NTNU – Trondheim
Norwegian University of
Science and Technology

Parametric Wavelet Estimation

Jacob Skauvold

Master of Science in Physics and Mathematics
Submission date: July 2014
Supervisor: Jo Eidsvik, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

A method for parametric estimation of seismic wavelets from well logs and seismic data is developed. Parameters include amplitude, skewness, length and fluctuation order, and the link between parameters and wavelet properties provides a user-friendly interpretation of the wavelet function. The method is set in a Bayesian framework, and is well-suited for addressing questions about uncertainty related to estimated wavelets. This is accomplished by sampling the posterior distribution using Markov Chain Monte Carlo methods. The estimation method is framed as a practical stepwise procedure. An extension of the model to enable joint wavelet estimation from seismic data with multiple incidence angles, is also described.

The method is tested on simulated data, and on well log and seismic amplitude data from the North Sea. The results in the synthetic case indicate that the method performs well under idealised conditions. When tested on real data, the method produces a realistic wavelet fit and uncertainty range. Uncertainty is substantially reduced from the prior to the posterior distribution, but in general, the shape of the posterior surface could make it hard to explore. A comparison with a wavelet estimator based on a Gaussian process indicates that the proposed parametric form gives a tighter wavelet, and is less prone to overfitting.

Sammendrag

En metode utarbeides for å estimere seismiske wavelets parametrisk fra brønnlogger og seismiske data. Metoden beskrives i et Bayesiansk rammeverk, og egner seg for å karakterisere usikkerhet tilknyttet estimerte wavelets. Dette gjennomføres ved at Markov Chain Monte Carlo-metoder tas i bruk for å trekke realisasjoner tilfeldig fra en aposteriorifordeling. Estimeringsmetoden presenteres som en praktisk stegvis prosedyre. I tillegg beskrives en modellutvidelse som gjør det mulig å estimere flere wavelets simultant utfra seismiske data med ulike innfallsvinkler.

Metoden testes på simulerte data, og på brønnlogger og amplitude-seismikk fra Nordsjøen. Resultatene fra syntetiske tester viser at metoden fungerer godt under ideelle forhold. I tester på reelle data gir metoden et realistisk estimat og usikkerhetsspenn. En markant reduksjon av usikkerhet registreres fra apriorifordelingen til aposteriorifordelingen. Generelt kan formen til aposteriorifordelingen gjøre den vanskelig å utforske. Sammenlikning med en alternativ wavelet-estimator basert på en Gaussisk prosess, indikerer at den parametriske modellen gir en mer kompakt wavelet, og er mindre utsatt for overtilpasning.

Acknowledgements

This thesis is submitted in fulfilment of the requirements for the Industrial Mathematics M.Sc. program at the Norwegian University of Science and Technology in Trondheim. The work was carried out at the Department of Mathematical Sciences in the spring semester of 2014.

I would like to thank my supervisor, Jo Eidsvik, for excellent guidance, inspiring discussions and helpful comments. I would also like to thank Ulrich Theune at Statoil Research Centre, Rotvoll, for sharing his geophysical expertise, and for offering advice and very useful feedback throughout the entire writing process. Thanks also to Statoil for providing data for test cases.

Trondheim, July 2014

Jacob Skauvold

Contents

1 Introduction

Wavelets play a central role in seismic data interpretation, because they link geology, represented by reflection coefficients, to seismic data, or traces. Since they are generally unknown, accurate estimates of wavelets are a prerequisite for reliable interpretation and imaging alike.

Without going into details about notation at this stage, the relationship between reflection coefficients, wavelets and seismic data can be expressed through the convolutional model

$$y = r * w + \epsilon, \tag{1}$$

where y is a seismic trace, w is the wavelet, r is a reflectivity series, or reflection coefficient series, and ϵ is a residual noise term. In other words, the seismic trace is modelled as the convolution of reflectivity, which depends on mass density and seismic wave velocity, with the wavelet. Any mismatch between the trace and this convolution is accounted for by a residual term. This could be random noise or processing and imaging artefacts of physical processes which are not explained by the convolutional model (1). Although y could be a raw field trace, and w the concomitant field wavelet, this work will consider the case where y is a processed trace, and w is a corresponding, unknown wavelet to be estimated from the data, which consists of y and the reflectivity series r.

This first section will define the wavelet estimation problem and outline the proposed estimation method, introducing necessary notation. Subsequent sections will discuss the estimation procedure in detail, and present results of numerical tests on real and synthetic data, to document the performance of the method.

1.1 Description of Problem

The goal is to obtain an estimate $\hat{w}(t)$ of the seismic wavelet as a function of time t, and to assess the uncertainty accompanying the estimate, using the available data. There will either be a single trace $y(t)$, or several traces $y_1(t)$, $y_2(t)$, ..., typically corresponding to different stacks, and associated with different angles of incidence. When considering the multi-stack case, we will assume throughout that there are three stacks; near, mid and far.

Also included in the data are well logs containing borehole measurements of mass density and P- and S-wave velocities, or equivalently, slownesses. From these, a time series $r(t)$ of reflection coefficients can be constructed for any given angle, using the Zoeppritz equations, or a suitable approximation of them (Aki and Richards, 1980).

Given $y(t)$ and $r(t)$, determining $w(t)$ is essentially a matter of deconvolving (1). The general strategy in wavelet estimation is to choose $w(t)$ by matching the convolution $r(t) * w(t)$ to the trace $y(t)$.

1.2 Existing solutions, literature review

Estimating the wavelet from seismic data alone, i.e. without well data, is a particularly ill-conditioned problem, which can be solved, for instance, by posing it as underdetermined, and looking for a solution which minimises carefully chosen norms of the parameters involved (Routh et al., 2003). Introducing well logs opens up a wider range of approaches to the wavelet estimation problem, such as cepstral stacking, where the wavelet and reflectivity are separated by computing the complex cepstrum of the seismic trace, or cumulant matching, a non-probabilistic optimisation-based technique (Ulrych et al., 1995). Perhaps the most widely used method for estimating seismic wavelets from seismic and well data is the one proposed by Walden and White (1998), which works by matching the convolution of the wavelet and reflectivity series to the seismic trace in the frequency domain. This makes the estimation very fast, but without a probabilistic setting it is not obvious how the method should be extended to handle situations where one wishes to simultaneously estimate multiple wavelets from seismic data with different incidence angles, gathered from different locations, or even at different times. The approach discussed here operates in the time domain and is built around a Bayesian framework. Thus it is more akin to the hierarchical model of Buland and Omre (2003) which uses a Gaussian process prior model for the wavelet, and where the posterior distribution

is explored using MCMC methods. Also similar is the comprehensive estimation procedure due to Gunning and Glinsky (2006), which searches for the best-fitting wavelet using numerical optimisation, and tackles the issue of determining wavelet length by framing it as a model selection problem. Aune et al. (2013) suggested a parametric form for the wavelet, and the current work adjusts this formulation, and frames it in a practical algorithm.

1.3 Proposed solution, outline of method

We represent the wavelet $w(t)$ by an analytical function $w(t;\boldsymbol{\beta})$ parametrised by a small number of parameters $\boldsymbol{\beta}$. Each of these parameters has an interpretable effect, such as controlling the amplitude or duration of the wavelet. This wavelet function, which is continuous in time, as well as having other desirable properties (to be described below), gives us a parsimonious representation of the wavelet. The wavelet function will be introduced in detail in section 3.

In addition to the wavelet parameters, we define a fourth parameter related to the magnitude of the residual term in (1), and we form a vector $\boldsymbol{\beta}$ from all the parameters. This vector then comprises all unknown parameters in the model, and the wavelet function can be estimated via the parameter vector.

All inference about $\boldsymbol{\beta}$ relies on the posterior distribution $\pi(\boldsymbol{\beta}|\boldsymbol{y})$ of the parameters conditional on all the data. For ease of notation, y refers to all the data in this case, i.e. both seismic data and well logs or reflection coefficient series. By the definition of conditional probability, this is (Gamerman and Lopes, 2006)

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{\beta},\boldsymbol{y})}{\pi(\boldsymbol{y})} \propto \pi(\boldsymbol{\beta},\boldsymbol{y}) \propto \underbrace{\pi(\boldsymbol{\beta})}_{\text{Prior}}\ \underbrace{\pi(\boldsymbol{y}|\boldsymbol{\beta})}_{\text{Likelihood}}. \tag{2}$$

The prior probability distribution of $\boldsymbol{\beta}$ is set using a spectrum smoothing technique to extract a tentative wavelet estimate from the amplitude spectrum of the seismic data. The data likelihood $\pi(\boldsymbol{y}|\boldsymbol{\beta})$ is defined using the parametric wavelet function, and the convolutional model (1). Maximising the posterior distribution with respect to $\boldsymbol{\beta}$ gives a maximum a posteriori (MAP) estimate of the parameters $\boldsymbol{\beta}$ and, by extension, the wavelet $w(t)$. Uncertainty associated with the estimate can be captured by exploring the posterior distribution of $\boldsymbol{\beta}$ around the mode, which is accomplished by sampling, using a Markov Chain Monte Carlo methods.

2 Geophysical model

The likelihood in the statistical deconvolution approach explored in this thesis, is based on the convolutional model of seismic wave propagation. It describes the recorded seismic trace as a convolution of the reflectivity series with a wavelet. The convolution operation has the effect of blurring or averaging out the details in the reflectivity, so that the seismic trace is much smoother than the reflectivity. Figure 1 illustrates the relation between elastic parameters, reflectivity, the wavelet and the trace.

2.1 Convolutional wave propagation model

To define the likelihood for a trace y, given the reflectivity series r and the wavelet w, we begin by introducing the discrete representation $\boldsymbol{y}$ of the trace. This is a vector containing discrete and equidistant samples of the continuous function $y(t)$. Similarly, the reflectivity function $r(t)$ is represented by the vector $\boldsymbol{r}$, and the wavelet function w is represented by the vector $\boldsymbol{w}$. All vectors are discretised with the same sampling time interval Δt. Details of the discretisation are provided in section 4.

The data likelihood for y given r and w, is defined by rewriting (1) with vector notation

$$\boldsymbol{y} = \boldsymbol{r} * \boldsymbol{w} + \boldsymbol{\epsilon} \tag{3}$$

and prescribing a Gaussian distribution for the residual term $\boldsymbol{\epsilon}$,

$$\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 I). \tag{4}$$

Hence, the probability distribution $\pi(\boldsymbol{y}|w,\sigma)$ of the data $\boldsymbol{y}$ conditional on the wavelet $\boldsymbol{w}$ and the standard deviation σ is Gaussian with mean $\boldsymbol{r} * \boldsymbol{w}$ and covariance matrix $\sigma^2 I$. The noise term could be modelled differently. Using a different covariance matrix we might, for instance, model correlated, or coloured, noise. In this thesis, however, the choice of covariance matrix is restricted to a multiple of the identity matrix. The scalar σ is a model parameter, to be estimated alongside the wavelet parameters $\boldsymbol{\beta}$.

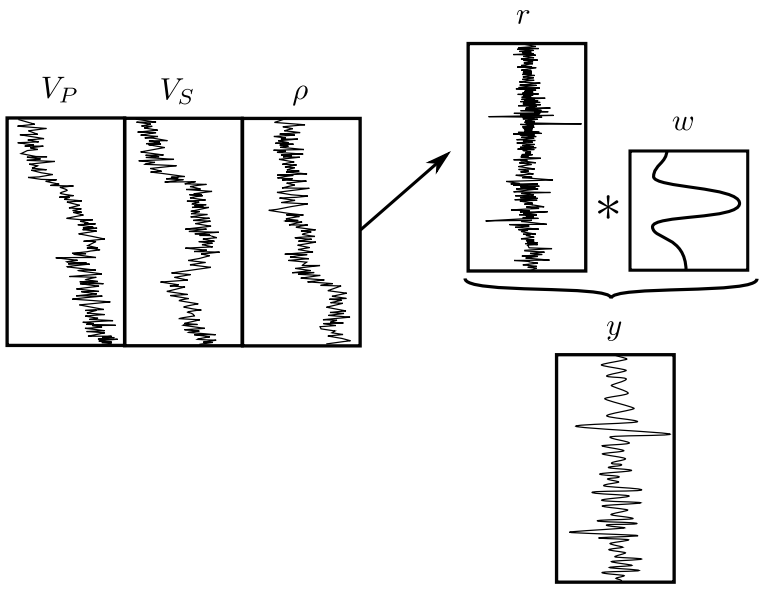


Figure 1: Components involved in the convolutional model. The elastic parameters V_P (P-wave velocity), V_S (S-wave velocity) and ρ (mass density), together with the angle of incidence θ, determine the series of reflection coefficients, or reflectivity, r. Convolving r with the wavelet w gives the trace y. In addition, we assume that the observed trace is contaminated by noise.

2.2 Data properties

In practice, seismic traces are not continuous functions of time, but series of samples taken at a regular time interval Δt. Because seismic traces tend to be coarser than borehole data in terms of sampling frequency, the seismic sampling time interval determines which time interval should be used for estimation. Therefore, after constructing reflection coefficient time series, these are downsampled, to match the seismic data. As a consequence, we do not expect to resolve very high frequency components of the wavelet.

Generally, we have $w = w(t,\theta)$, i.e. the wavelet is function, not only of time, but also of the angle of incidence θ of the seismic waves whose propagation it describes. If we have multiple seismic traces with different angles of incidence, then they will have different seismic wavelets associated with them. Reflectivity also depends on the angle of incidence, in addition to elastic parameters, so the same borehole data will give rise to a different reflectivity series depending on which angle of incidence is used. The exact reflectivity is given by the Zoeppritz equations, but we choose instead to use the approximation (Aki and Richards, 1980)

$$r(\theta) = \frac{1}{2}\left(1+\tan\theta^2\right)\frac{\Delta V_P}{\overline{V_P}} - 4\left(\frac{\overline{V_S}}{\overline{V_P}}\right)^2 \sin\theta^2 \frac{\Delta V_S}{\overline{V_S}} + \frac{1}{2}\left(1 - 4\frac{\overline{V_S}^2}{\overline{V_P}^2}\sin\theta^2\right)\frac{\Delta\rho}{\overline{\rho}}, \tag{5}$$

where V_P and V_S are the P- and S-wave velocities of the propagation medium, and ρ is its mass density. The bar and Δ symbols denote averaging and differencing over a boundary, so at sample point i, we have

$$\overline{\xi}_i = \frac{\xi_i + \xi_{i+1}}{2}, \quad \Delta\xi_i = \frac{\xi_{i+1} - \xi_i}{2}, \quad \xi = V_P, V_S, \rho. \tag{6}$$

2.3 Multi-angle formulation

Suppose that, instead of one trace $\boldsymbol{y}$ and one reflectivity series $\boldsymbol{r}$, we have three traces $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, $\boldsymbol{y}_3$ and three reflectivity series $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, $\boldsymbol{r}_3$. The situation is presented schematically in Figure 2. Rather than the single equation (1), we now have a separate equation

$$y_i = r_i * w_i + \epsilon_i \tag{7}$$

for each angle stack i. Using vector notation, and letting $\boldsymbol{w}_i$ represent w_i, this is

$$\boldsymbol{y}_i = \boldsymbol{r}_i * \boldsymbol{w}_i + \boldsymbol{\epsilon}_i. \tag{8}$$

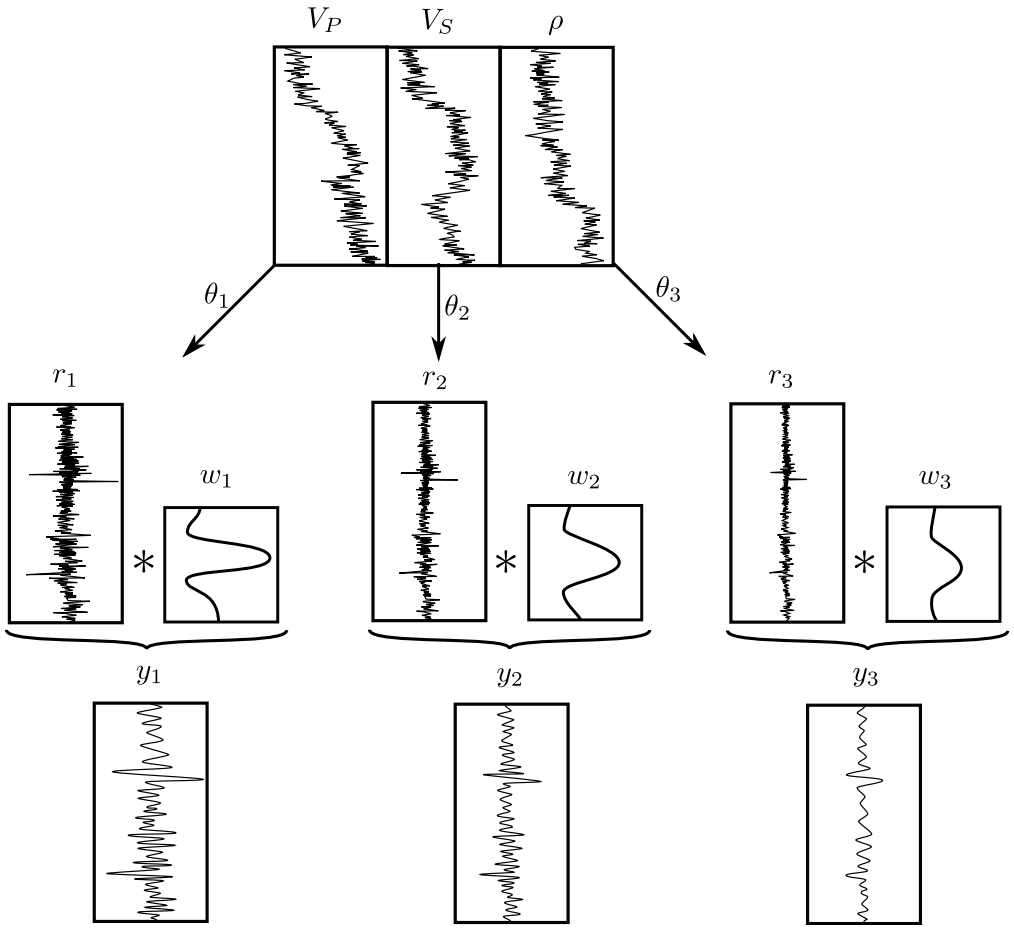


Figure 2: Multi-angle extension of the forward model. Compare with Figure 1. Different angles θ_i, $i = 1, 2, 3$, are used to compute the reflectivity for the near, mid and far traces. The convolution operations are applied independently to the respective reflectivity series.

Assumptions made about the relationship between the wavelets of different stacks, impact how the estimation problem should be approached. If dependence between stacks is negligible, then independent and parallel application of the single stack approach is the natural choice. On the other hand, if some correlation structure is imposed on wavelets of different angle stacks, then joint and simultaneous estimation becomes an option. It is of interest to assess whether, and in which cases, joint estimation offers advantages over independent estimation. The question of how to specify the prior between-angle correlation structure will be addressed in section 5. Note that the multi-angle case is not treated in the main presentation of the wavelet estimation method (section 4). Instead, the method is described there in a single-angle setting, and the discussion is then extended, in section 5, to cover multi-angle estimation.

3 Statistical model

This section explains the Bayesian model underlying the wavelet estimation procedure, describing the various model components, such as the parametric wavelet function, the prior distributions of the wavelet parameters, the likelihood, or conditional distribution of the data given the model parameters, and the posterior distribution, i.e. the conditional distribution of the parameters given the data.

3.1 Wavelet model

The four wavelet parameters, which will be explained in detail later, are the skewness parameter s, the amplitude parameter a, the dilation parameter v, and the wavelet order n. Note that the standard deviation σ is not a wavelet parameter, but rather a different model parameter which is not directly related to the wavelet. The parametric wavelet function is inspired by the parametrisation suggested in Aune et al. (2013), and has the form

$$w(t; s, a, v, n) = a \frac{\psi_{2n}(t/v)}{\psi_{2n}(0)} \frac{\operatorname{erf}(st/v) + 1}{2}, \tag{9}$$

where

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{10}$$

is the error function, and

$$\psi_n(x) = e^{-\frac{1}{2}x^2} He_n(x). \tag{11}$$

is a modified Hermite polynomial

$$He_n(x) = (-1)^n e^{\frac{1}{2}x^2} \frac{\partial^n}{\partial x^n} e^{-\frac{1}{2}x^2}. \tag{12}$$

We mentioned that the wavelet function takes four parameters; s, a, v and n, as seen in (9). Their interpretations are as follows.

Skewness parameter s
: $s \in (-\infty, \infty)$ controls wavelet skewness, and is related to phase rotation. Increasing s shifts the energy of the wavelet towards the right, while decreasing s shifts it to the left (Figure 3).

Amplitude parameter a
: $a \in (0, \infty)$ controls the wavelet amplitude. From (9) we have $w(0; s, a, v, n) = a\frac{\operatorname{erf}(0)+1}{2} = \frac{a}{2}$, since $\operatorname{erf}(0) = 0$.

Dilation parameter v
: $v \in (0, \infty)$ controls the width, or duration, of the wavelet. Increasing v elongates the wavelet, and decreasing v compresses it (Figure 4).

Wavelet order n
: $n \in \{0, 1, 2, \ldots\}$ determines the order of the Hermite polynomial used. Only even ordered polynomials are included, hence the $\psi_{2n}(\cdot)$ in (9). Increasing n effectively gives the wavelet additional sidelobes, shifting its peak frequency upward (Figure 5).

Figure 6 shows the first six Hermite polynomials, and Figure 7 shows the effect of damping, and then normalising, the Hermite polynomials. When n is odd, we have $He_n(0) = 0$, so the normalisation can only be applied to even-ordered wavelets. The shapes of odd-ordered wavelets can be attained by even-ordered wavelets with large (positive or negative) skewness, so excluding odd-ordered wavelets should not affect the model's flexibility.

Since s, a and v are continuous parameters and n is discrete, it is useful to regard these as distinct blocks of parameters for purposes of sampling or optimisation. We must also account for the fact that the parameters have different ranges. The skewness parameter s takes any real value,

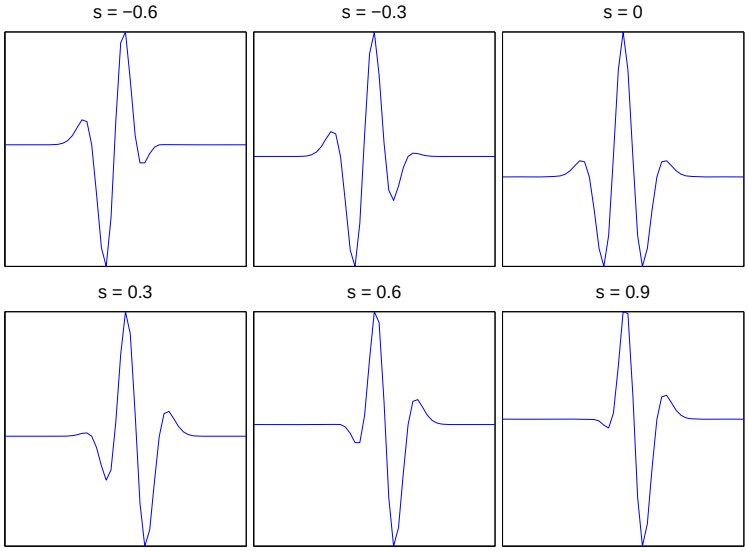


Figure 3: Wavelets with $n = 2$, and six different values of the skewness parameter s.

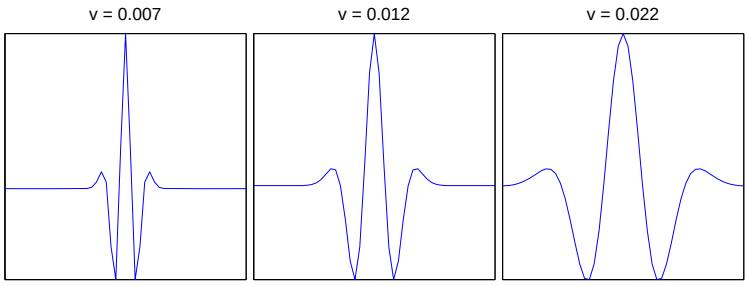


Figure 4: Wavelets with $n = 2$, and three different values of the dilation parameter v.

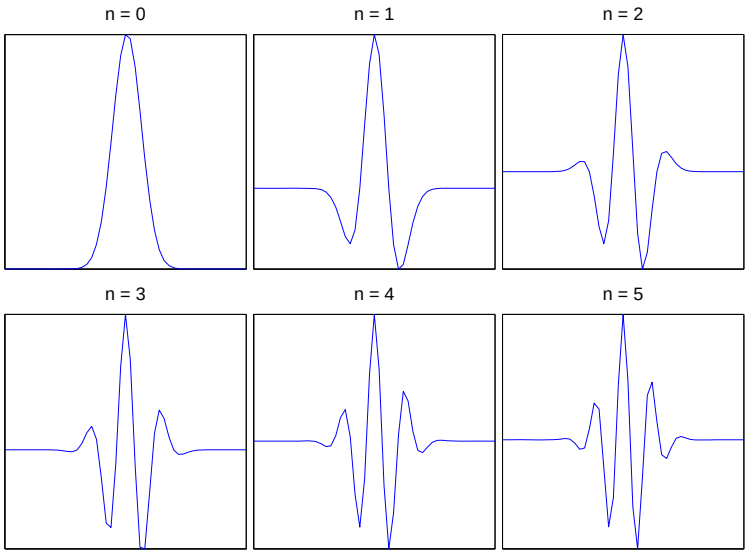


Figure 5: Wavelets with a small positive skewness ($s \approx 0.15$) and six different orders n.

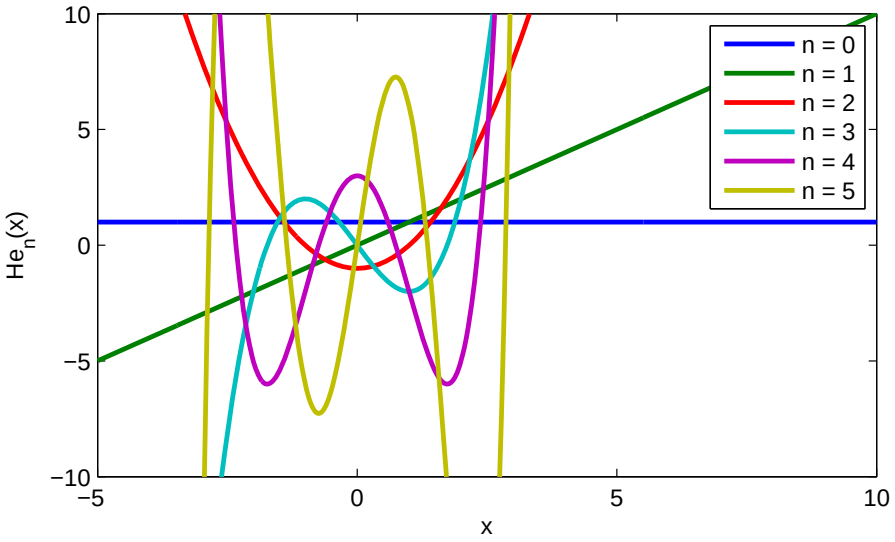


Figure 6: The first six probabilist's Hermite polynomials $He_n(x)$, $n = 0, \ldots, 5$.

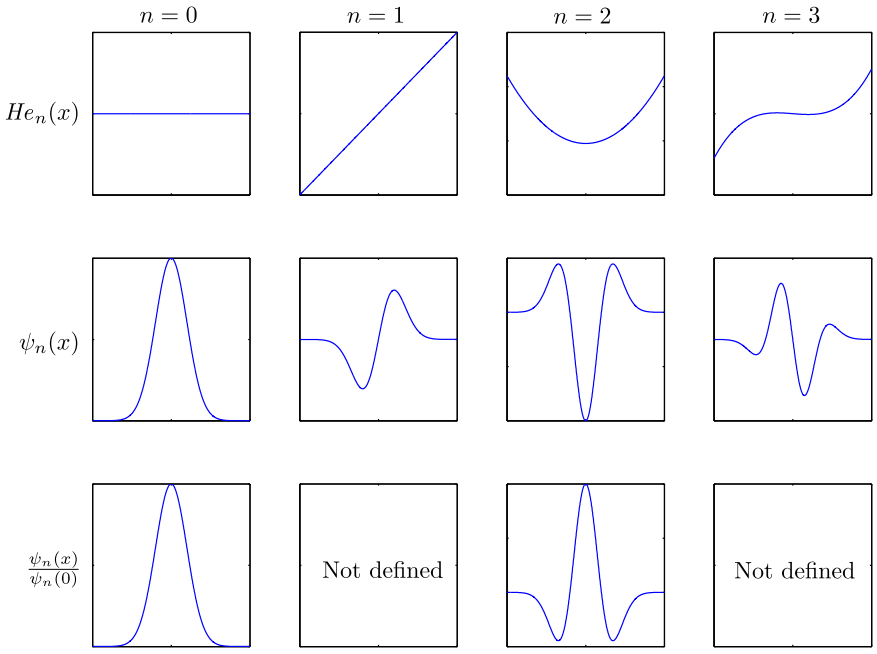


Figure 7: Hermite polynomials $He_n(x)$, damped Hermite polynomials $\psi_n(x) = e^{-x^2/2} He_n(x)$ and normalised damped Hermite polynomials, for polynomial orders $n = 0, 1, 2, 3$. Note that the normalised damped polynomials are not defined for odd n.

while a, v and the noise standard deviation σ are strictly positive, and n takes non-negative integer values only. Therefore, we define transformed parameters

$$\beta_1 = s, \qquad \beta_2 = \log a, \qquad \beta_3 = \log v, \qquad \beta_4 = \log \sigma, \tag{13}$$

and write $\boldsymbol{\beta} = (\beta_1 \ \beta_2 \ \beta_3 \ \beta_4)^T$. With this transformation, we have $\boldsymbol{\beta} \in \mathbb{R}^4$.

Note that the discrete parameter n is not included in $\boldsymbol{\beta}$, and must be handled separately. Also note that even though the wavelet function does not depend on β_4, we generally write $w(t; \boldsymbol{\beta}, n)$, or sometimes only $w(t; \boldsymbol{\beta})$, if the value of n is clear from the context.

3.2 Parameter prior

We give $\boldsymbol{\beta}$ a multivariate normal distribution with mean $\boldsymbol{\mu}_\beta$ and covariance matrix Σ_β,

$$\boldsymbol{\beta} \sim \mathcal{N}\left(\boldsymbol{\mu}_\beta, \Sigma_\beta\right) \tag{14}$$

so that the prior probability density function of $\boldsymbol{\beta}$ is

$$\pi(\boldsymbol{\beta}) = (2\pi)^{-2} \left(\det \Sigma_\beta\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta\right)^T \Sigma_\beta^{-1} \left(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta\right)\right\}. \tag{15}$$

As for the wavelet order parameter n, we can either use a fixed order, or we can treat n as a random variable, estimating $\boldsymbol{\beta}$ and n jointly. In the latter case, it will be useful to put a prior distribution on n. We therefore assign to it a Poisson distribution with mean λ, which gives the probability mass function

$$\pi(n) = \frac{\lambda^n}{n!} e^{-\lambda}. \tag{16}$$

3.3 Likelihood, parametrised

From (3) and the discussion of the data likelihood in section 2, we define the likelihood of $\boldsymbol{y}$ with respect to $\boldsymbol{\beta}$ and n by letting

$$\pi(\boldsymbol{y}|\boldsymbol{\beta}, n) = \pi\left(\boldsymbol{y}|w = w(t; \boldsymbol{\beta}, n), \sigma = e^{\beta_4}\right). \tag{17}$$

The probability density function of $\boldsymbol{y}$ given $\boldsymbol{\beta}$ and n is

$$\pi(\boldsymbol{y}|\boldsymbol{\beta}, n) = (2\pi\sigma^2)^{-\frac{n_y}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{y} - \boldsymbol{r} * \boldsymbol{w}(t; \boldsymbol{\beta}, n)\right)^T \left(\boldsymbol{y} - \boldsymbol{r} * \boldsymbol{w}(t; \boldsymbol{\beta}, n)\right)\right\}, \tag{18}$$

where n_y is the length of the data record, i.e. the number of samples in $\boldsymbol{y}$.

3.4 Posterior distribution

The posterior distribution of interest is

$$\pi(\boldsymbol{\beta}, n|\boldsymbol{y}) \propto \pi(\boldsymbol{\beta}, n, \boldsymbol{y}) \propto \pi(\boldsymbol{\beta})\pi(n)\pi(\boldsymbol{y}|\boldsymbol{\beta}, n). \tag{19}$$

when n is a random variable, and

$$\pi_n(\boldsymbol{\beta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\beta}, \boldsymbol{y}) \propto \pi(\boldsymbol{\beta})\pi(\boldsymbol{y}|\boldsymbol{\beta}) \tag{20}$$

when n is fixed.

Inserting (15), (16) and (18) into (19) gives

$$\begin{aligned} \pi(\boldsymbol{\beta}, n|\boldsymbol{y}) \propto \det(\Sigma_\beta)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)^T \Sigma_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)\right\} \\ \cdot \frac{\lambda^n}{n!} (\sigma^2)^{-\frac{n_y}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{r} * \boldsymbol{w}(\boldsymbol{\beta}, n))^T (\boldsymbol{y} - \boldsymbol{r} * \boldsymbol{w}(\boldsymbol{\beta}, n))\right\}. \end{aligned} \tag{21}$$

Similarly, inserting (15) and (18) into (20) gives

$$\begin{aligned}\pi_n(\boldsymbol{\beta}|\boldsymbol{y}) \propto \det(\Sigma_\beta)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta)^T\Sigma_\beta^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta)\right\} \\ \cdot (\sigma^2)^{-\frac{n_y}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{r}*\boldsymbol{w}(\boldsymbol{\beta},n))^T(\boldsymbol{y}-\boldsymbol{r}*\boldsymbol{w}(\boldsymbol{\beta},n))\right\}.\end{aligned} \tag{22}$$

4 Method

This section presents the estimation method in a single-angle setting. The multi-angle case is discussed in section 5. In order to provide a quick overview, the various steps of the estimation procedure are listed here. Each step will be explained in detail in sections 4.1–4.5.

Step 1: Load data $\boldsymbol{y}$ and $\boldsymbol{r}$. Unless the reflectivity series is already available, it must be computed from borehole data and matched to the sampling frequency of the seismic data. Depending on the properties of the seismic data, it may need to be regularised, to ensure that samples are equidistant.

Step 2: Extract a tentative wavelet estimate $\tilde{\boldsymbol{w}}$ from the seismic data.

- Compute the discrete Fourier transform of $\boldsymbol{y}$.
- Smooth the spectrum by convolving it with a window function.
- Apply the inverse DFT to the smoothed spectrum to recover time domain wavelet $\tilde{\boldsymbol{w}}$.
- Fit the parametric wavelet model $\boldsymbol{w}(t;\boldsymbol{\beta},n)$ to $\tilde{\boldsymbol{w}}$ to get preliminary estimates $\tilde{\boldsymbol{\beta}}$ and $\tilde{n}$.

Step 3: Set parameters for prior distributions based on the preliminary estimates $\tilde{\boldsymbol{\beta}}$ and $\tilde{n}$.

- Either fix $n = \tilde{n}$, or treat n as a random variable with prior mean $\lambda = \tilde{n}$.
- Set the prior standard deviation of each element of $\boldsymbol{\beta}$ separately, based on user specified bounds U_i and exceedance probabilities α_i, which define statements of the form "Peak wavelet amplitude exceeds U_1 with probability α_1".

Step 4: Locate the mode $\hat{\boldsymbol{\beta}}$ and, if applicable, $\hat{n}$, of the posterior distribution.

- Define an objective function $f_{\beta|y,n}(\boldsymbol{\beta}) = -\log\pi(\boldsymbol{\beta}|\boldsymbol{y},n)$ if n is fixed, or $f_{\beta,n|y}(\boldsymbol{\beta},n) = -\log\pi(\boldsymbol{\beta},n|\boldsymbol{y})$ if n is a random variable.
- Supplying $\tilde{\boldsymbol{\beta}}$ as an initial parameter vector, use a numerical optimisation algorithm to minimise the appropriate objective function either with respect to $\boldsymbol{\beta}$ or with respect to $\boldsymbol{\beta}$ and n. Denote the modal configurations $\hat{\boldsymbol{\beta}}$, $\hat{n}$.

Step 5: Sample parameter vectors from the posterior distribution.

- Compute the Hessian H of the negative log-posterior at $\hat{\boldsymbol{\beta}}$.
- Approximate the posterior distribution with a Gaussian distribution with mean $\hat{\boldsymbol{\beta}}$ and covariance matrix $\Sigma_{\hat{\beta}} = H^{-1}$.
- Sample parameter vectors $\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \ldots, \boldsymbol{\beta}^{n_s}$ from the posterior distribution $\pi(\boldsymbol{\beta}|\boldsymbol{y},n)$ by means of an independent-proposal Metropolis-hastings sampler using the Gaussian approximation as a proposal distribution. If n is fixed, then use $n = \tilde{n}$ as before. Otherwise, let $n \sim \mathrm{Po}(\lambda)$, with $\lambda = \tilde{n}$.

The input to the estimation algorithm is the seismic data $\boldsymbol{y}$ and either a reflection coefficient series $\boldsymbol{r}$ or borehole data from which $\boldsymbol{r}$ can be computed. These data records must have been tied, or aligned, in time prior to estimation. In other words, there should be a time vector $\boldsymbol{t}_y$, with sampling

time interval determined by the sampling frequency of $\boldsymbol{y}$, giving the two way time for each sample point of both $\boldsymbol{y}$ and $\boldsymbol{r}$.

Upon completion, the procedure should have produced (1) maximum a posteriori estimates of $\boldsymbol{\beta}$ and, if applicable, n, and (2) a sample of n_s parameter vectors drawn from the posterior distribution of $\boldsymbol{\beta}$. A maximum a posteriori estimate of the wavelet is most easily obtained by evaluating the wavelet function on the MAP parameter estimate $\hat{\boldsymbol{\beta}}$. The sample of parameter vectors can be converted into a sample of wavelets by evaluating the wavelet function on a time vector. These sample wavelets can then be used to construct confidence limits for the wavelet.

4.1 Data loading and processing

The seismic trace $y(t)$ is represented by the vector $\boldsymbol{y}$ indexed by the time vector $\boldsymbol{t}_y$.

Well logs contain records of the seismic P-wave velocity V_p, S-wave velocity V_s and mass density ρ of the rock, usually along with two-way time computed from measured depth and velocity. As mentioned in section 2, the reflection coefficient at a given time is a function of V_p, V_s and ρ, as well as the angle of incidence of the seismic waves. When calculating $\boldsymbol{r}$ from a well log, we can use the average angle given for the stack. The reflectivity series calculated in this way will be indexed by the time vector $\boldsymbol{t}_r$.

For the convolutional model (3) to apply, we must ensure that $\boldsymbol{y}$ and $\boldsymbol{r}$ are commensurate, i.e. the subsurface reflection points about which $\boldsymbol{y}$ contains information should correspond as closely as possible with the trajectory of the borehole providing the data for $\boldsymbol{r}$.

We also need to resample the data records, so that they are indexed by a common time vector $\boldsymbol{t}$, with equidistant samples. Generally, $\boldsymbol{t}_y$ will be much coarser than $\boldsymbol{t}_r$, so if the trace already has a regular sampling time interval, we can simply take $\boldsymbol{t} = \boldsymbol{t}_y$ and downsample $\boldsymbol{r}$ to the coarser time scale. Sometimes, however, $\boldsymbol{t}_y$ will be irregular, so that interpolation and resampling are necessary also for $\boldsymbol{y}$. A deviated well (Figure 8) is an example of a situation where samples will not be equidistant in time, because when seismic data are extracted along the well trajectory, the time interval between consecutive samples will depend on the angle of the trajectory.

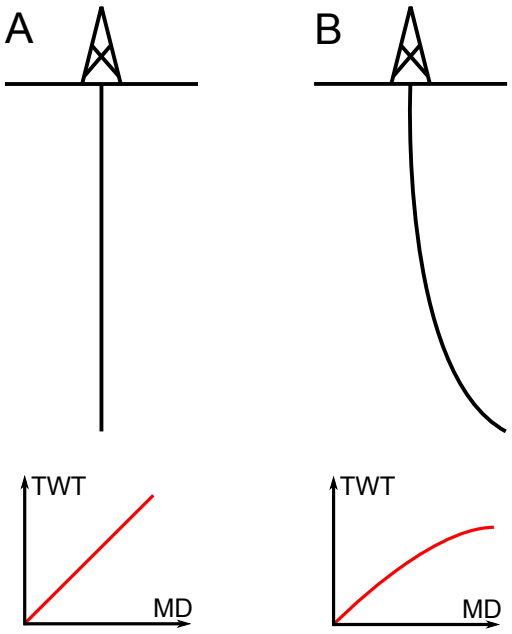


Figure 8: **A:** Straight well. Measured depth is proportional to Two-way time. Samples are equidistant. **B:** Deviated well. Measured depth is related to Two-way time in a non-linear way. A seismic trace extracted along the well trajectory will not have equidistant samples. Rather, the time interval between samples will be longer the greater the deviation from the vertical.

A time vector $\boldsymbol{t}_w$ for the wavelet is also required to evaluate the wavelet function. This can

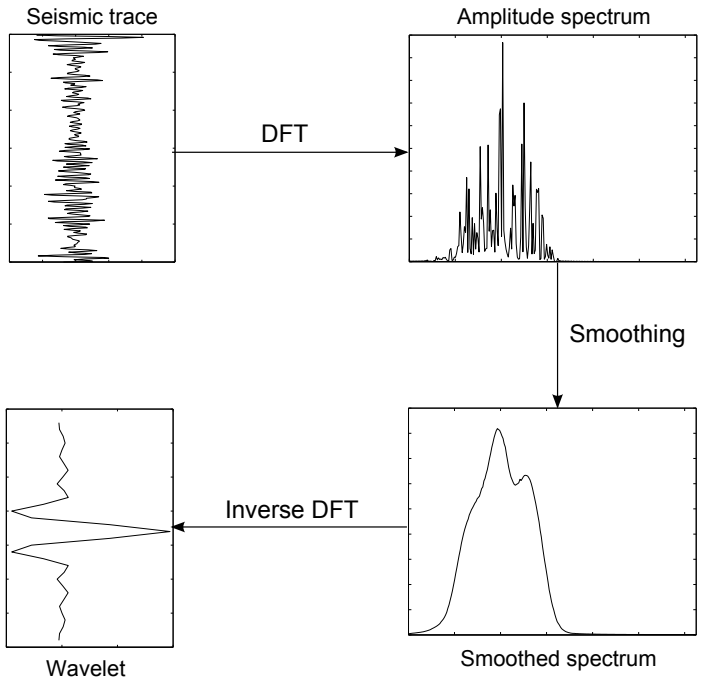


Figure 9: The four stages in the spectrum smoothing procedure: The seismic trace, $\boldsymbol{y}$, the amplitude spectrum, the smoothed spectrum and the wavelet $\tilde{w}$. The implementation of the procedure is based on the `smooth_spectrum` code from Sacchi (2014).

be specified from prior belief about wavelet duration. The time sampling intervals of the time vectors $\boldsymbol{t}$ and $\boldsymbol{t}_w$ must be identical to ensure that discrete convolution behaves as intended. After preprocessing, the three vectors $\boldsymbol{y}$, $\boldsymbol{r}$ and $\boldsymbol{t}$ should have n_y elements each, and we should have

$$y_i = y(t_i), \quad r_i = r(t_i), \qquad i = 1, \ldots, n_y \tag{23}$$

and

$$t_{i+1} - t_i = \Delta t, \qquad i = 1, \ldots, n_y - 1. \tag{24}$$

4.2 Preliminary estimation

To set up the prior distributions of the wavelet parameters, it is useful to have an initial "guess" at what the wavelet looks like. We can extract such a crude estimate from the frequency content of the seismic data, by taking the discrete Fourier transform, and smoothing the amplitude spectrum. This spectrum-smoothing process, which is illustrated in Figure 9, is taken from Sacchi (2014).

First, we compute the discrete Fourier transform of the trace $\boldsymbol{y}$ using an FFT algorithm

$$\boldsymbol{Y} = \mathrm{DFT}(\boldsymbol{y}).$$

Next, we convolve the squared absolute value $|\boldsymbol{Y}|^2$ of the transform, which is a vector of real numbers, with a window function $\boldsymbol{h}(f)$, such as a Hamming window. Let the *smoothed spectrum* $\boldsymbol{A}$ be the square root of this convolution,

$$\boldsymbol{A} = \sqrt{|\boldsymbol{Y}|^2 * \boldsymbol{h}(f)}.$$

Applying the inverse discrete Fourier transform to $\boldsymbol{A}$ turns it back into a time domain waveform. Finally, element-wise multiplication with another window function ensures that the estimated wavelet $\tilde{\boldsymbol{w}}$ has a reasonable shape, dominated by a central peak, and decaying quickly toward the edges,

$$\tilde{\boldsymbol{w}} = \mathrm{DFT}^{-1}(\boldsymbol{A})\boldsymbol{h}(t).$$

The window functions $\boldsymbol{h}(f)$ and $\boldsymbol{h}(t)$ control the degree of smoothing applied to the amplitude spectrum of $\boldsymbol{y}$ and the width of the preliminary wavelet estimate $\tilde{\boldsymbol{w}}$, respectively. The latter also defines a time vector $\tilde{\boldsymbol{t}}$ for the extracted wavelet $\tilde{\boldsymbol{w}}$.

Preliminary parameter estimates can be obtained by matching $\tilde{\boldsymbol{w}}$ to the wavelet function $w(t;\boldsymbol{\beta},n)$ evaluated on the time vector $\tilde{\boldsymbol{t}}$. To facilitate this, we define a distance measure $d(\boldsymbol{u},\boldsymbol{v})$ between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, both with length m, and with sampling time interval $\Delta\tau$,

$$d(\boldsymbol{u},\boldsymbol{v}) = \sum_{i=1}^{m}(u_i - v_i)^2\Delta\tau = (\boldsymbol{u}-\boldsymbol{v})^T(\boldsymbol{u}-\boldsymbol{v})\Delta\tau, \tag{25}$$

which approximates the integral

$$d_I(u(\tau),v(\tau)) = \int_0^{m\Delta\tau}(u(\tau)-v(\tau))^2 d\tau, \tag{26}$$

for continuous functions $u(\tau)$ and $v(\tau)$, $\tau\in\mathbb{R}$.

The preliminary parameter estimates corresponding to $\tilde{\boldsymbol{w}}$ are the parameters minimising the distance between it and $\boldsymbol{w}(\boldsymbol{\beta},n)$, which is $w(t;\boldsymbol{\beta},n)$ evaluated on the same time vector $\tilde{\boldsymbol{t}}$ as $\tilde{\boldsymbol{w}}$,

$$(\tilde{\boldsymbol{\beta}},\tilde{n}) = \arg\min_{(\boldsymbol{\beta},n)} d(\boldsymbol{w}(\boldsymbol{\beta},n),\tilde{\boldsymbol{w}}). \tag{27}$$

This determines $\tilde{n}$ and the first three elements of $\tilde{\boldsymbol{\beta}}$, but not $\beta_4 = \log\sigma$, since σ does not control any aspect of the wavelet. To set β_4, we compute the sample variance of the residual trace $\tilde{\boldsymbol{\epsilon}}$,

$$\tilde{\sigma}^2 = \tilde{\boldsymbol{\epsilon}}^T\tilde{\boldsymbol{\epsilon}} = \left(\boldsymbol{y} - \boldsymbol{r} * \boldsymbol{w}(\tilde{\boldsymbol{\beta}},\tilde{n})\right)^T\left(\boldsymbol{y} - \boldsymbol{r} * \boldsymbol{w}(\tilde{\boldsymbol{\beta}},\tilde{n})\right), \tag{28}$$

and let $\tilde{\beta}_4 = \log\tilde{\sigma} = \frac{1}{2}\log\tilde{\sigma}^2$.

4.3 Setting prior distribution parameters

Before parameters can be estimated from the posterior distribution, the prior distributions of $\boldsymbol{\beta}$ and n must be fully specified by setting the values of the parameters $\boldsymbol{\mu}_\beta$, Σ_β and λ. These values could be set arbitrarily, but basing the choice of prior parameters on the preliminary parameter estimates $\tilde{\boldsymbol{\beta}}$ and $\tilde{n}$ allows for a more structured approach, capable of adapting to the available data.

To set the mean $\boldsymbol{\mu}_\beta$, we use the preliminary estimate $\tilde{\boldsymbol{\beta}}$ directly, letting $\boldsymbol{\mu}_\beta = \tilde{\boldsymbol{\beta}}$.

For the prior covariance matrix Σ_β, we use a diagonal matrix $\Sigma_\beta = \mathrm{diag}(\boldsymbol{\sigma}_\beta^2)$ giving the variances of each element of $\boldsymbol{\beta}$. In other words, no between-parameter correlation is imposed a priori. The element-wise prior standard deviations of $\boldsymbol{\beta}$ are contained in the vector $\boldsymbol{\sigma}_\beta = (\sigma_{\beta_1}\ \sigma_{\beta_2}\ \sigma_{\beta_3}\ \sigma_{\beta_4})^T$.

There is no natural way to set these standard deviations from the preliminary wavelet and parameter estimates. Instead, we set them based on user supplied bounds and exceedance probabilities for wavelet properties controlled by the parameters.

β_1) The skewness parameter s can take any real value, hence we have $\beta_1 = s$ directly, unlike the other parameters. The preliminary wavelet estimate $\tilde{\boldsymbol{w}}$ will generally be symmetric, so that $\tilde{\beta}_1 = 0$. To express prior belief about the variability of s, users can specify a threshold or upper bound U_1 and an exceedance probability α_1, with the interpretation that

$$\mathrm{Prob}\{s > U_1\} = \alpha_1, \tag{29}$$

or equivalently, $\mathrm{Prob}\{\beta_1 > U_1\} = \alpha_1$. Because we assume that the marginal prior distribution of β_1 is $\mathcal{N}(\tilde{\beta}_1,\sigma_{\beta_1}^2)$, (29) can be rewritten as

$$1 - \Phi\left(\frac{U_1 - \tilde{\beta}_1}{\sigma_{\beta_1}}\right) = \alpha_1, \tag{30}$$

where $\Psi(\cdot)$ is the cumulative distribution function of the standard normal distribution. When solving for σ_{β_1}, this gives

$$\sigma_{\beta_1} = \frac{\log U_1 - \tilde{\beta}_1}{\Phi^{-1}(1-\alpha_1)}. \tag{31}$$

β_2) The amplitude parameter a is proportional to the amplitude of $w(t; s, a, v, n)$. Putting $t = 0$ in (9) gives a peak amplitude of

$$w(0; s, a, v, n) = a\frac{\psi_{2n}(0)}{\psi_{2n}(0)}\frac{\text{erf}(0)+1}{2} = \frac{a}{2} \tag{32}$$

for symmetric wavelets, where the central peak occurs at $t = 0$. Users can specify prior beliefs about the wavelet amplitude through an upper bound, or threshold U_2 and an exceedance probability α_2, to be interpreted as follows,

$$\text{Prob}\left\{\frac{a}{2} > U_2\right\} = \alpha_2. \tag{33}$$

Since $\beta_2 = \log a$, this is equivalent to

$$\text{Prob}\left\{\beta_1 > \log(2U_2)\right\} = \alpha_2. \tag{34}$$

And using that $\beta_2 \sim \mathcal{N}(\tilde{\beta}_2, \sigma^2_{\beta_2})$, this means that

$$1 - \Phi\left(\frac{\log(2U_2) - \tilde{\beta}_2}{\sigma_{\beta_2}}\right) = \alpha_2. \tag{35}$$

From this we find the standard deviation

$$\sigma_{\beta_2} = \frac{\log(2U_2) - \tilde{\beta}_2}{\Phi^{-1}(1-\alpha_2)}. \tag{36}$$

β_3) The dilation parameter v is proportional to the duration of the wavelet. For a precise definition of the duration of a wavelet $w(t; s, a, v, n)$, let $t_1(v)$ and $t_2(v)$ be the first and last time points, respectively, where the absolute value of the wavelet exceeds γ times the value at $t = 0$, that is

$$t_1(v) = \min\left\{t : \left|\frac{w(t; s, a, v, n)}{w(0, s, a, v, n)}\right| > \gamma\right\} \tag{37}$$

and

$$t_2(v) = \max\left\{t : \left|\frac{w(t; s, a, v, n)}{w(0, s, a, v, n)}\right| > \gamma\right\} \tag{38}$$

for some $\gamma \in (0, 1)$. Then the duration $\delta(v)$ of $w(t; s, a, v, n)$ is

$$\delta(v) = t_2(v) - t_1(v). \tag{39}$$

Taking advantage of the linear relationship between $\delta(v)$ and v, we can fit a line

$$\hat{\delta}(v) = \hat{b}_0 + \hat{b}_1 v. \tag{40}$$

When δ is measured in seconds, we estimate $\hat{b}_0 = -0.0005$ and $\hat{b}_1 = 7.6124$, with $\gamma = 0.05$. As for the other parameters, the prior standard deviation of $\beta_3 = \log v$ can now be set by choosing a threshold U_3 and an exceedance probability α_3, for the duration $\delta(v)$. These are then interpreted as

$$\text{Prob}\left\{\delta(v) > U_3\right\} = \alpha_3, \tag{41}$$

which, by (40), is approximately equivalent to

$$\begin{aligned} \text{Prob}\left\{\hat{b}_0 + \hat{b}_1 v > U_3\right\} &= \alpha_3 \\ \text{Prob}\left\{v > \frac{U_3 - \hat{b}_0}{\hat{b}_1}\right\} &= \alpha_3 \\ \text{Prob}\left\{\beta_3 > \log\left(\frac{U_3 - \hat{b}_0}{\hat{b}_1}\right)\right\} &= \alpha_3, \end{aligned} \tag{42}$$

which gives

$$1 - \Phi\left(\frac{\log\left(\frac{U_3-\hat{b}_0}{\hat{b}_1}\right) - \tilde{\beta}_3}{\sigma_{\beta_3}}\right) = \alpha_3, \tag{43}$$

and finally

$$\sigma_{\beta_3} = \frac{\log\left(\frac{U_3-\hat{b}_0}{\hat{b}_1}\right) - \tilde{\beta}_3}{\Phi^{-1}(1-\alpha_3)}. \tag{44}$$

β_4) The standard deviation σ of the residual $\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{r} * \boldsymbol{w}$ is interpretable enough that we take U_4 as a bound for σ directly. Again we have

$$\text{Prob}\{\sigma > U_4\} = \alpha_4, \tag{45}$$

and, using $\beta_4 = \log\sigma$ and $\beta_4 \sim \mathcal{N}(\tilde{\beta}_4, \sigma^2_{\beta_4})$, we end up with

$$\sigma_{\beta_4} = \frac{\log U_4 - \tilde{\beta}_4}{\Phi^{-1}(1-\alpha_4)} \tag{46}$$

by the same argument as for the other parameters.

To set the prior mean λ of n, we simply let $\lambda = \tilde{n}$. Alternatively, we can fix n at this stage, and estimate only $\boldsymbol{\beta}$, in which case $n = \tilde{n}$ is a reasonable choice.

4.4 Posterior distribution maximisation

The goal in this step is to obtain maximum a posteriori estimates of $\boldsymbol{\beta}$ and n, by locating the posterior mode, i.e. the values that maximise the posterior distribution $\pi(\boldsymbol{\beta}, n|\boldsymbol{y})$. If n is fixed, then the goal is to find the value of $\boldsymbol{\beta}$ that maximises $\pi_n(\boldsymbol{\beta}|\boldsymbol{y})$.

To locate the mode, a numerical optimisation algorithm will be applied to an objective function defined such that maximising the posterior distribution is equivalent to minimising the objective function. This allows us to frame the optimisation problem as a minimisation problem, and it lets the optimisation algorithm evaluate the posterior distribution in log scale.

From the posterior probability distribution functions (21) and (22), we define the objective functions

$$f_{\beta,n|y}(\boldsymbol{\beta}, n) = -\log\pi(\boldsymbol{\beta}, n|\boldsymbol{y}) \qquad \text{and} \qquad f_{\beta|y}(\boldsymbol{\beta}) = -\log\pi_n(\boldsymbol{\beta}|\boldsymbol{y}) \tag{47}$$

by taking the negative logarithms of those expressions. Whereas (21) and (22) express proportionality, or equality up to a multiplicative constant, the functions in (47) differ from the exact log-posteriors by additive constants. This is inconsequential for purposes of optimisation, however, because the functions' shapes are preserved.

There are two cases to consider, giving different optimisation problems. In the case where n is fixed in the prior specification step, say $n = \tilde{n}$, we seek the mode $\boldsymbol{\beta}$ of $\pi(\boldsymbol{\beta}|\tilde{n}, \boldsymbol{y})$. Since there are no constraints on $\boldsymbol{\beta}$, this can be found by solving the unconstrained minimisation problem

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} f_{\beta|y}(\boldsymbol{\beta}|\boldsymbol{y}). \tag{48}$$

In the other case, where n is not fixed, we are looking for the modal configuration $(\hat{\boldsymbol{\beta}}, \hat{n})$ of $\pi(\boldsymbol{\beta}, n|\boldsymbol{y})$, which is the solution of the mixed integer unconstrained minimisation problem

$$(\hat{\boldsymbol{\beta}}, \hat{n}) = \arg\min_{(\boldsymbol{\beta},n)} f_{\beta,n|y}(\boldsymbol{\beta}, n|\boldsymbol{y}). \tag{49}$$

In the first case, we locate $\hat{\boldsymbol{\beta}}$ by simply applying a numerical optimisation algorithm to (48). In the second case, we break the optimisation into several steps.

1. Choose a range of orders, $n_1, \ldots, n_k$. A reasonable choice is $1, 2, \ldots, k$, where k is 4 or 5.

2. Locate $\hat{\boldsymbol{\beta}}_i = \arg\min_{\boldsymbol{\beta}} f_{\beta,n|y}(\boldsymbol{\beta}, n_i)$, for $i = 1, \ldots, k$.

3. Let $i^* = \arg\min_i f_{\beta,n|y}(\hat{\boldsymbol{\beta}}_i, n_i)$. In other words, choose whichever i gives the smallest minimum of the objective function.

4. Let $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{i^*}$ and $\hat{n} = n_{i^*}$.

4.5 Sampling from posterior distribution

Once the modal configuration $(\hat{\boldsymbol{\beta}}, \hat{n})$ has been located, a maximum a posteriori estimate of the wavelet is available by evaluating (9),

$$\hat{w}(t) = w(t; \hat{\boldsymbol{\beta}}, \hat{n}). \tag{50}$$

For characterising the uncertainty in $\hat{w}(t)$, we can generate samples of $\boldsymbol{\beta}$ from the posterior distribution $\pi(\boldsymbol{\beta}|\hat{n}, \boldsymbol{y})$. Sampling will be more effective if we use a proposal distribution adjusted to the shape of the target posterior distribution. We accomplish this by using a Gaussian approximation $\pi_G(\boldsymbol{\beta}|\hat{n}, \boldsymbol{y})$ of $\pi(\boldsymbol{\beta}|\hat{n}, \boldsymbol{y})$. Let H be the Hessian matrix of $f_{\beta|n,y}$ at the mode $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{n}})$. Then we have

$$\begin{aligned} H &= \frac{\partial^2}{\partial \boldsymbol{\beta}^2} f_{\beta|n,y}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}},\ n=\hat{n}} \\ &= \frac{\partial^2}{\partial \boldsymbol{\beta}^2}\big(-\log \pi(\boldsymbol{\beta}|n, \boldsymbol{y})\big)\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}},\ n=\hat{n}} \\ &= -\frac{\partial^2}{\partial \boldsymbol{\beta}^2} \log \pi(\boldsymbol{\beta}|n, \boldsymbol{y})\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}},\ n=\hat{n}}, \end{aligned} \tag{51}$$

hence, the Gaussian approximation has mean $\hat{\boldsymbol{\beta}}$ and covariance matrix H^{-1}, and the probability density function of our proposal distribution is

$$\pi_G(\boldsymbol{\beta}|\hat{n}, \boldsymbol{y}) = \frac{1}{\sqrt{2\pi}} (\det H)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) H \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right\}. \tag{52}$$

There are several possibilities for generating a sample of parameter vectors. Drawing directly from $\pi_G(\boldsymbol{\beta}|\hat{\boldsymbol{n}}, \boldsymbol{y})$ is quick, but only yields a sample from approximately the right distribution. A more exact sample can be obtained by drawing candidate parameter combinations from the Gaussian approximation, and then accepting them with some acceptance probability α depending on the target distribution $\pi(\boldsymbol{\beta}|\hat{n}, \boldsymbol{y})$. This can be implemented either as a Metropolis-Hastings sampling algorithm, producing correlated samples, or as a rejection sampling algorithm, which produces independent samples. Rejection sampling requires that the scale of the target distribution be matched properly to the proposal distribution, and since the shape of the posterior distribution is unknown, it is difficult to perform this matching automatically. Therefore, we choose to implement a Metropolis-Hastings sampler, using the Gaussian approximation as the proposal distribution for $\boldsymbol{\beta}$. When n is a random variable, we can propose new values n' from its prior distribution before drawing the $\boldsymbol{\beta}$-proposal, i.e. $n' \sim \pi(n)$, and $\boldsymbol{\beta'} \sim \pi_G(\boldsymbol{\beta'}|n', \boldsymbol{y})$. The acceptance probability for a proposed transition from $(\boldsymbol{\beta}, n)$ to $(\boldsymbol{\beta'}, n')$ is

$$\begin{aligned} \alpha(\boldsymbol{\beta}, n \to \boldsymbol{\beta'}, n') &= \min\left\{1, \frac{\pi(\boldsymbol{\beta'}, n'|\boldsymbol{y})\pi_G(\boldsymbol{\beta}|n, \boldsymbol{y})\pi(n)}{\pi(\boldsymbol{\beta}, n|\boldsymbol{y})\pi_G(\boldsymbol{\beta'}|n', \boldsymbol{y})\pi(n')}\right\} \\ &= \min\left\{1, \frac{\pi(n')\pi(\boldsymbol{\beta'})\pi(\boldsymbol{y}|\boldsymbol{\beta'}, n')\pi_G(\boldsymbol{\beta}|n, \boldsymbol{y})\pi(n)}{\pi(n)\pi(\boldsymbol{\beta})\pi(\boldsymbol{y}|\boldsymbol{\beta}, n)\pi_G(\boldsymbol{\beta'}|n', \boldsymbol{y})\pi(n')}\right\} \\ &= \min\left\{1, \frac{\pi(\boldsymbol{\beta'})\pi(\boldsymbol{y}|\boldsymbol{\beta'}, n')\pi_G(\boldsymbol{\beta}|n, \boldsymbol{y})}{\pi(\boldsymbol{\beta})\pi(\boldsymbol{y}|\boldsymbol{\beta}, n)\pi_G(\boldsymbol{\beta'}|n', \boldsymbol{y})}\right\} \end{aligned} \tag{53}$$

since the factors $\pi(n)$ and $\pi(n')$ both appear once as a factor in the posterior distribution, and once as a proposal distribution for the order. If n is fixed, then we are only interested in sampling from

the posterior distribution of $\boldsymbol{\beta}$ conditional on the chosen value $\hat{n}$. The acceptance probability for a transition from $\boldsymbol{\beta}$ to $\boldsymbol{\beta'}$ is then

$$\alpha(\boldsymbol{\beta} \to \boldsymbol{\beta'}) = \min\left\{1, \frac{\pi(\boldsymbol{\beta'}|\hat{n}, \boldsymbol{y})\pi_G(\boldsymbol{\beta}|\hat{n}, \boldsymbol{y})}{\pi(\boldsymbol{\beta}|\hat{n}, \boldsymbol{y})\pi_G(\boldsymbol{\beta'}|\hat{n}, \boldsymbol{y})}\right\}. \tag{54}$$

When tested on synthetic and real data (details in section 6), the overall accept rate of the Metropolis-Hastings sampler tends to be quite high; usually in the 0.4–0.7 range.

Parameter vector samples can be used to create wavelet samples by evaluating the wavelet function. The variability of the wavelet samples then serves as a measure of the uncertainty of the MAP wavelet estimate.

5 Multi-angle extension

The method described in section 4 can be applied to the multi-angle or multi-stack case, where several seismic traces are available from different angle stacks. This section will describe the modifications necessary to extend the method in this way. In the multi-angle setting, we aim to estimate a separate wavelet, and hence a separate set of parameters, for each stack.

5.1 Data

Assume, for the sake of clarity, that there are three traces $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ and $\boldsymbol{y}_3$ from the near, mid and far stacks respectively, but covering the same position, i.e. a single well. Each stack has an average angle of incidence θ_i associated with it. Since the reflection coefficient depends on angle of incidence, we get a different reflection coefficient series $\boldsymbol{r}_i$ for each stack.

5.2 Preliminary estimation

The three wavelets to be estimated are denoted

$$w_1(t) = w(t; \boldsymbol{\beta}_1, n), \quad w_2(t) = w(t; \boldsymbol{\beta}_2, n) \quad \text{and} \quad w_3(t) = w(t; \boldsymbol{\beta}_3, n), \tag{55}$$

with parameter vectors $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$. For ease of notation, we combine these into one 12 element vector $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^T \ \boldsymbol{\beta}_2^T \ \boldsymbol{\beta}_3^T\right)^T$. The order n is the same for all three wavelets, reflecting the fact that all three wavelets are expected to have a similar shape.

As between-stack differences in wavelets are expected to be small, and since the preliminary wavelet estimate is only meant to give a rough indication of the wavelet shape, for the purpose of setting the parameters of the prior distribution, we apply the spectrum smoothing method discussed in section 4.2 only once to the multi-angle seismic data. This requires that all the traces be aggregated in some way. If, in addition to the traces $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, $\boldsymbol{y}_3$, from the near, mid and far stacks, a trace $\boldsymbol{y}$ from a full stack is available, then we can apply the spectrum smoothing method to that. Otherwise, the individual traces can be combined into an averaged trace $\overline{\boldsymbol{y}} = \sum_i \boldsymbol{y}_i/3$.

Either way, we extract a single preliminary wavelet estimate $\tilde{\boldsymbol{w}}$ from $\boldsymbol{y}$ using the spectrum smoothing approach. And to $\tilde{\boldsymbol{w}}$, we fit a single set of parameters $(\tilde{\boldsymbol{\beta}}, \tilde{n})$, as in the single-angle case.

5.3 Setting prior distributions

All three wavelets are expected to be qualitatively similar, the main points of difference being amplitude and dilation, and possibly the standard deviation parameter σ, although the latter is more closely related to the seismic trace than the wavelet. As a consequence of this expected similarity, and with a view to keeping the model simple and parsimonious, we use the same order n for all three wavelets. Moreover, we impose a high degree of between-stack correlation on the elements of the combined parameter vector $\boldsymbol{\beta}$.

To this end we define a 3×3 correlation matrix C_θ, containing the correlation coefficients ρ_{ij}, $i, j \in \{1, 2, 3\}$ between the sets of parameters corresponding to each of the three incidence angles θ_1, θ_2 and θ_3, which are the average incidence-angles of the near, mid and far traces, respectively,

$$C_\theta = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}. \tag{56}$$

There are many ways to choose the correlation coefficients ρ_{ij}. The simplest choice is probably to use the same value ρ for every coefficient, i.e.

$$\rho_{ij} = \rho, \quad i, j = 1, 2, 3, \quad i \neq j. \tag{57}$$

For a more nuanced approach, we could let the correlation between wavelet parameters associated with angles of incidence θ_i and θ_j, be a function of the angle difference $|\theta_i - \theta_j|$. For example, we could let

$$\rho_{ij} = \exp\left(-\frac{|\theta_i - \theta_j|^2}{\kappa^2}\right), \quad i, j = 1, 2, 3, \quad i \neq j, \tag{58}$$

where κ is a constant which determines the correlation range. More sophisticated correlation structures are also available, but in this thesis, we will focus on the simple case of constant correlation across all angles.

From C_θ, and the 4×4 covariance matrix $\Sigma_\beta = \text{diag}(\boldsymbol{\sigma}_\beta^2)$ defined as in section 4.3, we now define a new 12×12 covariance matrix $\Sigma_{\beta\theta}$ for the prior distribution of $\boldsymbol{\beta}$,

$$\Sigma_{\beta\theta} = C_\theta \otimes \Sigma_\beta = \begin{bmatrix} \Sigma_\beta & \rho_{12}\Sigma_\beta & \rho_{13}\Sigma_\beta \\ \rho_{12}\Sigma_\beta & \Sigma_\beta & \rho_{23}\Sigma_\beta \\ \rho_{13}\Sigma_\beta & \rho_{23}\Sigma_\beta & \Sigma_\beta \end{bmatrix}. \tag{59}$$

The standard deviations in $\boldsymbol{\sigma}_\beta$ are adjusted using bounds U_i and exceedance probabilities α_i as in the single angle case, but note that this is still only done for four parameters, and then replicated by taking the product in (59).

The mean of $\boldsymbol{\beta}$ is defined by repeating the preliminary parameter estimate $\tilde{\boldsymbol{\beta}}$ from $\tilde{\boldsymbol{w}}$ three times,

$$\boldsymbol{\mu}_{\beta\theta} = \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\beta}} \end{bmatrix}. \tag{60}$$

No changes from the single angle case are necessary concerning n. There is still the choice between fixing the order, by taking $n = \tilde{n}$, for instance, or viewing it as a random variable, in which case it is assigned a Poisson distribution with mean $\lambda = \tilde{n}$. It is also possible to let the orders of different wavelets be separate variables, obtaining individual estimates $\hat{n}_1$, $\hat{n}_2$, etc. of each wavelet's order. The simplest choice, however, is to use a single, common order n, and the current work is limited to that case.

5.4 Maximisation of posterior distribution

In the multi-angle case, the prior distribution of the parameter vector $\boldsymbol{\beta} \in \mathbb{R}^{12}$ is Gaussian with mean $\boldsymbol{\mu}_{\beta\theta}$ and covariance matrix $\Sigma_{\beta\theta}$,

$$\pi(\boldsymbol{\beta}) \propto \Sigma_{\beta\theta}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\beta\theta}\right)^T \Sigma_{\beta\theta}\left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\beta\theta}\right)\right\}. \tag{61}$$

The likelihood of each trace is defined in the same way as the likelihood in section 3.3, and the likelihood of all three traces is the product of the individual likelihoods, due to independence between the traces,

$$\pi(\boldsymbol{y}_\bullet|\boldsymbol{\beta}, n) = \pi(\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, n) = \pi(\boldsymbol{y}_1|\boldsymbol{\beta}_1, n)\pi(\boldsymbol{y}_2|\boldsymbol{\beta}_2, n)\pi(\boldsymbol{y}_3|\boldsymbol{\beta}_3, n). \tag{62}$$

Here, the bullet symbol ($\bullet$) signifies concatenation of vectors with different subscripts, so $\boldsymbol{y}_\bullet$ is the combination of $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ and $\boldsymbol{y}_3$. Combining the prior distributions of $\boldsymbol{\beta}$ and n with the likelihood (62) gives the posterior distributions

$$\pi(\boldsymbol{\beta}, n|\boldsymbol{y}_\bullet) \propto \pi(\boldsymbol{\beta})\pi(n)\pi(\boldsymbol{y}_\bullet|\boldsymbol{\beta}, n) \tag{63}$$

and

$$\pi_n(\boldsymbol{\beta}|\boldsymbol{y}_\bullet) \propto \pi(\boldsymbol{\beta})\pi(\boldsymbol{y}_\bullet|\boldsymbol{\beta}, n). \tag{64}$$

Objective functions $f_{\beta,n|y}(\boldsymbol{\beta}, n)$ and $f_{\beta|n,y}(\boldsymbol{\beta})$ are defined as the negative logarithms of $\pi(\boldsymbol{\beta}, n|\boldsymbol{y}_\bullet)$ and $\pi_n(\boldsymbol{\beta}|\boldsymbol{y}_\bullet)$ respectively.

The numerical optimisation is handled similarly as before, but since the dimension is increased threefold, the optimisation algorithm will take longer to converge to the minimum of the objective function. As before, the modal configuration is denoted $(\hat{\boldsymbol{\beta}}, \hat{n})$.

5.5 Sampling from posterior distribution

Sampling is similar to the single angle case, and the same sampling algorithm can be used with minimal modifications, whether it be a Metropolis random walk or a rejection sampling algorithm. Sampled parameter vectors will need to be "unpacked", or split up into three parts, so that each part can be evaluated to produce a separate wavelet. Sample properties should be interpreted the same way as in the single angle case.

In the single-angle situation, the posterior distribution might conceivably be explored using a non-sampling based technique, such as mapping out the functional shape of the distribution by evaluating it in a set of deterministically positioned points in parameter space. While this is viable in a single-angle setting where we have only four parameters, it quickly becomes infeasible in the multi-angle case, as the dimensionality increases. For multi-angle estimation, therefore, a sampling based approach to uncertainty characterisation is more generally applicable.

6 Examples and Test Cases

In this section, we present a number of test cases for the wavelet estimation procedure. Each case uses a different dataset, and serves as a usage example for the estimation procedure. In order to reveal strengths and weaknesses in the maximisation-based approach, we also take advantage of these test cases by comparing the performance of the proposed MAP estimation procedure to that of a linear Bayes estimator. This estimator is an example of a point-wise estimation method, because it represents the wavelet as a vector of samples, and essentially treats all the sample points as individual model parameters.

The test cases vary in complexity and realism from the fully synthetic and idealised first case, via a series of cases based on a semi-synthetic dataset, consisting of some real and some synthetic data, to a fully real case, with no synthetic data.

To study the properties of the MAP estimation procedure, as well as highlight some of its advantages and disadvantages, we introduce here an alternative wavelet estimation method to contrast it with. The Bayesian Least Squares (BLS) or Linear Minimum Mean Square Error (LMMSE) estimator is computed as a linear transformation of the samples in the observed seismic trace $\boldsymbol{y}$. It assumes a linear model which can be viewed as a simpler form of the hierarchical model suggested by Buland and Omre (2003).

Instead of a parametric function, let the wavelet be represented by a vector $\boldsymbol{w}$ of samples. Assume a Gaussian prior

$$\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{\mu}_w, C_w) \tag{65}$$

for the wavelet, and assume more specifically that the mean is $\boldsymbol{\mu}_w = 0$ and that the covarianc matrix C_w is given by one of two choices

$$(C_w)_{ij} = \begin{cases} \gamma^2 \exp\left\{-\frac{|i-j|^2}{2\tau^2}\right\} & \text{(homoscedastic prior)} \\ \gamma^2 \exp\left\{-\frac{|i-n_w/2|^2+|j-n_w/2|^2}{2\xi^2}\right\} \exp\left\{-\frac{|i-j|^2}{2\tau^2}\right\} & \text{(heteroscedastic prior)} \end{cases}. \tag{66}$$

With the homoscedastic prior distribution, prior wavelet uncertainty is independent of time, i.e. we are equally uncertain about the value of the wavelet at all n_w sample points. On the other hand, the heteroscedastic prior assigns smaller uncertainty on the edges of the wavelet than in the middle. The prior variance of wavelet sample i falls off as $|i - n_w/2|$ increases. This forces the wavelet to taper off and tend to zero at the edges, provided the parameter ξ is chosen appropriately.

To make use of the BLS estimator, we must rewrite the convolutional model (1) as

$$\boldsymbol{y} = R\boldsymbol{w} + \boldsymbol{e} \tag{67}$$

where R is a convolution matrix constructed by arranging the elements of $\boldsymbol{r}$ such that $R\boldsymbol{w} = \boldsymbol{r} * \boldsymbol{w}$. The noise vector $\boldsymbol{e}$ is assumed to be independently zero-mean Gaussian with variance σ^2,

$$\boldsymbol{e} \sim \mathcal{N}\left(0, \sigma^2 I\right) \Longrightarrow \boldsymbol{y}|\boldsymbol{w} \sim \mathcal{N}\left(R\boldsymbol{w}, \sigma^2 I\right). \tag{68}$$

With these assumptions, the LMMSE estimator of $\boldsymbol{w}$ is given by (Kay, 1993)

$$\hat{\boldsymbol{w}}_{\mathrm{BLS}} = \boldsymbol{\mu}_w + C_w R^T \left(RC_w R^T + \sigma^2 I\right)^{-1} \left(\boldsymbol{y} - R\boldsymbol{\mu}_w\right), \tag{69}$$

and the residual $\boldsymbol{\epsilon} = \boldsymbol{w} - \hat{\boldsymbol{w}}_{\mathrm{BLS}}$ is zero-mean Gaussian with covariance matrix

$$C_\epsilon = C_w - C_w R^T \left(RC_w R^T + \sigma^2 I\right)^{-1} RC_w. \tag{70}$$

6.1 Single reflection

The first test case assumes an earth model with two different geological layers, separated by a reflective boundary. This very simple test case lets us confirm that the estimation method is working as intended, and allows us to examine the properties of the estimation method, rather than those of the data.

A noiseless synthetic trace $\boldsymbol{y}^*$ is created by convolving a known wavelet $w(t)$ (Figure 10) with a reflectivity time series $\boldsymbol{r}$, representing the simple two-layer earth model (Figure 11). The resulting trace, which is also shown in Figure 11, contains a single reflected event at the layer boundary.

Since the reflectivity series $\boldsymbol{r}$ has a large value at the layer boundary, and is zero otherwise, the shape of the wavelet is captured exactly in $\boldsymbol{y}^*$, and perfect recovery would be expected in the noiseless case. Some low-intensity white Gaussian noise is added, however, to produce the perturbed trace $\boldsymbol{y} = \boldsymbol{y}^* + \boldsymbol{e}$, which is used as input data.

Figure 12 shows the prior wavelet estimate $\tilde{\boldsymbol{w}}$, and the MAP wavelet estimate $\hat{w}(t)$ together with the true wavelet $w(t)$. Note that the prior wavelet is represented in the seismic time resolution ($\Delta t = 4$ ms), which is coarser than the resolution used for plotting the parametric wavelets. This figure illustrates the difference between the preliminary wavelet estimate $\tilde{\boldsymbol{w}}$, and the final wavelet estimate, or maximum a posteriori estimate, $\hat{w}(t)$. By construction, $\tilde{\boldsymbol{w}}$ is symmetric, and normalised to unit peak amplitude. The preliminary wavelet estimate matches the true wavelet closely in the right side of the figure, but there is considerable mismatch between the two near the smaller, left sidelobe, and near the central peak. Since the true wavelet was generated with the same wavelet model used for estimation, it is no surprise that the posterior mode is close to it. What little discrepancy is present, can be ascribed to the influence of noise.

Figure 13 shows the prior distribution of n, which is a Poisson distribution with mean $\lambda = 2$, compared with the posterior distribution, computed by normalising the modal values of the posterior distribution when n takes values between 0 and 6. The posterior is maximised when $n = 2$, so this is the chosen order $\hat{n}$. In Figure 14, MAP and BLS wavelet estimates are shown surrounded by 95% confidence limits. The limits for the MAP estimate are computed empirically from samples of the posterior distribution, whereas the limits for the BLS wavelet are constructed from the posterior mean and standard deviation of the wavelet at each time point, specifically by taking $\hat{\boldsymbol{w}}_{\mathrm{BLS}} \pm 2\sqrt{\mathrm{diag}\left(C_\epsilon\right)}$. The MAP wavelet estimate here is the same as in Figure 12, and the narrow 95% confidence envelope reflects the small uncertainty associated with the estimate, which is expected, given the low noise level. The BLS wavelet was estimated with the heteroscedastic prior

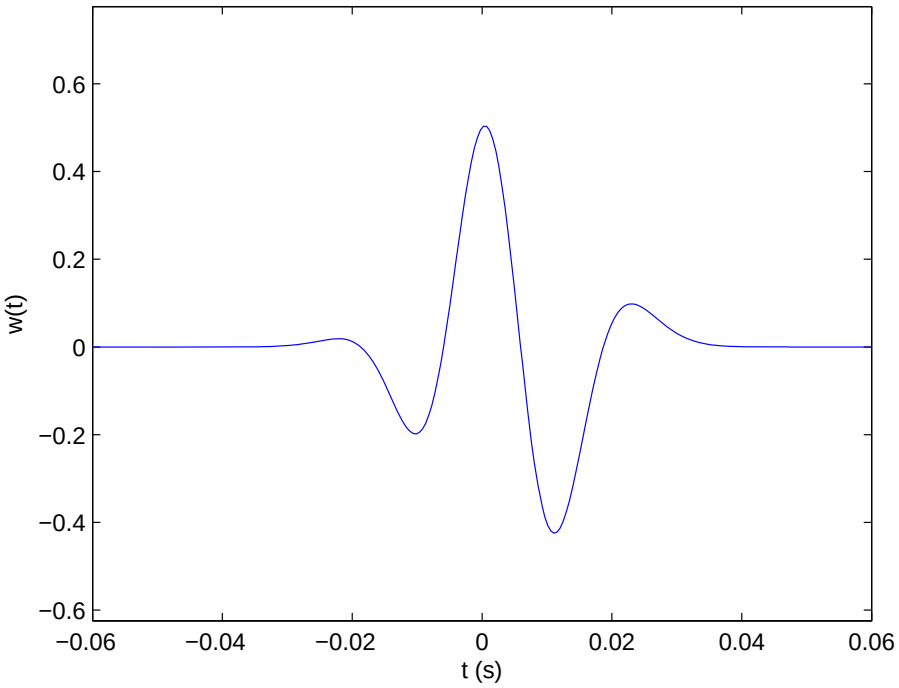


Figure 10: True wavelet used to create the synthetic seismic trace for the single reflection test case. This wavelet was generated using the parametric wavelet model with parameters $s = 0.25$, $a = 1$, $v = 0.008$ and $n = 2$.

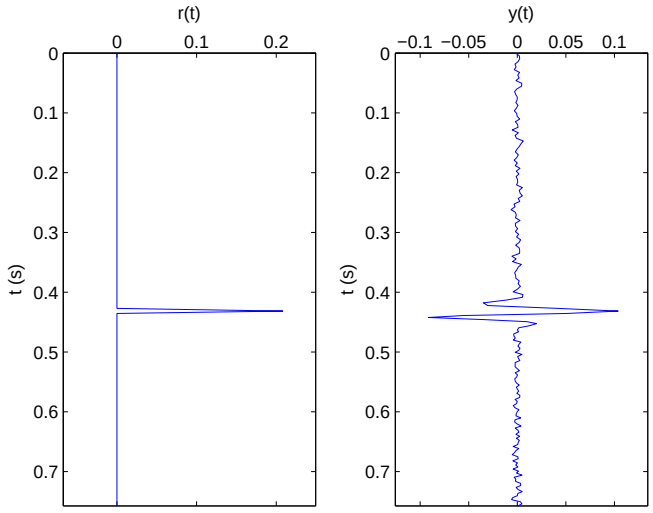


Figure 11: *Left:* Reflectivity for two-layer earth model. *Right:* Synthetic seismic trace $\boldsymbol{y}$ after addition of the vector $\boldsymbol{e}$ of white Gaussian noise.

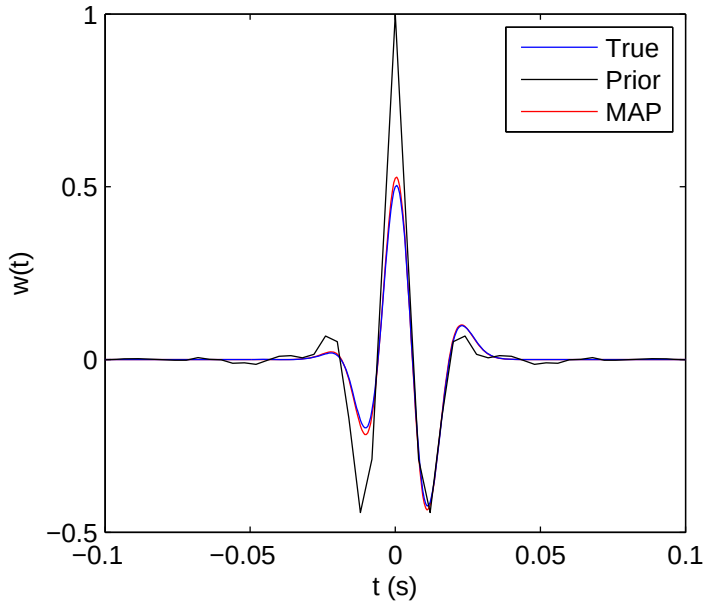


Figure 12: Prior mean wavelet (black), MAP wavelet estimate (red), and true wavelet (blue).

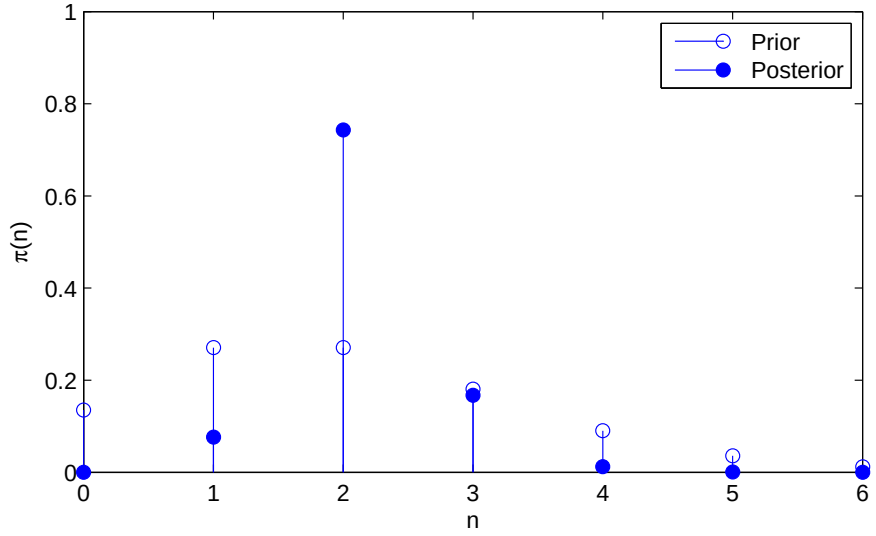
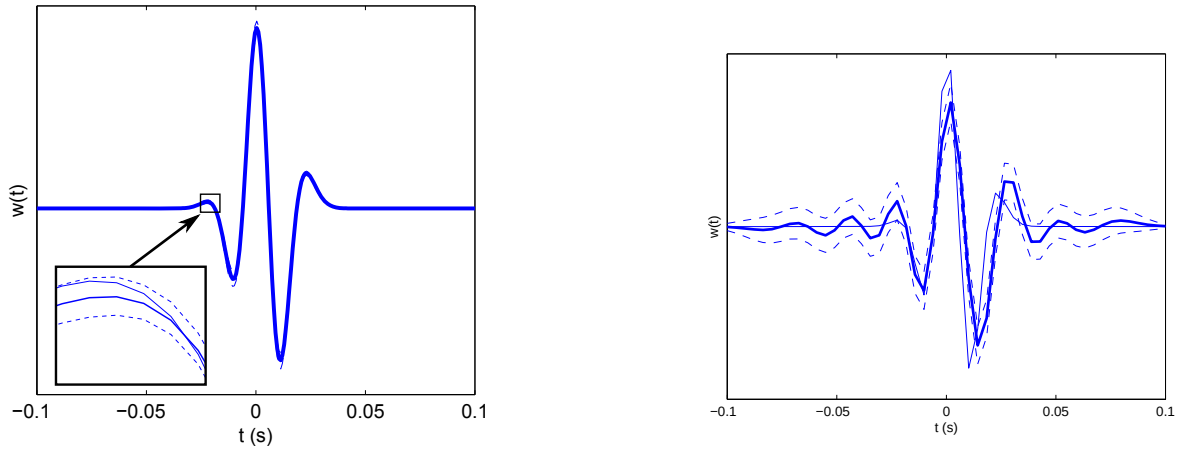


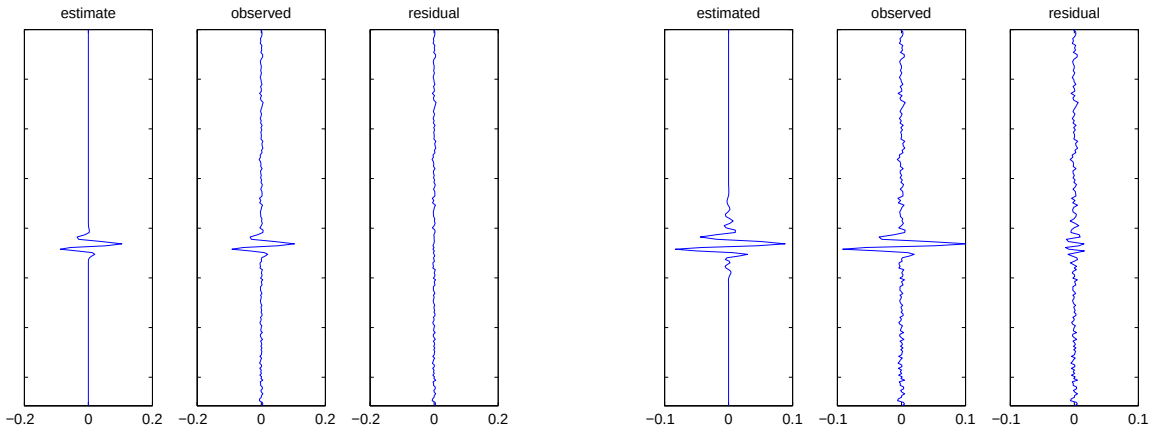
Figure 13: Prior and posterior distributions of the wavelet order n. A priori, n is Poisson distributed with mean $\lambda = 2$. The posterior distribution is approximated by normalising the values of the (unnormalised) posterior distribution at the mode associated with each order, within the range shown ($0 \leq n \leq 6$).



(a) MAP **(b)** BLS

Figure 14: Maximum a posteriori (MAP) and Bayesian least squares (BLS) wavelet estimates (thick solid lines) with approximate 95% confidence intervals (dashed lines) compared with true wavelet (thin solid lines). The curves in the MAP case are too close together to be easily distinguishable, so an enlarged view is included.

for $\boldsymbol{w}$, and the confidence envelope can be seen to taper off towards the edges of the figure. The wavelet estimate has largely succeeded in reconstructing the shape of the wavelet, but it has also picked up some oscillations which are not part of the true wavelet. This indicates that some degree of overfitting is taking place. Generally speaking, the point-wise sample representation approach of the BLS estimation procedure is much more generic and flexible than any parametric wavelet representation, but this flexibility also makes the estimate more sensitive to the influence of noise. This problem can be mitigated somewhat by choosing a smaller value of the parameter ξ in (66), to further penalise deviations from the mean. This cannot generally be done, however, without also affecting the estimation of the desired part of the wavelet.



(a) MAP **(b)** BLS

Figure 15: Estimated traces, $\hat{\boldsymbol{y}}$, observed traces $\boldsymbol{y}$ and residual traces $\boldsymbol{y} - \hat{\boldsymbol{y}}$ for MAP and BLS wavelet estimates.

Figure 15 shows the observed seismic trace $\boldsymbol{y}$, compared with the estimated seismic trace $\hat{\boldsymbol{y}}$, which is computed by convolving the reflectivity series with the estimated wavelet, i.e. by applying the forward model to the wavelet estimate. This is done separately for the MAP and BLS estimates.

Finally, the residual trace $\boldsymbol{y} - \hat{\boldsymbol{y}}$ is included in the figure. Ideally, we would like the residual trace to be as small as possible. For both models, minimising $\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2^2$ is equivalent to maximising the likelihood $\pi(\boldsymbol{y}|\boldsymbol{w})$. The residual trace of the MAP estimate has small values everywhere, while the BLS residual has larger values in a region near the reflector. This localised error is caused by the overfitting of the wavelet to the noisy signal.

6.2 Semi-synthetic data, Ricker and box wavelets

The data for the following test cases are semi-synthetic in the sense that the reflectivity series is computed from a real well log, but the seismic traces are synthesised by convolving $\boldsymbol{r}$ with known wavelets and adding noise. Unlike the true wavelet used in the first test case (section 6.1), these wavelets were not created using the parametric model. Instead, they were generated by computing the inverse discrete Fourier transform of pre-specified amplitude spectra, and then, in some of the cases, applying a phase rotation to the computed time-domain signal.

Figures 16 and 17 show MAP wavelet estimates and posterior wavelet samples. A Ricker wavelet was used to generate the seismic data for this test case, and it is shown in red in the plot. In Figure 16 the true wavelet is a zero phase, symmetric Ricker wavelet, and in Figure 17 it is phase rotated and skewed. The amplitudes of the estimated wavelets differ somewhat from the amplitudes of the true wavelets. To facilitate comparison between estimated and true wavelets in terms of other wavelet properties, such as length and skewness, some plots are given with actual estimated wavelet amplitudes, while others feature normalised wavelets. In other words, in some of the plots in Figures 16–19, the true wavelet and the MAP wavelet estimate have both been scaled to unit peak amplitude, and the wavelet samples from the posterior distribution have been scaled with the same scale factor as the MAP wavelet, to preserve dispersion relative to amplitude.

When estimating the zero-phase Ricker wavelet, the $n = 1$ wavelet gives the best match in terms of shape. Both the $n = 1$ and $n = 2$ wavelets fail to correctly match the amplitude of the Ricker wavelet, particularly at the peak, although the $n = 2$ wavelet does have approximately correct values in the troughs at $t = \pm 0.02$ s.

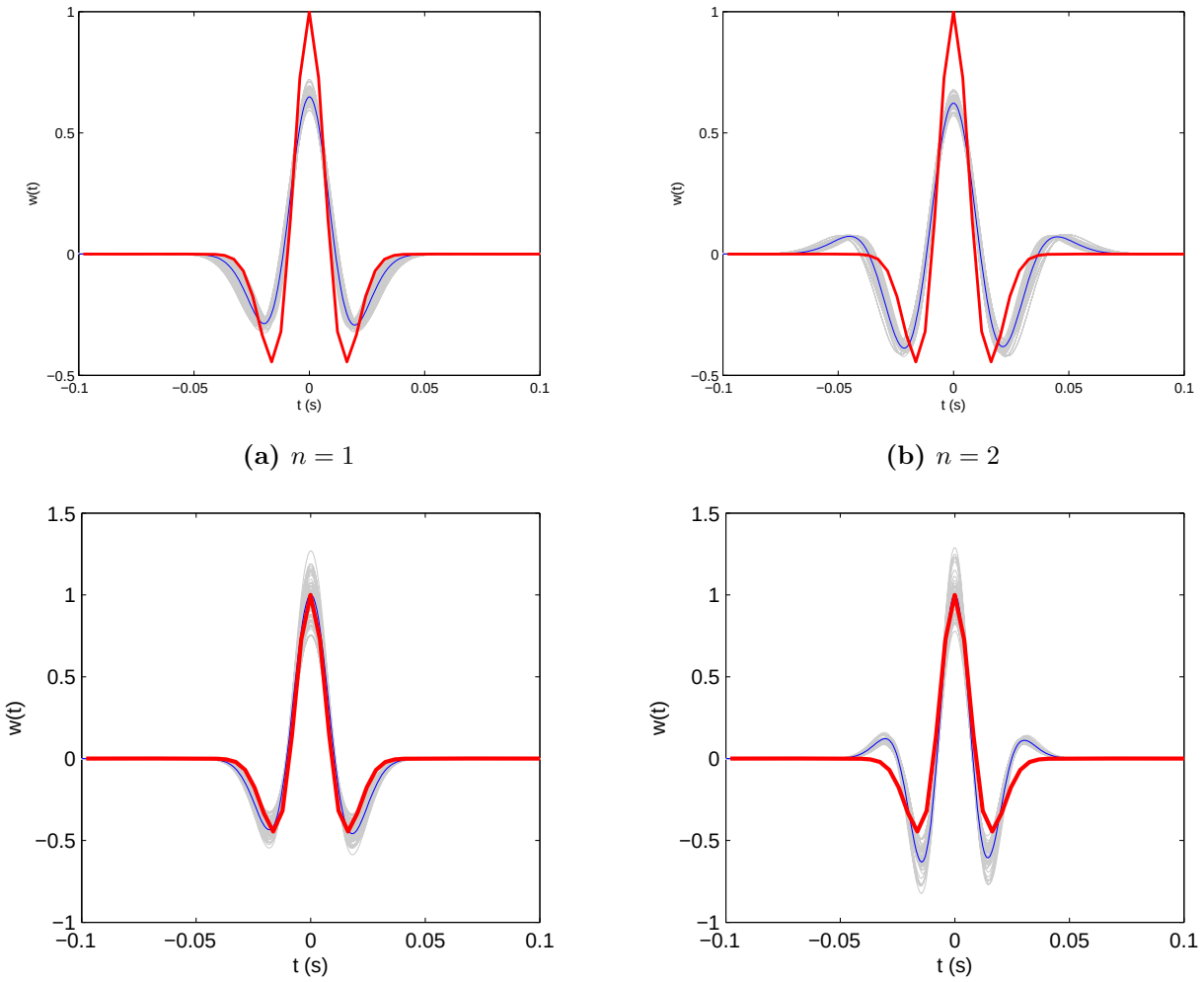
The situation for the phase-rotated Ricker wavelet in Figure 17 is similar, but here the left trough is shallower than the right one due to the skewness caused by the phase rotation. Both estimated wavelets ($n = 1$ and 2) fit the true wavelet more closely on the left side than on the right side of the central peak. The $n = 2$ estimate has the more correct shape to the right of the deepest trough, where the wavelet has a slight local maximum. On the other hand, it also introduces a similar local maximum on the left side, where no such maximum is present in the true wavelet. The amplitude and skewness parameters are both underestimated in this case.

Figures 18 and 19 show MAP wavelet estimates and posterior samples, estimated with various wavelet orders between $n = 1$ and $n = 3$. The true wavelet, shown in red, is a box wavelet. It has a box shaped amplitude spectrum, constant on some passband, with a relatively sharp (10–20 Hz) cut-off at either end. The time signal corresponding to this amplitude spectrum is dominated by a large central peak, surrounded by much smaller ripples. These ripples have a jagged appearance, with some sections looking approximately linear, while others are curvier.

As in the Ricker wavelet-case, two different wavelets are used to generate traces. In Figure 18, the box wavelet is zero-phase and symmetrical, whereas in Figure 19 the phase is rotated, and the wavelet is skewed.

The most salient characteristic of the wavelet estimates in Figures 18 and 19 is their strikingly small amplitude. If we think of the box wavelet as consisting of two components with different amplitudes, the large central peak, and the much smaller oscillations surrounding it, then the amplitudes of the wavelet estimates seem to fit the second, smaller component, but not the first and largest component. As a consequence, the overall quality of the fit suffers greatly.

Taking note of the ripples present in the box wavelet, we might try increasing n, since this adds more oscillations to the parametric wavelet. But using a too large value for n risks causing the dilation parameter to be grossly underestimated, so that all the oscillations of the wavelet are located within the timespan of the central peak of the box wavelet. Comparing the wavelet estimates with the box wavelet, we notice an important qualitative difference. The ripples in the box wavelet



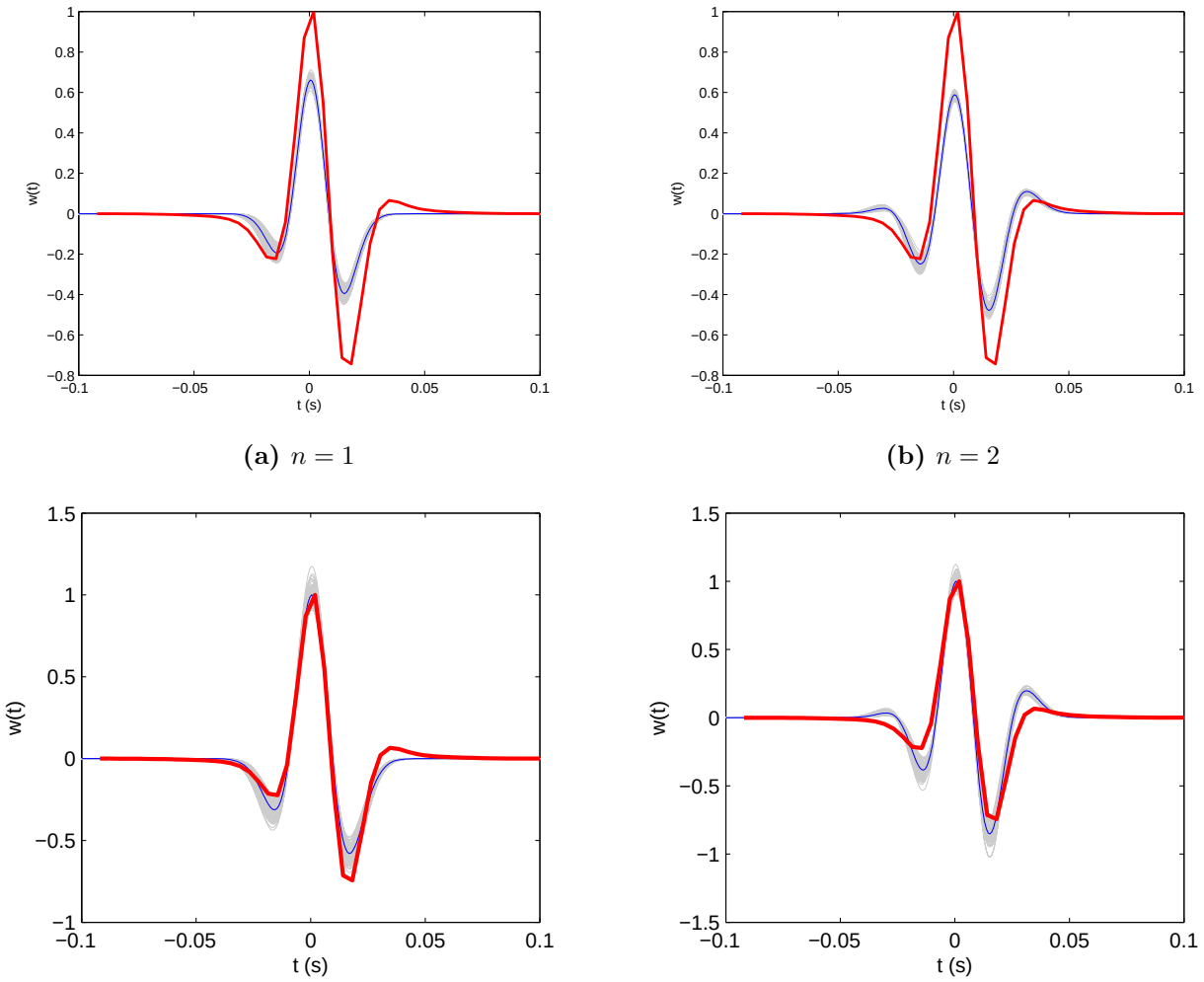
(a) $n = 1$

(b) $n = 2$

(c) $n = 1$, normalised

(d) $n = 2$, normalised

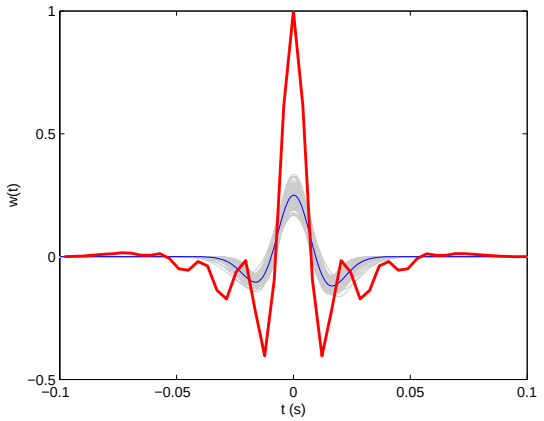
Figure 16: MAP wavelet estimates (blue) and posterior samples (grey) with $n = 1$ and $n = 2$. Seismic data for this test case was generated using the symmetric Ricker wavelet, shown in red. **Top:** Actual estimated amplitude. **Bottom:** Normalised amplitude. Note: The wavelet estimates in the top and bottom rows were estimated independently.



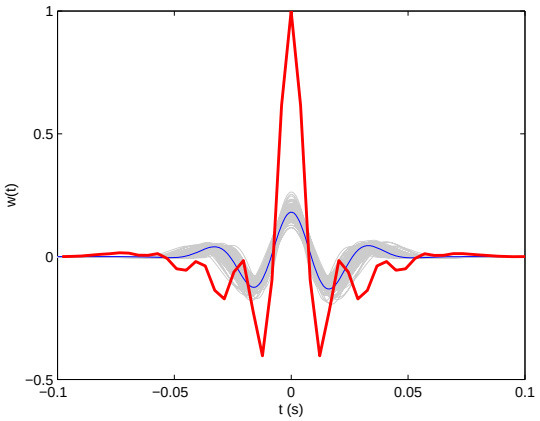
(a) $n = 1$ **(b)** $n = 2$

(c) $n = 1$, normalised **(d)** $n = 2$, normalised

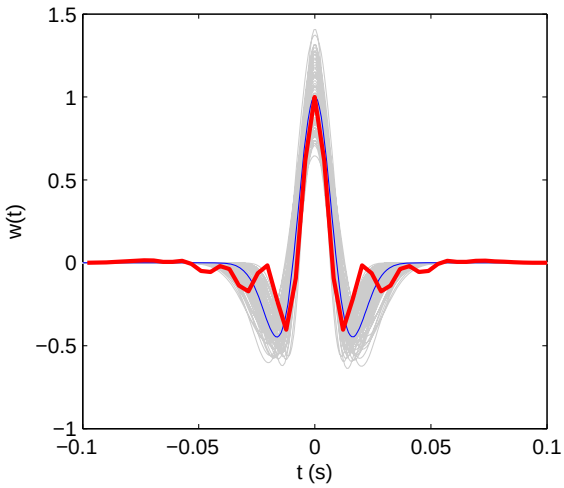
Figure 17: MAP wavelet estimates (blue) and posterior samples (grey) with $n = 1$ and $n = 2$. The seismic data for this test case was generated with a phase rotated Ricker wavelet, shown in red. **Top:** Actual estimated amplitude. **Bottom:** Normalised amplitude. Note: The top and bottom rows show independently estimated wavelets.



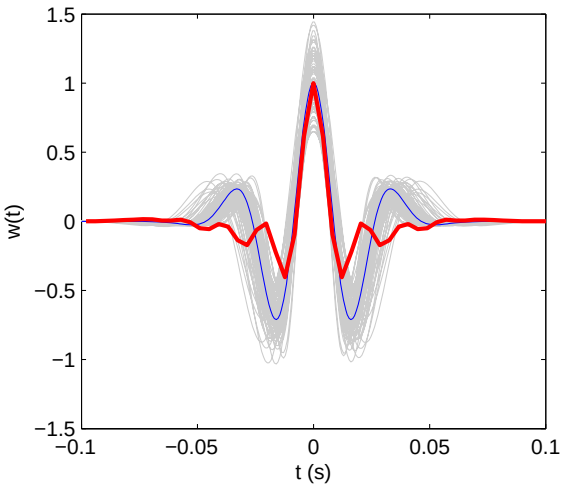
(a) $n = 1$



(b) $n = 3$

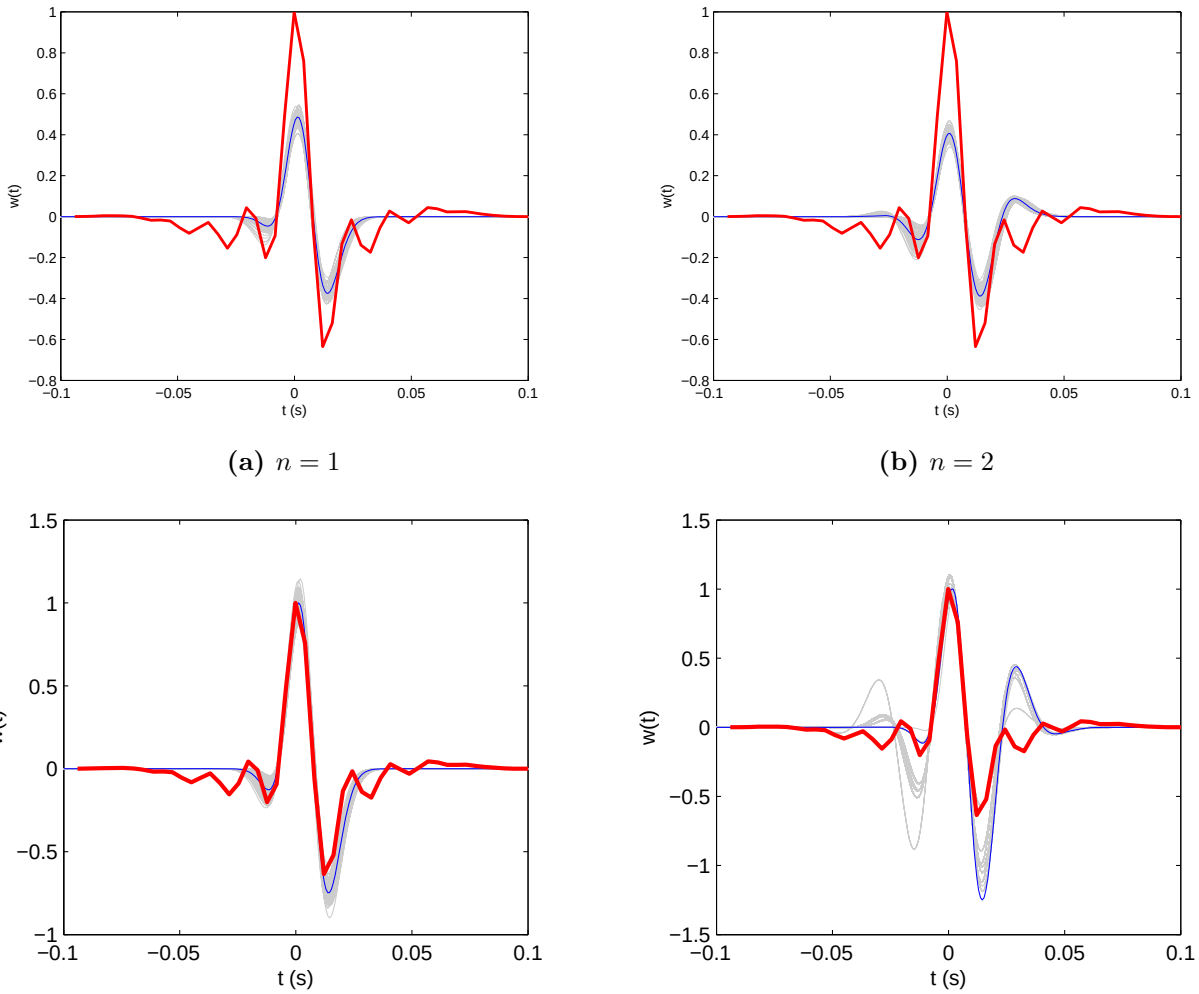


(c) $n = 1$, normalised



(d) $n = 3$, normalised

Figure 18: MAP wavelet estimates (blue), wavelet samples from posterior distribution (grey), and true, zero-phase box wavelet (red). Estimates are shown for $n = 1$ and $n = 3$. The wavelet model is unable to adapt well to the particular shape of the box wavelet, resulting in a relatively poor fit. **Top:** Wavelets with estimated amplitude. **Bottom:** Wavelets with normalised amplitude; wavelet amplitudes have been adjusted so that the maximum value is 1. Note: The normalised and unnormalised wavelets were estimated separately, in independent runs of the estimation procedure.



(a) $n = 1$ **(b)** $n = 2$

(c) $n = 1$, normalised **(d)** $n = 2$, normalised

Figure 19: MAP wavelet estimates (blue), posterior wavelet samples (grey) and true, phase-rotated box wavelet (red). Results are shown for wavelet orders $n = 1$ and $n = 2$. The wavelet estimates shown here suffer from the same problem as the ones in Figure 18, the parametric wavelet model is unable to match the shape of the box wavelet. **Top:** Wavelets shown with estimated amplitudes. **Bottom:** All wavelets normalised to unit peak amplitude. Note: The wavelets shown in the top and bottom rows were estimated independently.

are primarily found below zero, while the off-peak oscillations in the parametric wavelet are centred on zero, so that a significant portion of the lobes are positive. If the minima immediately adjacent to the central peak in a high-order parametric wavelet were to be matched to the corresponding troughs of the box wavelet, then the neighbouring local maxima of the parametric wavelet would necessarily become too large to fit the corresponding local maxima of the box wavelet, which are quite small. This is an example of model inflexibility.

These tests, using the Ricker and box wavelets, reveal an important limitation of the parametric wavelet model. Although relatively flexible considering the low number of parameters involved, the model is still only capable of representing a limited range of wavelet shapes. In cases like these, where a true wavelet exists, this target wavelet may either be within the range of the model, like the wavelet in Figure 10, or it may be outside that range. Judging by the lack of fit exhibited by the wavelet estimates in Figures 18 and 19, it appears that the box wavelet, phase rotated or not, falls into the latter category.

The Ricker wavelet, which is very smooth compared with the box wavelet, is certainly closer to the range of the parametric wavelet model than the box wavelet is. This is reflected by the superior quality of the fit in Figures 16 and 17 compared with the fit in Figures 18 and 19. Still, the estimates of the Ricker wavelet leave something to be desired, and it is clear that the wavelet model is unable to provide a complete and accurate representation of the Ricker wavelet, so that it, too, is ostensibly outside the model's range.

To create a test case for multi-angle estimation, we compute reflectivity series $\boldsymbol{r}(\theta)$ for three different angles θ_1, θ_2 and θ_3. Then we generate near, mid and far seismic traces $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ and $\boldsymbol{y}_3$ by convolving each reflectivity series with the same phase rotated Ricker wavelet used in the single-angle test case.

Figure 20 shows MAP wavelet estimates resulting from joint wavelet estimation from the near, mid and far traces. Also shown are wavelet samples from the posterior distribution. Figure 21 shows the prior mean used for estimation, and samples from the prior distribution. Finally, Figure 22 shows histograms of the amplitude and dilation parameters of the near wavelet, taken from the posterior samples shown in Figure 20.

The multi-angle wavelet estimates display many of the same problems as the single-angle estimates in Figure 17. They underestimate the amplitude, and do not match the shape of the true wavelet consistently throughout the estimation time interval. As was the case for the single-angle estimates, the match is significantly better near the left trough (at $t \approx -0.2$) than on the right side.

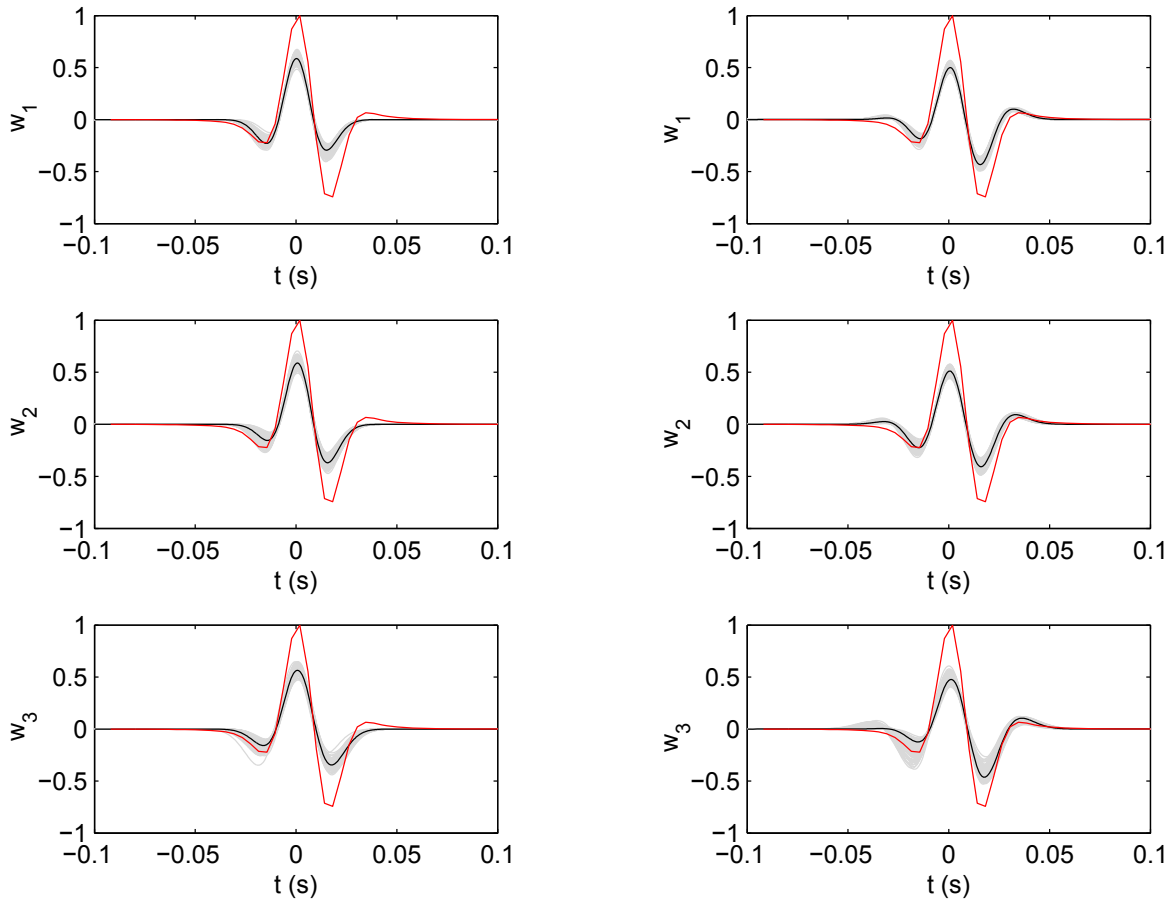
When generating the seismic data for the multi-angle test case, the same true wavelet was used for all three traces. In a realistic setting, one would expect to find differences between wavelets corresponding to different angles of incidence. Specifically, a larger angle of incidence implies that seismic waves traverse longer distances in the subsurface, and hence have their high-frequency components attenuated to a greater degree. Although this effect is corrected for in data processing, we might still expect the far wavelet to be smoother and less oscillatory, and perhaps longer, than the near and mid wavelets. Similarly, we might expect the mid wavelet to be smoother and longer than the near wavelet. In this test case, however, the near, mid and far angle stacks all have the same true wavelet. As is evident from the Figure 20, the wavelets do differ slightly from each other, mostly in terms of skewness. This variation is allowed, to some extent, since the between-angle correlation coefficient in the prior distribution of $\boldsymbol{\beta}$ is set to $\rho = 0.9$ in (56).

6.3 Real data example

In this final test case, the well log used to create the reflectivity series and the seismic data are both real, which means that no true wavelet is available for validation of estimates.

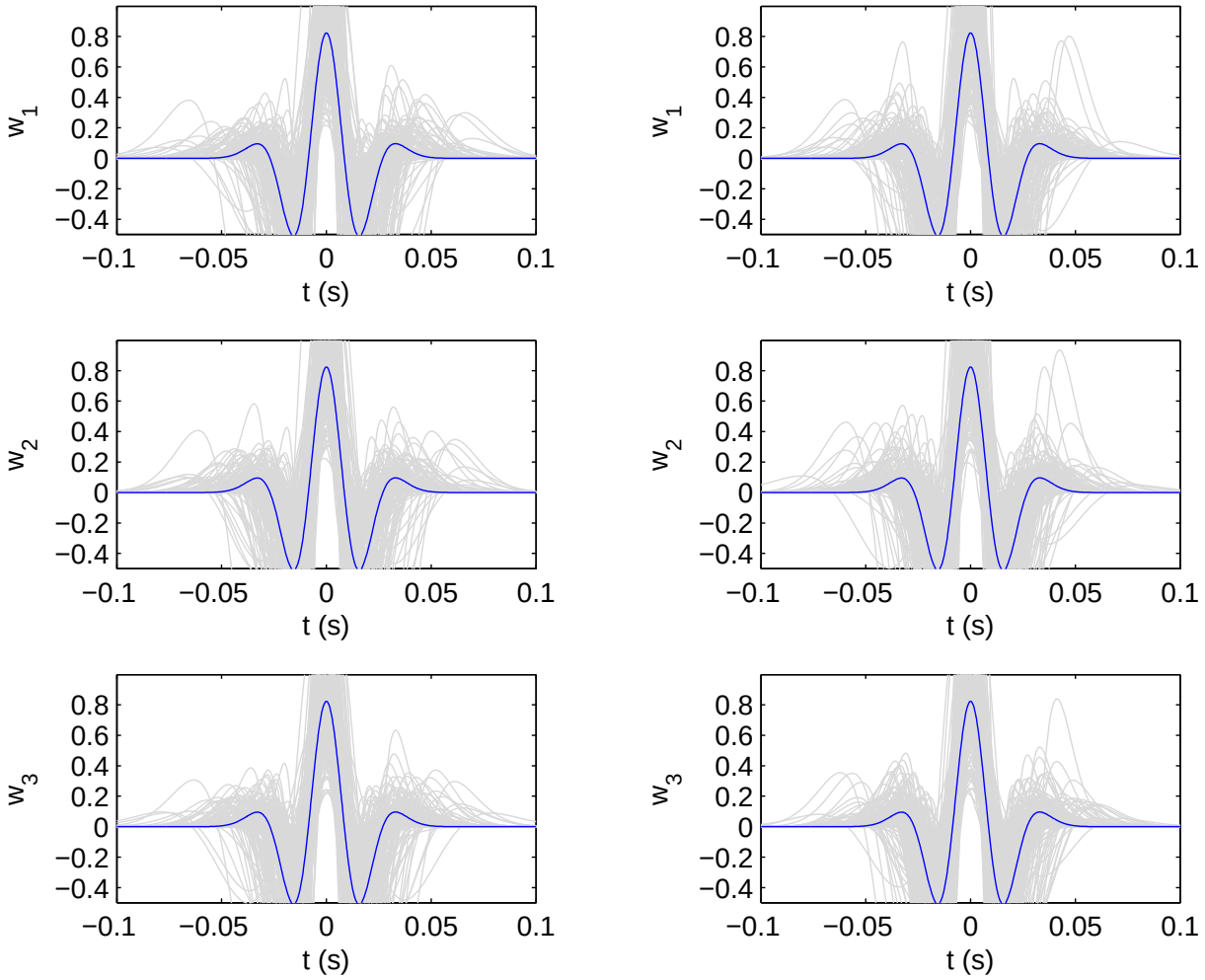
We estimate MAP wavelets using $n = 1$ and $n = 2$. The wavelet estimates, along with samples from the posterior distributions, are shown in Figure 23. Histograms of posterior samples of the skewness and amplitude parameters are plotted in Figure 24, together with prior probability densities.

A large prior variance was used for the amplitude, and the effect of this on the posterior distribution can be seen on the wavelet samples in Figure 23. Although in both the $n = 1$ and $n = 2$ case,



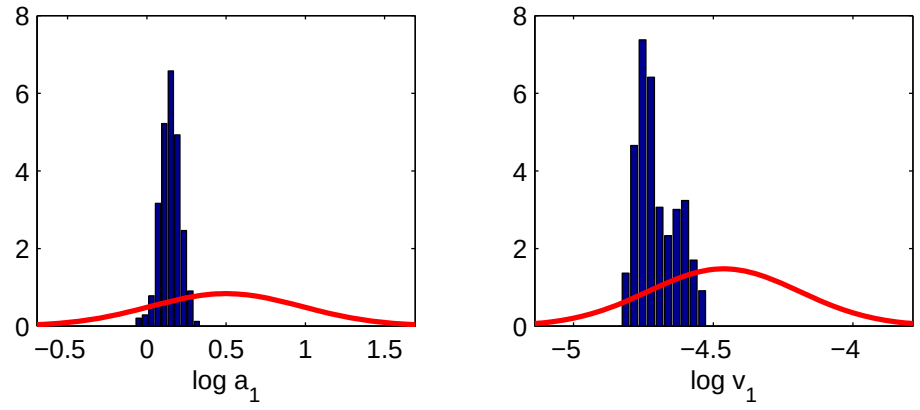
(a) $n = 1$ **(b)** $n = 2$

Figure 20: MAP wavelet estimates (black), posterior samples (grey) and true Ricker wavelet (red). Prior correlation between corresponding parameters of different wavelets is $\rho = 0.9$.

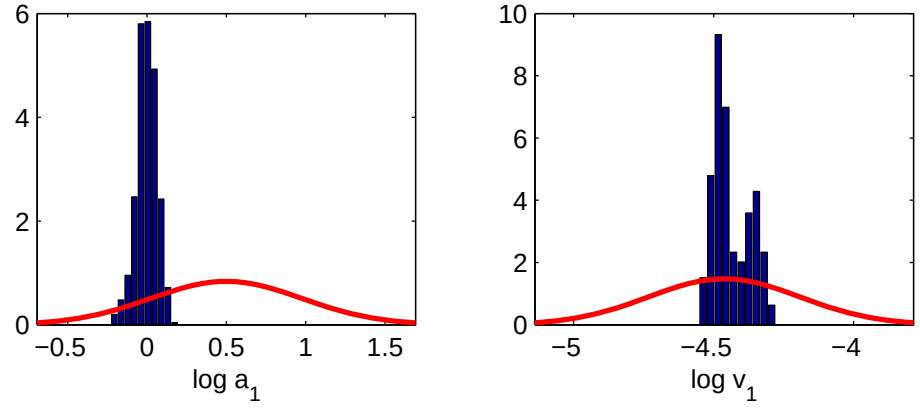


(a) $n = 1$ **(b)** $n = 2$

Figure 21: Prior mean wavelet (blue) and posterior samples from multi-angle distribution (grey).

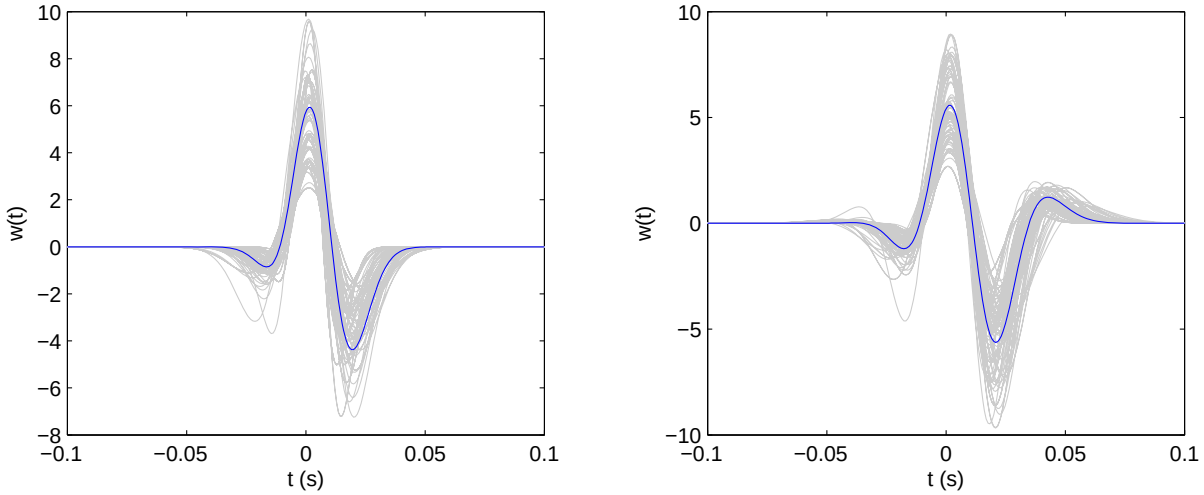


(a) $n = 1$



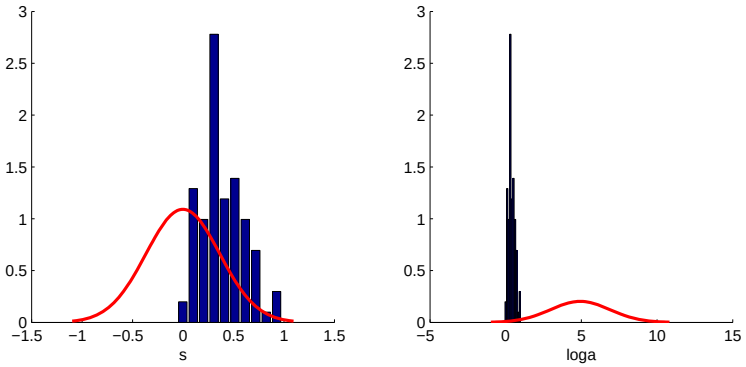
(b) $n = 2$

Figure 22: Prior densities (red) and histograms of posterior samples (blue bars) for the amplitude and dilation parameters of the near wavelet w_1, with $n = 1$ and $n = 2$.

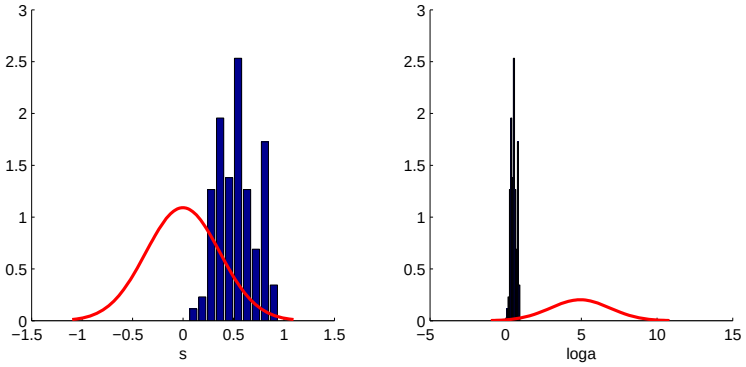


(a) $n = 1$ **(b)** $n = 2$

Figure 23: MAP wavelet estimates for the real test case (blue) with samples from the posterior distribution (grey), using $n = 1$ and $n = 2$.



(a) $n = 1$



(b) $n = 2$

Figure 24: Histograms of skewness parameter s and skewness parameter $\log a$ from posterior samples, with $n = 1$ and $n = 2$. Prior probability densities are shown as red lines.

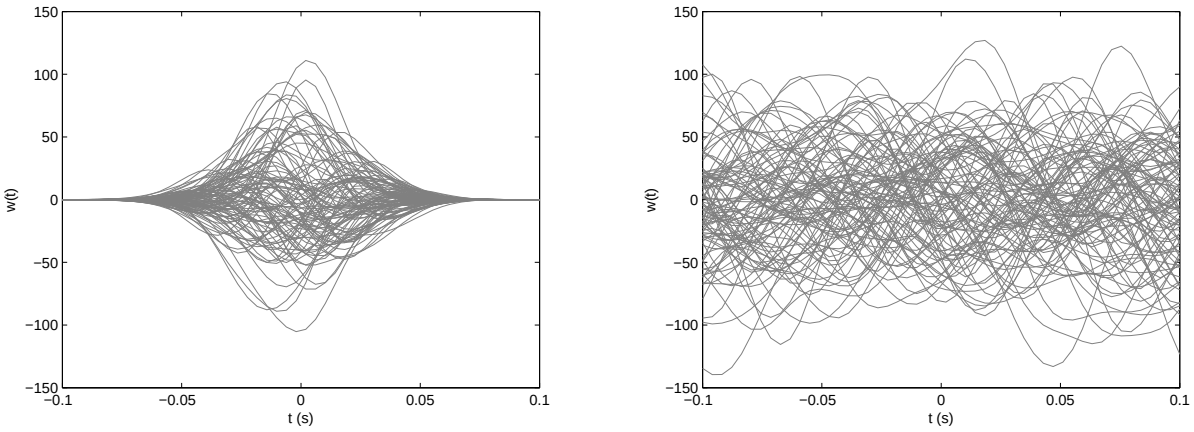
the posterior distribution is considerably more concentrated than the prior, as Figure 24 shows, the wavelet samples cover a relatively wide range of amplitudes.

Since no true wavelet is available to compare the wavelet estimates with, we estimate additional wavelets using the Bayesian Least Squares (BLS) estimator. The choice of prior distribution has a large impact on the appearance of the BLS estimate, so we estimate two BLS wavelets, one using the heteroscedastic prior with larger variance in the middle of the time interval, and another using the homoscedastic prior with constant variance. The covariance parameters in (66) are chosen as follows:

$$\gamma = 40, \quad \tau = 7, \quad \xi = \frac{n_w}{6} = \frac{50}{6},$$

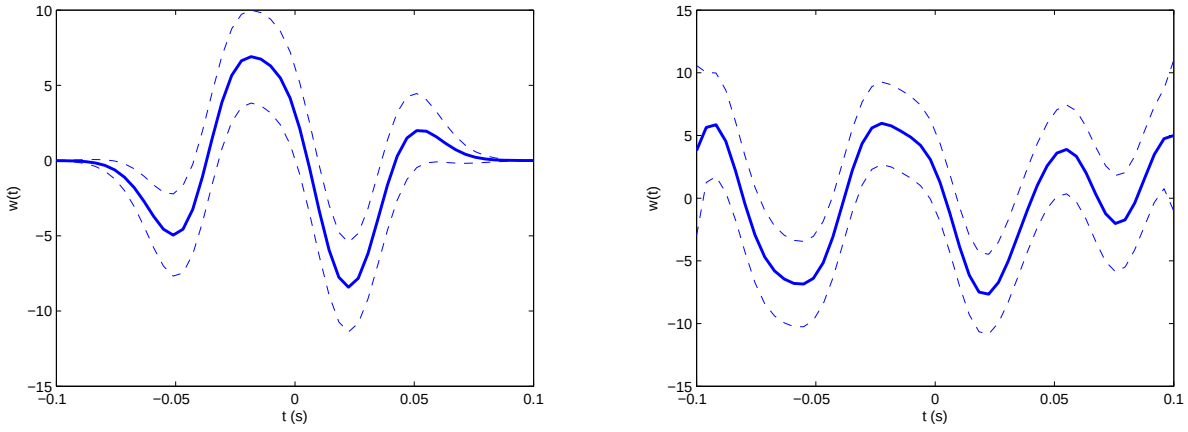
and the standard deviation of the noise in $\boldsymbol{y}$ is estimated as $\sigma = 0.05\|\boldsymbol{y}\|_2$.

The two prior distributions are visualised in Figure 25 by samples, and the wavelet estimates are shown in Figure 26, with two standard deviation-confidence envelopes.



(a) Heteroscedastic prior **(b)** Homoscedastic prior

Figure 25: Samples from the prior distribution of the Bayesian Least Squares model, using both types of prior distribution.



(a) BLS wavelet, heteroscedastic prior **(b)** BLS wavelet, homoscedastic prior

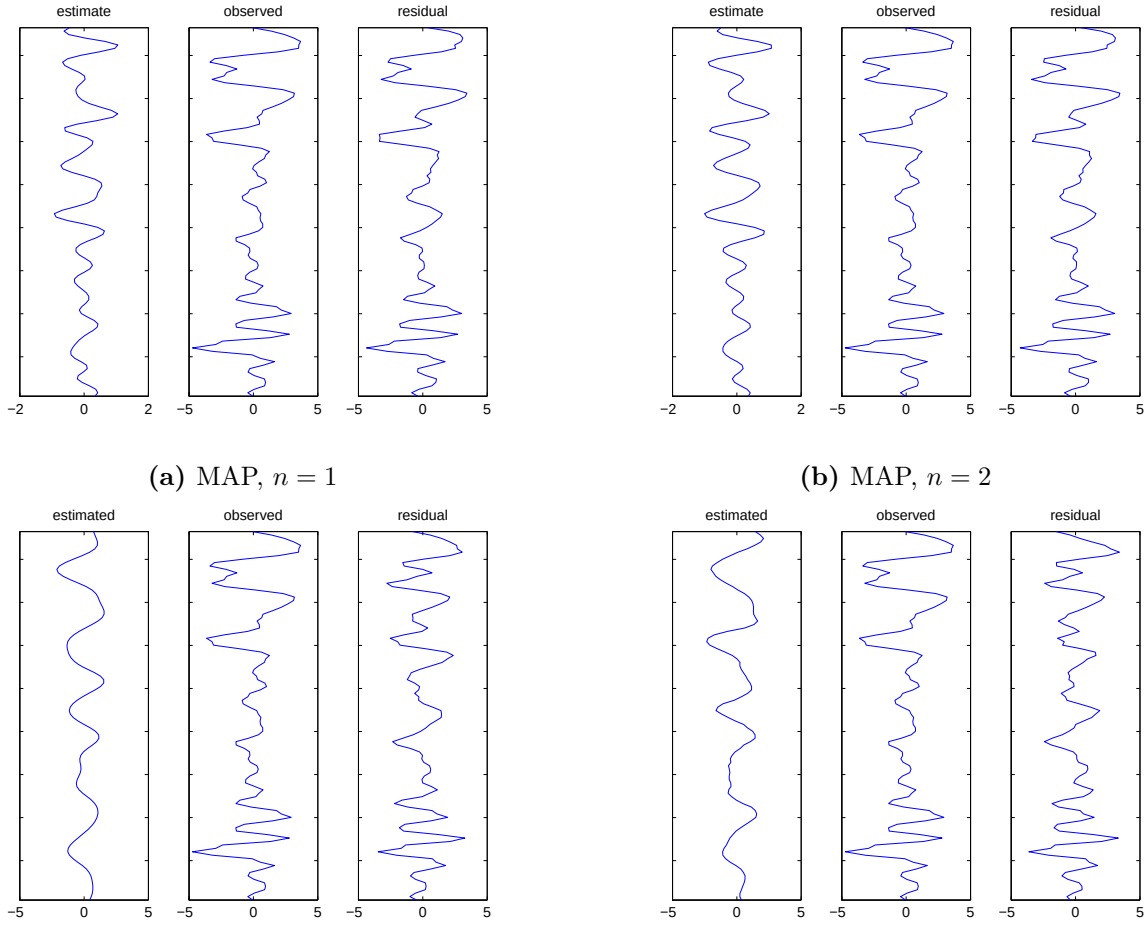
Figure 26: Bayesian Least Squares estimates of the real test case-wavelet (thick line), and two standard deviation-confidence limits (dotted lines), estimated using two different types of prior distribution.

There is a resemblance between the heteroscedastic prior-BLS estimate (25a), and the $n = 2$ MAP estimate 23b). It should be noted, however, that the BLS-wavelet is longer, its duration

is about 160 ms, whereas the duration of the MAP wavelet is about 120 ms. Nevertheless, the similarity in shape between the two independently estimated wavelets increases our confidence in the estimates.

Comparing the two BLS estimates, we see that the one using the homoscedastic prior features more oscillations, especially at the edges. Eliminating or damping these oscillations at the edges, while leaving the middle part of the wavelet relatively unchanged, is exactly the effect we expect the heteroscedastic prior to have. In one sense, using a prior distribution which essentially removes part of the estimated wavelet, may seem arbitrary. But on the other hand, as illustrated by the BLS estimate in Figure 14, the BLS estimator, with its large number of parameters, is sensitive to noise and overfitting. Using the heteroscedastic prior is an effective means of mitigating that sensitivity.

In Figure 27, we compare the seismic trace $\boldsymbol{y}$ to the synthetic seismic traces $\hat{\boldsymbol{y}}$ obtained by convolving the reflectivity series with the estimated wavelets. We get one synthetic trace for each wavelet estimate. As in Figure 15, we also plot the residual $\boldsymbol{y} - \hat{\boldsymbol{y}}$ for each estimate.



(a) MAP, $n = 1$

(b) MAP, $n = 2$

(c) BLS, heteroscedastic prior

(d) BLS, homoscedastic prior

Figure 27: Estimated, observed and residual traces for MAP wavelet estimates with $n = 1$ and $n = 2$, and for BLS wavelet estimates with the heteroscedastic and homoscedastic priors, in the real data test case.

The synthetic traces for the MAP wavelets have small amplitudes, suggesting that the wavelet amplitude may be underestimated. There are no major differences between the $n = 1$ and $n = 2$ traces, which is expected, since the wavelets are similarly shaped.

The traces for the BLS wavelets are smoother, with slower variation and less details. This observation can also be related to the wavelet shape, since the BLS wavelets are longer than the MAP wavelets.

In Figure 28, we apply the complete forward model to ensembles of wavelets from the posterior

distribution associated with each estimate. That is, we convolve each sampled wavelet with the reflectivity series, and add independent Gaussian noise according to (1), using the estimated standard deviation for the noise. This creates an ensemble of simulated seismic traces for each wavelet estimate, which we then compare to the actual observed trace.

In all four cases, the band of simulated traces covers the blue line, which represents the observed trace, well. The MAP bands have an almost constant width, while the BLS bands follow the variations in the observed trace. This difference is likely related to the difference in the synthetic traces in Figure 27, where the BLS traces have large amplitudes and slow variations, which are visible through noise, while the MAP traces have smaller amplitudes and faster oscillations, which may be masked by noise. The noise level itself is set based on the available estimate of the standard deviation of the noise in the observed trace. In the BLS case, this is simply estimated as five percent of the norm of $\boldsymbol{y}$, or 0.86. In the MAP case, it is estimated as a model parameter, and the estimated value is 1.61.

$\rho = 0$	Near	Mid	Far
s	0.000 (−0.826,1.047)	−0.001 (−1.016,0.949)	−0.001 (−0.988,0.916)
a	1.426 (0.546,4.189)	1.433 (0.477,3.745)	1.449 (0.552,3.768)
v	0.009 (0.004,0.023)	0.009 (0.004,0.020)	0.009 (0.005,0.021)
σ	3.309 (2.720,3.943)	2.825 (2.374,3.462)	2.328 (1.982,2.804)
$\rho = 0.5$	Near	Mid	Far
s	0.000 (−1.095,1.214)	−0.001 (−1.087,1.129)	−0.001 (−0.771,1.052)
a	1.434 (0.496,3.991)	1.438 (0.508,3.552)	1.447 (0.523,4.149)
v	0.009 (0.004,0.020)	0.009 (0.003,0.024)	0.009 (0.004,0.020)
σ	3.305 (2.772,4.029)	2.822 (2.389,3.435)	2.327 (1.970,2.841)
$\rho = 0.98$	Near	Mid	Far
s	0.000 (−1.003,1.098)	−0.001 (−0.915,1.099)	0.000 (−0.890,1.081)
a	1.450 (0.486,3.817)	1.450 (0.490,4.215)	1.450 (0.495,3.870)
v	0.010 (0.004,0.022)	0.010 (0.004,0.022)	0.010 (0.004,0.022)
σ	3.263 (2.722,3.979)	2.819 (2.342,3.457)	2.357 (1.977,2.802)

Table 1: MAP estimates and empirical 95% confidence intervals for the parameters s, a, v and σ of the near, mid and far wavelets using between-angle correlation coefficients $\rho = 0$, $\rho = 0.5$ and $\rho = 0.98$.

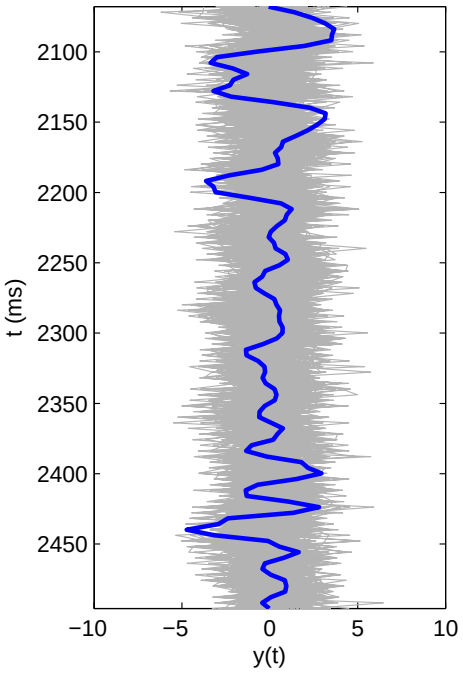
To investigate the effect of the choice of correlation coefficient, discussed in section 5.3, we jointly estimate parameters for near, mid and far wavelets, from multi-angle data with three seismic traces, and we repeat the estimation using three different values of the correlation coefficient ρ. The location, and the well log, are the same as before. MAP parameter estimates are given in Table 1, along with empirical 95% confidence intervals computed from samples. There is no clearly distinguishable pattern in the parameter estimates, nor in the widths of the confidence intervals. If the choice of correlation coefficient has an effect, it is hard to detect here. At any rate, it might be more interesting to consider other correlation structures.

7 Discussion and conclusions

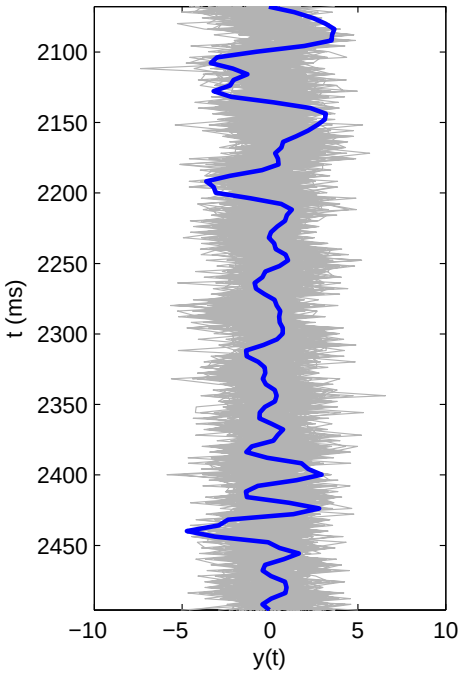
In this section, we discuss the results of the tests presented in section 6, offering comments on what works well and what doesn't. We also give brief mention to topics relevant to, but not covered in, this work, including possible extensions of the estimation procedure.

7.1 Model flexibility

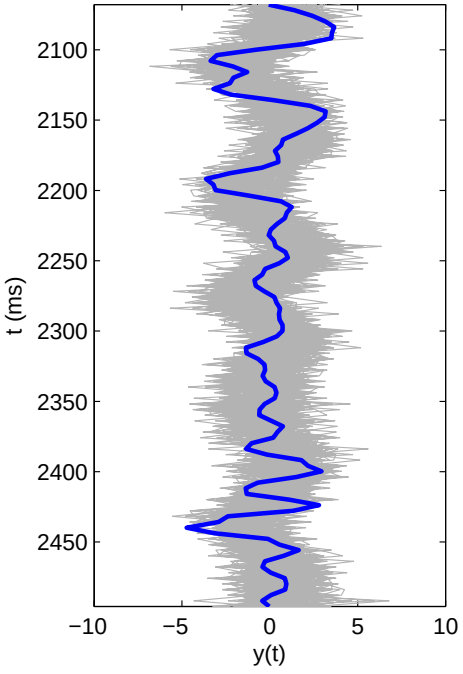
One lesson learned from the various test results, is that the model could benefit from being more flexible. In this context, flexibility means the model's ability to adapt to, or match, a wide range of



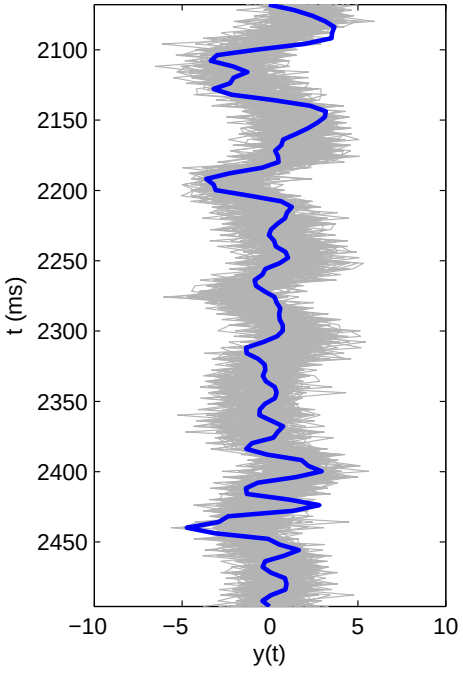
(a) MAP, $n = 1$



(b) MAP, $n = 2$



(c) Bayes LS, heteroscedastic prior



(d) Bayes LS, homoscedastic prior

Figure 28: Ensembles of traces generated from sampled wavelets by applying the forward model to each wavelet. Blue lines show observed trace.

different wavelet shapes. One way to think about this is to consider the parametric wavelet model as a function mapping points in "parameter space" to points in "wavelet space". Suppose we are using a time vector $\boldsymbol{t} \in \mathbb{R}^{n_w}$ to discretise the wavelet $w(t; \boldsymbol{\beta})$, where $\boldsymbol{\beta} \in \mathbb{R}^{n_p}$ is the parameter vector. Then the discretised wavelet is

$$\boldsymbol{w}(\boldsymbol{\beta}) = (w(t_1; \boldsymbol{\beta}), w(t_2; \boldsymbol{\beta}), \ldots, w(t_{n_w} \boldsymbol{\beta}))^T \in \mathbb{R}^{n_w},$$

where $t_1 \ldots, t_{n_w}$ are the elements of $\boldsymbol{t}$. If we regard the time vector as fixed, we can think of the parametrised wavelet as a function

$$w : \mathbb{R}^{n_p} \to \mathbb{R}^{n_w}$$

whose domain is the parameter space $\mathbb{R}^{n_p}$, containing all possible input combinations, and whose codomain is the wavelet space $\mathbb{R}^{n_w}$, containing all possible wavelets under this discretisation.

Perhaps the greatest strength of the parametric approach to the wavelet estimation problem, comes from the fact that the number of parameters, n_p, can be so much smaller than the number of wavelet samples n_w. In point-wise estimation, such as Bayesian Least Squares, each wavelet sample point is treated as a separate parameter to be fitted to the data, which gives the model many degrees of freedom. As demonstrated by some of the test cases, this can easily lead to overfitting, unless measures are taken specifically to counteract the model's sensitivity to noise. A related issue is the problem of determining the length of the wavelet, since a longer wavelet will typically achieve a better fit (Gunning and Glinsky, 2006). Ideally, belief about the wavelet length should be encoded as prior information.

Using a small number of parameters has a cost, however, in that not all points in wavelet space will be accessible to the wavelet function. In other words, the image of $w : \boldsymbol{\beta} \mapsto \boldsymbol{w}(\boldsymbol{\beta})$ will be a limited subset of $\mathbb{R}^{n_w}$. While a point-wise wavelet with each sample point as a separate parameter is able to take on almost any shape, a parametric representation is limited to a much smaller set of possible shapes; the fewer parameters used, the more limited it is.

Suppose now that there exists a true wavelet, which, using the discretisation implied by the time vector $\boldsymbol{t}$, is best represented by some vector $\boldsymbol{w}^* \in \mathbb{R}^{n_w}$. If this vector is in the image of the parametric wavelet function, then all is well, and we can hope to estimate the correct wavelet shape. Conversely, if there is no $\boldsymbol{\beta}$ such that $\boldsymbol{w}(\boldsymbol{\beta}) = \boldsymbol{w}^*$, then the best we can hope for is to estimate some closest point inside the model's range.

In practice, the range of the model is determined not only by the form of the wavelet parametrisation, but also by the prior distribution. When we specify the prior distribution, we are, in effect, choosing where to look for the best-fitting wavelet. If the prior distribution is very concentrated, so that the probability mass of the posterior distribution is essentially supported on only a small region of the parameter space, then the choice of prior distribution will contribute to limiting the model's range.

For any modelling approach, there is a trade-off between flexibility on the one hand, and parsimony and robustness to noise on the other hand. Comparing the parametric model to the BLS estimator provides a good example of this. Neither approach is obviously superior.

7.2 Stability of optimisation

Another finding from the test results is that the stability of the numerical optimisation of the posterior distribution depends on the prior distribution as well as on the data. By stability, we mean the ability of the optimisation algorithm to reliably converge to, and locate, the desired extreme point of the objective function.

While the need for model flexibility puts a limit on how concentrated the prior distribution can be, the need for stability effectively limits how diffuse it can be. This is because information from the prior distribution is necessary to give the posterior distribution a shape with a clearly isolated mode.

On synthetic data, with perfect alignment of reflectivity and seismic data, low noise level and a true wavelet which can be represented by the parametrisation, the data likelihood will have a clear maximum, and the shape of the prior distribution is not of critical importance for the success of the

optimisation. On realistic data, however, straightforward likelihood maximisation, where the prior distribution is non-informative, and hence does not contribute information to the posterior distribution, will not work in general. The optimisation will fail, either because the algorithm converges to a local maximum, or because the maximum number of iterations or function evaluations allowed by the algorithm is exceeded before any extreme point is located. Introducing a more informative prior distribution alleviates this problem by improving the shape of the posterior distribution, but only if it is concentrated enough. When specifying the prior distribution, therefore, we must find an acceptable compromise between concentrating the probability mass too much, so that the model range is inordinately limited, and making the distribution too diffuse, so that the optimisation problem becomes intractable, ultimately causing the estimation to fail.

In addition to the shape of the objective function, the outcome of optimisation depends on the initial position supplied to the optimisation algorithm. If one or more local maxima exist, then depending on the initial position, the algorithm may converge to one of those instead of the global maximum.

How sensitive the outcome of the optimisation is to the choice of prior distribution and starting position, is at least somewhat dependent on the dimensionality of the parameter space. In the test cases considered in section 6, the parameter space is $\mathbb{R}^4$ in the case of single-angle estimation, and $\mathbb{R}^{12}$ for multi-angle (three angles) estimation. This tripling of the dimensionality is enough to give a noticeable difference in stability, even when using the same method for automatically choosing a prior distribution and an initial parameter vector for optimisation. This is something to keep in mind when considering the feasibility of extending the parametric estimation method in ways which would involve increasing the dimensionality of the parameter space.

7.3 Validity of modelling assumptions

In the statistical framework of the multi-angle parametric wavelet estimation procedure, a central assumption is that the wavelets are relatively similar, i.e. that inter-angle differences between wavelets are small. In the model, this is represented by a large correlation between corresponding parameters of different wavelets, and, perhaps more importantly, by the fact that all three wavelets are assumed to be of the same order. While it does simplify the modelling effort, and makes the estimation problem easier, this assumption is not necessarily justified.

The way we model noise in the data also involves some basic modelling assumptions. In the current work, we include only a single noise term $\boldsymbol{e}$ in the seismic trace. For a more realistic and complete model, we might include similar additive noise terms for the reflectivity, and possibly also for the time vector indexing the reflectivity. Having separate model components which could account for other error sources (e.g. measurement errors, mis-tie), would enable us to more effectively isolate the uncertainty in the wavelet itself, as opposed to uncertainty induced by other sources of error (Buland and Omre, 2003).

The seismic noise term included in the model is also assumed to be white. That is, the covariance matrix in the zero-mean Gaussian distribution of the noise vector is a multiple of the identity matrix, and the elements of the noise vector are all independent. Again, the actual noise process could be described more realistically by allowing some of the noise to be coloured by the wavelet. Suppose that instead of modelling the noise term as $\boldsymbol{e} \sim \mathcal{N}(0, \sigma^2)$, we introduce two independent noise vectors

$$\boldsymbol{e}_1 \sim \mathcal{N}(0, \sigma_1^2) \quad \text{and} \quad \boldsymbol{e}_2 \sim \mathcal{N}(0, \sigma_2^2) \tag{71}$$

and let the total noise term be defined as

$$\boldsymbol{e} = W\boldsymbol{e}_1 + \boldsymbol{e}_2 \tag{72}$$

where the matrix W is defined such that $W\boldsymbol{e}_1 = \boldsymbol{e}_1 * \boldsymbol{w}$, and $\boldsymbol{w}$ is the current wavelet estimate. The first term is a coloured noise component, whose elements are correlated, whereas the second term is a white noise component, with independent elements, as in the current noise model. The correlation structure in the coloured noise term is determined by the wavelet convolution matrix W. From (72), the covariance matrix of the total noise term is

$$\mathrm{Cov}(\boldsymbol{e}) = W(\sigma_1^2 I)W^T + \sigma_2^2 I = \sigma_1^2 WW^T + \sigma_2^2 I. \tag{73}$$

In the context of the parametric estimation method, $\boldsymbol{w}$, and hence W, would depend on $\boldsymbol{\beta}$, so the wavelet parameters would enter into the likelihood function in both the mean (as before) and the covariance. With this noise model, there would be two model parameters σ_1 and σ_2 related to the noise, and the model would have five parameters in the single-angle case.

7.4 Omissions and suggestions

In this thesis, we have primarily been concerned with the form of the parametric wavelet model and the formulation of the Bayesian framework for the MAP estimation procedure. We have also discussed how the model may be extended to accommodate joint estimation of multiple wavelets from seismic data with different angles of incidence. This multi-angle extension is only one of several possible extensions.

Other options include multi-well estimation, where borehole data is available from multiple wells in an area covered by the same seismic survey. To exploit assumptions of lateral correlation between wavelets associated with different wells, a spatial component should be incorporated into the model, so that correlation between wavelets could be treated as a function of the distance between their respective wells.

Another possibility is extending the model to handle multi-vintage data, i.e. data from seismic surveys performed at different times. The correlation structure of interest is then temporal, rather than spatial, and could probably be handled in a similar manner as the incidence angle-correlation in the multi-angle case.

A consequence common to the multi-angle, multi-well and multi-vintage extensions, is that they all entail an increase in the number of model parameters, or equivalently, an increase in the dimensionality of the parameter space. As mentioned in section 7.2, this may adversely affect the stability of the numerical optimisation involved in locating the mode of the posterior distribution. When developing these model extensions, care should be taken to ensure that the posterior distribution remains somewhat well behaved.

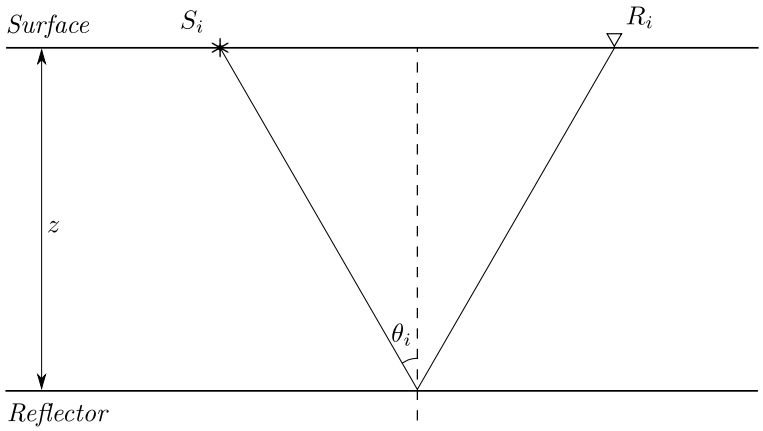


Figure 29: Schematic representation of a land seismic survey. Seismic waves are emitted from the source point S_i, and are incident on the subsurface reflector at depth z with angle of incidence θ_i. Reflected waves then travel along the indicated raypath and are recorded at the receiver location R_i.

Returning to the question of whether it is reasonable to impose the same order on all wavelets in multi-angle estimation, consider the situation depicted in Figure 29, where seismic waves emitted from a source S_i are reflected by a plane horizontal reflector at depth z, with angle of incidence θ_i, and are then recorded by a receiver R_i. The total distance travelled by the waves is $x_i = 2z\cos\theta_i$. In a realistic setting, the source wavelet associated with the seismic waves emitted at S_i will not be similar in appearance to the type of wavelet considered here, but suppose, in order to illustrate the effect of frequency-dependent attenuation, that the source wavelet looks like the top left plot in Figure 30. We compute the corresponding amplitude spectrum by taking the discrete Fourier

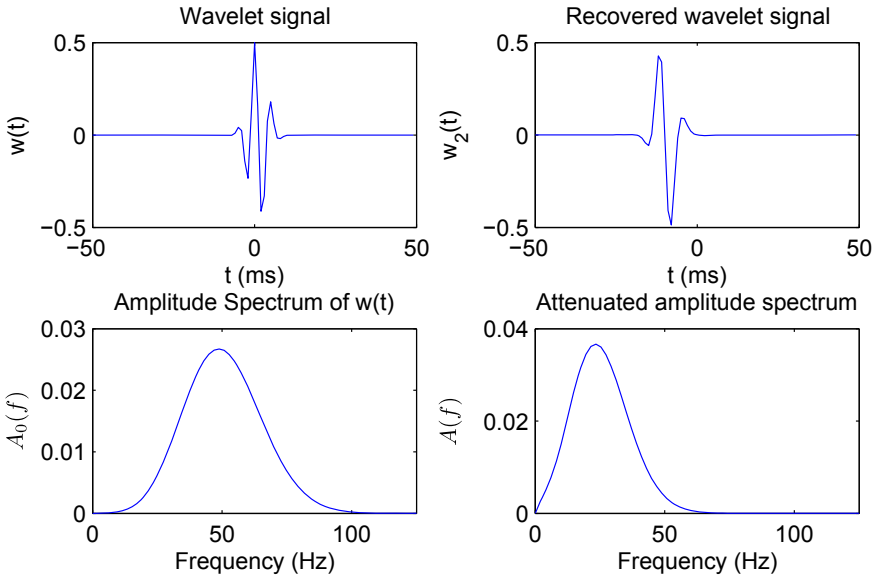


Figure 30: The severity of attenuation due to e.g. anelastic absorption, depends on the frequency of waves. Hence, the loss of high-frequency wavelet components may depend on the distance traversed, or equivalently, on the incidence-angle. **Top left:** Synthetic "source wavelet". **Bottom left:** Amplitude spectrum of the source wavelet. **Bottom right:** Simulated amplitude spectrum at receiver location, normalised. **Top right:** Simulated received near wavelet, recovered from attenuated amplitude spectrum.

transform of the source wavelet. This amplitude spectrum $A_0(f)$ is then modified by taking

$$A(f) = A_0(f) \exp\left(-\frac{f x_i}{Q}\right) \tag{74}$$

at each FFT sample point, where x_i is the total distance traversed by the seismic waves, and Q is a constant associated with the propagation medium. This simulates the frequency-dependent damping effect of anelastic absorption on a plane sine wave with frequency f (Sheriff and Geldart, 1995). The attenuated amplitude spectrum $A(f)$ has a lower peak frequency, and is less influenced by high-frequency components, than the original spectrum $A_0(f)$. Applying the inverse discrete Fourier transform to $A(f)$ gives us the waveform of the attenuated wavelet, which appears smoothed compared with the source wavelet. Since the attenuation depends on x_i, which in turn depends on the incidence angle θ_i, the far wavelet is smoothed more than the near and mid wavelets (Figure 31).

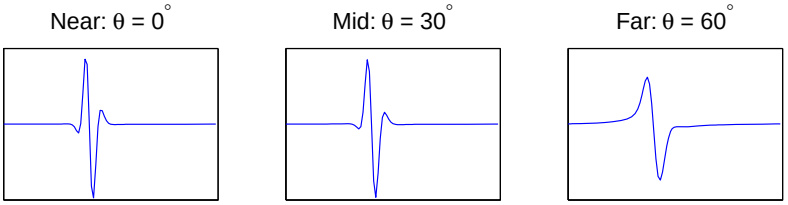


Figure 31: Near, mid and far wavelets after simulated frequency-dependent attenuation. Angles of incidence are $\theta_1 = 0°$, $\theta_2 = 30°$ and $\theta_3 = 60°$. The far wavelet is qualitatively different from the near and mid wavelets.

In terms of the parametric wavelet model studied here, the source wavelet in Figure 30 was generated with $n = 3$, and the near and mid wavelets in Figure 31 resemble wavelets with $n = 2$,

but the far wavelet has the shape of an $n = 1$ wavelet. When interpreting this difference, we should keep in mind that amplitude effects, including attenuation due to anelastic absorption, are corrected for in data processing. In order for the effect studied here to actually manifest itself in the processed data, it would have to be through artefacts present in the data after processing. Provided that this simplified numerical experiment is still somewhat representative, the conclusion is that forcing the same order on all wavelets in joint estimation is unjustified in the general case, although it should be noted that $\theta_3 = 60°$ is a very large angle of incidence, and is certainly at the high end of what one may reasonably expect to encounter in real data.

Having reached the conditional conclusion that using the same order for all wavelets in joint estimation is generally not justified, a natural modification to the multi-angle model would be to allow the wavelet orders to differ. With many angles, this would represent a substantive increase in the number of order combinations to search through, but this number may be reduced somewhat by requiring that for any pair (θ_i, θ_j) of non-negative incidence angles where $\theta_i \leq \theta_j$, we must have $n_i \geq n_j$, where n_k is the order of the wavelet associated with angle θ_k, $k = i, j$. This requirement is reasonable since a higher wavelet order implies a larger peak-frequency.

It is possible that the stability issues discussed in section 7.2, concerning numerical optimisation, may be mitigated by modifying, or "stabilising", the form of the wavelet parametrisation. For example, expanding the model range, i.e. the set of wavelet shapes which can be represented by the parametrisation, would allow more concentrated prior distributions to be used, which directly impacts the stability of optimisation. Additionally, expanding the model range may cause well-fitting, but previously inaccessible points in "wavelet space" to become feasible, which may also make the posterior distribution easier to maximise.

One way to effect such a stabilising modification could be to introduce a parameter controlling the rate at which the amplitude of the wavelet's oscillations decay as the distance from the central peak increases. The semi-synthetic test cases in section 6.2 demonstrate that the current envelope, which has a Gaussian form, is ill-suited for matching the shapes of some wavelets.

7.5 Concluding remarks

We have proposed a method for estimation of seismic wavelets and attending uncertainty from well logs and seismic data. The method is based on a parametric wavelet representation with a small number of model parameters, and set in a Bayesian statistical framework, where all inference concerning the wavelet and the uncertainty associated with it, are addressed through the posterior distribution of the model parameters conditional on the data. We have also described an extension of the model to facilitate joint estimation of multiple wavelets from seismic data with varying incidence angles.

The method has been tested on synthetic and real data, and has been compared to an alternative, point-wise estimation method. Test results indicate that the model is robust to noise, and has a low risk of overfitting the model parameters to the data. The tests also reveal that the performance of the estimation procedure is somewhat limited by a lack of flexibility in the wavelet parametrisation, and that because of its reliance on numerical optimisation, the method is prone to stability issues, and requires careful specification of the prior distribution in order to function reliably.

Other extensions beyond the multi-angle extension are possible with relatively straightforward additions to the model, but it is difficult to anticipate how the increase in dimensionality necessitated by such extensions would affect the stability and overall performance of the estimation method.

References

Keiiti Aki and Paul G Richards. *Quantitative seismology*. Freeman San Francisco, 1980.

Erlend Aune, Jo Eidsvik, and Bjørn Ursin. Three-dimensional non-stationary and non-linear isotropic AVA inversion. *Geophysical Journal International*, 194(2):787–803, 2013.

Arild Buland and Henning Omre. Bayesian wavelet estimation from seismic and well data. *Geophysics*, 68(6):2000–2009, 2003.

Dani Gamerman and Hedibert F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC, Second edition, 2006.

James Gunning and Michael E. Glinsky. Wavelet extractor: A bayesian well-tie and wavelet extraction program. *Computers & Geosciences*, 32, 2006.

Steven M. Kay. *Fundamentals of Statistical Signal Processing*. Prentice Hall, 1993.

Partha S Routh, Phil D Anno, Robert T Baumel, J Andres Chavarria, et al. Inversion for source wavelet and AVA parameters from prestack seismic data. In *2003 SEG Annual Meeting*. Society of Exploration Geophysicists, 2003.

Mauricio D. Sacchi. Seismic Lab, University of Alberta. `http://seismic-lab.physics.ualberta.ca/index.html`, 2014.

Robert E. Sheriff and Lloyd P. Geldart. *Exploration Seismology*. Cambridge University Press, Cambridge, Second edition, 1995.

Tadeusz J. Ulrych, Danilo R. Velis, and Mauricio D. Sacchi. Wavelet estimation revisited. *The Leading Edge*, 14(11):1139–1143, 1995.

Andrew T. Walden and Roy E. White. Seismic wavelet estimation: A frequency domain solution to a geophysical noisy input–output problem. *IEEE Transactions on Geoscience and Remote Sensing*, 36, 1998.