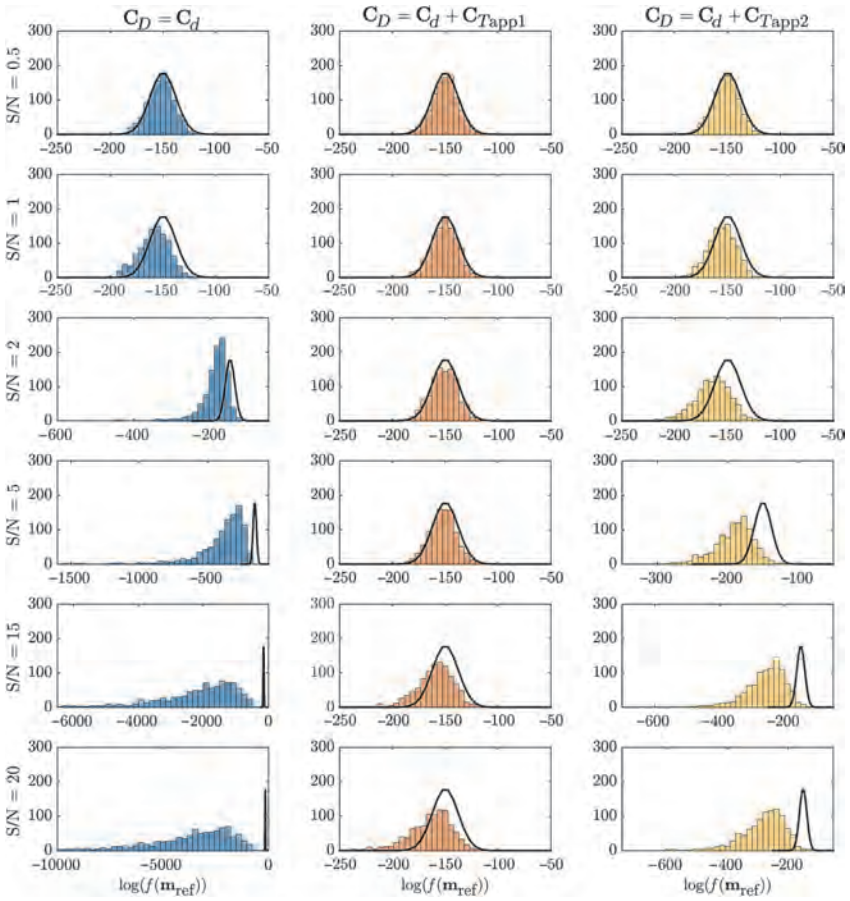


Figure C-1. Histograms of log-likelihood values according to equation 15 for a sample of 1000 prior realizations at different S/N levels. The linear Bayesian AVO inversion has been performed for three data uncertainty models $\mathbf{C}_D = \mathbf{C}_d$, $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_{T\text{app1}}$, and $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_{T\text{app2}}$.

models. The effect of the modeling error is therefore insignificant for S/N <= 0.5. This refers to unrealistic amounts of noise on the data and thus to where the resolution of the posterior models is poor. In other words, this demonstrates that the modeling error related to using the linearized Buland and Omre forward for reflection angle up to 50° can only be safely ignored if the S/N is extremely low.

REFERENCES

Aki, K., and P. G. Richards, 1980, Quantitative seismology, theory and methods: Vol I and II, 1st ed.: W.H. Freedman Publishing.

Alemie, W., and M. D. Sacchi, 2011, High-resolution three-term AVO inversion by means of a trivariate Cauchy probability distribution: Geophysics, **76**, no. 3, R43–R55, doi: 10.1190/1.3554627.

Armstrong, M., A. Galli, G. Le Loc'h, D. Renard, B. Deligez, F. Geffroy, and R. Eschard, 2011, Plurigaussian simulations in geosciences: Springer.

Aune, E., J. Eidsvik, and B. Ursin, 2013, Three-dimensional non-stationary and non-linear isotropic AVA inversion: Geophysical Journal International, **194**, 787–803, doi: 10.1093/gji/ggt127.

Ball, V., L. Tenorio, C. Schiott, J. P. Blangy, and M. Thomas, 2015, Uncertainty in inverted elastic properties resulting from uncertainty in the low-frequency model: The Leading Edge, **34**, 1028–1035, doi: 10.1190/tle34091028.1.

Barclay, F., A. Bruun, K. B. Rasmussen, J. C. Alfaro, A. Cooke, D. Cooke, D. Salter, R. Godfrey, D. Lowden, S. McHugo, H. Ozdemir, S. Pickering, F. G. Pineda, J. Herwanger, S. Volterrani, A. Murineddu, A. Rasmussen, and R. Roberts, 2008, Seismic inversion: Reading between the lines: Oilfield Review, **20**, 42–63.

Bosch, M., T. Mukerji, and E. F. Gonzalez, 2010, Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review: Geophysics, **75**, no. 5, 75A165–75A176, doi: 10.1190/1.3478209.

Buland, A., and H. Omre, 2003, Bayesian linearized AVO inversion: Geophysics, **68**, 185–198, doi: 10.1190/1.1543206.

Castagna, J. P., and M. M. Backus, 1993, Offset-dependent reflectivity: Theory and practice of AVO analysis: SEG.

Castagna, J. P., H. W. Swan, and D. J. Foster, 1998, Framework for AVO gradient and intercept interpretation: Geophysics, **63**, 948–956, doi: 10.1190/1.1444406.

Chen, M. A., M. Riedel, R. D. Hyndman, and S. E. Dosso, 2007, AVO inversions of BSRs in marine gas hydrate studies: Geophysics, **72**, no. 2, C31–C43, doi: 10.1190/1.2435604.

Claerbout, J. F., C. Green, and I. Green, 2004, Earth soundings analysis: Processing versus inversion: Stanford University.

Cooke, D., and W. A. Schneider, 1983, Generalized linear inversion of reflection seismic data: Geophysics, **48**, 665–676, doi: 10.1190/1.1441497.

Dosso, S. E., and C. W. Holland, 2006, Geoacoustic uncertainties from viscoelastic inversion of seabed reflection data: IEEE Journal of Oceanic Engineering, **31**, 657–671, doi: 10.1109/JOE.2005.858358.

Downton, J. E., 2005, Seismic parameter estimation from AVO inversion: Ph.D. thesis, University of Calgary.

Downton, J. E., and C. P. Ursenbach, 2006, Linearized amplitude variation with offset (AVO) inversion with supercritical angles: Geophysics, **71**, no. 5, E49–E55, doi: 10.1190/1.2227617.

Gerstoft, P., and C. F. Mecklenbräuker, 1998, Ocean acoustic inversion with estimation of a posteriori probability distributions: Journal of Acoustical Society America, **104**, 808–819, doi: 10.1121/1.423355.

Grana, D., 2016, Bayesian linearized rock-physics inversion: Geophysics, **81**, no. 6, D625–D641, doi: 10.1190/geo2016-0161.1.

Grana, D., T. Fjeldstad, and H. Omre, 2017, Bayesian Gaussian mixture linear inversion for geophysical inverse problems: Mathematical Geosciences, **49**, 1–37, doi: 10.1007/s11004-016-9658-6.

Hansen, T. M., K. S. Cordua, B. H. Jacobsen, and K. Mosegaard, 2014, Accounting for imperfect forward modeling in geophysical inverse problems: Exemplified for crosshole tomography: Geophysics, **79**, no. 3, H1–H21, doi: 10.1190/geo2013-0215.1.

Hansen, T. M., K. S. Cordua, M. C. Looms, and K. Mosegaard, 2013, SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information. Part 1: Methodology: Computers and Geosciences, **52**, 470–480, doi: 10.1016/j.cageo.2012.09.004.

Hansen, T. M., K. S. Cordua, A. Zunino, and K. Mosegaard, 2016, Probabilistic integration of geo-information, *in* M. Moorkamp, P. G. Lelièvre, N. Linde, and A. Khan, eds., Integrated imaging of the earth: Theory and applications, 1st ed.: John Wiley & Sons, 93–116.

Lay, T., and T. C. Wallace, 1995, Modern global seismology, 1st ed.: Academic Press.

Li, L., J. Caers, and P. Sava, 2015, Assessing seismic uncertainty via geostatistical velocity-model perturbation and image registration: An application to subsalt imaging: The Leading Edge, **34**, 1064–1070, doi: 10.1190/tle34091064.1.

Mavko, G., T. Mukerji, and J. Dvorkin, 2009, The rock physics handbook, 2nd ed.: Cambridge University Press.

Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, **100**, 12431–12447, doi: 10.1029/94JB03097.

Mosegaard, K., and A. Tarantola, 2002, 16 Probabilistic approach to inverse problems, *in* W. H. K. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger, eds., International handbook of earthquake and engineering seismology, Part A: Academic Press, 237–265.

Rabben, T. E., H. Tjelmeland, and B. Ursin, 2008, Non-linear Bayesian joint inversion of seismic reflection coefficients: Geophysical Journal International, **173**, 265–280, doi: 10.1111/j.1365-246X.2007.03710.x.

Riedel, M., S. E. Dosso, and L. Beran, 2003, Uncertainty estimation for amplitude variation with offset (AVO) inversion: Geophysics, **68**, 1485–1496, doi: 10.1190/1.1620621.

Sabeti, H., A. Moradzadeh, F. D. Ardejani, L. Azevedo, A. Soares, P. Pereira, and R. Nunes, 2017, Geostatistical seismic inversion for non-stationary patterns using direct sequential simulation and co-simulation: Geophysical Prospecting, doi: 10.1111/1365-2478.12502.

Sen, M. K., and P. L. Stoffa, 1996, Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion: Geophysical Prospecting, **44**, 313–350, doi: 10.1111/j.1365-2478.1996.tb00152.x.

Shuey, R., 1985, A simplification of the Zoeppritz equations: Geophysics, **50**, 609–614, doi: 10.1190/1.1441936.

Simmons, J. L., and M. M. Backus, 1996, Waveform-based AVO inversion and AVO prediction-error: Geophysics, **61**, 1575–1588, doi: 10.1190/1.1444077.

Singleton, S., 2009, The effects of seismic data conditioning on prestack simultaneous impedance inversion: The Leading Edge, **28**, 772–781, doi: 10.1190/1.3167776.

Stolt, R. H., and A. B. Weglein, 1985, Migration and inversion of seismic data: Geophysics, **50**, 2458–2472, doi: 10.1190/1.1441877.

Svendsen, J., E. Esmerode, H. Hansen, M. Lykker, H. Friis, and T. Stærmose, 2012, Unravelling the injected nature of a paleocene sand system: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts, 4.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation, 1st ed.: SIAM.

Tarantola, A., and B. Valette, 1982a, Generalized nonlinear inverse problems solved using the least squares criterion X = pt: Reviews of Geophysics and Space Physics, **20**, 219–232, doi: 10.1029/RG020i002p00219.

Tarantola, A., and B. Valette, 1982b, Inverse problems = quest for information: Journal of Geophysics, **50**, 159–170.

Thore, P., 2015, Uncertainty in seismic inversion: What really matters?: The Leading Edge, **34**, 1000–1004, doi: 10.1190/tle34091000.1.

Vecken, P. C. H., and M. Da Silva, 2004, Seismic inversion methods and some of their constraints: First Break, **22**, 47–70, doi: 10.3997/1365-2397.2004011.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, doi: 10.1190/1.3238367.

Wilson, A., 2010, Theory and methods of frequency-dependent AVO inversion: Ph.D. thesis, University of Edinburgh.

Zoeppritz, K., 1919, Über Reflexion und Durchgang seismischer Wellen durch Unstetigkeitsflächen: Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, **1919**, 66–84.

Zunino, A., K. Mosegaard, K. Lange, Y. Melnikova, and T. M. Hansen, 2015, Monte Carlo reservoir analysis combining seismic reflection data and informed priors: Geophysics, **80**, no. 1, R31–R41, doi: 10.1190/geo2014-0052.1.

H.5 Accounting for processing errors in AVO/AVA data

Rasmus Bødker Madsen, Egon Nørmark & Thomas Mejer Hansen, EAGE 80th annual meeting 2018, accepted 4 page expanded abstract

Introduction

Seismic data is an important source of information about geometry and properties of the subsurface. Rarely, if ever, are the raw seismic shot gathers used directly. Instead, some processing techniques are applied to transform the data into something useful for the specific seismic discipline (seismic interpretation, Amplitude Versus Offset (AVO) analysis, inversion, etc.). Seismic raw data are for instance typically stacked to reduce noise. Other processing techniques include; migration; gain correction; Normal Move-Out (NMO) correction; deconvolution; filtering; etc. In the following we focus on the processing of raw seismic data into Amplitude Versus Angle (AVA) seismic data for inversion of elastic parameters in the subsurface. An important, and perhaps the most important reason for processing raw seismic data into AVA/AVO data is that it allows the use of the popular convolutional model, were a seismic trace can be represented as a convolution between a reflectivity series and a wavelet (Yilmaz, 2001). Any successful processing of raw data would try to minimize the effect of features not explained by the convolutional model. Such features would in effect be noise/processing errors. We hereby refer to this the 'modeling error' from processing. For instance is the mapping from offset to angle prone to a large degree of uncertainty depending on conversion method (Mukhopadhyay and Mallick, 2011). In this extended abstract, we investigate the modeling error arising from processing raw seismic shot gathers to AVA gather in an ideal case with a known synthetic velocity model of the subsurface. Using the methodology presented in (Madsen and Hansen, 2018) the modeling error is quantified, estimated and used to account for processing errors in Bayesian linear inversion of seismic AVA data.

Method

To generate one realization of the modeling error we need two methods (forward models) of obtaining AVA data from an elastic log $\mathbf{m}$. For the present case, we have the processing sequence from raw seismic shot gathers to AVA data $\mathrm{g}_{\mathrm{proc}}$ and an approximate forward model consisting of a linearization of Zoeppritz equations plus a convolution with a wavelet $\mathrm{g}_{\mathrm{conv}}$. We have chosen the forward model $\mathrm{g}_{\mathrm{conv}} = \mathbf{G}$ similar to that presented in (Buland and Omre, 2003) as it allows a linear forward for subsequent linearized Bayesian inversion. The difference between these two results will be a realization of the modeling error $\mathbf{d}_{\mathrm{error}}$:

$$\mathbf{d}_{\mathrm{error}} = \mathbf{d}_{\mathrm{proc}} - \mathbf{d}_{\mathrm{conv}} = \mathrm{g}_{\mathrm{proc}}(\mathbf{m}) - \mathrm{g}_{\mathrm{conv}}(\mathbf{m}) = \mathrm{g}_{\mathrm{proc}}(\mathbf{m}) - \mathbf{G}\mathbf{m} \quad (1)$$

In order to generate a raw seismic shot gather from $\mathbf{m}$, a finite difference solver is used to do full waveform modeling. The solver is staggered grid fourth order space, second order time accuracy based on (Levander, 1988). Absorbing PML boundaries (Collino and Tsogka, 2001) are used for all boundaries, so a free surface is not considered. The reason for this is that a free surface boundary introduces significantly more modeling error (multiples, surface waves, etc.). Thus, what we present here will be an ideal case representing a lower limit of modeling errors (In practice modeling errors will be higher). The subsequent processing steps, and effect here of, can then be minimized. Furthermore, does the use of synthetic logs provide a known synthetic velocity model of the subsurface, which enable ideal conditions for processing. Finally, the wavelet can be extracted relatively exactly from the finite difference solver. The processing is done in Promax following the sequence described in table 1:

Processing sequence in Promax				
Load raw seismic data	Spherical divergence (constant velocity plus 6 dB/sec)	NMO-correction (with known background velocity model)	FK-filtering (apparent velocity = 1500 m/s)	Offset to angle conversion

Table 1 *Processing sequence (left to right) of raw seismic data to AVA data in Promax*

Notice that the processing sequence does not contain any noise suppression, and hence stacking routines, since the raw seismic data are basically noise free. Moreover, does the PML strategy allow an efficient suppression of the sea-bottom multiple, which further reduces the need for processing. The FK-filtering is done in order to remove S-wave energy within the raw data for larger offsets.

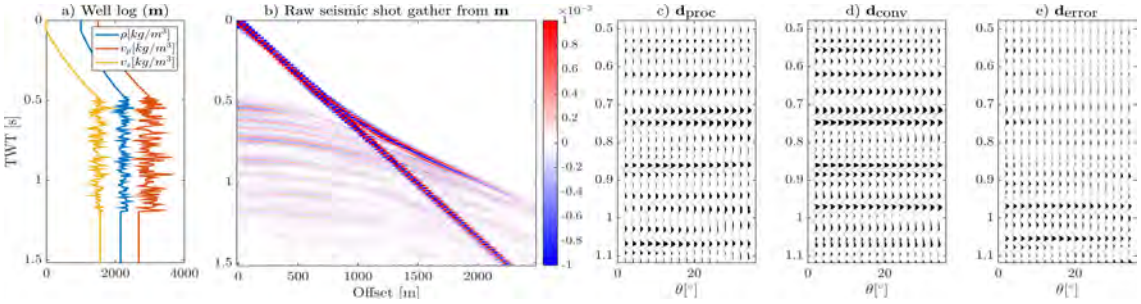


Figure 1 *a) Well log* **m** *of elastic parameters: v_p (yellow), v_s (red) and ρ (blue). b) Finite difference raw data. c-e)* $\mathbf{d}_{\text{proc}}$, $\mathbf{d}_{\text{conv}}$ *and* $\mathbf{d}_{\text{error}}$ *respectively, plotted as wiggle plots using the same scaling.*

Figure 1 shows an example of how one realization of $\mathbf{d}_{\text{error}}$ is calculated. First a log **m** is generated of the elastic parameters: P-wave velocity (v_p); S-wave velocity (v_s) and density (ρ). This is done combining a realization from the correlated Gaussian model described in (Buland and Omre, 2003) with a water column and a sediment layer down to around 0.5 ms which is depicted in Figure 1a. The log is slightly unrealistic in terms of the transition between sea-bottom and water/sediment. This has been done to avoid a large reflection from the sea-bottom and hence provide the best possible premise for subsequent processing. The raw seismic shot gather in Figure 1b is generated from **m** using the finite difference solver with absorbing boundary conditions. The effect of multiples is negligible as displayed in the resulting raw data. Figure 1c shows the AVA data $\mathbf{d}_{\text{proc}}$ from processing the raw data from figure 1b. The AVA data $\mathbf{d}_{\text{conv}}$ using the linear model convolutional model is shown in Figure 1d, where the wavelet is estimated directly from the finite difference solution. Finally, the modeling error realization $\mathbf{d}_{\text{error}}$, i.e. the difference between the two data is calculated using Eqn. 1 and is shown in Figure 1e. By comparing the standard deviation of all traces in figure 1c and 1e, the signal-to-noise ratio (S/N) between $\mathbf{d}_{\text{proc}}$ and $\mathbf{d}_{\text{error}}$ is calculated to 1,82.

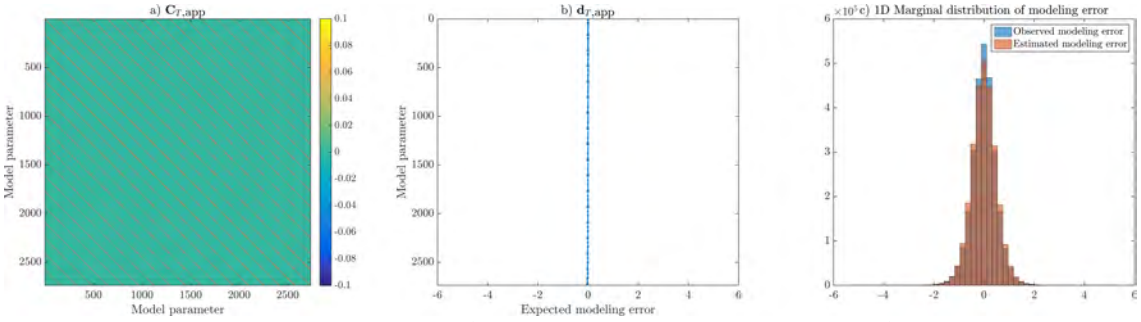


Figure 2 *a) Estimated covariance model for processing error:* $\mathbf{C}_{T,\text{app}}$. *b) Mean for estimated modeling error model:* $\mathbf{d}_{T,\text{app}}$ (blue) and variance (red). c) 1D marginal distribution of modeling error for observed and realizations from the estimated statistical model.

A statistical Gaussian model $\mathcal{N}\left(\mathbf{d}_{T,\text{app}}, \mathbf{C}_{T,\text{app}}\right)$ is estimated from a sample consisting of 1000 modeling error realizations using the methodology presented in (Hansen et al., 2014; Madsen and Hansen, 2018). For all 1000 realizations in the sample the average S/N ≈ 2 and the modeling error realization example in Figure 1 is a fair representation of the general level of modeling error. Figure 2a and 2b show the estimated covariance matrix $\mathbf{C}_{T,\text{app}}$ and mean vector $\mathbf{d}_{T,\text{app}}$ respectively. The estimated covariance show a strong correlated pattern of the modeling error between each of the angle traces. The modeling error variance is increasing with increasing reflection angle. The mean of the estimated modeling error is varying around zero. The 1D marginal distribution of the modeling error (Figure 2c) reveals that modeling errors of the estimated Gaussian model (orange) reasonably reflect the observed modeling errors (blue).

Inversion example

We invert a processed synthetic AVA dataset $\mathbf{d}_{\text{obs}} = \mathbf{g}_{\text{proc}}(\mathbf{m})$ using Bayesian linearized AVO

inversion equal to that presented in (Buland and Omre, 2003), where the posterior distribution $\tilde{\mathbf{m}}$ is described by a multivariate Gaussian distribution $\mathcal{N}(\tilde{\boldsymbol{\mu}}_M, \tilde{\mathbf{C}}_M)$. To assess the posterior distribution $\tilde{\mathbf{m}}$, three quantifiable measures are introduced. The entropy, H, describing the unpredictability of the posterior multivariate normal distribution can be calculated as:

$$\mathrm{H} = \frac{\mathrm{N}_M}{2} + \frac{\mathrm{N}_M}{2}\ln(2\pi) + \frac{1}{2}\ln(\det(\tilde{\mathbf{C}}_M)) \tag{2}$$

Where N_M is the number of model parameters (length of $\mathbf{m}$). The log-likelihood, logL, for the prediction $\tilde{\boldsymbol{\mu}}_M$ in terms of the posterior $\mathcal{N}(\tilde{\boldsymbol{\mu}}_M, \tilde{\mathbf{C}}_M)$ is calculated as:

$$\mathrm{logL} = -\frac{1}{2}(\mathbf{m} - \tilde{\boldsymbol{\mu}}_M)^{\mathrm{T}}\tilde{\mathbf{C}}_M^{-1}(\mathbf{m} - \tilde{\boldsymbol{\mu}}_M) \tag{3}$$

As argued in e.g. (Madsen and Hansen, 2018) the logL value should for large numbers of model parameters tend toward a value of $-\mathrm{N}_M/2$. For the current inversion case the value should be approximately -241.5. If so, the posterior distribution will be representative of the reference model. Finally, we use the root-mean-squared deviation (rmsd) to determine the accuracy of the prediction $\tilde{\boldsymbol{\mu}}_M$ compared with the true reference model $\mathbf{m}$. The processed data is inverted using three different noise models $\mathbf{C}_D$ and is shown in Figure 3. *Case 1, Estimated correlated noise:* We let the noise be described by the estimated Gaussian model ($\mathbf{C}_D = \mathbf{C}_{T,\mathrm{app}}$). In this inversion, the posterior distribution (Figure 3a-e) is representing the reference model (red line) and there are no apparent biases in the inversion result which is underlined by the low rmsd and logL value close to -241.5. Figure 3f shows the observed data (black) and the convoluted data (blue) for the given reference model. The posterior data response for ten posterior realizations (green) show that the data fit is poor with the observed data. Instead, the posterior realizations tend to follow the ideal convolution data response which is the forward model used for the inversion. I.e. the unwanted features from the processing are not fitted. *Case 2, Low uncorrelated noise:* An uncorrelated noise model is assumed, which is common inversion practice as it provides the highest entropy H for any Gaussian model. We set the level of uncorrelated noise with a S/N giving comparable results for the posterior entropy as for case 1. This corresponds to a S/N = 9. Here, some biases are apparent in the posterior distribution (Figure 3g-k). Figure 3l, show that the data misfit is very low as the posterior data (green) follows the observed data (black). This shows that unwanted processing features (noise) are fitted as data which results biases in the posterior. *Case 3, High uncorrelated noise:* Again, an uncorrelated Gaussian noise model is used, but now adjusted with increased noise level so the logL values is close to -241,5. This corresponds to a S/N = 0,4. Figure 3r shows that the increase in uncorrelated noise allow a larger misfit between the posterior data realizations and the observed data. This approach unfortunately lead to low posterior resolution (as seen in Figure 3m-q), and a much higher entropy than the corresponding two cases. The results from all three inversion cases are summarized in table 2.

	rmsd	H (Eqn. 2)	logL (Eqn. 3)
Case 1: $\mathbf{C}_D = \mathbf{C}_{T,\mathrm{app}}$	0,0306	-1135,08	-243,67
Case 2: $\mathbf{C}_D = \mathbf{C}_{d,1}$ (S/N=9)	0,0734	-1155.39	-50760,27
Case 3: $\mathbf{C}_D = \mathbf{C}_{d,2}$ (S/N=0.5)	0,0582	-789.98	-248,72

Table 2 *Root-mean squared deviation of prediction $\tilde{\boldsymbol{\mu}}_M$, Entropy of $\tilde{\mathbf{C}}_M$ and log-likelihood*

Conclusions

The inversion example suggests the need for reasonable handling of processing errors to avoid major biases and noise fitting in the inversion results as seen in Case 2. Case 3 demonstrate that in order to use an uncorrelated noise model, a S/N=0,4 must also be used just to account for the modeling error from processing. In this case, the low posterior resolution makes the inversion results practically unusable. In reality, a S/N would probably be chosen in between the two end-members of Case 2 and 3. Our proposed method allows handling of the processing error by estimating a statistical Gaussian model which can subsequently be used for inversion as in Case 1. By allowing for a larger misfit with

the observed processed data the inversion prediction is improved, which is suggested by the low rmsd value as compared with the two other cases. The posterior resolution is also more trustworthy than Case 2 as indicated by the logL value. Since the value is close to -241.5 the posterior distribution is representing the reference model. Our idealized logs with noise-free raw data allows us to consider a scenario with little processing needed to convert into AVA gathers. In a real-world case, the processing of actual seismic raw data would require more processing steps, especially to reduce levels of apparent noise and weakening of multiples. These additional processing steps would most probably increase the modeling error and hence further underpin conclusions above.

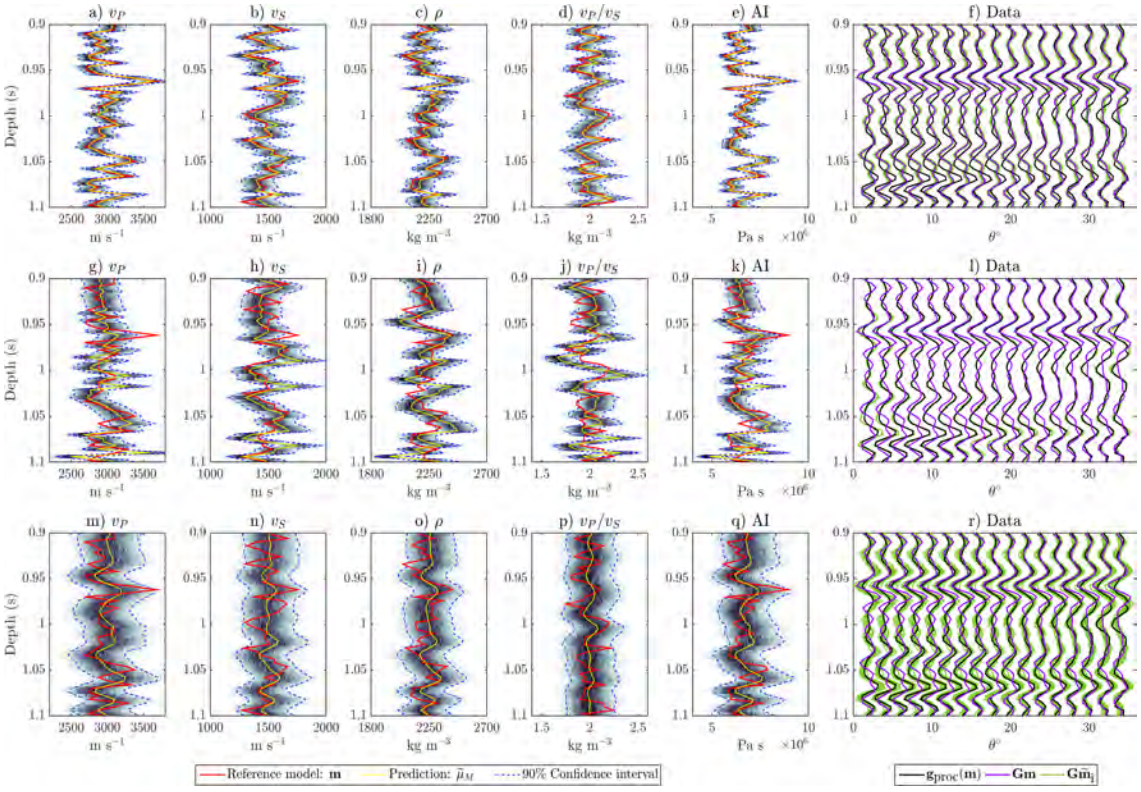


Figure 3 *Posterior distribution of elastic parameters, v_P, v_S, ρ, v_P/v_S and AI for the depth interval of 0,9 s – 1,1 s for three inversion cases. The mean model (yellow) is displayed along with the 90% confidence interval (dotted blue lines) and the reference model (red). On the right the data is shown as a wiggle plot with a-f) Case 1 g-l) Case 2 and m-r) Case 3*

References

Buland, A., Omre, H., 2003. Bayesian linearized AVO inversion. Geophysics 68, 185–198. https://doi.org/10.1190/1.1543206

Collino, F., Tsogka, C., 2001. Application of the PML absorbing layer model to the linear elastodynamic problem in anisotropic heteregeneous media. Geophysics 66, 294–307.

Hansen, T.M., Cordua, K.S., Jacobsen, B.H., Mosegaard, K., 2014. Accounting for imperfect forward modeling in geophysical inverse problems - Exemplified for crosshole tomography. Geophysics 79, H1–H21. https://doi.org/10.1190/GEO2013-0215.1

Levander, A.R., 1988. Fourth-order finite-difference P-W seismograms. Geophysics 53.

Madsen, R.B., Hansen, T.M., 2018. Estimation and accounting for the modeling error in probabilistic linearized AVO inversion. Geophysics 83, N15–N30. https://doi.org/10.1190/geo2017-0404.1

Mukhopadhyay, P.K., Mallick, S., 2011. Exact angle-mute pattern for a transversely isotropic medium with vertical symmetry axis and its implication in offset-to-angle transform. GEOPHYSICS 76, C41–C51. https://doi.org/10.1190/1.3628045

Yilmaz, O., 2001. Seismic Data Analysis, Tulsa: Society of exploration geophysicists. Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560801580

H.6 Prediction of acoustic impedance with non-stationary variance from seismic data

Rasmus Bødker Madsen, Thomas Mejer Hansen, & Henning Omre,
GEOPHYSICS (In review)

Prediction of acoustic impedance with non-stationary variance from seismic data

Rasmus Bødker Madsen[*], Thomas Mejer Hansen[*], and Henning Omre[†]

[*]*Niels Bohr Institute,*

University of Copenhagen,

Juliane Maries Vej 30,

2100, Copenhagen, Denmark

[†]*Department of Mathematical Sciences,*

Norwegian University of Science and Technology,

Alfred Getz' vei 1,

7034, Trondheim, Norway

(May 24, 2018)

Running head: **Non-stationary variance in seismic data**

ABSTRACT

Non-stationarity in statistical properties of the subsurface is often ignored. In a classical linear Bayesian inversion setting of seismic data, the prior distribution of physical parameters is often assumed to be stationary. Here we propose a new method of handling non-stationary in the variance of physical parameters in seismic data. We propose to infer the model variance prior to inversion using maximum likelihood estimators in a sliding window approach. A traditional, and a localized shrinkage estimator is defined for inferring the prior model variance. The estimators are assessed in a synthetic base case with heterogenous variance of the acoustic impedance in a zero-offset seismic cross section. Subsequently, this data is inverted for acoustic impedance using a non-stationary model set up with the inferred variances. Results indicate that prediction as well as posterior resolution is greatly improved using the non-stationary model compared with a common prior model with stationary variance. The localized shrinkage predictor is shown to be slightly more robust than the traditional estimator in terms of amplitude differences in the variance of acoustic impedance and size of local neighborhood. Finally, we apply the methodology to a real data set from the North Sea basin. Inversion results show a more realistic posterior model than using a conventional approach with stationary variance.

INTRODUCTION

Statistical properties of physical parameters for rocks (acoustic impedance, density, porosity, resistivity etc.) can generally be considered to be spatially non-stationary (Bouchedda et al., 2017; Sabeti et al., 2017). Seismic data are an important source of spatial information about the subsurface. Predictions of physical parameters from seismic data should ideally account for non-stationarity, as a correct prediction of the subsurface is of vital importance for locating and assessing hydrocarbon content, as well as possible migration paths. Recently, the incorporation of non-stationarity in seismic inversion has gained increasing interest in the academic literature. Aune et al. (2013) incorporates non-stationarity in the prior model by conditioning the model on nearby non-stationary well logs. Locally varying spatial correlation is also used to introduce non-stationarity in posterior solutions (Bongajum et al., 2013; Sabeti et al., 2017). By defining local likelihood models, Jullum and Kolbjørnsen (2016) approximates quantities related to rock properties in a Bayesian inversion setting. In the following we investigate spatial variability as represented by non-stationary variance.

The notation is as follows. Discrete matrices and vectors are denoted by bold fonts and their dimensions are written in front of the respective vector and matrix. The n-vector $\mathbf{i}_n$ is a unit-vector of dimension n and the diagonal $(n \times n)$-matrix $\mathbf{I}_n$ is an identity matrix of dimension $n \times n$. A probability density function (pdf) is denoted with $p(\cdot)$, $\sim$ means distributed as, $|$ means conditioned to, $|.|$ is the determinant and $\mathcal{N}_n$ is a normal distribution of dimension n (see Appendix A)

MODEL DESCRIPTION

In the current study we consider spatial non-stationarity in the statistical properties arising from heterogeneous variance in the spatial variable in the two-dimensional domain $\mathbb{R}^2$:

$$\{\mathrm{r}(\mathbf{x}) \in \mathfrak{D} \subset \mathbb{R}^2\} \tag{1}$$

discretized to a lattice $\mathfrak{L}_{\mathfrak{D}}$ of dimension $n_x \times n_y$ is,

$$\{\mathrm{r}(\mathbf{x}); \mathbf{x} \in \mathfrak{L}_{\mathfrak{D}}\} \tag{2}$$

represented by the n-vector $\mathbf{r}$ with $n = n_x \times n_y$. The observations (seismic data) are represented by the m-vector $\mathbf{d}$. We use a Bayesian spatial inversion framework to describe the problem of inferring information about $\mathbf{r}$ given $\mathbf{d}$. In this setting the characteristics of $\mathbf{r}$ is described by the posterior pdf $p(\mathbf{r}|\mathbf{d})$ and follows from the combination of prior pdf $p(\mathbf{r})$ and likelihood function $p(\mathbf{d}|\mathbf{r})$ through Bayes' theorem (see e.g. Box and Tiao (1992)),

$$p(\mathbf{r}|\mathbf{d}) = constant \times p(\mathbf{d}|\mathbf{r})p(\mathbf{r}). \tag{3}$$

The constant is a normalization constant. In the following we assume that the prior model is Gaussian and focus is on variability in the observable variable $\mathbf{r}$. Both the likelihood and prior models are dependent on the model parameters. The parameters of the likelihood are assumed known, while the prior model parameters $\boldsymbol{\theta}_p$ are considered to be unknown. We can then write Equation (3) as,

$$p(\mathbf{r}|\mathbf{d}; \boldsymbol{\theta}_p) = constant \times p(\mathbf{d}|\mathbf{r})p(\mathbf{r}; \boldsymbol{\theta}_p). \tag{4}$$

We assume that the prior model can be described by three model parameters $\boldsymbol{\mu}_r$, $\boldsymbol{\sigma}_r^2$, and $\rho_r(.)$. Both the expectation and the spatial correlation function $\rho_r(.)$ are assumed to be stationary known, i.e. independent of $\mathbf{x}$. The n-vector $\boldsymbol{\sigma}_r^2 = [\sigma_r^2(\mathbf{x}_1), \ldots, \sigma_r^2(\mathbf{x}_n)]$ is the

marginal variance at the nodes of the lattice $\mathfrak{L}_{\mathfrak{D}}$, is considered unknown and the spatial variability is represented by the non-stationary variance $\boldsymbol{\sigma}_r^2$. Consider the joint distribution of the variable $\mathbf{r}$ and data $\mathbf{d}$,

$$\begin{bmatrix} \mathbf{r} \\ \mathbf{d} \end{bmatrix} \sim p \left(\begin{matrix} \mathbf{r} \\ ;\boldsymbol{\theta}_p \\ \mathbf{d} \end{matrix} \right) = p(\mathbf{d}|\mathbf{r})p(\mathbf{r};\boldsymbol{\theta}_p) \tag{5}$$

with $\boldsymbol{\theta}_p = [\boldsymbol{\sigma}_r^2]$. There exist two challenges. Using the joint model we need to infer the model parameters $\boldsymbol{\theta}_p$ and to predict the physical parameters $[\mathbf{r}|\mathbf{d}]$. Maximum marginal likelihood (MML) estimators are used to infer $\boldsymbol{\theta}_p$:

$$\hat{\boldsymbol{\theta}}_p = \text{argmax}_{\hat{\boldsymbol{\theta}}_p}\{p(\mathbf{d};\boldsymbol{\theta}_p)\} \tag{6}$$

We formulate both a traditional MML and a hierarchical MML estimator. Prediction of $[\mathbf{r}|\mathbf{d}]$ is performed using the posterior model in Equation (4) with the inferred model parameter $\hat{\boldsymbol{\theta}}_p$ as a plugin estimate,

$$p(\mathbf{r}|\mathbf{d};\hat{\boldsymbol{\theta}}_p) = constant \times p(\mathbf{d}|\mathbf{r})p(\mathbf{r};\hat{\boldsymbol{\theta}}_p). \tag{7}$$

The strategy of inferring model parameters from data and subsequently use the data for prediction presents a philosophical issue. Prior model and observed data should in principle come form independent sources (Mosegaard and Tarantola, 2002). However, there is a long tradition in statistics for using data twice to both infer information of the statistical model and predict the outcome of an observable variable. The caveat here is, that this practice usually leads to underestimation of the variance.

Prior Model

We study the spatial variable $\{\mathrm{r}(\mathbf{x}) \in \mathfrak{D} \subset \mathbb{R}^2\}$, and focus on the effect of heterogeneous variances and consequently let the expected value be spatially stationary,

$$\mathrm{E}\{\mathrm{r}(\mathbf{x})\} = \mu_r(\mathbf{x}) = \mu_r; \quad \forall \mathbf{x} \in \mathfrak{D} \tag{8}$$

The variance is allowed to vary with the spatial reference,

$$\mathrm{Var}\{\mathrm{r}(\mathbf{x})\} = \sigma_r^2(\mathbf{x}); \quad \forall \mathbf{x} \in \mathfrak{D} \tag{9}$$

This allows assessing the effect of heterogenous variances in the prediction of $[\mathbf{r}|\mathbf{d}]$. The spatial correlation of the two points $\mathbf{x}_i$ and $\mathbf{x}_j$ is shift invariant and given by:

$$\mathrm{Corr}\{\mathrm{r}(\mathbf{x}_i), \mathrm{r}(\mathbf{x}_j)\} = \rho_r(\mathbf{x}_i - \mathbf{x}_j); \quad \forall \mathbf{x}_i, \mathbf{x}_j \in \mathfrak{D} \tag{10}$$

If the spatial correlation function is a positive definite function, the covariance function is also positive definite, since $\sigma_r(\mathbf{x}_i)$ and $\sigma_r(\mathbf{x}_j)$ are positive numbers,

$$\mathrm{Cov}\{\mathrm{r}(\mathbf{x}_i), \mathrm{r}(\mathbf{x}_j)\} = \sigma_r(\mathbf{x}_i)\sigma_r(\mathbf{x}_j)\rho_r(\mathbf{x}_i - \mathbf{x}_j); \quad \forall \mathbf{x}_i, \mathbf{x}_j \in \mathfrak{D}. \tag{11}$$

Consider the discrete representation of the spatial variable in the n-vector $\mathbf{r}$. The discretized Gaussian random field is then,

$$\mathbf{r} \sim p(\mathbf{r}; \boldsymbol{\theta}_p) = \mathcal{N}_n\{\mathbf{r}; \mu_r \mathbf{i}_n, \boldsymbol{\Sigma}_r\} \tag{12}$$

where μ_r is the constant expectation value and the covariance $(n \times n)$-matrix $\boldsymbol{\Sigma}_r$ represents the spatial correlation and variance. The covariance matrix $\boldsymbol{\Sigma}_r$ can be decomposed into the variance field and correlation matrix respectively,

$$\boldsymbol{\Sigma}_r = \boldsymbol{\Sigma}_r^{\sigma} \boldsymbol{\Sigma}_r^{\rho} \boldsymbol{\Sigma}_r^{\sigma} \tag{13}$$

where $\mathbf{\Sigma}_r^\sigma$ is a diagonal $(n \times n)$-matrix with diagonal values $\boldsymbol{\sigma}_r = [\sigma_r(\mathbf{x}_1), \cdots, \sigma_r(\mathbf{x}_n)]$. The correlation $(n \times n)$-matrix $\mathbf{\Sigma}_r^\rho$ is constructed by the spatial correlation function in Equation (10):

$$[\mathbf{\Sigma}_r^\rho]_{ij} = \rho_r(\mathbf{x}_i - \mathbf{x}_j) \quad \forall i, j = 1, 2, \cdots, n \tag{14}$$

Likelihood Model

We assume a Gauss-linear likelihood model describing the relationship between the data m-vector $\mathbf{d}$ and the n-vector $\mathbf{r}$:

$$[\mathbf{d}|\mathbf{r}] = \mathbf{H}\mathbf{r} + \mathbf{e}_d \sim p(\mathbf{d}|\mathbf{r}) = \mathcal{N}_m\{\mathbf{d}; \mathbf{H}\mathbf{r}, \mathbf{\Sigma}_{d|r}\} \tag{15}$$

where the $(m \times n)$-matrix $\mathbf{H}$ is the observation matrix (forward operator) while $\mathbf{e}_d$ is centered Gaussian with covariance $(m \times m)$-matrix $\mathbf{\Sigma}_{d|r}$. The covariance matrix is partitioned into a mixture of white and colored noise as proposed in Mosegaard and Tarantola (2002); Buland and Omre (2003); Madsen et al. (2017),

$$\mathbf{\Sigma}_{d|r} = \sigma_1^2 \mathbf{I}_m + \sigma_2^2 \mathbf{\Sigma}_{d|r}^\rho. \tag{16}$$

The first part corresponds to the uncorrelated white noise with variance σ_1^2. The second term represents the colored noise with variance σ_2^2 and correlation structure $\mathbf{\Sigma}_{d|r}^\rho = \mathbf{H}\mathbf{\Sigma}_r^\rho \mathbf{H}^\mathsf{T}$.

The linear forward operator $\mathbf{H}$ embodies the physics of the likelihood model. Seismic data are modeled as a linear convolution between a reflectivity series and a wavelet as described in e.g. Buland and Omre (2003). Here, we let the observable variable be relative acoustic impedances Z_P, i.e. the acoustic variable divided by the expected acoustic impedance. The relative acoustic impedances are henceforth simply known as acoustic impedances and consequently, $\mathrm{E}\{Z_P(\mathbf{x})\} = 1$ and unitless. The zero-offset plane-wave reflection coefficients are calculated using the procedure defined in Hampson et al. (2013),

which allows a linear relationship between the logarithm of acoustic impedance and the reflection coefficients at zero-offset as a 1D vertical convolution of dimension n_y. To obtain a full set of n_x seismic traces $\mathbf{d}$ from a 2D profile of $n_x \times n_y$ acoustic impedances, $\mathbf{H}$ is constructed as follows:

$$\mathbf{H} = \frac{1}{2}\begin{bmatrix} \mathbf{W} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{W} \end{bmatrix}\begin{bmatrix} \mathbf{D} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{D} \end{bmatrix} \tag{17}$$

where $\mathbf{W}$ is a matrix containing the wavelet and $\mathbf{D}$ is a differential operator, both set up for a single trace. The forward operator $\mathbf{H}$ then becomes an $(n \times n)$-matrix, meaning that the observable variable is available on the lattice $\mathfrak{L}_{\mathfrak{D}}$. The data is then a n-vector $\mathbf{d}$:

$$\{\mathbf{d}(\mathbf{x}); \mathbf{x} \in \mathfrak{L}_{\mathfrak{D}}\} \tag{18}$$

Joint Model

Since the prior and likelihood models are both Gaussian, the joint model (Appendix C) is given by,

$$\begin{bmatrix} \mathbf{r} \\ \mathbf{d} \end{bmatrix} \sim p\left(\begin{matrix} \mathbf{r} \\ ;\boldsymbol{\theta}_p \\ \mathbf{d} \end{matrix}\right) = p(\mathbf{d}|\mathbf{r})p(\mathbf{r};\boldsymbol{\theta}_p) \tag{19}$$

$$= \mathcal{N}_{n+m}\left\{\begin{bmatrix} \mathbf{r} \\ \mathbf{d} \end{bmatrix}; \begin{bmatrix} \mu_r \mathbf{i}_n \\ \mathbf{H}\mu_r \mathbf{i}_n \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_r & \boldsymbol{\Gamma}_{rd} \\ \boldsymbol{\Gamma}_{dr} & \boldsymbol{\Sigma}_d \end{bmatrix}\right\} \tag{20}$$

where the covariance $(n \times n)$-matrices $\boldsymbol{\Gamma}_{rd}$ and $\boldsymbol{\Gamma}_{dr}$ between the $\mathbf{r}$ and the observed data $\mathbf{d}$ are given by,

$$\boldsymbol{\Gamma}_{rd} = \boldsymbol{\Sigma}_r \mathbf{H}^{\mathsf{T}} \text{ and } \boldsymbol{\Gamma}_{dr} = \mathbf{H}\boldsymbol{\Sigma}_r \tag{21}$$

and the covariance of the data is given by,

$$\boldsymbol{\Sigma}_d = \mathbf{H}\boldsymbol{\Sigma}_r \mathbf{H}^{\mathsf{T}} + \boldsymbol{\Sigma}_{d|r}. \tag{22}$$

Using this joint model, one can infer the model parameter $\boldsymbol{\theta}_p = [\boldsymbol{\sigma}_r^2]$ and predict the physical parameter $[\mathbf{r}|\mathbf{d}]$.

SIMULATION SETUP DESIGN

We present the following synthetic Base Case (BC) for testing the ability to infer the variance of $\mathbf{r}$ and to predict $[\mathbf{r}|\mathbf{d}]$ given the inferred variances and the data. The BC consist of a grid with $n_x = 61$ seismic traces and $n_y = 81$ samples, giving $n = 61 \times 81 = 4941$ observable variables of acoustic impedance. This grid is considered the reference domain $\mathfrak{L}_{\mathfrak{D}}$.

Base case - Prior Model

For the BC the expectation is,

$$\mathrm{E}\{\mathrm{r}(\mathbf{x})\} = \mathrm{E}\{\ln(Z_P(\mathbf{x}))\} = \mu_{\mathrm{r}} = 0. \tag{23}$$

The reference variance $\sigma_r^2 = 0.001$. The standard deviation vector $\boldsymbol{\sigma}_r$ is scaled with a heterogenous scaling $(n_x \times n_y)$-matrix $\mathbf{S}$ as

$$\boldsymbol{\sigma}_r = [\sigma_r \mathbf{S}(\mathbf{x}_1), \cdots, \sigma_r \mathbf{S}(\mathbf{x}_n)]. \tag{24}$$

The chosen matrix $\mathbf{S}$ scales the reference standard deviation between one and five times and is displayed in Figure 1a. The spatial correlation is described by an anisotropic spherical correlation function,

$$\rho_r(\mathbf{x}_i - \mathbf{x}_j) = \rho_r(h) = \begin{cases} 1 - \frac{3h}{2a} - \frac{3h^3}{2a^3} & \text{if } h \leq a \\ 0 & \text{if } h > a \end{cases} \tag{25}$$

where a is the range parameter and h is the normalized distance between two locations. This model is positive definite. Geological layers in the subsurface are typically more con-

tinuous in the horizontal than vertical direction. To induce anisotropy into the spherical model we introduce an anisotropy factor κ between the horizontal (u) and vertical (v) component of h, so $h = \sqrt{(h_u)^2 + \kappa(h_v)^2} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)_u^2 + \kappa(\mathbf{x}_i - \mathbf{x}_j)_v^2}$. In the BC, $a = 5$ and the anisotropy factor $\kappa = 5$, creating elongated structures in the horizontal direction. A reference realization $\mathbf{r}$ (Equation 12) from the BC prior model is generated using Cholesky decomposition and is displayed in Figure 1b.

Base case - Forward Model

A Ricker-wavelet (Aki and Richards, 1980) with center frequency $f_c = 40$ Hz, with sampling rate 2 ms, wavelet length $n_w = 21$, and normalized amplitude is used to set up the wavelet matrix $\mathbf{W}$ which is displayed in Figure 1c. The resulting observed zero-offset seismic signal is calculated for all traces by $\mathbf{Hr}$ without adding noise, and can be seen in Figure 1d as a wiggle plot.

Base case - Noise Model

The synthetic data is added noise $\mathbf{e}_d$ with signal-to-noise ratios (S/N) of 15 and 3 for the uncorrelated ($\mathrm{S/N_{unc}}$) and correlated ($\mathrm{S/N_{cor}}$) noise respectively. The variance of the noise in Equation 16 is here computed by dividing the reference variance with the desired S/N,

$$\sigma_1^2 = \frac{\sigma_r^2}{\mathrm{S/N_{unc}}} \quad \text{and} \quad \sigma_2^2 = \frac{\sigma_r^2}{\mathrm{S/N_{cor}}}.$$

The low level of uncorrelated noise for the synthetic BC is a fairly realistic depiction of the general level of uncorrelated noise in seismic data. Uncorrelated noise can typically be filtered out during processing of seismic data using simple frequency filtering techniques (Vecken and Da Silva, 2004). The remaining noise in seismic data will therefore usually

$max(\mathbf{S})$	f_c(Hz)	Samp. rate (ms)	n_w	$\mathrm{S/N_{cor}}$	$\mathrm{S/N_{unc}}$
5	40	2	10	3	15

Table 1: Setup of parameters in BC.

tend to resemble the frequency content of the signal (Madsen et al., 2017). The shape of the colored noise is therefore assumed to imitate data as $\mathbf{H}\boldsymbol{\Sigma}_r\mathbf{H}^\mathsf{T}$. A noise realization $\mathbf{e}_d$ is drawn from the resulting noise distribution in Figure 1e. The final synthetic data set $\mathbf{d}$ is shown in Figure 1f. The parameters used in the BC are summarized in Table 1.

MODEL PARAMETER INFERENCE

The model parameters $\boldsymbol{\theta}_p = [\boldsymbol{\sigma}_r^2]$ are inferred using MML estimators. The variance should be estimated locally due to the imposed non-stationarity in the variance. In the following we assume that both mean and spatial correlation function are known a priori. This allows us to only assess the effect of spatial non-stationarity in the variance, $\{\sigma_r^2(\mathbf{x}); \mathbf{x} \in \mathfrak{D}\}$.

In order to locally estimate the variance we first define a local neighbourhood Δ for inference. Consider an arbitrary location $\mathbf{x}_0$ in $\mathfrak{D}$. In the data domain the local neighbourhood is symmetric around $\mathbf{x}_0$ including data points from nearby traces as displayed in Figure 2a. We denote the resulting subset of the data n_d^Δ-vector $\mathbf{d}_\Delta$ with $n_d^\Delta = 11$. Since each datum is affected by several spatial variables through the 1D convolution in the forward operator $\mathbf{H}$, the corresponding neighbourhood in $\mathbf{r}$ covers a larger area as displayed in Figure 2b. The size of this neighbourhood intrinsically depends on the width of the wavelet n_w. The resulting subset of the spatial variable n_r^Δ-vector is denoted $\mathbf{r}_\Delta$ with $n_r^\Delta = 121$.

Traditional estimator

In the following we define a traditional MML estimator within the local neighbourhood Δ. The prior and joint models follow analogous to the global prior and joint models, described earlier. The local prior model is,

$$\mathbf{r}_\Delta \sim p(\mathbf{r}_\Delta; \boldsymbol{\theta}_{p\Delta}) = \mathcal{N}_{n_r^\Delta}\{\mathbf{r}_\Delta; \mu_r \mathbf{i}_{n_r^\Delta}, \boldsymbol{\Sigma}_{r\Delta}\} \tag{26}$$

with the local joint model between data and observable variables,

$$\begin{bmatrix} \mathbf{r}_\Delta \\ \\ \mathbf{d}_\Delta \end{bmatrix} \sim p\left(\begin{matrix} \mathbf{r}_\Delta \\ ; \boldsymbol{\theta}_{p\Delta} \\ \mathbf{d}_\Delta \end{matrix} \right) = p(\mathbf{d}_\Delta|\mathbf{r}_\Delta) p(\mathbf{r}_\Delta; \boldsymbol{\theta}_{p\Delta}) \tag{27}$$

$$= \mathcal{N}_{n_r^\Delta + n_d^\Delta} \left\{ \begin{bmatrix} \mathbf{r}_\Delta \\ \\ \mathbf{d}_\Delta \end{bmatrix} ; \begin{bmatrix} \mu_r \mathbf{i}_{n_r^\Delta} \\ \\ \mathbf{H}_\Delta \mu_r \mathbf{i}_{n_r^\Delta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{r\Delta} & \boldsymbol{\Gamma}_{rd\Delta} \\ \\ \boldsymbol{\Gamma}_{dr\Delta} & \boldsymbol{\Sigma}_{d\Delta} \end{bmatrix} \right\}. \tag{28}$$

Within this Δ neighbourhood we assume stationarity of the variance σ_Δ^2. The local standard deviation $(n_r^\Delta \times n_r^\Delta)$-matrix $\boldsymbol{\Sigma}_{\mathrm{r}\Delta}^\sigma$ is then equal to $\sigma_\Delta \mathbf{I}_{n_r^\Delta}$. The local prior covariance follows from Equation (13) and is given by,

$$\boldsymbol{\Sigma}_{r\Delta} = \boldsymbol{\Sigma}_{r\Delta}^\sigma \boldsymbol{\Sigma}_{r\Delta}^\rho \boldsymbol{\Sigma}_{r\Delta}^\sigma = \sigma_\Delta^2 \boldsymbol{\Sigma}_{r\Delta}^\rho \tag{29}$$

and the local covariance of the data given by,

$$\boldsymbol{\Sigma}_{d\Delta} = \mathbf{H}_\Delta \boldsymbol{\Sigma}_{r\Delta} \mathbf{H}_\Delta^\mathsf{T} + \boldsymbol{\Sigma}_{d|r\Delta} \tag{30}$$

Combining Equation (30) with the noise model in Equation (16) yields,

$$\boldsymbol{\Sigma}_{d\Delta} = \mathbf{H}_\Delta \boldsymbol{\Sigma}_{r\Delta} \mathbf{H}_\Delta^\mathsf{T} + \sigma_1^2 \mathbf{I}_{n_d^\Delta} + \sigma_2^2 \boldsymbol{\Sigma}_{d|r\Delta}^\rho. \tag{31}$$

Since seismic data are usually mainly affected by correlated noise, the uncorrelated noise can be left out of Equation (31). If we furthermore assume a proportionality c between the

variance of the data σ_2^2 and the local prior variance σ_Δ^2,

$$\sigma_2^2 = c\sigma_\Delta^2 \tag{32}$$

the local covariance of the data can be described as,

$$\mathbf{\Sigma}_{d\Delta} = \sigma_\Delta^2(\mathbf{H}_\Delta \mathbf{\Sigma}_{r\Delta}^{\rho} \mathbf{H}_\Delta^{\mathsf{T}} + c\mathbf{\Sigma}_{d|r\Delta}^{\rho}) = \sigma_\Delta^2 \mathbf{\Omega} \tag{33}$$

where $\mathbf{\Omega} = \mathbf{H}_\Delta \mathbf{\Sigma}_{r\Delta}^{\rho} \mathbf{H}_\Delta^{\mathsf{T}} + c\mathbf{\Sigma}_{d|r\Delta}^{\rho}$. This assumption of proportionality basically means, that the amount of noise in the data is coupled to the variance in $\mathbf{r}$. If for example the variance of $\mathbf{r}$ is changing drastically due to a fault, we expect the noise on the date to change accordingly. The local marginal likelihood of the data, assuming μ_r and $\rho_r(.)$ known, is then Gaussian with the following distribution from Equation (28),

$$\mathbf{d}_\Delta \sim p(\mathbf{d}_\Delta; \sigma_\Delta^2) = \mathcal{N}_{n_d^\Delta}(\mathbf{d}_\Delta; \mathbf{H}_\Delta \mu_r \mathbf{i}_{n_r^\Delta}, \sigma_\Delta^2 \mathbf{\Omega}). \tag{34}$$

The variance can be inferred locally by finding the variance σ_Δ^2 that maximizes the local marginal likelihood,

$$\hat{\sigma}_\Delta^2 = \text{argmax}_{\sigma_\Delta^2}\{p(\mathbf{d}_\Delta; \sigma_\Delta^2)\}. \tag{35}$$

The maximum of the marginal likelihood $p(\mathbf{d}_\Delta; \sigma_\Delta^2)$ can be obtained analytically, and the final traditional estimator takes the following form,

$$\hat{\sigma}_\Delta^2 = \frac{1}{n_r^\Delta}(\mathbf{d}_\Delta - \mathbf{H}_\Delta \boldsymbol{\mu}_{r,\Delta})^{\mathsf{T}} \mathbf{\Omega}^{-1} (\mathbf{d}_\Delta - \mathbf{H}_\Delta \boldsymbol{\mu}_{r,\Delta}). \tag{36}$$

By applying the estimator across the whole data area (using a sliding window approach) a non-stationary variance can be obtained for all grid locations in $\mathfrak{L}_\mathfrak{D}$, except those at the boundaries where the neighbourhood is reaching outside the data area.

Hierarchical estimator

In addition to the localized traditional MML estimator, we also use a localized hierarchical MML. This estimator appears as a "shrinkage" estimator, meaning that the variance estimate is constrained by a hyper distribution on the variance. We introduce a random variable s_r^2 representing the variance σ_r^2. We let s_r^2 be inverse gamma distributed, which is a conjugate prior model for Gaussian models,

$$s_r^2 \sim p(s_r^2) = \mathcal{IG}\{s_r^2; \epsilon_\mathrm{s}, \gamma_\mathrm{s}\} \tag{37}$$

where ϵ_s and γ_s are the model parameters (shape and scale respectively) of the inverse gamma distribution. The inverse gamma distribution acts as a hierarchical distribution describing the variance, i.e. a hyperprior distribution. The parameters ϵ_s and γ_s are therefore also known as hyperparameters. This additional level in the stochastic model is visualized in Figure 3 as a direct acyclic graph. The rest of the stochastic model is unchanged, so for the local neighbourhood Δ the marginal likelihood of the data $\mathbf{d}_\Delta$ follows from Equation 34,

$$\mathbf{d}_\Delta \sim p(\mathbf{d}_\Delta | s_\Delta^2) = \mathcal{N}_{n_d^\Delta}(\mathbf{d}_\Delta; \mathbf{H}_\Delta \mu_r \mathbf{i}_{n_r^\Delta}, s_\Delta^2 \mathbf{\Omega}). \tag{38}$$

The local variance s_Δ^2 conditioned on the neighbourhood data $\mathbf{d}_\Delta$ follows from Bayes' theorem:

$$[s_r^2 | \mathbf{d}_\Delta] \sim p(s_r^2 | \mathbf{d}_\Delta; \epsilon_s, \gamma_s) = \mathcal{IG}\{s_r^2; \epsilon_{s|d}, \gamma_{s|d}\} \tag{39}$$

with the posterior expectation of the two hyperparameters, shape parameter $\epsilon_{s|d}$ and scale parameter $\gamma_{s|d}$,

$$\epsilon_{s|d} = \epsilon_s + \frac{n_\mathrm{d}^\Delta}{2} \tag{40}$$

$$\gamma_{s|d} = \gamma_s - \frac{1}{2}(\mathbf{d}_\Delta - \mathbf{H}_\Delta \mu_r \mathbf{i}_{n_r^\Delta})^\mathsf{T} \mathbf{\Omega}^{-1} (\mathbf{d}_\Delta - \mathbf{H}_\Delta \mu_r \mathbf{i}_{n_r^\Delta}). \tag{41}$$

The final hierarchical estimator is defined as,

$$\widetilde{\sigma}^2_{\Delta} = \mathrm{E}\{s^2_r|\mathbf{d}_{\Delta}, \epsilon_{\mathrm{s}}, \gamma_{\mathrm{s}})\} = \frac{\gamma_{s|d}}{\epsilon_{s|d} - 1}. \tag{42}$$

The hierarchical estimator defined above requires a prior selection of reasonable values for the hyperparameters ϵ_{s} and γ_{s}. We propose to estimate the properties of the assumed hyperprior using an empirical Bayesian approach similar to that of Asfaw and Omre (2016). Using the traditional estimator and the sliding window approach, the variance s^2_{Δ} can be estimated for all possible grid locations in $\mathfrak{L}_{\mathfrak{D}}$. This set of estimated variances is then treated as a super-population of s^2_{Δ} which allows the estimation of a global mean μ_{s} and variance σ^2_s of s^2_{Δ}. Let $\hat{\mu}_s$ and $\hat{\sigma}^2_s$ be the average and empirical variance of this super population, which are then used to infer the hyperparameters in the inverse gamma prior model,

$$\gamma_s = \hat{\mu}_s \left(\frac{\hat{\mu}^2_s}{\hat{\sigma}^2_s} + 1 \right) \tag{43}$$

$$\epsilon_{\mathrm{s}} = \frac{\hat{\mu}^2_s}{\hat{\sigma}^2_s} + 2. \tag{44}$$

POSTERIOR MODEL

The prediction of the physical parameters $[\mathbf{r}|\mathbf{d}]$ is performed using Bayesian spatial inversion. The posterior model follows from Bayes' theorem. The general posterior model for a Gaussian prior model and linear-Gaussian likelihood model, is:

$$[\mathbf{r}|\mathbf{d}] \sim p(\mathbf{r}|\mathbf{d}) = constant \times p(\mathbf{d}|\mathbf{r})p(\mathbf{r}) = \mathcal{N}_n\{\mathbf{r}; \boldsymbol{\mu}_{r|d}; \boldsymbol{\Sigma}_{r|d}\} \tag{45}$$

where the resulting mean $\boldsymbol{\mu}_{r|d}$ is given by,

$$\boldsymbol{\mu}_{r|d} = \mu_r \mathbf{i}_r + \boldsymbol{\Gamma}_{rd} \boldsymbol{\Sigma}_d^{-1} (\mathbf{d} - \mathbf{H} \mu_r \mathbf{i}_r) \tag{46}$$

and the posterior covariance $\boldsymbol{\Sigma}_{r|d}$,

$$\boldsymbol{\Sigma}_{r|d} = \boldsymbol{\Sigma}_r - \boldsymbol{\Gamma}_{rd}\boldsymbol{\Sigma}_d^{-1}\boldsymbol{\Gamma}_{dr}. \tag{47}$$

INFERENCE RESULTS

The neighbourhood configuration Δ presented in Figure 2 is contained within traces $n_x = [6; 56]$ and samples $n_y = [12; 69]$ in $\mathfrak{L}_{\mathfrak{D}}$. The variance estimates from the traditional estimator ($\hat{\boldsymbol{\sigma}}_r^2$) and the shrinkage estimator ($\widetilde{\boldsymbol{\sigma}}_r^2$) are obtained from $\mathbf{d}$ in Figure 1f. The histograms of $\hat{\boldsymbol{\sigma}}_r^2$ (orange) and $\widetilde{\boldsymbol{\sigma}}_r^2$ (yellow) are shown in Figure 4a along with the $\mathcal{IG}$-prior (black line) estimated using empirical Bayes. The estimated hyper-parameters are $\epsilon = 2.63$ and $\gamma = 3.89 \times 10^{-3}$. The traditional estimator, $\hat{\boldsymbol{\sigma}}_r^2$, tend to underestimate the reference variance $\sigma_r = 0.001$ slightly. The $\mathcal{IG}$-prior constraints the variance estimate of the hierarchical estimator $\widetilde{\boldsymbol{\sigma}}_r^2$ which lead to higher values. Simultaneously, some of the higher values of $\hat{\boldsymbol{\sigma}}_r^2$ are constrained in the $\widetilde{\boldsymbol{\sigma}}_r^2$ estimate. In general, the distribution of $\widetilde{\boldsymbol{\sigma}}_r^2$ is narrower than for $\hat{\boldsymbol{\sigma}}_r^2$, which aligns with the general properties of the shrinkage estimator.

The true variance $\boldsymbol{\sigma}_r^2$ is subtracted both estimated variances in Figure 4b-c. Preferably all values should be zero, indicating a perfect estimation of the prior variance. The estimate from the shrinkage estimator is smoother due to the a priori constraint on the variance. Both estimators are able to capture some of the heterogenous variance. Some areas of higher variance are lowered to zero, while others are only partially. The actual variance is on overall underestimated, which is expected. The biggest difference in the variance estimate is roughly 0.02. The striped appearance of the variance estimates are likely due to effects from the smooth wavelet and the neighbourhood configuration. Furthermore, the results are only based on a single realization of $\mathbf{r}$ and computed data set $\mathbf{d}$. For a different

prior realization $\mathbf{r}$ and noise $\mathbf{e}_d$, the $\mathbf{d}$ would be different and hence the inference result.

By repeating the inference for different prior and noise realizations $\mathbf{r}$ and $\mathbf{e}_d$, we asses whether the in-homogeneous variance pattern in the reference prior is obtainable on average. The average for 50 repetitions of both MML estimators are displayed in Figure 4d-e. The average pattern is smoother than for the single inference result for both estimators. The striped effect of the wavelet is diminished for the average. The biggest difference in the variance estimate is roughly 0.013. In summary, it is possible to infer the variance pattern of $\mathbf{r}$ to a reasonable degree. The traditional estimator is better at capturing some of the high values in the variance, but is prone to underestimate areas of the low variance for a single realization. The hierarchical estimator improves this underestimation, but is worse at capturing areas of high variance. The two inference results will ultimately be judged by their further application, since variance estimates of physical observables are usually not the "end goal".

INVERSION RESULTS

We invert the observed data set $\mathbf{d}$ in the BC (Figure 1f) using six different prior models (M). The BC noise model and forward model are used for all posterior models to enable a fair comparison between each prior model. **M0**: As a reference we set up an inversion case with the reference prior variance $\boldsymbol{\theta}_p = [\boldsymbol{\sigma}_r^2]$. This is an absolute best case scenario using linearized Bayesian inversion, as we use exactly the same setup for prediction of $[\mathbf{r}|\mathbf{d}]$ as to generate the synthetic data. This is sometimes also referred to as the 'inverse crime' in the literature (e.g. Wirgin (2004)). M0 thereby serves as a great comparison for alternative variance models.

M1: Here we use the inferred prior model $\hat{\boldsymbol{\theta}}_p = [\hat{\boldsymbol{\sigma}}_r^2]$ as a plug-in estimate of the variance. **M2**: Here we use the inferred prior model $\widetilde{\boldsymbol{\theta}}_p = [\widetilde{\boldsymbol{\sigma}}_r^2]$ as a plug-in estimate of the variance. **M3-M5**: Prior information regarding non-stationarity in the variance of observable variables are usually not available. A common approach is therefore to use stationary variance in the prior model. In M3-5 we use stationary variance in the prior models. M3-4 use a variance equivalent to the lowest and highest variance in the BC respectively. In M5, we propose to relax the constraints on the variance, i.e. we assume a stationary prior distribution with very high variance. The motivation for applying a 'broad' prior distribution comes from the notion that prior distributions do not always need to be realistically concentrated around the true values of the observable variables (Gelman et al., 2014). The main argument here is that if data carry sufficient information about the spatial variables, this should far outweigh the broad prior variance. The benefit of this strategy is that subjective (expert) elicitation is lessened and prior models are presumably more 'objective' (Simpson et al., 2017). In practice we assume a prior model with variance 3 times higher than in M4.

Qualitative assessment

We asses each of the resulting Gaussian posterior distributions qualitatively by generating posterior realizations $\hat{\mathbf{r}}^*$ of $[\mathbf{r}|\mathbf{d}]$ using Cholesky decomposition. Each of these realizations represent a solution to the probabilistic inverse problem. We denote solutions of M0 as $\hat{\mathbf{r}}^{*,\text{opt}}$ because M0 represents the best possible (optimal) solution. We can asses each other posterior model qualitatively by visually comparing how well $\hat{\mathbf{r}}^*$ from the other models resembles $\hat{\mathbf{r}}^{*,\text{opt}}$. Figure 5 shows three different posterior realizations ($\hat{\mathbf{r}}^1$, $\hat{\mathbf{r}}^2$, and $\hat{\mathbf{r}}^3$) drawn from the six respective posterior distributions. Each realization is drawn with the same

random seed for comparative purposes. The posterior realizations from M1 (b,h, and n) and M2 (c,i, and o) are almost impossible to discriminate visually. The M1 realizations are however showing slightly higher values than M2. This is in accordance with the inference results. Both models show a fair resemblance to the best solution, with a good match between areas of high and low variance. We do see less ability to match the highest values of M0, due to the lower prior variance in the two estimates than the true variance. The optimal posterior variance in M0 is overall visually well represented by both M1 and M2. The results from M3 (d,j and p), M4 (e,k and q) and M5 (f,l and r) show the weakness of applying a stationary prior variance model for a subsurface with non-stationary properties. In M3, the posterior realizations match with M0 in the areas of low variance, as they are similar. Unfortunately, the high variance areas are not represented very well, as the prior variance is lowering the resulting posterior variance. The opposite is experienced for M4, where areas of high variance are obtained better than in any of the other models. Unfortunately, areas of low variance are greatly misrepresented by the posterior result, since the data are not able to constrain the prior variance. This problem also occurs in the final model, where the broad prior variance is not outweighed by the data likelihood.

Quantitative assessment

We introduce the Kullback and Leibler (1951) divergence (D_{KL}) as a measure of the similarity between two distributions. The D_{KL} for each posterior model $[\mathbf{r}|\mathbf{d}]$ compared with the distribution of the optimal posterior model $[\mathbf{r}|\mathbf{d}]^{\text{opt}}$ is calculated as,

$$D_{KL}([\mathbf{r}|\mathbf{d}]^{\text{opt}}||[\mathbf{r}|\mathbf{d}]) = \frac{1}{2}\left(\text{tr}(\boldsymbol{\Sigma}_{r|d}^{-1}\boldsymbol{\Sigma}_{r|d}^{\text{opt}}) + (\boldsymbol{\mu}_{r|d} - \boldsymbol{\mu}_{r|d}^{\text{opt}})^{\mathsf{T}}\boldsymbol{\Sigma}_{r|d}^{-1}(\boldsymbol{\mu}_{r|d} - \boldsymbol{\mu}_{r|d}^{\text{opt}}) - n + \ln\left(\frac{|\boldsymbol{\Sigma}_{r|d}|}{|\boldsymbol{\Sigma}_{r|d}^{\text{opt}}|}\right)\right) \tag{48}$$

Low values of D_{KL} indicate a high resemblance between the two posterior distributions. We introduce the log-likelihood (logL) as a measure of how representative each posterior distribution is of the reference model. Because $\mathbf{r}$ is a realization from a multivariate Gaussian distribution and each posterior distribution is likewise Gaussian, the log-likelihood follows by taking the logarithm of the Gaussian likelihood function,

$$\text{logL} = \ln(\mathcal{N}_n(\mathbf{r}; \boldsymbol{\mu}_{r|d}, \boldsymbol{\Sigma}_{r|d})) = -\frac{1}{2}\left(n\ln(2\pi) + \ln|\boldsymbol{\Sigma}_{r|d}| + (\mathbf{r}-\boldsymbol{\mu}_{r|d})^{\mathsf{T}}\boldsymbol{\Sigma}_{r|d}^{-1}(\mathbf{r}-\boldsymbol{\mu}_{r|d})\right) \quad (49)$$

Higher logL-values indicate that $\mathbf{r}$ is more likely to be considered a possible realization from the given posterior model.

We construct 50 new data sets $\mathbf{d}$ using the BC setup. We repeat the method of inferring the variance and subsequently invert these 50 data sets for acoustic impedance. Using Equation (48) and (49) we calculate the D_{KL} and logL, and present the average value for all 50 inversions in Table 2. The results in Table 2, maintain and emphasize the visual conclusions drawn from Figure 4. For the prior models with stationary variance, the logL of M4 is larger than M5, and especially M3, meaning that M4 is better at capturing the reference model while still having a smaller posterior variance. On average, M4 can therefore be considered the best stationary prior model for the BC. However, in general, the inferred non-stationary variance models are superior to any of the stationary models. Of the two, the variance estimate $\widetilde{\boldsymbol{\sigma}}_r^2$ obtained using the hierarchical estimator is slightly preferably in both quantifiable measures.

Sensitivity analysis

The BC represents a specific synthetic setup for which the estimatiors are assessed. In the following we investigate the sensitivity of the inversion results in terms of the amplitude

	M1	M2	M3	M4	M5
avg. D_{KL}	640,6	$\mathbf{530,9}$	5337,9	994,3	2304,7
avg. logL	4618,1	$\mathbf{4727,6}$	-100,4	4266,0	2956,3

Table 2: Summary of 50 BC inversions. Displayed is the average Kullback-Leibler divergence (D_{KL}), and log-likelihood (logL)

of variance difference ($\max(\mathbf{S})$), size of neigbourhood (n_d^{Δ}), and correlated noise level in the observed data ($\mathrm{S/N_{cor}}$). The sensitivity for these parameters is quantified using the approach used in Table 2, i.e. generating 50 data sets and calculate the average values of D_{KL}, and logL. The results are shown in Figure 6.

Because there is an underlying assumption of non-stationarity in the variance of the observable variables, it is important to test the effect of the difference between areas of high and low variance. In the BC we assumed this difference to be 5 times as high ($\max(\mathbf{S}) = 5$). We investigate the sensitivity of $\max(\mathbf{S})$ in the interval between 1 and 9, where $\max(\mathbf{S}) = 1$ is equal to a reference model with homogeneous variance. For $\max(\mathbf{S}) = 1$, M3 (yellow) and M4 (purple) are identical to the reference prior. M5 (green) with the broad prior variance is substantially outperformed by all other models for $\max(\mathbf{S}) = 1$. More importantly, the sensitivity analysis indicate that the inversion results for M2 (red) using the variance inferred with the hierarchical estimator, are more reliable and precise, already when the difference is only $\max(\mathbf{S}) = 1.5$. I.e. very little difference in the variance of physical parameters is needed for the non-stationary estimates to be feasible. For inversion using the variance obtained with the traditional estimator in M1 (blue), the point of feasibility is $\max(\mathbf{S}) = 2.5$. For

variance differences larger than $\max(\mathbf{S}) = 2.5$, the two non-stationary models are superior to the stationary models for both quantifiable measures.

Noise levels on seismic data are usually not fully disclosed, in fact the opposite is true in many instances. For this reason we investigate the effect of the noise level $\mathrm{S/N_{cor}}$ on the inversion results. For all $\mathrm{S/N_{cor}}$, the two non-stationary variance estimates are significantly providing better results than any of the stationary models. Of the two, M2 is slightly preferable for low $\mathrm{S/N_{cor}}$, whereas the results are similar for high $\mathrm{S/N_{cor}}$. Thus, the inversion results using the non-stationary variance models are significantly better than using the stationary models, regardless of noise level on the data.

Finally, we investigate the effect of neighbourhood size n_d^Δ in the inversion results. As seen in Figure 2, the n_d^Δ is a measure of how many data points and hence traces are included in the local window. M3-M5 are obviously unaffected by the change in neighbourhood size of the two localized predictors, and show the BC value for all three quantifiable measures. Both M1 and M2 show general improved results from higher n_d^Δ, i.e. including more traces in the prediction. However, the pattern of improvement flattens out for $n_d^\Delta > 17$ and will get worse for even larger neighbourhood sizes as the neighbourhood goes towards the global solution. M2 has superior results for all neighbourhood sizes in all measures compared with the best possible stationary model (M4). For M1, when only few traces are used in the estimate, the results become worse than for the stationary models. This is especially visible for the results of D_{KL} and logL which show extreme values for small n_d^Δ. For $n_d^\Delta > 13$, M1 is slightly better than M2. Despite this small preeminence for larger neighbourhood sizes, M2 is probably more robust than M1 in terms of neighbourhood size, due to the shrinkage properties of the hierarchical model.

REAL DATA CASE - NINI FIELD

We apply our proposed methodology and work-flow for a real data case from the Nini exploration field in the danish North Sea. Hydrocarbons in the Nini field are associated with re-mobilized sands (Tyr sand) above an active salt diapir (Svendsen et al., 2010). The Tyr sand is overlyain by the Vile shale, trapping the hydrocarbons. The expected large differences in subsurface variability, coupled to the presence of hydrocarbon rich sands, as well as the salt diapir, presents a justifiable real world application for the methodology. The zero-offset section of inline 1283 cuts across the salt diapir and is used as observed data $\mathbf{d}$. The observed data is shown in figure 7a. For inferring the variance of acoustic impedance, we use the likelihood model in equation 15. We use the accompanying estimated wavelet from the seismic section, to setup the linear forward operator $\mathbf{H}$. From the nearby Nini-1a well, we establish an a priori variance of $\sigma_{\mathrm{r}}^2 = 0,0138$. We choose an anistropic spherical model as the stationary correlation function $\rho_{\mathrm{r}}(.)$ as described in the BC. The range parameter is set to $a = 2$, and the anisotropy factor is set to $\kappa = 5$. This creates relatively thin and elongated structures in the prior model realizations. The noise model is set up as in the BC. The noise level is assumed to be $\mathrm{S/N_{unc}} = 7$ and $\mathrm{S/N_{cor}} = 1.5$, hence the main part of the noise is assumed to be correlated. The value of $\mathrm{S/N_{cor}} = 1.5$ is a conservative choice for the noise level, in order to lower the risk of overfitting the data. This conservative choice might lead to slight overestimation of the variance of the acoustic impedance in the posterior model. The variance is inferred using the hierarchical MML estimator. The ratio between variance of model parameters and noise is approximated with equation D-2 and the assumed signal-to-noise ratios. The local neighbourhood configuration is similar to that in Figure 2, i.e. $n_d^{\Delta} = 11$. We choose this configuration based on two criteria. Firstly, the BC showed that subsequent inversion benefits from a certain minimum neighbourhood

size. Secondly, to minimize the smoothing from the estimator, the neighbourhood must not become to large.

The inferred prior model $\widetilde{\boldsymbol{\theta}}_p = [\widetilde{\boldsymbol{\sigma}}_r^2]$ is presented in Figure 7b. The estimated variance is generally higher around the salt diapir. Since the variance estimate is coupled to the data, it is sensitive to changes in the data which can not be described by the noise model. In practice this means that if the data in certain areas does not show any variability, neither will the variance estimate. If a feature in the data, like a "bad trace" with lower data variance, is used in the variance estimate, then it drags the result in a direction of less variance. This can for instance be seen around crossline 6940 where the high variance estimate is lowered in a region due to low variance in the data. If we consider this variance to arise from the same layer, it would be geologically sound to suggest that the properties of this layer are not suddenly changing. The drop in variance is most likely an effect of "bad" data. For hydrocarbon exploration purposes we are probably more interested in regions close to the salt diapir where the variance is high and the effect is not as strong. Yet, the variance estimate's susceptibility to unmodeled features in data is still a weakness of the proposed methodology.

We invert the data presented in Figure 7a using equation 46 and 47 for two different prior models. **MA**: Here we use a prior model with stationary variance based on the a priori variance ($\boldsymbol{\sigma}_r^2 = \sigma_r^2 \mathbf{i}_n$) obtained from the well log. Having a global stationary behavior of the model parameters is the conventional geostatistical methodology for inverse problems (Sabeti et al., 2017). **MB**: Here we use the inferred prior model with non-stationary variance $\widetilde{\boldsymbol{\theta}}_p = [\widetilde{\boldsymbol{\sigma}}_r^2]$ as a plug-in estimate of the prior variance.

Figure 8 shows 3 different posterior realizations of relative acoustic impedance ($\mathrm{AI}^1 =$

$\exp(\hat{\mathbf{r}}^1)$, $\mathrm{AI}^2 = \exp(\hat{\mathbf{r}}^2)$, and $\mathrm{AI}^3 = \exp(\hat{\mathbf{r}}^3)$) drawn of the posterior distribution from both models using the same random seed for comparison. We stress the point that all realizations are probable representations of the subsurface. The diversity between each of the three realizations is a visual measure of the posterior variance, i.e. the final uncertainty on the relative acoustic impedance. In MA, the posterior variance is stationary, as data is mainly affecting the mean value. The biggest differences between MA and MB are found around the salt diapir. The non-stationary prior variance allows more variability in the posterior solutions, as seen by the darker colors in MB than MA. The internal variance between the realizations of MB is also clearly higher around the salt diapir, than in MA. Meanwhile, the variability of areas adjacent to the salt diapir is generally lower, seen by the dominantly yellow color of the realizations in MB. For all realizations of MB, some areas are generally showing a higher relative acoustic impedance than in MA, as for instance seen by the recurring structures on top of the salt diapir. This sudden change in acoustic impedance could potentially be linked with the presence of hydrocarbons in the area, which are generally found on the top of the salt diapir.

Using the inferred non-stationary variance for inversion of the Nini data, reveals that areas with potential hydrocarbons are highlighted. The posterior variability is larger around these areas, whereas less important areas show less variability. The final posterior variance is probably more realistic than the stationary variance in MA.

CONCLUSIONS

Spatial non-stationarity in observable variables are usually not possible to account for in standardized linearized Bayesian inversion of seismic data, as prior estimates of the variance are usually assumed to be stationary. Clearly, the validity of this assumption is low for

complex geological scenarios. The variance estimate, is usually derived from nearby seismic well logs. These logs may even be placed far from the actual seismic acquisition area, which should provide some additional uncertainty on the prior variance estimate. Our proposed methodology allow spatial non-stationarity in the prior variance. An additional benefit, is the consistency of the inferred variances. Rarely, one knows whether the stationary prior estimate is close or far from the actual variance in the subsurface, i.e. if the prior estimate is a Model 3 or Model 4, or maybe even a Model 5 scenario. Estimating the non-stationary variance prior to inversion limits this issue, as variance estimates are generally more consistent in a scenario with heterogenous variance of physical parameters in the subsurface.

Inversion results suggest that uncertainty estimates becomes more trustworthy, i.e. higher logL and lower D_{KL}. Trustworthy estimates of the posterior variance are essential when uncertainties are propagated further to other seismic disciplines. If false posterior uncertainties are propagated further in e.g. rock physics, the resulting rock physics parameters would probably lead to serious under- or overestimation of potential hydrocarbons in the subsurface. Our results indicate that the affectability of posterior variance is coupled to prior variance in the posterior model. Data instead provides more information on the mean model. In other words, the variance asked for, is roughly the variance obtained. The hierarchical MML estimator is generally the best performing and more robust of the two estimators in terms noise and variance levels.

The estimated prior model (Model B) demonstrates the applicability of the proposed methodology for a real data set. The field example does display one potential weakness of the proposed method. The variance estimates are susceptible to areas with features that are not explained in the noise model, such as "bad traces". This method can potentially be

further improved by a better description of the noise in the data.

Finally we comment on the philosophical issue mentioned in the introduction. In the proposed work-flow of applying the inferred variances as prior information, both inference and prediction are performed using the same data set. The distinction between prior information and data thereby becomes undesirably fuzzy (Scales and Snieder, 1997). However, a clear distinction is perhaps rarely the case in any real world application. In accordance with the expectations outlined in the introduction, the variance is slightly underestimated using the proposed work-flow. From a practitioners point of view, our results indicate that a more correct posterior distribution can be obtained in a scenario with heterogeneous variance in the subsurface, than e.g. applying a traditional stationary variance for inversion.

REFERENCES

Aki, K., and P. G. Richards, 1980, Quantitative seismology, theory and methods: Vol I and II, 1 ed.: W.H. Freedman Publishing.

Asfaw, Z. G., and H. Omre, 2016, Localized/Shrinkage Kriging Predictors: Mathematical Geosciences, **48**, 595–618.

Aune, E., J. Eidsvik, and B. Ursin, 2013, Three-dimensional non-stationary and non-linear isotropic AVA inversion: Geophysical Journal International, **194**, 787–803.

Bongajum, E. L., J. Boisvert, and M. D. Sacchi, 2013, Bayesian linearized seismic inversion with locally varying spatial anisotropy: Journal of Applied Geophysics, **88**, 31–41.

Bouchedda, A., G. Bernard, and E. Gloaguen, 2017, Constrained electrical resistivity tomography Bayesian inversion using inverse Matérn covariance matrix: Geophysics, **82**, E129–E141.

Box, G. E., and G. C. Tiao, 1992, Bayesian Inference in Statistical Analysis: A Wiley-Interscience Publication, 608.

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

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2014, Bayesian Data Analysis, 3 ed.: CRC Press.

Ghosh, M., and C. P. Robert, 2007, The Bayesian Choice: A Decision-Theoretic Motivation, 2 ed.: Springer.

Hampson, D., B. H. Russell, and A. C. G. G. Company, 2013, Joint simultaneous inversion of PP and PS angle gathers: CSEG Recorder, **38**, 32–38.

Johnson, R. A., and D. W. Wichern, 2007, Applied Multivariate Statistical Analysis, 6th ed.: Pearson Prentice Hall.

Jullum, M., and O. Kolbjørnsen, 2016, A Gaussian-based framework for local Bayesian inversion of geophysical data to rock properties: Geophysics, **81**, R75–R87.

Kullback, S., and R. A. Leibler, 1951, On information and sufficiency: Annals of Mathematical Statistics, **22**, 79–86.

Madsen, R. B., A. Zunino, and T. M. Hansen, 2017, On inferring the noise in probabilistic seismic AVO inversion using hierarchical Bayes: SEG Technical Program Expanded Abstracts, 601–604.

Mosegaard, K., and A. Tarantola, 2002, 16 Probabilistic approach to inverse problems, *in* International Handbook of Earthquake and Engineering Seismology, Part A: Academic Press, 237–265.

Sabeti, H., A. Moradzadeh, F. D. Ardejani, L. Azevedo, A. Soares, P. Pereira, and R. Nunes, 2017, Geostatistical seismic inversion for non-stationary patterns using direct sequential simulation and co-simulation: Geophysical Prospecting, 1–24.

Scales, J. A., and R. Snieder, 1997, To Bayes or not to Bayes . . . : Geophysics, **62**, 1045–1046.

Simpson, D. P., H. Rue, T. G. Martins, A. Riebler, and S. H. Sørbye, 2017, Penalising model component complexity: A principled, practical approach to constructing priors: Statistical Science, **32**, 1–28.

Svendsen, J. B., H. J. Hansen, T. Stærmose, and M. K. Engkilde, 2010, Sand remobilization and injection above an active salt diapir: The Tyr sand of the Nini Field, Eastern North Sea: Basin Research, **22**, 548–561.

Vecken, P. C. H., and M. Da Silva, 2004, Seismic inversion methods and some of their constraints: First Break, **22**, 47–70.

Wirgin, A., 2004, The inverse crime: arXiv preprint math-ph/0401050, 1–10.

APPENDIX A

THE GAUSSIAN DISTRIBUTION

The multivariate Gaussian probability density function $p(\mathbf{r})$ for a variable $\mathbf{r}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is given by (Johnson and Wichern, 2007)

$$p(\mathbf{r}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{r}-\boldsymbol{\mu})\right\} \tag{A-1}$$

where n is the dimension of the vector $\mathbf{r}$ and $|\boldsymbol{\Sigma}|$ denotes the determinant of the covariance matrix $\boldsymbol{\Sigma}$. A compact notation form is $\mathbf{r} \sim \mathcal{N}_n(\mathbf{r}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

APPENDIX B

THE INVERSE-GAMMA DISTRIBUTION

The univariate inverse gamma distribution for a variable x with shape parameter ϵ and scale parameter γ is defined by the probability density function (Ghosh and Robert, 2007).

$$f(x;\epsilon,\gamma) = \frac{\gamma^{\epsilon}}{\Gamma(\epsilon)}x^{-(\epsilon-1)}\exp\left\{-\frac{\gamma}{x}\right\} \tag{B-1}$$

where $\Gamma(.)$ is the Gamma function and $\epsilon \in \mathbb{R}^1_+ + 2$ and $\gamma \in \mathbb{R}^1_+$. The expectation and variance of x for the $\mathcal{IG}$ distribution, in terms of ϵ, and γ is described by:

$$\mathrm{E}\{x\} = \frac{\gamma}{\epsilon-1} \tag{B-2}$$

$$\mathrm{Var}\{x\} = \frac{\gamma^2}{(\epsilon-1)^2(\epsilon-2)} \tag{B-3}$$

A compact notation form is $x \sim \mathcal{IG}(x;\gamma,\epsilon)$.

APPENDIX C

THE JOINT DISTRIBUTION

The joint multivariate Gaussian variables of $\mathbf{r}_1 \sim \mathcal{N}_{n_1}(\mathbf{r}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathbf{r}_2 \sim \mathcal{N}_{n_2}(\mathbf{r}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ takes the following form (Johnson and Wichern, 2007):

$$\begin{bmatrix} \mathbf{r}_1 \\ \\ \mathbf{r}_2 \end{bmatrix} \sim \mathcal{N}_{n_1+n_2} \left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{21} \\ \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) \tag{C-1}$$

and the conditional distribution of $\mathbf{r}_1$ given $\mathbf{r}_2$ is also Gaussian

$$\mathbf{r}_1 | \mathbf{r}_2 \sim \mathcal{N}_{n_1}(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}) \tag{C-2}$$

where the expected mean and covariance are given by:

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{r}_2 - \boldsymbol{\mu}_2) \tag{C-3}$$

and

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \tag{C-4}$$

APPENDIX D

THE CHOICE OF C

Both MML estimators require a choice of c, which reflects the assumed ratio between the variance of noise and model parameters ($\sigma_2^2 = c\sigma_\Delta^2$). The a priori choice of c represents a challenge of itself. For the BC, the realization $\mathbf{r}$ from the reference prior distribution is

known. This allows us to estimate a global value of c for the specific realization in the BC as:

$$c = \frac{\sigma_2^2}{\sigma_\Delta^2} \approx \frac{\mathrm{Var}(\mathbf{e}_d)}{\mathrm{Var}(\mathbf{r})} \tag{D-1}$$

For the realization of $\mathbf{e}_d$ and $\mathbf{r}$ presented in Figure 1 the c-value is equal to 0.1766. For a sample of 1000 realizations of $\mathbf{e}$ and $\mathbf{r}$, we obtain the probability density distribution of c with the BC S/N using equation D-1. The resulting histogram of the sample is shown in Figure D-1a. The distribution is slightly skewed which is attributed to the non-stationarity in the variance of $\mathbf{e}_d$ and $\mathbf{r}$. The value of c for the current setup is varying between 0.1 and 0.5 with a sample mean = 0.229 (blue dot) and mode = 0.208 (red dot). Because $\mathbf{r}$ is only available due to the synthetic nature of the BC, we suggest to use the ratio between the a priori levels of σ_r^2 and σ_e^2 as an approximation of c. Letting σ_e^2 be the average of the noise variances σ_1^2 and σ_2^2 in equation 16, we obtain an approximate $\tilde{c}$ for the BC:

$$\tilde{c} = \frac{\sigma_\mathrm{e}^2}{\sigma_\mathrm{r}^2} \approx \frac{\left(\sigma_1^2 + \sigma_2^2\right)}{2\sigma_\mathrm{r}^2} = \frac{\left(\frac{\sigma_\mathrm{r}^2}{\mathrm{S/N_{unc}}} + \frac{\sigma_\mathrm{r}^2}{\mathrm{S/N_{cor}}}\right)}{2\sigma_\mathrm{r}^2} = \frac{1}{2}\left(\frac{1}{\mathrm{S/N_{unc}}} + \frac{1}{\mathrm{S/N_{cor}}}\right) \tag{D-2}$$

For the S/N in the BC, we obtain an approximate $\tilde{c}$ equal to 0.192 (black dot). $\tilde{c}$ is reasonably close to the mean value from the sample (blue dot) and almost identical to the mode. In order to further asses the robustness of the approximation in equation D-2 we propose a high and low S/N case. A histogram corresponding to Figure D-1a) is produced for both a high S/N case ($\mathrm{S/N_{unc}} = 30$ and $\mathrm{S/N_{cor}} = 10$) and a low S/N case ($\mathrm{S/N_{unc}} = 3$ and $\mathrm{S/N_{cor}} = 1$) in Figure D-1b-c respectively. Both histograms show the same shape of the distribution as in Figure D-1a. For the high S/N case the values of c are however lower and vice versa for the low S/N case. In both cases the approximate $\tilde{c}$ is fairly close to the mean and mode of the sample. The two alternative cases provide evidence of robustness of the c-estimate in terms of noise level on the data. In all three cases, $\tilde{c}$ is slightly underestimating the mean of the sample. Regardless, we consider $\tilde{c}$ a reasonable approximation for c based

only on a priori information available about the S/N. For a real data case the estimate of c-estimate depend on good a priori knowledge of noise levels.

LIST OF FIGURES

8 Posterior realization of Nini inline 1283. Results are shown for the two prior models, MA (a,c and e) and MB (b,d and f).

D-1 Histograms of c-values obtained from a sample of using equation D-1. a) S/N is equal to that of the BC ($S/N_{unc} = 20$ and $S/N_{cor} = 3$), b) High S/N ($S/N_{unc} = 30$ and $S/N_{cor} = 10$), and c) Low S/N ($S/N_{unc} = 3$ and $S/N_{cor} = 1$)

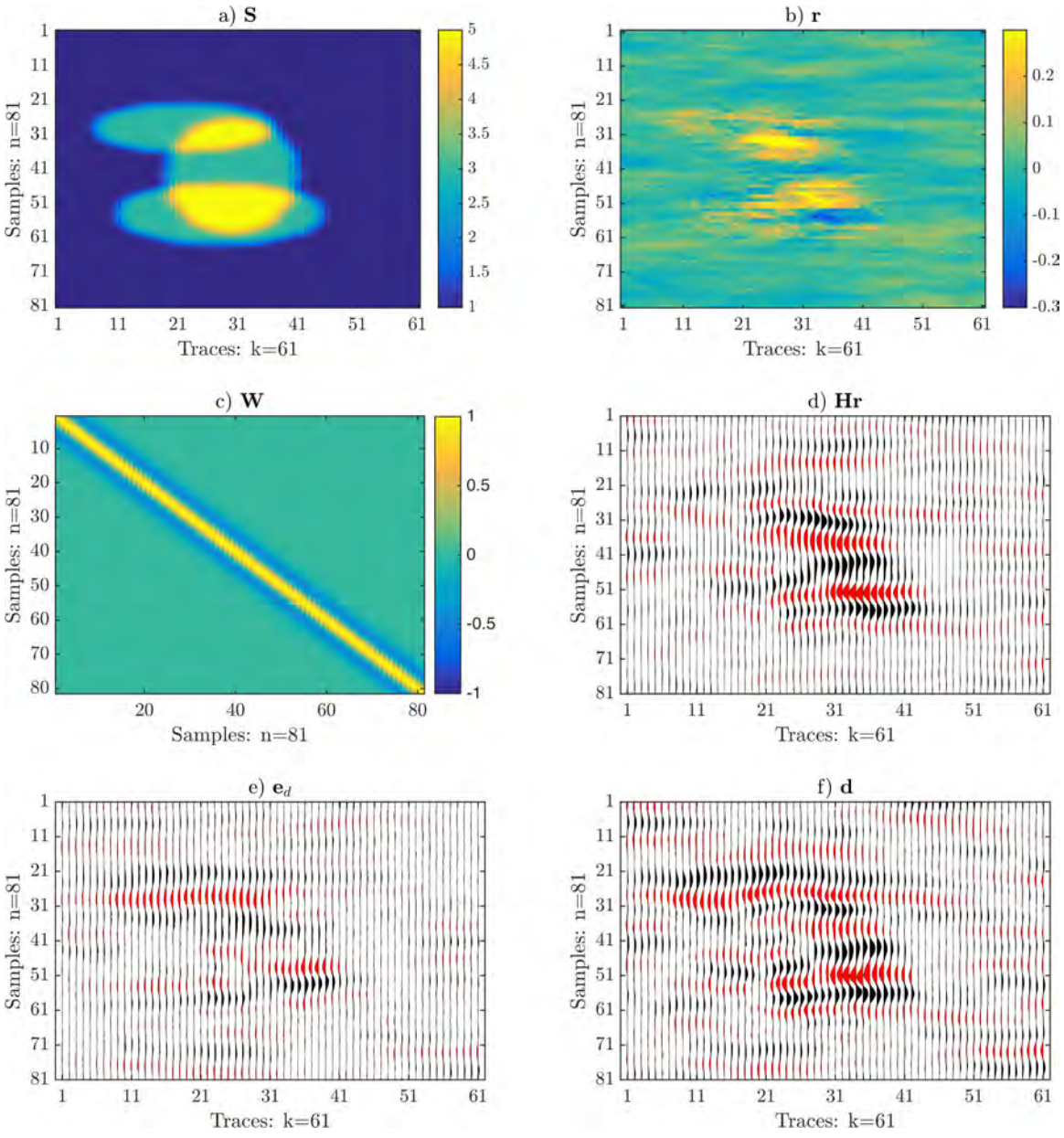


Figure 1: Base case. a) **S**: Scaling matrix, b) **r**: Realization from prior distribution, c) **W**: Wavelet matrix used for the BC, d) **Hr**: Seismic response from **r**, e) $\mathbf{e}_d$: Error realization, f) **d**: Synthetic seismic data set

–

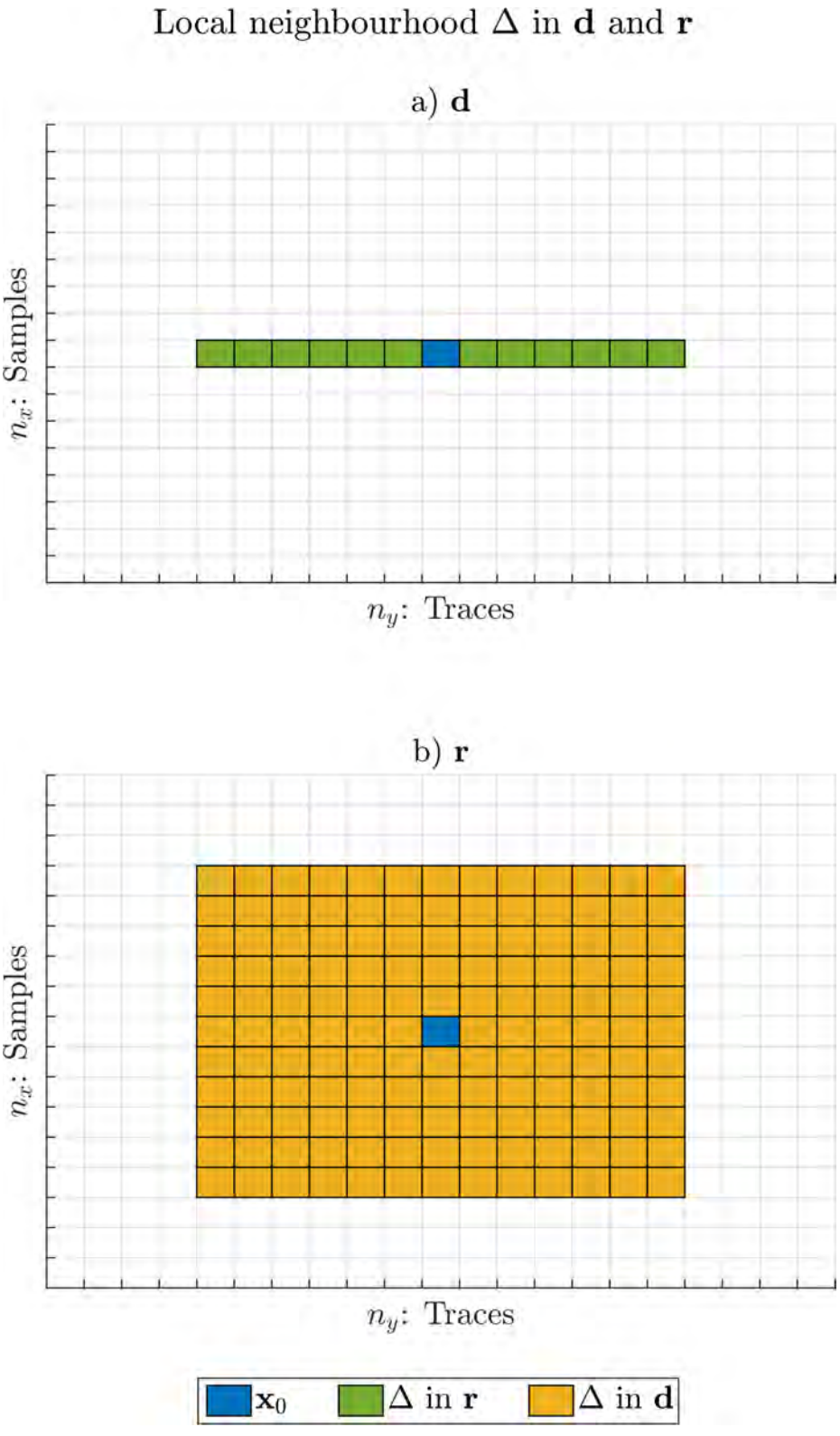


Figure 2: Local neighbourhood (Δ) configuration in $\mathbf{d}$ and $\mathbf{r}$ around location $\mathbf{x}_0$ (blue). –

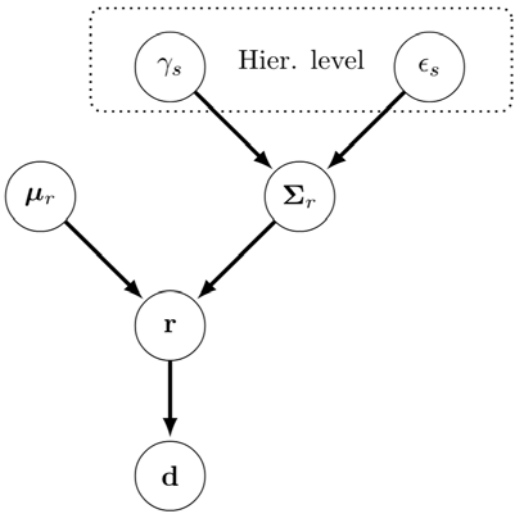


Figure 3: The stochastic model as a directed acyclic graph. The nodes are representing stochastic variables and the black arrows show probability dependencies. The hierarchical level of the inverse gamma prior is marked with a dotted rectangle.

–

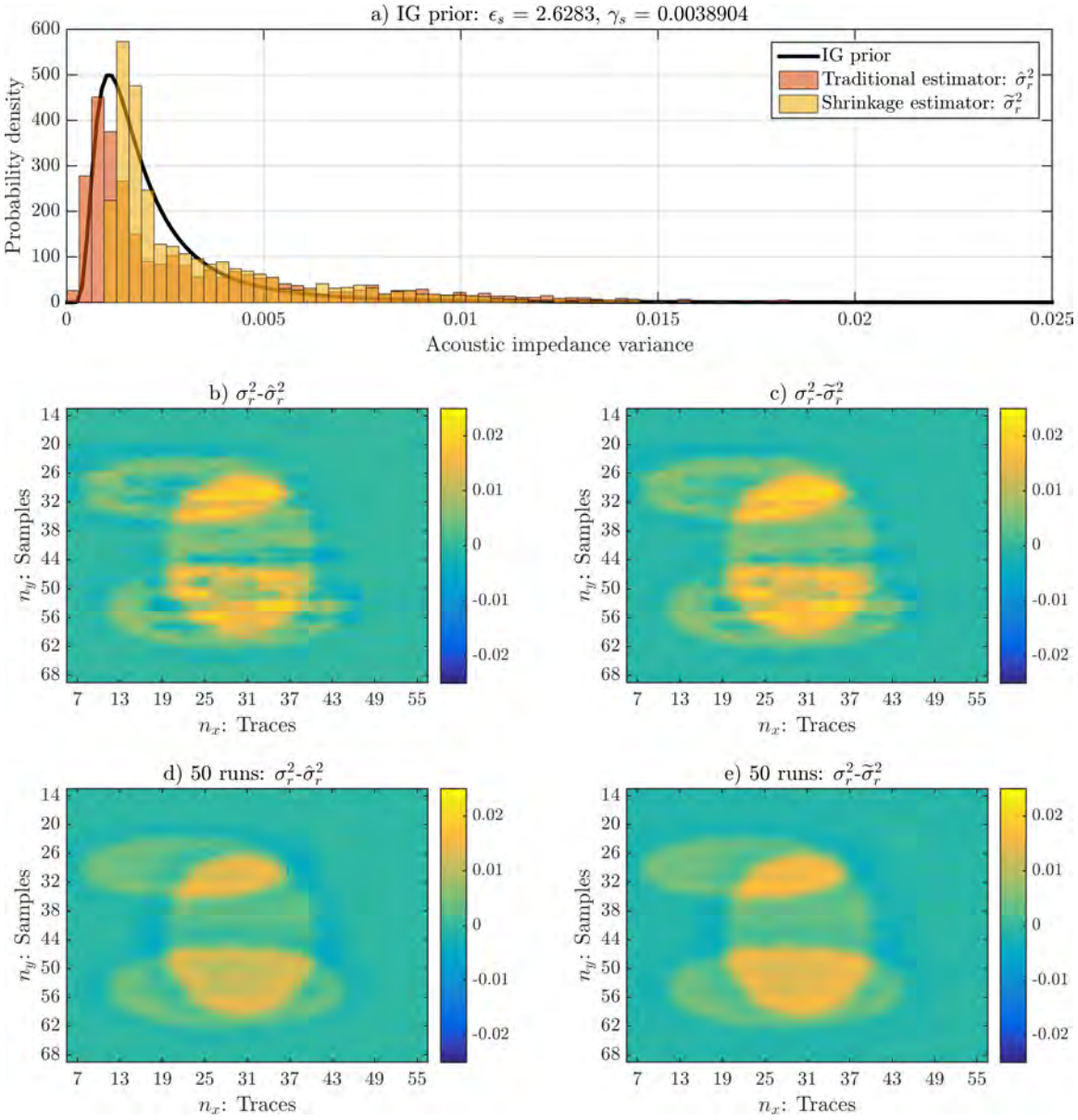


Figure 4: Results from inference of prior variance in the base case using the two MML estimators. a) Distribution of acoustic impedance variance for $\hat{\boldsymbol{\sigma}}_r^2$ (red) and $\widetilde{\boldsymbol{\sigma}}_r^2$ (yellow) obtained using the two localized estimators. The black line indicate the $\mathcal{IG}$ prior distribution with estimated ϵ_s and γ_s obtained using empirical Bayes. b) Difference between inference result and true variance: $\boldsymbol{\sigma}_r^2 - \hat{\boldsymbol{\sigma}}_r^2$. Yellow colors indicate underestimation of the variance and blue colors indicate overestimation of the variance. c) Difference between inference result and true variance: $\boldsymbol{\sigma}_r^2 - \widetilde{\boldsymbol{\sigma}}_r^2$. d-e) : Same plot as in b) and c), but for an average of

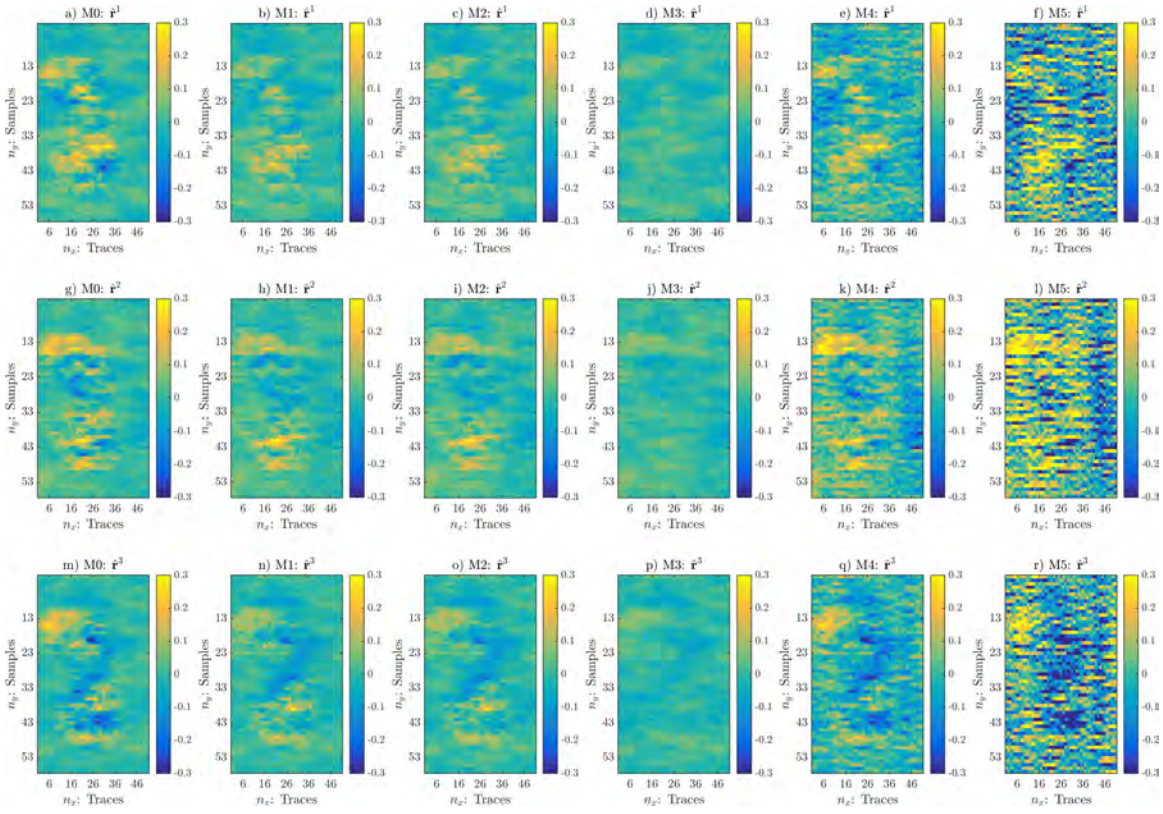


Figure 5: Posterior realizations $\hat{\mathbf{r}}^*$ (a-r) for prior models M0-M6 –

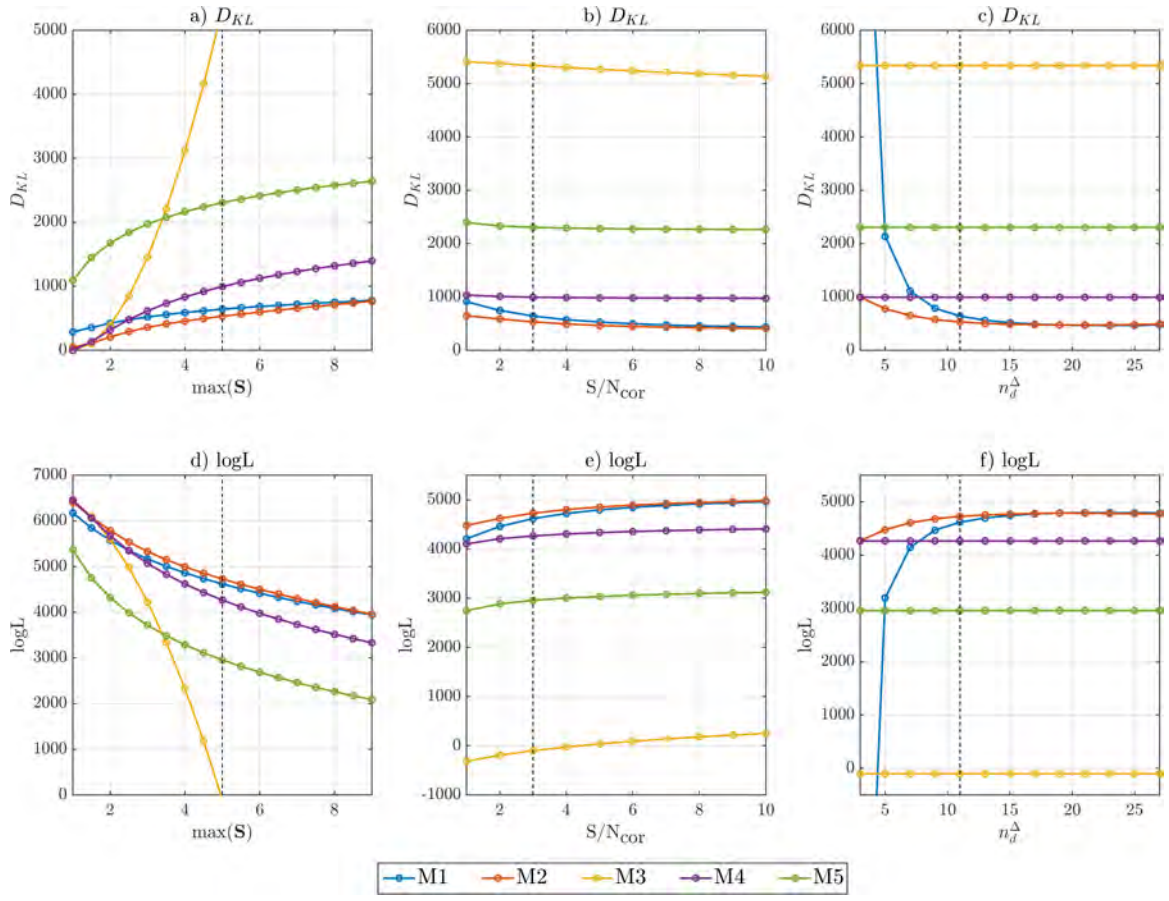


Figure 6: Summary of sensitivity analysis in terms of D_{KL} (top row), and logL (bottom row). The sensitivity is shown for three parameters max(**S**) (a and d), S/N_{cor} (b and e) and size of neighbourhood n_d^{Δ} (c and f). The results are displayed for the two heterogeneous variance models: M1 (blue) and M2 (red), and the three homogeneous variance models: M3 (yellow), M4 (purple), and M5 (green). The BC value of each parameter are marked with a stippled line

–

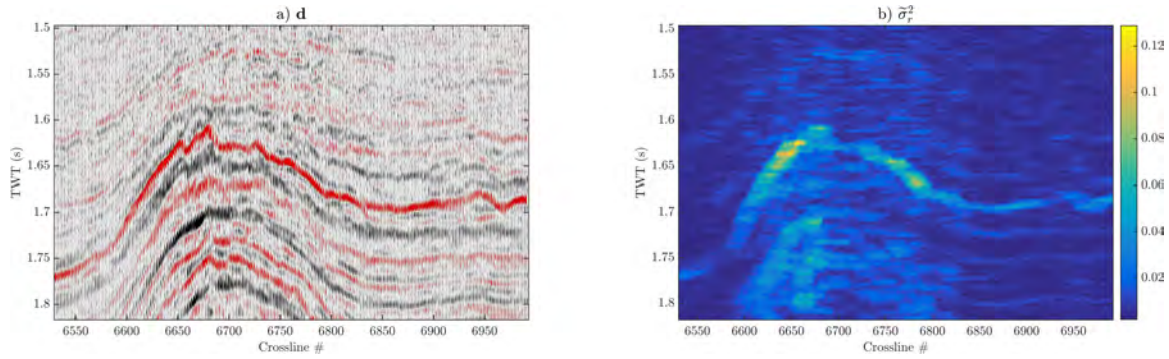


Figure 7: a) Observed data (**d**): Nini inline 1283. b) Inferred variance estimate $\widetilde{\boldsymbol{\sigma}}_r^2$. –

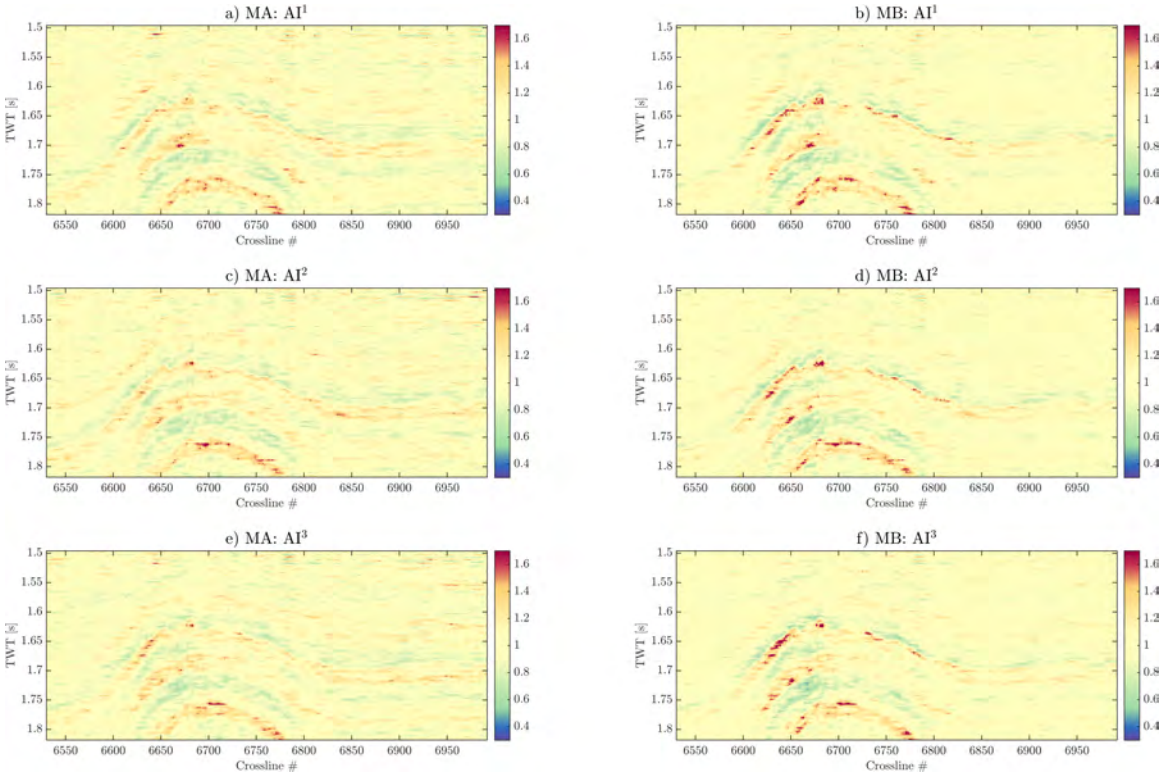


Figure 8: Posterior realization of Nini inline 1283. Results are shown for the two prior models, MA (a,c and e) and MB (b,d and f).

–

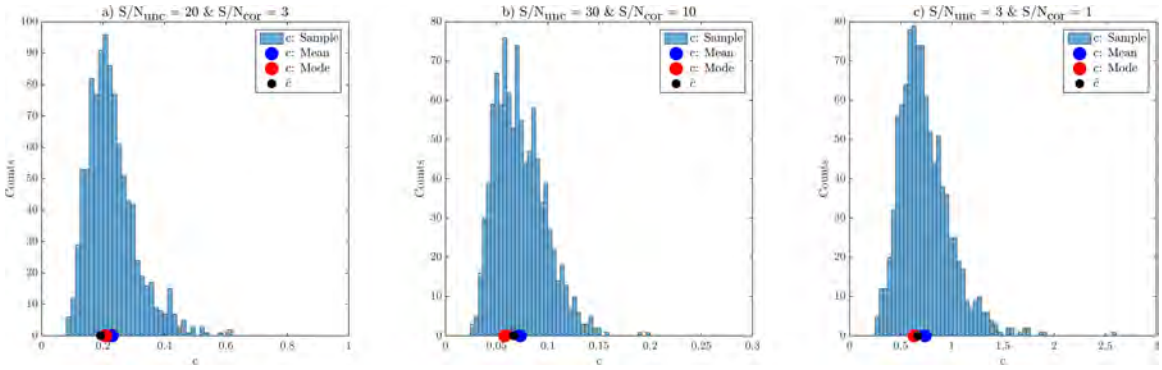


Figure D-1: Histograms of c-values obtained from a sample of using equation D-1. a) S/N is equal to that of the BC ($S/N_{unc} = 20$ and $S/N_{cor} = 3$), b) High S/N ($S/N_{unc} = 30$ and $S/N_{cor} = 10$), and c) Low S/N ($S/N_{unc} = 3$ and $S/N_{cor} = 1$)

–