Bayesian seismic AVO inversion

Arild Buland

Thesis for the degree of

DOKTOR INGENIØR

Department of Mathematical Sciences
Norwegian University of Science and Technology
N-7491 Trondheim
Norway

2002

Preface

This thesis is submitted in partial fulfillment for the requirements of the degree of "Doktor Ingeniør" at the Norwegian University of Science and Technology (NTNU). The work has been carried out partly at the Department of Mathematical Sciences at NTNU and partly at Statoil Research Centre in Trondheim.

I am indebted to my supervisor Professor Henning Omre for encouragement and continuous support and guidance through numerous meetings, and for the patience to carefully read the manuscripts. I would also like to thank Statoil for the financial support, and all my colleagues at Statoil Research Centre for many interesting discussions and social company. Especially, I would like to thank my former and current leaders at Statoil, Lasse Amundsen and Peder Garten, for allowing me to work with these interesting topics. I would also like to thank Kåre Horpestad at the Sleipner office and Lars Haakon Nordby at the Heidrun office for help and discussions regarding the real data sets used in the examples. I also thank Odd Kolbjørnsen for a pleasant collaboration with the FFT paper in Chapter 5. Finally, I would like to thank my family: my mum and dad, my wife Brita Louise and our sons Anders and Birger, most of all for representing another and more important dimension to life than geophysics and statistics.

Trondheim, September 2002

Arild Buland

Contents

Chapter 1

Introduction

Seismic analysis is a key element in successful exploration and production of natural resources. During the last decades, seismic methodology has had a significant progress with respect to both acquisition, processing and analysis. Despite all the new technology, the uncertainty related to seismic analysis is still large, and even worse, the uncertainty is often not systematically assessed.

In this thesis, the uncertainty aspect of seismic amplitude versus offset (AVO) inversion is assessed using a Bayesian approach to inversion. The main objective is to estimate elastic material parameters with associated uncertainty from large seismic data sets, but the inversion problem also includes estimation of seismic wavelets and the noise level. State of the art statistical methodology is applied to attack these current and crucial geophysical problems. The core part of the work is presented in four separate papers written for geophysical journals, constituting Chapter 2 through 5 in this thesis. Each of the papers is self-contained, with exception of the references which are placed in a separate bibliography chapter.

This introduction starts with a presentation of required basic concepts within spatial random fields, Bayesian inference and stochastic simulation, and continues with a presentation of classical and Bayesian seismic inversion. Finally, an outline of the four papers is given.

1.1 Spatial random fields

Geophysical problems are often described in a spatial setting. Usually, geophysical models represent simplifications of the reality, and typical geophysical data suffer from measurement uncertainty and noise. This calls for spatial random field models.

A spatial random field $\{r(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{D}\}$ is a function whose values are random for any location $\boldsymbol{x} \in \mathcal{D}$, where $\mathcal{D}$ is a subset of the d-dimensional Euclidean space, $\mathcal{D} \subset \mathbb{R}^d$, see for example Cressie (1991); Christakos (1992). In the following, we consider continuous random fields, i.e. random fields where $r(\boldsymbol{x})$ is a continuous variable.

A random field is fully specified by the cumulative distribution functions (cdf)

$$P(r_1, \ldots, r_n) = \text{Prob}\{r(\boldsymbol{x}_1) \leq r_1, \ldots, r(\boldsymbol{x}_n) \leq r_n\}, \tag{1.1}$$

for any positive integer n, and any $\boldsymbol{x}_i \in \mathcal{D}$. The generic cdf $P(\cdot)$ is right-continuous and nondecreasing. For continuous distribution functions, the probability density function (pdf) is, if it exists, obtained from the partial derivatives as

$$p(r_1, \ldots, r_n) = \frac{\partial^n P(r_1, \ldots, r_n)}{\partial r_1 \cdots \partial r_n}. \tag{1.2}$$

Expectation and covariance

The expectation of the random field in location $\boldsymbol{x}$ can be defined by the probability density function by

$$\mu_r(\boldsymbol{x}) = \text{E}\{r(\boldsymbol{x})\} = \int r(\boldsymbol{x})\, p(r(\boldsymbol{x}))\, dr(\boldsymbol{x}), \tag{1.3}$$

and the covariance for locations $\boldsymbol{x}$ and $\boldsymbol{y}$ can be defined by

$$\begin{aligned} \Sigma_r(\boldsymbol{x}, \boldsymbol{y}) &= \text{Cov}\{r(\boldsymbol{x}), r(\boldsymbol{y})\} \\ &= \int\int r(\boldsymbol{x}) r(\boldsymbol{y})\, p(r(\boldsymbol{x}), r(\boldsymbol{y}))\, dr(\boldsymbol{x})\, dr(\boldsymbol{y}) - \mu_r(\boldsymbol{x})\mu_r(\boldsymbol{y}). \end{aligned} \tag{1.4}$$

The variance function is defined from the covariance function as

$$\sigma_r^2(\boldsymbol{x}) = \Sigma_r(\boldsymbol{x}, \boldsymbol{x}), \tag{1.5}$$

and the spatial correlation function for the random field is defined by

$$\nu_r(\boldsymbol{x}, \boldsymbol{y}) = \frac{\Sigma_r(\boldsymbol{x}, \boldsymbol{y})}{\sigma_r(\boldsymbol{x})\sigma_r(\boldsymbol{y})}. \tag{1.6}$$

Any valid covariance function must be positive definite, which means that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \Sigma_r(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0, \tag{1.7}$$

for any positive integer n, any set of weights $c_i \in \mathbb{R}^1$, and any $\boldsymbol{x}_i \in \mathcal{D}$. Correlation functions must also be in the class of positive definite functions, and further have the property that $\nu_r(\boldsymbol{x}, \boldsymbol{x}) = 1$. The class of valid correlation functions is closed under multiplication and addition, in the sense that if $\nu_1(\boldsymbol{x}, \boldsymbol{y})$ and $\nu_2(\boldsymbol{x}, \boldsymbol{y})$ are valid correlation functions, then

$$\nu_r(\boldsymbol{x}, \boldsymbol{y}) = \nu_1(\boldsymbol{x}, \boldsymbol{y})\nu_2(\boldsymbol{x}, \boldsymbol{y}), \tag{1.8}$$

is a valid correlation function, and

$$\nu_r(\boldsymbol{x}, \boldsymbol{y}) = a_1\nu_1(\boldsymbol{x}, \boldsymbol{y}) + a_2\nu_2(\boldsymbol{x}, \boldsymbol{y}) \tag{1.9}$$

is a valid correlation function if $a_1, a_2 \geq 0$ and $a_1 + a_2 = 1$.

A random field may be multidimensional, that is a vector random field

$$\boldsymbol{r}(\boldsymbol{x}) = [r_1(\boldsymbol{x}), \ldots, r_m(\boldsymbol{x})]^T, \tag{1.10}$$

with expectation functions

$$\mu_{r_i}(\boldsymbol{x}) = \mathrm{E}\{r_i(\boldsymbol{x})\}, \tag{1.11}$$

and cross-covariance functions

$$\Sigma_{r_i,r_j}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{Cov}\{r_i(\boldsymbol{x}), r_j(\boldsymbol{y})\}, \tag{1.12}$$

for $i, j = 1, \ldots, m$. Special cases of expression (1.12) are the spatial covariance functions

$$\Sigma_{r_i}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{Cov}\{r_i(\boldsymbol{x}), r_i(\boldsymbol{y})\}, \tag{1.13}$$

and the variable covariance functions

$$\Sigma_{r_i,r_j}(\boldsymbol{x}, \boldsymbol{x}) = \mathrm{Cov}\{r_i(\boldsymbol{x}), r_j(\boldsymbol{x})\}. \tag{1.14}$$

Symmetry properties

A random field may have symmetry properties defined by invariance of statistical properties under some transformations. Two important symmetry properties are homogeneity (stationarity for time series) related to invariance under translations, and isotropy related to invariance under translations and rotations.

A random field is homogeneous in the strict sense (first order or strong homogeneity) if all finite-dimensional distributions are invariant under translation:

$$\text{Prob}\{r(\boldsymbol{x}_1+\boldsymbol{s}) \leq r_1, \ldots, r(\boldsymbol{x}_n+\boldsymbol{s}) \leq r_n\} = \text{Prob}\{r(\boldsymbol{x}_1) \leq r_1, \ldots, r(\boldsymbol{x}_n) \leq r_n\} \tag{1.15}$$

for $\boldsymbol{s} \in \mathbb{R}^d$. Homogeneity in the strict sense implies homogeneity in the wide sense (second order or weak homogeneity), defined by homogeneity for the first and second moments:

$$\mu_r(\boldsymbol{x}) = \mu_r \quad \text{and} \quad \Sigma_r(\boldsymbol{x}, \boldsymbol{y}) = \Sigma_r(\boldsymbol{\xi}), \tag{1.16}$$

where $\boldsymbol{\xi} = \boldsymbol{x} - \boldsymbol{y}$ is the separation vector. Generally, homogeneity in the wide sense does not imply homogeneity in the strict sense.

The covariance function for a homogeneous random field is said to be a homogeneous covariance function. A homogeneous covariance function has constant variance

$$\Sigma_r(\boldsymbol{\xi}) = \sigma_r^2 \nu_r(\boldsymbol{\xi}). \tag{1.17}$$

A homogeneous random field is an isotropic random field if the covariance depends on distance only

$$\Sigma_r(\boldsymbol{x}, \boldsymbol{y}) = \Sigma_r(\boldsymbol{\xi}) = \Sigma_r(\xi), \tag{1.18}$$

where ξ is the distance between the locations $\boldsymbol{x}$ and $\boldsymbol{y}$. An isotropic random field is also a homogeneous random field.

Derivatives of random fields

Consider the random field $r(\boldsymbol{x})$ and assume that it is differentiable with respect to $\boldsymbol{x}$. The gradient field (potential vector field) $\nabla r(\boldsymbol{x})$ is defined by the

components (Christakos, 1992)

$$\frac{\partial r}{\partial x_i}(\boldsymbol{x}) = \lim_{\epsilon \to 0} \frac{r(\boldsymbol{x} + \epsilon \boldsymbol{e}_i) - r(\boldsymbol{x})}{\epsilon}, \tag{1.19}$$

where $\boldsymbol{e}_i$ is a unit vector in direction i.

The expectation of a gradient component is

$$\mu_{\partial_i r}(\boldsymbol{x}) = \mathrm{E}\left\{ \frac{\partial r}{\partial x_i}(\boldsymbol{x}) \right\} = \frac{\partial}{\partial x_i} \mathrm{E}\{r(\boldsymbol{x})\}. \tag{1.20}$$

The cross-covariance function between $\partial r(\boldsymbol{x})/\partial x_i$ and $r(\boldsymbol{y})$ is defined by

$$\Sigma_{\partial_i r, r}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{Cov}\left\{ \frac{\partial r}{\partial x_i}(\boldsymbol{x}), r(\boldsymbol{y}) \right\} = \frac{\partial}{\partial x_i} \Sigma_r(\boldsymbol{x}, \boldsymbol{y}), \tag{1.21}$$

and the cross-covariance function between $\partial r(\boldsymbol{x})/\partial x_i$ and $\partial r(\boldsymbol{y})/\partial y_j$ is

$$\Sigma_{\partial_i r, \partial_j r}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{Cov}\left\{ \frac{\partial r}{\partial x_i}(\boldsymbol{x}), \frac{\partial r}{\partial y_j}(\boldsymbol{y}) \right\} = \frac{\partial}{\partial x_i} \frac{\partial}{\partial y_j} \Sigma_r(\boldsymbol{x}, \boldsymbol{y}). \tag{1.22}$$

If $r(\boldsymbol{x})$ is a homogeneous random field, the expectation of the gradient components are zero, $\mu_{\partial_i r}(\boldsymbol{x}) = 0$. Further, since $\Sigma_r(\boldsymbol{x}, \boldsymbol{y}) = \Sigma_r(\boldsymbol{\xi})$, the covariance of $\partial r/\partial x_i$ is $-\partial^2 \Sigma_r(\boldsymbol{\xi})/\partial \xi_i^2$, where the minus sign is related to the minus sign in $\boldsymbol{\xi} = \boldsymbol{x} - \boldsymbol{y}$. For a homogeneous random field, the covariance attains maximum at $\boldsymbol{\xi} = \boldsymbol{0}$, and therefore $\partial \Sigma_r(\boldsymbol{\xi} = \boldsymbol{0})/\partial \xi_i = 0$.

Gaussian random fields

A large number of natural phenomena can be modeled with sufficient accuracy by Gaussian random fields. Gaussian random fields have convenient mathematical properties, and many problems can only be solved analytically under Gaussian assumptions.

A Gaussian random field $r(\boldsymbol{x})$ is a random field where all the finite-dimensional distributions $P(r_1, \ldots, r_n) = \mathrm{Prob}\{r(\boldsymbol{x}_1) \leq r_1, \ldots, r(\boldsymbol{x}_n) \leq r_n\}$ are multivariate Gaussian distributions for any n and any $\boldsymbol{x}_i \in \mathcal{D}$. Let $\boldsymbol{r} = [r_1, \ldots, r_n]^T$. The multi Gaussian probability density for $\boldsymbol{r}$ is (Anderson, 1984)

$$p(\boldsymbol{r}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_r|^{1/2}} \exp\left[-\frac{1}{2} (\boldsymbol{r} - \boldsymbol{\mu}_r)^T \boldsymbol{\Sigma}_r^{-1} (\boldsymbol{r} - \boldsymbol{\mu}_r) \right], \tag{1.23}$$

which is completely specified by the expectation vector $\boldsymbol{\mu}_r$ and the covariance matrix $\boldsymbol{\Sigma}_r$. A compact notation is

$$\boldsymbol{r} \sim \mathcal{N}_n(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r), \tag{1.24}$$

where $\sim$ means "distributed as". The expectation can be arbitrarily chosen, but the covariance function must be positive definite to ensure the existence of all finite-dimensional distributions. Since a Gaussian random field is completely specified by its expectation and covariance, homogeneity in the wide and strict sense are equivalent.

Linear combinations of the components of $\boldsymbol{r}$ and all subsets of the components of $\boldsymbol{r}$ are (multivariate) Gaussian. If $\boldsymbol{r} \sim \mathcal{N}_n(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$ and $\boldsymbol{M}$ is an arbitrary $m \times n$ matrix, then $\boldsymbol{M}\boldsymbol{r} \sim \mathcal{N}_m(\boldsymbol{M}\boldsymbol{\mu}_r, \boldsymbol{M}\boldsymbol{\Sigma}_r\boldsymbol{M}^T)$. If $r(\boldsymbol{x})$ is a differentiable Gaussian random field, then the components $\partial r(\boldsymbol{x})/\partial x_i$ are also Gaussian random fields since the differential operator is defined as the limit of linear combinations of Gaussian variables.

A well-known result which is frequently used in this thesis is the formula for a conditional Gaussian distribution. Consider two multivariate Gaussian random vectors $\boldsymbol{r}_1 \sim \mathcal{N}_{n_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\boldsymbol{r}_2 \sim \mathcal{N}_{n_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ with joint distribution

$$\begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \end{bmatrix} \sim \mathcal{N}_{n_1+n_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right). \tag{1.25}$$

The conditional distribution of $\boldsymbol{r}_1$ given $\boldsymbol{r}_2$ is Gaussian

$$\boldsymbol{r}_1|\boldsymbol{r}_2 \sim \mathcal{N}_{n_1}(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}), \tag{1.26}$$

with conditional expectation

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{r}_2 - \boldsymbol{\mu}_2), \tag{1.27}$$

and conditional covariance

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{1.28}$$

Note that the conditional covariance does not depend on the given $\boldsymbol{r}_2$, but is solely determined from the covariance matrix of the joint distribution.

1.2 Bayesian inference

For a given problem, let $\boldsymbol{\theta}$ be a vector of unknown parameters with a specific meaning and interest for the problem, and let $\boldsymbol{o}$ be some observations related to $\boldsymbol{\theta}$. The observations are in general uncertain, and statistically linked to $\boldsymbol{\theta}$ via the pdf $p(\boldsymbol{o}|\boldsymbol{\theta})$. In the Bayesian paradigm, the unknown parameter vector $\boldsymbol{\theta}$ is also regarded as stochastic, and it is therefore natural to represent the current knowledge and uncertainty about $\boldsymbol{\theta}$ by a probability distribution. The distribution for $\boldsymbol{\theta}$ assigned before $\boldsymbol{o}$ is observed is called the prior distribution, denoted $p(\boldsymbol{\theta})$. At this point, the Bayesian philosophy differs from classical frequentistic statistics, where $\boldsymbol{\theta}$ is regarded as an unknown, but fixed nonstochastic quantity (Robert, 1994).

Statistical inference can be formulated as drawing conclusions or inference about $\boldsymbol{\theta}$ through the quantitative observations $\boldsymbol{o}$. As opposed to statistical modeling, statistical inference can be regarded as an inversion process, since the objective is to retrieve the "causes" $\boldsymbol{\theta}$ from the observed "effects" $\boldsymbol{o}$. When $\boldsymbol{o}$ is an observed and constant vector, the pdf $p(\boldsymbol{o}|\boldsymbol{\theta})$ defines a function of $\boldsymbol{\theta}$ known as the likelihood function. Some authors use the notation $L(\boldsymbol{\theta}|\boldsymbol{o}) = p(\boldsymbol{o}|\boldsymbol{\theta})$ to stress the inversion aspect of the likelihood function. In classical frequentistic statistics, a natural and common point estimator for $\boldsymbol{\theta}$ is the maximum likelihood estimator (MLE)

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\{p(\boldsymbol{o}|\boldsymbol{\theta})\}. \tag{1.29}$$

In the Bayesian approach to inference, the prior model expressed by $p(\boldsymbol{\theta})$ is combined with the information about $\boldsymbol{\theta}$ provided by $\boldsymbol{o}$ through the likelihood model. While $p(\boldsymbol{\theta})$ expresses the probability distribution before $\boldsymbol{o}$ is observed, the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{o})$ expresses the updated probability distribution for $\boldsymbol{\theta}$ after observing the value of $\boldsymbol{o}$. The posterior distribution can be expressed by Bayes' law,

$$p(\boldsymbol{\theta}|\boldsymbol{o}) = \frac{p(\boldsymbol{o}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{o})}, \tag{1.30}$$

where

$$p(\boldsymbol{o}) = \int p(\boldsymbol{o}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{1.31}$$

is the marginal density for $\boldsymbol{o}$. Since $p(\boldsymbol{o})$ does not depend on the unknown $\boldsymbol{\theta}$, it can be regarded as a normalizing constant for the posterior distribution in expression (1.30). For real problems, the marginal density $p(\boldsymbol{o})$, and hence the

posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{o})$ are generally not available on closed analytical forms. Historically, this has been one of the most serious drawbacks using Bayesian methodology on real problems.

Bayes' law can be regarded as an inversion formula for $\boldsymbol{\theta}$, where the posterior distribution represents the complete solution, including the uncertainty. A convenient point estimator for $\boldsymbol{\theta}$ is the maximum posterior (MAP) solution

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\{p(\boldsymbol{\theta}|\boldsymbol{o})\}, \tag{1.32}$$

while the classical Bayes estimator (under quadratic loss function) is the posterior expectation

$$\widehat{\boldsymbol{\theta}} = \mathrm{E}\{\boldsymbol{\theta}|\boldsymbol{o}\}. \tag{1.33}$$

Bayesian inference is completely based on the posterior distribution. The posterior information is usually summarized by quantities as posterior mean and variance, maximum posterior distribution, quantiles and credibility regions. Many summary quantities can be represented by integrals on the form

$$\mathrm{E}\{g(\boldsymbol{\theta})|\boldsymbol{o}\} = \int g(\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{o})d\boldsymbol{\theta}, \tag{1.34}$$

where for example $g(\boldsymbol{\theta}) = \theta_i$ for the posterior mean μ_i of element θ_i in $\boldsymbol{\theta}$, and $g(\boldsymbol{\theta}) = (\theta_i - \mu_i)(\theta_j - \mu_j)$ for the posterior covariance of θ_i and θ_j. A $(1-\epsilon)100\%$ credible interval C for θ_i is such that

$$\int \mathrm{I}\{\theta_i \in C\}p(\boldsymbol{\theta}|\boldsymbol{o})d\boldsymbol{\theta} = 1 - \epsilon, \tag{1.35}$$

where $\mathrm{I}\{\mathcal{A}\}$ is the indicator function being 1 when $\mathcal{A}$ is true and 0 otherwise. A widely used $(1-\epsilon)100\%$ credible interval is defined such that the probability that θ_i is below the lower limit of the interval is equal to the probability that θ_i is above the upper limit, that is $\epsilon/2$. A $(1-\epsilon)100\%$ highest posterior distribution (HPD) interval for θ_i is

$$C = \{\theta_i; p(\theta_i|\boldsymbol{o}) \geq p_\epsilon\}, \tag{1.36}$$

where $p(\theta_i|\boldsymbol{o})$ is the posterior marginal density for θ_i, and p_ϵ is the largest constant such that C is a $(1-\epsilon)100\%$ credible interval.

1.3 Markov chain Monte Carlo

A practical difficulty with Bayesian analysis is that the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{o})$ is generally not available on a closed analytical form, and integrals as in expression (1.34) may be hard to evaluate. This problem can be solved approximately if a number of posterior samples $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(m)}$ from $p(\boldsymbol{\theta}|\boldsymbol{o})$ can be provided, such that the integral in expression (1.34) can be evaluated by Monte Carlo integration (Robert and Casella, 1999):

$$\mathrm{E}\{g(\boldsymbol{\theta})|\boldsymbol{o}\} \approx \frac{1}{m}\sum_{i=1}^{m} g(\boldsymbol{\theta}^{(i)}). \tag{1.37}$$

The law of large numbers ensures that the result of the Monte Carlo integration converges almost surely to the correct value when $m \to \infty$. A credible interval for $g(\boldsymbol{\theta})$ can be approximated by

$$[g^{[\frac{\epsilon}{2}m]}, g^{[(1-\frac{\epsilon}{2})m]}], \tag{1.38}$$

where $g^{[i]}$ represents an increasing sorted sequence of $g(\boldsymbol{\theta}^{(i)})$. A HPD interval may be more difficult to obtain, see e.g., Chen et al. (2000).

Generation of samples from the posterior distribution requires a numerical simulation algorithm, usually based on pseudo random numbers generated by a computer. For several well-known distributions, algorithms which generate independent samples are available, see Ripley (1997). For more complex distributions, however, viable methods for direct generation of independent samples do generally not exist. In such cases, Markov chain Monte Carlo (MCMC) simulation can be applied. MCMC simulation has wide applicability and has during the last decade gone into mainstream statistics, especially within the fields of spatial statistics, image analysis and Bayesian statistics.

The idea of MCMC is to generate samples from a sampling distribution $p_s(\boldsymbol{\theta})$, which in our setting is the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{o})$, by simulating a cleverly constructed Markov chain, see e.g., Gilks et al. (1996); Robert and Casella (1999). This is done by constructing a Markov chain with stationary distribution equal to the desired sampling distribution. When the Markov chain has converged to its stationary distribution, each new state of the Markov chain is then also a sample from the desired sampling distribution. The Metropolis-Hastings algorithm, which was proposed by Metropolis et al. (1953) and later generalized by Hastings (1970), gives a very general recipe on con-

struction of Markov chains with the required properties. The pseudo-code of the Metropolis-Hastings algorithm is simple:

Initiate: Set arbitrary $\boldsymbol{\theta}^{(0)}$ where $p_s(\boldsymbol{\theta}^{(0)}) > 0$.

Iterate: $i = 1, 2, \ldots$

- Draw a possible new sample candidate $\tilde{\boldsymbol{\theta}}$ from $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i-1)})$
- Compute acceptance probability

$$p_a = \min\left\{1, \frac{p_s(\tilde{\boldsymbol{\theta}})}{p_s(\boldsymbol{\theta}^{(i-1)})}\frac{q(\boldsymbol{\theta}^{(i-1)}|\tilde{\boldsymbol{\theta}})}{q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^{(i-1)})}\right\}, \tag{1.39}$$

- With probability p_a set $\boldsymbol{\theta}^{(i)} = \tilde{\boldsymbol{\theta}}$, otherwise $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$.

The Metropolis-Hastings algorithm simulates a Markov chain by proposing a sample $\tilde{\boldsymbol{\theta}}$ as candidate for the next element $\boldsymbol{\theta}^{(i)}$ in the Markov chain. The proposed sample is drawn from a proposal density $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i-1)})$, and accepted with probability p_a such that $\boldsymbol{\theta}^{(i)} = \tilde{\boldsymbol{\theta}}$, otherwise $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$. The samples $\boldsymbol{\theta}^{(i)}$ generated by the Metropolis-Hastings algorithm form a Markov chain which converges to $p_s(\boldsymbol{\theta})$ as stationary distribution for any legal initial state $\boldsymbol{\theta}^{(0)}$ and almost any choice of proposal density $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i-1)})$. For a wide range of practical problems, the choice of simple proposal densities give satisfactory results, see e.g., Gilks et al. (1996). For more complicated distributions, for example multi modal distributions with modes separated by low or zero probability zones, careful tuning of the proposal densities may be necessary (Tjelmeland and Hegstad, 2001; Tjelmeland and Eidsvik, 2002).

The convergence rate of the Markov chain for an actual $p_s(\boldsymbol{\theta})$ depends on both the initial state and on the choice of proposal density, but is generally hard to assess. In practice, the convergence of an actual simulation problem can be evaluated by monitoring the involved variables using different initial values (Gilks et al., 1996). The main objective is to determine the burn-in period of the Markov chain, that is the initial period before the Markov chain has converged to its stationary distribution. The samples generated before convergence are usually not used in the Monte Carlo integration. To be strictly correct, if burn-in samples are abandoned, the samples generated by the Metropolis-Hastings algorithm must be re-indexed compared to the Monte Carlo integration in expression (1.37).

An important property of the Metropolis-Hastings algorithm is that the sampling distribution $p_s(\boldsymbol{\theta})$ only appears in the ratio in expression (1.39). The

algorithm can therefore be used to draw samples from the posterior distribution without knowing the exact value of the normalizing constant.

Different classes of algorithms can be defined from the choice of the proposal distribution. If the proposal density is symmetric, $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i-1)}) = q(\boldsymbol{\theta}^{(i-1)}|\boldsymbol{\theta})$, the algorithm reduces to the Metropolis algorithm (Metropolis et al., 1953). If the proposal density does not depend on the current state $\boldsymbol{\theta}^{(i-1)}$, such that $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i-1)}) = q(\boldsymbol{\theta})$, the algorithm is called an independence sampler. The algorithms above suggest a simultaneously sampling of the complete vector $\boldsymbol{\theta}$, but it is often a better procedure to use single-component Metropolis-Hastings to update single elements of $\boldsymbol{\theta}$ one by one, or alternatively blocks of elements.

The popular Gibbs sampler algorithm is also a special case of the general Metropolis-Hastings algorithm, where the full conditional distributions for the elements in $\boldsymbol{\theta}$ given the current state of the remaining elements are used as proposal distributions, such that the proposal density for element θ_j is

$$q(\theta_j|\boldsymbol{\theta}_{-j}^{(i-1)}) = p_s(\theta_j|\boldsymbol{\theta}_{-j}^{(i-1)}), \tag{1.40}$$

where $\boldsymbol{\theta}_{-j}^{(i-1)} = \{\theta_1^{(i)}, \ldots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i-1)}, \ldots, \theta_n^{(i-1)}\}$. In this case, the acceptance probability is $p_a = 1$, that means that a proposed candidate sampled from the full conditional distribution is always accepted. One iteration of the Gibbs sampler algorithm consists of drawing a new sample for each of the elements in $\boldsymbol{\theta}$. Once an element is drawn, it goes into the current state of $\boldsymbol{\theta}$.

The pseudo-code of the Gibbs sampling algorithm can be written

Initiate: Set arbitrary $\boldsymbol{\theta}^{(0)}$ where $p_s(\boldsymbol{\theta}^{(0)}) > 0$.

Iterate: $i = 1, 2, \ldots$

- Draw $\theta_j^{(i)}$ from $p_s(\theta_j|\boldsymbol{\theta}_{-j}^{(i-1)})$ for $j = 1, \ldots, n$.

Instead of drawing single elements from $\boldsymbol{\theta}$ as shown above, a more general algorithm is obtained by allowing for simultaneously drawing a block of elements. It is well known that the performance of the Gibbs sampler may be poor if there are strong correlations between some of the elements in $\boldsymbol{\theta}$. If samples from the full conditional joint distribution for a block of correlated elements can be provided, such blocking may significantly improve the Gibbs sampling. There is no restrictions on the size of a block as long as samples can be generated from the full conditional joint distribution.

1.4 Seismic inversion

The problem of finding the cause of an observed effect is generally termed inversion (Tarantola, 1987; Scales and Snieder, 2000). The opposite problem is to calculate the effect of a given cause, that is the direct or forward modeling problem. In seismic exploration, a seismic wave is initiated by firing a seismic source, and the waves reflected from the subsurface are recorded on a set of receivers. In marine seismic acquisition, a seismic vessel tows the source, usually an air gun array, and one or several streamers containing hydrophones. The wave propagation through the subsurface rocks is physically described by the wave equation. The calculation of synthetic seismic data for a known earth model is the seismic forward modeling problem, while seismic inversion aims at estimating material properties of the earth model from the seismic data.

The wave equation

Seismic wave propagation in an anelastic, anisotropic, and inhomogeneous earth medium can be described by a differential vector equation which gives the direction and quantity of the energy transport at every location $\boldsymbol{x}$ in the medium as a function of time t, see e.g., Aki and Richards (1980); Tarantola (1988). The fundamental basis for the wave equation is Newton's law of motion, represented by

$$\rho(\boldsymbol{x})\frac{\partial^2 u_i}{\partial t^2}(\boldsymbol{x},t) = \frac{\partial \sigma_{ij}}{\partial x_j}(\boldsymbol{x},t) + f_i(\boldsymbol{x},t), \tag{1.41}$$

in coordinate direction i, and a stress-strain relation on the form

$$\sigma_{ij}(\boldsymbol{x},t) = \int c_{ijkl}(\boldsymbol{x},t-\tau)\,\epsilon_{kl}(\boldsymbol{x},\tau)\,d\tau, \tag{1.42}$$

where ρ is the density, u_i is particle displacement, σ_{ij} is the stress field, f_i is body force density, c_{ijkl} is the viscoelastic Hooke's tensor, and ϵ_{ij} is the strain tensor defined by

$$\epsilon_{ij}(\boldsymbol{x},t) = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}(\boldsymbol{x},t) + \frac{\partial u_j}{\partial x_i}(\boldsymbol{x},t)\right). \tag{1.43}$$

Repeated latin indexes are summed over according to Einstein's sum convention. Combination of expressions (1.41) through (1.43) gives the wave equation for particle displacement in an anelastic, anisotropic, and inhomogeneous

medium:

$$\rho(\boldsymbol{x})\frac{\partial^2 u_i}{\partial t^2}(\boldsymbol{x},t) = \frac{\partial}{\partial x_j}\int c_{ijkl}(\boldsymbol{x},t-\tau)\,\frac{\partial u_k}{\partial x_l}(\boldsymbol{x},\tau)\,d\tau + f_i(\boldsymbol{x},t). \tag{1.44}$$

In an elastic medium, the stress-strain relation in expression (1.42) is reduced to Hooke's law:

$$\sigma_{ij}(\boldsymbol{x},t) = c_{ijkl}(\boldsymbol{x})\,\epsilon_{kl}(\boldsymbol{x},t). \tag{1.45}$$

The general elastic Hooke's tensor $c_{ijkl}(\boldsymbol{x})$ has $3^4 = 81$ components, but the number of independent elastic coefficients can be reduced to 21 for an arbitrary anisotropic medium (Aki and Richards, 1980).

Combination of expressions (1.41) and (1.45) gives the elastic wave equation for particle displacement

$$\rho(\boldsymbol{x})\frac{\partial^2 u_i}{\partial t^2}(\boldsymbol{x},t) = \frac{\partial}{\partial x_j}\left(c_{ijkl}(\boldsymbol{x})\frac{\partial u_k}{\partial x_l}(\boldsymbol{x},t)\right) + f_i(\boldsymbol{x},t). \tag{1.46}$$

For an isotropic, elastic medium, the elasticity is determined by only two elastic parameters, and the Hooke's tensor can be written

$$c_{ijkl}(\boldsymbol{x}) = \lambda(\boldsymbol{x})\delta_{ij}\delta_{kl} + \mu(\boldsymbol{x})(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}), \tag{1.47}$$

where $\lambda(\boldsymbol{x})$ and $\mu(\boldsymbol{x})$ are the Lamé parameters and δ_{ij} is the Kronecker symbol. In this case, a complete description of the material properties is given by $\{\lambda(\boldsymbol{x}), \mu(\boldsymbol{x}), \rho(\boldsymbol{x})\}$. Several other parameterizations can also be used, for example $\{\alpha(\boldsymbol{x}), \beta(\boldsymbol{x}), \rho(\boldsymbol{x})\}$, where

$$\alpha = \sqrt{\frac{\lambda + 2\mu}{\rho}} \tag{1.48}$$

is the velocity of a P-wave (compressional wave), and

$$\beta = \sqrt{\frac{\mu}{\rho}} \tag{1.49}$$

is the velocity of an S-wave (shear wave). The material properties can also be described by impedances, $\{Z_P(\boldsymbol{x}), Z_S(\boldsymbol{x}), \rho(\boldsymbol{x})\}$, where $Z_P = \alpha\rho$ is the acoustic impedance, and $Z_S = \beta\rho$ is the shear impedance. The shear impedance is often substituted by the P-wave velocity to S-wave velocity ratio α/β.

Seismic forward modeling

When an earth model is described, for example by $\{\alpha(\boldsymbol{x}), \beta(\boldsymbol{x}), \rho(\boldsymbol{x})\}$ for elastic isotropic rocks, and the seismic source signature is known, a synthetic seismogram can be computed by modeling the wave propagation according to the wave equation. The solution of the wave equation is non-trivial except for simple, idealized situations, but several numerical techniques can be used to find approximate solutions. The finite difference (FD) method (Madariaga, 1986; Virieux, 1984, 1986; Levander, 1988; Emmerich and Korn, 1987; Carcione et al., 1988) and the finite element method (FEM) (Lysmer and Drake, 1972; Marfurt, 1984) are brute force techniques giving the complete solution of the wave equation. The accuracy of the solution is primarily related to the discretization of a continuum by a grid with finite extent.

Ray tracing methods (Červený et al., 1977) do not give the complete solution of the wave equation, but the sum of prespecified events, for example primary reflected waves reflected from a set of selected reflectors. The ray path depends on the initial ray direction and the wave propagation velocity. When a ray passes an interface between two layers with different wave velocity, the ray bending at the interface is given by Snell's law. When a wave hits an interface between two layers, some of the energy of the incident wave is reflected. The reflection coefficient is usually defined as the ratio of the amplitude of the reflected wave to the amplitude of the incident wave, and depends on the material properties and the ray angle relative to the interface normal. The exact expression for the reflection coefficient was derived by Zoeppritz (1919), and is known as the Zoeppritz equation.

In the special case of a horizontally layered earth model, the reflectivity method (Kennett, 1983; Ursin, 1983, 1987) can be used. The reflectivity method models a synthetic seismogram for a stack of layers in a 1-D earth model with laterally constant material properties.

The modeling of synthetic seismic data is time consuming, especially 3-D FD and FEM modeling, and the relationship between the earth model and the corresponding seismic data is strongly nonlinear. Usually, the recorded seismic data go through an extensive processing before the data are used in analysis and interpretation. Modern true amplitude processing aims at providing geometrically correct images of the subsurface structures where the amplitudes are proportional to the reflection strength if interference effects are neglected.

Unwanted effects from the acquisition system, different kind of noise, and unwanted wave propagation effects are removed or accounted for by seismic processing. Examples are the removal of multiples and diffractions, and amplitude recovery due to geometrical spreading and attenuation. The data are sorted and migrated towards their correct spatial positions, including correction for offset dependent traveltime. Usually, the depth coordinate is represented by the vertical seismic two-way traveltime, but depth migration will probably be more common in the future. The migration process effectively transforms the 3-D relationship between the earth model and the seismic data to vertical 1-D relationships for each surface location. This is approximate for time migration, but still an adequate assumption, except for very complex earth models with strong velocity variations.

A forward modeling algorithm can be constructed such that processed seismic data are simulated directly. Such modeling are often used for a target zone of special interest, often with a simplified model for the overburden. True amplitude processed seismic data in a target zone are usually represented by the convolutional model. The convolutional model in continuous form for a single seismic time trace can be written

$$d_{obs}(t) = \int s(\tau)\, c(t-\tau)\, d\tau + e(t), \tag{1.50}$$

where $d_{obs}(t)$ is the seismic data as function of the vertical two-way traveltime t, s is the wavelet, c is the reflectivity, and e is an error term. The wavelet may be estimated from the seismic data in the target zone under certain assumptions, or from seismic and well log data. The reflectivity c is determined by the nonlinear Zoeppritz equation, but if c is linearized, a linear relationship between the seismic data and the earth model is established.

Over the years, a number of approximations to the nonlinear Zoeppritz equation has been made. One of the first was Bortfeld (1961), and further modifications were done by Richards and Frasier (1976) and by Aki and Richards (1980). The approximate reflection coefficient given by Aki and Richards (1980) is linear with respect to relative contrasts in the material parameters under weak contrast assumption. The reflection coefficient is a function of the angle of incidence relative to the interface normal. On seismic prestack data, this can be observed as amplitude versus offset (AVO) variation (Ostrander, 1984), where offset is the distance between the source and the receiver. AVO analysis is alternatively termed AVA (amplitude versus angle) analysis.

Classical seismic inversion

An earth model may be defined by a vector field $\boldsymbol{m}(\boldsymbol{x})$, for example related to $\{\alpha(\boldsymbol{x}), \beta(\boldsymbol{x}), \rho(\boldsymbol{x})\}$, or represented by a discretized model parameter vector $\boldsymbol{m}$. The seismic data are available on discrete form, represented by the vector $\boldsymbol{d}_{obs}$. The seismic forward model can generally be written

$$\boldsymbol{d}_{obs} = \boldsymbol{g}(\boldsymbol{m}) + \boldsymbol{e}, \tag{1.51}$$

where $\boldsymbol{g}$ is a (possible nonlinear) modeling operator and $\boldsymbol{e}$ is an error term. The error term represents both random noise and systematic errors related to effects which is not included in the modeling operator (Scales and Snieder, 1998).

The aim of seismic inversion is to estimate $\boldsymbol{m}$ from $\boldsymbol{d}_{obs}$, see e.g., Tarantola (1987). Classical seismic inversion is often formulated as optimization of an objective function E, for example the squared data error

$$E = \|\boldsymbol{d}_{obs} - \boldsymbol{g}(\boldsymbol{m})\|^2. \tag{1.52}$$

An optimal solution may be the $\boldsymbol{m}$ which minimizes E.

Seismic inverse problems are often ill-posed, and usually need to be stabilized. The stabilization can be achieved by constraining the solution. The constraints may be imposed by including a penalty function to the objective function, for example

$$E = \|\boldsymbol{d}_{obs} - \boldsymbol{g}(\boldsymbol{m})\|^2 + \lambda\|\boldsymbol{m} - \boldsymbol{m}_0\|^2, \tag{1.53}$$

where $\boldsymbol{m}_0$ is a prior guess on a reasonable solution, and λ is the weight of the penalty term relative to the data misfit term.

The objective function is often related to the maximum likelihood or the maximum posterior concepts. If the error term $\boldsymbol{e}$ in expression (1.51) is assumed to be zero-mean Gaussian with covariance $\boldsymbol{\Sigma}_e$,

$$\boldsymbol{e} \sim \mathcal{N}_{n_d}(0, \boldsymbol{\Sigma}_e), \tag{1.54}$$

the maximum likelihood solution is defined by minimizing

$$E = \frac{1}{2}(\boldsymbol{d}_{obs} - \boldsymbol{g}(\boldsymbol{m}))^T \boldsymbol{\Sigma}_e^{-1}(\boldsymbol{d}_{obs} - \boldsymbol{g}(\boldsymbol{m})). \tag{1.55}$$

This is identical to a (possible nonlinear) weighted least square problem. If the error term represents white noise such that $\boldsymbol{\Sigma}_e = \sigma_e^2 \boldsymbol{I}$, the problem reduces to an ordinary (nonlinear) least square problem.

If the prior knowledge and uncertainty about $\boldsymbol{m}$ are summarized by a Gaussian prior distribution,

$$\boldsymbol{m} \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \tag{1.56}$$

the maximum posterior solution is defined by minimizing

$$E = \frac{1}{2}(\boldsymbol{d}_{obs} - \boldsymbol{g}(\boldsymbol{m}))^T \boldsymbol{\Sigma}_e^{-1}(\boldsymbol{d}_{obs} - \boldsymbol{g}(\boldsymbol{m})) + \frac{1}{2}(\boldsymbol{m} - \boldsymbol{\mu}_m)^T \boldsymbol{\Sigma}_m^{-1}(\boldsymbol{m} - \boldsymbol{\mu}_m). \tag{1.57}$$

This is equivalent to a (nonlinear) stabilized weighted least square problem, where the prior model acts as the stabilizer. If $\boldsymbol{\Sigma}_e = \sigma_e^2 \boldsymbol{I}$ and $\boldsymbol{\Sigma}_m = \sigma_m^2 \boldsymbol{I}$, the maximum posterior problem equals minimizing E in expression (1.53) with $\boldsymbol{m}_0 = \boldsymbol{\mu}_m$ and $\lambda = \sigma_e^2/\sigma_m^2$.

In classical seismic inversion, the optimal solution with respect to an objective function is estimated by iterative search methods such as steepest descent, conjugated gradients, Gauss-Newton or the Levenberg-Marquard method (Lines and Treitel, 1984). A serious problem with these local search methods is that the search could be trapped in local optima. A possible way around this problem is provided by global optimization methods such as simulated annealing (Metropolis et al., 1953; Kirkpatrick et al., 1983; Geman and Geman, 1984; Rothman, 1986) and genetic algorithms (Goldberg, 1989; Sen and Stoffa, 1996), but these methods may be extremely time consuming. However, both the local and the global optimization methods only provide a point estimate of the solution, and do not assess the associated uncertainty.

Bayesian seismic inversion

Seismic data are usually strongly affected by noise and measurement uncertainty, and seismic inverse problems are in general multidimensional and ill-posed. An ill-posed inverse problem may not have any solutions, which is typical for an over-determined problem, or the solution may not be unique, which is typical for under-determined problems. Nonuniqueness means that several models fit the measured data equally well. The earth parameters vary continuously in three spatial directions, leading to an infinite dimensional problem. In a realistic seismic exploration situation, only a limited amount of data are available, making the inverse problem inherently nonunique. Another aspect of nonuniqueness is related to identifiability of the parameters from the data. This problem arises when more than one parameter set has the same forward response. Further, an ill-posed problem may be unstable, that means that the solution is highly sensitive to slight perturbations in the data.

A Bayesian setting is a natural choice for many geophysical inverse problems (Tarantola and Valette, 1982; Tarantola, 1987; Duijndam, 1988a,b; Omre and Tjelmeland, 1997; Ulrych et al., 2001; Scales and Tenorio, 2001; Kolbjørnsen, 2002). The solution of a Bayesian inverse problem is represented by the complete posterior distribution, which means that the solution is not limited to a single best-fitting set of model parameters, but it does also characterize the uncertainty of the inversion results. In Bayesian inversion, it is possible to combine available prior knowledge with the information contained in the measured data. The use of a prior model has been criticized by non-Bayesians where the primary argument is that subjective prior information may lead to a nonobjective solution. In many geophysical inversion problems, however, the prior distribution serves as a necessary stabilizator to ensure a physically reasonable solution. Further, when prior knowledge is available, for example in form of well log measurements or general geological knowledge, this prior knowledge can naturally be included via the prior distribution.

For nonlinear problems, even with Gaussian prior and Gaussian likelihood models, the posterior distribution can only be found by use of stochastic simulation techniques (Monte Carlo) (Mosegaard and Tarantola, 1995; Mosegaard, 1998; Eide et al., 2002). However, stochastic simulation of inverse problems involving seismic data is in most cases impractical due to high computational costs. For linear problems, or problems that can be linearized, analytical solutions may in some cases be found. A solution with an explicit analytical form is usually computationally superior to iterative search and simulation based solutions.

1.5 Outline of the papers

The main part of the thesis consists of four separate self-contained papers written for geophysical journals. The papers may be read independently, but the order listed below is strongly recommended. The main subject is related to the uncertainty aspects of seismic amplitude versus offset (AVO) inversion, where the uncertainty is assessed using a Bayesian approach to inversion. The objective of the AVO inversion is to estimate elastic material parameters from large seismic data sets, but the complete inversion problem also includes estimation of seismic wavelets and the noise level. To be able to invert large seismic data sets, a crucial aspect is the computational efficiency of the inversion algorithm.

Properly processed seismic data are used such that the AVO inversion problem can be linearized, and the inversion algorithm utilizes the convenient mathematical properties obtained by Gaussian assumptions. The wavelet estimation and noise level estimation involve well log data and seismic data only at the well locations. This allows for more advanced and time consuming techniques like MCMC. In the examples, only one or two wells are used to illustrate the methods, but more wells should be used to increase the accuracy. Real data examples are shown in all papers.

The paper in Chapter 2, **Bayesian linearized AVO inversion**, presents a linearized AVO inversion technique based on the convolutional model and log-Gaussian prior and Gaussian noise assumptions. The wavelet and the noise covariance are in this paper assumed to be known. The objective of the AVO inversion is to obtain posterior distributions for P-wave velocity, S-wave velocity, and density from processed seismic data. A new element with this method is related to the implicit trace integration which transforms the reflectivity information to material properties in a stochastic setting. The solution is represented by a Gaussian posterior distribution with explicit expressions for the posterior expectation and covariance, which provides a computationally fast inversion method. The inversion algorithm is applied to a real 3-D dataset from the Sleipner Field.

The paper in Chapter 3, **Bayesian wavelet estimation from seismic and well data**, presents a new approach to wavelet and noise level estimation, where the uncertainties are an integral part of the solution. The method is based on the convolutional model, where the reflectivity is calculated from the well logs. Possible mistie between the seismic traveltimes and the time axis of the well logs, errors in the log measurements, and seismic noise are included in the model. The stochastic model is illustrated by a directed acyclic graph (DAG). The solution is obtained by Markov chain Monte Carlo simulation, and the method is illustrated through a real data example.

The paper in Chapter 4, **Joint AVO inversion, wavelet estimation, and noise level estimation using a spatially coupled hierarchical Bayesian model**, presents a method for joint AVO inversion, wavelet estimation, and estimation of the noise level, and is a combination and extension of the two previous papers in chapters 2 and 3. A spatial coupling of the model parameters is now imposed via the prior distribution by a spatial correlation function. The stochastic model includes uncertainty of both the elastic parameters, the wavelet, and the seismic and well log data. The posterior distribution is

explored by Markov chain Monte Carlo simulation using the Gibbs sampler algorithm. The inversion algorithm has been tested on a seismic line from the Heidrun Field with two wells located on the line.

The paper in Chapter 5, **Rapid spatially coupled AVO inversion in the Fourier domain**, presents a fast and exact method to solve high dimensional spatially coupled inverse problems sampled on a regular grid. The method utilizes the fact that a circulant matrix can be diagonalized by the Fourier transform. Under certain assumptions, we show how the spatially coupled inversion problem can be decoupled in the Fourier domain, such that the inversion problem can be solved independently for each frequency component. The method is applied to a 3-D spatially coupled AVO inversion problem. The computing time for a 3-D inversion example from the Sleipner Field with 12 million unknown model parameters was only some few minutes.

Chapter 2

Bayesian linearized AVO inversion

Arild Buland and Henning Omre

Paper accepted for publication in Geophysics, 2002. Presented at the 62nd EAGE conference, Glasgow 2000 (Extended abstract).

Abstract

A new linearized AVO inversion technique is developed in a Bayesian framework. The objective is to obtain posterior distributions for P-wave velocity, S-wave velocity, and density. Distributions for other elastic parameters can also be assessed, for example acoustic impedance, shear impedance, and P-wave to S-wave velocity ratio. The inversion algorithm is based on the convolutional model and a linearized weak contrast approximation of the Zoeppritz equation. The solution is represented by a Gaussian posterior distribution with explicit expressions for the posterior expectation and covariance, hence exact prediction intervals for the inverted parameters can be computed under the specified model. The explicit analytical form of the posterior distribution provides a computationally fast inversion method. Tests on synthetic data show that all inverted parameters were almost perfectly retrieved when the noise approached zero. With realistic noise levels, acoustic impedance was the best determined parameter, while the inversion provided practically no information about the density. The inversion algorithm has also been tested on a real 3-D dataset from the Sleipner Field. The results show good agreement with well logs, but the uncertainty is high.

2.1 Introduction

The objectives of geophysical inverse problems are to make inferences about model parameters based on general knowledge and a set of geophysical measurements. In general, these inverse problems are multidimensional and ill-posed, and they are often strongly affected by noise and measurement uncertainty.

In a statistical perspective, the solution of an inverse problem is not limited to a single set of predicted parameters, but is represented by a probability density function (pdf) on the model space. The aim of inversion is not only to find a best-fitting set of model parameters, but also to characterize the uncertainty in the inversion results. A Bayesian setting is a natural choice for many geophysical inverse problems, where it is possible to combine available prior knowledge with the information contained in the measured data (Tarantola and Valette, 1982; Tarantola, 1987; Duijndam, 1988a,b; Ulrych et al., 2001; Scales and Tenorio, 2001). The solution of a Bayesian inverse problem is represented by the posterior distribution. Analysis of nonuniqueness and uncertainty in the solutions of inverse problems is often called resolution analysis. In a Bayesian setting, all questions about resolution are addressed by the posterior distribution. The Bayesian concept is general and recognized both in statistics and in geophysics, but the specific model definition and the associated solution of an actual problem may be complicated.

A general geophysical inverse problem has a nonlinear relationship between the model parameters and the measured data, and the posterior distribution can usually be found only by use of stochastic simulation techniques (Monte Carlo) (Mosegaard and Tarantola, 1995; Sen and Stoffa, 1996; Mosegaard, 1998; Eide et al., 2002). However, stochastic simulation of inverse problems involving seismic data is in most cases impractical due to high computational costs. For linear problems, or problems that can be linearized, analytical solutions may in some cases be found. A solution with an explicit analytical form is usually computationally superior to iterative search and simulation based solutions.

Amplitude versus offset (AVO) inversion is a seismic prestack inversion technique for estimating elastic subsurface parameters. AVO inversion utilizes the fact that the reflection strength from subsurface interfaces depends on the reflection angle and on the material properties where the reflections take place. The AVO inversion methods can be separated into nonlinear methods (Dahl

and Ursin, 1991; Buland et al., 1996; Gouveia and Scales, 1997, 1998), and linearized inversion (Smith and Gidlow, 1987; Lörtzer and Berkhout, 1993). Gouveia and Scales (1998) defined a Bayesian nonlinear model, and optimized the posterior distribution via a nonlinear conjugated gradient procedure to find the maximum a posteriori model (MAP). The uncertainty analysis was performed by making a Gaussian approximation of the posterior distribution centered at the MAP solution. This approach is much faster than using Monte Carlo simulation, but since the problem is complex and strongly nonlinear, the method is still computationally intensive compared to linearized AVO inversion. In the linearized AVO inversion methods, the Zoeppritz equation for the reflection coefficient is linearized assuming weak contrasts. The seismic data must be processed prior to linear AVO inversion to remove nonlinear relations between the model parameters and the seismic response, such as offset dependent traveltime and geometrical spreading.

In this paper, a Bayesian linearized AVO inversion method is presented, where the objective is to obtain posterior distributions for the P-wave velocity, S-wave velocity, and density. Closest to our approach is the method presented in Lörtzer and Berkhout (1993). Their inversion algorithm, however, was formulated for relative contrasts of the elastic parameters, and not for the elastic parameters themselves. Transformation to the elastic parameters may not be simple in a stochastic setting. Lörtzer and Berkhout used standard AVO techniques based on single interface theory, which are then applied independently on a sample by sample basis. We will base our inversion algorithm on a weak contrast reflectivity function defined for continuous seismic traveltime. The inversion problem is solved simultaneously for all times in a given time window, which makes it possible to include the wavelet by convolution, and temporal correlation between model parameters close in time. The solution is represented by a Gaussian posterior distribution with explicit expressions for the posterior expectation and covariance. The explicit analytical form of the posterior distribution provides a computationally fast inversion method. A continuous definition of the elastic parameters makes it possible to integrate various types of observations with varying sampling density and sampling support. Moreover, generalization to a full 3-D model with spatial dependencies is in principle straightforward, but an efficient numerical technique is required.

The inversion result is obtained from true amplitude processed seismic prestack data. Important steps in the processing are the removal of multiples and corrections for the effects of geometrical spreading and absorption. We assume

that wave mode conversions, interbed multiples, and anisotropy effects can be neglected after processing. The data should be prestack migrated, such that dip related effects and the moveouts are removed. After prestack migration, we assume that each single bin gather can be regarded as the response of a local 1-D earth model. Finally, the gathers must be transformed from offsets to angles, either by common-angle migration, or by a transform after the migration.

In the following sections, the methodology, synthetic tests, and the application of the inversion method on a real 3-D dataset are presented.

2.2 Methodology

The seismic reflection coefficients depend on the material properties of the subsurface. An isotropic, elastic medium is completely described by three material parameters. We have chosen $\{\alpha(t), \beta(t), \rho(t)\}$, where $\alpha(t)$, $\beta(t)$, and $\rho(t)$ are P-wave velocity, S-wave velocity, and density, and t is the two-way vertical seismic traveltime. Several other parameterizations can also be used, for example $\{Z_P(t), Z_S(t), \rho(t)\}$, where $Z_P = \alpha\rho$ is the acoustic impedance, and $Z_S = \beta\rho$ is the shear impedance. The shear impedance is often substituted by the P-wave velocity to S-wave velocity ratio α/β. When inversion results are obtained for a chosen parameter set, the corresponding results can be generated for another parameter set, including uncertainty.

The inversion method is based on a weak contrast approximation to the PP reflection coefficient (Aki and Richards, 1980):

$$c_{PP}(\theta) = a_\alpha(\theta)\frac{\Delta\alpha}{\bar{\alpha}} + a_\beta(\theta)\frac{\Delta\beta}{\bar{\beta}} + a_\rho(\theta)\frac{\Delta\rho}{\bar{\rho}}, \tag{2.1}$$

where

$$a_\alpha(\theta) = \frac{1}{2}\left(1 + \tan^2\theta\right), \tag{2.2}$$

$$a_\beta(\theta) = -4\frac{\bar{\beta}^2}{\bar{\alpha}^2}\sin^2\theta, \tag{2.3}$$

$$a_\rho(\theta) = \frac{1}{2}\left(1 - 4\frac{\bar{\beta}^2}{\bar{\alpha}^2}\sin^2\theta\right), \tag{2.4}$$

and $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\rho}$ are average P-wave velocity, S-wave velocity, and density over the reflecting interface, $\Delta\alpha$, $\Delta\beta$ and $\Delta\rho$ are the corresponding contrasts, and

θ is the reflection angle. Only PP reflectivity is discussed in this paper, but the extension to include converted waves is straightforward.

The single interface reflection coefficient in equation (2.1) can be extended to a time continuous reflectivity function (Stolt and Weglein, 1985):

$$c_{PP}(t,\theta) = a_\alpha(t,\theta)\,\frac{\partial}{\partial t}\ln\alpha(t) + a_\beta(t,\theta)\,\frac{\partial}{\partial t}\ln\beta(t) + a_\rho(t,\theta)\,\frac{\partial}{\partial t}\ln\rho(t), \quad (2.5)$$

where $a_\alpha(t,\theta)$, $a_\beta(t,\theta)$, and $a_\rho(t,\theta)$ are generalizations of the coefficients in expressions (2.2)-(2.4) with time dependent velocities $\bar{\alpha}(t)$ and $\bar{\beta}(t)$. We assume that $\bar{\alpha}(t)$ and $\bar{\beta}(t)$ can be represented by a constant or slowly varying known background model, such that $\bar{\alpha}(t)$ and $\bar{\beta}(t)$ are the average or moving average of $\alpha(t)$ and $\beta(t)$ in a time window. The inversion algorithm requires that this background model is known prior to the inversion, such that the coefficients can be precalculated.

In the formulas above, the reflection angle θ is used as an independent variable. However, the seismic data are recorded as a function of source-receiver distance h (offset). The transform of the data from the (t,h)-domain to the (t,θ)-domain depends on the velocity function. This transform can be performed by common-angle migration, ray-tracing, or standard approximate offset-angle relations based on a smooth background model.

The prior model

In a Bayesian setting, the prior model defines a statistical model for the prior information of the material parameters, expressed mathematically by probability density functions (pdf). The prior information must be independent of the seismic data, and is established from other available information and knowledge. It is in a good statistical tradition to be vague when specifying the prior model.

The material parameters $\{\alpha(t),\beta(t),\rho(t)\}$ are a priori assumed to be log-Gaussian, which implies that the parameters are restricted to take positive values. This assumption is required for later analytical treatment due to expression (2.5). The logarithm of these material parameters defines a continuous Gaussian vector field

$$\boldsymbol{m}(t) = [\ln\alpha(t), \ln\beta(t), \ln\rho(t)]^T, \quad (2.6)$$

with expectation

$$\mathrm{E}\{\boldsymbol{m}(t)\} = \boldsymbol{\mu}(t) = [\mu_\alpha(t), \mu_\beta(t), \mu_\rho(t)]^T, \tag{2.7}$$

where the elements $\mu_\alpha(t)$, $\mu_\beta(t)$, and $\mu_\rho(t)$ are the expectations of $\ln\alpha(t)$, $\ln\beta(t)$, and $\ln\rho(t)$, respectively. We assume that the expectation functions are smooth. The covariances between $\ln\alpha$, $\ln\beta$, and $\ln\rho$ at times t and s are denoted

$$\mathrm{Cov}\{\boldsymbol{m}(t), \boldsymbol{m}(s)\} = \boldsymbol{\Sigma}(t, s). \tag{2.8}$$

A simple example is the stationary covariance function

$$\boldsymbol{\Sigma}(t, s) = \boldsymbol{\Sigma}_0 \; \nu_t(\tau), \tag{2.9}$$

where $\nu_t(\tau)$ is a temporal correlation function, and τ is the time-lag $\tau = |t - s|$. The time invariant covariance matrix $\boldsymbol{\Sigma}_0$ is

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_\alpha^2 & \sigma_\alpha\sigma_\beta\nu_{\alpha\beta} & \sigma_\alpha\sigma_\rho\nu_{\alpha\rho} \\ \sigma_\alpha\sigma_\beta\nu_{\alpha\beta} & \sigma_\beta^2 & \sigma_\beta\sigma_\rho\nu_{\beta\rho} \\ \sigma_\alpha\sigma_\rho\nu_{\alpha\rho} & \sigma_\beta\sigma_\rho\nu_{\beta\rho} & \sigma_\rho^2 \end{bmatrix}, \tag{2.10}$$

where the diagonal elements are the variances, and $\nu_{\alpha\beta}$, $\nu_{\alpha\rho}$, and $\nu_{\beta\rho}$, are the correlations between $\ln\alpha(t)$, $\ln\beta(t)$ and $\ln\rho(t)$, respectively. The temporal correlation function $\nu_t(\tau)$ must be a positive definite function, take values in the interval $[-1, 1]$, and have the property that $\nu_t(0) = 1$. One such correlation function is the second order exponential function

$$\nu_t(\tau) = \exp\left[-\left(\frac{\tau}{d}\right)^2\right], \tag{2.11}$$

where d is a range parameter characterizing the temporal dependency.

The time derivative of $\boldsymbol{m}(t)$ defines a new vector field

$$\boldsymbol{m}'(t) = \left[\frac{\partial}{\partial t}\ln\alpha(t), \frac{\partial}{\partial t}\ln\beta(t), \frac{\partial}{\partial t}\ln\rho(t)\right]^T, \tag{2.12}$$

where the elements can be recognized in the reflectivity function, expression (2.5). This is also a Gaussian vector field due to the linearity of the differentiation process (Christakos, 1992). The expectation of $\boldsymbol{m}'(t)$,

$$\mathrm{E}\{\boldsymbol{m}'(t)\} = \frac{\partial}{\partial t}\boldsymbol{\mu}(t) = \boldsymbol{\mu}'(t), \tag{2.13}$$

and the covariance

$$\mathrm{Cov}\{\boldsymbol{m}'(t), \boldsymbol{m}'(s)\} = \frac{\partial^2}{\partial t\, \partial s}\boldsymbol{\Sigma}(t,s) = \boldsymbol{\Sigma}''(t,s), \tag{2.14}$$

are both derived from the specified prior model for $\boldsymbol{m}(t)$. The stochastic model for the differentiated field $\boldsymbol{m}'(t)$ is needed later, but discussed here due to the close relationship to $\boldsymbol{m}(t)$. The cross-covariance between $\boldsymbol{m}'(t)$ and $\boldsymbol{m}(s)$ is

$$\mathrm{Cov}\{\boldsymbol{m}'(t), \boldsymbol{m}(s)\} = \frac{\partial}{\partial t}\boldsymbol{\Sigma}(t,s) = \boldsymbol{\Sigma}'(t,s). \tag{2.15}$$

For the stationary covariance function in expression (2.9), the covariance in expression (2.14) and the cross-covariance in expression (2.15) are

$$\mathrm{Cov}\{\boldsymbol{m}'(t), \boldsymbol{m}'(s)\} = -\boldsymbol{\Sigma}_0 \frac{\partial^2}{\partial \tau^2}\nu_t(\tau), \tag{2.16}$$

and

$$\mathrm{Cov}\{\boldsymbol{m}'(t), \boldsymbol{m}(s)\} = \mathrm{sign}(t-s)\boldsymbol{\Sigma}_0 \frac{\partial}{\partial \tau}\nu_t(\tau), \tag{2.17}$$

respectively, where $\mathrm{sign}(\cdot)$ returns the sign of the argument.

The continuous form of the Gaussian field $\boldsymbol{m}(t)$ makes it possible to give a proper definition of the time differentiated Gaussian field $\boldsymbol{m}'(t)$. In a computer program, the continuous fields are represented on a grid. The grid density should be determined by the temporal variability of the elastic parameters, and not by the sampling density of the seismic data. A discrete representation of $\boldsymbol{m}(t)$ in a time interval is Gaussian:

$$\boldsymbol{m} = [\ln \boldsymbol{\alpha}^T, \ln \boldsymbol{\beta}^T, \ln \boldsymbol{\rho}^T]^T \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \tag{2.18}$$

where the vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\rho}$ are discrete representations of $\alpha(t)$, $\beta(t)$, and $\rho(t)$, respectively. The logarithm operates element wise on the vector elements, $\sim$ means "distributed as", and $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes an n-dimensional Gaussian distribution with expectation $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, see Appendix 2.A. The dimension of $\boldsymbol{m}$ is n_m, and $\boldsymbol{\mu}_m$ and $\boldsymbol{\Sigma}_m$ are defined from the continuous analogs $\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(t,s)$, respectively. A discrete representation of the differentiated field $\boldsymbol{m}'(t)$ in a time interval is Gaussian:

$$\boldsymbol{m}' \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}'_m, \boldsymbol{\Sigma}''_m), \tag{2.19}$$

with $\boldsymbol{\mu}'_m$ and $\boldsymbol{\Sigma}''_m$ being defined from the continuous analogs in expressions (2.13) and (2.14), respectively. A discrete version of the cross-covariance in expression (2.15) is denoted $\boldsymbol{\Sigma}'_m$.

Seismic forward modeling

In Bayesian AVO inversion, the likelihood function can be defined through a pdf for the observed seismic data $\boldsymbol{d}_{obs}$ given a specific parameter vector $\boldsymbol{m}$, denoted $p(\boldsymbol{d}_{obs}|\boldsymbol{m})$. The likelihood function describes how likely the observed data are for these specific parameters. The unconditional pdf for the seismic data, denoted $p(\boldsymbol{d}_{obs})$, is related to the likelihood function by

$$p(\boldsymbol{d}_{obs}) = \int p(\boldsymbol{d}_{obs}|\boldsymbol{m})p(\boldsymbol{m})\, d\boldsymbol{m}, \tag{2.20}$$

where $p(\boldsymbol{m})$ is the prior model for the parameter vector $\boldsymbol{m}$. Below, the pdf $p(\boldsymbol{d}_{obs})$ is defined and will later be used to derive the posterior distribution $p(\boldsymbol{m}|\boldsymbol{d}_{obs})$.

A discrete version of the reflectivity function $c_{PP}(t,\theta)$, expression (2.5), for a given time interval and a set of reflection angles can be written

$$\boldsymbol{c} = \boldsymbol{A}\boldsymbol{m}', \tag{2.21}$$

where the matrix $\boldsymbol{A}$ is defined by the coefficients $a_\alpha(t,\theta)$, $a_\beta(t,\theta)$ and $a_\rho(t,\theta)$, see Appendix 2.B. The corresponding seismic angle gather $\boldsymbol{d}_{obs}$ is represented by the convolutional model,

$$\boldsymbol{d}_{obs} = \boldsymbol{S}\boldsymbol{c} + \boldsymbol{e} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{m}' + \boldsymbol{e}, \tag{2.22}$$

where the convolution is formulated by the matrix-vector multiplication $\boldsymbol{S}\boldsymbol{c}$. The matrix $\boldsymbol{S}$ represents one wavelet for each reflection angle, see Appendix 2.B. We assume that the error term $\boldsymbol{e}$ is zero mean Gaussian

$$\boldsymbol{e} \sim \mathcal{N}_{n_d}(\boldsymbol{0}, \boldsymbol{\Sigma}_e), \tag{2.23}$$

and independent of $\boldsymbol{m}$, then $\boldsymbol{d}_{obs}$ is Gaussian

$$\boldsymbol{d}_{obs} \sim \mathcal{N}_{n_d}(\boldsymbol{\mu}_{d_{obs}}, \boldsymbol{\Sigma}_{d_{obs}}), \tag{2.24}$$

where

$$\boldsymbol{\mu}_{d_{obs}} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{\mu}'_m, \tag{2.25}$$

$$\boldsymbol{\Sigma}_{d_{obs}} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{\Sigma}''_m\boldsymbol{A}^T\boldsymbol{S}^T + \boldsymbol{\Sigma}_e, \tag{2.26}$$

using Result 1 in Appendix 2.A.

The posterior model

Combination of equations (2.18) and (2.24), gives the joint distribution

$$\begin{bmatrix} \boldsymbol{m} \\ \boldsymbol{d}_{obs} \end{bmatrix} \sim \mathcal{N}_{n_m+n_d}\left(\begin{bmatrix} \boldsymbol{\mu}_m \\ \boldsymbol{\mu}_{d_{obs}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_m & \boldsymbol{\Sigma}_{m,d_{obs}} \\ \boldsymbol{\Sigma}_{d_{obs},m} & \boldsymbol{\Sigma}_{d_{obs}} \end{bmatrix} \right), \tag{2.27}$$

where $\boldsymbol{\Sigma}_{d_{obs},m}$ is the cross-correlation between $\boldsymbol{d}_{obs}$ and $\boldsymbol{m}$,

$$\boldsymbol{\Sigma}_{d_{obs},m} = \mathrm{Cov}\{\boldsymbol{d}_{obs}, \boldsymbol{m}\} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{\Sigma}'_m, \tag{2.28}$$

and $\boldsymbol{\Sigma}_{m,d_{obs}}$ is the transpose of $\boldsymbol{\Sigma}_{d_{obs},m}$.

The posterior distribution for $\boldsymbol{m}$ given $\boldsymbol{d}_{obs}$ is Gaussian

$$\boldsymbol{m}|\boldsymbol{d}_{obs} \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}_{m|d_{obs}}, \boldsymbol{\Sigma}_{m|d_{obs}}), \tag{2.29}$$

where the posterior expectation and covariance are

$$\boldsymbol{\mu}_{m|d_{obs}} = \boldsymbol{\mu}_m + (\boldsymbol{S}\boldsymbol{A}\boldsymbol{\Sigma}'_m)^T \boldsymbol{\Sigma}_{d_{obs}}^{-1}(\boldsymbol{d}_{obs} - \boldsymbol{\mu}_{d_{obs}}), \tag{2.30}$$

$$\boldsymbol{\Sigma}_{m|d_{obs}} = \boldsymbol{\Sigma}_m - (\boldsymbol{S}\boldsymbol{A}\boldsymbol{\Sigma}'_m)^T \boldsymbol{\Sigma}_{d_{obs}}^{-1} \boldsymbol{S}\boldsymbol{A}\boldsymbol{\Sigma}'_m, \tag{2.31}$$

using the general formula for a conditional Gaussian distribution, see Appendix 2.A. The posterior distribution, which is found on an explicit analytical form, contains the complete solution of the inverse problem, including uncertainty. A set of possible solutions can be generated by stochastic simulation from the posterior distribution: $\boldsymbol{m}^{(1)}, \boldsymbol{m}^{(2)}, \ldots, \boldsymbol{m}^{(n)}$. Since $\boldsymbol{m}$ represents the logarithm of the elastic material parameters, the corresponding set of simulated solutions of the P-wave velocity, S-wave velocity, and density are obtained by the inverse transform $\exp[\boldsymbol{m}^{(i)}]$, $i = 1, \ldots, n$.

The maximum aposteriori (MAP) solution for $\boldsymbol{m}$ is equal to the posterior expectation, $\mathrm{MAP}\{\boldsymbol{m}\} = \boldsymbol{\mu}_{m|d_{obs}}$, since the posterior distribution is Gaussian. The MAP solution is generally smoother than a single realization $\boldsymbol{m}^{(i)}$. When the noise level in expression (2.22) increases, the posterior distribution converges to the prior distribution. In the case where the noise level is high, the MAP solution will be defined mainly from the maximum of the prior distribution, which we have assumed to be smooth. A single realization, however, will be a possible solution with the full variability defined mainly from the prior distribution. In the opposite case, when the noise approaches zero, the solution is determined mainly from the seismic data, and the influence of the prior model decreases.

The posterior distributions for the P-wave velocity, S-wave velocity, and density are log-Gaussian. The MAP solution for the elastic parameters, $\exp[\boldsymbol{m}]$, can be determined component wise by

$$\text{MAP}\{\exp[m]\} = \exp[\mu_{m|d_{obs}} - \sigma^2_{m|d_{obs}}], \tag{2.32}$$

while the posterior expectation is

$$\text{E}\{\exp[m]\} = \exp[\mu_{m|d_{obs}} + \sigma^2_{m|d_{obs}}/2], \tag{2.33}$$

where m is a component of $\boldsymbol{m}$, $\mu_{m|d_{obs}}$ is the posterior expectation for m, and $\sigma^2_{m|d_{obs}}$ is the posterior variance. Typically, $\sigma^2_{m|d_{obs}} \ll \mu_{m|d_{obs}}$, such that the log-Gaussian distribution is close to symmetric, and $\exp[\mu_{m|d_{obs}}]$ is a good approximation for both $\text{MAP}\{\exp[m]\}$ and $\text{E}\{\exp[m]\}$.

A $(1-\epsilon)$ prediction interval for a specific parameter m, for example $\ln \alpha(t)$, is given by

$$\mu_{m|d_{obs}} \pm z_{\frac{\epsilon}{2}} \sigma_{m|d_{obs}}, \tag{2.34}$$

where z_ϵ is the ϵ-quantile in the standard Gaussian distribution. A 0.95 interval for $\alpha(t)$ is $\exp[\mu_{m|d_{obs}} \pm 1.96\sigma_{m|d_{obs}}]$. Prediction intervals for Z_P, Z_S, α/β, and other combinations of α, β, and ρ can be obtained by variable transforms or by stochastic simulation from the posterior distribution.

2.3 A synthetic example

The inversion method is tested on two synthetic earth profiles, well A and B, shown in Figures 2.1 and 2.2. The two profiles are simulated from a stationary prior distribution with constant expectation

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu} = [\mu_\alpha, \mu_\beta, \mu_\rho]^T = [8.006, 7.313, 7.719]^T, \tag{2.35}$$

corresponding to the logarithm of 3000 m/s P-wave velocity, 1500 m/s S-wave velocity, and 2250 kg/m^3 density, and a stationary covariance function with $\sigma^2_\alpha(t) = 0.0074$, $\sigma^2_\beta(t) = 0.0074$, and $\sigma^2_\rho(t) = 0.0024$. A second order exponential correlation function, equation (2.11), with range $d = 5$ ms is used for both wells. The two wells have different correlations between $\ln \alpha$, $\ln \beta$ and $\ln \rho$: The logs in well A are uncorrelated, while the logs in well B have a relatively strong correlation, $\nu_{\alpha\beta} = \nu_{\alpha\rho} = \nu_{\beta\rho} = 0.7$.

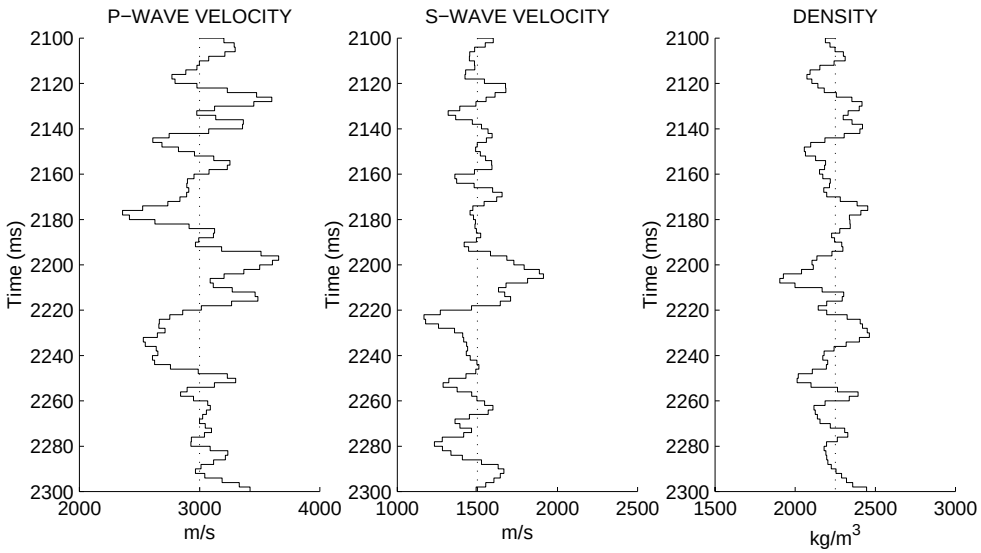


Figure 2.1: The P-wave velocity, S-wave velocity, and density in Well A (solid lines). The constant background model is dotted.

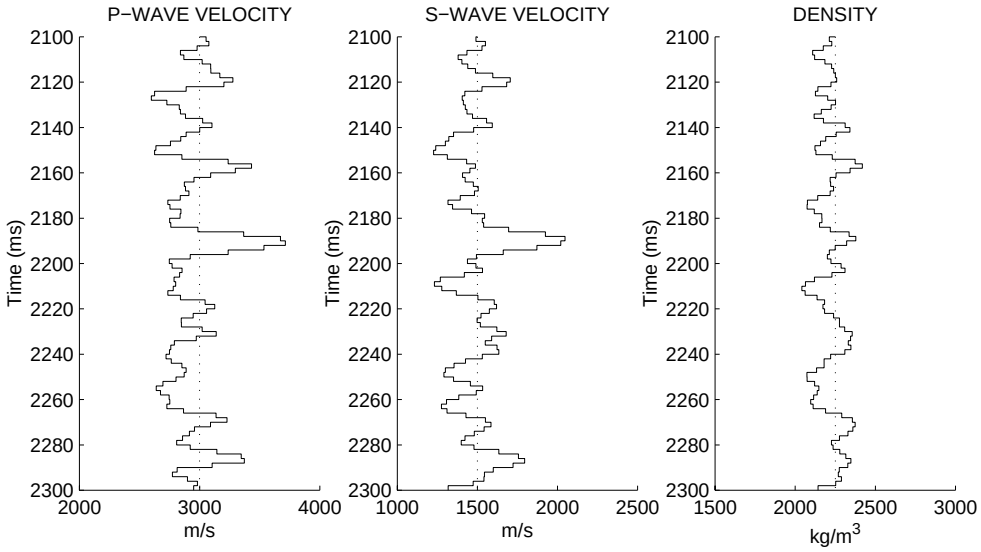


Figure 2.2: The P-wave velocity, S-wave velocity, and density in Well B (solid lines). The constant background model is dotted.

Modeling test

The weak contrast reflection coefficient approximation in equation (2.1) is adequate for moderate reflection angles. A further approximation is utilized in this inversion algorithm, where the coefficients $a_\alpha(t,\theta)$, $a_\beta(t,\theta)$ and $a_\rho(t,\theta)$, equations (2.2) to (2.4), are calculated from a constant or slowly varying prior known background model. The seismic forward modeling is performed by convolution directly in the time-angle domain. Real seismic data are recorded in the time-offset domain, and transformed to the time-angle domain prior to inversion.

Modeling tests are shown in Figures 2.3 and 2.4 for well A and B, respectively. The wavelet is a Ricker wavelet with 25 Hz center frequency and normalized amplitude. The CDP gathers to the left are calculated using weak contrast modeling directly in the time-angle domain. The coefficients a_α, a_β and a_ρ are calculated from the constant background model defined in equation (2.35). The CDP gathers in the middle of the figures are produced by ray-tracing in the time-offset domain using the full Zoeppritz equation, and the gathers are then transformed to the time-angle domain. The gathers to the right show the differences between these two modeling methods. The relative squared data errors are only 0.0019 and 0.0027 for the two examples, which justifies the use of the weak contrast reflectivity function and convolution directly in the time-angle domain. Compared to typical noise levels in real seismic data, the modeling errors in these examples are negligible.

Inversion test

The AVO inversion is tested on synthetic seismic data sets with different noise levels. A mixture of white and colored noise of the form

$$\boldsymbol{e} = \boldsymbol{e}_1 + \boldsymbol{S}\boldsymbol{e}_2 \tag{2.36}$$

was added to the synthetic data. The first term represents the white Gaussian noise with variance σ_1^2, $\boldsymbol{e}_1 \sim \mathcal{N}_{n_d}(\boldsymbol{0}, \sigma_1^2\boldsymbol{I})$. The error term $\boldsymbol{S}\boldsymbol{e}_2$ represents source generated noise, for example remaining multiples, and are assumed to be correlated between different angle traces by an exponential correlation function, $\nu_\theta = \exp[-|\theta_i - \theta_j|/d_\theta]$, where $d_\theta = 20^\circ$ is the correlation range. For each angle θ_i, the elements in the error term $\boldsymbol{e}_2$ corresponding to θ_i are white Gaussian with variance σ_2^2. Noise gathers with four different levels were simulated

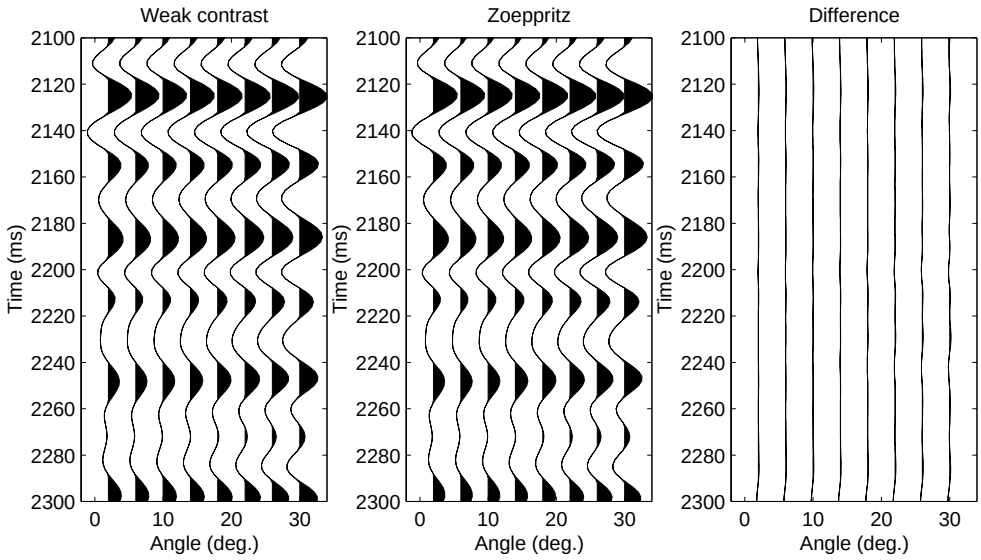


Figure 2.3: Weak contrast modeling in well A with constant background (left), raytracing with full Zoeppritz equation (middle) and residual (right).

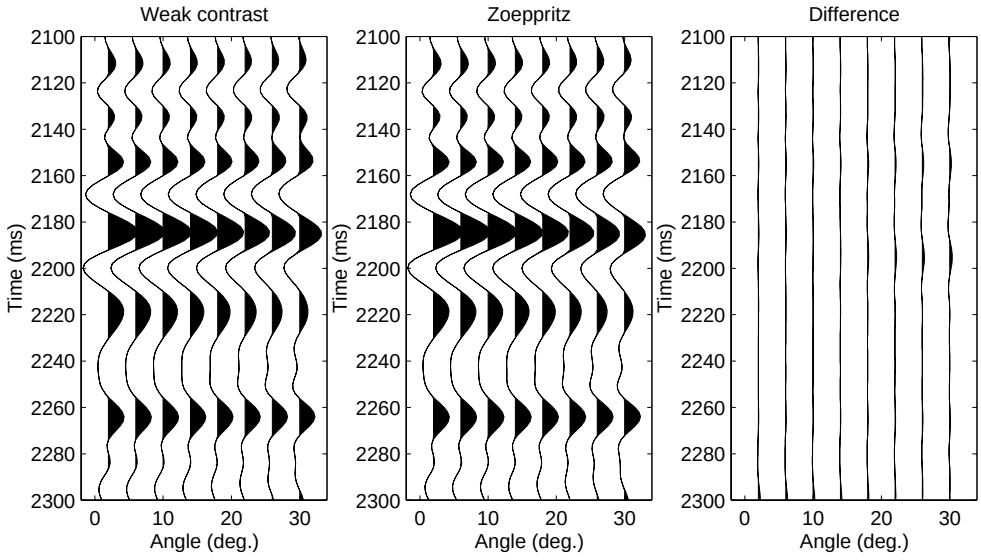


Figure 2.4: Weak contrast modelling in well B with constant background (left), raytracing with full Zoeppritz equation (middle) and residual (right).

with variances $\sigma_1^2 = \sigma_2^2 = 0.00005^2$, 0.008^2, 0.015^2, and 0.03^2, corresponding to signal to noise ratios (S/N) 10^5, 15, 5, and 1, respectivley. The CDP gathers for model A with noise are shown in Figure 2.5.

The inversion results of dataset A with S/N ratio 10^5 are displayed in Figure 2.6, showing the MAP solutions with 0.95 prediction intervals for P- and S-wave velocities, density, acoustic impedance, shear impedance, and P- to S-wave velocity ratio. The solutions for the P- and S-wave velocities, and the density are analytically obtainable, while stochastic simulation from the posterior distribution is used for the acoustic impedance, shear impedance, and P- to S-wave velocity ratio. The prior model applied is optimal in the sense that it is the same statistical model which was used to make the synthetic well logs. The 0.95 intervals of the prior model are shown in Figure 2.6 with dotted lines. Also the correct noise covariance and the correct wavelet have been used. With this low noise level, all inverted parameters are retrieved almost perfectly with low uncertainty.

The inversion results of dataset A with S/N ratio 5 are displayed in Figure 2.7. The confidence regions are much wider than in Figure 2.6, and the MAP solutions are smoother. Realizations from the posterior distribution for the data set with S/N ratio 5 are shown in Figure 2.8. These realizations have a higher, but more realistic, variability than the smooth MAP solution. Each of the realizations is a possible solution with this noise level. The large variations among these realizations explain the wide confidence regions.

A comparison of the 0.95 confidence regions for the prior and the posterior model makes a picture of the information content in the seismic prestack data. For model A, the seismic data provide mostly information about the acoustic impedance, the P- to S-wave velocity ratio, and the shear impedance for the case with S/N ratio 10^5, where the width of the confidence regions is reduced by 75% to 79%, see Table 2.1. In the case with an S/N ratio of 1, the seismic data provide mostly information about the acoustic impedance, the P-wave velocity and the P- to S-wave velocity ratio, but the intervals are only reduced by about 20%. In this case with strong noise, the seismic data provide practically no information about the S-wave velocity, the density or the shear impedance (3-4% reduction of the interval width).

In well A, the P-wave velocity, S-wave velocity, and density are uncorrelated, but this is unrealistic. Well B is simulated with a strong correlation (0.7) between these three logs. The results in Table 2.2 shows that the confidence

intervals for most parameters have been reduced more than in case A.

S/N-ratio	α	β	ρ	Z_P	Z_S	α/β
10^5	71	66	58	79	75	78
15	37	20	12	46	26	39
5	33	11	8	40	11	30
1	21	3	4	25	3	17

Table 2.1: Percentwise decrease of the width of the 0.95 interval from the prior model to the posterior model for model A.

S/N-ratio	α	β	ρ	Z_P	Z_S	α/β
10^5	77	73	69	81	78	75
15	45	29	33	49	36	22
5	39	21	28	41	26	16
1	25	12	20	27	17	7

Table 2.2: Percentwise decrease of the width of the 0.95 interval from the prior model to the posterior model for model B.

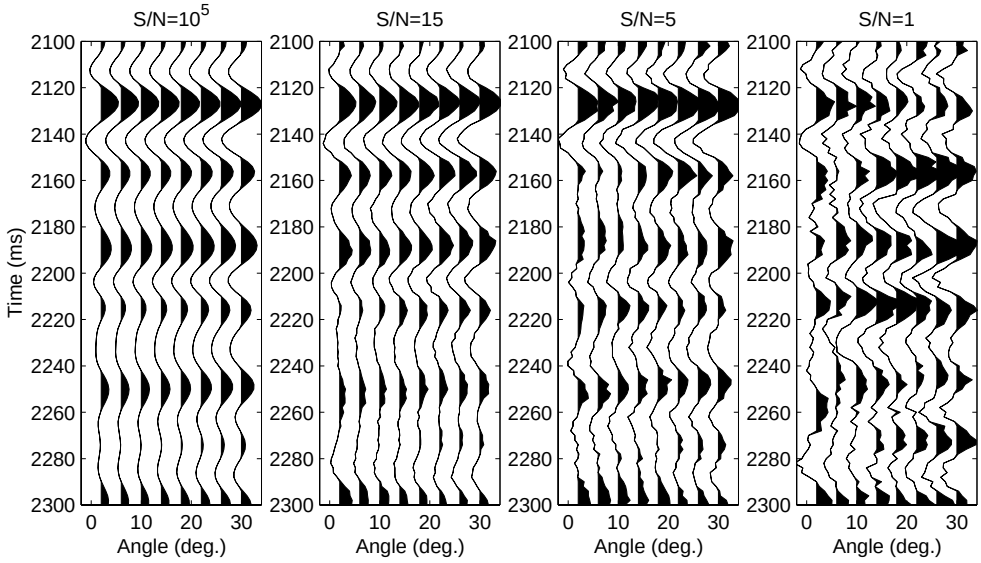


Figure 2.5: Synthetic CDP gathers for model A with different noise levels.

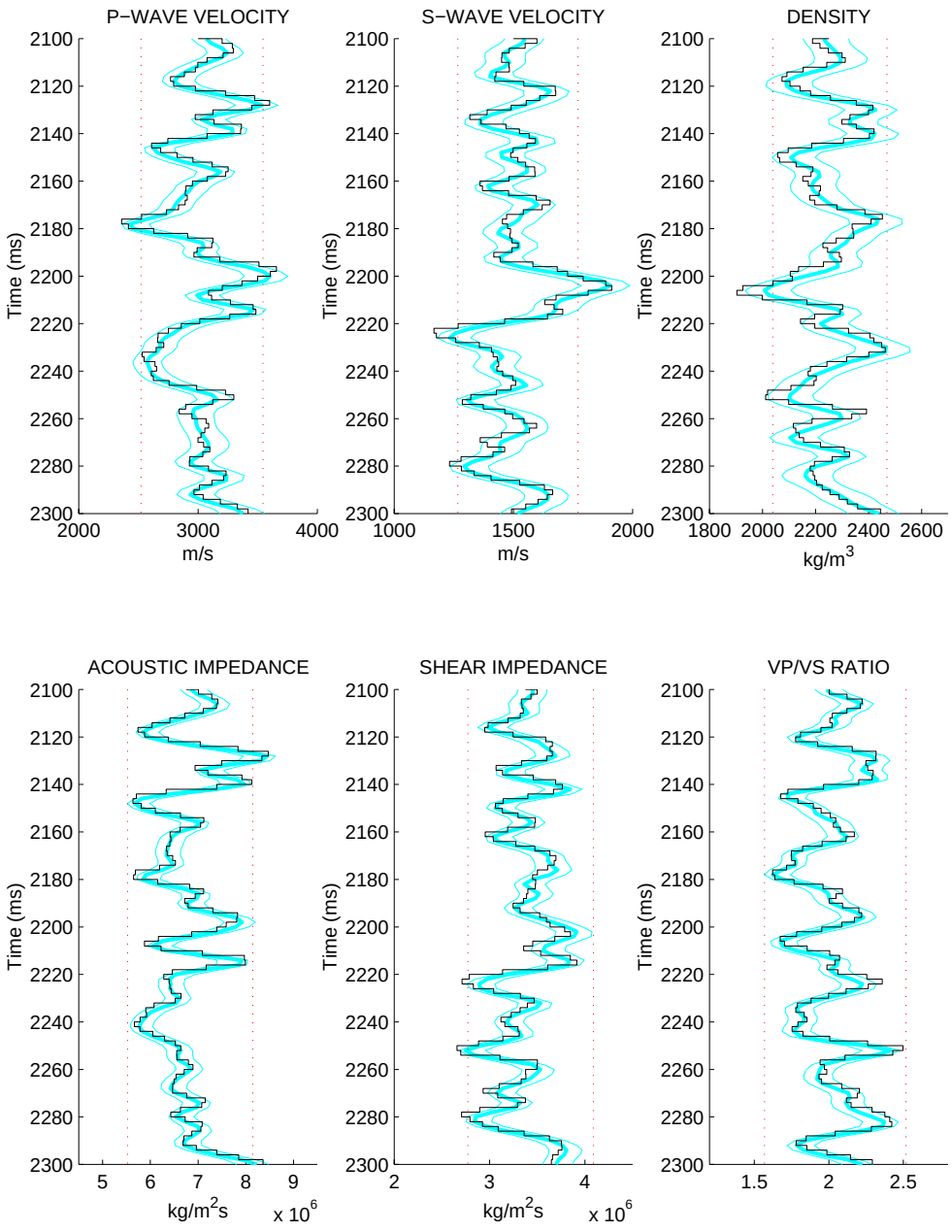


Figure 2.6: The MAP solution (thick blue line) of model A with S/N ratio 10^5 with 0.95 prediction interval (thin blue lines), the true earth profile (black line), and 0.95 prior model interval (red dotted lines).

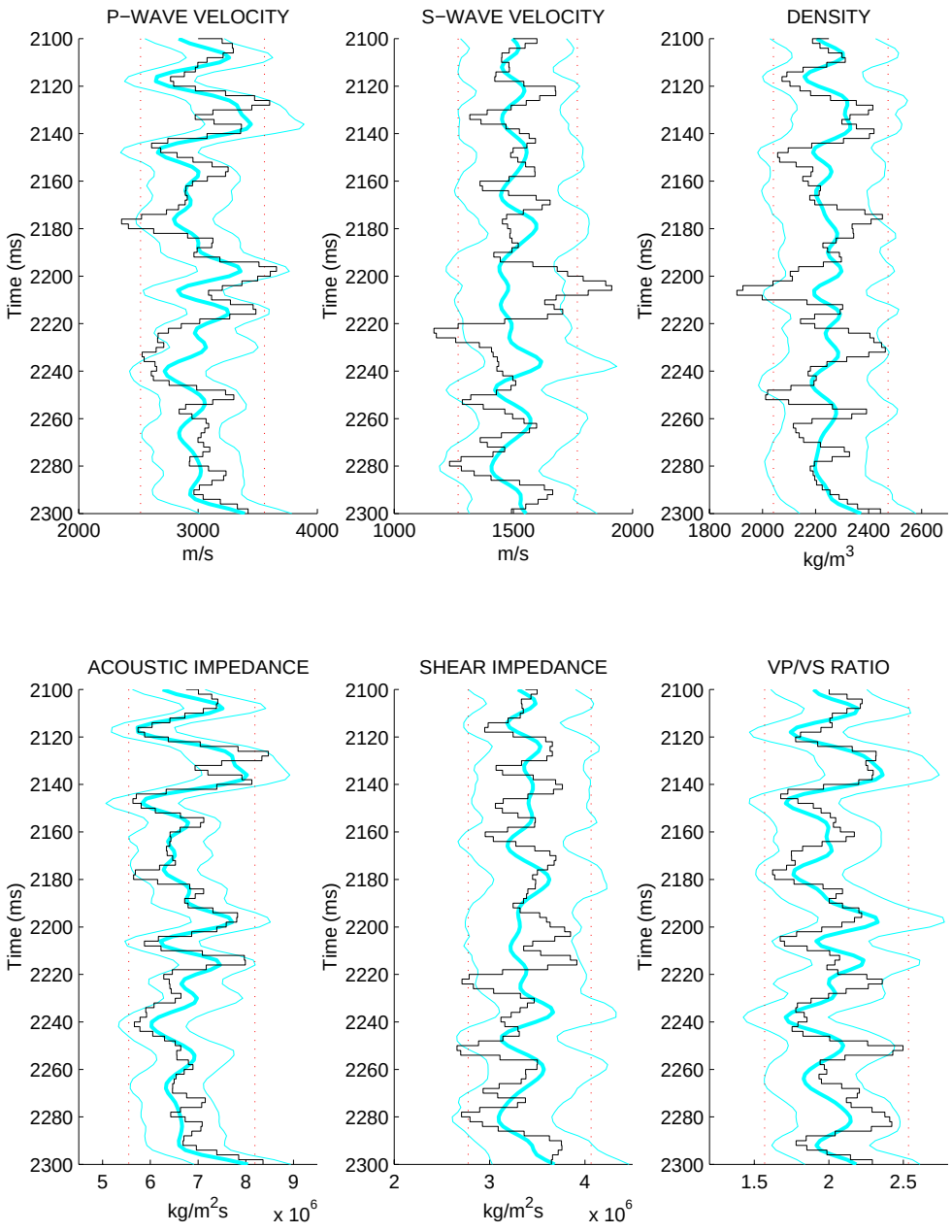


Figure 2.7: The MAP solution (thick blue line) of model A with S/N ratio 5 with 0.95 prediction interval (thin blue lines), the true earth profile (black line), and 0.95 prior model interval (red dotted lines).

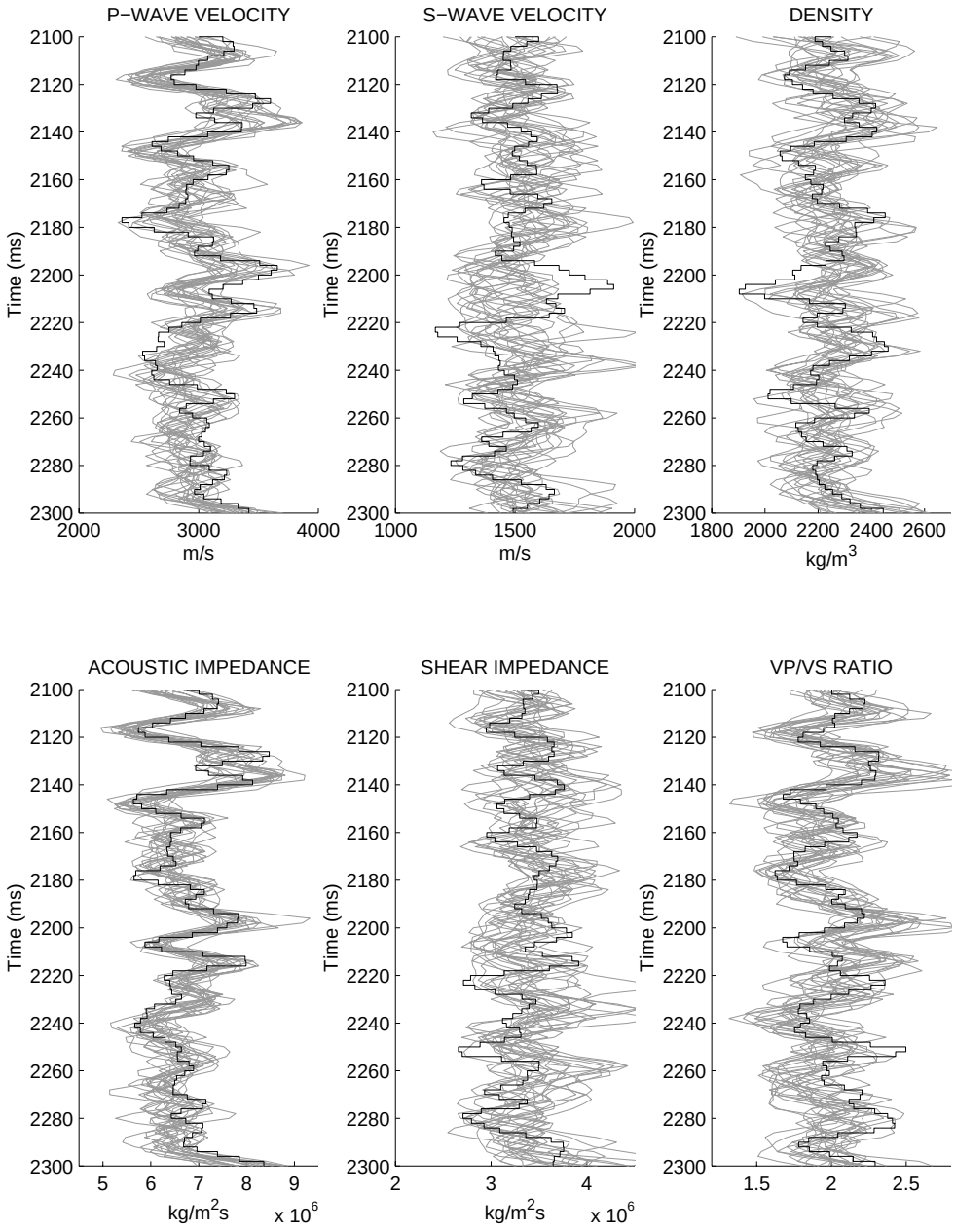


Figure 2.8: Realizations from the posterior distribution (gray lines) for model B with S/N ratio 5. The true earth profile is shown in black.

2.4 Inversion of Sleipner data

The Sleipner Øst Field is located on the eastern side of the South Viking Graben in the Norwegian Block 15/9. The gas/condensate is trapped in a submarine-fan sandstone complex of the early Tertiary Ty Formation. The depth of the reservoir sand is in the range of 2270-2500 m sub-sea with the gas-water contact at 2415 m sub-sea. The reservoir underlies a thick shale package. The Ty Formation is composed of mainly clean sandstone with some thin shale and siltstone layers. The base reservoir is a sand/chalk interface (Top Shetland Formation).

A rectangular portion of the seismic survey, defined from inlines 1411 to 1751, and from crosslines 1225 to 1400, is used in this inversion. The inversion area covers 9.3 km^2, representing 12% of the total survey. The seismic data were processed by a contractor applying the processing sequence shown in Table 2.3. The processing sequence was defined such that the final prestack amplitudes should image the reflection strength of the subsurface interfaces as correctly as possible. The stack section of inline 1627 is shown in Figure 2.9, where the strong continuous reflector appearing between 2350 and 2450 ms is the Top Shetland horizon (base reservoir).

A well is located at inline 1627 and crossline 1291. The well logs are displayed as function of depth in Figure 2.10, where the impedance logs and the P- to S-wave velocity ratio log are calculated from the P-wave velocity, S-wave velocity, and density logs.

The top reservoir is characterized by an increase in P-wave velocity (20%), a large increase in S-wave velocity (60%), and a small decrease in density (-8%). The acoustic impedance has only a minor increase from the shale to the gas sand (12%). The increase in the S-wave velocity may therefore be a better indicator for the top reservoir.

The base reservoir is well delineated on the seismic stack by the strong Top Shetland reflector. In the well, the P-wave velocity increases from about 3000 m/s in the sand reservoir to about 5000 m/s in the Top Shetland chalk formation, the S-wave velocity increases from about 1400 to 2600 m/s, and the density increases from 2200 to 2550 $\mathrm{kg/m}^3$, which gives a reflection coefficient of about 0.3. As the inversion algorithm is based on a weak contrast assumption, the inversion window is defined as a 250 ms window ending 32 ms above the Top Shetland (a half wavelet). At the well, the inversion window is from

1. SEGD read and data editing
2. Navigation merge
3. Gun, cable and filter delay correction (-38.6 ms)
4. Deterministic zero phasing
5. Lowcut filter 6Hz
6. Gain correction t^2
7. Tidal correction
8. Swell noise and interference attenuation
9. Forward $\tau - p$ transformation
10. Predicted deconvolution 64/300 ms gap/operator
11. Inverse $\tau - p$ transformation
12. Old velocity input
13. NMO
14. k-filter spatial resampling
15. CMPs to 75 m trace spacing
16. Sort to 39 offsetvolumes, 50% xline binning
17. $f - x$ trace interpolation
18. Static correction
19. DMO, reduction to 19 offsets
20. Prestack time-migration with minimum velocity
21. Inverse NMO with old velocities
22. New velocity picking
23. NMO with new velocities
24. Transform to time-angle by ray-tracing

Table 2.3: Processing sequence

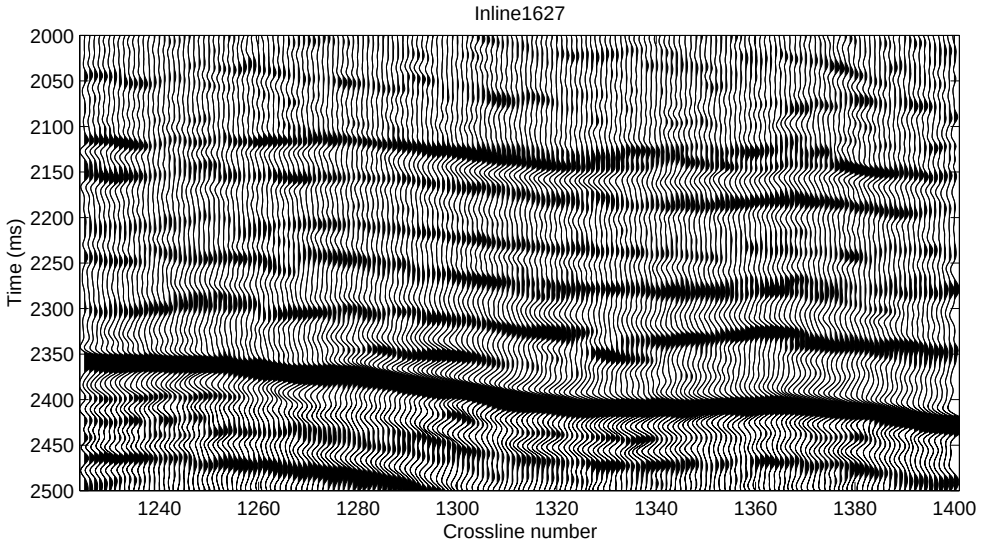


Figure 2.9: Stack section of inline 1627. The data are stacked after the processing listed in Table 2.3. The strong continuous reflector appearing between 2350 and 2450 ms is the Top Shetland horizon (base reservoir)

2100 ms to 2350 ms.

Prior model estimation

The specification of the prior model is a controversal part of the Bayesian analysis. The extensive use of mathematically convenient prior distributions and the precision level of the prior distributions are often matters of discussions. Often, the available prior information is not sufficient to define a unique parametric prior distribution. This problem can partly be handled by non-informative prior models, or by including the uncertainty in the prior model by a hierarchical prior model. When some prior observations are available, a pragmatic approach is to select a parametric distribution, and then estimate the neccessary parameters from the available prior observations. The most obvious pitfall with this approach is that a prior distribution with too low variance may be specified.

In the current study, the most important source of prior information is the available well logs. Ideally, the prior model should be estimated from several

wells in the area in order to include lateral heterogeneity. Unfortunately, only one well is available in the portion of the survey selected for this inversion test.

The parameter vector $\boldsymbol{m}$ is a priori assumed to be Gaussian. This assumption can be graphically evaluated in the well position by a Gaussian probability plot. If the parameters at this well are exactly Gaussian, the plot will be linear. Gaussian probability plots of $\ln\alpha$, $\ln\beta$, and $\ln\rho$ calculated from the well logs in the time interval 2100-2350 ms are shown in Figure 2.11. Despite some curvature in these plots, we consider the Gaussian a priori assumption to be acceptable. The computational cost of rejecting the Gaussian assumption is dramatic, since we loose the analytical form of the solution.

From the well logs, estimates of the elements in $\boldsymbol{\Sigma}_0$ are obtained by standard estimators. The estimated variances for $\ln\alpha$, $\ln\beta$ and $\ln\rho$ are $\widehat{\sigma_\alpha^2} = 0.0037$, $\widehat{\sigma_\beta^2} = 0.0116$, and $\widehat{\sigma_\rho^2} = 0.0004$, and the estimated correlation coefficients are $\widehat{\nu_{\alpha\beta}} = 0.65$, $\widehat{\nu_{\alpha\rho}} = 0.11$, $\widehat{\nu_{\beta\rho}} = -0.07$.

The temporal correlation function are estimated for certain time lags from the well logs, see Figure 2.12. The temporal correlation function is modeled by an analytic correlation function,

$$\begin{aligned}\nu(\tau; d_1, d_2) &= \frac{1}{2}\exp\left[-\left(\frac{\tau}{d_1}\right)^2\right] \\ &\quad + \frac{1}{2}\left(1 - \frac{2\tau^2}{d_2^2}\right)\exp\left[-\left(\frac{\tau}{d_2}\right)^2\right], \end{aligned} \tag{2.37}$$

defined by the sum of an exponential second order correlation function with range $d_1 = 1.8$ ms, and a normalized second derivative of an exponential second order correlation function with range $d_2 = 9$ ms. The fit to the estimated correlation function is considered to be good, see Figure 2.12.

Estimation of wavelets and noise covariance

The last step before inversion is to estimate wavelets and the noise covariance matrix from the well-logs and the CDP gather at the well position. A Bayesian wavelet estimation method has been used, where seismic noise, possible mistie between the time axis of the well logs and the seismic traveltimes, and errors in the log measurements are included, see Buland and Omre (2002b). The possible mistie between the time axis of the seismic data and the well-logs is

handled by allowing for shift, stretch and squeeze of the time axis of the well-logs. The noise term is assumed to be a mixture of white and colored noise as in expression (2.36). The solution of the estimation problem is obtained by Markov chain Monte Carlo (MCMC) simulation, using 2000 iterations. The estimated wavelets are shown in Figure 2.13. The two variance terms of $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are estimated to $\widehat{\sigma_1^2} = 0.0114$ and $\widehat{\sigma_2^2} = 0.0001$, and the estimated correlation range for $\boldsymbol{e}_2$ is $\widehat{d_\theta} = 10°$.

Inversion results

The inversion results in the well position are displayed in Figure 2.14, showing the MAP solutions and 0.95 prediction intervals. With the estimated noise level in this data set, the prediction intervals are only marginally reduced compared to the prior model. The acoustic impedance and the P-wave velocity are the best determined parameters, but the prediction intervals are only reduced by 31% and 27%, respectively. The prediction intervals for the P- to S-wave velocity ratio, the shear impedance and the S-wave velocity are reduced by about 15%, while the interval for the density is reduced by only 5%. Top reservoir is located at 2300 ms, and is characterized by an increase in both P- and S-wave velocity. This increase in the two velocities is seen in the inversion results, but the thin low velocity layer at 2317 ms is not retrieved by the inversion.

The MAP solutions of inline 1627 are shown in Figure 2.15. The bottom of the inversion window is shown by a thick black line 32 ms above the Top Shetland horizon. The well logs are plotted for comparison, and show a good agreement with the inversion results. Time slices (2320 ms) of the P- and S-wave velocities are shown in Figure 2.16. The boundary of the top reservoir, interpreted from stacked data, is shown by a black line.

The inversion algorithm is fast. The inversion of the 3-D test cube was finished in less than 3 minutes on a single 400 MHz Mips R12000 CPU. The algorithm is suitable for parallelization. On a 8 CPU parallel machine, the whole survey (77.5 km^2) could be inverted in approximately 3 minutes.

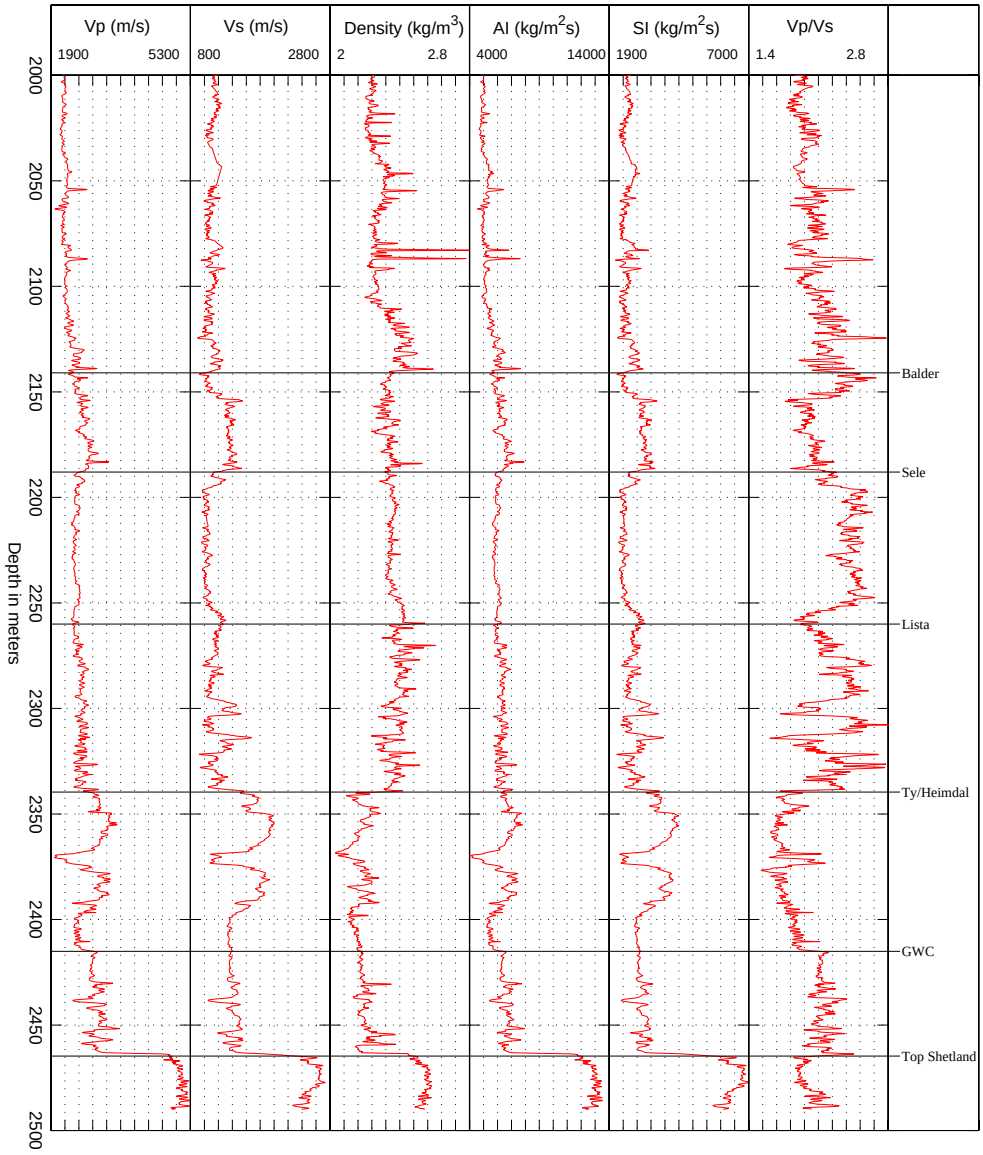


Figure 2.10: The well logs plotted in the depth interval 2000-2500 m. The Top reservoir is at 2340 m, the gas-water contact at 2415 m, and the Base reservoir at 2465 m.

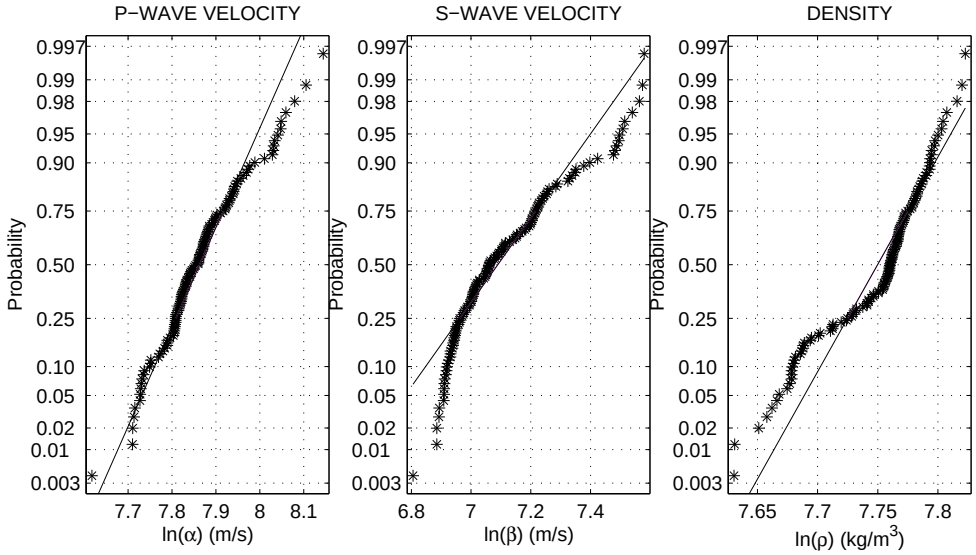


Figure 2.11: Gaussian probability plot of the logarithm of the P-wave velocity (left), S-wave velocity (middle), and density (right).

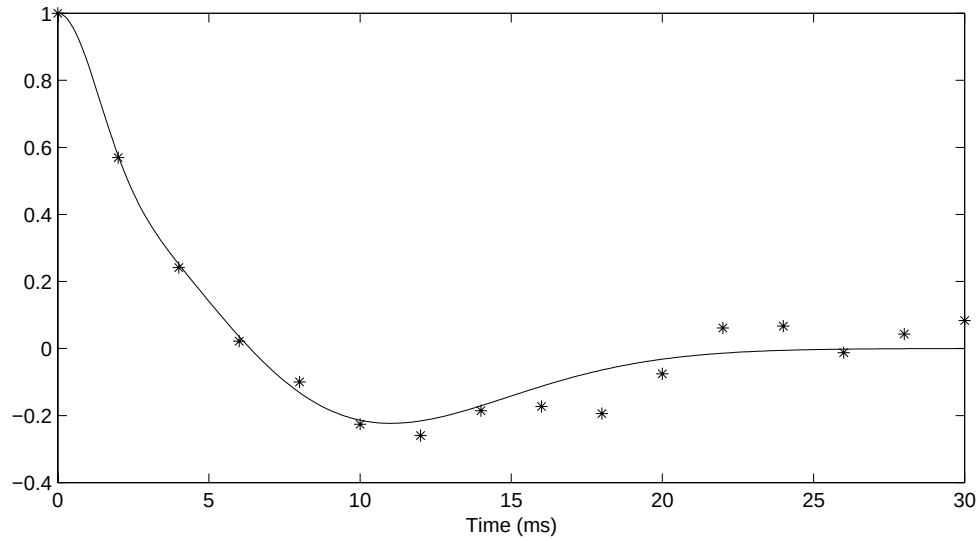


Figure 2.12: Correlation function estimated from well logs (stars), and an analytical correlation function (black line) derived from two exponential second order correlation functions.

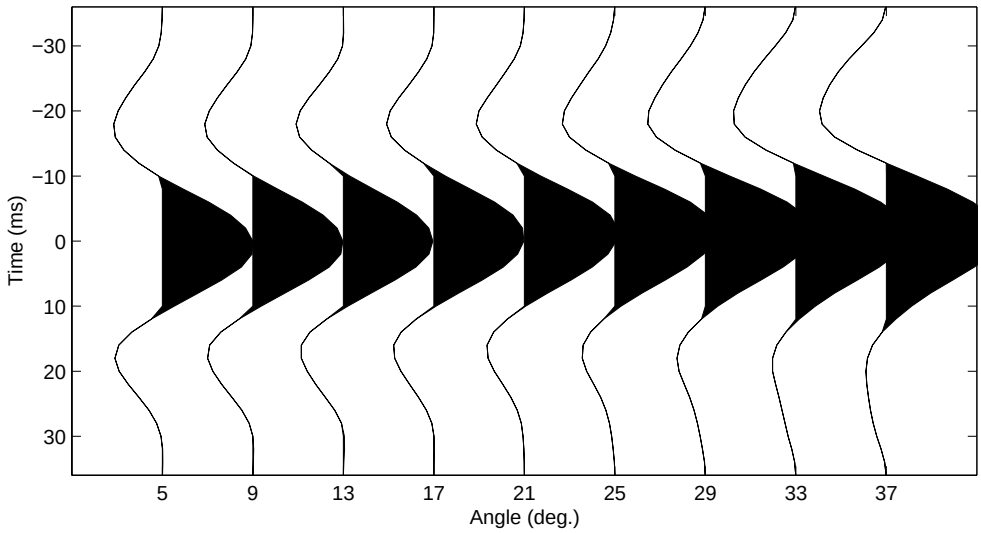


Figure 2.13: Estimated wavelets at reservoir level.

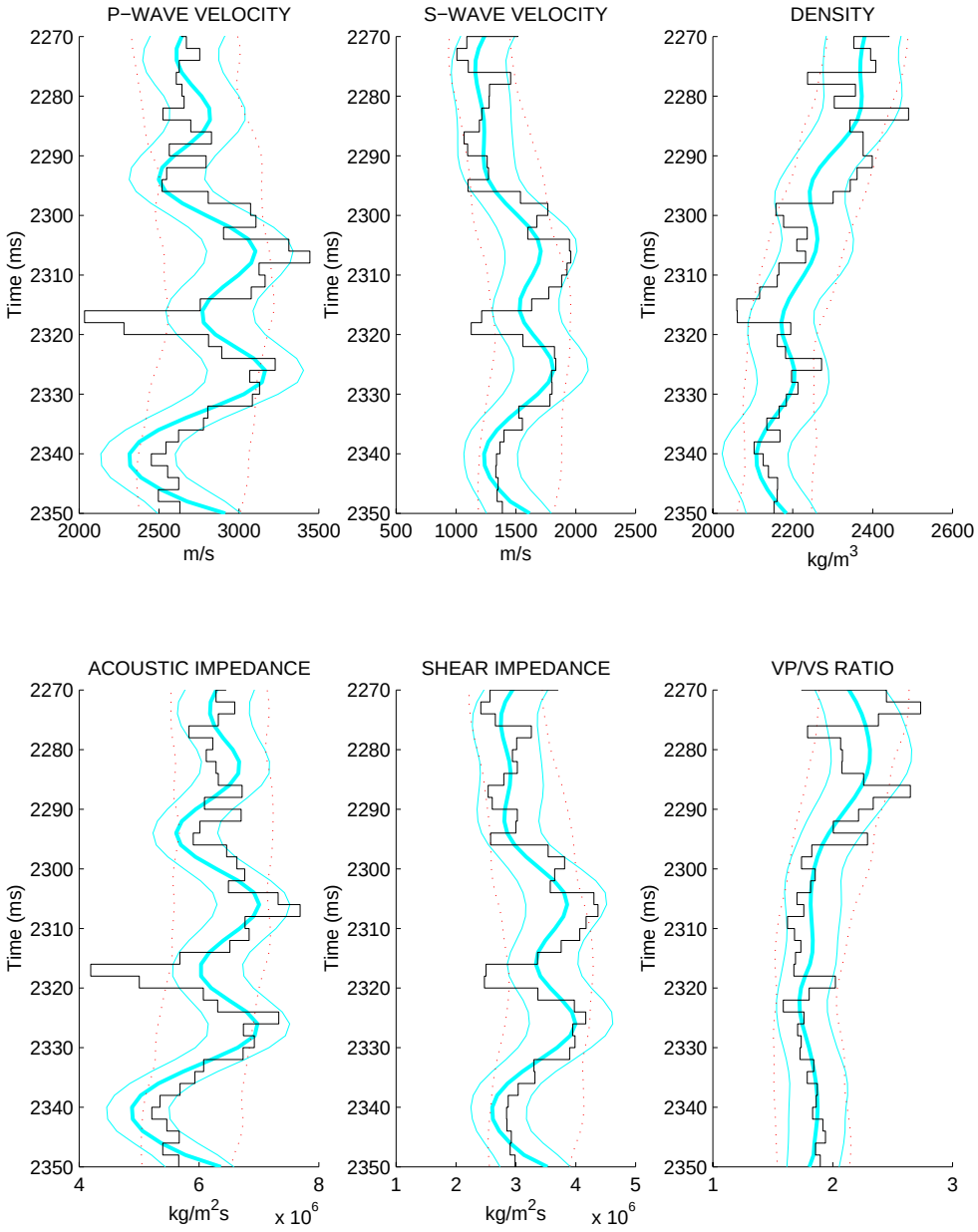


Figure 2.14: The MAP solution (thick blue line) in the well position with 0.95 prediction interval (thin blue lines), the well log (black line), and 0.95 prior model interval (red dotted lines).

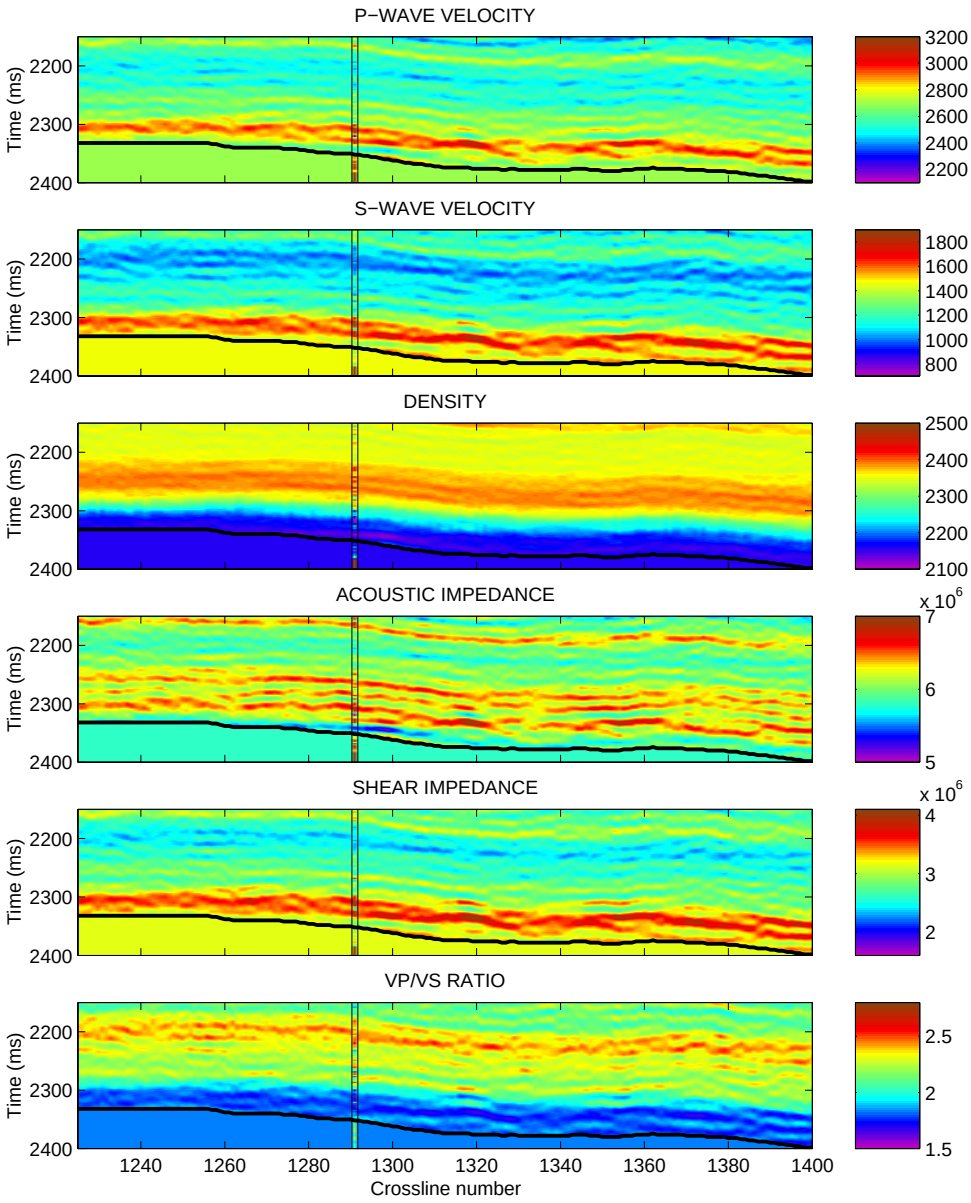


Figure 2.15: MAP solution of inline 1627.

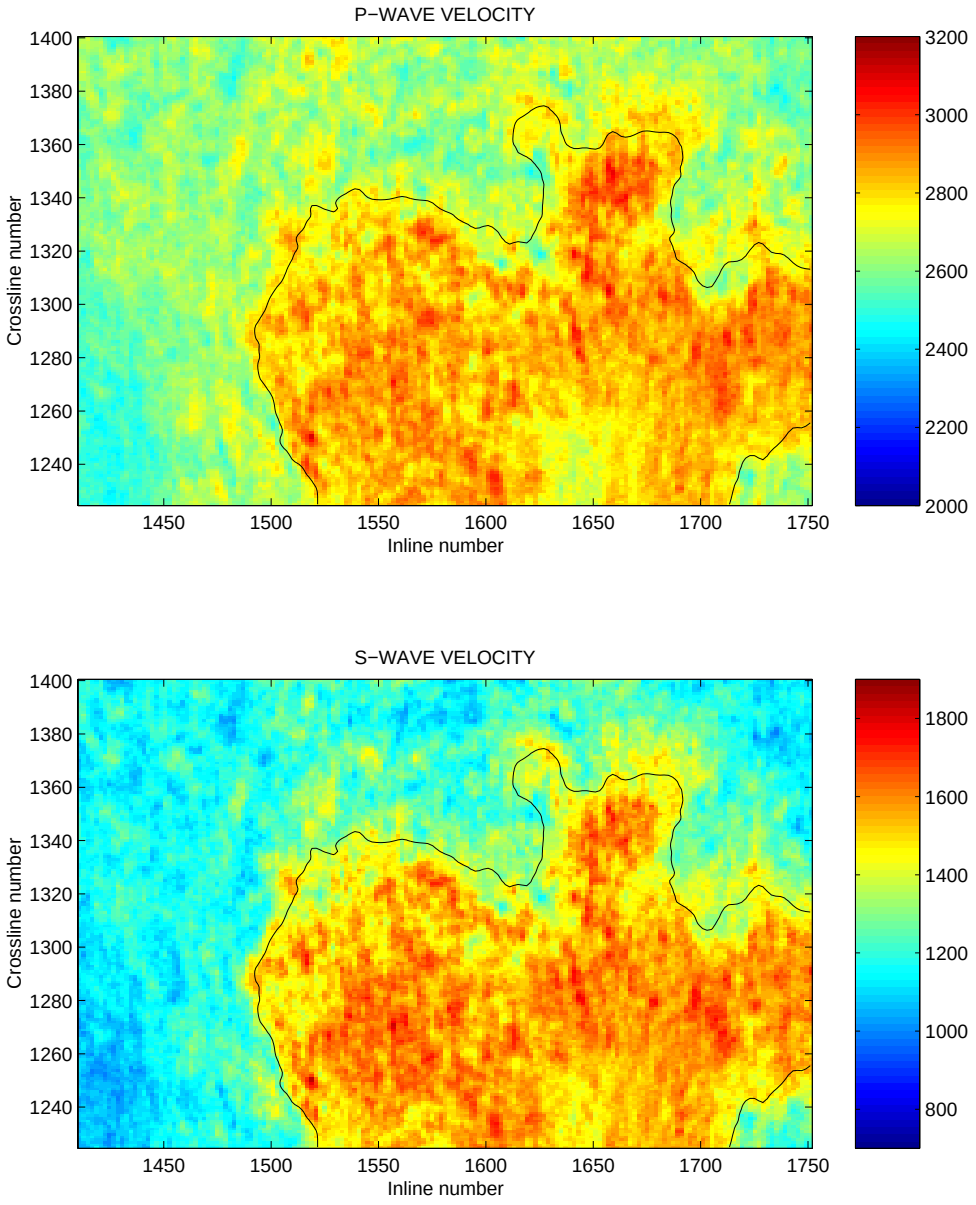


Figure 2.16: Time slice of MAP solution. The interpreted boundary of the top reservoir is shown by black line.

2.5 Discussion and Conclusions

We have developed a Bayesian AVO inversion method where the objective is to obtain posterior distributions for P-wave velocity, S-wave velocity, and density. The solution of the AVO inversion is given by a Gaussian posterior distribution. The explicit analytical form of the posterior distribution provides a fast inversion method. The uncertainty in the inverted parameters is an integral part of the solution.

Inversion tests on synthetic data with S/N=10^5 show high agreement between the estimated and the correct model, but weak noise (S/N=15) have a dramatic effect on the uncertainty of the predicted parameters. The inversion of the field data from Sleipner show good agreement with well logs, but the uncertainty is high. Acoustic impedance is the best determined parameter, while the AVO inversion provides practically no information about the density. Generally, the resolution of the different parameters depends on the prior model and the noise covariance.

2.6 Acknowledgments

We thank Statoil and the Sleipner licence (Statoil, Exxon/Mobil, Norsk Hydro and TotalFinaElf) for permission to publish this paper. We will also thank the reviewers for a thorough review and useful comments.

2.A Gaussian distribution

A random field $r(t)$ is a spatially correlated function where the function value is random for any t. A Gaussian random field $r(t)$ is a random field where all multi-dimensional distribution for vectors $\boldsymbol{r} = [r(t_1), \ldots, r(t_n)]^T$ are Gaussian for any n and any configuration $\{t_1, \ldots, t_n\}$. The Gaussian random field is completely specified by the expectation function $\mu(t)$ and the covariance function $\Sigma(t,s) = \mathrm{Cov}\{r(t), r(s)\}$. The expectation can be arbitrary, but the covariance must be symmetric, $\Sigma(t,s) = \Sigma(s,t)$, and positive definite. The Gaussian probability density function for $\boldsymbol{r}$ is

$$p(\boldsymbol{r}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{r}-\boldsymbol{\mu})\right], \qquad (2.A.1)$$

where n is the dimension of the vector $\boldsymbol{r}$, and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the expectation vector and the covariance matrix defined from $\mu(t)$ and $\Sigma(t,s)$, respectively. A compact notation is $\boldsymbol{r} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. For further reading on multivariate statistical analysis, see e.g., Anderson (1984), where also the two basic results below are presented.

Result 1 *If the vector* $\boldsymbol{r}$ *is Gaussian,* $\boldsymbol{r} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, and* $\boldsymbol{M}$ *is an* $m \times n$ *matrix, then*

$$\boldsymbol{M}\boldsymbol{r} \sim \mathcal{N}_m(\boldsymbol{M}\boldsymbol{\mu}, \boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}^T). \tag{2.A.2}$$

Result 2 *Consider two multivariate Gaussian variables* $\boldsymbol{r}_1 \sim \mathcal{N}_{n_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ *and* $\boldsymbol{r}_2 \sim \mathcal{N}_{n_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ *with joint distribution*

$$\begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \end{bmatrix} \sim \mathcal{N}_{n_1+n_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right). \tag{2.A.3}$$

Then the conditional distribution of $\boldsymbol{r}_1$ *given* $\boldsymbol{r}_2$ *is Gaussian*

$$\boldsymbol{r}_1 | \boldsymbol{r}_2 \sim \mathcal{N}_{n_1}(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}), \tag{2.A.4}$$

where the conditional expectation and covariance are

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{r}_2 - \boldsymbol{\mu}_2), \tag{2.A.5}$$

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{2.A.6}$$

2.B The forward modeling

A discrete version of the continuous reflectivity function

$$c_{PP}(t,\theta) = a_\alpha(t,\theta)\frac{\partial}{\partial t}\ln\alpha(t) + a_\beta(t,\theta)\frac{\partial}{\partial t}\ln\beta(t) + a_\rho(t,\theta)\frac{\partial}{\partial t}\ln\rho(t), \tag{2.B.1}$$

in a time interval and for a set of reflection angles can be written as

$$\boldsymbol{c} = \boldsymbol{A}\boldsymbol{m}'. \tag{2.B.2}$$

The sparse matrix $\boldsymbol{A}$ is defined by

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_\alpha(\theta_1) & \boldsymbol{A}_\beta(\theta_1) & \boldsymbol{A}_\rho(\theta_1) \\ \vdots & \vdots & \vdots \\ \boldsymbol{A}_\alpha(\theta_{n_\theta}) & \boldsymbol{A}_\beta(\theta_{n_\theta}) & \boldsymbol{A}_\rho(\theta_{n_\theta}) \end{bmatrix}, \tag{2.B.3}$$

where $\boldsymbol{A}_\alpha(\theta_i)$, $\boldsymbol{A}_\beta(\theta_i)$, and $\boldsymbol{A}_\rho(\theta_i)$ are $\frac{n_m}{3} \times \frac{n_m}{3}$ diagonal matrices containing discrete time samples of $a_\alpha(t,\theta_i)$, $a_\beta(t,\theta_i)$, and $a_\rho(t,\theta_i)$, respectively, n_θ is the number of reflection angles, and n_m is the dimension of $\boldsymbol{m}$ and $\boldsymbol{m}'$.

The convolution of the reflection coefficients $\boldsymbol{c}$ with the wavelets can be formulated as a matrix-vector multiplication

$$\boldsymbol{d}_{obs} = \boldsymbol{S}\boldsymbol{c} + \boldsymbol{e}, \tag{2.B.4}$$

where $\boldsymbol{S}$ is a block-diagonal matrix containing one wavelet for each reflection angle. In an expanded form, expression (2.B.4) can be written

$$\begin{bmatrix} \boldsymbol{d}_{obs}(\theta_1) \\ \vdots \\ \boldsymbol{d}_{obs}(\theta_{n_\theta}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{S}(\theta_1) & & \\ & \ddots & \\ & & \boldsymbol{S}(\theta_{n_\theta}) \end{bmatrix} \begin{bmatrix} \boldsymbol{c}(\theta_1) \\ \vdots \\ \boldsymbol{c}(\theta_{n_\theta}) \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}(\theta_1) \\ \vdots \\ \boldsymbol{e}(\theta_{n_\theta}) \end{bmatrix}, \tag{2.B.5}$$

where $\boldsymbol{d}_{obs}(\theta_i)$ is the seismic time trace for angle θ_i, and $\boldsymbol{c}(\theta_i)$ and $\boldsymbol{e}(\theta_i)$ are the corresponding reflection coefficients and error samples, respectively. The block matrix $\boldsymbol{S}(\theta_i)$ represents the wavelet for angle θ_i

$$\boldsymbol{S}(\theta_i) = \begin{bmatrix} s_1(\theta_i) & & & & & & \\ s_2(\theta_i) & s_1(\theta_i) & & & & & \\ \vdots & & \ddots & & & & \\ s_{n_s}(\theta_i) & \cdots & \cdots & s_1(\theta_i) & & & \\ & \ddots & & & \ddots & & \\ & & s_{n_s}(\theta_i) & \cdots & \cdots & s_1(\theta_i) & \\ & & & \ddots & & \vdots & \\ & & & & \ddots & \vdots & \\ & & & & & s_{n_s}(\theta_i) & \end{bmatrix}, \tag{2.B.6}$$

where $(s_1(\theta_i), \ldots, s_{n_s}(\theta_i))$ are the samples of the wavelet for angle θ_i. In this example, the sampling of the wavelet is equal to the sampling of the seismic data. If the sampling of $\boldsymbol{c}$ and $\boldsymbol{d}_{obs}$ are different, the rows of $\boldsymbol{S}$ contains wavelets corresponding to the sampling of $\boldsymbol{c}$, and the rows are shifted relatively according to the sampling of $\boldsymbol{d}_{obs}$.

Chapter 3

Bayesian wavelet estimation from seismic and well log data

Arild Buland and Henning Omre

Paper submitted for publication. Presented at the 63rd EAGE conference, Amsterdam 2001 (Extended abstract).

Abstract

A Bayesian method for wavelet estimation from seismic and well data is developed. The method works both on stacked data, and prestack data in form of angle gathers. The seismic forward model is based on the convolutional model, where the reflectivity is calculated from the well logs. The estimated wavelets are given as probability density functions such that uncertainties of the wavelets are an integral part of the solution. Possible mistie between the seismic traveltimes and the time axis of the well logs, errors in the log measurements, and seismic noise are included in the model. The solution is obtained by Markov chain Monte Carlo simulation.

3.1 Introduction

The seismic wavelet is a required input to many seismic processing, modeling and inversion algorithms, and a misspecified wavelet may lead to erroneous

results. By quantifying the uncertainty related to detailed seismic analysis, for example by including confidence bounds on inversion results, see e.g. Buland and Omre (2002a), more robust interpretations of the reservoir characteristics can be made. Uncertainty bounds on the provided wavelet may contribute to more realistic uncertainty models.

There exist several methods and approaches for wavelet extraction from seismic data. The methods fall into two main classes depending on whether well log information is used or not. In this paper, wavelet estimation from seismic and well data is discussed. Wavelet estimation from seismic and well data is routinely done in seismic analysis, and discussed by several authors, see Danielson and Karlsson (1984); Lines and Treitel (1985); Nyman et al. (1987); Richard and Brac (1988); Poggiagliolmi and Allred (1994). In these papers different least square and spectral division methods are proposed and it is assumed that the seismic data can be explained by the convolutional model

$$d_{obs}(t) = \int c(\tau)s(t-\tau)d\tau + e_d(t), \tag{3.1}$$

where $d_{obs}(t)$ represents the seismic trace, $c(t)$ is the reflectivity derived from the well logs, $s(t)$ is the unknown wavelet, and $e_d(t)$ is an error term. The seismic data are assumed to be properly processed by a true amplitude processing sequence, which means that unwanted effects like multiples, transmission losses, absorption, geometrical spreading, focusing, changes in source strength and receiver sensitivity, etc., are removed or accounted for by the processing. It is further assumed that the reflectivity along the well trace can be calculated from the available well logs, where the well logs have been edited, calibrated and transformed from depth to seismic travel time. Moreover, to fulfill the underlying stationarity assumption in the convolutional model, the wavelet estimation should be restricted to a limited time window covering the target zone.

A possible time varying mistie between the seismic traveltimes and the time axis of the well log reflectivity is a complicating factor in the wavelet estimation. Danielson and Karlsson (1984) concluded that such mistie may have serious effects on the wavelet estimation. If uncertainty in the well position and measurement errors are included, the true reflectivity can be represented by

$$c(t) = c_{obs}(t + e_t(t)) + e_c(t), \tag{3.2}$$

where $e_t(t)$ and $e_c(t)$ are error terms. The time error function $e_t(t)$ allows for shift, stretch and squeeze of the time axis, while $e_c(t)$ represents errors in the

reflection coefficient values. These are related to measurement uncertainty, but also to a possible change in the reflectivity related to uncertainty of the well position relative to the seismic data.

The wavelet extraction is often done from stacked data using the common assumption that a stack is a good approximation for normal incidence data. The normal incidence reflectivity is determined by the acoustic impedance, $Z(t) = \alpha(t)\rho(t)$, calculated from the P-wave velocity and density logs. A more general problem is estimation of angle dependent wavelets $s(t, \theta)$, where θ is the reflection angle. These wavelets are estimated from a seismic reflection angle gather $d_{obs}(t, \theta)$ and the corresponding angle dependent well log reflectivity $c(t, \theta)$. In this case, also the error terms $e_d(t, \theta)$ and $e_c(t, \theta)$ are angle dependent. The angle dependent reflectivity can be calculated by the Zoeppritz equation or by approximations when they are valid. The S-wave velocity information is now needed in addition to the P-wave velocity and density logs. If a reliable S-wave velocity log is not available, an estimated S-wave velocity log can be used, but this will in general increase the uncertainty of the wavelet estimation.

In the following, a Bayesian wavelet estimation method based on the convolutional model is presented. Mistie between the seismic traveltimes and the time axis of the well logs, errors in the log measurements, and seismic noise are included in the model such that the uncertainty of the estimated wavelet becomes an integral part of the solution. The solution is obtained by Markov chain Monte Carlo (MCMC) simulation, see for example Gilks et al. (1996); Chen et al. (2000). Stochastic simulation by MCMC techniques has a large potential for realistic statistical modeling, and it has during the last decade gone into mainstream statistics, especially within the fields of spatial statistics, image analysis and Bayesian statistics.

3.2 Methodology

The methodology will be described in a discrete setting. A discrete version of the convolutional model in expression (3.1) can be written

$$\boldsymbol{d}_{obs} = \boldsymbol{c} \star \boldsymbol{s} + \boldsymbol{e}_d, \tag{3.3}$$

where the vector $\boldsymbol{d}_{obs}$ represents the seismic data in a limited time window, $\boldsymbol{c}$ is the corresponding reflectivity, $\boldsymbol{s}$ is the wavelet and $\boldsymbol{e}_d$ is the error term.

If the vectors $\boldsymbol{d}_{obs}$, $\boldsymbol{c}$, $\boldsymbol{s}$ and $\boldsymbol{e}_d$ represent several traces, for example different offsets or reflection angles, the same notation is used where the convolution is applied trace by trace. The discrete representation of expression (3.2) is

$$\boldsymbol{c} = \boldsymbol{c}_{obs}(\boldsymbol{e}_t) + \boldsymbol{e}_c, \tag{3.4}$$

where $\boldsymbol{c}_{obs}(\boldsymbol{e}_t)$ is a discrete representation of the well log reflectivity after modification of the time axis by the time error vector $\boldsymbol{e}_t$. In practice, $\boldsymbol{e}_t$ may be defined from a set of well markers $\{t_1, t_2, \ldots, t_m\}$, which may correspond to major changes in the well logs or interfaces between major layers in a layered subsurface model. A time shift Δt of one of the well markers then corresponds to a change in the layer thicknesses (in time) of the layers just above and below this marker.

Combining expressions (3.3) and (3.4) gives the generalized convolutional model

$$\boldsymbol{d}_{obs} = \{\boldsymbol{c}_{obs}(\boldsymbol{e}_t) + \boldsymbol{e}_c\} \star \boldsymbol{s} + \boldsymbol{e}_d. \tag{3.5}$$

The error term $\boldsymbol{e}_c \star \boldsymbol{s}$ is related to well log errors, but components of the seismic error term $\boldsymbol{e}_d$ may have a similar form, for example remaining multiples. These two error terms are not separable without additional information.

Stochastic model

The problem will now be cast in a Bayesian framework, such that the relevant statistical properties of the problem can be explored. The variables and the problem structure are displayed in Figure 3.1 as a directed acyclic graph (DAG), see for example Spiegelhalter et al. (1996); Lauritzen (1996). The nodes in the graph represent stochastic variables, and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote expectations and covariances which will be defined later. The arrows represent the causal dependency structure, where thin lines indicate probability dependencies and thick lines represent deterministic relationships. The DAG can be interpreted as a family tree where the parent nodes point to its children. Before any data are observed, the marginal distributions of the parents are independent, but when their child is known, the properties of the parents become dependent. An example is the reflectivity $\boldsymbol{c}$ and the wavelet $\boldsymbol{s}$ which are apriori independent, but become dependent when the observed seismic data are included in the model.

When probabilistic dependencies are considered, nodes connected by deterministic links merge into a single node, so the parents of the seismic data are

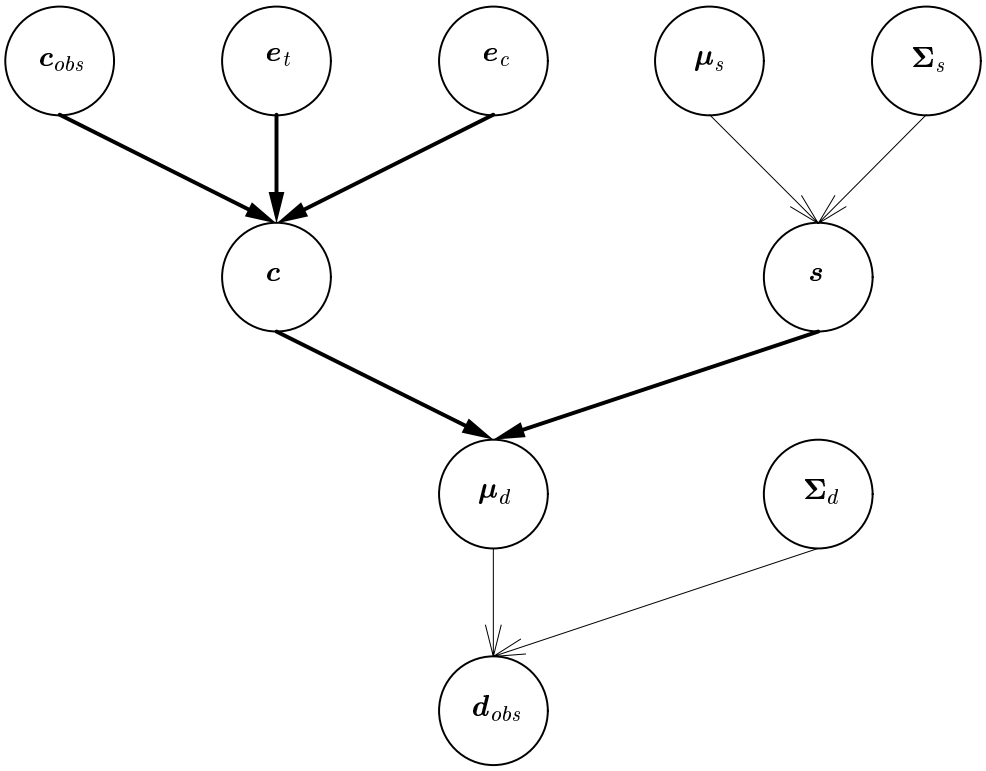


Figure 3.1: The convolutional model represented by a directed acyclic graph.

$\mathrm{Pa}(\boldsymbol{d}_{obs}) = \{\boldsymbol{c}_{obs}, \boldsymbol{e}_t, \boldsymbol{e}_c, \boldsymbol{s}, \boldsymbol{\Sigma}_d\}$, and the parents of the wavelet are $\mathrm{Pa}(\boldsymbol{s}) = \{\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s\}$, where the notation $\mathrm{Pa}(v)$ denotes the parents of a node v. The complete stochastic model can be specified by likelihood terms for the nodes with parents, and prior distributions for the nodes without parents.

Likelihood model

The seismic error term $\boldsymbol{e}_d$ in expression (3.5) is assumed to be Gaussian with zero expectation and covariance $\boldsymbol{\Sigma}_d$, so the likelihood model for the seismic data is Gaussian,

$$\boldsymbol{d}_{obs} | \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d \sim \mathcal{N}_{n_d}(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d), \tag{3.6}$$

where n_d is the dimension of $\boldsymbol{d}_{obs}$, and $\boldsymbol{\mu}_d$ is the expectation vector defined by the convolution $\boldsymbol{\mu}_d = \boldsymbol{c} \star \boldsymbol{s}$, where $\boldsymbol{c}$ is the reflectivity vector defined by expression (3.4). The multi-Gaussian distribution is defined in Appendix 3.A.

The wavelet $\boldsymbol{s}$ is modeled by a Gaussian likelihood,

$$\boldsymbol{s}|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s), \tag{3.7}$$

where n_s is the dimension of $\boldsymbol{s}$, and $\boldsymbol{\mu}_s$ and $\boldsymbol{\Sigma}_s$ are the expectation vector and the covariance matrix, respectively. The prior knowledge of the expectation and covariance are defined by prior distributions. The length of the wavelet is fixed. In general, the wavelet length is determined by a compromise between minimizing the data residual and avoiding a too long wavelet. By increasing the wavelet length, the data residual will in general decrease since an increasing amount of local noise is explained by the wavelet. In principle, the wavelet length could be included in the wavelet estimation algorithm as an unknown variable, but this is not done in this study.

Alternatively, the wavelet could have been represented by a parameterized model, for example a Ricker wavelet with unknown amplitude, frequency and time shift, but this is not considered here.

Prior model

To obtain a fully specified stochastic model, prior distributions must be specified for the nodes without parents, that is the two error terms $\boldsymbol{e}_t$ and $\boldsymbol{e}_c$, the wavelet expectation and covariance $\boldsymbol{\mu}_s$ and $\boldsymbol{\Sigma}_s$, and the seismic noise covariance $\boldsymbol{\Sigma}_d$. A prior model for $\boldsymbol{c}_{obs}$ is not needed since it only appears in the model as a constant vector since it is observed.

The prior model for the time error vector $\boldsymbol{e}_t$ can be defined from a set of well markers $\{t_1, t_2, \ldots, t_m\}$. If a time shift Δt of a well marker can be restricted to a specified interval, a vague prior model is defined by the uniform distribution on this interval. The intervals for the well markers must be defined such that time order for the well markers remain unchanged. This is ensured if the legal time intervals are small compared to the distance between the well markers. Other distributions may be more appropriate than the uniform distribution, for example the triangular distribution

$$\Delta t \sim \mathcal{T}r\,(t_l, t_c, t_u), \tag{3.8}$$

where the probability is maximum in the center point t_c, and decreases linearly to zero at the low and upper limits t_l and t_u. If we consider the depth to time conversion of the well logs to be optimal based on our current knowledge, the

top of the triangular distribution then corresponds to no time shift, such that $t_c = 0$.

The reflectivity error is assumed to be zero mean Gaussian

$$\boldsymbol{e}_c \sim \mathcal{N}_{n_c}(\mathbf{0}, \boldsymbol{\Sigma}_c), \tag{3.9}$$

where $\boldsymbol{\Sigma}_c$ is the covariance matrix. Here, the reflectivity error is assumed to be white noise such that $\boldsymbol{\Sigma}_c = \sigma_c^2 \boldsymbol{I}$, with σ_c^2 being the variance of the errors. In prestack analysis, it is reasonable to assume high correlation between different reflection angles.

The wavelet is defined by a Gaussian likelihood, expression (3.7), with prior distributions on the expectation vector $\boldsymbol{\mu}_s$ and the covariance matrix $\boldsymbol{\Sigma}_s$. Usually, the expectation and covariance define a smooth wavelet with samples approaching zero at the ends. The likelihood expression for the wavelet can alternatively be regarded as the first level of a hierarchical prior model, the structural portion, while the distributions of the parameters $\boldsymbol{\mu}_s$ and $\boldsymbol{\Sigma}_s$ form the second level, the subjective portion. A prior model is often defined by a pdf specified by some fixed parameters, for example a Gaussian distribution with fixed expectation and variance. In a hierarchical prior model, the parameters which specify the prior pdf are regarded as uncertain. Each of these parameters is modeled by a pdf specified by hyperparameters, which gives a more flexible prior model. A general introduction to hierarchical Bayesian theory can be found in e.g., Robert (1994); Carlin (1996), but a deep insight into this concept is not required for this paper.

A mathematically convenient class of prior distributions is the conjugate prior distributions. These are distributions which have the same parametrical prior and posterior forms, but with different parameters. The conjugate distributions for expectation and variance in a Gaussian model are the Gaussian and the inverse gamma distributions, respectively. The expectation vector $\boldsymbol{\mu}_s$ is therefore modeled by a conjugate Gaussian distribution,

$$\boldsymbol{\mu}_s \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_{\mu_s}, \boldsymbol{\Sigma}_{\mu_s}), \tag{3.10}$$

where $\boldsymbol{\mu}_{\mu_s}$ and $\boldsymbol{\Sigma}_{\mu_s}$ are hyperparameters.

If we have sufficient prior knowledge about the structure of the covariance matrix $\boldsymbol{\Sigma}_s$ such that it is known up to an unknown multiplicative variance factor σ_s^2, the covariance matrix can be factorized as

$$\boldsymbol{\Sigma}_s = \sigma_s^2 \boldsymbol{\Sigma}_{0_s}. \tag{3.11}$$

The uncertainty of the unknown variance factor σ_s^2 is modeled by the inverse gamma distribution

$$\sigma_s^2 \sim \mathcal{IG}(\gamma_s, \lambda_s), \tag{3.12}$$

which is the conjugate distribution and the standard choice for unknown variances in hierarchical Bayesian models (Robert, 1994). The hyperparameters γ_s and λ_s in expression (3.12) should be defined in accordance with the prior knowledge. The expectation of the inverse gamma distribution is $\lambda/(\gamma-1)$, valid for $\gamma > 1$, and the variance is $\lambda^2/((\gamma-1)^2(\gamma-2))$, valid for $\gamma > 2$. A precise prior distribution with low variance is obtained by specifying large values for γ and λ such that the expectation becomes correct. A vague prior distribution is obtained by $\gamma \to 1$ and $\lambda \to 0$. In this case, the variance is undefined (infinite). See Appendix 3.B for more details about the inverse gamma distribution.

Similarly, if we have sufficient prior knowledge about the structure of the seismic noise covariance, the covariance matrix can be factorized as

$$\boldsymbol{\Sigma}_d = \sigma_d^2 \boldsymbol{\Sigma}_{0_d}, \tag{3.13}$$

where the uncertainty in the unknown variance factor σ_d^2 is inverse gamma distributed

$$\sigma_d^2 \sim \mathcal{IG}(\gamma_d, \lambda_d). \tag{3.14}$$

If the covariance matrices are completely unknown, the factorization in expressions (3.11) and (3.13) will not be possible. It is still possible to obtain conjugate prior distributions for the unknown covariances, however, by using the Wishart distribution (Anderson, 1984). This is not used in this study.

Posterior model

The objective of this wavelet estimation method is to find the posterior distributions for the involved variables based on the observed seismic data $\boldsymbol{d}_{obs}$ and the well log reflectivity $\boldsymbol{c}_{obs}$. Let $\boldsymbol{v} = \{v_1, \ldots, v_n\}$ represents the unknown quantities, where $\boldsymbol{v} = \{\boldsymbol{e}_t, \boldsymbol{e}_c, \boldsymbol{s}, \boldsymbol{\mu}_s, \sigma_s^2, \sigma_d^2\}$ in this problem.

The posterior distribution for $\boldsymbol{v}$ given $\boldsymbol{d}_{obs}$ and $\boldsymbol{c}_{obs}$ can be written

$$p(\boldsymbol{v}|\boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}) = \frac{p(\boldsymbol{v}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs})}{p(\boldsymbol{d}_{obs}, \boldsymbol{c}_{obs})} \tag{3.15}$$

where $p(\boldsymbol{v}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs})$ is the complete joint distribution for the problem. The pdf $p(\boldsymbol{d}_{obs}, \boldsymbol{c}_{obs})$ does not contain unknown variables, and can therefore be regarded as a constant. The complete joint distribution for a model represented by a DAG can be written (Spiegelhalter et al., 1996)

$$p(\boldsymbol{v}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}) = \prod_{v_i \in \{\boldsymbol{v}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}\}} p(v_i | \mathrm{Pa}(v_i)). \tag{3.16}$$

From expressions (3.15) and (3.16), and the DAG in Figure 3.1, the posterior distribution can now be written

$$\begin{aligned} p(\boldsymbol{e}_t, \boldsymbol{e}_c, \boldsymbol{s}, \boldsymbol{\mu}_s, \sigma_s^2, \sigma_d^2 | \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}) \propto \\ p(\boldsymbol{d}_{obs} | \boldsymbol{c}_{obs}, \boldsymbol{e}_t, \boldsymbol{e}_c, \boldsymbol{s}, \sigma_d^2) \\ p(\boldsymbol{e}_t)\ p(\boldsymbol{e}_c)\ p(\boldsymbol{s} | \boldsymbol{\mu}_s, \sigma_s^2)\ p(\boldsymbol{\mu}_s)\ p(\sigma_s^2)\ p(\sigma_d^2), \end{aligned} \tag{3.17}$$

where the proportionality is caused by unknown constant probability density functions involving $\boldsymbol{d}_{obs}$ and $\boldsymbol{c}_{obs}$.

Bayesian inference is completely based on the posterior distribution, where quantities as posterior mean and variance, maximum posterior distribution, quantiles and confidence (credibility) regions may be used. Interesting quantities can often be expressed as the posterior expectations of a function $g(\boldsymbol{v})$,

$$\mathrm{E}\{g(\boldsymbol{v}) | \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}\} = \int g(\boldsymbol{v})\ p(\boldsymbol{v} | \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs})\ d\boldsymbol{v}, \tag{3.18}$$

where for example $g(\boldsymbol{v}) = v$ for the posterior mean μ_v of variable v, and $g(\boldsymbol{v}) = (v - \mu_v)^2$ for the posterior variance.

The posterior distribution in expression (3.17) is not analytically obtainable, but can be explored by stochastic simulation. Several MCMC algorithms are available for this purpose, for example the Gibbs sampler and the Metropolis-Hastings algorithm, see e.g. Gilks et al. (1996); Chen et al. (2000). These algorithms provide a set of samples $\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \ldots, \boldsymbol{v}^{(n_{mc})}$ from the posterior distribution $p(\boldsymbol{v} | \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs})$. The integral in expression (3.18) can now be estimated by Monte Carlo integration:

$$\widehat{\mathrm{E}}\{g(\boldsymbol{v}) | \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}\} = \frac{1}{n_{mc}} \sum_{i=1}^{n_{mc}} g(\boldsymbol{v}^{(i)}). \tag{3.19}$$

A $(1 - \epsilon)100\%$ confidence interval for $g(\boldsymbol{v})$ can be approximated by

$$[g^{[\frac{\epsilon}{2} n_{mc}]}, g^{[(1 - \frac{\epsilon}{2}) n_{mc}]}], \tag{3.20}$$

where $g^{[i]}$ represents an increasing sorted sequence of $g^{[i]} = g(\boldsymbol{v}^{(\cdot)})$.

The MCMC algorithms are iterative, and may need a number of iterations before the first sample from the correct distribution can be provided. These initial iterations, termed burn-in iterations, should not be used in the statistical analysis. In general, the convergence rate is complicated to assess, but it can be monitored during the iteration procedure.

Gibbs sampler algorithm

The Gibbs sampler is perhaps the best known and most popular of the MCMC algorithms. The name of the algorithm was introduced by Geman and Geman (1984) who worked with Gibbs distributions on lattices, but the name is misleading as the application of the algorithm is general, and not restricted to Gibbs distributions.

The Gibbs sampler algorithm works on the complete joint distribution and requires conditional distributions for each single variable given all the others. One iteration consists of drawing new samples for the elements in $\boldsymbol{v}$. An element may be a single scalar, for example σ_d^2, or it may be a group of values, for example the vector $\boldsymbol{s}$. A new sample is drawn conditioned on the current state of the other elements in $\boldsymbol{v}$. Once an element is drawn, it goes into the current state of $\boldsymbol{v}$. The pseudo-code of the Gibbs sampling algorithm can be written

Initiate: Set arbitrary $\boldsymbol{v}^{(0)}$ where $p(\boldsymbol{v}^{(0)}|\boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}) > 0$.

Iterate: For $i = 1, 2, \ldots$, draw

$$\begin{aligned} v_1^{(i)} &\sim p(v_1|\boldsymbol{v}_{-1}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}), \\ &\vdots \\ v_j^{(i)} &\sim p(v_j|\boldsymbol{v}_{-j}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}), \\ &\vdots \\ v_n^{(i)} &\sim p(v_n|\boldsymbol{v}_{-n}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}), \end{aligned}$$

where $\boldsymbol{v}_{-j} = \{v_1^{(i)}, \ldots, v_{j-1}^{(i)}, v_{j+1}^{(i-1)}, \ldots, v_n^{(i-1)}\}$. The conditional distributions are specified in Appendix 3.C.

3.3 Examples

The method is applied on a data set from the Sleipner Øst Field in the Norwegian Block 15/9. Well logs and the corresponding CDP gather are shown in Figure 3.2, where the CDP gather has been transformed from offset to a set of reflection angles (5°, 13°, 21°, 29°, and 37°). A set of well markers are displayed by horizontal lines, numbered from 1 to 8. The last well marker corresponds to a strong sand/chalk contrast with a reflection coefficient of about 0.3.

In the following, wavelet estimation is first done from a CDP stack and zero incidence well log reflectivity, and then from the angle gather in Figure 3.2 and the corresponding angle dependent well log reflectivity.

Wavelet estimation from stacked data

As a first test on wavelet estimation from stacked data, assume that there are no errors in the well log information, such that the generalized convolutional model in expression (3.5) reduces to

$$\boldsymbol{d}_{obs} = \boldsymbol{c}_{obs} \star \boldsymbol{s} + \boldsymbol{e}_d. \tag{3.21}$$

Prior distributions are required for the wavelet expectation $\boldsymbol{\mu}_s$, the wavelet variance σ_s^2, and the seismic noise variance σ_d^2. This first test example has a simple model, but the problem has no simple analytical solution since σ_s^2 and σ_d^2 are unknown and independent.

A priori, we know that the maximum reflection coefficient in the target zone is about 0.3. Further, we know that the maximum amplitude of the seismic data is in the order of 10, and that the seismic data has been zero phased in an early step in the processing sequence. A vague prior model for the unknown wavelet is defined. The wavelet expectation has prior distribution

$$\boldsymbol{\mu}_s \sim \mathcal{N}_{n_s}(\mathbf{0}, 10^4 \boldsymbol{\Sigma}_{0_s}), \tag{3.22}$$

where $n_s = 35$. Initial tests showed that this was sufficient for a wavelet with one center loop and two side loops. The expected smoothness of the wavelet is imposed through $\boldsymbol{\Sigma}_{0_s}$, which is defined by a second order exponential correlation function, $\exp[-(\tau/d)^2]$, with range d=10 ms. The variance of the

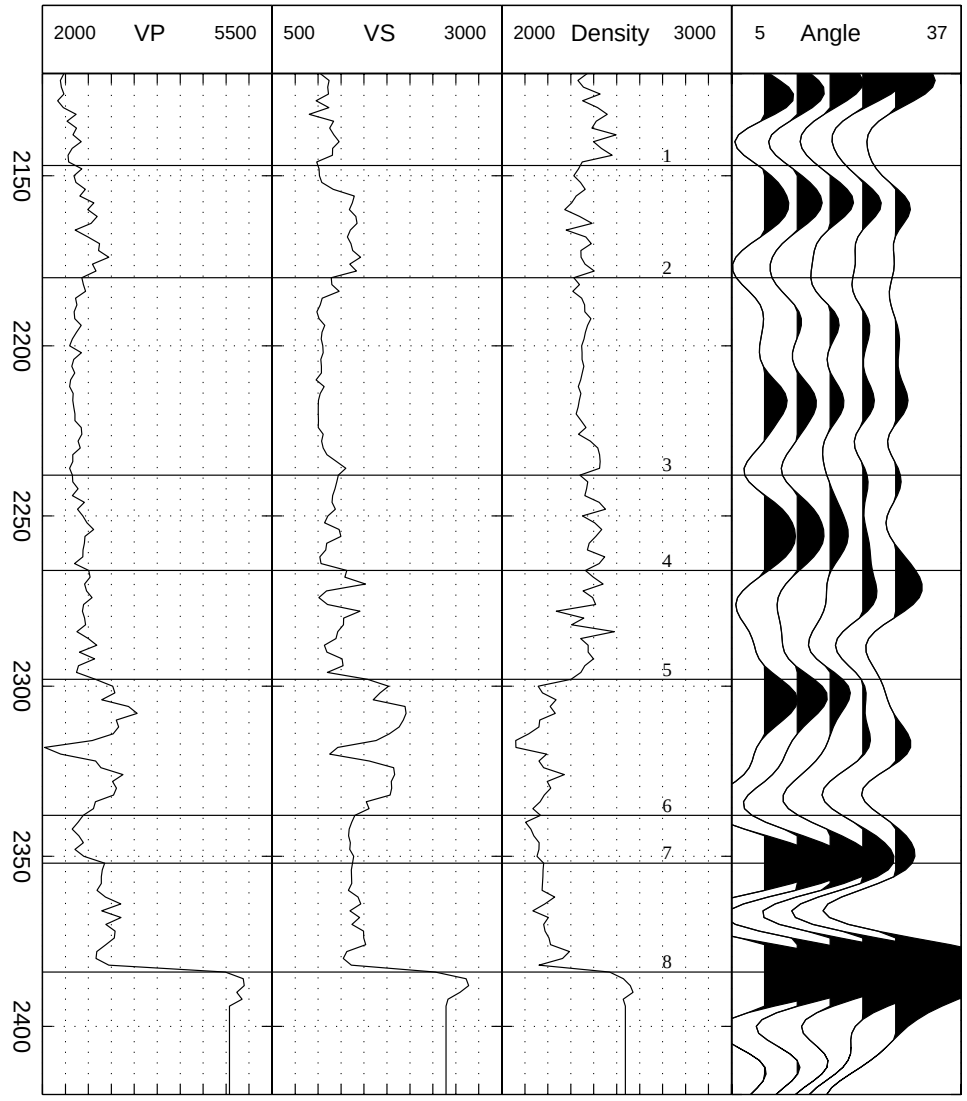


Figure 3.2: The well logs and the corresponding offset to angle transformed CDP gather.

wavelet expectation, $\sigma^2_{\mu_s} = 10^4$, is large compared to the wavelet amplitude. The prior distribution for the wavelet covariance is

$$\begin{aligned} \boldsymbol{\Sigma}_s &= \sigma_s^2 \boldsymbol{\Sigma}_{0_s}, & (3.23) \\ \sigma_s^2 &\sim \mathcal{IG}(2,5), & (3.24) \end{aligned}$$

where the variance factor σ_s^2 has expectation 5 and infinite variance.

The unknown data variance, σ_d^2, is also given a vague inverse gamma distributed prior

$$\sigma_d^2 \sim \mathcal{IG}(2,5). \quad (3.25)$$

The Gibbs sampler algorithm was used to draw samples from the posterior distribution, and 2000 samples were generated. The convergence of the Gibbs simulation can be monitored by plotting variable values for each iteration. A monitor plot of the maximum wavelet amplitude and the seismic noise variance is shown in Figure 3.3. The initial values are 0 and 5, respectively, and both quantities converge within some few iterations. Tests with other initial values show the same desirable convergence properties. To be sure, the number of burn-in iterations is set to 100, and these samples are not used in subsequent calculations.

The posterior expectation and 95 % confidence bounds for the wavelet are displayed in Figure 3.4. The posterior distribution of the unknown seismic noise variance σ_d^2 is shown in Figure 3.5. The estimated maximum wavelet amplitude is 26.5, and estimated mean data variance is 2.2.

When errors in the well log reflectivity are ignored as in the example above, the estimated seismic noise variance, σ_d^2, is too high. The relatively large misfit between the seismic well log response and the observed seismic data has a damping effect on the wavelet estimation. The error terms $\boldsymbol{e}_t$ and $\boldsymbol{e}_c$ are now included in the model as defined in expression (3.5). The strong contrast at well marker 8 is used as a reference time, hence well marker 8 is fixed at this reference time to avoid a complete translation of the log. The time shifts of the other well markers are triangular distributed

$$\Delta t_i \sim \mathcal{T}r\,(-8\,\text{ms}, 0\,\text{ms}, 8\,\text{ms}), \quad (3.26)$$

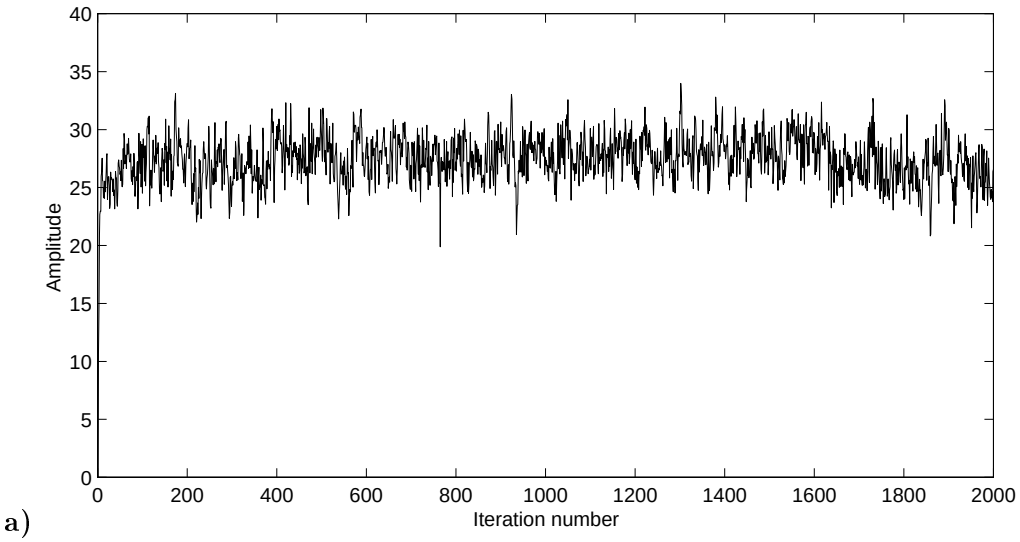
and the reflectivity errors are zero mean white Gaussian

$$\boldsymbol{e}_c \sim \mathcal{N}_{n_c}(\boldsymbol{0}, 10^{-5}\boldsymbol{I}). \quad (3.27)$$

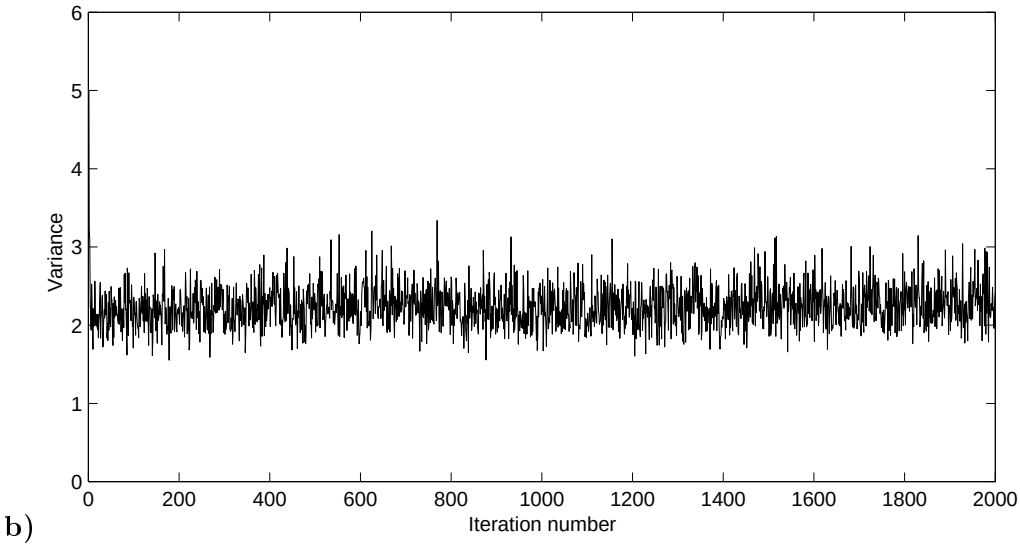
In this example, the variance is set to $\sigma_c^2 = 10^{-5}$, such that 95 % of the reflectivity errors are within ± 0.0063, corresponding to two standard deviations from the expectation value. For comparison, the mean well log reflection coefficient is 0.02 (absolute value), while the maximum reflection coefficient is about 0.3.

The posterior expectation and 95 % confidence bounds for the wavelet are displayed in Figure 3.6, while the posterior distribution of the unknown seismic noise variance, σ_d^2, is shown in Figure 3.7. By including well log errors in the model, the estimated maximum wavelet amplitude is now increased from 26.5 to 32.3, while the estimated mean data variance is reduced from 2.2 to 0.26. The uncertainty of the wavelet is decreased since the well tie is improved by adjusting the well logs. Note that this improved wavelet is not within the 95% confidence interval in Figure 3.4. The reason is that the confidence region in Figure 3.4 is based on a simplified model neglecting the well log errors. The posterior distributions of the well markers are shown in Figure 3.8, where the position of well marker 8 is fixed. The result shows that the total time window of the log is slightly stretched, while the layer between marker 5 and 6 is squeezed.

The real stack trace $\boldsymbol{d}_{obs}$, the expected synthetic trace $\boldsymbol{c} \star \boldsymbol{s}$, the corresponding seismic error $\boldsymbol{e}_d$, and the expected reflectivity error response $\boldsymbol{e}_c \star \boldsymbol{s}$ are shown in Figure 3.9. In principle, also the variance σ_c^2 of the well log reflectivity error $\boldsymbol{e}_c$, could be regarded as an unknown quantity with a vague prior distribution, similarly as σ_s^2 and σ_d^2. The problem with this approach is that the error terms $\boldsymbol{e}_d$ and $\boldsymbol{e}_c \star \boldsymbol{s}$ are not completely separable. The prior model for the variance of the well log error must therefore be specified exactly, or at least relatively precise. The effect of different σ_c^2 is shown in Figure 3.10. For reflection error variances below 10^{-4}, the estimation of the wavelet is stable, also with respect to uncertainty. Variances larger than 10^{-4} allows for reflectivity errors larger than the typical well log reflection coefficient, and this has a dramatic damping effect on the wavelet. As expected, the estimated seismic noise variance decreases with increasing reflection error variance. Reflection error variances below 10^{-6} only have a marginal effect.



a)



b)

Figure 3.3: Monitoring maximum wavelet amplitude (a) and seismic noise variance (b).

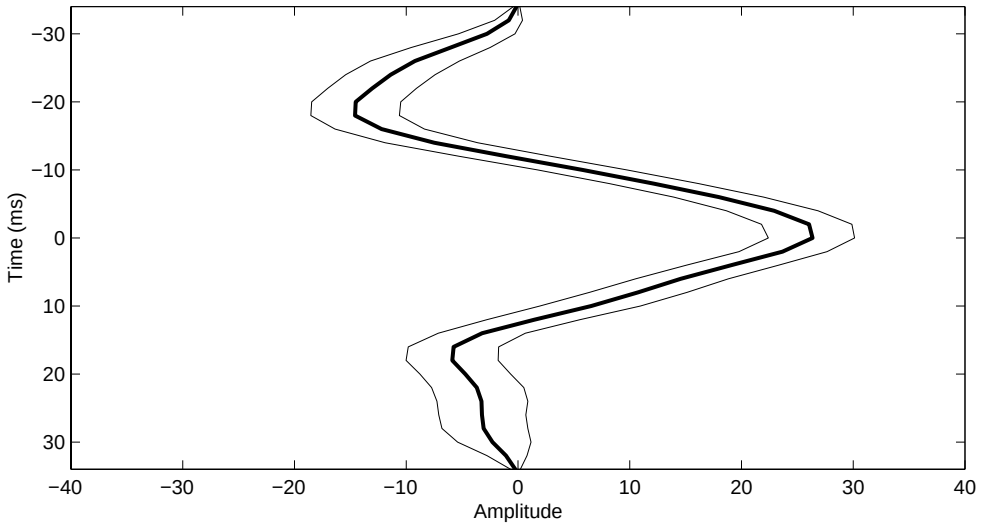


Figure 3.4: Estimated wavelet from stacked data assuming no well log errors.

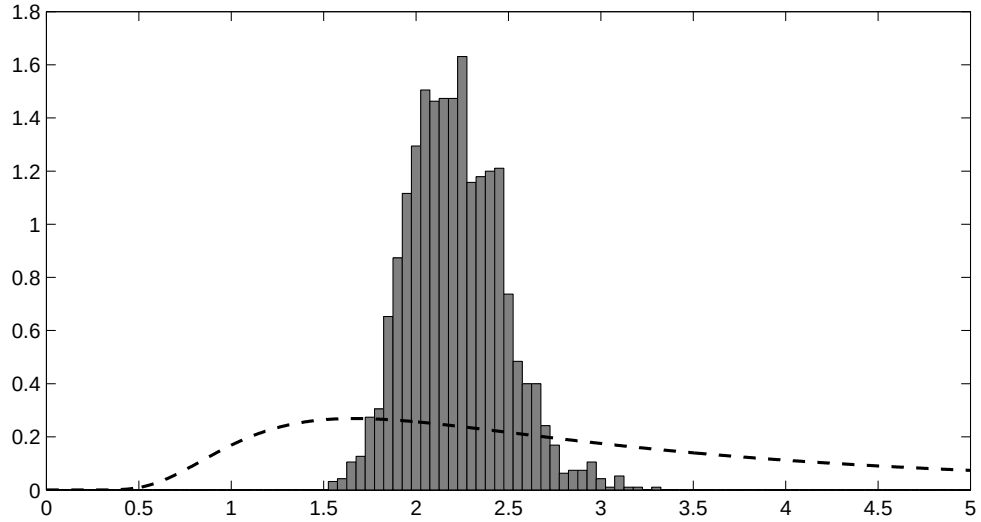


Figure 3.5: Estimated seismic noise variance from stacked data assuming no well log errors. The prior pdf is dashed.

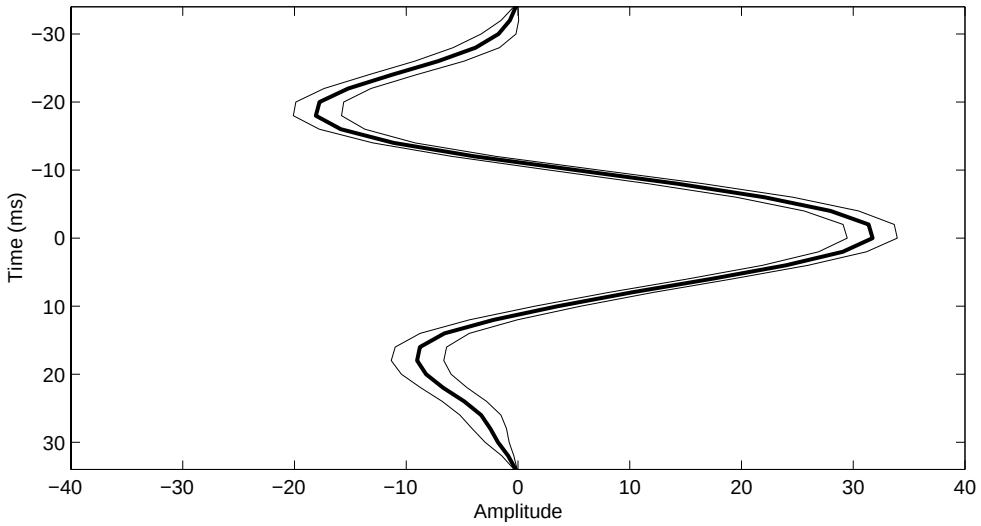


Figure 3.6: Estimated wavelet from stacked data including well log stretching and well log errors.

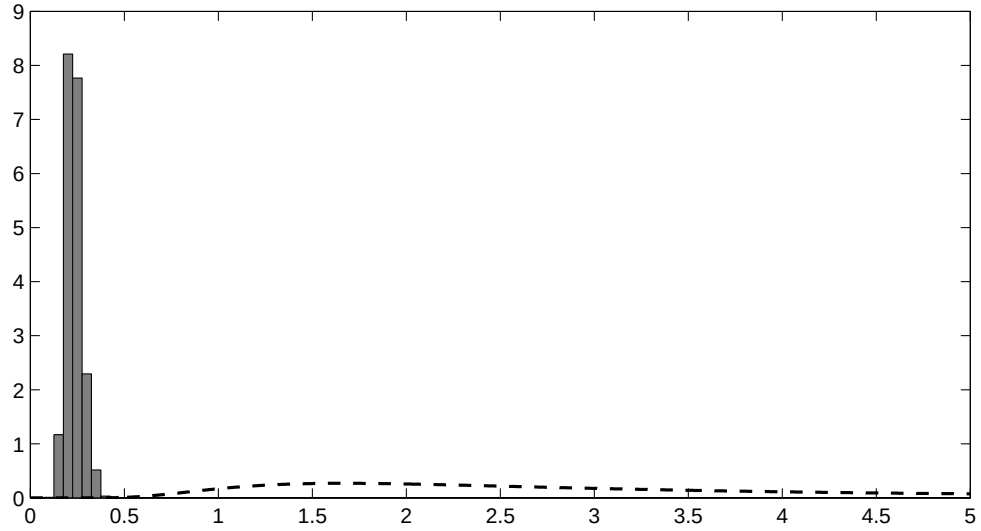


Figure 3.7: Estimated seismic noise variance from stacked data including well log stretching and well log errors. The prior pdf is dashed.

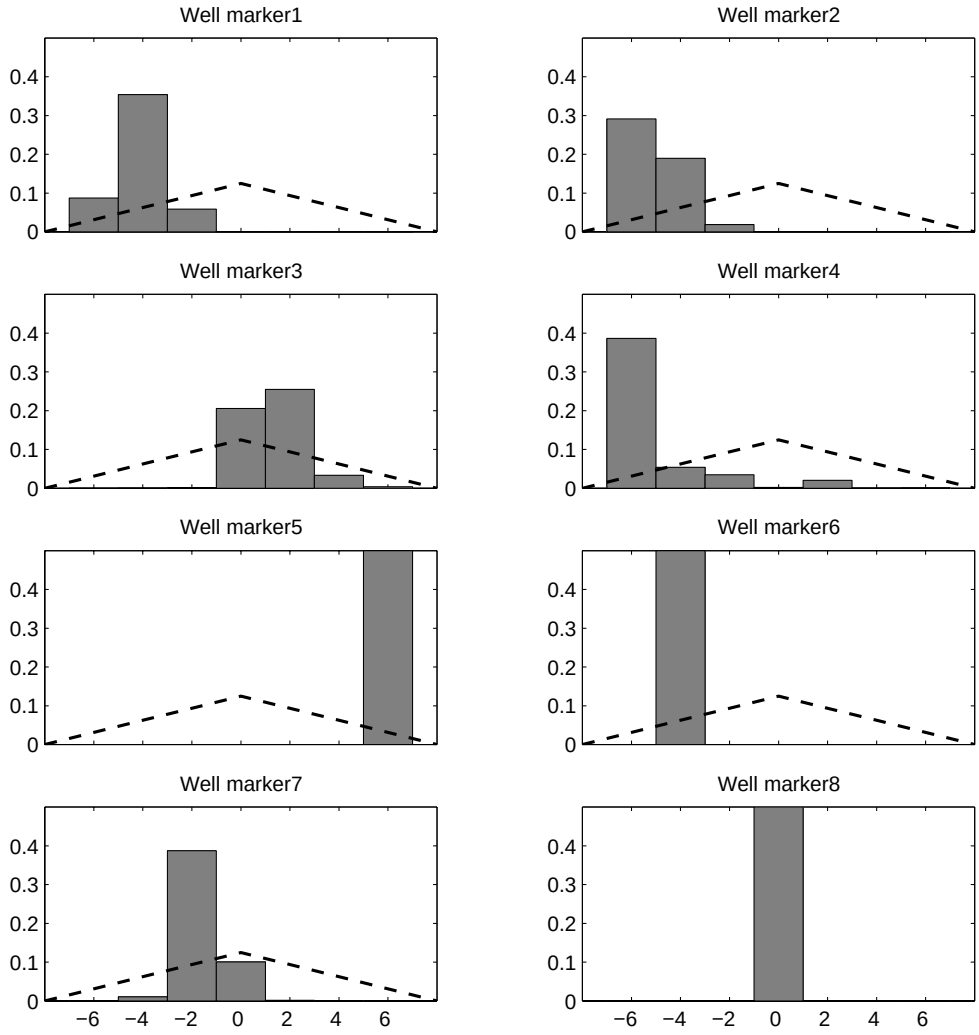


Figure 3.8: Estimated time shifts (ms) of the well markers from stacked data. Note that well marker 8 is fixed. The prior triangular distribution is dashed.

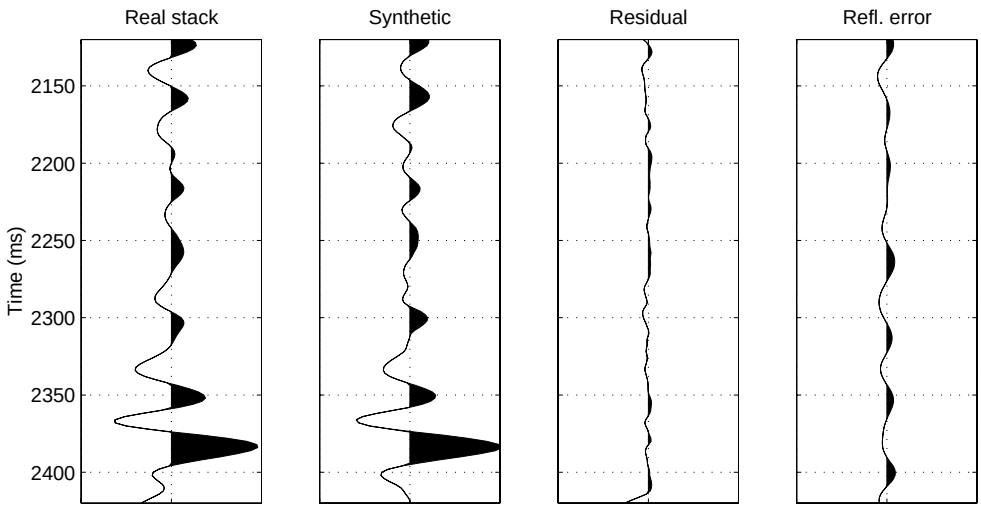


Figure 3.9: The real stack trace $\boldsymbol{d}_{obs}$, the synthetic trace $\boldsymbol{c} \star \boldsymbol{s}$, the seismic error $\boldsymbol{e}_d$, and the reflectivity error response $\boldsymbol{e}_c \star \boldsymbol{s}$.

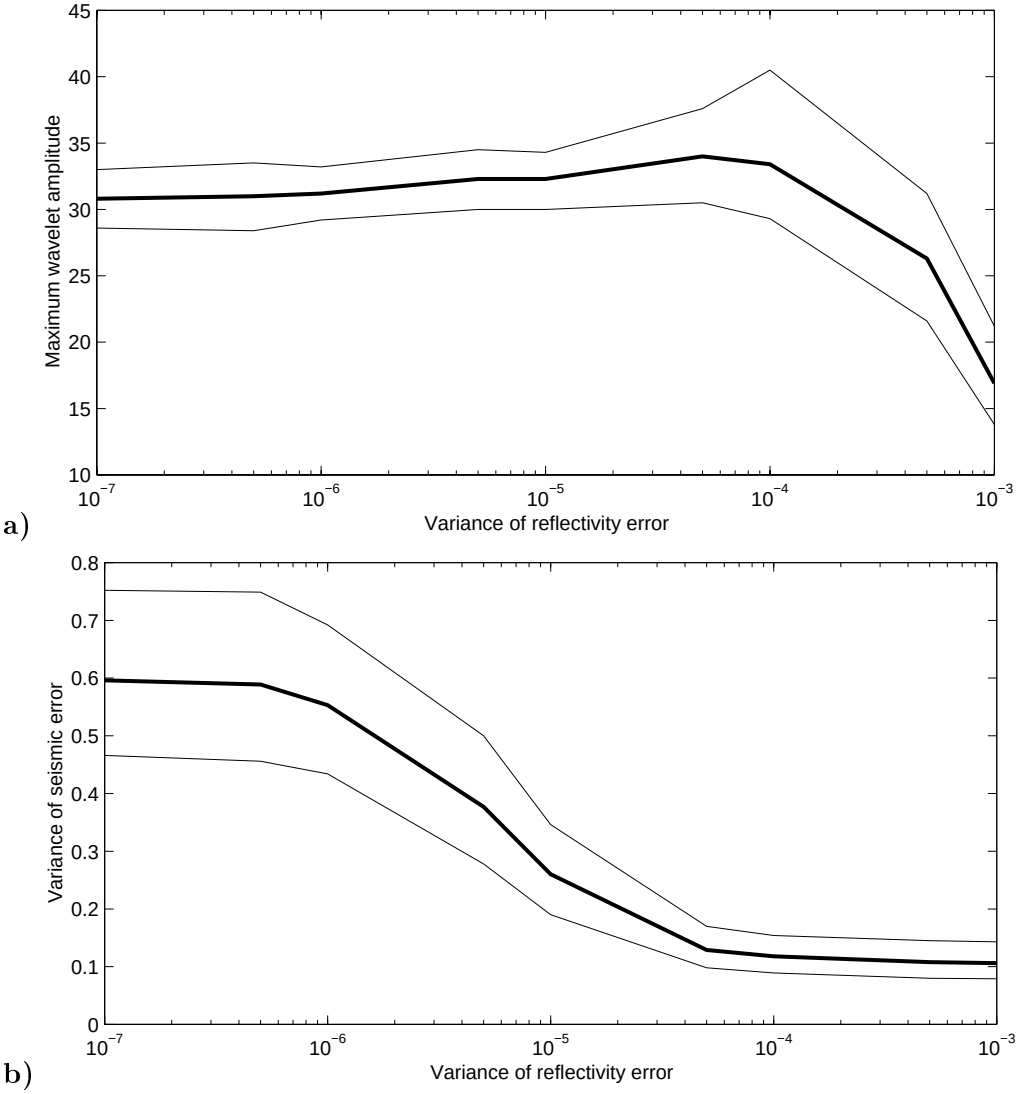


Figure 3.10: Maximum wavelet amplitude (a) and seismic noise variance (b) as function of reflectivity error variance, σ_c^2. The thin lines show the 95 % confidence intervals.

Wavelet estimation from prestack data

The extension of the wavelet estimation to prestack data in form of angle gathers is straightforward. In this prestack example, a complete model is used including well log errors, and the prior distributions are as in the poststack example. It is reasonable to assume a high degree of similarity between the wavelets corresponding to different reflection angles. This is obtained by an exponential correlation function, $\nu_\theta = \exp[-|\theta_i - \theta_j|/d_\theta]$, where $d_\theta = 20^\circ$ is the choosen correlation range. The log errors are still assumed to be white as function of time, but a strong correlation between different angles are imposed by the second order correlation function $\exp[-(\theta_i - \theta_j)^2/80^\circ]$.

From Gibbs sampling, the posterior expectation and 95 % confidence bounds for the five angle wavelets are displayed in Figure 3.11. The maximum amplitudes are slightly higher than the poststack amplitude, ranging from 34 for the lowest angle to 43 for the largest angle. Note also that the frequency content of the wavelets decreases with increasing reflection angle. The posterior distribution of the seismic noise variance, σ_d^2, is shown in Figure 3.12. The estimated mean value of the noise variance is 0.50, which is about twice the stack noise variance. The posterior distributions of the time shifts are more peaked than in the poststack example, see Figure 3.13.

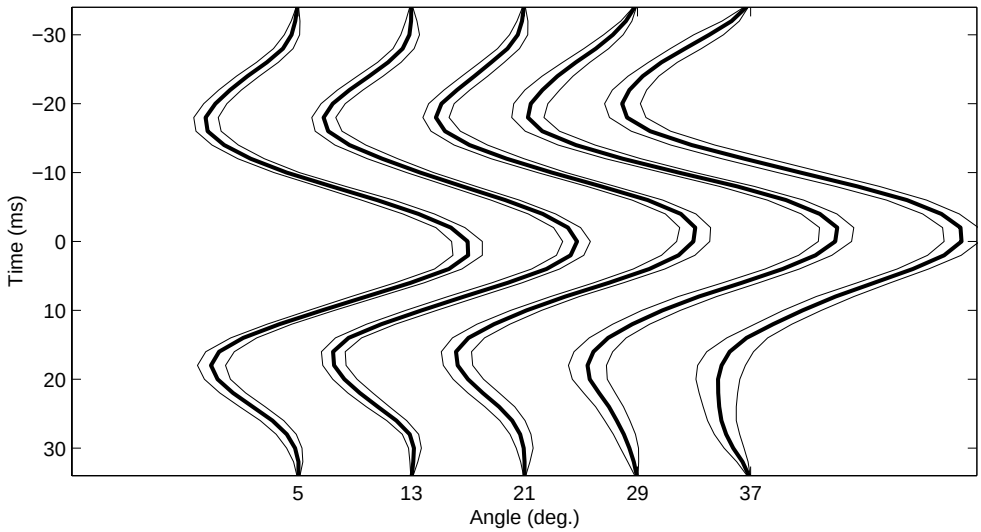


Figure 3.11: Estimated wavelets from prestack data including well log stretching and well log errors.

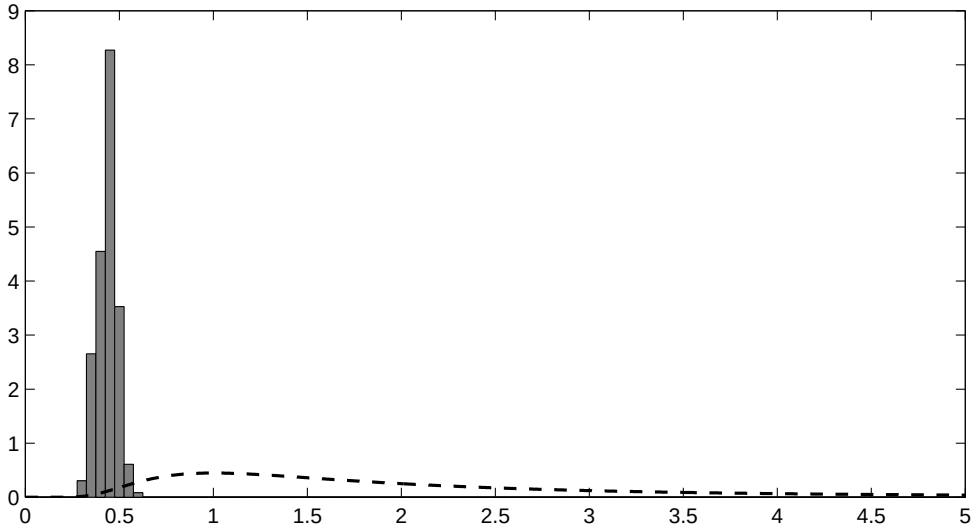


Figure 3.12: Estimated seismic noise variance from prestack data including well log stretching and well log errors. The prior pdf is dashed.

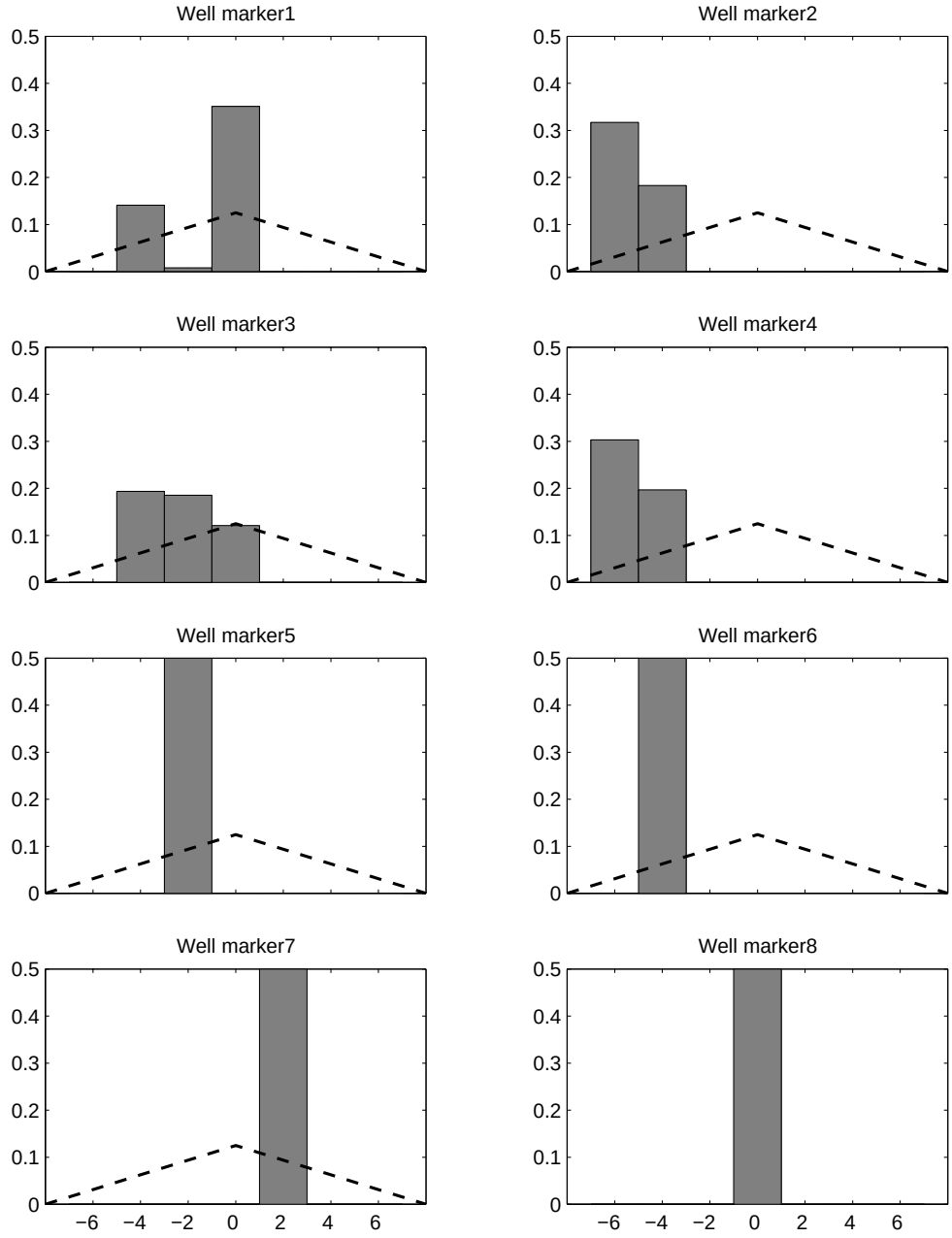


Figure 3.13: Estimated time shifts (ms) of the well markers from prestack data. Note that well marker 8 is fixed. The prior triangular distribution is dashed.

3.4 Discussion

Based on the convolutional model, the estimated wavelet can be interpreted as the link between the measured well log and the seismic data. In contrast to the physical source signature at the source, a wavelet estimated from seismic and well data is not a physical source wavelet. In addition to the source signature it also contains components from wave propagation in the overburden and from the processing sequence. When the estimated wavelet is used in inversion or by other tools for detecting spatial variation of the geology, it is usually assumed that the wavelet is laterally constant. The validness of this assumption depends on the acquisition, the wave propagation through the overburden, and the processing of the seismic data. These problems were discussed in Ziolkowski (1991), where also the value of wavelet estimation from seismic and well data was questioned. As an alternative, a deterministic approach was proposed where the seismic signature at the source is measured, and the seismic prestack response is calculated by a complete forward modeling including all kinds of propagating effects like mode conversions, interbed multiples, etc.. However, this strategy has yet not become a routinely viable method. Ziolkowski (1991) initiated an intensive debate where the front line was between strict theory and more practical approaches. Despite some theoretical objections, there is no doubt that wavelet estimation from well log data and processed seismic data based on the convolutional model has shown to be a very powerful technique in detailed seismic analysis.

3.5 Conclusions

A Bayesian method for wavelet estimation from seismic and well data is developed, where uncertainty in the wavelet estimation process is included. The problem is nonlinear, and the solution is explored by MCMC simulation. The applied Gibbs sampler has a fast convergence. The method works both on stacked data, and prestack data in form of angle gathers. The forward model is based on the convolutional model, where the reflectivity is calculated from the well logs. Large misfits between the synthetic well log seismogram and the real seismic data have a damping effect on the wavelet estimation. When well log errors are included in the model, the uncertainty of the wavelet estimation decreases and the amplitude of the estimated wavelets increase. Uncertainty bounds on the estimated seismic wavelet may contribute to more realistic un-

certainty models related to detailed seismic analysis where a seismic wavelet is a required input.

3.6 Acknowledgments

We thank Statoil for permission to publish this paper. We will also thank the reviewers for a thorough review and useful comments.

3.A Gaussian distribution

The multi-Gaussian probability density is

$$p(\boldsymbol{r}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{r}-\boldsymbol{\mu})\right], \tag{3.A.1}$$

where n is the dimension of $\boldsymbol{r}$, and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the expectation vector and the covariance matrix, respectively.

Consider two multivariate Gaussian variables $\boldsymbol{r}_1 \sim \mathcal{N}_{n_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\boldsymbol{r}_2 \sim \mathcal{N}_{n_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ with joint distribution

$$\begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \end{bmatrix} \sim \mathcal{N}_{n_1+n_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right), \tag{3.A.2}$$

where n_1 and n_2 are the dimensions. Then the conditional distribution of $\boldsymbol{r}_1$ given $\boldsymbol{r}_2$ is Gaussian with expectation

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{r}_2 - \boldsymbol{\mu}_2) \tag{3.A.3}$$

and covariance

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{3.A.4}$$

3.B Gamma and inverse gamma distribution

The gamma distribution $\mathcal{G}(\gamma, \lambda)$ is defined by the probability density function

$$p(x|\gamma, \lambda) = \frac{\lambda^\gamma}{\Gamma(\gamma)} x^{\gamma-1} \exp[-\lambda x], \tag{3.B.1}$$

for $x \geq 0,\ \gamma > 0,\ \lambda > 0$. The expectation and variance are

$$E\{x|\gamma,\lambda\} = \frac{\gamma}{\lambda}, \tag{3.B.2}$$

$$\mathrm{Var}\{x|\gamma,\lambda\} = \frac{\gamma}{\lambda^2}. \tag{3.B.3}$$

Two special cases of the gamma distribution are the exponential distribution $\mathcal{E}(\lambda)$ for $\gamma = 1$, and the chi-squared distribution $\mathcal{X}^2_\nu$ for $\gamma = \nu/2$ and $\lambda = 1/2$.

If $y \sim \mathcal{G}(\gamma,\lambda)$, then $x = 1/y$ has the inverse gamma distribution $\mathcal{IG}(\gamma,\lambda)$ defined by the probability density function

$$p(x|\gamma,\lambda) = \frac{\lambda^\gamma}{\Gamma(\gamma)}\left(\frac{1}{x}\right)^{\gamma+1}\exp\left[-\frac{\lambda}{x}\right], \tag{3.B.4}$$

for $x \geq 0,\ \gamma > 0,\ \lambda > 0$. The expectation and variance are

$$E\{x|\gamma,\lambda\} = \frac{\lambda}{\gamma-1}, \qquad \gamma > 1, \tag{3.B.5}$$

$$\mathrm{Var}\{x|\gamma,\lambda\} = \frac{\lambda^2}{(\gamma-1)^2(\gamma-2)}, \qquad \gamma > 2. \tag{3.B.6}$$

The expectation and variance in expressions (3.B.5) and (3.B.6) approach infinity when $\gamma \to 1$ and $\gamma \to 2$ (from above), respectively. For $\gamma \leq 1$ and $\gamma \leq 2$ ($\gamma > 0$), the expectation and variance are not defined.

For a Gaussian distribution $\mathcal{N}_1(\mu,\sigma^2)$ with unknown variance, the inverse gamma distribution is a conjugate distribution for the variance. That means that if the prior distribution for the variance is inverse gamma $\sigma^2 \sim \mathcal{IG}(\gamma,\lambda)$, then the posterior distribution is also inverse gamma, but with modified parameters. For $x_1,\ldots,x_n$ iid $\mathcal{N}_1(\mu,\sigma^2)$, the posterior distribution for the unknown variance is

$$\sigma^2|s^2 \sim \mathcal{IG}(\gamma+\frac{n}{2}, \lambda+\frac{ns^2}{2}), \tag{3.B.7}$$

where

$$s^2 = \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{n}. \tag{3.B.8}$$

3.C Conditional distributions in Gibbs sampling

The Gibbs sampler algorithm requires conditional distributions for each single variable given all the others. For a DAG model, the joint distribution is

$$p(\boldsymbol{v}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}) = \prod_{v_i \in \{\boldsymbol{v}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}\}} p(v_i | \mathrm{Pa}(v_i)), \tag{3.C.1}$$

and the full conditional distribution for a variable v_i can be constructed by using only the terms $p(\cdot|\cdot)$ which involve the actual variable v_i. In the following, all the necessary conditional distributions are defined.

Conditional distribution for $\boldsymbol{e}_t$

The conditional distribution for the time error term $\boldsymbol{e}_t$ given the other parameters is not analytical available, but if the time error function is defined by a set of well markers taking discrete time shift values, the total conditional distribution for $\boldsymbol{e}_t$ can be calculated. For each well marker, the following algorithm can be applied:

- For all possible discrete time shifts Δt, calculate the corresponding
  - $\tilde{\boldsymbol{e}}_t$,
  - $\tilde{\boldsymbol{c}} = \boldsymbol{c}_{obs}(\tilde{\boldsymbol{e}}_t) + \boldsymbol{e}_c$.
  - $\tilde{\boldsymbol{e}}_d = \boldsymbol{d}_{obs} - \boldsymbol{s} \star \tilde{\boldsymbol{c}}$
  - $p(\Delta t|\cdot) = \exp[-\frac{1}{2}\tilde{\boldsymbol{e}}_d{}^T \boldsymbol{\Sigma}_d^{-1} \tilde{\boldsymbol{e}}_d]\, p(\Delta t)$
- Normalize $p(\Delta t|\cdot)$
- Draw a time shift Δt from $p(\Delta t|\cdot)$, and calculate the updated $\boldsymbol{e}_t$.

Conditional distribution for $\boldsymbol{e}_c$

When the wavelet $\boldsymbol{s}$ is given, the convolution can be reformulated to a matrix-vector multiplication

$$\boldsymbol{d}_{obs} = \boldsymbol{S}\{\boldsymbol{c}_{obs}(\boldsymbol{e}_t) + \boldsymbol{e}_c\} + \boldsymbol{e}_d, \tag{3.C.2}$$

where the matrix $\boldsymbol{S}$ contains the given wavelet. The conditional joint distribution for the reflectivity error and the seismic data is

$$\left[\begin{array}{c} \boldsymbol{e}_c \\ \boldsymbol{d}_{obs} \end{array} \middle| \boldsymbol{c}_{obs}, \boldsymbol{e}_t, \boldsymbol{s}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_d \right] \sim \mathcal{N}_{n_c+n_d} \left(\left[\begin{array}{c} \boldsymbol{0} \\ \boldsymbol{S}\boldsymbol{c}_{obs}(\boldsymbol{e}_t) \end{array} \right], \left[\begin{array}{cc} \boldsymbol{\Sigma}_c & \boldsymbol{\Sigma}_c\boldsymbol{S}^T \\ \boldsymbol{S}\boldsymbol{\Sigma}_c & \boldsymbol{S}\boldsymbol{\Sigma}_c\boldsymbol{S}^T + \boldsymbol{\Sigma}_d \end{array} \right] \right). \quad (3.C.3)$$

The full conditional distribution for $\boldsymbol{e}_c$ is

$$\boldsymbol{e}_c | \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}, \boldsymbol{e}_t, \boldsymbol{s}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_d \sim \mathcal{N}_{n_c}(\boldsymbol{\mu}_{e_c|\cdot}, \boldsymbol{\Sigma}_{e_c|\cdot}), \quad (3.C.4)$$

where

$$\boldsymbol{\mu}_{e_c|\cdot} = \boldsymbol{\Sigma}_c\boldsymbol{S}^T(\boldsymbol{S}\boldsymbol{\Sigma}_c\boldsymbol{S}^T + \boldsymbol{\Sigma}_d)^{-1}(\boldsymbol{d}_{obs} - \boldsymbol{S}\boldsymbol{c}_{obs}(\boldsymbol{e}_t)), \quad (3.C.5)$$

$$\boldsymbol{\Sigma}_{e_c|\cdot} = \boldsymbol{\Sigma}_c - \boldsymbol{\Sigma}_c\boldsymbol{S}^T(\boldsymbol{S}\boldsymbol{\Sigma}_c\boldsymbol{S}^T + \boldsymbol{\Sigma}_d)^{-1}\boldsymbol{S}\boldsymbol{\Sigma}_c. \quad (3.C.6)$$

Conditional distribution for $\boldsymbol{s}$

The convolution can be reformulated to a matrix-vector multiplication

$$\boldsymbol{d}_{obs} = \boldsymbol{C}\boldsymbol{s} + \boldsymbol{e}_d, \quad (3.C.7)$$

where the matrix $\boldsymbol{C}$ contains the given reflectivity $\boldsymbol{c} = \boldsymbol{c}_{obs}(\boldsymbol{e}_t) + \boldsymbol{e}_c$. The conditional joint distribution for the wavelet and the seismic data is

$$\left[\begin{array}{c} \boldsymbol{s} \\ \boldsymbol{d}_{obs} \end{array} \middle| \boldsymbol{c}, \boldsymbol{\Sigma}_d \right] \sim \mathcal{N}_{n_s+n_d} \left(\left[\begin{array}{c} \boldsymbol{\mu}_s \\ \boldsymbol{C}\boldsymbol{\mu}_s \end{array} \right], \left[\begin{array}{cc} \boldsymbol{\Sigma}_s & \boldsymbol{\Sigma}_s\boldsymbol{C}^T \\ \boldsymbol{C}\boldsymbol{\Sigma}_s & \boldsymbol{C}\boldsymbol{\Sigma}_s\boldsymbol{C}^T + \boldsymbol{\Sigma}_d \end{array} \right] \right). \quad (3.C.8)$$

The full conditional distribution for $\boldsymbol{s}$ is

$$\boldsymbol{s} | \boldsymbol{d}_{obs}, \boldsymbol{c}, \boldsymbol{\Sigma}_d \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_{s|\cdot}, \boldsymbol{\Sigma}_{s|\cdot}), \quad (3.C.9)$$

where

$$\boldsymbol{\mu}_{s|\cdot} = \boldsymbol{\mu}_s + \boldsymbol{\Sigma}_s\boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{\Sigma}_s\boldsymbol{C}^T + \boldsymbol{\Sigma}_d)^{-1}(\boldsymbol{d}_{obs} - \boldsymbol{C}\boldsymbol{\mu}_s), \quad (3.C.10)$$

$$\boldsymbol{\Sigma}_{s|\cdot} = \boldsymbol{\Sigma}_s - \boldsymbol{\Sigma}_s\boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{\Sigma}_s\boldsymbol{C}^T + \boldsymbol{\Sigma}_d)^{-1}\boldsymbol{C}\boldsymbol{\Sigma}_s. \quad (3.C.11)$$

Conditional distribution for $\boldsymbol{\mu}_s$

The conditional distribution for the wavelet expectation $\boldsymbol{\mu}_s$ depends on the wavelet $\boldsymbol{s}$ and the covariance $\boldsymbol{\Sigma}_s$. Both $\boldsymbol{\mu}_s$ and $\boldsymbol{s}$ are Gaussian, and the conditional joint distribution is

$$\left[\begin{array}{c} \boldsymbol{\mu_s} \\ \boldsymbol{s} \end{array} \middle| \boldsymbol{\Sigma}_s \right] \sim \mathcal{N}_{n_s+n_s} \left(\left[\begin{array}{c} \boldsymbol{\mu}_{\mu_s} \\ \boldsymbol{\mu}_{\mu_s} \end{array} \right], \left[\begin{array}{cc} \boldsymbol{\Sigma}_{\mu_s} & \boldsymbol{\Sigma}_{\mu_s} \\ \boldsymbol{\Sigma}_{\mu_s} & \boldsymbol{\Sigma}_{\mu_s} + \boldsymbol{\Sigma}_s \end{array} \right] \right). \tag{3.C.12}$$

The full conditional distribution for $\boldsymbol{\mu}_s$ is

$$\boldsymbol{\mu}_s | \boldsymbol{s}, \boldsymbol{\Sigma}_s \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_{\mu_s|\cdot}, \boldsymbol{\Sigma}_{\mu_s|\cdot}), \tag{3.C.13}$$

where

$$\boldsymbol{\mu}_{\mu_s|\cdot} = \boldsymbol{\mu}_{\mu_s} + \boldsymbol{\Sigma}_{\mu_s}(\boldsymbol{\Sigma}_{\mu_s} + \boldsymbol{\Sigma}_s)^{-1}(\boldsymbol{s} - \boldsymbol{\mu}_{\mu_s}), \tag{3.C.14}$$

$$\boldsymbol{\Sigma}_{\mu_s|\cdot} = \boldsymbol{\Sigma}_{\mu_s} - \boldsymbol{\Sigma}_{\mu_s}(\boldsymbol{\Sigma}_{\mu_s} + \boldsymbol{\Sigma}_s)^{-1}\boldsymbol{\Sigma}_{\mu_s}. \tag{3.C.15}$$

Conditional distribution for σ_s^2

The conditional distribution for the wavelet variance σ_s^2 depends on the wavelet $\boldsymbol{s}$ and the expectation $\boldsymbol{\mu}_s$. When the prior distribution for σ_s^2 is inverse gamma

$$\sigma_s^2 \sim \mathcal{IG}(\gamma_s, \lambda_s), \tag{3.C.16}$$

then the full conditional distribution for σ_s^2 is inverse gamma

$$\sigma_s^2 | s_s^2 \sim \mathcal{IG}(\gamma_s + \frac{n_s}{2}, \lambda_s + \frac{n_s s_s^2}{2}), \tag{3.C.17}$$

where

$$s_s^2 = \frac{(\boldsymbol{s} - \boldsymbol{\mu}_s)^T \boldsymbol{\Sigma}_{0_s}^{-1}(\boldsymbol{s} - \boldsymbol{\mu}_s)}{n_s}. \tag{3.C.18}$$

Conditional distribution for σ_d^2

The conditional distribution for the data variance σ_d^2 depends on the seismic data $\boldsymbol{d}_{obs}$ and the expectation $\boldsymbol{\mu}_d = \boldsymbol{s} \star \boldsymbol{c}$, where $\boldsymbol{c} = \boldsymbol{c}_{obs}(\boldsymbol{e}_t) + \boldsymbol{e}_c$. When the prior distribution for σ_d^2 is inverse Gamma

$$\sigma_d^2 \sim \mathcal{IG}(\gamma_d, \lambda_d), \tag{3.C.19}$$

then the full conditional distribution for σ_d^2 is inverse gamma

$$\sigma_d^2 | s_d^2 \sim \mathcal{IG}(\gamma_d + \frac{n_d}{2}, \lambda_d + \frac{n_d s_d^2}{2}), \tag{3.C.20}$$

where

$$s_d^2 = \frac{(\boldsymbol{d}_{obs} - \boldsymbol{\mu}_d)^T \boldsymbol{\Sigma}_{0_d}^{-1} (\boldsymbol{d}_{obs} - \boldsymbol{\mu}_d)}{n_d}. \tag{3.C.21}$$

Chapter 4

Joint AVO inversion, wavelet estimation, and noise level estimation using a spatially coupled hierarchical Bayesian model

Arild Buland and Henning Omre

Paper submitted for publication.

Abstract

The main objective of the AVO inversion is to obtain posterior distributions for P-wave velocity, S-wave velocity, and density from specified prior distributions, seismic data, and well log data. The inversion problem also involves estimation of a seismic wavelet and the seismic noise level. The noise model is represented by a zero mean Gaussian distribution specified by a covariance matrix. A method for joint AVO inversion, wavelet estimation, and estimation of the noise level is developed in a Bayesian framework. The stochastic model includes uncertainty of both the elastic parameters, the wavelet, and the seismic and well log data. The posterior distribution is explored by Markov chain Monte Carlo simulation using the Gibbs sampler algorithm. The inversion

algorithm has been tested on a seismic line from the Heidrun Field with two wells located on the line. The use of a colored seismic noise model resulted in about 10% lower uncertainties for the P-wave velocity, S-wave velocity, and density compared to a white noise model. The uncertainty of the estimated wavelet is low. In the Heidrun example, the effect of including uncertainty of the wavelet and the noise level was marginal with respect to the AVO inversion results.

4.1 Introduction

Geophysical measurements are of crucial importance to make models of the structures and the physical properties of the subsurface. The geophysical measurements are often strongly affected by noise and measurement uncertainty, and the established subsurface models may be highly uncertain. Quantification of the uncertainty is important to correctly appreciate these subsurface models. In a Bayesian setting, available prior knowledge is combined with the information contained in the measured data (Tarantola and Valette, 1982; Duijndam, 1988a,b; Malinverno, 2000; Ulrych et al., 2001; Scales and Tenorio, 2001). The prior knowledge about the model parameters is specified by probability density functions where the prior belief and the corresponding uncertainty is defined. The relationship between the model parameters and the measured data is described by the likelihood model. The solution of a Bayesian inverse problem is represented by the posterior distribution, which provides both the most probable solution and information about the corresponding uncertainty.

Amplitude versus offset (AVO) inversion can be used to extract information about the elastic subsurface parameters utilizing the angle dependency of the reflection coefficient (Smith and Gidlow, 1987; Hampson and Russell, 1990; Buland et al., 1996; Gouveia and Scales, 1998). The AVO inversion problem can be linearized if an appropriate processing sequence is applied to the seismic data prior to the inversion. Important elements in such processing are the removal of the moveout, multiples, and the effect of geometrical spreading and absorption. The seismic data should be prestack migrated and transformed from offset to reflection angle. A Bayesian linearized AVO inversion is defined in Buland and Omre (2002a), where an explicit analytical form of the posterior distribution is derived under Gaussian model assumptions. The explicit analytical form of the posterior distribution provides a computationally fast inversion method suitable for inversion of large 3-D seismic datasets. However,

the elastic parameters are not laterally coupled, so the inversion is performed independently for each bin gather. Further, the seismic wavelet and the noise covariance are assumed to be known prior to the inversion. A Bayesian method for estimation of the wavelet and the noise covariance is presented in Buland and Omre (2002b).

In this paper, a more realistic and complex statistical model is defined for the linearized AVO inversion problem. Firstly, a spatial coupling of the model parameters is imposed by a spatial correlation function. This ensures a spatially consistent solution. Secondly, the solution is obtained both from seismic prestack data and well logs. The spatial coupling of the model parameters is required for the integration of seismic and well log data. Thirdly, the AVO inversion, the wavelet estimation, and the estimation of the noise level are done simultaneously. The uncertainty of the wavelet and the noise level, and the corresponding effect on the estimated elastic parameters are integrated in the algorithm. A general hierarchical Bayesian model is defined. That means that also the statistical parameters specifying probability density functions are regarded as uncertain, for example the mean vector and the covariance matrix for a Gaussian variable. These statistical parameters are assigned prior distributions specified by hyperparameters. For trivial Bayesian problems, analytical expressions for the posterior distributions can often be found. In the current case, no analytical solution exists, but the posterior distribution can be explored by Markov chain Monte Carlo (MCMC) simulation (Gilks et al., 1996; Chen et al., 2000). The methodology is presented in the following sections, firstly in rather general terms, and then illustrated by an inversion example of a real dataset from the Heidrun Field.

4.2 Methodology

An isotropic, elastic medium is completely described by the material parameters $\{\alpha(\boldsymbol{x},t),\beta(\boldsymbol{x},t),\rho(\boldsymbol{x},t)\}$, where α, β, and ρ are P-wave velocity, S-wave velocity, and density, $\boldsymbol{x}$ is the lateral location, and t is the two-way vertical seismic traveltime. A weak contrast reflectivity function for PP reflections is (Aki and Richards, 1980; Stolt and Weglein, 1985)

$$\begin{aligned} c(\boldsymbol{x},t,\theta) = a_\alpha(\boldsymbol{x},t,\theta)\,\frac{\partial}{\partial t}\ln\alpha(\boldsymbol{x},t) \\ +a_\beta(\boldsymbol{x},t,\theta)\,\frac{\partial}{\partial t}\ln\beta(\boldsymbol{x},t) + a_\rho(\boldsymbol{x},t,\theta)\,\frac{\partial}{\partial t}\ln\rho(\boldsymbol{x},t), \end{aligned} \tag{4.1}$$

where θ is the reflection angle, $a_\alpha = (1 + \tan^2\theta)/2$, $a_\beta = -4(\beta/\alpha)^2 \sin^2\theta$, and $a_\rho = (1 - 4(\beta/\alpha)^2 \sin^2\theta)/2$. The inversion algorithm requires that a_α, a_β, and a_ρ are defined from a prior known background model. Motivated by the form of the reflectivity function in expression (4.1), let the unknown model parameter vector be

$$\boldsymbol{m}(\boldsymbol{x}, t) = [\ln \alpha(\boldsymbol{x}, t), \ln \beta(\boldsymbol{x}, t), \ln \rho(\boldsymbol{x}, t)]^T, \tag{4.2}$$

where T denotes transpose.

The seismic data are represented by the convolutional model

$$d_{obs}(\boldsymbol{x}, t, \theta) = \int s(\tau, \theta)\, c(\boldsymbol{x}, t - \tau, \theta)\, d\tau + e_d(\boldsymbol{x}, t, \theta), \tag{4.3}$$

where s is the wavelet, and e_d is an error term. The wavelet is allowed to be angle dependent, but should be independent of the lateral location $\boldsymbol{x}$. The wavelet is assumed to be stationary within a limited target window.

The seismic data $d_{obs}(\boldsymbol{x}, t, \theta)$ are available on a discrete form, denoted $\boldsymbol{d}_{obs}$. In general, the discretization of the subsurface parameters $\boldsymbol{m}(\boldsymbol{x}, t)$ should be determined by the intrinsic variability and future use of the inverted parameters. An identical lateral and temporal sampling of the model parameters and the seismic data is often chosen. Usually, this is an adequate choice, but the methodology is not restricted to equal sampling. Let a discrete representation of $\boldsymbol{m}(\boldsymbol{x}, t)$ be written $\boldsymbol{m}$. Further, let $\boldsymbol{s}$ be a discrete representation of the seismic wavelet and let $\boldsymbol{w}_{obs}$ be a discrete representation of the well log data.

The stochastic dependency model

The variables and the problem structure are graphically displayed by a directed acyclic graph (DAG) in Figure 4.1, see for example Spiegelhalter et al. (1996); Lauritzen (1996). All nodes in the graph represent stochastic variables, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote expectation vectors and covariance matrices. The arrows represent the causal dependency structure, where thin lines indicate probability dependencies and thick lines represent deterministic relationships. The DAG can be interpreted as a family tree where the parent nodes point to its children. When probabilistic dependencies are considered, nodes connected by a deterministic link merge into a single node. The expectation vectors $\boldsymbol{\mu}_w$ and $\boldsymbol{\mu}_d$ are deterministically determined from $\boldsymbol{m}$ and from $\boldsymbol{m}$ and $\boldsymbol{s}$, respectively. The parents of the well log data are therefore $\mathrm{Pa}(\boldsymbol{w}_{obs}) = \{\boldsymbol{m}, \boldsymbol{\Sigma}_w\}$, and the

parents of the seismic data are $\mathrm{Pa}(\boldsymbol{d}_{obs}) = \{\boldsymbol{m}, \boldsymbol{s}, \boldsymbol{\Sigma}_d\}$, where the notation $\mathrm{Pa}(v)$ denotes the parents of a node v. Before any data are observed, the marginal distributions of the parents are independent, but when their child is observed, the properties of the parents become dependent. An example is $\boldsymbol{m}$ and $\boldsymbol{s}$ which are apriori independent, but become dependent when the observed seismic data $\boldsymbol{d}_{obs}$ are included in the model.

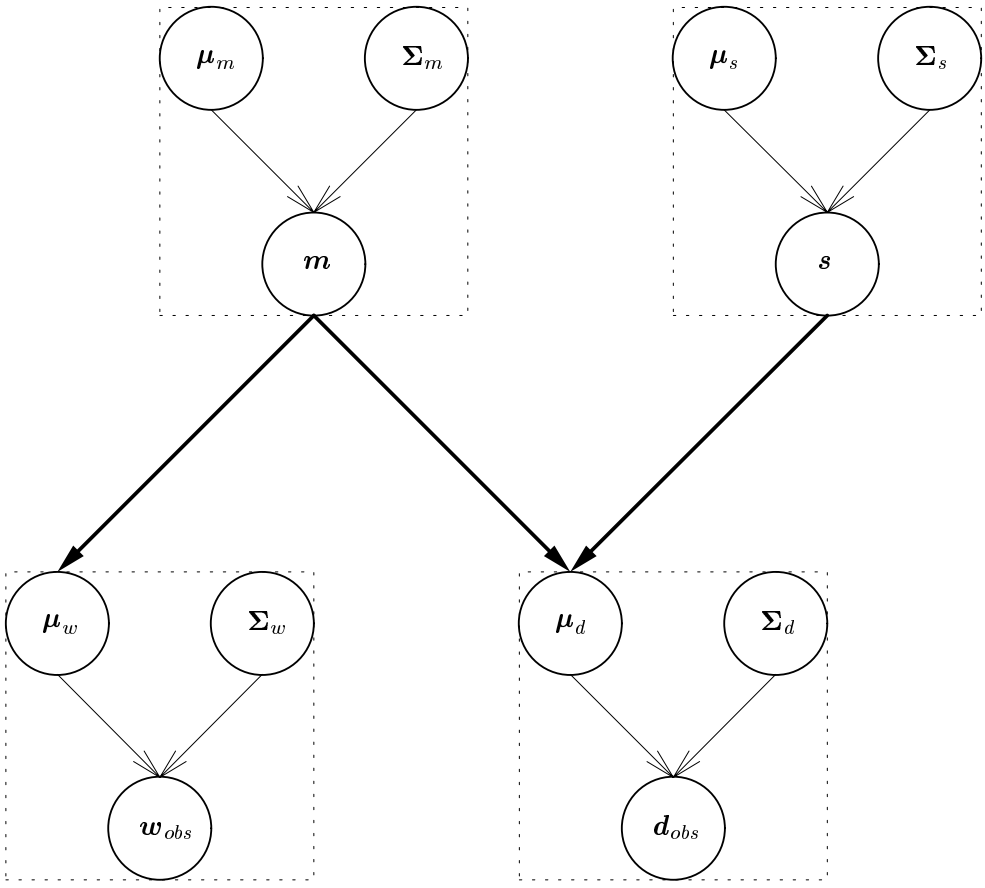


Figure 4.1: The stochastic model represented by a directed acyclic graph. The nodes represent stochastic variables, the thin arrows represent probability dependencies, and the thick arrows represent deterministic relationships.

The likelihood model

The error term in the convolutional model in expression (4.3) is assumed to be zero mean Gaussian with covariance $\boldsymbol{\Sigma}_d$. The likelihood model for the seismic data $\boldsymbol{d}_{obs}$ is then Gaussian, compactly denoted

$$\boldsymbol{d}_{obs}|\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d \sim \mathcal{N}_{n_d}(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d), \tag{4.4}$$

where n_d is the dimension of the seismic data vector $\boldsymbol{d}_{obs}$. A definition of the multi-Gaussian distribution is given in Appendix 4.A. The expectation vector $\boldsymbol{\mu}_d$ is deterministically determined by the convolution of the reflection coefficients computed from $\boldsymbol{m}$ with the wavelet $\boldsymbol{s}$.

The relation between the well observations $\boldsymbol{w}_{obs}$ and the total model parameter vector $\boldsymbol{m}$ can be written

$$\boldsymbol{w}_{obs} = \boldsymbol{P}\boldsymbol{m} + \boldsymbol{e}_w, \tag{4.5}$$

where $\boldsymbol{P}$ is a design matrix of zeros and ones which defines the locations of the wells relative to $\boldsymbol{m}$, and $\boldsymbol{e}_w$ is an error term. If we assume that the well log error $\boldsymbol{e}_w$ is zero mean Gaussian with covariance $\boldsymbol{\Sigma}_w$, then the likelihood model for the well log information is Gaussian

$$\boldsymbol{w}_{obs}|\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w \sim \mathcal{N}_{n_w}(\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w), \tag{4.6}$$

where n_w is the dimension of $\boldsymbol{w}_{obs}$ and $\boldsymbol{\mu}_w = \boldsymbol{P}\boldsymbol{m}$. The well locations relative to the seismic data are assumed to be correct. Also the conversion of the well logs from depth to seismic traveltime is assumed to be correct. In principle, this uncertainty related to the depth to time transform of the well logs could have been included in the present method by simulation of possible shift, stretch and squeeze of the log time axis (Buland and Omre, 2002b).

The prior model

The model parameters $\boldsymbol{m}$ are apriori assumed to be Gaussian

$$\boldsymbol{m}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \tag{4.7}$$

where n_m is the dimension of $\boldsymbol{m}$, and $\boldsymbol{\mu}_m$ and $\boldsymbol{\Sigma}_m$ are the expectation vector and covariance matrix, respectively. The expectation is usually constant or slowly varying for each of the parameters $\ln\alpha$, $\ln\beta$, and $\ln\rho$ in $\boldsymbol{m}$. The covariance defines the variances and the correlations between the different elements

in $\boldsymbol{m}$. A spatial coupling of the parameters is imposed through the covariance matrix by a spatial correlation function.

Also the wavelet $\boldsymbol{s}$ is apriori modeled with a Gaussian distribution,

$$\boldsymbol{s}|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s), \tag{4.8}$$

where n_s is the dimension of $\boldsymbol{s}$, and $\boldsymbol{\mu}_s$ and $\boldsymbol{\Sigma}_s$ are the expectation vector and the covariance matrix, respectively. The length of the wavelet is fixed. In general, the length of the wavelet should be as short as possible, but still long enough to represent the link between the model parameters and the seismic data as defined by the convolutional model. Expected smoothness of the wavelet can be imposed through the covariance matrix by a temporal correlation function.

The DAG in Figure 4.1 represents a hierarchical Bayesian model, where the expectations and covariances are stochastic. The prior expressions for $\boldsymbol{m}$ and $\boldsymbol{s}$ in expressions (4.7) and (4.8) can be regarded as the first level of an hierarchical prior model, the structural portion, while the prior distributions for the expectation vectors and covariance matrices form the second level, the subjective portion of the prior, see Robert (1994); Carlin (1996).

A mathematically convenient class of prior distributions for the second level is the conjugate prior distributions (Robert, 1994). These distributions have the same parametric prior and posterior forms, but with different parameters. The conjugate distributions for expectation and variance in a Gaussian model are the Gaussian and the inverse gamma distributions, respectively. The uncertainties of $\boldsymbol{\mu}_m$ and $\boldsymbol{\mu}_s$ are therefore modeled by Gaussian distributions,

$$\boldsymbol{\mu}_m \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}_{\mu_m}, \boldsymbol{\Sigma}_{\mu_m}), \tag{4.9}$$

$$\boldsymbol{\mu}_s \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_{\mu_s}, \boldsymbol{\Sigma}_{\mu_s}), \tag{4.10}$$

where the expectations $\boldsymbol{\mu}_{\mu_m}$ and $\boldsymbol{\mu}_{\mu_s}$, and the covariances $\boldsymbol{\Sigma}_{\mu_m}$ and $\boldsymbol{\Sigma}_{\mu_s}$ are fixed hyperparameters. The expectations $\boldsymbol{\mu}_w$ and $\boldsymbol{\mu}_d$ are deterministically determined and do not need prior distributions.

If the covariance matrices $\boldsymbol{\Sigma}_m$, $\boldsymbol{\Sigma}_s$, $\boldsymbol{\Sigma}_w$, and $\boldsymbol{\Sigma}_d$ are known up to unknown multiplicative variance factors, the covariances can be written

$$\boldsymbol{\Sigma}_i = \sigma_i^2 \boldsymbol{\Sigma}_{0,i}, \tag{4.11}$$

where i is one of $\{m, s, w, d\}$. The uncertainty of the unknown variance factor

σ_i^2 is modeled by the inverse gamma distribution,

$$\sigma_i^2 \sim \mathcal{IG}(\gamma_i, \lambda_i), \tag{4.12}$$

where the hyperparameters γ_i and λ_i define the prior expectation and variance, see Appendix 4.B. If a covariance matrix is completely unknown, a conjugate prior distribution can be obtained using the inverted Wishart distribution. Then the conditional distribution given the corresponding sample covariance matrix is also inverted Wishart distributed, see Appendix 4.C.

The posterior model

The objective of the inversion is to obtain the posterior distributions for for the involved variables based on the observed seismic data $\boldsymbol{d}_{obs}$ and the well logs $\boldsymbol{w}_{obs}$. Let the vector $\boldsymbol{v}$ represent the unknown quantities, that is all variables in the DAG in Figure 4.1 except $\boldsymbol{d}_{obs}$ and $\boldsymbol{w}_{obs}$. Further, let $\boldsymbol{o}$ be a vector containing the observed data,

$$\boldsymbol{o} = \begin{bmatrix} \boldsymbol{w}_{obs} \\ \boldsymbol{d}_{obs} \end{bmatrix}. \tag{4.13}$$

The posterior distribution for $\boldsymbol{v}$ given $\boldsymbol{d}_{obs}$ and $\boldsymbol{w}_{obs}$ can be written

$$p(\boldsymbol{v}|\boldsymbol{o}) = \frac{p(\boldsymbol{v}, \boldsymbol{o})}{p(\boldsymbol{o})} \tag{4.14}$$

where $p(\boldsymbol{v}, \boldsymbol{o})$ is the complete joint distribution. The pdf $p(\boldsymbol{o})$ does not contain unknown variables, and can therefore be regarded as a constant. The complete joint distribution for a model represented by a DAG can generally be written (Spiegelhalter et al., 1996)

$$p(\boldsymbol{v}, \boldsymbol{o}) = \prod_{x_i \in \{\boldsymbol{v}, \boldsymbol{o}\}} p(x_i|\mathrm{Pa}(x_i)). \tag{4.15}$$

An expression for the posterior distribution is now defined from expressions (4.14) and (4.15), and the DAG in Figure 4.1,

$$\begin{aligned} p(\boldsymbol{m}, \boldsymbol{s}, \boldsymbol{\mu}_m, \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_m, \boldsymbol{\Sigma}_s, \boldsymbol{\Sigma}_w, \boldsymbol{\Sigma}_d|\boldsymbol{d}_{obs}, \boldsymbol{w}_{obs}) \propto \\ p(\boldsymbol{d}_{obs}|\boldsymbol{m}, \boldsymbol{s}, \boldsymbol{\Sigma}_d)\, p(\boldsymbol{w}_{obs}|\boldsymbol{m}, \boldsymbol{\Sigma}_w)\, p(\boldsymbol{m}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \\ p(\boldsymbol{s}|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)\, p(\boldsymbol{\mu}_m)\, p(\boldsymbol{\mu}_s)\, p(\boldsymbol{\Sigma}_m)\, p(\boldsymbol{\Sigma}_s)\, p(\boldsymbol{\Sigma}_w)\, p(\boldsymbol{\Sigma}_d), \end{aligned} \tag{4.16}$$

where the proportionality is caused by the unknown constant probability density functions involving $\boldsymbol{d}_{obs}$ and $\boldsymbol{w}_{obs}$.

The Gibbs sampler algorithm

The posterior distribution can be explored by MCMC simulation. The Gibbs sampler is perhaps the best known and most popular of the MCMC algorithms. The name of the algorithm was introduced by Geman and Geman (1984) who worked with Gibbs distributions on lattices, but the name is misleading as the application of the algorithm is general, and not restricted to Gibbs distributions. A general presentation of MCMC can be found in Gilks et al. (1996); Chen et al. (2000).

The Gibbs sampler algorithm works on the complete joint distribution and requires full conditional distributions for each single variable given all the others. One iteration consists of drawing new samples for the unknown quantities. A new sample is drawn conditioned on the current state of the other elements in $\boldsymbol{v}$. Once an element is drawn, it goes into the current state of $\boldsymbol{v}$.

The pseudo-code of the Gibbs sampling algorithm can be written

Initiate: Set arbitrary $\boldsymbol{v}^{(0)}$ where $p(\boldsymbol{v}^{(0)}|\boldsymbol{o}) > 0$.

Iterate: For $i = 1, 2, \ldots$, draw

$$\begin{aligned} v_1^{(i)} &\sim p(v_1|\boldsymbol{v}_{-1}, \boldsymbol{o}), \\ &\vdots \\ v_j^{(i)} &\sim p(v_j|\boldsymbol{v}_{-j}, \boldsymbol{o}), \\ &\vdots \\ v_n^{(i)} &\sim p(v_{n_v}|\boldsymbol{v}_{-n}, \boldsymbol{o}), \end{aligned}$$

where $\boldsymbol{v}_{-j} = \{v_1^{(i)}, \ldots, v_{j-1}^{(i)}, v_{j+1}^{(i-1)}, \ldots, v_{n_v}^{(i-1)}\}$. Instead of drawing single elements from $\boldsymbol{v}$ as shown above, a slightly more general algorithm is obtained by allowing for simultaneously drawing of a group of elements. An example is to draw the complete wavelet $\boldsymbol{s}$ instead of updating single samples of the wavelet. The elements in the model parameter vector $\boldsymbol{m}$ can also be grouped, for example into the complete vector $\boldsymbol{m}$, or some partition of $\boldsymbol{m}$. The full conditional distributions needed in the Gibbs sampling are specified in Appendix 4.D.

4.3 Inversion example of Heidrun data

The Heidrun Field is an oil and gas field offshore Mid-Norway. Geologically, the field comprises a heavily faulted and eroded horst block of early to mid-Jurassic clastic deposits. The reservoirs are located in the shallow, clean Fangst Group sandstones, and in the deeper and more heterogeneous sandstones of the Tilje and Åre Formations.

A seismic line passing through two wells, well A and B, is used in this inversion example. The seismic data consist of $n_\theta = 3$ angle stacks with average angles 11°, 22°, and 33°, see Figure 4.2. Well A is located at CDP 1191 and well B at CDP 1227. An interpretation of top reservoir (Fangst Group) is shown by a black line. A 300 ms time window is inverted, where the top and bottom are parallel to the interpreted top reservoir horizon. Typical seismic peak amplitudes are in the order of 0.1.

The prior model

Among the hierarchical parameters, we consider the unknown seismic error covariance matrix $\boldsymbol{\Sigma}_d$ to be the most interesting. The estimation of both the wavelet $\boldsymbol{s}$ and the model parameter vector $\boldsymbol{m}$ with corresponding uncertainties depend on $\boldsymbol{\Sigma}_d$. Unfortunately, the estimation of $\boldsymbol{\Sigma}_d$ is usually not a trivial problem. The dominating noise in processed seismic data is usually source generated noise, for example remaining multiples, or processing artifacts. Such noise components usually have a smooth waveform similar to the waveform of the primary events. However, if well logs are available, the misfit between the synthetic seismic well log response and the real seismic data can be used to estimate $\boldsymbol{\Sigma}_d$, or a set of parameters which determines the complete $\boldsymbol{\Sigma}_d$. Three different seismic noise models are defined below, numbered from 1 to 3. The covariance matrices are here defined for a single CDP gather with $n_\theta = 3$ traces, and we assume no correlation between different CDP positions. In contrast, $\boldsymbol{\Sigma}_d$ in expression (4.4) is the covariance for the complete seismic dataset. Accordingly, n_d is here reduced to be the dimension of a CDP gather.

Noise model 1 is a simple white noise model with unknown variance,

$$\boldsymbol{\Sigma}_{d,1} = \sigma_d^2 \boldsymbol{I}_{n_d}, \tag{4.17}$$

where $\boldsymbol{I}_{n_d}$ is an $n_d \times n_d$ identity matrix. Our apriori knowledge about the

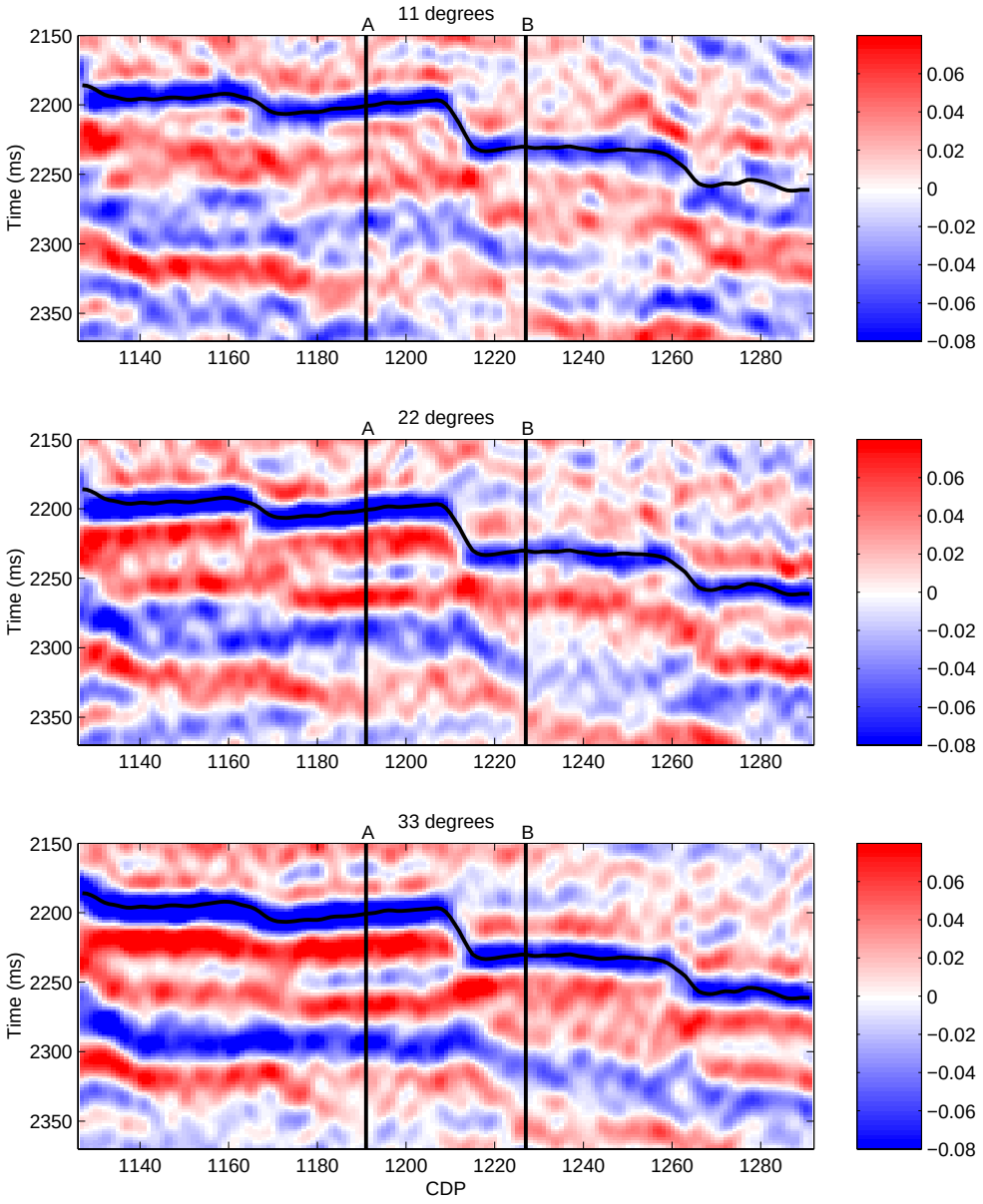


Figure 4.2: The seismic data represented by angle stacks, 11° (top), 22° (middle), and 33° (bottom).

seismic noise level is vague. The apriori uncertainty is described by the inverse gamma distribution

$$\sigma_d^2 \sim \mathcal{IG}(2, 0.0005), \tag{4.18}$$

where the apriori expectation of σ_d^2 is 0.0005, and the variance is undefined (infinite), see Appendix 4.B.

Noise model 2 has colored covariance with fixed correlation structure. The expected temporal smoothness of the seismic error is modeled by a second order exponential correlation function with range 14 ms. In addition, 10% white noise is added. Let the corresponding discrete correlation matrix for a seismic error trace be denoted $\boldsymbol{\Upsilon}_d$. A block-diagonal colored covariance matrix is defined by the Kronecker product

$$\boldsymbol{\Sigma}_{d,2} = \sigma_d^2 \boldsymbol{I}_{n_\theta} \otimes \boldsymbol{\Upsilon}_d, \tag{4.19}$$

where each of the elements in the 3×3 identity matrix $\boldsymbol{I}_{n_\theta}$ is multiplied with the temporal $n_t \times n_t$ correlation matrix $\boldsymbol{\Upsilon}_d$, with $n_t = 151$ being the number of time samples. This means that we assume that the error variance is equal for the three angle stacks, and that there is no correlation between them. The unknown variance σ_d^2 is modeled by the inverse gamma distribution as in expression (4.18). Note that noise model 1 can be written on a similar form, $\boldsymbol{\Sigma}_{d,1} = \sigma_d^2 \boldsymbol{I}_{n_\theta} \otimes \boldsymbol{I}_{n_t}$.

Noise model 3 has colored covariance with a separate unknown variance factor for each angle stack and unknown correlation coefficients between the angles. The covariance matrix is obtained by substituting the diagonal matrix $\sigma_d^2 \boldsymbol{I}_{n_\theta}$ in expression (4.19) with an unknown $n_\theta \times n_\theta$ covariance matrix $\boldsymbol{\Sigma}_\theta$, such that

$$\boldsymbol{\Sigma}_{d,3} = \boldsymbol{\Sigma}_\theta \otimes \boldsymbol{\Upsilon}_d. \tag{4.20}$$

The covariance matrix $\boldsymbol{\Sigma}_\theta$ is apriori assigned an inverted Wishart distribution with 5 degrees of freedom,

$$\boldsymbol{\Sigma}_\theta \sim \mathcal{IW}_{n_\theta}(0.0001 \boldsymbol{I}_{n_\theta}, 5), \tag{4.21}$$

see Appendix 4.C. With only 5 degrees of freedom, the prior distribution will only have a marginal effect on the conditional distribution for $\boldsymbol{\Sigma}_\theta$. Here, the relative influence of the prior distribution versus the sample covariance is about 1 to 60. If the prior knowledge about $\boldsymbol{\Sigma}_\theta$ is more precise, the relative influence of the prior distribution can be increased by increasing the degree of freedom.

The wavelet is apriori assumed to be Gaussian with unknown expectation and covariance. The wavelet length is set to 80 ms, which is sufficient for a main center loop and two side loops. A separate wavelet is estimated for each of the three angle stacks, such that $n_s = 3 \times 41$ with 2 ms sampling. The expected temporal smoothness is imposed through a covariance matrix $\boldsymbol{\Sigma}_{0,s}$ by a second order exponential correlation function with range 14 ms. Note that this correlation function is identical to the correlation used to model the smoothness of the colored seismic noise. Further, it is imposed that the wavelets should smoothly approach zero towards the first and last wavelet samples. The wavelets are also assumed to vary smoothly as function of reflection angle. This is imposed through $\boldsymbol{\Sigma}_{0,s}$ by a second order exponential correlation function with range 30°. The prior distribution for the wavelet expectation is

$$\boldsymbol{\mu}_s \sim \mathcal{N}_{n_s}(\mathbf{0}, \boldsymbol{\Sigma}_{\mu_s}), \tag{4.22}$$

where the prior expectation is a zero vector, and the covariance $\boldsymbol{\Sigma}_{\mu_s}$ is a fixed hyperparameter. From the seismic amplitudes, see Figure 4.2, we expect the maximum wavelet amplitude to be in the order of 1, so $\boldsymbol{\Sigma}_{\mu_s} = 1^2\boldsymbol{\Sigma}_{0,s}$. The unknown wavelet covariance $\boldsymbol{\Sigma}_s$ is

$$\boldsymbol{\Sigma}_s = \sigma_s^2\boldsymbol{\Sigma}_{0,s}, \tag{4.23}$$

where the apriori model for the unknown variance is

$$\sigma_s^2 \sim \mathcal{IG}(2, 0.001), \tag{4.24}$$

with expectation 0.001, and undefined (infinite) variance. To illustrate the defined prior model, a set of wavelets simulated from the prior distribution is shown in Figure 4.3. Except the imposed smoothness and the amplitude decay towards -40 ms and 40 ms, the prior model is flexible with respect to the shape of the wavelets.

The prior model for $\boldsymbol{m}$ is modeled by a Gaussian distribution specified by the expectation $\boldsymbol{\mu}_m$ and the covariance $\boldsymbol{\Sigma}_m$. A thorough presentation of the prior model for $\boldsymbol{m}$ can be found in Buland and Omre (2002a). In this example, the prior model for $\boldsymbol{m}$ is determined from the well logs and then fixed. The expectation $\boldsymbol{\mu}_m$ can be considered to be a slowly varying background model for $\ln\alpha(\boldsymbol{x}, t)$, $\ln\beta(\boldsymbol{x}, t)$, and $\ln\rho(\boldsymbol{x}, t)$. The temporal variation of $\boldsymbol{\mu}_m$ should be smooth, but sufficiently high frequent to cover the lacking low frequencies in the band-limited seismic data. In this example, the prior expectation is vertically slowly varying but laterally constant, that means constant along

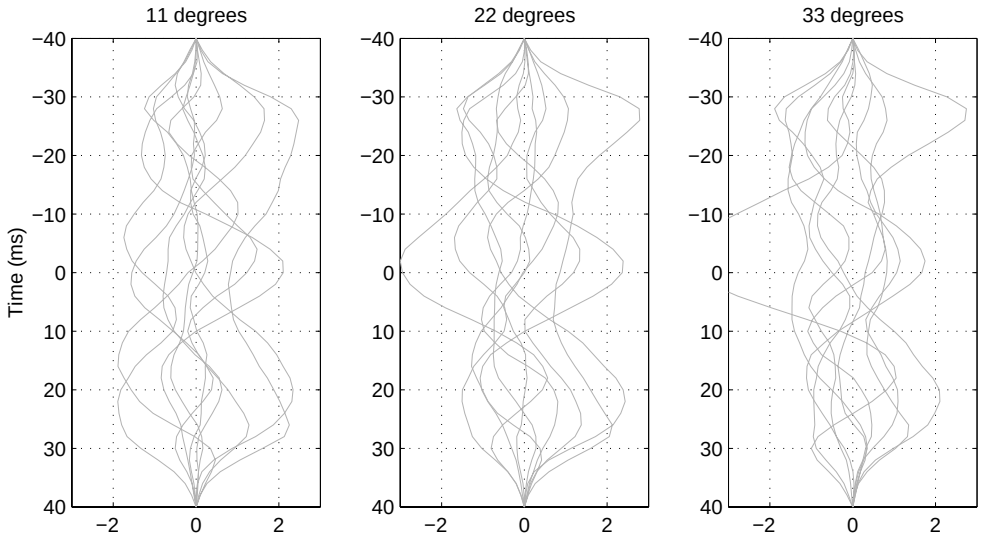


Figure 4.3: Wavelets simulated from the prior distribution.

structures parallel to the interpreted top reservoir horizon. The covariance function is stationary and homogeneous, and can be factorized as

$$\boldsymbol{\Sigma}_m(\boldsymbol{x}_1, t_1; \boldsymbol{x}_2, t_2) = \boldsymbol{\Sigma}_{0,m}\, \nu_m(\xi, \tau), \tag{4.25}$$

where $\nu_m(\xi, \tau)$ is a spatial correlation function, ξ is the distance between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, τ is the time lag between t_1 and t_2, and

$$\boldsymbol{\Sigma}_{0,m} = \begin{bmatrix} \sigma_\alpha^2 & \sigma_\alpha\sigma_\beta\nu_{\alpha\beta} & \sigma_\alpha\sigma_\rho\nu_{\alpha\rho} \\ \sigma_\alpha\sigma_\beta\nu_{\alpha\beta} & \sigma_\beta^2 & \sigma_\beta\sigma_\rho\nu_{\beta\rho} \\ \sigma_\alpha\sigma_\rho\nu_{\alpha\rho} & \sigma_\beta\sigma_\rho\nu_{\beta\rho} & \sigma_\rho^2 \end{bmatrix}. \tag{4.26}$$

The diagonal elements of $\boldsymbol{\Sigma}_{0,m}$ are the variances, and $\nu_{\alpha\beta}$, $\nu_{\alpha\rho}$, and $\nu_{\beta\rho}$, are the correlations between $\ln\alpha(\boldsymbol{x}, t)$, $\ln\beta(\boldsymbol{x}, t)$ and $\ln\rho(\boldsymbol{x}, t)$, respectively. From the well logs, estimates of the elements in $\boldsymbol{\Sigma}_{0,m}$ are obtained by standard estimators. The estimated variances for $\ln\alpha$, $\ln\beta$ and $\ln\rho$ are $\widehat{\sigma_\alpha^2} = 0.0026$, $\widehat{\sigma_\beta^2} = 0.0034$, and $\widehat{\sigma_\rho^2} = 0.0020$, and the estimated correlation coefficients are $\widehat{\nu_{\alpha\beta}} = 0.76$, $\widehat{\nu_{\alpha\rho}} = 0.77$, $\widehat{\nu_{\beta\rho}} = 0.81$. The vertical variability is modeled by a sum of an exponential second order correlation function with range 2 ms, and a normalized second derivative of an exponential second order correlation function with range 10 ms (Buland and Omre, 2002a). The lateral coupling of the model parameters is imposed by a first order exponential correlation function with range 125 m, that is 10 CDP positions.

In the well positions, the well logs represent measurements of the model parameters $\boldsymbol{m}$. The covariance of the possible well log errors is set to

$$\boldsymbol{\Sigma}_w = \sigma_w^2 \boldsymbol{I}, \tag{4.27}$$

with $\sigma_w^2 = 0.0001$, such that the variance is fixed and identical for the $\ln \alpha$, $\ln \beta$, and the $\ln \rho$ logs. A noninformative prior distribution should not be specified for σ_w^2 since the seismic error term in the convolutional model is not completely separable from errors corresponding to the well log errors. The prior model for σ_w^2 should therefore be specified exactly, or at least relatively precise (Buland and Omre, 2002b).

Simulation results

The posterior distribution is explored by 500 iterations with the Gibbs sampler algorithm. MCMC algorithms generally need some initial iterations before convergence to the posterior distribution is reached. The convergence can be evaluated by monitoring the involved variables using different initial values. In this problem, the convergence is fast, only some few burnin iterations are needed. To be sure, the first 20 iterations are considered to be burnin iterations, and these are not used in the calculations. For noise model 1 in expression (4.17), the simulated seismic error variance is shown as function of iteration number in Figure 4.4. The initial variance was 0.0001, but the convergence to the level between 0.0002 and 0.0003 is reached after the first iteration. Tests with other start values do also show good convergence properties. The posterior distribution of σ_d^2 is represented by a histogram of the simulated variance values in Figure 4.5. The posterior mean of the seismic white noise variance is estimated from the MCMC samples to 0.00024. For noise model 2 in expression (4.19), the posterior distribution for the variance factor is shown in Figure 4.6, with posterior mean about 0.0001. A lower variance factor is sufficient to represent the seismic error with the colored covariance compared to the white covariance. For noise model 3 in expression (4.20), the posterior mean for $\boldsymbol{\Sigma}_\theta$ is estimated to

$$\boldsymbol{\Sigma}_\theta = 10^{-4} \begin{bmatrix} 0.91 & 0.14 & -0.21 \\ 0.14 & 1.03 & 0.56 \\ -0.21 & 0.56 & 0.99 \end{bmatrix}. \tag{4.28}$$

The variance factors on the diagonal of $\boldsymbol{\Sigma}_\theta$ are about 0.0001 for the three angle stacks as for noise model 2. The correlations are 0.15 between the error traces

for 11° and 22°, 0.5 between 22° and 33°, and -0.2 between 11° and 33°. The posterior uncertainty of $\mathbf{\Sigma}_\theta$ represented by two standard deviations is

$$10^{-4} \begin{bmatrix} \pm 0.21 & \pm 0.16 & \pm 0.15 \\ \pm 0.16 & \pm 0.24 & \pm 0.21 \\ \pm 0.15 & \pm 0.21 & \pm 0.24 \end{bmatrix}. \tag{4.29}$$

The uncertainty of variance factors on the diagonal is about 23%, while the off diagonal covariance factors are relatively far more uncertain.

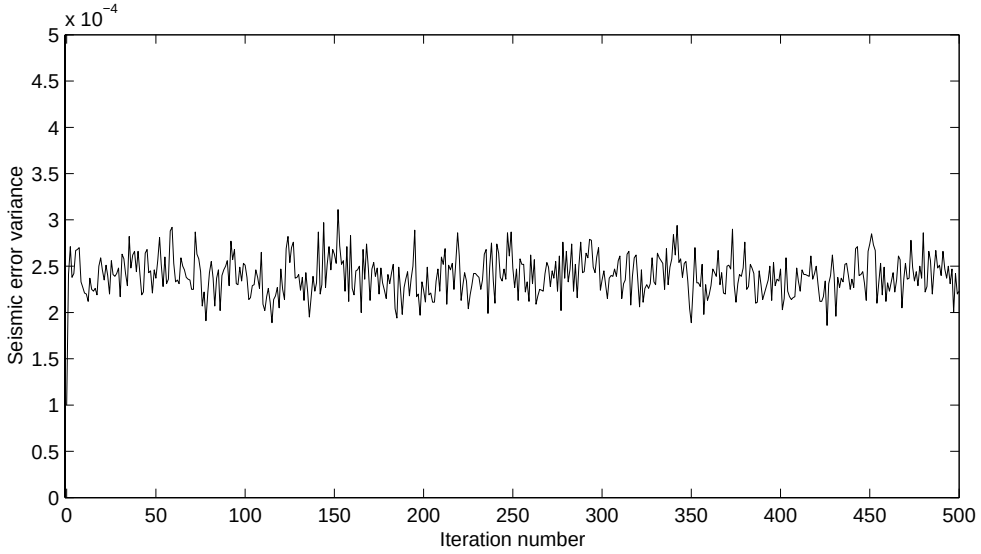


Figure 4.4: Monitor plot of the seismic white noise variance σ_d^2.

The simulated wavelets for the three different seismic noise models in expressions (4.17), (4.19), and (4.20) are plotted in Figures 4.7, 4.8, and 4.9 with gray lines. The gray clouds illustrate the wavelet uncertainty. The estimated posterior mean wavelets are shown with black lines. The differences between the wavelets for the three seismic noise models are marginal.

A simulated solution for the P-wave velocity, the S-wave velocity, and the density is shown in Figure 4.10 using seismic noise model 2 in expression (4.19). The simulated solution represents one possible laterally consistent solution with high vertical variability. In some cases, for example in fluid flow simulation, a set of simulated solutions with realistic variability is needed. Usually, however, it is more convenient to summarize the results in a best solution with corresponding uncertainty, for example the posterior mean or the maximum

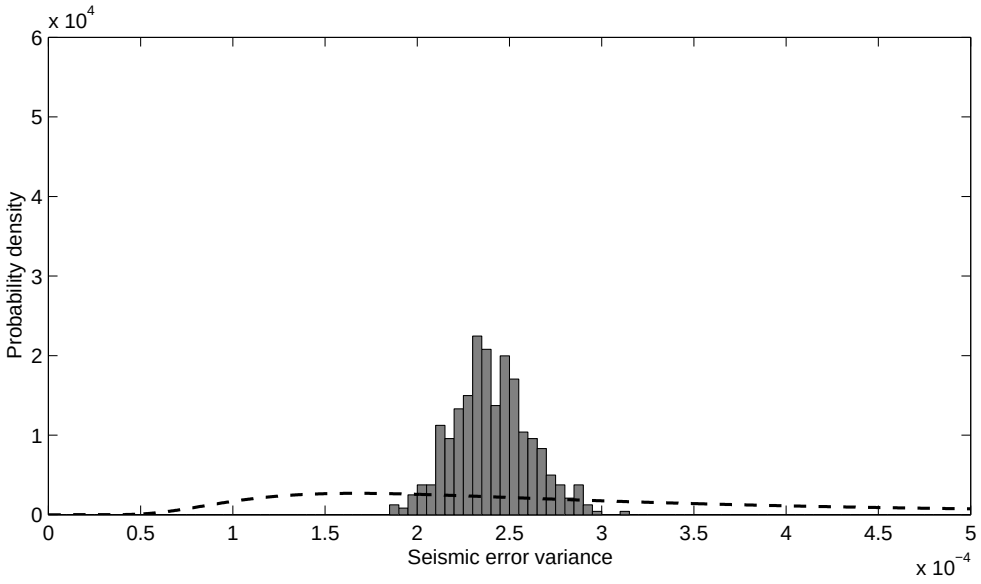


Figure 4.5: Posterior distribution for the seismic white noise variance σ_d^2. The apriori distribution is dotted.

posterior solution. The posterior mean solution obtained from the simulations is shown in Figure 4.11. The corresponding solutions using the two other seismic covariance models differ only marginally. Note that the solution has highest resolution near the two wells. The additional information obtained from the well logs reduces the uncertainty near the wells. The vertically averaged posterior uncertainty represented by two standard deviations is shown in Figures 4.12, 4.13, and 4.14 for each CDP position along the seismic line for the three different seismic noise models. In comparison, the average apriori uncertainty is 316 m/s for the P-wave velocity, 181 m/s for the S-wave velocity, and 200 kg/m^3 for the density. The average uncertainty at well B is higher compared to the uncertainty at well A since the well log data in B do not cover the complete inversion window. The jagged form of the curves is related to the approximate screening algorithm presented in Appendix 4.E. However, we consider the deviations from a smooth curve to be small compared to the total uncertainty. The posterior uncertainty of the estimated elastic parameters is about 10-15% higher for the white noise model than the two colored noise models.

If the wavelet and the noise covariance are estimated prior to the inversion,

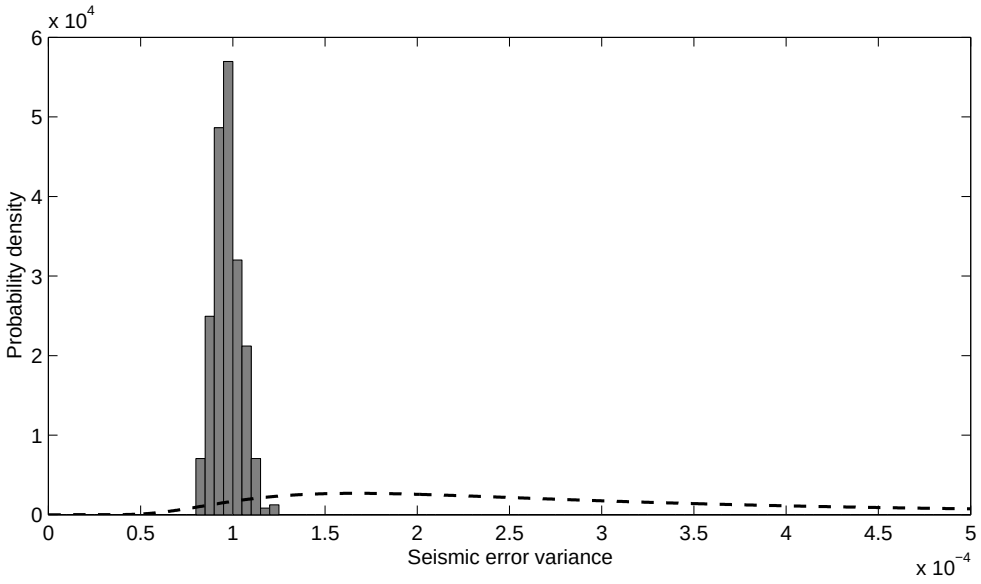


Figure 4.6: Posterior distribution for the seismic error variance σ_d^2 for colored noise. The apriori distribution is dotted.

and then plugged into the inversion as fixed quantities, the solution of the linearized AVO inversion problem is given on an explicit analytical form, and MCMC simulation is not required. In this paper, a stochastic model is defined which includes the uncertainty of $\boldsymbol{s}$ and $\boldsymbol{\Sigma}_d$. However, in the Heidrun example the effect of including the uncertainty of the wavelet and the noise covariance was marginal with respect to the estimated $\boldsymbol{m}$ and its uncertainty. With the peaked posterior distribution for the noise variance in Figure 4.6 and the low wavelet uncertainty in Figure 4.8, this is not a big surprise. In cases where the wavelet and the covariance uncertainty are larger, the corresponding effect on the estimated P-wave velocity, the S-wave velocity, and the density may be more distinct.

If the lateral coupling of the model is abandoned, the AVO inversion can not be conditioned to the well log data, except exactly in the well positions. The posterior mean solution with no lateral coupling and with fixed wavelet and noise covariance is shown in Figure 4.15, and the corresponding posterior uncertainty is shown in Figure 4.16. The differences between the MCMC solution in Figure 4.11 and the the analytically obtained solution in Figure 4.15 are minor, except near the wells. The uncertainty level is close to the maximum

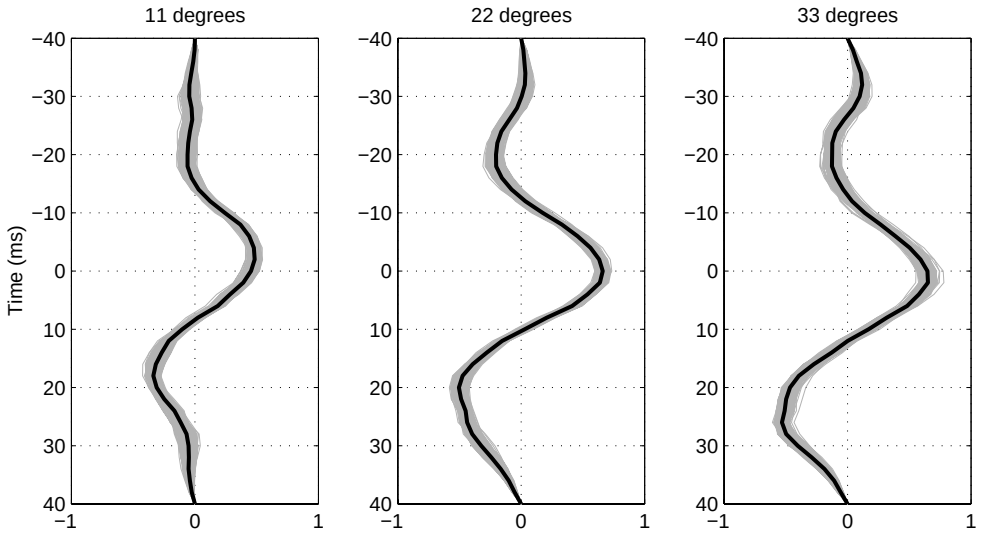


Figure 4.7: Simulated wavelets (gray lines), and the posterior mean wavelet (black line) for noise model 1.

uncertainty in Figure 4.13. The difference in computer time is dramatic, however. The MCMC solution shown in Figure 4.11 was obtained in about 12 hours, while the solution shown in Figure 4.15 was completed within some few seconds on a 400 MHz single CPU machine. The difference in the computer time is mainly caused by the 500 iterations of the Gibbs sampler algorithm, but also by the lateral coupling of the model parameters.

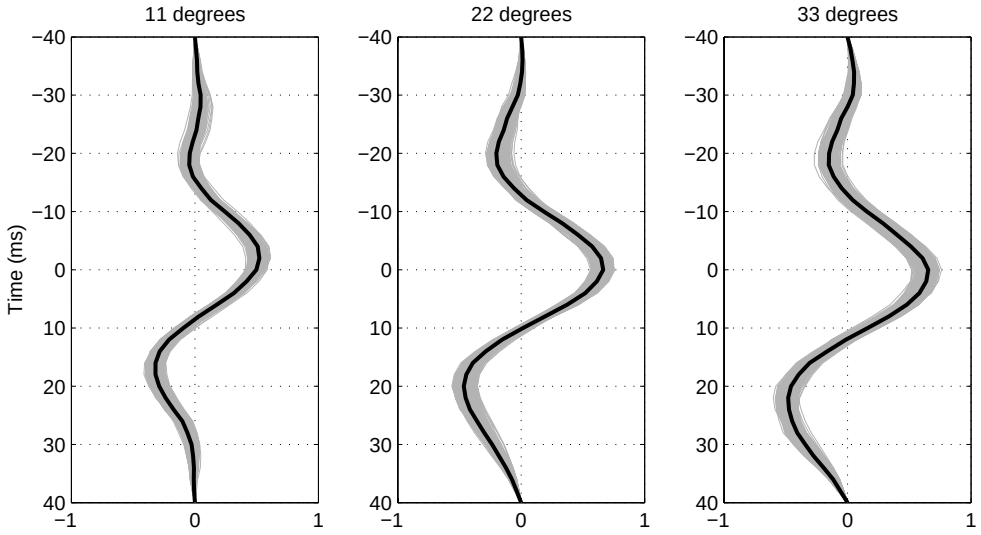


Figure 4.8: Simulated wavelets (gray lines), and the posterior mean wavelet (black line) for noise model 2.

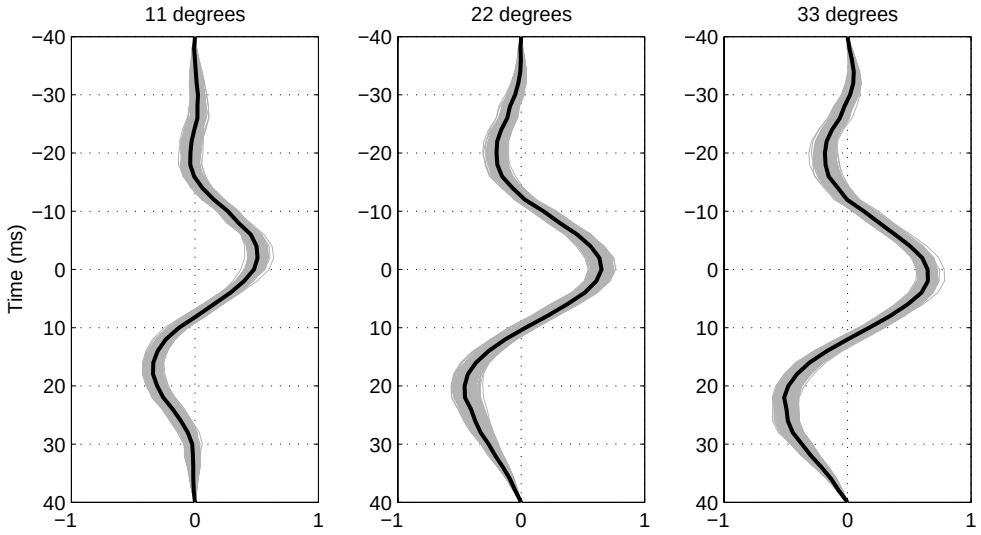


Figure 4.9: Simulated wavelets (gray lines), and the posterior mean wavelet (black line) for noise model 3.

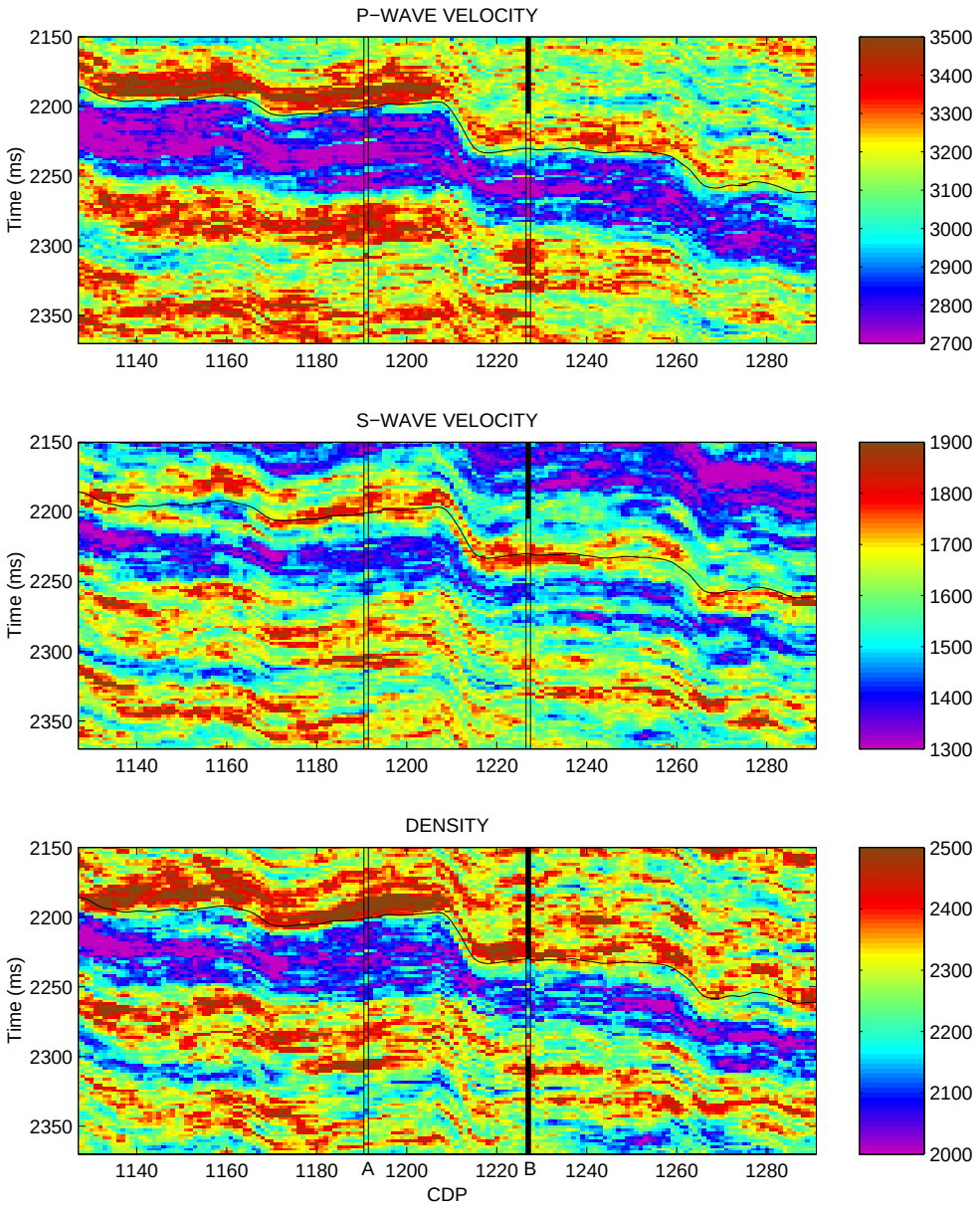


Figure 4.10: Simulated solution conditioned to seismic and well log data.

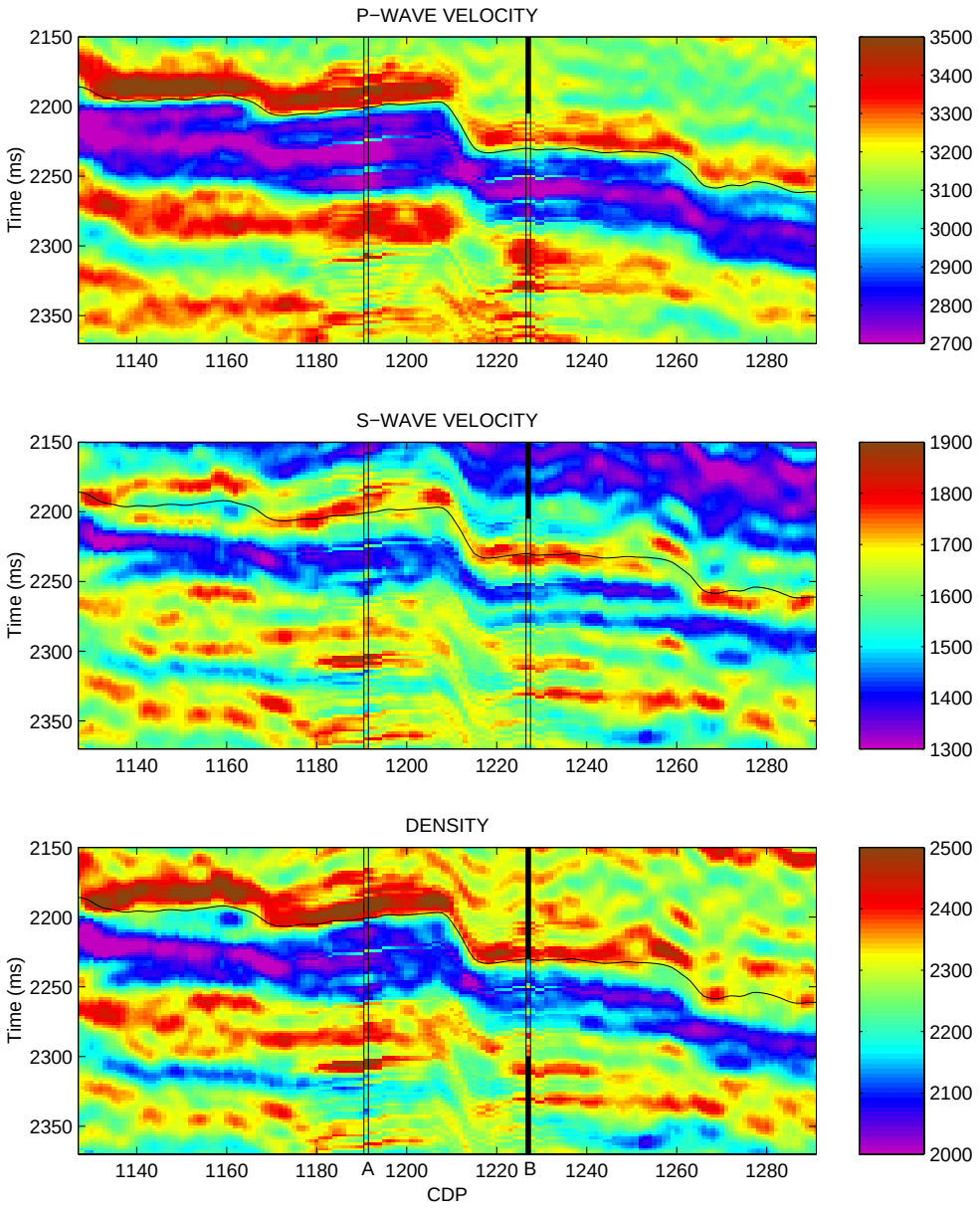


Figure 4.11: Posterior mean solution conditioned to seismic and well log data.

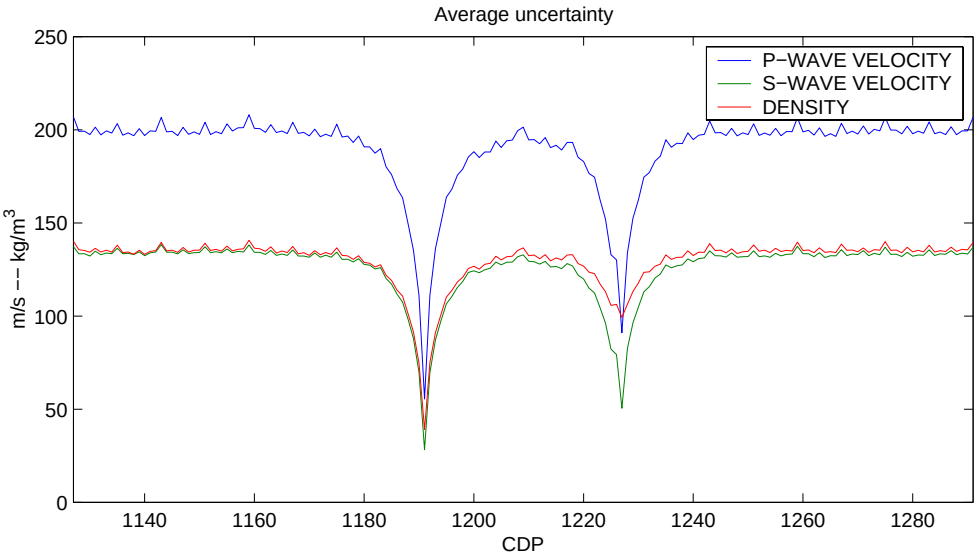


Figure 4.12: Average posterior uncertainty for P-wave velocity (blue), S-wave velocity (green), and density (red) for noise model 1.

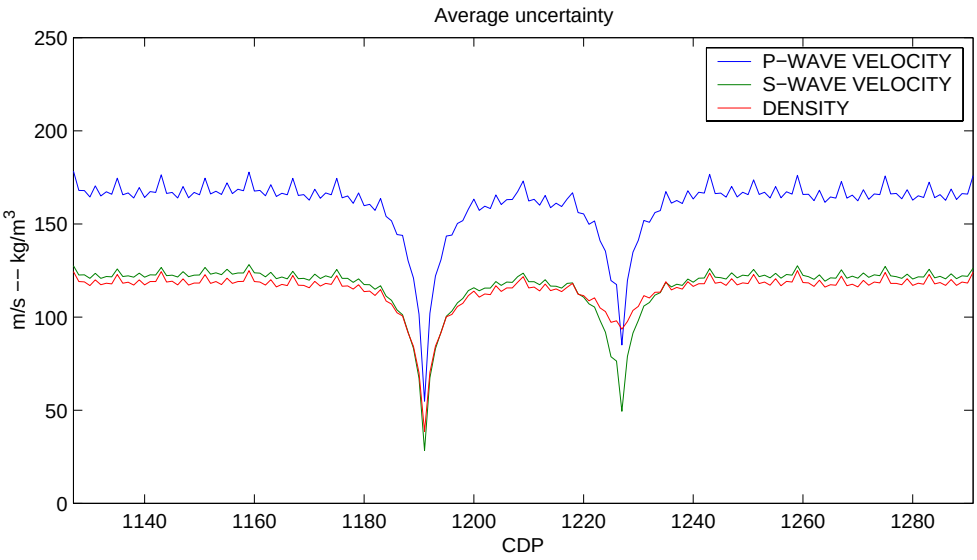


Figure 4.13: Average posterior uncertainty for P-wave velocity (blue), S-wave velocity (green), and density (red) for noise model 2.

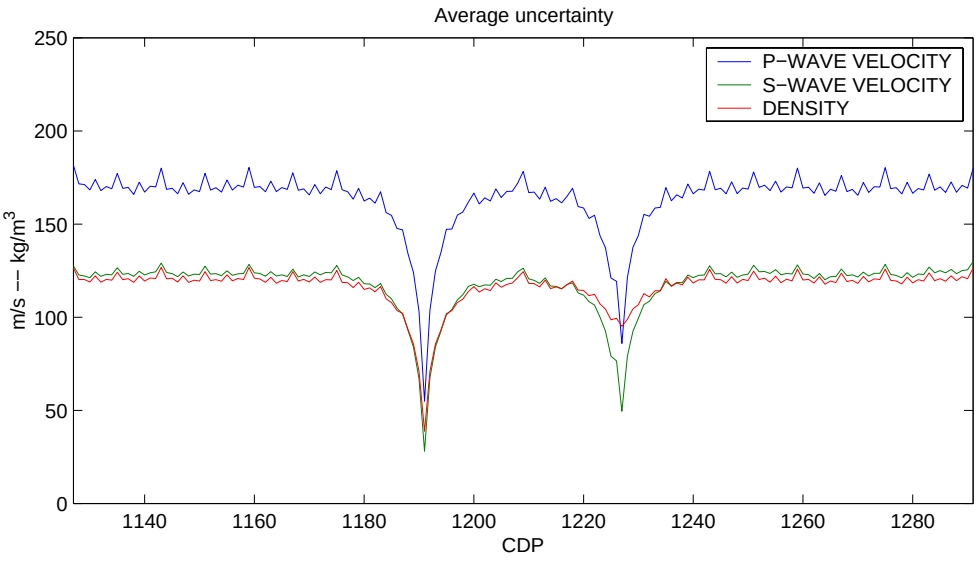


Figure 4.14: Average posterior uncertainty for P-wave velocity (blue), S-wave velocity (green), and density (red) for noise model 3.

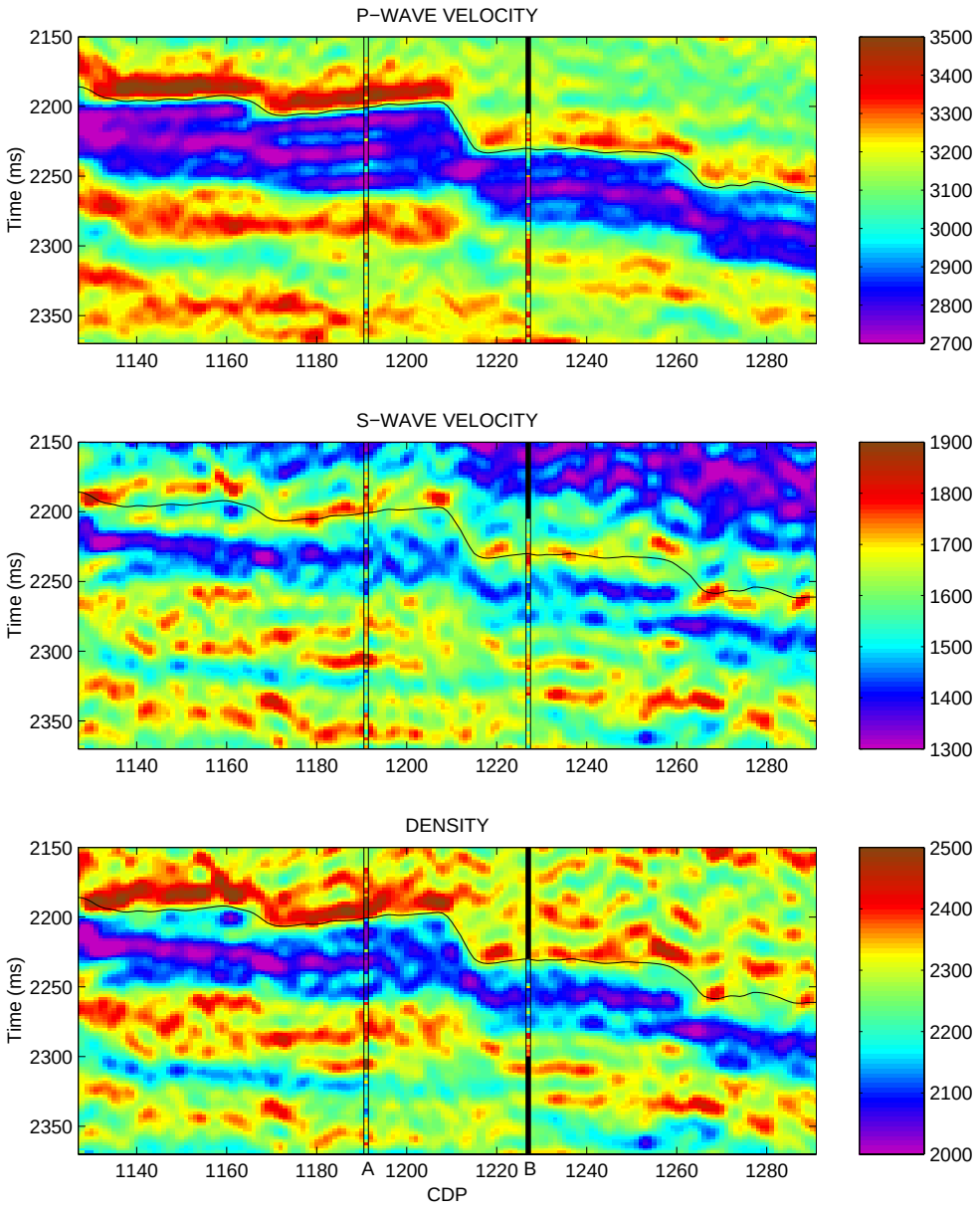


Figure 4.15: Posterior mean solution conditioned to seismic data.

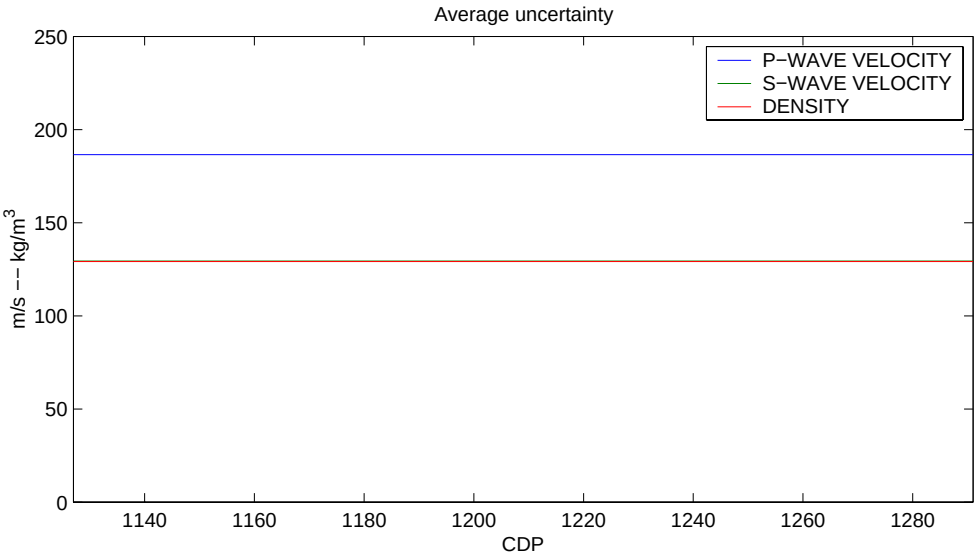


Figure 4.16: Average posterior uncertainty for P-wave velocity (blue), S-wave velocity (green), and density (red) using fixed wavelet and fixed colored noise covariance.

4.4 Conclusions

A spatially coupled AVO inversion method has been defined in a hierarchical Bayesian framework. The stochastic model includes uncertainty of both the elastic parameters, the wavelet, and the seismic and well log data. The seismic noise model is of special interest since both the AVO inversion and the wavelet estimation depend on the noise covariance, and since the estimation of a noise model may be highly uncertain itself. One white and two colored seismic noise models with stochastic noise levels are tested. On a real data example from the Heidrun Field, the use of colored seismic noise models gave better results with about 10% lower uncertainty than the white noise model. The uncertainty of the estimated wavelet was low for all three noise models. If the wavelet and the noise covariance are estimated prior to the AVO inversion and used as fixed nonstochastic quantities, the inversion problem has an explicit analytic solution. When the wavelet and the noise level are stochastic, the posterior solution can be obtained by Monte Carlo simulation. In the Heidrun example, the effect of including the uncertainty of the wavelet and the noise level was negligible with respect to the elastic parameters and the corresponding uncertainty.

4.5 Acknowledgments

We thank Statoil and the Heidrun license (Statoil, Conoco, and Fortum) for permission to publish this paper.

4.A Gaussian distribution

The multi-Gaussian probability density is (Anderson, 1984)

$$p(\boldsymbol{r}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{r}-\boldsymbol{\mu})\right], \tag{4.A.1}$$

where n is the dimension of $\boldsymbol{r}$, and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the expectation vector and the covariance matrix, respectively.

Consider two multivariate Gaussian variables $\boldsymbol{r}_1 \sim \mathcal{N}_{n_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\boldsymbol{r}_2 \sim$

$\mathcal{N}_{n_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ with joint distribution

$$\begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \end{bmatrix} \sim \mathcal{N}_{n_1+n_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right), \tag{4.A.2}$$

where n_1 and n_2 are the dimensions. Then the conditional distribution of $\boldsymbol{r}_1$ given $\boldsymbol{r}_2$ is Gaussian with expectation

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{r}_2 - \boldsymbol{\mu}_2) \tag{4.A.3}$$

and covariance

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{4.A.4}$$

4.B Gamma and inverse gamma distribution

The univariate gamma distribution $\mathcal{G}(\gamma, \lambda)$ is defined by the probability density function (Robert, 1994)

$$p(x|\gamma, \lambda) = \frac{\lambda^\gamma}{\Gamma(\gamma)} x^{\gamma-1} \exp[-\lambda x], \tag{4.B.1}$$

for $x \geq 0,\ \gamma > 0,\ \lambda > 0$. The expectation and variance are

$$E\{x|\gamma, \lambda\} = \frac{\gamma}{\lambda}, \tag{4.B.2}$$

$$\mathrm{Var}\{x|\gamma, \lambda\} = \frac{\gamma}{\lambda^2}. \tag{4.B.3}$$

Two special cases of the gamma distribution are the exponential distribution $\mathcal{E}(\lambda)$ for $\gamma = 1$, and the chi-squared distribution $\mathcal{X}^2_\nu$ for $\gamma = \nu/2$ and $\lambda = 1/2$.

If $y \sim \mathcal{G}(\gamma, \lambda)$, then $x = 1/y$ has the inverse gamma distribution $\mathcal{IG}(\gamma, \lambda)$ defined by the probability density function

$$p(x|\gamma, \lambda) = \frac{\lambda^\gamma}{\Gamma(\gamma)} \left(\frac{1}{x}\right)^{\gamma+1} \exp\left[-\frac{\lambda}{x}\right], \tag{4.B.4}$$

for $x \geq 0,\ \gamma > 0,\ \lambda > 0$. The expectation and variance are

$$E\{x|\gamma, \lambda\} = \frac{\lambda}{\gamma - 1}, \quad \gamma > 1, \tag{4.B.5}$$

$$\mathrm{Var}\{x|\gamma, \lambda\} = \frac{\lambda^2}{(\gamma-1)^2(\gamma-2)}, \quad \gamma > 2. \tag{4.B.6}$$

The expectation and variance in expressions (4.B.5) and (4.B.6) approach infinity when $\gamma \to 1$ and $\gamma \to 2$ (from above), respectively. For $\gamma \leq 1$ and $\gamma \leq 2$ ($\gamma > 0$), the expectation and variance are not defined.

For a Gaussian distribution $\mathcal{N}_1(\mu, \sigma^2)$ with unknown variance, the inverse gamma distribution is a conjugate distribution for the variance. That means that if the prior distribution for the variance is inverse gamma $\sigma^2 \sim \mathcal{IG}(\gamma, \lambda)$, then the posterior distribution is also inverse gamma, but with modified parameters. For $x_1, \ldots, x_n$ iid $\mathcal{N}_1(\mu, \sigma^2)$, the posterior distribution for the unknown variance is

$$\sigma^2 | s^2 \sim \mathcal{IG}(\gamma + \frac{n}{2}, \lambda + \frac{ns^2}{2}), \tag{4.B.7}$$

where

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{n}. \tag{4.B.8}$$

4.C Wishart and inverted Wishart distribution

Let $\boldsymbol{\Sigma}$ be an $n \times n$ covariance matrix. The Wishart probability density function for a symmetric positive definite $n \times n$ matrix $\boldsymbol{W}$ is (Anderson, 1984)

$$p(\boldsymbol{W} | \boldsymbol{\Sigma}, m) = \frac{|\boldsymbol{W}|^{(m-n-1)/2} \exp[-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{W})]}{2^{mn/2} |\boldsymbol{\Sigma}|^{m/2} \Gamma_n(m/2)}, \tag{4.C.1}$$

for $m > n$ where m is the degrees of freedom, $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, that is the sum of the diagonal elements, and

$$\Gamma_n(\gamma) = \pi^{n(n-1)/4} \prod_{i=1}^{n} \Gamma[\gamma - (i-1)/2]. \tag{4.C.2}$$

In the special case with $n = 1$, the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{W}$ are reduced to scalars. If $\boldsymbol{\Sigma} = 1/2\lambda$, $\boldsymbol{W} = x$, and $m/2 = \gamma$, the Wishart distribution is reduced to the gamma distribution.

Let the Wishart distribution be denoted $\mathcal{W}_n(\boldsymbol{\Sigma}, m)$. If $\boldsymbol{r}_i$ is independent Gaussian $\boldsymbol{r}_i \sim \mathcal{N}_n(\boldsymbol{0}, \boldsymbol{\Sigma})$ for $i = 1, \ldots, m$, then

$$\boldsymbol{W} = \sum_{i=1}^{m} \boldsymbol{r}_i \boldsymbol{r}_i^T \sim \mathcal{W}_n(\boldsymbol{\Sigma}, m), \tag{4.C.3}$$

and

$$\boldsymbol{W}^{-1} \sim \mathcal{IW}_n(\boldsymbol{\Sigma}^{-1}, m), \tag{4.C.4}$$

where $\mathcal{IW}_n(\boldsymbol{\Sigma}^{-1}, m)$ is the inverted Wishart distribution with m degrees of freedom. The expectations are

$$\mathrm{E}\{\boldsymbol{W}\} = m\boldsymbol{\Sigma}, \tag{4.C.5}$$

and

$$\mathrm{E}\{\boldsymbol{W}^{-1}\} = \frac{1}{m-n-1}\boldsymbol{\Sigma}^{-1}. \tag{4.C.6}$$

Consider $\boldsymbol{\Sigma}$ to be unknown and stochastic. If $\boldsymbol{\Sigma}$ is assigned an inverted Wishart prior distribution,

$$\boldsymbol{\Sigma} \sim \mathcal{IW}_n(\boldsymbol{\Sigma}_0, m_0), \tag{4.C.7}$$

then the conditional distribution of $\boldsymbol{\Sigma}$ given $\boldsymbol{W}$ is

$$\boldsymbol{\Sigma}|\boldsymbol{W} \sim \mathcal{IW}_n(\boldsymbol{\Sigma}_0 + \boldsymbol{W}, m_0 + m). \tag{4.C.8}$$

4.D The full conditional distributions

The full conditional distribution for an unknown variable conditioned on all the other variables can be constructed only by the terms in expression (4.16) containing the actual variable. In the following, all the necessary conditional distributions are defined.

The full conditional distribution for $\boldsymbol{m}$

From expression (4.16), the full conditional distribution for the model parameter vector $\boldsymbol{m}$ can be written

$$p(\boldsymbol{m}|\cdot) \propto p(\boldsymbol{d}_{obs}|\boldsymbol{m}, \boldsymbol{s}, \boldsymbol{\Sigma}_d)\; p(\boldsymbol{w}_{obs}|\boldsymbol{m}, \boldsymbol{\Sigma}_w)\; p(\boldsymbol{m}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \tag{4.D.1}$$

where the two first terms are the Gaussian likelihoods defined in expressions (4.4) and (4.6), and the last term is the Gaussian prior defined in expression (4.7). Below, the full conditional distribution for $\boldsymbol{m}$ is calculated by firstly

defining $p(\boldsymbol{d}_{obs}|\boldsymbol{s}, \boldsymbol{\Sigma}_d)$ and $p(\boldsymbol{w}_{obs}|\boldsymbol{\Sigma}_w)$. Following Buland and Omre (2002a), let the derivative of $\boldsymbol{m}(\boldsymbol{x}, t)$ be

$$\boldsymbol{m}'(\boldsymbol{x}, t) = \left[\frac{\partial}{\partial t}\ln\alpha(\boldsymbol{x}, t), \frac{\partial}{\partial t}\ln\beta(\boldsymbol{x}, t), \frac{\partial}{\partial t}\ln\rho(\boldsymbol{x}, t)\right]^T, \tag{4.D.2}$$

with discrete representation $\boldsymbol{m}'$. The corresponding discrete version of the reflectivity is

$$\boldsymbol{c} = \boldsymbol{A}\boldsymbol{m}'. \tag{4.D.3}$$

where $\boldsymbol{A}$ is a sparse matrix defined by the coefficients a_α, a_β and a_ρ in expression (4.1).

Since $\boldsymbol{m}$ is assumed to be Gaussian with expectation vector $\boldsymbol{\mu}_m$ and covariance matrix $\boldsymbol{\Sigma}_m$, expression (4.7), then $\boldsymbol{m}'$ is Gaussian,

$$\boldsymbol{m}'|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}'_m, \boldsymbol{\Sigma}''_m), \tag{4.D.4}$$

see Christakos (1992) for details on differentiation of Gaussian variables. The discrete expectation vector $\boldsymbol{\mu}'_m$ is defined by

$$\mathrm{E}\{\boldsymbol{m}'(\boldsymbol{x}, t)\} = \frac{\partial}{\partial t}\boldsymbol{\mu}_m(\boldsymbol{x}, t), \tag{4.D.5}$$

and the covariance matrix $\boldsymbol{\Sigma}''_m$ is defined by

$$\mathrm{Cov}\{\boldsymbol{m}'(\boldsymbol{x}, t), \boldsymbol{m}'(\boldsymbol{y}, s)\} = \frac{\partial^2}{\partial t \partial s}\boldsymbol{\Sigma}_m(\boldsymbol{x}, t; \boldsymbol{y}, s). \tag{4.D.6}$$

The discrete cross-covariance matrix between $\boldsymbol{m}'$ and $\boldsymbol{m}$ is denoted $\boldsymbol{\Sigma}'_m$, and is defined by

$$\mathrm{Cov}\{\boldsymbol{m}'(\boldsymbol{x}, t), \boldsymbol{m}(\boldsymbol{y}, s)\} = \frac{\partial}{\partial t}\boldsymbol{\Sigma}_m(\boldsymbol{x}, t; \boldsymbol{y}, s). \tag{4.D.7}$$

The relationships between the observed data $\boldsymbol{o} = [\boldsymbol{w}_{obs}, \boldsymbol{d}_{obs}]^T$ and the model parameter vector $\boldsymbol{m}$ are given by

$$\boldsymbol{w}_{obs} = \boldsymbol{P}\boldsymbol{m} + \boldsymbol{e}_w, \tag{4.D.8}$$

and the convolutional model

$$\boldsymbol{d}_{obs} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{m}' + \boldsymbol{e}_d, \tag{4.D.9}$$

here written in matrix-vector form where $\boldsymbol{S}$ is a matrix representation of the wavelet $\boldsymbol{s}$. The conditional joint distribution for $\boldsymbol{m}$ and $\boldsymbol{o}$ is

$$\left[\begin{array}{c} \boldsymbol{m} \\ \boldsymbol{o} \end{array} \middle| \boldsymbol{s}, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m, \boldsymbol{\Sigma}_d, \boldsymbol{\Sigma}_w \right] \sim \mathcal{N}_{n_m+n_w+n_d} \left(\left[\begin{array}{c} \boldsymbol{\mu}_m \\ \boldsymbol{\mu}_o \end{array} \right], \left[\begin{array}{cc} \boldsymbol{\Sigma}_m & \boldsymbol{\Sigma}_{m,o} \\ \boldsymbol{\Sigma}_{o,m} & \boldsymbol{\Sigma}_o \end{array} \right] \right), \tag{4.D.10}$$

where

$$\boldsymbol{\mu}_o = \left[\begin{array}{c} \boldsymbol{P}\boldsymbol{\mu}_m \\ \boldsymbol{S}\boldsymbol{A}\boldsymbol{\mu}'_m \end{array} \right], \tag{4.D.11}$$

$$\boldsymbol{\Sigma}_o = \left[\begin{array}{cc} \boldsymbol{P}\boldsymbol{\Sigma}_m\boldsymbol{P}^T + \boldsymbol{\Sigma}_w & \boldsymbol{P}{\boldsymbol{\Sigma}'_m}^T\boldsymbol{A}^T\boldsymbol{S}^T \\ \boldsymbol{S}\boldsymbol{A}\boldsymbol{\Sigma}'_m\boldsymbol{P}^T & \boldsymbol{S}\boldsymbol{A}\boldsymbol{\Sigma}''_m\boldsymbol{A}^T\boldsymbol{S}^T + \boldsymbol{\Sigma}_d \end{array} \right], \tag{4.D.12}$$

$$\boldsymbol{\Sigma}_{o,m} = \left[\begin{array}{c} \boldsymbol{P}\boldsymbol{\Sigma}_m \\ \boldsymbol{S}\boldsymbol{A}\boldsymbol{\Sigma}'_m \end{array} \right], \tag{4.D.13}$$

and

$$\boldsymbol{\Sigma}_{m,o} = \boldsymbol{\Sigma}_{o,m}^T. \tag{4.D.14}$$

The full conditional distribution for $\boldsymbol{m}$ is

$$\boldsymbol{m}|\boldsymbol{o}, \boldsymbol{s}, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m, \boldsymbol{\Sigma}_d, \boldsymbol{\Sigma}_w \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}_{m|\cdot}, \boldsymbol{\Sigma}_{m|\cdot}), \tag{4.D.15}$$

where the conditional expectation and covariance are

$$\boldsymbol{\mu}_{m|\cdot} = \boldsymbol{\mu}_m + \boldsymbol{\Sigma}_{m,o}\boldsymbol{\Sigma}_o^{-1}(\boldsymbol{o} - \boldsymbol{\mu}_o), \tag{4.D.16}$$

$$\boldsymbol{\Sigma}_{m|\cdot} = \boldsymbol{\Sigma}_m - \boldsymbol{\Sigma}_{m,o}\boldsymbol{\Sigma}_o^{-1}\boldsymbol{\Sigma}_{o,m}. \tag{4.D.17}$$

For real size problems, the dimension of $\boldsymbol{m}$ is usually large, and an efficient numerical technique is required. For model parameters and seismic data sampled on a regular grid, the spatially coupled inversion problem can be solved efficiently and exactly in the Fourier domain when the involved covariance matrices are homogeneous and stationary (Buland et al., 2002). For more general cases, an approximative solution can be obtained by the sequential screening algorithm (Omre et al., 1993), see Appendix 4.E.

The full conditional distribution for $\boldsymbol{s}$

From expression (4.16), the full conditional distribution for the wavelet $\boldsymbol{s}$ can be written

$$p(\boldsymbol{s}|\cdot) \propto p(\boldsymbol{d}_{obs}|\boldsymbol{m}, \boldsymbol{s}, \boldsymbol{\Sigma}_d)\, p(\boldsymbol{s}|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s), \tag{4.D.18}$$

where the first term is the Gaussian likelihood defined in expression (4.4) and the last term is the Gaussian prior defined in expression (4.8). Firstly, $p(\boldsymbol{d}_{obs}|\boldsymbol{m}, \boldsymbol{\Sigma}_d)$ is calculated. The convolutional model can be written

$$\boldsymbol{d}_{obs} = \boldsymbol{C}\boldsymbol{s} + \boldsymbol{e}_d, \tag{4.D.19}$$

where the matrix-vector multiplication $\boldsymbol{C}\boldsymbol{s}$ represents the convolution of the reflectivity vector $\boldsymbol{c} = \boldsymbol{A}\boldsymbol{m}'$ with the wavelet vector $\boldsymbol{s}$. The conditional joint distribution for $\boldsymbol{s}$ and $\boldsymbol{d}_{obs}$ is

$$\left[\begin{array}{c} \boldsymbol{s} \\ \boldsymbol{d}_{obs} \end{array} \middle| \boldsymbol{m}, \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \boldsymbol{\Sigma}_d \right] \sim \mathcal{N}_{n_s+n_d} \left(\left[\begin{array}{c} \boldsymbol{\mu}_s \\ \boldsymbol{C}\boldsymbol{\mu}_s \end{array} \right], \left[\begin{array}{cc} \boldsymbol{\Sigma}_s & \boldsymbol{\Sigma}_s\boldsymbol{C}^T \\ \boldsymbol{C}\boldsymbol{\Sigma}_s & \boldsymbol{C}\boldsymbol{\Sigma}_s\boldsymbol{C}^T + \boldsymbol{\Sigma}_d \end{array} \right] \right). \tag{4.D.20}$$

The full conditional distribution for $\boldsymbol{s}$ is then

$$\boldsymbol{s}|\boldsymbol{d}_{obs}, \boldsymbol{m}, \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \boldsymbol{\Sigma}_d \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_{s|\cdot}, \boldsymbol{\Sigma}_{s|\cdot}), \tag{4.D.21}$$

where the conditional expectation and covariance are

$$\boldsymbol{\mu}_{s|\cdot} = \boldsymbol{\mu}_s + \boldsymbol{\Sigma}_s\boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{\Sigma}_s\boldsymbol{C}^T + \boldsymbol{\Sigma}_d)^{-1}(\boldsymbol{d}_{obs} - \boldsymbol{C}\boldsymbol{\mu}_s), \tag{4.D.22}$$

$$\boldsymbol{\Sigma}_{s|\cdot} = \boldsymbol{\Sigma}_s - \boldsymbol{\Sigma}_s\boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{\Sigma}_s\boldsymbol{C}^T + \boldsymbol{\Sigma}_d)^{-1}\boldsymbol{C}\boldsymbol{\Sigma}_s. \tag{4.D.23}$$

The full conditional distribution for $\boldsymbol{\mu}$

The conditional distribution for the wavelet expectation $\boldsymbol{\mu}_s$ depends on the wavelet $\boldsymbol{s}$ and the covariance matrix $\boldsymbol{\Sigma}_s$. Similarly, the conditional distribution for $\boldsymbol{\mu}_m$ depends on $\boldsymbol{m}$ and $\boldsymbol{\Sigma}_m$. The derivation of the conditional distribution for $\boldsymbol{\mu}_s$ and $\boldsymbol{\mu}_m$ follows the same pattern and is shown below for $\boldsymbol{\mu}_s$.

Both $\boldsymbol{s}$ and its expectation $\boldsymbol{\mu}_s$ are Gaussian, and the conditional joint distribution is

$$\left[\begin{array}{c} \boldsymbol{\mu_s} \\ \boldsymbol{s} \end{array} \middle| \boldsymbol{\Sigma}_s \right] \sim \mathcal{N}_{n_s+n_s} \left(\left[\begin{array}{c} \boldsymbol{\mu}_{\mu_s} \\ \boldsymbol{\mu}_{\mu_s} \end{array} \right], \left[\begin{array}{cc} \boldsymbol{\Sigma}_{\mu_s} & \boldsymbol{\Sigma}_{\mu_s} \\ \boldsymbol{\Sigma}_{\mu_s} & \boldsymbol{\Sigma}_{\mu_s} + \boldsymbol{\Sigma}_s \end{array} \right] \right). \tag{4.D.24}$$

The full conditional distribution for $\boldsymbol{\mu}_s$ is

$$\boldsymbol{\mu}_s|\boldsymbol{s}, \boldsymbol{\Sigma}_s \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_{\mu_s|\cdot}, \boldsymbol{\Sigma}_{\mu_s|\cdot}), \tag{4.D.25}$$

where

$$\boldsymbol{\mu}_{\mu_s|\cdot} = \boldsymbol{\mu}_{\mu_s} + \boldsymbol{\Sigma}_{\mu_s}(\boldsymbol{\Sigma}_{\mu_s} + \boldsymbol{\Sigma}_s)^{-1}(\boldsymbol{s} - \boldsymbol{\mu}_{\mu_s}), \tag{4.D.26}$$

$$\boldsymbol{\Sigma}_{\mu_s|\cdot} = \boldsymbol{\Sigma}_{\mu_s} - \boldsymbol{\Sigma}_{\mu_s}(\boldsymbol{\Sigma}_{\mu_s} + \boldsymbol{\Sigma}_s)^{-1}\boldsymbol{\Sigma}_{\mu_s}. \tag{4.D.27}$$

The full conditional distribution for σ^2

The covariance matrices are assumed to be known up to unknown variance factors. When the prior distribution for σ^2 is inverse Gamma

$$\sigma^2 \sim \mathcal{IG}(\gamma, \lambda), \tag{4.D.28}$$

then the full conditional distribution for σ^2 is

$$\sigma^2 | s^2 \sim \mathcal{IG}(\gamma + \frac{n}{2}, \lambda + \frac{ns^2}{2}), \tag{4.D.29}$$

where s^2 is the sample variance, for example

$$s_s^2 = \frac{1}{n_s}(\boldsymbol{s} - \boldsymbol{\mu}_s)^T \boldsymbol{\Sigma}_{0,s}^{-1}(\boldsymbol{s} - \boldsymbol{\mu}_s), \tag{4.D.30}$$

for the wavelet variance σ_s^2, where $\boldsymbol{\Sigma}_{0,s}$ is the known structural factor of the covariance matrix $\boldsymbol{\Sigma}_s$.

4.E The sequential screening algorithm

The full conditional distribution for the complete spatially coupled model parameter vector $\boldsymbol{m}$, expressions (4.D.15)-(4.D.17), can be factorized as

$$\begin{aligned} p(\boldsymbol{m}|\cdot) &= p(\boldsymbol{m}(\boldsymbol{x}_1)|\cdot) \\ &\times \; p(\boldsymbol{m}(\boldsymbol{x}_2)|\boldsymbol{m}(\boldsymbol{x}_1), \cdot) \\ &\vdots \\ &\times \; p(\boldsymbol{m}(\boldsymbol{x}_n)|\boldsymbol{m}(\boldsymbol{x}_1), \ldots, \boldsymbol{m}(\boldsymbol{x}_{n-1}), \cdot), \end{aligned} \tag{4.E.1}$$

where $\boldsymbol{x}_1$ to $\boldsymbol{x}_n$ are the involved surface locations for the inversion in some arbitrary order, and $\boldsymbol{m}(\boldsymbol{x}_i)$ is a sub-vector of $\boldsymbol{m}$ containing the model parameters at location $\boldsymbol{x}_i$. Expression (4.E.1) defines an exact sequential simulation algorithm, but the dimensions of the pdfs on the right hand side is rapidly growing, and the algorithm can not be used in real size problems.

The sequential screening algorithm is an approximate version of the exact sequential simulation algorithm defined above. The sequential screening algorithm is based on the assumption that the solution at a location $\boldsymbol{x}_i$ depends only on a subset of the model and data. The full conditional distribution for

$\boldsymbol{m}(\boldsymbol{x}_i)$ is obtained by conditioning to a subset of $\{\boldsymbol{m}(\boldsymbol{x}_1), \ldots, \boldsymbol{m}(\boldsymbol{x}_{i-1})\}$, defined by nearest neighbors to $\boldsymbol{x}_i$ in specified directions, and a subset of $\boldsymbol{w}_{obs}$ and $\boldsymbol{d}_{obs}$. The locations of the wells should define the first locations. Next, the inversion should be run for the boundary locations of the survey, and then a halving procedure is used to select new inversion locations, see Omre et al. (1993).

Chapter 5

Rapid spatially coupled AVO inversion in the Fourier domain

Arild Buland, Odd Kolbjørnsen and Henning Omre

Paper submitted for publication.

Abstract

Spatial coupling of the model parameters in an inversion problem provides lateral consistence and robust solutions. We have defined the inversion problem in a Bayesian framework, where the solution is represented by a posterior distribution obtained from a prior distribution and a likelihood model for the recorded data. The spatial coupling of the model parameters is imposed via the prior distribution by a spatial correlation function. In the Fourier domain, the spatially correlated model parameters can be decoupled, and the inversion problem can be solved independently for each frequency component. For a spatial model parameter represented on n grid nodes, the computing time for the inversion in the Fourier domain follows a linear function of the number of grid nodes, while the computing time for the fast Fourier transform follows an $n \log n$ function. We have developed a 3-D linearized AVO inversion method with spatially coupled model parameters, where the objective is to obtain posterior distributions for P-wave velocity, S-wave velocity, and density. The inversion algorithm has been tested on a 3-D dataset from the Sleipner Field with 4 million grid nodes, each with three unknown model parameters. The computing time was less than 3 minutes on the inversion in the Fourier domain,

while each 3-D Fourier transform used about 30 seconds on a single 400 MHz Mips R12000 CPU.

5.1 Introduction

Many geophysical inverse problems can naturally be cast in a Bayesian framework, where it is possible to combine available prior knowledge with the information contained in the measured data, see e.g., Tarantola and Valette (1982); Tarantola (1987); Duijndam (1988a,b). The solution of a Bayesian inverse problem is represented by the posterior distribution. From the posterior distribution, the best estimate of the solution and the corresponding uncertainty can be extracted. A set of plausible solutions can also be drawn directly from the posterior distribution.

Amplitude versus offset (AVO) inversion can be used to extract information about the elastic subsurface parameters from the angle dependency in the reflectivity, see e.g., Hampson and Russell (1990); Lörtzer and Berkhout (1993); Pan et al. (1994); Buland et al. (1996); Gouveia and Scales (1998). In practice, and especially for 3-D surveys, linearized AVO inversion is attractive since it can be performed with use of moderate computer resources. Prior to a linearized AVO inversion, the seismic data must be processed to remove nonlinear relations between the model parameters and the seismic response. Important steps in the processing are the removal of the moveout, multiples, and the effects of geometrical spreading and absorption. The seismic data should be prestack migrated, such that dip related effects are removed. After prestack migration, it is reasonable to assume that each single bin-gather can be regarded as the response of a local 1-D earth model. The benefits of prestack migration before AVO analysis is discussed in Brown (1992); Mosher et al. (1996); Buland and Landrø (2001). We further assume that wave mode conversions, interbed multiples and anisotropy effects can be neglected after processing. Finally, the prestack gathers must be transformed from offsets to reflection angles.

Under Gaussian model assumptions, an explicit analytical solution of a Bayesian linearized AVO inversion problem is worked out for a single angle gather, see Buland and Omre (2002a). The objective of this method is to obtain posterior distributions for the P-wave velocity, S-wave velocity, and density. The solution is fast to compute and the method is therefore suitable for inversion of seismic 3-D data. However, the model parameters are not laterally coupled,

so each CDP gather is inverted independently of the neighbor CDPs.

In the current paper, a spatially coupled model is defined to obtain a spatial consistent and robust solution of the linearized AVO inversion problem. The consequence of the spatial coupling is that the solution in each location depends on the solutions in all other locations. Even for small data sets, this results in an enormous system of equations. For example, a small 3-D inversion problem may have dimension $100 \times 100 \times 100$, that is $n = 10^6$ grid nodes. The computing time for inversion of the corresponding equation system is proportional to n^3, denoted $\mathcal{O}(n^3)$. An obvious approximate approach to this problem is to assume that the solution in a specific location only depends on the solutions at the nearest neighbor locations, see e.g., Omre et al. (1993); Rue (2000). Domain decomposition constitutes another approximate technique, where the inversion area is divided into several subareas, each limited to a size which efficiently can be handled by the actual computer. The problems with this method are related to boundary effects and the final coupling of the inverted subareas. In this paper we present a Bayesian AVO inversion method where the spatial coupling can be handled exactly under certain assumptions. The method utilizes the fact that the covariance matrix for a homogeneously correlated spatial variable sampled on a regular grid can be diagonalized by a Fourier transform, see Wood (1995). In the Fourier domain, the inversion problem can be solved independently for each frequency component. The computing time for the inversion in the Fourier domain is then $\mathcal{O}(n)$, which is the optimal scaling property for an inversion algorithm. However, the computing time for a fast Fourier transform is $\mathcal{O}(n \log n)$, such that the Fourier and inverse Fourier transforms will dominate the computing time asymptotically.

5.2 Methodology

An isotropic, elastic medium is completely described by three material parameters $\{\alpha(\boldsymbol{x}, t), \beta(\boldsymbol{x}, t), \rho(\boldsymbol{x}, t)\}$, where α, β, and ρ are P-wave velocity, S-wave velocity, and density, $\boldsymbol{x}$ is the lateral position, and t is the vertical seismic traveltime. A weak contrast approximation to the seismic reflectivity function $c(\boldsymbol{x}, t, \theta)$ is (Buland and Omre, 2002a)

$$\begin{aligned} c(\boldsymbol{x}, t, \theta) = a_\alpha(\boldsymbol{x}, t, \theta) \frac{\partial}{\partial t} \ln \alpha(\boldsymbol{x}, t) \\ + a_\beta(\boldsymbol{x}, t, \theta) \frac{\partial}{\partial t} \ln \beta(\boldsymbol{x}, t) + a_\rho(\boldsymbol{x}, t, \theta) \frac{\partial}{\partial t} \ln \rho(\boldsymbol{x}, t), \end{aligned} \quad (5.1)$$

where θ is the reflection angle, and

$$\begin{aligned} a_\alpha(\boldsymbol{x},t,\theta) &= \frac{1}{2}\left(1+\tan^2\theta\right), \\ a_\beta(\boldsymbol{x},t,\theta) &= -4\frac{\beta^2(\boldsymbol{x},t)}{\alpha^2(\boldsymbol{x},t)}\sin^2\theta, \\ a_\rho(\boldsymbol{x},t,\theta) &= \frac{1}{2}\left(1-4\frac{\beta^2(\boldsymbol{x},t)}{\alpha^2(\boldsymbol{x},t)}\sin^2\theta\right). \end{aligned} \tag{5.2}$$

Motivated by the form of the reflectivity function in expression (5.1), let

$$\boldsymbol{m}(\boldsymbol{x},t) = [\ln\alpha(\boldsymbol{x},t), \ln\beta(\boldsymbol{x},t), \ln\rho(\boldsymbol{x},t)]^T, \tag{5.3}$$

where T denotes transpose such that $\boldsymbol{m}(\boldsymbol{x},t)$ is a column vector. Further, let $\boldsymbol{a}(\boldsymbol{x},t,\theta)$ be a row vector

$$\boldsymbol{a}(\boldsymbol{x},t,\theta) = [a_\alpha(\boldsymbol{x},t,\theta), a_\beta(\boldsymbol{x},t,\theta), a_\rho(\boldsymbol{x},t,\theta)]. \tag{5.4}$$

For zero incidence reflections, the reflectivity function $c(\boldsymbol{x},t,\theta)$ reduces to

$$c(\boldsymbol{x},t,0) = \frac{1}{2}\frac{\partial}{\partial t}\ln Z_P(\boldsymbol{x},t), \tag{5.5}$$

where $Z_P = \alpha\rho$ is the acoustic impedance. In this case, $\boldsymbol{a}(\boldsymbol{x},t,0)$ reduces to $1/2$, and $\boldsymbol{m}(\boldsymbol{x},t)$ reduces to $\ln Z_P(\boldsymbol{x},t)$. Inversion for acoustic impedance from zero incidence data can be defined by a simple reformulation of the AVO inversion problem, and is therefore not further discussed in this paper.

The seismic data are represented by the convolutional model

$$d_{obs}(\boldsymbol{x},t,\theta) = \int s(\tau,\theta)\, c(\boldsymbol{x},t-\tau,\theta)\, d\tau + e(\boldsymbol{x},t,\theta), \tag{5.6}$$

where s is the wavelet, and e is an error term. Note that the wavelet is allowed to be angle dependent, but independent of the lateral position $\boldsymbol{x}$. The wavelet is assumed to be stationary within a limited target window.

The Fourier transform

The spatial coupled inversion problem can be decoupled in the Fourier domain. The inversion problem can then be solved independently for each frequency component. Let the Fourier transform be defined as

$$\tilde{f}(\boldsymbol{k},\omega) = \iiint f(\boldsymbol{x},t)\exp[-i(\boldsymbol{k}\cdot\boldsymbol{x}+\omega t)]\, d\boldsymbol{x}\, dt, \tag{5.7}$$

with inverse transform

$$f(\boldsymbol{x},t) = \frac{1}{(2\pi)^3} \iiint \tilde{f}(\boldsymbol{k},\omega) \exp[i(\boldsymbol{k}\cdot\boldsymbol{x} + \omega t)]\, d\boldsymbol{k}\, d\omega, \tag{5.8}$$

where $i = \sqrt{-1}$, ω is the temporal frequency, and $\boldsymbol{k}$ is the spatial frequency vector with components k_x and k_y. In the following, frequency means a $(\boldsymbol{k},\omega)$ pair. Note that geophysicists often call k_x and k_y wavenumbers, and the sign of the temporal frequency ω is often defined opposite of the definition above.

The Fourier transform of the convolutional model in expression (5.6) is

$$\tilde{d}_{obs}(\boldsymbol{k},\omega,\theta) = \tilde{s}(\omega,\theta)\, \tilde{c}(\boldsymbol{k},\omega,\theta) + \tilde{e}(\boldsymbol{k},\omega,\theta). \tag{5.9}$$

We use a constant β/α ratio in expression (5.2), such that $\boldsymbol{a}(\boldsymbol{x},t,\theta) = \boldsymbol{a}(\theta)$, then the Fourier transform of the convolutional model can be written

$$\tilde{d}_{obs}(\boldsymbol{k},\omega,\theta) = \boldsymbol{g}(\omega,\theta)\cdot\tilde{\boldsymbol{m}}(\boldsymbol{k},\omega) + \tilde{e}(\boldsymbol{k},\omega,\theta), \tag{5.10}$$

where $\boldsymbol{g}$ is a row vector defined by

$$\boldsymbol{g}(\omega,\theta) = i\omega\, \tilde{s}(\omega,\theta)\, \boldsymbol{a}(\theta), \tag{5.11}$$

and $\tilde{\boldsymbol{m}}(\boldsymbol{k},\omega)$ is the Fourier transform of the elements in $\boldsymbol{m}(\boldsymbol{x},t)$, that is $\ln\alpha(\boldsymbol{x},t)$, $\ln\beta(\boldsymbol{x},t)$, and $\ln\rho(\boldsymbol{x},t)$, see expression (5.3). Although a constant β/α ratio is used in an approximative expression for the reflection coefficient, the solution will in general have a varying β/α ratio. Further, note that the differentiations in equation (5.1) now appear as an $i\omega$ term in expression (5.11). In Buland and Omre (2002a), it was assumed that $\boldsymbol{m}(\boldsymbol{x},t)$ was differentiable with respect to time, but here it is sufficient to assume that the convolution of $s(t,\theta)$ and $\boldsymbol{a}(\theta)\cdot\boldsymbol{m}(\boldsymbol{x},t)$ is differentiable.

For a set of n_θ specified reflection angles, the Fourier transformed seismic data can be written in the vector form

$$\tilde{\boldsymbol{d}}_{obs}(\boldsymbol{k},\omega) = \boldsymbol{G}(\omega)\, \tilde{\boldsymbol{m}}(\boldsymbol{k},\omega) + \tilde{\boldsymbol{e}}(\boldsymbol{k},\omega), \tag{5.12}$$

where $\boldsymbol{G}(\omega)$ is an $n_\theta \times 3$ matrix defined by

$$\boldsymbol{G}(\omega) = \begin{bmatrix} \boldsymbol{g}(\omega,\theta_1) \\ \vdots \\ \boldsymbol{g}(\omega,\theta_{n_\theta}) \end{bmatrix}, \tag{5.13}$$

and $\tilde{\boldsymbol{d}}_{obs}(\boldsymbol{k},\omega)$ and $\tilde{\boldsymbol{e}}(\boldsymbol{k},\omega)$ are n_θ-dimensional vectors.

The prior model

The elastic parameters $\alpha(\boldsymbol{x},t)$, $\beta(\boldsymbol{x},t)$, and $\rho(\boldsymbol{x},t)$ are assumed to be log-Gaussian random fields, hence the vector field $\boldsymbol{m}(\boldsymbol{x},t)$, which contains the logarithm of these parameters, is Gaussian with expectation

$$\boldsymbol{\mu}_m(\boldsymbol{x},t) = [\mu_\alpha(\boldsymbol{x},t), \mu_\beta(\boldsymbol{x},t), \mu_\rho(\boldsymbol{x},t)]^T, \tag{5.14}$$

where the elements are the expectations of $\ln\alpha(\boldsymbol{x},t)$, $\ln\beta(\boldsymbol{x},t)$, and $\ln\rho(\boldsymbol{x},t)$, respectively, and with covariance

$$\boldsymbol{\Sigma}_m(\boldsymbol{x}_1,t_1;\boldsymbol{x}_2,t_2) = \mathrm{Cov}\{\boldsymbol{m}(\boldsymbol{x}_1,t_1), \boldsymbol{m}(\boldsymbol{x}_2,t_2)\}. \tag{5.15}$$

We assume that the covariance function is stationary and homogeneous, and can be factorized as

$$\boldsymbol{\Sigma}_m(\boldsymbol{x}_1,t_1;\boldsymbol{x}_2,t_2) = \boldsymbol{\Sigma}_{0,m}\,\nu_m(\boldsymbol{\xi},\tau), \tag{5.16}$$

where $\nu_m(\boldsymbol{\xi},\tau)$ is a spatial correlation function, $\boldsymbol{\xi} = [|x_2 - x_1|, |y_2 - y_1|]^T$, $\tau = |t_2 - t_1|$, and

$$\boldsymbol{\Sigma}_{0,m} = \begin{bmatrix} \sigma_\alpha^2 & \sigma_\alpha\sigma_\beta\nu_{\alpha\beta} & \sigma_\alpha\sigma_\rho\nu_{\alpha\rho} \\ \sigma_\alpha\sigma_\beta\nu_{\alpha\beta} & \sigma_\beta^2 & \sigma_\beta\sigma_\rho\nu_{\beta\rho} \\ \sigma_\alpha\sigma_\rho\nu_{\alpha\rho} & \sigma_\beta\sigma_\rho\nu_{\beta\rho} & \sigma_\rho^2 \end{bmatrix}. \tag{5.17}$$

The diagonal elements of $\boldsymbol{\Sigma}_{0,m}$ are the variances, and $\nu_{\alpha\beta}$, $\nu_{\alpha\rho}$, and $\nu_{\beta\rho}$, are the correlations between $\ln\alpha(\boldsymbol{x},t)$, $\ln\beta(\boldsymbol{x},t)$ and $\ln\rho(\boldsymbol{x},t)$, respectively. A more general covariance function is also allowed, where the covariance function is composed of a sum of terms with the form on the right-hand side of expression (5.16).

The model parameters and the seismic data are so far defined for continuous $\boldsymbol{x}$ and t. In practice, the seismic data are available in a discrete form. In the following we assume an identical sampling of the model parameters and the seismic data on a regular grid in space and time. This is a required assumption for this method. Let the discrete representation of the model parameter field $\boldsymbol{m}(\boldsymbol{x},t)$ in a time window and for a set of lateral positions be written $\boldsymbol{m}$. The discrete model parameter vector $\boldsymbol{m}$ is Gaussian with expectation vector $\boldsymbol{\mu}_m$ and covariance matrix $\boldsymbol{\Sigma}_m$, shortly denoted

$$\boldsymbol{m} \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \tag{5.18}$$

where n_m is the dimension of $\boldsymbol{m}$. For a 3-D problem on a regular grid with $n = n_x n_y n_t$ grid nodes, the dimensions of $\boldsymbol{m}$ and $\boldsymbol{\mu}_m$ are $n_m = 3n$, while the dimension of $\boldsymbol{\Sigma}_m$ is $n_m \times n_m$. Since the covariance function can be written as in expression (5.16), the complete covariance matrix for $\boldsymbol{m}$ can be written as a Kronecker product

$$\boldsymbol{\Sigma}_m = \boldsymbol{\Sigma}_{0,m} \otimes \boldsymbol{\Upsilon}_m, \tag{5.19}$$

where each of the elements in the 3×3 constant matrix $\boldsymbol{\Sigma}_{0,m}$, defined in expression (5.17), are multiplied with the $n \times n$ correlation matrix $\boldsymbol{\Upsilon}_m$, defined from $\nu_m(\boldsymbol{\xi}, \tau)$.

The spatial dependency can be decoupled by Fourier transforming the problem. The Fourier transform of $\boldsymbol{m}$, denoted $\tilde{\boldsymbol{m}}$, is Gaussian with Fourier transformed expectation vector $\tilde{\boldsymbol{\mu}}_m$ and covariance matrix

$$\tilde{\boldsymbol{\Sigma}}_m = \boldsymbol{\Sigma}_{0,m} \otimes \tilde{\boldsymbol{\Lambda}}_m, \tag{5.20}$$

where $\tilde{\boldsymbol{\Lambda}}_m$ is the diagonal eigenvalue matrix of $\boldsymbol{\Upsilon}_m$ scaled by the dimension of the discrete Fourier transform, see Appendix 5.A. The important consequence of this diagonalization is that the frequency components of $\tilde{\boldsymbol{m}}$ are independent, with each component being Gaussian

$$\tilde{\boldsymbol{m}}_k \sim \mathcal{N}_3(\tilde{\boldsymbol{\mu}}_{m,k}, \tilde{\boldsymbol{\Sigma}}_{m,k}), \tag{5.21}$$

with index k corresponding to a specific discrete $(\boldsymbol{k}, \omega)$ pair. The covariance matrix for frequency component k is a 3×3 matrix defined by

$$\tilde{\boldsymbol{\Sigma}}_{m,k} = \tilde{\lambda}_{m,k} \boldsymbol{\Sigma}_{0,m}, \tag{5.22}$$

with $\tilde{\lambda}_{m,k}$ being the corresponding diagonal element in the scaled diagonal eigenvalue matrix $\tilde{\boldsymbol{\Lambda}}_m$. Further, and of crucial importance, is that the $n = n_x n_y n_t$ eigenvalues can be calculated efficiently by a 3-D Fourier transform of n discrete samples of the correlation function $\nu_m(\boldsymbol{\xi}, \tau)$ extended to a circulant form, see Appendix 5.A. That means that the complete $n \times n$ correlation matrix $\boldsymbol{\Upsilon}_m$ and the even larger covariance matrix $\boldsymbol{\Sigma}_m$ are not involved in the computations.

The statistical model for the seismic data

We assume that the error term $e(\boldsymbol{x}, t, \theta)$, introduced in the convolutional model in expression (5.6), is zero mean colored Gaussian noise. The covariance of the

error vector $\boldsymbol{e}(\boldsymbol{x},t) = [e(\boldsymbol{x},t,\theta_1),\ldots,e(\boldsymbol{x},t,\theta_{n_\theta})]^T$ is

$$\boldsymbol{\Sigma}_e(\boldsymbol{x}_1,t_1;\boldsymbol{x}_2,t_2) = \boldsymbol{\Sigma}_{0,e}\,\nu_e(\boldsymbol{\xi},\tau), \tag{5.23}$$

where $\boldsymbol{\Sigma}_{0,e}$ is an $n_\theta \times n_\theta$ covariance matrix containing the noise variances for the different reflection angles and the correlations between the angles, and $\nu_e(\boldsymbol{\xi},\tau)$ is a spatial and temporal correlation function. Again we allow sums of terms with the form on the right-hand side of expression (5.23). Note that white noise is a special case, where $\boldsymbol{\Sigma}_{0,e}$ is diagonal, and $\nu_e(\boldsymbol{\xi},\tau) = 0$ except for $\nu_e(\boldsymbol{0},0) = 1$.

As for the prior model above, the frequency components of the discrete Fourier transformed error vector $\tilde{\boldsymbol{e}}$ are now independent Gaussian

$$\tilde{\boldsymbol{e}}_k \sim \mathcal{N}_{n_\theta}(\boldsymbol{0},\tilde{\boldsymbol{\Sigma}}_{e,k}). \tag{5.24}$$

From expressions (5.12), (5.21), and (5.24), each frequency component of the seismic data is then apriori Gaussian

$$\tilde{\boldsymbol{d}}_{obs,k} \sim \mathcal{N}_{n_\theta}(\tilde{\boldsymbol{\mu}}_{d,k},\tilde{\boldsymbol{\Sigma}}_{d,k}), \tag{5.25}$$

where

$$\tilde{\boldsymbol{\mu}}_{d,k} = \boldsymbol{G}_k\tilde{\boldsymbol{\mu}}_{m,k}, \tag{5.26}$$

$$\tilde{\boldsymbol{\Sigma}}_{d,k} = \boldsymbol{G}_k\tilde{\boldsymbol{\Sigma}}_{m,k}\boldsymbol{G}_k^* + \tilde{\boldsymbol{\Sigma}}_{e,k}, \tag{5.27}$$

and $*$ denotes the conjugate transpose (adjoint).

The cross-covariance between the Fourier transform of seismic data and the model parameters is

$$\mathrm{Cov}\{\tilde{\boldsymbol{d}}_{obs,k},\tilde{\boldsymbol{m}}_k\} = \boldsymbol{G}_k\tilde{\boldsymbol{\Sigma}}_{m,k}. \tag{5.28}$$

The cross-covariance is needed to compute the posterior distribution.

The posterior model

The posterior distribution is defined by a Gaussian conditional distribution. A general presentation of Gaussian and conditional Gaussian distributions can be found in Anderson (1984). Using expressions (5.21) and (5.25)-(5.28), the posterior distribution for frequency component k is given by the Gaussian conditional distribution

$$\tilde{\boldsymbol{m}}_k|\tilde{\boldsymbol{d}}_{obs,k} \sim \mathcal{N}_3(\tilde{\boldsymbol{\mu}}_{m|d_{obs},k},\tilde{\boldsymbol{\Sigma}}_{m|d_{obs},k}), \tag{5.29}$$

where

$$\tilde{\boldsymbol{\mu}}_{m|d_{obs},k} = \tilde{\boldsymbol{\mu}}_{m,k} + (\boldsymbol{G}_k \tilde{\boldsymbol{\Sigma}}_{m,k})^* \tilde{\boldsymbol{\Sigma}}_{d,k}^{-1} (\tilde{\boldsymbol{d}}_{obs,k} - \tilde{\boldsymbol{\mu}}_{d,k}) \quad (5.30)$$

$$\tilde{\boldsymbol{\Sigma}}_{m|d_{obs},k} = \tilde{\boldsymbol{\Sigma}}_{m,k} - (\boldsymbol{G}_k \tilde{\boldsymbol{\Sigma}}_{m,k})^* \tilde{\boldsymbol{\Sigma}}_{d,k}^{-1} \boldsymbol{G}_k \tilde{\boldsymbol{\Sigma}}_{m,k}. \quad (5.31)$$

The core part of the inversion is the calculation of the 3 elements in the posterior mean vector in expression (5.30) and the 3×3 posterior covariance matrix in expression (5.31) for all frequency components k, that is for all discrete $(\boldsymbol{k}, \omega)$ pairs. The solution is transformed back to the $(\boldsymbol{x}, t)$ domain by 3-D inverse Fourier transforms, 3 for the posterior mean, and 6 for the posterior covariance since the covariance is symmetrical. The posterior distribution of the model parameters is represented by the posterior mean $\boldsymbol{\mu}_{m|d_{obs}}$ and the posterior covariance $\boldsymbol{\Sigma}_{m|d_{obs}}$. The posterior covariance is stationary and homogeneous and hence can be represented by six cubes of size n.

A set of possible solutions can be generated by simulation from the posterior distribution. This can be done efficiently in the Fourier domain: For each frequency component k, draw $\tilde{\boldsymbol{m}}_k$ from the posterior distribution, and then transform $\tilde{\boldsymbol{m}}$ to $\boldsymbol{m}$ by an inverse 3-D Fourier transform for each of the three model parameters in $\boldsymbol{m}$. Since $\boldsymbol{m}$ represents the logarithm of the elastic material parameters, see expression (5.3), the corresponding set of simulated solutions of the P-wave velocity, S-wave velocity, and density are obtained by the inverse transform $\exp[\boldsymbol{m}]$. Since the posterior distribution for the model parameters can be represented explicitly by the posterior mean and covariance, the inversion results can be merged with a set of well logs to refine the solution around wells. This can be done both for the conditional mean and the conditional simulations using Kriging, see e.g., Cressie (1991).

The inversion procedure

The inversion procedure is summarized by the following steps :

1. Define the prior model for the model parameters based on the available knowledge, that is $\boldsymbol{\mu}_m(\boldsymbol{x}, t)$, $\boldsymbol{\Sigma}_{0,m}$, and $\nu_m(\boldsymbol{\xi}, \tau)$, see expressions (5.14)-(5.17).
2. Estimate the wavelet $\boldsymbol{s}(t, \theta)$.
3. Estimate the noise covariance, that is $\boldsymbol{\Sigma}_{0,e}$ and $\nu_e(\boldsymbol{\xi}, \tau)$, see expression (5.23).

4. Calculate the discrete 3-D Fourier transform of $\boldsymbol{\mu}_m(\boldsymbol{x}, t)$, $\nu_m(\boldsymbol{\xi}, \tau)$, $\nu_e(\boldsymbol{\xi}, \tau)$, and the 1-D Fourier transform of $\boldsymbol{s}(t, \theta)$. Sort the seismic data $d_{obs}(\boldsymbol{x}, t, \theta)$ into common angle cubes, and 3-D Fourier transform each of these angle cubes to $\tilde{d}_{obs}(\boldsymbol{k}, \omega, \theta)$.

5. For each frequency component k, calculate the posterior expectation $\tilde{\boldsymbol{\mu}}_{m|d_{obs},k}$ and the posterior covariance $\tilde{\boldsymbol{\Sigma}}_{m|d_{obs},k}$, see expressions (5.30) and (5.31).

6. Inverse Fourier transform the solution represented by the posterior mean and covariance.

5.3 Inversion example of Sleipner data

A rectangular portion of a seismic survey from the Sleipner Øst Field is used in this inversion example. This is the same dataset which was used in Buland and Omre (2002a), where a more detailed presentation of the dataset can be found, including seismic processing, prior model definition, wavelet estimation and estimation of the noise covariance. More on wavelet estimation and the estimation of the noise covariance can be found in Buland and Omre (2002b). The main focus in this paper is on the lateral coupling of the model parameters.

The inversion area is defined from inlines 1411 to 1751, and from crosslines 1225 to 1400, covering 9.3 km^2, or 12% of the total survey. Every second line is used, such that $n_x = 176$ and $n_y = 171$. The seismic data set is reduced to three angle stacks, $n_\theta = 3$, representing 9°, 21°, and 33°. The thickness of the target area is 250 ms in two-ways traveltime, such that $n_t = 126$ with sampling interval 2 ms. The time window follows an interpretation of the main layering in this target zone. The corresponding number of frequency components are $n_\omega = 318$, $n_{k_x} = 350$, and $n_{k_y} = 340$, that is 38 million frequency components. Compared to the grid size, this increase is caused by extending the problem to a circulant form, see Appendix 5.A. The grid size is not optimal with respect to fast Fourier transform, so the fast radix-2 Fourier transform can not be applied.

The spatial coupling of the model parameters is imposed through the spatial correlation function in expression (5.16). The lateral correlation of the model parameters is estimated from the seismic data and found to be adequately

fitted by a first order exponential correlation function,

$$\nu_{m,\xi}(\boldsymbol{\xi}) = \exp\left[-\frac{\sqrt{\xi_x^2+\xi_y^2}}{d_r}\right], \tag{5.32}$$

with lateral range $d_r = 250\,\text{m}$. The temporal correlation of the model parameters can not be directly estimated from the seismic data since they are blurred by the seismic wavelet. However, a temporal correlation function can be estimated directly from the well logs, here modeled by the composite correlation function

$$\nu_{m,\tau}(\tau) = \frac{1}{2}\left\{\exp\left[-\left(\frac{\tau}{d_{t_1}}\right)^2\right] + \left(1-\frac{2\tau^2}{d_{t_2}^2}\right)\exp\left[-\left(\frac{\tau}{d_{t_2}}\right)^2\right]\right\}, \tag{5.33}$$

with temporal range parameters $d_{t_1} = 1.8\,\text{ms}$ and $d_{t_2} = 9\,\text{ms}$, see Buland and Omre (2002a). The complete spatial correlation function $\nu_m(\boldsymbol{\xi},\tau)$ is the product of the lateral and the temporal correlation functions defined in expressions (5.32) and (5.33).

The error term $e(\boldsymbol{x},t,\theta)$ in the convolutional model, expression (5.6), includes both seismic noise and errors related to the inversion methodology. We have assumed that the error is zero-mean Gaussian with covariance function on the form given in expression (5.23). The simplest form of the covariance function is obtained for white noise, that is noise with no spatial correlation. However, in seismic inversion, the most serious noise is usually source generated noise, where remaining multiples are an important example. Such noise components have a smooth waveform similar to the waveform of the primary events. The estimated temporal correlation of this noise can be modeled by a scaled second derivative of a second order exponential correlation function,

$$\nu_{e,\tau}(\tau) = \left(1-\frac{2\tau^2}{d_t^2}\right)\exp\left[-\left(\frac{\tau}{d_t}\right)^2\right], \tag{5.34}$$

where the temporal range is estimated to $d_t = 13\,\text{ms}$. Note that this correlation function can be recognized as a Ricker wavelet with center frequency $f_c = 25\,\text{Hz}$, using the relation $d_t = 1/(\pi f_t)$. Further, we model the lateral correlation of the seismic coherent noise with same correlation function as the model parameters, see expression (5.32). A first order exponential correlation function is also used to model the correlation between the different reflection angles, with range estimated to $d_\theta = 10^\circ$. The variance of the coherent noise is estimated to be 0.1, and the variance of the white noise component is estimated

to be 0.01. In the frequency domain, the coherent seismic noise colored by the seismic wavelet is band-limited, lacking the lowest and the highest frequencies. In contrast, the white noise distributes equally to all frequencies.

The complete solution of the inversion is represented by the posterior distribution, defined by the posterior mean and covariance, see expressions (5.30) and (5.31). In Figure 5.1, the P-wave velocity, S-wave velocity, and density corresponding to the exponent of the posterior mean of $\boldsymbol{m}$, $\exp[\boldsymbol{\mu}_{m|d_{obs}}]$, are shown for inline 1627. A well is located at crossline 1291, and the well logs are plotted for comparison, showing good agreement with the inversion results for the velocities. A constant time slice of the P-wave velocity and the S-wave velocity at 2320 ms are shown in Figure 5.2. The real data of inline 1627, the synthetic data computed from the posterior mean solution, and the corresponding residual are shown in Figures 5.3, 5.4, and 5.5 for 9°, 21°, and 33°. It is important to realize that the objective of the inversion is not only to minimize the data residual, but to estimate a solution which honor both the seismic data and our prior knowledge. The residual could have been reduced by erroneously altering the error covariance, e.g., by erroneously reducing the variance or the spatial correlation range.

A set of possible solutions for the P-wave velocity, S-wave velocity, and density can be found by drawing a set of vectors $\tilde{\boldsymbol{m}}$ from the posterior distribution, inverse Fourier transform them to $\boldsymbol{m}$, and then calculate $\exp[\boldsymbol{m}]$. One such simulated solution is shown in Figure 5.6 for inline 1627. The simulated solution differs significantly from the smoother posterior mean solution in Figure 5.1, but both give a good explaination of the real seismic data.

The prior model specifies the prior values for the variances of $\ln\alpha$, $\ln\beta$, and $\ln\rho$. These values are defined on the diagonal of $\boldsymbol{\Sigma}_{0,m}$, see expression (5.17). In this example, the prior variances are estimated from the well logs to be

$$\mathrm{Diag}(\boldsymbol{\Sigma}_{0,m}) = 10^{-4} \times [39, 123, 4]. \tag{5.35}$$

After inversion, the corresponding posterior variances are

$$\mathrm{Diag}(\boldsymbol{\Sigma}_{0,m|d_{obs}}) = 10^{-4} \times [22, 85, 4]. \tag{5.36}$$

The variance of $\ln\alpha$ has the relatively strongest decrease, followed by the variance of $\ln\beta$. The variance of $\ln\rho$ is hardly changed, that means that the inversion does not provide significant new information which reduces the uncertainty about this parameter.

In Figures 5.1 and 5.6, the well logs are plotted for comparison with the inversion result. However, the well logs can be included in a refined solution by Kriging, see e.g. Cressie (1991). This requires specification of a covariance matrix for the well log errors. For simplicity, the covariance matrix is here set to zero, meaning that the well logs are defined to be exact. Including the well logs to the sections in Figures 5.1 and 5.6 by Kriging gives the corresponding updated solutions shown in Figure 5.7 and 5.8. Since the well logs are defined to be exact, the solutions updated by Kriging are equal to the well logs in the well position. The influence of the well logs decreases with increasing distance to the well position, and the uncertainty of the refined solution decreases near the well. The effect of merging the well log information with the seismic inversion results is most distinct for the posterior mean density in Figure 5.7. The reason is that seismic data provides little information about the density, resulting in a smooth posterior mean solution. After merging with the density log, the solution is far more detailed near the well. Also the simulated solutions in Figure 5.8 are updated near the well and are practically unchanged far from the well. The average of a large number of simulated solutions updated by Kriging will approach the solution in Figure 5.7.

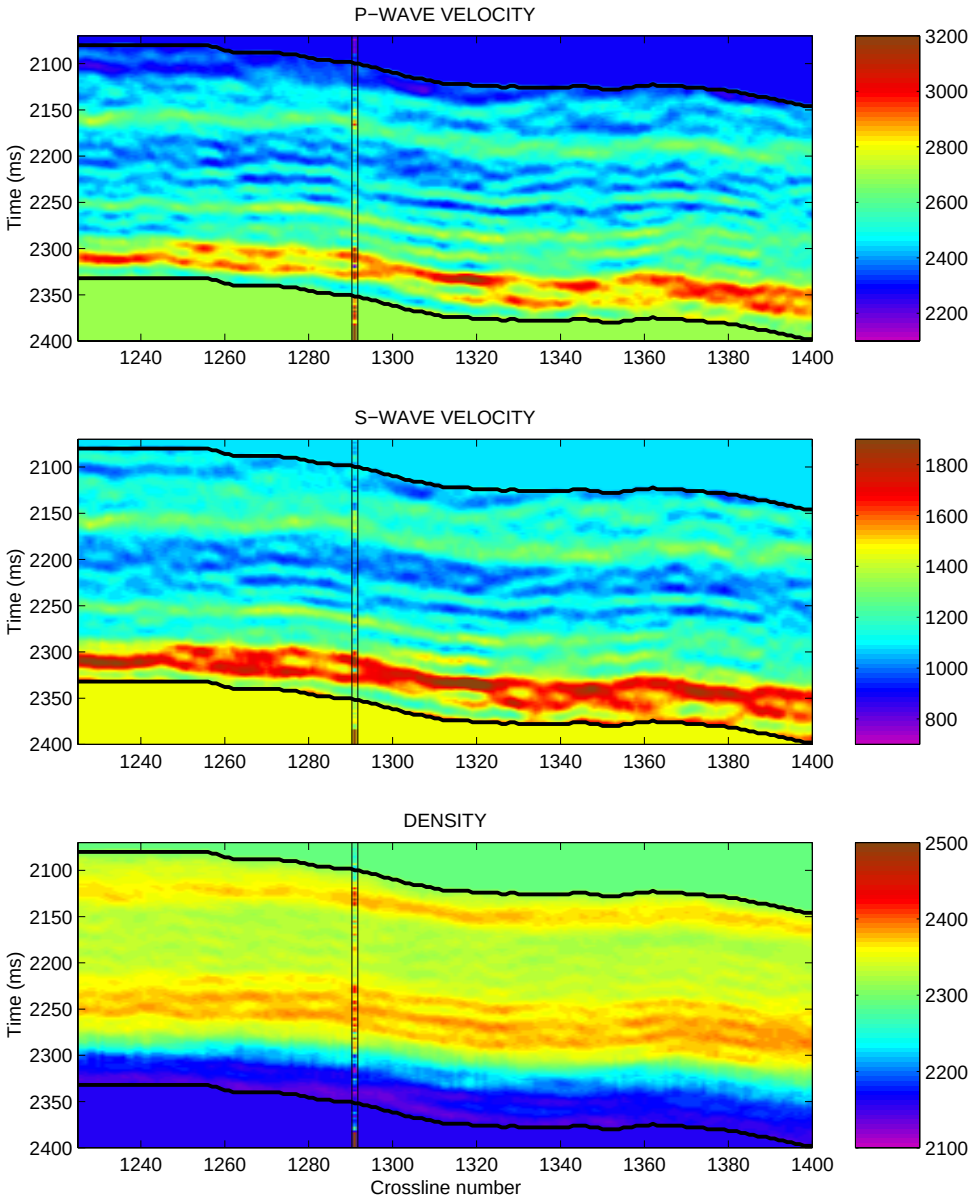


Figure 5.1: P-wave velocity (top), S-wave velocity (middle), and density (bottom) corresponding to the posterior mean.

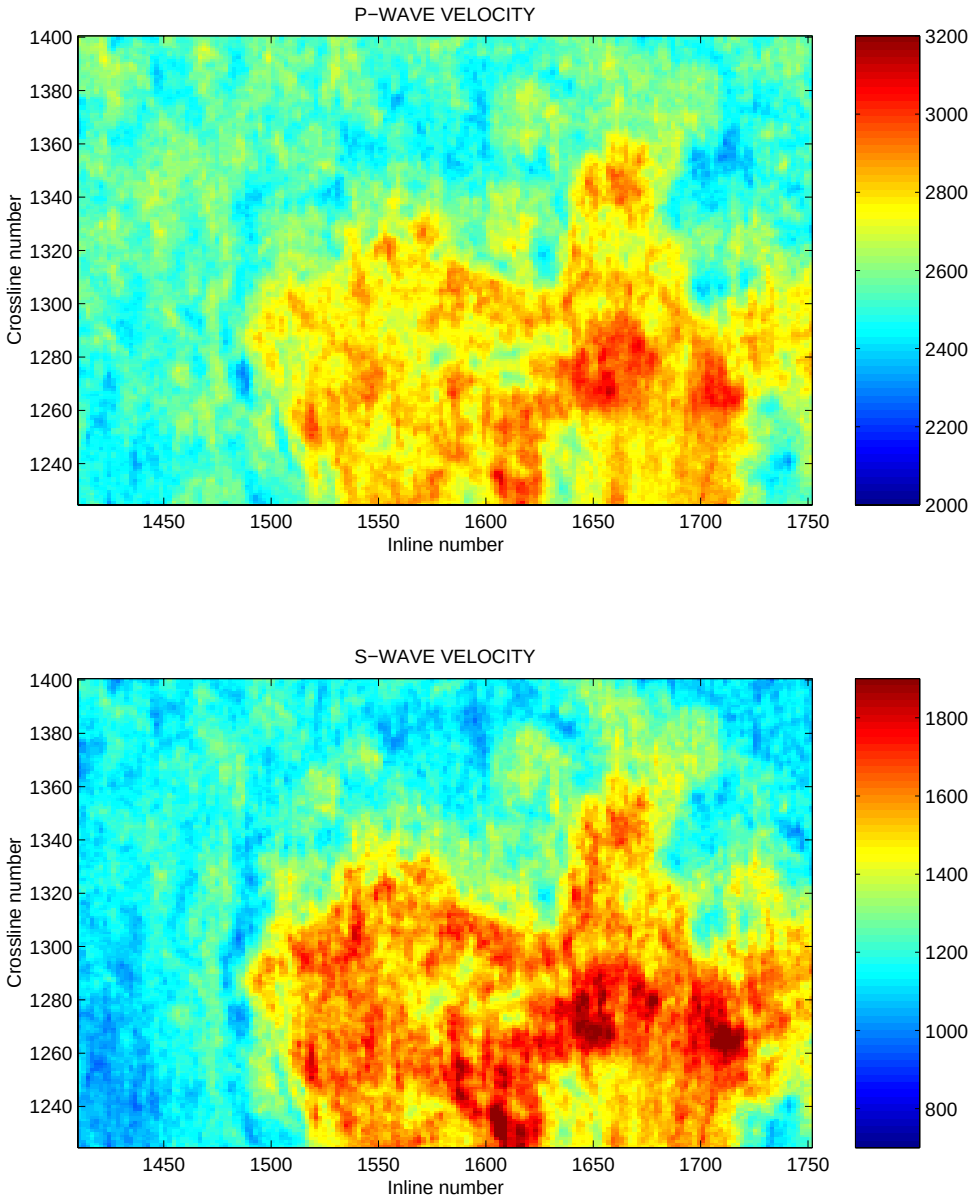


Figure 5.2: Time slice of the P-wave velocity (top) and the S-wave velocity (bottom) at 2320 ms.

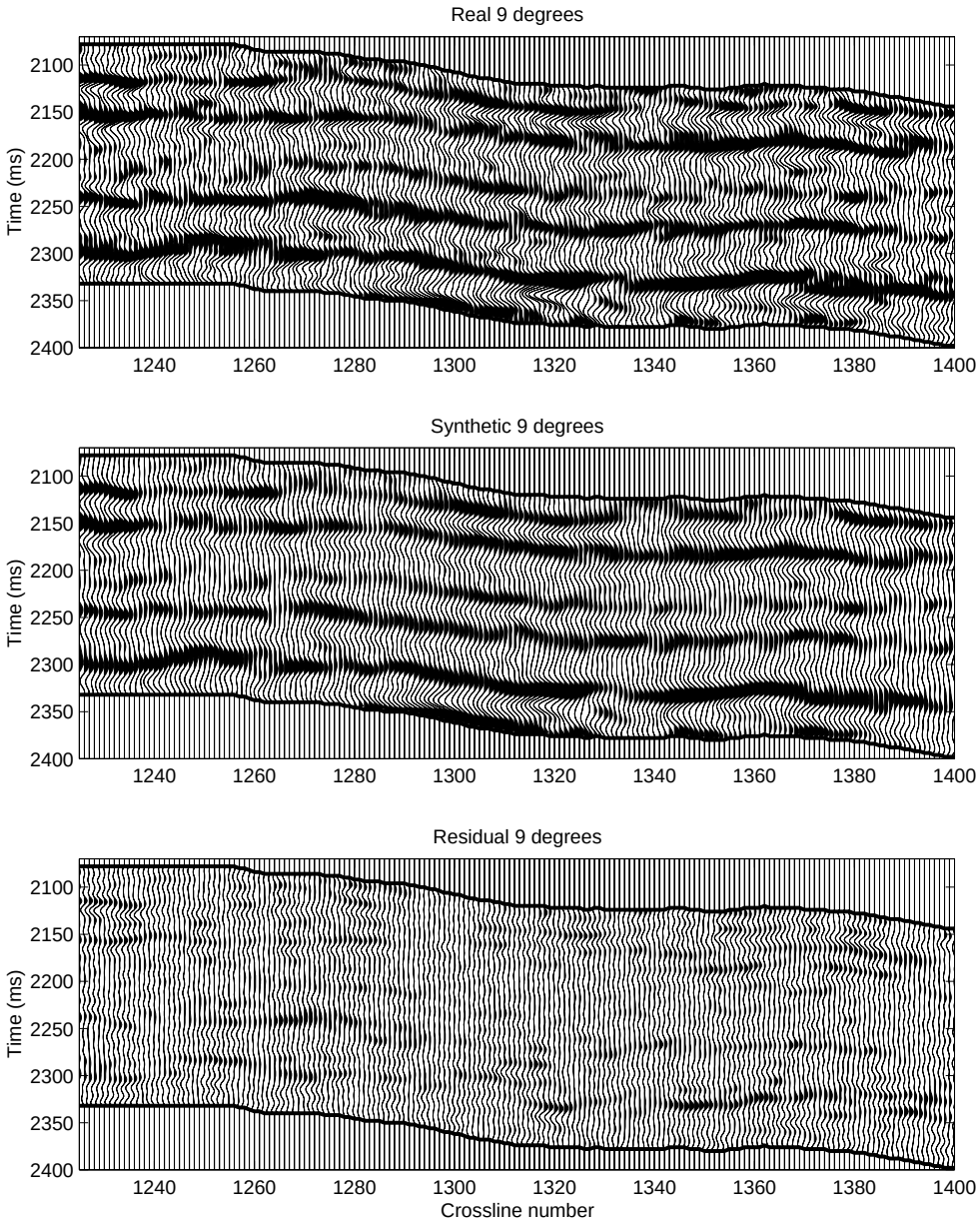


Figure 5.3: Real data (top), synthetic (middle), and residual (bottom) for 9°.

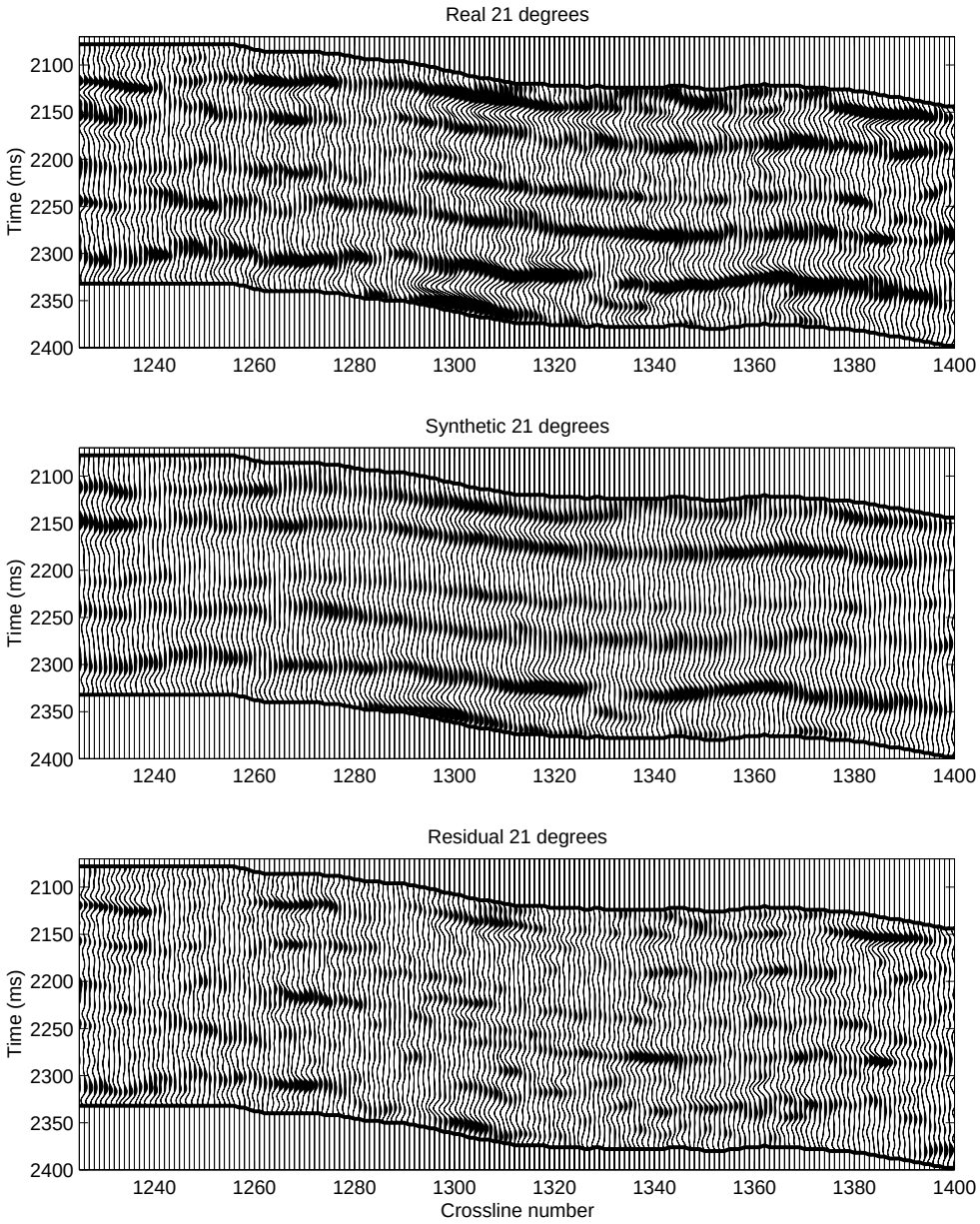


Figure 5.4: Real data (top), synthetic (middle), and residual (bottom) for 21°.

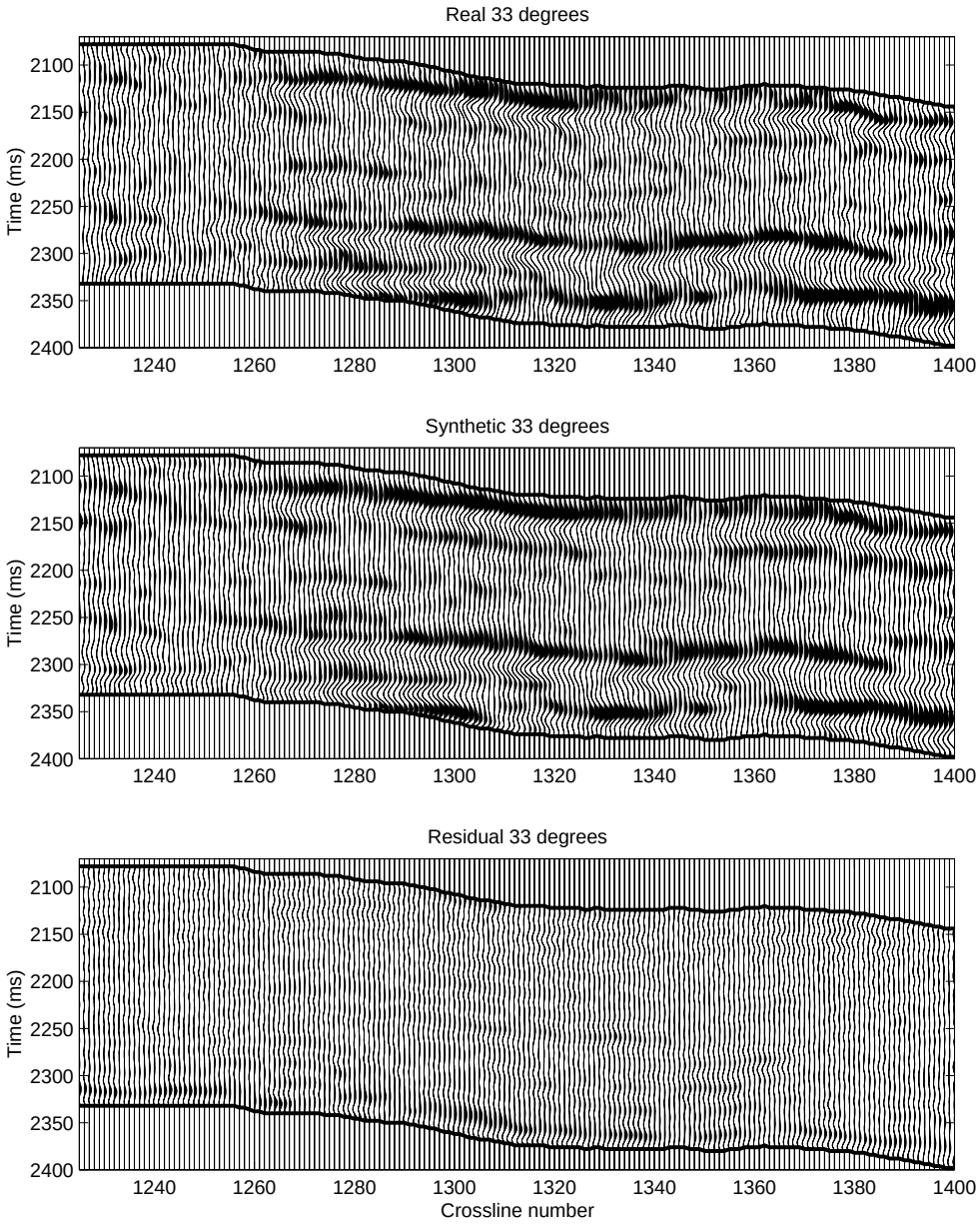


Figure 5.5: Real data (top), synthetic (middle), and residual (bottom) for 33°.

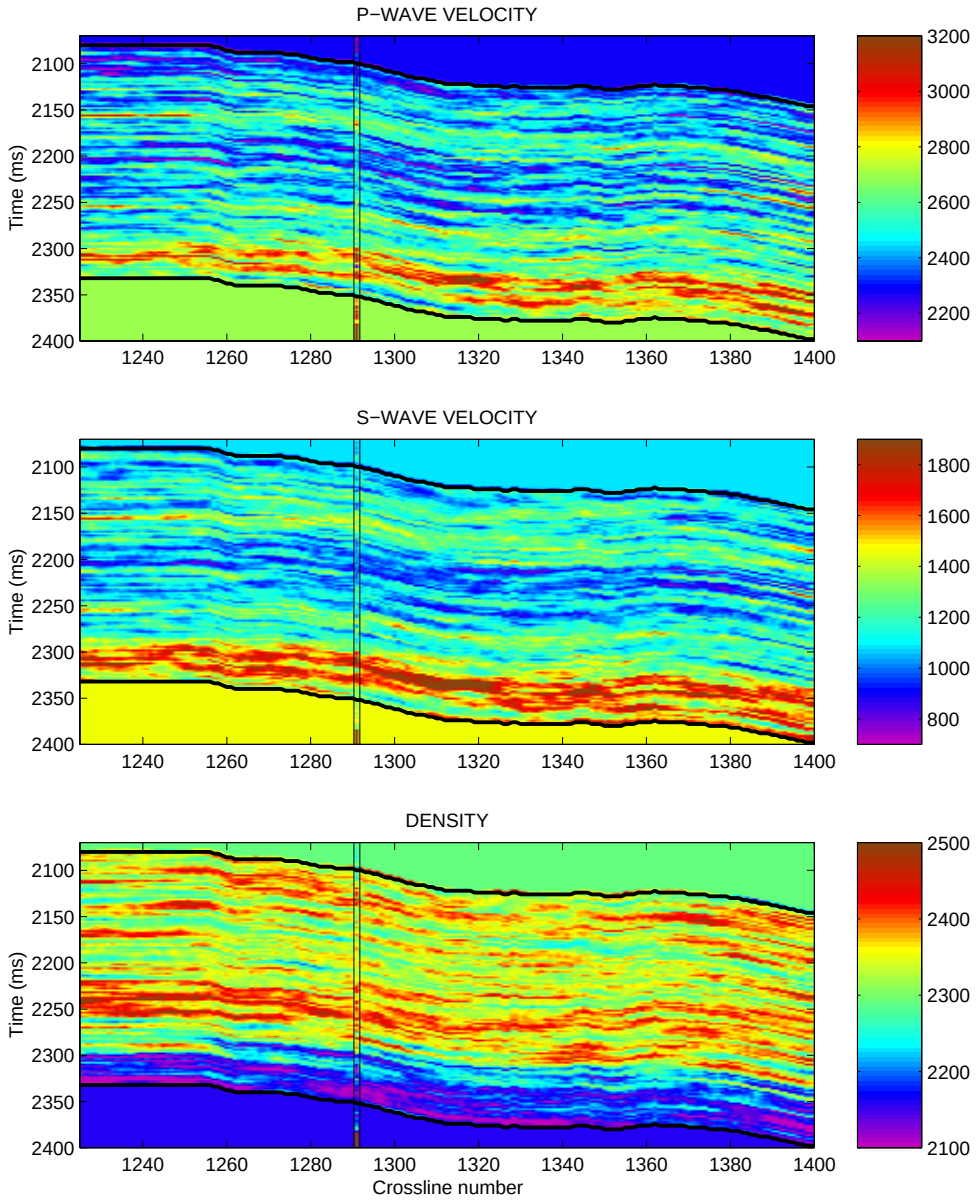


Figure 5.6: Simulated P-wave velocity (top), S-wave velocity (middle), and density (bottom).

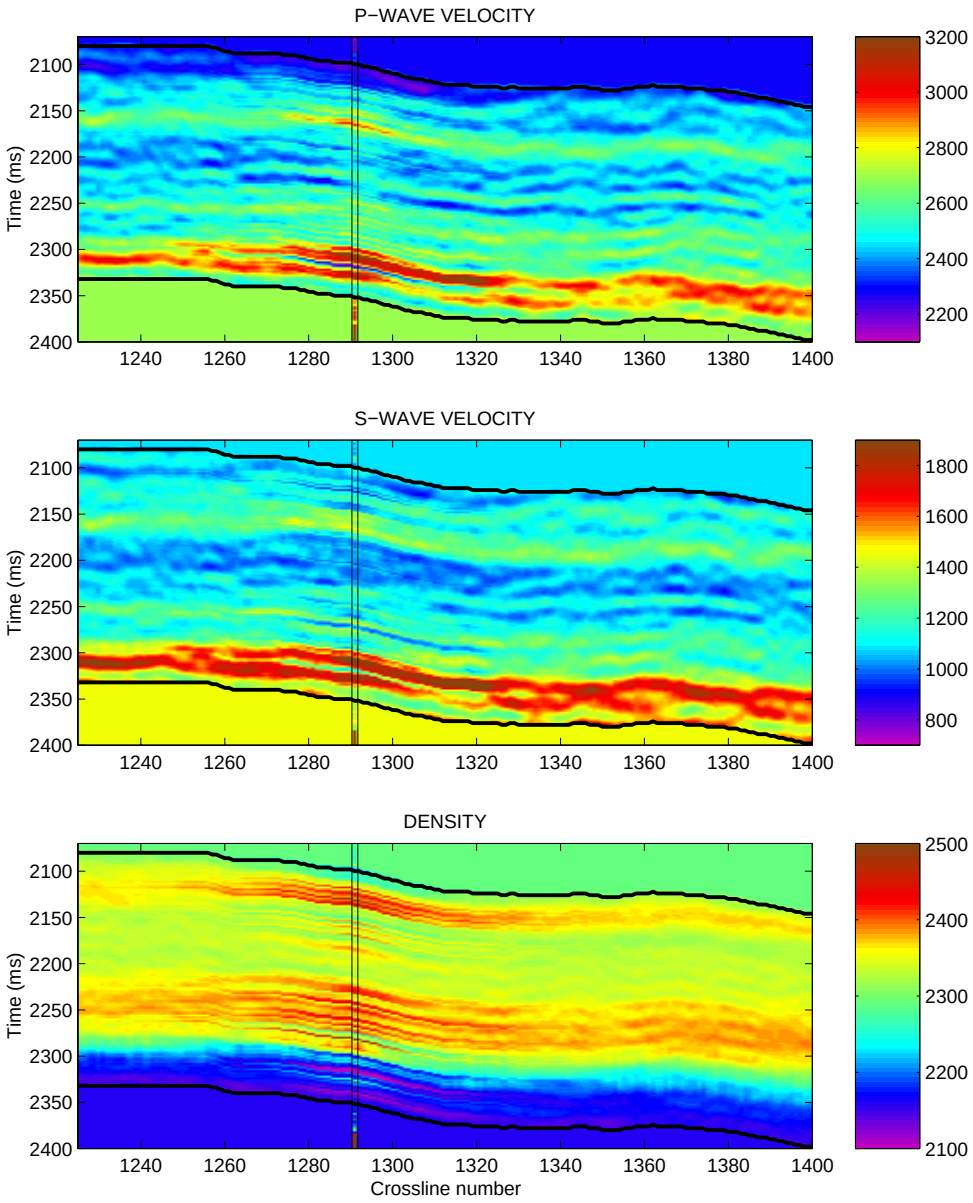


Figure 5.7: The posterior mean solution conditioned to the well logs.

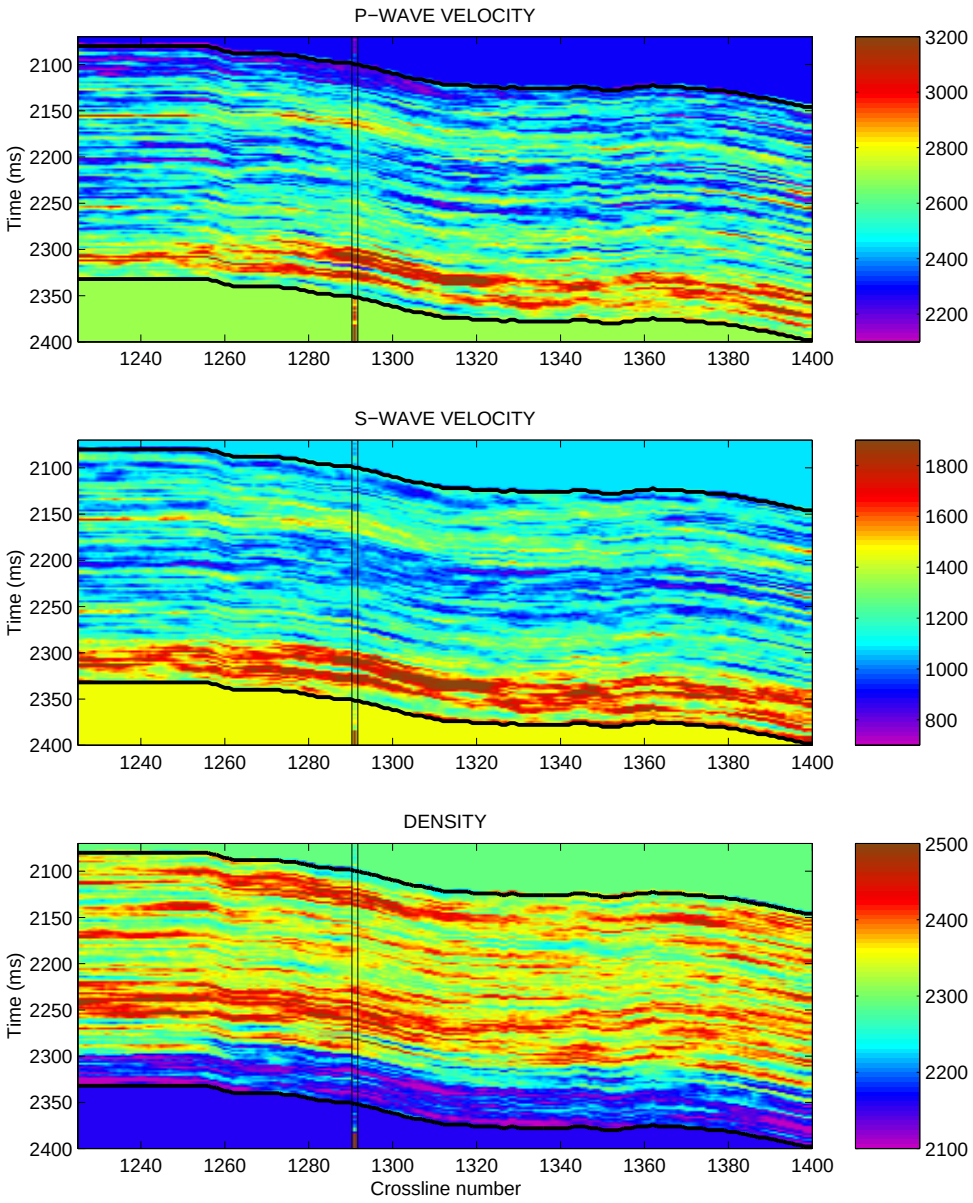


Figure 5.8: The simulated solution conditioned to the well logs.

5.4 Discussion and conclusions

We have developed an efficient AVO inversion technique where the spatially correlated model parameters are decoupled in the Fourier domain. The seismic data and the model parameters are assumed to be represented on an identical, regularly sampled grid. Further, the covariance functions for the model parameters and the data errors are assumed to be homogeneous and stationary (i.e., translationally invariant). When the range of the spatial dependency is shorter than the total spatial extension of the grid, the inversion technique is exact with respect to the spatial coupling.

The solution of the inversion problem is represented by a Gaussian posterior distribution with explicit matrix expressions for the posterior mean and covariance. The posterior mean can be interpreted as a smooth best estimate of the solution, while the posterior covariance contains the uncertainty and the correlation structures of the solution. The posterior covariance is homogeneous and stationary, such that the estimated uncertainty of the solution is equal for all positions $(\boldsymbol{x}, t)$. The uncertainty at the boundary of the inversion area is in general underestimated, most severely at the corners. This problem is related to the assumed symmetry in the spatial coupling of the model parameters. At the boundary of the inversion area, this symmetry is lacking. The thickness of the influenced boundary zone depends on the correlation range.

The computing time for the inversion in the Fourier domain follows a linear function of the total number of grid nodes, $\mathcal{O}(n)$, while the computing time for the fast Fourier transform follows an $\mathcal{O}(n \log n)$ function. A 3-D dataset from the Sleipner Field represented by 3 angle stacks on a grid with 4 million grid cells, each with 3 unknown model parameters, used less than 3 minutes on the inversion on a single 400 MHz Mips R12000 CPU. In addition, each Fourier transform used about 30 seconds, but asymptotically the Fourier transforms will dominate the computing time when n approaches infinity. The inversion method is suitable for parallellization, since the inversion problem can be solved independently for each frequency component. Utilizing that the seismic data are band limited, a further speedup can be obtained by inverting only the significant frequencies.

5.5 Acknowledgments

We thank Statoil for permission to publish this paper.

5.A Diagonalization of a covariance matrix by DFT

In the following, the relationship between the discrete Fourier transform (DFT) and the eigen-values and eigen-vectors of a circulant matrix is presented. Further, it is shown how this can be used to diagonalize a homogeneous covariance function sampled on a regular grid. For simplicity, the presentation is limited to 1-D, but the extension to higher dimensions is straightforward. More details on these topics can be found in Brockwell and Davis (1987); Wood and Chan (1994); Wood (1995).

The DFT

The 1-D discrete Fourier transform (DFT) of the sequence $f(k),\ k = 0, \ldots, n-1$, can be written

$$\tilde{f}(l) = \sum_{k=0}^{n-1} f(k) \exp\left[-2\pi i \frac{kl}{n}\right], \quad l = 0, \ldots, n-1, \tag{5.A.1}$$

with inverse transform (IDFT)

$$f(k) = \frac{1}{n} \sum_{l=0}^{n-1} \tilde{f}(l) \exp\left[2\pi i \frac{kl}{n}\right], \quad k = 0, \ldots, n-1. \tag{5.A.2}$$

The DFT can alternatively be written as a matrix-vector product

$$\tilde{\boldsymbol{f}} = \boldsymbol{F}\boldsymbol{f}, \tag{5.A.3}$$

where $\boldsymbol{f} = [f(0), \ldots, f(n-1)]^T$, and

$$\boldsymbol{F} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & w^1 & \cdots & w^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & w^{n-1} & \cdots & w^{(n-1)^2} \end{bmatrix}, \tag{5.A.4}$$

where $w = \exp[-2\pi i/n]$. The matrix corresponding to the IDFT is $\boldsymbol{F}^{-1} = n^{-1}\boldsymbol{F}^*$, where $*$ denotes the conjugate transpose. If the dimension n is a power of 2, the fast radix-2 Fourier transform (FFT) can be used.

Circulant matrices

An $n \times n$ matrix $\boldsymbol{M}$ is a circulant matrix if the elements m_{kl} are defined by a function $m(\cdot)$ with period n such that $m_{kl} = m_{l-k} = m(l-k)$, that is

$$\boldsymbol{M} = \begin{bmatrix} m_0 & m_1 & \cdots & m_{n-1} \\ m_{n-1} & m_0 & \cdots & m_{n-2} \\ \vdots & \vdots & & \vdots \\ m_1 & m_2 & \cdots & m_0 \end{bmatrix}, \tag{5.A.5}$$

see Brockwell and Davis (1987). Note that a circulant matrix is Toeplitz, but the opposite is generally not true. The eigenvalues of a circulant matrix $\boldsymbol{M}$ are

$$\lambda_l = \sum_{k=0}^{n-1} m(k) \exp\left[-2\pi i \frac{kl}{n}\right], \quad l = 0, \ldots, n-1, \tag{5.A.6}$$

with orthonormal eigenvectors

$$\boldsymbol{v}_l = n^{-1/2} \begin{bmatrix} 1 \\ w^l \\ \vdots \\ w^{(n-1)l} \end{bmatrix}. \tag{5.A.7}$$

The circulant matrix $\boldsymbol{M}$ can be diagonalized by

$$\boldsymbol{V}\boldsymbol{M}\boldsymbol{V}^* = \boldsymbol{\Lambda}_M, \tag{5.A.8}$$

where $\boldsymbol{\Lambda}_M = \mathrm{diag}\{\lambda_0, \lambda_1, \ldots, \lambda_{n-1}\}$, and $\boldsymbol{V}$ is the unitary eigenvector matrix

$$\boldsymbol{V} = [\boldsymbol{v}_0, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}]. \tag{5.A.9}$$

For the following, it is important to recognize that the eigenvalues of a circulant matrix $\boldsymbol{M}$ are equal to the DFT of the first row, and that $\boldsymbol{F} = n^{1/2}\boldsymbol{V}$.

Diagonalization of a circulant covariance matrix

Let $r(x)$ be a zero mean Gaussian variable with homogeneous covariance function

$$\Sigma(x_1, x_2) = \sigma^2 \nu(\xi), \tag{5.A.10}$$

where $\xi = |x_2 - x_1|$. Let $\boldsymbol{r}$ be a discrete representation of $r(x)$ sampled on a regular grid, $x_k = k\Delta x$, where $k = 0, \ldots, n_x - 1$. The corresponding $n_x \times n_x$ covariance matrix is then symmetric Toeplitz,

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} \nu_0 & \nu_1 & \cdots & \nu_{n_x-1} \\ \nu_1 & \nu_0 & \cdots & \nu_{n_x-2} \\ \vdots & \vdots & & \vdots \\ \nu_{n_x-1} & \nu_{n_x-2} & \cdots & \nu_0 \end{bmatrix}, \tag{5.A.11}$$

where $\nu_k = \nu(k\Delta x)$. This covariance matrix is not circulant, but it can be embedded in a symmetric circulant $n \times n$ matrix

$$\boldsymbol{\Sigma}_c = \sigma^2 \begin{bmatrix} \nu_0 & \nu_1 & \cdots & \nu_{n/2} & \cdots & \nu_1 \\ \nu_1 & \nu_0 & \cdots & \nu_{n/2-1} & \cdots & \nu_2 \\ \vdots & \vdots & & \vdots & & \vdots \\ \nu_{n/2} & \nu_{n/2-1} & \cdots & \vdots & & \nu_{n/2-1} \\ \vdots & \vdots & & \vdots & & \vdots \\ \nu_1 & \nu_2 & \cdots & \nu_{n/2-1} & \cdots & \nu_0 \end{bmatrix}, \tag{5.A.12}$$

where $n \geq 2(n_x - 1)$, and such that the top left $n_x \times n_x$ sub matrix of $\boldsymbol{\Sigma}_c$ is equal to $\boldsymbol{\Sigma}$. The circulant matrix $\boldsymbol{\Sigma}_c$ is a legal covariance matrix if and only if it is positive definite. A sufficient, but not neccessary condition for positive definiteness is that $\nu_k = 0$ for all $k > k_0$, where $k_0 < n_x$, see Wood (1995). Strictly, this excludes many of the most common correlation functions, for example exponential correlation functions with order $(1, 2]$. However, the range of a correlation function will often be much shorter than the total spatial dimension, such that $\nu_k \approx 0$ for all $k > k_0$. In such cases a truncation of the correlation function may be adequate.

Let now $\boldsymbol{r}_c$ be an extension of $\boldsymbol{r}$ with dimension n and covariance matrix $\boldsymbol{\Sigma}_c$. While $\boldsymbol{r}$ is sampled on a line, $\boldsymbol{r}_c$ can be interpreted to be sampled on a circle. Then the Fourier transform of $\boldsymbol{r}_c$, $\tilde{\boldsymbol{r}}_c = \boldsymbol{F}\boldsymbol{r}_c$, has a diagonal covariance matrix

$$\tilde{\boldsymbol{\Sigma}}_c = \boldsymbol{F}\boldsymbol{\Sigma}_c\boldsymbol{F}^* = n\boldsymbol{V}\boldsymbol{\Sigma}_c\boldsymbol{V}^{-1} = n\boldsymbol{\Lambda}_\Sigma = \tilde{\boldsymbol{\Lambda}}_\Sigma, \tag{5.A.13}$$

where $\mathbf{\Lambda}_\Sigma$ is the eigenvalue matrix of $\mathbf{\Sigma}_c$ with real nonnegative eigenvalues. This means that the correlated variables in $\boldsymbol{r}$ are transformed to independent variables in the Fourier domain. From above, we know that $\mathbf{\Lambda}_\Sigma$ can simply be calculated by a DFT of the first row of $\mathbf{\Sigma}_c$. This means that it is not necessary to compute the matrix products $\boldsymbol{F}\mathbf{\Sigma}_c\boldsymbol{F}^*$. In fact, the complete matrix $\mathbf{\Sigma}_c$ is not involved in the computations.

The extension to 2-D and 3-D problems is straightforward. Let $\boldsymbol{r}$ be a discrete representation of a zero mean Gaussian variable with homogeneous covariance function, sampled on a regular 2-D or 3-D grid. The corresponding covariance matrices are block Toeplitz in 2-D and nested block Toeplitz in 3-D, and they can be embedded in block or nested block circulant matrices. Similarly to the 1-D case, the n eigenvalues can be found by a 2-D or 3-D DFT of a circulant discrete representation of the correlation function.

Bibliography

Aki, K., and Richards, P. G., 1980, Quantitative seismology: W.H. Freeman and Co.

Anderson, T. W., 1984, An introduction to multivariate statistical analysis: John Wiley and Sons Inc.

Bortfeld, R., 1961, Approximation to the reflection and transmission coefficients of plane longitudinal and transverse wave: Geophys. Prosp., **9**, 485–502.

Brockwell, P. J., and Davis, R. A., 1987, Time series: Theory and methods: Springer Verlag.

Brown, R. L., 1992, Estimation of AVO in the presence of noise using prestack migration: 62nd Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 855–888.

Buland, A., and Landrø, M., 2001, The impact of common offset migration on porosity estimation by AVO inversion: Geophysics, **66**, 755–762.

Buland, A., and Omre, H., 2002a, Bayesian linearized AVO inversion: Accepted for publication in Geophysics. Chapter 2 in this thesis.

——— 2002b, Bayesian wavelet estimation from seismic and well data: Submitted for publication, March 2001. Chapter 3 in this thesis.

Buland, A., Landrø, M., Andersen, M., and Dahl, T., 1996, AVO inversion of Troll Field data: Geophysics, **61**, 1589–1602.

Buland, A., Kolbjørnsen, O., and Omre, H., 2002, Rapid spatially coupled AVO inversion in the Fourier domain: Submitted for publication, April 2002. Chapter 5 in this thesis.

Carcione, J., Kosloff, D., and Kosloff, R., 1988, Viscoacoustic wave propagation simulation in the earth: Geophysics, **53**, 769–777.

Carlin, B. P., 1996, Hierarchical longitudinal modelling, *in* Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., Eds., Markov Chain Monte Carlo in practice: Chapman & Hall, 303–319.

Červený, V., Molotkov, I., and Psencik, I., 1977, Ray methods in seismology: Univerzita Karlova.

Chen, M., Shao, Q., and Ibrahim, J., 2000, Monte Carlo methods in Bayesian computation: Springer.

Christakos, G., 1992, Random field models in earth sciences: Academic Press Inc., San Diego.

Cressie, N., 1991, Statistics for spatial data: John Wiley and Sons Inc.

Dahl, T., and Ursin, B., 1991, Parameter estimation in a one-dimensional anelastic medium: J. Geophys. Res., **96**, 20217–20233.

Danielson, V., and Karlsson, T. V., 1984, Extraction of signatures from seismic and well data: First Break, **2**, no. 4, 15–21.

Duijndam, A. J. W., 1988a, Bayesian estimation in seismic inversion. Part I: Principles: Geophys. Prosp., **36**, 878–898.

——— 1988b, Bayesian estimation in seismic inversion. Part II: Uncertainty analysis: Geophys. Prosp., **36**, 899–918.

Eide, A. L., Omre, H., and Ursin, B., 2002, Prediction of reservoir variables based on seismic data and well observations: J. Am. Statistical Assn., **97**, no. 457, 18–28.

Emmerich, H., and Korn, M., 1987, Incorporation of attenuation into time-domain computations of seismic wavwfields: Geophysics, **52**, 1252–1264.

Geman, S., and Geman, D., 1984, Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images: IEEE Trans. Pattern. Anal. Mach. Intell., **6**, 721–741.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., 1996, Markov Chain Monte Carlo in practice: Chapman & Hall.

Goldberg, D. E., 1989, Genetic algorithms in search, optimization, and machine learning: Addison-Wesley.

Gouveia, W. P., and Scales, J. A., 1997, Resolution of seismic waveform inversion: Bayes versus Occam: Inverse Problems, **13**, 323–349.

Gouveia, W. P., and Scales, J. A., 1998, Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis: J. Geophys. Res., **103**, 2759–2779.

Hampson, D., and Russell, B., 1990, AVO inversion: theory and practice: 60th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1456–1458.

Hastings, W. K., 1970, Monte Carlo sampling methods using Markov chains and their applications: Biometrika, **57**, 97–109.

Kennett, B. L. N., 1983, Seismic wave propagation in stratified media: Cambridge Univ. Press.

Kirkpatrick, S. C., Gelatt, D., and Vecchi, M. P., 1983, Optimization by simulated annealing: Science, **220**, 671–780.

Kolbjørnsen, O., 2002, Nonlinear topics in the Bayesian approach to inverse problems with applications to seismic inversion: Ph.D. thesis, Norwegian University of Science and Technology.

Lauritzen, S. L., 1996, Graphical models: Oxford University Press.

Levander, A. R., 1988, Fourth-order finite-difference $P - SV$ seismograms: Geophysics, **53**, 1425–1436.

Lines, L. R., and Treitel, S., 1984, Tutorial: A review of least-squares inversion and its application to geophysical problems: Geophys. Prosp., **32**, 159–186.

Lines, L. R., and Treitel, S., 1985, Wavelets, well logs and Wiener filters: First Break, **3**, no. 8, 9–14.

Lörtzer, G. J. M., and Berkhout, A. J., 1993, Linearized AVO inversion of multicomponent seismic data, *in* Castagna, J., and Backus, M., Eds., Offset-dependent reflectivity - theory and practice of AVO analysis: Soc. Expl. Geophys., 317–332.

Lysmer, J., and Drake, L. A., 1972, A finite element method for seismology, *in* Alder, B., Fernbach, S., and Bolt, B. A., Eds., Methods in computional physics 11, Seismology: Academic Press.

Madariaga, R., 1986, Dynamics of an expanding circular fault: Bull., Seis. Soc. Am., **66:3**, 639–666.

Malinverno, A., 2000, A Bayesian criterion for simplicity in inverse problem parametrization: Geophys. J. Int., **140**, 267–285.

Marfurt, K. J., 1984, Seismic modeling: A frequency-domain finite element approach: 54th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts.

Metropolis, N., Rosenbluth, M. N., Rosenbluth, A. W., Teller, A. H., and Teller, E., 1953, Equation of state calculations by fast computing machines: J. Chem. Phys., **21**, 1087–1092.

Mosegaard, K., and Tarantola, A., 1995, Monte Carlo sampling of solutions to inverse problems: J. Geophys. Res., **100**, 12431–12447.

Mosegaard, K., 1998, Resolution analysis of general inverse problems through inverse Monte Carlo sampling: Inverse problems, **14**, 405–426.

Mosher, C. C., Keho, T. H., Weglein, A. B., and Foster, D. J., 1996, The impact of migration on AVO: Geophysics, **61**, 1603–1615.

Nyman, D. C., Parry, M. J., and Knight, R. D., 1987, Seismic wavelet estimation using well control: 57th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 535–537.

Omre, H., and Tjelmeland, H., 1997, Petroleum geostatistics *in* Baafi, E. Y., and Schofield, N. A., Eds., Geostatistics Wollongong '96:: Kluwer Acad. Publ., 41–52.

Omre, H., Sølna, K., and Tjelmeland, H., 1993, Simulation of random functions on large lattices, *in* Soares, A., Ed., Geostatistics Tróia '92: Kluwer Academic Publisher, 179–199.

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

Pan, G. S., Young, C. Y., and Castagna, J. P., 1994, An integrated target-oriented prestack elastic waveform inversion: Sensitivity, calibration, and application: Geophysics, **59**, 1392–1404.

Poggiagliolmi, E., and Allred, R. D., 1994, Detailed reservoir definition by integration of well and 3-D seismic data using space adaptive wavelet processing: The Leading Edge, **13**, no. 7, 749–754.

Richard, V., and Brac, J., 1988, Wavelet analysis using well log information: 58th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 946–949.

Richards, P. G., and Frasier, C. W., 1976, Scattering of elastic waves from depth dependent inhomogeneities: Geophysics, **41**, 441–458.

Ripley, B. D., 1997, Stochastic simulation: John Wiley and Sons Inc.

Robert, C. P., and Casella, G., 1999, Monte Carlo statistical methods: Springer-Verlag.

Robert, C. P., 1994, The Bayesian choice: Springer-Verlag.

Rothman, D. H., 1986, Automatic estimation of large residual static corrections: Geophysics, **51**, 332–346.

Rue, H., 2000, Fast sampling of Gaussian Markov random fields with applications: Submitted for publication, http://www.math.ntnu.no/preprint/-statistics/2000.

Scales, J. A., and Snieder, R., 1998, What is noise?: Geophysics, **63**, 1122–1124.

Scales, J. A., and Snieder, R., 2000, The anatomy of inverse problems: Geophysics, **65**, 1708–1710.

Scales, J. A., and Tenorio, L., 2001, Prior information and uncertainty in inverse problems: Geophysics, **66**, 389–397.

Sen, M. K., and Stoffa, P. L., 1996, Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion: Geophys. Prosp., **44**, 313–350.

Smith, G. C., and Gidlow, P. M., 1987, Weighted stacking for rock property estimation and detection of gas: Geophys. Prosp., **35**, 993–1014.

Spiegelhalter, D. J., Best, N. G., Gilks, W. R., and Inskip, H., 1996, Hepatitis B: A case study in MCMC methods, *in* Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., Eds., Markov Chain Monte Carlo in practice: Chapman & Hall, 21–43.

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

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

Tarantola, A., 1987, Inverse problem theory: Elsevier Science Publ. Co. Inc.

Tarantola, A., 1988, Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation: Pure and Appl. Geophys., **128**, 365–399.

Tjelmeland, H., and Eidsvik, J., 2002, On the use of local optimisations within Metropolis-Hastings: http://www.math.ntnu.no/preprint/statistics/2002.

Tjelmeland, H., and Hegstad, B. K., 2001, Mode jumping proposals in MCMC: Scandinavian Journal of Statistics, **28**, 205–223.

Ulrych, T. J., Sacchi, M. D., and Woodbury, A., 2001, A Bayes tour of inversion: a tutorial: Geophysics, **66**, 55–69.

Ursin, B., 1983, Review of elastic and electromagnetic wave propagation in layered media: Geophysics, **48**, 1063–1081.

Ursin, B., 1987, The plane-wave reflection and transmission response of a vertically inhomogeneous acoustic medium *in* Bernabini, M., Carrion, P., Jacovitti, G., Rocca, F., Treitel, S., and Worthington, M. H., Eds., Deconvolution and inversion:: Blackwell Scientific Publications, Inc.

Virieux, J., 1984, SH-wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **49**, 1933–1942.

Virieux, J., 1986, $P - SV$ wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, 889–901.

Wood, A. T. A., and Chan, G., 1994, Simulation of stationary Gaussian processes in $[0, 1]^d$: Jour. of Computational and Graphical Statistics, **3**, 409–432.

Wood, A. T. A., 1995, When is a truncated covariance function on the line a covariance function on the circle?: Statistics & probability letters, **24**, 157–164.

Ziolkowski, A., 1991, Why don't we measure seismic signatures?: Geophysics, **56**, 190–201.

Zoeppritz, K., 1919, Über Reflexion und Durchgang seismischer Wellen durch Unstetigkeitsflächen: Über Erdbebenwellen, VII B, Nachrichten Göttingen, Matematisch physikalische Klasse, pages 57–84.